

Technical Notes and Correspondence

The Wonham Filter With Random Parameters: Rate of Convergence and Error Bounds

X. Guo and G. Yin

Abstract—Let $\alpha(t)$ be a finite-state continuous-time Markov chain with generator $Q = (q^{ij}) \in \mathbb{R}^{m \times m}$ and state space $\mathcal{M} = \{z^1, \dots, z^m\}$, where z^i for $i \leq m$ are distinct real numbers. When the state-space and the generator are known *a priori*, the best estimator of $\alpha(t)$ (in terms of mean square error) under noisy observation is the classical Wonham filter. This note addresses the estimation issue when values of the state-space or values of the generator are *unknown a priori*. In each case, we propose a (suboptimal) filter and prove its convergence to the desired Wonham filter under simple conditions. Moreover, we obtain the rate of convergence using both the mean square and the higher moment error bounds.

Index Terms—Approximation, error bounds, rate of convergence, Wonham filter.

I. INTRODUCTION

Given a probability space (Ω, \mathcal{F}, P) and $t \in [0, T]$ for some $T > 0$. Suppose that $\alpha(t)$ is a finite-state continuous-time Markov process with generator $Q = (q^{ij}) \in \mathbb{R}^{m \times m}$ and state space $\mathcal{M} = \{z^1, \dots, z^m\}$, where z^i for $i \leq m$ are distinct real numbers, so that the transition probabilities are

$$P^{ij}(h) = P(\alpha(t+h) = z^j | \alpha(t) = z^i) \quad (1)$$

$$P^{ij}(h) = \begin{cases} 1 - q^i h + o(h), & i = j \quad h \rightarrow 0 \\ q^{ij} h + o(h), & i \neq j \quad h \rightarrow 0 \end{cases} \quad (2)$$

where $q^i = \sum_{j \neq i} q^{ij}$.

Assume that the Markov process $\alpha(t)$ is observed with the observation process $y(t)$ such that

$$\begin{cases} dy(t) = \alpha(t)dt + \sigma(t)dw(t) \\ y(0) = 0 \text{ w.p. } 1 \end{cases} \quad (3)$$

where $w(\cdot)$ is a standard one-dimensional Brownian motion that is independent of $\alpha(t)$, and that $\sigma(\cdot) : \mathbb{R} \mapsto \mathbb{R}$, with $\sigma(t) \geq c$ for all $t \in [0, T]$ and some $c > 0$, is a continuously differentiable function. Note that (3) indicates that the basic observation model has the form of “signal plus noise,” and that the distribution of $y(t)$ is non-Gaussian but is a Gaussian mixture due to the jump processes $\alpha(t)$.

In this framework, one of the classical results known as the *Wonham filter* concerns estimating $\alpha(t)$ based on the observation $y(\cdot)$. When the values of the states z^1, \dots, z^m and the generator Q are *known a*

priori and fixed, the Wonham filter [9] provides the optimal filter in the sense of mean square error. It was the first finite-dimensional filter for non-Gaussian processes, and remains as one of the very few known finite-dimensional filters to-date. (The first rigorous development of nonlinear filters for diffusion-type processes was [5]. For a more detailed treatment of filtering problems, see [7, Vols. 1, 2]; for a more recent reference on filtering, see [2]; for the use of Wonham filter in adaptive control and estimation, see [1] and references therein; for the use of Clark’s transformation, see [8]; for some recent references on filtering and estimation, see [10] and [12].

It is natural to take a step further and consider the estimation problem when there are additional uncertainties in the value of the state space z^1, \dots, z^m or of the generator (q^{ij}) . For instance, one may consider the case when z^i ’s are *unknown a priori*, and their values are known only up to a certain distribution. This problem arises frequently in for example, Bayesian statistics, where priors could be drawn from a distribution. One possible solution is to apply a Monte Carlo approach which entails the replacement of states in the Wonham filter by their simulated or approximated values. Such an approach raises important questions such as the rate of convergence and error estimates. To some extent, many of these problems may be viewed as a robustness issues—this is the central topic of this note.

Our first main result is the construction of approximating filters when only noisy or simulated values z^i or (q^{ij}) are available. We then prove that these suboptimal filters converge to the desired Wonham filter under simple ergodicity conditions. We also evaluate the accuracy of the approximations and derive the approximation error bounds. These bounds, including both mean square (or L_2 bounds) as well as bounds based on higher moments, provide results on the rate of convergence and enable one to assess the quality of the approximations.

The rest of the note is organized as follows. Section II contains the main results. Proofs of the main theorems are in Section III. Section IV concludes the note with some remarks.

II. MAIN RESULTS

A. Wonham Filter

Let $\alpha(t)$ be a finite-state continuous time Markov process with state-space $\mathcal{M} = \{z^1, \dots, z^m\}$ and generator $Q = (q^{ij}) \in \mathbb{R}^{m \times m}$, as defined earlier. Given (3), define

$$\begin{aligned} p(t) &= (p^1(t), \dots, p^m(t)) \in \mathbb{R}^{1 \times m} \quad \text{with} \\ p^i(t) &= P(\alpha(t) = z^i | y(s), 0 \leq s \leq t), \quad i = 1, \dots, m, p^i(0) = p_0^i. \end{aligned}$$

It was proved in [9] that this conditional density satisfies the following system of stochastic differential equations:

$$\begin{aligned} dp^i(t) &= \sum_{j=1}^m p^j(t) q^{ji} dt - \sigma^{-2}(t) \bar{\alpha}(t) [z^i - \bar{\alpha}(t)] p^i(t) dt \\ &\quad + \sigma^{-2}(t) [z^i - \bar{\alpha}(t)] p^i(t) dy(t), \quad i = 1, \dots, m \end{aligned} \quad (4)$$

where

$$\bar{\alpha}(t) = \langle p(t), z \rangle \stackrel{\text{def}}{=} \sum_{i=1}^m z^i p^i(t), \quad z = (z^1, \dots, z^m)'$$

Manuscript received December 10, 2003; revised May 21, 2005. Recommended by Associate Editor L. Gerencser. The work of X. Guo was supported in part by the National Security Agency under Grant H98230-05-1-0084. The work of G. Yin is supported by the National Science Foundation under Grant DMS-0304928.

X. Guo is with School of ORIE, Cornell University, Ithaca, NY 14853 USA (e-mail: xinguo@orie.cornell.edu).

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

Digital Object Identifier 10.1109/TAC.2005.864192

and v' denotes the transpose of v . Adopting a vector notation, define

$$A(t) \stackrel{\text{def}}{=} \text{diag}(z^1 - \bar{\alpha}(t), \dots, z^m - \bar{\alpha}(t)) \\ = \begin{pmatrix} z^1 - \bar{\alpha}(t) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & z^m - \bar{\alpha}(t) \end{pmatrix}.$$

Then the Wonham filter can be rewritten as shown in (5) at the bottom of the page.

B. Approximate Wonham Filter Using $\{\hat{z}_n\}$

Now, let us assume that z^i 's are not available, and that only their noise corrupted measurements or observations or distributional information are at our disposal. We assume further that (q^{ij}) remains unchanged and known *a priori*.

In particular, we assume that a sequence of observations of the form

$$\hat{z}_n = (\hat{z}_n^1, \dots, \hat{z}_n^m)' \in \mathbb{R}^{m \times 1} \text{ such that } E\hat{z}_n = z$$

can be obtained. For example, $\hat{z}_n = z + \xi_n$, where $\{\xi_n\}$ is a sequence of \mathbb{R}^m -valued zero mean observation noise satisfying appropriate conditions. This essentially means that although the information about states of the Markov chain is inaccurate, one can repeat many experiments to get samples of the states. For instance, such a sequence may be obtained by simulation using Markov chain Monte Carlo (MCMC). Based on this assumption, we proceed to construct the approximate filter. First, define

$$\bar{z}_n = \frac{1}{n} \sum_{j=1}^n \hat{z}_j. \quad (6)$$

Then in lieu of (5), we have a sequence of approximations $p_n(t)$ given by (7), as shown at the bottom of the page, where

$$\bar{\alpha}_n(t) = \langle p_n(t), \bar{z}_n \rangle \quad A_n(t) \\ = \text{diag}(\bar{z}_n^1 - \bar{\alpha}_n(t), \dots, \bar{z}_n^m - \bar{\alpha}_n(t)).$$

Since $\{p_n(t)\}$ is a sequence of approximations of the posterior density $p(t)$, we may appropriately normalize it to ensure its boundedness [9]. Define

$$e_n(t) = p_n(t) - p(t).$$

Then, $e_n(t)$ satisfies

$$\begin{cases} de_n(t) = e_n(t)Qdt - \sigma^{-2}(t)[\bar{\alpha}_n(t) - \bar{\alpha}(t)]p_n(t)A_n(t)dt \\ - \sigma^{-2}(t)\bar{\alpha}(t)e_n(t)A_n(t)dt \\ - \sigma^{-2}(t)\bar{\alpha}(t)p(t)[A_n(t) - A(t)]dt \\ + \sigma^{-2}(t)e_n(t)A_n(t)dy(t) \\ + \sigma^{-2}(t)p(t)[A_n(t) - A(t)]dy(t) \\ e_n(0) = 0. \end{cases} \quad (8)$$

To obtain the desired limit result, we impose the following conditions.

- A1) $\{\hat{z}_n\}$ is a stationary ergodic sequence that satisfies $E\hat{z}_n = z$ and is uniformly bounded. The sequence $\{\hat{z}_n\}$ is independent of the Markov chain $\alpha(\cdot)$ and the Brownian motion $w(\cdot)$.

Remark 1: The uniform boundedness of the sequence $\{\hat{z}_n\}$ is not a restriction since one may use, for example, truncated normal random distributions. Although no independence is assumed for the sequence $\{\hat{z}_n\}$, in simulation one often uses i.i.d. sequences for simplicity. Ergodicity implies that $\bar{z}_n \rightarrow z$ w.p.l. The conditions cover a large class of processes. Examples of interests include the case when $\{\hat{z}_n\}$ is a stationary ϕ -mixing sequence, which in turn is ergodic (see [4, pp. 488–489]). That is, $\bar{z}_n \rightarrow z$ w.p.l.

Theorem 2: Under assumption A1), $\sup_{0 \leq t \leq T} E|e_n(t)|^2 \rightarrow 0$ as $n \rightarrow \infty$.

It is well known that convergence in L_2 implies convergence in probability. Thus, the following is immediate.

Corollary 3: Under assumption A1), for any $\eta > 0$, $\lim_{n \rightarrow \infty} P(|e_n(t)| \geq \eta) = 0$.

Next, define $e_n^\kappa(t) = n^\kappa e_n(t)$ for any $0 < \kappa \leq 1/2$. Then, the following estimates hold.

Theorem 4: Under assumption A1), as $n \rightarrow \infty$

$$\sup_{0 \leq t \leq T} E|e_n^\kappa(t)|^2 = \begin{cases} o(1), & 0 < \kappa < 1/2 \\ O(1), & \kappa = 1/2. \end{cases} \quad (9)$$

In this, $o(1)$ means that $E|e_n^\kappa(t)|^2 \rightarrow 0$ uniformly in t , whereas $O(1)$ indicates that $E|e_n^\kappa(t)|^2$ is bounded uniformly. To proceed, we obtain higher moment bounds.

Theorem 5: Assume assumption A1).

- i) For any positive integer $\ell > 1$, as $n \rightarrow \infty$

$$\sup_{0 \leq t \leq T} E|e_n^\kappa|^{2\ell} = \begin{cases} o(1), & 0 < \kappa < 1/2 \\ O(1), & \kappa = 1/2. \end{cases}$$

- ii) For $\kappa = 1/2$, denote $\tilde{e}_n(t) = e_n^{1/2}(t)$. Then as $n \rightarrow \infty$

$$\sup_{0 \leq t \leq T} E \exp(|\tilde{e}_n(t)|) = O(1).$$

C. Approximate Wonham Filter Using $\{\hat{Q}_n\}$

Now, we assume uncertainty in (q^{ij}) with its sequence of noise corrupted observations $\hat{Q}_n = (\hat{q}_n^{ij}) \in \mathbb{R}^{m \times m}$. As in the previous section, the approximate filter can be built accordingly by defining

$$\bar{Q}_n = \frac{1}{n} \sum_{i=1}^n \hat{Q}_i \quad (10)$$

and by redefining a sequence of approximations $p_n(t)$ with (11), as shown at the bottom of the next page, where

$$\bar{\alpha}_n(t) = \langle p_n(t), z \rangle \quad A_n(t) = \text{diag}(z^1 - \bar{\alpha}_n(t), \dots, z^m - \bar{\alpha}_n(t)).$$

$$\begin{cases} dp(t) = p(t)Qdt - \sigma^{-2}(t)\bar{\alpha}(t)p(t)A(t)dt + \sigma^{-2}(t)p(t)A(t)dy(t) \\ p(0) = p_0 \end{cases} \quad (5)$$

$$\begin{cases} dp_n(t) = p_n(t)Qdt - \sigma^{-2}(t)\bar{\alpha}_n(t)p_n(t)A_n(t)dt + \sigma^{-2}(t)p_n(t)A_n(t)dy(t) \\ p_n(0) = p_0 \end{cases} \quad (7)$$

Let $e_n(t) \stackrel{\text{def}}{=} p_n(t) - p(t)$. Then

$$\begin{cases} de_n(t) = e_n(t)\bar{Q}_n dt + p(t)(\bar{Q}_n - Q)dt \\ -\sigma^{-2}(t)[\bar{\alpha}_n(t) - \bar{\alpha}(t)]p_n(t)A_n(t)dt \\ -\sigma^{-2}(t)\bar{\alpha}(t)e_n(t)A_n(t)dt \\ -\sigma^{-2}(t)\bar{\alpha}(t)p(t)[A_n(t) - A(t)]dt \\ +\sigma^{-2}(t)e_n(t)A_n(t)dy(t) \\ +\sigma^{-2}(t)p(t)[A_n(t) - A(t)]dy(t) \\ e_n(0) = 0. \end{cases} \quad (12)$$

A2) $\{\hat{Q}_n\}$ is a stationary ergodic sequence such that it is uniformly bounded, $E\hat{Q}_n = Q$, and $\{\hat{Q}_n\}$ is independent of the Markov chain $\alpha(\cdot)$ and the Brownian motion $w(\cdot)$.

In a similar fashion, we can derive the following results.

Theorem 6: Assume A2). Then, $\sup_{0 \leq i \leq T} E|e_n(t)|^2 \rightarrow 0$ as $n \rightarrow \infty$.

Remark 7: For ease of exposition, the assumption A2) is stronger than needed. For instance, Theorem 6 holds when the uniform boundedness of $\{\hat{Q}_n\}$ is replaced by $E|\hat{Q}_n|^2 < \infty$. This remark applies to the following theorems as well when the uniform boundedness is replaced by appropriate higher moment conditions.

Theorem 8: Assume A2).

i) For any positive integer $\ell > 1$, as $n \rightarrow \infty$

$$\sup_{0 \leq t \leq T} E|e_n^\kappa(t)|^{2\ell} = \begin{cases} o(1), & 0 < \kappa < 1/2 \\ O(1), & \kappa = 1/2. \end{cases} \quad (13)$$

ii) For $\kappa = 1/2$, denote $\tilde{e}_n(t) = e_n^{1/2}(t)$. Then, as $n \rightarrow \infty$

$$\sup_{0 \leq t \leq T} E \exp(|\tilde{e}_n(t)|) = O(1).$$

Remark 9: Theorems 4 and 8 provide a rate of convergence result. It asserts that the convergence speed is of the order $n^{1/2}$. For any $\kappa < 1/2$, a trivial limit is obtained. Note that the convergence is in the sense of mean square convergence. Thus, Theorems 4 and 8 tell us how the mean square error depends on the size of the sample.

Finally, the same techniques can be used to build Wonham filters using both $\{\hat{z}_n\}$ and $\{\hat{Q}_n\}$. Formally, the approximate filters so designed have the same form as (11). Assuming A1) and A2), and the independence of $\{\hat{z}_n\}$ and $\{\hat{Q}_n\}$, then the conclusions of Theorem 8 continue to hold.

III. PROOFS

This section presents the proofs of results. Since the proof of Theorem 4 uses similar techniques as that of Theorem 5, in order not to dwell on it, the proof of Theorem 5 proceeds that of Theorem 4.

A. Notation

Throughout the rest of this note, we use the notation $K, K_T, K_{T,\ell}$ to denote some generic positive constants, in which the subscripts highlight their dependence on the indicated quantities. These constants are generic in the sense their values may be different for different usage. That is, the convention $K + K = K$ and $KK = K$ is used. Before going through the proofs, here are some basic inequalities used throughout.

B. Review of Some Inequalities

Lemma 10: Suppose that $f(t)$ is \mathcal{F}_t -measurable and that it satisfies $\int_0^T E|f(t)|^2 dt < \infty$, where \mathcal{F}_t is the σ -algebra generated by $\{w(s), \alpha(s) : s \leq t\}$. Then, with $y(s)$ is given by (3)

$$E \left| \int_0^t f(s) dy(s) \right|^2 \leq K_T \int_0^t E|f(s)|^2 ds.$$

Proof: It is straightforward that

$$\begin{aligned} E \left| \int_0^t f(s) dy(s) \right|^2 &= E \left| \int_0^t f(s)[\alpha(s) ds + \sigma(s) dw(s)] \right|^2 \\ &\leq KE \left| \int_0^t f(s)\alpha(s) ds \right|^2 + KE \left| \int_0^t f(s)\sigma(s) dw(s) \right|^2 \\ &\leq KE \int_0^t |f(s)|^2 ds + K \int_0^t E|f(s)\sigma(s)|^2 ds \\ &\leq K_T \int_0^t E|f(s)|^2 ds. \end{aligned}$$

The lemma follows. \blacksquare

Lemma 11 (Gronwall's Inequality): Let $u(t)$ and $g(t)$ be nonnegative continuous functions on $[0, T]$ for which the inequality

$$u(t) \leq C + \int_0^t g(s)u(s) ds, \quad t \in [0, T]$$

holds, where C is a nonnegative constant. Then

$$u(t) \leq C \exp\left(\int_0^t g(s) ds\right), \quad t \in [0, T].$$

Proof: See [3, p. 36]. \blacksquare

Lemma 12: For any real numbers J_1, J_2, J_3, J_4, J_5 , and $r \geq 1$

$$|J_1 + J_2 + J_3 + J_4 + J_5|^r \leq K_r (|J_1|^r + |J_2|^r + |J_3|^r + |J_4|^r + |J_5|^r)$$

where K_r is a positive constant depending on r .

Proof: This is obtained via repeated applications of the following inequality: For any real numbers a, b , and $r \geq 1$, $|a + b|^r \leq 2^{r-1}(|a|^r + |b|^r)$.

C. Proof of Theorem 2

Writing (8) in a variational form and noting $e_n(0) = 0$, we obtain

$$\begin{aligned} e_n(t) &= - \int_0^t \sigma^{-2}(s)[\bar{\alpha}_n(s) - \bar{\alpha}(s)]p_n(s)A_n(s)e^{Q(t-s)} ds \\ &\quad - \int_0^t \sigma^{-2}(s)\bar{\alpha}(s)e_n(s)A_n(s)e^{Q(t-s)} ds \\ &\quad - \int_0^t \sigma^{-2}(s)\bar{\alpha}(s)p(s)[A_n(s) - A(s)]e^{Q(t-s)} ds \\ &\quad + \int_0^t \sigma^{-2}(s)e_n(s)A_n(s)e^{Q(t-s)} dy(s) \\ &\quad + \int_0^t \sigma^{-2}(s)p(s)[A_n(s) - A(s)]e^{Q(t-s)} dy(s). \end{aligned} \quad (14)$$

$$\begin{cases} dp_n(t) = p_n(t)\bar{Q}_n dt - \sigma^{-2}(t)\bar{\alpha}_n(t)p_n(t)A_n(t)dt + \sigma^{-2}(t)p_n(t)A_n(t)dy(t) \\ p_n(0) = p_0 \end{cases} \quad (11)$$

For a suitable function $g(\cdot)$ (e.g., \mathcal{F}_t -measurable and $\int_0^T E|g(t)|^2 dt < \infty$), by the Cauchy-Schwarz inequality

$$\begin{aligned} E \left| \int_0^t g(s) ds \right|^2 &\leq \left(\int_0^t ds \right) \left(\int_0^t E|g(s)|^2 ds \right) \\ &\leq K_T \int_0^t E|g(s)|^2 ds. \end{aligned} \quad (15)$$

Thus, we arrive at

$$\begin{aligned} E|e_n(t)|^2 &\leq K_T \left\{ \int_0^t E|\sigma^{-2}(s)|^2 |\bar{\alpha}_n(s) - \bar{\alpha}(s)|^2 p_n(s)|^2 \right. \\ &\quad \times |A_n(s)|^2 |e^{Q(t-s)}|^2 ds \\ &\quad + \int_0^t E|\sigma^{-2}(s)|^2 |\bar{\alpha}(s)|^2 |e_n(s)|^2 |A_n(s)|^2 |e^{Q(t-s)}|^2 ds \\ &\quad + \int_0^t E|\sigma^{-2}(s)|^2 |\bar{\alpha}(s)|^2 |p(s)|^2 |A_n(s) - A(s)|^2 |e^{Q(t-s)}|^2 ds \\ &\quad + \int_0^t E|\sigma^{-2}(s)|^2 |e_n(s)|^2 |A_n(s)|^2 |e^{Q(t-s)}|^2 ds \\ &\quad \left. + \int_0^t E|\sigma^{-2}(s)|^2 |p(s)|^2 |A_n(s) - A(s)|^2 |e^{Q(t-s)}|^2 ds \right\}. \end{aligned} \quad (16)$$

Since we are working with the finite horizon $[0, T]$, the uniform boundedness of $\sigma^{-2}(s)$, $p_n(s)$, $A_n(s)$, and $e^{Q(t-s)}$ on $0 \leq s \leq t \leq T$ then implies

$$\begin{aligned} E|e_n(t)|^2 &\leq K_T \left\{ \int_0^t E|\bar{\alpha}_n(s) - \bar{\alpha}(s)|^2 ds \right. \\ &\quad \left. + \int_0^t E|A_n(s) - A(s)|^2 ds + \int_0^t E|e_n(s)|^2 ds \right\}. \end{aligned} \quad (17)$$

Indeed, note that

$$\bar{\alpha}_n(s) - \bar{\alpha}(s) = \langle p_n(s) - p(s), \bar{z}_n \rangle + \langle p(s), \bar{z}_n - z \rangle.$$

Thus

$$\begin{aligned} \int_0^t E|\bar{\alpha}_n(s) - \bar{\alpha}(s)|^2 ds &\leq K_T \left\{ \int_0^t E|e_n(s)|^2 ds \right. \\ &\quad \left. + \int_0^t E|\bar{z}_n - z|^2 ds \right\}. \end{aligned} \quad (18)$$

Likewise

$$\begin{aligned} \int_0^t E|\bar{A}_n(s) - A(s)|^2 ds &\leq K_T \left\{ \int_0^t E|e_n(s)|^2 ds \right. \\ &\quad \left. + \int_0^t E|\bar{z}_n - z|^2 ds \right\}. \end{aligned} \quad (19)$$

Substituting (18) and (19) into (17), by Gronwall's inequality

$$E|e_n(t)|^2 \leq K_T \delta_n \exp(K_T t) \leq K_T \delta_n \exp(K_T T)$$

where

$$\delta_n = \int_0^T E|\bar{z}_n - z|^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Recall that K_T is a generic positive constant depending on T . We have

$$\sup_{0 \leq t \leq T} E|e_n(t)|^2 \leq K_T \delta_n \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The desired result then follows. \blacksquare

D. Proof of Theorem 5

1) *Part i)*: Using (14) and Lemma 12, we obtain that for a positive integer ℓ with $r = 2\ell > 2$

$$E|e_n^\kappa(t)|^{2\ell} \leq K_\ell (E|J_1|^{2\ell} + E|J_2|^{2\ell} + E|J_3|^{2\ell} + E|J_4|^{2\ell} + E|J_5|^{2\ell}) \quad (20)$$

where

$$\begin{aligned} J_1 &= \int_0^t n^\kappa \sigma^{-2}(s) [\bar{\alpha}_n(s) - \bar{\alpha}(s)] p_n(s) A_n(s) e^{Q(t-s)} ds \\ J_2 &= \int_0^t \sigma^{-2}(s) \bar{\alpha}(s) e_n^\kappa(s) A_n(s) e^{Q(t-s)} ds \\ J_3 &= \int_0^t n^\kappa \sigma^{-2}(s) \bar{\alpha}(s) p(s) [A_n(s) - A(s)] e^{Q(t-s)} ds \\ J_4 &= \int_0^t n^\kappa \sigma^{-2}(s) e_n(s) e^{Q(t-s)} dy(s) \\ J_5 &= \int_0^t n^\kappa \sigma^{-2}(s) p(s) [A_n(s) - A(s)] e^{Q(t-s)} dy(s). \end{aligned}$$

Similar to the estimates for $E|e_n^\kappa(t)|^2$ and in particular to (15), we have, from the Hölder's inequality

$$\begin{aligned} E \left| \int_0^t g(s) ds \right|^{2\ell} &\leq \left[\left(\int_0^t ds \right)^{1/(1-(1/(2\ell)))} \left(\int_0^t E|g(s)|^{2\ell} ds \right)^{1/(2\ell)} \right]^{2\ell} \\ &\leq K_{T,\ell} \int_0^t E|g(s)|^{2\ell} ds \end{aligned} \quad (21)$$

where $K_{T,\ell} > 0$ is a generic positive constant depending on T and ℓ .

This inequality, together with the fact that $\sigma^{-2}(\cdot)$, $p_n(\cdot)$, and $A_n(\cdot)$ are uniformly bounded in the interval $[0, T]$, and that $e^{Q(t-s)}$ is bounded for $0 \leq s \leq t \leq T$, suggests

$$\begin{aligned} E|J_1|^{2\ell} &\leq K_{T,\ell} \int_0^t n^{2\kappa\ell} E|\sigma^{-2}(s)|^{2\ell} |\bar{\alpha}_n(s) \\ &\quad - \bar{\alpha}(s)|^{2\ell} |p_n(s)|^2 |A_n(s)|^{2\ell} |e^{Q(t-s)}|^{2\ell} ds \\ &\leq K_{T,\ell} \int_0^t n^{2\kappa\ell} E|\bar{\alpha}_n(s) - \bar{\alpha}(s)|^{2\kappa\ell} ds. \end{aligned} \quad (22)$$

Likewise, we obtain

$$\begin{aligned} E|J_2|^{2\ell} &\leq K_{T,\ell} \int_0^t E|e_n^\kappa(s)|^{2\ell} ds \\ E|J_3|^{2\ell} &\leq K_{T,\ell} \int_0^t n^{2\kappa\ell} E|A_n(s) - A(s)|^{2\ell} ds. \end{aligned} \quad (23)$$

Let $f(\cdot)$ be a function satisfying $\int_0^T E|f(t)|^{2\ell} dt < \infty$. Similar to the proof of Lemma 10, with the use of [7, Vol. I, p. 131, Lemma 4.12], we obtain

$$E \left| \int_0^t f(s) dy(s) \right|^{2\ell} \leq K_{T,\ell} \int_0^t E|f(s)|^{2\ell} ds.$$

Thus

$$\begin{aligned} E|J_4|^{2\ell} &\leq K_{T,\ell} \int_0^t E|e_n^\kappa(s)|^{2\ell} ds \\ E|J_5|^{2\ell} &\leq K_{T,\ell} \int_0^t E|A_n(s) - A(s)|^{2\ell} ds. \end{aligned} \quad (24)$$

Combining the estimates obtained thus far and separating $e_n^\kappa(s)$ and $|\bar{z}_n - z|$ as done in (18) and (19), we arrive at

$$E |e_n^\kappa|^{2\ell} \leq K_{T,\ell} \left\{ \int_0^t n^{2\kappa\ell} E |\bar{z}_n - z|^{2\ell} + \int_0^t E |e_n^\kappa(s)|^{2\ell} ds \right\}. \quad (25)$$

Note that

$$E |\bar{z}_n - z|^{2\ell} \leq \frac{1}{n} \sum_{l=0}^n E |\hat{z}_l - z|^{2\ell}.$$

An application of the Gronwall's inequality yields that

$$E |e_n^\kappa(t)|^{2\ell} \leq K_{T,\ell} \zeta_n^\kappa \exp(K_{T,\ell} T)$$

and in addition

$$\sup_{0 \leq t \leq T} E |e_n^\kappa(t)|^{2\ell} \leq K_{T,\ell} \zeta_n^\kappa \exp(K_{T,\ell} T)$$

where

$$\zeta_n^\kappa = \int_0^T n^{2\kappa\ell} E |\bar{z}_n - z|^{2\ell} ds.$$

Therefore

$$\zeta_n^\kappa = \begin{cases} O(1), & \kappa = 1/2 \\ o(1), & 0 < \kappa < 1/2. \end{cases}$$

2) *Part ii*): In part i), we have shown that $|\tilde{e}_n(t)|^{2\ell} = O(1)$ and that the bound is uniform in $t \in [0, T]$. Similar to [6, p. 142], $E|\tilde{e}_n(t)|^{2\ell-1} \leq E^{(2\ell-1)/(2\ell)}|\tilde{e}_n(t)|^{2\ell} = O(1)$ uniformly in $t \in [0, T]$. In view of (21), we can make $K_{T,\ell} \leq K_T^{2\ell}$. Thus

$$E \exp(|\tilde{e}_n(t)|) = \sum_{l=0}^{\infty} \frac{E|\tilde{e}_n(t)|^l}{l!} \leq K \sum_{l=0}^{\infty} \frac{K_T^l}{l!} < \infty. \quad (26)$$

The proof is concluded. ■

Proof of Theorem 4: Similar to (17), we obtain

$$E |e_n^\kappa(t)|^2 \leq K_T \left\{ \int_0^t E(n^{2\kappa} |\bar{\alpha}_n(s) - \bar{\alpha}(s)|^2) ds + \int_0^t E(n^{2\kappa} |A_n(s) - A(s)|^2) ds + \int_0^t E |e_n^\kappa(s)|^2 ds \right\}.$$

Note that by virtue of Remark 1

$$E(n^{2\kappa} |\bar{\alpha}_n(t) - \alpha(t)|^2) = \begin{cases} o(1), & 0 < \kappa < 1/2 \\ O(1), & \kappa = 1/2 \end{cases}$$

$$E(n^{2\kappa} |A_n(t) - A(t)|^2) = \begin{cases} o(1), & 0 < \kappa < 1/2 \\ O(1), & \kappa = 1/2 \end{cases}$$

and the above bounds hold uniformly in $t \in [0, T]$. The rest of the proof is similar to the proof of previous theorems. ■

IV. FURTHER REMARKS

An approximation algorithm has been developed in this work. For simplicity, it was setup as a scalar problem. Extensions to vector-valued problems are straightforward. For example, one may consider the observation of the form

$$dy(t) = g(\alpha(t))dt + \sigma(t)dw(t), \quad y(0) = 0 \quad \text{w.p.1}$$

where $y(t) \in \mathbb{R}^r$, $g(\cdot) : \mathcal{M} \rightarrow \mathbb{R}^r$, $\sigma(\cdot) : \mathbb{R}^{r \times r} \rightarrow \mathbb{R}^r$, $w(\cdot)$ is standard r -dimensional Brownian motion, $\alpha(t)$ is a continuous-time Markov chain taking values in $\mathcal{M} = \{1, \dots, m\}$. Then one can proceed with the corresponding Wonham filter and approximation when the state values are observed with noise. The results can be carried over with few modifications. Another interesting problem is to construct approximation of Wonham filter based on observations taken at discrete time and/or by noisy observation of the underlying system instead of the noisy observation of the unknown system parameters. For a recent paper on this direction, the reader is referred to [11].

ACKNOWLEDGMENT

The authors would like to thank A. Dembo for stimulating discussions and references. They would also like to thank the referees for carefully reading the manuscript and for many constructive suggestions.

REFERENCES

- [1] P. E. Caines and J. F. Zhang, "On the adaptive control of jump parameter systems via nonlinear filtering," *SIAM J. Control Optim.*, vol. 33, pp. 1758–1777, 1995.
- [2] R. J. Elliott, *Stochastic Calculus and Applications*. New York: Springer-Verlag, 1982.
- [3] J. K. Hale, *Ordinary Differential Equations*, 2nd ed. Malabar, FL: R.E. Krieger, 1980.
- [4] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed. New York: Academic, 1975.
- [5] H. J. Kushner, "On the differential equations satisfied by conditional probability densities of Markov processes with applications," *SIAM J. Control*, vol. 2, pp. 106–119, 1964.
- [6] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, 2nd ed. New York: Springer-Verlag, 2003.
- [7] R. S. Liptser and A. N. Shiriyayev, *Statistics of Random Processes I & II*. New York: Springer-Verlag, 2001.
- [8] W. P. Malcome, R. J. Elliott, and J. van der Hoek, "On the numerical stability of time-discretized state estimation via Clark transformations," in *Proc. 42nd IEEE Conf. Decision Control*, 2003, pp. 1406–1412.
- [9] W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," *SIAM J. Control*, vol. 2, pp. 347–369, 1965.
- [10] G. Yin and S. Dey, "Weak convergence of hybrid filtering problems involving nearly completely decomposable hidden Markov chains," *SIAM J. Control Optim.*, vol. 41, pp. 1820–1842, 2003.
- [11] G. Yin, Q. Zhang, and Y. J. Liu, "Discrete-time approximation of Wonham filters," *J. Control Theory Appl.*, vol. 2, pp. 1–10, 2004.
- [12] O. Zeitouni and A. Dembo, "Exact filters for the estimation of the number of transitions of finite-state continuous-time Markov processes," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 890–893, Jul. 1988.