

## WEAK CONVERGENCE OF HYBRID FILTERING PROBLEMS INVOLVING NEARLY COMPLETELY DECOMPOSABLE HIDDEN MARKOV CHAINS\*

G. YIN<sup>†</sup> AND S. DEY<sup>‡</sup>

**Abstract.** Concentrating on a class of hybrid discrete-time filtering problems that are modulated by a Markov chain, this work aims to reduce the complexity of the underlying problems. Since the Markov chain has a large state space, the solution of the problem relies on solving a large number of filtering equations. Exploiting the hierarchical structure of the system, it is noted that the transition probability matrix of the Markov chain can be viewed as a nearly decomposable one. It is shown that a reduced system of filtering equations can be obtained by aggregating the states of each recurrent class into one state. Extensions to inclusion of transient states and nonstationary cases are also treated.

**Key words.** Markov chain, filtering, near complete decomposability, weak convergence

**AMS subject classifications.** 60F05, 60G35, 60J10, 93E11

**PII.** S0363012901388464

**1. Introduction.** In this work, we concern ourselves with hybrid filtering problems in discrete time. Since a wide variety of problems arising in target tracking, speech recognition, telecommunication, and manufacturing requires solutions of filtering problems involving a hidden Markov chain, in addition to the usual random system disturbances and observation noise, we assume that the system under consideration is influenced by a hidden Markov chain with finite state space. Due to the rapid advances in science and technology, various systems tend to be rather complex and large-scale in nature. As a result, although the state space of the Markov chain is finite, it inevitably contains a large number of states. Our main effort is devoted to reducing the complexity of such filtering problems involving large-scale hidden Markov chains.

In a recent paper, linear systems with coefficients driven by a hidden Markov chain were considered [21]. Discrete-time systems were studied in [1, 6, 12, 28] among others. In [33], Zhang studied hybrid filters in continuous time and treated problems involving non-Gaussian noise. Our study is motivated by these recent developments and stems from the needs in many applications mentioned above.

In the seminal paper [26], Simon and Ando pointed out that various large-scale systems have hierarchical structures. Some of the states vary rapidly, and others change slowly. In addition, these states are also naturally decomposable into different layers or a hierarchy. Such a hierarchy allows one to take advantage and to organize and reorganize the systems accordingly. Based on such ideas, Courtois dealt with the so-called *nearly completely decomposable* Markov chain models [7]. Recently, Dey derived reduced-complexity filtering results for hidden Markov models, in which the underlying Markov chains are nearly completely decomposable [9]. Such hierarchical

---

\*Received by the editors April 16, 2001; accepted for publication (in revised form) June 26, 2002; published electronically February 6, 2003.

<http://www.siam.org/journals/sicon/41-6/38846.html>

<sup>†</sup>Department of Mathematics, Wayne State University, Detroit, MI 48202 (gyin@math.wayne.edu). The research of this author was supported in part by the National Science Foundation under grant DMS-9877090.

<sup>‡</sup>Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3010, Australia (sdey@ee.mu.oz.au).

Markov chains have numerous applications in queueing and computer systems [7], multiple time-scale heterogeneous traffic modelling (e.g., variable bit rate video traffic [27]), manufacturing systems, operations research, and many other biological and physical systems in which a multiple time-scale or hierarchical behavior is involved; see also related work in [3, 5, 14, 18, 22, 23, 25] and the references therein. Taking the approaches of [7] and [9] as our point of departure, to reduce the complexity of the underlying problem, we introduce a small parameter  $\varepsilon > 0$  into the system. Note that the small parameter is used to reflect the high contrast of the transition rates of the Markov chain. For the subsequent asymptotic analysis, to obtain the desired results, it is necessary to send  $\varepsilon \rightarrow 0$ , which can serve as a guideline for various applications and for approximation and heuristics. In real applications, however,  $\varepsilon$  might be a fixed constant, and only the relative order of magnitude of this parameter matters. In our setup, we also consider a nearly completely decomposable Markovian model, in which the hidden Markov chain has a large state space. The transition probability matrix is a sum of a completely decomposable transition matrix and a generator of a continuous-time Markov chain. Following our systematic studies on singularly perturbed Markov chains in both continuous time and discrete time [16, 29, 30, 31, 34], we investigate the asymptotic properties of the filtering problem by means of weak convergence methods. We show that a limit filtering problem can be derived in which the underlying Markov chain is replaced by an averaged chain and the system coefficients are averaged out with respect to the stationary measures of each ergodic class. The reduction of complexity is particularly pronounced when the transition matrix of the Markov chain consists of only one ergodic class. In this case, the limit filtering problem becomes a standard Kalman filter free of Markovian jump processes.

The rest of the paper is arranged as follows. Section 2 presents the precise formulation of the problem and a number of preliminary results that are to be used in our study. Section 3 is concerned with weak convergence analysis and the derivation of limit filtering problems or reduced systems. In order not to disrupt the flow of presentation, all proofs are placed in an appendix. Section 4 proceeds with numerical experiments and simulation studies that demonstrate the relationship between the original system and that of a reduced system. Section 5 gives remarks and a few extensions.

Throughout the paper, we use  $K$  to denote a generic positive constant, whose values may be different for different usage. For any  $z \in \mathbb{R}^{\ell_1 \times \ell_2}$  with some positive integers  $\ell_1$  and  $\ell_2$ ,  $z'$  denotes its transpose. For a suitable function  $f$ ,  $f_x$  and  $f_{xx}$  denote its first-order and second-order partial derivatives with respect to  $x$ .

**2. Formulation and preliminaries.** This section gives the precise formulation of the problem to be studied. It also presents some preliminary results needed in the analysis to follow.

**2.1. Hybrid filtering problem.** Let  $\varepsilon > 0$  be a small parameter, and let  $\{\alpha_n^\varepsilon\}$  be a (time) homogeneous singularly perturbed Markov chain in discrete time with a finite state space  $\mathcal{M}$  having  $m$  elements and a transition matrix

$$(2.1) \quad P^\varepsilon = \tilde{P} + \varepsilon Q,$$

where  $\tilde{P}$  is an  $m \times m$  transition matrix and  $Q = (q_{\iota\ell})$  is a generator of a continuous-time homogeneous Markov chain, i.e.,  $q_{\iota\ell} \geq 0$  for  $\iota \neq \ell$  and  $\sum_\ell q_{\iota\ell} = 0$  for each  $\iota$ .

Suppose that, for some  $T > 0$  and  $0 \leq n \leq \lfloor T/\varepsilon \rfloor$  (where  $\lfloor z \rfloor$  denotes the largest integer part of  $z$ ),  $x_n^\varepsilon \in \mathbb{R}^r$  is the state to be estimated,  $y_n^\varepsilon$  is the corresponding observation, and  $A(\iota)$ ,  $C(\iota)$ ,  $\sigma_w(\iota)$ , and  $\sigma_v(\iota)$  are well defined for each  $\iota \in \mathcal{M}$  (i.e., they are finite for each  $\iota \in \mathcal{M}$ ). With initial data  $x_0$  and  $y_0$ , the hybrid filtering problem is concerned with the linear system of equations

$$(2.2) \quad \begin{aligned} x_{n+1}^\varepsilon &= x_n^\varepsilon + \varepsilon A(\alpha_n^\varepsilon)x_n^\varepsilon + \sqrt{\varepsilon}\sigma_w(\alpha_n^\varepsilon)w_n, \\ y_{n+1}^\varepsilon &= y_n^\varepsilon + \varepsilon C(\alpha_n^\varepsilon)x_n^\varepsilon + \sqrt{\varepsilon}\sigma_v(\alpha_n^\varepsilon)v_n, \end{aligned}$$

where  $\{w_n\}$  and  $\{v_n\}$  are the system disturbance and the observation noise, respectively. For ease of presentation, in what follows, we will suppress the floor-function notation  $\lfloor \cdot \rfloor$  and write it as  $0 \leq n \leq T/\varepsilon$  throughout. The use of the  $\sqrt{\varepsilon}$  in the noise terms stems from the central limit scaling. Precise conditions on the noises will be provided later. In what follows, we will show that, as  $\varepsilon \rightarrow 0$ , the above filtering problem has a limit. The limit filtering problem is still modulated by a Markov chain. However, the total number of states of the limit Markov chain is equal to the number of recurrent groups or clusters  $l$ . As mentioned before, typically  $l \ll m$ , and by considering this limit filtering problem, substantial computational savings can be obtained. Although (2.2) is a discrete-time filtering problem, the limit under appropriate scaling is a continuous-time hybrid filtering problem. In the rest of the paper, our main effort is devoted to deriving the limit filtering problem. For solutions of continuous-time hybrid filtering problems involving jump Markov processes, see [4, 8, 11, 12, 21]; see also [2, 10] and the references therein for discrete-time results.

**2.2. Nearly completely decomposable Markov chain  $\alpha_n^\varepsilon$ .** In view of (2.1), the transition probabilities of  $\alpha_n^\varepsilon$  are dominated by  $\tilde{P}$ . The structure of  $\tilde{P}$  is thus important. Since  $\alpha_n^\varepsilon$  is a finite-state Markov chain, the Markov chain corresponding to the transition matrix  $\tilde{P}$  either consists of all recurrent states or includes transient states in addition to recurrent states (see [15]). We first consider the case of inclusion of recurrent states only. Later we will discuss a generalization to the case in which transient states are also included. Suppose that the matrix  $\tilde{P}$  is given by

$$(2.3) \quad \tilde{P} = \text{diag}(\tilde{P}^1, \dots, \tilde{P}^l) = \begin{pmatrix} \tilde{P}^1 & & \\ & \ddots & \\ & & \tilde{P}^l \end{pmatrix},$$

where each  $\tilde{P}^i \in \mathbb{R}^{m_i \times m_i}$  is itself a transition matrix and  $\sum_{i=1}^l m_i = m$ . Here and henceforth, by  $\text{diag}(Z^1, \dots, Z^l)$ , we mean a diagonal block matrix with matrix entries  $Z^1$  through  $Z^l$  of appropriate dimensions. It is clear that, for sufficiently small  $\varepsilon > 0$ ,  $P^\varepsilon$  is close to  $\tilde{P}$ , and so  $P^\varepsilon$  is a nearly completely decomposable transition matrix (see [7]). Note that, typically for large scale Markovian systems,  $l \ll m$ , and therein lies the motivation for reducing computational complexity. Concerning the Markov chain, we assume the following condition.

- (A1) The transition probability matrix of the Markov chain  $\alpha_n^\varepsilon$  is given by (2.1) with  $\tilde{P}$  specified in (2.3), and the state space of the Markov chain is

$$(2.4) \quad \begin{aligned} \mathcal{M} &= \mathcal{M}_1 \cup \mathcal{M}_2 \cap \dots \cup \mathcal{M}_l \\ &= \{s_{11}, \dots, s_{1m_1}\} \cup \dots \cup \{s_{l1}, \dots, s_{lm_l}\}. \end{aligned}$$

For each  $i = 1, \dots, l$ ,  $\mathcal{M}_i = \{s_{i1}, \dots, s_{im_i}\}$  is the state space corresponding to the transition matrix  $\tilde{P}^i$ , and  $\tilde{P}^i$  is irreducible and aperiodic.

Note that the probability vector

$$p_n^\varepsilon = (P(\alpha_n^\varepsilon = s_{11}), \dots, P(\alpha_n^\varepsilon = s_{1m_1}), \dots, P(\alpha_n^\varepsilon = s_{l1}), \dots, P(\alpha_n^\varepsilon = s_{lm_l})) \in \mathbb{R}^{1 \times m}$$

satisfies

$$(2.5) \quad p_{n+1}^\varepsilon = p_n^\varepsilon P^\varepsilon, \quad p_0^\varepsilon = p_0,$$

such that  $p_0$  is the initial probability distribution. By (A1), the result in [30] yields the following lemma.

LEMMA 2.1. *Assume condition (A1). Then the following assertions hold:*

- (1) *Denote by  $\nu^i$  the stationary distribution corresponding to the transition matrix  $\tilde{P}_i$  for each  $i = 1, \dots, l$ . Then, for some  $0 < \lambda < 1$ ,*

$$(2.6) \quad p_n^\varepsilon = \theta(t) \operatorname{diag}(\nu^1, \dots, \nu^l) + O(\varepsilon + \lambda^n),$$

where  $\theta(t) = (\theta_1(t), \dots, \theta_l(t)) \in \mathbb{R}^{1 \times l}$  (with  $t = \varepsilon n$ ) satisfies

$$\frac{d\theta(t)}{dt} = \theta(t)\bar{Q}, \quad \theta_i(0) = x_0^i \mathbf{1}_{m_i},$$

with

$$(2.7) \quad \begin{aligned} \bar{Q} &= \operatorname{diag}(\nu^1, \dots, \nu^l) Q \tilde{\mathbf{1}}, \\ \tilde{\mathbf{1}} &= \operatorname{diag}(\mathbf{1}_{m_1}, \dots, \mathbf{1}_{m_l}), \end{aligned}$$

where  $\mathbf{1}_\ell$  denotes an  $\ell$ -dimensional column vector with all entries being 1.

- (2) *For  $n \leq T/\varepsilon$ , the  $n$ -step transition probability matrix  $(P^\varepsilon)^n$  satisfies*

$$(2.8) \quad (P^\varepsilon)^n = \Phi(t) + O(\varepsilon + \lambda^n),$$

where

$$(2.9) \quad \begin{aligned} P^0 \Phi(t) &= \tilde{\mathbf{1}} \Theta(t) \operatorname{diag}(\nu^1, \dots, \nu^l), \\ \frac{d\Theta(t)}{dt} &= \Theta(t)\bar{Q}, \quad \Theta(0) = I. \end{aligned}$$

Remark 2.2. Since we are primarily concerned with the form of the limit distribution, only the leading terms are presented in the lemma, although a full asymptotic expansion can be obtained. See [30] for more details.

Starting from the Markov chain  $\alpha_n^\varepsilon$ , define an aggregated process  $\bar{\alpha}_n^\varepsilon$  by setting  $\bar{\alpha}_n^\varepsilon = i$  if  $\alpha_n^\varepsilon \in \mathcal{M}_i$ . Define piecewise constant interpolated processes  $\alpha^\varepsilon(\cdot)$  and  $\bar{\alpha}^\varepsilon(\cdot)$  by

$$\alpha^\varepsilon(t) = \alpha_n, \quad \bar{\alpha}^\varepsilon(t) = \bar{\alpha}_n^\varepsilon, \quad t \in [n\varepsilon, (n+1)\varepsilon).$$

Lemma 2.1 is mainly deterministic, whereas the following lemma is a weak convergence result on the aggregated process. Its proof is provided in [32]; a continuous-time counterpart can be found in [29, pp. 170–171].

LEMMA 2.3. Under (A1), as  $\varepsilon \rightarrow 0$ ,  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$ , which is a continuous-time Markov chain with state space  $\bar{\mathcal{M}} = \{1, \dots, l\}$  and generator  $\bar{Q}$  given by (2.7). Moreover, for the occupation measures defined by

$$o_{n,ij}^\varepsilon = \varepsilon \sum_{k=0}^n [I_{\{\alpha_k^\varepsilon = s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \text{ for each } i = 1, \dots, l, j = 1, \dots, m_i,$$

the following mean square estimates hold:

$$(2.10) \quad \sup_{0 \leq n \leq T/\varepsilon} E|o_{n,ij}^\varepsilon|^2 = O(\varepsilon).$$

To proceed, we give additional conditions needed for the filtering problem.

- (A2)  $E|x_0|^2 < \infty$  and  $E|y_0|^2 < \infty$ . For each  $\iota \in \mathcal{M}$ ,  $A(\iota)$ ,  $C(\iota)$ ,  $\sigma_w(\iota)$ , and  $\sigma_v(\iota)$  are finite;  $\sigma_w(\iota)\sigma_w'(\iota)$  and  $\sigma_v(\iota)\sigma_v'(\iota)$  are positive definite matrices.
- (A3) The sequences  $\{w_n\}$  and  $\{v_n\}$  are independent of  $\{\alpha_n^\varepsilon\}$  and independent of each other. The  $\{w_n\}$  and  $\{v_n\}$  are stationary martingale difference sequences (with zero mean) such that

$$Ew_n w_n' = I, \quad E v_n v_n' = I, \\ E|w_n|^{2+\Delta} < \infty, \quad \text{and} \quad E|v_n|^{2+\Delta} < \infty \text{ for some } \Delta > 0.$$

Remark 2.4. For simplicity and ease of presentation, we assume that the noises are stationary martingale difference sequences and that the covariance of  $w_n$  and  $v_n$  is the identity matrix. Even though no Gaussian assumption is used, as a result of the scaling, these noise processes will be asymptotically normal thanks to the functional central limit theorem.

In what follows, we use the weak convergence method to establish the desired results. Further details on the weak convergence method, which is an extension of convergence in distribution, can be found in, for example, [13, Chapter 3] or [19, Chapters 7 and 8].

**3. Limit filtering problem.** This section is devoted to the derivation of the limit filtering problem. In lieu of treating the discrete-time iterates, our analysis focuses on suitable continuous-time interpolations of piecewise constant processes.

For  $0 \leq n \leq T/\varepsilon$ , define the interpolations  $x^\varepsilon(\cdot)$  and  $y^\varepsilon(\cdot)$  as

$$(3.1) \quad x^\varepsilon(t) = x_n^\varepsilon, \quad y^\varepsilon(t) = y_n^\varepsilon, \quad t \in [n\varepsilon, n\varepsilon + \varepsilon),$$

where  $x_n^\varepsilon$  and  $y_n^\varepsilon$  are given in (2.2). Then  $x^\varepsilon(\cdot)$  and  $y^\varepsilon(\cdot) \in D^r[0, T]$ , which is the space of  $\mathbb{R}^r$ -valued functions that are right continuous and have left limits, endowed with the Skorohod topology [13, p. 122]. Using weak convergence methods, we will show that the interpolated processes converge weakly to  $x(\cdot)$  and  $y(\cdot)$ , which satisfy continuous-time hybrid Kalman filtering equations. Following the approach of weak convergence methods [13, 17], we first show that the sequences of interests are tight, and then we characterize the limit processes by using martingale averaging techniques.

Owing to the assumption on the system and observation noise and  $\sqrt{\varepsilon}$  scaling, the following lemma, known as the functional central limit theorem or Donsker's invariance theorem, holds. Its proof is standard; see, for example, [13, Theorem 3.1, p. 351].

LEMMA 3.1. *Define*

$$(3.2) \quad w^\varepsilon(t) = \sqrt{\varepsilon} \sum_{j=0}^{t/\varepsilon-1} w_j \quad \text{and} \quad v^\varepsilon(t) = \sqrt{\varepsilon} \sum_{j=0}^{t/\varepsilon-1} v_j.$$

Under (A3),  $w^\varepsilon(\cdot)$  and  $v^\varepsilon(\cdot)$  converge weakly to standard  $r$ -dimensional Brownian motions  $w(\cdot)$  and  $v(\cdot)$ , respectively.

In fact, correlated  $\varphi$ -mixing noises may be dealt with, and the corresponding central limit result can be obtained, but the notation will be much more complex for the subsequent averaging. Thus we decide to work with the martingale difference sequences  $\{w_n\}$  and  $\{v_n\}$ . In the analysis to follow, we need the a priori bounds on  $\{x_n^\varepsilon\}$  and  $\{y_n^\varepsilon\}$ , which are presented in the form of the following lemma. The proof is provided in the appendix.

LEMMA 3.2. *Assume (A1)–(A3). For  $\{x_n^\varepsilon\}$  and  $\{y_n^\varepsilon\}$  defined in (2.2), the following bounds hold:*

$$(3.3) \quad \sup_{0 \leq n \leq T/\varepsilon} E|x_n^\varepsilon|^2 < \infty \quad \text{and} \quad \sup_{0 \leq n \leq T/\varepsilon} E|y_n^\varepsilon|^2 < \infty.$$

**3.1. Tightness and weak convergence.** To proceed, let  $\mathcal{F}_n$  be the  $\sigma$ -algebra generated by  $\{\alpha_j^\varepsilon, w_j, v_j : j \leq n\}$ , and let  $E_n$  be the conditional expectation with respect to  $\mathcal{F}_n$ ; let  $\mathcal{F}_t^\varepsilon$  be the  $\sigma$ -algebra generated by  $\{\alpha^\varepsilon(s), w^\varepsilon(s), v^\varepsilon(s) : s \leq t\}$ , and let  $E_t^\varepsilon$  be the conditional expectation with respect to  $\mathcal{F}_t^\varepsilon$ . We are to derive the tightness of  $\{x^\varepsilon(\cdot)\}$  and  $\{y^\varepsilon(\cdot)\}$ . This is a compactness result, which is established by verifying a tightness criterion; the proof is in the appendix.

THEOREM 3.3. *Assume (A1)–(A3). Then  $\{x^\varepsilon(\cdot)\}$  is tight in  $D^r[0, T]$ , and so is  $\{y^\varepsilon(\cdot)\}$ , where  $D^r[0, T]$  is the space of  $\mathbb{R}^r$ -valued functions that are right continuous and have left limits, endowed with the Skorohod topology.*

We are now in a position to obtain the weak convergence of the sequences  $\{x^\varepsilon(\cdot)\}$  and  $\{y^\varepsilon(\cdot)\}$ . To prove the assertion, we use a martingale problem formulation. Thus our task becomes to figure out the limit by characterizing the operator of the limit martingale problem. The technique used is essentially an averaging approach. Different from the diffusion approximation in wideband noise systems [17], the limit  $\bar{\alpha}^\varepsilon(\cdot)$  also contributes to the limit process and adds further complication. The result is recorded in the following theorem, whose proof is in the appendix as well.

THEOREM 3.4. *Suppose the conditions of Theorem 3.3 hold. Then  $x^\varepsilon(\cdot)$  and  $y^\varepsilon(\cdot)$  converge weakly to  $x(\cdot)$  and  $y(\cdot)$ , respectively, such that  $x(\cdot)$  and  $y(\cdot)$  are solutions of the filtering equations*

$$(3.4) \quad \begin{aligned} dx &= \bar{A}(\bar{\alpha}(t))xdt + \bar{\sigma}_w(\bar{\alpha}(t))dw, \\ dy &= \bar{C}(\bar{\alpha}(t))xdt + \bar{\sigma}_v(\bar{\alpha}(t))dv, \end{aligned}$$

where  $w(\cdot)$  and  $v(\cdot)$  are the independent  $r$ -dimensional standard Brownian motions given by Lemma 3.1,

$$(3.5) \quad \bar{A}(i) = \sum_{j=1}^{m_i} \nu_j^i A(s_{ij}), \quad \bar{B}(i) = \sum_{j=1}^{m_i} \nu_j^i B(s_{ij}) \quad \text{for each } i \in \bar{\mathcal{M}},$$

and, for each  $i \in \overline{\mathcal{M}}$ ,  $\overline{\sigma}_w(i)$  and  $\overline{\sigma}_v(i)$  satisfy

$$(3.6) \quad \begin{aligned} \overline{\sigma}_w(i)\overline{\sigma}'_w(i) &= \sum_{j=1}^{m_i} \nu_j^i \sigma_w(s_{ij})\sigma'_w(s_{ij}), \\ \overline{\sigma}_v(i)\overline{\sigma}'_v(i) &= \sum_{j=1}^{m_i} \nu_j^i \sigma_v(s_{ij})\sigma'_v(s_{ij}). \end{aligned}$$

**3.2. Markov chains with one ergodic class.** The reduction of complexity is particularly pronounced if the transition matrix (2.3) consists of only one ergodic class (i.e.,  $P$  in (2.3) consists of only one block). That is,  $P^\varepsilon = P + \varepsilon Q$  such that  $P$  is irreducible and aperiodic. It is easily seen that, for sufficiently small  $\varepsilon > 0$ ,  $P^\varepsilon$  is also irreducible. Consider the filtering problem (2.2). Similarly to the previous case, define  $x^\varepsilon(\cdot)$  and  $y^\varepsilon(\cdot)$  as the piecewise constant interpolations of  $x_k^\varepsilon$  and  $y_k^\varepsilon$ , respectively. Replace  $\overline{A}(\cdot)$  and  $\overline{\sigma}_w(\cdot)$  by

$$(3.7) \quad \overline{A}^0 = \sum_{j=1}^m A(j)\nu_j \quad \text{and} \quad \overline{\sigma}_w^0(\overline{\sigma}_w^0)' = \sum_{j=1}^m \nu_j \sigma_w(j)\sigma'_w(j),$$

with  $\nu = (\nu_1, \dots, \nu_m)$  denoting the stationary distribution of  $P$ . Similarly replace  $\overline{C}(\cdot)$  and  $\overline{\sigma}_v(\cdot)$  by  $\overline{C}^0$  and  $\overline{\sigma}_v^0$ , respectively. The weak convergence of  $(x^\varepsilon(\cdot), y^\varepsilon(\cdot))$  will still be obtained. The proofs are similar to the previous case. In fact, it is readily seen that Lemma 3.2 and Theorem 3.3 continue to hold. Lemma 2.1 still holds with obvious modifications, and (2.10) (in Lemma 2.3) is changed to

$$\sup_{0 \leq n \leq T/\varepsilon} E \left[ \varepsilon \sum_{k=0}^n [I_{\{\alpha_k^\varepsilon=j\}} - \nu_j] \right]^2 = O(\varepsilon).$$

Using this mean square estimate and similar arguments as before, we can show that Theorem 3.4 continues to hold. It is interesting to note that the limit filtering problem becomes a standard Kalman filter, in which the jump process effect has been completely averaged out. We state this as the following result.

**COROLLARY 3.5.** *Consider the filtering problem (2.2) such that  $P$  is irreducible and aperiodic. Then  $(x^\varepsilon(\cdot), y^\varepsilon(\cdot))$  converges weakly to  $(x(\cdot), y(\cdot))$ , that is, the solution of the filtering problem*

$$(3.8) \quad \begin{aligned} dx(t) &= \overline{A}^0 x(t)dt + \overline{\sigma}_w^0 dw(t), \\ dy(t) &= \overline{C}^0 x(t)dt + \overline{\sigma}_v^0 dv(t). \end{aligned}$$

**4. Simulation studies.** In this section, we demonstrate the relationship between the full-order discrete-time system (2.2) and the reduced-order limit filtering equations (3.4) through simulation examples. All results are averaged over 50 trials.

For (2.2), we simulate a discrete-time Markov chain with four states (with two blocks,  $m_1 = m_2 = 2$ ) for which the transition probability matrices are

$$\tilde{P}^1 = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}, \quad \tilde{P}^2 = \begin{pmatrix} 0.25 & 0.75 \\ 0.68 & 0.32 \end{pmatrix}$$

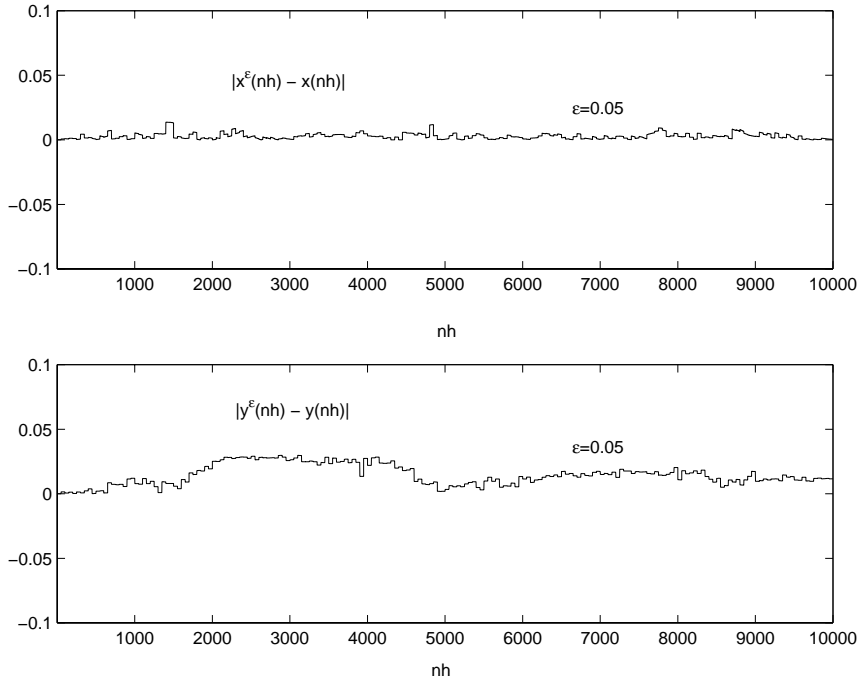


FIG. 4.1. Absolute error between the piecewise constant interpolated full-order system and the reduced-order limit filtered system,  $\varepsilon = 0.05$ .

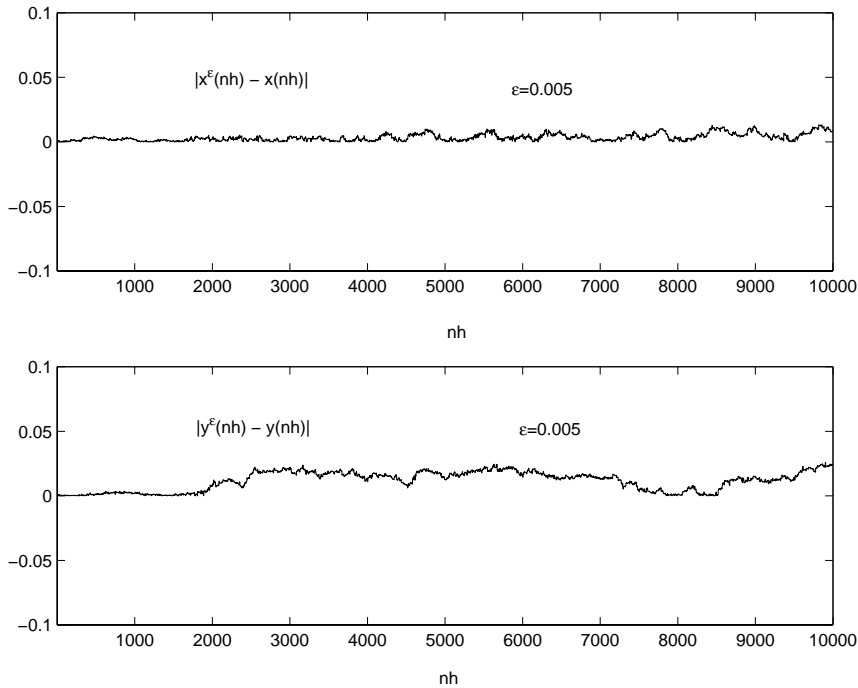


FIG. 4.2. Absolute error between the piecewise constant interpolated full-order system and the reduced-order limit filtered system,  $\varepsilon = 0.005$ .



and the generator is

$$Q = \begin{pmatrix} -0.6 & 0.4 & 0.1 & 0.1 \\ 0.05 & -0.4 & 0.05 & 0.3 \\ 0.1 & 0.2 & -0.7 & 0.4 \\ 0.15 & 0.05 & 0.1 & -0.3 \end{pmatrix}.$$

We take  $(A(1) A(2) A(3) A(4)) = (-4.0 -1.0 -2.0 -3.0)$ ,  $(C(1) C(2) C(3) C(4)) = (0.2 0.5 0.1 1.0)$ . Also,  $(\sigma_w(1) \sigma_w(2) \sigma_w(3) \sigma_w(4)) = (0.2 0.5 0.1 1.0)$ ,  $\sigma_v(i) = \sigma_w(i)$ , for  $i = 1, 2, 3, 4$ . The noise sequences  $\{w_n\}$  and  $\{v_n\}$  are simulated as Gaussian random variables with zero mean and unity variance. The piecewise constant interpolated processes  $x^\varepsilon(t)$  and  $y^\varepsilon(t)$  are constructed from (3.1). The time horizon is taken to be  $T = 10$ . To simulate (3.4), a continuous-time Markov chain is used with a generator  $\bar{Q}$ . This is then discretized with a discretization interval  $h = 0.001$ . Figures 4.1 and 4.2 show the difference  $|x^\varepsilon(\cdot) - x(\cdot)|$  and  $|y^\varepsilon(\cdot) - y(\cdot)|$  for  $\varepsilon = 0.05$  and  $\varepsilon = 0.005$ , respectively.

**5. Remarks and extensions.** This section is devoted to several remarks regarding the approximation issue. They include reduction of complexity as well as ramifications of the results we have obtained thus far.

**Reduction of complexity.** One of the main motivations of the current study is the effort of reduction of complexity. Regarding (2.2), note that the time horizon we are working with is  $0 \leq n \leq \lfloor T/\varepsilon \rfloor$ . As pointed out in [24], if we treat the discrete-time case directly, it can be reduced to an  $m^{\lfloor T/\varepsilon \rfloor}$ -dimensional recursive system of equations, where  $m$  is the total number of states of the Markov chain. For us,  $m$  is a fairly large number. As a result, the amount of computation becomes practically untrackable. One cannot complete the computation in polynomial time. By weak convergence methods, we have obtained a reduced or limit system of filtering equations. This limit system of equations allows us to find nearly optimal filtering, and the limit system has reduced complexity. In particular, if the transition matrix  $P$  given in (2.1) is irreducible, the limit becomes a Kalman filter (see Proposition 3.5).

For continuous-time Kalman filter problems with Markovian switching, it has been recognized (see [4, 11, 21]) that, in general, the problem is an infinite-dimensional one just as in the nonlinear filter case [20]. Nevertheless, Björk [4] proved that a finite-dimensional filter exists for a linear hybrid system if and only if the observation is independent of the state variable. For the filtering problem considered in this paper, this requires the observation process in the limit problem being independent of state. Corresponding to such a requirement, we can consider

$$(5.1) \quad \begin{aligned} x_{n+1}^\varepsilon &= x_n^\varepsilon + \varepsilon A(\alpha_n^\varepsilon)x_n^\varepsilon + \sqrt{\varepsilon}\sigma_w(\alpha_n^\varepsilon)w_n, \\ y_{n+1}^\varepsilon &= y_n^\varepsilon + \varepsilon C(\alpha_n^\varepsilon) + \sqrt{\varepsilon}\sigma_v v_n. \end{aligned}$$

Similar to the derivation of Theorem 3.4, we obtain the limit filtering equations

$$(5.2) \quad \begin{aligned} dx &= \bar{A}(\bar{\alpha}(t))xdt + \bar{\sigma}_w(\bar{\alpha}(t))dw, \\ dy &= \bar{C}(\bar{\alpha}(t))dt + \sigma_v dv. \end{aligned}$$

Note that the calculation of (5.1) leads to recursive filters of dimension  $m^{\lfloor T/\varepsilon \rfloor}$ , whereas (5.2) yields a finite-dimensional filtering problem.

**Inclusion of transient states.** In the previous sections, the main ingredient is the aggregation of states in each recurrent classes. The results obtained can be extended to the case in which the Markov chain has finite state space with inclusion of transient states. To be more specific, Let the transition probability be of the form (2.1). However, in lieu of (2.3), suppose the transition matrix  $\tilde{P}$  in (2.1) is given by

$$(5.3) \quad \tilde{P} = \begin{pmatrix} \tilde{P}^1 & & & \\ & \ddots & & \\ & & \tilde{P}^l & \\ \tilde{P}_*^1 & \dots & \tilde{P}_*^l & \tilde{P}_* \end{pmatrix}.$$

In lieu of (A1), assume (A1').

(A1')  $\alpha_n^\varepsilon$  is a Markov chain with a transition probability matrix given by (2.1) and (5.3), and with state space

$$(5.4) \quad \mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cap \dots \cup \mathcal{M}_l \cup \mathcal{M}_* \\ = \{s_{11}, \dots, s_{1m_1}\} \cup \dots \cup \{s_{l1}, \dots, s_{lm_l}\} \cup \{s_{*1}, \dots, s_{*m_*}\},$$

where, for each  $i = 1, \dots, l$ ,  $\mathcal{M}_i = \{s_{i1}, \dots, s_{im_i}\}$  is the state space corresponding to the transition matrix  $\tilde{P}^i$  and where the subspace  $\mathcal{M}_* = \{s_{*1}, \dots, s_{*m_*}\}$  collects the transient states. Moreover,  $\tilde{P}^i$  is irreducible for each  $i = 1, \dots, l$ , and all eigenvalues of  $\tilde{P}_*$  are inside the unit disk.

To obtain the desired asymptotics, we still use aggregations. However, we aggregate only the states in each recurrent class. Partition the matrix  $Q$  as

$$(5.5) \quad Q = \begin{pmatrix} Q^{11} & Q^{12} \\ Q^{21} & Q^{22} \end{pmatrix},$$

where

$$Q^{11} \in \mathbb{R}^{(m-m_*) \times (m-m_*)}, \quad Q^{12} \in \mathbb{R}^{(m-m_*) \times m_*}, \\ Q^{21} \in \mathbb{R}^{m_* \times (m-m_*)}, \quad \text{and} \quad Q^{22} \in \mathbb{R}^{m_* \times m_*}.$$

Set

$$(5.6) \quad \bar{Q}_* = \text{diag}(\nu^1, \dots, \nu^l)(Q^{11}\tilde{\mathbf{1}} + Q^{12}A_*),$$

with

$$(5.7) \quad A_* = (a_1, \dots, a_l) \in \mathbb{R}^{m_* \times l} \quad \text{and} \\ a_i = -(\tilde{P}_* - I)^{-1}\tilde{P}_*^i \mathbf{1}_{m_i} \quad \text{for } i = 1, \dots, l.$$

Let  $U$  be a random variable uniformly distributed over  $[0, 1]$ . For each  $j = 1, \dots, m_*$ , define an integer-valued random variable  $\xi_j$  by

$$\xi_j = I_{\{0 \leq U \leq a_{m_1, j}\}} + 2I_{\{a_{1, j} < U \leq a_{1, j} + a_{2, j}\}} + \dots + lI_{\{a_{1, j} + \dots + a_{l-1, j} < U \leq 1\}}.$$

Define the aggregated process and its interpolation by

$$(5.8) \quad \bar{\alpha}_n^\varepsilon = \begin{cases} i & \text{if } \alpha_n \in \mathcal{M}_i, \\ U_j & \text{if } \alpha_n^\varepsilon = s_{*j}, \end{cases} \\ \bar{\alpha}^\varepsilon(t) = \bar{\alpha}_n^\varepsilon \quad \text{for } t \in [n\varepsilon, (n+1)\varepsilon).$$

Then we can show that  $\bar{\alpha}^\varepsilon(\cdot)$  converges weakly to  $\bar{\alpha}(\cdot)$  and that the limit is still a Markov chain with state space  $\bar{\mathcal{M}}$ . Furthermore, we can obtain similar limit results for the filtering problems. The notation is more involved, but the main idea and the averaging techniques are as in the previous case. Loosely, the transient states are asymptotically negligible. In the limit (reduced) system, only the states in the recurrent states are important. The limit is still an average with respect to the stationary measures of each recurrent class.

**THEOREM 5.1.** *Assume (A1'), (A2), and (A3). Then the conclusions of Theorems 3.3 and 3.4 continue to hold with  $\bar{Q}$  replaced by  $\bar{Q}_*$  defined in (5.6).*

**Nonstationary Markov chains.** Generally, nonstationary or time-inhomogeneous cases are much more difficult to deal with. However, for a class of problems, it can be worked out. The main setup is similar to that of [30]. In lieu of (2.1), assume that the transition probability matrix is nonstationary and given by

$$P^\varepsilon(\varepsilon n) = \tilde{P}(\varepsilon n) + \varepsilon Q(\varepsilon n),$$

where  $\tilde{P}(\varepsilon n)$  is the dominating part of the transition matrix. In this case, we can carry out the analysis as in the previous case, although the details and notation are more involved.

**Continuous-time problems.** So far, we have considered discrete-time filtering problems exclusively. There is also a continuous-time analogue of the hybrid filtering problems. In place of (2.2), for  $t \in [0, T]$ , consider

$$(5.9) \quad \begin{aligned} dx^\varepsilon(t) &= A(\alpha^\varepsilon(t))x^\varepsilon(t)dt + \sigma_w(\alpha^\varepsilon(t))dw, \\ dy^\varepsilon(t) &= C(\alpha^\varepsilon(t))x^\varepsilon(t)dt + \sigma_v(\alpha^\varepsilon(t))dv, \end{aligned}$$

where  $w(\cdot)$  and  $v(\cdot)$  are independent standard Brownian motions, and where  $\alpha^\varepsilon(\cdot)$  is a continuous-time singularly perturbed Markov chain with finite state space  $\mathcal{M}$  and with generator

$$(5.10) \quad Q^\varepsilon(t) = \frac{\tilde{Q}(t)}{\varepsilon} + \hat{Q}(t),$$

where both  $\tilde{Q}(t)$  and  $\hat{Q}(t)$  are generators. The state space  $\mathcal{M}$  can be of the form of either (2.4) (with recurrent states only) or (5.4) (inclusion of transient states). We can follow our approach of averaging and aggregation to reduce the complexity of the underlying system and obtain a limit system with much reduced state space. Various results on the asymptotic properties of  $\alpha^\varepsilon(\cdot)$  can be found in [29, 31] among others. The proof of the following result is similar to the discrete-time case; we omit the details. Note that, since the problem is in continuous time, no interpolations are needed, however. For definiteness, we state the result for decomposition of the form (5.4). The matrix  $\tilde{Q}(t)$  has the form

$$(5.11) \quad \tilde{Q}(t) = \begin{pmatrix} \tilde{Q}^1(t) & & & \\ & \ddots & & \\ & & \tilde{Q}^l(t) & \\ \tilde{Q}_*^1(t) & \dots & \tilde{Q}_*^l(t) & \tilde{Q}_*(t) \end{pmatrix}.$$

For each  $i \in \{1, \dots, l\}$ , let  $\tilde{Q}_*^i(t) = B(t)\tilde{Q}_{*,c}^i$ ,  $\tilde{Q}_*(t) = B(t)\tilde{Q}_{*,c}$ , where  $B(t)$  is an  $\mathbb{R}^{m_* \times m_*}$  matrix-valued function, and  $\tilde{Q}_{*,c}^i \in \mathbb{R}^{m_* \times m_i}$  and  $\tilde{Q}_{*,c} \in \mathbb{R}^{m_* \times m_*}$  are

constant matrices. It is readily seen that  $B(t)$  is invertible for each  $t \in [0, T]$ , and, for each  $i$ ,

$$(5.12) \quad a_i(t) \stackrel{\text{def}}{=} -\tilde{Q}_*^{-1}(t)\tilde{Q}_*^i(t)\mathbf{1}_{m_i} = -\tilde{Q}_{*,c}^{-1}\tilde{Q}_{*,c}^i\mathbf{1}_{m_i} = a_i$$

is a time-independent vector. Define  $\bar{Q}_*(t)$  and  $\bar{\alpha}^\varepsilon(t)$  as in (5.6) and (5.8) with  $Q^{\ell}$  (the partition of  $Q$ ) replaced by  $\hat{Q}^{\ell}(t)$  (the partition of  $\hat{Q}(t)$ ) and with  $a_i$  in (5.7) replaced by (5.12). Then we have the following theorem and its corollary.

**THEOREM 5.2.** *Suppose that, for each  $i \in \{1, \dots, l\}$ ,  $\tilde{Q}^i(t)$  is weakly irreducible (see [29, pp. 21–22] for a definition of weak irreducibility and quasi-stationary distribution), that  $\tilde{Q}_*(t)$  has all of its eigenvalues on the left half of the complex plan, that  $\tilde{Q}(\cdot)$  and  $\hat{Q}(\cdot)$  are bounded and Borel measurable, and that  $\tilde{Q}(\cdot)$  is Lipschitz continuous on  $[0, T]$ . Then  $(x^\varepsilon(\cdot), y^\varepsilon(\cdot))$  converges weakly to  $(x(\cdot), y(\cdot))$  such that  $(x(\cdot), y(\cdot))$  is a solution of the averaged filtering equations (3.4), where  $\bar{A}(\cdot)$ ,  $\bar{\sigma}_w(\cdot)$ ,  $\bar{C}(\cdot)$ , and  $\bar{\sigma}_v(\cdot)$  are defined as before with time-dependent quasi-stationary distributions  $\nu^i(t)$  used.*

**COROLLARY 5.3.** *Suppose that  $Q^\varepsilon(t)$  is given by (5.10) such that  $\tilde{Q}(t)$  is weakly irreducible. Suppose that all other conditions in Theorem 5.2 are satisfied. Then  $(x^\varepsilon(\cdot), y^\varepsilon(\cdot))$  converges weakly to  $(x(\cdot), y(\cdot))$ , satisfying*

$$(5.13) \quad \begin{aligned} dx(t) &= \bar{A}^0(t)x(t)dt + \bar{\sigma}_w^0(t)dw(t), \\ dy(t) &= \bar{C}^0(t)x(t)dt + \bar{\sigma}_v^0(t)dv(t), \end{aligned}$$

where  $\bar{A}^0$ ,  $\bar{C}^0$ ,  $\bar{\sigma}_w^0$ , and  $\bar{\sigma}_v^0$  are defined as in (3.7) with the time-dependent quasi-stationary distribution  $\nu(t) = (\nu_1(t), \dots, \nu_m(t))$  used.

**Appendix. Proofs of results.**

*Proof of Lemma 3.2.* We first work with  $x_n^\varepsilon$ . Iterating on the first equation in (2.2), for  $0 \leq n \leq T/\varepsilon$ ,

$$x_{n+1}^\varepsilon = x_0^\varepsilon + \varepsilon \sum_{j=0}^n A(\alpha_j^\varepsilon)x_j^\varepsilon + \sqrt{\varepsilon} \sum_{j=0}^n \sigma_w(\alpha_j^\varepsilon)w_j.$$

Note that, for any  $z \in \mathbb{R}^r$ ,  $|z|^2 = \text{tr}(zz')$ , where  $\text{tr}(zz')$  denotes the trace of  $zz'$ . Consequently (recall that  $K$  is a generic positive constant),

$$(A.1) \quad \begin{aligned} E|x_{n+1}^\varepsilon|^2 &\leq K \left( E|x_0^\varepsilon|^2 + \varepsilon^2 E \left| \sum_{j=0}^n A(\alpha_j^\varepsilon)x_j^\varepsilon \right|^2 + \varepsilon E \left| \sum_{j=0}^n \sigma_w(\alpha_j^\varepsilon)w_j \right|^2 \right) \\ &\leq KE|x_0^\varepsilon|^2 + K\varepsilon \sum_{j=0}^n E|x_j^\varepsilon|^2 + \varepsilon K \sum_{j=0}^n \sum_{k=0}^n E\text{tr}(\sigma_w(\alpha_j^\varepsilon)w_j w_k' \sigma_w'(\alpha_k^\varepsilon)). \end{aligned}$$

Using the independence of  $\{\alpha_n^\varepsilon\}$  and  $\{w_n\}$  and the boundedness of  $\sigma_w(\ell)$  for each

$\iota \in \mathcal{M}$  and noting that  $Ew_jw'_k = 0$  if  $j \neq k$ ,

$$\begin{aligned}
 \text{(A.2)} \quad & \varepsilon \sum_{j=0}^n \sum_{k=0}^n \text{tr} (E\sigma_w(\alpha_j^\varepsilon)w_jw'_k\sigma'_w(\alpha_k^\varepsilon)) \\
 & \leq \varepsilon K \left| \sum_{j=0}^n \sum_{k=0}^n E\{\sigma_w(\alpha_j^\varepsilon)[Ew_jw'_k]\sigma'_w(\alpha_k^\varepsilon)\} \right| \\
 & \leq \varepsilon K \sum_{k=0}^n |Ew_kw'_k| \\
 & \leq \varepsilon K \frac{T}{\varepsilon} \leq K < \infty.
 \end{aligned}$$

Combining this with (A.1), an application of Gronwall’s inequality yields

$$E|x_{n+1}^\varepsilon|^2 \leq K + K\varepsilon \sum_{j=0}^n E|x_j^\varepsilon|^2 \leq K \exp(K\varepsilon n) \leq K < \infty.$$

Moreover, the bound holds uniformly in  $n$  for  $0 \leq n \leq T/\varepsilon$ .

As for  $y_n^\varepsilon$ , using the bound of  $\sup_{0 \leq n \leq T/\varepsilon} E|x_n^\varepsilon|^2$ , we have

$$\begin{aligned}
 \text{(A.3)} \quad & E|y_{n+1}^\varepsilon|^2 \leq K \left( E|y_0^\varepsilon|^2 + \varepsilon^2 E \left| \sum_{j=0}^n C(\alpha_j^\varepsilon)x_j^\varepsilon \right|^2 + \varepsilon E \left| \sum_{j=0}^n \sigma_v(\alpha_j^\varepsilon)v_j \right|^2 \right) \\
 & \leq KE|y_0^\varepsilon|^2 + K\varepsilon \sum_{j=0}^n E|x_j^\varepsilon|^2 + \varepsilon K \sum_{j=0}^n \sum_{k=0}^n E\text{tr} (\sigma_v(\alpha_j^\varepsilon)v_jv'_k\sigma'_v(\alpha_k^\varepsilon)) \\
 & \leq K < \infty.
 \end{aligned}$$

Moreover, the bound holds uniformly in  $0 \leq n \leq T/\varepsilon$ . □

*Proof of Theorem 3.3.* Let us first deal with the sequence  $\{x^\varepsilon(\cdot)\}$ . For any  $\delta > 0$ ,  $t > 0$ , and  $s > 0$  with  $s \leq \delta$ , consider

$$\begin{aligned}
 \text{(A.4)} \quad & E_t^\varepsilon |x^\varepsilon(t+s) - x^\varepsilon(t)|^2 = E_t^\varepsilon \left| \varepsilon \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} A(\alpha_j^\varepsilon)x_j^\varepsilon + \sqrt{\varepsilon} \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sigma_w(\alpha_j^\varepsilon)w_j \right|^2 \\
 & = \varepsilon^2 \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E_t^\varepsilon \text{tr}[A(\alpha_j^\varepsilon)x_j^\varepsilon x_k^{\varepsilon'} A'(\alpha_k^\varepsilon)] \\
 & \quad + 2\sqrt{\varepsilon^3} \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E_t^\varepsilon \text{tr}[A(\alpha_j^\varepsilon)x_j^\varepsilon w'_k \sigma'_w(\alpha_k^\varepsilon)] \\
 & \quad + \varepsilon \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E_t^\varepsilon \text{tr}[\sigma_w(\alpha_j^\varepsilon)w_j w'_k \sigma'_w(\alpha_k^\varepsilon)]
 \end{aligned}$$

$$\text{(A.5)} \quad \stackrel{\text{def}}{=} I_1^\varepsilon(t, s) + I_2^\varepsilon(t, s) + I_3^\varepsilon(t, s),$$

where  $I_\ell^\varepsilon(t, s)$  for  $\ell = 1, 2, 3$  are defined in an obvious manner.

Consider each of the terms on the right-hand side of (A.4) separately as follows. First, by the finiteness of  $A(\iota)$  for each  $\iota \in \mathcal{M}$ ,

$$\begin{aligned} I_1^\varepsilon(t, s) &= \varepsilon^2 \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \text{tr} \left( E_t^\varepsilon [A(\alpha_j^\varepsilon) x_j^\varepsilon x_k^{\varepsilon'} A'(\alpha_k^\varepsilon)] \right) \\ &\leq K \varepsilon^2 \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E_t^\varepsilon |x_j^\varepsilon| |x_k^\varepsilon|. \end{aligned}$$

By virtue of Lemma 3.2, an application of the Cauchy–Schwarz inequality then yields

$$\begin{aligned} EI_1^\varepsilon(t, s) &\leq K \varepsilon^2 \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E^{1/2} |x_j^\varepsilon|^2 E^{1/2} |x_k^\varepsilon|^2 \\ &\leq K \varepsilon^2 \left( \frac{t+s}{\varepsilon} - \frac{t}{\varepsilon} \right)^2 \leq K s^2 = O(\delta^2). \end{aligned}$$

Thus

$$(A.6) \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} EI_1^\varepsilon(t, s) = \lim_{\delta \rightarrow 0} O(\delta^2) = 0.$$

As for the second term on the right-hand side of (A.4), note that  $x_j^\varepsilon$  and  $A(\alpha_j^\varepsilon)$  are  $\mathcal{F}_j$ -measurable. Since, for  $j < k$ ,  $E_j w_k = 0$ , the independence of  $\{\alpha_n^\varepsilon\}$  and  $\{w_n\}$  in (A3) and the finiteness of  $A(\iota)$  and  $\sigma_w(\iota)$  for each  $\iota \in \mathcal{M}$  lead to

$$\begin{aligned} I_2^\varepsilon(t, s) &= 2\sqrt{\varepsilon^3} \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \text{tr} E_t^\varepsilon [A(\alpha_j^\varepsilon) x_j^\varepsilon w_k' \sigma_w'(\alpha_k^\varepsilon)] \\ &\leq K \sqrt{\varepsilon^3} \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k \geq j} |\text{tr} [E_t^\varepsilon A(\alpha_j^\varepsilon) x_j^\varepsilon (E_j w_k') (E_j \sigma_w'(\alpha_k^\varepsilon))]| \\ &\leq K \sqrt{\varepsilon^3} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \sqrt{E_t^\varepsilon |x_k^\varepsilon|^2} \sqrt{E_t^\varepsilon |w_k|^2}. \end{aligned}$$

Therefore, an application of the Cauchy–Schwarz inequality yields

$$(A.7) \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} EI_2^\varepsilon(t, s) = \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} O(\sqrt{\varepsilon}) = 0.$$

Next, we consider the last term of (A.4). Using the martingale difference property, the independence of  $\{\alpha_n^\varepsilon\}$  and  $\{w_n\}$ , and  $E_j w_k = 0$  for  $j < k$  and  $E_k w_j = 0$  for  $k < j$ , we obtain

$$\begin{aligned} I_3^\varepsilon(t, s) &= \varepsilon \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \text{tr} \left( E_t^\varepsilon \sigma_w(\alpha_j^\varepsilon) w_j w_k' \sigma_w'(\alpha_k^\varepsilon) \right) \\ &= \varepsilon \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} |\text{tr} [E_t^\varepsilon \sigma_w(\alpha_k^\varepsilon) w_k w_k' \sigma_w'(\alpha_k^\varepsilon)]|, \end{aligned}$$

and so

$$EI_3^\varepsilon(t, s) \leq K\varepsilon \left( \frac{t+s}{\varepsilon} - \frac{t}{\varepsilon} \right) = O(\delta).$$

As a result,

$$(A.8) \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} EI_3^\varepsilon(t, s) = \lim_{\delta \rightarrow 0} O(\delta) = 0.$$

Combining (A.6), (A.7), and (A.8), we obtain

$$\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} E|x^\varepsilon(t+s) - x^\varepsilon(t)|^2 = 0.$$

The criteria due to Kurtz [17, p. 47] then yields that  $\{x^\varepsilon(\cdot)\}$  is tight in  $D^r[0, T]$ .

As far as the estimates of  $y^\varepsilon(\cdot)$  are concerned, we merely note that

$$\begin{aligned} E_t^\varepsilon |y^\varepsilon(t+s) - y^\varepsilon(t)|^2 &= E_t^\varepsilon \left| \varepsilon \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} C(\alpha_k^\varepsilon) x_k^\varepsilon + \sqrt{\varepsilon} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \sigma_v(\alpha_k^\varepsilon) v_k \right|^2 \\ &\leq KE_t^\varepsilon \left| \varepsilon \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} C(\alpha_k^\varepsilon) x_k^\varepsilon \right|^2 + KE_t^\varepsilon \left| \sqrt{\varepsilon} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \sigma_v(\alpha_k^\varepsilon) v_k \right|^2. \end{aligned}$$

The rest of the estimates are all similar to the previous case. Thus we also have that  $\{y^\varepsilon(\cdot)\}$  is tight in  $D^r[0, T]$ .  $\square$

*Proof of Theorem 3.4.* Consider  $\{x^\varepsilon(\cdot)\}$  first. In fact, we work with the pair  $(x^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$ . Owing to the tightness of  $\{x^\varepsilon(\cdot)\}$  and the weak convergence of  $\{\bar{\alpha}^\varepsilon(\cdot)\}$ ,  $\{(x^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))\}$  is tight. By virtue of the Prohorov theorem [13, p. 104], we can extract a weakly convergent subsequence. Select such a subsequence, and still denote it by  $\{(x^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))\}$  for simplicity. Denote the limit of the sequence by  $(x(\cdot), \bar{\alpha}(\cdot))$ . By the Skorohod representation [13, p. 102], we may assume without loss of generality that  $(x^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$  converges to  $(x(\cdot), \bar{\alpha}(\cdot))$  with probability one (w.p.1). Moreover, the convergence is uniform on each bounded time interval. We proceed to use martingale averaging techniques to figure out the limit.

To obtain the desired limit, it suffices to show that the limit  $(x(\cdot), \bar{\alpha}(\cdot))$  is the solution of a martingale problem with operator  $\mathcal{L}$  given by

$$(A.9) \quad \mathcal{L}f(x, i) = f'_x(x, i)\bar{A}(i)x + \frac{1}{2}\text{tr}[f_{xx}(x, i)\bar{\sigma}_w(i)\bar{\sigma}_w'(i)] + \bar{Q}f(x, \cdot)(i), \quad i \in \bar{\mathcal{M}},$$

where

$$\bar{Q}f(x, \cdot)(i) = \sum_{j \in \bar{\mathcal{M}}} \bar{q}_{ij}f(x, j) = \sum_{j \in \bar{\mathcal{M}}, j \neq i} \bar{q}_{ij}(f(x, j) - f(x, i))$$

for each  $i \in \bar{\mathcal{M}}$ , and  $f(\cdot, i) \in C_0^2$  (twice continuously differentiable function with compact support). Since the filtering equation is linear in the state variable, by using a similar argument to that in [29, Lemma 7.18], the corresponding martingale problem with operator  $\mathcal{L}$  given in (A.9) has a unique solution.

To obtain the desired results, it suffices to show (see [19, Chapters 7 and 8]), for any positive integer  $k_0$ , any bounded and continuous function  $h_\kappa(\cdot)$  with  $\kappa \leq k_0$ , any  $t, s > 0$ , and  $t_\kappa \leq t \leq t + s$ , that the following equation holds:

$$(A.10) \quad E \prod_{\kappa=1}^{k_0} h_\kappa(x(t_\kappa), \bar{\alpha}(t_\kappa)) \left( f(x(t+s), \bar{\alpha}(t+s)) - f(x(t), \bar{\alpha}(t)) - \int_t^{t+s} \mathcal{L}f(x(u), \bar{\alpha}(u)) du \right) = 0.$$

To obtain (A.10), we begin with the pair  $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ . For each  $x$ , define  $\check{f}$  by

$$(A.11) \quad \check{f}(x, \alpha) = \sum_{i=1}^l f(x, i) I_{\{\alpha \in \mathcal{M}_i\}} \text{ for each } \alpha \in \mathcal{M}.$$

Note that, for each  $\alpha = s_{ij} \in \mathcal{M}_i$ ,  $\check{f}(x, \alpha)$  takes a constant value  $f(x, i)$ . Note also that, at any time instant  $t$ ,  $\alpha^\varepsilon(t) = \alpha_{i/\varepsilon}^\varepsilon$  takes on one of the  $m$  possible values from  $\mathcal{M}$ .

Note that  $\check{f}(x_k^\varepsilon, \alpha_k^\varepsilon) = f(x_k^\varepsilon, \bar{\alpha}_k^\varepsilon)$  for each  $k$ . Choose a sequence of positive integers  $\{n_\varepsilon\}$  such that  $n_\varepsilon \rightarrow \infty$  but  $\delta_\varepsilon = \varepsilon n_\varepsilon \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . The piecewise constant interpolation implies that

$$(A.12) \quad \begin{aligned} & \check{f}(x^\varepsilon(t+s), \alpha^\varepsilon(t+s)) - \check{f}(x^\varepsilon(t), \alpha^\varepsilon(t)) \\ &= \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} [\check{f}(x_{ln_\varepsilon+n_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon+n_\varepsilon}^\varepsilon) - \check{f}(x_{ln_\varepsilon+n_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon)] \\ & \quad + \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} [\check{f}(x_{ln_\varepsilon+n_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon) - \check{f}(x_{ln_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon)], \end{aligned}$$

and hence

$$(A.13) \quad \begin{aligned} & \lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) [\check{f}(x^\varepsilon(t+s), \alpha^\varepsilon(t+s)) - \check{f}(x^\varepsilon(t), \alpha^\varepsilon(t))] \\ &= \lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} [\check{f}(x_{ln_\varepsilon+n_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon+n_\varepsilon}^\varepsilon) - \check{f}(x_{ln_\varepsilon+n_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon)] \\ & \quad + \lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} [\check{f}(x_{ln_\varepsilon+n_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon) - \check{f}(x_{ln_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon)] \\ & \stackrel{\text{def}}{=} \lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) [g_1^\varepsilon + g_2^\varepsilon]. \end{aligned}$$

In the above,  $\sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon}$  can also be written as  $\sum_{ln_\varepsilon=t/\varepsilon}^{((t+s)/\varepsilon)-1}$ . We proceed to obtain the desired limit by examining  $g_i^\varepsilon$  ( $i = 1, 2$ ) in (A.13).



By virtue of a Taylor expansion, rewrite  $g_2^\varepsilon$  as

$$\begin{aligned}
 g_2^\varepsilon &= \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \check{f}'_x(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}^\varepsilon)[x_{ln_\varepsilon+n_\varepsilon}^\varepsilon - x_{ln_\varepsilon}^\varepsilon] \\
 &\quad + \frac{1}{2} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} [x_{ln_\varepsilon+n_\varepsilon}^\varepsilon - x_{ln_\varepsilon}^\varepsilon]' \check{f}_{xx}(x_{ln_\varepsilon}^+, \alpha_{ln_\varepsilon}^\varepsilon)[x_{ln_\varepsilon+n_\varepsilon}^\varepsilon - x_{ln_\varepsilon}^\varepsilon] \\
 &= \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \check{f}'_x(x_{ln_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon)[\varepsilon A(\alpha_k^\varepsilon)x_k^\varepsilon + \sqrt{\varepsilon}\sigma_w(\alpha_k^\varepsilon)w_k] \\
 (A.14) \quad &+ \frac{1}{2} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} [\varepsilon A(\alpha_k^\varepsilon)x_k^\varepsilon + \sqrt{\varepsilon}\sigma_w(\alpha_k^\varepsilon)w_k]' \check{f}_{xx}(x_{ln_\varepsilon}^+, \alpha_{ln_\varepsilon}^\varepsilon) \\
 &\times \sum_{k_1=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} [\varepsilon A(\alpha_{k_1}^\varepsilon)x_{k_1}^\varepsilon + \sqrt{\varepsilon}\sigma_w(\alpha_{k_1}^\varepsilon)w_{k_1}]
 \end{aligned}$$

$$(A.15) \stackrel{\text{def}}{=} \left[ g_{2,1}^\varepsilon + \frac{1}{2}g_{2,2}^\varepsilon \right],$$

where  $x_{ln_\varepsilon}^+$  is on the line segment joining  $x_{ln_\varepsilon}^\varepsilon$  and  $x_{ln_\varepsilon+n_\varepsilon}^\varepsilon$ .

Then we have

$$\begin{aligned}
 &E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \left[ \sqrt{\varepsilon} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \check{f}'_x(x_{ln_\varepsilon}^\varepsilon, \alpha_{ln_\varepsilon}^\varepsilon) \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \sigma_w(\alpha_k^\varepsilon)w_k \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \left[ \sqrt{\varepsilon} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \check{f}'_x(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}^\varepsilon) \right. \\
 &\quad \left. \times \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} E_{ln_\varepsilon} \sigma_w(\alpha_k^\varepsilon) E_{ln_\varepsilon} w_k \right].
 \end{aligned}$$

In the above, the second line is a consequence of the independence of  $\{\alpha_n^\varepsilon\}$  and  $\{w_n\}$  and the measurability of  $x_{ln_\varepsilon}^\varepsilon$  and  $\alpha_{ln_\varepsilon}^\varepsilon$  with respect to  $\mathcal{F}_{ln_\varepsilon}$ . In view of the boundedness of  $h_\kappa(\cdot)$  and  $\check{f}_x(\cdot)$  and the finiteness of  $\sigma_w(\alpha_k^\varepsilon)$ , we obtain

$$\begin{aligned}
 &E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \left[ \sqrt{\varepsilon} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \check{f}'_x(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}^\varepsilon) E_{ln_\varepsilon} \sigma_w(\alpha_k^\varepsilon) E_{ln_\varepsilon} w_k \right] \\
 &\rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \\
 (A.16)
 \end{aligned}$$

Next, let us treat the term on the second line of (A.14). First note that

$$\begin{aligned}
 &E \left| \varepsilon \sum_{i=1}^l \sum_{j=1}^{m_i} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} A(s_{ij})x_k^\varepsilon [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right| \\
 &\leq K \sum_{i=1}^l \sum_{j=1}^{m_i} E \left| \varepsilon \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} A(s_{ij})x_k^\varepsilon [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right|.
 \end{aligned}$$

Thus we need only examine the terms with fixed indices  $i$  and  $j$ . By a partial summation,

$$\begin{aligned} & E \left| \varepsilon \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} A(s_{ij})x_k^\varepsilon [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right| \\ & \leq KE \left| \varepsilon A(s_{ij})x_{ln_\varepsilon+n_\varepsilon-1}^\varepsilon \sum_{k=0}^{ln_\varepsilon+n_\varepsilon-1} [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right| \\ & \quad + KE \left| \varepsilon A(s_{ij})x_{ln_\varepsilon-1}^\varepsilon \sum_{k=0}^{ln_\varepsilon-1} [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right| \\ & \quad + KE \left| \varepsilon \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-2} (x_k^\varepsilon - x_{k+1}^\varepsilon) \sum_{k_0=0}^k [I_{\{\alpha_{k_0}^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_{k_0}^\varepsilon \in \mathcal{M}_i\}}] \right|. \end{aligned}$$

Using the mean square estimates on the occupation measures (2.10) and Lemma 3.2,

$$\begin{aligned} & E \left| \varepsilon A(s_{ij})x_{ln_\varepsilon+n_\varepsilon-1}^\varepsilon \sum_{k=0}^{ln_\varepsilon+n_\varepsilon-1} [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right| \\ & \leq KE^{1/2} |x_{ln_\varepsilon+n_\varepsilon-1}^\varepsilon|^2 E^{1/2} \left| \varepsilon \sum_{k=0}^{ln_\varepsilon+n_\varepsilon-1} [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right|^2 \\ & = O(\sqrt{\varepsilon}) \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \end{aligned}$$

Similarly,

$$E \left| \varepsilon A(s_{ij})x_{ln_\varepsilon-1}^\varepsilon \sum_{k=0}^{ln_\varepsilon-1} [I_{\{\alpha_k^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}}] \right| \rightarrow 0 \text{ as } \varepsilon \rightarrow 0.$$

Using (2.2), the Cauchy–Schwarz inequality, and the mean square estimates (2.10),

$$\begin{aligned} & E \left| \varepsilon \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-2} (x_k^\varepsilon - x_{k+1}^\varepsilon) \sum_{k_0=0}^k [I_{\{\alpha_{k_0}^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_{k_0}^\varepsilon \in \mathcal{M}_i\}}] \right| \\ & \leq \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-2} E^{1/2} |x_k^\varepsilon - x_{k+1}^\varepsilon|^2 E^{1/2} \left| \varepsilon \sum_{k_0=0}^k [I_{\{\alpha_{k_0}^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_{k_0}^\varepsilon \in \mathcal{M}_i\}}] \right|^2 \\ & \leq \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-2} E^{1/2} |\varepsilon A(\alpha_k^\varepsilon)x_k^\varepsilon + \sqrt{\varepsilon}\sigma_w(\alpha_k^\varepsilon)w_k|^2 \\ & \quad \times E^{1/2} \left| \varepsilon \sum_{k_0=0}^k [I_{\{\alpha_{k_0}^\varepsilon=s_{ij}\}} - \nu_j^i I_{\{\alpha_{k_0}^\varepsilon \in \mathcal{M}_i\}}] \right|^2 \\ & \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \end{aligned}$$

Using the above estimates and the continuity of  $\check{f}_x(\cdot, \alpha)$  for each  $\alpha \in \mathcal{M}$ ,

$$\begin{aligned}
 & E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \left[ \varepsilon \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} \check{f}'_x(x_{ln_{\varepsilon}}^{\varepsilon}, \alpha_{ln_{\varepsilon}}^{\varepsilon}) \sum_{k=ln_{\varepsilon}}^{ln_{\varepsilon}+n_{\varepsilon}-1} A(\alpha_k^{\varepsilon}) x_k^{\varepsilon} \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \left[ \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} \sum_{i=1}^l \sum_{j=1}^{m_i} \check{f}'_x(x_{ln_{\varepsilon}}^{\varepsilon}, \alpha_{ln_{\varepsilon}}^{\varepsilon}) \right. \\
 & \quad \left. \times \frac{\delta_{\varepsilon}}{n_{\varepsilon}} \sum_{k=ln_{\varepsilon}}^{ln_{\varepsilon}+n_{\varepsilon}-1} A(s_{ij}) x_{ln_{\varepsilon}}^{\varepsilon} v_j^i I_{\{\alpha_k^{\varepsilon} \in \mathcal{M}_i\}} \right] + o(1),
 \end{aligned}
 \tag{A.17}$$

where  $o(1) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . Then, as  $\varepsilon \rightarrow 0$ , letting  $\varepsilon ln_{\varepsilon} \rightarrow u$ , and using the techniques of [19, Chapter 8], (A.17) together with (A.16) leads to

$$\begin{aligned}
 & E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) g_{2,1}^{\varepsilon} \\
 & \tag{A.18} \rightarrow E \prod_{\kappa=1}^{k_0} h_{\kappa}(x(t_{\kappa}), \bar{\alpha}(t_{\kappa})) \left( \int_t^{t+s} f'_x(x(u), \bar{\alpha}(u)) A(\bar{\alpha}(u)) x(u) du \right) \text{ as } \varepsilon \rightarrow 0.
 \end{aligned}$$

As for  $g_{2,2}^{\varepsilon}$ , we have, by the continuity of  $f_{xx}(\cdot, \alpha)$  for each  $\alpha \in \mathcal{M}$ ,  $x_{ln_{\varepsilon}}^+ - x_{ln_{\varepsilon}}^{\varepsilon} \rightarrow 0$  in probability as  $\varepsilon \rightarrow 0$ . Consequently,

$$E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) g_{2,2}^{\varepsilon} \stackrel{\text{def}}{=} \tilde{g}_{2,2}^{\varepsilon} + o(1),$$

where  $o(1) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  uniformly in  $t$ , and

$$\begin{aligned}
 \tilde{g}_{2,2}^{\varepsilon} &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \\
 & \times \left[ \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} \sum_{k=ln_{\varepsilon}}^{ln_{\varepsilon}+n_{\varepsilon}-1} [\varepsilon A(\alpha_k^{\varepsilon}) x_k^{\varepsilon} + \sqrt{\varepsilon} \sigma_w(\alpha_k^{\varepsilon}) w_k] \check{f}_{xx}(x_{ln_{\varepsilon}}^{\varepsilon}, \alpha_{ln_{\varepsilon}}^{\varepsilon}) \right. \\
 & \quad \left. \times \sum_{k_1=ln_{\varepsilon}}^{ln_{\varepsilon}+n_{\varepsilon}-1} [\varepsilon A(\alpha_{k_1}^{\varepsilon}) x_{k_1}^{\varepsilon} + \sqrt{\varepsilon} \sigma_w(\alpha_{k_1}^{\varepsilon}) w_{k_1}] \right].
 \end{aligned}$$

It then follows that

$$\begin{aligned}
 \tilde{g}_{2,2}^\varepsilon &= E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \\
 &\times \left[ \varepsilon^2 \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (A(\alpha_k^\varepsilon) x_k^\varepsilon)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sum_{k_1=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} A(\alpha_{k_1}^\varepsilon) x_{k_1}^\varepsilon \right. \\
 &+ \sqrt{\varepsilon^3} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (A(\alpha_k^\varepsilon) x_k^\varepsilon)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sum_{k_1=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} \sigma_w(\alpha_{k_1}^\varepsilon) w_{k_1} \\
 &+ \sqrt{\varepsilon^3} \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (\sigma_w(\alpha_k^\varepsilon) w_k)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sum_{k_1=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} A(\alpha_{k_1}^\varepsilon) x_{k_1}^\varepsilon \\
 &\left. + \varepsilon \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (\sigma_w(\alpha_k^\varepsilon) w_k)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sum_{k_1=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} \sigma_w(\alpha_{k_1}^\varepsilon) w_{k_1} \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \\
 &\times \left[ \varepsilon \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (\sigma_w(\alpha_k^\varepsilon) w_k)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sigma_w(\alpha_k^\varepsilon) w_k \right] + o(1),
 \end{aligned}$$

where  $o(1) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . Furthermore, using the idea of the estimates leading to (A.17) and the mean square estimates (2.10), it can be shown that

$$\begin{aligned}
 &\varepsilon \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (\sigma_w(\alpha_k^\varepsilon) w_k)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sigma_w(\alpha_k^\varepsilon) w_k \\
 &= \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{i=1}^l \sum_{j=1}^{m_i} \frac{\delta_\varepsilon}{n_\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} \text{tr}[\check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, s_{ij}) \sigma_w(s_{ij}) w_k w_k' \sigma_w'(s_{ij})] I_{\{\alpha_k^\varepsilon = s_{ij}\}} \\
 &= \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{i=1}^l \sum_{j=1}^{m_i} \frac{\delta_\varepsilon}{n_\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} \text{tr}[\check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sigma_w(s_{ij}) w_k w_k' \sigma_w'(s_{ij})] \\
 &\quad \times \nu_j^i I_{\{\alpha_k^\varepsilon \in \mathcal{M}_i\}} + o(1),
 \end{aligned}$$

where  $o(1) \rightarrow 0$  in probability as  $\varepsilon \rightarrow 0$  uniformly in  $t$ . It then follows that

$$\begin{aligned}
 \text{(A.19)} \quad &\lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_\kappa(x^\varepsilon(t_\kappa), \bar{\alpha}^\varepsilon(t_\kappa)) \\
 &\times \left[ \varepsilon \sum_{l:t \leq l\delta_\varepsilon \leq (t+s)-\varepsilon} \sum_{k=l n_\varepsilon}^{l n_\varepsilon + n_\varepsilon - 1} (\sigma_w(\alpha_k^\varepsilon) w_k)' \check{f}_{xx}(x_{l n_\varepsilon}^\varepsilon, \alpha_{l n_\varepsilon}^\varepsilon) \sigma_w(\alpha_k^\varepsilon) w_k \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_\kappa(x(t_\kappa), \bar{\alpha}(t_\kappa)) \left[ \int_t^{t+s} \text{tr}[f_{xx}(x(u), \bar{\alpha}(u)) \sigma_w(\bar{\alpha}(u)) \sigma_w'(\bar{\alpha}(u))] du \right].
 \end{aligned}$$

Next, we consider the term  $g_1^\varepsilon$ . Using the continuity of  $\check{f}(\cdot, \alpha)$  for each  $\alpha \in \mathcal{M}$ , the Markov property of  $\alpha_n^\varepsilon$ , the mean square estimate (2.10) of the occupation measures,

(2.1), and Lemma 2.1, we have

$$\begin{aligned}
 \text{(A.20)} \quad & E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) g_1^{\varepsilon} \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \left[ \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} (\check{f}(x_{l n_{\varepsilon}}^{\varepsilon}, \alpha_{l n_{\varepsilon}+n_{\varepsilon}}^{\varepsilon}) - \check{f}(x_{l n_{\varepsilon}}^{\varepsilon}, \alpha_{l n_{\varepsilon}}^{\varepsilon})) \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \left[ \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} \sum_{k=l n_{\varepsilon}}^{l n_{\varepsilon}+n_{\varepsilon}-1} \sum_{i_1=1}^l \sum_{j_1=1}^{m_{i_1}} \left[ \sum_{i=1}^l \sum_{j=1}^{m_i} \check{f}(x_{i n_{\varepsilon}}^{\varepsilon}, s_{ij}) \right. \right. \\
 &\quad \left. \left. \times P(\alpha_{k+1}^{\varepsilon} = s_{ij} | \alpha_k^{\varepsilon} = s_{i_1 j_1}) - \check{f}(x_{i n_{\varepsilon}}^{\varepsilon}, s_{i_1 j_1}) \right] I_{\{\alpha_k^{\varepsilon} = s_{i_1 j_1}\}} \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \left[ \varepsilon \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} \sum_{k=l n_{\varepsilon}}^{l n_{\varepsilon}+n_{\varepsilon}-1} (P - I + \varepsilon Q) \check{f}(x_{i n_{\varepsilon}}^{\varepsilon}, \cdot)(\alpha_k^{\varepsilon}) \right] \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) \left[ \varepsilon \sum_{l:t \leq l\delta_{\varepsilon} \leq (t+s)-\varepsilon} \sum_{k=l n_{\varepsilon}}^{l n_{\varepsilon}+n_{\varepsilon}-1} Q \check{f}(x_{i n_{\varepsilon}}^{\varepsilon}, \cdot)(\alpha_k^{\varepsilon}) \right] \\
 &\rightarrow \int_t^{t+s} \bar{Q} f(x(u), \bar{\alpha}(u)) du \quad \text{as } \varepsilon \rightarrow 0.
 \end{aligned}$$

Combining (A.18), (A.19), and (A.20),

$$\begin{aligned}
 \text{(A.21)} \quad & \lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) [\check{f}(x^{\varepsilon}(t+s), \bar{\alpha}^{\varepsilon}(t+s)) - \check{f}(x^{\varepsilon}(t), \bar{\alpha}^{\varepsilon}(t))] \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x(t_{\kappa}), \bar{\alpha}(t_{\kappa})) \left[ \int_t^{t+s} \mathcal{L} f(x(u), \bar{\alpha}(u)) du \right].
 \end{aligned}$$

On the other hand, by the weak convergence of  $(x^{\varepsilon}(\cdot), \bar{\alpha}^{\varepsilon}(\cdot))$  to  $(x(\cdot), \bar{\alpha}(\cdot))$ , the Skorohod representation, and the definition of  $\check{f}(\cdot)$ , we have

$$\begin{aligned}
 \text{(A.22)} \quad & \lim_{\varepsilon \rightarrow 0} E \prod_{\kappa=1}^{k_0} h_{\kappa}(x^{\varepsilon}(t_{\kappa}), \bar{\alpha}^{\varepsilon}(t_{\kappa})) [\check{f}(x^{\varepsilon}(t+s), \alpha^{\varepsilon}(t+s)) - \check{f}(x^{\varepsilon}(t), \alpha^{\varepsilon}(t))] \\
 &= E \prod_{\kappa=1}^{k_0} h_{\kappa}(x(t_{\kappa}), \bar{\alpha}(t_{\kappa})) [f(x(t+s), \bar{\alpha}(t+s)) - f(x(t), \bar{\alpha}(t))].
 \end{aligned}$$

By (A.21) and (A.22), (A.10) holds. Using the same techniques, detailed estimates yield the second equation in (3.4). Thus the desired results follow.  $\square$

REFERENCES

[1] Y. BAR-SHALOM AND X. R. LI, *Estimation and Tracking: Principles, Techniques, and Software*, Artech House Publishers, Norwood, MA, 1996.  
 [2] D. P. BERTSEKAS, *Dynamic Programming and Stochastic Control*, Academic Press, New York, 1976.

- [3] T. R. BIELECKI AND L. STETTNER, *Ergodic control of a singularly perturbed Markov process in discrete time with general state and compact action spaces*, Appl. Math. Optim., 38 (1998), pp. 261–281.
- [4] T. BJÖRK, *Finite-dimensional optimal filters for a class of Itô processes with jumping parameters*, Stochastics, 4 (1980), pp. 167–183.
- [5] G. BLANKENSHIP, *Singularly perturbed difference equations in optimal control problems*, IEEE Trans. Automat. Control, 26 (1981), pp. 911–917.
- [6] O. L. V. COSTA, *Linear minimum mean square error estimation for discrete-time Markov jump linear systems*, IEEE Trans. Automat. Control, 39 (1994), pp. 1685–1689.
- [7] P. J. COURTOIS, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, 1977.
- [8] D. P. DE FARIAS, J. C. GEROMEL, J. B. R. DO VAL, AND O. L. V. COSTA, *Output feedback control of Markov jump linear systems in continuous time*, IEEE Trans. Automat. Control, 45 (2000), pp. 944–949.
- [9] S. DEY, *Reduced-complexity filtering for partially observed nearly completely decomposable Markov chains*, IEEE Trans. Signal Process., 48, (2000), pp. 3334–3344.
- [10] A. DOUCET, N. J. GORDON, AND V. KRISHNAMURTHY, *Particle filtering for state estimation for jump Markov linear systems*, IEEE Trans. Signal Process., 49 (2001), pp. 613–624.
- [11] F. DUFOUR AND P. BERTRAND, *The filtering problem for continuous-time linear systems with Markovian switching coefficients*, Systems Control Lett., 23 (1994), pp. 453–461.
- [12] F. DUFOUR AND R. J. ELLIOTT, *Adaptive control of linear systems with Markov perturbations*, IEEE Trans. Automat. Control, 43 (1997), pp. 351–372.
- [13] S. N. ETHIER AND T. G. KURTZ, *Markov Processes: Characterization and Convergence*, John Wiley, New York, 1986.
- [14] A. IL'IN, R. Z. KHASHMINSKII, AND G. YIN, *Singularly perturbed switching diffusions: Rapid switchings and fast diffusions*, J. Optim. Theory Appl., 102 (1999), pp. 555–591.
- [15] M. IOSIFESCU, *Finite Markov Processes and Their Applications*, John Wiley, Chichester, UK, 1980.
- [16] R. Z. KHASHMINSKII, G. YIN, AND Q. ZHANG, *Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains*, SIAM J. Appl. Math., 56 (1996), pp. 277–293.
- [17] H. J. KUSHNER, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.
- [18] H. J. KUSHNER, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser Boston, Boston, 1990.
- [19] H. J. KUSHNER AND G. YIN, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, 1997.
- [20] R. S. LIPTSER AND A. N. SHIRYAYEV, *Statistics of Random Processes I & II*, Springer-Verlag, New York, 2001.
- [21] B. M. MILLER AND W. J. RUNGALDIER, *Kalman filtering for linear systems with coefficients driven by a hidden Markov jump process*, Systems Control Lett., 31 (1997), pp. 93–102.
- [22] A. A. PERVOZVANSKII AND V. G. GAITSGORI, *Theory of Suboptimal Decisions: Decomposition and Aggregation*, Kluwer, Dordrecht, The Netherlands, 1988.
- [23] R. G. PHILLIPS AND P. V. KOKOTOVIC, *A singular perturbation approach to modelling and control of Markov chains*, IEEE Trans. Automat. Control, 26 (1981), pp. 1087–1094.
- [24] W. J. RUNGALDIER AND C. VISENTIN, *Combined filtering and parameter estimation: Approximation and robustness*, Automatica J. IFAC, 26 (1990), pp. 401–404.
- [25] S. P. SETHI AND Q. ZHANG, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser Boston, Boston, 1994.
- [26] H. A. SIMON AND A. ANDO, *Aggregation of variables in dynamic systems*, Econometrica, 29 (1961), pp. 111–138.
- [27] D. N. C. TSE, R. G. GALLAGER, AND J. N. TSITSIKLIS, *Statistical multiplexing of multiple time-scale Markov streams*, IEEE J. Selected Areas Comm., 13 (1995), pp. 1028–1038.
- [28] C. YANG, Y. BAR-SHALOM, AND C.-F. LIN, *Discrete-time point process filter for mode estimation*, IEEE Trans. Automat. Control, 37 (1992), pp. 1812–1816.
- [29] G. YIN AND Q. ZHANG, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*, Springer-Verlag, New York, 1998.
- [30] G. YIN AND Q. ZHANG, *Singularly perturbed discrete-time Markov chains*, SIAM J. Appl. Math., 61 (2000), pp. 834–854.
- [31] G. YIN, Q. ZHANG, AND G. BADOWSKI, *Asymptotic properties of a singularly perturbed Markov chain with inclusion of transient states*, Ann. Appl. Probab., 10 (2000), pp. 549–572.

- [32] G. YIN, Q. ZHANG, AND G. BADOWSKI, *Decomposition and aggregation of large-dimensional Markov chains in discrete time*, in Proceedings of the 40th IEEE Conference on Decision and Control, IEEE Control Systems Society, Piscataway, NJ, 2001, pp. 1687–1692.
- [33] Q. ZHANG, *Hybrid filtering for linear systems with non-Gaussian disturbances*, IEEE Trans. Automat. Control, 45 (2000), pp. 50–61.
- [34] Q. ZHANG AND G. YIN, *On nearly optimal controls of hybrid LQG problems*, IEEE Trans. Automat. Control, 44 (1999), pp. 2271–2282.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.