Persistent tracking and identification of regime-switching systems with structural uncertainties: unmodeled dynamics, observation bias, and nonlinear model mismatch

Shaobai Kan1,*,†, George Yin2 and Le Yi Wang3

1Department of Mathematics and Computer Science, John Jay College of Criminal Justice, CUNY, New York, NY 10019, USA
2Department of Mathematics, Wayne State University, Detroit, MI 48202, USA
3Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA

SUMMARY

This work focuses on tracking and system identification of systems with regime-switching parameters, which are modeled by a Markov process. It introduces a framework for persistent identification problems that encompass many typical system uncertainties, including parameter switching, stochastic observation disturbances, deterministic unmodeled dynamics, sensor observation bias, and nonlinear model mismatch. In accordance with the ‘frequency’ of the parameter switching process, we divide the problems into two classes. For fast-switching systems, the switching parameters are stochastic processes modeled by irreducible and aperiodic Markov chains. Because accurately tracking real-time parameters in such systems is not possible because of the uncertainty principles, the effect of parameter switching is evaluated on their average by the stationary distribution of the Markovian chain and estimated by the least squares algorithms. We derive upper and lower bounds on identification errors, which characterize how identification accuracy depends on the earlier uncertainty terms. When the system parameters switch their values infrequently in a probabilistic sense, their values can be tracked based on input/output observations. Stochastic approximation algorithms with adaptive step sizes are used for such systems. Simulation studies are carried out to demonstrate that slowly varying parameters could be tracked with reasonable accuracy. Copyright © 2012 John Wiley & Sons, Ltd.

Received 29 June 2011; Revised 15 January 2012; Accepted 4 March 2012

KEY WORDS: parameter switching, unmodeled dynamics, nonlinear model mismatch, observation bias, exogenous noise, tracking, persistent identification

1. INTRODUCTION

Stochastic hybrid systems are of essential importance in modeling real-world systems such as networked control of mobile agents with switching communication topologies, automotive powertrain control with gear shifting, hybrid vehicle control strategies with battery charging and discharging, and power grids with fault occurrence and clearance. System identification of hybrid systems must take into consideration switching system parameters in such systems. In addition, practical systems encounter measurement noises, modeling errors caused by model simplifications, and inaccuracy in sensor/actuator characteristics.

Significant research efforts have been made to the identification of stochastic hybrid systems during the past decade. Most frameworks in such a pursuit include typical observation noises in tracking switching parameters. This paper aims to develop a more comprehensive framework that
accommodates stochastic observation noises and worst-case uncertainties from unmodeled dynamics, the model mismatches, and bias. This will allow an integrated approach in evaluating combined impact of various sources of uncertainty on system identification and characterizing identification errors more accurately and comprehensively.

In [1], the concept of persistent identification of time-varying system was first presented. This idea was further investigated in [2] for systems with unmodeled dynamics and exogenous disturbances, and in [3] for closed-loop systems. In [4], regime-switching systems with unmodeled dynamics was first introduced. It turns out that to track and identify such regime-switching systems whose parameters can be modeled by discrete-time Markov chains, one needs to divide them into two different classes of problems: fast-switching systems and infrequent-switching systems. In the fast-switching systems, the parameter switching is relatively fast in comparison with the system dynamics and estimation convergence speed, which makes parameter intractable [5]. Instead of tracking the instantaneous variations of the Markovian parameter, it is more practical to track the average behavior of these systems [4]. In the infrequent-switching systems, the system parameter changes its states relatively slowly. It allows one to track its moment-by-moment changes with desired accuracy by developing effective stochastic approximation algorithms. This identification framework was later exploited in the regime-switching systems (with/without unmodeled dynamics) equipped with only binary-valued sensors [6, 7].

Continuing along this line of work, this paper further investigates identification of a time-varying process whose vector-valued parameters can be represented by a discrete-time Markov chain switching among a finite set of possible values. In addition, the systems that we aim to identify are subject not only to stochastic measurement noise but also to various structural uncertainties such as unmodeled dynamics, model mismatch, and observation bias. In reality, the bias term comes either from a sensor bias or a nonzero mean from random disturbance, and the model mismatch term stems from the consideration of approximation of nonlinearity by means of linear parametrization, see [8] for details. Compared with the existing literature, in our setup, the parameter is not a constant but a randomly varying stochastic processes. In addition, added difficulties come from additional structure uncertain factors including observation bias and model mismatch. The challenges are not just caused by the impact of these terms themselves on the parameter identification, but also their joint impacts together with the Markovian regime-switching, stochastic noise, and unmodeled dynamics. Moreover, in order to cover a variety of situations, we assume only little information being known about these uncertain terms in this work except that they are of polynomial decay with respect to the discrete time, which makes the identification and tracking in such regime-switching settings even more challenging. Similar to [4, 6, 7], in this paper, we investigate two classes of regime-switching systems: fast-switching systems and infrequent-switching systems, see also [9, 10] and references therein for related work.

We would like to mention that in the literature, under the setup of $H^\infty$ framework, simultaneous identification of nominal model, parametric and unstructured uncertainties were treated in [11]. The authors aimed to find the smallest model set that is described by nominal model, parametric uncertainty bound, unstructured uncertainty bound, and consistent with a number of noise-free input/output data. In our work, identification with randomness, in fact, Markovian parameters are considered. The systems under consideration include, in addition to modeled part and observation noise, the unmodeled dynamics, the bias, and the model mismatch. The unmodeled dynamics are caused by finite-order model parametrization; the bias term, often comes from either a sensor bias or caused by a nonzero mean from random disturbance; the model mismatch may be caused by the consideration of approximation of nonlinearity by means of linear parametrization.

In the fast-switching systems, the probability of a parameter staying at the current value is much smaller than that of jumping to a different value. The existing literature has shown that, independent of identification algorithms, fast-switching systems cannot be identified with reasonable accuracy from observation data owing to the uncertainty principle [12], see also [4, 5]. Thus, a viable alternative suggests to track the average behavior of these systems, which is where our effort lies. This is particularly suited for applications in mobile agent control, distributed control, hierarchical systems, and supervisory and management systems, see [4, 6, 7] for details. By using the concepts of persistent identification, control-oriented system modeling, and stochastic analysis, the central issues of
irreducible identification errors and time complexity in such identification problems are investigated. Upper and lower bounds on estimation errors are established to reveal the joint impacts of Markovian switching, unmodeled dynamics, model mismatch, observation bias, and stochastic noise on the identification accuracy. A clear understanding of the dependence of identification accuracy on the preceding terms can potentially provide guidelines on model order selection, sensor accuracy specifications, sensor drifting compensations, and identification algorithm development.

In the slow or infrequent switching systems, the probability of a parameter staying at the current value is near one. Because the states of the Markov chain switch relatively infrequently, we are allowed to track the instantaneous variations of the system parameters with substantial accuracy. To carry out this task, we first present an adaptive algorithm with variable step sizes and then use it to identify and track the time-varying parameters in several numerical experiments. It is worthwhile to mention that the idea of this adaptive step size algorithm evolved from the fixed step size algorithm. It was first suggested in [5] for slowly varying parameters, and further exploited in [4, 13–15]. Simulation studies demonstrate that in the presence of Markovian regime-switching, unmodeled dynamics, observation bias, nonlinear model mismatch, and stochastic noise, this proposed algorithm can be well performed in the parameter tracking and identification. In fact, this is the first time for us to investigate the simulation studies of such regime-switching systems in the presence of all these structural uncertainties. Compared with our previous work [4, 7], where we presented numerical experiments without any deterministic structural uncertainties, the tracking performance in this work should be a little bit worse. This is reasonable and unavoidable because the existence of the added uncertain terms will surely undermine the effect of the tracking algorithm.

The remainder of the paper is organized as follows. In Section 2, we give the precise formulation of the identification problems of regime-switching parameter processes with various structural uncertainties. Section 3 proceeds with long-time average behavior of parameter estimates for fast-switching systems. The notation of the persistent tracking and identification is recalled briefly, and a modified version of these concepts is presented. Upper and lower bounds on estimation errors are established to characterize joint impact of Markovian regime-switching, unmodeled dynamics, observation bias, model mismatch, and stochastic observation noise. Section 4 is devoted to tracking system parameters of infrequent switching types, based on adaptive step size algorithm. Simulation study is carried out to demonstrate the performance of the proposed algorithm, including both one-dimensional and two-dimensional cases. Section 5 concludes the paper with a few further remarks. For clarity and readability, the proofs of the main results are included in the Appendix at the end of the paper.

2. FORMULATION

2.1. Markovian systems with structural uncertainties

Consider a single-input single-output (SISO), discrete-time, linear-time-varying system

\[ y_k = \sum_{l=0}^{\infty} a_l(k)u_{k-l} + b_k + \Delta_k + d_k, \quad k = i_0, i_0 + 1, \ldots, \quad (2.1) \]

where \( i_0 \) is the starting time of observations, \( u_k \) is the probing input signal, \( b_k \) is a time-dependent bias term, \( \Delta_k \) represents the nonlinear model mismatch, and \( d_k \) is a sequence of random disturbances. In this paper, for a vector \( v \) (either finite or infinite dimensional), \( \|v\|_1 \) and \( \|v\|_\infty \) denote its \( \ell_1 \) and \( \ell_\infty \) norms, respectively. The system is bounded-input bounded-output stable with \( \sup_k \sum_{l=0}^{\infty} |a_l(k)| < \infty \). The probing input \( u \) is deterministic and can be selected arbitrarily by the designer except that \( u \) is uniformly bounded, that is, \( \|u\|_{\infty} \leq \kappa_u < \infty \).

For a given model order \( n \), the system parameters can be decomposed into two parts, that is, the modeled part \( \theta_k = (a_0(k), \ldots, a_{n-1}(k))' \) and the unmodeled dynamics \( \bar{\theta}_k = (a_n(k), a_{n+1}(k), \ldots)' \). Thus, the system input/output relationship can be represented by

\[ y_k = \phi'_k \theta_k + \tilde{\phi}'_k \bar{\theta}_k + b_k + \Delta_k + d_k, \quad (2.2) \]
where $\gamma'$ denotes the transpose of $\gamma \in \mathbb{R}^{1 \times 2}$ for $t_1, t_2 \geq 1$, and
\[
\phi_k = (u_k, \ldots, u_{k-n+1})', \quad \hat{\phi}_k = (u_{k-n}, u_{k-n-1}, \ldots)'.
\] (2.3)

Note that here, $\theta_k \in \mathbb{R}^n$ is assumed to be the unknown Markovian parameter that needs to be identified, $\phi_k$ is an $n$-dimensional vector, and $\hat{\phi}_k$ is an infinite-dimensional vector associated with the unmodeled dynamics $\hat{\theta}_k$. There is no detailed information given on the unmodeled dynamics, but only the uniform upper bound is provided, namely $\sup_k \|\hat{\theta}_k\|_1 \leq \hat{\epsilon}$. This is in line with the past study on persistent identification [2, 4, 6, 16].

Throughout the paper, the following assumptions are imposed:

(C1) $\sum_{i=0}^{N-1} \phi_i \phi_i' / N \to \Xi$ as $N \to \infty$, where $\Xi$ is positive definite.

(C2) There exist a positive integer $K$ and positive real numbers $\kappa_1$ and $\kappa_2$ such that
- $\sup_{-K \leq i \leq K} |b_{i} + i| \leq \kappa_1 (i + 1)^{-\alpha_1}$ with probability one, for some $\alpha_1 > 0$,
- $\sup_{-K \leq i \leq K} |\Delta_{i} + i| \leq \kappa_2 (i + 1)^{-\alpha_2}$ with probability one for some $\alpha_2 > 0$.

(C3) The disturbance $\{d_k\}$ is a sequence of independent and identically distributed random variables, with a common distribution $\rho$ that is symmetric with respect to the origin and $0 < \rho d^2 = \sigma_d^2 < \infty$. In addition, the moment generating function $G(z) = E \exp(zd)$ exists.

We collect the output observations in such a scenario: after applying an input sequence $\{u_k\}$, starting at $i_0 (-K \leq i_0 \leq K)$, the output is observed during $k = i_0, \ldots, i_0 + N - 1$ with observation length $N \gg n$.

Remark 2.1

It was pointed out in [4] that assumption (C1) requires that the selected input sequence possess a certain persistent excitation property, which is satisfied by all full-rank $n$-periodic signals in open-loop identification problems. This condition, in fact, is required and unique to persistent identification with structural uncertainties such as unmodeled dynamics, see [2, 4] for details.

Assumption (C2) is imposed to overcome conservativeness of estimation errors. The nonlinear model mismatch term stems from local linearization, and as such contains higher-order polynomial terms of inputs in its expressions. In control applications of system identification, the identified models are usually used in regulation problems with constant nominal operating points or tracking control for slow time-varying systems. In both cases, the system will be operating near a fixed operating point during active identification/control intervals. The model structure used in this paper is, in fact, a perturbation model around the operating point. As a result, the input $u_k$ is actually a small dither added to the nominal (constant) input to provide excitation for system identification. To understand this, suppose $U = \{u_k\}$ is an $n$-periodic and full-rank input with a normalized magnitude $\|u_k\|_\infty = 1$, and $\mu$ is a scaling factor on the input. The input/output relationship under the scaled input $\{\mu u_k\}$ is
\[
y_k^0 = \sum_{j=0}^{\infty} a_j \mu u_{k-j} + \Delta_k^0 (\mu u_k, \ldots)
\]
which implies that the normalized input/output relationship
\[
y_k = y_k^0 / \mu = \sum_{j=0}^{\infty} a_j u_{k-j} + \Delta_k^0 (\mu u_k, \ldots) / \mu.
\]

To reduce the adversary effect of the dither on the output, the scaling factor $\mu$ is gradually reduced. Here, owing to higher-order terms in $\Delta_k^0$, we have
\[
\Delta_k^0 (\mu u_k, \ldots) / \mu \to 0 \text{ as } \mu \to 0
\]
leading to the assumptions that we use in this paper. Our results indicate that if $\mu$ is reduced gradually in a polynomial speed, convergence of the parameter estimates can be achieved. It is noted that if we use a constant $\mu$, then the model mismatch term will be an unknown but bounded uncertainty, which will induce an irreducible identification error, similar to the unmodeled dynamics.
In this case, the error bounds can be derived similarly to unmodeled dynamics. But this will result in a conservative error bound.

Assumption (C3) is concerned with the random noise sequence. This condition is satisfied for many random processes that are symmetric with respect to the origin such as stationary Gaussian processes and noises with uniform distributions.

Remark 2.2
Note that compared with the framework given in [8], here, we only specify the form of unmodeled dynamics, that is, $\dot{\theta}_k\hat{\theta}_k$, but we do not specify the form of the observation bias and nonlinear model mismatch. The bias term, $b_k$ in (2.2) often comes from either a sensor bias or caused by a nonzero mean from random disturbance; and the model mismatch term, $\Delta_k$ in (2.2), stems from the consideration of approximation of nonlinearity by means of linear parametrization. In general, there are two types of observation bias commonly used in system identification: persistent bias and transient bias. Assumption (C2) indicates that, in our setup, the bias is assumed to be transient. To be more specific, the observation bias here is assumed to be of polynomial decay, and similarly for the model mismatch. As in [8], only their upper bounds are given in (C2) so that it facilitates a general treatment of the identification problem and reflects many real-world scenarios, in which we have some ideas on the approximation bounds but not necessarily the detailed and precise descriptions of deviations from the true system.

3. IDENTIFICATION OF FAST SWITCHING SYSTEMS: LONG-RUN AVERAGE BEHAVIOR AND ESTIMATION ERROR BOUNDS

We first investigate the case that the system parameter $\theta_k$ is a discrete-time Markov chain that changes its states in a fast pace. It is well known that if the jump parameter switches back and forth too frequently, it would be impossible to identify the instantaneous jumps [4–7]. Instead, one can examine its average behavior. This is quite meaningful in reality because when a system performance is measured by some averaged outputs, as in most performance indices for optimal or adaptive control, the net effect of fast-switching parameters on the system performance can be approximated by using their average values [4, 6, 7]. Thus, in this section, we aim to identify the averaged system for a fast-switching system. To proceed, in this section, we further impose the following assumption as a general regime-switching case.

(C4) The discrete-time Markov chain $\{\theta_k\}$ is irreducible and aperiodic with finite state space $\mathcal{M} = \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}\}$ and transition probability matrix $P$. Here, the transition matrix $P$ and state space $\mathcal{M}$ are both unknown. $\theta_k$ and $d_k$ are independent.

Because the Markov chain varies at a fast pace, within a short period of time, it should settle at a stationary or steady state frame, in which the underlying system is a weighted average with the weighting factors being the stationary distribution of the Markovian chain. Mathematically, this could be shown by introducing and minimizing such a cost function

$$J(\bar{\theta}) = \lim_{k \to \infty} E(\theta_k - \bar{\theta})^2.$$  

By the well-known ergodicity of the Markov chain, the average behavior of the Markovian parameter can be characterized by the minimizer of $J(\bar{\theta})$, that is,

$$\bar{\theta} = \sum_{i=1}^{m} \theta^{(i)} v_i,$$  

where $\nu = (v_1, v_2, \ldots, v_m)$ is the stationary distribution of the parameter $\theta_k$ [4].

Remark 3.1
Note that the system parameter $\theta_k$ is assumed to be an irreducible and aperiodic discrete-time Markov chain in our setup. Its stationary distribution thus exists. Moreover, because in our assumption (C4), the transition probability matrix $P$ and state space $\mathcal{M}$ are both unknown, the average $\bar{\theta}$
cannot be calculated directly. Alternatively, in this work, we will employ a standard least squares estimation to estimate the average of the system parameter

$$\hat{\theta}_{i_0+N-1} = \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}' \right)^{-1} \sum_{i=0}^{N-1} \phi_{i_0+i} y_{i_0+i}. \quad (3.2)$$

The resulting estimation error bounds, including upper and lower bounds, will be established to reveal how identification accuracy depends on the aforementioned structural uncertainties such as unmodeled dynamics, nonlinear model mismatch, and observation bias.

3.1. Persistent identification

To facilitate the understanding of the results presented in the coming sections, we recall the definitions and the notations of the general persistent identification that were first presented by Wang and Yin in [2]. Define $L = \{L_{i_0} \in \mathbb{R}^{n \times n} : L_{i_0}(\phi_{i_0}', \ldots, \phi_{i_0+N-1}') = I_n \}$, where $I_n$ is the $n$-dimensional identity matrix. For any $L_{i_0} \in L$, denote

$$\eta = \hat{\theta}_N(i_0) - \hat{\theta} = L_{i_0}(y_{i_0}, \ldots, y_{i_0+N-1})' - \hat{\theta}$$

to be the estimation error.

**Definition 3.2 (General)**

For a given tolerable identification error $\beta > 0$, the optimal persistent identification error is given as

$$Q_{(N,\tilde{\beta})}(\beta) = \inf_{\|u\|_\infty \leq \kappa u} \inf_{i_0} \sup_{L \in \mathbb{L}} \sup_{\|\hat{\theta}_N\|_1 \leq \tilde{\beta}} P_{(N,L,u,i_0,\hat{\theta})}(\|\eta\|_1 \geq \beta), \quad (3.3)$$

where $\hat{\theta}$ denotes the infinite-dimensional vector of unmodeled dynamics. For a selected confidence level $0 \leq \alpha \leq 1$, the minimal observation length $N$ is defined as

$$N_{\alpha}(\tilde{\beta}, \beta) = \inf\{N : Q_{(N,\tilde{\beta})}(\beta) \leq \alpha\}.$$

To investigate the central issues of irreducible identification errors and time complexity, we will use these concepts of persistent identification in our problem setting. However, because of the restrictions on the regime-switching setting and assumptions on our transient bias and model mismatch, we need to modify slightly the preceding definitions. First, in this paper, we only pay attention to the periodic signals in this work. We would only consider $N = \kappa n$ for positive integer $\kappa$, that is, the observation length is a multiple of the model order. Denote by $\mathbb{S}$ the following class of input signals: $\mathbb{S} := \{u \in L^\infty : \|u\|_\infty \leq \kappa u, u$ is $n$-periodic and full rank}. When the input is limited to $\mathbb{S}$ and the identification mapping is specified to be the LS estimation, with a slight modification of (3.3), we can have the following definitions:

**Definition 3.3 (Modified)**

For a given tolerable identification error $\beta > 0$ and a selected confidence level $0 \leq \alpha \leq 1$, the optimal persistent identification error and the minimal observation length are given as

$$Q^0_{(N,\tilde{\beta})}(\beta) = \inf_{u \in \mathbb{S}} \sup_{-K \leq i_0 \leq K} \sup_{\|\hat{\theta}_N\|_1 \leq \tilde{\beta}} P_{(N,L,u,i_0,\hat{\theta})}(\|\eta\|_1 \geq \beta)$$

and

$$N^0_{\alpha}(\tilde{\beta}, \beta) = \inf\{N : Q^0_{(N,\tilde{\beta})}(\beta) \leq \alpha\}.$$
3.2. Asymptotic upper bounds on estimation errors

Continuing on our asymptotic study, we aim to evaluate the quality of our LS estimate and then obtain the asymptotic upper bounds on estimation errors, that is, \( \hat{\theta}_{i_0+N-1} - \theta \). Before proceeding further, we first present an auxiliary result for the observation bias and nonlinear model mismatch. Then we obtain the desired upper bounds on identification errors. At the end of this section, the minimum requirement for the sample size to achieve desired accuracy would be provided. As in [2] and [4], we select the input \( u \) to be the \( n \)-periodic signal with the first \( n \) components \( 1, 0, \ldots, 0 \). Then we can derive the following lemma:

**Lemma 3.4**

Under assumptions (C1)–(C3), as \( \kappa \to \infty \) (i.e., \( N \to \infty \)),

\[
\left\| \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} b_{i_0+i} \right) \right\|_1 \leq \kappa_1 n^{1-\alpha_1},
\]

(3.4)

and

\[
\left\| \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \Delta_{i_0+i} \right) \right\|_1 \leq \kappa_2 n^{1-\alpha_2}.
\]

(3.5)

To proceed with the upper bounds, we need to recall a result that was presented in Yin et al. [4]. Consider the vector

\[
\rho_N(t) = (\rho_N^1(t), \ldots, \rho_N^m(t)), \quad \text{with}
\]

\[
\rho_N^j(t) = \frac{1}{\sqrt{N}} \sum_{i=0}^{[Nt]-1} [1_{\{\theta_{i_0+i} = \theta(j)\}} - v_j], \quad j = 1, \ldots, m,
\]

where \( 1_{\{\theta_{i_0+i} = \theta(j)\}} \) is the indicator function defined by

\[
1_{\{\theta_{i_0+i} = \theta(j)\}} = \begin{cases} 1, & \text{if } \theta_{i_0+i} = \theta(j), \\ 0, & \text{otherwise}. \end{cases}
\]

It was shown in [4] that under assumption (C4), the following assertions hold:

(i) \( \rho_N(\cdot) \) converges weakly to a Brownian motion whose mean is zero and whose variance function is given by \( \Sigma \) with

\[
\Sigma = \sum_{l=0}^{\infty} [\operatorname{diag}(v_1, \ldots, v_m) \Psi(l) + \Psi'(l) \operatorname{diag}(v_1, \ldots, v_m)],
\]

where \( \operatorname{diag}(v_1, \ldots, v_m) \) denotes the diagonal matrix with entries \( v_1, \ldots, v_m \) and \( \Psi(l) \) is the solution of

\[
\begin{cases} \Psi(l) = \Psi(l-1) P, \\ \Psi(0) = I - 1_m v, \end{cases}
\]

where \( 1_m = (1, \ldots, 1)' \in \mathbb{R}^{m \times 1} \) (see [17] for the notion of weak convergence).

(ii) for each \( j = 1, \ldots, m \), \( \rho_N^j(\cdot) \) converges weakly to a one-dimensional Brownian motion \( w_j(\cdot) \) with 0 mean and variance \( \sigma_j^2 t \), where \( \sigma_j^2 = \Sigma_{jj} \). (see [4, Lemma 3.6] for detail)

Denote

\[
g(\tau) = \inf_{z} E \exp(z (d - \tau)) = \inf_{z} \exp(-z \tau) G(z),
\]
where $G(z)$ is the moment generating function of $d$, which is assumed to exist by virtue of (C3). Then we can obtain the following theorem. This result characterizes the joint impact of random noise, stochastic unknown parameters, unmodeled dynamics, observation bias, and model mismatch in a worst-case sense for all possible unmodeled dynamics bounded by $\hat{e}$, on probabilities of identification errors.

**Theorem 3.5**
Assume (C1)–(C4). Suppose $u \in S$ and the identification algorithm is the LS estimation (3.2). Then for each $\varepsilon > 0$, the following upper bound holds as $\kappa \to \infty$ (i.e., $N \to \infty$),

$$Q_{(\hat{e},N)}^0(\hat{e} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon) \leq 2m \exp \left( -\frac{\varepsilon^2 n \kappa}{2\sigma^2 m^2 c_1^2} \right) + 2n \left[ g \left( \frac{\varepsilon}{2n} \right) \right]^\kappa,$$

where

$$c_1 = \max(||\hat{\theta}(1)||_1, \ldots, ||\hat{\theta}(m)||_1), \quad \sigma^2 = \max(\sigma_1^2, \ldots, \sigma_m^2).$$

**Remark 3.6**
The left-hand side of (3.6) is given as

$$Q_{(\hat{e},N)}^0(\hat{e} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon) = \inf_{u \in S} \sup_{-K \leq \hat{\theta}_k \leq K} \sup_{||\hat{\theta}_k||_1 \leq \hat{e}} P_{(N,1,S,u,\hat{\theta}_k)}(||\hat{\theta}_k - \hat{\theta}||_1 \geq \hat{e} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon).$$

The preceding result, together with the lower bounds established in Theorem 3.9, reveals how the identification error bounds depend on the size $\hat{e}$ of unmodeled dynamics, the rate $\alpha_1$ of bias decaying, the rate $\alpha_2$ of model mismatch decreasing, and the fixed model order $n$ selected by the designer.

It can be observed from (3.6) that to have

$$P(\|\hat{\theta}_k - \hat{\theta}||_1 \geq \hat{e} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon) \leq \alpha,$$

it suffices that

$$2m \exp \left( -\frac{\varepsilon^2 n \kappa}{2\sigma^2 m^2 c_1^2} \right) + 2n \left[ g \left( \frac{\varepsilon}{2n} \right) \right]^\kappa \leq \alpha.$$

Then the following corollary could be derived immediately. Note that the proof of Corollary 3.7 can be carried out similarly as in [4, Corollary 3.9]. The details are thus omitted.

**Corollary 3.7**
Assume (C1)–(C4) and $u \in S$. For each $\varepsilon > 0$, an upper bound of

$$N_{(\hat{e},N)}^0(\hat{e}, \hat{e} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon) \leq \tilde{k}$$

is given by

$$\tilde{k} = \inf \left\{ \kappa : \left( \exp(-\kappa_0 \varepsilon^2 \kappa) \vee \left[ g \left( \frac{\varepsilon}{2n} \right) \right]^\kappa \right) \leq \frac{\alpha}{4(m \lor n)} \right\},$$

where $\kappa_0 = n/(2\sigma^2 m c_1^2)$ and $(a \lor b) = \max(a, b)$. 

---

3.3. Asymptotic lower bounds on estimation errors

For simplicity of analysis, we define

\[ \eta_b = \left( \sum_{i=0}^{N-1} \phi_{i0+i} \phi_{i0+i}' \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i0+i} b_{i0+i} \right), \]

and

\[ \eta_\Delta = \left( \sum_{i=0}^{N-1} \phi_{i0+i} \phi_{i0+i}' \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i0+i} \Delta_{i0+i} \right). \]

It is easily seen from (A.3), (A.4), and (A.5) in the Appendix that the estimation error \( \hat{\theta}_{t_0+N-1} - \overline{\theta} \) can be decomposed into five parts, that is, \( \hat{\theta}_{t_0+N-1} - \overline{\theta} = \eta_m + \eta_d + \eta_b + \eta_\Delta + \eta_s \). Here,

\[ \eta_m = \left( \sum_{i=0}^{N-1} \phi_{i0+i} \phi_{i0+i}' \right)^{-1} \left[ \sum_{i=0}^{N-1} \phi_{i0+i} \phi_{i0+i}' (\theta_{i0+i} - \overline{\theta}) \right] \]

is the error caused by the Markov chain, \( \eta_d \) denotes the error caused by the unmodeled dynamics, \( \eta_b \) and \( \eta_\Delta \) are the errors attributable to the observation bias and nonlinear model mismatch, respectively, and \( \eta_s \) is the error caused by the stochastic disturbances. To obtain the lower bound, similar to [2,4], we define \( \xi_d = \text{sign}(\eta_d) := (\text{sign}(\eta^{(1)}_d), ..., \text{sign}(\eta^{(N)}_d))^t \), where \( \eta^{(i)}_d \) is the \( i \)th component of the vector \( \eta_d \) and

\[
\text{sign}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0. \end{cases}
\]

Then we can obtain the following lemma. Note that the proof of the result (ii) in Lemma 3.8 is quite similar to [2, Lemma 3], and the details are thus omitted. We only give the proof of (i) in the Appendix.

**Lemma 3.8**
Suppose that \( \hat{\theta}_{t_0+N-1} - \overline{\theta} = \eta_m + \eta_d + \eta_b + \eta_\Delta + \eta_s \in \mathbb{R}^n \) with input \( u \in \Sigma \).

(i) If \( |\eta^{(i)}_m \xi_d| \leq \varepsilon \) and \( |\eta^{(i)}_d \xi_d| \geq 2\varepsilon + 2\kappa_1 n^{1-\alpha_1} + 2\kappa_2 n^{1-\alpha_2} + 2\varepsilon \), then \( \|\hat{\theta}_{t_0+N-1} - \overline{\theta}\|_1 \geq \tilde{\varepsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon \).

(ii) Moreover, there exists a \( \lambda \in \mathbb{R}^n \) with \( \|\lambda\|_1 \geq (1/\kappa_u) \) such that

\[
P \left( \|\hat{\theta}_{t_0+k\kappa}-\overline{\theta}\|_1 \geq \tilde{\varepsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon \right) \geq P \left( \|\lambda^t D\| \geq 2\varepsilon + 2\kappa_1 n^{1-\alpha_1} + 2\kappa_2 n^{1-\alpha_2} + 2\varepsilon, |\eta^{(i)}_m \xi_d| \leq \varepsilon \right),
\]

where \( D = (d_{i0}, ..., d_{i0+N-1})^t \in \mathbb{R}^N \).

Now, we are ready to present the estimation lower bounds in the following theorem. The desired lower bounds should reflect the impacts of all those stochastic and deterministic uncertain factors on parameter identification.

**Theorem 3.9**
Assume (C1)–(C4). Then

\[
Q_0^{0}(\tilde{\varepsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon) \geq P(d \geq 2(\tilde{\varepsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \epsilon) \kappa_u | d \geq 0) \left[ 1 - 2 \exp \left( \frac{-2\varepsilon^2 N}{\sigma^2 m^2 c_1^2} \right) \right],
\]

(3.10)
where \( c_1 = \max(\|\theta^{(1)}\|_1, \ldots, \|\theta^{(m)}\|_1) \).

**Remark 3.10**

As a matter of fact, with regard to the lower bounds, one can also consider the same persistent identification errors as given in [2]:

\[
Q^0_{(\bar{e}, N)}(\bar{e} + \bar{e}) = \inf_{u \in \mathcal{U}} \sup_{-K \leq s \leq K} \sup_{\|\theta_k\|_1 \leq \bar{e}} P_{(N, L, S, \mu, I_0, \hat{\theta})} (\|\hat{\theta}_{t_0 + \kappa n - 1} - \mu\|_1 \geq \bar{e} + \bar{e}).
\]

It can be shown in the similar way that the identification error \( P(\|\hat{\theta}_{t_0 + \kappa n - 1} - \mu\|_1 \geq \bar{e} + \bar{e}) \), and hence \( Q^0_{(\bar{e}, N)}(\bar{e} + \bar{e}) \) have a different but similar lower bound

\[
Q^0_{(\bar{e}, N)}(\bar{e} + \bar{e}) \geq P(d \geq (2\bar{e} + \kappa_1 n^{1-a_1} + \kappa_2 n^{1-a_2} + 2\bar{e})N\kappa_u | d \geq 0) \left[ 1 - 2m \exp \left(-\frac{2\bar{e}^2 N}{\sigma^2 m^2 c_1^2}\right) \right].
\]

**Corollary 3.11**

Suppose that the conditions of Theorem 3.9 are satisfied and noise \( d \) follows a normal distribution with mean 0 and variance \( \sigma_d^2 \). Then

\[
Q^0_{(\bar{e}, N)}(\bar{e} + \kappa_1 n^{1-a_1} + \kappa_2 n^{1-a_2} + \bar{e})
\]

\[
\geq \sqrt{\frac{2}{\pi}} \left[ \frac{L}{\sqrt{N}} - \left( \frac{L}{\sqrt{N}} \right)^3 \right] \exp \left(-\frac{N}{2L^2}\right) \left[ 1 - 2m \exp \left(-\frac{2\bar{e}^2 N}{\sigma^2 m^2 c_1^2}\right) \right],
\]

where

\[
L = \frac{\sigma_d}{2(\bar{e} + \kappa_1 n^{1-a_1} + \kappa_2 n^{1-a_2} + \bar{e})\kappa_u}.
\]

4. MARKOVIAN PARAMETERS WITH INFREQUENT SWITCHING

This section will be devoted to the tracking and identification of the instantaneous variation of infrequent switching Markovian parameter processes. We will perform simulation studies of such regime switching systems in the presence of all these structural uncertainties, including unmodeled dynamics, observation bias, and nonlinear model mismatch. Compared with our previous work [4, 6, 7], where we presented numerical experiments without any structural uncertainties, the tracking performance in this work would be a little bit worse. This is reasonable and unavoidable because the existence of the added bias terms will surely undermine the effect of the identification algorithm. It was shown in [4] (also [7], [6]) that an infrequent switching Markovian system can be characterized by the form of its transition probability matrix

\[
P^\Delta = I + \Delta Q,
\]

where \( I \) is the identity matrix, \( \Delta > 0 \) is a small real number, and the row sums of \( Q \) are all equal to zero. The use of (4.1) is motivated by the multiscale formulation [18]. Note that in (4.1), a value of \( \Delta \) close to zero indicates that it is a relatively infrequent switching system.

In this section, we will show that such infrequent switching systems with various structural uncertainties can be tracked with desired accuracy by using stochastic adaptive step size algorithms. To proceed, we first present the adaptive step size algorithm and then use it to carry out our simulation studies.

**4.1. Adaptive step size algorithm**

In the classical setup, to track a slow time-varying parameter process, one would choose to use the following fixed step size algorithm with step size \( \delta \): \n
\[
\hat{\theta}^\delta_{k+1} = \hat{\theta}^\delta_k + \delta \phi_k [y_k - \phi_k' \hat{\theta}^\delta_k].
\]
where \( \{y_k, \phi_k\} \) is the available sequence of outputs and regressors. Using such a fixed step size algorithm, the selection of the step size \( \delta \) depends on the details of the physical model. That is, the optimal value of the step size \( \delta \) depends not only on the ‘rates of variation’ of the true parameter \( \theta_k \) but also on the probability distribution of \( d_k, y_k, \) and \( \phi_k \). Nevertheless, owing to the lack of prior information, the exact form of the model and the relevant probability distributions are not always known in applications. Thus, an easily implementable algorithm can be built, in which in lieu of selecting a constant step size value, one chooses to construct a secondary stochastic approximation algorithm to approximate the ‘best possible step size’, in addition to the primary tracking algorithm (4.2). The adaptive-step size algorithm is then given as follows:

\[
\begin{align*}
\tilde{\theta}_{k+1} &= \tilde{\theta}_k + \delta_k \phi_k e_k, \\
\delta_{k+1} &= \Pi_{[\delta_-, \delta_+]} \left[ \delta_k + \mu \epsilon_k \phi_k' V_k \right], \\
V_{k+1} &= V_k - \delta_k c \phi_k \phi_k' V_k + \phi_k [y_k - \phi_k' \tilde{\theta}_k], \quad V_0 = 0,
\end{align*}
\]

where \( e_k = y_k - \phi_k' \tilde{\theta}_k \) is the error sequence and \( \Pi_{[\delta_-, \delta_+]} \) is the projection operator defined by

\[
\Pi_{[\delta_-, \delta_+]} z = \begin{cases} 
  z, & \text{if } z \in (\delta_-, \delta_+), \\
  \delta_-, & \text{if } z \leq \delta_-, \\
  \delta_+, & \text{if } z \geq \delta_+.
\end{cases}
\]

Note that in algorithm (4.3), \( V_k \) is the derivative \((\partial/\partial \delta) \theta_k^0\) in the mean squares sense and the recursion for \( V_k \) can be obtained by differentiating (4.2) with respect to \( \delta \) at time \( k \). Here we omit the formal discussions of adaptive step size algorithm and refer the reader to [4] and [15, Section 3.2] for further details.

4.2. Numerical experiments

To include the impacts of all the structural uncertain factors in our numerical experiments, we need to first determine the explicit forms of our unmodeled dynamics, observation bias, and nonlinear model mismatch. By (2.1) and (2.2), one could see that, in general, unmodeled dynamics term \( \tilde{\phi}_k' \tilde{\theta}_k \) should take the infinite form \( \sum_{i=0}^{\infty} a_i(k) u_{k-i} \) (bounded above by a small positive number \( \tilde{c} \)). For simplicity, in what follows, we assume that the unmodeled dynamics \( \tilde{\phi}_k' \tilde{\theta}_k \) only takes the finite form \( \epsilon_1 u_{k-n} + \epsilon_2 u_{k-n-1} \), where \( \epsilon_1 \) and \( \epsilon_2 \) are small real numbers to be determined in the following examples.

We also assume that the observation bias has the form \( b_k = (0.2 + 0.1 \sin k)(k + 2)^{-0.5} \) and nonlinear model mismatch takes the form \( \Delta_k = 3 \cos u_k (k + 1)^{-0.22} \). It is obvious that these two specific forms that we assumed here are consistent with the general model assumption (C2). To proceed, we will present several numerical examples, including one-dimensional and two-dimensional cases, to demonstrate the tracking performance of the proposed adaptive step size algorithm in our problem setting.

Example 4.1

Let the true parameter \( \{\theta_k, k = 1, 2, 3 \ldots\} \) be a discrete-time Markov chain with state space \( \mathcal{M} = \{2, 6\} \), the initial state \( \theta_0 = 6 \), and the transition probability matrix given in (4.1) with \( \Delta = 0.01 \) and \( Q = \begin{pmatrix} -5 & 5 \\ 1 & -1 \end{pmatrix} \). The input is \( u_k = 1 \), and hence \( \phi_k = 1 \). With respect to the unmodeled dynamics, here, we assume that \( \epsilon_1 = 0.1 \) and \( \epsilon_2 = -0.02 \). It implies the unmodeled dynamics term \( |\tilde{\phi}_k' \tilde{\theta}_k| \approx 0.12 \). To carry out the tracking task, we set \( \delta_+ = 0.3 \), \( \delta_- = 0.02 \), \( \mu = 0.001 \), and initial data \( \delta_0 = 0.01 \) in (4.3).

Figure 1 consists of two parts. Part (a) shows a typical sample path for the iterates \( \tilde{\theta}_k \) with the correct initial estimate, that is, \( \tilde{\theta}_0 = \theta_0 = 6 \). The figure illustrates that the tracking algorithm (4.3) tracks the true parameter very well. Part (b) shows a sample path for \( \tilde{\theta}_k \) with the wrong initial estimate, that is, \( \tilde{\theta}_0 = 2 \) but \( \theta_0 = 6 \); as illustrated by the figure, the estimate \( \tilde{\theta}_k \) catches up very fast with
Figure 1. Sample paths of $\hat{\theta}_k$ generated by using algorithm (4.3): horizontal axis, iteration number $k$; vertical axis, $\hat{\theta}_k$.

The true parameter $\theta_k$. There are only a few jumps among 1000 iterations in both cases, and this is a consequence of infrequently switching systems. The figure also shows that the time that the system parameter staying at the second state is much longer than the first state, which is a consequence of the asymmetric generator $Q$ that we applied here.

**Example 4.2**

Let $\{\theta_k, k = 1, 2, 3, \ldots\}$ be a discrete-time Markov chain with state space $\mathcal{M} = \{(2, 3)', (6, 9)\}'$, the initial state $\theta_0 = (6, 9)'$, and the transition probability matrix given by (4.1) with $\Delta = 0.01$ and $Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$.

Figure 2 part (a) shows us the tracking performance of adaptive algorithm in two-dimensional cases and part (b) gives the sample paths for each component of $\hat{\theta}_k$.
5. CONCLUDING REMARKS

This work developed persistent tracking and identification of regime-switching systems that are subject to measurement noise and structural uncertainties such as unmodeled dynamics, sensor non-linear mismatch, and bias. We considered two classes of problems. In the first class, the switching parameters are stochastic processes modeled by irreducible and aperiodic Markov chains with transition rates much faster than adaptation rates of the identification algorithms. Instead of tracking real-time parameters by output observations, we investigated the average behavior of the parameter process. Identification error bounds were established and analyzed for their dependence on these structural uncertainties. In the second class of problems, the system parameters vary infrequently. An adaptive algorithm with variable step sizes was introduced for tracking the time-varying parameters. Numerical results were presented to illustrate the performance of the algorithm.

This work focuses on identification algorithms. For references on identification, we refer to the work, [19–21], and references therein. Future work can be pursued in several directions. In this paper, we focus on tracking and system identification for plants that are equipped with regular sensors. One may also investigate the identification problems for such systems equipped with only binary sensors, which is valuable because binary sensors are commonly employed in practical systems, and they are far more cost effective than regular sensors [22]. Furthermore, this paper is confined to open-loop identification problems. One may also consider the associated closed-loop systems in conjunction with the approach provided in [3]. Investigating impacts of noise, unmodeled dynamics, observation bias, and model mismatch on such regime-switching systems in the closed-loop setting will be beneficiary from both theoretical and practical considerations.

APPENDIX A: PROOFS OF RESULTS

Proof of Lemma 3.4

Note that the input $u$ is the $n$-periodic signal with the first $n$ components $1, 0, \ldots, 0$. Direct computation shows that

$$
\left( \sum_{i=0}^{N-1} \phi_{l0+i} \phi'_{l0+i} \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{l0+i} b_{l0+i} \right) = (\bar{b}_N, \bar{b}_{N-1}, \ldots, \bar{b}_1)'$

where

$$\bar{b}_j = \frac{1}{\kappa} \sum_{i=1}^{\kappa} b_{l0+i n-j}, \text{ for } j = 1, \ldots, n;$$

see [2] for detail. Thus, we obtain

$$
\left\| \left( \sum_{i=0}^{N-1} \phi_{l0+i} \phi'_{l0+i} \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{l0+i} b_{l0+i} \right) \right\|_1 = \sum_{j=1}^{n} |\bar{b}_j| . \tag{A.1}
$$

By assumption (C3), we derive that

$$
\sum_{j=1}^{n} |\bar{b}_j| \leq \frac{1}{\kappa} \sum_{j=1}^{n} \sum_{i=1}^{\kappa} \kappa_1 (i n - j + 1)^{-\alpha_1} .
$$

Note that for each \( j = 1, 2, \ldots, n \), \((in - j + 1)^{-\alpha_1} \leq (in - n + 1)^{-\alpha_1}\). Detailed calculation reveals that

\[
\sum_{j=1}^{n} |\tilde{\theta}_j| \leq \frac{\kappa_1 n^{-\alpha_1}}{\kappa} \sum_{i=1}^{\kappa} \left( i - \frac{n-1}{n} \right)^{-\alpha_1} \\
\leq \frac{\kappa_1 n}{\kappa} + \frac{\kappa_1 n^{-\alpha_1}}{\kappa} \sum_{i=2}^{\kappa} \left( i - \frac{n-1}{n} \right)^{-\alpha_1} \\
\leq \frac{\kappa_1 n}{\kappa} + \frac{\kappa_1 n^{-\alpha_1}}{\kappa} \frac{\kappa - 1}{\kappa}.
\]

The last inequality holds because \( i - \frac{n-1}{n} \geq 1 \) for \( i = 2, 3, \ldots, \kappa \). By (A.1), it is easy to see that result (3.4) follows as \( \kappa \to \infty \). The desired result (3.5) can be derived in the similar way.

**Proof of Theorem 3.5**

Direct computation yields that

\[
\tilde{\theta}_{i_0 + N - 1} - \bar{\theta} = \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \left( \tilde{\theta}_{i_0+i} - \bar{\theta} \right) \right)
\]

\[
+ \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \tilde{\phi}_{i_0+i}^t \tilde{\theta}_{i_0+i} \right)
\]

\[
+ \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \Delta_{i_0+i} \right)
\]

\[
+ \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} d_{i_0+i} \right).
\]

Define

\[
\eta_d = \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \tilde{\phi}_{i_0+i}^t \tilde{\theta}_{i_0+i} \right)
\]

and

\[
\eta_s = \left( \sum_{i=0}^{N-1} \phi_{i_0+i} \phi_{i_0+i}^t \right)^{-1} \left( \sum_{i=0}^{N-1} \phi_{i_0+i} d_{i_0+i} \right),
\]

where \( \eta_d \) and \( \eta_s \) denote the deterministic (associated with unmodeled dynamics) and stochastic parts of identification errors, respectively.

By selecting the input \( u \) to be the \( n \)-periodic signals with the first \( n \) components being \( 1, 0, \ldots, 0 \), one can obtain

\[
\sup_{|\tilde{\theta}| \leq \tilde{\varepsilon}} \| \eta_d \|_1 = \tilde{\varepsilon} \quad \text{and} \quad \sup_{0 \leq i_0 < \infty} \| \eta_s \|_1 = \sum_{j=1}^{n} |\tilde{d}_j|,
\]

where

\[
\tilde{d}_j = \frac{1}{\kappa} \sum_{i=0}^{\kappa} d_{i_0+i-n-j}, \quad \text{for} \quad j = 1, \ldots, n;
\]

\[\text{DOI: 10.1002/acs}\]
see [2, 4] for details.

Using (A.3), (3.4), and (3.5), we have the following inequality with regard to the identification error:

\[
\| \hat{\theta}_{t_0+N-1} - \overline{\theta} \|_1 \leq \left\| \sum_{i=0}^{N-1} \frac{\phi_{t_0+i} \phi'_{t_0+i}}{\sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} (\theta_{t_0+i} - \overline{\theta})} \right\|_1 \nonumber
\]

\[+ \overline{\epsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \sum_{j=1}^{n} |\overline{d}_j| \nonumber.
\]

Note that the preceding inequality also implies that for any \( \varepsilon > 0 \), the following set inclusion holds

\[
\{ (\theta_{t_0}, \ldots, \theta_{t_0+N-1}, d_0, \ldots, d_{t_0+N-1}) : \| \hat{\theta}_{t_0+N-1} - \overline{\theta} \|_1 \geq \overline{\epsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon \} \nonumber
\]

\[
\subseteq \{ (\theta_{t_0}, \ldots, d_{t_0+N-1}) : \| \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} \|_1 \left( \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} (\theta_{t_0+i} - \overline{\theta}) \right) \|_1 + \sum_{j=1}^{n} |\overline{d}_j| \geq \varepsilon \} \nonumber.
\]

Thus, we have

\[
P \left( \| \hat{\theta}_{t_0+N-1} - \overline{\theta} \|_1 \geq \overline{\epsilon} + \kappa_1 n^{1-\alpha_1} + \kappa_2 n^{1-\alpha_2} + \varepsilon \right) \nonumber
\]

\[
\leq P \left( \left\| \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} \right\|_1 \left( \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} (\theta_{t_0+i} - \overline{\theta}) \right) \|_1 + \sum_{j=1}^{n} |\overline{d}_j| \geq \varepsilon \right) \nonumber
\]

\[
\leq P \left( \left\| \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} \right\|_1 \left( \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} (\theta_{t_0+i} - \overline{\theta}) \right) \|_1 \geq \varepsilon/2 \right) + P \left( \sum_{j=1}^{n} |\overline{d}_j| \geq \varepsilon/2 \right) \nonumber.
\]

It has been shown in [4] that under assumptions (C1)–(C4), we obtain the inequalities

\[
P \left( \left\| \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} \right\|_1 \left( \sum_{i=0}^{N-1} \phi_{t_0+i} \phi'_{t_0+i} (\theta_{t_0+i} - \overline{\theta}) \right) \|_1 \geq \varepsilon/2 \right) \nonumber
\]

\[
\leq 2m \exp \left( - \frac{\varepsilon^2 N}{2 \sigma^2 m^2 c_1^2} \right) \quad \text{as} \quad N \to \infty, \quad (A.7)
\]

and

\[
P \left( \sum_{j=1}^{n} |\overline{d}_j| \geq \varepsilon/2 \right) \leq n P \left( |\overline{d}_1| \geq \frac{\varepsilon/2}{n} \right) \leq 2n \left[ g \left( \frac{\varepsilon}{2n} \right) \right]^\kappa, \quad (A.9)
\]

where \( c_1 \) and \( \sigma^2 \) are given by (3.7). Thus, (3.6) follows by noting the definition of \( Q_{\hat{\theta},N}^0 \) and combining (A.7), (A.8), and (A.9) together.

**Proof of Lemma 3.8**

(i) Consider the total estimation error \( \hat{\theta}_{t_0+N-1} - \overline{\theta} = \eta_m + \eta_d + \eta_b + \eta_\Delta + \eta_s \). It has been shown in [2] that with the condition input \( u \in \mathcal{S} \), we can have one upper bound on \( \eta_d \), that is, \( \sup_{\| \hat{\theta}_t \|_1 \leq \overline{\theta}} \| \eta_d \|_1 = \overline{\theta} \). Thus, the following inequalities hold

\[
\| \hat{\theta}_{t_0+N-1} - \overline{\theta} \|_1 \geq \| \eta_m \|_1 - \| \eta_d \|_1 - \| \eta_b \|_1 - \| \eta_\Delta \|_1 \nonumber
\]

\[\geq \| \eta_m \|_1 - \| \eta_s \|_1 - \overline{\theta} - \kappa_1 n^{1-\alpha_1} - \kappa_2 n^{1-\alpha_2}. \quad (A.10)
\]
The last inequality earlier follows directly from Lemma 3.4. As a result, because $\xi_d$ contains only 0 or $\pm 1$
\[ \|\eta_m + \eta_s\|_1 \geq |(\eta_m + \eta_s)'\xi_d| \geq |\eta_m'\xi_d| - |\eta_s'\xi_d|. \] (A.11)

Thus, it follows from (A.10) and (A.11) that a sufficient condition for $\|\hat{\theta}_{i_0+N-1} - \hat{\theta}\|_1 \geq \bar{\epsilon} + \kappa_1 n^{1 - \alpha_1} + \kappa_2 n^{1 - \alpha_2} + \varepsilon$ is $|\eta_m'\xi_d| \leq \varepsilon$ and $|\eta_s'\xi_d| \geq \bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon$.

**Proof of Theorem 3.9**

It is readily seen from (A.10) and Lemma 3.8 (ii) that
\[ Q^0_{(\bar{\epsilon}, N)} (\bar{\epsilon} + \kappa_1 n^{1 - \alpha_1} + \kappa_2 n^{1 - \alpha_2} + \varepsilon) \geq P \left( |\lambda' D| \geq 2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon, |\eta_m'\xi_d| \leq \varepsilon \right) \] (A.12)
\[ = P \left( |\lambda' D| \geq 2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon \right) P \left( |\eta_m'\xi_d| \leq \varepsilon \right) \]
\[ + P \left( |\lambda' D| \leq -2\bar{\epsilon} - 2\kappa_1 n^{1 - \alpha_1} - 2\kappa_2 n^{1 - \alpha_2} - 2\varepsilon \right) P \left( |\eta_m'\xi_d| \leq \varepsilon \right). \]

One can also show that
\[ P \left( |\lambda' D| \geq 2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon \right) \geq \frac{P \left( d \geq (2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon)|d \geq 0) \right.}{2} \] (A.13)
and
\[ P \left( |\lambda' D| \leq -2\bar{\epsilon} - 2\kappa_1 n^{1 - \alpha_1} - 2\kappa_2 n^{1 - \alpha_2} - 2\varepsilon \right) \geq \frac{P \left( d \geq (2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon)|d \geq 0) \right.}{2}. \] (A.14)

The proofs of (A.13) and (A.14) are similar to [2, Theorem 4]. Moreover, because $\|\eta_m\|_1 \geq |\eta_m'\xi_d|$, we can have
\[ P \left( |\eta_m'\xi_d| \leq \varepsilon \right) = 1 - P \left( |\eta_m'\xi_d| \geq \varepsilon \right) \geq 1 - P \left( |\eta_m|_1 \geq \varepsilon \right) \] (A.15)
\[ \geq 1 - 2m \exp \left( -\frac{2\bar{\epsilon}^2 N}{\sigma^2 d^2 c_1^2} \right). \]

Here, we have used (A.8) in the last inequality. Therefore, the desired result follows from (A.12)–(A.15).

**Proof of Corollary 3.11**

By Lemma 3.8, there exists a $\lambda \in \mathbb{R}^N$ such that
\[ Q^0_{(\bar{\epsilon}, N)} (\bar{\epsilon} + \kappa_1 n^{1 - \alpha_1} + \kappa_2 n^{1 - \alpha_2} + \varepsilon) \geq P \left( |\lambda' D| \geq 2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon, |\eta_m'\xi_d| \leq \varepsilon \right) \] (A.16)
\[ = 2P \left( |\lambda' D| \geq 2\bar{\epsilon} + 2\kappa_1 n^{1 - \alpha_1} + 2\kappa_2 n^{1 - \alpha_2} + 2\varepsilon \right) P \left( |\eta_m'\xi_d| \leq \varepsilon \right). \]

Note that the last equality holds because of the fact that $\lambda' D$ is normally distributed with mean 0 and variance $\lambda' \lambda \sigma_d^2$. Moreover, we can derive that the variance $\lambda' \lambda \sigma_d^2 \geq \left( 1/\kappa_0^2 N \right) \sigma_d^2$ because $\|\lambda\|_1 \geq 1/\kappa_0$ and $\|\lambda\|_2^2 \leq N \lambda' \lambda$. Thus, we have
\[ Q^0_{(\bar{\epsilon}, N)} (\bar{\epsilon} + \kappa_1 n^{1 - \alpha_1} + \kappa_2 n^{1 - \alpha_2} + \varepsilon) \geq 2P \left( \frac{|\lambda' D|}{\sqrt{\lambda' \lambda \sigma_d}} \geq \sqrt{\frac{N}{L}} \right) P \left( |\eta_m'\xi_d| \leq \varepsilon \right) \] (A.17)
\[ = 2 \left( 1 - \Phi \left( \sqrt{\frac{N}{L}} \right) \right) P \left( |\eta_m'\xi_d| \leq \varepsilon \right), \]
where $\Phi(\cdot)$ is the standard normal distribution function. Then the desired result follows from (A.15) and [23, Lemma 2, p.175].

ACKNOWLEDGEMENTS
Research of Shaobai Kan was supported in part by PSC-CUNY Research Award # 60135-39 40. Research of George Yin was supported in part by the National Science Foundation under DMS-0907753. Research of Le Yi Wang was supported in part by the National Science Foundation under CNS-1136007.

REFERENCES