Decision-Based System Identification and Adaptive Resource Allocation

Jin Guo, Biqiang Mu, Le Yi Wang, *Fellow, IEEE*, George Yin, *Fellow, IEEE*, and Lijian Xu, *Member, IEEE*

***Abstract*—System identification extracts information from a system's operational data to derive a representative model for the system so that a decision can be made with desired accuracy and reliability. When resources are limited, especially for networked systems sharing data and communication power and bandwidth, identification must consider complexity as a critical limitation. Focusing on optimal resource allocation under a given reliability requirement, this paper studies identification complexity and its relations to decision making. Dynamic resource assignments are investigated. Algorithms are developed and their convergence properties are established, including strong convergence, almost sure convergence rate, and asymptotic normality. By a suitable design of resource updating step sizes, the algorithms are shown to achieve the CR lower bound asymptotically, and hence are asymptotically efficient. Illustrative examples demonstrate significant advantages of our real-time and individualized resource allocation methodologies over population-based worst-case strategies.**

***Index Terms*—Complexity, decision, resource allocation, system identification.**

I. Introduction

SYSTEM identification supports decisions by extracting information from a system's operational data to derive a representative model for the system. Studies of system identification have been concentrated on model structure selection, model parametrization, estimation algorithms, experimental design, consistency and convergence of parameter estimates, rate of convergence, etc. [1]–[4]. In these aspects, system identification is a well developed field.

When resources are limited, especially for networked systems sharing data and communication power and bandwidth, complexity may become a critical limitation. This paper studies identification complexity and its relations to decision making. The term "decision" is used in a broad sense, including controller design, fault detection, status monitoring, communication network management, medical outcome prediction, to name just a few.

As an integrated part of a decision process, goals of identification, especially estimation accuracy, are dependent on its targeted decisions. More concretely, it is observed that identification accuracy requirements change dramatically over different operating ranges. For the ranges in which parameter accuracy is not highly required, one may select to use a model of lower complexity, to use less observation data, to reduce data acquisition rates, and to request less computational resources. For example, a robust controller can tolerate parameter variations. The more robust the controller is, the less accurate the identification needs to be. When system parameters drift outside the robust region of the controller, the controller must be adapted. Depending on how far the parameters are from the robustness boundary of the controller, identification accuracy requirement varies, and the resources should be assigned accordingly. In this case identification accuracy depends directly on the capability of robust control (a decision process) and the operating points of the system. In decision-based identification, identified models only need to be sufficiently accurate for making decisions. In such applications, it is not necessary to identify systems to a uniform precision, or to establish parameter convergence, or even to be identifiable over the entire parameter space.

In this paper, the designated "resources" will be represented by the data amount in a pre-determined time interval, which is problem specific. Typical examples of such resources include: (1) allocated bandwidths in communication networks; (2) data acquisition speed in process control problems; (3) frequencies of information exchange between management and its subordinates; (4) rates of workload re-assignment in parallel computing; among others.

To illustrate, consider an example of TDMA (time division multiple access) protocols in Fig. 1, in which m users are sharing a communication channel [5]. All users update their estimates of

Manuscript received January 12, 2015; revised May 3, 2016, July 20, 2015, December 21, 2015, and December 22, 2015; accepted August 11, 2016. Date of publication September 21, 2016; date of current version April 24, 2017. This research was supported in part by the Air Force Office of Scientific Research under FA9550-15-1-0131, in part by the National Natural Science Foundation of China under grants 61403027 and 61603379, and in part by the President Fund of Academy of Mathematics and Systems Science, CAS under Grant No. 2015-hwyxqnrc-mbq and the National Key Basic Research Program of China (973 program) under Grant 2014CB845301. Recommended by Associate Editor M. Verhaegen.

J. Guo is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: guojin@amss.ac.cn).

B. Mu is with the Key Laboratory of Systems and Control of CAS, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: bqmu@amss.ac.cn).

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

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

L. Xu is with the Department of Electrical and Computer Engineering Technology, Farmingdale State College, the State University of New York, Farmingdale, NY 11735 USA (e-mail: xul@farmingdale.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2612483

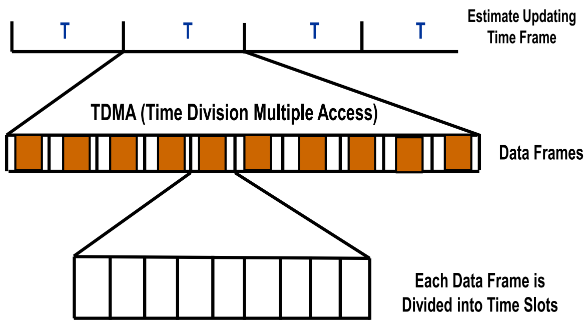


Fig. 1. Communication resource sharing by multiple users.

their own parameters once every T seconds. Since users' parameters can potentially change with time, data in a time window are not re-used in subsequent windows. Resource allocation aims to assign time slots according to certain strategies. In this case, the assigned time slots to each user determine its data throughput and is a measure of identification complexity.

The idea of "complexity versus required accuracy" has been discussed conceptually and diversely in many applications. Complexity issues in modeling and identification have been pursued by many researchers. The concepts of ε-net and ε-dimension in the Kolmogorov sense [6] were first employed by Zames [7], [8] in studies of model complexity and system identification. For certain classes of continuous-time and discrete-time systems, the n-widths and ε-dimensions in the l^1 kernel norm and the H^∞ norm were obtained in [7], [9], [10]. The notion of identification n-width was introduced in [11] to characterize intrinsic complexities in worst-case identification problems. Complexity issues in system identification were studied in [12], [13]. n-widths of many other classes of functions and operators were summarized in the books by Pinkus [14] and Vitushkin [15]. A general framework of information based complexity was comprehensively developed in [16]. Complexity issues in estimation and feedback control problems have recently attracted great attention. These include estimation with communication uncertainties [17], [18], stabilization with limited data rates [19]–[21], and networked control systems [22]. This paper explores new issues of resource management in this direction.

In our earlier work [23], complexity relationships between system identification and control robustness were explored in an information-based complexity setting. The work reported in this paper explores more concretely and more generally this concept. A closely related problem of resource saving in system identification is quantization. Comprehensive studies of system identification under limited measurement information can be found in [24]–[28] and the references therein.

This paper focuses on dynamic resource assignments for a group of systems, in which we seek "optimal" resource allocation so that the decision reliability in terms of probability of correct decisions is uniform for the entire group. To achieve this goal, the resource assignment is dynamic and individualized within the group. Adaptive optimal resource assignment algorithms are devised and their convergence performance established. The problems are technically challenging due to a mixture of continuous variable and discrete variables and interaction between the data size and estimation accuracy, leading to a stochastic approximation algorithm with mixed random noises.

For clarity and simplicity, the paper is structured to present the technical details in the scalar case first. It is shown that the algorithms are strongly convergent. The asymptotic normality is also established. By a suitable design of resource updating step sizes, the algorithms are shown to achieve asymptotically the CR (Cramer-Rao) lower bound, and hence are asymptotically efficient. The results are then extended to multivariate systems. By selecting proper input signals and decision reliability requirement we can make full use of the techniques and results in the scalar case and obtain the corresponding results. Although this paper presents the technical results on the basic "gain" systems, there is an important link to more general FIR and ARMAX systems. By using suitable periodic inputs, identification of FIR and ARMAX models can be reduced equivalently to a collection of the basic problems of identifying "gain" systems discussed in this paper, by using the similar approaches as in our previous work; see for example [28].

The rest of the paper is organized into the following sections. Section II formulates the decision-based identification problem as a set identification problem with a reliability requirement. Technical developments start with scalar identification problems in Section III. Adaptive resource allocation algorithms are introduced. The algorithms are shown to converge to the optimal resource allocation by employing the ODE approach in stochastic approximation methodologies. Convergence rates, asymptotic normality, and asymptotic efficiency of the algorithms are established. The algorithms are then generalized to higher-dimensional cases in Section IV. Convergence and convergence rates are established. Finally, findings of the paper are summarized in Section V, together with some open issues.

II. Problem Formulation

We start with a description of decision-based identification and related resource assignment problems. Consider a discrete-time, linear, time-invariant, single-input-single-output system $y = G_\theta u + d$ where u is the input, y is the output, and d represents the observation noise. θ is to be identified. In this paper, we consider the following discrete-time finite impulse response (FIR) system:

$$y_k = a_0 u_k + \cdots + a_{n-1} u_{k-n+1} + d_k = \phi_k' \theta + d_k, \qquad (1)$$

where $\theta = [a_0, \ldots, a_{n-1}]'$ is the unknown parameter vector, $\phi_k' = [u_k, \ldots, u_{k-n+1}]$ is the regressor, and d_k is the observation noise. An important special case is the gain system $(n = 1)$

$$y_k = \theta u_k + d_k. \qquad (2)$$

The primary scenario of this paper is: A group of systems are to be identified with some shared resource. Its members may change their parameter values. As a result, only the "most recent" data, defined as a time window of a given length T, can be used in estimation, and past data will not be re-used. Suppose that the maximum permissable number of observations in a pre-designated time interval of length T, is $N_{\max}$. Dynamic resource allocation aims to determine how much resource should be assigned to a member for its identification. Let θ_k denote an estimate of θ by using N_k observations during the kth updating interval $[kT, (k+1)T)$. In this paper, the observation length N_k represents the resource consumption. Dependence of the observation length on k indicates a dynamic resource assignment problem in which the data size in each updating interval is dynamically assigned.

In traditional system identification, the question is "what is the value of θ^*?" and convergence of the estimates to the true parameter θ^*, when $N \to \infty$, is the ultimate goal. In contrast, in decision-based identification the question becomes "is $\theta^* \in S$?" since the set S is associated with a control or decision, and we want to answer this question within an acceptable probabilistic reliability.

Due to limited resources, to ensure decision accuracy we designate a suitable subset $M \subset S$ for reliability assessment. Implicitly, it is assumed that the decision robustness can cover the uncertainty set S. For example, in a robust control problem for systems with gain uncertainty, if one robust controller F_1 can cover gain uncertainty from $[1, 2)$ and the second robust controller F_2 can cover $[1.5, 3)$, then $M = [1, 1.5)$ and $S = [1, 2)$. Consequently, controller selection is wrong only when the true parameter is in M but its estimate leaves S, resulting in using the wrong controller F_2. It is noted that if the true parameter is actually in $[1.5, 2)$, both controllers are valid and there will be no reliability issue here. We also observe that in this case, there is a similar set identification problem and its reliability: When the true parameter is in $M = [2, 3)$ but its estimate leaves $S = [1.5, 3)$ to be in $[1, 1.5)$, resulting in using the wrong controller F_1. Since these two problems are identical, we focus on algorithms and their convergence properties for one.

In a stochastic setting, decision accuracy requirements are stated as

$$\text{if } \theta^* \in M,\ \Pr\{\theta_k \in S\} \geq 1 - \alpha,$$

where $\Pr\{\cdot\}$ denotes probability and $0 < \alpha < 1$ is the decision error bound. In this framework, identification aims to determine if $\theta^* \in M$. In other words, this is a set identification problem.

In its connection to a decision, a pre-designed decision (a controller, a diagnosis, etc.) will be made if $\theta_k \in S$. Implicitly, when $\theta_k \notin S$, some other decisions (a different robust controller, another diagnosis, etc.) will be made, which will be another decision set or sets. So, the above decision reliability problem is generic.

We remark that unlike traditional system identification, here we seek minimum resource consumption for each individual system within a population, under the condition that the required decision reliability is uniformly met for all members in the population. If θ^* is known, then the optimal N^* can be obtained by the standard statistical analysis. On the other hand, since the true θ^* is unknown, the standard practice in statistical hypothesis testing (which is also a set identification problem) is to use the worst-case strategy for the population to determine the sample size. As a result, the optimality of resource allocation for each individual within the population is lost. The main question is: Can we devise an online resource allocation algorithm that achieves convergence to N^*? When N^* is obtained, the parameter estimation in the subsequent intervals will be sufficiently accurate to support the decision and consume the minimum resource.

III. Basic Adaptation Schemes for Resource Assignment: Scalar Cases

The generic identification algorithm structure, leaving out the actual estimation scheme, is the following iterative resource updating structure:

Generic Identification and Resource Assignment Structure:

1) At the kth updating step, assign the resource $N_{k+1}(\theta_k)$ based on the current estimate θ_k.
2) The parameter estimate is updated to θ_{k+1} by using N_{k+1} observations during the $(k+1)$th interval $[(k+1)T, (k+2)T)$.

In algorithm development, we seek "optimal" or "minimum" resource assignments in the following sense. Let the estimate θ_N and the true value θ^* be related by

$$\theta_N = \theta^* + e_N$$

where e_N is the estimation error whose distribution function depends on the data size N. For a given decision error bound $0 < \alpha < 1$, the minimum resource assignment N^* is defined as: Under the condition that $\theta^* \in M$

$$N^*(\theta^*) = \min\{N : \text{such that } \Pr_{e_N}\{\theta_N \in S\} \geq 1 - \alpha\}. \quad (3)$$

Remark 1: For implementation, it is understood that the integer roundoff of N^*, namely, $\lceil N^* \rceil$ where $\lceil z \rceil$ denotes the smallest integer greater than or equals to $z \in \mathbb{R}$, will be used. With this understanding, for convenience of rigorous analysis, we will use (3) in all subsequent development.

For clarity and simplicity, we start with the basic scenario of scalar cases and simple decision sets to convey the key issues and main ideas of the algorithms. Generalization will follow in the subsequent sections. We first describe the main ideas without technical details.

Typical monitoring or diagnosis problems for a given θ concern the problem: Is $\theta \leq C$? The threshold C can be the systolic or diastolic blood pressures in hypertension monitoring; the SOC (State of Charge) upper bound to avoid battery overcharge and thermal runaway; the load limit on a transmission line; among many others. Mathematically, this is a set identification problem: Evaluate if $\theta \in (-\infty, C]$ or $\theta \in [C, \infty)$. Since $\theta \geq C$ if and only if $2C - \theta \leq C$, without loss of generality, we consider only the problem $\theta \in (-\infty, C]$.[1]

If θ is estimated by $\widehat{\theta}$ with an estimation error e whose probability density function is symmetric such as Gaussian random variables, it is clear that no matter how many data points are

[1] The classical hypothesis testing contains two disjoint sets M_1 and M_2, representing "normal" vs. "fault", or "legal" vs. "illegal", etc. In our setting, this can be simply represented by $M_1 = (-\infty, C^*)$ and $M_2 = (C, \infty)$. The subsequent development of our algorithms can be applied as two parallel resource assignment problems. In classical hypothesis testing, one pre-determines the sample size for the entire possible population (a priori information), and then runs the test to reach a posterior conclusion. The sample size is not adapted or individualized.

taken and how small is the variance of the estimation error, the worst-case scenario for decision reliability is

$$\max_{\theta \le C} \Pr\{\widehat{\theta} > C\} = 0.5,$$

implying an intractable problem for decision reliability analysis. For this reason, we designate a value $C^* < C$ such that when the maximum data size $N_{\max}$ is used, the decision reliability is met. Then, we have $M = (-\infty, C^*)$ and $S = (-\infty, C]$. The main goal of real-time adaptive resource allocation is to ensure that when the true parameter is below C^*, only the minimum data size $N^* < N_{\max}$ is consumed without compromising the required decision reliability.

In an off-line setting, this may be viewed as a hypothesis testing problem with $H_0 : \theta \in (-\infty, C^*)$ and $H_1 : \theta \in [C, \infty)$. The decision reliability can be defined as the Type I error. Accordingly, the selection of N^* falls in the field of sample size determination. It should be noted that θ^* or $C - \theta^*$ need to be known for the computation of the acceptance region, rejection region and the sample size in the classical statistical hypothesis testing [29], [30]. The iterative procedure in this paper will deal with the case that θ^* and $C - \theta^*$ are unknown.

A. Set Identification and Optimal Resource Assignment

Consider the basic estimation scheme for a scalar θ^*. For the gain system (2), if $u_k \equiv 1$, then the system becomes $y_k = \theta + d_k$, which is a case of estimating a scalar. In common vital sign monitoring problems, a patient's heart rate can be measured by processing EKG signals. In battery management systems, voltage or current measurements with noise corruption are also such estimation problems. In a given time interval of length T, noise-corrupted measurements of θ^* are obtained

$$y_j = \theta^* + d_j, j = 1, \dots, N,$$

where d_j is i.i.d., Gaussian distributed, mean zero and variance σ^2. It is well known that the minimum-variance (or maximum likelihood) estimate of θ^* is given by

$$\widehat{\theta} = \frac{1}{N}\sum_{j=1}^{N} y_j = \theta^* + e,$$

where e is Gaussian, mean zero and variance σ^2/N. To monitor θ^* persistently for its potential drifting outside M, in the kth interval $[kT, (k+1)T)$, $k = 0, 1, \dots$, the data of length N_k within that interval are used for parameter estimation, leading to the estimate

$$\theta_k = \theta^* + e_k.$$

Suppose that the maximum data length for this user is $N_{\max}$. C^* and C must have the sufficient gap so that by using the maximum resource $N_{\max}$, the decision reliability can be met. This is calculated as follows. The density function of the standard Gaussian distribution $\mathcal{N}(0,1)$ is denoted by $f(\cdot)$. Define the Gaussian tail function as

$$Q(x) = \int_x^{\infty} f(y)dy.$$

Since α is usually close to 0, we set $0 < \alpha < 1/2$ in the following derivations. Let x_0 satisfy $Q(x_0) = \alpha$. Under the maximum resource assignment, $e \sim \mathcal{N}(0, \sigma^2/N_{\max})$. The maximum θ^* that can ensure the decision error α can be calculated as

$$C^* = \max\left\{\theta^* < C : Q\left(\frac{C-\theta^*}{\sigma/\sqrt{N_{\max}}}\right) \le \alpha\right\} = C - \frac{x_0\sigma}{\sqrt{N_{\max}}}. \tag{4}$$

Consequently, the identification decision set is $M = (-\infty, C^*)$. Here the open set is used for conciseness in the subsequent derivations.

Theorem 1: For a given $\theta^* \in M = (-\infty, C^*)$ and the decision error α, the minimum resource N^* is

$$N^* = \frac{x_0^2\sigma^2}{(C-\theta^*)^2}. \tag{5}$$

Proof: Since $\theta^* \in M$, we have $\theta^* < C^*$. From $Q(x_0) = \alpha$, by (3) the minimum positive real number to satisfy the bound α is obtained by

$$N^* = \min\left\{x > 0 : Q\left(\frac{C-\theta^*}{\sigma/\sqrt{x}}\right) \le \alpha\right\} = \frac{x_0^2\sigma^2}{(C-\theta^*)^2},$$

which implies the theorem. ∎

In view of Theorem 1, it can be seen that N^* will go to infinity as $C - \theta^*$ goes to zero. To ensure $N^* \le N_{\max}$, there must exist a gap between C and θ^*. This again explains the reason for two sets S and M. In addition, N^* depends on the particular value of θ^* and $N^* \equiv 1$ for $\theta^* \in (-\infty, C - x_0\sigma]$. This case will not be discussed and only the non-trivial case of $C > \theta^* > C - x_0\sigma$ is covered in the following development.

B. Algorithms and Convergence

In reality, θ^* is unknown. The main question is: Can N^* (from (5)) be asymptotically estimated from observation data? We introduce an adaptive resource assignment algorithm to find N^* in real time. At the kth iteration, θ_k is obtained based on N_k observations in the time interval $[(k-1)T, kT)$. The estimate satisfies

$$\theta_k = \theta^* + e_k \tag{6}$$

with $e_k \sim \mathcal{N}(0, \sigma^2/N_k)$. For convenience of algorithm development, we use $\mu_k = N_k/N_{\max}$ instead of N_k in adaptation. This allows us to develop algorithms for a continuous variable in $(0, 1)$, rather than integers. The actual resource assignment will be $N_k = \lceil \mu_k N_{\max} \rceil$. When, $N_{\max}$ is sufficiently large, we ignore the quantization error.

Noticing that $1 \le N_k \le N_{\max}$, one can have $\mu_k \in (-\delta + \frac{1}{N_{\max}}, 1 + \delta) := (\underline{\mu}, \overline{\mu})$, where δ can be any one in $(0, \frac{1}{N_{\max}})$. The resource updating algorithm is

$$\mu_{k+1} = \Pi_{\underline{\mu},\overline{\mu}}\left(\mu_k + \tau_k\left(\theta_k - C + \frac{x_0\sigma}{\sqrt{\mu_k N_{\max}}}\right)\right), \tag{7}$$

where $\Pi_{(\underline{\mu},\overline{\mu})}$ is the projection to a fixed point $\mu_0 \in (\underline{\mu}, \overline{\mu})$, i.e., $\Pi_{(\underline{\mu},\overline{\mu})}(x) = x$ if $x \in (\underline{\mu}, \overline{\mu})$, and $\Pi_{(\underline{\mu},\overline{\mu})}(x) = \mu_0$ if $x \notin (\underline{\mu}, \overline{\mu})$ and τ_k is the updating stepsize to be specified later. For convergence analysis, by denoting

$$h(\mu) = \theta^* - C + \frac{x_0\sigma}{\sqrt{\mu N_{\max}}}, \tag{8}$$

the algorithm (7) can be rewritten as

$$\mu_{k+1} = \Pi_{(\underline{\mu},\overline{\mu})}\Big(\mu_k + \tau_k\left(h(\mu_k) + e_k\right)\Big). \tag{9}$$

Thus, the algorithm (7) is a stochastic approximation algorithm [4]. Consider the the ordinary differential equation (ODE)

$$\dot{\mu} = h(\mu). \tag{10}$$

Let μ^* be the stationary point of (10), which is unique and given by

$$\mu^* = \frac{x_0^2\sigma^2}{(C-\theta^*)^2 N_{\max}} = \frac{N^*}{N_{\max}}. \tag{11}$$

Theorem 2: Suppose that the stepsize τ_k satisfies $\tau_k > 0$, $\tau_k \to 0$ as $k \to \infty$, and $\sum_{k=1}^{L}\tau_k \to \infty$ as $L \to \infty$. If $\theta^* \in M$, then μ_k from (7) converges strongly

$$\mu_k \to \mu^* = N^*/N_{\max},\ w.p.1,\ k \to \infty.$$

Furthermore, if N^* is not an integer, then N_k converges strongly to N^*, i.e.,

$$N_k \to \lceil N^* \rceil,\ w.p.1,\ k \to \infty.$$

Proof: We note that if $\mu_k \in (\underline{\mu},\overline{\mu})$, (7) can be expressed as

$$\mu_{k+1} = \mu_k + \tau_k\left(\theta_k - C + \frac{x_0\sigma}{\sqrt{\mu_k N_{\max}}}\right).$$

We define $t_k = \sum_{j=0}^{k-1}\tau_j,\ m(t) = \max\{k : t_k \le t\}$, and the piecewise constant interpolation $\mu^0(t) = \mu_k$ for $t \in [t_k, t_{k+1})$. The time-shifted sequence $\mu^k(t) = \mu^0(t + t_k)$ can be shown to be equicontinuous in the extended sense w.p.1. (see [4, p. 102]). This implies that we can extract a convergent subsequence $\mu^{k_\ell}(\cdot)$. Then the Arzela-Ascoli theorem concludes that $\mu^{k_\ell}(\cdot)$ converges w.p.1 to a function $\mu(\cdot)$ that is the unique solution of (10).

Under the hypothesis, the convergence of μ_k is determined by its limit ODE (10). In addition

$$\left.\frac{\partial h(\mu)}{\partial \mu}\right|_{\mu^*} = -\frac{1}{2}\frac{x_0\sigma}{(\mu^*)^{3/2}\sqrt{N_{\max}}} < 0. \tag{12}$$

This implies that the stationary point θ^* of (10) is locally asymptotically stable. By [4], we have $\mu_k \to \mu^*$ w.p.1 as $k \to \infty$. ■

Remark 2: N^* is obtained by solving an optimization problem of (3), and its explicit expression is given by Theorem 1. But N_k $(= \lceil \mu_k N_{\max} \rceil)$ is derived from the algorithm (7), which is constructed directly by using the information and available measurements, rather than minimizing a cost function or solving an estimation criterion.

This paper adopts a common assumption in system identification that noises are normal distributed. Our results can be potentially extended to cover non-Gaussian noises by the Central Limit Theorem with additional explicit error bounds (Berry-Esseen Theorem). In that case, instead of the precise convergence rates and estimation errors, we can obtain lower and upper bounds on the optimal resource N^* as a function of the decision error bound α.

If the noises are not i.i.d., namely colored noises, since the resource updating algorithms are of stochastic approximation types, their convergence is valid for a much larger class of random processes with appropriate moment conditions, such as martingale difference and ϕ-mixing sequences.

In addition, the condition that N^* is not an integer can hold almost everywhere since

$$\mathcal{L}\left(\{(x_0,\sigma,C,\theta^*) \in \mathbb{R}^4 : N^* \text{ is an integer}\}\right) = 0,$$

where $\mathcal{L}(\cdot)$ means the Lebesgue measure on $\mathbb{R}^4$.

C. Convergence Rate

This subsection provides the convergence rate of the resource allocation algorithm. For that, denote $\widetilde{\mu}_k = \mu_k - \mu^*$, $k = 1, 2, \ldots$, and

$$\beta = \frac{1}{2}\frac{x_0\sigma}{(\mu^*)^{3/2}\sqrt{N_{\max}}}. \tag{13}$$

Lemma 1: Suppose that $\{a_k, k \ge 1\}$ is a sequence of real numbers such that for $k \ge k_0$

$$a_{k+1} = \left(1 - \frac{\lambda_1}{k}\right)a_k + \frac{\lambda_2}{k}v_k, \tag{14}$$

where $\{v_k, \mathcal{G}_k\}$ is a martingale difference sequence (MDS) with $\sup_{k\ge1}\mathrm{E}[|v_{k+1}|^\gamma|\mathcal{G}_k] < \infty$, w.p.1. $\gamma > 2$, $\lambda_1 > 0$ and $\lambda_2 > 0$ are two constants. Then, in the almost sure sense we have

$$a_k = \begin{cases} O\left(\dfrac{1}{k^{\lambda_1}}\right), & \lambda_1 < \frac{1}{2}; \\ O\left(\dfrac{\sqrt{\log k}(\log\log k)^\epsilon}{\sqrt{k}}\right),\ \forall\epsilon > \dfrac{1}{2}, & \lambda_1 = \frac{1}{2}; \\ O\left(\dfrac{(\log k)^\epsilon}{\sqrt{k}}\right), \forall\epsilon > \dfrac{1}{2}, & \lambda_1 > \frac{1}{2}. \end{cases}$$

The rate of strong convergence is given in the following theorem.

Theorem 3: Under the conditions of Theorem 2, if $\tau_k = \tau/k$ and $\tau > (1/2\beta)$, then the algorithm (7) has the convergence rate

$$\widetilde{\mu}_k = O\left(\frac{(\log k)^\epsilon}{\sqrt{k}}\right),\ \text{w.p.1}, \forall\epsilon > \frac{1}{2}.$$

Proof: From Theorem 2, $\mu_k \to \mu^*$ w.p.1 as $k \to \infty$. By (9), there exists k_0 such that $\mu_{k+1} = \mu_k + \tau_k\left(h(\mu_k) + e_k\right)$ for all $k \ge k_0$. From (8) and (12), $h(\mu_k) = -\beta\widetilde{\mu}_k + o(\widetilde{\mu}_k)$, where $o(\widetilde{\mu}_k)/\widetilde{\mu}_k \to 0$, as $\widetilde{\mu}_k \to 0$. As a result, we have

$$\begin{aligned}\widetilde{\mu}_{k+1} &= \widetilde{\mu}_k + \frac{\tau}{k}\left(-\beta\widetilde{\mu}_k + o(\widetilde{\mu}_k) + e_k\right) \\ &= \left(1 - \frac{\tau\beta}{k}(1 + o(\widetilde{\mu}_k)/\widetilde{\mu}_k)\right)\widetilde{\mu}_k + \frac{\tau}{k}e_k,\ k \ge k_0. \end{aligned} \tag{15}$$

In the time interval $[(k-1)T, kT)$, we denote the observation as y_i^k with $y_i^k = \theta^* + d_i^k, i = 1, 2, \ldots, N_k$. By (7), it can be seen that $N_{k+1} = \lceil \mu_{k+1} N_{\max} \rceil \in \mathcal{F}_k$ and

$$\begin{aligned} \mathrm{E}[e_{k+1}|\mathcal{F}_k] &= \mathrm{E}\left[\frac{1}{N_{k+1}} \sum_{i=1}^{N_{k+1}} d_i^{k+1} \middle| \mathcal{F}_k\right] \\ &= \frac{1}{N_{k+1}} \sum_{i=1}^{N_{k+1}} \mathrm{E}[d_i^{k+1}|\mathcal{F}_k] = 0 \end{aligned}$$

where $\mathcal{F}_k = \sigma\left(d_i^j, i \le N_{\max}, j \le k\right)$. Thus, $\{e_k, \mathcal{F}_k\}$ is an MDS. From Lemma 5 in Appendix, we have

$$\begin{aligned} \mathrm{E}[|e_{k+1}|^3|\mathcal{F}_k] &\le \mathrm{E}\left(\sum_{i=1}^{N_{\max}} |d_i^{k+1}|\right)^3 \\ &\le N_{\max}^2 \sum_{i=1}^{N_{\max}} \mathrm{E}|d_i^{k+1}|^3 = \frac{4N_{\max}^3\sigma^3}{\sqrt{2\pi}}, \end{aligned}$$

which implies that $\sup_{k\ge1} \mathrm{E}[|e_{k+1}|^3|\mathcal{F}_k] \le 4N_{\max}^3\sigma^3/\sqrt{2\pi} < \infty$. Hence, by Lemma 1 the theorem follows. ■

D. Asymptotic Normality

We now establish asymptotic normality of the estimates μ_k. A lemma is presented first, whose proof is contained in Appendix.

Lemma 2: Consider the sequence $\{a_k, k \ge k_0\}$ in Lemma 1. If $\lambda_1 > 1/2$, the distribution of v_{k+1} conditioned on $\mathcal{G}_k$ is $\mathcal{N}(0, \phi_k^2)$, and there exist two constants $\Phi > 0$ and $\phi > 0$ such that $\sup_{k\ge1} \phi_k^2 \le \Phi < \infty$ and $\lim_{k\to\infty} \phi_k^2 = \phi^2 < \infty$, w.p.1, then we have

$$\sqrt{k} a_k \xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda_2^2\phi^2}{2\lambda_1 - 1}\right), \text{ as } k \to \infty \tag{16}$$

and

$$k\mathrm{E}a_k^2 \to \frac{\lambda_2^2\phi^2}{2\lambda_1 - 1}, \text{ as } k \to \infty. \tag{17}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Theorem 4: Under the conditions of Theorem 3, if $N_k \to N^*$ w.p.1 as $k \to \infty$, then the following centered and scaled sequence of μ_k from the algorithm (7) is asymptotically normal, i.e.

$$\sqrt{k}\left(\mu_k - \mu^*\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau^2\sigma^2}{(2\tau\beta - 1)N^*}\right), \text{ as } k \to \infty.$$

Proof: The distribution of e_{k+1} conditioned on $\mathcal{F}_k$ is $\mathcal{N}(0, \sigma^2/N_{k+1})$, from which it can be seen that $\sup_{k\ge1}(\sigma^2/N_{k+1}) \le \sigma^2$ and $\sigma^2/N_{k+1} \to \sigma^2/N^*$ w.p.1 as $k \to \infty$. In Lemma 2, let $v_k = e_k$, $\lambda_1 = \tau\beta$ and $\lambda_2 = \tau$. By (15) and (16), one obtains the desired result. ■

Corollary 1: Under the conditions of Theorem 4, if $\tau = 1/\beta$, then the limit distribution of $\sqrt{k}\left(\mu_k - \mu^*\right)$ has the minimum variance $\sigma^2/(N^*\beta^2)$.

Proof: Note that

$$\frac{\tau^2\sigma^2}{(2\tau\beta - 1)N^*} = \frac{\sigma^2}{N^*} \frac{1}{\beta^2 - (\frac{1}{\tau} - \beta)^2}. \tag{18}$$

The corollary follows. ■

E. Asymptotic Efficiency

Lemma 3: In every time interval $[(k-1)T, kT)$, let $N_k \equiv N^*$ and $y_1^k, \ldots, y_{N^*}^k$ denote the observations, $k = 1, 2, \ldots$. Then, the Cramér-Rao lower bound for estimating μ^* based on $\{\mathrm{y}_1, \ldots, \mathrm{y}_k\}$ is

$$\sigma_{CR}^2(k) = \frac{1}{k} \cdot \frac{4N_{\max}(\mu^*)^3}{x_0^2 N^*},$$

where $\mathrm{y}_i = \{y_1^i, \ldots, y_{N^*}^i\}$, $i = 1, \ldots, k$.

Proof: By (11), the likelihood function of y_i conditioned on μ^* is given by

$$\begin{aligned} &f\left(y_1^i, \ldots, y_{N^*}^i; \mu^*\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{N^*} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{N^*}(y_j^i - \theta^*)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{N^*} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{N^*}\left(y_j^i - C + \frac{x_0\sigma}{\sqrt{\mu^* N_{\max}}}\right)^2\right\}, \end{aligned}$$

where y_j^i denotes the random variable as well as the value it takes, $j = 1, \ldots, N^*$, $i = 1, \ldots, k$. Consequently, we have

$$\begin{aligned} \ell_i &:= \log f\left(y_1^i, \ldots, y_{N^*}^i; \mu^*\right) \\ &= N^* \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{j=1}^{N^*}\left(y_j^i - C + \frac{x_0\sigma}{\sqrt{\mu^* N_{\max}}}\right)^2 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ell_i}{\partial \mu^*} &= \frac{x_0(\mu^*)^{-3/2}}{2\sigma\sqrt{N_{\max}}} \sum_{j=1}^{N^*}\left(y_j^i - C + \frac{x_0\sigma}{\sqrt{\mu^* N_{\max}}}\right) \\ &= \frac{x_0(\mu^*)^{-3/2}}{2\sigma\sqrt{N_{\max}}} \sum_{i=1}^{N^*}\left(y_j^i - \theta^*\right) \end{aligned}$$

and

$$\mathrm{E}\left(\frac{\partial \ell_i}{\partial \mu^*}\right)^2 = \frac{x_0^2 N^*}{4N_{\max}(\mu^*)^3},$$

which gives the lemma since $\mathrm{y}_1, \ldots, \mathrm{y}_k$ are mutually independent. ■

Theorem 5: Under the condition of Theorem 4, if $\tau = 1/\beta$, then algorithm (7) is asymptotically efficient in the sense that $\lim_{k\to\infty} k\left(E(\mu_k - \mu^*)^2 - \sigma_{CR}^2(k)\right) = 0$.

Proof: By (17) and (18), we know that $k\mathrm{E}(\mu_k - \mu^*)^2 \to \sigma^2/(N^*\beta^2)$, which together with Lemma 3 and (13) gives the theorem. ■

Remark 3: Theorem 5 is just a theoretical result since β is unknown. In practical applications, a possible way may be to use an approximate asymptotically efficient algorithm, which can be obtained by replacing $\tau_k = 1/(k\beta)$ with its estimate $1/(k\beta_k)$ in the algorithm (7), where $\beta_k = \frac{1}{2}\frac{x_0\sigma}{(\mu_k)^{3/2}\sqrt{N_{\max}}}$.

F. Bounded Interval

This subsection discusses the case that the identification set is a bounded interval. System identification aims to determine if the true parameter is in a set $[C_1, C_2]$ with a decision error bound α. Suppose that the maximum data length is $N_{\max}$. To satisfy the required decision confidence level, we designate a subset $M \subset [C_1, C_2]$ such that by using the maximum resource $N_{\max}$, the design confidence can be met. This is calculated as follows.

In order to calculate conveniently, we use the symmetric confidence level $\mathcal{Q}(\cdot)$ defined by

$$\mathcal{Q}(x) = \int_x^{\infty} f(y)dy + \int_{-\infty}^{-x} f(y)dy = 2\int_x^{\infty} f(y)dy.$$

This provides a systematic characterization of optimal resource assignment for different M. For a given decision error bound $0 < \alpha < 1$, let s_0 satisfy $\mathcal{Q}(s_0) = \alpha$. Under the maximum resource assignment, the estimation error $e_k \sim \mathcal{N}(0, \sigma^2/N_{\max})$. The maximum and minimum θ^* that can ensure the decision error α can be calculated as

$$\begin{aligned} C_1^* &= \max\left\{\theta^* : \mathcal{Q}\left(\frac{\theta^* - C_1}{\sigma/\sqrt{N_{\max}}}\right) \le \alpha\right\} = C_1 + \frac{s_0\sigma}{\sqrt{N_{\max}}}; \\ C_2^* &= \max\left\{\theta^* : \mathcal{Q}\left(\frac{C_2 - \theta^*}{\sigma/\sqrt{N_{\max}}}\right) \le \alpha\right\} = C_2 - \frac{s_0\sigma}{\sqrt{N_{\max}}}. \end{aligned} \tag{19}$$

Consequently, the identification decision set is $M = (C_1^*, C_2^*)$.

Theorem 6: For a given $\theta^* \in M$ and the given decision error α, the minimum resource N^*

$$N^* = \max\left\{\frac{s_0^2\sigma^2}{(\theta^* - C_1)^2}, \frac{s_0^2\sigma^2}{(C_2 - \theta^*)^2}\right\}. \tag{20}$$

Proof: For a given $\theta^* \in M$, the minimum resource to satisfy the bound α is obtained by considering its distance to the boundary points

$$\begin{aligned} N^1 &= \min\left\{N : \mathcal{Q}\left(\frac{\theta^* - C_1}{\sigma/\sqrt{N}}\right) \le \alpha\right\} = \frac{s_0^2\sigma^2}{(\theta - C_1)^2}; \\ N^2 &= \min\left\{N : \mathcal{Q}\left(\frac{C_2 - \theta^*}{\sigma/\sqrt{N}}\right) \le \alpha\right\} = \frac{s_0^2\sigma^2}{(C_2 - \theta^*)^2}. \end{aligned}$$

Then, $N^* = \max\{N^1, N^2\}$. ■

By the theorem above and (19), we have $N^* < \max\left\{\frac{s_0^2\sigma^2}{(C_1^* - C_1)^2}, \frac{s_0^2\sigma^2}{(C_2 - C_2^*)^2}\right\} = N_{\max}$ and

$$\begin{aligned} N^* &> \max\left\{\frac{s_0^2\sigma^2}{(C_2^* - C_1)^2}, \frac{s_0^2\sigma^2}{(C_2 - C_1^*)^2}\right\} \\ &= \frac{s_0^2\sigma^2}{(C_2 - C_1 - \frac{s_0\sigma}{\sqrt{N_{\max}}})^2} := \underline{\mu}. \end{aligned}$$

The resource updating algorithm is

$$\mu_{k+1}^{(1)} = \Pi_{(\underline{\mu}, 1)}\left(\mu_k^{(1)} + \tau_k\left(C_1 - \theta_k + \frac{s_0\sigma}{\sqrt{\mu_k^{(1)} N_{\max}}}\right)\right) \tag{21}$$

$$\mu_{k+1}^{(2)} = \Pi_{(\underline{\mu}, 1)}\left(\mu_k^{(2)} + \tau_k\left(\theta_k - C_2 + \frac{s_0\sigma}{\sqrt{\mu_k^{(2)} N_{\max}}}\right)\right) \tag{22}$$

$$\mu_{k+1} = \max\left\{\mu_{k+1}^{(1)}, \mu_{k+1}^{(2)}\right\}. \tag{23}$$

The actual resource assignment is $\lceil \mu_k N_{\max} \rceil$ in the kth interval.

Theorem 7: Suppose that the step size τ_k satisfies $\tau_k > 0$, $\tau_k \to 0$ as $k \to \infty$, and $\sum_{k=1}^{L} \tau_k \to \infty$ as $L \to \infty$. If $\theta^* \in M$, then μ_k from (21)–(23) follows

$$\mu_k \to \mu^* = N^*/N_{\max}, \text{ w.p.1 as } k \to \infty.$$

Proof: By use of the method in Theorem 2, we have

$$\mu_k^{(i)} \to \mu_{(i)}^* := \frac{s_0^2\sigma^2}{(\theta^* - C_i)^2 N_{\max}} \text{ w.p.1 as } k \to \infty, \ i = 1, 2, \tag{24}$$

which together with (20) implies

$$\mu_k \to \max\left\{\frac{s_0^2\sigma^2}{(\theta^* - C_1)^2 N_{\max}}, \frac{s_0^2\sigma^2}{(C_2 - \theta^*)^2 N_{\max}}\right\} = \frac{N^*}{N_{\max}}.$$

Thus, the assertion is proved. ■

Theorem 8: Under the conditions of Theorem 7, if $\tau_k = \tau/k$ and $\tau > 1/(2\overline{\beta})$, then the algorithm (21)–(23) has the convergence rate

$$\mu_k - \mu^* = O\frac{(\log k)^{\epsilon}}{\sqrt{k}} \text{ w.p.1}, \forall \epsilon > \frac{1}{2},$$

where $\overline{\beta} = \min_{i=1,2}\{\frac{1}{2}\frac{s_0\sigma}{(\mu_i^*)^{3/2}\sqrt{N_{\max}}}\}$ and $\mu_{(i)}^*$ is the one in (24) for $i = 1, 2$.

Proof: According to Theorem 3, it is known that if $\tau > (1/\beta_i)$, we have

$$\mu_k^{(i)} - \mu_{(i)}^* = O\left(\frac{(\log k)^{\epsilon}}{\sqrt{k}}\right) \text{ w.p.1}, \forall \epsilon > \frac{1}{2}, \ i = 1, 2,$$

which gives the theorem. ■

G. Examples

Example 1: System identification aims to determine if the true parameter θ^* is in the set $(-\infty, C]$ with a decision error bound $\alpha = 1\%$ and $C = 9$. The measurement noise $d_j \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 2$, $j = 1, 2, \ldots$ It can be verified that $x_0 = 2.3263$. The maximum data size is $N_{\max} = 100$ and we can get $C^* = 8.5347$ by (4).

If $\theta^* = 8.5$, then $N^*(8.5) = 86.5867$ by (5), and the estimation error $e_{N^*(8.5)} \sim \mathcal{N}(0, 0.2149^2)$ can ensure the decision error bound 1%. If $\theta^* = 8$, then $N^*(8) = 21.6467$ and $e_{N^*(8)} \sim \mathcal{N}(0, 0.4299^2)$ can ensure the decision error bound 1%. Their comparison is shown by Fig. 2, which gives the density functions of the estimate.

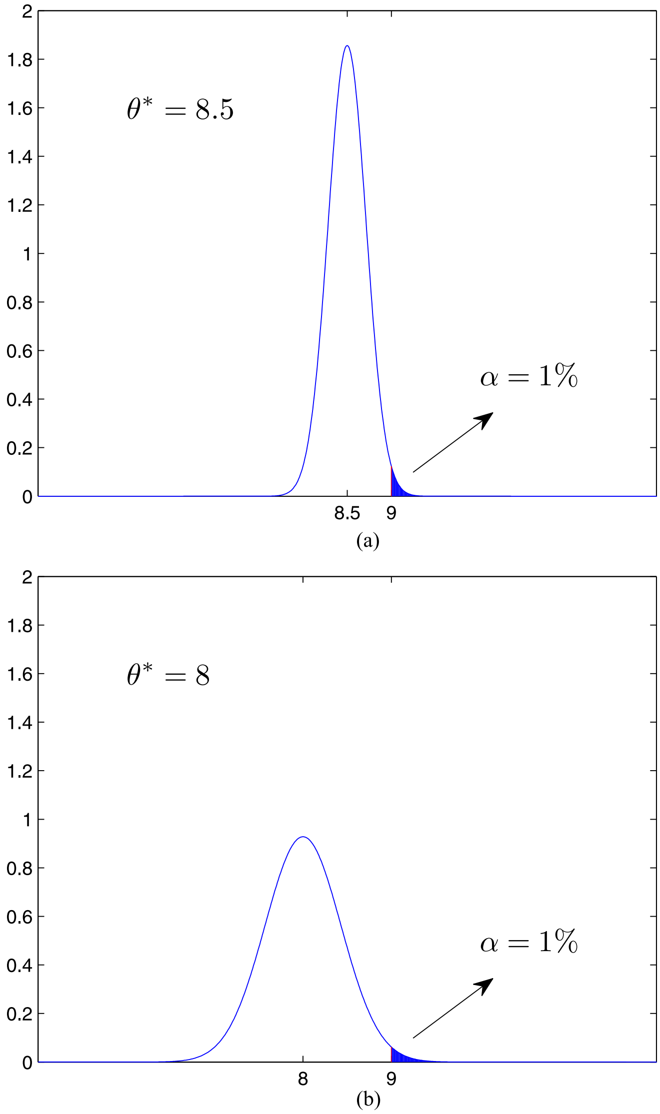


Fig. 2. A comparison of the minimum resource for two different real parameters. (a) $\theta^* = 8.5$, (b) $\theta^* = 8$.

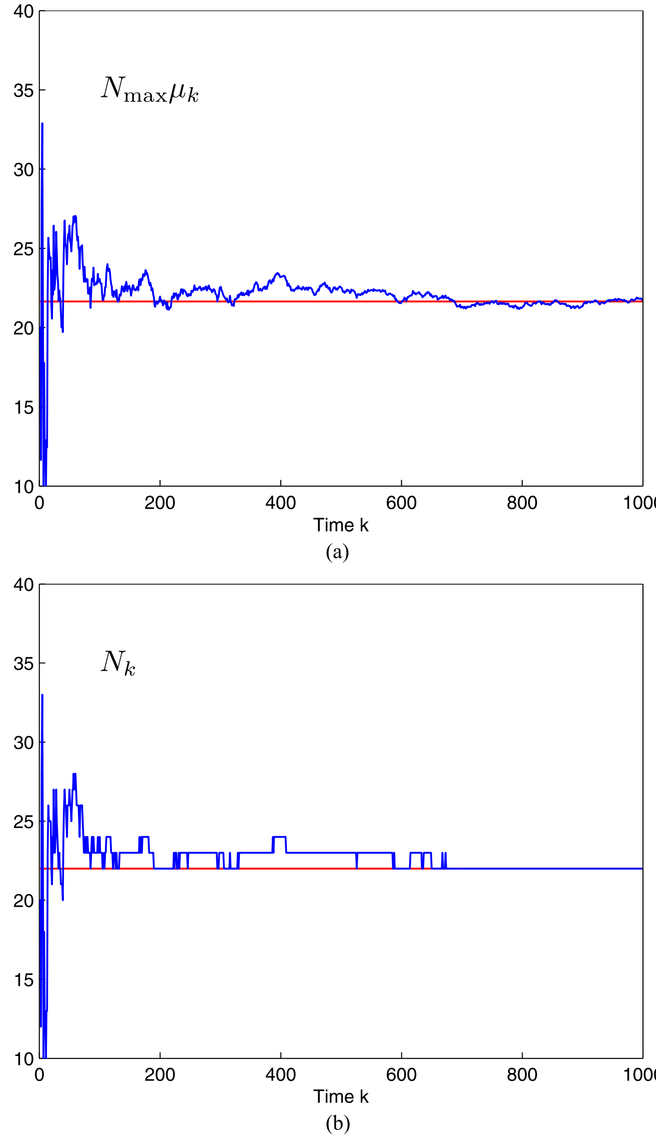


Fig. 3. Convergence of the resource updating algorithm. (a) Convergence of $N_{\max}\mu_k$, (b) Convergence of N_k.

If the real unknown parameter $\theta^* = 8$, then we have $\mu^* = N^*(8)/N_{\max} = 0.216467$, $\beta = 2.3098$ by (13). Algorithm (7) is simulated with $\delta = 10^{-3}$, $\tau_k = 1/k$ and the initial value $\mu_0 = 0.2$. Fig. 3 demonstrates the convergence of $N_{\max}\mu_k$ and N_k by their trajectories. Furthermore, it is known that $\mu_k - \mu^* = O\left((\log k)^\epsilon/\sqrt{k}\right)$, w.p.1., $\forall \epsilon > \frac{1}{2}$ by Theorem 3 and $\tau = 1 > 1/\beta = 0.4329$, which is shown in Fig. 4 with $\epsilon = 1$.

To illustrate the asymptotic property of $E(\mu_k - \mu^*)^2$, we employ the average of 100 trajectories of $N_{\max}(\mu_k - \mu^*)^2$ by Monte Carlo simulation, which is shown by Fig. 5, where τ takes three difference values $0.4329 = 1/\beta$, 0.3, and 2, respectively. By $F(0.4329) < F(0.3) < F(2)$ with $F(\tau) = \frac{\tau^2\sigma^2}{(2\tau\beta-1)N^*}$, we know that the height of the curve corresponds to the size of $F(\tau)$, which illustrates Theorem 4. Especially, we can also see that the curve with $\tau = 1/\beta$ converges to the line $y = 3.41 = N_{\max}k\sigma_{CR}^2(k)$ by Lemma 3, which accords with the result of Theorem 5.

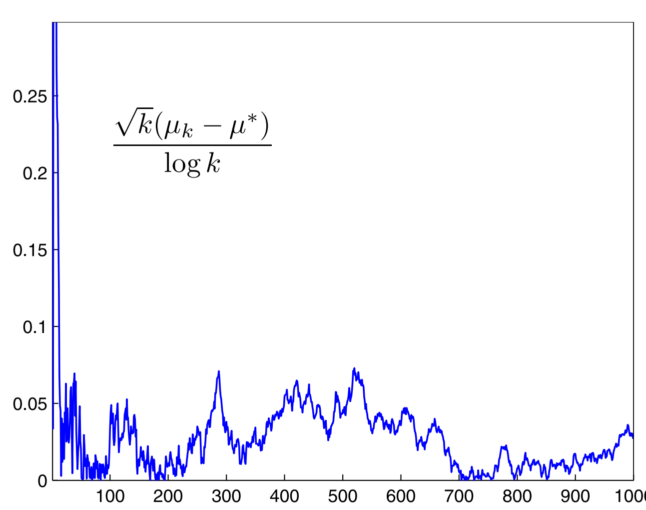


Fig. 4. Convergence rate of the resource updating algorithm.

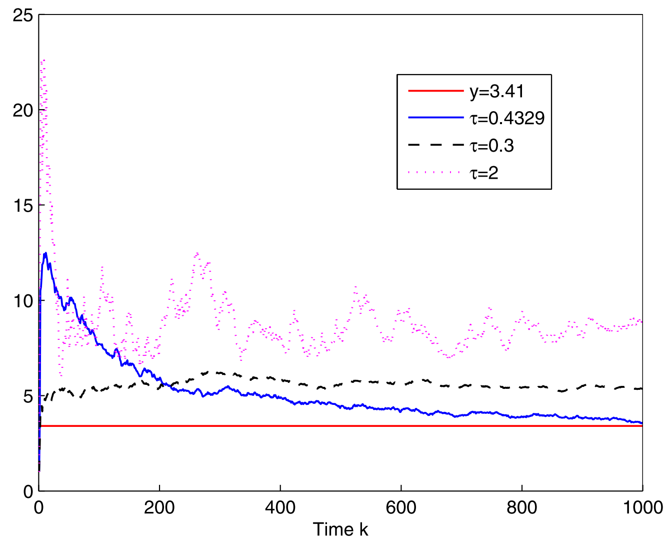


Fig. 5. Asymptotic efficiency of the resource updating algorithm.

Since β is unknown, an asymptotically efficient algorithm cannot be used in practical applications. But an approximate one may be available. It can be obtained by using $1/(k\beta_k)$ to substitute for $\tau_k = 1/(k\beta) = 0.4329/k$ in the algorithm (7), where $\beta_k = \frac{1}{2}\frac{x_0\sigma}{(\mu_k)^{3/2}\sqrt{N_{\max}}}$. Its convergence and asymptotic efficiency are illustrated by Fig. 6.

Example 2: We consider a batched resource allocation problem of 150 subjects. Suppose that θ^* (when it is in the decision set) takes 150 possible values $1, 2, \ldots, 150$ and is uniformly distributed. For each given θ^*, system identification aims to determine if it is in the set $(-\infty, C]$ with a decision error bound $\alpha = 1\%$, where $C = 151$. The measurement noise $d_j \sim \mathcal{N}\left(0, \sigma^2\right)$ with $\sigma = 5$, $j = 1, 2, \ldots$

When an off-line, population-based, and worst-case strategy is used, the total required resource is

$$150 \times \left\lceil \frac{x_0^2\sigma^2}{(C-150)^2} \right\rceil = 20400.$$

However, when our individualized and dynamic resource allocation strategy is employed, the total resource in the average sense is

$$150 \times EN^* = 150 \times \left\lceil \frac{1}{150}\sum_{i=1}^{150}\frac{x_0^2\sigma^2}{(C-i)^2} \right\rceil = 300,$$

representing a drastic reduction on resource consumption.

IV. Multivariate Systems

In this section, we extend the results of scalar cases to systems with multiple variables. The noise-corrupted measurements of $\theta^* = (\theta_1^*, \ldots, \theta_n^*)' \in \mathbb{R}^n$ are obtained

$$y_j = \theta^* + d_j,\ j = 1, 2, \ldots$$

where $\{d_j\}$ is a sequence of i.i.d. random vectors with Gaussian distribution, mean zero, and nonsingular covariance matrix Σ. Given the data size N_k in the kth interval, the parameter estimate is

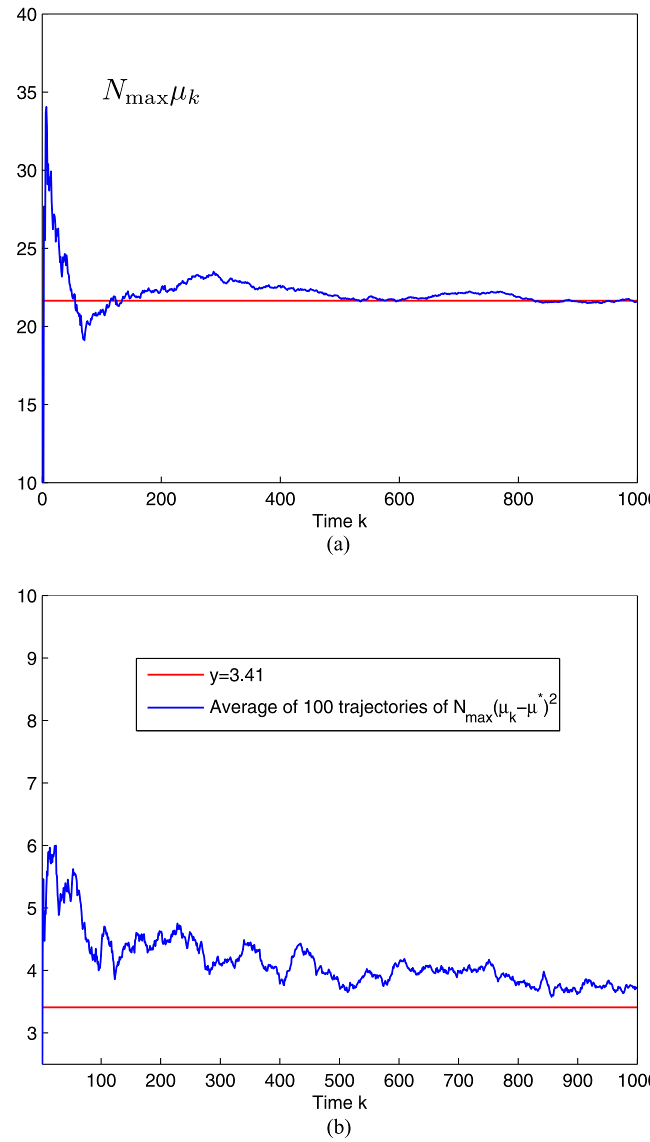


Fig. 6. Performance of the approximate asymptotically efficient algorithm. (a) Convergence, (b) Asymptotic efficiency.

$$\theta_k = (\theta_{k,1}, \ldots, \theta_{k,n})' = \frac{1}{N_k}\sum_{j=1}^{N_k} y_j^k = \theta^* + e_k$$

where e_k is the estimation error and $e_k \sim \mathcal{N}(0, \Sigma/N_k)$.

To motivate this problem in a system setting, consider the FIR system (1). Suppose that the input probing signal is selected to be n-periodic with its one-period values $u_0 = p_1,\ u_1 = p_2, \ldots, u_{n-1} = p_n$. For $l = 1, 2, \ldots$, define $Y_l^0 = (y_{ln-1}, \ldots, y_{(l+1)n-2})'$, $\theta = (a_0, \ldots, a_{n-1})'$, $D_l^0 = (d_{ln-1}, \ldots, d_{(l+1)n-2})'$. It follows that $Y_l^0 = \Phi\theta + D_l^0$, where Φ is the circulant matrix generated by $\{p_n, \ldots, p_1\}$. If Φ is full rank, then we have

$$Y_l = \theta + D_l,\ l = 1, 2, \ldots,$$

with $Y_l = \Phi^{-1}Y_l^0$ and $D_l = \Phi^{-1}D_l^0$. If Φ is a tall matrix and column full rank, then we can let $Y_l = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y_l^0$ and

$D_l = \left(\Phi^T\Phi\right)^{-1}\Phi^T D_l^0$, which also gives $Y_l = \theta + D_l$. This makes the input information to be involved in the covariance matrix of D_l, which indicates that the optimal input design problem can be investigated based on an asymptotically efficient resource updating algorithm.

A. Optimal Resource Allocation

In a decision-based identification, the goal is to determine if θ^* is in a set $(-\infty, C) = \{x \in \mathbb{R}^n : x < C\}$ with the decision error bound $\alpha = (\alpha_1, \ldots, \alpha_n)' \in \mathbb{R}^n$, where $C = (c_1, \ldots, c_n)' \in \mathbb{R}^n$ is a given vector and $0 < \alpha = (\alpha_1, \ldots, \alpha_n)' < \mathbb{1}/2 = (1, 1, \ldots 1)'/2 = (1/2, 1/2, \ldots, 1/2)' \in \mathbb{R}^n$. And, the decision reliability requirement is that when $\theta^* < C^*$,

$$\Pr(\theta_{k,i} \le c_i) \ge 1 - \alpha_i,\ i = 1, 2, \ldots, n, \tag{25}$$

where $C^* < C$ will be given by (26). Here and henceforth, one writes $y = (y_1, \ldots, y_n)' \le z = (z_1, \ldots, z_n)'$ if and only if $y_i \le z_i, i = 1, \ldots, n$. So does '<'. Also, denote $\frac{1}{\sqrt{y}} = (\frac{1}{\sqrt{y_1}}, \ldots, \frac{1}{\sqrt{y_n}})'$.

Under the maximum resource assignment, $e_k \sim \mathcal{N}(0, \Sigma/N_{\max})$ and $\theta_{k,i} \sim \mathcal{N}(\theta_i^*, \sigma_{i,i}^2/N_{\max})$, where $\sigma_{i,i}^2$ is the ith diagonal element of Σ, $i = 1, 2, \ldots, n$. Similar to the scalar case, the maximum C^* that can ensure the decision reliability $\mathbb{1} - \alpha$ can be achieved by

$$C^* = C - \frac{1}{\sqrt{N_{\max}}}\begin{pmatrix} x_1\sigma_{1,1} \\ \vdots \\ x_n\sigma_{n,n}\end{pmatrix} \tag{26}$$

where x_i is given by $Q(x_i) = \alpha_i, i = 1, \ldots, n$. Then, the identification decision set is $M = (-\infty, C^*)$.

Theorem 9: For $\theta^* \in M$ and the decision reliability requirement (25), the minimum resource N^* for the given decision error α is $\lceil \|N_v^*\| \rceil$, where

$$N_v^* = \left(\frac{x_1^2\sigma_{1,1}^2}{(c_1 - \theta_1^*)^2}, \ldots, \frac{x_n^2\sigma_{n,n}^2}{(c_n - \theta_n^*)^2}\right)' \in \mathbb{R}^n, \tag{27}$$

and $\|\cdot\|$ means the l_∞ norm of a vector.

Proof: According to the decision requirement (25), it is known that

$$N_v^* = \max_{1\le i\le n}\ \min_{N \le N_{\max}} \left\{N : Q\left(\frac{c_i - \theta_i^*}{\sigma_{i,i}/\sqrt{N}}\right) \le \alpha_i\right\}$$

which yields the theorem. ■

B. Algorithms and Convergence

For the same reason in Section III, only the case of $\theta^* > (c_1 - x_1\sigma_{1,1}, \ldots, c_n - x_n\sigma_{n,n})'$ is considered in this section. Hence, it follows that $\mathbb{1} \le N_v^* \le N_{\max}\mathbb{1}$, which implies that

$$N_v^*/N_{\max} \in \left((\frac{1}{N_{\max}} - \delta_v)\mathbb{1}, (1 + \delta_v)\mathbb{1}\right) := (\underline{\mu}, \overline{\mu}),$$

where δ_v can be any one in $(0, \frac{1}{N_{\max}})$.

The resource updating algorithm is

$$\mu_{k+1} = \Pi_{(\underline{\mu}, \overline{\mu})}\left(\mu_k + \tau_k\Gamma(\theta_k - C + \nu_k)\right) \tag{28}$$

$$\nu_k = \frac{\mathrm{diag}(x_1\sigma_{1,1}, \ldots, x_n\sigma_{n,n})}{\sqrt{N_{\max}}}\frac{1}{\sqrt{\mu_k}}, \tag{29}$$

where Γ is a given matrix to be used for adjusting the asymptotic property, and $\Pi_{(\underline{\mu}, \overline{\mu})}(\cdot)$ defined on $\mathbb{R}^n$ is the projection to a fixed point $(\mu_{0,1}, \ldots, \mu_{0,n})' \in (\underline{\mu}, \overline{\mu})$, given by $\Pi_{(\underline{\mu}, \overline{\mu})}((y_1, \ldots, y_n)') = (z_1, \ldots, z_n)'$ with $z_i = y_i$ if $y_i \in (\underline{\mu}_i, \overline{\mu}_i)$ and $z_i = \mu_{0,i}$ if $y_i \notin (\underline{\mu}_i, \overline{\mu}_i)$, $i = 1, \ldots, n$. In the kth interval, the actual resource assignment will be $N_k = \lceil \|\mu_k\| N_{\max} \rceil$.

Theorem 10: Suppose that Γ is an identity matrix, and the step size τ_k satisfies $\tau_k > 0$, $\tau_k \to 0$ as $k \to \infty$, and $\sum_{k=1}^L \tau_k \to \infty$ as $L \to \infty$. If $\theta^* \in M$, then μ_k from (28)–(29) follows:

$$\mu_k \to \mu^* = (\mu_1^*, \ldots, \mu_n^*)' = N_v^*/N_{\max},\ \text{w.p.1 as } k \to \infty,$$

where N_v^* is the vector given by (27).

Proof: Similar to the proof of Theorem 2, the corresponding ODE equation of (28)–(29) is

$$\dot{\mu} = \theta^* - C + \frac{\mathrm{diag}(x_1\sigma_{1,1}, \ldots, x_n\sigma_{n,n})}{\sqrt{N_{\max}}}\frac{1}{\sqrt{\mu}}. \tag{30}$$

By letting the right side of the above equal to zero and verifying the negative definiteness of

$$\frac{\partial}{\partial\mu}\left(\frac{\mathrm{diag}(x_1\sigma_{1,1}, \ldots, x_n\sigma_{n,n})}{\sqrt{N_{\max}}}\frac{1}{\sqrt{\mu}}\right)\Bigg|_{\mu^*}$$
$$= -\frac{1}{2}\mathrm{diag}\left(\frac{x_1\sigma_{1,1}}{(\mu_1^*)^{3/2}\sqrt{N_{\max}}}, \ldots, \frac{x_1\sigma_{n,n}}{(\mu_n^*)^{3/2}\sqrt{N_{\max}}}\right),$$

the theorem is proved. ■

Theorem 11: Under the conditions of Theorem 10, if $\tau_k = \tau/k$ and $\tau > (1/2\overline{\beta})$, then the algorithm (28)–(29) has the convergence rate

$$\|\mu_k - \mu^*\| = O\left(\frac{(\log k)^\epsilon}{\sqrt{k}}\right)\ \text{w.p.1}, \forall\epsilon > \frac{1}{2}$$

where $\tau > 0$ is a constant and $\overline{\beta} = \min_{1\le i\le n}\beta_i$ with $\beta_i = \dfrac{1}{2}\dfrac{x_i\sigma_{i,i}}{(\mu_i^*)^{3/2}\sqrt{N_{\max}}}$.

Proof: With $\mu_k = (\mu_{k,1}, \ldots, \mu_{k,n})'$, (28)–(29) can be rewritten as the component form

$$\mu_{k+1,i} = \Pi_{(\underline{\mu}_i, \overline{\mu}_i)}\left(\mu_{k,i} + \tau_k\left(\theta_{k,i} - c_i + \frac{x_i\sigma_{i,i}}{\sqrt{\mu_{k,i}N_{\max}}}\right)\right)$$

for $i = 1, \ldots, n$. Noticing that $\theta_{k,i} = \theta_i^* + e_{k,i}$ with $e_{k,i} \sim \mathcal{N}(0, \sigma_{i,i}^2/N_k)$, by Theorem 3 it can be seen that if $\tau > (1/2\beta_i)$, we have

$$\mu_{k,i} - \mu_i^* = O\left(\frac{(\log k)^\epsilon}{\sqrt{k}}\right)\ \text{w.p.1}, \forall\epsilon > \frac{1}{2},\ i = 1, \ldots, n.$$

Hence, the theorem is true. ■

Theorem 12: If $\theta^* \in M$, $\tau_k = 1/k$, $\Gamma = \text{diag}(\tau_1, \ldots, \tau_n)$ with $\tau_i > (1/2\beta_i)$ for $1 \le i \le n$, and $N_k \to N^*$ w.p.1 as $k \to \infty$, then the centered and scaled sequence of the estimation error from algorithm (28)–(29) is asymptotically normal, i.e.

$$\sqrt{k}(\mu_k - \mu^*) \xrightarrow{d} \mathcal{N}(0, \ \Sigma \circ S/N^*) \text{ as } k \to \infty,$$

where $S = \left(\frac{\tau_i \tau_j}{\tau_i \beta_i + \tau_j \beta_j - 1}\right)_{n \times n}$ and $\circ$ means the Hadamard product of matrices (component-wise multiplication).

Proof: Under the condition of the theorem, we know that Theorem 10 is true. Thus, there exists k_0 such that (28)–(29) can be rewritten as

$$\widetilde{\mu}_{k+1} = \left(I - \left(\frac{1}{k}B + o(\widetilde{\mu}_k)/\widetilde{\mu}_k\right)\right)\widetilde{\mu}_k + \frac{\Gamma}{k}e_k, \ k \ge k_0,$$

where $B := \text{diag}(\tau_1\beta_1, \ldots, \tau_n\beta_n)$. Noticing $\mathrm{E}e_k e_k' \to \Sigma/N^*$ and similar to Theorem 4, one can prove the theorem by [35, Lemma 3.3.1]. ■

C. Asymptotic Efficiency

Lemma 4: In every time interval $[(k-1)T, kT)$, let $N_k \equiv N^*$ and $y_1^k, \ldots, y_{N^*}^k$ denote the observations, $k = 1, 2, \ldots$ Then, the Cramér-Rao lower bound for estimating μ^* based on $\{\mathrm{y}_1, \ldots, \mathrm{y}_k\}$ is

$$\Sigma_{CR}(k) = \frac{1}{k}\frac{S^*\Sigma S^*}{N^*},$$

where $S^* = \text{diag}\left(\frac{1}{\beta_1}, \ldots, \frac{1}{\beta_n}\right)$ with β_i being the ones in Theorem 11, and $\mathrm{y}_i = \{y_1^i, \ldots, y_{N^*}^i\}$, $i = 1, \ldots, k$.

Proof: For $i = 1, \ldots, k$, the likelihood function of y_i conditioned on μ^* is given by

$$f(\mathrm{y}_i; \mu^*) = \left(\frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}}\right)^{N^*} \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{N^*}(y_j^i - \theta^*)'\Sigma^{-1}(y_j^i - \theta^*)\right\},$$

where $\theta^* = \theta^*(\mu^*) = C - \frac{\text{diag}(x_1\sigma_1, \ldots, x_n\sigma_n)}{\sqrt{N_{\max}}}\frac{1}{\sqrt{\mu^*}}$ by (30). It follows that:

$$\ell := \log f(\mathrm{y}_i; \mu^*) = N^* \log \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} - \frac{1}{2}\sum_{j=1}^{N^*}(y_j^i - \theta^*)'\Sigma^{-1}(y_j^i - \theta^*),$$

and $\frac{\partial \ell}{\partial \mu^*} = \sum_{j=1}^{N^*}(y_j^i - \theta^*)'\Sigma^{-1}\frac{\partial \theta^*}{\partial \mu^*}$. Furthermore, we have

$$\mathrm{E}\left(\frac{\partial \ell}{\partial \mu^*}\right)'\left(\frac{\partial \ell}{\partial \mu^*}\right) = \sum_{j=1}^{N^*}\mathrm{E}\left(\frac{\partial \theta^*}{\partial \mu^*}\Sigma^{-1}(y_j^i - \theta^*)(y_j^i - \theta^*)'\Sigma^{-1}\frac{\partial \theta^*}{\partial \mu^*}\right) = N^*\frac{\partial \theta^*}{\partial \mu^*}\Sigma^{-1}\frac{\partial \theta^*}{\partial \mu^*}$$

which together with $\frac{\partial \theta^*}{\partial \mu^*} = \text{diag}(\beta_1, \ldots, \beta_n)$ and the independence of $\mathrm{y}_1, \ldots, \mathrm{y}_n$ implies the lemma. ■

Theorem 13: Under the conditions of Theorem 12, if $\Gamma = S^*$, then the algorithm (28)–(29) is asymptotically efficient in the sense that

$$\lim_{k \to \infty} k\left(E(\mu_k - \mu^*)(\mu_k - \mu^*)' - \Sigma_{CR}(k)\right) = 0.$$

Proof: If $\Gamma = S^*$ which indicates $\tau_i = 1/\beta_i$ for $i = 1, \ldots, n$, then we have $S = (\frac{1}{\beta_i}\frac{1}{\beta_j})_{n \times n}$. It follows that $\Sigma \circ S/N^* = S^*\Sigma S^*/N^*$. By Lemma 4 and Theorem 12, we obtain the assertion of the theorem. ■

V. Conclusion

In this age of information explosion, system identification consumes precious resources and carries a price. This new trend demands a revisit of the traditional identification paradigm, which has been focused on accuracy and convergence. Recent advances in system identification under sampling and quantization constraints have laid a necessary foundation for development of a complexity-based identification paradigm. However, the state-of-the-art in system identification remains in their infancy in dealing with this new reality of complexity-based methodologies.

By considering complexity as a fundamental constraint and a design variable, this paper introduces the concept of decision-based identification in which the goal of identification is to achieve required reliability with minimum resource consumption. As a new direction, there are many potential open issues along this line of research such as different types of systems, noise characterizations, decision sets, and uncertainties.

This paper presents technical results on gain systems. By using full rank periodic inputs, identification of FIR and ARMAX models can be reduced equivalently to a set of identification problems for gain systems, see our previous work [28] on this approach. Extensions to general systems under general inputs are open problems and worth investigation.

Appendix

Lemma 5 ([31, pp. 132]): If $b_i \ge 0$, then the following equation hold,

$$\left(\sum_{i=1}^n b_i\right)^r \le \begin{cases} n^{r-1}\sum_{i=1}^n b_i^r, & r \ge 1, \\ \sum_{i=1}^n b_i^r, & 0 \le r \le 1. \end{cases}$$

Lemma 6 ([32]): If the positive real number sequences $\{\tau_i^1, i \ge 1\}$ and $\{\tau_i^2, i \ge 1\}$ satisfy $\sum_{i=1}^\infty \tau_i^1 = \infty$, $\sum_{i=1}^\infty \tau_i^2 = \infty$, and $\tau_i^1 \simeq \tau_i^2$, then $\sum_{i=1}^k \tau_i^1 \simeq \sum_{i=1}^k \tau_i^2$ as $k \to \infty$, where "$\tau_i^1 \simeq \tau_i^2$" means that "$\lim_{i\to\infty} \tau_i^1/\tau_i^2 = 1$".

Lemma 7 ([33]): For the MDS $\{v_k, \mathcal{G}_k\}$ given by Lemma 1 and an adapted process $\{w_k, \mathcal{G}_k\}$, we have

$$\sum_{i=1}^k w_i v_{i+1} = O\left(W_k(\log W_k)^\epsilon\right), \text{ w.p.1., } \forall \epsilon > \frac{1}{2}.$$

with $W_k = \left(\sum_{i=1}^k w_i^2\right)^{1/2}$.

Lemma 8 ([34]): Consider an MDS $\{\xi_i, \mathcal{G}_i, i \geq 1\}$ and a double subscript real number sequence $\{r_{ki} : 1 \leq i \leq k\}$. If $\mathrm{E}\xi_i^2 < \infty$, $\mathrm{E}[\xi_i^2|\mathcal{G}_{i-1}] = \rho_i^2$ w.p.1 for $i \geq 1$,

$$\lim_{b\to\infty} \sup_{i\geq 1} \mathrm{E}[\xi_i^2 I_{|\xi_i|>b}|\mathcal{G}_{i-1}] = 0 \text{ w.p.1}, \tag{A.1}$$

$$\lim_{k\to\infty} \sum_{i=1}^{k} r_{ki}^2 \rho_i^2 = 1, \tag{A.2}$$

$$\sup_{k\geq 1} \sum_{i=1}^{k} r_{ki}^2 < \infty \text{ and } \lim_{k\to\infty} \max_{1\leq i\leq k} |r_{ki}| = 0, \tag{A.3}$$

then $\sum_{i=1}^k r_{ki}\xi_i \xrightarrow{d} \mathcal{N}(0,1)$.

Proof of Lemma 1: By (14), it can be verified that

$$\begin{aligned} a_{k+1} &= \prod_{i=k_0}^{k}\left(1-\frac{\lambda_1}{i}\right)a_{k_0} + \sum_{i=k_0}^{k}\prod_{j=i+1}^{k}\left(1-\frac{\lambda_1}{j}\right)\frac{\lambda_2}{i}v_i \\ &:= I_{1k} + I_{2k}. \end{aligned}$$

Noticing that $\prod_{j=i}^k \left(1-\frac{\lambda_1}{j}\right) = O\left(\left(\frac{i}{k}\right)^{\lambda_1}\right)$, we have

$$I_{1k} = O\left(k^{-\lambda_1}\right), \tag{A.4}$$

and

$$I_{2k} = O\left(k^{-\lambda_1}\sum_{i=k_0}^{k} i^{\lambda_1-1}v_i\right). \tag{A.5}$$

According to Lemma 7, one can get

$$\begin{aligned} &\sum_{i=k_0}^{k} i^{\lambda_1-1}v_i \\ &= \begin{cases} O(1), & \lambda_1 < 1/2; \\ O\left(\sqrt{\log k}(\log\log k)^{\epsilon}\right), \text{ w.p.1., } \forall\epsilon > \frac{1}{2}, & \lambda_1 = 1/2; \\ O\left(k^{\lambda_1-\frac{1}{2}}(\log k)^{\epsilon}\right), \text{ w.p.1., } \forall\epsilon > \frac{1}{2}, & \lambda_1 > 1/2. \end{cases} \end{aligned}$$

This, together with (A.4) and (A.5), proves the lemma. ■

Proof of Lemma 2: In view of (14), we have

$$\begin{aligned} &\sqrt{k+1}a_{k+1} \\ &= \left(1-\frac{\lambda_1}{k}\right)\sqrt{\frac{k+1}{k}}\sqrt{k}a_k + \frac{\lambda_2\sqrt{k+1}}{k}v_k \\ &= \prod_{i=k_0}^{k}\left(1-\frac{\lambda_1}{i}\right)\sqrt{\frac{i+1}{i}}\sqrt{k_0}a_{k_0} \\ &\quad +\lambda_2\sum_{i=k_0}^{k}\prod_{j=i+1}^{k}\left(1-\frac{\lambda_1}{j}\right)\sqrt{\frac{j+1}{j}}\frac{\sqrt{i+1}}{i}v_i \\ &:= S_{1k} + \frac{\lambda_2\phi}{\sqrt{2\lambda_1-1}}S_{2k}, \end{aligned} \tag{A.6}$$

i.e., $S_{2k} = \sum_{i=k_0}^k r_{ki}v_i$ with $r_{ki} = \frac{\sqrt{2\lambda_1-1}}{\phi}\prod_{j=i+1}^k\left(1-\frac{\lambda_1}{j}\right)\sqrt{\frac{j+1}{j}}\frac{\sqrt{i+1}}{i}$.

Using $\sum_{j=1}^k \frac{1}{j} = \log k + \gamma + O(1/k)$ with γ being the Euler constant, one can get

$$\begin{aligned} &\prod_{j=i}^{k}\left(1-\frac{\lambda_1}{j}\right)\sqrt{\frac{j+1}{j}} \\ &= \exp\left\{\sum_{j=i}^{k}\left[\log\left(1-\frac{\lambda_1}{j}\right)+\frac{1}{2}\log\left(1+\frac{1}{j}\right)\right]\right\} \\ &= \exp\left\{\left(\frac{1}{2}-\lambda_1\right)\sum_{j=i}^{k}\frac{1}{j} + O\left(\sum_{j=i}^{k}\frac{1}{j^2}\right)\right\} \\ &= \exp\left\{\left(\frac{1}{2}-\lambda_1\right)(\log k - \log i) + O\left(\frac{1}{k}\right) + O\left(\frac{1}{i}\right)\right\} \\ &= \exp\{O(1/k)+O(1/i)\}\left(\frac{k}{i}\right)^{\frac{1}{2}-\lambda_1}. \end{aligned}$$

Then, it follows that

$$S_{1k} = O\left(k^{\frac{1}{2}-\lambda_1}\right) \tag{A.7}$$

and

$$r_{ki} = \frac{\sqrt{2\lambda_1-1}}{\phi}k^{\frac{1}{2}-\lambda_1}\exp\{O(1/k)+O(1/i)\}\frac{(i+1)^{\lambda_1}}{i}. \tag{A.8}$$

By $\lambda_1 > 1/2$ and (A.7), we have $S_{1k} \to 0$ as $k \to \infty$. Thus, to prove (16) it is sufficient to show $S_{2k} \to \mathcal{N}(0,1)$, which will be established by Lemma 8 with the following steps.

Since the conditional distribution of v_k is $\mathcal{N}(0, \phi_k^2)$ and $\sup_{k\geq 1}\phi_k^2 \leq \Phi < \infty$, it can be verified that

$$\lim_{b\to\infty}\sup_{i\geq 1}\mathrm{E}[v_i^2 I_{|v_i|>b}|\mathcal{G}_{i-1}] = 0, \text{ w.p.1.}$$

According to Lemma 6 and (A.8), we have

$$\begin{aligned} &\sum_{i=k_0}^{k} r_{ki}^2\phi_i^2 \\ &= \frac{2\lambda_1-1}{\phi^2}k^{1-2\lambda_1} \\ &\quad \times\sum_{i=k_0}^{k}\left(\exp\{O(1/k)+O(1/i)\}\frac{(i+1)^{\lambda_1}}{i}\right)^2\phi_i^2 \\ &\simeq (2\lambda_1-1)k^{1-2\lambda_1}\sum_{i=k_0}^{k} i^{2\lambda_1-2} \\ &\simeq (2\lambda_1-1)k^{1-2\lambda_1}\cdot\frac{1}{2\lambda_1-1}\left(k^{2\lambda_1-1}-k_0^{2\lambda_1-1}\right) \\ &= 1-\left(\frac{k_0}{k}\right)^{2\lambda_1-1}, \end{aligned}$$

which implies

$$\lim_{k\to\infty}\sum_{i=k_0}^{k} r_{ki}^2\phi_i^2 = 1, \text{ w.p.1.} \tag{A.9}$$

Likewise, it is also known that

$$\sup_{k\geq k_0}\sum_{i=k_0}^{k} r_{ki}^2 < \infty. \tag{A.10}$$

Furthermore, noticing (A.8) and

$$\max_{k_0\leq i\leq k}\left\{\frac{(i+1)^{\lambda_1}}{i}\right\} = \begin{cases} \dfrac{(k+1)^{\lambda_1}}{k}, & \lambda_1 > 1; \\ \dfrac{(k_0+1)^{\lambda_1}}{k_0}, & 1/2 < \lambda_1 \leq 1, \end{cases}$$

we have $\lim_{k\to\infty}\max_{k_0\leq i\leq k} r_{ki} = 0$. Hence, (A.1)–(A.3) are true for $\{v_i\}$ and $\{r_{ki}\}$. By Lemma 8, we have $S_{2k}\to\mathcal{N}(0,1)$ as $k\to\infty$, and (16) is proved.

To prove (17), by (A.6) we have

$$\begin{aligned} &\mathrm{E}\left(\sqrt{k+1}a_{k+1}\right)^2 \\ &= \mathrm{E}S_{1k}^2 + \frac{\lambda_2^2\phi^2}{2\lambda_1-1}\mathrm{E}S_{2k}^2 + \frac{2\lambda_2\phi}{\sqrt{2\lambda_1-1}}\mathrm{E}S_{1k}S_{2k}. \end{aligned}$$

From $\sup_{k\geq 1}\phi_k^2 \leq \Phi < \infty$ and (A.10), we know that $\{\sum_{i=k_0}^{k} r_{ki}^2\phi_i^2, k\geq k_0\}$ is uniformly integrable. So, $\mathrm{E}S_{2k}^2 = \mathrm{E}\sum_{i=k_0}^{k} r_{ki}^2\phi_i^2 \to 1$ by (A.9). Moreover, it is known that $\mathrm{E}S_{1k}^2\to 0$ and $\mathrm{E}S_{1k}S_{2k}\to 0$ by (A.7) and $\mathrm{E}S_{1k}S_{2k} \leq \sqrt{\mathrm{E}S_{1k}^2}\sqrt{\mathrm{E}S_{2k}^2}$. Hence, (17) is also true. ■

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, 1987.

[2] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Boston, MA: Birkhauser, 1991.

[3] L. Ljung, A. Vicino, *IEEE Trans. Autom. Control: Special Issue Identification*, Oct. 2005.

[4] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. 2nd Ed. Springer-Verlag, New York, 2003.

[5] D. D. Falconer, F. Adachi, and B. Gudmundson, "Time division multiple access methods for wireless personal communications," *IEEE Commun. Mag.*, vol. 33, no. 1, pp. 50–57, 1995.

[6] A. N. Kolmogorov, "On some asymptotic characteristics of completely bounded spaces," *Dokl. Akad. Nauk SSSR*, vol. 108, no. 3, pp. 385–389, 1956.

[7] G. Zames, "On the metric complexity of causal linear systems: ε-entropy and ε-dimension for continuous time," *IEEE Trans. Autom. Control*, vol. AC-24, pp. 222–230, 1979.

[8] G. Zames, "Information-based theory of identification and adaptation," *Plenary Speech, 1996 CDC Conference*, Kobe, Japan, Dec., 1996.

[9] L. Y. Wang and L. Lin, "On metric complexity of discrete-time systems," *Syst. Control Lett.*, vol. 19, pp. 287–291, 1992.

[10] L. Y. Wang, "Uncertainty, information and complexity in identification and control," *Int. J. Robust Nonlin. Control*, vol. 10, pp. 857–874, 2000.

[11] G. Zames, L. Lin and L. Y. Wang, "Fast identification n-widths and uncertainty principles for LTI and slowly varying systems," *IEEE Trans. Autom. Control*, vol. AC-39, pp. 1827–1838, 1994.

[12] 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, no. 3, 1993.

[13] K. Poolla and A. Tikku, "On the time complexity of worst-case system identification," *IEEE Trans. Autom. Control*, vol. AC-39, pp. 944–950, 1994.

[14] A. Pinkus, *N-widths in Approximation Theory*. Springer-Verlag, Berlin, Germany, 1985.

[15] A. G. Vitushkin, *Theory of the Transmission and Processing of Information*. Pergamon Press, Oxford, U.K., 1961.

[16] J. F. Traub, G. W. Wasilkowski, and H. Wozniakowski, *Information-Based Complexity*. New York: Academic Press, 1988.

[17] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, 2004.

[18] D. E. Quevedo, A. Ahlén, A. S. Leong, and S. Dey, "On Kalman filtering over fading wireless channels with controlled transmission powers," *Automatica*, vol. 48, pp. 1306–1316, 2012.

[19] K. You and L. Xie, "Minimum data rate for mean square stabilizability of linear systems with Markovian packet losses," *IEEE Trans. Autom. Control*, vol. 56, no. 4, pp. 772–785, 2011.

[20] L. Qiu, G. Gu, and W. Chen, "Stabilization of networked multi-input systems with channel resource allocation," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 554–568, 2013.

[21] W.-H. Chen and W. X. Zheng, "An improved stabilization method for sampled-data control systems with control packet loss," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2378–2384, 2012.

[22] T. Li, M. Fu, L. Xie, and J. F. Zhang, "Distributed consensus with limited communication data rate," *IEEE Trans. Autom. Control*. vol. 56, no. 2, pp. 279–292, 2011.

[23] L. Y. Wang and L. Lin, "Information-based complexity of uncertainty sets in feedback control," *IEEE Trans. Autom. Control*, vol. 46, no. 4, pp. 519–533, 2001.

[24] L. Y. Wang, J. F. Zhang, and G. Yin, "System identification using binary sensors," *IEEE Trans. Autom. Control*, vol. 48, pp. 1892–1907, 2003.

[25] B. Godoy, G. Goodwin, J. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no. 9, pp. 1905–1915, 2011.

[26] M. Casini, A. Garulli and A. Vicino, "Input design in worst-case system identification with quantized measurements," *Automatica*, vol. 48, no. 12, pp. 2997–3007, 2012.

[27] D. Marelli, K. You, and M. Fu, "Identification of ARMA models using intermittent and quantized output observations," *Automatica*, vol. 49, pp. 360–369, 2013.

[28] L. Y. Wang, G. Yin, J. F. Zhang, Y. L. Zhao, *System Identification With Quantized Observations*. Birkhäuser, Boston, MA, 2010.

[29] D. Wackerly, W. Mendenhall, and R. L. Scheaffer, *Mathematical Statistics With Applications*. 7th Ed., Cengage Learning, Boston, MA, 2007.

[30] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. 2nd Ed., Routledge, London, U.K., 1988.

[31] J. Kuang, *Applied Inequation*. 3rd Ed., Shandong Science and Technology Press, Shandong, China, 2004.

[32] J. Guo and Y. Zhao, "Identification of the gain system with quantized observations and bounded persistent excitations," *Science China Inform. Sci.*, vol. 57, pp. 012205:1–012205:15, 2014.

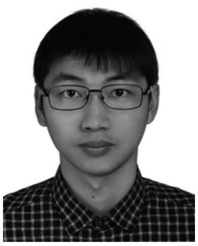
[33] C. Z. Wei, "Asymptotic properties of least-squares estimates in stochastic regression models," *Annu. Statist.*, vol. 13, no. 4, pp. 1498–1508, 1985.

[34] S. Hu, "Central limit theorem for weighted sum of martingale difference," *Acta Mathematicae Applicatae Sinica*, vol. 24, no. 4, pp. 539–546, 2001.

[35] H. F. Chen, *Stochastic Approximation and Its Application*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

Jin Guo received the B.S. degree in mathematics from Shandong University, China, in 2008, and the Ph.D. degree in system modeling and control theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2013.

He is currently an Associate Professor in the School of Automation and Electrical Engineering, University of Science and Technology, Beijing, China. His research interests are identification and control of set-valued output systems and systems biology.



Biqiang Mu was born in Sichuan, China, 1986. He received the B.Eng. degree in material formation and control engineering from Sichuan University in 2008, and the Ph.D. degree in operations research and cybernetics from the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2013.

He was a Postdoctoral Fellow at the Wayne State University, Detroit, MI, from 2013 to 2014. He is currently a Postdoctoral Fellow at the University of Western Sydney, Sydney, Australia. His research interests include system identification and applications.

Le Yi Wang (F'12) 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, Michigan, where he is currently a Professor in the Department of Electrical and Computer Engineering. His research interests are in the areas of 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, power systems, and computer applications of control methodologies. He was a keynote speaker in several international conferences. He serves on the IFAC Technical Committee on Modeling, Identification and Signal Processing. He was an Associate Editor of the IEEE Transactions on Automatic Control and several other journals. He was a Visiting Faculty at University of Michigan in 1996 and Visiting Faculty Fellow at University of Western Sydney in 2009 and 2013. He is a member of the Core International Expert Group at Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and an International Expert Adviser at Beijing Jiao Tong University. He is a Fellow of IEEE.

George Yin (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 Ph.D. degree in applied mathematics from Brown University, Providence, RI, in 1987.

He joined Wayne State University in 1987, and became a Professor in 1996. His research interests include stochastic systems and applications.

Dr. Yin is a Fellow of IFAC and a Fellow of SIAM. He was Chair of SIAM Activity Group on Control and Systems Theory; he has been serving on the Board of Directors of American Automatic Control Council. He is an Associate Editor of SIAM Journal on Control and Optimization, and on the editorial board of a number of other journals. He was an Associate Editor of Automatica (2005–2011) and IEEE Transactions on Automatic Control (1994–1998). He served as a member of the program committee for many IEEE Conference on Control and Decision; he also severed on the IFAC Technical Committee on Modeling, Identification and Signal Processing; he was Co-Chair of SIAM Conference on Control & Its Application, 2011, and Co-Chair of two AMS-IMS-SIAM Summer Research Conferences; he also chaired a number SIAM prize selection committees.

Lijian Xu (M'14) received the B.S. degree from the University of Science and Technology, Beijing, China, in 1998, the M.S. degree from the University of Central Florida, Orlando, in 2001, and the Ph.D. degree from Wayne State University, Detroit, MI, in 2014.

He worked for AT&T, Florida, USA and Telus Communication, Calgary, Alberta, Canada as an Engineer and Engineering Manager from 2001 to 2007. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering Technology, Farmingdale State College, the State University of New York. His research interests are in the areas of networked control systems, digital wireless communications, consensus control, vehicle platoon and safety.

Dr. Xu received the Best Paper Award from 2012 IEEE International Conference on Electro/Information Technology.