Identification Input Design for Consistent Parameter Estimation of Linear Systems With Binary-Valued Output Observations

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

Abstract—Input design is of essential importance in system identification for providing sufficient probing capabilities to guarantee convergence of parameter estimates to their true values. This paper presents conditions on input signals that characterize their probing richness for strongly consistent parameter estimation of linear systems with binary-valued output observations. Necessary and sufficient conditions on periodic signals are derived for sufficient richness. These conditions are further studied under different system configurations including open-loop and feedback systems, and different scenarios of noises including actuator noise, input measurement noise, and output measurement noise. In addition to system parameter estimation, essential properties of identifiability and input conditions are also derived when sensor thresholds or noise distribution functions are unknown. The findings of this paper provide a foundation to study identification of systems that either use binary-valued or quantized sensors or involve communication channels, which mandate quantization of signals.

Index Terms—Binary-valued observation, distribution function, identification, input design, parameter estimation, sensor threshold, sufficient excitation.

I. INTRODUCTION

INPUT DESIGN is of essential importance in system identification for providing sufficient probing capabilities to guarantee convergence of parameter estimates to their true values, namely, consistent estimation. Input conditions for consistent estimation depend on sensor characteristics, system configurations, noise locations and distributions, and identification algorithms. In traditional identification problems with linear sensors, such conditions are collectively called persistent excitation conditions. Several typical forms of persistent excitation conditions are now standard [4], [23]. This paper studies conditions and design of input signals that characterize their probing richness for consistent parameter estimation of linear systems with binary-valued output observations. We introduce sufficiently rich conditions to distinguish them from traditional persistent excitation conditions. These conditions are then studied with different system configurations including open-loop and feedback systems, and different scenarios of noises including actuator noise, input measurement noise, and output measurement noise. In addition to system parameter estimation, essential properties of identifiability and richness conditions are also derived when sensor thresholds or noise distribution functions are unknown.

System identification of plants with binary-valued observations is of importance in understanding modeling capability for systems with limited sensor information, establishing relationships between communication resource limitations and identification complexity, and studying sensor networks. There are practical systems in which binary-valued sensors are much cheaper than regular sensors, or are the only ones available [34]. Our motivation here is more toward the new paradigm of sensor networks, networked systems and control, e-health systems for remote monitoring, diagnosis, etc. When a signal must be sent over a communication network, the signal must be quantized. A quantized output measurement can be represented by a cascade of binary-valued sensors. In other words, pursuing identification of systems that involve communication channels will need, as a foundation, identification and complexity analysis of the identification problem with binary-valued sensors.

A linear plant combined with a binary-valued or quantized sensor is a structure of Wiener systems, in which the switching sensor represents the memoryless nonlinearity. However, the output of such a sensor takes only a finite number of values, and hence is inherently not invertible anywhere. In this sense, it contains far less information about the system output than the traditional piecewise continuously invertible nonlinearities such as piecewise-linear functions with nonzero slopes. Consequently, the problems studied in this paper require methodologies that depart from most existing methods for identifying Wiener systems. Previous identification methodologies used for Wiener structures include those that deal with piecewise continuous nonlinearities such as...
as iterative algorithms [16], [18], correlation techniques [3], least-squares estimation and singular value decomposition methods [1], [21], etc. The stochastic recursive algorithms [5], [15] can deal with switching nonlinearities with different methods and input design. Furthermore, the algorithms developed in [28] and [29] address Wiener/Hammerstein systems with piece-wise linear functions with jumps and dead zones. Our method involves a two-step algorithm that consists of empirical measures followed by nonlinear mappings using distribution information. The algorithms are uniquely designed for binary-valued or quantized output observations with output disturbances. It has been shown that the algorithms are asymptotically efficient, and hence are asymptotically optimal in terms of convergence speed [31]. Furthermore, the algorithms have been extended to identification of Wiener systems with binary-valued observations [37], which compounds system nonlinearity with sensor nonlinearity. This paper is focused on identification of linear systems.

Our work along this direction was first reported in [34], where a framework was introduced so that the identification of linear systems with binary-valued output observations can be rigorously pursued either in a stochastic setting or in a deterministic worst-case scenario. Extensions to rational systems, unknown noise distribution functions, quantized observations, and communication resource allocations have been recently reported in [32], [33].

This paper presents conditions on input ensembles that provide sufficiently rich probing power for convergence of parameter estimates. Certain results of this paper are extensions of that of [30], [32], and [34]. In particular, the basic sufficient condition of periodic inputs for identifying finite-impulse response (FIR) system was given in [34] and extended to rational systems in [32]. This paper broadens the sufficient richness definition to include also necessity, namely conditions under which the input is not sufficiently rich. The concept of joint identifiability when the noise distribution is unknown was introduced in [32] without detailed analysis on input design or comprehensive recursive algorithms. This paper completes input design, sufficient richness analysis, and general recursive algorithms for this problem. Closed-loop identification problems were studied in [30] under regular sensors. This paper is for binary-valued sensors and covers more scenarios of system configurations and disturbance types. It is shown that sufficient richness of inputs depends essentially on system configurations, disturbance locations, and prior information on parameters.

Classical control theory of Bode and Nyquist characterizes systems by using periodic input signals (frequency responses). They are relatively easy to apply, and there are many special devices for obtaining system frequency responses [23], [26], [27]. This paper is focused on periodic inputs as probing inputs for the following technical reasons. 1) Periodic inputs are uniformly bounded. In contrast, typical stochastic identification methods use Gaussian-distributed signals that are unbounded and more difficult to apply in practical systems. Truncation of unbounded signals due to input saturation may cause bias in system identification. 2) As shown in this paper, essential features for a periodic signal to be rich for identification are certain rank conditions, rather than the magnitudes of the signals. As a result, one may use small probing inputs for identification with the benefit of contained perturbation to system operations. 3) Periods and ranks of periodic signals are shift invariant. As such, they are natural choices for achieving “persistent identification” for time-varying systems [30], [35]. 4) Periods and ranks of periodic signals are invariant after passing through a linear time invariant system (with some mild coprime conditions). Consequently, an externally applied periodic signal can be easily designed for identification of a plant in a closed-loop setting [30]. 5) As shown in this paper, under periodic inputs identification of a system with multiple parameters under quantized sensors can often be reduced to a number of much simplified identification problems for gains. 6) Under periodic inputs, our algorithms have been shown to be asymptotically optimal [31]. This has not been established for other probing inputs.

This paper is organized as follows. Section II begins with a problem formulation for system identification with binary-valued output observations and introduces the basic definition of sufficient richness conditions under this framework. The main results are first presented in Section III for the scenario in which the sensor threshold and noise distribution function are known. Characterizations of Toeplitz matrices and their frequency-domain features are used to establish most results. Under such input signals, causal and recursive algorithms are derived. A key property on invariance of periodicity and rank when a signal passes through stable systems is established in Section IV. By applying this property, sufficient richness conditions are extended to different system configurations including open-loop and feedback systems.

Section V deals with the problem of input noises. Since input disturbances enter the unknown plant to affect the plant output, impact of input noise is technically more challenging than output noises. It is shown that the basic method of empirical measures can still be applied to derive convergent estimates after distributions are modified to reflect impacts from both input and output noises.

The situation when the sensor threshold is unknown is investigated in Section VI. In this situation, the threshold itself becomes part of unknown parameters to be identified, leading to an augmented parameter vector of dimension \( n + 1 \). Although identification algorithms are different in this case from those for plants with \( n + 1 \) parameters, we show that sufficient richness conditions are similar. Identification problems with unknown noise distribution functions are more complicated and are investigated in Section VII. In identification algorithms that are based on empirical measures, as is the case in binary-valued identification problems, distribution functions must be jointly identified with plant parameters by using certain interpolation equations. A concept of joint identifiability is introduced to characterize the fundamental requirement in such problems. Sufficient richness conditions for this problem and recursive algorithms are developed.

Two illustrative examples are presented in Section VIII to demonstrate input design, identification algorithms, and convergence results of the methodologies discussed in this paper. Finally, Section IX provides extensions and further remarks.
Traditional system identification using linear sensors is a relatively mature research area that bears a vast body of literature. There are numerous textbooks and monographs on the subject in a stochastic or worst-case framework such as [20], [23], and [25]. Many significant results have been obtained for identification and adaptive control involving random disturbances in the past decades [4], [6], [14], [19], [20], [23]. The utility of set-valued observations carries a flavor that is related to many branches of signal processing problems. Gradient algorithms for an adaptive filtering using quantized data were studied in [36].

One class of adaptive filtering problems that has recently drawn considerable attention uses “hard limiters” to reduce the computational complexity. The idea, sometimes referred to as binary reinforcement [12], employs the sign operator in the error and/or 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 [7], [9], and [10].

II. PROBLEM FORMULATION

Consider a system

\[ y_k = \sum_{i=0}^{n-1} a_i u_{k-i} + d_k, \quad k = 1, \ldots \]

where \( \{d_k\} \) is a sequence of disturbances. The order \( n \) of the system is known.

Remark 1: When one uses the FIR models, implicitly the system is assumed to be stable. The FIR model is also suitable for approximating exponentially stable systems that can be represented by infinite impulse response (IIR) models or rational models. However, for unstable or marginally stable systems, FIR or IIR models are no longer suitable. There are fundamental issues of model structure selection in this case. 1) Is it reasonable to use a low-order autoregressive (AR), autoregressive model with exogenous inputs (ARX), or an autoregressive and moving average (ARMA) model to represent a practical system? 2) What are the implications of such an approximation on subsequent control design? These issues were raised [11] and discussed in detail in [24]. It was shown in [24] that certain classes of unstable or marginally stable systems defy low-order AR model approximations, and general two-operator coprime factorization models are needed if feedback stabilization of the system is required. In other words, the ubiquitously implied parsimony principle (that a black-box system with input–output data can be represented by low-order models that explain data) for system modeling may not be valid in some systems. For these systems, one must search for high fidelity models of high orders to achieve stabilization, it will not be stated explicitly in what follows.

Define \( \theta = [a_0, \ldots, a_{n-1}]^T \). Then, the system input–output relationship becomes

\[ y_k = \phi_k^T \theta + d_k \]  

where \( \phi_k = [u_k, u_{k-1}, \ldots, u_{k-n+1}]^T \). Further, by using the vector notation, for \( j = 1, 2, \ldots, Y_j = [y_{j-1}]_{n+1}, \ldots, y_{jn}]^T \in \mathbb{R}^n \), \( \Phi_j = [\phi_{j-1}^T, \ldots, \phi_{jn}^T] \in \mathbb{R}^{n \times n} \), \( D_j = [d_{j-1}]_{n+1}, \ldots, d_{jn}]^T \in \mathbb{R}^n \), \( S_j = [s_{j-1}]_{n+1}, \ldots, s_{jn}]^T \in \mathbb{R}^n \), the system output can be rewritten in a block form as

\[ Y_j = \Phi_j \theta + D_j \quad S_j = \mathcal{S}(Y_j). \]  

Note that \( \{\Phi_j\} \) is a sequence of \( n \times n \) Toeplitz matrices obtained from the input \( u \). Under a selected input sequence \( u = \{u_1, u_2, \ldots\} \), the output \( s_k \) is measured for \( k = 1, \ldots, N \).

Estimates of \( \theta \) will be derived from the input–output observations on \( u_k \) and \( s_k \). Denote \( \theta_N \) as an estimate of \( \theta \) on the basis of \( N \) observations on \( s_k \).

Assumption A1: The noise \( \{d_k\} \) is a sequence of independent identically distributed (i.i.d.) random variables with \( E d_k = 0 \) and \( \sigma^2_d = E|d_k|^2 < \infty \). Its distribution function \( F(\cdot) \) is continuously differentiable with a bounded density \( f(\cdot) \) and a continuous inverse \( F^{-1}(\cdot) \).

Remark 2: The i.i.d. assumption on the disturbance can be relaxed to appropriate mixing conditions, compromising only simplicity and clarity of the identification algorithms. Our approach relies on estimating a scalar \( \theta \) from the estimates of the probability \( p = F(C - \theta u_0) \) by inverting the distribution function. Consequently, invertibility of \( F^{-1}(\cdot) \) is required around the point \( C - \theta u_0 \), not necessarily everywhere. This occurs when the noise is uniformly distributed whose density functions will have only a finite support. Continuity of \( f(\cdot) \) is necessary for establishing convergence properties.

In this paper, we derive conditions on the inputs under which we can construct a sequence of consistent estimates of \( \theta \) in the sense of convergence with probability 1 (w.p.1.). To proceed, we introduce the concepts of sufficient richness and information sufficiency.

Definition 1: 1) An input sequence \( u = \{u_k\} \) is sufficiently rich, if under \( u \), one can construct an estimator \( \theta_N \) of \( \theta \) from observations on \( s_k \), \( k \leq N \) such that \( \theta_N \rightarrow \theta \) w.p.1 as \( N \rightarrow \infty \). 2) An input sequence \( u = \{u_k\} \) is information insufficient, if under \( u \), there exist two distinct parameter vectors \( \theta_1 \) and \( \theta_2 \) such that the corresponding output sample paths \( s_k(\theta_1) \) and \( s_k(\theta_2) \) are identical for all \( k \).

Remark 3: Note that this definition is information theoretic. Sufficient richness ensures that the input can provide sufficient probing capability for strong convergence under binary-valued observations. It does not mandate a specific algorithm. On the other hand, if a sequence is information insufficient, then one cannot distinguish \( \theta_1 \) and \( \theta_2 \) from observing \( s_k \), regardless of what algorithms are used. Apparently, if \( u \) is information insufficient, it is not sufficiently rich. However, an information sufficient input may not be sufficiently rich that requires strong convergence.
III. BASIC RICHNESS CONDITIONS UNDER OUTPUT DISTURBANCES

We first establish some essential properties of periodic signals, which will play an important role in subsequent development.

A. Toeplitz Matrices

Recall that an \( n \times n \) Toeplitz matrix [13] is any matrix with constant values along each (top-left to lower-right) diagonal. That is, a Toeplitz matrix has the form

\[
T = \begin{bmatrix}
v_n & \cdots & v_2 & v_1 \\
v_{n-1} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots \\
v_2 & \cdots & v_{n-1} & v_n
\end{bmatrix}.
\]

It is clear that a Toeplitz matrix is completely determined by its entries in the first row and the first column \( \{v_1, \ldots, v_{2n-1}\} \), which is referred to as the symbol of the Toeplitz matrix.

Consider the system (3) and the infinite Toeplitz matrix \( \Phi^\infty = [\Phi_1^T, \Phi_2^T, \ldots]^T \), which will be called the Toeplitz matrix of input \( u \).

Lemma 1: If the Toeplitz matrix \( \Phi^\infty \) of an input \( u \) is not full rank, then \( u \) is information insufficient.

Proof: If \( \Phi^\infty \) is not full rank, then there exists \( \zeta \neq 0 \) such that \( \Phi^\infty \zeta = 0 \). Let \( \theta_1 \) be the true parameter and \( \theta_2 = \theta_1 + \zeta \neq \theta_1 \). Then, \( \gamma_j = \Phi_j(\theta_2) = \Phi_j(\theta_1 + D_j) = \Phi_j(\theta_1) + D_j \), for \( j = 1, 2, \ldots \), which implies that \( u \) is information insufficient. \( \square \)

B. Circulant Toeplitz Matrices and Periodic Signals

\( T \) is said to be a circulant matrix if its symbol satisfies \( v_k = v_{k-n} \) for \( k = n+1, \ldots, 2n-1 \); see [8]. Or in terms of the matrix entries \( T(i, j) \) of \( T \) at the \( i \)th row and \( j \)th column, \( T(i, j) = T(1, n-j+i) \) for \( i = 2, \ldots, n \). A circulant matrix [22] is completely determined by its entries in the first row \( \{v_n, \ldots, v_1\} \), so we denote it by \( T(\{v_n, \ldots, v_1\}) \).

Moreover, \( T \) is said to be a generalized circulant matrix [22] if \( v_k = q^k \) for \( k = n+1, \ldots, 2n-1 \), where \( q > 0 \), which is denoted by \( T(q, \{v_n, \ldots, v_1\}) \).

Definition 2: An \( n \)-periodic signal generated from its one-period values \( v = (v_1, \ldots, v_n) \) is said to be full rank if the circulant matrix \( T(\{v_n, \ldots, v_1\}) \) is full rank.

An important property of circulant matrices is the following frequency-domain criterion.

Lemma 2: If \( T = T(q, \{v_n, \ldots, v_1\}) \) is a generalized circulant matrix, then the eigenvalues of \( T \) are \( \{q^j \gamma_k, k = 1, \ldots, n\} \), and the determinant of \( T \) is \( \det(T) = \prod_{k=1}^{n} q^j \gamma_k \). Here, \( \gamma_k \) is the discrete Fourier transform (DFT) of \( v_j q^{-j/n} e^{-i\omega k} \), \( \omega_k = 2\pi k/n \), \( k = 1, \ldots, n \). Hence, \( T \) is full rank if and only if \( \gamma_k \neq 0, k = 1, \ldots, n \).

Proof: Let \( P = \begin{bmatrix} 0 & I_{n-1} \\ q & 0 \end{bmatrix} \) whose characteristic polynomial is \( \lambda^2 - q \) and eigenvalues are \( q^{1/n} e^{i\omega k} \), \( k = 1, \ldots, n \). Then, \( T \) can be represented by \( T = \sum_{j=1}^{n} v_j P^{n-j} \).

For \( P = 1 \), \( T \) is a full rank matrix.

For \( P \) and \( k = 1, \ldots, n \), if \( x_k \) is the corresponding eigenvector of \( q^{1/n} e^{i\omega_k} \), then \( T x_k = \sum_{j=1}^{n} v_j P^{n-j} x_k = \sum_{j=1}^{n} v_j (q^{j/n} e^{i\omega_k})^{n-j} x_k = \rho x_k \). Therefore, \( \rho \gamma_k \) is an eigenvalue of \( T \), and the expression for \( \det(T) \) is confirmed. By hypothesis, \( q > 0 \). Hence, \( T \) is full rank if and only if \( \gamma_k \neq 0, k = 1, \ldots, n \).

For the special case when \( q = 1 \), we have the following property.

Corollary 1: An \( n \)-periodic signal generated from \( v = (v_1, \ldots, v_n) \) is full rank if and only if its DFT \( \gamma_k = V(\omega_k) = \sum_{j=1}^{n} v_j e^{-i\omega k} \) is nonzero at \( \omega_k = 2\pi k/n \), \( k = 1, \ldots, n \).

Recall that \( T = \{\gamma_1, \ldots, \gamma_n\} = F[v] \) is the frequency sample of the \( n \)-periodic signal \( v \), where \( F[\cdot] \) is the DFT. Hence, Definition 2 may be equivalently stated as "an \( n \)-periodic signal \( v \) is said to be full rank if its frequency samples do not contain 0." In other words, the signal contains \( n \) nonzero frequency components.

C. Basic Sufficient Richness Conditions

We use the following notation for element-wise vector functions. For the distribution function \( F(\cdot) \) and a vector \( x = [x_1, \ldots, x_n]^T \in \mathbb{R}^n \), we define

\[
\begin{align*}
F(x) &= [F(x_1), \ldots, F(x_n)]^T \in \mathbb{R}^n \\
G(x) &= [F^{-1}(x_1), \ldots, F^{-1}(x_n)]^T \in \mathbb{R}^n.
\end{align*}
\]

Similarly, for \( \alpha = [\alpha_1, \ldots, \alpha_n]^T \) and \( c = [c_1, \ldots, c_n]^T \) in \( \mathbb{R}^n \), we use \( I_{\{x \leq \alpha\}} = \sum_{x \leq \alpha} I_{\{x \leq \alpha\}} \) to denote column vectors with all components being 1 and 0, respectively. For a given threshold \( C \), \( C_n = C I_n \in \mathbb{R}^n \). Let \( \xi_N = \frac{1}{N} \sum_{j=1}^{N} S_j \), and \( X_N = G(\xi_N) \).

The sufficiency of the following theorem was first proved in [34] although the term "sufficiently rich" was not used. The necessity, however, is new.

Theorem 1: Under Assumption A1, suppose \( u \) is \( n \)-periodic. Then, \( u \) is sufficiently rich if and only if \( u \) is full rank.

Proof: When \( u \) is \( n \)-periodic, we have \( \Phi_1 = \Phi_2 = \cdots = \Phi \) in (3), where \( \Phi \) is the circulant matrix with symbol \( u \).

Sufficiency: By hypothesis, \( u \) is full rank. Hence, \( \Phi \) is invertible. An estimate of \( \theta \) is defined as \( \theta_N = \Phi^{-1} (C_n - X_N) \). We will show that \( \theta_N \to \theta \) w.p.1. We claim that \( \lim_{N} \xi_N = F(C_n - \Phi \theta) \) w.p.1. To verify this, note that by the well-known strong law of large numbers, as \( N \to \infty \), \( \xi_N - F(C_n - \Phi \theta) = \frac{1}{N} \sum_{j=1}^{N} I_{\{d_j \leq C_n - \Phi \theta\}} - F(C_n - \Phi \theta) \to 0 \) w.p.1. Hence, \( \xi_N \to F(C_n - \Phi \theta) \) w.p.1. Now, by the continuity of \( F(\cdot) \) and \( F^{-1}(\cdot) \), convergence of \( \xi_N \) implies that \( X_N = G(\xi_N) \to G(F(C_n - \Phi \theta)) = C_n - \Phi \theta \) w.p.1. It follows that as \( N \to \infty \), \( \theta_N = \Phi^{-1} (C_n - X_N) \to \theta \) w.p.1. This proves that \( u \) is sufficiently rich.

Necessity: If \( \Phi \) is not full rank, then \( \Phi^\infty = [\Phi^T, \Phi^T, \ldots]^T \) is also not full rank. By Lemma 1, \( u \) is information insufficient, which implies that \( u \) is not sufficiently rich. \( \square \)
IV. SUFFICIENT RICHNESS CONDITIONS IN FILTERING, REGULATION, AND TRACKING PROBLEMS

Consider some typical system configurations illustrated in Fig. 1. Filtering configuration is an open-loop system, where $M$ is linear, time invariant, and stable, but may be unknown. The feedback configuration is a general structure of 2-degree-of-freedom controllers, where $K$ and $F$ are linear time invariant, may be unstable, but are stabilizing for the closed-loop system. The mapping from $r$ to $u$ is the stable system $M = K/(1 + PKF)$. When $K = 1$, it is a regulator structure, and when $F = 1$, it is a servo-mechanism or tracking structure. Note that system components $M$, $K$, and $F$ are usually designed for achieving other goals and cannot be tuned for identification experiment design.

In these configurations, the input $u$ to the plant $P$ can be measured, but cannot be directly selected. Only the external input $r$ can be designed. By Theorem 1, a sufficient condition for $u$ to provide sufficient richness is that $u$ is $n$-periodic and full rank. Here, we would like to establish relationships between periodicity and rank properties of the external signal $r$ and those of $u$.

A. Invariance of Input Periodicity and Rank in Open and Closed-Loop Configurations

Let $H$ be a linear time invariant and stable system with impulse response $\{h_k\}$. Suppose that $u = Hr$, or in the time domain

$$u_k = \sum_{l=0}^{\infty} h_l r_{k-l}. \quad (6)$$

Suppose that the discrete-time Fourier transform (DTFT) of $H$ is $H(e^{i\omega}) = \sum_{l=-\infty}^{\infty} h_l e^{-i\omega l}$.

**Theorem 2:** Suppose that $r$ is $n$-periodic and full rank. Then, $u$ is also $n$-periodic and full rank if and only if $H(e^{i\omega}) \neq 0$, for $\omega = \omega_k := \frac{2\pi k}{n}, k = 1, \ldots, n$.

**Proof:** Since $r$ is $n$-periodic and full rank, by Corollary 1, the frequency samples of $r$ are $R_k = \sum_{l=1}^{n} r_l e^{-i\omega_k l} \neq 0, k = 1, \ldots, n$. Since $r$ is $n$-periodic, it is easy to verify from (6) that $u$ is also $n$-periodic. Furthermore, the frequency samples $U_k$ of $u$ are related to $R_k$ by

$$U_k = \sum_{l=1}^{n} u_l e^{-i\omega_k l} = \sum_{l=1}^{n} h_l r_{l-\omega_k} e^{-i\omega_k l} = \sum_{l=0}^{\infty} h_l e^{-i\omega_k l} \sum_{l=1}^{n} r_{l-\omega_k} e^{-i\omega_k (l-1)} = H(e^{i\omega_k}) R_k.$$

Here, the cyclic property of the DFT is applied: $R_k = \sum_{l=1}^{n} r_l e^{-i\omega_k l} = \sum_{l=1}^{n} r_{l-\omega_k} e^{-i\omega_k (l-1)}$. By Corollary 1, $u$ is full rank if and only if $U_k \neq 0, k = 1, \ldots, n$. However, by hypothesis, $R_k \neq 0, k = 1, \ldots, n$. As a result, $U_k \neq 0$ if and only if $H(e^{i\omega_k}) \neq 0, k = 1, \ldots, n$. □

**Example 1:** The necessity of the condition of Theorem 2 can be verified by examining the following second-order system $u_k = r_k + r_{k-1}$. When $r$ is a 2-periodic signal and full rank, $u_k$ is a constant, and hence is not rank 2. This is due to the fact that $H(e^{i\omega}) = 1 + e^{i\omega}$ and for $\omega = \omega_2 = 2\pi/2 = \pi, H(e^{i\omega}) = 0$.

**Remark 4:** Theorem 2 implies that, for any system $H$ that does not have annihilating zeros at $n$ points $e^{i\omega_k}, \omega_k = \frac{2\pi k}{n}, k = 1, \ldots, n$, on the unit circle, sufficient richness capability of the signal $r$, established by Theorem 1, is always preserved after passing through $H$. In particular, for the feedback configuration in Fig. 1, we have the following result that indicates that input richness properties are invariant under a feedback mapping.

**Assumption A2:** Consider the feedback configuration Fig. 1(b). Assume that, for $\omega_k = \frac{2\pi k}{n}, k = 1, \ldots, n$, $K(e^{i\omega_k})$ does not have zeros at $\omega_k$; and $P(e^{i\omega})$ and $F(e^{i\omega})$ do not have singularities (such as poles) at $\omega_k$.

**Corollary 2:** Under Assumption A2, $M = K/(1 + PKF)$ does not have annihilating zeros at $\omega_k = \frac{2\pi k}{n}, k = 1, \ldots, n$. As a result, if $r$ is $n$-periodic and full rank, so is $u$.

**Proof:** From $M(e^{i\omega}) = \frac{1}{1 + P(e^{i\omega}) K(e^{i\omega}) F(e^{i\omega})}$, it is clear that the zeros of $M$ are either the zeros of $K$ or the singularities (such as poles) of $P$ or $F$. By Assumption A2, $K(e^{i\omega}) \neq 0$, and $\omega_k$ is not a singularity point of $P(e^{i\omega})$ or $F(e^{i\omega})$. Hence, $M(e^{i\omega}) \neq 0, k = 1, \ldots, n$. Now by Theorem 2, $u$ is $n$-periodic and full rank whenever $r$ is $n$-periodic and full rank. □

B. Periodically Perturbed Input Signals

Consider the tracking configuration in Fig. 2. When the desired output is $r_0$, usually $r = r_0$ is the set point. However, a constant $r_0 \neq 0$ is 1-periodic. It is only good for identification of a gain system (namely, $n = 1$). To enhance the probing capability, one may add a small dither $w_k$ to $r_0$, leading to $e_k = w_k + r_0$. Since $u_k = Mr_k = Mw_k + Mr_0 = v_k + e_k$, where $v_k$ is an $n$-periodic signal and $e_k = Mr_0$ becomes a constant $\mu$ after a short transient. We need to establish rank conditions on $u_k$.

Generally, consider an input signal $u$: $u_k = v_k + e_k$, which is a perturbation from $v$. Suppose that $v_k$ is $n$-periodic and full rank. We would like to establish conditions under which $u_k$ is also $n$-periodic and full rank.

**Assumption A3:** Both $v$ and $e$ are $n$-periodic.

Under Assumption A3, the Toeplitz matrices for $v$, $e$, and $u$, denoted by $\Phi_v, \Phi_e$, and $\Phi_u$, respectively, are circulant matrices. Let their corresponding frequency samples be
\[ \Gamma_u = \mathcal{F}[u] = \{\gamma^u_k, k = 1, \ldots, n\}, \quad \Gamma_v = \mathcal{F}[v] = \{\gamma^v_k, k = 1, \ldots, n\}, \quad \Gamma_e = \mathcal{F}[e] = \{\gamma^e_k, k = 1, \ldots, n\}. \]

Theorem 3: Under Assumption A3, \(u\) is full rank if and only if \(\gamma^u_k \neq 0, k = 1, \ldots, n\).

Proof: This follows immediately from the fact \(\gamma^w_k = \gamma^e_k + \gamma^v_k\), and that \(\Phi_u\) is full rank if and only if its frequency samples do not contain 0.

We now consider the special case when \(e_k \equiv \mu\), which is a typical case in tracking problems, as shown before.

Corollary 3: Suppose \(v_k\) is \(n\)-periodic and full rank, and \(e_k = \mu\). Then, \(u_k\) is \(n\)-periodic. Let \(\eta = \frac{1}{n} \sum_{j=1}^{\nu} v_j\). \(u\) is full rank if and only if \(\mu \neq -\eta\).

Proof: Since \(v_k\) is full rank, by Corollary 1, we have \(\gamma^v_k \neq 0, k = 1, \ldots, n\). In particular, \(\gamma^v_n = \sum_{j=1}^{\nu} v_j = n\eta\). Moreover, the frequency samples of \(e_k \equiv \mu\) are \(\gamma^e_k = 0, k = 1, \ldots, n - 1\), and \(\gamma^e_n = n\mu\). Consequently, by Theorem 3, \(u_k\) is full rank if and only if \(\gamma^v_n + \gamma^e_n \neq 0\). That is, \(n\eta + n\mu \neq 0, \mu \neq -\eta\), as claimed.

Corollary 3 may be verified directly by matrix manipulations. Toeplitz matrices \(\Phi_u, \Phi_v, \phi_e\) for \(u, v, e\), respectively, are

\[ \Phi_u = \Phi_v + \Phi_e. \]

by adding the second through \(n\)th rows to the first row, followed by subtracting the first column from the second to \(n\)th columns. The last matrix is full rank since \(\eta + \mu \neq 0\), and the lower-right \((n - 1) \times (n - 1)\) submatrix, which is obtained by elementary operations from \(\Phi_v\), is full rank.

V. SUFFICIENT RICHNESS CONDITIONS UNDER INPUT NOISES

Under the system configurations in Fig. 1, \(u = Mr\) is generated from \(r\) by a possibly unknown system \(M\). In the previous sections, \(u\) is assumed to be accurately measured. When \(u\) is further corrupted by noise, it can no longer be exactly measured. Furthermore, the actual values of \(u\) cannot be directly derived from \(r\) since \(M\) is unknown. Sufficient richness conditions and identification algorithms under this scenario will be explored in this section.

We will consider the following two cases of input noises shown in Fig. 3.

1) Input measurement noise: When \(u\) is measured by a regular sensor, the measured values are related to \(u\) by \(w_k = u_k + \epsilon_k\), where \(\epsilon_k\) is the measurement noise.

2) Actuator noise: In this case, the actual input to the plant is \(u_k = v_k + \epsilon_k\), where \(v_k = Mr\) and \(\epsilon_k\) is the actuator noise.

As a result, the measured input is \(w_k = v_k + \epsilon_k + \epsilon_k\), and the identification of the plant must be performed from \(w_k\) and \(s_k = S(y_k)\).
\[ \theta_{N+1} = \begin{cases} \Phi^{-1}_{N+1}(C_n - G(\xi_{N+1})) & \text{if } \Phi_{N+1} \text{ is nonsingular} \\ \theta_N & \text{if } \Phi_{N+1} \text{ is singular}. \end{cases} \]

**Theorem 4:** Under Assumption A4, \( \theta_N \to \theta \) w.p.1 as \( N \to \infty. \)

**Proof:** Since the true input to the plant is \( u, \xi_N \to \xi = F'(C_n - \Phi \theta) \) w.p.1. Then, \( \theta_N - \theta = \Phi_{N+1}(G(\xi) - G(\xi_N)) + (\Phi^{-1}_{N+1} - \Phi^{-1})(C_n - G(\xi)). \) By the strong law of large numbers, the convergence \( \theta_N - \theta \to 0 \) follows from \( \Phi_N \to \Phi, \xi_N \to \xi \) w.p.1, continuity of \( F^{-1} \), and the invertibility of \( \Phi. \)

\[ \square \]

**B. Actuator Noise**

Unlike the measurement noise \( \varepsilon_k \) that affects measured input values but does not enter the plant, actuator noise \( e_k \) affects the output of the plant \( y_k. \) Now, consider the case \( u_k = v_k + e_k \) and \( w_k = u_k. \) To understand the impact of \( e_k \), express the regressor in (2) by \( \psi^\phi_k \) or \( \psi^\nu_k \), depending on which signal is used in the regressor. Under Assumption A4, \( v \) is \( n \)-periodic and full rank, but \( u \) is not periodic. However, by Lemma 3, \( \frac{1}{N} \sum_{j=0}^{N-1} \Phi^j v \to \Phi^N v \) w.p.1 as \( N \to \infty. \)

Since \( u_k = v_k + e_k \), we have \( y_k = \left( \psi^\phi_k \right)^T \theta + d_k = \left( \psi^\nu_k \right)^T \theta + z_k. \) Observe that the equivalent noise \( z_k \) is \( z_k = \left( \psi^\theta_k \right)^T \theta + d_k = a_0 e_k + \cdots + a_{n-1} e_{k-n+1} + d_k. \)

\[ \text{Under Assumption A4, although } \{z_k\} \text{ may not be independent, it is strictly stationary.} \]

Recall that \( \{z_k\} \) is strictly stationary if for any positive integer \( n \), points \( t_1, \ldots, t_n \in \mathbb{Z}^+ \) and \( l \in \mathbb{Z}^+ \), the joint distribution of \( \{z_{t_1}, \ldots, z_{t_n}\} \) is the same as that of \( \{z_{t_1+l}, \ldots, z_{t_n+l}\} \). Its finite dimensional distributions are translation invariant; see [17, p. 443]). Denote the distribution function by \( F_z(x; \theta). \) A moment of reflection reveals that the sequence is \( (n-1) \)-dependent. A precise definition of \( (n-1) \)-dependence can be found in [2, p. 167, Example 1]. Since an \( (n-1) \)-dependent sequence belongs to the class of \( \phi \)-mixing signals, whose remote past and distant future are asymptotically independent, the sequence is strongly ergodic [17, p. 488]. That is, a strong law of large numbers still holds.

Following (5), define \( \xi_N = 1/N \sum_{j=1}^{N} S_j. \) Let \( \theta_N \) be the solution to

\[ \xi_N = F_z(C_n - \Phi \theta_N; \theta_N). \]  

(7)

For any \( \theta \), define the Jacobian matrix \( J(\theta) = \frac{\partial F_z(C_n - \Phi \theta; \theta)}{\partial \theta}. \) A condition for invertibility of the function in (7) is that \( J(\theta) \) is full rank. In this case, by denoting the inverse of \( \xi = F_z(C_n - \Phi \theta; \theta) \) as \( \theta \to H(\xi) \), the estimate \( \theta_N \) in (7) may be symbolically written as \( \theta_N = H(\xi_N). \)

**Proposition 1:** If \( H(\cdot) \) exists and is continuous, then \( \theta_N \to \theta \) w.p.1.

**Proof:** By the strong law of large numbers, \( \xi_N \to \xi = F_z(C_n - \Phi \theta; \theta) \) w.p.1. Since \( H(\cdot) \) exists and is continuous, \( \theta_N = H(\xi_N) \to H(\xi) = \theta \) w.p.1.

\[ \square \]

For a given \( \theta \), denote the inverse of \( F_z(x; \theta) \) (with respect to \( x \)) by

\[ G_z(x; \theta) = F_z^{-1}(x; \theta). \]  

(8)

Computationally, it is observed that, for a given \( \xi \), the implicit function \( \xi = F_z(C_n - \Phi \theta; \theta) \) of \( \theta \) may be expressed as a fixed-point equation \( \theta = \Phi(C_n - G_z(\xi; \theta)). \)

Next, a special case will be considered when \( \{e_k\} \) and \( \{d_k\} \) are both normal random variables. Suppose that \( \{e_k\} \) is a sequence of i.i.d. normal random variables with mean zero and variance \( \sigma^2 \), and \( \{d_k\} \) is a sequence of i.i.d. normal random variables with mean zero and variance \( \sigma'^2 \). Then, \( z = a_0 e_k + \cdots + a_{n-1} e_{k-n+1} + d_k \) is also normally distributed, has zero mean and variance \( \sigma^2 + \sigma'^2 \). Let \( F_0(x) \) be the normal distribution function of zero mean and variance one. Then, \( F_z(x; \theta) = F_0(C_n - \Phi \theta/\sigma(\theta)). \)

It follows that \( F_z(C_n - \Phi \theta; \theta) = F_0(C_n - \Phi \theta/\sigma(\theta)). \)

The Jacobian matrix is

\[ J(\theta) = \frac{dF_z(C_n - \Phi \theta; \theta)}{d\theta} = -\frac{1}{\sigma(\theta)} \left[ F(z - \sigma^2 \theta^T / \sigma'^2) + \sigma^2 C_n \theta^T / \sigma'^2 \right] \]

where \( x = C_n - \Phi \theta/\sigma(\theta). \)

\[ \text{In the following derivations, the norms of matrices and vectors are: for matrix } A \in \mathbb{R}^{n \times n} \text{ and vector } x \in \mathbb{R}^n, \|A\| = \sqrt{\lambda_{\text{max}}(A^T A)}, \text{ where } \lambda_{\text{max}}(\cdot) \text{ is the largest eigenvalue of the matrix;} \]
\[ \text{and } \|x\| = \sqrt{\sum_{i=1}^{n} |x_i|^2}. \]

It is obvious that for a vector \( \theta, \|\theta^T \| = \theta^T \theta. \)

**Remark 5:** It is easily verified that if \( A \) is an \( n \) dimension matrix with \( \|A\| < 1 \), then \( I_n + A \) is invertible, where \( I_n \) denotes the \( n \times n \) identity matrix. Moreover, suppose \( A \) is an \( n \) dimension invertible matrix. If \( \|B\| < \|A^{-1}\|^{-1} \), then \( A + B \) is invertible.

**Theorem 5:** If

\[ \|\Phi(\theta) - \Phi(\theta')\|^2 < \frac{2\sigma^2}{C\sigma^2 \sqrt{\lambda_{\text{max}}(\sigma^2 \theta^T \theta + \sigma'^2)}} \]

then \( \theta_N = H(\xi_N) \to \theta \) w.p.1.

**Proof:** Noting that \( \|\sigma^2 \theta^T \theta / \sigma'^2\| = \|\sigma^2 \theta^T \theta + \sigma'^2\| \leq \|\sigma^2 \theta^T \theta + \sigma'^2\| \leq \sigma^2 C \sqrt{\lambda_{\text{max}}(\theta^T \theta)}} / (2\lambda_{\text{max}}(A)) \]

\[ \text{we have } \|\sigma^2 \theta^T \theta / \sigma'^2\|^2 \leq \|\sigma^2 \theta^T \theta / \sigma'^2\|^2 \leq \|\sigma^2 \theta^T \theta / \sigma'^2\|^2 \leq \|\Phi(\theta) - \Phi(\theta')\|^2 < \|\Phi(\theta) - \Phi(\theta')\|^2. \]

By Remark 5, \( \Phi + \sigma^2 C \theta^T / \sigma'^2 \) is invertible. So, \( J(\theta) \) is invertible. Hence, Proposition 1 confirms that \( \theta_N \to \theta \) w.p.1.

**Remark 6:** Condition (9) can be used to design input signals. Indeed, suppose that the prior information on the unknown parameters is that \( \|\theta\| \leq \beta \). By using \( \beta^2 \) in place of \( \|\theta\|^2 \), one can design an input such that \( \Phi \) satisfies (9). Consequently, consistency of the estimates will be guaranteed for any \( \theta \in \{\theta : \|\theta\| \leq \beta\}. \)
Example 2: Suppose the true system is $y_k = 0.9 u_k + 1.1 u_{k-1} + d_k$. Hence, the true parameters are $\theta = [0.9, 1.1]^T$ and $\|\theta\|^2 = 1.93$. Assume that the prior information on $\theta$ is that $\|\theta\|^2 \leq 2$. The output measurement noise $d_k$ is i.i.d., normally distributed with zero mean and variance $\sigma_d^2 = 4$. The input signal $u_k = v_k + e_k$, where $v_k$ is two-periodic with its one-period values $\mu_1 = 3$, $\mu_2 = 15$, and $e_k$ is an i.i.d. normally distributed noise of zero mean and variance $\sigma_e^2 = 1$. By direct calculation, $|\Phi^{-1}| = 0.083$. For $C = 20$, and the prior information $|\theta|^2 \leq 2$, the right-hand side of (9) is 0.094. Hence, the input satisfies condition (9). In fact under this input, (9) is satisfied for all $\theta \in \{\theta : \|\theta\|^2 \leq 2\}$.

An identification algorithm is devised for this example. At each step $N$, $\xi_N$ is calculated from (5). Then the estimate $\hat{\theta}_N$ is derived by solving (7). The inverse function of normal distribution is calculated by the Matlab function norminv. The simulation illustrates the convergence of parameter estimates. The relative estimation error $\|\theta_N - \theta\|/\|\theta\|$ is used to evaluate accuracy and convergence of the estimates. Fig. 4 shows parameter convergence of this algorithm.

VI. SUFFICIENT RICHNESS: UNKNOWN THRESHOLD $C$

The main relationship in computing estimates is the $n$ limiting equations of empirical measures $\xi = F(C_n - \Phi \theta)$. When $C$ is unknown, this relationship is not sufficient to determine $\theta$ and $C$, since it has $n$ equations but $n + 1$ unknowns. We introduce the following modified algorithm to estimate $C$ and $\theta$ collectively.

We will use the configuration of Fig. 3 to carry out our discussions: The input is subject to measurement noise (no input actuator noise), and the output has measurement noise, namely $u_k = v_k, w_k = u_k + e_k$, and $y_k = \theta_k + d_k$, where $d_k$ satisfies Assumption A1. Other cases can be similarly derived and will not be detailed here.

A. Sufficient Richness Conditions

Assumption A5: Suppose that $\{v_k\}$ is $(n + 1)$-periodic and full rank, and that $\{e_k\}$ is an i.i.d. and zero mean sequence.
Proof: 1) Recall that $\Psi_N^w = [I_{n+1}, -\tilde{\Psi}_N^w]$. Under Assumption A5, $\Psi_N^w \rightarrow w.p.1$. Under Assumption A1, by the strong law of large numbers, $\tilde{\xi}_N \rightarrow \xi = F(\Psi \Theta)$ w.p.1 as $N \rightarrow \infty$. This implies, by continuity of $F^{-1}(\cdot)$, $G(\xi_N) \rightarrow \Psi \Theta$ w.p.1 as $N \rightarrow \infty$. As a result, by Lemma 4, $\Theta_N = (\Psi_N^{-1})^T G(\xi_N) \rightarrow \Theta$ w.p.1 as $N \rightarrow \infty$.

2) Under the hypothesis, by Lemma 4, $\Psi$ is not full rank. Hence, there exists $\delta \neq 0$ such that $\Psi \delta = 0$. Suppose $C_1$ and $\theta_1$ are true parameters, and $[C_2, \theta_2]^T = [C_1, \theta_1]^T + \delta$. Then, $y_k(\theta_1) = \phi_1^T \theta_1 + d_k \leq C_1$ if $y_k(\theta_2) = \phi_1^T \theta_2 + d_k \leq C_2 \ \forall k$. It follows that the output sequences satisfy $s_k(C_1, \theta_1) = s_k(C_2, \theta_2)$. In other words, $v_k$ is information insufficient, which implies that $v_k$ is not sufficiently rich. □

B. Recursive Algorithms

A causal and recursive algorithm for computing $\Theta_N$ can be constructed as follows.

1) Initial conditions: $\tilde{\xi}_1 = \tilde{S}_1, \tilde{\Psi}_1 = \tilde{\Phi}_1^w$, and $\Theta_0 = 0$.

2) Recursion: Suppose that at $N$, $\xi_N$, $\tilde{\Psi}_N^w$, and $\Theta_N = [C_N, \theta_N]^T$ have been obtained. Then, at $N + 1$, we update

\[
\begin{align*}
\tilde{\xi}_{N+1} &= \tilde{\xi}_N - \frac{1}{N+1} \tilde{\xi}_N + \frac{1}{N+1} \tilde{S}_{N+1} \\
\tilde{\Psi}_{N+1}^w &= \tilde{\Psi}_N^w - \frac{1}{N+1} \tilde{\Psi}_N^w + \frac{1}{N+1} \tilde{\Psi}_{N+1}^w \\
\Psi_{N+1}^w &= [I_{n+1}, -\tilde{\Psi}_{N+1}^w] \\
\Theta_{N+1} &= \begin{cases} 
\Theta_N, & \text{if } \Psi_{N+1}^w \text{ is singular} \\
(I_{n+1})^{-1} G(\xi_{N+1}), & \text{otherwise}.
\end{cases}
\end{align*}
\]

The following theorem claims convergence of $\Theta_N$, whose proof is similar to that of Theorem 4 and is omitted.

Theorem 7: Under Assumptions A1 and A5, $\Theta_N \rightarrow \Theta w.p.1$ as $N \rightarrow \infty$.

VII. SUFFICIENT RICHNESS: UNKNOWN DISTRIBUTION FUNCTION

The identification algorithms and sufficient richness conditions derived so far rely on the knowledge of the distribution function $F(\cdot)$ or its inverse. However, in most applications, the noise distributions are not known, or only limited information is available. On the other hand, input–output data from the system contain information about the noise distribution. Hence, $F(\cdot)$ can be potentially estimated, together with system parameter $\theta$. In this section, we will derive sufficient richness conditions under which $\theta$ and $F$ can be jointly identifiable. This problem and the concept of joint identifiability were first introduced in [32], together with the basic sufficient conditions for identifiability, a recursive algorithm, and its convergence. The input design and its sufficient richness presented in this section are new.

A. Parametrization of $F$

To estimate the distribution function $F(x)$, one needs interpolation equations in the form of $\xi_i = F(x_i)$, for $i = 1, 2, \ldots, L$. When $F(\cdot)$ is not parameterized, estimation on $F$ can become sufficiently accurate only if the data points $\{x_i\}$ are sufficiently dense, rendering an estimation problem of high complexity. Here, we adopt a parametrization approach for $F(\cdot)$.

Suppose that $F(x)$ is parameterized by a vector $a$ of dimension $m$. To emphasize this parametrization, $F$ will be written as $F(x; a)$. For example, for normal distributions, $a = [\mu, \sigma]^T$. Given a set of $L$ points $X_L = [x_1, \ldots, x_L]^T$, suppose that interpolation values of the distribution at these points are $p^l = F(x_i; a)$, $l = 1, \ldots, L$. Define $F(X_L; a) = F(x_1; a), \ldots, F(x_L; a)$ and $P_L = [p^1, \ldots, p^L]^T$. Hence, the interpolation relationship for the given data pair $(X_L, P_L)$ can be written as $P_L = F(X_L; a)$.

Assumption A6: The function $F(x; a)$ has continuous partial derivatives with respect to both variables $x$ and $a$.

Definition 3: $F(x; a)$ is said to be jointly identifiable if for any set of $m + 1$ nonzero distinct points $p = [p_1, \ldots, p_{m+1}]^T$, $F(C_{m+1} - pa; a)$ is invertible as a function of $a$ (a scalar) and $a$.

Jointly identifiable functions guarantee that there is a unique solution $a$ and $\alpha$ to the equation $\xi = F(C_{m+1} - pa; a)$. Joint identifiability is an essential property. Otherwise, the parameterized distribution function $F(x; a)$ and $\theta$ may not be uniquely determined from interpolation equations, which are the foundation of system identification with binary-valued observations. The following example highlights the main reasons for this property.

Example 3: Let $F_0(x)$ be the distribution function of the standard normal random variable (with zero mean and variance 1). Then, a normal random variable with mean $\mu$, variance $\sigma^2$, and distribution function $F(x)$, can be expressed as $F(x; [\mu, \sigma]) = F_0(x - \mu/\sigma)$. Suppose that the system is $y_k = au_k + d_k$, $k = 1, 2, \ldots$, namely, a gain system with unknown parameter $a$. Then, $F(C - au_k; [\mu, \sigma]) = F_0(C - au_k - \mu)/\sigma = F_0(c_1 - c_2 u_k)$, where $c_1 = (C - \mu)/\sigma$ and $c_2 = a/\sigma$. Since this is a two-parameter function, one cannot identify three parameters $a$, $\mu$, and $\sigma$ uniquely, regardless of how many interpolation points $u_k$ are used.

The main issue here is that, although the class of distribution functions is uniquely parameterized by $[\mu, \sigma]$, when they are combined with unknown parameters of the system, the parameter set $(\theta, a)$ is not identifiable from input–output relationships, motivating the notion of joint identifiability. A remedy of this situation will require acquisition of partial information on the distribution function to reduce the dimension of its parameter vector. For example, if $\mu$ is known as $\mu_0$, then the class of distribution functions $F(x; \sigma) = F_0(x - \mu_0/\sigma)$ can be shown to be jointly identifiable. Indeed, take any $u_1 \neq u_2$. Let $\xi_i = F_0(c - au_i - \mu)/\sigma$, $i = 1, 2$. Then, $x_i = F_0^{-1}(\xi_i) = (c - au_i - \mu)/\sigma$, $i = 1, 2$, which have a unique solution since $u_1 \neq u_2$. Similarly, if $\sigma = \sigma_0$ is known, $F(x; \mu) = F_0(x - \mu)/\sigma_0$ is jointly identifiable.

B. Sufficient Richness Conditions

For notational simplicity, we shall use the basic configuration $y_k = \phi_1^T \theta + d_k$ for developing algorithms, where $u_k$ is periodic and has no input disturbance. Other cases can be readily derived.
from the same principles of these algorithms. Suppose that the threshold \( C \) is known, first, we derive a special class of inputs \( u \) that will provide sufficient probing capability to identify both \( \theta \) and \( \alpha \).

**Definition 4:** A \( 2n(m + 1) \)-periodic signal \( u \) is called a scaled full rank signal if its one-period values are \((v, v, \ldots, v,\rho_1 v,\ldots,\rho_n v)\), where \( v = (v_1, \ldots, v_n) \) is full rank, i.e., \( 0 \not\in F[v] \), and \( \rho_j \neq 0 \) and \( \rho_j \neq 1 \), \( j = 1, \ldots, m \), and \( \rho_j \neq \rho_i, i \neq j \). Let \( U \) denote the class of such signals.

Let \( \xi_N \) be defined as in (5), with the dimension changed from \( n \) to \( n(m + 1) \). Under the strong law of large numbers

\[
\xi_N \rightarrow \xi = F(C_{n(m+1)} - \Phi \theta; \alpha) \text{ w.p.1 as } N \to \infty
\]

for some \( 2n(m+1) \times n \) matrix \( \Phi \). Partition \( \Phi \) into \( 2(m+1) \) submatrices of dimension \( n \times n \), \( \Phi = [\Phi_1, \Phi_2, \ldots, \Phi_{2(m+1)}] \).

Let \( u \in U \), it can be directly verified that \( \Phi_1 \) is the \( n \times n \) circulant matrix of symbol \( v, \Phi_1 = \mathbf{T}\{[v_n, \ldots, v_1]\} \), and the odd-indexed block matrices are expressed as \( \Phi_3 = \rho_1 \Phi_1, \ldots, \Phi_{2m+1} = \rho_m \Phi_1 \).

Under this input, the limit \( \xi \) in (11) for the system \( y_k = \phi_k \theta + d_k, s_k(\theta) = s_k(y_k) \) contains the following equations by extracting the odd-indexed blocks \( \xi^{(2l+1)} = F(C_n - \rho_j \Phi_j \theta; \alpha) \) for \( j = 1, \ldots, m \). We now show that these equations are sufficient to determine \( \theta \) and \( \alpha \) uniquely.

**Theorem 8:** Suppose that \( u \in U \) and \( F(x; \alpha) \) satisfies Assumption A6 and is jointly identifiable. Then, \( \xi = F(C_{n(m+1)} - \Phi \theta; \alpha) \) has a unique solution \( \theta^* \) and \( \alpha^* \).

**Proof:** Consider the first block \( \Phi_1 \theta \) of \( \Phi \theta \). Since \( v \) is full rank, \( \Phi_1 \) is a full rank matrix. It follows that, for any \( \theta \), \( \Phi_1 \theta \neq 0 \). Without loss of generality, suppose that the \( \nu \)th element \( \delta \) of \( \Phi_1 \theta \) is nonzero. By construction of \( \Phi_1 \), we can extract the following \( m+1 \) nonzero elements from \( \Phi_1 \): the \( (2n+\nu) \)th element, \( l = 1, \ldots, m \), is \( \rho_\nu \). Extracting these rows from the equation \( \xi = F(C_{n(m+1)} - \Phi \theta; \alpha) \) leads to a set of \( m+1 \) equations that will be denoted by

\[
\xi^0 = F(C_{m+1} - \rho \delta; \alpha)
\]

where \( \delta = [1, \rho_1, \ldots, \rho_m]^T \). Since \( \delta \neq 0 \) and \( \Pi \) has distinct elements, \( C_{m+1} - \rho \delta \) has distinct elements. By hypothesis, \( F(x; \alpha) \) is jointly identifiable. It follows that (12) has a unique solution \( \theta^* \) and \( \alpha^* \). Now, using the already obtained \( \alpha^* \), let the first \( n \) equations of \( \xi = F(C_{n(m+1)} - \Phi \theta; \alpha) \) be denoted by \( \xi^{1} = F(C_n - \Phi_1 \theta; \alpha^*) \). By Assumption A6, \( G(x; \alpha^*) \) exists. As a result, \( \theta^* = \Phi_1^{-1} (C_n - G(\xi^{1}; \alpha^*)) \) is the unique solution. This completes the proof.

**C. Exponentially Scaled Signals**

A particular choice of the scaling factors \( \rho_j = q^j, j = 1, \ldots, m \) for some \( q \neq 0 \) and \( q \neq 1 \). In this case, the period of input \( u \) can be shortened to \( n(m+1) \) under a slightly different condition.

**Definition 5:** An \( n(m+1) \)-periodic signal \( u \) is called an exponentially scaled full rank signal if its one-period values are \((v, qv, \ldots, q^m v)\), where \( q \neq 0 \) and \( q \neq 1 \), and \( v = (v_1, \ldots, v_n) \) satisfies that \( \Phi = T(q, [v_n, \ldots, v_1]) \) is full rank. We use \( U \) to denote this class of input signals.

Let \( \xi_N \) be defined as in (5), with dimension changed from \( n \) to \( n(m+1) \). By the strong law of large numbers, \( \xi_N \rightarrow \xi = F(C_{n(m+1)} - \Phi \theta; \alpha) \) w.p.1, as \( N \to \infty \), for some \( (n(m+1)) \times n \) Toeplitz matrix \( \Phi \). Partition \( \Phi \) into \( (m+1) \) submatrices of dimension \( n \times n \), \( \Phi = [\Phi_1^T, \Phi_2^T, \ldots, \Phi_{m+1}^T] \).

If \( u \in U \), then it can be directly verified that \( \Phi_1 \theta = \xi \) is the \( n \times n \) general circulant matrix defined in Definition 5 and \( \Phi_1 = q^{-\delta} \Phi_1^T, l = 2, \ldots, m+1 \). We have the following result, whose proof is similar to that of Theorem 8 and is omitted.

**Theorem 9:** Suppose that \( u \in U \), and \( F \) satisfies Assumption A6 and is jointly identifiable. Then, \( \xi = F(C_{n(m+1)} - \Phi \theta; \alpha) \) has a unique solution \( \theta^* \) and \( \alpha^* \).

**D. Recursive Algorithms**

A recursive algorithm for computing estimates of \( \theta \) and \( \alpha \) can be constructed as follows. For notational simplicity, the dimensions of the matrices, which will become clear from context, are suppressed. Recall that, for any fixed \( \alpha, G(:, \alpha) \) is the inverse of \( F(:, \alpha) \). For a fixed \( \theta \) and a fixed \( \xi \), starting from \( \alpha_0 \), we wish to construct estimates of \( \alpha \) by solving a nonlinear least-squares problem \( \min_{\alpha} (\xi - F(C - \Phi \theta; \alpha))^T (\xi - F(C - \Phi \theta; \alpha)) \).

Nevertheless, we do not really have a fixed constant \( \xi \). Rather it is a sequence of empirical measures. Thus, the problem is not purely deterministic, but involves random processes.

In what follows, we outline a recursive algorithm with multiple levels of updates. There are several estimates involved. First, we still use the empirical measures since the binary data are the only measurements available. Second, we construct a stochastic algorithm for recursively estimating \( \alpha \). Third, we carry out an inversion to obtain an estimate of \( \theta \). Taking into consideration the frequencies of updates, it appears to be more productive that we do not perform the inversion at every iteration. This is the rationale for using a two-level procedure.

To begin, for a sequence \( \{x_k\} \) (real numbers, or vectors, or matrices with appropriate dimensions), denote \( \pi_k^{(N)} = \pi_{N+k} \).

The procedure consists of inner and outer iterations. In the inner iteration, we update the estimates of the empirical measures as well as that of \( \alpha \); in the outer loop, we update \( \theta_k \) that is kept as a constant during the inner iteration. For the inner iterations, we also solve an optimization of the form \( \min_{\alpha} K(\theta, \alpha) = E((\xi^{(N)} - F(C - \Phi \theta; \alpha))^T (\xi^{(N)} - F(C - \Phi \theta; \alpha))) \).

Note that the expectation is not available. Instead we use its noise-corrupted observed values \((\partial/\partial \alpha) K(\xi^{(N)}, \theta, \alpha) \).

The construction of the estimates are recursive. Suppose that \( (\xi^{(N)}_k, \theta^{(N)}_k, \alpha^{(N)}_k) \) has been constructed. Then, the recursion is defined by the following algorithm

\[
\xi_{k+1}^{(N)} = \xi_k^{(N)} + \frac{1}{\ell N + k + 1} \xi_k^{(N)} + \frac{1}{\ell N + k + 1} S_{k+1}^{(N)}, k < N
\]

\[
\alpha_{k+1}^{(N)} = \alpha_k^{(N)} + \frac{\theta_k^{(N)}}{\partial \alpha} \frac{\partial K(\xi_k^{(N)}, \theta_k^{(N)}, \alpha_k^{(N)})}{\partial \alpha}, k < N
\]

For simplicity, we assume that the partial derivatives can be observed. Otherwise, we can use the finite difference to approximate the gradient.
In the algorithm above, \( \{\beta_k\} \) is a sequence of step-sizes satisfying \( \beta_k \geq 0, \beta_k \to 0 \) as \( k \to \infty \), and \( \sum_k \beta_k = \infty \).

As demonstrated in the previous sections, a sufficient condition for the sequence of empirical measures to converge is the ergodicity. Thus, we simply assume that \( \{\xi_k\} \), the sequence of empirical measures is stationary and ergodic. Suppose also that for each \( \varepsilon \), the function \( K(\varepsilon, \alpha) \) has a unique minimizer \( \alpha \). Then, using the ordinary differential equation (ODE) methods [20], we can show that \( \theta_k^N \to \theta \) w.p.1 as \( N \to \infty \). The results in the previous section reveal that \( \xi_k^N \) also converges. Finally, similar to the previous sections, the inversion leads to \( \theta_k^N \to \theta \) w.p.1 as desired.

### VIII. ILLUSTRATIVE EXAMPLES

In this section, we will use two examples to demonstrate the algorithms developed in this paper. Example 4 illustrates the case when the switching threshold is unknown. It shows that when the input is \( n + 1 \) full rank, both the threshold \( C \) and system parameters \( \theta \) can be estimated simultaneously. Example 5 covers the scenario of unknown noise distributions. The input design and joint identification algorithms are shown to lead to consistent estimates.

**Example 4:** Suppose that the threshold \( C \) is unknown and the input has measure noise. Consider a third-order system: \( y_k = \phi_k^T \theta + d_k \), where the output is measured by a binary-valued sensor with unknown threshold \( C \). Suppose that the true parameters are \( C = 28 \) and \( \theta = [2.1, 2.7, 3.6]^T \). \( \{d_k\} \) is a sequence of i.i.d. normal variables with mean zero and variance \( \sigma^2 = 4 \). The noise-free input \( v \) is four-periodic with one period values \( [3.1, 4.3, 2.3, 3.5] \), which is full rank. The actual input is \( u_k = u_k + \xi_k \), where \( \{\xi_k\} \) is a sequence of i.i.d. normal variables with mean zero and variance \( \sigma^2 = 1 \).

For \( n = 3 \), define \( \tilde{\Phi}_j = [y_{4(j-1)+1}, \ldots, y_{4j}]^T \in \mathbb{R}^4 \), \( \tilde{\Phi}_j = [\phi_{4(j-1)+1}, \ldots, \phi_{4j}]^T \in \mathbb{R}^{4 \times 3} \), \( \tilde{D}_j = [d_{4(j-1)+1}, \ldots, d_{4j}]^T \in \mathbb{R}^4 \), \( \tilde{S}_j = [s_{4(j-1)+1}, \ldots, s_{4j}]^T \in \mathbb{R}^4 \). It follows that \( \tilde{Y}_j = \tilde{\Phi}_j \theta + \tilde{D}_j \), for \( j = 1, 2, \ldots \). Since \( \{\tilde{D}_j\} \) is a sequence of i.i.d. normal variable vectors, we have \( \tilde{\xi}_N = \frac{1}{N} \sum_{j=1}^N \tilde{S}_j \to \mathcal{N}(\tilde{\Phi}(\tilde{\Psi})) \).

Since \( v \) is full rank, \( \tilde{\Psi} \) is invertible. If \( \tilde{\Psi} \) is known, by the continuity of \( \tilde{\Phi} \) and \( \tilde{G} \), an estimate of \( \theta \) can be constructed as \( (\tilde{\Psi})^{-1} \tilde{G}(\tilde{\xi}_N) \). Due to the input measure noise, \( \tilde{\Psi} \) is not measured directly. What we can use is \( \tilde{\Psi}_N \). Theorem 6 confirms that \( \tilde{\Theta}_N = (\tilde{\Psi}_N)^{-1} \tilde{G}(\tilde{\xi}_N) \) will be a consistent estimate of \( \Theta \).

Set initial conditions as \( \tilde{\xi}_1 = \tilde{S}_1 = [1, 1, 1, 1]^T \), \( \tilde{\Phi}_1 = \tilde{\Phi}_1 \), and \( \Theta_0 = 0 \). We construct a causal and recursive algorithm as in Section VI-B. The relative estimation error \( \|\tilde{\Theta}_N - \Theta\|/\|\Theta\| \) is used to evaluate accuracy and convergence of the estimates. Fig. 5 shows that \( \tilde{\Theta}_N \) converges to the true parameters \( \tilde{\Theta} = [C, \theta^T]^T \).

**Example 5:** When the noise distribution function is unknown, joint identification is used to estimate jointly the system parameters and noise distribution function. Consider a gain system \( (n = 1) \): \( y_k = au_k + d_k \), where the true value \( a = 2 \), and \( \{d_k\} \) is a sequence of i.i.d. normal variables. The sensor has threshold \( C = 12 \). Let \( F_0(x) \) be the normal distribution function of known mean and variance \( 1 \), and \( G_0(x) \) be the inverse of \( F_0(x) \). Then the distribution function of \( d_k \) is \( F(x; [\mu, \sigma]) = F_0((x - \mu)/\sigma) \).

Let \( \mu = 3 \) be given, and the true value of variance \( \sigma = 3 \). By Example 3, if \( \mu \) is known, \( F(x; [\mu, \sigma]) \) is jointly identifiable. Let \( v = 4 \). For \( k = 1, 2, \ldots \), the scaled input is defined as \( u_{2k-1} = v \); \( u_{2k} = kv \), where \( q = 1.05 \). It is easy to verify that \( v \) is an exponentially scaled full rank signal (Definition 5). Set \( U = [4, 4.2]^T \) and \( \xi_N = \frac{1}{N} \sum_{i=1}^N s_{2i-1} - \frac{1}{N} \sum_{i=1}^N s_{2i} \).

Then, \( \xi_N \to \xi \) w.p.1. and \( G_0(\xi_N) \to [1 - (\mu - 1)2] - aU]/\sigma \).

Since \( F(x, \alpha) \) is jointly identifiable, we obtain the estimates of \( \alpha \) and \( \sigma = (\tilde{a}_N, \tilde{\sigma}_N) = [U, G_0(\xi_N)]^{-1}[8, 8]^T \). Fig. 6 illustrates that the estimated values of the system parameter and distribution function parameter converge to the actual ones.

### IX. FURTHER REMARKS

#### A. Extensions

**Rational models:** We now summarize some results from [32] that show input conditions ensuring strong convergence of parameter estimates for rational systems. Consider the following system \( y_k = G(q)u_k + d_k \), where \( G(q) \) is a stable rational transfer function of \( q \), \( G(q) = \frac{B(q)}{1 - A(q)} = \frac{b_0 + b_1 q + \cdots + b_N q^N}{1 - a_1 q - \cdots - a_N q^N} \). The rational parameters \( \theta = [a_1, \ldots, a_N, b_0, b_1, \ldots, b_N]^T \) need to be identified. Suppose that \( u_k \) is 2n-periodic and the observation length \( N = 2nL \) for some positive integer \( L \). Then the noise-free system output \( x_k = G(q)u_k \) is also 2n-periodic, after a short transient period. Hence, for some unknown real numbers \( c_j, j = 1, \ldots, 2n, x_j = c_j \), \( j = 1, \ldots, 2n, x_{j+2n} = x_j \), for any positive integer \( L \). Then, for a given \( j \in \{1, \ldots, 2n\}, y_{j+2n} = c_j + d_j + 2\alpha \), \( \ell = 0, 1, \ldots, L - 1 \). Here, we need to identify \( c_j \).
Next, we establish a mapping from $[c_1, \ldots, c_{2n}]$ to $\theta = [a_1, \ldots, a_n, b_1, \ldots, b_n]^T$. Recall that $x_k = G(q) u_k = \frac{1}{1 - \phi_1 q^{-1} - \cdots - \phi_p q^{-p}} u_k$, or in a regression form $x_k = \phi_k^T \theta$, $k = 1, \ldots, N$, where $\phi_k^T = [x_{k-1}, \ldots, x_{k-n}, u_{k-1}, \ldots, u_{k-n}]$, and $\theta = [a_1, \ldots, a_n, b_1, \ldots, b_n]^T$. For any starting time $k_0$, define $\Phi = [\phi_{k_0}, \ldots, \phi_{k_0+2n-1}]^T$ and $X = [x_{k_0}, \ldots, x_{k_0+2n-1}]^T$. Then, $X = \Phi \theta$. Apparently, if $\Phi$ is invertible, $\theta = \Phi^{-1} X$ defines a mapping from $[c_1, \ldots, c_{2n}]$ to $\theta$. Consequently, the identification of $\theta$ is reduced to that of $[c_1, \ldots, c_{2n}]$. Moreover, identifying the rational transfer function can now be reduced to identifying the set of gains. In the following theorem, Statements i) and ii) are in [32, Th. 2], and Statement iii) is in [32, Th. 3]. Statement iv) is new, whose proof is omitted since it is similar to that of Theorem 1.

**Theorem 10:** Suppose that the pair $D(q) = 1 - A(q)$ and $B(q)$ are coprime polynomials, i.e., they do not have common roots. If $u_k$ is $2n$-periodic and full rank, then

i) $\Phi$ is invertible for all $k_0$;

ii) $\|\Phi^{-1}\|_2$ is independent of $k_0$, where $\|\cdot\|_2$ is the largest singular value;

iii) $u_k$ is sufficiently rich for identifying $\theta$;

iv) If $u_k$ is $2n$ periodic but not full rank, then it is not sufficiently rich for identifying $\theta$.

**Correlated noises:** Up until now, we have assumed the noise $\{d_k\}$ to be uncorrelated. This condition can be much weakened. In fact, the i.i.d. condition is mainly for convenience and notational simplicity. Under this condition, we have highlighted the main issues in the input design for binary-valued output observations without undue technical complications. To illustrate, let us suppose that there is a sequence $\{d_k\}$ of i.i.d. normal random variables with mean zero and variance $\sigma^2$ such that $d_k = \sum_{i=0}^p c_i d_{k-i}$, a moving average process. Then, it is easily seen that $d_k$ is still a normal random variable with mean zero and variance $\sum_{i=0}^p c_i^2 \sigma^2$. If $c_k$ has a common distribution function $F(\cdot)$, then $d_k$ has a distribution function $F(x/\sqrt{\sum_{i=0}^p c_i^2})$. That is, only the scale is changed. All previous discussions still carry over.

The moving average processes present a scenario of finitely correlated noise. Next, in lieu of the finite correlated noise, if $d_k$ is a $\phi$-mixing sequence, which assumes the remote past and distant future being asymptotically independent, then as pointed out in Section V-B, it is well-known that the sequence is strongly ergodic (see [17, p. 488]). Thus the limit of the empirical measures as well as the centered and scaled sequence of errors leading to the Brownian bridge limit still hold [2]. As a result, we can push the envelop to include such infinitely corrected noises for the input designs with binary-valued output.

**B. Concluding Remarks**

Conditions and input signal designs for system identification using binary-valued observations are developed for different system configurations (open- and closed-loop systems), scenarios of noises (input measurement noise, input actuator noise, and output noise), structural uncertainties (unknown sensor threshold), and noise distributional uncertainty (unknown distribution functions). The concept of sufficient richness conditions is introduced to capture the essential requirements on input signals for consistent estimation in these circumstances. There are many potential extensions of the results in this paper. For example, when system models contain unmodeled dynamics, sufficient richness conditions will become more involved. These issues will be reported elsewhere.

**REFERENCES**


Le Yi Wang (S’85–M’89–SM’01) received the Ph.D. degree in electrical engineering from McGill University, Montreal, Canada, in 1990.
Since 1990, he has been with Wayne State University, Detroit, MI, where he is currently a Professor in the Department of Electrical and Computer Engineering. He is currently an Editor of the *Journal of System Sciences and Complexity* and an Associate Editor of the *Journal of Control Theory and Applications*. His current research interests include complexity and information, system identification, robust control, H-infinity optimization, time-varying systems, adaptive systems, hybrid and nonlinear systems, information processing and learning, as well as medical, automotive, communications, and computer applications of control methodologies.

Prof. Wang was an Associate Editor of the *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*. He was a keynote speaker in several international conferences. He was also on the International Federation of Automatic Control (IFAC) Technical Committee on Modeling, Identification and Signal Processing.

G. George Yin (S’87–M’87–SM’96–F’02) received the B.S. degree in mathematics from the University of Delaware, Newark, in 1983, the M.S. degree in electrical engineering, and the Ph.D. in applied mathematics both from Brown University, Providence, RI, in 1987.
He joined the Department of Mathematics, Wayne State University, Detroit, MI, where he has been a Professor since 1996. He is an Associate Editor of Automatica and the SIAM Journal on Control and Optimization, and is or was on the editorial board of a number of other journals.
Prof. Yin was an Associate Editor of the *IEEE TRANSACTIONS ON AUTOMATIC CONTROL* from 1994 to 1998. He was with the Mathematical Review Database Committee, International Federation of Automatic Control (IFAC) Technical Committee on Modeling, Identification and Signal Processing, and various conference program committees. He was the editor of the Society for Industrial and Applied Mathematics (SIAM) Activity Group on Control and Systems Theory Newsletters, the Co-Chair of the 1996 American Mathematical Society (AMS)-SIAM Summer Seminar in Applied Mathematics, and the Co-Chair of the 2003 AMS-Institute of Mathematical Statistics (IMS)-SIAM Summer Research Conference: Mathematics of Finance, Co-Organizer of the 2005 IMA Workshop on Wireless Communications, and Co-Organizer of the 2006 IMA PI Conference.

Yanlong Zhao (M’07) was born in Jinan, China, in 1980. He received the B.S. degree in mathematics from Shandong University, Jinan, in 2002, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2007.
From August 2005 to July 2006, he was with Wayne State University, Detroit, MI, as a visiting scholar. Since 2007, he has been with AMSS, CAS, where he is currently a Research Associate. His current research interests include identification and control of quantized systems.
Ji-Feng Zhang (M’92–SM’97) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 1985, and the Ph.D. degree in operation research and control theory from the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), Beijing, China, in 1991.

Since 1985, he has been with ISS, CAS, where he is currently a Professor at the Academy of Mathematics and Systems Science and also the Vice-Director of the ISS. He is a Deputy Editor-in-Chief of the journals Acta Automatica Sinica, Journal of Systems Science and Mathematical Sciences, a Managing Editor of the Journal of Systems Science and Complexity, and an Associate Editor of several other journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and SIAM Journal on Control and Optimization. His current research interests include system modeling and identification, adaptive control, stochastic systems, and descriptor systems.

Prof. Zhang was the recipient of the Distinguished Young Scholar Fund from the National Natural Science Foundation of China in 1997 and the First Prize of the Young Scientist Award of CAS in 1995. He is currently the Vice-General Secretary of the Chinese Association of Automation.