Continuous-Time Tracking Algorithms Involving Two-Time-Scale Markov Chains

George Yin*, Fellow, IEEE*, Qing Zhang, *Senior Member, IEEE*, John B. Moore, *Fellow, IEEE*, and Yuan Jin Liu

Abstract—This work is concerned with least-mean-squares (LMS) algorithms in continuous time for tracking a time-varying parameter process. A distinctive feature is that the true parameter process is changing at a fast pace driven by a finite-state Markov chain. The states of the Markov chain are divisible into a number of groups. Within each group, the transitions take place rapidly; among different groups, the transitions are infrequent. Introducing a small parameter into the generator of the Markov chain leads to a two-time-scale formulation. The tracking objective is difficult to achieve. Nevertheless, a limit result is derived yielding algorithms for limit systems. Moreover, the rates of variation of the tracking error sequence are analyzed. Under simple conditions, it is shown that a scaled sequence of the tracking errors converges weakly to a switching diffusion. In addition, a numerical example is provided and an adaptive step-size algorithm developed.

Index Terms—Adaptive filtering, continuous-time Markov chain, two-time scale.

I. INTRODUCTION

THIS paper is concerned with least-mean-squares (LMS) algorithms in continuous time for tracking a time-varying parameter process, which is under the influence of a continuous-time Markov chain. For analyzing LMS algorithms with time-varying parameters, by assuming that the parameter varies continuously but slowly over time with small amount of changes, the performance of tracking algorithms is studied in [1], [5], [9], [11], [12], [16], among others. Most literature has focused on discrete-time problems (see [1], [4], [11], [12], [17], and the references therein). In contrast, our method here deals with rapidly changing parameters in a continuous-time setting. Our results are good approximations to discrete-time algorithms when the sampling frequency is very high. Note that it is important to examine algorithms when the sampling rate becomes very high (see the motivation and description given in [18] for least-squares estimation schemes for continuous-time systems).

G. Yin and Y. J. Liu are with the Department of Mathematics, Wayne State University, Detroit, MI 48202 USA (e-mail: gyin@math.wayne.edu; yuanjin@math.wayne.edu).

Q. Zhang is with the Department of Mathematics, University of Georgia, Athens, GA 30602-7403 USA (e-mail: qingz@math.uga.edu).

J. B. Moore is with the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia (e-mail: john.moore@anu.edu.au).

Although least-squares-type algorithms have been studied extensively and applied to a wide range of problems in control, optimization, learning, and related fields, emerging applications have been found in signal processing, wireless communication, and network problems (see [7], [14], and [15]). These applications, in turn, require in-depth understanding and further development of more efficient procedures.

To the best of the authors' knowledge, most of the existing results are concerned with slowly varying parameter processes, whereas not as much attention has been drawn for systems with fast variations. In this paper, we consider the parameter process under a regime switching and subject to fast variations. The regime switching is modeled as a Markovian time-varying dependence and involves two time scales. Suppose that the Markov chain has a finite-state space whose states are divided into a number of recurrent groups so that within each recurrent group, the transitions take place frequently, and among different groups, the transitions occur less frequently. This time-scale separation is one of the novelties of this paper. The rationale is: For a fast-changing jump process, the tracking aim or computation objective is difficult to achieve. Nevertheless, due to the features of the fast-varying jumps in relation to the slow-changing jumps, we may consider a computationally simpler limit system associated with much slower jump changes. In lieu of the original tracking algorithms, we can use the limit systems to carry out the investigation.

Once the limit system is obtained, the next question of interest is, How does the tracking error of the system evolve as a function of time? We handle the rate of variation problem by means of a weak convergence approach. We show that a properly scaled sequence of tracking errors converges to a nonzero stochastic process limit. Distinct from the usual analysis of tracking problems, the limit is not a diffusion process but a switching diffusion or diffusions modulated by a continuous-time Markov chain. The limit dynamic system delineates the evolution of the scaled tracking error process. By working with a joint pair of processes, the proof of this result uses a combination of martingale averaging and perturbed test function methods and is interesting in its own right. Finally, we use the limit tracking algorithm and the idea of step-size adaptation to build adaptive modifications leading to efficient performance.

The remainder of the paper is organized as follows. Section II provides the formulation and the algorithm. Asymptotic properties of the algorithm are presented in Section III. Using the recently developed results on two-time-scale Markov chains [20], [21], we derive algorithms for certain limit systems, which are also of LMS type. The main techniques are the weak convergence methods. Section IV is devoted to the asymptotic anal-

ysis of the tracking error sequence. Based on the limit system, Section V proceeds with the design of adaptive step-size algorithms. Further remarks are given in Section VI to conclude the paper. Technical details and proofs are deferred to an appendix to better preserve the flow of presentation. Throughout the paper, we use z' to denote the transpose of $z \in \mathbb{R}^{i \times j}$ for $i, j \geq 1$ and use $|z|$ to denote the norm of z. For notational simplicity, K denotes a generic positive constant whose values may vary for different usages. For a square matrix B, by $B > 0$, we mean that it is positive definite.

II. TWO-TIME-SCALE LMS TRACKING ALGORITHM

Suppose that $\{y(t)\}$ is a real-valued sequence of signals with $y(t)$ representing the observation obtained at time t and that $\{\theta(t)\}$ is a time-varying true parameter, an $\mathbb{R}^r$-valued random process. Assume the true parameter is observed in noise as

$$y(t) = \varphi'(t)\theta(t) + e(t) \qquad (1)$$

where $\varphi(t) \in \mathbb{R}^r$ is the regression vector and $e(t) \in \mathbb{R}$ represents a zero-mean observation noise. We construct an algorithm so that weighted tracking errors are made as small as possible in an appropriate sense. Here, we use the mean-squares tracking error criterion. That is, we choose Θ so that $E|y(t) - \varphi'(t)\Theta|^2$ is minimized.

The model given by (1) stems from a large class of applications in signal process and adaptive estimation. It arises from noise cancellation, adaptive qualization, and adaptive filter among others. (For various such processes in discrete time, see [1] and references therein.) The model also arises in tracking performance of an adaptive linear multiuser detector in a cellular digital signal/code-division multiple-access (DS/CDMA) wireless network when the profile of active users changes due to an admission or access (scheduling) controller at the base station. A discrete-time version of a finite-state Markovian model for a linear minimum-mean-square-error (LMMSE) multiuser detector can be found in [19]. The continuous-time signal model describes the case when the sampling frequency is high. We assume the parameter process $\theta(t) = \widehat{\theta}(t, \alpha(t))$, where $\widehat{\theta}(\cdot)$ is a function of (t, α) and $\alpha(t)$ is a continuous-time Markov chain. Note that if $\varphi(t) \equiv 1$ for all $t \geq 0$, then it becomes a hidden Markov model in continuous time. Our tasks are to track the time-varying parameter and to figure out the bounds of the tracking errors.

In many applications, the state space of the Markov chain is inevitably large. This raises the issue and concern on computation complexity. One of the ways to reduce the complexity is to use a two-time-scale model. The rationale is that in a large system, not all subsystems or components or parts or states change at the same speed. Some of them vary rapidly, and others change relatively slowly. Taking advantage of the contrast of the different speeds of variation, we can put the states with similar changing rate into one group. Then, we arrive at a model with a total of number of groups much less than the total number of states of the Markov chain. Mathematically, this is done as follows. We introduce a small parameter $\varepsilon > 0$ to highlight different rates of changes. Then, $\alpha(t) = \alpha^\varepsilon(t)$, which we assume

to be a stationary or time-homogeneous Markov chain with finite state space

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_l \qquad (2)$$

where $\mathcal{M}_i = \{a_1^i, \ldots, a_{m_i}^i\}$, and generator

$$Q^\varepsilon = \frac{1}{\varepsilon}\mathrm{diag}(\widetilde{Q}^1, \ldots, \widetilde{Q}^l) + \widehat{Q} \qquad (3)$$

such that $\widehat{Q}$ and $\widetilde{Q}^i$ (for each $i = 1, \ldots, l$) are themselves generators (see [20, pp. 47–49] for a procedure of obtaining a generator of such a canonical form (3)). In the above, $\mathrm{diag}(A^1, \ldots, A^l)$ denotes a block diagonal matrix with entries A^1 through A^l of appropriate dimensions. Note that in the asymptotic analysis, we let $\varepsilon \to 0$. In the actual application, ε is merely a constant that highlights the different rates of variations among the states. The idea is, to achieve the reduction of complexity, we lump all the states in each group into one state, and then the process becomes one with only l states. If l, the number of groups, is much less than the total number of Markov states, a substantial reduction of effort is achieved.

With the signal model given above, to track the parameter $\{\theta(t)\}$, we construct $\{\Theta^\varepsilon(t)\}$ which depends on the scale parameter ε. The algorithm takes the form

$$\dot{\Theta}^\varepsilon(t) = \mu\varphi(t)(y(t) - \varphi'(t)\Theta^\varepsilon(t)), \ \Theta^\varepsilon(0) = \Theta_0 \qquad (4)$$

where $\mu > 0$ is a small constant step size of the algorithm.

A0) The parameter process is $\theta(t) = \widehat{\theta}(t, \alpha^\varepsilon(t))$, where $\varepsilon > 0$, $\alpha^\varepsilon(\cdot)$ is a continuous-time Markov chain with generator Q^ε given by (3), and for each α, $\widehat{\theta}(\cdot, \alpha)$ is a continuous function.

A1) For each $i = 1, \ldots, l$, $\widetilde{Q}^i$ is irreducible.

A2) The $\{\varphi(t), e(t)\}$ is a sequence of bounded signals that is stationary and independent of $\alpha^\varepsilon(t)$ such that $\left|\int_t^\infty E[(\varphi(s)\varphi'(s) - A)|\mathcal{F}_t^{\varphi,e}]ds\right| \leq K$, and in $\left|\int_t^\infty E[\varphi(s)e(s)|\mathcal{F}_t^{\varphi,e}]ds\right| \leq K$, for a symmetric and positive definite matrix $A \in \mathbb{R}^{r \times r}$, where $\mathcal{F}_t^{\varphi,e}$ denotes the σ-algebra generated by $\{\varphi(u), e(u) : u \leq t\}$.

Remark 2.1: Condition A1) specifies the structure of the underlying Markov model. Assumption A2) is essentially a condition on the mixing property of the signals. The inequalities in A2) are modeled after the well-known mixing inequalities (see [8, p. 82]). They indicate that although correlated, the signals are asymptotically independent, and the correlations decay in a certain rates.

III. TRACKING ALGORITHM FOR THE LIMIT SYSTEM

A. Preliminary Results

First, recall the definition of weak convergence. Suppose that X_k and X are $\mathbb{R}^r$-valued random variables. X_k is said to converge weakly to X if for any bounded and continuous function $\widetilde{g}(\cdot)$, $E\widetilde{g}(X_k) \to E\widetilde{g}(X)$ as $k \to \infty$; $\{X_k\}$ is said to be tight if for each $\eta > 0$, there is a compact set K_η such that $P(X_k \in K_\eta) \geq 1 - \eta$ for all k. The above definitions extend to random variables in a metric space. The notion of weak convergence is a substantial generalization of convergence in distribution. On a complete separable metric space, tightness is equiv-

alent to sequential compactness, which is known as Prohorov's theorem. Due to this theorem, we can extract convergent subsequences once tightness is verified. For terminologies and results of weak convergence such as Skorohod topology, Skorohod representation, and Prohorov's theorem, etc., we refer the reader to [11, Ch. 7].

For our signal model, we aggregate the states of the Markov chain $\alpha^\varepsilon(\cdot)$ according to the following rule. Define $\overline{\alpha}^\varepsilon(t) = i$ if $\alpha^\varepsilon(t) \in \mathcal{M}_i$. Generally, the aggregated process $\overline{\alpha}^\varepsilon(\cdot)$ is not necessarily Markov, but it yields a Markovian limit process. The proof of the first assertion can be found in [20, Th. 7.4] (see also [21] for time-varying generators being measurable only). The proofs of the second and the third assertions are modifications of that of [20, Th. 7.2 and Lemma 7.12] with $\gamma(\tau)$ being random but independent of $\alpha^\varepsilon(t)$. The details are omitted. Using these results, we then study asymptotic properties of the tracking algorithm.

Lemma 3.1: Under A0) and A1), the following assertions hold.

- $\overline{\alpha}^\varepsilon(\cdot)$ converges weakly to $\overline{\alpha}(\cdot)$, a Markov chain with state space $\overline{\mathcal{M}} = \{1, \ldots, l\}$ and generator

$$\overline{Q} = \mathrm{diag}(\nu^1, \ldots, \nu^l)\widehat{Q}\mathrm{diag}(\mathbb{1}_{m_1}, \ldots, \mathbb{1}_{m_l}) \qquad (5)$$

  where for each $i = 1 \ldots, l$, ν^i is the stationary distribution corresponding to the generator $\widetilde{Q}^i$, and $\mathbb{1}_{m_i}$ is a column vector in $\mathbb{R}^{m_i \times 1}$ with all components being 1.

- For each $i = 1, \ldots, l$ and $j = 1, \ldots, m_i$, and for any bounded and measurable process $\gamma(t)$ that is independent of $\alpha^\varepsilon(t)$

$$\sup_{0 \leq t \leq T} E \left| \int_0^t \gamma(s)[I_{\{\alpha^\varepsilon(s)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(s)=i\}}]ds \right|^2 = O(\varepsilon)$$

  where ν_j^i is the jth component o ν^i, and $\gamma(\cdot)$ can be either a scalar or a vector-valued process.

- Define $\widetilde{\eta}^\varepsilon(t)$ by

$$\widetilde{\eta}^\varepsilon(t) = E \left(\int_t^{t+s} \gamma(\tau)[I_{\{\alpha^\varepsilon(\tau)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(\tau)=i\}}]d\tau \Big| \widetilde{\mathcal{F}}_t \right)^2$$

  where $\widetilde{\mathcal{F}}_t$ is the σ-algebra generated by $\{\alpha^\varepsilon(u), \gamma(u) : u \leq t\}$. Then, $\widetilde{\eta}^\varepsilon(t)/\varepsilon = O(s)$.

B. Moment Bounds

We use the perturbed Liapunov function method (see [11]) to prove the following moment bounds. The basic idea is to construct small perturbation of the Liapunov function associated with the differential equation to enable cancellation of unwanted terms due to correlated signals.

Theorem 3.2: Define $V(\Theta) = \Theta'\Theta/2$. *Assume* A0)–A2). Then, $EV(\Theta^\varepsilon(t)) = O(1)$.

Remark 3.3: Note that the theorem above is a stability result. A direct consequence of Theorem 3.2 is as follows: For $T > 0$, $\sup_{0 \leq t \leq T} E|\Theta^\varepsilon(t)|^2 = O(1)$.

C. Weak Convergence

Theorem 3.4: Assume A0)–A2). Then, for each $\mu > 0$, as $\varepsilon \to 0$, $(\Theta^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(\Theta(\cdot), \overline{\alpha}(\cdot))$ in which $\overline{\alpha}(\cdot)$ is a reduced Markov chain generated by $\overline{Q}$ given in (5) and $\Theta(\cdot)$ is a solution of

$$\dot{\Theta}(t) = \mu\varphi(t)(\overline{y}(t) - \varphi'(t)\Theta(t)), \ \Theta(0) = \Theta_0 \qquad (6)$$

where

$$\overline{y}(t) = \varphi'(t)\overline{\theta}(t, \overline{\alpha}(t)) + e(t),$$
$$\overline{\theta}(t, i) = \sum_{j=1}^{m_i} \nu_j^i \widehat{\theta}(t, a_j^i), \ \text{ for } \ i = 1, \ldots, l. \qquad (7)$$

Remark 3.5: A particular form of the function $\widehat{\theta}(\cdot)$ that we have in mind is $\widehat{\theta}(t, \alpha) = \widehat{\theta}(\alpha)$. That is, it assigns a value to each of the Markovian states. Nevertheless, our formulation allows us to deal with more general setup.

The Markov chain $\alpha^\varepsilon(\cdot)$ is fast varying; it may not be sensible from a computational point of view to track its variation at any given instance since its value will jump to another state in a short duration. Nevertheless, due to the above limit result, the system under consideration is closely related to a limit system, in which the parameter and the observations are that of the original system averaged out with respect to the stationary distributions. Thus, in lieu of the original system, we can examine the limit system.

Define

$$\overline{y}^\varepsilon(t) = \varphi'(t)\overline{\theta}(t, \overline{\alpha}^\varepsilon(t)) + e(t).$$

The corresponding algorithm with $\overline{y}^\varepsilon(t)$ used is

$$\frac{d}{dt}\overline{\Theta}^\varepsilon(t) = \mu\varphi(t)(\overline{y}^\varepsilon(t) - \varphi'(t)\overline{\Theta}^\varepsilon(t)), \ \overline{\Theta}^\varepsilon(0) = \Theta_0.$$

The proof of the following lemma is essentially the same as that of Theorem 3.2 and is omitted. The proof of the corollary is relegated to the appendix .

Lemma 3.6: Under the conditions of Theorem 3.4, $EV(\overline{\Theta}^\varepsilon(t)) = O(1)$.

Corollary 3.7: Under the conditions of Theorem 3.4, for $0 < T < \infty$, $\sup_{0 \leq t \leq T} E|\Theta^\varepsilon(t) - \overline{\Theta}^\varepsilon(t)|^2 \to 0$ as $\varepsilon \to 0$.

Remark 3.8: The results obtained thus far are under the conditions that the Markov chain $\alpha^\varepsilon(t)$ is time homogeneous. These results can be extended to time-inhomogeneous Markov chains, in which the notation of weak irreducibility and quasi-station distribution are needed (see [20, Ch. 2] for definitions of weak irreducible and quasi-stationary distributions). We present the following result but omit the verbatim proof, which can be carried out by modifying the argument for that of the time-homogeneous case.

Theorem 3.9: Assume A0)–A2) with the following modifications. Suppose that the generators in (3) all depend on time. Suppose that for each i, $\widetilde{Q}^i(t)$ is weakly irreducible. Then, the conclusions of Lemmas 3.1 and 3.6, Theorems 3.2, 3.4, and 3.7, and Corollary 3.7 continue to hold with ν^i and $\overline{Q}$ replaced by time-dependent $\nu^i(t)$ and $\overline{Q}(t)$, respectively.

IV. RATES OF TRACKING ERROR VARIATION

This section is devoted to the analysis of tracking errors. The desired rate results are obtained by a combined approach of martingale averaging and perturbed test functions for a pair of processes. The proof is different from our previous paper [20] and is interesting in its own right.

Define $\widetilde{\Theta}^\varepsilon(t) = \Theta^\varepsilon(t) - \overline{\Theta}^\varepsilon(t)$. Then, we have $\widetilde{\Theta}^\varepsilon(0) = 0$ and

$$
\begin{aligned}
&\frac{d}{dt}\widetilde{\Theta}^\varepsilon(t) \\
&= \mu\varphi(t)(y(t) - \overline{y}(t)) - \mu\varphi(t)\varphi'(t)\widetilde{\Theta}^\varepsilon(t) \\
&= \mu\sum_{i=1}^{l}\sum_{j=1}^{m_i}\varphi(t)\varphi'(t)\widehat{\theta}(t, a_j^i)[I_{\{\alpha^\varepsilon(t)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(t)=i\}}] \\
&\quad - \mu\varphi(t)\varphi'(t)\widetilde{\Theta}^\varepsilon(t).
\end{aligned}
$$

Define a sequence of scaled tracking errors $u^\varepsilon(t) = \widetilde{\Theta}^\varepsilon(t)/\sqrt{\varepsilon}$. Then, $u^\varepsilon(t)$ satisfies

$$
\begin{aligned}
u^\varepsilon(t) &= \mu\int_0^t \varphi(\tau)\varphi'(\tau)\frac{\widehat{\theta}(\tau, \alpha^\varepsilon(\tau) - \overline{\theta}(\tau, \overline{\alpha}^\varepsilon(\tau))}{\sqrt{\varepsilon}} \\
&\quad - \mu\int_0^t \varphi(\tau)\varphi'(\tau)u^\varepsilon(\tau)d\tau \\
&= \mu\int_0^t \varphi(\tau)\varphi'(\tau)\sum_{i=1}^{l}\sum_{j=1}^{m_i}\widehat{\theta}(t, a_j^i) \\
&\quad \cdot \left(\frac{I_{\{\alpha^\varepsilon(\tau)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(\tau)=i\}}}{\sqrt{\varepsilon}}\right)d\tau \\
&\quad - \mu\int_0^t \varphi(\tau)\varphi'(\tau)u^\varepsilon(\tau)d\tau. \quad (8)
\end{aligned}
$$

We proceed to prove that $\{u^\varepsilon(\cdot)\}$ has a weak limit that is a switching diffusion process. The following theorem reveals the rates of tracking error variations.

Theorem 4.1: Under the conditions of Theorem 3.4, $(u^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(u(\cdot), \overline{\alpha}(\cdot))$, a switching diffusion process such that $u(\cdot)$ is a solution of

$$
u(t) = -\mu\int_0^t \varphi(\tau)\varphi'(\tau)u(\tau)d\tau + \mu\int_0^t \overline{\Sigma}(\tau)dw(\tau) \quad (9)
$$

$\overline{\Sigma}(t) = (1/2)(\widetilde{\Sigma}(t) + \widetilde{\Sigma}'(t))$, and $\widetilde{\Sigma}(t)/2$ is the limit in probability of $(1/\varepsilon)\int_t^T E_t w(\tau, \alpha^\varepsilon(\tau))w'(t, \alpha^\varepsilon(t))d\tau$.

Note that $\widetilde{\Sigma}(t)$ depends on $\overline{\alpha}(t)$. Up to now, we have used Assumption A0) to specify the parameter process. The result obtained can be generalized to the case of "hidden" Markovian driving models. We present the corresponding results in the following theorem with the detailed proofs omitted since the proofs are essentially the same as before.

Theorem 4.2: Assume the conditions of Theorem 3.4 with the following modification: In A0), assume $\theta(t) = \widehat{\theta}(t, \alpha^\varepsilon(t)) + \widetilde{\zeta}(t)$, where $\widetilde{\zeta}(t)$ is stationary and is independent of $\alpha^\varepsilon(\cdot)$, $\varphi(\cdot)$, and $e(\cdot)$ such that $E\widetilde{\zeta}(t) = 0$ and $E|\widetilde{\zeta}(t)|^{2+\gamma} < \infty$. Then, the conclusions of Theorems 3.2, 3.4, 3.9, and 4.1 continue to hold.

V. EXAMPLES

A. Numerical Results

In this section, we give simulation results. Suppose the continuous-time Markov chain $\alpha^\varepsilon(\cdot)$ has four states $\mathcal{M} = \{1, 2, 3, 4\}$ and the generator Q^ε is given by (3), where

$$
\widetilde{Q} = \mathrm{diag}(\widetilde{Q}^1, \widetilde{Q}^2), \quad \widetilde{Q}^1 = \begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix}
$$

$$
\widetilde{Q}^2 = \begin{pmatrix} -3 & 3 \\ 4 & -4 \end{pmatrix}, \quad \widehat{Q} = \begin{pmatrix} -3 & 1 & 1 & 1 \\ 2 & -4 & 1 & 1 \\ 1 & 2 & -6 & 3 \\ 1 & 2 & 2 & -5 \end{pmatrix}.
$$

Consider

$$
\begin{aligned}
\varphi(t) &= 5e^{-0.1t} + 10(1 - e^{-0.1t})\sqrt{0.1}N(0, 1) \\
\theta(t) &= \widehat{\theta}(t, \alpha^\varepsilon(t)) = \alpha^\varepsilon(t), \ e(t) = \sqrt{0.1}N(0, 1).
\end{aligned}
$$

Fig. 1(a)–(c) shows the sample paths of the second components $\Theta_2^\varepsilon(t)$ versus that of $\Theta_2(t)$. All other components display similar behavior. the horizontal axis represents the time t. As ε is getting smaller, the sample paths of the original systems are well approximated by that of the averaged (or limit) systems. Although the boundedness condition of the signal $(\varphi(t), e(t))$ is violated, the tracking algorithm still works well. Fig. 1(d) displays the sample paths of $u_2^\varepsilon(t)$.

B. Adaptive Step-Size Algorithm

In this section, we design an adaptive step-size algorithm in continuous time. The motivation comes from previous work reported in [1] and [2]. The idea of step-size adaptation is suggested in [1], in which one superimposes an adaptive algorithm for selecting the "best" step size. This is further developed in [2] with many worked out numerical examples for problems in signal processing. The analysis of such algorithms in discrete time is analyzed in [10] (see also [11]).

Here, we work with continuous-time algorithms, and we also seek an adaptation scheme so as to achieve the goal of selecting the best step size μ. In lieu of a fixed μ, we choose a time-dependent $\mu(t)$. Then, we construct a second-level continuous-time stochastic approximation algorithm for $\mu(t)$ adaptation.

We let the step size μ in (6) be confined in $[\mu_1, \mu_2]$ with $0 < \mu_1 < \mu_2$ being sufficiently small. Suppose for each $\mu \in [\mu_1, \mu_2]$, $(\varphi(t), \overline{y}(t), \Theta(t))$ is a stationary process. We regard the observation noise associate with $\overline{y}(t)$ as a function of μ and write

$$
\overline{e}(t, \mu) = \overline{y}(t) - \varphi'(t)\Theta(t). \quad (10)
$$

Define $\Theta^\pm(t)$ by

$$
\dot{\Theta}^\pm(t) = (\mu(t) \pm \Delta)\varphi(t)(\overline{y}(t) - \varphi'(t)\Theta^\pm(t)),
$$

and define

$$
\delta\ell(t, \mu) = \frac{1}{2}[(\overline{y}(t) - \varphi'(t)\Theta^+(t))^2 - (\overline{y}(t) - \varphi'(t)\Theta^-(t))^2]
$$

and the central finite difference $\delta\Theta(t) = (\Theta^+(t) - \Theta^-(t))/2\Delta$. Similar to [10, p. 1406], as
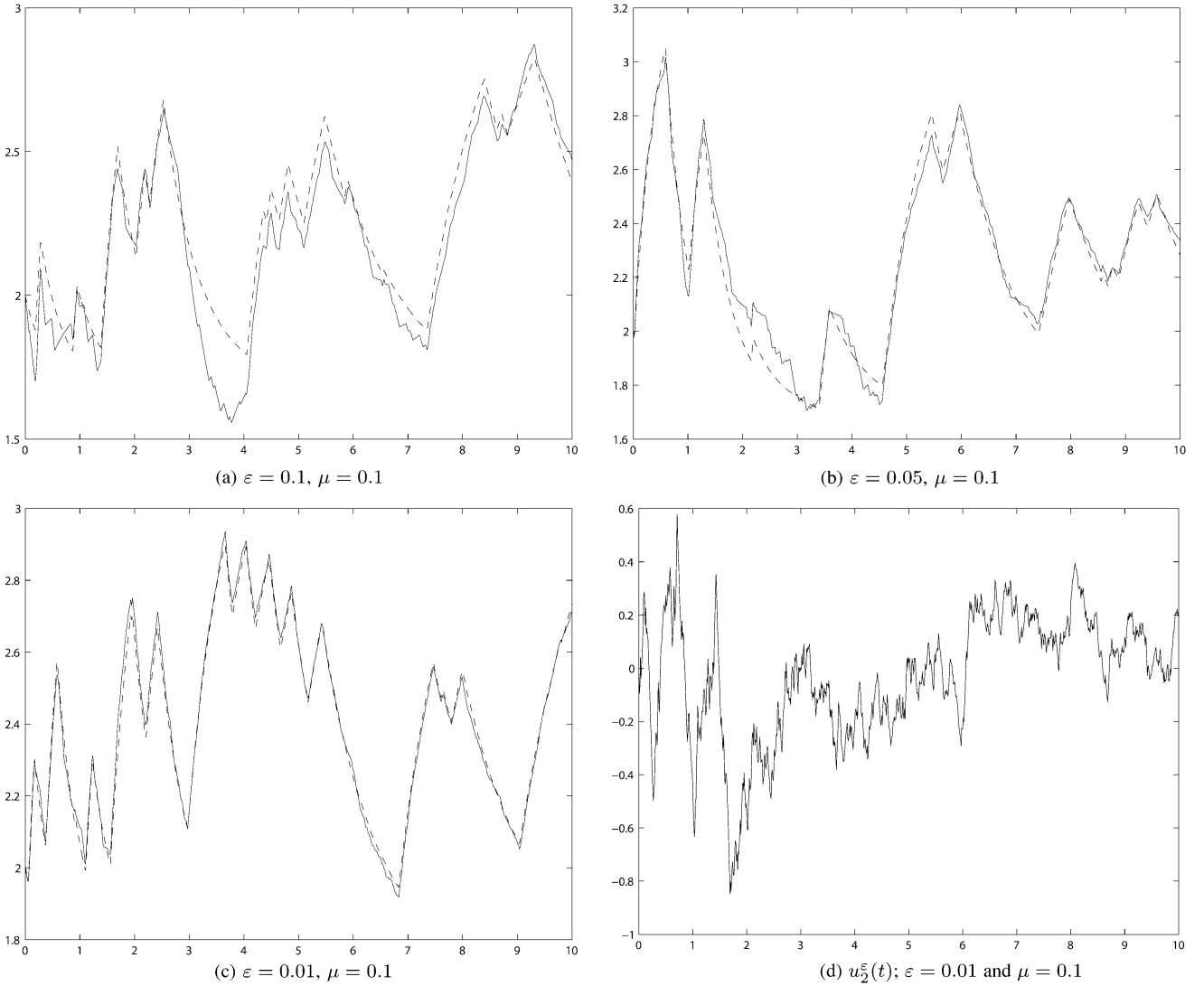
Fig. 1.   (a)–(c) Sample paths of $\Theta_2^\varepsilon(t)$ versus $\Theta_2(t)$ (the second components): solid line—$\Theta_2^\varepsilon(t)$; dashed line—$\Theta_2(t)$. (d) Sample path of $u_2^\varepsilon(t)$.

$\Delta \rightarrow 0$, $\delta\Theta(t)$ has the interpretation of mean-squares derivative. Note that

$$\delta\ell(t,\mu) = -(\overline{y}(t) - \varphi'(t)\Theta^-(t))\varphi'(t)(\delta\Theta(t))$$
$$+\Delta(\delta\Theta(t))'\varphi(t)\varphi'(t)(\delta\Theta(t)).$$

Then, we use an adaptive step-size algorithm of the following form:

$$\dot{\Theta}(t) = \mu(t)\varphi(t)(\overline{y}(t) - \varphi'(t)\Theta(t))$$
$$\dot{\mu}(t) = [-\beta(\delta\ell(t,\mu))]_{\mu_-}^{\mu_+}. \qquad (11)$$

Suppose that $\Delta = \Delta(\beta)$ such that $\Delta \rightarrow 0$ as $\beta \rightarrow 0$ and $\beta/\Delta \rightarrow 0$ as $\beta \rightarrow 0$. That is, β goes to 0 much faster than Δ.

Using techniques in stochastic approximation [11], we proceed to analyze the adaptive-step size tracking algorithm. Define $\widetilde{\mu}^\beta(\cdot)$ by $\widetilde{\mu}^\beta(t) = \mu(t/\beta)$. In other words, we use a time-scale transformation. It follows from (11), $\{\widetilde{\mu}^\beta(\cdot)\}$ is bounded due to its truncation. Then, it is easily seen that for any $t > 0$, $s > 0$, $\delta > 0$ with $s < \delta$, $E_t|\widetilde{\mu}^\beta(t+s) - \widetilde{\mu}^\beta(t)|^2 \leq \zeta(\beta,s)$, where $\zeta(\beta,s)$ is a random variable satisfying $E\zeta(\beta,s) \leq O(s^2) = O(\delta^2)$. It follows that $\lim_{\delta\rightarrow 0}\lim\sup_{\beta\rightarrow 0} E\zeta(\beta,s) = 0$, hence the tightness follows.

The tightness of $\{\widetilde{\mu}^\beta(\cdot)\}$ and the Prohorov's theorem enable us to extract a convergent subsequence. For notational simplicity, still denote the subsequence by $\{\widetilde{\mu}^\beta(\cdot)\}$ with limit $\widetilde{\mu}(\cdot)$. By the Skorohod representation, without changing notation, we may assume that $\widetilde{\mu}^\beta(\cdot)$ converges to $\widetilde{\mu}(\cdot)$ with probability 1. Also for simplicity, we drop the projection in the following discussion but retain it in the limit. Using the definition of $\widetilde{\mu}^\beta(\cdot)$, we have for any $0 < t, s \leq T$

$$\widetilde{\mu}^\beta(t+s) - \widetilde{\mu}^\beta(t)$$
$$= -\beta \int_{t/\beta}^{(t+s)/\beta} \overline{e}(\tau, \widetilde{\mu})\varphi'(\tau)(\delta\Theta(\tau))d\tau$$
$$-\beta \int_{t/\beta}^{(t+s)/\beta} (\Theta(\tau) - \Theta^-(\tau))'\varphi(\tau)\varphi'(\tau)(\delta\Theta(\tau))d\tau$$
$$-\beta\Delta \int_{t/\beta}^{(t+s)/\beta} (\delta\Theta(\tau))'\varphi(\tau)\varphi'(\tau)(\delta\Theta(\tau))d\tau.$$

TABLE I
COMPARISONS OF MEAN-SQUARES ERRORS

μ	MSE_1	AMSE_1	MSE_2	AMSE_2
0.1	0.869127	0.837454	0.843072	0.836821
0.05	0.979705	0.806152	0.910635	0.870448
0.01	1.508524	0.861338	1.005044	0.864732

μ	MSE_3	AMSE_3	MSE_4	AMSE_4
0.1	0.848370	0.838358	0.892907	0.853415
0.05	0.924675	0.869267	1.007786	0.825081
0.01	1.071338	0.843864	1.694935	0.918858

First, it can be shown that

$$-\beta\Delta\int_{t/\beta}^{(t+s)/\beta}(\delta\Theta(\tau))'\varphi(\tau)\varphi'(\tau)(\delta\Theta(\tau))d\tau\to 0$$

in probability. In addition, $|\Theta(\tau) - \Theta^-(\tau)| \le K\Delta(\delta\Theta(\tau))$. Thus, we also have $\beta\int_{t/\beta}^{(t+s)/\beta}(\Theta(\tau) - \Theta^-(\tau))'\varphi(\tau)\varphi'(\tau)(\delta\Theta(\tau))d\tau \to 0$ in probability. It follows that

$$\widetilde{\mu}^\beta(t+s)-\widetilde{\mu}^\beta(t)=-\beta\int_{t/\beta}^{(t+s)/\beta}\overline{e}(\tau,\widetilde{\mu})\varphi'(\tau)(\delta\Theta(\tau))d\tau+o(1)$$

where $o(1) \to 0$ in probability uniformly in $t \in [0,T]$. By using the perturbed test function methods (see [8] and [11]), we characterize the limit process $\widetilde{\mu}(\cdot)$ as

$$\widetilde{\mu}(t + s) - \widetilde{\mu}(t) = \int_t^{t+s} g(\widetilde{\mu}(\tau))d\tau, \ \text{ for } \widetilde{\mu} \in (\mu_-, \mu_+)$$

where $g(\widetilde{\mu}) = -(\partial/\partial\widetilde{\mu})E(\overline{e}(t,\widetilde{\mu}))^2/2$. That is, as its discrete-time counter part (see [10]), there is a mean ordinary differential equation associated with (11), which is given by

$$\frac{d}{dt}\widetilde{\mu}(t) = g(\widetilde{\mu}(t))|_{\mu_-}^{\mu_+}.$$

To demonstrate, we consider a numerical example. Use $[\mu_-, \mu_+] = [0.001, 0.3]$, and consider the time interval [0,10]. Take $\Delta = 0.18$ and $\beta = 0.05$, and keep all other parameters the same as in the previous section. Table I presents the mean-squares errors obtained using (6) and that of the adaptive algorithm. In the table, corresponding to the initial state i, MSE_i and AMSE_i denote the mean-squares errors using (6) and (11), respectively. The first column gives the value of μ used in (6). We run 50 sample paths and partition the time interval [0,10] into 1000 subintervals using step size 0.01 (i.e., $0 = t_0 < t_1 < \cdots < t_{1000} = 10$). Denote the solution in the jth sample (jth simulation run) using (6) or (11) by $\Theta^{\{j\}}(t)$. Then, we compute $(1/50)\sum_{j=1}^{50}(1/1001)\sum_{i=0}^{1000}(\Theta^{\{j\}}(t_i) - \theta(t_i))^2$, where $\theta(t_i)$ is the true parameter.

As can be seen from Table I, the mean-squares errors of using the adaptive step-size algorithm are consistently smaller than that of (6). As the value of μ decreases, the corresponding mean-squares errors increase. From a computational point of view, we would not want to choose μ to be too small, since it may result in updates of a small step at each time, which may not be desirable.

VI. FURTHER DISCUSSIONS

Continuous-time tracking algorithms for Markov regime switching models have been developed and analyzed in this paper. One of the main features of the algorithm is that the driving random process is fast varying. The resulting system is thus difficult to track. Our approach is to use averaging methods to ignore microvariations and concentrate on a macrosystem that is an average with respect to invariant measures.

Developing further on this idea, we have suggested adaptive step-size algorithms. As can be seen, the adaptive step-size algorithm relies on finite-difference approximation. The added stochastic approximation algorithm aims at given "descent" direction of the step sizes. As in [10], we may consider algorithms with the step-size adaptation built based on derivatives in the L_2 sense.

Out focus has been on the case that all of the states of the Markov chain are recurrent. Similar methods can be used to analyze Markov chains, including transient states. In this case, we can partition the states of the Markov chain as $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cdots \cup \mathcal{M}_l \cup \mathcal{M}_*$, where $\mathcal{M}_*$ denotes the group of transient states. The corresponding generator $\widetilde{Q}$ now takes the form

$$\widetilde{Q} = \begin{pmatrix} \widetilde{Q}^1 & & & \\ & \widetilde{Q}^2 & & \\ & & \ddots & \\ & & & \widetilde{Q}^l \\ \widetilde{Q}_*^1 & \widetilde{Q}_*^2 & \cdots & \widetilde{Q}_* \end{pmatrix}$$

where Q_* is an $m_* \times m_*$ matrix with all of its eigenvalues on the left half of the complex plane. In this case, we only aggregate the states in each of the recurrent classes. Thus, we define

$$\overline{\alpha}^\varepsilon(t) = \begin{cases} i, & \text{if } \alpha^\varepsilon(t) \in \mathcal{M}_i \text{ for } i = 1,\ldots,l, \\ U_j, & \text{if } \alpha^\varepsilon(t) = a_j^*, \text{ for } j = 1,\ldots,m_* \end{cases}$$

where U_j is given by

$$U_j = I_{\{0 \le U \le \xi_j^1\}} + 2I_{\{\xi_j^1 < U \le \xi_j^1 + \xi_j^2\}} + \cdots + lI_{\{\xi_j^1 + \cdots + \xi_j^{l-1} < U \le 1\}}$$

$\xi^i = -\widetilde{Q}_*^{-1}\widetilde{Q}_*^i\mathbb{1}_{m_i}$, ξ_j^i denotes the jth component of ξ^i for $i = 1,\ldots,l$ and $j = 1,\ldots,m_i$, and U is a random variable uniformly distributed on $[0,1]$, independent of the Markov chain $\alpha^\varepsilon(t)$. Note that ξ_j^i can be viewed as the probability of entering the ith recurrent class starting from transient state a_j^* (see [22]). Then, we can proceed to study the tracking property and ascertain the rates of tracking error variations.

Finally, there is a fairly complete treatment of the hidden Markov model (HMM) (see [6]). The filtering schemes suggested in this reference can be used to treat the averaged systems. Thus, we could use a combined approach of averaging and HMM to track the time-varying signals driven by fast Markov chains.

APPENDIX

In this section, we give the proofs of Theorems 3.2, 3.4, and 4.1 and Corollary 3.7.

Proof or Theorem 3.2: The function $V(\Theta)$ chosen above is a Liapunov function associated with the differential equation (4). Differentiating $V(\Theta^\varepsilon(t))$ along the solution leads to

$$
\begin{aligned}
\frac{d}{dt}&V(\Theta^\varepsilon(t))\\
&= \mu\Theta^{\varepsilon,\prime}(t)\left[\varphi(t)\varphi'(t)\widehat{\theta}(t,\alpha^\varepsilon(t))+\varphi(t)e(t)-\varphi(t)\varphi'(t)\Theta^\varepsilon(t)\right]\\
&\quad + O(\mu^2)(1+V(\Theta^\varepsilon(t)))\\
&= \mu\Theta^{\varepsilon,\prime}(t)\Big[-A\Theta^\varepsilon(t)+(A-\varphi(t)\varphi'(t))\Theta^\varepsilon(t)\\
&\qquad\qquad +\varphi(t)\varphi'(t)\widehat{\theta}(t,\alpha^\varepsilon(t))+\varphi(t)e(t)\Big]\\
&\quad + O(\mu^2)(1+V(\Theta^\varepsilon(t))).
\end{aligned}
$$

To proceed, we use the perturbed Liapunov function method (see [11]). Define

$$
V_1^\mu(\Theta,t) = \mu\int_t^\infty E_t\Theta'(A-\varphi(u)\varphi'(u))\Theta\,du
$$

where E_t denotes the conditional expectation with respect to the σ-algebra $\mathcal{F}_t = \{\alpha^\varepsilon(u),\varphi(u),e(u):u\le t\}$. Then, using A2)

$$
\begin{aligned}
|V_1^\mu(\Theta,t)| &\le \mu|\Theta|^2\int_t^\infty|E_t[A-\varphi(u)\varphi'(u)]|du\\
&\le O(\mu)(1+V(\Theta)). \qquad\qquad (12)
\end{aligned}
$$

Moreover, direct calculation yields

$$
\begin{aligned}
\frac{d}{dt}V_1^\mu(\Theta^\varepsilon(t),t) =&-\mu\Theta^{\varepsilon,\prime}(t)[A-\varphi(t)\varphi'(t)]\Theta^\varepsilon(t)\\
&+\mu\int_t^\infty\dot{\Theta}^{\varepsilon,\prime}(t)[A-\varphi(u)\varphi'(u)]\Theta^\varepsilon(t)du\\
&+\mu\int_t^\infty\Theta^{\varepsilon,\prime}(t)[A-\varphi(u)\varphi'(u)]\dot{\Theta}^\varepsilon(t)du\\
=&-\mu\Theta^{\varepsilon,\prime}(t)[A-\varphi(t)\varphi'(t)]\Theta^\varepsilon(t)\\
&+O(\mu^2)(1+V(\Theta^\varepsilon(t))). \qquad (13)
\end{aligned}
$$

Define

$$
V^\mu(\Theta,t) = V(\Theta)+V_1^\mu(\Theta,t).
$$

Then, for some $\lambda>0$, $\Theta^{\varepsilon,\prime}(t)A\Theta^\varepsilon(t)\ge\lambda V(\Theta^\varepsilon(t))$, and as result

$$
\begin{aligned}
\frac{d}{dt}V^\mu(\Theta^\varepsilon(t),t) =&-\mu\lambda V(\Theta^\varepsilon(t))+\mu V_\Theta'(\Theta^\varepsilon(t))\\
&\cdot[\varphi(t)\varphi'(t)\widehat{\theta}(t,\alpha^\varepsilon(t))+\varphi(t)e(t)]\\
&+O(\mu^2)(1+V(\Theta^\varepsilon(t)))
\end{aligned}
$$

where $V_\Theta(\Theta)$ denotes the gradient of $V(\Theta)$ with respect to Θ. It is readily seen that $|V_\Theta(\Theta)|\le K(1+V(\Theta))$. The boundedness of $\{\varphi(t),e(t)\}$ implies that

$$
\begin{aligned}
|V_\Theta'(\Theta^\varepsilon(t))[\varphi(t)\varphi'(t)\widehat{\theta}&(t,\alpha^\varepsilon(t))+\varphi(t)e(t)]|\\
&\le K(1+V(\Theta^\varepsilon(t))).
\end{aligned}
$$

By virtue of (12), the $V(\Theta^\varepsilon(t))$ on the right-hand side above can be replaced by $V^\mu(\Theta^\varepsilon(t),t)$ with a small error term added. Therefore

$$
\begin{aligned}
\frac{d}{dt}V^\mu(\Theta^\varepsilon(t),t) \le&-\mu\lambda V^\mu(\Theta^\varepsilon(t),t)\\
&+O(\mu)(1+V(\Theta^\varepsilon(t),t))+O(\mu^2)(1+V^\mu(\Theta^\varepsilon(t),t)).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
EV^\mu(\Theta^\varepsilon(t),t) \le&\, e^{-\mu\lambda t}EV^\mu(\Theta^\varepsilon(0),0)\\
&+K\mu\int_0^t e^{-\mu\lambda(t-u)}V^\mu(\Theta^\varepsilon(u),u)du\\
&+K\mu\int_0^t e^{-\mu\lambda(t-u)}du.
\end{aligned}
$$

An application of Gronwall's inequality yields that $EV^\mu(\Theta^\varepsilon(t),t) = O(1)$. Replacing $V^\mu(\Theta^\varepsilon(t),t)$ by $V(\Theta^\varepsilon(t))$ and using (12), we arrive at $EV(\Theta^\varepsilon(t)) = O(1)$. The proof is concluded. $\square$

Proof of Theorem 3.4: We first prove the tightness of $\{\Theta^\varepsilon(\cdot)\}$. For fixed $\mu>0$, using (4) and Remark 3.3, for any $\delta>0$, and $0<s,t<T$ with $s\le\delta$ and $s+t\le T$, by the continuity of $\widehat{\theta}(\cdot,\alpha)$ for each $\alpha\in\mathcal{M}$

$$
\begin{aligned}
E_t&|\Theta^\varepsilon(t+s)-\Theta^\varepsilon(t)|^2\\
=&\,E_t\left|\mu\int_t^{t+s}[\varphi(u)\varphi'(u)\widehat{\theta}(u,\alpha^\varepsilon(u))\right.\\
&\qquad\qquad\left.-\varphi(u)\varphi'(u)\Theta^\varepsilon(u)+\varphi(u)e(u)]du\right|^2\\
\le&\,K\int_t^{t+s}\int_t^{t+s}E_t|\varphi(u)\varphi'(u)||\varphi(\tau)\varphi'(\tau)|dud\tau\\
&+K\int_t^{t+s}\int_t^{t+s}E_t|\varphi(u)\varphi'(u)||\varphi(\tau)\varphi'(\tau)||\Theta^\varepsilon(u)||\Theta^\varepsilon(\tau)|dud\tau\\
&+K\int_t^{t+s}\int_t^{t+s}E_t|\varphi(u)e(u)||\varphi(\tau)e(\tau)|dud\tau\\
\le&\,Ks^2+\int_t^{t+s}\int_t^{t+s}[E_t|\Theta^\varepsilon(\tau)|^2+E_t|\Theta^\varepsilon(u)|^2]du\\
\le&\,K\zeta(\varepsilon,s)
\end{aligned}
$$

where $\zeta(\varepsilon,s)$ is random and $E\zeta(\varepsilon,s)\le O(\delta^2)$ by virtue of Theorem 3.2. It follows that

$$
\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}E\zeta(\varepsilon,s)=0.
$$

Thus, $\{\Theta^\varepsilon(\cdot)\}$ is tight by the tightness criterion [8, p. 47]. The weak convergence of $\{\overline{\alpha}^\varepsilon(\cdot)\}$ implies that it is also tight. Then, $(\Theta^\varepsilon(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ is tight in $D([0,T]:\mathbb{R}^r\times\mathcal{M})$, which is the space of functions defined on $[0,T]$, taking values in $\mathbb{R}^r\times\mathcal{M}$ and

being right continuous and having limit limits, endowed with the Skorohod topology.

Since $(\Theta^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ is tight, by Prohorov's theorem, we can extract a weakly convergent subsequence. For simplicity, still denote the subsequence by $(\Theta^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$, whose limit is denoted by $(\Theta(\cdot), \overline{\alpha}(\cdot))$. Using (4), we write $\Theta^\varepsilon(t)$ as

$$
\Theta^\varepsilon(t) = \Theta_0 + \mu \int_0^t [\varphi(s)\varphi'(s)\widehat{\theta}(s, \alpha^\varepsilon(s))
$$
$$
- \varphi(s)\varphi'(s)\Theta^\varepsilon(s) + \varphi(s)e(s)]ds
$$
$$
= \Theta_0 + \mu \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \varphi(s)\varphi'(s)\widehat{\theta}(s, a_j^i)\nu_j^i I_{\{\alpha^\varepsilon(s) \in \mathcal{M}_i\}} ds
$$
$$
- \mu \int_0^t \varphi(s)\varphi'(s)\Theta^\varepsilon(s)ds + \mu \int_0^t \varphi(s)e(s)ds
$$
$$
+ \mu \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \varphi(s)\varphi'(s)\widehat{\theta}(s, a_j^i)
$$
$$
\times [I_{\{\alpha^\varepsilon(s)=a_j^i\}} - \nu_j^i I_{\{\alpha^\varepsilon(s) \in \mathcal{M}_i\}}]ds. \quad (14)
$$

Focusing on the last term, we have by virtue of Lemma 3.1, as $\varepsilon \to 0$

$$
E\left|\mu \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \varphi(s)\varphi'(s)\widehat{\theta}(s, a_j^i)\right.
$$
$$
\left. [I_{\{\alpha^\varepsilon(s)=a_j^i\}} - \nu_j^i I_{\{\alpha^\varepsilon(s) \in \mathcal{M}_i\}}]ds\right|^2 \to 0.
$$

Thus

$$
\Theta^\varepsilon(t) = \Theta_0 + \mu \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \varphi(s)\varphi'(s)\widehat{\theta}(s, a_j^i)\nu_j^i I_{\{\alpha^\varepsilon(s) \in \mathcal{M}_i\}} ds
$$
$$
- \mu \int_0^t \varphi(s)\varphi'(s)\Theta^\varepsilon(s)ds + \mu \int_0^t \varphi(s)e(s)ds + o(1) \quad (15)
$$

where $o(1) \to 0$ in probability as $\varepsilon \to 0$ uniformly in $t \in [0, T]$. By virtue of the weak convergence of $(\Theta^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ to $(\Theta(\cdot), \overline{\alpha}(\cdot))$, $I_{\{\alpha^\varepsilon(s) \in \mathcal{M}_i\}} = I_{\{\overline{\alpha}^\varepsilon(s)=i\}} \to I_{\{\overline{\alpha}(s)=i\}}$ in distribution. By the Skorohod representation (with a slight abuse of notation), we may assume that $I_{\{\alpha^\varepsilon(s) \in \mathcal{M}_i\}} = I_{\{\overline{\alpha}^\varepsilon(s)=i\}} \to I_{\{\overline{\alpha}(s)=i\}}$ and $\Theta^\varepsilon(\cdot) \to \Theta(\cdot)$ with probability 1, and the convergence is uniform on any bounded time interval. Using this fact and taking limit in (15) as $\varepsilon \to 0$, we obtain

$$
\Theta(t) = \Theta_0 + \mu \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \varphi(s)\varphi'(s)\widehat{\theta}(s, a_j^i)\nu_j^i I_{\{\overline{\alpha}(s)=i\}} ds
$$
$$
- \mu \int_0^t \varphi(s)\varphi'(s)\Theta(s)ds + \mu \int_0^t \varphi(s)e(s)ds
$$
$$
= \Theta_0 + \mu \int_0^t \sum_{i=1}^l \varphi(s)\varphi'(s)\overline{\theta}(s, i)I_{\{\overline{\alpha}(s)=i\}} ds
$$
$$
- \mu \int_0^t \varphi(s)\varphi'(s)\Theta(s)ds + \mu \int_0^t \varphi(s)e(s)ds. \quad (16)
$$

Equivalently, what has been proved so far is that the pair of processes $(\Theta(\cdot), \overline{\alpha}(\cdot))$ solves a martingale problem with operator L_0 defined by

$$
L_0 F(\Theta, i)
$$
$$
= \mu F'_\Theta(\Theta, i)[\varphi\varphi'\overline{\theta}(i) + \varphi e - \varphi\varphi'\Theta] + QF(\Theta, \cdot)(i)
$$

where

$$
QF(\Theta, \cdot)(i)
$$
$$
= \sum_j q_{ij} F(\Theta, j) = \sum_{j \neq i} q_{ij}[F(\Theta, j) - F(\Theta, i)]
$$

for an appropriate real-valued function $F(\cdot)$. We claim that the solution of the martingale problem associated with the operator L_0 is unique in the sense of in distribution. To see this, using the characteristic function, define $\xi_{ij}(t) = E\left[I_{\{\overline{\alpha}(t)=i\}} \exp(\mathbf{i}[\widehat{t}'\Theta + \widehat{t}_0 j])\right]$, $t \geq 0$, where $\mathbf{i}^2 = -1$, $\widehat{t} \in \mathbb{R}^r$, $\widehat{t}_0 \in \mathbb{R}$, and $F(\cdot, i)$ is a continuously differentiable, real-valued function. Define the vector $\xi(t) = (\xi_{ij}(t) : i = 1, \ldots, l, j = 1, \ldots, m_i)$. Then, it is easily verified that $\xi(t)$ satisfies a linear ordinary differential equation, which has a unique solution. It follows that $E \exp(\mathbf{i}[\widehat{t}'\Theta(t) + \widehat{t}_0\overline{\alpha}(t)])$ is uniquely determined for all $\widehat{t} \in \mathbb{R}^r$ and all $\widehat{t}_0 \in \mathbb{R}$. As a consequence, the distribution of $(\Theta(t), \overline{\alpha}(t))$ is uniquely determined by virtue of the well-known inversion theorem of the characteristic function. The proof of the theorem is thus completed. $\square$

Proof of Corollary 3.7: Subtracting $\overline{\Theta}^\varepsilon(t)$ from $\Theta^\varepsilon(t)$ yields

$$
\Theta^\varepsilon(t) - \overline{\Theta}^\varepsilon(t)
$$
$$
= \mu \int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} \varphi(s)\varphi'(s)\widehat{\theta}(s, a_j^i)[I_{\{\alpha^\varepsilon(s)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(s)=i\}}]ds
$$
$$
- \mu \int_0^t \varphi(s)\varphi'(s)[\Theta^\varepsilon(s) - \overline{\Theta}^\varepsilon(s)]ds. \quad (17)
$$

It follows that

$$
E|\Theta^\varepsilon(t) - \overline{\Theta}^\varepsilon(t)|^2
$$
$$
\leq K\rho^\varepsilon + K\mu^2 \int_0^t E|\Theta^\varepsilon(s) - \overline{\Theta}^\varepsilon(s)|^2 ds
$$

where

$$
\rho^\varepsilon = \mu^2 E\left(\int_0^t \sum_{i=1}^l \sum_{j=1}^{m_i} [I_{\{\alpha^\varepsilon(s)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(s)=i\}}]ds\right)^2.
$$

Note that $\rho^\varepsilon \to 0$ as $\varepsilon \to 0$ by virtue of Lemma 3.1. By virtue of Gronwall's inequality

$$
E|\Theta^\varepsilon(t) - \overline{\Theta}^\varepsilon(t)|^2 \leq K\rho^\varepsilon \exp(K\mu^2 T) \to 0 \text{ as } \varepsilon \to 0.
$$

The assertion is proved. $\square$

Proof of Theorem 4.1: The proof is divided into four steps.

Step 1: We first show that $\sup_{0 \leq t \leq T} E|u^\varepsilon(t)|^2 < \infty$. By virtue of (8) and Lemma 3.1

$$E|u^\varepsilon(t)|^2 \leq \frac{K}{\varepsilon} \sum_{i=1}^{l} \sum_{j=1}^{m_i} \left| \int_0^t \varphi(\tau)\varphi'(\tau)\widehat{\theta}(t, a_j^i) \right.$$

$$\left. [I_{\{\alpha^\varepsilon(\tau)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(\tau)=i\}}]d\tau \right|^2$$

$$+ \int_0^t Eu^{\varepsilon,\prime}(s)\varphi(s)\varphi'(s)\varphi(\tau)\varphi'(\tau)u^\varepsilon(\tau)dsd\tau$$

$$\leq K + K \int_0^t E|u^\varepsilon(s)|^2 ds.$$

The desired bound follows from Lemma 3.1 and Gronwall's inequality.

Step 2: We proof the tightness of $\{u^\varepsilon(\cdot)\}$. For any $\delta > 0$, and $t, s > 0$ with $s \leq \delta$ and $t + s \leq T$, we have

$$E[|u^\varepsilon(t+s) - u^\varepsilon(t)|^2 | \mathcal{F}_t]$$

$$\leq \frac{K}{\varepsilon} E \left[\left| \int_t^{t+s} \varphi(\tau)\varphi'(\tau) \sum_{i=1}^{l} \sum_{j=1}^{m_i} \widehat{\theta}(t, a_j^i) \right. \right.$$

$$\left. \left. \cdot [I_{\{\alpha^\varepsilon(\tau)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(\tau)=i\}}]d\tau \right|^2 \right| \mathcal{F}_t \right]$$

$$+ Ks \int_t^{t+s} E[|u^\varepsilon(\tau)|^2 | \mathcal{F}_t]d\tau$$

$$\leq K\Gamma^\varepsilon(s)$$

where $\Gamma^\varepsilon(s) = O(s) + s\int_t^{t+s} E[|u^\varepsilon(\tau)|^2 | \mathcal{F}_t]d\tau$ by Lemma 3.1, and $E\Gamma^\varepsilon(s) = O(s) = O(\delta)$. Thus, the tightness of $\{u^\varepsilon(\cdot)\}$ follows from the above estimate by taking expectations and $\limsup_{\varepsilon \to 0}$ followed by $\lim_{\delta \to 0}$.

Step 3: We proceed to figure out the limit process. For convenience, set

$$x^\varepsilon(t) = \frac{1}{\sqrt{\varepsilon}} \int_0^t w(s, \alpha^\varepsilon(s))ds \qquad (18)$$

where

$$w(t, \alpha) = \varphi(t)\varphi'(t)[\widehat{\theta}(t, \alpha) - \overline{\theta}(t, \overline{\alpha})].$$

As done in Step 2, we can show that $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ is tight in $D([0, T] : \mathbb{R}^r \times \overline{\mathcal{M}})$. Select a weakly convergent subsequence and still denote it by $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$. Using the Skorohod representation without changing notation, we may assume the convergence is in the sense of with probability one and the convergence takes place uniformly on compact subsets.

For each $\alpha \in \mathcal{M}$ and any $F(\cdot, \cdot, \alpha) \in C_0^{1,2}$ (the collection of $C^{1,2}$ functions with compact support), define the operator

$$L^\varepsilon F(t, x, \alpha) = \frac{\partial}{\partial t} F(t, x, \alpha) + \frac{1}{\sqrt{\varepsilon}} F_x'(t, x, \alpha)w(t, \alpha)$$

$$+ Q^\varepsilon F(t, x, \cdot)(\alpha). \qquad (19)$$

For each $i \in \overline{\mathcal{M}} = \{1, \ldots, l\}$ and any $f(t, x, i) \in C_0^{1,2}$, define

$$\widehat{f}(t, x, \alpha) = \sum_{i=1}^{l} f(t, x, i)I_{\{\alpha \in \mathcal{M}_i\}}.$$

It is well known that

$$\widehat{f}(t, x^\varepsilon(t), \alpha^\varepsilon(t)) - \int_0^t L^\varepsilon \widehat{f}(\tau, x^\varepsilon(\tau), \alpha^\varepsilon(\tau))d\tau$$

$$= f(t, x^\varepsilon(t), \overline{\alpha}^\varepsilon(t)) - \int_0^t L^\varepsilon f(\tau, x^\varepsilon(\tau), \overline{\alpha}^\varepsilon(\tau))d\tau$$

is a martingale. $\qquad (20)$

We will use a combination of perturbed test function and martingale averaging to figure out the limit process. In view of the definition (19)

$$L^\varepsilon \widehat{f}(t, x^\varepsilon(t), \alpha^\varepsilon(t)) = \frac{\partial}{\partial t}\widehat{f}(t, x^\varepsilon(t), \alpha^\varepsilon(t))$$

$$+ \frac{1}{\sqrt{\varepsilon}}\widehat{f}_x'(t, x^\varepsilon(t), \alpha^\varepsilon(t))w(t, \alpha^\varepsilon(t))$$

$$+ Q^\varepsilon \widehat{f}(t, x^\varepsilon(t), \cdot)(\alpha^\varepsilon(t)) \qquad (21)$$

where $\widehat{f}_x(\cdot)$ denotes the gradient of $\widehat{f}(\cdot)$ with respect to x. Noting $\widetilde{Q}\widehat{f}(t, x^\varepsilon(t), \cdot)(\alpha^\varepsilon(t)) = 0$

$$Q^\varepsilon \widehat{f}(t, x^\varepsilon(t), \cdot)(\alpha^\varepsilon(t)) = \widehat{Q}\widehat{f}(t, x^\varepsilon(t), \cdot)(\alpha^\varepsilon(t)). \qquad (22)$$

To get rid of the term of the order $O(1/\sqrt{\varepsilon})$, we introduce the following perturbation:

$$f_1^\varepsilon(t, x, \alpha) = \frac{1}{\sqrt{\varepsilon}} \int_t^T E_t \widehat{f}_x'(t, x, \alpha)w(\tau, \alpha^\varepsilon(\tau))d\tau \qquad (23)$$

where E_t denotes the conditional expectation with respect to the σ-algebra $\mathcal{F}_t$ generated by $\{\alpha^\varepsilon(\tau), \varphi(\tau), e(\tau) : \tau \leq t\}$. By the Markov property

$$E_t[I_{\{\alpha^\varepsilon(\tau)=a_j^i\}} - \nu_j^i I_{\{\overline{\alpha}^\varepsilon(\tau)=i\}}]$$

$$= P(\alpha^\varepsilon(\tau) = a_j^i | \mathcal{F}_t) - \nu_j^i P(\overline{\alpha}^\varepsilon(\tau) = i | \mathcal{F}_t)$$

$$= \sum_{i_1=1}^{l} \sum_{j_1=1}^{m_i} \left[P(\alpha^\varepsilon(\tau)=a_j^i | \alpha^\varepsilon(t)=a_{j_1}^{i_1}) \right.$$

$$\left. -\nu_j^i \sum_{j_2=1}^{m_i} P(\alpha^\varepsilon(\tau)=a_{j_2}^i | \alpha^\varepsilon(t)=a_{j_1}^{i_1}) \right]$$

$$= O\left(\varepsilon + \exp\left(-\frac{(\tau - t)}{\varepsilon}\right)\right). \qquad (24)$$

This implies that (by the independence of $\varphi(t)$ and $\alpha^\varepsilon(\cdot)$)

$$E|f_1^\varepsilon(t, x^\varepsilon(t), \alpha^\varepsilon(t))|$$

$$\leq \frac{1}{\sqrt{\varepsilon}} \int_t^T O\left(\varepsilon + \exp\left(-\frac{(\tau - t)}{\varepsilon}\right)\right)d\tau = O(\sqrt{\varepsilon}). \qquad (25)$$

Thus, the perturbation is small.

Similar to (22)

$$\widetilde{Q}\widehat{f}'_x(t,x^\varepsilon(t),\cdot)(\alpha^\varepsilon(t)) = 0$$

direct calculation yields that

$$
\begin{aligned}
&L^\varepsilon f_1^\varepsilon(t,x^\varepsilon(t),\alpha^\varepsilon(t)) \\
&= -\frac{1}{\sqrt{\varepsilon}}\widehat{f}'(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(t,\alpha^\varepsilon(t)) \\
&\quad + \frac{1}{\varepsilon}\int_t^T E_t \mathrm{tr}[\widehat{f}_{xx}(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(\tau,\alpha^\varepsilon(\tau))w'(t,\alpha^\varepsilon(t))]d\tau \\
&\quad + \int_t^T E_t^\varepsilon \widehat{Q}\widehat{f}'_x(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(t,\alpha^\varepsilon(t))d\tau.
\end{aligned}
\tag{26}
$$

To proceed, define

$$f^\varepsilon(t,x,\alpha) = \widehat{f}(t,x,\alpha) + f_1^\varepsilon(t,x,\alpha).$$

By virtue of (25)

$$E|f^\varepsilon(t,x^\varepsilon(t),\alpha^\varepsilon(t)) - \widehat{f}(t,x^\varepsilon(t),\alpha^\varepsilon(t))| \to 0 \ \text{ as } \ \varepsilon \to 0.
\tag{27}$$

Using (21) and (26)

$$
\begin{aligned}
&L^\varepsilon f^\varepsilon(t,x^\varepsilon(t),\alpha^\varepsilon(t)) \\
&= \frac{\partial}{\partial t}\widehat{f}(t,x^\varepsilon(t),\alpha^\varepsilon(t)) + \widehat{Q}\widehat{f}(t,x^\varepsilon(t),\alpha^\varepsilon(t)) \\
&\quad + \frac{1}{\varepsilon}\int_t^T E_t \mathrm{tr}[\widehat{f}_{xx}(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(\tau,\alpha^\varepsilon(\tau))w'(t,\alpha^\varepsilon(t))]d\tau \\
&\quad + \int_t^T E_t \widehat{Q}\widehat{f}'_x(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(\tau,\alpha^\varepsilon(\tau))d\tau.
\end{aligned}
$$

By virtue of Lemma 3.1, and the boundedness of $\widehat{f}_x(\cdot,\cdot,\alpha)$ for each α, as $\varepsilon \to 0$

$$
E\left|\int_t^T E_t^\varepsilon \widehat{Q}\widehat{f}'_x(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(\tau,\alpha^\varepsilon(\tau))d\tau\right|^2
$$

$$
\leq K\left|\int_t^T w(\tau,\alpha^\varepsilon(\tau))d\tau\right|^2 \to 0.
$$

Moreover, in view of (24), the limit (in probability) of $(1/\varepsilon)\int_t^T E_t w(\tau,\alpha^\varepsilon(\tau))w'(t,\alpha^\varepsilon(t))d\tau$ exists and is denoted by $\widetilde{\Sigma}(t)/2$. Consequently

$$
\begin{aligned}
&\frac{1}{\varepsilon}\int_t^T E_t \mathrm{tr}[\widehat{f}_{xx}(t,x^\varepsilon(t),\alpha^\varepsilon(t))w(\tau,\alpha^\varepsilon(\tau))w'(t,\alpha^\varepsilon(t))]d\tau \\
&\qquad\qquad \to \mathrm{tr}f_{xx}(t,x(t),\overline{\alpha}(t))\overline{\Sigma}(t) \text{ in probability.}
\end{aligned}
$$

Noting the use of $f(\cdot)$ and $\widehat{f}(\cdot)$, the with probability 1 convergence of $(x^\varepsilon(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ to $(x(\cdot),\overline{\alpha}(\cdot))$ (via the weak convergence and the Skorohod representation), we have that for any positive integer k_0, any $j_1 \leq k_0$, any bounded and continuous function $h_{j_1}(\cdot,i)$ for $i \in \overline{\mathcal{M}}$, see (28), shown at the bottom of the page, where for notational simplicity, we have used

$$\pi_{k_0}^\varepsilon = \prod_{j_1=1}^{k_0} h_{j_1}(x^\varepsilon(t_{j_1}),\overline{\alpha}^\varepsilon(t_{j_1})), \pi_{k_0} = \prod_{j_1=1}^{k_0} h_{j_1}(x(t_{j_1}),\overline{\alpha}(t_{j_1})).$$

In view of the definition of the perturbation and (25), see the equation that is shown at the bottom of the page, where the limit operator is defined by

$$Lf(t,x,i) = \frac{\partial}{\partial t}f(t,x,i) + \mathrm{tr}[f_{xx}(t,x,i)\overline{\Sigma}(t)] + \overline{Q}f(t,x,\cdot)(i),$$
$$\text{for } i \in \overline{\mathcal{M}}.$$

Thus, the limit process $(x(\cdot),\overline{\alpha}(\cdot))$ solves a martingale problem with operator L. Moreover, via the use of characteristic functions as in the last part of the proof of Theorem 3.4, the martingale problem is unique in the sense in distribution. Thus, we have a representation for the limit $x^\varepsilon(\cdot)$, namely $x(t) = \int_0^t \overline{\Sigma}(u)dw(u)$, where $w(\cdot)$ is an $\mathbb{R}^r$-dimensional standard Brownian motion. Note that in the above, the diffusion coefficient $\overline{\Sigma}(t)$ in fact depends on the limit Markov chain $\overline{\alpha}(\cdot)$. Thus, in lieu of a diffusion process, we have a limit switching diffusion.

Step 4: Using the result of Step 3 and (8), the weak convergence of $(u^\varepsilon(\cdot),x^\varepsilon(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ to $(u(\cdot),x(\cdot),\overline{\alpha}(\cdot))$ and the Skorohod representation then imply that the limit $u(\cdot)$ is given by (9). Furthermore, the limit is unique in distribution by the same reason as in Step 3. □

$$
\begin{aligned}
&E\pi_{k_0}^\varepsilon \left[\int_t^{t+s}\left[\frac{\partial}{\partial\tau}\widehat{f}(\tau,x^\varepsilon(\tau),\alpha^\varepsilon(\tau)) + \widehat{Q}\widehat{f}(\tau,x^\varepsilon(\tau),\cdot)(\alpha^\varepsilon(\tau))\right.\right. \\
&\qquad\qquad\left.\left. + \frac{1}{\varepsilon}\int_\tau^T E_\tau \mathrm{tr}[\widehat{f}_{xx}(\tau,x^\varepsilon(\tau),\alpha^\varepsilon(\tau))w(u,\alpha^\varepsilon(u))w'(\tau,\alpha^\varepsilon(\tau))]du\right]d\tau\right] \\
&\to E\pi_{k_0}\left[\int_t^{t+s}\left[\frac{\partial}{\partial\tau}f(\tau,x(\tau),\overline{\alpha}(\tau)) + \overline{Q}f(\tau,x(\tau),\cdot)(\overline{\alpha}(\tau)) + \mathrm{tr}\left[f_{xx}(\tau,x(\tau),\overline{\alpha}(\tau))\overline{\Sigma}(\tau)\right]\right]d\tau\right]
\end{aligned}
\tag{28}
$$

$$
\begin{aligned}
&\lim_{\varepsilon\to 0}E\pi_{k_0}^\varepsilon\left[f(t+s,x^\varepsilon(t+s),\overline{\alpha}^\varepsilon(t+s)) - f(t,x^\varepsilon(t),\overline{\alpha}^\varepsilon(t)) - \int_t^{t+s}L^\varepsilon f(\tau,x^\varepsilon(\tau),\overline{\alpha}^\varepsilon(\tau))d\tau\right] \\
&= E\pi_{k_0}\left[f(t+s,x(t+s),\overline{\alpha}(t+s)) - f(t,x(t),\overline{\alpha}(t)) - \int_t^{t+s}Lf(\tau,x(\tau),\overline{\alpha}(\tau))d\tau\right]
\end{aligned}
$$

REFERENCES

[1] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.

[2] J. M. Brossier, "Egalization Adaptive et Estimation de Phase: Application aux Communications Sous-Marines ," Ph.D, Institut National Polytechnique de Grenoble, Grenoble, France, 1992.

[3] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley, 1968.

[4] H.-F. Chen, *Stochastic Approximation and its Applications*. Dordrecht, The Netherlands: Kluwer Academic, 2003.

[5] L. Guo and L. Ljung, "Performance analysis of general tracking algorithms," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1388–1402, Aug. 1995.

[6] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models*. New York: Springer-Verlag, 1995.

[7] M. L. Honig, U. Madhow, and S. Verdu, "Adaptive blind multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, no. , pp. 944–960, 1995.

[8] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*. Cambridge, MA: MIT Press, 1984.

[9] H. J. Kushner and A. Shwartz, "Weak convergence and asymptotic properties of adaptive filters with constant gains," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 177–182, Mar. 1984.

[10] H. J. Kushner and J. Yang, "Analysis of adaptive step-size SA algorithms for parameter tracking," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1403–1410, Aug. 1995.

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

[12] L. Ljung and S. Gunnarsson, "Adaptive tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–22.

[13] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 161–176, Mar. 1984.

[14] H. V. Poor and X. Wang, "Code-aided interference suppression for DS/CDMA communications—Part I: Interference suppression capability," *IEEE Trans. Commun.*, vol. 45, no. 9, pp. 1101–1111, Sep. 1997.

[15] ——, "Code-aided interference suppression for DS/CDMA communications—Part II: Parallel blind adaptive implementations," *IEEE Trans. Commun.*, vol. 45, no. 9, pp. 1112–1122, Sep. 1997.

[16] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms – Stability and Performance* . Englewood Cliffs, NJ: Prentice-Hall, 1995.

[17] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[18] J. B. Moore, "Convergence of continuous time stochastic ELS parameter estimation," *Stoch. Process. Appl.*, vol. 27, pp. 195–215, 1988.

[19] S. Verdu, *Multiuser Detection* . Cambridge, U.K.: Cambridge Univ. Press, 1998.

[20] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*. New York: Springer-Verlag, 1998.

[21] G. Yin, Q. Zhang, and G. Badowski, "Singularly perturbed Markov chains: Convergence and aggregation," *J. Multivariate Anal.*, vol. 72, pp. 208–229, 2000.

[22] ——, "Asymptotic properties of a singularly perturbed Markov chain with inclusion of transient states," *Ann. Appl. Probab.*, vol. 10, pp. 549–572, 2000.

George Yin (S'87–M'87–SM'96–F'02) received the B.S. degree in mathematics from the University of Delaware, Newark, in 1983 and the M.S. degree in electrical engineering and the Ph.D. degree in applied mathematics from Brown University, Providence, RI, in 1987.

Subsequently, he joined the Department of Mathematics, Wayne State University, Detroit, MI, where he became a Professor in 1996.

Dr. Yin served on the Mathematical Review Date Base Committee; the IFAC Technical Committee on Modeling, Identification and Signal Processing; and various conference program committees. He was the Editor of the SIAM Activity Group on Control and Systems Theory Newsletters, the SIAM Representative to the 34th CDC, Co-Chair of 1996 AMS-SIAM Summer Seminar in Applied Mathematics and Co-Chair of 2003 AMS-IMS-SIAM Summer Research Conference: Mathematics of Finance, and co-organizer of 2005 IMA Workshop on Wireless Communications. He is an Associate Editor of the *SIAM Journal on Control and Optimization*, an Associate Editor of *Automatica*, was formerly an Associate Editor of IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 1994 to 1998, and is (or was) on the editorial board of four other journals.

Qing Zhang (M'98–SM'01) received the B.Sc. degree from Nankai University in 1983 and the M.Sc. and Ph.D. degrees from Brown University, Providence, RI, in 1985 and 1988, respectively.

He is a Professor of Mathematics at University of Georgia, Athens. His research interests include stochastic system and control, nonlinear filtering, and applications in manufacturing and finance. He has published more than 80 journal papers and four monographs on two-time-scale Markovian systems and application in manufacturing and co-edited four books.

Dr. Zhang received a Creative Research Medal from the University of Georgia for outstanding accomplishment in research and creativity in 2000. He was an Associate Editor of IEEE TRANSACTIONS ON AUTOMATIC CONTROL and is currently Associate Editor of the *SIAM Journal on Control and Optimization*. He also served on a number of international conference organizing committees, including Co-Chair of 1996 AMS-SIAM Summer Seminar in Manufacturing and Co-Chair of 2003 AMS-IMS-SIAM Summer Research Conference in Mathematics of Finance.
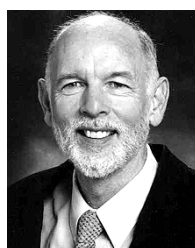
John B. Moore (S'66–M'68–SM'77–F'79) was born in China in 1941. He received the Bachelor's and Master's degrees, both in electrical engineering, in 1963 and 1964, respectively, and the Ph.D. degree in electrical engineering from the University of Santa Clara, Santa Clara, CA, in 1967.

In 1967, he was appointed Senior Lecturer at the Electrical Engineering Department, University of Newcastle, U.K., and promoted to Associate Professor in 1968, and Full Professor (personal chair) in 1973. He was the Department Head from 1975 to 1979. In 1982, he was appointed as a Professorial Fellow in the Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, and was promoted to Professor in 1990. He has been head of the department from 1992 to 1999 and from 2002 to present. The department, now the Information Engineering Department, is located in the Research School of Information Sciences and Engineering. He is currently also a Researcher in the Systems Engineering and Complex Systems Program within National Information and Communication Technologies (ICT) Australia, based in Canberra.

Prof. Moore is a Fellow of the the Australian Academy of Technological Sciences and the Australian Academy of Science.

Yuan Jin Liu received the B.S. degree in mathematics and the M.A. degree in mathematical statistics from Tongji University, Shanghai, China, in 1996 in 1999, respectively. He is currently working toward the Ph.D. degree in mathematics at Wayne State University, Detroit, MI.

His research interests include stochastic approximation, optimization, singular perturbation, and mathematics of finance.

Mr. Liu received the Outstanding Graduate of Shanghai Award when he received the B.S. degree in 1996 and the M.A. degree in 1999.