MULTISCALE MODEL. SIMUL.
Vol. 12, No. 3, pp. 1193–1229
© 2014 Society for Industrial and Applied Mathematics

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

FILTERING THE MAXIMUM LIKELIHOOD FOR MULTISCALE PROBLEMS*

ANDREW PAPANICOLAOU† AND KONSTANTINOS SPILIOPOULOS‡

Abstract. Filtering and parameter estimation under partial information for multiscale diffusion problems are studied in this paper. The nonlinear filter converges in the mean-square sense to a filter of reduced dimension. Based on this result, we establish that the conditional (on the observations) log-likelihood process has a correction term given by a type of central limit theorem. We prove that an appropriate normalization of the log-likelihood minus a log-likelihood of reduced dimension converges weakly to a normal distribution. In order to achieve this we assume that the operator of the (hidden) fast process has a discrete spectrum and an orthonormal basis of eigenfunctions. We then propose to estimate the unknown model parameters using the reduced log-likelihood, which is beneficial because reduced dimension means that there is significantly less runtime for this optimization program. We also establish consistency and asymptotic normality of the maximum likelihood estimator. Simulation results illustrate our theoretical findings.

Key words. ergodic filtering, fast mean reversion, homogenization, Zakai equation, maximum likelihood estimation, central limit theory

AMS subject classifications. 93E10, 93E11, 93C70

DOI. 10.1137/140952648

1. Introduction.

In this paper we consider the problem of filtering and parameter estimation for stochastic differential equations (SDEs) with multiple time scales. The model has parameter $0 < \delta \ll 1$ that separates the slow and fast scales of the system, and it is assumed that δ is known a priori. The filtering problem involves two SDEs: a hidden ergodic diffusion process X^δ, whose solution is known to be a path from an SDE with a fast time scale of $1/\delta$, and an observation Y^δ that depends on X^δ but evolves in a slow time scale that is of order 1. The parameter estimation problem arises when the SDE satisfied by (Y^δ, X^δ) has an unknown parameter $\theta \in \Theta$ where $\Theta \subset \mathbb{R}^d$.

Under the appropriate conditions, the nonlinear filter converges in a mean-square sense to a homogenized filter of reduced dimension. Based on this result and under the additional assumption that the infinitesimal generator of the fast process has a discrete spectrum with an orthonormal basis of eigenfunctions, we establish a central limit theorem (CLT) for the (conditional) log-likelihood. In particular, we prove that the difference of the log-likelihood (in other words, the log of the solution to the Zakai equation with input test function of $f \equiv 1$) minus a log-likelihood of reduced dimension, normalized by $\sqrt{\delta}$, converges weakly to a centered normal distribution with a variance that is a function of the model parameters. To the best of the authors' knowledge, the CLT proven in this paper is the first of its kind. We also establish consistency and asymptotic normality of the maximum likelihood estimator (MLE) of

*Received by the editors January 14, 2014; accepted for publication (in revised form) May 30, 2014; published electronically September 3, 2014.
http://www.siam.org/journals/mms/12-3/95264.html
†Department of ORFE, Princeton University, Sherrerd Hall, Charlton Street, Princeton, NJ 08544 (apapanic@princeton.edu). The work of this author was partially supported by NSF grant DMS-0739195.
‡Department of Mathematics & Statistics, Boston University, 111 Cummington Street, Boston, MA 02215 (kspiliop@math.bu.edu). The work of this author was partially supported by NSF grant DMS-1312124.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

the reduced log-likelihood. Compared to the original log-likelihood, the computation of the MLE based on the reduced log-likelihood is simpler and faster.

This work is related to other works in filtering, wherein the observed process evolves in a slower scale than the hidden process. In Kushner (1990), it is shown that the difference of the unnormalized actual filter and its homogenized counterpart goes to zero in distribution for fixed test functions. Bensoussan and Blankenship (1986) and Ichihara (2004) study homogenization of nonlinear filtering based on asymptotic analysis of a dual representation of the filtering equation. Park et al. (2008, 2011, 2010) and Imkeller et al. (2013) prove convergence in probability and in the *pth*-norm (in the latter article) of the nonlinear filter to its homogenized version. Notably, Imkeller et al. (2013) use a formulation through backward SDEs and make use of the estimates for the related transition probability densities of Pardoux and Veretennikov (2003); they also obtain rates of convergence in L^p. Kleptsina et al. (1997) prove convergence of the filter in the mean-square sense and in a quite general setting; they assume convergence of the total variation norm of (Y^δ, X^δ) and also assume convergence in probability of the slow part of the hidden component.

Parameter estimation problems for partially observed processes also have been studied elsewhere in the literature, e.g., Kutoyants (2004) and James and Gland (1995), although the effect of multiple scales was not studied there. Moreover, Papavasiliou et al. (2009) study maximum likelihood estimation for fully observed systems (not partially observed as in our case) of multiscale processes where the fast process takes values on a compact set.

The aforementioned existing literature has focused on proving convergence of the nonlinear filter to a filter of reduced dimension, namely to understand the dominant limiting behavior. In this paper, we are interested in parameter estimation for such models. Thus, for statistical inference purposes we need to prove that the filter will be close to a filter of reduced dimension for any parameter value (and not just for the true parameter value), with closeness referring to either convergence in probability or mean square under the measure parameterized by the true parameter value. We establish that this result is true in the L^2-sense and also show that convergence results in the existing literature can be extended to a class of unbounded test functions that have more than two moments. Then, we obtain a CLT for the difference between the log-likelihood function and the log-likelihood from the filter of reduced dimension. To obtain the CLT, we further assume that the infinitesimal generator of the fast process has a discrete spectrum with an orthonormal basis of eigenfunctions. The difference in the log-likelihood functions is of order $\sqrt{\delta}$, and we are able to state explicitly the variance of the limiting centered normal distribution. We emphasize that the filter of reduced dimension uses the original observations, which are the only available observations, and hence the results justify using the reduced log-likelihood for purposes of statistical inference. For computational purposes, it is simpler and much faster to implement the filter of reduced dimension than it is for the original log-likelihood.

Filtering is a well established area and some general references for stochastic nonlinear filtering are Bain and Crisan (2009), Kallianpur (1980), Kushner (1990), and Rozovskii (1990). Our motivation for studying parameter estimation for partially observed multiscale diffusion models comes from financial applications, e.g., convenience yield in commodities markets or estimation of latent states in markets with high frequency trading (HFT). For example, nonpredatory HFTs lead to increased liquidity and faster price discovery. Hence, a change-point detection algorithm on HFT data can be used to determine when price discovery has occurred. Another application

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

could be the detection of an increased bid-ask spread which may correspond to increased volatility. We refer the reader to Brogaard et al. (2012) and Zhang (2010) for related discussions.

The rest of the paper is organized as follows. Section 2 presents the system of equations that we consider, states our main assumptions, and restates fundamental results from filtering theory. Section 3 presents our results on the asymptotic properties of the filter and of the log-likelihood. In particular, in subsection 3.1 we discuss the L^2-convergence of the nonlinear filter, a result which is used in subsection 3.2 to establish the CLT for the log-likelihood; the CLT is the main result. These results are then used in section 4 to justify the claim that parameter estimation can be based on the reduced system, where we prove consistency and asymptotic normality of the MLE of the reduced log-likelihood. A simulation study illustrating the theoretical results is presented in section 5. Conclusions are in section 6. For presentation purposes, most of the proofs are deferred to Appendices A and B.

2. Formulation of problem and known preliminary results. On a probability space $(\Omega, (\mathcal{F}_t)_{t\le T}, \mathbb{P})$ with $T < \infty$, for positive integers m, n we consider the $(m+n)$-dimensional process $(X^\delta, Y^\delta) = \{(X_t^\delta, Y_t^\delta) \in \mathbb{R}^m \times \mathbb{R}^n, 0 \le t \le T\} \in \mathcal{C}([0,T]; \mathbb{R}^m \times \mathbb{R}^n)$, which satisfies a system of SDEs

$$\begin{aligned} dY_t^\delta &= h_\theta\left(X_t^\delta\right) dt + dW_t && \text{(observed)},\\ dX_t^\delta &= \frac{1}{\delta} b_\theta\left(X_t^\delta\right) dt + \frac{1}{\sqrt{\delta}} \sigma_\theta\left(X_t^\delta\right) dB_t && \text{(hidden)}, \end{aligned} \tag{2.1}$$

where $(W_t)_{t\le T}$ and $(B_t)_{t\le T}$ are (unobserved) independent Wiener processes in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Our general assumptions on the functions h_θ, b_θ, and σ_θ are given in subsection 2.1, but some of our theorems will require a stronger assumption on the spectrum of the infinitesimal generator of the X-process given in subsection 2.2. We assume that the parameter θ is also unknown, but takes values in a set $\Theta \subset \mathbb{R}^d$, with d being a positive integer. Initially, the process X_0^δ is distributed according to a given prior distribution, and from here forward we take $Y_0 = 0$. We denote the probability measure with $\mathbb{P}$, but we work with the parameterized family $(\mathbb{P}_\theta)_{\theta\in\Theta}$ in order to denote probabilities that are conditional on the parameter value,

$$\mathbb{P}_\theta((X^\delta, Y^\delta) \in \mathcal{B}) = \mathbb{P}\left((X^\delta, Y^\delta) \in \mathcal{B} \middle| \, \theta \text{ is parameter in (2.1)}\right) \qquad \forall \theta \in \Theta,$$

for any Borel set $\mathcal{B} \subset C([0,T], \mathbb{R}^m \times \mathbb{R}^n)$, and we let $\mathbb{E}_\theta$ denote its expectation operator. The parameter value to be estimated is the true (but unknown) value of θ; we denote the true value by $\alpha \in \Theta$.

Our goal for this paper is to develop a theoretical framework allowing statistical inference on the unknown parameter θ given an observed path $(Y_s^\delta)_{s\le t}$ and assuming that $0 < \delta \ll 1$. In particular, our goal in this paper is twofold:

(i) Obtain the limiting behavior and a CLT type correction for the posterior (on the observed path $(Y_s^\delta)_{s\le t}$) likelihood function as $\delta \downarrow 0$.

(ii) Use the asymptotic behavior of the likelihood function to develop a framework for statistical inference for the unknown parameter θ given an observed path $(Y_s^\delta)_{s\le t}$, assuming that $0 < \delta \ll 1$.

In subsection 2.1 we establish notation and conditions guaranteeing ergodicity and that the filtering problem is well posed. Then, in subsection 2.2 we introduce a more specific framework wherein the infinitesimal generator of the fast process X^δ with $\delta = 1$ has a discrete spectrum with an orthonormal basis of eigenfunctions, which

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

allows us to establish the CLT of Theorem 3.3. Then, in subsection 2.3 we review some known, useful results from filtering theory.

2.1. Notation and general assumptions. Let a, b be two vectors in some Euclidean space, say $\mathbb{R}^n$. For notational convenience we shall often write $a \cdot b$ or simply ab for their inner product, and we will denote by $|\cdot|$ the standard Euclidean norm.

Moreover, we denote by $\mathcal{X} = \mathbb{R}^m$ the state space of the fast component X. For any $f \in \mathcal{C}^2(\mathcal{X})$, we define the set of operators $(\mathcal{L}_\theta)_{\theta\in\Theta}$ such that

$$\mathcal{L}_\theta f(x) = b_\theta(x) \cdot D_x f(x) + \frac{1}{2}\mathrm{tr}\left[\sigma_\theta(x)\sigma_\theta^T(x) D_x^2 f(x)\right], \tag{2.2}$$

where D_x is the gradient operator. From (2.1) it follows that $\frac{1}{\delta}\mathcal{L}_\theta$ is the infinitesimal generator of X_t^δ.

We will make several assumptions on the growth and smoothness of the coefficients in order to guarantee that (2.1) has a well-defined strong solution, that the fast component X_t^δ is ergodic, that the slow component Y_t^δ has a well-defined homogenization limit as $\delta \downarrow 0$ in the appropriate sense, and that the filtering equations make sense. A set of assumptions guaranteeing these properties is contained in the following condition (see Pardoux and Veretennikov (2003) for ergodic theory, where they consider parts (i)–(iv) given below, and also Chapter 3 of Bain and Crisan (2009) for filtering).

Condition 2.1.

(i) *In order to guarantee the existence of an invariant measure $\mu_\theta(dx)$ for X^1 (i.e., for the process X_t^δ with $\delta = 1$) we assume that*

$$\lim_{|x|\to\infty} \sup_{\theta\in\Theta} b_\theta(x) \cdot x = -\infty.$$

(ii) *To guarantee uniqueness of the invariant measure for X^1, we assume that $\sigma_\theta(x)\sigma_\theta^T(x)$ is uniformly nondegenerate in θ; i.e., there exist constants $c(\theta) > 0$ such that for all $x \in \mathcal{X}$,*

$$|\xi\sigma_\theta(x)|^2 \geq c(\theta)|\xi|^2 \quad \forall\ (\theta, \xi) \in \Theta \times \mathbb{R}^n \text{ and } \forall x \in \mathbb{R}^n.$$

(iii) *$\sigma_\theta(x)\sigma_\theta^T(x)$ is bounded in $(\theta, x) \in \Theta \times \mathcal{X}$ and $\sigma_\theta(x)$ is globally Lipschitz in $x \in \mathcal{X}$ uniformly in $\theta \in \Theta$.*

(iv) *$b_\theta(x)$ is locally bounded and globally Lipschitz in $x \in \mathcal{X}$, uniformly in $\theta \in \Theta$.*

(v) *$h_\theta \in C(\mathcal{X})$ is locally bounded and globally Lipschitz in $x \in \mathcal{X}$, uniformly in $\theta \in \Theta$.*

(vi) *$X_0^\delta = X_0$ is a continuous random variable such that $\mathbb{E}|X_0|^3 < \infty$.*

(vii) *The functions $h_\theta, b_\theta, \sigma_\theta$ are Lipschitz continuous in $\theta \in \Theta$, and $\Theta \subset \mathbb{R}^d$ is compact.*

Remark 1. A typical example of a process X that satisfies Condition 2.1 is the Ornstein–Uhlenbeck process of Example 2.1 that we present below. One can verify that our results also hold for certain degenerate processes, such as the square root Cox–Ingersol–Ross (CIR) process of Example 2.2, where $\sigma(x) = \sqrt{x}$, i.e., it degenerates at $x = 0$ but nevertheless is ergodic; we do not analyze these special cases in this paper.

For any function $f \in L^2(\mathcal{X}, \mu_\theta)$, denote its average with respect to the invariant measure $\mu_\theta(dx)$ as

$$\bar{f}_\theta = \int_{\mathcal{X}} f(x)\mu_\theta(dx).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

It is a well-known result that $Y^\delta_\cdot$ converges in distribution in $C([0,T];\mathbb{R}^n)$ to the process $\overline{Y}_\cdot$ (see, e.g., Bensoussan et al. (1978), Pardoux and Veretennikov (2003)), where

$$\overline{Y}_t = \bar{h}_\theta t + W_t. \tag{2.3}$$

Actually, due to the fact that the observation process Y^δ_t has constant diffusion, Condition 2.1 and the ergodic theorem guarantee that a stronger result holds for any $\theta \in \Theta$; i.e., for every $\varepsilon > 0$,

$$\mathbb{P}_\theta\left(\sup_{0\le t\le T}\left|Y^\delta_t - \overline{Y}_t\right| \ge \varepsilon\right) \to 0 \text{ as } \delta \downarrow 0 \qquad \forall\ \theta \in \Theta. \tag{2.4}$$

2.2. Spectral decomposition. A stronger assumption than Condition 2.1 is that the operator $\mathcal{L}_\theta$ has a discrete spectrum with an orthonormal basis of eigenfunctions. Some of the theorems in this paper do not require such strong assumptions on the operator's spectrum (e.g., Theorems 3.1, 4.1, and 4.2 do not rely on discrete spectrum and orthonormal eigenfunctions), but the proof of the CLT in Theorem 3.3 relies on $\mathcal{L}_\theta$'s spectrum having these properties.

The steps taken in proving Theorem 3.3 utilize the spectral expansion of functions $f \in L^2(\mathcal{X}, \mu_\theta)$ with respect to the eigenfunctions of the operator $\mathcal{L}_\theta$. We say that the class of operators $\{\mathcal{L}_\theta\}_{\theta\in\Theta}$ has a discrete spectrum if for each $\theta \in \Theta$ there are eigenvalues $(-\lambda^\theta_i)_{i=0,1,2,3,\ldots}$ such that

$$0 = \lambda^\theta_0 > -\lambda^\theta_1 \ge -\lambda^\theta_2 \ge \cdots .$$

For each $i \ge 0$ we denote the ith eigenfunction as $\psi^\theta_i(x)$ such that

$$\mathcal{L}_\theta \psi^\theta_i = -\lambda^\theta_i \psi^\theta_i,$$

and we assume for each $\theta \in \Theta$ that the eigenfunctions form an orthonormal basis of $L^2(\mathcal{X}, \mu_\theta)$ so that

$$\int \psi^\theta_i(x)\psi^\theta_j(x)\mu_\theta(dx) = 1_{[i=j]},$$

and any square-integrable function $f \in L^2(\mathcal{X}, \mu_\theta)$ can be written as

$$f(x) = \sum_{i=0}^{\infty} \psi^\theta_i(x)\left\langle f, \psi^\theta_i\right\rangle_\theta,$$

where $\langle f, \psi^\theta_i\rangle_\theta = \int f(x')\psi^\theta_i(x')\mu_\theta(dx')$. Notice that because $\mathcal{L}_\theta$ is a differential operator and the spectral elements are assumed to be an orthonormal basis, we get that $\psi^\theta_0 \equiv 1$. This means

$$\left\langle \psi^\theta_i, 1\right\rangle_\theta = \left\langle \psi^\theta_i, \psi^\theta_0\right\rangle_\theta = 0 \qquad \text{for } i = 1, 2, 3, \ldots . \tag{2.5}$$

Below we consider some examples of processes whose operators have discrete spectrum with an orthonormal basis of eigenfunctions.

Example 2.1. A nondegenerate ergodic process with a discrete spectrum is the 1-dimensional OU process,

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{2}dB_t,$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $\theta \in \Theta \subset \mathbb{R}$ and $\sigma, \kappa > 0$. The eigenvalues of $\mathcal{L}_\theta$ are $0, -1, -2, -3, \ldots$, and the Hermite polynomials form an orthonormal basis. Moreover, this process is ergodic with invariant measure Gaussian and in particular $\mu_\theta(dx) = \sqrt{\frac{\kappa}{2\pi\sigma^2}} e^{-\frac{\kappa(x-\theta)^2}{2\sigma^2}} dx$.

Example 2.2. A degenerate ergodic process with a discrete spectrum is the 1-dimensional CIR process,

$$dX_t = \kappa(\theta - X_t)dt + \sqrt{2\sigma^2 X_t} dB_t,$$

where $\theta \in \Theta \subset \mathbb{R}^+$ and $\kappa > 0$. The eigenvalues of $\mathcal{L}_\theta$ are $0, -1, -2, -3, \ldots$, and the (generalized) Laguerre polynomials form an orthonormal basis. Moreover, if $\kappa\theta > \sigma^2$, then this process is ergodic with invariant measure, the measure for a gamma distribution, and in particular $\mu_\theta(dx) = \frac{a^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-ax} dx$, where $\Gamma(\cdot)$ is the gamma function, $a = \kappa/\sigma^2$, and $\beta = \kappa\theta/\sigma^2$. Even though this SDE does not satisfy Condition 2.1(ii)–(iii), the SDE has a unique strong solution which is ergodic, and thus one expects the results of this paper to hold.

We conclude with a multidimensional example.

Example 2.3. A nondegenerate ergodic process with a discrete spectrum is the m-dimensional linear SDE,

$$dX_t = -AX_t dt + \Gamma dB_t,$$

where A is $m \times m$ positive definite and Γ is a matrix of appropriate dimensions such that (A, Γ) is a controllable pair. This process is ergodic and its infinitesimal generator has discrete spectrum. The orthonormal basis can be constructed by taking products of the modified Hermite functions for each variable; see Liberzon and Brockett (2000) and Linetsky (2007) for more details and analysis.

2.3. Filtering equations. Our data are contained in the filtration generated by the observed path, which is the σ-algebra $\mathcal{Y}_t^\delta \doteq \mathcal{F}_t^{Y^\delta} = \sigma\{(Y_s^\delta)_{s \leq t}\}$. The filtration $\mathcal{Y}_t^\delta$ does not reveal the true but unknown parameter value $\alpha \in \Theta$. However, we can compute a posterior distribution conditional on a given parameter value, and then perform further statistical inference such as maximum likelihood in order to estimate the true parameter value. For a general introduction to stochastic filtering we refer the reader to classical texts such as Bain and Crisan (2009), Kallianpur (1980), Kushner (1990), and Rozovskii (1990).

For any $\theta \in \Theta$ (and not just the true parameter value, $\alpha \in \Theta$, that has generated the data in $\mathcal{Y}_t^\delta$), let us define the exponential martingale $Z_T^{\delta,\theta}$, which gives a new measure $\mathbb{P}_\theta^*$ on $(\Omega, \mathcal{F})$ such that

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}_\theta^*} \doteq Z_T^{\delta,\theta} = \exp\left\{\int_0^T h_\theta(X_s^\delta) dY_s^\delta - \frac{1}{2}\int_0^T \left|h_\theta(X_s^\delta)\right|^2 ds\right\}. \tag{2.6}$$

By Girsanov's theorem on the absolutely continuous change of measure in the space of trajectories in $\mathcal{C}([0,T], \mathbb{R}^m)$, the probability measures $\mathbb{P}_\theta$ and $\mathbb{P}_\theta^*$ are absolutely continuous with respect to each other, and the distribution of X^δ is the same under both $\mathbb{P}_\theta$ and $\mathbb{P}_\theta^*$. Furthermore, the process Y^δ is a $\mathbb{P}_\theta^*$-Brownian motion independent of X^δ, and $Z^{\delta,\theta}$ is a $\mathbb{P}_\theta^*$-martingale.

Next, for $f : \mathcal{X} \to \mathbb{R}$ such that $\mathbb{E}_\theta^* |f(X_t^\delta)|^2 < \infty$, we define the measure valued process $\phi_t^{\delta,\theta}$ acting on f as

$$\phi_t^{\delta,\theta}[f] \doteq \mathbb{E}_\theta^* \left[Z_t^{\delta,\theta} f(X_t^\delta) \middle| \mathcal{Y}_t^\delta \right], \tag{2.7}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

a process which, for $f \in C_c^2(\mathcal{X})$, is well known to be the unique solution (see Rozovsky (1991)) to the following equation:

$$d\phi_t^{\delta,\theta}[f] = \frac{1}{\delta}\phi_t^{\delta,\theta}[\mathcal{L}_\theta f]dt + \phi_t^{\delta,\theta}[h_\theta f]dY_t^\delta, \quad \mathbb{P}_\theta^*\text{-a.s.}, \quad \phi_0^\theta[f] = \mathbb{E}_\theta f(X_0^\delta). \tag{2.8}$$

Equation (2.8) is the Zakai equation for nonlinear filtering. In the literature, the term "filter" refers to a posterior measure on X_t^δ given $\mathcal{Y}_t^\delta$, and so $\phi_t^{\delta,\theta}$ is also a filter. Specifically, the process $\phi_t^{\delta,\theta}$ is an unnormalized probability measure with $\phi_t^{\delta,\theta}[1]$ being the likelihood function, and the maximizer of $\phi_t^{\delta,\theta}[1]$ is the MLE. In other words, given the observation $(Y_s^\delta)_{s\le t}$, the MLE is

$$\theta_t^\delta \doteq \arg\max_{\theta\in\Theta} \phi_t^{\delta,\theta}[1] \ . \tag{2.9}$$

Furthermore, we can apply the Kallianpur–Striebel formula to obtain the normalized filter,

$$\pi_t^{\delta,\theta}[f] \doteq \mathbb{E}_\theta\left[f(X_t^\delta)\middle|\mathcal{Y}_t^\delta\right] = \frac{\phi_t^{\delta,\theta}[f]}{\phi_t^{\delta,\theta}[1]} \quad \mathbb{P}_\theta, \mathbb{P}_\theta^*\text{-a.s.} \ . \tag{2.10}$$

An important case is $f(x) = x$ because X_t^δ is often tracked with the posterior mean, $\widehat{X}_t^{\delta,\theta} \doteq \mathbb{E}_\theta[X_t^\delta|\mathcal{Y}_t^\delta]$. The posterior mean can be given by the Kalman filter when σ_θ does not depend on x and there is linearity in x for both h_θ and b_θ. Another important case is $f(x) = h_\theta(x)$ because of the *innovations process*,

$$\nu_t^{\delta,\theta} \doteq Y_t^\delta - \int_0^t \pi_s^{\delta,\theta}[h_\theta]ds \quad \forall t \in [0,T]$$

(recall we assumed that $Y_0 = 0$). The process $\nu_t^{\delta,\theta}$ is a $\mathbb{P}_\theta$-Brownian motion under the filtration $\mathcal{Y}_t^\delta$, but it will only be observable as Brownian motion if $\theta = \alpha$, i.e., when the true parameter value is taken. For suitable test functions $f : \mathcal{X} \to \mathbb{R}$, the innovation is used in the nonlinear Kushner–Stratonovich equation to describe the evolution of $\pi_t^{\delta,\theta}[f]$,

$$d\pi_t^{\delta,\theta}[f] = \frac{1}{\delta}\pi_t^{\delta,\theta}[\mathcal{L}_\theta f]dt + \left(\pi_t^{\delta,\theta}[fh_\theta] - \pi_t^{\delta,\theta}[f]\pi_t^{\delta,\theta}[h_\theta]\right)d\nu_t^{\delta,\theta} \quad \mathbb{P}_\theta\text{-a.s.} \ . \tag{2.11}$$

The innovations Brownian motion will be used in later sections where we consider asymptotics of the log-likelihood function.

3. Asymptotic results of the filter and of the likelihood function. In this section we establish some results on the filter's convergence. In subsection 3.1 we use the convergence results found in Imkeller et al. (2013) (see also Park et al. (2008, 2011, 2010), Imkeller et al. (2013)) to prove convergence in probability of the filter for a class of unbounded test functions (e.g., for the eigenfunctions of the operator $\mathcal{L}_\theta$). Then, in subsection 3.2 we will use these results to derive a CLT for the log-likelihood function, which is the main result of the paper.

Consider the "averaged" exponentials

$$\bar{Z}_t^{\delta,\theta} \doteq \exp\left\{\bar{h}_\theta Y_t^\delta - \frac{1}{2}\left|\bar{h}_\theta\right|^2 t\right\}, \quad \bar{Z}_t^\theta \doteq \exp\left\{\bar{h}_\theta \overline{Y}_t - \frac{1}{2}\left|\bar{h}_\theta\right|^2 t\right\}. \tag{3.1}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

In fact the solution to the Zakai equation of (2.8) is close in the mean-square sense to a limiting filter based on $\bar{Z}_T^{\delta,\theta}$. For $f \in \mathcal{C}_c^2(\mathcal{X})$, we define new posterior measures $\bar{\phi}_t^{\delta,\theta}[f]$ and $\bar{\phi}_t^{\theta}[f]$, which satisfy the stochastic evolution equations

$$d\bar{\phi}_t^{\delta,\theta}[f] = \frac{1}{\delta}\bar{\phi}_t^{\delta,\theta}[\mathcal{L}_\theta f]dt + \bar{\phi}_t^{\delta,\theta}[f]\bar{h}_\theta dY_t^\delta, \quad \bar{\phi}_0^{\delta,\theta}[f] = \mathbb{E}_\theta\{f(X_0^\delta)\}, \tag{3.2}$$

$$d\bar{\phi}_t^{\theta}[f] = \bar{\phi}_t^{\theta}[f]\bar{h}_\theta d\overline{Y}_t\,, \quad \bar{\phi}_0^{\delta,\theta}[f] = \bar{f}_\theta. \tag{3.3}$$

It is straightforward to verify with Itô's lemma that the "average" Zakai equations (3.2) and (3.3) have solutions

$$\bar{\phi}_t^{\delta,\theta}[f] = \mathbb{E}_\theta^*\left[f(X_t^\delta)\bar{Z}_t^{\delta,\theta}\middle|\mathcal{Y}_t^\delta\right] = \mathbb{E}_\theta[f(X_t^\delta)]\bar{Z}_t^{\delta,\theta}\ , \tag{3.4}$$

$$\bar{\phi}_t^{\theta}[f] = \bar{f}_\theta\bar{Z}_t^{\theta}. \tag{3.5}$$

We also define $\bar{\pi}_t^{\delta,\theta}[f] = \frac{\bar{\phi}_t^{\delta,\theta}[f]}{\bar{\phi}_t^{\delta,\theta}[1]} = \mathbb{E}_\theta f(X_t^\delta)$ and $\bar{\pi}_t^{\theta}[f] = \frac{\bar{\phi}_t^{\theta}[f]}{\bar{\phi}_t^{\theta}[1]} = \bar{f}_\theta$.

Remark 2. The results of this section (namely Theorems 3.1 and 3.3 and Corollaries 3.2 and 3.4) will justify the approximation of $\phi^{\delta,\theta}[1]$ by $\bar{\phi}^{\delta,\theta}[1]$ for statistical inference purposes. Notice that $\bar{\phi}^{\delta,\theta}[1]$ is associated with the actual data; i.e., it is associated with Y_t^δ and not with $\overline{Y}_t$. $\overline{Y}_t$ is only used as a vehicle to obtain the necessary convergence results. Issues related to statistical inference are explored in section 4.

3.1. Convergence of the filter and of the likelihood function. At this point we need to impose an additional assumption on $Z_t^{\delta,\theta}$. In particular, we assume the following.

CONDITION 3.1. *For any $\theta \in \Theta$, there is a $q \in (1,\infty)$ such that*

$$\sup_{t\in[0,T]}\sup_{\delta\in(0,1)}\mathbb{E}_\theta^*|Z_t^{\delta,\theta}|^q + \sup_{t\in[0,T]}\sup_{\delta\in(0,1)}\mathbb{E}_\theta|Z_t^{\delta,\theta}|^{-q} < \infty.$$

Let us consider the $q \in (1,\infty)$ from Condition 3.1, and let $p \in (1,\infty)$ be such that $1/q + 1/p = 1$. Now let $\eta > 2(p^2-1)$, and define the following class of test functions:

$$\mathcal{A}_\eta^\theta \doteq \left\{ f \in C^4(\mathcal{X}) \cap L^2(\mathcal{X},\mu_\theta) : \sup_{t\in[0,T]}\sup_{\delta\in(0,1)}\mathbb{E}_\theta\left|f(X_t^\delta)\right|^{2+\eta} < \infty \right\}. \tag{3.6}$$

Before stating the convergence results, we make some remarks related to Condition 3.1 and the set $\mathcal{A}_\eta^\theta$.

Remark 3. Notice that because δ is a time scale, we could have written the definition in (3.6) with only a supremum over $t \geq 0$, and it would be an equivalent definition. That is, X_t^δ equals in distribution to $X_{t/\delta}^1$, so $\sup_{t\in[0,T]}\sup_{\delta\in(0,1)}\mathbb{E}_\theta|f(X_t^\delta)|^{2+\eta} = \sup_{t\in[0,T]}\sup_{\delta\in(0,1)}\mathbb{E}_\theta|f(X_{t/\delta}^1)|^{2+\eta} = \sup_{t\geq 0}\mathbb{E}_\theta|f(X_t^1)|^{2+\eta}$.

Remark 4. Condition 3.1 holds automatically for any finite $q > 1$ if $h_\theta(x)$ is bounded; see, e.g., Lemma 6.7 in Imkeller et al. (2013). Moreover, any $f \in \mathcal{C}_b^4(\mathcal{X})$ will also satisfy $f \in \mathcal{A}_\eta^\theta$ for any $\eta \geq 0$.

Remark 5. Suppose X_0 is distributed according to its invariant distribution. Then $\mathcal{A}_\eta^\theta$ consists of all functions $f \in C^4(\mathcal{X})$ such that $\int |f(x)|^{2+\eta}\mu_\theta(x)dx < \infty$. However, the orthonormal basis of eigenfunctions $(\psi_i^\theta)_{i=0}^\infty$ associated with the operator $\mathcal{L}_\theta$ (as described in section 2.2) are not generally contained in $\mathcal{A}_\eta^\theta$ if $\eta > 0$, but the examples

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

given earlier qualify. Examples 2.1, 2.2, and 2.3 also have $\psi_i^\theta \in \mathcal{A}_\eta^\theta$ for $\eta > 0$, because ψ_i^θ are polynomials with moments of all order, and so there are certainly $2+\eta$ moments of $\psi_i^\theta(X_t^\delta)$.

The first result of this section holds without the assumption of spectral expansions, and is stated in the following theorem.

THEOREM 3.1. *Assume Conditions* 2.1 *and* 3.1. *For any $\alpha, \theta \in \Theta$, we have that, uniformly in $t \in [0, T]$, the following are true:*

(i) *Let $f \in C_b^4(\mathcal{X})$. Then, for every $\varepsilon > 0$*

$$\lim_{\delta\downarrow 0} \mathbb{P}_\alpha \left(\left| \phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f] \right| \geq \varepsilon \right) = 0.$$

(ii) *Assume that there is $\eta > 0$ such that $f \in \mathcal{A}_\eta^\theta$. Then, we have convergence of the filters in mean square*

$$\lim_{\delta\downarrow 0} \mathbb{E}_\alpha \left| \pi_t^{\delta,\theta}[f] - \bar{\pi}_t^{\delta,\theta}[f] \right|^2 = 0,$$

and, moreover,

$$\lim_{\delta\downarrow 0} \left| \bar{\pi}_t^{\delta,\theta}[f] - \bar{\pi}_t^{\theta}[f] \right| = 0 \qquad \textit{in } \mathbb{P}_\alpha\textit{-probability.}$$

Proof. The proof of this theorem is in Appendix A. ∎

In statistical inference, a useful corollary of Theorem 3.1 is the convergence of likelihood functions.

COROLLARY 3.2. *Assume Conditions* 2.1 *and* 3.1. *For any $\alpha, \theta \in \Theta$ and each $t \geq 0$, we have*

$$\phi_t^{\delta,\theta}[1] - \bar{\phi}_t^{\delta,\theta}[1] \to 0 \qquad \textit{in } \mathbb{P}_\alpha\textit{-probability as } \delta \to 0.$$

We note that results similar to Theorem 3.1 appear elsewhere in the literature, e.g., Kleptsina et al. (1997), Ichihara (2004), Park et al. (2008, 2011, 2010), Imkeller et al. (2013), but with slightly different assumptions and setup. The main difference is that Theorem 3.1, when compared to the previous works, states the convergence result under the measure parameterized by the true parameter value (i.e., the measure under which the observations are made, where $\theta = \alpha$) with the filters converging for *any* parameter value. In other words, we will "observe" the filters converging to the reduced filter. Moreover, the convergence of the filters in Theorem 3.1 is for test functions that belong to the space $\mathcal{A}_\eta^\theta$, which can include unbounded functions such as the eigenfunctions of the OU processes in Examples 2.1 and 2.3 (see Remark 5). By assuming that $\psi_i^\theta \in \mathcal{A}_\eta^\theta$ for some $\eta > 0$, we are able to prove the results in subsection 3.2.

3.2. Asymptotic normality of likelihood function. We proceed to the statement and proof of the CLT for the log-likelihood function. In particular, we find that the difference in the original log-likelihood minus the log-likelihood of reduced dimension, divided by $\sqrt{\delta}$, yields a quantity that is asymptotically normal. In proving the CLT, we make extensive use of the discrete spectrum and eigenfunction basis. In this section we shall also assume the following.

CONDITION 3.2. *For any $i, j \in \mathbb{N}$ and any $\theta \in \Theta$, we assume the following:*

(i) *There exists $C_h > 0$ independent of θ such that $\|h_\theta\|_\infty \leq C_h$.*

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(ii) *$\mathcal{L}_\theta$ has discrete spectrum with orthonormal basis functions (as prescribed in section 2.2).*

(iii) *There exists $\eta > 0$ such that $\psi_i^\theta \in \mathcal{A}_\eta^\theta$ for all $\theta \in \Theta$ and $i \in \mathbb{N}$.*

(iv) *$\pi_0^\theta[\psi_i^\theta] < \infty$ for all $\theta \in \Theta$ and $i \in \mathbb{N}$.*

It is worth noting that Condition 3.2 subsumes Condition 3.1 because it places a bound on h_θ (see Remark 4). Moreover, the assumption that $\mathcal{L}_\theta$ has discrete spectrum with orthonormal basis functions is useful because the Zakai equation for the eigenfunctions ψ_i^θ simplifies to

$$d\phi_t^{\delta,\theta}[\psi_i^\theta] = -\frac{\lambda_i^\theta}{\delta}\phi_t^{\delta,\theta}[\psi_i^\theta]dt + \phi_t^{\delta,\theta}[h_\theta\psi_i^\theta]dY_t^\delta. \tag{3.7}$$

Applying Itô's lemma to $\frac{\phi_t^{\delta,\theta}[\psi_i^\theta]}{\phi_t^{\delta,\theta}[1]}$ we have the Kushner–Stratonovich equation

$$d\left(\frac{\phi_t^{\delta,\theta}[\psi_i^\theta]}{\phi_t^{\delta,\theta}[1]}\right) = -\frac{\lambda_i^\theta}{\delta}\frac{\phi_t^{\delta,\theta}[\psi_i^\theta]}{\phi_t^{\delta,\theta}[1]}dt + \underbrace{\left(\frac{\phi_t^{\delta,\theta}[h_\theta\psi_i^\theta]}{\phi_t^{\delta,\theta}[1]} - \frac{\phi_t^{\delta,\theta}[h_\theta]\phi_t^{\delta,\theta}[\psi_i^\theta]}{\left(\phi_t^{\delta,\theta}[1]\right)^2}\right)}_{=cov^{\delta,\theta}\left(h_\theta(X_t^\delta),\psi_i^\theta(X_t^\delta)|\mathcal{Y}_t^\delta\right)} d\nu_t^{\delta,\theta}, \tag{3.8}$$

where $d\nu_t^{\delta,\theta} = dY_t^\delta - \frac{\phi_t^{\delta,\theta}[h_\theta]}{\phi_t^{\delta,\theta}[1]}dt = dY_t^\delta - \mathbb{E}^{\delta,\theta}[h_\theta(X_t^\delta)|\mathcal{Y}_t^\delta]dt$ is the innovations Brownian motion under $\mathbb{P}_\theta$. By Duhamel's principle the solution is

$$\frac{\phi_t^{\delta,\theta}[\psi_i^\theta]}{\phi_t^{\delta,\theta}[1]} = e^{-\frac{\lambda_i^\theta t}{\delta}}\phi_0^\theta[\psi_i^\theta] + \int_0^t e^{-\frac{\lambda_i^\theta(t-s)}{\delta}} cov^{\delta,\theta}\left(h_\theta(X_s^\delta), \psi_i^\theta(X_s^\delta)\middle|\mathcal{Y}_s^\delta\right)d\nu_s^{\delta,\theta}. \tag{3.9}$$

Equivalently, we can write

$$\pi_t^{\delta,\theta}[\psi_i^\theta] = e^{-\frac{\lambda_i^\theta t}{\delta}}\pi_0^\theta[\psi_i^\theta] + \int_0^t e^{-\frac{\lambda_i^\theta(t-s)}{\delta}}\left(\pi_s^{\delta,\theta}[h_\theta\psi_i^\theta] - \pi_s^{\delta,\theta}[h_\theta]\pi_s^{\delta,\theta}[\psi_i^\theta]\right)d\nu_s^{\delta,\theta}. \tag{3.10}$$

Equations (3.9) and (3.10) are the key identities used to prove the CLT. However, there are some ergodic properties of the filter that are required to do the proof. Appendix B has these results; subsection 3.2.1 states and proves the CLT.

3.2.1. Statement of CLT and proof. In this section, we quantify the estimation error which occurs if the reduced log-likelihood is used in place of the full version. In particular, we establish that the error in the log-likelihood function will be normally distributed with standard deviation of order $O(\sqrt{\delta})$.

By Lemma 3.9 in Bain and Crisan (2009) we have

$$\log\left(\phi_t^{\delta,\theta}[1]\right) = \int_0^t \pi_s^{\delta,\theta}[h_\theta]dY_s^\delta - \frac{1}{2}\int_0^t \left|\pi_s^{\delta,\theta}[h_\theta]\right|^2 ds.$$

Let us write $\tilde{h}_\theta(x) = h_\theta(x) - \bar{h}_\theta$ and note that $\langle\tilde{h}_\theta, 1\rangle_\theta = 0$. Then we write

$$\begin{aligned}
&\frac{1}{\sqrt{\delta}}\left(\log\left(\phi_t^{\delta,\theta}[1]\right) - \log\left(\bar{\phi}_t^{\delta,\theta}[1]\right)\right)\\
&= \frac{1}{\sqrt{\delta}}\left(\int_0^t \left(\pi_s^{\delta,\theta}[h_\theta] - \bar{h}_\theta\right)dY_s^\delta - \frac{1}{2}\left(\int_0^t \left|\pi_s^{\delta,\theta}[h_\theta]\right|^2 ds - \int_0^t \left|\bar{h}_\theta\right|^2 ds\right)\right)\\
&= J_1^\delta + J_2^\delta,
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where we have defined J_1^δ and J_2^δ as

$$J_1^\delta = \frac{1}{\sqrt{\delta}}\int_0^t \left(\pi_s^{\delta,\theta}[h_\theta] - \bar{h}_\theta\right) dY_s^\delta$$
$$= \frac{1}{\sqrt{\delta}}\int_0^t \pi_s^{\delta,\theta}[\tilde{h}_\theta] d\nu_s^{\delta,\theta} + \frac{1}{\sqrt{\delta}}\int_0^t \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds + \frac{\bar{h}_\theta}{\sqrt{\delta}}\int_0^t \pi_s^{\delta,\theta}[\tilde{h}_\theta] ds$$

and

$$J_2^\delta = -\frac{1}{2\sqrt{\delta}}\left(\int_0^t \left|\pi_s^{\delta,\theta}[h_\theta]\right|^2 ds - \int_0^t \left|\bar{h}_\theta\right|^2 ds\right)$$
$$= -\frac{1}{2\sqrt{\delta}}\int_0^t \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds - \frac{\bar{h}_\theta}{\sqrt{\delta}}\int_0^t \pi_s^{\delta,\theta}[\tilde{h}_\theta] ds\ .$$

Hence, we obtain the representation

$$\begin{aligned}(3.11)\quad &\frac{1}{\sqrt{\delta}}\left(\log\left(\phi_t^{\delta,\theta}[1]\right) - \log\left(\bar{\phi}_t^{\delta,\theta}[1]\right)\right)\\ &= \underbrace{\int_0^t \frac{1}{\sqrt{\delta}}\left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) d\nu_s^{\delta,\alpha}}_{(*)} + \underbrace{\int_0^t \frac{1}{\sqrt{\delta}}\left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right)\left(\pi_s^{\delta,\theta}[h_\theta] - \pi_s^{\delta,\alpha}[h_\alpha]\right) ds}_{(**)}\\ &\qquad + \underbrace{\frac{1}{2\sqrt{\delta}}\int_0^t \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds}_{(\dagger)},\end{aligned}$$

where $\nu_t^{\delta,\alpha}$ is a $\mathbb{P}_\alpha$-Brownian motion (i.e., it is Brownian motion under the true parameter), but not for $\mathbb{P}_\theta$ with $\theta \neq \alpha$. Now recall that by Condition 3.2, for every $i \in \mathbb{N}$ we have $\psi_i^\theta \in \mathcal{A}_\eta^\theta$. This implies that there exist finite constants that may depend on i, T, and θ such that

$$\sup_{\delta\in(0,1),\rho\in[0,T]} \mathbb{E}_\theta\left[|\psi_i^\theta(X_\rho^\delta)|^2\right] \leq C(\psi_i, T, \theta), \tag{3.12}$$

from which we define another constant

$$C_{i,j}(T,\theta) \doteq \left(\frac{\left|\pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]\right|}{\lambda_i^\theta + \lambda_j^\theta} + \left(C(\psi_i,T,\theta) + C(\psi_j,T,\theta)\right)\left(\frac{1}{\lambda_i^\theta} + \frac{1}{\lambda_j^\theta}\right)\right). \tag{3.13}$$

If the infinite sum of these constants converges, then we can prove the following CLT for the log-likelihood function.

THEOREM 3.3 (likelihood CLT). *Assume Conditions* 2.1 *and* 3.2. *Moreover, assume that there exists constants* $C(\psi_i, T, \theta)$ *that satisfy* (3.12) *such that for all* $\theta \in \Theta$,

$$\sum_{i,j=1}^\infty |\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta | C_{i,j}(T,\theta) < \infty,$$

where $C_{i,j}(T,\theta)$ *is given by* (3.13). *Denote* $u_\theta^2(h_\theta) \doteq \sum_{i,j=1}^\infty \frac{\langle h_\theta,\psi_i^\theta\rangle_\theta \langle h_\theta,\psi_j^\theta\rangle_\theta \pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]}{\lambda_i^\theta+\lambda_j^\theta} < \infty$ *and* $v_\theta^2(h_\theta) \doteq \sum_{i,j=1}^\infty \frac{|\langle h_\theta,\psi_i^\theta\rangle_\theta \langle h_\theta,\psi_j^\theta\rangle_\theta|^2}{\lambda_i^\theta+\lambda_j^\theta}$, *with* $v_\theta^2(h_\theta) < \infty$ *by Parseval's identity (see*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Remark 6*). Then, and under* $\mathbb{P}_\alpha$ *and for any fixed* $t \in (0,T]$*, we have*

$$\frac{1}{\sqrt{\delta}}\left(\log\left(\phi_t^{\delta,\theta}[1]\right) - \log\left(\bar{\phi}_t^{\delta,\theta}[1]\right)\right) \Rightarrow \mathcal{W}\left(u_\theta^2(h_\theta) + tv_\theta^2(h_\theta)\right) \quad \text{as } \delta \to 0$$

in distribution, where $\mathcal{W}\left(u_\theta^2(h_\theta) + tv_\theta^2(h_\theta)\right)$ *is a normal random variable with mean zero and variance* $u_\theta^2(h_\theta) + tv_\theta^2(h_\theta)$*.*

If X_0 starts in its invariant distribution, then $C(\psi_i, T, \theta) = 1$ and $\pi_0^\theta[\psi_i^\theta] = 0$ for all $i \geq 0$ and $C_{i,j}(T,\theta) = \frac{2}{\lambda_i^\theta} + \frac{2}{\lambda_j^\theta}$ for all $i, j \geq 0$, and we have the following corollary from Theorem 3.3.

Corollary 3.4 (likelihood CLT for paths). *Assume Conditions* 2.1 *and* 3.2*. Moreover, assume that for all* $\theta \in \Theta$*,*

$$\sum_{i,j=1}^{\infty} |\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta| \left(\frac{1}{\lambda_i^\theta} + \frac{1}{\lambda_j^\theta}\right) < \infty.$$

If X_0 *is distributed according to the invariant measure* μ_θ *(i.e.,* $\pi_0^\theta[f] = \bar{f}_\theta$ *for all* $f \in \mathcal{A}_\eta^\theta$ *and any* $\theta \in \Theta$*), then under* $\mathbb{P}_\alpha$ *we have*

$$\frac{1}{\sqrt{\delta}}\left(\log\left(\phi_\cdot^{\delta,\theta}[1]\right) - \log\left(\bar{\phi}_\cdot^{\delta,\theta}[1]\right)\right) \Rightarrow \sqrt{v_\theta^2(h_\theta)}\,\mathcal{W}(\cdot) \quad \text{as } \delta \to 0$$

in distribution on $C([0,T],\mathbb{R})$*, where* $\mathcal{W}$ *is a Brownian motion and* $v_\theta^2(h_\theta)$ *is as defined in Theorem* 3.3*.*

Before continuing with the proofs of Theorem 3.3 and Corollary 3.4, we make some remarks related to the conditions that appear in the statement of the CLT.

Remark 6. The orthonormal basis of eigenfunctions that was assumed in Condition 3.2 is enough to ensure that the variance $v_\theta^2(h_\theta) < \infty$. Indeed, by Parseval's identity we have

$$\begin{aligned} v_\theta^2(h_\theta) &= \sum_{i,j=1}^{\infty} \frac{\left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right|^2}{\lambda_i^\theta + \lambda_j^\theta} \leq \frac{1}{2\lambda_1^\theta} \sum_{i,j=1}^{\infty} \left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right|^2 \\ &= \frac{1}{2\lambda_1^\theta}\left(\sum_{i=1}^{\infty} \left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta\right|^2\right)^2 = \frac{1}{2\lambda_1^\theta}\left(\int_{\mathcal{X}} |h_\theta(x)|^2 \mu_\theta(dx)\right)^2 \\ &= \frac{\|h_\theta\|_{L^2(\mathcal{X},\mu_\theta)}^4}{2\lambda_1^\theta} < \frac{C_h^4}{2\lambda_1^\theta} < \infty, \end{aligned}$$

where C_h is the constant from Condition 3.2. Finiteness of $u_\theta^2(h_\theta)$ follows from (3.20) in the proof of Theorem 3.3.

Remark 7 (absolutely summable h_θ). The function $h_\theta(x)$ is said to be an *absolutely summable* function if

$$\sum_{i=0}^{\infty} |\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta| < \infty.$$

Absolute summability is sufficient for Corollary 3.4 to hold. Indeed, notice that

$$\sum_{i,j=1}^{\infty} |\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta| \left(\frac{1}{\lambda_i^\theta} + \frac{1}{\lambda_j^\theta}\right) \leq \frac{2}{\lambda_1^\theta}\left(\sum_{i=1}^{\infty} |\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta|\right)^2.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A similar treatment applies to the more general summability constraint that appears in Theorem 3.3. For more on functions whose eigencoefficients decay fast enough to ensure absolute convergence, see the conditions/examples given in Boyd (2000, 1984).

Remark 8 (converging initial distributions). Corollary 3.4 could be generalized to the case where the initial distribution depends on δ and converges to the invariant distribution. That is, assuming a priori the limit

$$\pi_0^{\delta,\theta}[f] \to \bar{f}_\theta \quad \text{as } \delta \to 0 \quad \forall\, \theta \in \Theta, \quad \forall f \in \mathcal{A}_\eta^\theta\ ,$$

one expects that the same pathwise limit remains as stated in the corollary. However, generalization of the proofs in this paper will require verification that the initial filters $\pi_0^{\delta,\theta}$ satisfy (3.12) and allow for the limit to pass into the sum in (3.20).

Remark 9. We could also combine Theorem 3.3 and Corollary 3.4 by writing

$$\frac{1}{\sqrt{\delta}}\left(\log\left(\phi_\cdot^{\delta,\theta}[1]\right) - \log\left(\bar{\phi}_\cdot^{\delta,\theta}[1]\right) - \sqrt{\delta} R_\cdot^{1,\delta}\right) \Rightarrow \sqrt{v_\theta^2(h_\theta)}\mathcal{W}(\cdot) \quad \text{as } \delta \to 0$$

in $C([0,T],\mathbb{R})$, where $R_t^{1,\delta}$ is given by (3.16) in the proof of Theorem 3.3.

Proof of Theorem 3.3. The proof of this theorem involves showing that ($\dagger$) and ($**$) from (3.11) converge to zero in probability uniformly in $t \in [0,T]$, and then showing that ($*$) converges weakly to the appropriate normal distribution. Then, the result follows by Slutzky's theorem (see Billingsley (1968)).

First we consider the term ($\dagger$). By Lemma B.3 we have that there exists a constant $C < \infty$ such that

$$\sup_{\delta\in(0,1)} \mathbb{E}_\alpha \sup_{t\in[0,T]} \int_0^t \left[\frac{1}{\delta}\left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2\right] ds \le \sup_{\delta\in(0,1)} \mathbb{E}_\alpha \int_0^T \left[\frac{1}{\delta}\left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2\right] ds < C.$$

Therefore, the conclusion

$$\lim_{\delta\downarrow 0} \mathbb{E}_\alpha \frac{1}{2\sqrt{\delta}} \sup_{t\in[0,T]} \int_0^t \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds = 0$$

follows, implying the claimed convergence of the term ($\dagger$) in $\mathbb{P}_\alpha$-probability, uniformly in $t \in [0,T]$. Convergence to zero in $\mathbb{P}_\alpha$-probability of the ($**$) term follows by Lemma B.5.

Now we turn our attention toward ($*$) and define the integrated process

$$I_t^\delta \doteq \int_0^t \frac{1}{\sqrt{\delta}}\left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) d\nu_s^{\delta,\alpha}\ ,$$

which is a $\mathbb{P}_\alpha$-martingale. Since h_θ is bounded, we clearly have that $\tilde{h}_\theta \in L^2(\mathcal{X},\mu_\theta)$, and hence we have the representation

$$\tilde{h}_\theta(x) = \sum_{i=0}^\infty \left\langle \tilde{h}_\theta, \psi_i^\theta\right\rangle_\theta \psi_i^\theta(x) = \sum_{i=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \psi_i^\theta(x).$$

Thus, we get

$$\frac{1}{\sqrt{\delta}}\pi_s^{\delta,\theta}[\tilde{h}_\theta] = \sum_{i=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \frac{1}{\sqrt{\delta}}\pi_s^{\delta,\theta}[\psi_i^\theta]. \tag{3.14}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

From this and (3.10), it follows that

$$
\begin{aligned}
\frac{1}{\sqrt{\delta}}\pi_s^{\delta,\theta}[\tilde{h}_\theta] &= \frac{1}{\sqrt{\delta}}\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta e^{-\frac{\lambda_i^\theta s}{\delta}}\pi_0^\theta[\psi_i^\theta] \\
&+ \frac{1}{\sqrt{\delta}}\int_0^s\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}} \\
&\times\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right)d\nu_\rho^{\delta,\theta} \\
&+\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\frac{\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta}{\sqrt{\delta}}\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[h_\theta]-\pi_\rho^{\delta,\alpha}[h_\alpha]\right)d\rho \\
&+\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\frac{\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta}{\sqrt{\delta}}\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}d\nu_\rho^{\delta,\alpha}.
\end{aligned}
\tag{3.15}
$$

Hence, we have

$$
\begin{aligned}
I_t^\delta &= \sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\pi_0^\theta[\psi_i^\theta]\frac{1}{\sqrt{\delta}}\int_0^t e^{-\frac{\lambda_i^\theta s}{\delta}}d\nu_s^{\delta,\alpha} \\
&+\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\frac{1}{\sqrt{\delta}} \\
&\times\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right)d\nu_\rho^{\delta,\theta}\right]d\nu_s^{\delta,\alpha} \\
&+\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\frac{\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta}{\sqrt{\delta}}\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[h_\theta]-\pi_\rho^{\delta,\alpha}[h_\alpha]\right)d\rho\right]d\nu_s^{\delta,\alpha} \\
&+\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\frac{\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta}{\sqrt{\delta}}\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}d\nu_\rho^{\delta,\alpha}\right]d\nu_s^{\delta,\alpha} \\
&= R_t^{1,\delta}+R_t^{2,\delta}+R_t^{3,\delta}+R_t^{4,\delta},
\end{aligned}
\tag{3.16}
$$

where $R_t^{j,\delta}$ for $j=1,2,3,4$ are defined by the four lines in (3.16). We treat each of the $R_t^{j,\delta}$ terms separately. By Lemmas B.6 and B.7, we have that

$$
\lim_{\delta\downarrow 0}\left\{\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|R_t^{2,\delta}\right|^2+\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|R_t^{3,\delta}\right|^2\right\}=0.
$$

Thus, we have established that uniformly in $t\in[0,T]$,

$$
I_t^\delta-(R_t^{1,\delta}+R_t^{4,\delta})\to 0,\ \text{in } \mathbb{P}_\alpha\text{-probability as } \delta\downarrow 0.
\tag{3.17}
$$

It remains to treat the first and the last term, i.e., the term $R_t^{1,\delta}$ and the term $R_t^{4,\delta}$. Recall that

$$
\begin{aligned}
R_t^{1,\delta}+R_t^{4,\delta} = \int_0^t\Bigg(&\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\frac{1}{\sqrt{\delta}}e^{-\frac{\lambda_i^\theta s}{\delta}}\pi_0^\theta[\psi_i^\theta] \\
&+\sum_{i=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta^2\frac{1}{\sqrt{\delta}}\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}d\nu_\rho^{\delta,\alpha}\Bigg)d\nu_s^{\delta,\alpha}.
\end{aligned}
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The solution to the linear SDE

$$\Xi_t^{\delta,i} = -\frac{\lambda_i^\theta}{\delta}\int_0^t \Xi_s^{\delta,i} ds + \frac{1}{\sqrt{\delta}}\nu_t^{\delta,\alpha} \tag{3.18}$$

is simply

$$\Xi_t^{\delta,i} = \frac{1}{\sqrt{\delta}}\int_0^t e^{-\frac{\lambda_i^\theta (t-s)}{\delta}} d\nu_s^{\delta,\alpha}. \tag{3.19}$$

So, by the martingale representation theorem, there is an appropriate Wiener process $\mathcal{W}$ such that we have in distribution (see Theorem 4.6 on page 174 of Karatzas and Shreve (1991))

$$\begin{aligned}
R_t^{1,\delta} + R_t^{4,\delta} &= \int_0^t \left(\sum_{i=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \frac{1}{\sqrt{\delta}} e^{-\frac{\lambda_i^\theta s}{\delta}} \pi_0^\theta[\psi_i^\theta] + \sum_{i=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta^2 \Xi_s^{\delta,i}\right) d\nu_s^{\delta,\alpha} \\
&= \mathcal{W}\left(\int_0^t \left(\sum_{i=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \frac{1}{\sqrt{\delta}} e^{-\frac{\lambda_i^\theta s}{\delta}} \pi_0^\theta[\psi_i^\theta] + \sum_{i=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta^2 \Xi_s^{\delta,i}\right)^2 ds\right) \\
&= \mathcal{W}\left(\int_0^t \left(\sum_{i,j=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta \frac{1}{\delta} e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)s}{\delta}} \pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]\right.\right. \\
&\quad + \sum_{i,j=1}^\infty \left|\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta\right|^2 \Xi_s^{\delta,i}\Xi_s^{\delta,j} \\
&\quad \left.\left. + 2\sum_{i,j=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left|\left\langle h_\theta, \psi_j^\theta\right\rangle_\theta\right|^2 \frac{1}{\sqrt{\delta}} e^{-\frac{\lambda_i^\theta s}{\delta}} \pi_0^\theta[\psi_i^\theta]\Xi_s^{\delta,j}\right) ds\right) \\
&= \mathcal{W}\left(\sum_{i,j=1}^\infty \frac{\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta}{\lambda_i^\theta + \lambda_j^\theta} \pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta] \left(1 - e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)t}{\delta}}\right)\right. \\
&\quad + \sum_{i,j=1}^\infty \left|\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta\right|^2 \int_0^t \Xi_s^{\delta,i}\Xi_s^{\delta,j} ds \\
&\quad \left. + 2\sum_{i,j=1}^\infty \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left|\left\langle h_\theta, \psi_j^\theta\right\rangle_\theta\right|^2 \pi_0^\theta[\psi_i^\theta] \frac{1}{\sqrt{\delta}} \int_0^t e^{-\frac{\lambda_i^\theta s}{\delta}} \Xi_s^{\delta,j} ds\right) \\
&= \mathcal{W}\left(J_t^{1,\delta} + J_t^{2,\delta} + J_t^{3,\delta}\right),
\end{aligned}$$

where $J_t^{\ell,\delta}$ is the ℓth term in the variance of $\mathcal{W}(\cdot)$. So in order to find where $R_t^{1,\delta}+R_t^{4,\delta}$ converges to in distribution, we need to find the limit in probability of $J_t^{1,\delta}+J_t^{2,\delta}+J_t^{3,\delta}$. For each fixed $t \in (0,T]$ we have

$$\begin{aligned}
\lim_{\delta\downarrow 0} J_t^{1,\delta} &= \lim_{\delta\downarrow 0} \sum_{i,j=1}^\infty \frac{\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta}{\lambda_i^\theta + \lambda_j^\theta} \pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta] \left(1 - e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)t}{\delta}}\right) \\
&= \sum_{i,j=1}^\infty \frac{\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta}{\lambda_i^\theta + \lambda_j^\theta} \pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta] = u_\theta^2(h_\theta).
\end{aligned} \tag{3.20}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For $J_t^{2,\delta}$ we use ergodicity of the pair $(\Xi_t^{\delta,i}, \Xi_t^{\delta,j})$. Clearly, for any $i \geq 1$, $\Xi_t^{\delta,i}$ is ergodic (it is a 1-dimensional OU process). Also, one can check the Fokker–Planck equation for the pair $(\Xi_t^{\delta,i}, \Xi_t^{\delta,j})$ to see that for $\lambda_i^\theta \neq \lambda_j^\theta$, $(\Xi_t^{\delta,i}, \Xi_t^{\delta,j}) =_d (\Xi_{t/\delta}^{1,i}, \Xi_{t/\delta}^{1,j})$ is jointly Gaussian and ergodic, and for every $t \in [0,T]$ converges as $\delta \downarrow 0$ in distribution to a pair of jointly Gaussian random variables (Z_i, Z_j) with mean zero and invertible covariance matrix. Thus, by the ergodic theorem we have for every $t \geq 0$,

$$\lim_{\delta\downarrow 0} \mathbb{E}\left|\int_0^t \Xi_s^{\delta,i}\Xi_s^{\delta,j} ds - t\beta^{i,j}\right| = \lim_{\delta\downarrow 0} \mathbb{E}\left|\delta\int_0^{t/\delta} \Xi_s^{i,1}\Xi_s^{j,1} ds - t\beta^{i,j}\right| = 0, \tag{3.21}$$

where $\beta^{i,j} = \mathbb{E}[Z_i Z_j] = \lim_{\delta\downarrow 0} \mathbb{E}[\Xi_t^{\delta,i}\Xi_t^{\delta,j}] = \frac{1}{\lambda_i^\theta + \lambda_j^\theta}$. Since by assumption we have $v_\theta^2(h_\theta) < \infty$, we have for every $t \geq 0$

$$\lim_{\delta\downarrow 0} J_t^{2,\delta} = t\ v_\theta^2(h_\theta) \quad \text{in probability as } \delta \downarrow 0. \tag{3.22}$$

For similar reasons, we also obtain that for every $t \geq 0$,

$$\lim_{\delta\downarrow 0} J_t^{3,\delta} = 0 \quad \text{in probability as } \delta \downarrow 0. \tag{3.23}$$

Hence, we get that for every fixed $t \in (0,T]$,

$$R_t^{1,\delta} + R_t^{4,\delta} \Rightarrow \mathcal{W}\left(u_\theta^2(h_\theta)\mathbb{1}_{[t>0]} + tv_\theta^2(h_\theta)\right) \tag{3.24}$$

as $\delta \to 0$, which then implies that for every fixed $t \in (0,T]$,

$$I_t^\delta \Rightarrow \mathcal{W}\left(u_\theta^2(h_\theta) + tv_\theta^2(h_\theta)\right) \tag{3.25}$$

as $\delta \to 0$. ∎

Proof of Corollary 3.4. The proof of the CLT for paths requires identifying the weak limit of $I_\cdot^\delta$, which we do using the martingale CLT stated in Theorem 1.4 on page 339 of Ethier and Kurtz (1986). In particular, the process I_t^δ is a martingale and takes values in the space $C([0,T];\mathbb{R})$ with probability one, so it follows that

$$\lim_{\delta\downarrow 0} \mathbb{E}_\alpha\left[\sup_{t\in[0,T]} \left|I_t^\delta - I_{t-}^\delta\right|\right] = 0. \tag{3.26}$$

Given (3.26), if the quadratic variation of I_t^δ converges to a constant multiple of t for each $t \in [0,T]$, then the martingale CLT says that $I_\cdot^\delta$ converges weakly to a Brownian motion multiplied by the limiting quadratic variation.

Convergence of the quadratic variation was shown in the proof of Theorem 3.3 by showing that terms $J_t^{1,\delta}$, $J_t^{2,\delta}$, and $J_t^{3,\delta}$ converge in probability as $\delta \to 0$. Indeed, if X_0 follows the invariant distribution, then $\pi_0^\theta[\psi_i^\theta] = 0$ and $C(\psi_i, T, \theta) = 1$ for all $i \in \mathbb{N}$. This means that

$$C_{i,j}(T,\theta) = \frac{2}{\lambda_i^\theta} + \frac{2}{\lambda_j^\theta}$$

and that the terms $J_t^{1,\delta} = J_t^{3,\delta} = 0$ in the proof of Theorem 3.3. Hence, the quadratic variation of I_t^δ converges to $tv_\theta(h_\theta)$ in $\mathbb{P}_\alpha$-probability for all $t \in [0,T]$, and so we get that in distribution

$$I_\cdot^\delta \Rightarrow \sqrt{v_\theta^2(h_\theta)}\mathcal{W}(\cdot) \qquad \text{under } \mathbb{P}_\alpha. \tag{3.27}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The remaining terms $(**)$ and $(\dagger)$ from (3.11) were shown in the proof of Theorem 3.3 to go to zero in $\mathbb{P}_\alpha$-probability uniformly for all $t \in [0,T]$, and therefore they also (both) converge pathwise to zero in probability. Hence, all three terms in (3.11) converge pathwise, two of which in probability to zero, and the other weakly to a $\sqrt{v_\theta^2(h_\theta)}\mathcal{W}(\cdot)$. Therefore, by Slutzky's theorem the sum of all three terms converges weakly to $\sqrt{v_\theta^2(h_\theta)}\mathcal{W}(\cdot)$. $\square$

4. On statistical inference.

In subsection 3.1, and in particular in Corollary 3.2, we proved that the likelihood function $\phi_T^{\delta,\theta}[1]$ is close in probability to the reduced likelihood $\bar{\phi}_T^{\delta,\theta}[1]$ when δ is small. In this section, we use these results to do statistical inference for the unknown true parameter $\alpha \in \Theta$ based on the MLE of the log-likelihood function.

Corollary 3.2 suggests that for parameter estimation, we can approximate the log-likelihood

$$\rho_T^\delta(\theta) = \log \phi_T^{\delta,\theta}[1] = \log \mathbb{E}_\theta^* \left[Z_T^{\delta,\theta} \middle| \mathcal{Y}_T^\delta \right] \tag{4.1}$$

by the "reduced" log-likelihood

$$\bar{\rho}_T^\delta(\theta) = \log \bar{\phi}_T^{\delta,\theta}[1] = \log \mathbb{E}_\theta^* \left[\bar{Z}_T^{\delta,\theta} \middle| \mathcal{Y}_T^\delta \right] = \bar{h}_\theta Y_T^\delta - \frac{1}{2}|\bar{h}_\theta|^2 T. \tag{4.2}$$

Clearly, $\bar{\rho}_T^\delta(\theta)$ is of reduced dimension and easier to work with, as long as one can compute or approximate the invariant measure of the fast dynamics and thus compute or approximate $\bar{h}_\theta$. Based on the full log-likelihood (4.1), one would need to compute $\rho_T^\delta(\theta)$ and thus rely on methods such as particle filters or sequential Monte Carlo (e.g., Chapter 9 of Bain and Crisan (2009)). However, such methods can be computationally expensive due to high-dimensionality issues.

With this mind, we prove that the MLE based on (4.2) is, in fact, under the appropriate identifiability condition, asymptotically consistent when the time horizon is large enough.

CONDITION 4.1.

(i) *The mapping $\bar{h}_\theta$ from $\Theta \mapsto \mathbb{R}^m$ is a one-to-one function of θ.*

(ii) *There are constants $C > 0$, $p \geq 1$, and $q > 1$ such that for any $\theta_1, \theta_2 \in \Theta$,*

$$|\bar{h}_{\theta_1} - \bar{h}_{\theta_2}|^{2p} \leq C|\theta_1 - \theta_2|^q.$$

Recall the definition of MLE from (2.9), and let us equivalently define the reduced estimator as

$$\bar{\theta}_T^\delta \doteq \arg\max_{\theta \in \Theta} \bar{\rho}_T^\delta(\theta). \tag{4.3}$$

Continuity of $\bar{\rho}_T^\delta(\cdot)$ that is ensured by Condition 4.1, together with compactness of Θ, implies that the corresponding maximizer exists almost surely.

Next, we prove consistency of the reduced log-likelihood.

THEOREM 4.1. *Assume Conditions* 2.1 *and* 4.1. *Let α be the true parameter value. Let us denote by $\bar{\Theta}_T^\delta$ the equivalence class of maximizers of $\bar{\rho}_T^\delta(\theta)$. The maximum likelihood estimator based on* (4.2)*, i.e., any $\bar{\theta}_T^\delta \in \bar{\Theta}_T^\delta$, is strongly consistent as first $\delta \downarrow 0$ and then $T \to \infty$, i.e., for any $\varepsilon > 0$*

$$\lim_{T\to\infty} \lim_{\delta \downarrow 0} \mathbb{P}_\alpha \left(\left| \bar{\theta}_T^\delta - \alpha \right| > \varepsilon \right) = 0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Let us denote

$$\bar{\rho}_T^\delta(\theta,\alpha) = \bar{h}_\theta \int_0^T h_\alpha(X_s^\delta)ds + \bar{h}_\theta W_T - \frac{1}{2}|\bar{h}_\theta|^2 T.$$

Then, we have

$$\begin{aligned}\mathbb{E}_\alpha \left|\bar{\rho}_T^\delta(\theta_1,\alpha) - \bar{\rho}_T^\delta(\theta_2,\alpha)\right|^{2p} &\le C|\bar{h}_{\theta_1} - \bar{h}_{\theta_2}|^{2p}\left(1 + \mathbb{E}_\alpha \int_0^T |h_\alpha(X_s^\delta)|^{2p} ds\right)\\ &\le C|\theta_1-\theta_2|^q,\end{aligned}$$

where we used Condition 4.1. The constant C might change from line to line, but we do not indicate this in the notation. Next, the ergodic theorem guarantees that the finite-dimensional distributions of $\bar{\rho}_T^\delta(\cdot,\alpha)$ converge with probability 1, as $\delta\downarrow 0$, to those of

$$\bar{\rho}_T(\theta,\alpha) = \bar{h}_\theta\bar{h}_\alpha T + \bar{h}_\theta W_T - \frac{1}{2}|\bar{h}_\theta|^2 T = -\frac{1}{2}|\bar{h}_\theta - \bar{h}_\alpha|^2 T + \frac{1}{2}|\bar{h}_\alpha|^2 T + \bar{h}_\theta W_T.$$

Therefore, by Theorem 12.3 in Billingsley (1968), we have weak convergence of the measure $\bar{\rho}_T^\delta(\cdot,\alpha)$ to that of $\bar{\rho}_T(\cdot,\alpha)$. Hence, we have obtained (in a similar manner to Theorem 2.25 on page 161 of Kutoyants (2004))

$$\begin{aligned}\lim_{\delta\downarrow 0}\mathbb{P}_\alpha\left(\left|\bar{\theta}_T^\delta - \alpha\right| > \varepsilon\right) &= \lim_{\delta\downarrow 0}\mathbb{P}_\alpha\left(\sup_{|\theta-\alpha|>\varepsilon}\frac{1}{T}\bar{\rho}_T^\delta(\theta,\alpha) > \sup_{|\theta-\alpha|\le\varepsilon}\frac{1}{T}\bar{\rho}_T^\delta(\theta,\alpha)\right)\\ &= \mathbb{P}_\alpha\left(\sup_{|\theta-\alpha|>\varepsilon}\frac{1}{T}\bar{\rho}_T(\theta,\alpha) > \sup_{|\theta-\alpha|\le\varepsilon}\frac{1}{T}\bar{\rho}_T(\theta,\alpha)\right).\end{aligned}$$

Hence, if we now define $\bar{\bar{\rho}}(\theta,\alpha) = -\frac{1}{2}|\bar{h}_\theta - \bar{h}_\alpha|^2 + \frac{1}{2}|\bar{h}_\alpha|^2$, we then get

$$\lim_{T\to\infty}\lim_{\delta\downarrow 0}\mathbb{P}_\alpha\left(\left|\bar{\theta}_T^\delta - \alpha\right| > \varepsilon\right) = \mathbb{1}_{\left[\sup_{|\theta-\alpha|>\varepsilon}\bar{\bar{\rho}}(\theta,\alpha) > \sup_{|\theta-\alpha|\le\varepsilon}\bar{\bar{\rho}}(\theta,\alpha)\right]} = 0,$$

where the last computation used the fact that $\bar{\bar{\rho}}(\theta,\alpha)$ has a unique maximum at $\theta=\alpha$, which follows from part (i) of Condition 4.1. With this, we conclude the proof of the theorem. ∎

Solving the equation $\frac{\partial}{\partial\theta}\bar{\rho}_T^\delta(\theta) = 0$ for $\theta\in\Theta$, we define $\tilde{\theta}_T^\delta$ to be the solution (if it exists) to

$$\bar{h}_\theta = \frac{1}{T}Y_T^\delta. \tag{4.4}$$

It is clear that (4.3) and (4.4) are not equivalent; (4.3) contains all local minima and local maxima of $\bar{\rho}_T^\delta(\theta)$, which may be more than one. Also (4.4) may not even have a solution in Θ with positive probability. For example, letting $\tilde{\theta}_T^\delta$ be a solution to (4.4) and assuming $\theta\in(\theta_\ell,\theta_u)$, then

$$\bar{\theta}_T^\delta = \tilde{\theta}_T^\delta\mathbb{1}_{\left[\{\tilde{\theta}_T^\delta\in(\theta_\ell,\theta_u)\}\right]} + \theta_\ell\mathbb{1}_{\left[\tilde{\theta}_T^\delta\le\theta_\ell\right]} + \theta_u\mathbb{1}_{\left[\tilde{\theta}_T^\delta\ge\theta_u\right]}.$$

By Theorem 4.1, and based on smoothness of $\bar{h}_\theta$ as a function of θ, asymptotic normality of the MLE corresponding to the reduced log-likelihood holds.

THEOREM 4.2. *Assume Conditions* 2.1 *and* 4.1, *that* $\dot{\bar{h}}_\theta \doteq \frac{\partial\bar{h}_\theta}{\partial\theta}$ *is continuous, and that for every* $\theta\in\mathbb{R}^d$ *the matrix* $\dot{\bar{h}}_\theta^*\dot{\bar{h}}_\theta$ *is positive definite. The maximum likelihood*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

estimator based on (4.2) *is asymptotically normal under* $\mathbb{P}_\alpha$, *i.e.,*

$$\sqrt{T}\left(\bar{\theta}^{\delta}_{T}-\alpha\right) \Rightarrow N\left(0, \left(\dot{\bar{h}}^{*}_{\alpha}\dot{\bar{h}}_{\alpha}\right)^{-1}\right) \quad \textit{first as } \delta \downarrow 0 \textit{ and then } T \to \infty. \tag{4.5}$$

Proof. The proof is similar to that of Proposition 1.34 of Kutoyants (2004), even though there are no multiscale effects there. Below, we present the proof, emphasizing the differences due to the multiscale aspect of the present problem. Based on (4.4) for $\theta = \bar{\theta}^{\delta}_{T}$ we write

$$\bar{h}_{\alpha} + \left(\bar{\theta}^{\delta}_{T}-\alpha\right)\dot{\bar{h}}_{\alpha^{*}} = \frac{1}{T}Y^{\delta}_{T},$$

where $|\alpha^{*}-\alpha| \leq |\bar{\theta}^{\delta}_{T}-\alpha|$. Rearranging the latter expression we get

$$\sqrt{T}\left(\bar{\theta}^{\delta}_{T}-\alpha\right) = \sqrt{T}\left[\frac{1}{T}Y^{\delta}_{T}-\bar{h}_{\alpha}\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1}.$$

Now under the measure $\mathbb{P}_\alpha$, we have that $Y^{\delta}_{T} = \int_0^T h_{\alpha}(X^{\delta}_{s})ds + W_T$. Hence, we can continue the latter expression as

$$\begin{aligned}\sqrt{T}\left(\bar{\theta}^{\delta}_{T}-\alpha\right) &= \sqrt{T}\left[\frac{1}{T}\int_0^T h_{\alpha}(X^{\delta}_{s})ds-\bar{h}_{\alpha}\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1} + \left[\frac{1}{\sqrt{T}}W_T\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1}\\ &= \sqrt{T}\left[\frac{1}{T/\delta}\int_0^{T/\delta} h_{\alpha}(X^{1}_{s})ds-\bar{h}_{\alpha}\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1} + \left[\frac{1}{\sqrt{T}}W_T\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1},\end{aligned} \tag{4.6}$$

where we also used that $X^{\delta}_{\cdot} = X^{1}_{\cdot/\delta}$ in distribution. By taking $\delta \downarrow 0$ we have by the L^1 ergodic theorem that

$$\lim_{\delta\downarrow 0}\mathbb{E}\left|\frac{1}{T/\delta}\int_0^{T/\delta} h_{\alpha}(X^{1}_{s})ds-\bar{h}_{\alpha}\right| = 0 \quad \text{for any } T \in (0,\infty).$$

Since $|\alpha^{*}-\alpha| \leq |\bar{\theta}^{\delta}_{T}-\alpha|$ we can apply the consistency of Theorem 4.1 to get

$$\lim_{T\to\infty}\lim_{\delta\downarrow 0}\mathbb{P}_{\alpha}\left(|\alpha^{*}-\alpha|>\varepsilon\right) \leq \lim_{T\to\infty}\lim_{\delta\downarrow 0}\mathbb{P}_{\alpha}\left(\left|\bar{\theta}^{\delta}_{T}-\alpha\right|>\varepsilon\right) = 0 \quad \text{for any } \varepsilon > 0,$$

and hence by continuity we have $\dot{\bar{h}}_{\alpha^{*}} \to \dot{\bar{h}}_{\alpha}$ in probability as $\delta \downarrow 0$ and then $T \to \infty$. Therefore, by the positive definiteness of $\dot{\bar{h}}_{\alpha}$ we have the limit

$$\sqrt{T}\left[\frac{1}{T/\delta}\int_0^{T/\delta} h_{\alpha}(X^{1}_{s})ds-\bar{h}_{\alpha}\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1} \to 0$$

in probability as $\delta \downarrow 0$ and then $T \to \infty$. For similar reasons, Slutsky's theorem implies

$$\left[\frac{1}{\sqrt{T}}W_T\right]\left(\dot{\bar{h}}_{\alpha^{*}}\right)^{-1} \Rightarrow N\left(0, \left(\dot{\bar{h}}^{*}_{\alpha}\dot{\bar{h}}_{\alpha}\right)^{-1}\right) \quad \text{first as } \delta \downarrow 0 \text{ and then } T \to \infty.$$

Finally, using Slutsky's theorem on the combined expression in (4.6) yields the statement of the theorem. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

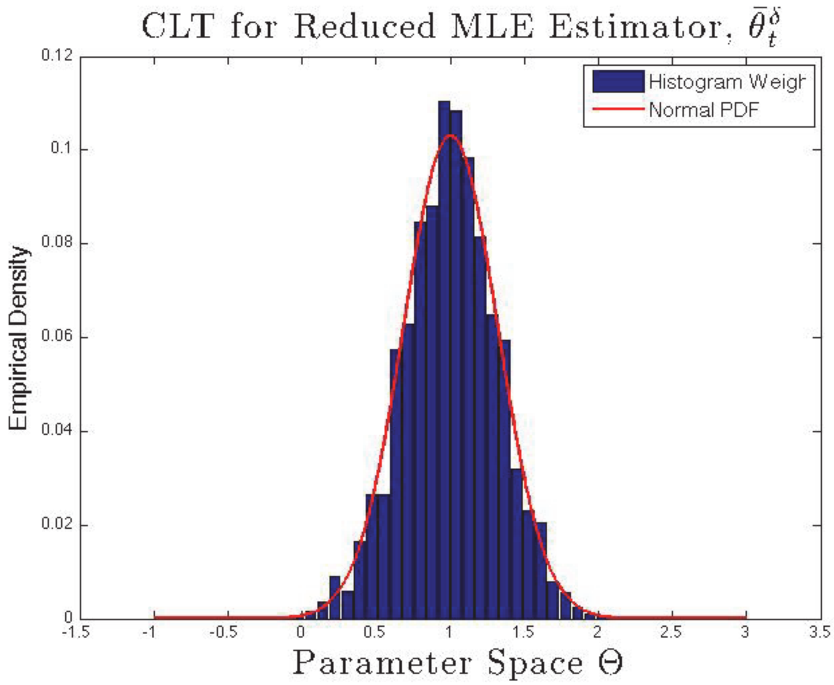

FIG. 1. *The empirical distribution of the reduced estimator $\bar{\theta}_t^\delta$, for which the asymptotic distribution is close to Gaussian. We run the system* 2,000 *times and each time compute $\bar{\theta}_t^\delta$. For these trials, the MLE has an empirical error of* 0.3180*, which is close to the* $\frac{1}{\sqrt{T}} = 0.3162$ *that is the standard error predicted by* (4.5) *in the CLT of Theorem* 4.2 *with* $\bar{h}_\theta = \theta + \frac{1}{2\sqrt{\pi}}$ *and* $\dot{\bar{h}}_\theta = 1$.

5. Simulation example. In this section, we present a simulation example, illustrating the theoretical findings. As an example, we consider the parameter space $\Theta \subset \mathbb{R}$ and take the true parameter value to be $\alpha = 1$. We consider the model

$$\begin{aligned} dX_t^\delta &= \frac{1}{\delta}\left(\theta - X_t^\delta\right) dt + \sqrt{\frac{2}{\delta}} dB_t, \\ dY_t^\delta &= \max(X_t^\delta, \theta) dt + dW_t \end{aligned} \tag{5.1}$$

for $t \leq T = 5$ and $\delta = .01$. By Theorem 2.9 of Karatzas and Shreve (1991), there exists a unique strong solution to the SDE for (X^δ, Y^δ). For the purposes of the numerical example, we assume that the initial distribution of X is its invariant law. If we run the system 2,000 times and each time compute $\bar{\theta}_t^\delta$, we get the histogram shown in Figure 1. For these trials, the MLE has an empirical error of 0.3180, which is close to the $\frac{1}{\sqrt{T}} = 0.3162$ that is the standard error predicted by (4.5) in the CLT of Theorem 4.2 with $\bar{h}_\theta = \theta + \frac{1}{2\sqrt{\pi}}$ and $\dot{\bar{h}}_\theta = 1$.

To show the effect of Theorem 3.3, we compare the full log-likelihood to the reduced log-likelihood. The generator of the OU process in (5.1) has a discrete set of eigenvalues such that $\lambda_i^\theta = -i$ for $i = 0, 1, 2, 3, \ldots$ for any $\theta \in \mathbb{R}$ and admits an orthonormal basis that is given (up to a normalizing constant) by the Hermite

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 1

The eigencoefficients for $h_\theta(x) = \max(x, \theta)$ using the (normalized) Hermite polynomials. The limiting variance as predicted by Theorem 3.3 is given in the last column, and is well-approximated by the first 15–20 basis elements. For this example, changes in θ only affect the first eigenmode.

θ	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$v^2(h_\theta)$
.5	0.8989	0.5000	0.2821	0	-0.0814	0	0.0446	.04723
1	1.3989	0.5000	0.2821	0	-0.0814	0	0.0446	.04723
1.5	1.8989	0.5000	0.2821	0	-0.0814	0	0.0446	.04723

polynomials

$$\psi_i^\theta(x) = \frac{1}{C_i} H_i(x-\theta) = \frac{(-1)^i}{C_i} e^{\frac{(x-\theta)^2}{2}} \frac{d^i}{dx^i} e^{-\frac{(x-\theta)^2}{2}},$$

where ψ is an eigenfunction as defined in subsection 2.2, H_i is the ith (probabilists') Hermite polynomial (see Abramowitz and Stegun (1965)), and $C_i \doteq \sqrt{i!}$ is a normalizing constant. The eigencoefficients of the function $h_\theta(x) = \max(x, \theta)$ are computed as follows:

$$\begin{aligned}
\langle h_\theta, \psi_i^\theta \rangle_\theta &= \frac{1}{\sqrt{2\pi} C_i} \int_{-\infty}^{\infty} \max(x,\theta) H_i(x-\theta) e^{-\frac{(x-\theta)^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi} C_i} \int_{-\infty}^{\infty} (\theta + \max(x-\theta, 0)) H_i(x-\theta) e^{-\frac{(x-\theta)^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi} C_i} \int_{-\infty}^{\infty} (\theta + \max(u, 0)) H_i(u) e^{-\frac{u^2}{2}} dx \\
&= \theta \cdot \mathbb{1}_{[i=0]} + \frac{1}{\sqrt{2\pi} C_i} \int_0^\infty u H_i(u) e^{-\frac{u^2}{2}} dx \ ,
\end{aligned}$$

and so only the zero order term depends on θ (the last computation used (2.5)). The eigencoefficients are given in Table 1. There is relatively fast decay among these coefficients, and hence the limiting variance function $v^2(\theta)$ from Theorem 3.3 can be well-approximated by the first 15–20 basis elements.

The simulations and the analysis that follow demonstrate two things:

- On one hand, $\rho_t^\delta(\theta)$ needs to be approximated based on methods such as Monte Carlo. As δ gets smaller one needs more samples in order to compute $\rho_t^\delta(\theta)$ accurately.
- On the other hand, the computation of $\bar\rho_t^\delta(\theta)$ is straightforward with no Monte Carlo errors. Theorem 3.3 quantifies the deviation of $\bar\rho_t^\delta(\theta)$ from $\rho_t^\delta(\theta)$.

To compute $\rho_t^\delta(\theta)$ we use sequential Monte Carlo (SMC). Namely, we take independent samples $(X^{\delta,\ell})_{\ell=1}^N$ for some $N < \infty$ where each $X^{\delta,\ell} =_d X^\delta$, and our full log-likelihood is approximated as

$$\rho_t^\delta(\theta) \approx \log\left(\frac{1}{N} \sum_{\ell=1}^N e^{-\frac{1}{2}\int_0^t h_\theta^2(X_s^{\delta,\ell}) ds + \int_0^t h_\theta(X_s^{\delta,\ell}) dY_s^\delta} \right).$$

Estimation using SMC samples will have an error that is of order $1/\sqrt{N}$, and with an asymptotically normal distribution (see Del Moral et al. (2001), Cappé et al. (2005))

$$\sqrt{N}\left(\log\left(\frac{1}{N} \sum_{\ell=1}^N e^{-\frac{1}{2}\int_0^t h_\theta^2(X_s^{\delta,\ell}) ds + \int_0^t h_\theta(X_s^{\delta,\ell}) dY_s^\delta} \right) - \rho_t^\delta(\theta) \right) \Rightarrow \mathcal{Z}(Y^\delta)$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

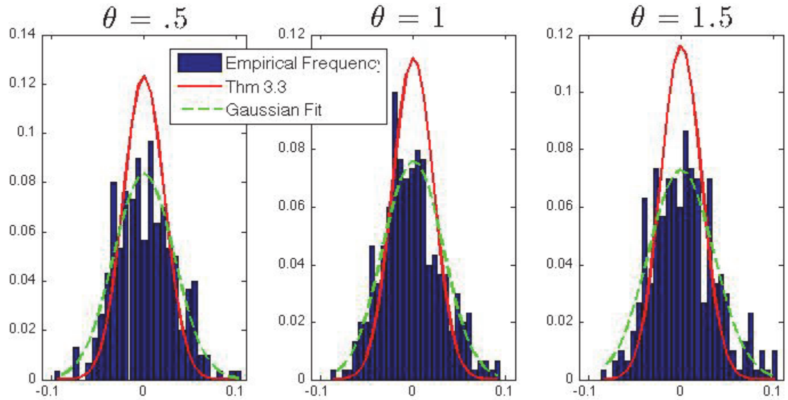

FIG. 2. *Histograms of the quantity* $\frac{1}{\sqrt{t}}\left(\rho_t^\delta(\theta) - \bar{\rho}_t^\delta(\theta)\right)$ *for* $\theta = .5, 1, 1.5$ *with the true parameter being* $\alpha = 1$*. The solid line is the limiting Gaussian distribution of Theorem* 3.3*, and the dashed line is a Gaussian fit to the histogram. The dashed line has a slightly greater standard deviation because* $\rho_t^\delta(\theta)$ *needs to be approximated with Monte Carlo sampling and a discrete time scheme, and hence the empirical distribution has some additional variance. However, the Kolmogorov–Smirnoff test does not reject the hypothesis that the histogram is a Gaussian distribution.*

as $N \to \infty$, where $\mathcal{Z}(Y^\delta)$ is a normal random variable whose variance depends on the data Y^δ.

In Figure 2 we see the histograms and fitted normal distributions obtained by looking at $\frac{1}{\sqrt{t}}\left(\rho_t^\delta(\theta) - \bar{\rho}_t^\delta(\theta)\right)$. The solid line is the density suggested by the CLT of Theorem 3.3, namely a normal density with mean zero and variance $\delta v^2(\theta)$, and the dashed line is a Gaussian density with mean zero and the empirical standard deviation. The Kolmogorov–Smirnoff test does not reject any of the empirical histogram fits to the dashed lines (at the 99.9% confidence level), and the test rejects the histogram fits to the solid lines for low confidence values and for different parameters. Heuristically, the difference in these standard errors should be $O(1/\sqrt{N})$:

$$\begin{aligned}\text{empirical standard error} &= \sqrt{var\left(\log\left(\frac{1}{N}\sum_{\ell=1}^{N} e^{-\frac{1}{2}\int_0^t h_\theta^2(X_s^{\delta,\ell})ds + \int_0^t h_\theta(X_s^{\delta,\ell})dY_s^\delta}\right) - \bar{\rho}_t^\delta(\theta)\right)}\\ &\leq \sqrt{var\left(\log\left(\frac{1}{N}\sum_{\ell=1}^{N} e^{-\frac{1}{2}\int_0^t h_\theta^2(X_s^{\delta,\ell})ds + \int_0^t h_\theta(X_s^{\delta,\ell})dY_s^\delta}\right) - \rho_t^\delta(\theta)\right)}\\ &\quad + \sqrt{var\left(\rho_t^\delta(\theta) - \bar{\rho}_t^\delta(\theta)\right)}\\ &\simeq O\left(\frac{1}{\sqrt{N}}\right) + \sqrt{\delta v^2(\theta)}.\end{aligned}$$

Indeed, from Table 2 we see that the difference between the standard error of the CLT of Theorem 3.3 and the empirical standard error is of order $1/\sqrt{N}$, which indicates the strong possibility that the aforementioned error due to approximation via SMC is significant when estimating the log-likelihood.

Figure 2 indicates the following: not only is the reduced estimate of the log-likelihood close to the full likelihood, but it also might be a better estimate than a Monte Carlo approximation of the full log-likelihood. The enlarged Monte Carlo error

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Table 2

For the model in (5.1), 300 *simulations of the quantity* $\frac{1}{\sqrt{t}}\left(\rho_t^\delta(\theta) - \bar{\rho}_t^\delta(\theta)\right)$ *computed with* $N =$ 2,000, *this table shows the standard error predicted by Theorem* 3.3, *the empirical standard error, and the difference between the two. It turns out that* $\frac{1}{\sqrt{N}} = \frac{1}{\sqrt{2{,}000}} = .0224$, *which is of the same order as the entries in the* 4*th column, and so we conclude that the dashed line in Figure* 2 *has extra variance that is due to the SMC sampling error.*

Statistics for simulations of $\frac{1}{\sqrt{t}}\left(\rho_t^\delta(\theta) - \bar{\rho}_t^\delta(\theta)\right)$ with $\delta = .01$.

θ	$\sqrt{\delta v^2(\theta)}$	Empirical std-err.	Empirical std-err. $- \sqrt{\delta v^2(\theta)}$
.5	.02174	.0346	.0128
1	.02174	.0322	.0105
1.5	.02174	.0354	.0137

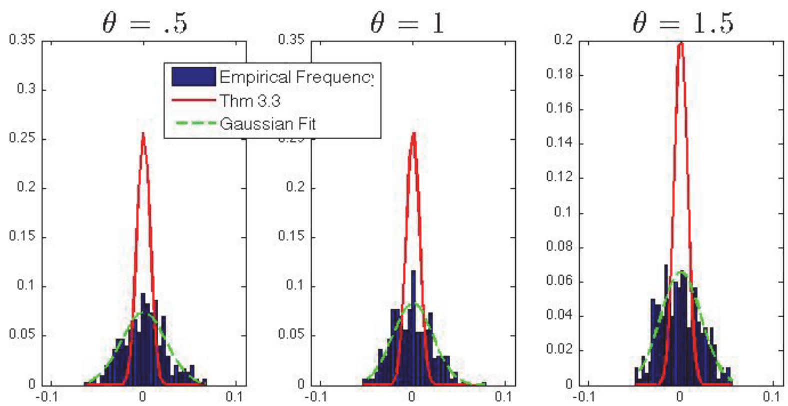

Fig. 3. *The same histograms as those in Figure* 2, *except for the experiment run with* $\delta = .001$. *Notice how the Monte Carlo error is a greater proportion of the total variance. This illustrates how for* δ *small, the reduced likelihood can be more accurate than a Monte Carlo approximation.*

in the computation of $\rho_t^\delta(\theta)$ can be seen in Figure 3, which is the same experiment, except with $\delta = .001$ (i.e., the same number of particles at $N = 2{,}000$). In Figure 3 it is important to notice how the Monte Carlo error is a significantly greater proportion of the total empirical error. If we want Figure 3 to look similar to Figure 2, then we would need to increase N by a factor of 10. Such an increase in the number of particles would significantly increase the computation time. Hence, the reduced filter outperforms the direct Monte Carlo filter for $\delta \ll 1$, which is a motivation for this paper.

6. Conclusions and future work. This paper studies parameter estimation with partially observed diffusions of models with multiple time scales. This problem is primarily an application of ergodic theory to nonlinear filtering. We prove convergence in probability of the nonlinear filter and of the conditional (on the observations) log-likelihood. Furthermore, we prove a CLT for the log-likelihood. These results justify the use of a log-likelihood of reduced dimension for the purposes of parameter estimation, which is simpler to implement and has faster runtime in computations. Consistency and asymptotic normality for the MLE of the reduced log-likelihood are also obtained, and simulation studies are presented to show how the reduced log-likelihood can outperform a direct Monte Carlo filter when $\delta \ll 1$.

It is plausible that some of the results presented in this paper can be generalized.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For instance, it is possible that the CLT can be proven with the removal of the assumption of h_θ being bounded, which is also supported by the simulation example of section 5. Regarding the generalization of Theorem 3.3, it may be possible to prove a version of the theorem using generalized spectral theory rather than assuming a discrete spectrum with orthonormal eigenfunctions, but modifications to the techniques developed in this paper will be needed.

Appendix A. Proof of Theorem 3.1. The proof of Theorem 3.1 follows by the results of Imkeller et al. (2013) (see also Park et al. (2008, 2011, 2010)) after we adjust for the parameter mismatch. In particular, the main difference that Theorem 3.1 has when compared to the previous works is that under the measure parameterized by the true parameter value (i.e., the measure under which the observations are made) the filters will converge for *any* parameter value. Moreover, we also need to prove that the convergence of the filters is for test functions in the space $\mathcal{A}_\eta^\theta$, whereas the results in Imkeller et al. (2013) use bounded and smooth test functions.

LEMMA A.1. *Let us consider* $f \in C_b^4(\mathcal{X})$ *and assume Conditions* 2.1 *and* 3.1. *For any* $\theta, \alpha \in \Theta$, *we have uniformly in* $t \in [0,T]$

$$\mathbb{E}_\alpha \left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right| \to 0 \quad \textit{as } \delta \to 0.$$

Proof. By Hölder's inequality, for $p, q < 0$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have

$$\begin{aligned}
\mathbb{E}_\alpha \left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right| &= \mathbb{E}_\alpha^* \left[Z_t^{\delta,\alpha}\left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right|\right] \\
&\le \left(\mathbb{E}_\alpha^* \left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}_\alpha^* \left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right|^p\right)^{1/p} \\
&= \left(\mathbb{E}_\alpha^* \left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}_\theta^* \left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right|^p\right)^{1/p} \\
&\le \left(\mathbb{E}_\alpha^* \left|Z_t^{\delta,\alpha}\right|^q\right)^{1/p}\left(\mathbb{E}_\theta \left|Z_t^{\delta,\theta}\right|^{-q}\right)^{1/(pq)}\left(\mathbb{E}_\theta \left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right|^{p^2}\right)^{1/p^2},
\end{aligned}$$

which goes to zero as $\delta \downarrow 0$ by Condition 3.1 and Lemma 6.6 in Imkeller et al. (2013). The third line, i.e., $\mathbb{E}_\alpha^*|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]|^p = \mathbb{E}_\theta^*|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]|^p$, follows because both $\phi_t^{\delta,\theta}$ and $\bar{\phi}_t^{\delta,\theta}$ are functionals of $Y_\cdot^\delta$ (and no other random variable), and Y^δ is a Brownian motion under both measures $\mathbb{P}_\alpha^*$ and $\mathbb{P}_\theta^*$. This concludes the proof of the lemma. ∎

We conclude with the proof of Theorem 3.1.

Proof of Theorem 3.1. Lemma A.1 implies convergence in probability:

$$\mathbb{P}_\alpha\left(\left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right| > \varepsilon\right) \le \frac{1}{\varepsilon}\mathbb{E}_\alpha\left|\phi_t^{\delta,\theta}[f] - \bar{\phi}_t^{\delta,\theta}[f]\right| \to 0 \quad \forall \varepsilon > 0$$

for bounded f. Let us now prove the second part of the theorem. We prove it first for $f \in C_b^4(\mathcal{X})$. Then, we prove it under the assumption that there exists $\eta > 0$ such that $f \in \mathcal{A}_\eta^\theta$. So, let us assume that $f \in C_b^4(\mathcal{X})$. It is clear that by ergodicity we have

$$\lim_{\delta\downarrow 0}\left|\bar{\pi}_t^{\delta,\theta}[f] - \bar{\pi}_t^\theta[f]\right| = 0 \qquad \text{in } \mathbb{P}_\alpha\text{-probability},$$

so it remains to prove that

$$\lim_{\delta\downarrow 0}\mathbb{E}_\alpha\left(\pi_t^{\delta,\theta}[f] - \bar{\pi}_t^{\delta,\theta}[f]\right)^2 = 0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

For this purpose, Hölder's inequality gives

$$\begin{aligned}\mathbb{E}_\alpha\left|\pi_t^{\delta,\theta}[f]-\bar{\pi}_t^{\delta,\theta}[f]\right|^2 &= \mathbb{E}^*_\alpha\left[Z_t^{\delta,\alpha}\left|\pi_t^{\delta,\theta}[f]-\bar{\pi}_t^{\delta,\theta}[f]\right|^2\right]\\ &\le \left(\mathbb{E}^*_\alpha\left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}^*_\alpha\left|\pi_t^{\delta,\theta}[f]-\bar{\pi}_t^{\delta,\theta}[f]\right|^{2p}\right)^{1/p}\\ &\le \left(\mathbb{E}^*_\alpha\left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}^*_\theta\left|\pi_t^{\delta,\theta}[f]-\bar{\pi}_t^{\delta,\theta}[f]\right|^{2p}\right)^{1/p}\\ &\le \left(\mathbb{E}^*_\alpha\left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}_\theta\left|Z_t^{\delta,\theta}\right|^{-q}\right)^{1/(pq)}\left(\mathbb{E}_\theta\left|\pi_t^{\delta,\theta}[f]-\bar{\pi}_t^{\delta,\theta}[f]\right|^{2p^2}\right)^{1/p^2},\end{aligned}$$

which goes to zero as $\delta\downarrow 0$ by Condition 3.1 and Corollary 6.9 in Imkeller et al. (2013). The third line, i.e., $\mathbb{E}^*_\alpha|\phi_t^{\delta,\theta}[f]-\bar{\phi}_t^{\delta,\theta}[f]|^{2p}=\mathbb{E}^*_\theta|\phi_t^{\delta,\theta}[f]-\bar{\phi}_t^{\delta,\theta}[f]|^{2p}$, follows because both $\phi_t^{\delta,\theta}$ and $\bar{\phi}_t^{\delta,\theta}$ are functionals of $Y^\delta_\cdot$ (and no other random variable), and Y^δ is a Brownian motion under both measures $\mathbb{P}^*_\alpha$ and $\mathbb{P}^*_\theta$. This completes the proof for $f\in C_b^4(\mathcal{X})$.

Let us complete the proof of the theorem by assuming that there exists an $\eta>0$ such that $f\in\mathcal{A}^\theta_\eta$. For $n\in\mathbb{N}$, define

$$u_n(x)=\begin{cases}x, & |x|\le n,\\ n\ \mathrm{sign}(x), & |x|>n\end{cases}$$

and set $f_n(x)=u_n(f(x))$. Analogously define

$$\pi_t^{\delta,\theta}[f_n]\doteq\mathbb{E}_\theta\left[f_n(X_t^\delta)\middle|\mathcal{Y}_t^\delta\right],\quad \bar{f}_{n,\theta}=\int_{\mathcal{X}}f_n(x)\mu_\theta(dx).$$

Since f_n is bounded, we already know that $\lim_{\delta\downarrow 0}\mathbb{E}_\alpha|\pi_t^{\delta,\theta}[f_n]-\bar{\pi}_t^{\theta}[f_n]|^2=0$. So, it is enough to prove that

$$\lim_{n\to\infty}\limsup_{\delta\downarrow 0}\mathbb{E}_\alpha\left|\pi_t^{\delta,\theta}[f]-\pi_t^{\delta,\theta}[f_n]\right|^2=0$$

and

$$\lim_{n\to\infty}\left|\bar{\pi}_t^{\theta}[f]-\bar{\pi}_t^{\theta}[f_n]\right|^2=0.$$

Both of these statements follow from the observation

$$|f(x)-f_n(x)|^{2p^2}\le|f(x)|^{2p^2}\mathbb{1}_{[|f(x)|>n]}\le|f(x)|^{2+\eta}\mathbb{1}_{[|f(x)|>n]}\le n^{-\eta}|f(x)|^{2+\eta}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In particular, we have

(A.1)

$$\begin{aligned}
&\lim_{n\to\infty}\limsup_{\delta\downarrow 0}\mathbb{E}_\alpha\left(\pi_t^{\delta,\theta}[f]-\pi_t^{\delta,\theta}[f_n]\right)^2=\lim_{n\to\infty}\limsup_{\delta\downarrow 0}\mathbb{E}_\alpha\left(\mathbb{E}_\theta\left[f(X_t^\delta)-f_n(X_t^\delta)\Big|\mathcal{Y}_t^\delta\right]\right)^2\\
&\quad\le\lim_{n\to\infty}\limsup_{\delta\downarrow 0}\mathbb{E}_\alpha\mathbb{E}_\theta\left[\left|f(X_t^\delta)-f_n(X_t^\delta)\right|^2\Big|\mathcal{Y}_t^\delta\right]\\
&\quad\le\lim_{n\to\infty}\limsup_{\delta\downarrow 0}\left(\mathbb{E}^*_\alpha\left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}_\theta\left|Z_t^{\delta,\theta}\right|^{-q}\right)^{1/(pq)}\left(\mathbb{E}_\theta\left|f(X_t^\delta)-f_n(X_t^\delta)\right|^{2p^2}\right)^{1/p^2}\\
&\quad\le 2\lim_{n\to\infty}\limsup_{\delta\downarrow 0}n^{-\eta/p^2}\left(\mathbb{E}^*_\alpha\left|Z_t^{\delta,\alpha}\right|^q\right)^{1/q}\left(\mathbb{E}_\theta\left|Z_t^{\delta,\theta}\right|^{-q}\right)^{1/(pq)}\left(\mathbb{E}_\theta\left|f(X_t^\delta)\right|^{2+\eta}\right)^{1/p^2}\\
&\quad=0
\end{aligned}$$

and clearly $\lim_{n\to\infty}\left(\bar\pi_t^\theta[f]-\bar\pi_t^\theta[f_n]\right)^2=\lim_{n\to\infty}\left(\bar f_\theta-\bar f_{n,\theta}\right)^2=0$. This concludes the proof of the theorem. ∎

Appendix B. Some convergence results for the posterior expectation of the eigenfunctions. In this appendix, we collect a number of results associated with the asymptotic behavior of the posterior for the eigenfunctions and their correlation as $\delta\downarrow 0$. Recall that α denotes the true parameter value.

LEMMA B.1. *Suppose h_θ is uniformly bounded over θ by a constant $C_h<\infty$ such that $\sup_{\theta\in\Theta}\|h_\theta\|_\infty\le C_h$. Then there exists another constant $C_\mu<\infty$ such that*

$$\sup_{\alpha,\theta\in\Theta}\mathbb{E}^*_\theta\left[\left(\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]}\right)^2\right]\le C_\mu,$$

and for any $f\in\mathcal{A}^\eta_\theta$ with $f\ge 0$ we have

$$\mathbb{E}_\alpha\pi_t^{\delta,\theta}[f]\le\frac{\mathbb{E}_\theta\left[f(X_t^\delta)\right]}{2}\left(e^{tC_h^2}+C_\mu\right)$$

for any $\theta,\alpha\in\Theta$ and for any $t\in[0,T]$.

Proof. From the Cauchy inequality (i.e., $ab\le a^2/2+b^2/2$ for all $a,b\in\mathbb{R}$), we have the following uniform bound:

$$\begin{aligned}
\mathbb{E}^*_\theta\left[\left(\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]}\right)^2\right]&\le\frac{1}{2}\mathbb{E}^*_\theta\left[\left(\phi_t^{\delta,\alpha}[1]\right)^4\right]+\frac{1}{2}\mathbb{E}^*_\theta\left[\left(\frac{1}{\phi_t^{\delta,\theta}[1]}\right)^4\right]\\
&\le\frac{1}{2}\mathbb{E}^*_\theta\left[\left(Z_t^{\delta,\alpha}\right)^4\right]+\frac{1}{2}\mathbb{E}^*_\theta\left[\left(Z_t^{\delta,\theta}\right)^{-4}\right]\\
&=\frac{1}{2}\mathbb{E}^*_\theta\mathbb{E}^*_\theta\left[\left(Z_t^{\delta,\alpha}\right)^4\Big|(X_s^\delta)_{s\le t}\right]+\frac{1}{2}\mathbb{E}^*_\theta\mathbb{E}^*_\theta\left[\left(Z_t^{\delta,\theta}\right)^{-4}\Big|(X_s^\delta)_{s\le t}\right]\\
&=\frac{1}{2}\mathbb{E}^*_\theta\left[e^{6\int_0^t|h_\alpha(X_s^\delta)|^2ds}\right]+\frac{1}{2}\mathbb{E}^*_\theta\left[e^{12\int_0^t|h_\theta(X_s^\delta)|^2ds}\right]\\
&\le\frac{e^{6TC_h^2}+e^{12TC_h^2}}{2}\\
&<\infty\,.
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

This proves the first statement of the lemma, with the constant being $C_\mu \doteq \frac{1}{2}(e^{6TC_h^2} + e^{12TC_h^2})$. To prove the lemma's second statement, we take any $f \in \mathcal{A}_\theta^\eta$ with $f \geq 0$ and proceed as follows:

$$\begin{aligned}
\mathbb{E}_\alpha \pi_t^{\delta,\theta}[f] &= \frac{\mathbb{E}_\alpha^* \left[Z_t^{\delta,\alpha} \pi_t^{\delta,\theta}[f] \right]}{\mathbb{E}_\alpha^* \left[Z_t^{\delta,\alpha} \right]} \\
&= \mathbb{E}_\alpha^* \left[Z_t^{\delta,\alpha} \pi_t^{\delta,\theta}[f] \right] = \mathbb{E}_\alpha^* \left[\mathbb{E}_\alpha^* \left[Z_t^{\delta,\alpha} \pi_t^{\delta,\theta}[f] \middle| \mathcal{Y}_t^\delta \right] \right] \\
&= \mathbb{E}_\alpha^* \left[\mathbb{E}_\alpha^* \left[Z_t^{\delta,\alpha} \middle| \mathcal{Y}_t^\delta \right] \pi_t^{\delta,\theta}[f] \right] = \mathbb{E}_\alpha^* \left[\phi_t^{\delta,\alpha}[1] \pi_t^{\delta,\theta}[f] \right],
\end{aligned}$$

and because $(Y_t^\delta)_{t \leq T}$ is Brownian motion under both $\mathbb{P}_\alpha^*$ and $\mathbb{P}_\theta^*$, we have that the last display continues as

$$\begin{aligned}
= \mathbb{E}_\theta^* \left[\phi_t^{\delta,\alpha}[1] \pi_t^{\delta,\theta}[f] \right] &= \mathbb{E}_\theta^* \left[\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} \phi_t^{\delta,\theta}[f] \right] \\
&= \mathbb{E}_\theta^* \left[\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} \mathbb{E}_\theta^* \left[Z_t^{\delta,\theta} f(X_t^\delta) \middle| \mathcal{Y}_t^\delta \right] \right] \\
&= \mathbb{E}_\theta^* \left[\mathbb{E}_\theta^* \left[\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} Z_t^{\delta,\theta} f(X_t^\delta) \middle| \mathcal{Y}_t^\delta \right] \right] \\
&= \mathbb{E}_\theta^* \left[\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} Z_t^{\delta,\theta} f(X_t^\delta) \right] \\
&= \mathbb{E}_\theta^* \left[f(X_t^\delta) \mathbb{E}_\theta^* \left[\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} Z_t^{\delta,\theta} \middle| (X_s^\delta)_{s \leq t} \right] \right] \\
&\leq \frac{1}{2} \mathbb{E}_\theta^* \left[f(X_t^\delta) \mathbb{E}_\theta^* \left[\left(Z_t^{\delta,\theta} \right)^2 + \left(\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} \right)^2 \middle| (X_s^\delta)_{s \leq t} \right] \right] \\
&= \frac{1}{2} \mathbb{E}_\theta^* \left[f(X_t^\delta) \exp\left(\int_0^t |h_\theta^\delta(X_s^\delta)|^2 ds \right) \right] \\
&\qquad + \frac{1}{2} \mathbb{E}_\theta^* \left[f(X_t^\delta) \left(\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} \right)^2 \right] \\
&= \frac{1}{2} \mathbb{E}_\theta^* \left[f(X_t^\delta) \exp\left(\int_0^t |h_\theta^\delta(X_s^\delta)|^2 ds \right) \right] \\
&\qquad + \frac{1}{2} \mathbb{E}_\theta^* \left[f(X_t^\delta) \right] \underbrace{\mathbb{E}_\theta^* \left[\left(\frac{\phi_t^{\delta,\alpha}[1]}{\phi_t^{\delta,\theta}[1]} \right)^2 \right]}_{\leq C_\mu} \\
&\leq \frac{\mathbb{E}_\theta^* \left[f(X_t^\delta) \right]}{2} \left(e^{tC_h^2} + C_\mu \right) .
\end{aligned}$$

This concludes the proof of the lemma. $\square$

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LEMMA B.2. *Assume Conditions* 2.1 *and* 3.2. *For any* $s \in (0,T]$, *for any* $\alpha, \theta \in \Theta$, *we have that there exists a constant* $C_{i,j}(T,\theta) < \infty$ *that may depend on* i,j,T,θ *but does not depend on* δ *such that for* δ *small enough,*

$$\sup_{i,j\geq 1} \frac{1}{\delta}\left|\mathbb{E}_\alpha \pi_s^{\delta,\theta}[\psi_i^\theta]\pi_s^{\delta,\theta}[\psi_j^\theta]\right| \leq C_{i,j}(T,\theta).$$

In addition, we also have that

$$\lim_{\delta\downarrow 0}\frac{1}{\delta}\mathbb{E}_\alpha\left[\pi_s^{\delta,\theta}[\psi_i^\theta]\pi_s^{\delta,\theta}[\psi_j^\theta]\right] = \frac{\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta}{\lambda_i^\theta + \lambda_j^\theta}.$$

Proof. Based on (3.10), we can write

(B.1)
$$\begin{aligned}
&\frac{1}{\delta}\mathbb{E}_\alpha\left[\pi_s^{\delta,\theta}[\psi_i^\theta]\pi_s^{\delta,\theta}[\psi_j^\theta]\right] \\
&= \frac{1}{\delta}e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}s}\pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta] \\
&\quad + \frac{1}{\delta}\int_0^s \mathbb{E}_\alpha\left\{\left[e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}\pi_\rho^{\delta,\theta}[\psi_j^\theta]\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta] - \pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right)\right.\right. \\
&\quad \left.\left. + e^{-\frac{\lambda_j^\theta}{\delta}(s-\rho)}\pi_\rho^{\delta,\theta}[\psi_i^\theta]\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta] - \pi_\rho^{\delta,\theta}[\psi_j^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right)\right]\left(\pi_\rho^{\delta,\theta}[h_\theta] - \pi_\rho^{\delta,\alpha}[h_\alpha]\right)\right\}d\rho \\
&\quad + \frac{1}{\delta}\int_0^s e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}(s-\rho)} \\
&\times \mathbb{E}_\alpha\left[\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta] - \pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right)\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta] - \pi_\rho^{\delta,\theta}[\psi_j^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right)\right]d\rho\ ,
\end{aligned}$$

and then taking absolute values inside the integrals, applying the Cauchy inequality ($ab \leq a^2/2 + b^2/2$ for all $a, b \in \mathbb{R}$), and applying Lemma B.1, we have the following bound:

(B.2)
$$\begin{aligned}
&\frac{1}{\delta}\mathbb{E}_\alpha\left|\pi_s^{\delta,\theta}[\psi_i^\theta]\pi_s^{\delta,\theta}[\psi_j^\theta]\right| \\
&\leq \frac{1}{\delta}e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}s}\left|\pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]\right| \\
&\quad + \frac{1}{\delta}\int_0^s \mathbb{E}_\alpha\left[\left(e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}\left|\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right|\left|\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta] - \pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|\right.\right. \\
&\quad \left.\left. + e^{-\frac{\lambda_j^\theta}{\delta}(s-\rho)}\left|\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right|\left|\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta] - \pi_\rho^{\delta,\theta}[\psi_j^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|\right)\left|\pi_\rho^{\delta,\theta}[h_\theta] - \pi_\rho^{\delta,\alpha}[h_\alpha]\right|\right]d\rho \\
&\quad + \frac{1}{\delta}\int_0^s e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}(s-\rho)}\mathbb{E}_\alpha\left[\left|\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta] - \pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|\right. \\
&\quad \left.\times\left|\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta] - \pi_\rho^{\delta,\theta}[\psi_j^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|\right]d\rho
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$
\begin{aligned}
&\leq \frac{1}{\delta} e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}s} \left|\pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]\right| \\
&\quad + \frac{1}{2\delta}\int_0^s \mathbb{E}_\alpha \Bigg[e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)} \Bigg[\left|\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right|^2 + \left|\pi_\rho^{\delta,\theta}[h_\theta \psi_i^\theta] - \pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|^2 \\
&\qquad + e^{-\frac{\lambda_j^\theta}{\delta}(s-\rho)} \left|\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right|^2 + \left|\pi_\rho^{\delta,\theta}[h_\theta \psi_j^\theta] - \pi_\rho^{\delta,\theta}[\psi_j^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|^2 \Bigg] 2C_h \Bigg] d\rho \\
&\quad + \frac{1}{2\delta}\int_0^s e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}(s-\rho)} \mathbb{E}_\alpha \Big[\left|\pi_\rho^{\delta,\theta}[h_\theta \psi_i^\theta] - \pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|^2 \\
&\quad + \left|\pi_\rho^{\delta,\theta}[h_\theta \psi_j^\theta] - \pi_\rho^{\delta,\theta}[\psi_j^\theta]\pi_\rho^{\delta,\theta}[h_\theta]\right|^2 \Big] d\rho \\
&\leq C_0 \Bigg(\frac{1}{\lambda_i^\theta+\lambda_j^\theta} \frac{\lambda_i^\theta+\lambda_j^\theta}{\delta} e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}s} \left|\pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]\right| \\
&+ \frac{1}{\delta}\int_0^s \left[e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)} + e^{-\frac{\lambda_j^\theta}{\delta}(s-\rho)} \right] \mathbb{E}_\theta \left[|\psi_i^\theta(X_\rho^\delta)|^2 + |\psi_j^\theta(X_\rho^\delta)|^2 \right] d\rho \Bigg),
\end{aligned}
$$

where C_0 is a constant not dependent on δ. Recall now that by assuming Condition 3.2, for every $i \in \mathbb{N}$ we have $\psi_i^\theta \in \mathcal{A}_\eta^\theta$. This implies that there exists finite constants that may depend on i, T, and θ such that

$$
\sup_{\delta\in(0,1),\rho\in[0,T]} \mathbb{E}_\theta \left[|\psi_i^\theta(X_\rho^\delta)|^2 \right] \leq C(\psi_i, T, \theta).
$$

Noticing that

$$
\frac{1}{\delta}\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)} d\rho = \frac{1}{\lambda_i^\theta}\left(1 - e^{-\frac{\lambda_i^\theta}{\delta}s}\right) \leq \frac{1}{\lambda_i^\theta}
$$

and that for δ sufficiently small $\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta} e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}s} \leq 1$, and recalling Condition 3.2, it follows that the required bound for the first statement follows with the constant

$$
C_{i,j}(T,\theta) = C_0 \left(\frac{\left|\pi_0^\theta[\psi_i^\theta]\pi_0^\theta[\psi_j^\theta]\right|}{\lambda_i^\theta+\lambda_j^\theta} + (C(\psi_i,T,\theta) + C(\psi_j,T,\theta)) \left(\frac{1}{\lambda_i^\theta} + \frac{1}{\lambda_j^\theta} \right) \right). \tag{B.3}
$$

The second statement is obtained by adding and subtracting the terms $\langle h_\theta, \psi_i^\theta \rangle_\theta$ and $\langle h_\theta, \psi_j^\theta \rangle_\theta$ in the products of the last integral of (B.1) and then using Theorem 3.1. ∎

Lemma B.3. *Assume Conditions* 2.1 *and* 3.2 *and* $\sum_{i,j=1}^\infty |\langle h_\theta, \psi_i^\theta\rangle_\theta \langle h_\theta, \psi_j^\theta\rangle_\theta| C_{i,j}(T,\theta) < \infty$, *where* $C_{i,j}(T,\theta)$ *is given by* (B.3) *in the proof of Lemma* B.2. *For any* $0 < T < \infty$ *and for any* $\theta \in \Theta$, *we have that there exists a constant* $C < \infty$ *that does not depend on* δ *and* $\delta_0 < \infty$ *such that*

$$
\sup_{\delta\in(0,\delta_0)} \mathbb{E}_\alpha \left[\frac{1}{\delta}\int_0^T \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds \right] < C.
$$

Proof. Recalling that

$$
\pi_s^{\delta,\theta}[\tilde{h}_\theta] = \sum_{i=1}^\infty \langle h_\theta, \psi_i^\theta \rangle_\theta \pi_s^{\delta,\theta}[\psi_i^\theta],
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

we obtain

$$
\begin{aligned}
&\sup_{\delta\in(0,\delta_0)} \frac{1}{\delta}\int_0^T \mathbb{E}_\alpha \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds\\
&\quad = \sup_{\delta\in(0,\delta_0)} \int_0^T \sum_{i,j=1}^{\infty} \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta \frac{1}{\delta}\mathbb{E}_\alpha\left[\pi_s^{\delta,\theta}[\psi_i^\theta]\pi_s^{\delta,\theta}[\psi_j^\theta]\right] ds\\
&\quad \le \sum_{i,j=1}^{\infty} \left|\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta\right| \sup_{\delta\in(0,\delta_0)} \int_0^T \frac{1}{\delta}\left|\mathbb{E}_\alpha \pi_s^{\delta,\theta}[\psi_i^\theta]\pi_s^{\delta,\theta}[\psi_j^\theta]\right| ds\\
&\quad \le T\sum_{i,j=1}^{\infty} \left|\left\langle h_\theta, \psi_i^\theta\right\rangle_\theta \left\langle h_\theta, \psi_j^\theta\right\rangle_\theta\right| C_{i,j}(T,\theta)\\
&\quad < \infty,
\end{aligned}
$$

and so the constant is $C = T\sum_{i,j=1}^{\infty} |\langle h_\theta, \psi_i^\theta\rangle_\theta \langle h_\theta, \psi_j^\theta\rangle_\theta| C_{i,j}(T,\theta)$. ∎

LEMMA B.4. *Assume Conditions* 2.1 *and* 3.2. *For any* $0 < T < \infty$ *and* $\theta \in \Theta$ *we have*

$$
\sup_{t\in[0,T]} \mathbb{E}_\alpha \left| \frac{1}{\sqrt{\delta}} \int_0^t e^{-\frac{\lambda_i^\theta(t-s)}{\delta}} \left(\pi_s^{\delta,\theta}[h_\theta\psi_i^\theta] - \left\langle h_\theta, \psi_i^\theta\right\rangle_\theta - \pi_s^{\delta,\theta}[h_\theta]\pi_s^{\delta,\theta}[\psi_i^\theta]\right) d\nu_s^{\delta,\alpha}\right|^2 \to 0
$$

as $\delta \downarrow 0$.

Proof. Due to Itô isometry we have

$$
\begin{aligned}
&\mathbb{E}_\alpha\left|\frac{1}{\sqrt{\delta}}\int_0^t e^{-\frac{\lambda_i^\theta(t-s)}{\delta}}(\pi_s^{\delta,\theta}[h_\theta\psi_i^\theta] - \langle h_\theta,\psi_i^\theta\rangle_\theta - \pi_s^{\delta,\theta}[h_\theta]\pi_s^{\delta,\theta}[\psi_i^\theta])d\nu_s^{\delta,\alpha}\right|^2\\
&= \mathbb{E}_\alpha\left|\frac{1}{\sqrt{\delta}}\int_0^t e^{-\frac{\lambda_i^\theta(t-s)}{\delta}}(\pi_s^{\delta,\theta}[h_\theta\psi_i^\theta] - \langle h_\theta,\psi_i^\theta\rangle_\theta\right.\\
&\qquad\qquad \left. - (\pi_s^{\delta,\theta}[h_\theta] - \bar{h}_\theta)\pi_s^{\delta,\theta}[\psi_i^\theta] - \bar{h}_\theta\pi_s^{\delta,\theta}[\psi_i^\theta])d\nu_s^{\delta,\alpha}\right|^2\\
&\le 3\left[\frac{1}{\delta}\int_0^t e^{-2\frac{\lambda_i^\theta(t-s)}{\delta}}\mathbb{E}_\alpha|\pi_s^{\delta,\theta}[h_\theta\psi_i^\theta] - \langle h_\theta,\psi_i^\theta\rangle_\theta|^2 ds + \frac{C_h^2}{\delta}\int_0^t e^{-2\frac{\lambda_i^\theta(t-s)}{\delta}}\mathbb{E}_\alpha\left|\pi_s^{\delta,\theta}[\psi_i^\theta]\right|^2 ds\right.\\
&\qquad \left. + \frac{1}{\delta}\int_0^t e^{-2\frac{\lambda_i^\theta(t-s)}{\delta}}\mathbb{E}_\alpha\|\pi_s^{\delta,\theta}[h_\theta] - \bar{h}_\theta|^2 ds\right].
\end{aligned}
\tag{B.4}
$$

Noticing that

$$
\sup_{t\in[0,T]} \frac{1}{\delta}\int_0^t e^{-2\frac{\lambda_i^\theta(t-s)}{\delta}} ds = \sup_{t\in[0,T]} \frac{1}{\lambda_i^\theta}\left(1 - e^{-2\frac{\lambda_i^\theta}{\delta}t}\right) \le \frac{1}{\lambda_i^\theta},
$$

the statement of the lemma follows by Theorem 3.1 and the dominated convergence theorem to (B.4). Notice that the dominated convergence theorem is applicable since we can apply Lemma B.1 to the integrands and notice that the integrands are expectations of functions in $\mathcal{A}_\eta^\theta$. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

LEMMA B.5. *Assume the conditions of Lemma* B.3. *For any* $0 < T < \infty$ *and for any* $\theta \in \Theta$, *we have in* $\mathbb{P}_\alpha$*-probability and uniformly in* $t \in [0, T]$ *that*

$$\int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) \left(\pi_s^{\delta,\theta}[h_\theta] - \pi_s^{\delta,\alpha}[h_\alpha]\right) ds \to 0 \qquad \text{as } \delta \downarrow 0.$$

Proof. First we notice that

$$\pi_s^{\delta,\theta}[h_\theta] - \pi_s^{\delta,\alpha}[h_\alpha] = \left(\bar{h}_\theta - \bar{h}_\alpha\right) + \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta] - \pi_s^{\delta,\alpha}[\tilde{h}_\alpha]\right).$$

Using the Cauchy inequality ($ab \leq a^2/2 + b^2/2$ for all $a, b \in \mathbb{R}$), this implies that

(B.5)
$$\begin{aligned}
&\mathbb{E}_\alpha \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) \left(\pi_s^{\delta,\theta}[h_\theta] - \pi_s^{\delta,\alpha}[h_\alpha]\right) ds \right| \\
&\leq \left|\bar{h}_\theta - \bar{h}_\alpha\right| \mathbb{E}_\alpha \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) ds \right| \\
&\qquad\qquad + \mathbb{E}_\alpha \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta] - \pi_s^{\delta,\alpha}[\tilde{h}_\alpha]\right) ds \right| \\
&\leq \left|\bar{h}_\theta - \bar{h}_\alpha\right| \mathbb{E}_\alpha \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) ds \right| \\
&\qquad + \mathbb{E}_\alpha \sup_{t\in[0,T]} \int_0^t \frac{2}{\sqrt{\delta}} \left|\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right|^2 ds + \mathbb{E}_\alpha \sup_{t\in[0,T]} \int_0^t \frac{1}{\sqrt{\delta}} \left|\pi_s^{\delta,\alpha}[\tilde{h}_\alpha]\right|^2 ds \\
&\leq \left|\bar{h}_\theta - \bar{h}_\alpha\right| \mathbb{E}_\alpha \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) ds \right| + 3C\sqrt{\delta},
\end{aligned}$$

where in the last step we used the bound from Lemma B.3. Since $3C\sqrt{\delta}$ clearly goes to zero as $\delta \downarrow 0$, it remains to show that the first term will also go to zero as $\delta \downarrow 0$. Namely, it remains to show that

$$\lim_{\delta\downarrow 0} \mathbb{E}_\alpha \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{\sqrt{\delta}} \left(\pi_s^{\delta,\theta}[\tilde{h}_\theta]\right) ds \right| = 0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Using similar computations as in the proof of Lemma B.3, we notice that

$$\begin{aligned}
&\mathbb{E}_\alpha \sup_{t\in[0,T]}\left|\frac{1}{\sqrt{\delta}}\int_0^t \pi_s^{\delta,\theta}[\tilde{h}_\theta]ds\right| \leq \left|\frac{1}{\sqrt{\delta}}\delta\sum_{i=1}^{\infty}\frac{\langle h_\theta,\psi_i^\theta\rangle_\theta}{\lambda_i^\theta}\pi_0^\theta[\psi_i^\theta]\sup_{t\in[0,T]}\frac{\lambda_i^\theta}{\delta}\int_0^t e^{-\frac{\lambda_i^\theta}{\delta}s}ds\right|\\
&\quad+\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\sum_{i=1}^{\infty}\langle h_\theta,\psi_i^\theta\rangle_\theta\frac{1}{\sqrt{\delta}}\int_0^t\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}\right.\\
&\qquad\qquad\left.\times(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta])d\nu_\rho^{\delta,\alpha}ds\right|\\
&\quad+\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\sum_{i=1}^{\infty}\langle h_\theta,\psi_i^\theta\rangle_\theta\frac{1}{\sqrt{\delta}}\int_0^t\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}\right.\\
&\qquad\qquad\left.\times(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta])(\pi_\rho^{\delta,\theta}[h_\theta]-\pi_\rho^{\delta,\alpha}[h_\alpha])d\rho ds\right|\\
&\leq 2\sqrt{\delta}\sum_{i=1}^{\infty}\left|\frac{\langle h_\theta,\psi_i^\theta\rangle_\theta}{\lambda_i^\theta}\pi_0^\theta[\psi_i^\theta]\right|\\
&\quad+\sum_{i=1}^{\infty}|\langle h_\theta,\psi_i^\theta\rangle_\theta|\int_0^T\mathbb{E}_\alpha\frac{1}{\sqrt{\delta}}\left|\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}\right.\\
&\qquad\qquad\left.\times(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\langle h_\theta,\psi_i^\theta\rangle_\theta-\pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta])d\nu_\rho^{\delta,\alpha}\right|ds\\
&\quad+\sum_{i=1}^{\infty}|\langle h_\theta,\psi_i^\theta\rangle_\theta|^2\frac{1}{\sqrt{\delta}}\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\int_0^t\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}d\nu_\rho^{\delta,\alpha}ds\right|\\
&\quad+\sqrt{\delta}\sum_{i=1}^{\infty}\frac{|\langle h_\theta,\psi_i^\theta\rangle_\theta|}{\lambda_i^\theta}\frac{\lambda_i^\theta}{\delta}\int_0^T\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}\mathbb{E}_\alpha\big|(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\pi_\rho^{\delta,\theta}[\psi_i^\theta]\pi_\rho^{\delta,\theta}[h_\theta])\\
&\qquad\qquad\times(\pi_\rho^{\delta,\theta}[h_\theta]-\pi_\rho^{\delta,\alpha}[h_\alpha])\big|d\rho ds.
\end{aligned}$$

Clearly, the first term goes to zero as $\delta\downarrow 0$. Similarly, the fourth term also goes to zero as $\delta\downarrow 0$ and this follows by Condition 3.2(i)–(iii). By Lemma B.4, the second term can also be shown to go to zero. So, it essentially remains to treat the third term. For this purpose, we recall that the solution to (3.18), $\Xi_t^{\delta,i}$, is given by (3.19), which is normally distributed with mean zero and variance $\frac{1}{2\lambda_i^\theta}(1-e^{-\frac{\lambda_i^\theta t}{\delta}})$. Hence, the third term in question can be written as

$$\begin{aligned}
&\sum_{i=1}^{\infty}\left|\langle h_\theta,\psi_i^\theta\rangle_\theta\right|^2\frac{1}{\sqrt{\delta}}\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\int_0^t\int_0^s e^{-\frac{\lambda_i^\theta}{\delta}(s-\rho)}d\nu_\rho^{\delta,\alpha}ds\right|\\
&=\sum_{i=1}^{\infty}\left|\langle h_\theta,\psi_i^\theta\rangle_\theta\right|^2\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\int_0^t\Xi_s^{\delta,i}ds\right|\\
&\leq\sqrt{\delta}\left\{\sum_{i=1}^{\infty}\frac{\left|\langle h_\theta,\psi_i^\theta\rangle_\theta\right|^2}{\lambda_i^\theta}\left[\sqrt{\delta}\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\Xi_t^{\delta,i}\right|+\mathbb{E}_\alpha\sup_{t\in[0,T]}\left|\nu_t^{\delta,\alpha}\right|\right]\right\},
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and it is easy to see that this term goes to zero as $\delta \downarrow 0$. This completes the proof of the lemma. ∎

LEMMA B.6. *Assume Conditions* 2.1 *and* 3.2 *and that*

$$\sum_{i,j=1}^{\infty} \frac{\left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right| \left(C(\psi_i, T, \theta) + C(\psi_j, T, \theta)\right)}{\lambda_i^\theta + \lambda_j^\theta} < \infty,$$

where $C(\psi_i, T, \theta)$ *is a constant such that*

$$\sup_{\delta \in (0,1), \rho \in [0,T]} \mathbb{E}_\theta \left[|\psi_i^\theta(X_\rho^\delta)|^2 \right] \leq C(\psi_i, T, \theta),$$

which is the statement of (3.12)*. Then, the term* $R_t^{2,\delta}$ *from* (3.16) *converges to zero in the mean-square sense uniformly in* $t \in [0, T]$*,*

$$\lim_{\delta \downarrow 0} \mathbb{E}_\alpha \sup_{t \in [0,T]} \left| R_t^{2,\delta} \right|^2 = 0.$$

Proof. Since $R_t^{2,\delta}$ is a martingale, by Doob's inequality we have

$$\mathbb{E}_\alpha \sup_{t \in [0,T]} \left| R_t^{2,\delta} \right|^2 \leq 4 \mathbb{E}_\alpha \left[R_T^{2,\delta} \right]^2 ,$$

and it follows by the Cauchy inequality (i.e., $ab \leq a^2/2 + b^2/2$ for any $a, b \in \mathbb{R}$) and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

then Itô isometry that

(B.6)

$$\begin{aligned}
&\mathbb{E}_\alpha\left[R_T^{2,\delta}\right]^2 = \sum_{i,j=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\\
&\times\frac{1}{\delta}\mathbb{E}_\alpha\left[\int_0^T\left(\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right)d\nu_\rho^{\delta,\theta}\right)\right.\\
&\qquad\left.\times\left(\int_0^s e^{-\frac{\lambda_j^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta]-\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right)d\nu_\rho^{\delta,\theta}\right)ds\right]\\
&\le\sum_{i,j=1}^{\infty}\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\\
&\qquad\times\frac{1}{\delta}\left[\int_0^T\left(\int_0^s e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)(s-\rho)}{\delta}}\mathbb{E}_\alpha\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right)\right.\right.\\
&\qquad\left.\left.\times\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta]-\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right)d\rho\right)ds\right]\\
&+|2C_h|^2\sum_{i,j=1}^{\infty}\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\right|\\
&\qquad\times\frac{1}{\delta}\left[\int_0^T\left(\int_0^s e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)(s-\rho)}{\delta}}\mathbb{E}_\alpha\left|\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right|\right.\right.\\
&\qquad\qquad\left.\left.\times\left|\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta]-\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right|d\rho\right)ds\right]\\
&\le\frac{1+|C_h|^2}{2}\sum_{i,j=1}^{\infty}\int_0^T\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\right|\\
&\quad\times\frac{1}{\delta}\left[\int_0^s e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)(s-\rho)}{\delta}}\left(\mathbb{E}_\alpha\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right)^2\right.\right.\\
&\qquad\left.\left.+\mathbb{E}_\alpha\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta]-\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right)^2d\rho\right)ds\right].
\end{aligned}$$

Now we want to apply dominated convergence theorem equation (B.6) in order to argue that the upper bound of the last inequality goes to zero as $\delta\downarrow 0$. First, we notice that by Lemma B.1, we have the following bound for the integrand:

(B.7)

$$\begin{aligned}
&\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\right|\frac{1}{\delta}\left(\int_0^s e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)(s-\rho)}{\delta}}\mathbb{E}_\alpha\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_i^\theta]-\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_i^\theta]\right)^2\right.\\
&\qquad\left.+\mathbb{E}_\alpha\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta]-\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta-\pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right)^2d\rho\right)\\
&\le C_0\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\right|\frac{1}{\delta}\int_0^s e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)(s-\rho)}{\delta}}\left[\mathbb{E}_\theta\left(\left|\psi_i^\theta\left(X_\rho^\delta\right)\right|^2+\left|\psi_j^\theta\left(X_\rho^\delta\right)\right|^2\right)\right]d\rho.
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Recall now that by assuming Condition 3.2, for every $i \in \mathbb{N}$ we have $\psi_i^\theta \in \mathcal{A}_\eta^\theta$. This implies that there exist finite constants that may depend on i, T, and θ such that

$$\sup_{\delta\in(0,1),\rho\in[0,T]} \mathbb{E}_\theta\left[|\psi_i^\theta(X_\rho^\delta)|^2\right] \le C(\psi_i, T, \theta).$$

Noticing that

$$\frac{1}{\delta}\int_0^s e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}(s-\rho)}d\rho = \frac{1}{\lambda_i^\theta+\lambda_j^\theta}\left(1-e^{-\frac{\lambda_i^\theta+\lambda_j^\theta}{\delta}s}\right) \le \frac{1}{\lambda_i^\theta+\lambda_j^\theta},$$

we can then continue bounding (B.7) by the term

$$\text{(B.8)}\qquad \begin{aligned} &\le C_0\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta \left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\right| \frac{1}{\delta}\int_0^s e^{-\frac{\left(\lambda_i^\theta+\lambda_j^\theta\right)(s-\rho)}{\delta}}\left[C(\psi_i,T,\theta)+C(\psi_j,T,\theta)\right]d\rho \\ &\le C_0\left[\frac{\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta \left\langle h_\theta,\psi_j^\theta\right\rangle_\theta\right|}{\lambda_i^\theta+\lambda_j^\theta}\left(C(\psi_i,T,\theta)+C(\psi_j,T,\theta)\right)\right].\end{aligned}$$

Hence, the summands in (B.6) are dominated by terms that are summable and are finite irrespective of $\delta \in (0,1)$. Second, by Theorem 3.1 we know that for each $i \ge 1$ there is the limit

$$\lim_{\delta\downarrow 0}\mathbb{E}_\alpha\left(\pi_\rho^{\delta,\theta}[h_\theta\psi_j^\theta] - \left\langle h_\theta,\psi_j^\theta\right\rangle_\theta - \pi_\rho^{\delta,\theta}[h_\theta]\pi_\rho^{\delta,\theta}[\psi_j^\theta]\right)^2 = 0.$$

Hence, by dominated convergence we have established that (B.6) goes to zero in probability, and then it follows that

$$\lim_{\delta\downarrow 0}\mathbb{E}_\alpha \sup_{t\in[0,T]}\left|R_t^{2,\delta}\right|^2 = 0. \qquad \square$$

LEMMA B.7. *Assume the conditions of Lemma* B.3. *Then, the term* $R_t^{3,\delta}$ *from* (3.16) *converges to zero in the mean-square sense uniformly in* $t \in [0,T]$,

$$\lim_{\delta\downarrow 0}\mathbb{E}_\alpha \sup_{t\in[0,T]}\left|R_t^{3,\delta}\right|^2 = 0.$$

Proof. We have

(B.9)

$$\begin{aligned} R_t^{3,\delta} &= \sum_{i=1}^\infty \left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\right|^2 \frac{1}{\sqrt{\delta}}\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[h_\theta]-\pi_\rho^{\delta,\alpha}[h_\alpha]\right)d\rho\right]d\nu_s^{\delta,\alpha} \\ &= (\bar h_\theta-\bar h_\alpha)\sum_{i=1}^\infty \left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\right|^2 \frac{1}{\sqrt{\delta}}\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}d\rho\right]d\nu_s^{\delta,\alpha} \\ &\quad + \sum_{i=1}^\infty \left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\right|^2 \frac{1}{\sqrt{\delta}}\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[\tilde h_\theta]-\pi_\rho^{\delta,\alpha}[\tilde h_\alpha]\right)d\rho\right]d\nu_s^{\delta,\alpha} \\ &= \sqrt{\delta}(\bar h_\theta-\bar h_\alpha)\sum_{i=1}^\infty \frac{\left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\right|^2}{\lambda_i^\theta}\int_0^t\left(1-e^{-\frac{\lambda_i^\theta s}{\delta}}\right)d\nu_s^{\delta,\alpha} \\ &\quad + \sum_{i=1}^\infty \left|\left\langle h_\theta,\psi_i^\theta\right\rangle_\theta\right|^2 \frac{1}{\sqrt{\delta}}\int_0^t\left[\int_0^s e^{-\frac{\lambda_i^\theta(s-\rho)}{\delta}}\left(\pi_\rho^{\delta,\theta}[\tilde h_\theta]-\pi_\rho^{\delta,\alpha}[\tilde h_\alpha]\right)d\rho\right]d\nu_s^{\delta,\alpha}.\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

By applying Doob's inequality followed by the Cauchy inequality and then Jensen's inequality, we have

$$\begin{aligned}
&\mathbb{E}_\alpha \sup_{t\in[0,T]} \left|R_t^{3,\delta}\right|^2 \\
&\leq \delta 2(\bar{h}_\theta - \bar{h}_\alpha)^2 \sum_{i,j=1}^{\infty} \frac{\left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right|^2}{\lambda_i^\theta + \lambda_j^\theta} \int_0^T \left(1 - e^{-\frac{\lambda_i^\theta s}{\delta}}\right)\left(1 - e^{-\frac{\lambda_j^\theta s}{\delta}}\right) ds \\
&\quad + 2 \sum_{i,j=1}^{\infty} \left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right|^2 \frac{1}{\delta} \mathbb{E}_\alpha \int_0^T \left(\int_0^s e^{-\frac{\lambda_i^\theta (s-\rho)}{\delta}} \left(\pi_\rho^{\delta,\theta}[\tilde{h}_\theta] - \pi_\rho^{\delta,\alpha}[\tilde{h}_\alpha]\right) d\rho\right) \\
&\qquad\qquad \times \left(\int_0^s e^{-\frac{\lambda_j^\theta (s-\rho)}{\delta}} \left(\pi_\rho^{\delta,\theta}[\tilde{h}_\theta] - \pi_\rho^{\delta,\alpha}[\tilde{h}_\alpha]\right) d\rho\right) ds \\
&\leq \delta 2T(\bar{h}_\theta - \bar{h}_\alpha)^2 \sum_{i,j=1}^{\infty} \frac{\left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right|^2}{\lambda_i^\theta + \lambda_j^\theta} \\
&\quad + 2 \sum_{i,j=1}^{\infty} \left|\left\langle h_\theta, \psi_i^\theta \right\rangle_\theta \left\langle h_\theta, \psi_j^\theta \right\rangle_\theta\right|^2 \int_0^T \left(\int_0^s e^{-\frac{2\lambda_1^\theta (s-\rho)}{\delta}} \frac{1}{\delta} \mathbb{E}_\alpha \left(\pi_\rho^{\delta,\theta}[\tilde{h}_\theta] - \pi_\rho^{\delta,\alpha}[\tilde{h}_\alpha]\right)^2 d\rho\right) ds.
\end{aligned}$$

Then, by the fact that $v_\theta^2(h_\theta) < \infty$ and $\sum_{i,j=1}^{\infty} |\langle h_\theta, \psi_i^\theta \rangle_\theta \langle h_\theta, \psi_j^\theta \rangle_\theta|^2 < \infty$ (see Remark 6) and the uniform bound from Lemma B.3, we obtain that

$$\lim_{\delta\downarrow 0} \mathbb{E}_\alpha \sup_{t\in[0,T]} \left|R_t^{3,\delta}\right|^2 = 0,$$

concluding the proof of the lemma. ∎

REFERENCES

M. Abramowitz and I. Stegun (1965), *Handbook of Mathematical Functions*, Dover Publications, Mineola, NY.

A. Bain and D. Crisan (2009), *Fundamentals of Stochastic Filtering*, Springer, New York.

A. Bensoussan and G. L. Blankenship (1986), *Nonlinear filtering with homogenization*, Stochastics, 17, pp. 67–90.

A. Bensoussan, J. Lions, and G. Papanicolaou (1978), *Asymptotic Analysis for Periodic Structures*, Stud. Math. Appl. 5, North–Holland, Amsterdam.

P. Billingsley (1968), *Convergence of Probability Measures*, John Wiley, New York.

J. Boyd (1984), *Asymptotic coefficients of Hermite function series*, J. Comput. Phys., 54, pp. 382–410.

J. Boyd (2000), *Chebyshev and Fourier Spectral Methods*, 2nd ed., Dover Publications, Mineola, NY.

J. Brogaard, T. J. Hendershott, and R. Riordan (2012), *High Frequency Trading and Price Discovery*, Technical report, Berkeley University.

O. Cappé, E. Moulines, and T. Ryden (2005), *Inference in Hidden Markov Models*, Springer Ser. Statist., Springer-Verlag, New York, Secaucus, NJ.

P. Del Moral, J. Jacod, and P. Protter (2001), *The Monte-Carlo method for filtering with discrete-time observations*, Probab. Theory Related Fields, 120, pp. 346–368.

S. Ethier and T. Kurtz (1986), *Markov Processes: Characterization and Convergence*, John Wiley, Hoboken, NJ.

N. Ichihara (2004), *Homogenization problem for stochastic partial differential equations of Zakai type*, Stochastics Stochastics Rep., 76, pp. 243–266.

P. Imkeller, N. S. Namachchivaya, N. Perkowski, and H. C. Yeong (2013), *Dimensional reduction in nonlinear filtering: A homogenization approach*, Ann. Appl. Probab., 23, pp. 2290–2326.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/26/20 to 128.178.69.223. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

M. R. James and F. L. Gland (1995), *Consistent parameter estimation for partially observed diffusions with small noise*, Appl. Math. Optim., 32, pp. 47–72.

G. Kallianpur (1980), *Stochastic Filtering Theory*, Springer, Berlin.

I. Karatzas and S. E. Shreve (1991), *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, New York.

M. L. Kleptsina, R. S. Liptser, and A. Serebrovski (1997), *Nonlinear filtering problem with contamination*, Ann. Appl. Probab., 7, pp. 917–934.

H. J. Kushner (1990), *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser, Boston, Basel, Berlin.

Y. Kutoyants (2004), *Statistical Inference for Ergodic Diffusion Processes*, Springer, London.

D. Liberzon and R. W. Brockett (2000), *Spectral analysis of Fokker–Planck and related operators arising from linear stochastic differential equations*, SIAM J. Control Optim., 38, pp. 1453–1467.

V. Linetsky (2007), *Spectral methods in derivative pricing*, in Handbooks in Operations Research and Management Science: Financial Engineering, Volume 15, Elsevier B.V., Amsterdam, pp. 223–299.

A. Papavasiliou, G. Pavliotis, and A. Stuart (2009), *Maximum likelihood drift estimation for multiscale diffusions*, Stochast. Process. Appl., 119, pp. 3173–3210.

E. Pardoux and A. Veretennikov (2003), *On Poisson equation and diffusion approximation* II, Ann. Probab., 31, pp. 1066–1092.

J. Park, B. Rozovsky, and R. Sowers (2011), *Efficient nonlinear filtering of a singularly perturbed stochastic hybrid system*, LMS J. Comput. Math., 14, pp. 254–270.

J. Park, R. Sowers, and N. S. Namachchivaya (2008), *A problem in stochastic averaging of nonlinear filters*, Stoch. Dyn., 8, pp. 543–560.

J. Park, R. Sowers, and N. S. Namachchivaya (2010), *Dimensional reduction in nonlinear filtering*, Nonlinearity, 23, pp. 305–324.

L. Rozovskii (1990), *Stochastic Evolution System: Linear Theory and Applications to Non-linear Filtering*, Kluwer Academic Publishers, Dordrecht.

B. Rozovsky (1991), *A simple proof of uniqueness for Kushner and Zakai equations*, in Stochastic Analysis, E. Mayer-Wolf, ed., Academic Press, Boston, pp. 449–458.

F. Zhang (2010), *High-Frequency Trading, Stock Volatility, and Price Discovery*, SSRN eLibrary, http://ssrn.com/abstract=1691679.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.