Least Mean Square Algorithms With Markov Regime-Switching Limit

G. George Yin, *Fellow, IEEE,* and Vikram Krishnamurthy, *Fellow, IEEE*

Abstract—We analyze the tracking performance of the least mean square (LMS) algorithm for adaptively estimating a time varying parameter that evolves according to a finite state Markov chain. We assume the Markov chain jumps infrequently between the finite states at the same rate of change as the LMS algorithm. We derive mean square estimation error bounds for the tracking error of the LMS algorithm using perturbed Lyapunov function methods. Then combining results in two-time-scale Markov chains with weak convergence methods for stochastic approximation, we derive the limit dynamics satisfied by continuous-time interpolation of the estimates. Unlike most previous analyzes of stochastic approximation algorithms, the limit we obtain is a system of ordinary differential equations with regime switching controlled by a continuous-time Markov chain. Next, to analyze the rate of convergence, we take a continuous-time interpolation of a scaled sequence of the error sequence and derive its diffusion limit. Somewhat remarkably, for correlated regression vectors we obtain a jump Markov diffusion. Finally, two novel examples of the analysis are given for state estimation of hidden Markov models (HMMs) and adaptive interference suppression in wireless code division multiple access (CDMA) networks.

Index Terms—Adaptive filtering, adaptive multiuser detection, diffusion process, hidden Markov model (HMM), jump Markov linear diffusion, limit theorems, Markov chain, stochastic approximation, tracking.

I. INTRODUCTION

THIS paper deals with analyzing how well the least mean square (LMS) algorithm can track a time-varying parameter process that undergoes infrequent jump changes with possibly large jump sizes. The values of the time-varying parameter process belong to a finite state space. At any given instance, the parameter process takes one of the possible values in this finite state space. It then sojourns in this state for a random duration. Next, the process jumps into a new location (i.e., a switching of regime takes place). Then, the process sojourns in this new state in a random duration and so on. Naturally the parameter can be modeled by a finite-state Markov chain. We assume that the regime changes take place infrequently. Thus, the time-varying parameter is a discrete-time Markov chain with

"near identity" transition matrix. Henceforth, for brevity, we often call it a slowly varying Markov chain or a slow Markov chain. The slowness is meant to be in the sense of infrequent jumps (transitions). Due to practical concerns arising from many applications, it is crucial to analyze the performance (such as asymptotic error bounds and evolution of the scaled tracking error sequence) of the LMS algorithm. In this paper, we derive mean squares error bounds, treat an interpolated sequence of the centered estimation errors, and examine a suitably scaled sequence via martingale problem formulation.

A. Motivation

A number of examples in fault diagnosis and change detection fit into the slow Markovian model (see [2]), as do problems in target tracking, econometrics (see [11]) and the references therein. As will be seen later, such problems also appear in emerging applications of adaptive interference suppression in wireless code division multiple access (CDMA) networks and state estimation of hidden Markov models (HMMs). To the best of our knowledge, in analyzing LMS algorithms with time-varying parameters, most works up to date have assumed that the parameter varies continuously but slowly over time with small amount of changes; see [2], [7], [8], [15], [16], [20], and [28]. In contrast, we deal with the case when the parameter is constant over long periods of time and then jump changes by possibly a large amount. By considering the Markovian time-varying parameter dependence, we formulate it as a Markov chain with two-time scales. Using recent results on two-time-scale Markov chains [33], [34], we examine the asymptotic properties of the tracking algorithm. A salient feature of the analysis in this paper is: It allows for the parameter to evolve as fast (i.e., on the same time scale) as the LMS tracking algorithm, which is more relevant to applications involving jump change parameters. Two of our recent papers [31], [32] examine similar tracking problems—but yield different results; see description titled "Context" at the end of this section.

B. Main Results

We assume that the true parameter θ_n (called the hypermodel in [2]) evolves according to a slow finite-state Markov chain with transition probability matrix $I + \varepsilon Q$ where $\varepsilon > 0$ is a small parameter, and that the LMS algorithm operates with a step size μ. We consider both unscaled and scaled sequences of tracking errors. The main results of the paper are as follows.

1) In Section III, our focus is on the unscaled tracking error sequence. It is further divided into two parts. First, we present a mean square stability analysis of the LMS algorithm using perturbed Liapunov function

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

V. Krishnamurthy is with Department of Electrical Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vikramk@ece.ubc.ca).

methods in Section III-A, and show the mean square tracking error being $O(\mu + \varepsilon + \varepsilon^2/\mu)$. Thus, as long as the true parameter (a slow Markov chain with transition probability matrix $I + \varepsilon Q$) evolves on the same time scale as the stochastic approximation algorithm with step size μ ($\varepsilon = O(\mu)$), the mean square error is $O(\mu)$.

Then, in Section III-B we examine the limit of an interpolated sequence of tracking errors of the LMS algorithm. One of the novelties of this work is the derivation of switching limit dynamics. For stochastic approximation algorithms, a standard and widely used technique is the so-called ordinary differential equation (ODE) method [14], [16], [18], which combines analysis with probability theory. In most of the results up to date, the limit of the interpolated sequence of estimates generated by a stochastic approximation algorithm results in an autonomous ODE. In certain cases (see [16]), one gets a nonautonomous system of ODEs but the equations are still deterministic. In this paper, due to the Markovian parameter, the existing results of stochastic approximation cannot be applied. In Section III-B, by using martingale averaging, we derive a hybrid limit system, namely, ordinary differential equations modulated by a continuous-time Markov chain. A remarkable feature of our result is that the limit is no longer a deterministic ODE, but rather a system of ODEs with regime switching (i.e., a system of ODEs modulated by a continuous-time Markov chain). Due to the random varying nature of the parameter, the results of stochastic approximation in [16] cannot be directly applied to the current case. Thus, we use a martingale formulation to treat the (estimate, parameter) pair of processes and derive the desired limit.

2) In comparison to Section III-A, Section IV is devoted to suitably scaled tracking errors. Based on the system of switching ODEs obtained in Section III-B, we examine the limiting behavior of the normalized tracking errors of the LMS algorithm. It is well known that for a stochastic approximation algorithm, if the true parameter is a fixed constant, then a suitably scaled sequence of estimation errors has a Gaussian diffusion limit. We show that if the regression vector is independent of the parameter then although the limit system involves Markovian regime switching ODEs, the scaled sequence of tracking errors of the LMS algorithm has a Gaussian diffusion limit.

3) The Gaussian diffusion limit implies that the iterate averaging can be used to accelerate the convergence of the tracking algorithm. In Section IV-B, we show that iterate averaging reduces the asymptotic covariance of the estimate of the LMS algorithm. Originally proposed in [23] for accelerating the convergence of stochastic approximation algorithms, it is well known [16] that for a constant true parameter and decreasing step size, iterate averaging results in asymptotically optimal convergence rate (the same rate as the recursive least squares), which use a matrix step sizes, with an order of magnitude lower computational complexity than RLS.

In the tracking case for a random walk time-varying parameter, it has recently been shown in [19] that the fixed step size LMS algorithm with iterate averaging has similar properties to a recursive least squares algorithm with a forgetting factor. Section IV-B shows that if $\varepsilon = O(\mu)$ (parameter evolves as fast as LMS tracking algorithm), and the averaging window width is $O(1/\mu)$ (where μ denotes the step size of the LMS algorithm), then iterate averaging still results in an asymptotically optimal tracking algorithm. To our knowledge, apart from [31] where we showed a similar result for the more restrictive case with $\varepsilon = o(\mu)$, this is the first example of a case where iterate averaging results in a constant step size LMS algorithm with optimal tracking properties.

4) In Section V, we present expressions for the probability of error of the quantized state estimate of the LMS algorithm when tracking the state of a slow HMM Also an example in adaptive multiuser detection in wireless CDMA networks is given.

5) In Section VI, we analyze the LMS algorithm when the regression vector depends on the parameter (in contrast to Section IV with the regression vector being independent of the parameter). Examples of such dependent regression vector models include autoregressive models with Markov regime that have been widely used in econometrics, failure detection, and maneuvering target tracking; see [9], [11], [12], and the references therein. Unlike, the independent case studied in Section IV, somewhat remarkably, for the dependent case the scaled tracking error sequence generated by the LMS algorithm does not have a diffusion limit. Instead, the limit is a system of diffusions with Markovian regime switching. In the limit system, the diffusion coefficient depends on the modulating Markov chain, which reveals the distinctive time-varying nature of the underlying system and provides new insight on Markov modulated stochastic approximation problems. This result is in stark contrast to the traditional analysis of the LMS algorithm for constant parameter where the scaled sequence of estimation errors has a Gaussian diffusion limit.

C. Context

In Sections III-B, IV–VI, we assume that the dynamics of the true parameter (a slow Markov chain with transition matrix $I + \varepsilon Q$) evolve on the same time scale as the adaptive algorithm with step size μ, i.e., $\varepsilon = O(\mu)$. We note that the case $\varepsilon = O(\mu)$ addressed in this paper is more difficult to handle than $\varepsilon = o(\mu)$ (e.g., $\varepsilon = O(\mu^2)$), which is widely used in the analysis of tracking algorithms [2]. The meaning of $\varepsilon = o(\mu)$ is that the true parameter evolves much slower than the adaptation speed of the stochastic optimization algorithm and is more restrictive than $\varepsilon = O(\mu)$ considered in this paper.

Recently, in [31] and [32], we have considered two related algorithms with different tracking properties. In [32], we studied the tracking performance of an adaptive algorithm for updating the occupation measure of a finite-state Markov chain when $\varepsilon = O(\mu)$. One of its applications is the so-called discrete stochastic optimization in which gradient information cannot be incorporated. The resulting analysis yields a system of Markovian switching differential equations and also a switching diffusion limit for the scaled errors of the occupation measure estimate. In such a case, due to the switching Markov diffusion, iterate averaging does not improve the performance. In [31], we considered the LMS algorithm tracking a slow Markov parameter, where the parameter evolves according to a Markov chain being an order of magnitude slower than the dynamics of the LMS algorithm, i.e., $\varepsilon = o(\mu)$. In that reference, we obtained a standard ODE, whereas with $\varepsilon = O(\mu)$ we show for the first time in this paper that one obtains a randomly switching system of ODEs. In [31], with $\varepsilon = o(\mu)$, iterate averaging was shown to result in an asymptotically optimal tracking algorithm. Here, we show that even with $\varepsilon = O(\mu)$, i.e., much faster evolving parameter, iterate averaging still results in an asymptotically optimal performance. Finally, in this paper, we also consider the case where the regression vector depends on the parameter yielding a Markov switching diffusion limit.

Throughout this paper, for $z \in \mathbb{R}^{\ell \times r}$, z' denotes its transpose. We use K to denote a generic positive constant. The convention $K + K = K$ and $KK = K$ will be used without further notice. Detailed proofs appear in the Appendix .

II. FORMULATION

Let $\{y_n\}$ be a sequence of real-valued signals representing the observations obtained at time n, and $\{\theta_n\}$ be the time-varying true parameter, an $\mathbb{R}^r$-valued random process. Suppose that

$$y_n = \varphi_n' \theta_n + e_n \qquad (1)$$

where $\varphi_n \in \mathbb{R}^r$ is the regression vector, $\theta_n \in \mathbb{R}^r$ is the vector-valued parameter process and $e_n \in \mathbb{R}$ represents the zero mean observation noise. Throughout this paper, we assume that θ_n is a discrete-time Markov chain.

A. Assumptions on the Markov Chain θ_n

(A1): Suppose that there is a small parameter $\varepsilon > 0$ and that θ_n is a discrete-time homogeneous Markov chain, whose state–space is

$$\mathcal{M} = \{a_1, \ldots, a_m\}, \; a_i \in \mathbb{R}^r, \qquad i = 1, \ldots, m \qquad (2)$$

and whose transition probability matrix is given by

$$P^\varepsilon = I + \varepsilon Q \qquad (3)$$

where I is an $\mathbb{R}^{m \times m}$ identity matrix and $Q = (q_{ij}) \in \mathbb{R}^{m \times m}$ is an irreducible generator (i.e., Q satisfies $q_{ij} \geq 0$ for $i \neq j$ and $\sum_{j=1}^m q_{ij} = 0$ for each $i = 1, \ldots, m$) of a continuous-time Markov chain. For simplicity, assume the initial distribution $P(\theta_0 = a_i) = p_{0,i}$ to be independent of ε for each $i = 1, \ldots, m$, where $p_{0,i} \geq 0$ and $\sum_{i=1}^m p_{0,i} = 1$.

B. LMS Algorithm

The LMS algorithm is a constant step size stochastic approximation algorithm that recursively operates on the observation sequence $\{y_n\}$ to generate a sequence of parameter estimates $\{\widehat{\theta}_n\}$ according to

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + \mu \varphi_n (y_n - \varphi_n' \widehat{\theta}_n) \qquad (4)$$

where $\mu > 0$ denotes the small constant step size.

Using (1) with $\widetilde{\theta}_n = \widehat{\theta}_n - \theta_n$ denoting the estimation error, we obtain

$$\widetilde{\theta}_{n+1} = \widetilde{\theta}_n - \mu \varphi_n \varphi_n' \widetilde{\theta}_n + \mu \varphi_n e_n + (\theta_n - \theta_{n+1}). \qquad (5)$$

Our task is to figure out the bounds on the deviation $\widetilde{\theta}_n$.

Remark 2.1: The parameter θ_n is termed a hypermodel in [2]. Although the dynamics (3) of the hypermodel θ_n will be used in our analysis, the implementation of the LMS algorithm (4), does not require any explicit knowledge of these dynamics.

C. Assumptions on the Signals

Let $\mathcal{F}_n$ be the σ-algebra generated by $\{(\varphi_j, e_j), \theta_j : j < n; \theta_n\}$, and denote the conditional expectation with respect to $\mathcal{F}_n$ by E_n. We will use the following conditions on the signals.

(A2): The sequences $\{\varphi_n\}$ and $\{e_n\}$ are independent of θ, the parameter. Either i) $\{\varphi_n, e_n\}$ is a sequence of bounded signals such that there is a symmetric and positive–definite matrix $A \in \mathbb{R}^{r \times r}$ satisfying that for each $0 < T < \infty$

$$\left| \sum_{j=n}^\infty E_n[\varphi_j \varphi_j' - A] \right| \leq K, \quad \left| \sum_{j=n}^\infty E_n \varphi_j e_j \right| \leq K \qquad (6)$$

or ii) $\{\varphi_n, e_n\}$ is a martingale difference sequence satisfying $E|\varphi_n|^{4+\Delta} < \infty$ and $E|\varphi_n e_n|^{2+\Delta} < \infty$ for some $\Delta > 0$.

Remark 2.2: The signal models include a large class of practical applications. Conditions (6) are modeled after mixing processes. This allows us to work with correlated signals whose remote past and distant future are asymptotically independent. To obtain the desired result, the distribution of the signal need not be known. The boundedness is a mild restriction, for example, one may consider truncated Gaussian process, etc. Moreover, dealing with recursive procedures in practice, in lieu of (4), one often uses a projection or truncation algorithm of the form

$$\widehat{\theta}_{n+1} = \pi_H[\widehat{\theta}_n + \mu \varphi_n (y_n - \varphi_n' \widehat{\theta}_n)] \qquad (7)$$

where π_H is a projection operator and H is a bounded set. When the iterates is outside of H, it will be projected back to the constrained set H. Extensive discussions for such projection algorithms can be found in [16]. On the other hand, for the possibly unbounded signals, we can treat martingale difference sequences. With some modification, such an approach can also be used to treat moving average type signals; see [30] for such an approach. In the subsequent development, we will deal with processes satisfying (6). The proof for the unbounded martingale difference sequence is slightly simpler. For brevity, henceforth, we will omit the verbatim proof for such processes.

III. UNSCALED TRACKING ERRORS

A. Mean Square Error Bounds

This section establishes a mean square error estimate, $E|\widehat{\theta}_n - \theta_n|^2$, for the LMS algorithm. We use a stability argument involving perturbed Liapunov function methods [16].

Theorem 3.1: Under conditions (A1) and (A2), for sufficiently large n

$$E|\widetilde{\theta}_n|^2 = E|\widehat{\theta}_n - \theta_n|^2 = O\left(\mu + \varepsilon + \frac{\varepsilon^2}{\mu}\right). \qquad (8)$$

In view of Theorem 3.1, to balance the terms in $O(\mu + \varepsilon + \varepsilon^2/\mu)$, we need to have $\varepsilon = O(\mu)$. Therefore, we arrive at the following corollary.

Corollary 3.2: Under the conditions of Theorem 3.1, if $\varepsilon = O(\mu)$, then for sufficiently large n, $E|\widetilde{\theta}_n|^2 = O(\mu)$.

Thus, the mean square tracking error is of the order $O(\mu)$ given that the step size $\mu = O(\varepsilon)$, i.e., given that the true parameter evolves at the rate as the adaptation speed of the LMS algorithm. For notational simplicity, in the rest of this paper we assume $\mu = \varepsilon$. Naturally, all the following results hold for $\mu = O(\varepsilon)$ (with obvious modifications), i.e., the main assumption is that the LMS algorithm and true parameter process $\{\theta_n\}$ evolve on the same time scale.

B. Limit System of Switching ODE

As is well known since the 1970s, the limiting behavior of a stochastic approximation algorithm (such as the LMS algorithm) is typically captured by a system of deterministic ODEs—indeed, this is the basis of the widely used "ODE approach" for convergence analysis of stochastic approximation algorithms—see [16]. Here we show the somewhat unusual result that the limiting behavior for the LMS algorithm (4) when the true parameter $\{\theta_n\}$ evolves according to (3) is not a deterministic ODE—instead it is a regime switching ODE modulated by a continuous-time Markov chain. To obtain the result, we use weak convergence methods, which require that first the tightness of the sequence be verified and then the limit be characterized via the martingale problem formulation.

We work with a piecewise interpolated sequence $\widehat{\theta}^\mu(\cdot)$ defined by $\widehat{\theta}^\mu(t) = \widehat{\theta}_n$ on $[\mu n, \mu n + \mu)$. Then $\widehat{\theta}^\mu(\cdot) \in D([0,T];\mathbb{R}^r)$, the space of $\mathbb{R}^r$-valued functions defined on $[0,T]$ that have right continuous sample paths and left-hand limits endowed with the Skorohod topology (see [3], [5], and [16]). Similarly, we define the interpolated process $\theta^\mu(\cdot)$ for the time-varying parameter by $\theta^\mu(t) = \theta_n$ for $t \in [\mu n, \mu n + \mu)$. The rest of the section is divided into three parts. The first subsection presents preliminary results concerning the asymptotics of $\theta^\mu(\cdot)$; the second subsection derives a tightness result; the third subsection obtains the limit switching ODE.

1) Asymptotic Results of the Interpolated Process $\theta^\mu(\cdot)$: This section is concerned with limit results of the interpolation $\theta^\mu(\cdot)$ of the Markov parameter process θ_n. The main ideas follow from the recent progress in two-time-scale Markov chains; see [33]–[35] and the references therein.

Lemma 3.3: Assume (A1) and $\varepsilon = \mu$. For the Markovian parameter θ_n. The following assertions hold: i) Denote $p_n^\mu = (P(\theta_n = a_1), \ldots, P(\theta_n = a_m))$. Then

$$p_n^\mu = z(t) + O\left(\mu + \exp\left(\frac{-k_0 t}{\mu}\right)\right), \ z(t) \in \mathbb{R}^{1 \times m}$$

$$\frac{dz(t)}{dt} = z(t)Q, z(0) = p_0$$

$$(P^\mu)^n = Z(t) + O\left(\mu + \exp\left(\frac{-k_0 t}{\mu}\right)\right)$$

$$\frac{dZ(t)}{dt} = Z(t)Q, \ Z(0) = I. \qquad (9)$$

ii) The continuous-time interpolation $\theta^\mu(\cdot)$ converges weakly to $\theta(\cdot)$, a continuous-time Markov chain generated by Q.

Proof: The proof of the first assertion is in [34] and the second one can be derived from [35]. $\qquad\square$

2) Tightness: We work with the pair $(\widehat{\theta}^\mu(\cdot), \theta^\mu(\cdot))$. The desired tightness is essentially a compactness result. This is stated next followed by the proof.

Theorem 3.4: Assume (A1) and (A2). Then, the pair $(\widehat{\theta}^\mu(\cdot), \theta^\mu(\cdot))$ is tight in $D([0,T];\mathbb{R}^r \times \mathcal{M})$.

Proof: By virtue of Lemma 3.3, $\{\theta^\mu(\cdot)\}$ is tight. In view of the Crámer and Wold Lemma [3], it suffices to verify the tightness of $\widehat{\theta}^\mu(\cdot)$. Applying the tightness criterion [13, p. 47], it suffices that for any $\delta > 0$ and $t, s > 0$ with $s < \delta$ such that $\lim_{\delta \to 0} \limsup_{\mu \to 0} E|\widehat{\theta}^\mu(t+s) - \widehat{\theta}^\mu(t)|^2 = 0$. Direct calculation leads to

$$E|\widehat{\theta}^\mu(t+s) - \widehat{\theta}^\mu(t)|^2 \leq K\mu^2\left(\frac{t+s}{\mu} - \frac{t}{\mu}\right)^2 = O(s^2). \ (10)$$

By first taking $\limsup_{\mu \to 0}$ and then $\lim_{\delta \to 0}$ in (10), the desired tightness follows. $\qquad\square$

3) Switching ODE Limit: We have demonstrated that $(\widehat{\theta}^\mu(\cdot), \theta^\mu(\cdot))$ is tight. Since tightness is equivalent to sequential compactness on any complete separable metric space, by virtue of the Prohorov's theorem, we may extract a weakly convergent subsequence. For notational simplicity, still denote the subsequence by $(\widehat{\theta}^\mu(\cdot), \theta^\mu(\cdot))$ with limit denoted by $(\widehat{\theta}(\cdot), \theta(\cdot))$. We proceed to identify the limit process. Assume the following condition.

(A2'): Condition (A2) holds with (6) replaced by

$$\frac{1}{n}\sum_{j=\ell}^{\ell+n} E_\ell \varphi_j \varphi_j' \to A, \quad \frac{1}{n}\sum_{j=\ell}^{\ell+n} E_\ell \varphi_j e_j \to 0 \qquad (11)$$

both in probability as $n \to \infty$ uniformly in ℓ.

Theorem 3.5: Assume (A1) and (A2'). Then, $(\widehat{\theta}^\mu(\cdot), \theta^\mu(\cdot))$ converges weakly to $(\widehat{\theta}(\cdot), \theta(\cdot))$ such that $\theta(t)$ is the continuous-time Markov chain given in Lemma 3.3 and $\widehat{\theta}(\cdot)$ is the solution of the following regime switching ODE:

$$\frac{d\widehat{\theta}(t)}{dt} = A(\theta(t) - \widehat{\theta}(t)), \ \widehat{\theta}(0) = \widehat{\theta}_0 \qquad (12)$$

modulated by the continuous-time Markov chain $\theta(t)$.

Proof: The proof is relegated to the Appendix . $\qquad\square$

Remark 3.6: A distinctive feature of the limit dynamic system (12) is that it is not deterministic—due to the presence

of the modulating Markov jump process $\theta(\cdot)$ the dynamic system randomly changes its regime. This is different from most results in the stochastic approximation literature, where the limit is a deterministic ODE.

IV. LIMIT DISTRIBUTION OF A SCALED TRACKING ERROR SEQUENCE

Here, we analyze the asymptotic distribution of the sequence of tracking errors $\{\widehat{\theta}_n - \theta_n\}$. We wish to find suitable scaling factor γ so that $(\widehat{\theta}_n - \theta_n)/\mu^\gamma$ converges to a nontrivial limit. In view of Corollary 3.2, the natural scaling is $\gamma = 1/2$. Again, we work with a continuous-time interpolation to better describe the evolution of the tracking errors. Note that our analysis focuses on the evolution of $\widehat{\theta}_n$ and the quantity being centered is the true parameter process θ_n, which is time varying and is stochastic. The scaling factor together with the asymptotic covariance matrix can be viewed as a "figure of merit" for tracking. That is, not only does it tell how close the algorithm tracks the true parameter, but also yields the asymptotic convergence rate of the tracking algorithm in terms of the asymptotic distribution of the tracking errors. Indeed, we show that $(\widehat{\theta}_n - \theta_n)/\sqrt{\mu}$ is asymptotically normal from which the error covariance can be obtained.

A. Limit Distribution

Recall $\widetilde{\theta}_n = \widehat{\theta}_n - \theta_n$ and denote $u_n = \widetilde{\theta}_n/\sqrt{\mu}$. Owing to (5), we obtain

$$u_{n+1} = u_n - \mu\varphi_n\varphi_n' u_n + \sqrt{\mu}\varphi_n e_n - \frac{\theta_{n+1} - \theta_n}{\sqrt{\mu}}. \quad (13)$$

By virtue of Theorem 3.1 and Corollary 3.2, and noting the quadratic structure of the Liapunov function, there is a $K_\mu > 0$ such that $\{u_n : n \geq K_\mu\}$ is tight. To further reveal the asymptotic properties of the tracking error, define the piecewise constant interpolation $u^\mu(t) = u_n$, for $t \in [\mu(n - K_\mu), \mu(n - K_\mu + 1))$. It is easily seen that for any $t, s > 0$, by telescoping

$$u^\mu(t) = u_{K_\mu} - \sum_{k=K_\mu}^{t/\mu - 1} [\mu\varphi_k\varphi_k' u_k - \sqrt{\mu}\varphi_k e_k] - \frac{\theta_{t/\mu} - \theta_{K_\mu}}{\sqrt{\mu}}. \quad (14)$$

To examine the asymptotic distribution of $u^\mu(\cdot)$, we state a Lemma, which is essentially a second moment bounds for the iterates and the interpolation. Its proof together with the proofs of Lemma 4.3 and Theorem 4.4 is in the Appendix.

Lemma 4.1: The following moment bounds hold:

$$\sup_{0 \leq n \leq T/\mu} E|u_n|^2 < \infty, \quad \sup_{0 \leq t \leq T} E|u^\mu(t)|^2 < \infty. \quad (15)$$

With the previous *a priori* estimates at hand, we proceed to establish the tightness of $\{u^\mu(\cdot)\}$. We first give another condition. Then, we state a lemma concerning the tightness.

(A3): The $\{\varphi_n\varphi_n'\}$ and $\{\varphi_n e_n\}$ are stationary uniform ϕ-mixing signals with mixing measure γ_k (see [5, Ch. 7]) satisfying $E\varphi_n e_n = 0$ and $E\varphi_n\varphi_n' = A$. The mixing measure γ_k satisfies $\sum_k \gamma_k^{1/2} < \infty$.

Remark 4.2: By (A3), the sequence $\sqrt{\mu}\sum_{k=0}^{t/\mu - 1}\varphi_k e_k$ converges to $\Sigma^{1/2}w(t)$, where $w(\cdot)$ is a standard Brownian motion and Σ is a positive–definite matrix.

Lemma 4.3: Assume (A1)–(A3). Then, $\{u^\mu(\cdot)\}$ is tight in $D([0, T]; \mathbb{R}^r)$. Moreover, any limit process has continuous sample paths w.p. 1.

Since $\{u^\mu(\cdot)\}$ is tight, by Prohorov's theorem, we may extract weakly convergent subsequences. Select such a sequence, still index it by μ for simplicity, and denote the limit process by $u(\cdot)$. By Skorohod representation, we may assume that $u^\mu(\cdot)$ converges to $u(\cdot)$ w.p. 1 and the convergence is uniform on any bounded time interval. In addition, the limit process has continuous sample paths w.p. 1. Our task is to identify the limit process. We will show that the limit process turns out to be a diffusion process.

Theorem 4.4: Under the conditions of Lemma 4.3, $u^\mu(\cdot)$ converges weakly to $u(\cdot)$ such that $u(\cdot)$ is the solution of the stochastic differential equation

$$du = -Au\,dt + \Sigma^{1/2}dw \quad (16)$$

where $w(\cdot)$ is a standard Brownian motion and

$$\Sigma = E\varphi_0\varphi_0' e_0^2 + \sum_{k=1}^{\infty} E\varphi_k\varphi_0' e_k e_0 + \sum_{k=1}^{\infty} E\varphi_0\varphi_k' e_k e_0.$$

Note that (16) is linear so it admits a unique solution for each initial condition. The stationary covariance of the aforementioned diffusion process is the solution of the Lyapunov equation

$$AS + SA = -\Sigma. \quad (17)$$

Note also $S = \int_0^\infty \exp(-At)\Sigma\exp(-At)dt$. In view of Theorem 4.4, the tracking errors $\{\widehat{\theta}_n - \theta_n\}$ is asymptotically normal with mean 0 and covariance μS.

B. Iterate Averaging and Minimal Window of Averaging

In this section, we illustrate the use of iterate averaging for tracking the Markov parameter θ_n. Iterate averaging was originally proposed in [23] for accelerating the convergence of stochastic approximation algorithms. It is well known [16] that for a constant parameter, i.e., $\varepsilon = 0$ in (3), and decreasing step size [e.g., $\mu = \mu_n = 1/n^\gamma$ with $\gamma < 1$ in (4)], iterate averaging results in asymptotically optimal convergence rate, i.e., identical to recursive least squares (which is a matrix step size algorithm). In the tracking case for a random walk time-varying parameter, in general iterate averaging does not result in an optimal tracking algorithm [19]. In light of Theorem 4.4, it is shown below for the slow Markov chain parameter that iterate averaging results in an asymptotically optimal tracking algorithm.

The rationale in using iterate averaging is to reduce the stationary covariance. To see how we may incorporate this into the current setup, we begin with a related algorithm

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + \frac{\Gamma}{n}\varphi_n(y_n - \varphi_n'\widehat{\theta}_n)$$

where Γ is an $r \times r$ matrix. Redefine $u_n = \sqrt{n}(\widehat{\theta}_n - \theta_n)$ for the previous algorithm. Set $t_n = \sum_{j=1}^{n}(1/j)$ and let $u^0(t)$ be the piecewise constant interpolation of u_n on $[t_n, t_{n+1})$ and $u^n(t) = u^0(t + t_n)$. Then, using an analogous argument, we arrive at $u^n(\cdot)$ converges weakly to $u(\cdot)$, which is the solution of the stochastic differential equation

$$ du = \left(-\Gamma A + \frac{I}{2} \right) u\,dt + \Gamma\Sigma^{1/2}dw. \tag{18} $$

Note that $(-\Gamma A + I/2)$ and $\Gamma\Sigma^{1/2}$ replace $-A$ and $\Sigma^{1/2}$ in (16). The additional term $I/2$ is due to the use of step size $O(1/n)$ [16, p. 329]. Minimizing the stationary covariance

$$ \int_0^\infty \exp\left[\left(-\Gamma A + \frac{I}{2} \right) t \right] \Gamma A \Gamma' \exp\left[\left(-A\Gamma' + \frac{I}{2} \right) t \right] dt $$

of the diffusion given in (18) with resect to the matrix parameter Γ leads to the "optimal" covariance $A^{-1}\Sigma A^{-1}$. In view of the previous discussion, we consider the iterate averaging algorithm with fixed step size $\mu > 0$ as

$$ \widehat{\theta}_{n+1} = \widehat{\theta}_n + \mu\varphi_n(y_n - \varphi'_n\widehat{\theta}_n) \quad \Theta_n = \frac{\mu}{t}\sum_{j=n_0}^{n_1+t/\mu-1}\widehat{\theta}_j. \tag{19} $$

We are taking the averaging for a sufficiently large n_0. Note that the averaging is taken with a window of width $O(1/\mu)$, so-called minimal window width of averaging (see [16, Ch. 11]). To analyze the algorithm, we let t_1 and t_2 be nonnegative real numbers that satisfy $t_1 + t_2 = t$, define

$$ \Theta_n = \frac{\mu}{t}\sum_{j=t_\mu-t_2/\mu}^{t_\mu+t_1/\mu-1}\widehat{\theta}_j, \text{ and } \overline{\Theta}_n = \frac{\mu}{t}\sum_{j=t_\mu-t_2/\mu}^{t_\mu+t_1/\mu-1}(\widehat{\theta}_j - \theta_j) \tag{20} $$

where $t_\mu \to \infty$ as $\mu \to 0$. Then, consider the scaled cumulative tracking error

$$ \overline{u}^\mu(t) = \frac{\sqrt{\mu}}{t}\sum_{j=t_\mu-t_2/\mu}^{t_\mu+t_1/\mu-1}(\widehat{\theta}_j - \theta_j). \tag{21} $$

Using a similar argument as [16, p. 379], we obtain the following theorem.

Theorem 4.5: Assume (A1)–(A3) hold. For each fixed $t > 0$, $\overline{u}^\mu(t)$ converges in distribution to a normal random vector $\overline{u}(t)$ with mean 0 and covariance $A^{-1}\Sigma A^{-1}/t + O(1/t^2)$.

Remark 4.6: Note that $A^{-1}\Sigma A^{-1}/t$ is the optimal asymptotic covariance of recursive least squares when estimating a constant parameter. Thus, iterate averaging over a window of $O(1/\mu)$ when $\varepsilon = O(\mu)$ results in an tracking algorithm with asymptotically optimal covariance.

V. EXAMPLES: STATE ESTIMATION OF HMMs AND ADAPTIVE INTERFERENCE SUPPRESSION IN CDMA WIRELESS NETWORKS

A. HMM With Infrequent Jumps

We analyze the performance of the LMS algorithm with step size μ for tracking the state of an HMM [4], where the underlying m-state *slowly* varying Markov chain's transition probability is of the form $I + \varepsilon Q$ with $\varepsilon = O(\mu)$ being a small pa-

rameter. The optimal HMM state filter (which yields the conditional mean estimate of the state) requires $O(m^2)$ computations at each time instant and, hence, intractable for very large m. For sufficiently small ε, it might be expected that LMS would do a reasonable job tracking the underlying Markov chain since the states change infrequently. The LMS algorithm requires $O(1)$ computational complexity for an m-state HMM (i.e., the complexity is independent of m). Recently, an $O(m)$ complexity asymptotic (steady-state) HMM state filter was proposed in [6]; see also [26]. It is therefore of interest to analyze the performance of an LMS algorithm (in terms of error probabilities) for tracking a time-varying HMM with infrequent jump changes.

A conventional HMM [4], [25] comprising of a finite-state Markov chain observed in noise is of the form (1) where $\varphi_n = 1$ for all n and the states of the Markov chain a_i, $1 \le i \le m$ are scalars. For this HMM case, the LMS algorithm (4) has complexity $O(1)$, i.e., independent of m. With $\widehat{\theta}_n$ denoting of (4) denoting the "soft" valued estimate of the state of the Markov chain θ_n, let $\widehat{\theta}_n^H$ denote the "hard" valued estimate of θ_n obtained by quantizing $\widehat{\theta}_n$ to the nearest Markov state, i.e.,

$$ \widehat{\theta}_n^H = a_{i^*} \quad \text{where } i^* = \arg\min_{1 \le i \le m}|a_i - \widehat{\theta}_n|. \tag{22} $$

Assume that the zero mean scalar noise process $\{e_n\}$ in (1) has finite variance σ_e^2.

Error Probability for Slow Hidden Markov Model: For notational convenience assume that the m Markov states of the above HMM are arranged in ascending order and are equally spaced, i.e., $a_1 < a_2 < \cdots < a_m$, and $d = a_{i+1} - a_i$ is a positive constant. Equation (17) implies that $\widetilde{\Sigma} = \sigma_e^2/2$ (that is a scalar). The probability of error can be computed as follows:

$$ P(\widehat{\theta}_n^H \ne \theta_n) $$
$$ = \sum_{i=1}^{m} P(\widehat{\theta}_n^H \ne \theta_n \mid \theta_n = a_i)P(\theta_n = a_i) $$
$$ = P\left(\widehat{\theta}_n - \theta_n > \frac{d}{2} \mid \theta_n = a_1 \right) P(\theta_n = a_1) $$
$$ + \sum_{i=2}^{m-1} P\left(|\widehat{\theta}_n - \theta_n| > \frac{d}{2} \mid \theta_n = a_i \right) P(\theta_n = a_i) $$
$$ + P\left(\widehat{\theta}_n - \theta_n < -\frac{d}{2} \mid \theta_n \ne a_m \right) P(\theta_n = a_m) $$
$$ = z^1(t)\Phi^c\left(\frac{d}{2\sqrt{\widetilde{\Sigma}\mu}} \right) + 2(z^2(t) + \cdots + z^{m-1}(t)) $$
$$ \times \Phi^c\left(\frac{d}{2\sqrt{\widetilde{\Sigma}\mu}} \right) + z^m(t)\Phi^c\left(\frac{d}{2\sqrt{\widetilde{\Sigma}\mu}} \right) $$
$$ = (2 - z^1(t) - z^m(t))\Phi^c\left(\frac{d}{2\sqrt{\widetilde{\Sigma}\mu}} \right) $$

where $\Phi^c(\cdot)$ is the complementary Gaussian distribution function, $t = \mu n$, $z^i(t)$ denotes the probabilities of the continuous-time Markov chain $\theta(t) = a_i$ for $i = 1, \ldots, m$; see (9).

Error Probability and Iterate Averaging: Iterate averaging (Section IV-B) can be used for vector state HMMs to reduce

the error probability of estimates generated by the LMS algorithm. Recall that the purpose of iterate averaging is to construct a scalar step size LMS algorithm that leads to the same asymptotic convergence rate as a matrix step size algorithm for estimating a vector parameter. (Iterate averaging is, thus, only useful when the parameter θ_n is vector valued). Thus, we consider a vector state HMM.

As is well known in digital communications, for more than two vector states, it is difficult to construct an explicit expression for the error probability as it involves multidimensional Gaussian integrals over half spaces (apart from special cases, e.g., if the vectors are symmetric about the origin). Thus, for convenience, we assume the underlying Markov chain has two states ($\mathcal{M} = \{a_1, a_2\}$), $a_i \in \mathbb{R}^q$, and $\varphi_n \in \mathbb{R}^q$ is a sequence of i.i.d. random variables independent of θ_n and e_n.

Define $d = a_2 - a_1$ and define the hard valued estimates $\widehat{\theta}_n^H = a_{i^*}$, where $i^* = \arg\min_{1 \le i \le m} \|a_i - \widehat{\theta}_n\|$. The probability of error of the LMS algorithm in tracking this vector state HMM is

$$
\begin{aligned}
P(\widehat{\theta}_n^H \ne \theta_n) &= P\left(|\widehat{\theta}_n - a_1| > |\widehat{\theta}_n - a_2|\Big|\theta_n = a_1\right) \\
&\quad \cdot P(\theta_n = a_1) + P\left(|\widehat{\theta}_n - a_2| > |\widehat{\theta}_n - a_1|\Big|\theta_n = a_2\right) \\
&\quad \cdot P(\theta_n = a_2).
\end{aligned}
\tag{23}
$$

However, since $|\widehat{\theta}_n - a_1|^2 - |\widehat{\theta}_n - a_2|^2 = 2(\widehat{\theta}_n - a_1)'d - d'd$, we have $P(|\widehat{\theta}_n - a_1| > |\widehat{\theta}_n - a_2||\theta_n = a_1) = P((\widehat{\theta}_n - a_1)'d > d'd/2)$.

First, let us compute the error probability for the state estimate $\widehat{\theta}_n^H$ generated by the LMS algorithm without using iterate averaging. The weak convergence result Theorem 4.4 implies that $[\widehat{\theta}_n - \theta_n]$ is asymptotically normal with mean 0 and variance $\mu\widetilde{\Sigma}$. This implies that conditional on $\theta_n = a_1$, $(\widehat{\theta}_n - \theta_n)'d \sim N(0, \mu d'\widetilde{\Sigma}d)$. Substituting in (23) yields the following error probability for the state estimates generated by the LMS algorithm (without iterate averaging)

$$
P(\widehat{\theta}_n^H \ne \theta_n) = \Phi^c\left(\frac{\frac{d'd}{2}}{\sqrt{\mu d'\widetilde{\Sigma}d}}\right).
\tag{24}
$$

Now, consider computing the error probability when iterate averaging is used. With iterate averaging, using Theorem 4.5, the covariance $\widetilde{\Sigma}$ in (24) is replaced by iterate averaged covariance $A^{-1}\Sigma A^{-1}$ where A is defined in (A3) of Section IV – thus the error probability for state estimates generated by the iterate averaged LMS algorithm satisfies

$$
P(\widehat{\theta}_n^H \ne \theta_n) = \Phi^c\left(\frac{\frac{d'd}{2}}{\sqrt{\mu d'A^{-1}\Sigma A^{-1}d}}\right).
\tag{25}
$$

Since $\widetilde{\Sigma} - A^{-1}\Sigma A^{-1} > 0$ (positive definite) and Φ^c is monotonically decreasing, the error probability of the state estimates generated by LMS with iterate averaging (25) is lower than that for the LMS algorithm without iterate averaging (24).

B. Effect of Admission/Access Control on Adaptive Multiuser Detector

In this section, we examine the tracking performance of an adaptive linear multiuser detector in a cellular DS/CDMA wireless network when the profile of active users changes due to an admission or access (scheduling) controller at the base station. The main point is to show that in many cases, the optimal linear minimum mean square error (LMMSE) multiuser detector [29] varies according to a finite Markov chain—hence, the previous weak convergence analysis for the LMS algorithm directly applies to the corresponding adaptive linear multiuser detector which aims to track the LMMSE detector coefficients.

Consider a synchronous DS-CDMA system with a maximum of K users and an additive white Gaussian noise channel. Here, user 1 is the one of interest. Let $\mathcal{P}(X)$ denote the power set of an arbitrary finite set X. For user 1, the set of all possible combinations of active users (interferers) is (where $\emptyset$ denotes the null set, i.e., no interferer) $\mathcal{P}(\{2, \ldots, K\}) = \{\emptyset, \{2\}, \{2, 3\}, \ldots \{2, 3, 4, \ldots, K\}\}$. Let $\overline{K}_n$ denote a finite-state (set-valued) discrete time process that evolves on the state–space $\mathcal{P}(\{2, \ldots, K\})$. Thus at each time instant n, $\overline{K}_n$ denotes the set of active users.

After the received continuous-time signal is preprocessed and sampled at the CDMA receiver (the received signal is passed through a chip-matched filter followed by a chip-rate sampler), the resulting discrete-time received signal at time n, denoted by r_n, is given by (see [22] for details)

$$
r_n = A(1)b_n(1)s(1) + \sum_{k \in \overline{K}_n} A(k)b_n(k)s(k) + \varpi_n.
\tag{26}
$$

Here, r_n is an N-dimensional vector; N is called the processing (spreading) gain; $s(k)$ is an N-vector denoting the normalized signature sequence of the kth user, so that $s'(k)s(k) = 1$; $b_n(k)$ denotes the data bit of the kth user transmitted at time n; $A(k)$ is the received amplitude of the kth user; ϖ_n is a sequence of white Gaussian vectors with mean zero and covariance matrix σI where I denotes the $N \times N$ identity matrix and $\sigma > 0$ is a scalar. It is assumed that the discrete-time stochastic processes $\{b_n(k)\}$, and $\{\varpi_n\}$ are mutually independent, and that $\{b_n(k)\}$ is a collection of independent equiprobable ± 1 random variables.

Specification of Active Users: We assume that the network admission/access controller operates on a slower time scale (e.g., multiframe by multiframe basis) than the bit duration, i.e., the finite-state process $\{\overline{K}_n\}$ evolves according to a slower time scale than the bits $\{b_n[k]\}$. This is usual in DS/CDMA systems where typically a user arrives or departs after several multiframes (i.e., several hundreds of bits). Then, $\overline{K}_n$ can be modeled as a slow finite-state Markov chain with transition probability matrix $I + \varepsilon Q$ in the following examples.

i) Consider a single class of users (e.g., voice) with Poisson arrival rate λ and exponential departure rate μ. Then, the active users form a continuous time Markov chain (birth death process) with state space $\mathcal{P}(\{2, \ldots, K\})$ and generator Q. The time sampled version, sampled at the chip-rate ε, is then a slow Markov chain with $P^\varepsilon = I + \varepsilon Q$.

ii) Markov decision based admission control of multiclass users: The formulation in [27] considers admission control in a multiservice CDMA network comprising voice and data users. Assuming a linear multiuser detector at the receiver, the admission controller aims to minimize the blocking probability of users seeking to access the network subject to signal to interference ratio (quality-of-service) constraints on the active users $\overline{K}_n$. Assuming that the arrival rate of voice and data users are Poisson and departure rates of active users are exponential, the problem of devising the optimal admission policy is formulated in [27] as a semi-Markov decision process with exponential holding times. Again, assuming that the arrival and departure of users are at a slower time scale (e.g., several frames) than the bit duration, a time sampled version of this continuous-time process at chip rate ε results in a slow Markov chain.

Here, we consider the effect of the above Markovian or periodic admission/access control strategies on two types of adaptive multiuser detectors—decision directed receiver and precombining receiver. In both cases, the optimal receiver weight coefficients evolve according to a finite-state Markov chain and the adaptive linear receiver is a LMS algorithm which attempts to track this Markovian weight vector.

Adaptive Decision Directed Multiuser Detection: We assume that user 1 is the user of interest. Assuming knowledge of the active user set $\overline{K}_n$, the optimal linear multiuser detector seeks to compute the weight vector c_n^* such that

$$c_n^* = \arg\min_c E_{\overline{K}_n} \{Ab_n(1) - c'r_n\}^2 \qquad (27)$$

where $b_n(1)$ is a training data sequence (or else the estimates of the bits when the receiver is operating in the decision directed mode). As shown in [24], $c_n^* = R_n^{-1}s_1$ where $R_n = E_{\overline{K}_n}\{r_n r_n'\}$. Given $\overline{K}_n$ is a slow Markov chain, it follows from (26) that R_n and, thus, the optimal weight vector c_n^* are also $2^{(K-1)}$ state slow finite-state Markov chains, respectively. It is clear that the above formulation is identical to the signal model (1) with $y_n = Ab_n(1)$ (observation), $\theta_n = c_n^*$ (slow Markov chain parameter), $\varphi_n = r_n$ (regression vector). Indeed $\{e_n\}$, with $e_n = y_n - \varphi_n'\theta_n$, is a sequence of i.i.d. random variables due to the orthogonality principle of Wiener filters.

Now, consider the *adaptive* multiuser detection problem where the active user set $\overline{K}_n$ is not known. Thus $\{y_n\}$ is the observation sequence of an HMM. Hence, in principle, the optimal (conditional mean) estimate of $\overline{K}_n$ and therefore c_n^* given the observation history $(y_1, \ldots, y_n)$ can be computed using the HMM state filter. However, due to the large state–space (exponential in the number of users K), this is computationally prohibitive. For this reason, the adaptive linear multiuser detector [29] uses the LMS algorithm (4) to minimize (27) without taking into account the Markovian dynamics of $\theta_n = c_n^*$, i.e., θ_n is the hypermodel (see Remark 2.1). For such an adaptive linear multiuser detector, the weak convergence analysis in Theorem 4.4 implies that if $\varepsilon = \mu$, the estimate $\widehat{c}_n$ of the adaptive multiuser detector is approximately normally distributed with mean c_n^* and covariance $\widetilde{\Sigma}$.

Precombining Adaptive Multiuser Detectors for Fading Channels: A performance analysis of MMSE receivers for frequency selective fading channels is presented in [21]. In general, the optimal receiver weight coefficient c_n^* of the LMMSE receiver varies rapidly in time depending on the instantaneous channel values. Here, we consider a receiver structure, developed in [17], called a precombining LMMSE receiver (also called LMMSE-RAKE receiver) which results in the optimal receiver weight vector c_n^* evolving according to a slow finite-state Markov chain.

The continuous-time received signal for a frequency selective fading channel has the form

$$r(t) = \sum_{n=0}^{N_b-1} \sum_{k \in \overline{K}_n} \sum_{l=1}^{L} A(k)b_n(k)c_n(k,l)s_{t-nT-\tau_{k,l}}(k) + \varpi(t)$$

where T denotes the symbol interval, L is the number of propagation paths, $\varpi(t)$ is a complex-valued zero mean additive white Gaussian noise with variance σ^2, $c_n(k,l)$ is the complex attenuation factor for the kth user and lth path, and $\tau_{k,l}$ is the propagation delay. The received discrete-time signal over a data block of N_b symbols after antialias filtering and sampling at the rate $T_s = T/(SG)$ (where S is the number of samples per chip, G is the number of chips per symbol) is (see [17] for details) $r_n = \mathbf{S}\mathbf{C}_n\mathbf{A}_n\mathbf{b}_n + \mathbf{n}_n \in \mathbb{C}^{SGN_b}$, where $\mathbf{S}$ is the sampled spread sequence matrix, $\mathbf{C}_n$ is the channel coefficient matrix, $\mathbf{A}_n$ is a matrix of received amplitudes (the time variation in the notation is because inactive users, i.e., $k \notin \overline{K}_n$ are considered to have zero amplitude, thus, $\mathbf{A}_n$ is a slow finite-state Markov chain), $\mathbf{b}_n$ is the data vector and $\mathbf{n}_n$ is the complex-valued channel noise vector. Assuming knowledge of the active users $\overline{K}_n$, the precombining LMMSE receiver seeks to find c_n^* to minimize $E_{\overline{K}_n}\{\mathbf{C}_n\mathbf{A}_n\mathbf{b}_n - c'r_n\}^2$. The optimal receiver is $c_n^* = \mathbf{S}(\mathbf{S}'\mathbf{S} + \sigma^2\Sigma_n^{-1})^{-1} \in \mathbb{R}^{SGN_b \times KLN_b}$ where Σ_n is the covariance matrix of $\mathbf{C}_n\mathbf{A}_n\mathbf{b}_n$ which consists of transmitted user powers and average channel tap powers. As remarked in [17], this shows that the precombining LMMSE receiver no longer depends on the instantaneous values of the channel complex coefficients but on the average power profiles of the channel Σ_n. Thus c_n^* is a finite-state Markov chain.

In the case when the active users $\overline{K}_n$ are unknown, i.e., θ_n is the hypermodel (see Remark 2.1), the adaptive precombining LMMSE receiver uses the LMS algorithm to optimize $E\{\mathbf{C}_n\mathbf{A}_n\mathbf{b}_n - c'r_n\}^2$. This is again of the form (4) and the weak convergence tracking analysis of Theorem 4.4 applies.

Iterate Averaging: In all three cases with $\mu = O(\varepsilon)$, iterate averaging (Section IV-B) over a window of $O(1/\mu)$ results in an adaptive receiver with asymptotically optimal convergence rate.

VI. Switching Diffusion Limit for θ-Dependent Regression Vector

Up to this point, we have derived a number of properties for the tracking algorithms. One of the premise is that the signals are independent of the parameter process θ_n, which does cover important cases such as HMM arising in many applications. One of the interesting aspects of the results is that the scaled sequence

of tracking errors leads to a diffusion limit. However, there are other important cases involving regression of $y_{n-1}, y_{n-2}, \ldots$. In such a case, φ_n is θ-dependent.

A. Example: Linear Autoregressive Processes With Markov Regime

A linear rth order autoregressive process with Markov regime is of the form (1), where the regressor $\varphi_n = (y_{n-1}, \ldots, y_{n-r})$ and the jump Markov parameter vector $\theta_n \in \mathcal{M}$ with each $a_i \in \mathbb{R}^r$ being the coefficients of the AR model for regime i, $i = 1, \ldots, m$. We refer the reader to [9], [11], [12], and the references therein for extensive examples of such processes in modeling business cycles in econometrics, failure detection, and target tracking.

In this section, we demonstrate how to generalize the results obtained thus far to such parameter-dependent processes. One of the main techniques is the use of fixed θ-process. We assume that

$$P(\varphi_n \in \cdot \mid \varphi_j, e_j, \theta_j : j < n, \theta_n) = P(\varphi_n \in \cdot \mid \varphi_j, e_j, : j < n, \theta_n).$$

Denote $E_j \varphi_j \varphi_j' = \eta_j$. Define a fixed-$\theta$ process $\{\eta_j(\theta), j \geq n\}$ for each n with initial condition $\eta_n(\theta_n) = \eta_n$. That is, the process starts at value η_n at time n and then evolves as if the parameter were fixed at θ_n for all $n \leq j \leq T/\varepsilon$. Similarly, define $\zeta_j = E_j \varphi_j e_j$ and $\zeta_j(\theta)$ to be $\zeta_n(\theta_n) = \zeta_n$.

In what follows, we provide the conditions needed, give the main results, and illustrate how the proofs of previous theorems can be modified for this more general case. We will use the following assumptions.

(A4): $\{\varphi_n\}$ and $\{e_n\}$ are sequences of bounded signals such that: i) $\{e_n\}$ is uniform mixing and is independent of the parameter process; ii) for each $\theta \in \mathcal{M}$, there is a symmetric and positive–definite matrix $A(\theta) \in \mathbb{R}^{r \times r}$ satisfying that for each $0 < T < \infty$

$$\left| \sum_{j=n}^{\infty} E_n [\eta_j(\theta) - A(\theta)] \right| \leq K \quad \left| \sum_{j=n}^{\infty} E_n \zeta_j(\theta) \right| \leq K \quad (28)$$

and iii) as $n \to \infty$, the following limits in probability exist uniformly in ℓ:

$$\frac{1}{n} \sum_{j=\ell}^{\ell+n} E_\ell \eta_j(\theta) \to A(\theta) \quad \frac{1}{n} \sum_{j=\ell}^{\ell+n} E_\ell \zeta_j(\theta) \to 0. \quad (29)$$

With the conditions given previously, we can derive the following results; see Appendix for its proof.

Theorem 6.1: Assume (A1) and (A4). Then, i) the conclusion of Theorem 3.1 continue to hold; ii) the conclusion of Theorem 3.5 continue to hold with the ODE (12) replaced by

$$\frac{d\widehat{\theta}(t)}{dt} = A(\theta(t))(\theta(t) - \widehat{\theta}(t)), \ \widehat{\theta}(0) = \widehat{\theta}_0 \quad (30)$$

and iii) the conclusion of Theorem 4.4 continues to hold with (16) replaced by the jump Markov linear diffusion

$$du = -A(\theta(t))u dt + \Sigma^{1/2}(\theta(t)) dw \quad (31)$$

where, for $\theta \in \mathcal{M}$

$$\Sigma(\theta) = E\zeta_0(\theta)\zeta_0'(\theta) + \sum_{k=1}^{\infty} E\zeta_k(\theta)\zeta_0'(\theta) + \sum_{k=1}^{\infty} E\zeta_0(\theta)\zeta_k'(\theta).$$

Lemma 6.2: Suppose that there exist m symmetric positive–definite matrices $B_i, i = 1, 2, \ldots, m$, and positive constant $s > 0$ satisfying

$$-u'B_i A(a_i)u + \frac{1}{s} u'B_i u \sum_{j=1}^{m} q_{ij} \left(\frac{u'B_j u}{u'B_i u} \right)^{s/2} \leq \gamma|u|^2 \quad (32)$$

for all $u \in \mathbb{R}^r$, $u \neq 0$, $i = 1, \ldots, m$. Then, there exists an invariant distribution ν for the switching diffusion u_t defined in (31), such that as $t \to \infty$, the law of u_t converges to ν independently of the initial condition u_0. Moreover, all moments $\kappa > 0$ of ν are finite.

Proof: According to [1, Th. 3.1], a sufficient condition for the existence of an invariant distribution with finite moment $\kappa > 0$ is that (32) holds. Choose $B_i = I$ for all i. Then, since $\sum_j q_{ij} = 0$, the left-hand side becomes $-u'A(a_i)u$. Since $A(a_i)$, $i = 1, \ldots, m$ are positive–definite, clearly $-u'A(a_i)u < -\gamma|u|^2$ for some $\gamma > 0$ for all $s > 0$. $\square$

Comparing with the results in the previous sections, the signals depending on θ results in different asymptotic behavior. In Section IV, the approximation yields a diffusion limit. This picture changes for θ-dependent signals. In lieu of a diffusion limit, a switching diffusion limit is obtained. This reveals the fundamental difference of θ-dependent and θ-independent processes. Note that when the signals are independent of the parameter process, Theorem 6.1 reduces to the previous diffusion limit in Theorem 4.4.

VII. Conclusion

We analyzed the tracking properties of the least mean square algorithm LMS algorithm with step size μ for tracking a parameter that evolves according to a Markov chain with transition probability matrix $I + \varepsilon Q$, where $\varepsilon = O(\mu)$, i.e., the parameter evolves as fast as the LMS algorithm with infrequent jumps but possibly large jump sizes. To the best of our knowledge, the results we derived such as switching ODE limit and switching diffusions associated with the tracking problem appear to be the first one of their types. We illustrated these results in low complexity HMM state estimation (when the underlying Markov chain evolves slowly) and adaptive multiuser detection in DS-CDMA multiuser detection. These results are also useful in other applications of recursive estimation in signal processing and communication networks. While we have focused exclusively on the LMS algorithm in this paper, the ideas can be extended to adaptive estimation via the recursive least squares (RLS) algorithm with forgetting factor for tracking a slow Markov chain. It is a worthwhile extension to see if with forgetting factor λ such that $\varepsilon = O(\lambda)$, whether the limiting behavior of the RLS algorithm also exhibits Markov modulated regime switching. Finally, while the application presented in Section V-B dealt with adaptive multiuser detection, the analysis can also be applied to equalization of Markov modulated communication channels. In particular, using a similar analysis

to this paper it can be shown that blind adaptive equalization algorithms such as the constant modulus algorithm (CMA) [10] also exhibit a switching regime ODE limit for a Markov modulated channel.

APPENDIX
PROOFS OF RESULTS

A. Proof of Theorem 3.1

Define $V(x) = (x'x)/2$. Direct calculation leads to

$$
\begin{aligned}
&E_n V(\widetilde{\theta}_{n+1}) - V(\widetilde{\theta}_n) \\
&= E_n \widetilde{\theta}'_n \left[-\mu\varphi_n\varphi'_n\widetilde{\theta}_n + \mu\varphi_n e_n + (\theta_n - \theta_{n+1}) \right] \\
&\quad + E_n| -\mu\varphi_n\varphi'_n\widetilde{\theta}_n + \mu\varphi_n e_n + (\theta_n - \theta_{n+1})|^2. \quad (33)
\end{aligned}
$$

In view of the Markovian assumption, the independence of the Markov chain with the signals $\{(\varphi_n, e_n)\}$, and the structure of the transition probability matrix given by (3)

$$
\begin{aligned}
E_n(\theta_n - \theta_{n+1}) &= \sum_{i=1}^{m} E(a_i - \theta_{n+1} \mid \theta_n = a_i) I_{\{\theta_n = a_i\}} \\
&= \sum_{i=1}^{m} \left[a_i - \sum_{j=1}^{m} a_j(\delta_{ij} + \varepsilon q_{ij}) \right] I_{\{\theta_n = a_i\}} = O(\varepsilon).
\end{aligned}
$$
(34)

Note that we are taking conditional expectation w.r.t. $\mathcal{F}_n$, and $I_{\{\theta_n = a_i\}}$ is $\mathcal{F}_n$-measurable. Similarly, we also have

$$
\begin{aligned}
&E_n|\theta_n - \theta_{n+1}|^2 \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} |a_i - a_j| \, I_{\{\theta_n = a_i\}} P(\theta_{n+1} = a_j|\theta_n = a_i) \\
&= \sum_{j=1}^{m}\sum_{i=1}^{m} |a_i - a_j| I_{\{\theta_n = a_i\}}(\delta_{ij} + \varepsilon q_{ij}) = O(\varepsilon). \quad (35)
\end{aligned}
$$

Using an elementary inequality $ab \leq (a^2 + b^2)/2$ for two real numbers a and b, we have $|\widetilde{\theta}_n| = |\widetilde{\theta}_n| \cdot 1 \leq (|\widetilde{\theta}_n|^2 + 1)/2$, so

$$
O(\varepsilon)|\widetilde{\theta}_n| \leq O(\varepsilon)(V(\widetilde{\theta}_n) + 1). \quad (36)
$$

By virtue of the boundedness of the signal $\{(\varphi_n, e_n)\}$

$$
\begin{aligned}
&E_n| -\mu\varphi_n\varphi'_n\widetilde{\theta}_n + \mu\varphi_n e_n + (\theta_n - \theta_{n+1})|^2 \\
&\quad = E_n|\theta_n - \theta_{n+1}|^2 + O(\mu^2 + \mu\varepsilon)(V(\widetilde{\theta}_n) + 1). \quad (37)
\end{aligned}
$$

Using (35)–(37), we obtain

$$
\begin{aligned}
&E_n V(\widetilde{\theta}_{n+1}) - V(\widetilde{\theta}_n) \\
&= E_n \widetilde{\theta}'_n \left[-\mu\varphi_n\varphi'_n\widetilde{\theta}_n + \mu\varphi_n e_n + (\theta_n - \theta_{n+1}) \right] \\
&\quad + E_n|\theta_n - \theta_{n+1}|^2 + O(\mu^2 + \mu\varepsilon)(V(\widetilde{\theta}_n) + 1). \quad (38)
\end{aligned}
$$

To proceed, we need to "average out" the terms in the next to the last line and the first term on the last line of (38). This is

accomplished by using perturbed Liapunov functions. To do so, define a number of perturbations of the Lyapunov function by

$$
V_1^\varepsilon(\widetilde{\theta}, n) = -\mu \sum_{j=n}^{\infty} E_n \widetilde{\theta}'(\varphi_j\varphi'_j - A)\widetilde{\theta}
$$

$$
V_2^\varepsilon(\widetilde{\theta}, n) = \mu \sum_{j=n}^{\infty} \widetilde{\theta}' E_n \varphi_j e_j
$$

$$
V_3^\varepsilon(\widetilde{\theta}, n) = \sum_{j=n}^{\infty} \widetilde{\theta}' E_n(\theta_j - \theta_{j+1})
$$

$$
V_4^\varepsilon(n) = \sum_{j=n}^{\infty} E_n(\theta_n - \theta_{n+1})'(\theta_j - \theta_{j+1}). \quad (39)
$$

For each $\widetilde{\theta}$, by virtue of (A2), it is easily verified that $\mu \left| \sum_{j=n}^{\infty} [E_n\varphi_j\varphi'_j - A] \right| |\widetilde{\theta}|^2 \leq O(\mu)(V(\widetilde{\theta}) + 1)$, so

$$
|V_1^\varepsilon(\widetilde{\theta}, n)| \leq O(\mu)(V(\widetilde{\theta}) + 1). \quad (40)
$$

Similarly, for each $\widetilde{\theta}$

$$
|V_2^\varepsilon(\widetilde{\theta}, n)| \leq O(\mu)(V(\widetilde{\theta}) + 1). \quad (41)
$$

Note that the irreducibility of Q implies that of $I + \varepsilon Q$. Thus, there is an N_ε such that for all $n \geq N_\varepsilon, |(I + \varepsilon Q)^n - \mathbb{1}\nu_\varepsilon| \leq K\varepsilon$, where ν_ε denotes the stationary distribution associated with the transition matrix $I + \varepsilon Q$. Using telescoping and the aforementioned estimates, for all $N_1 \geq n \geq N_\varepsilon$

$$
\left| \sum_{j=n}^{N_1} \widetilde{\theta}' E_n(\theta_j - \theta_{j+1}) \right| = \left| \widetilde{\theta}' E_n(\theta_n - \theta_{N_1}) \right| \leq O(\varepsilon)(V(\widetilde{\theta}) + 1)
$$

and, hence

$$
|V_3^\varepsilon(\widetilde{\theta}, n)| \leq O(\varepsilon)(V(\widetilde{\theta}) + 1). \quad (42)
$$

Likewise, it can be shown that

$$
|V_4^\varepsilon(n)| = O(\varepsilon). \quad (43)
$$

Note also that

$$
\begin{aligned}
&E_n V_1^\varepsilon(\widetilde{\theta}_{n+1}, n + 1) - V_1^\varepsilon(\widetilde{\theta}_n, n) \\
&= E_n V_1^\varepsilon(\widetilde{\theta}_{n+1}, n + 1) - E_n V_1^\varepsilon(\widetilde{\theta}_n, n + 1) \\
&\quad + E_n V_1^\varepsilon(\widetilde{\theta}_n, n + 1) - V_n^\varepsilon(\widetilde{\theta}_n, n). \quad (44)
\end{aligned}
$$

It follows that

$$
E_n V_1^\varepsilon(\widetilde{\theta}_n, n + 1) - V_1^\varepsilon(\widetilde{\theta}_n, n) = \mu E_n \widetilde{\theta}'_n(\varphi_n\varphi'_n - A)\widetilde{\theta}_n \quad (45)
$$

by virtue of (A2). In addition

$$
\begin{aligned}
&E_n V_1^\varepsilon(\widetilde{\theta}_{n+1}, n + 1) - E_n V_1^\varepsilon(\widetilde{\theta}_n, n + 1) \\
&= -\mu \sum_{j=n+1}^{\infty} E_n(\widetilde{\theta}_{n+1} - \widetilde{\theta}_n)'[E_{n+1}\varphi_j\varphi'_j - A]\widetilde{\theta}_{n+1} \\
&\quad - \mu \sum_{j=n+1}^{\infty} E_n\widetilde{\theta}'_n[E_{n+1}\varphi_j\varphi'_j - A](\widetilde{\theta}_{n+1} - \widetilde{\theta}_n). \quad (46)
\end{aligned}
$$

Using (5), similar estimates as that of (35) yields

$$E_n|\widetilde{\theta}_{n+1} - \widetilde{\theta}_n| \leq \mu E_n|\varphi_n\varphi_n'||\widetilde{\theta}_n| + \mu E_n|\varphi_n e_n| + O(\varepsilon)$$
$$= O(\mu)(V(\widetilde{\theta}_n) + 1) + O(\varepsilon). \quad (47)$$

Moreover

$$\left|\mu \sum_{j=n+1}^{\infty} E_n\widetilde{\theta}_n'[E_{n+1}\varphi_j\varphi_j' - A](\widetilde{\theta}_{n+1} - \widetilde{\theta}_n)\right|$$
$$\leq O(\mu^2 + \mu\varepsilon)(V(\widetilde{\theta}_n) + 1) \quad (48)$$

and

$$\left|-\mu \sum_{j=n+1}^{\infty} E_n(\widetilde{\theta}_{n+1} - \widetilde{\theta}_n)'[E_{n+1}\varphi_j\varphi_j' - A]\widetilde{\theta}_{n+1}\right|$$
$$\leq O(\mu^2 + \mu\varepsilon)(V(\widetilde{\theta}_n) + 1). \quad (49)$$

Therefore, we obtain

$$E_n V_1^{\varepsilon}(\widetilde{\theta}_{n+1}, n+1) - V_1^{\varepsilon}(\widetilde{\theta}_n, n)$$
$$= \mu E_n\widetilde{\theta}_n'(\varphi_n\varphi_n' - A)\widetilde{\theta}_n + O(\mu^2 + \mu\varepsilon)(V(\widetilde{\theta}_n) + 1). \quad (50)$$

Analogous estimate for $V_2^{\varepsilon}(\widetilde{\theta}_n, n)$ leads to

$$E_n V_2^{\varepsilon}(\widetilde{\theta}_{n+1}, n+1) - V_2^{\varepsilon}(\widetilde{\theta}_n, n)$$
$$= -\mu E_n\widetilde{\theta}_n'\varphi_n e_n + O(\mu^2 + \mu\varepsilon)(V(\widetilde{\theta}_n) + 1). \quad (51)$$

Likewise, it can be shown that

$$E_n[V_3^{\varepsilon}(\widetilde{\theta}_{n+1}, n+1) - V_3^{\varepsilon}(\widetilde{\theta}_n, n+1)] = O(\varepsilon^2 + \mu^2)(V(\widetilde{\theta}_n) + 1)$$
$$E_n V_3^{\varepsilon}(\widetilde{\theta}_n, n+1) - V_3^{\varepsilon}(\widetilde{\theta}_n, n) = -E_n\widetilde{\theta}_n'(\theta_n - \theta_{n+1})$$
$$(52)$$

and

$$E_n[V_4^{\varepsilon}(n+1) - V_4^{\varepsilon}(n)] = -E_n|\theta_n - \theta_{n+1}|^2 + O(\varepsilon^2). \quad (53)$$

Define

$$W(\widetilde{\theta}, n) = V(\widetilde{\theta}) + V_1^{\varepsilon}(\widetilde{\theta}, n) + V_2^{\varepsilon}(\widetilde{\theta}, n) + V_3^{\varepsilon}(\widetilde{\theta}, n) + V_4^{\varepsilon}(n).$$

Then, using (38) and (50)–(53), upon cancellation, we obtain

$$E_n W(\widetilde{\theta}_{n+1}, n+1) - W(\widetilde{\theta}_n, n)$$
$$\leq -\mu\widetilde{\theta}_n'A\widetilde{\theta}_n + O(\mu^2 + \varepsilon^2)(V(\widetilde{\theta}_n) + 1)$$
$$\leq -\lambda\mu V(\widetilde{\theta}_n) + O(\mu^2 + \varepsilon^2)(V(\widetilde{\theta}_n) + 1)$$
$$\leq -\lambda\mu W(\widetilde{\theta}_n, n) + O(\mu^2 + \varepsilon^2)(W(\widetilde{\theta}_n, n) + 1) \quad (54)$$

for some $\lambda > 0$. In (54), from the second line to the third line, since A is positive definite, there is a $\lambda > 0$ such that $\widetilde{\theta}'A\widetilde{\theta} \geq \lambda V(\widetilde{\theta})$. From the third to the last line, we used estimates (40)–(43) and replaced $V(\widetilde{\theta}_n)$ by $W(\widetilde{\theta}_n, n)$, which results in an $O(\mu\varepsilon)$ term by the boundedness of $\widetilde{\theta}_n$; we also used $O(\mu\varepsilon) = O(\mu^2 + \varepsilon^2)$ via an elementary inequality.

Choose μ and ε small enough so that there is a $\lambda_0 > 0$ satisfying $\lambda_0 \leq \lambda$ and $-\lambda\mu + O(\mu^2) + O(\varepsilon^2) \leq -\lambda_0\mu$. Then we obtain $E_n W(\widetilde{\theta}_{n+1}, n+1) \leq (1 - \lambda_0\mu)W(\widetilde{\theta}_n, n) + O(\mu^2 + \varepsilon^2)$.

Taking this expectation and iterating on the resulting inequality yield

$$EW(\widetilde{\theta}_{n+1}, n+1) \leq (1 - \lambda_0\mu)^{n - N_\varepsilon} EW(\widetilde{\theta}_{N_\varepsilon}, N_\varepsilon) + O\left(\mu + \frac{\varepsilon^2}{\mu}\right).$$

By taking n large enough, we can make $(1 - \lambda_0\mu)^{n - N_\varepsilon} \leq O(\mu)$. Thus $EW(\widetilde{\theta}_{n+1}, n+1) \leq O(\mu + \varepsilon^2/\mu)$. Finally, applying (40)–(43) again, we also obtain $EV(\widetilde{\theta}_{n+1}) \leq O(\mu + \varepsilon + \varepsilon^2/\mu)$. Thus, the desired result follows. $\qquad\square$

B. Proof of Theorem 3.5

We will show that the pair of stochastic processes $(\widehat{\theta}^{\mu}(\cdot), \theta^{\mu}(\cdot))$ converges weakly to $(\widehat{\theta}(\cdot), \theta(\cdot))$ that is a solution of the martingale problem associated with the operator D defined by

$$Df(\widehat{\theta}, a_i) = f_{\widehat{\theta}}'(\widehat{\theta}, a_i)A(a_i - \widehat{\theta}) + Qf(\widehat{\theta}, \cdot)(a_i) \quad (55)$$

for each $a_i \in \mathcal{M}$, where

$$Qf(\widehat{\theta}, \cdot)(a_i) = \sum_{j=1}^{m} q_{ij}f(\widehat{\theta}, a_j) = \sum_{j \neq i} q_{ij}(f(\widehat{\theta}, j) - f(\widehat{\theta}, a_i))$$
$$(56)$$

and $f(\cdot, a_i)$ is a real-valued and twice continuously differentiable test function with compact support. Note that D is the generator for the process $(\widehat{\theta}(t), \theta(t))$.

We note that by virtue an argument similar to [33, Lemma 7.18], it can be verified that the martingale problem associated with the operator D has a unique solution. To complete the proof, we need only prove that the limit $(\widehat{\theta}(\cdot), \theta(\cdot))$ is indeed the solution of the martingale problem, i.e., for each $a_i \in \mathcal{M}$, for any twice continuously differentiable function $f(\cdot, a_i)$, $f(\widehat{\theta}(t), \theta(t)) - \int_0^t Df(\widehat{\theta}(v), \theta)v))dv$ is a continuous-time martingale. To verify this, it suffices to show that for any positive integer k_0, any $t, s \in [0, T]$ and $t + s \in [0, T]$, and $0 < t_j \leq t$ with $j \leq k_0$, and any bounded and continuous function $\rho_j(\cdot, a_i)$ for each $a_i \in \mathcal{M}$

$$E\prod_{j=1}^{k_0} \rho_j(\widehat{\theta}(t_j), \theta(t_j))\left[f(\widehat{\theta}(t+s), \theta(t+s)) - f(\widehat{\theta}(t), \theta(t))\right.$$
$$\left. - \int_t^{t+s} Df(\widehat{\theta}(u), \theta(u))du\right] = 0. \quad (57)$$

Indeed, as described in [5, p. 174], $(\widehat{\theta}(\cdot), \theta(\cdot))$ is a solution to the martingale problem if and only if (57) is satisfied. To verify (57), in the rest of the proof we work with the processes indexed by μ and derive (57) by taking the limit as $\mu \to 0$. This task is divided into several sub-tasks.

Step 1: For notational simplicity, denote

$$L_{k_0} = \prod_{j=1}^{k_0} \rho_j(\widehat{\theta}(t_j), \theta(t_j)), \quad L_{k_0}^{\mu} = \prod_{j=1}^{k_0} \rho_j(\widehat{\theta}^{\mu}(t_j), \theta^{\mu}(t_j)).$$
$$(58)$$

By the weak convergence of $(\widehat{\theta}^{\mu}(\cdot), \theta^{\mu}(\cdot))$ to $(\widehat{\theta}(\cdot), \theta(\cdot))$ and the Skorohod representation [3], [5], [16], as $\mu \to 0$

$$EL_{k_0}^{\mu}[f(\widehat{\theta}^{\mu}(t+s), \theta^{\mu}(t+s)) - f(\widehat{\theta}^{\mu}(t), \theta^{\mu}(t))]$$
$$\to EL_{k_0}[f(\widehat{\theta}(t+s), \theta(t+s)) - f(\widehat{\theta}(t), \theta(t))]. \quad (59)$$

Step 2: To proceed, choose a sequence n_μ such that $n_\mu \to \infty$ as $\mu \to 0$, but $\delta_\mu = \mu n_\mu \to 0$, and divide $[t, t+s]$ into intervals of width δ_μ. Then, the following is obtained:

$$EL_{k_0}^\mu [[f(\widehat{\theta}^\mu(t+s), \theta^\mu(t+s)) - f(\widehat{\theta}^\mu(t), \theta^\mu(t))]$$

$$= EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} [f(\widehat{\theta}_{ln_\mu+n_\mu}, \theta_{ln_\mu+n_\mu}) - f(\widehat{\theta}_{ln_\mu+n_\mu}, \theta_{ln_\mu})] \right.$$

$$\left. + \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} [f(\widehat{\theta}_{ln_\mu+n_\mu}, \theta_{ln_\mu}) - f(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})] \right]. \quad (60)$$

Step 3a: To complete the proof, we work on the last two lines of the right-hand side of (60) separately. By virtue of the smoothness and boundedness of $f(\cdot, \theta)$ (for each $\theta \in \mathcal{M}$), it can be seen that as $\mu \to 0$

$$EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} [f(\widehat{\theta}_{ln_\mu+n_\mu}, \theta_{ln_\mu+n_\mu}) - f(\widehat{\theta}_{ln_\mu+n_\mu}, \theta_{ln_\mu})] \right]$$

$$= EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} [f(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu+n_\mu}) - f(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})] \right] + o_\mu$$

where $o_\mu \to 0$ as $\mu \to 0$. Thus, we need only work with the remaining term on the right-hand side of the aforementioned equation. Letting $\mu \to 0$ and $\mu ln_\mu \to u$

$$EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} [f(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu+n_\mu}) - f(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})] \right]$$

$$= EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu \left[\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} Qf(\widehat{\theta}_{ln_\mu}, \cdot)(\theta_{ln_\mu}) \right] \right] + o_\mu$$

$$\to EL_{k_0} \left[\int_t^{t+s} Qf(\widehat{\theta}(u), \theta(u)) du \right]. \quad (61)$$

Step 3b: Since $\widehat{\theta}_{ln_\mu}^\mu$ and θ_{ln_μ} are $\mathcal{F}_{ln_\mu}$-measurable, by virtue of the continuity and boundedness of $f_{\widehat{\theta}}'(\cdot, \theta)$; see the equation

at the bottom of the page, where $o_\mu \to 0$ as $\mu \to 0$. We claim that as $\mu \to 0$

$$EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\delta_\mu f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k y_k \right]$$

$$\to EL_{k_0} \int_t^{t+s} f_{\widehat{\theta}}'(\widehat{\theta}(u), \theta(u)) A\theta(u) du \quad (62)$$

and

$$EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\delta_\mu f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k \varphi_k' \widehat{\theta}_k \right]$$

$$\to EL_{k_0} \int_t^{t+s} f_{\widehat{\theta}}'(\widehat{\theta}(u), \theta(u)) A\widehat{\theta}(u) du. \quad (63)$$

Assume (62) and (63) hold temporarily. Then, (59) and (61), together with (62) and (63), imply (57), which completes the proof of Theorem 3.5.

Thus, it only remains to verify (62) and (63). To this end, replacing y_k by use of (1), we obtain

$$EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \delta_\mu \left[\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k y_k \right] \right]$$

$$= EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \delta_\mu \right.$$

$$\left. \times \left[\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k \varphi_k' \theta_k \right] \right] + o_\mu \quad (64)$$

since by virtue of (11)

$$EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \frac{\delta_\mu}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k e_k \to 0.$$

Using a partial summation

$$\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi_k' \theta_k = \frac{1}{n_\mu} \left[\sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi_k' \right] \theta_{ln_\mu+n_\mu-1}$$

$$+ \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-2} \left[\sum_{j=ln_\mu}^{k} \varphi_j \varphi_j' \right] (\theta_k - \theta_{k+1}). \quad (65)$$

$$EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} [f(\widehat{\theta}_{ln_\mu+n_\mu}, \theta_{ln_\mu}) - f(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})]$$

$$= EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\mu f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k (y_k - \varphi_k' \widehat{\theta}_k) \right] + o_\mu$$

$$= EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\frac{\delta_\mu f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k y_k \right]$$

$$- EL_{k_0}^\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\delta_\mu f_{\widehat{\theta}}'(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu}) \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k \varphi_k' \widehat{\theta}_k \right] + o_\mu$$

It follows that

$$EL_{k_0}^\mu \left[\sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 2} \left[\sum_{j=ln_\mu}^k \varphi_j \varphi'_j \right] \right.$$
$$\left. \times (\theta_k - \theta_{k+1}) \right]$$
$$= EL_{k_0}^\mu \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu}$$
$$\times \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 2} E_{ln_\mu} \left[\sum_{j=ln_\mu}^k \varphi_j \varphi'_j \right] E_k(\theta_k - \theta_{k+1})$$
$$\to 0 \quad \text{as} \quad \mu \to 0 \tag{66}$$

by virtue of (34). In similar spirit

$$\left| EL_{k_0}^\mu \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \right.$$
$$\left. \times \left[\sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \right] [\theta_{ln_\mu + n_\mu - 1} - \theta_{ln_\mu}] \right|$$
$$\leq \left| E|L_{k_0}^\mu| \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{|f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})|}{n_\mu} \right.$$
$$\left. \times E_{ln_\mu}^{1/2} \left| \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \right|^2 E_{ln_\mu}^{1/2} |\theta_{ln_\mu + n_\mu - 1} - \theta_{ln_\mu}|^2 \right|$$
$$\to 0 \quad \text{as} \quad \mu \to 0. \tag{67}$$

In addition, sending $l\delta_\mu \to u$, by the weak convergence of $(\widehat{\theta}^\mu(\cdot), \theta^\mu(\cdot))$ to $(\widehat{\theta}(\cdot), \theta(\cdot))$ and the Skorohod representation

$$EL_{k_0}^\mu \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \left[\sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \right] \theta_{ln_\mu}$$
$$= EL_{k_0}^\mu \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu$$
$$\times \sum_{i=1}^m \left[\frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \right] a_i I_{\{\theta^\mu(u) = a_i\}}$$
$$\to EL_{k_0} \int_t^{t+s} \sum_{i=1}^m f'_{\widehat{\theta}}(\widehat{\theta}(u), a_i) A a_i I_{\{\theta(u) = a_i\}} du$$
$$= EL_{k_0} \int_t^{t+s} f'_{\widehat{\theta}}(\widehat{\theta}(u), \theta(u)) A\theta(u) du. \tag{68}$$

Thus, (62) holds. To verify (63), note that

$$\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \widehat{\theta}_k = \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \widehat{\theta}_{ln_\mu}$$
$$+ \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k (\widehat{\theta}_k - \widehat{\theta}_{ln_\mu}). \tag{69}$$

Thus

$$EL_{k_0}^\mu \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \left[\sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \right]$$
$$\times (\widehat{\theta}_k - \widehat{\theta}_{ln_\mu}) \to 0 \quad \text{as} \quad \mu \to 0$$

by using (4) and the boundedness of the signals.

Finally, as $\mu \to 0, l\delta_\mu \to u$

$$\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \widehat{\theta}_{ln_\mu} = \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \widehat{\theta}^\mu(l\delta_\mu) \tag{70}$$

and, hence

$$EL_{k_0}^\mu \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} \delta_\mu \frac{f'_{\widehat{\theta}}(\widehat{\theta}_{ln_\mu}, \theta_{ln_\mu})}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi'_k \widehat{\theta}_{ln_\mu}$$
$$\to EL_{k_0} \int_t^{t+s} f'_{\widehat{\theta}}(\widehat{\theta}(u), \theta(u)) A\widehat{\theta}(u) du.$$

Thus, (63) is verified and the proof of Theorem 3.5 is concluded. □

C. Proof of Lemma 4.1

Using (13), we obtain

$$u_{n+1} = u_0 - \sum_{k=0}^n [\mu \varphi_k \varphi'_k u_k - \sqrt{\mu} \varphi_k e_k] - \frac{\theta_{n+1} - \theta_0}{\sqrt{\mu}}.$$

Then, for $n + 1 \leq T/\mu$

$$E|u_{n+1}|^2 \leq K \left[E|u_0|^2 + \mu^2 E \left| \sum_{k=0}^n \varphi_j \varphi'_k u_k \right|^2 \right.$$
$$\left. + \mu E \left| \sum_{k=0}^n \varphi_k e_k \right|^2 + \frac{E|\theta_{n+1} - \theta_0|^2}{\mu} \right].$$

By virtue of the boundedness of the signal $\{\varphi_k\}$

$$\mu^2 E \left| \sum_{k=0}^n \varphi_j \varphi'_k u_k \right|^2 \leq K \mu^2 \sum_{k=0}^n \sum_{j=0}^n E|u_j||u_k|$$
$$\leq K\mu \sum_{k=0}^n E|u_k|^2. \tag{71}$$

Using the mixing property of $\{\varphi_k e_k\}$

$$\mu E \left| \sum_{k=0}^n \varphi_k e_k \right|^2 = \mu \sum_{k=0}^n \sum_{j=0}^n |E e_k \varphi'_k \varphi_j e_j - E e_k \varphi'_k E \varphi_j e_j|$$
$$\leq K. \tag{72}$$

Similar to (35), $E|\theta_{n+1} - \theta_0|^2 = O(\mu)$, so

$$\frac{E|\theta_{n+1} - \theta_0|^2}{\mu} \leq K. \tag{73}$$

Combining (71)–(73), $E|u_{n+1}|^2 \leq K + K\mu \sum_{k=0}^n E|u_k|^2$. An application of the well-known Gronwall's inequality leads to $E|u_{n+1}|^2 \leq K \exp(n\mu) \leq K \exp(T) \leq K$. Moreover, taking sup over n yields $\sup_{0 \leq n \leq T/\mu} E|u_n|^2 \leq K$. The *a priori* bound is obtained. □

D. Proof of Lemma 4.3

Step 1: Corresponding to (13), define

$$U_{n+1} = U_n - \mu \varphi_n \varphi_n' U_n + \sqrt{\mu} \varphi_n e_n, \ n \geq K_\mu$$

$$U_{K_\mu} = u_{K_\mu}$$

$$U^\mu(t) = U_n, \ t \in [\mu(n - K_\mu), \mu(n - K_\mu + 1)). \quad (74)$$

It can be shown that there is a K_μ (relabeling when needed) such that both $\{u_n : n \geq K_\mu\}$ and $\{U_n : k \geq K_\mu\}$ are tight. In addition, $\sup_{0 \leq n \leq T/\mu} E|U_n|^2 < \infty$. Note that

$$U^\mu(t) = U_{K_\mu} - \mu \sum_{k=K_\mu}^{t/\mu - 1} \varphi_k \varphi_k' U_k + \sqrt{\mu} \sum_{k=K_\mu}^{t/\mu - 1} \varphi_k e_k. \quad (75)$$

Using (14) and (75) and noting the boundedness of $\{\varphi_k\}$

$$E|U^\mu(t) - u^\mu(t)|$$

$$= E\left| (U_{K_\mu} - u_{K_\mu}) - \mu \sum_{k=K_\mu}^{t/\mu - 1} \varphi_k \varphi_k' (U_k - u_k) + \frac{\theta_{t/\mu} - \theta_{K_\mu}}{\sqrt{\mu}} \right|$$

$$\leq K\mu \sum_{k=K_\mu}^{t/\mu - 1} E|U_k - u_k| + O(\sqrt{\mu}). \quad (76)$$

Using (13) and (74) and applying Gronwall's inequality, we obtain

$$E|U^\mu(t) - u^\mu(t)| = O(\sqrt{\mu}). \quad (77)$$

Moreover, (77) holds uniformly for $t \in [0, T]$. Thus, to obtain the tightness of $\{u^\mu(\cdot)\}$, it suffices to consider $\{U^\mu(\cdot)\}$.

Step 2: Working with $\{U^\mu(\cdot)\}$, the rest of the proof uses techniques similar to that of Theorem 3.4. For any $\delta > 0$ and $t, s > 0$ with $s < \delta$, by virtue of the boundedness of the signal $\{\varphi_k\}$, Lemma 4.1, and the inequality $E|U_j||U_k| \leq E^{1/2}|U_j|^2 E^{1/2}|U_k|^2$, it is easily verified that

$$\lim_{\delta \to 0} \limsup_{\mu \to 0} E\left| \mu \sum_{k=t/\mu}^{(t+s)/\mu - 1} \varphi_k \varphi_k' U_k \right|^2$$

$$\leq K \lim_{\delta \to 0} \limsup_{\mu \to 0} \mu^2 \sum_{k=t/\mu}^{(t+s)/\mu - 1} \sum_{j=t/\mu}^{(t+s)/\mu - 1} E|U_j||U_k|$$

$$\leq K \lim_{\delta \to 0} \limsup_{\mu \to 0} \mu^2 \left[\frac{(t+s)}{\mu} - \frac{t}{\mu} \right]^2 = 0. \quad (78)$$

Next, using the mixing inequality [13, Lemma 4.4]

$$E\left| \sqrt{\mu} \sum_{k=t/\mu}^{(t+s)/\mu - 1} \varphi_k e_k \right|^2$$

$$= \mu \sum_{k=t/\mu}^{(t+s)/\mu - 1} \sum_{j=t/\mu}^{(t+s)/\mu - 1} E e_k \varphi_k' \varphi_j e_j$$

$$\leq \mu \sum_{k=t/\mu}^{(t+s)/\mu - 1} \sum_{j=t/\mu}^{(t+s)/\mu - 1} |E e_k \varphi_k' \varphi_j e_j - E e_k \varphi_k' E \varphi_j e_j|$$

$$\leq K\mu \sum_{j=t/\mu}^{(t+s)/\mu - 1} \sum_{k \geq j} \gamma_{k-j} \leq Ks \leq K\delta. \quad (79)$$

Thus, $\lim_{\delta \to 0} \limsup_{\mu \to 0} E\left| \sqrt{\mu} \sum_{k=t/\mu}^{(t+s)/\mu - 1} \varphi_k e_k \right|^2 = 0$, and as a result, $\lim_{\delta \to 0} \limsup_{\mu \to 0} E|u^\mu(t+s) - u^\mu(t)|^2 = 0$. By the tightness criterion, (78)–(79) yield the tightness of $\{U^\mu(\cdot)\}$. Moreover, similar as in the proof of (78)–(79), we can also show that $E|U^\mu(t+s) - U^\mu(t)|^4 \leq O(\delta^2)$. Thus, all limit of sample paths are continuous w.p. 1. Furthermore, Step 1 of the proof implies that the same assertion holds with $\{U^\mu\}$ replaced by $\{u^\mu(\cdot)\}$. The lemma is proved. $\square$

E. Proof of Theorem 4.4

By virtue of step 1) in the proof of Lemma 4.3, we need only work with the sequence $\{U^\mu(\cdot)\}$ since $u^\mu(\cdot)$ and $U^\mu(\cdot)$ have the same weak limit. By virtue of the paragraph preceding Theorem 4.4, it suffices to characterize the limit process $U(\cdot)$ owing to Lemma 4.3. We will show that $U(\cdot)$ is a solution of the martingale problem with the operator

$$Lh(u) = -h_u'(u) A u + \frac{1}{2} \text{tr}(h_{uu}(u) \Sigma) \quad (80)$$

for any twice continuously differentiable $h(\cdot)$ with compact support. The linearity in (16) implies that it has a unique solution for each initial condition so the solution of the martingale problem associated with the operator L defined in (80) is also unique.

Similar to the proof of Theorem 3.5 (although no jump terms are involved and a diffusion is added) we need to show that for any positive integer k_0, and $j \leq k_0$, and any real-valued bounded and continuous function $\rho_j(\cdot)$

$$E \prod_{j=1}^{k_0} \rho_j(U(t_j)) \Big[h(U(t+s)) - h(U(t))$$

$$- \int_t^{t+s} Lh(U(v)) dv \Big] = 0. \quad (81)$$

Again working with the process indexed by μ, choose a sequence $n_\mu \to \infty$ as $\mu \to 0$ and $\delta_\mu = \mu n_\mu \to 0$. We obtain

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} [h(U_{ln_\mu + n_\mu}) - h(U_{ln_\mu})] \right]$$

$$= E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left\{ \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} h_U'(U_{ln_\mu}) G_1 \right.$$

$$\left. + \sum_{ln_\mu = t/\mu}^{(t+s)/\mu - 1} G_1' h_{UU}(U_{ln_\mu}) G_1 \right\} \quad (82)$$

where

$$G_1 = -\mu \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k \varphi_k' U_k + \sqrt{\mu} \sum_{k=ln_\mu}^{ln_\mu + n_\mu - 1} \varphi_k e_k.$$

Note that in (82), the term involving $h_U'(U) G_1$ yields the limit drift, whereas the term consisting $G_1' h_{UU}(U) G_1$ leads to the limit diffusion coefficient.

For the drift term, using averaging procedure, as $\mu \to 0$

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[-\mu \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} h'_U(U_{ln_\mu}) \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right]$$

$$\to E \prod_{j=1}^{k_0} \rho_j(U(t_j)) \left[\int_t^{t+s} h'_U(U(v)) A U(v) dv \right] \quad (83)$$

and

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} h'_U(U_{ln_\mu}) \sqrt{\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k e_k \right]$$

$$= E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} h'_U(U_{ln_\mu}) \sqrt{\mu} \right.$$

$$\left. \times \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k e_k \right] \to 0. \quad (84)$$

For the diffusion term, as $\mu \to 0$, we have the estimates

$$E \left| \mu^2 \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right]' \left[\sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_j \varphi'_j U_j \right] \right|$$

$$\leq K \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu^2$$

$$\times \left[\frac{1}{n_\mu^2} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \frac{E|U_k|^2 + E|U_j|^2}{2} \right]$$

$$\leq K \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu^2 \to 0$$

and

$$E \left| \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\mu \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right]' \left[\sqrt{\mu} \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_j e_j \right] \right|$$

$$\leq \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} E \left| \mu \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right| \left| \sqrt{\mu} \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_j e_j \right|$$

$$\leq \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \frac{\delta_\mu^{3/2}}{2} \left[E \left| \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right|^2 \right.$$

$$\left. + E \left| \frac{1}{\sqrt{n_\mu}} \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_j e_j \right|^2 \right]$$

$$\leq K \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu^{3/2} \to 0.$$

Thus

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right]' \right.$$

$$\left. \times h_{UU}(U_{ln_\mu}) \left[\sqrt{\mu} \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_j \varphi'_j U_j \right] \right] \to 0$$

and

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \left[\mu \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k U_k \right]' \right.$$

$$\left. \times h_{UU}(U_{ln_\mu}) \left[\sqrt{\mu} \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_j e_j \right] \right] \to 0.$$

Therefore, in view of Remark 4.2, we arrive at

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \mu \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} e_k \varphi'_k \right.$$

$$\left. \times h_{UU}(U_{ln_\mu}) \varphi_j e_j \right]$$

$$= E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \mu \sum_{j=ln_\mu}^{ln_\mu+n_\mu-1} \right.$$

$$\left. \times \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \mathrm{tr}\{h_{UU}(U_{ln_\mu}) e_k \varphi_k \varphi'_j e_j\} \right]$$

$$\to E \prod_{j=1}^{k_0} \rho_j(U(t_j)) \left[\int_t^{t+s} \mathrm{tr}(h_{UU}(U(v))\Sigma) dv \right]$$

and, hence

$$E \prod_{j=1}^{k_0} \rho_j(U^\mu(t_j)) \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} G'_1 h_{UU}(U_{ln_\mu}) G_1 \right]$$

$$\to E \prod_{j=1}^{k_0} \rho_j(U(t_j)) \left[\int_t^{t+s} \mathrm{tr}(h_{UU}(U(v))\Sigma) dv \right]. \quad (85)$$

Combining (82)–(85), (81) holds. Finally, by virtue of (77), $u^\mu(\cdot)$ converges to the same limit as that of $U^\mu(\cdot)$. Thus, the desired result follows. □

F. Ideas of Proof of Theorem 6.1

Since the proofs are similar in spirit to the previous sections, we will be brief with only the distinct features noted and the details omitted. To prove i), as in Theorem 3.1, we still use perturbed Liapunov function methods. With $\theta_n = \theta$, modify the definition of $V_1^\varepsilon(\tilde{\theta}, n)$ by $V_1^\varepsilon(\tilde{\theta}, n) = -\mu \sum_{j=n}^\infty E_n \tilde{\theta}'[\eta_j(\theta) - A(\theta)]\tilde{\theta}$. Note that $E_n \eta_n(\theta_n) = \eta_n = E_n \varphi_n \varphi'_n$, which facilitates the desired cancellation. Note also since $\theta \in \mathcal{M}$, we can

still find a single λ such that $\widetilde{\theta}' A(\theta)\widetilde{\theta} \geq \lambda V(\widetilde{\theta})$. The rest of the proof is similar to the previous case.

To prove ii), we need to average out terms of the form $(1/n_\mu) \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k \widehat{\theta}_k$, and $(1/n_\mu) \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k \theta_k$. We illustrate how this can be done for the first expression shown previously. In view of our notation, $\eta_k = \eta_k(\theta_k) = E_k \varphi_k \varphi'_k$ and

$$
\frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k \varphi'_k \widehat{\theta}_k
$$
$$
= \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \varphi_k \varphi'_k \widehat{\theta}_{ln_\mu} + o(1)
$$
$$
= E_{ln_\mu} \frac{1}{n_\mu} \left[\sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \eta_k(\theta_{ln_\mu}) \right.
$$
$$
\left. + \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} [\eta_k(\theta_k) - \eta_k(\theta_{ln_\mu})] \right] \widehat{\theta}_{ln_\mu} + o(1) \quad (86)
$$

where $o(1) \to 0$ in probability. By virtue of the Markov property of $\{\theta_k\}$ and the boundedness of $\{\varphi_k \varphi'_k\}$

$$
E|\eta_k(\theta_k) - \eta_k(\theta_{ln_\mu})|
$$
$$
= \sum_{i,\ell}^{m} E|\eta(a_\ell) - \eta(a_i)| I_{\{\theta_k=a_\ell, \theta_{ln_\mu}=a_i\}}
$$
$$
= \sum_{i=1}^{m} \sum_{\ell \neq i}^{m} E|\eta(a_\ell) - \eta(a_i)| I_{\{\theta_k=a_\ell, \theta_{ln_\mu}=a_i\}}
$$
$$
\leq K \sum_{i=1}^{m} \sum_{\ell \neq i} \varepsilon q_{i\ell} = O(\varepsilon) = O(\mu).
$$

Thus, the last term in (86) does not contribute anything to the limit. Moreover, we obtain

$$
EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \eta_k(\theta_{ln_\mu}) \widehat{\theta}_{ln_\mu} \right]
$$
$$
= EL_{k_0}^\mu \left[\sum_{i=1}^{m} \sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \eta_k(a_i) \widehat{\theta}_{ln_\mu} \right.
$$
$$
\left. \times I_{\{\theta^\mu(\mu ln_\mu)=a_i\}} \right]
$$
$$
\to EL_{k_0} \left[\sum_{i=1}^{m} \int_t^{t+s} A(a_i) \widehat{\theta}(u) I_{\{\theta(u)=a_i\}} du \right]
$$
$$
= EL_{k_0} \left[\sum_{i=1}^{m} \int_t^{t+s} A(\theta(u)) \widehat{\theta}(u) du \right], \quad \text{and}
$$
$$
EL_{k_0}^\mu \left[\sum_{ln_\mu=t/\mu}^{(t+s)/\mu-1} \delta_\mu \frac{1}{n_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} E_{ln_\mu} \sum_{k=ln_\mu}^{ln_\mu+n_\mu-1} \varphi_k \varphi'_k \theta_k \right]
$$
$$
\to EL_{k_0} \left[\sum_{i=1}^{m} \int_t^{t+s} A(\theta(u)) \theta(u) du \right].
$$

The rest of the proof is similar to that of Theorem 3.5 with similar modifications as before. Part iii) can be proved using analogs modifications for the fixed-θ process shown previously.

REFERENCES

[1] G. K. Basak, A. Bisi, and M. K. Ghosh, "Stability of random diffusion with linear drift," *J. Math. Anal. Appl.*, vol. 202, pp. 604–622, 1996.

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

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

[4] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.

[5] S. N. Ethier and T. G. Kurtz, *Markov Processes—Characterization and Covergence*. New York: Wiley, 1986.

[6] G. Golubev and R. Khasminskii, "Asymptotic optimal filtering for a hidden Markov model," *Math. Meth. Statist.*, vol. 7, no. 2, pp. 192–208, 1998.

[7] L. Guo, "Stability of recursive stochastic tracking algorithms," *SIAM J. Control Optim.*, vol. 32, no. 5, pp. 1195–1225, Sep. 1994.

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

[9] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.

[10] C. R. Johnson Jr. *et al.*, "Blind equalization using the constant modulus algorithm: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[11] V. Krishnamurthy and T. Rydén, "Consistent estimation of linear and nonlinear autoregressive models with Markov regime," *J. Time Ser. Anal.*, vol. 19, pp. 291–308, May 1998.

[12] V. Krishnamurthy and G. Yin, "Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 458–476, Feb. 2002.

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

[14] H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. New York: Springer-Verlag, 1978.

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

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

[17] M. Latva-Aho, "LMMSE receivers for DS-CDMA systems in frequency selective fading channels," in *CDMA Techniques for Third Generaion Mobile Systems*. Norwell, MA: Kluwer, 1999, pp. 145–183.

[18] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Auto. Control*, vol. AC-22, no. 4, pp. 551–575, Apr. 1977.

[19] ——, "Recursive least-squares and accelerated convergence in stochastic approximation schemes," *Int. J. Adapt. Control Signal Process.*, vol. 15, no. 2, pp. 169–178, Mar. 2001.

[20] L. Ljung and S. Gunnarsson, "Adaptive tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–22, 1990.

[21] S. Miller, H. Honig, and L. B. Milstein, "Performance analysis of MMSE receivers for DS-CDMA in frequency selective fading channels," *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 1919–1929, Nov. 2000.

[22] T. Oskiper and H. V. Poor, "Online activity detection in a multiuser environment using the matrix CUSUM algorithm," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 477–491, Feb. 2002.

[23] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM J. Control Optim.*, vol. 30, no. 4, pp. 838–855, Jul. 1992.

[24] 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.

[25] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.

[26] L. Shue, B. D. O. Anderson, and F. De Bruyne, "Asymptotic smoothing errors for hidden Markov models," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3289–3302, Dec. 2000.

[27] S. Singh, V. Krishnamurthy, and H. V. Poor, "Integrated voice/data call admission control for wireless DS-CDMA systems with fading," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1483–1495, Jun. 2002.

[28] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms – Stability and Performance*. Upper Saddle River, NJ: Prentice-Hall, 1995.

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

[30] G. Yin, "Adaptive filtering with averaging," in *IMA Volumes in Mathematics and its Applications*, G. Goodwin, K. Aström, and P. R. Kumar, Eds. New York: Springer-Verlag, 1995, vol. 74, pp. 375–396.

[31] G. Yin and V. Krishnamurthy, "Analysis of LMS algorithm for Markovian parameters with infrequent jumps—Applications to tracking hidden Markov models and adaptive multiuser detection in DS/CDMA," *IEEE Trans. Inform. Theory*, Jul. 2005, to be published.

[32] G. Yin, V. Krishnamurthy, and C. Ion, "Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization," *SIAM J. Optim.*, vol. 14, pp. 1187–1215, 2004.

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

[34] ——, "Singularly perturbed discrete-time Markov chains," *SIAM J. Appl. Math.*, vol. 61, pp. 833–854, 2000.

[35] G. Yin, Q. Zhang, and G. Badowski, "Discrete-time singularly perturbed Markov chains: Aggregation, occupation measures, and switching diffusion limit," *Adv. Appl. Probab.*, vol. 35, no. 2, pp. 449–476, 2003.

Vikram Krishnamurthy (S'90–M'91–SM'99–F'05) was born in 1966. He received the B.S. degree from the University of Auckland, New Zealand, in 1988, and the Ph.D. degree from the Australian National University, Canberra, in 1992.

He is currently a Professor and Canada Research Chair at the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada. Prior to 2002, he was a Chaired Professor with the Department of Electrical and Electronic Engineering, University of Melbourne, Australia, where he also served as Deputy Head of the Department. His current research interests include ion channels and biological nanotubes, networked sensor scheduling and control, statistical signal processing, nonlinear filtering, and wireless telecommunications.

Dr. Krishnamurthy is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS AEROSPACE AND ELECTRONIC SYSTEMS, and *Systems and Control Letters*. He is Guest Editor of a special issue of the IEEE TRANSACTIONS ON NANOBIOSCIENCE (March 2005) on bio-nanotubes. He currently serves on the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society, and the International Federation of Automatic Control (IFAC) Technical Committee on Modeling, Identification, and Signal Processing. He has served on the technical program committee of several conferences in signal processing, telecommunications, and control.

G. George Yin (S'87–M'87–SM'96–F'02) received the B.S. degree in mathematics from the University of Delaware, Newark, in 1983, 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, and became a Professor in 1996. He severed on the Mathematical Review Date Base Committee, 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 Conference on Decision and Control, 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 *SIAM Journal on Control and Optimization*, and is (or was) on the editorial board of four other journals.

Dr. Yin was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 1994 to 1998.