STATISTICAL INFERENCE FOR PERTURBATIONS OF MULTISCALE DYNAMICAL SYSTEMS

SIRAGAN GAILUS AND KONSTANTINOS SPILIOPOULOS

DEPARTMENT OF MATHEMATICS & STATISTICS, BOSTON UNIVERSITY

111 CUMMINGTON MALL, BOSTON, MA 02215

E-MAIL (SG): SIRAGAN@MATH.BU.EDU (KS): KSPILIOP@MATH.BU.EDU

ABSTRACT. In this paper, we study statistical inference for small-noise perturbations of multiscale dynamical systems. We prove the asymptotic consistency and asymptotic normality of an appropriately constructed maximum likelihood estimator (MLE) for a parameter of interest, identifying precisely its limiting variance. We allow unbounded coefficients in the equation for the slow process and assume neither periodicity nor that the fast process is compact. Ergodicity of the fast process is guaranteed by imposing a recurrence condition. Moreover, we allow full dependence of the coefficients on the slow and fast components and allow correlation between the driving noises of the slow and fast components. The results provide a theoretical basis for calibration of small-noise perturbed multiscale stochastic dynamical systems and related diffusion processes. In the course of the proof we also derive exponential bounds and ergodic theorems that may be of independent interest. Data from numerical simulations are provided to supplement and illustrate the theory.

1 Introduction

In many cases, data from physical dynamical systems exhibit multiple characteristic length- and time-scales. It is of interest in such cases to develop models that capture the large-scale dynamics without losing sight of the small scales. Meanwhile, noise may be introduced in order to account for uncertainty or as an essential part of a particular modelling problem. Therefore, multiscale SDE models are widely deployed in applied fields as diverse as physics, chemistry, and biology [5, 11, 27], neuroscience [12], meteorology [16], and econometrics and mathematical finance [9, 26] to capture stochastic perturbations of multiscale dynamical systems with two or more different space- or time-scales.

The noise in such models is often taken to be of small magnitude. This may be because one is interested in studying rare transition events among different equilibrium states of a system [6, 11, 27], a small stochastic perturbation of an underlying dynamical system [10, 14, 12], or small-time asymptotics [7, 8, 23]. Manuscript [14] is devoted to statistical inference problems for small noise perturbations of dynamical systems, although multiple scales are not considered there.

The mathematical problem of parameter estimation in the context of general small-noise perturbed multiscale dynamical systems is of practical interest due to this wide range of applications; it is at the same time a challenging problem due to the interaction of the different scales. Our goal is to develop the theoretical framework for maximum likelihood estimation of the parameter \( \theta \in \Theta \subset \mathbb{R}^n \) in a family of \( m+(d-m) \) dimensional processes \((X^\varepsilon, Y^\varepsilon)_T = \{(X^\varepsilon_t, Y^\varepsilon_t)\}_{0 \leq t \leq T}\) satisfying stochastic differential equations

\[
\begin{align*}
    dX^\varepsilon_t &= c_\theta(X^\varepsilon_t, Y^\varepsilon_t)dt + \sqrt{\tau_2}(X^\varepsilon_t, Y^\varepsilon_t)dB_t, \\
    dY^\varepsilon_t &= \frac{1}{\delta}f(X^\varepsilon_t, Y^\varepsilon_t)dt + \frac{1}{\sqrt{\delta}}(X^\varepsilon_t, Y^\varepsilon_t)dW_t + \frac{1}{\sqrt{\delta}}(X^\varepsilon_t, Y^\varepsilon_t)dW_t, \\
    X^\varepsilon_0 &= x_0 \in \mathcal{X} = \mathbb{R}^m, Y^\varepsilon_0 = y_0 \in \mathcal{X} = \mathbb{R}^{d-m},
\end{align*}
\]

1

arXiv:1504.07645v1 [math.PR] 28 Apr 2015
where $W_t$ and $B_t$ are independent Wiener processes, and $\varepsilon = (\epsilon, \delta)$ is a pair of small positive parameters $0 < \epsilon \ll 1, 0 < \delta \ll 1$. It is important to remember that throughout this paper, $\varepsilon = (\epsilon, \delta) \in \mathbb{R}_+^2$; when we write $\varepsilon \to 0$, we mean $\epsilon + \delta \to 0$. Conditions on the coefficients are described in Conditions 1 and 2 in Section 3. A model such as (1) results, for example, when one has a dynamical system that is subject to periodicity or explicit compactness assumptions for the fast process are [3, 19].

In this paper, we exploit the relatively recent results of [20, 21] and tools from [25] to complete a series of averaging techniques under periodicity or explicit compactness assumptions for the fast process are [3, 19].

A key technique in the theory of averaging for SDE models is that of exploiting bounds on the solutions of Poisson equations involving the differential operators associated to the SDEs. Classical manuscripts on statistical inference in SDE models such as (1) when $\varepsilon$ is compact - both processes range in full Euclidean spaces. Lastly, we emphasize that the MLE uses a single appendix). To the best of our knowledge, this is the first paper that proves both asymptotic consistency and asymptotic normality of the MLE in the noncompact case.

The rest of this paper is structured as follows. In Section 2 we discuss generalities on the MLE and on statistical inference in SDE models such as (1) when $\varepsilon$ is compact - both processes range in full Euclidean spaces. Lastly, we emphasize that the MLE uses a single appendix). To the best of our knowledge, this is the first paper that proves both asymptotic consistency and asymptotic normality of the MLE in the noncompact case.

Let us conclude the introduction with a bit of methodology. In considering the fast process we must pass to the limit as $\delta \to 0$. The limiting behavior of the system is then described by the theory of averaging. A key technique in the theory of averaging for SDE models is that of exploiting bounds on the solutions of Poisson equations involving the differential operators associated to the SDEs. Classical manuscripts on averaging techniques under periodicity or explicit compactness assumptions for the fast process are [3, 19]. In this paper, we exploit the relatively recent results of [20, 21] and tools from [25] to complete a series of delicate analytic estimates, thereby extending the theory to a fairly general model in the noncompact case.

Maximum likelihood estimation in multiscale models with noise of order $O(1)$ has been studied in [1, 2, 13, 17, 18]. In particular, the authors of [13] study semiparametric estimation with linear dependence in $\theta$, and the authors of [17, 18] prove consistency of the MLE based on the limit of the slow process $X^\varepsilon$ in the model (1) with $\epsilon = 1$ under the assumptions that all coefficients are uniformly bounded and that the fast process is ergodic. It is important to point out that the situation in the present work is different in that the diffusion coefficient $\sqrt{\varepsilon \sigma}$ vanishes in the limit as $\epsilon \to 0$. Hence, as made precise by Theorem 1, the limiting behavior of $X^\varepsilon$ in our model is described by the solution of an ODE and not an SDE. Consequently, the limit has no well-defined likelihood function and thus one works directly with the likelihood function of the multiscale model. Besides [17, 18], perhaps most closely related to the present work is [24], wherein the authors prove asymptotic consistency and asymptotic normality of the MLE in the special case of the model (1) where $Y^\varepsilon_t = X^\varepsilon_t / \delta$ for all $t$, under the assumptions that all coefficients are uniformly bounded and periodic in $y$; such assumptions, it is important to point out, greatly simplify the analysis relative to the present work. More generally, we refer the interested reader to [4, 15, 22] for a treatment of the problem of statistical inference in SDE models such as (1) when $\epsilon = \delta = 1$, i.e., noise of order $O(1)$ without multiple scales, and to [14] for when $0 < \epsilon \ll 1$ but $\delta = 1$, i.e., small noise without multiple scales.

In light of the existing literature, the contribution of this paper is threefold. Firstly, in the averaging regime, we prove not only that the maximum likelihood estimator is asymptotically consistent (i.e., that it consistently estimates the true parameter), but also that it is asymptotically normal - we derive a central limit theorem that identifies precisely the limiting variance of the estimator (i.e., the Fisher information matrix). Secondly, we allow unbounded coefficients in the equation for the slow process $X^\varepsilon$ and assume neither periodicity nor that the fast process $Y^\varepsilon$ is compact - both processes range in full Euclidean spaces. Although $c_\theta$ and $\sigma$ are still subject to growth conditions and $f$ is subject to a recurrence condition to guarantee ergodicity of $Y^\varepsilon$, certain necessary bounds which would otherwise be straightforward to obtain now demand a delicate series of analytic estimates. Thirdly, at a technical level, in the course of the proof we obtain exponential bounds and ergodic theorems that may be of independent interest (see Section 9, the appendix). To the best of our knowledge, this is the first paper that proves both asymptotic consistency and asymptotic normality of the MLE in the context of small-noise perturbed multiscale models with general coefficients and taking values in the full Euclidean space. Lastly, we emphasize that the MLE uses a single time-series of the data, which is usually the available form of the data.

Let us conclude the introduction with a bit of methodology. In considering the fast process we must pass to the limit as $\delta \to 0$. The limiting behavior of the system is then described by the theory of averaging. A key technique in the theory of averaging for SDE models is that of exploiting bounds on the solutions of Poisson equations involving the differential operators associated to the SDEs. Classical manuscripts on averaging techniques under periodicity or explicit compactness assumptions for the fast process are [3, 19]. In this paper, we exploit the relatively recent results of [20, 21] and tools from [25] to complete a series of delicate analytic estimates, thereby extending the theory to a fairly general model in the noncompact case.

The rest of this paper is structured as follows. In Section 2 we discuss generalities on the MLE and on our results. In Section 3 we specify basic conditions on the coefficients in our model and describe precisely
the limiting behavior of the slow process in Theorem 1, a proof of which may be found in the appendix. In Section 4, we present our asymptotic consistency theorem, Theorem 3. In Section 5, we present our asymptotic normality theorem, Theorem 4. In Section 6, data from numerical simulations are provided to supplement and illustrate the theory. In Section 7, we sketch a straightforward extension of our results to models with a greater plurality of time-scales. Section 8 is reserved for acknowledgements. Finally, in Section 9, the appendix, we state and prove the auxiliary theorems and lemmas to which we appeal in the rest of the paper.

2 The Maximum Likelihood Estimator

In this paper we suppose that we observe continuous data; that is, we observe a continuous trajectory \((X, Y)_T = \{(x_t, y_t)\}_{0 \leq t \leq T}\) of the model (1).

The MLE \(\hat{\theta}^\epsilon\) is defined as the maximizer of the likelihood function, which is nothing else than the Girsanov density of the measure induced by the model (1) with respect to the measure induced by the same model with \(c_\theta \equiv 0\), see for example [14, 15]. Denoting these measures respectively by \(P_\theta\) and \(P_0\), we have equivalently

\[
\hat{\theta}^\epsilon = \arg \max_{\theta \in \Theta} Z^\epsilon_{\theta, T},
\]

where \(Z^\epsilon_{\theta, T}\) is the rescaled log-likelihood defined by the equation

\[
\frac{dP_\theta}{dP_0}((X, Y)_T) = e^{\frac{1}{2} Z^\epsilon_{\theta, T}((X, Y)_T)}.
\]

Let us see what \(Z^\epsilon_{\theta, T}((X, Y)_T)\) looks like in our case. Let us for brevity use \((a, b)_\kappa\) to denote the weighted inner product \((\kappa a, \kappa b) = a^T \kappa^T \kappa b\), where

\[
\kappa^T \kappa = (\sigma \sigma^T)^{-1} \sigma (I + \tau_1^T (\tau_2 \tau_2^T)^{-1} \tau_1) \sigma (\sigma \sigma^T)^{-1}.
\]

With this notation, the rescaled log-likelihood function is

\[
Z^\epsilon_{\theta, T}((X, Y)_T) = \int_0^T (c_\theta, dx_t)_\kappa(x_t, y_t) - \frac{1}{2} \int_0^T ||c_\theta||^2_\kappa(x_t, y_t)dt + (\epsilon/\delta)^{1/2} \int_0^T \langle \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta, (\tau_2 \tau_2^T)^{-1} f \rangle(x_t, y_t)dt - (\epsilon\delta)^{1/2} \int_0^T \langle \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta, (\tau_2 \tau_2^T)^{-1} dy_t \rangle(x_t, y_t).
\]

Under the assumptions of Section 3, we prove asymptotic consistency and asymptotic normality of \(\hat{\theta}^\epsilon\) in Sections 4 and 5. However, while (2) and (3) are indeed the true MLE and likelihood function in terms of which we state and prove our results in Theorems 2, 3, and 4, it is important to observe that the theory readily accommodates some degree of flexibility. In particular, the last two integral terms in (4) are problematic in practice, where exact values of \(\epsilon\) and \(\delta\) are often difficult to obtain. Estimation of \(\epsilon, \delta \ll 1\) in fact is a difficult problem, and one would like to have estimation procedures for \(\theta\) that avoid having to use \(\epsilon\) and \(\delta\), e.g., [17, 23]. In this connection, notice that the asymptotic variance of Theorem 4, i.e., the Fisher information matrix, depends only on \(c_\theta\) and on the matrix \(\kappa\). This is a hint that, asymptotically as \(\epsilon \downarrow 0\), the last two problematic terms in (4) are insignificant for statistical inference.

One can in fact avoid the problematic terms by defining the quasi-MLE instead as

\[
\hat{\theta} = \arg \max_{\theta \in \Theta} \hat{Z}_{\theta, T},
\]

3
where $\tilde{Z}_{\theta,T}$ is the quasi rescaled log-likelihood function
\[
\tilde{Z}_{\theta,T}((X, Y)_T) = \int_0^T \langle c_\theta, dx_i \rangle \kappa(x_i, y_i) - \frac{1}{2} \int_0^T ||c_\theta||_2^2 (x_i, y_i) dt. \tag{6}
\]

One can fairly easily see that the asymptotic consistency results extend to the quasi-MLE as follows. Considering firstly that, as in Section 4, $Z$ where $\tilde{Z}$ we obtain immediately that $\leq \gamma$ and secondly that by Lemma 8, for any $1 - \epsilon$ parameters vanish and the true MLE and quasi-MLE coincide. The first example in Section 6 is an illustration of this $(4)$ is in a simple form from the beginning. In this case $(4)$ and $(6)$ are the same, i.e. the problematic terms $\kappa$ extend at once to the quasi-MLE $\tilde{Z}$ the numerical results that the gap between the empirical variances closes as $\epsilon$ gets smaller. These conclusions are of course to be expected. More details may be found in Remark 1.

Thus, a cursory inspection of Theorems 2 and 3 is enough to see that the asymptotic consistency results extend to the quasi-MLE $\tilde{Z}$ that is, for practical purposes one can use $(5)$ and $(6)$ instead of $(2)$ and $(4)$. Since the true log-likelihood is given by $(4)$ (see \[14, 15, 22\]), we state and prove the theorems for $\tilde{Z}$ and we exploit averaging results and related estimates from \[20, 21\] to derive a deterministic $L^p$ limit. Precisely, we prove that
\[
\lim_{\epsilon \to 0} E \sup_{0 \leq t \leq T, \theta \in \Theta} |Z_{\theta,L}((X^\epsilon, Y^\epsilon)_t) - \tilde{Z}_{\theta,L}((X^\epsilon, Y^\epsilon)_t)|^p = 0.
\]

Direct theoretical analysis of the properties of the estimator is complicated by the presence of the small parameters $\epsilon$ and $\delta$ in the expression for the rescaled log-likelihood evaluated at random data. To proceed, we exploit averaging results and related estimates from \[20, 21\] to derive a deterministic $L^p$ limit. Precisely, we prove that
\[
\lim_{\epsilon \to 0} E \sup_{\theta \in \Theta} \left| Z_{\theta,T}((X^\epsilon, Y^\epsilon)_T) - \tilde{Z}_{\theta,0,T}((X)_T) \right|^p = \lim_{\epsilon \to 0} E \sup_{\theta \in \Theta} \left| Z_{\theta,T}((X^\epsilon, Y^\epsilon)_T) - \tilde{Z}_{\theta,0,T}((X)_T) \right|^p = 0,
\]

where $\tilde{Z}$ is an appropriately defined rescaled limiting log-likelihood function and $X$ is the limit of the slow process as described in Theorem 1. It is by exploiting these limits that we are able to prove asymptotic consistency and asymptotic normality, and moreover to identify precisely the limiting variance.

We would like to emphasize that it is more difficult to establish the necessary limits once one does away with uniform bounds, periodicity, and compactness. In many instances in the present work, bounds that would otherwise be standard must be replaced with delicate estimates. These estimates essentially must be made using only polynomial-type growth conditions on the coefficients and the recurrence condition that we impose to guarantee ergodicity of the fast process.
3 Preliminaries and Assumptions

Throughout this paper we work with a canonical probability space \((\Omega, \mathcal{F}, P_\theta)\) equipped with a filtration \(\{\mathcal{F}_t\}_{0 \leq t \leq T}\) satisfying the usual conditions. That is, \(\mathcal{F}_t\) is right continuous and \(\mathcal{F}_0\) contains all \(P_\theta\)-negligible sets.

We impose the following regularity and growth conditions on the coefficients in the model (1). These conditions guarantee that (1) is well-posed and has a strong solution, and that the limit results that follow are valid.

**Condition 1.**

*Conditions on \(c_\theta\)*

1. \(\forall \theta \in \Theta, \exists K, q > 0; \sum_{j=0}^{2} \left| \frac{\partial^j}{\partial x^j} c_\theta(x, y) \right| \leq K(1 + |y|^q)\)

2. \(\forall \theta \in \Theta, c_\theta \in C^{2+\alpha}; \) namely, \(c_\theta\) has two continuous derivatives in \(x\), Hölder \(\alpha\) in \(y\) uniformly in \(x\)

3. \(\forall \theta \in \Theta, \frac{\partial^2}{\partial y^2} c_\theta(x, y) \in C(X, Y)\)

4. \(c_\theta(x, y)\) has two bounded derivatives in \(\theta\)

*Conditions on \(\sigma\)*

1. \(\forall N > 0, \exists C(N); \forall x_1, x_2 \in X, \forall y \in Y \text{ with } |y| \leq N, |\sigma(x_1, y) - \sigma(x_2, y)| \leq C(N)|x_1 - x_2|\)

2. \(\exists K > 0, q > 0, r \in [0, \frac{1}{2}) \); \(|\sigma(x, y)| \leq K(1 + |x|^r)(1 + |y|^q)\)

3. \(\sigma \sigma^T\) is uniformly nondegenerate

*Conditions on \(f, \tau_1, \tau_2\)*

1. \(f, \tau_1, \tau_2, \tau_2^T \in C^{2+\alpha}(X, Y)\); namely, these have two bounded derivatives in \(x\) and two derivatives in \(y\), and all derivatives \(\partial^i \partial^j_y f, \partial^i \partial^j_y \tau_1, \partial^i \partial^j_y \tau_2, 0 \leq i \leq 2, 0 \leq j \leq 2\) are Hölder \(\alpha\) in \(y\) uniformly in \(x\)

2. \(\tau_2 \tau_2^T\) is uniformly nondegenerate.

We further impose the following recurrence condition to guarantee ergodicity of the fast process \(Y^\varepsilon\).

**Condition 2.**

\[
\lim_{|y| \to \infty} \sup_{x \in X} f(x, y) \cdot y = -\infty.
\]

Along with the nondegeneracy of \(\tau_1 \tau_1^T + \tau_2 \tau_2^T\) in Condition 1, Condition 2 guarantees the existence, for each fixed \(\theta = \theta_0\) and \(x \in X\), of a unique invariant measure \(\mu_x\) associated with the operator

\[
\mathcal{L}_x = f(x, \cdot) \cdot \nabla_y + \frac{1}{2} (\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, \cdot) : \nabla^2_y.
\]

Under Conditions 1 and 2 we have the following averaging theorem, which is essentially the law of large numbers for the slow process \(X^\varepsilon\) in the model 1.
**Theorem 1.** Assume Conditions 1 and 2. For any fixed $\theta_0 \in \Theta$, initial condition $(x_0, y_0) \in X \times Y$, and $0 < p \leq \infty$,

$$\lim_{\varepsilon \to 0} E \sup_{0 \leq t \leq T} |X^\varepsilon_t - \bar{X}_t|^p = 0,$$

where $\bar{X}_t$ is the deterministic solution of the integral equation

$$\bar{X}_t = x_0 + \int_0^t \int_Y c_{\theta_0}(\bar{X}_s, y) \mu_{\bar{X}_s}(dy)ds.$$

A proof is given in the appendix.

We remark here that although convergence of $X^\varepsilon$ to $\bar{X}$ is generally expected (see [19, 20, 21, 24]), the statement of Theorem 1 is stronger than what is to be found in the literature of which we are aware - we prove that $X^\varepsilon \to \bar{X}$ in $L^p$ uniformly in $t \in [0, T]$, in the general case with the fast process taking values in the full Euclidean space. Of course, we use Conditions 1 and 2, as well as the fact that $\bar{X}$ is deterministic.

Theorem 1 is a key element in our proof of the asymptotic consistency of the MLE (Theorem 3), where it is used in conjunction with averaging over the invariant distribution of the fast process to establish a limiting value for the rescaled log-likelihood.

### 4 Asymptotic Consistency of the MLE

In this section we show that if an appropriate smoothness condition and identifiability condition are assumed, then the MLE $\hat{\theta}^\varepsilon$ for $\theta$ is asymptotically $L^p$-consistent as $\varepsilon \to 0$; that is, $\lim_{\varepsilon \to 0} E|\hat{\theta}^\varepsilon - \theta_0|^p \to 0$ (Theorem 3).

As mentioned in Section 2, direct theoretical analysis of the properties of the estimator is complicated by the presence of the small parameters $\varepsilon$ and $\delta$ in the expression for the rescaled log-likelihood evaluated at random data. To proceed, we derive a deterministic $L^p$ limit for this random variable.

Let us recall from Section 2 that we use $(a, b)_\kappa$ to denote the weighted inner product $(\kappa a, \kappa b) = a^T \kappa^T \kappa b$, where

$$\kappa^T \kappa = (\sigma \sigma^T)^{-1} \sigma (I + \tau_1^T (\tau_2 \tau_2^T)^{-1} \tau_1) \sigma^T (\sigma \sigma^T)^{-1}.$$ 

Let us further recall that with this notation, the rescaled log-likelihood function evaluated at random data can be written

$$Z_{\delta, T}((X^\varepsilon, Y^\varepsilon)_T) = \int_0^T \langle c_\theta, dX^\varepsilon_t \rangle_\kappa (X^\varepsilon_t, Y^\varepsilon_t) - \frac{1}{2} \int_0^T ||c_\theta||^2_\kappa (X^\varepsilon_t, Y^\varepsilon_t) dt$$

$$- \sqrt{\varepsilon} \int_0^T (\tau_1 \sigma (\sigma \sigma^T)^{-1} c_\theta, (\tau_2 \tau_2^T)^{-1} (\tau_1 dW_t + \tau_2 dB_t)) (X^\varepsilon_t, Y^\varepsilon_t).$$

This suggests the following limit as $\varepsilon \to 0$.

**Definition 1.** Fix $\theta_0 \in \Theta$. For a trajectory $Z_T = \{z_t\}_{0 \leq t \leq T} \subset X$ and $\theta \in \Theta$, we define the limiting rescaled log-likelihood $Z_{\theta_0, \theta, T}$ by the equation

$$Z_{\theta_0, \theta_0, T}(Z_T) = \int_0^T \int_Y \langle c_\theta, c_{\theta_0} \rangle_\kappa (z_t, y) \mu_{z_t}(dy) dt - \frac{1}{2} \int_0^T \int_Y ||c_\theta||^2_\kappa (z_t, y) \mu_{z_t}(dy) dt.$$
Completing the square, we see that the limiting rescaled log-likelihood can also be expressed as

\[
\bar{Z}_{0,\theta_0,T}(Z_T) = \frac{1}{2} \int_0^T \int_y |c_{\theta_0}|^2(\bar{Z}_t, y) \mu_{\bar{X}_t}(dy) dt - \frac{1}{2} \int_0^T \int_y |c_0 - c_{\theta_0}|^2(\bar{Z}_t, y) \mu_{\bar{X}_t}(dy) dt.
\]

This makes it clear that \(\bar{Z}_{0,\theta_0,T}(Z_T)\) attains a maximum at \(\theta = \theta_0\).

We are now ready to state and prove Theorem 2, which establishes the deterministic \(L^p\) limit of the rescaled log-likelihood evaluated at random data. We shall impose the following smoothness condition on the matrix \(\kappa\).

**Condition 3.** (Smoothness Condition) \(\kappa \in C^{2,\alpha}(\mathcal{X}, \mathcal{Y})\); namely, \(\kappa\) has two continuous derivatives in \(x\), Hölder \(\alpha\) in \(y\) uniformly in \(x\).

**Theorem 2.** Let \(\theta_0\) be the true value of the parameter. Assume Conditions 1, 2, and 3. Then for any \(1 \leq p < \infty\),

\[
\lim_{\varepsilon \to 0} E \sup_{\theta \in \Theta} |Z_{\theta,T}((X^\varepsilon, Y^\varepsilon)_T) - \bar{Z}_{\theta,\theta_0,T}((\bar{X})_T)|^p = 0,
\]

where \(\bar{X}\) is the solution to the limiting ODE as in Theorem 1.

**Proof.** By Lipschitz dependence in \(\theta\) and the assumption that \(\Theta\) is totally bounded, it will suffice to show that for any fixed \(\theta \in \Theta\),

\[
\lim_{\varepsilon \to 0} E |Z_{0,T}((X^\varepsilon, Y^\varepsilon)_T) - \bar{Z}_{\theta,\theta_0,T}((\bar{X})_T)|^p = 0.
\]

On the one hand, we have almost surely

\[
Z_{\theta,T}((X^\varepsilon, Y^\varepsilon)_T) = \int_0^T \langle c_\theta, dX^\varepsilon_t \rangle \kappa(X^\varepsilon_t, Y^\varepsilon_t) - \frac{1}{2} \int_0^T |c_\theta|^2(X^\varepsilon_t, Y^\varepsilon_t) dt
\]

\[
- \sqrt{\epsilon} \int_0^T \langle \tau_1 \sigma^T(\sigma \sigma^T)^{-1}c_\theta, (\tau_2 \tau_2^T)^{-1}(\tau_1 dW_t + \tau_2 dB_t) \rangle (X^\varepsilon_t, Y^\varepsilon_t)
\]

\[
= \int_0^T \langle c_\theta, c_{\theta_0} \rangle \kappa(X^\varepsilon_t, Y^\varepsilon_t) dt - \frac{1}{2} \int_0^T |c_\theta|^2(X^\varepsilon_t, Y^\varepsilon_t) dt + \sqrt{\epsilon} \int_0^T \langle c_\theta, dW_t \rangle \kappa(X^\varepsilon_t, Y^\varepsilon_t)
\]

\[
- \sqrt{\epsilon} \int_0^T \langle \tau_1 \sigma^T(\sigma \sigma^T)^{-1}c_\theta, (\tau_2 \tau_2^T)^{-1}(\tau_1 dW_t + \tau_2 dB_t) \rangle (X^\varepsilon_t, Y^\varepsilon_t).
\]

On the other hand, by definition

\[
\bar{Z}_{0,\theta_0,T}((\bar{X})_T) = \int_0^T \int_Y \langle c_\theta, c_{\theta_0} \rangle \kappa(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt - \frac{1}{2} \int_0^T \int_Y |c_\theta|^2(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt.
\]
Combining these and applying the triangle inequality,

\[
|Z^\varepsilon_{\theta,0,T}((X^\varepsilon,Y^\varepsilon)_T) - \bar{Z}^\varepsilon_{\theta,0,T}((\bar{X})_T)| \leq \left| \int_0^T \langle c_\theta, c_{\theta_0}\rangle_N(X^\varepsilon_t, Y^\varepsilon_t) - \langle c_\theta, c_{\theta_0}\rangle_N(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt \right|
\]

\[
+ \frac{1}{2} \int_0^T \|c_\theta\|^2_N(X^\varepsilon_t, Y^\varepsilon_t) - \langle c_\theta, c_{\theta_0}\rangle_N(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt
\]

\[
+ \sqrt{\varepsilon} \int_0^T \langle c_\theta, dW_t\rangle_N(X^\varepsilon_t, Y^\varepsilon_t)
\]

\[
+ \sqrt{\varepsilon} \int_0^T \langle \tau_1 \sigma^T(\sigma^T)^{-1} c_\theta, (\tau_2 \tau_2^T)^{-1}(\tau_1 dW_t + \tau_2 dB_t) \rangle(X^\varepsilon_t, Y^\varepsilon_t)
\]

Hence, for some constant \(C_p < \infty\),

\[
|Z_{\theta,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T) - \bar{Z}_{\theta,0,T}((\bar{X})_T)| \leq C_p \left[ \int_0^T \langle c_\theta, c_{\theta_0}\rangle_N(X^\varepsilon_t, Y^\varepsilon_t) - \langle c_\theta, c_{\theta_0}\rangle_N(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt \right]^

\]

\[
+ \frac{C_p}{2^p} \int_0^T \|c_\theta\|^2_N(X^\varepsilon_t, Y^\varepsilon_t) - \langle c_\theta, c_{\theta_0}\rangle_N(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt
\]

\[
+ C_p \varepsilon^{1/2} \int_0^T \langle c_\theta, dW_t\rangle_N(X^\varepsilon_t, Y^\varepsilon_t)
\]

\[
+ C_p \varepsilon^{1/2} \int_0^T \langle \tau_1 \sigma^T(\sigma^T)^{-1} c_\theta, (\tau_2 \tau_2^T)^{-1}(\tau_1 dW_t + \tau_2 dB_t) \rangle(X^\varepsilon_t, Y^\varepsilon_t)
\]

It will suffice to show that the expectation of each term on the right tends to 0.

By Conditions 1 and 3 we may apply Corollary 1 (see Section 9.2) to obtain

\[
\lim_{\varepsilon \to 0} E \left[ C_p \left[ \int_0^T \langle c_\theta, c_{\theta_0}\rangle_N(X^\varepsilon_t, Y^\varepsilon_t) - \langle c_\theta, c_{\theta_0}\rangle_N(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt \right] \right] = 0
\]

and

\[
\lim_{\varepsilon \to 0} E \left[ \frac{C_p}{2^p} \left[ \int_0^T \|c_\theta\|^2_N(X^\varepsilon_t, Y^\varepsilon_t) - \langle c_\theta, c_{\theta_0}\rangle_N(\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt \right] \right] = 0,
\]

whereas by Lemma 8 (see Section 9.2 in the appendix)

\[
\lim_{\varepsilon \to 0} E \left[ C_p \varepsilon^{1/2} \int_0^T \langle c_\theta, dW_t\rangle_N(X^\varepsilon_t, Y^\varepsilon_t) \right] = 0
\]

and

\[
\lim_{\varepsilon \to 0} E \left[ C_p \varepsilon^{1/2} \int_0^T \langle \tau_1 \sigma^T(\sigma^T)^{-1} c_\theta, (\tau_2 \tau_2^T)^{-1}(\tau_1 dW_t + \tau_2 dB_t) \rangle(X^\varepsilon_t, Y^\varepsilon_t) \right] = 0.
\]

Thus, by applying Theorem 2, which ties the rescaled log-likelihood, which determines the MLE, to its limiting value. If this limiting value is sufficiently regular in its dependence on \(\theta\), we can use Theorem 2 to extend this regularity.
to the rescaled log-likelihood and thereby establish asymptotic consistency of the MLE. Precisely, we will assume the following identifiability condition. Here it is useful to recall that \( Z_{\theta,0,T} \) be the true value of the parameter.

**Condition 4. (Identifiability Condition)** For all \( \eta > 0 \),

\[
\sup_{|u| > \eta} (Z_{\theta_0+u,0,T}((X)_{T}) - Z_{\theta_0,0,T}((X)_{T})) \leq -\eta.
\]

We are now ready to state and prove our asymptotic consistency theorem.

**Theorem 3.** Let \( \hat{\theta}^\varepsilon = \arg \max_{\theta \in \Theta} Z_{\theta,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T) \) be the MLE and let \( \theta_0 \) be the true value of the parameter. Assume Conditions 1, 2, 3, and 4. Then for any \( 1 \leq p < \infty \),

\[
\lim_{\varepsilon \to 0} E_{\theta_0} |\hat{\theta}^\varepsilon - \theta_0|^p = 0.
\]

**Proof.** Let us suppress the subscripts \( \theta_0 \) and \( T \) in the notation for the rescaled log-likelihood and limiting rescaled log-likelihood; we write \( Z_{\theta}^\varepsilon \) for \( Z_{\theta,T}^\varepsilon \) and \( \bar{Z}_{\theta}^\varepsilon \) for \( Z_{\theta,0,T}^\varepsilon \).

For any \( \eta > 0 \) and any \( 1 \leq r < \infty \),

\[
P_{\theta_0} \left( |\hat{\theta}^\varepsilon - \theta_0| > \eta \right)
\leq P_{\theta_0} \left( \sup_{|u| > \eta} (Z_{\theta_0+u,0,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T) - Z_{\theta_0,0,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T)) \geq 0 \right)
\leq P_{\theta_0} \left( \sup_{|u| > \eta} ((Z_{\theta_0+u,0,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T) - Z_{\theta_0,0,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T)) - (Z_{\theta_0+u}((X)_{T}) - Z_{\theta_0}((X)_{T})) \geq 0 \right)
\geq - \sup_{|u| > \eta} (Z_{\theta_0+u}((X)_{T}) - Z_{\theta_0}((X)_{T}))
\leq P_{\theta_0} \left( \sup_{|u| > \eta} ((Z_{\theta_0+u,0,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T) - Z_{\theta_0,0,T}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T)) - (Z_{\theta_0+u}((X)_{T}) - Z_{\theta_0}((X)_{T})) \geq \frac{1}{2} \eta \right)
\leq P_{\theta_0} \left( \sup_{|u| > \eta} |Z_{\theta_0+u}((X^\varepsilon,Y^\varepsilon)_T) - \bar{Z}_{\theta_0}((X)_{T})| \geq \frac{1}{2} \eta \right) + P_{\theta_0} \left( |Z_{\theta_0}((X^\varepsilon,Y^\varepsilon)_T) - \bar{Z}_{\theta_0}((X)_{T})| \geq \frac{1}{2} \eta \right)
\leq 2^{r+1} \eta^r \cdot E_{\theta_0} \sup_{\theta \in \Theta} |Z_{\theta}^\varepsilon((X^\varepsilon,Y^\varepsilon)_T) - \bar{Z}_\theta((X)_{T})|^r,
\]

where the third inequality follows by Condition 4.

It follows by the Markov inequality and the convergence of Theorem 2 that for \( \varepsilon \) sufficiently small,

\[
P_{\theta_0} \left( |\hat{\theta}^\varepsilon - \theta_0| > \eta \right) \leq \frac{C(\varepsilon)}{|\eta|^r},
\]

9
where \( \lim_{\varepsilon \to 0} C(\varepsilon) = 0 \). Hence
\[
E|\hat{\theta}^\varepsilon - \theta_0|^p \leq \sum_{\eta=0}^\infty \left[ (\eta + 1)^p \cdot P_{\theta_0} \left( \eta < |\hat{\theta}^\varepsilon - \theta_0| \leq \eta + 1 \right) \right]
\leq \sum_{\eta=0}^\infty \left[ (\eta + 1)^p \cdot P_{\theta_0} \left( \eta < |\hat{\theta}^\varepsilon - \theta_0| \right) \right]
\leq \sum_{\eta=0}^\infty \left[ (\eta + 1)^p \cdot \frac{C(\varepsilon)}{|\eta|^r} \right].
\]
Choosing \( r \) large enough relative to \( p \), we are done.

5 Asymptotic Normality of the MLE

In this section we show that if appropriate conditions are assumed, then the MLE is asymptotically normal with converging moments (Theorem 4). We obtain this result by establishing the conditions of Theorem 1.6 in [14]. Having this description of the distribution of the error of the MLE allows one to translate the point estimate into a confidence interval, greatly increasing its practical value.

Sufficient conditions for the asymptotic normality of the MLE for stochastic models appear in the absence of multiple scales in [14, 15] and for the case \( Y_\varepsilon^t = X_\varepsilon^t/\delta \) for all \( t \), with bounded and periodic dependence of the coefficients in the second variable, in [24]. As far as we are aware, this paper contains the first such result for a small-noise multiscale model in the full space and with general dependence of the coefficients on the slow and fast processes. The lack of such assumptions amounts to a lack of compactness, which in turn means that a number of delicate estimates must be obtained in order to guarantee that certain quantities of interest remain bounded. With a compact state space or periodic dependence, such quantities are automatically bounded, but in the present case one must expend some effort to derive the fact.

We now define the Fisher information and state the conditions that we will need to assume. Recall that
\[
\kappa^T\kappa = \left( \sigma\sigma^T \right)^{-1} \sigma (I + \tau_1^T(\tau_2^T)^{-1}\tau_1)\sigma^T(\sigma\sigma^T)^{-1},
\]
\( \kappa = \left( \kappa^T\kappa \right)^{1/2} \) and that \( \mu_x \) denotes the invariant measure.

**Definition 2.** We define the quantity \( q \) and the Fisher information matrix \( I(\theta) \) by the equations
\[
q(x,\theta) = \int_Y \| \nabla_\theta c_\theta \|^2_c (x,y) \mu_x(dy)
\]
\[
I(\theta) = \int_0^T q(\tilde{X}_t,\theta)dt.
\]

We shall impose the following nondegeneracy condition on the Fisher information matrix \( I(\theta) \).

**Condition 5.**
1. \( I(\theta) \) is positive definite uniformly in \( \theta \), i.e.,
\[
\exists c_0 > 0; c_0 \leq \inf_{\theta \in \Theta} \inf_{|\lambda| = 1} \langle I(\theta)\lambda,\lambda \rangle
\]
2. The vector process \( q^{1/2}(X_\varepsilon^t,\theta), t \in [0,T] \) is continuous in probability, uniformly on \( \theta \in \Theta \) in \( L^2[0,T] \) in \( \theta \) and in \( X \) at the point \( \theta = \theta_0 \).
We now state the three lemmas that establish the conditions of Theorem 1.6 in [14], followed by the asymptotic normality theorem (Theorem 4). Let us define the function
\[ \phi = \phi(\epsilon, \theta) = \sqrt{\epsilon} I^{-1/2}(\theta) \]
and normalized difference
\[ \Delta c_{\theta} = \frac{c_0 + \phi u - c_{\theta}}{\sqrt{\epsilon}}, \]
and set
\[ M_{\epsilon}(\theta, u) = \log \frac{dP_{\theta + \phi(\epsilon, \theta) u}}{dP_0}((X^\epsilon, Y^\epsilon)_T) \]
\[ = \int_0^T \langle \kappa \Delta c_{\theta}, d(W, B)_t(X^\epsilon_t, Y^\epsilon_t) - \frac{1}{2} \int_0^T ||\kappa \Delta c_{\theta}||^2(X^\epsilon_t, Y^\epsilon_t)dt. \]

**Lemma 1.** Assume Conditions 1, 2, 4, and 5. Then the family \( \{P_{\theta^\epsilon} : \theta \in \Theta\} \) is uniformly asymptotically normal with normalizing matrix \( \phi(\epsilon, \theta) = \sqrt{\epsilon} I^{-1/2}(\theta) \).

A proof is given in the appendix.

**Lemma 2.** Assume Conditions 1, 2, 4, and 5. Then there exist constants \( m > d/2 \) and \( C < \infty \) such that for every \( \epsilon \in (0, 1)^2 \) and compact \( \tilde{\Theta} \subset \Theta \),
\[ \sup_{\theta \in \tilde{\Theta}} \sup_{|u_2 - u_1| < r} |u_2 - u_1|^{-2m} E_{\theta} \left| e^{\frac{1}{2m} M_{\epsilon}(\theta, u_2)} - e^{\frac{1}{2m} M_{\epsilon}(\theta, u_1)} \right|^{2m} \leq C. \]

A proof is given in the appendix.

**Lemma 3.** Assume Conditions 1, 2, 3, 4, and 5. Then for any \( p \in (0, 1) \) and compact \( \tilde{\Theta} \subset \Theta \), there exists a function \( g_{\tilde{\Theta}, p}(|u|) \) with the property that for any \( u \in \mathbb{N} \),
\[ \lim_{u \to \infty} u^n e^{-g_{\tilde{\Theta}, p}(|u|)} = 0, \]
and which is also such that
\[ \sup_{\theta \in \tilde{\Theta}} E_{\theta} e^{pM_{\epsilon}(\theta, u)} \leq e^{-g_{\tilde{\Theta}, p}(|u|)}. \]

A proof is given in the appendix.

**Theorem 4.** Assume Conditions 1, 2, 3, 4, and 5. Then uniformly on compact subsets \( \tilde{\Theta} \subset \Theta \), in distribution under \( P_0 \),
\[ \frac{1}{\sqrt{\epsilon}} I^{1/2}(\theta) (\hat{\theta}^\epsilon - \theta) \Rightarrow N(0, I) \]
as \( \epsilon \to 0 \); moreover, \( \hat{\theta}^\epsilon \) has converging moments for all \( p > 0 \), i.e.,
\[ \lim_{\epsilon \to 0} \sup_{\theta \in \Theta} |E_{\theta} |I^{1/2}(\theta) (\hat{\theta}^\epsilon - \theta)|^p e^{-p/2} - E|Z|^p | = 0, \]
where \( Z \) is a standard \( N(0, I) \) random vector.

**Proof.** This follows by Theorem 1.6 in [14]. Lemmas 1, 2 and 3 establish the conditions of that theorem. □
6 Numerical Examples

In this section we present data from numerical simulations to supplement and illustrate the theory. Let us start by considering the system of equations

\[ \begin{align*}
\frac{dX_t^\varepsilon}{\sqrt{\varepsilon}} &= \theta_0 (\sin(X_t^\varepsilon))(Y_t^\varepsilon)^2 dt + \sqrt{\varepsilon} dW_t \\
\frac{dY_t^\varepsilon}{\sqrt{\varepsilon}} &= -\frac{1}{\delta}Y_t^\varepsilon dt + \frac{1}{\sqrt{\delta}} dB_t
\end{align*} \] (7)

for \( t \in [0, T = 1] \) with \( X_0^\varepsilon = Y_0^\varepsilon = 1 \in \mathbb{R} \).

Before discussing the simulations we remark that it is easy to see that the limit \( \bar{X} \) as in Theorem 1 for the slow process \( X^\varepsilon \) in the model (7) is the solution of

\[ \bar{X}_t = 1 + \frac{\theta_0}{2} \int_0^t \sin(\bar{X}_s) ds \]

and that the Fisher information used to compute the theoretical standard deviation in Theorem 4 is

\[ I(\theta) = \frac{3}{4} \int_0^1 \sin^2(\bar{X}_t) dt. \]

We simulate trajectories of the model (7) using an Euler scheme. Precisely, we use the approximation

\[ \begin{align*}
X_{t_{k+1}}^\varepsilon &= \theta_0 (\sin(X_{t_k}^\varepsilon))(Y_{t_k}^\varepsilon)^2(t_{k+1} - t_k) + \sqrt{\varepsilon}(W_{t_{k+1}} - W_{t_k}) \\
Y_{t_{k+1}}^\varepsilon &= -\frac{1}{\delta}Y_{t_k}^\varepsilon(t_{k+1} - t_k) + \frac{1}{\sqrt{\delta}}(B_{t_{k+1}} - B_{t_k})
\end{align*} \]

where \( k = 0, \ldots, n - 1 \), where \( n \) is the number of discrete time steps.

Let us fix \( \varepsilon = 0.1, \delta = 0.001, n = 10^6 \), and suppose that the discrete time steps are evenly spaced (i.e. \( t_{k+1} - t_k = \Delta t = 10^{-6} \)). We remark here that \( \delta \) influences the error of the numerical approximation done by the Euler scheme. In particular, in a similar fashion to [24], one can derive that the error of the discrete approximation of the dynamics scales like \( O(\Delta t/\delta) \), which implies that with the choice \( \delta = 10^{-3} \) and \( \Delta t = 10^{-6} \) one has an approximation error of the order \( 10^{-3} \).

The rescaled log-likelihood function for (7) given by (4) (or equivalently (6)) is

\[ Z_{\theta,T=1}(X,Y)=1 = \int_0^1 \theta \sin(x_t) y_t^2 dx_t - \frac{1}{2} \int_0^1 \theta^2 \sin^2(x_t) y_t^4 dt. \]

The MLE, therefore, is

\[ \hat{\theta}^\varepsilon = \frac{\int_0^1 \sin(x_t) y_t^2 dx_t}{\int_0^1 \sin^2(x_t) y_t^4 dt}. \]

Discretizing the MLE for our simulated trajectories we have the approximation

\[ \hat{\theta}^\varepsilon = \frac{\sum_{k=0}^{n-1} \sin(x_{t_k}) y_{t_k}^2 (x_{t_{k+1}} - x_{t_k})}{\sum_{k=0}^{n-1} \sin^2(x_{t_k}) y_{t_k}^4 \Delta t}. \]

Evidently, we are using a single time-series of the data to compute \( \hat{\theta}^\varepsilon \). We simulate the trajectories and MLE \( 10^4 \) times for each of \( \theta_0 = 2, 1, 0.1 \). Table 4 presents the mean MLE in each case along with confidence
intervals based on the empirical standard deviations and theoretical standard deviations based on Theorem 4. The histograms that follow in Figures 1-3 compare the empirical distribution of the MLE in each case with the theoretical density curve of Theorem 4.

In the table below and all tables that follow in this section, the theoretical standard deviation reported is that given by Theorem 4, i.e., \( \sqrt{\epsilon/I(\theta_0)} \).

| True Value of \( \theta_0 \) | Mean Estimator | 68% Confidence Interval | 95% Confidence Interval | Theoretical SD |
|-----------------------------|----------------|------------------------|------------------------|---------------|
| 2                           | 2.003          | (1.604, 2.403)         | (1.204, 2.803)         | 0.381         |
| 1                           | 0.975          | (0.559, 1.390)         | (0.143, 1.806)         | 0.391         |
| 0.1                         | 0.049          | (-0.418, 0.517)        | (-0.885, 0.985)        | 0.428         |

Table 1: Estimates of \( \theta_0 \) with empirical confidence intervals and theoretical standard deviations.

Let us look at another example to illustrate the case of dependent noise (i.e., \( \tau_1 \neq 0 \)) and the difference between the true MLE \( \hat{\theta}^\epsilon \) and the quasi-MLE \( \tilde{\theta}^\epsilon \). Instead of (7), consider

\[
dX_t^\epsilon = \theta_0 (\sin(X_t^\epsilon)) (Y_t^\epsilon)^2 dt + \sqrt{\epsilon} dW_t
\]

\[
dY_t^\epsilon = -\frac{1}{\delta} Y_t^\epsilon dt + \frac{1}{\sqrt{2\delta}} dW_t + \frac{1}{\sqrt{2\delta}} dB_t
\]

for \( t \in [0, T = 1] \) with \( X_0^\epsilon = Y_0^\epsilon = 1 \in \mathbb{R} \).

This time we consider both \( \epsilon = 0.1 \) and \( \epsilon = 0.01 \), but otherwise again fix \( \delta = 0.001 \) and simulate trajectories using an Euler scheme with \( n = 10^6 \) discrete time steps. We simulate the trajectories and discretized true MLE \( \hat{\theta}^\epsilon \) and quasi-MLE \( \tilde{\theta} \) \( 10^4 \) times for \( \theta_0 = 1 \). Table 2 below presents the mean MLEs.
with $\epsilon = 0.1$ and $\epsilon = 0.01$ along with confidence intervals based on the empirical standard deviations and theoretical standard deviations based on our asymptotic normality theorem. The histograms that follow in Figures 4-7 compare the empirical distribution of the MLE in each case with the theoretical density curve for the true MLE $\hat{\theta}^\epsilon$ of Theorem 1.

| $\epsilon$ | MLE   | Mean Estimator | 68% Confidence Interval | 95% Confidence Interval | Theoretical SD ($\hat{\theta}^\epsilon$) |
|------------|-------|----------------|-------------------------|-------------------------|----------------------------------------|
| 0.1        | $\theta^\epsilon$ | 0.985          | (0.688, 1.282)          | (0.391, 1.579)          | 0.276                                  |
| 0.1        | $\theta$       | 0.972          | (0.551, 1.393)          | (0.130, 1.814)          | -                                      |
| 0.01       | $\theta^\epsilon$ | 0.998          | (0.907, 1.090)          | (0.815, 1.182)          | 0.087                                  |
| 0.01       | $\theta$       | 0.996          | (0.876, 1.115)          | (0.757, 1.235)          | -                                      |

Table 2: Estimates of $\theta_0 = 1$ with empirical confidence intervals for the true MLE $\hat{\theta}^\epsilon$ and the quasi-MLE $\tilde{\theta}$, with $\epsilon = 0.1$ and $\epsilon = 0.01$, with $\delta = 0.001$.

**Remark 1.** A few remarks are in order here. We notice that when $\epsilon = 0.01$ the variance of the estimator is smaller than when $\epsilon = 0.1$ and as a result the confidence bounds are tighter. This conclusion is true for both estimators $\hat{\theta}^\epsilon$ and $\hat{\theta}$. This is of course consistent with our asymptotic theory. We also notice that the variance of $\tilde{\theta}$ is larger than the variance of $\hat{\theta}^\epsilon$. This is to be expected since $\hat{\theta}^\epsilon$ is the true MLE whereas $\tilde{\theta}$ is a quasi-MLE, based on omitting the problematic factors, as discussed in Section 2. However, Table 2 suggests that asymptotically this does not matter, since the gap in the empirical variances of $\hat{\theta}^\epsilon$ and $\tilde{\theta}$ closes as $\epsilon$ gets smaller.

We conclude the section with data from one more example to illustrate what happens when Condition 1 is not satisfied in that one has linear dependence of $c_0(x, y)$ on $x$. The point of this example is to illustrate
that the results of this paper are valid under weaker conditions on the coefficients than Condition 1, even though we can prove our theorems only assuming the validity of this condition. Consider
\[
\begin{align*}
\frac{dX_t}{\epsilon} &= \theta_0 X_t^\epsilon (Y_t^\epsilon)^2 dt + \sqrt{\epsilon} dW_t \\
\frac{dY_t}{\epsilon} &= -\frac{1}{\delta} Y_t^\epsilon dt + \frac{\sqrt{\delta}}{2\sqrt{\delta}} dW_t + \frac{1}{2\sqrt{\delta}} dB_t
\end{align*}
\]
for \( t \in [0, T = 1] \) with \( X_0^\epsilon = Y_0^\epsilon = 1 \in \mathbb{R} \).

As in the last example, we consider both \( \epsilon = 0.1 \) and \( \epsilon = 0.01 \), fix \( \delta = 0.001 \), and simulate trajectories using an Euler scheme with \( n = 10^6 \) discrete time steps. We simulate the trajectories and discretized true MLE \( \hat{\theta}^\epsilon \) and quasi-MLE \( \tilde{\theta}^{0.1} \) times for \( \theta_0 = 1 \). Table 3 presents the mean MLEs with \( \epsilon = 0.1 \) and \( \epsilon = 0.01 \) along with confidence intervals based on the empirical standard deviations and theoretical standard deviations based on our asymptotic normality theorem. The histograms that follow in Figures 8-9 compare the empirical distribution of the true MLE in each case with the theoretical density curve computed according to the formula of Theorem 4.

| \( \epsilon \) | MLE | Mean Estimator | 68% Confidence Interval | 95% Confidence Interval | Theoretical SD (\( \hat{\theta}^\epsilon \)) |
|---|---|---|---|---|---|
| 0.1 | \( \theta^\epsilon \) | 0.988 | (0.839, 1.136) | (0.690, 1.285) | 0.139 |
| 0.1 | \( \theta \) | 0.956 | (0.659, 1.253) | (0.361, 1.551) | - |
| 0.01 | \( \theta^\epsilon \) | 0.999 | (0.955, 1.043) | (0.911, 1.088) | 0.044 |
| 0.01 | \( \theta \) | 0.996 | (0.909, 1.082) | (0.822, 1.169) | - |

Table 3: Estimates of \( \theta_0 = 1 \) with empirical confidence intervals for the true MLE \( \hat{\theta}^\epsilon \) and the quasi-MLE \( \tilde{\theta} \), with \( \epsilon = 0.1 \) and \( \epsilon = 0.01 \), with \( \delta = 0.001 \).

![Figure 8: \( \hat{\theta}^\epsilon \), \( \epsilon = 0.1 \)](image1)

![Figure 9: \( \hat{\theta}^\epsilon \), \( \epsilon = 0.01 \)](image2)

**7 Possible Extension to models with multiple time-scales**

We remark here that the conclusions of this paper can be extended immediately to models with a greater plurality of time-scales. In particular, consider the model
\[
\begin{align*}
\frac{dX_t^\epsilon}{\epsilon} &= c_\theta(X_t^\epsilon, Y_t^\epsilon) + h_1(\epsilon, \delta) d_\theta(X_t^\epsilon, Y_t^\epsilon) dt + \sqrt{\epsilon} \sigma(X_t^\epsilon, Y_t^\epsilon) dW_t \\
\frac{dY_t^\epsilon}{\epsilon} &= \frac{1}{\delta} f(X_t^\epsilon, Y_t^\epsilon) + \frac{1}{h_2(\epsilon, \delta)} g(X_t^\epsilon, Y_t^\epsilon) dt + \frac{1}{\sqrt{\delta}} \tau_1(X_t^\epsilon, Y_t^\epsilon) dW_t + \frac{1}{\sqrt{\delta}} \tau_2(X_t^\epsilon, Y_t^\epsilon) dB_t,
\end{align*}
\]
where \( d_\theta \) satisfies the same conditions as \( c_\theta, g \) satisfies the same conditions as \( f \), \( \lim_{\epsilon \to 0, \delta \to 0} h_i(\epsilon, \delta) = 0 \) for \( i \in \{1, 2\} \), and \( \delta \leq h_2(\epsilon, \delta) \) for \( \epsilon, \delta \) sufficiently small.
If \( \lim_{\varepsilon \to 0} \frac{\delta}{\kappa_{(\varepsilon, \delta)}} = 0, \) then it is relatively easy to see that Theorems 1, 2, 3, and 4 hold exactly as they do for our original model (1).

If on the other hand \( \lim_{\varepsilon \to 0} \frac{\delta}{\kappa_{(\varepsilon, \delta)}} = \gamma \in (0, \infty), \) then replacing Condition 2 with

\[
\lim_{|y| \to \infty x \in X} \sup (\gamma f + g)(x, y) \cdot y = -\infty,
\]

Theorems 1, 2, 3, and 4 hold after replacing the invariant measure \( \mu_x(dy) \) with the invariant measure associated with the operator

\[
L_x^\gamma = (\gamma f + g)(x, \cdot) \cdot \nabla_y + \frac{1}{2} (\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, \cdot) : \nabla^2_y.
\]

The key assumption is the boundedness of \( g \), which allows us to bound the extra terms that will appear in the proofs in the same way as we do the terms involving \( f \).

8 Acknowledgements

The present work was partially supported by NSF grant DMS-1312124.

9 Appendix

We gather here in the appendix the technical results to which we appeal in the proofs of the main results of this paper. In Section 9.1 we establish a series of estimates and bounds that are used in the proofs of the main results of the paper and then prove Theorem 1. In Section 9.2 we establish intermediate lemmas that are used in the proof of asymptotic consistency of the MLE. In Section 9.3 we prove several further bounds of potential independent interest and then complete the proof of the asymptotic normality of the MLE.

9.1 Exponential Bound for the Fast Motion and Convergence of the Slow Motion

The first main result of this section is the existence of an absolute bound, uniform in \( \varepsilon \) sufficiently small, on the expectation of a random variable of the form \( e^{K \int_0^T |Y^\varepsilon_t|^q dt} \) (see Lemma 4). Expressions such as this one arise, for example, when one uses the Gronwall inequality to bound the deviation of the process \( X_\varepsilon \) from its small-\( \varepsilon \) limit \( \bar{X} \) in the proof of Theorem 1. We remark that if \( Y^\varepsilon \) takes values in a compact space, e.g. \( [24] \), then such a uniform bound is trivial.

The second main result of this section is a proof of Theorem 1 which makes use of the bound in Lemma 4. Let us therefore begin with Lemma 4.

Lemma 4. (Exponential Bound for Y) Let \( V_t \) be either of the Wiener processes \( W_t, B_t \). Assume Conditions 2 and 4. Then for any \( \theta \in \Theta \) and \( K, q > 0 \), there exists a positive constant \( \hat{K} \) such that for all \( \varepsilon \) sufficiently small,

\[
E_\theta \int_0^T e^{K |Y^\varepsilon_t|^q} dt \leq \hat{K};
\]

it follows that it is possible to choose \( \hat{K} \) such that

\[
E_\theta e^{K \int_0^T |Y^\varepsilon_t|^q dt} \leq \hat{K},
\]

\[
E_\theta e^{K \int_0^T |Y^\varepsilon_t|^q dV_t} \leq \hat{K}.
\]
Proof. We may assume without loss of generality that \( \tau_1 = 0 \) and that \( q > 2 \). Fixing \( X = \xi \), consider the process

\[
dY_t^\delta = \frac{1}{\delta} f(\xi, Y_t^\delta) dt + \frac{1}{\sqrt{\delta}} \tau_2(\xi, Y_t^\delta) dB_t;
\]

for any \( \delta \), the time-rescaled process \( Y_t^\delta \) is equal in distribution to \( Y_t^\xi \), where

\[
dY_t^\xi = f(\xi, Y_t^\xi) dt + \tau_2(\xi, Y_t^\xi) dB_t.
\]

For the \( q \)th absolute power of \( Y_t^\xi \) we have the equation

\[
d|Y_t^\xi|^q = (\nabla |y|^q)(Y_t^\xi) dY_t^\xi + \frac{1}{2} (\nabla^2 |y|^q)(Y_t^\xi) : \tau_2^T(\xi, Y_t^\xi) dt,
\]

where

\[
\nabla |y|^q = q |y|^{q-2} y,
\]

\[
(\nabla^2 |y|^q)_{ij} = q |y|^{q-4}((q-2)y_i y_j + \delta_{ij}|y|^2);
\]

in particular, the drift coefficient for the process \( |Y|^q \) is

\[
b^{\xi,q}(y) = |y|^{q-2}qy \cdot f(\xi, y) + \frac{1}{2} q |y|^{q-4}((q-2)y_i y_j + \delta_{ij}|y|^2) : \tau_2^T(\xi, y) = |y|^{q-2}(qy \cdot f(\xi, y) + O(1)).
\]

Condition 2 guarantees

\[
\lim_{|y| \to \infty} \left( f(\xi, y) \cdot y \right) = -\infty;
\]

this and with the expression for \( b^{\xi,q}(y) \) that we have just computed imply

\[
\lim \sup_{y \to \infty} \left( b^{\xi,q}(y) \cdot \frac{y}{|y|} \right) = -\infty.
\]

In other words, for any fixed \( \gamma > 0 \), there is a value of \( r \) such that for all \( |y| \geq r \), one has

\[
b^{\xi,q}(y) \cdot y \leq -\gamma |y|.
\]

Moreover, since the diffusion coefficient of \( |Y|^q \) is degenerate only at 0 and uniformly nondegenerate outside any neighborhood of 0, we have, for \( \beta_t = q |Y|^q \tau_2(\xi, Y_t^\xi) \), that the functional \( \int_0^t |\beta_s|^2 ds \) is increasing in \( t \) with probability one. Hence, in light of (15), Lemma 2 in [25] guarantees that for any \( \gamma > 0 \) there are constants \( R, C \) such that for all \( p \in (0, \gamma) \) and all \( 0 \leq t \leq T \),

\[
E e^{p \cdot \text{dist}(D_R, |Y_t^\xi|^q)} \leq C,
\]

where \( D_R \) is the compact disk of radius \( R \) and dist the Euclidean distance.

Hence, for some sufficiently large \( R \),

\[
E_\theta \int_0^T e^{K |Y_t^\delta|^q} dt = E \left( \delta \int_0^{T/\delta} e^{KT |Y_t^\xi|^q} dt \right) 
\]

\[
\leq \delta \int_0^{T/\delta} E e^{KTR + KT \cdot \text{dist}(D_R, |Y_t^\xi|^q)} dt
\]

\[
\leq \delta \int_0^{T/\delta} \sqrt{e^{2KTR} e^{2KT \cdot \text{dist}(D_R, |Y_t^\xi|^q)}} dt
\]

\[
\leq T \sqrt{e^{2KTR} C} = K.
\]
Due to the uniformity in \( x \in \mathcal{X} \) of Condition 2 and the uniform boundedness assumed in Condition 1, the last display establishes the validity of (9).

(10) can be reduced to (9) by Jensen’s inequality:

\[
E e^{K \int_0^T |Y_t^x|^q dt} \leq \frac{1}{T} \int_0^T E e^{K T |Y_t^x|^q} dt.
\]

(11) can be reduced to (10) by the argument

\[
\begin{align*}
E_{\theta} e^{K \int_0^T |Y_t^x|^q dV_t} & \leq E_{\theta} \left[ e^{K \int_0^T |Y_t^x|^q dV_t} - K^2 \int_0^T |Y_t^x|^{2q} dt + K^2 \int_0^T |Y_t^x|^{2q} dt \right] \\
& \leq \frac{1}{2} E_{\theta} e^{2K \int_0^T |Y_t^x|^q dV_t} - \frac{2K^2}{2} \int_0^T |Y_t^x|^{2q} dt + \frac{1}{2} E_{\theta} e^{K^2 \int_0^T |Y_t^x|^{2q} dt} \\
& \leq \frac{1}{2} + \frac{1}{2} \tilde{K},
\end{align*}
\]

where we used the submartingale property of the integrand \( e^{2K \int_0^T |Y_t^x|^q dV_t} - \frac{2K^2}{2} \int_0^T |Y_t^x|^{2q} dt \).

In addition to Lemma 4, a number of lemmas will be used in the proof of Theorem 1 that we present in this section; let us state and prove these lemmas now as Lemmas 5, 6, and 7.

**Lemma 5.** Assume Conditions 1 and 2. For any \( K, p > 0 \), there is a constant \( \tilde{K} \) such that uniformly in \( \varepsilon \) sufficiently small,

\[
E \sup_{0 \leq t \leq T} |X_t^\varepsilon|^p < \tilde{K},
\]

\[
E \sup_{0 \leq t \leq T} |X_t^\varepsilon - \tilde{X}_t|^p < \tilde{K}.
\]

**Proof.** As \( \tilde{X} \) is deterministic and continuous, the two statements are equivalent; we prove the first one. It is enough to prove the lemma for \( p \geq 2 \). Recall that

\[
X_t^\varepsilon = x_0 + \int_0^t c_0(X_s^\varepsilon, Y_s^\varepsilon) ds + \sqrt{\varepsilon} \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s;
\]

hence, for some constants \( C_j \) and \( q \),

\[
E \sup_{0 \leq t \leq T} |X_t^\varepsilon|^p \leq C_1 E \left( |x_0|^p + \int_0^T |c_0(X_t^\varepsilon, Y_t^\varepsilon)|^p dt + \left( \sup_{0 \leq t \leq T} \left| \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s \right| \right)^p \right) \\
\leq C_2 \left( |x_0|^p + E \int_0^T (1 + |Y_t^\varepsilon|^q)^p dt + E \int_0^T (1 + |Y_t^\varepsilon|^q)(1 + |X_t^\varepsilon|^{1/2})^p dt \right) \\
\leq C_2 \left( |x_0|^p + E \int_0^T (1 + |Y_t^\varepsilon|^q)^p dt + E \int_0^T (1 + |Y_t^\varepsilon|^q)^p C_p (1 + |X_t^\varepsilon|^{p/2}) dt \right) \\
\leq C_2 \left( |x_0|^p + E \int_0^T (1 + |Y_t^\varepsilon|^q)^p dt + E \int_0^T C_p (1 + |Y_t^\varepsilon|^q)^p dt \right. \\
\left. + \frac{1}{2} E \int_0^T C_p^2 (1 + |Y_t^\varepsilon|^q)^2 dt + \frac{1}{2} E \int_0^T |X_t^\varepsilon|^p dt \right) \\
\leq C_3 \left( 1 + \int_0^T E \sup_{0 \leq s \leq t} |X_s^\varepsilon|^p dt \right),
\]

where we appeal to the bound in Lemma 1 of [21] in passing to the last inequality. The proof is complete upon applying the Gronwall inequality. \( \square \)
Lemma 6. Assume Conditions 1 and 2. For any \( q > 0 \), there is a constant \( K \) such that for \( 0 \leq t \leq T \),
\[
\int_{\mathcal{Y}} (1 + |y|^q)|\mu_{X_t} - \mu_{\tilde{X}_t}|(dy) \leq K|X_t^\varepsilon - \tilde{X}_t|; \tag{17}
\]
in particular, if \( \varphi \) is a function of \( x \) and \( y \) satisfying \( |\varphi(x,y)| \leq K(1 + |y|^q) \) for some \( K,q > 0 \), then \( K \) can be chosen so that for \( 0 \leq t \leq T \),
\[
\int_{\mathcal{Y}} |\varphi(X_t^\varepsilon,y)||\mu_{X_t} - \mu_{\tilde{X}_t}|(dy) \leq K|X_t^\varepsilon - \tilde{X}_t|. \tag{18}
\]

Proof. By Theorem 1 in [21], the densities \( m_x \) of the measures \( \mu_x \) admit, for any \( p \), a constant \( C_p \) such that \( |\frac{d}{dx} m_x(y)| \leq \frac{C_p}{1 + |y|^p} \). Choosing \( p \) large enough that \( \int_{\mathcal{Y}} \frac{1 + |y|^q}{1 + |y|^p}dy < \infty \),
\[
\int_{\mathcal{Y}} (1 + |y|^q)|m_{X_t} - m_{\tilde{X}_t}|(dy) \leq \int_{\mathcal{Y}} C_p |X_t^\varepsilon - \tilde{X}_t| \frac{1 + |y|^q}{1 + |y|^p}dy \leq K|X_t^\varepsilon - \tilde{X}_t|,
\]
which is (17). Clearly, (18) follows immediately, which concludes the proof. \( \square \)

Lemma 7. Assume Conditions 1 and 2. For any \( q > 0 \), there is a constant \( K \) such that for \( 0 \leq t \leq T \),
\[
\int_{\mathcal{Y}} (1 + |y|^q)|\mu_{\tilde{X}_t}(dy) \leq K; \tag{19}
\]
in particular, if \( \varphi \) is a function of \( x \) and \( y \) satisfying \( |\frac{d}{dx} \varphi(x,y)| \leq K(1 + |y|^q) \) for some \( K,q > 0 \), then \( K \) can be chosen so that for \( 0 \leq t \leq T \),
\[
\int_{\mathcal{Y}} |\varphi(X_t^\varepsilon,y) - \varphi(\tilde{X}_t,y)||\mu_{\tilde{X}_t}(dy) \leq K|X_t^\varepsilon - \tilde{X}_t|. \tag{20}
\]

Proof. By Theorem 1 in [21], the densities \( m_x \) of the measures \( \mu_x \) admit, for any \( p \), a constant \( C_p \) such that \( |m_x(y)| \leq \frac{C_p}{1 + |y|^p} \). Choosing \( p \) large enough that \( \int_{\mathcal{Y}} \frac{1 + |y|^q}{1 + |y|^p}dy < \infty \), we see that \( \int_{\mathcal{Y}} (1 + |y|^q)|m_x(dy) \) takes finite values. (19) follows immediately by continuity in \( x \) and compactness of \( \{\tilde{X}_t\}_{0 \leq t \leq T} \subset \mathcal{X} \).

As for (20), we have
\[
\int_{\mathcal{Y}} |\varphi(X_t^\varepsilon,y) - \varphi(\tilde{X}_t,y)||\mu_{\tilde{X}_t}(dy) \leq \int_{\mathcal{Y}} K(1 + |y|^q)|X_t^\varepsilon - \tilde{X}_t||\mu_{\tilde{X}_t}(dy) \leq K|X_t^\varepsilon - \tilde{X}_t|.
\]

Proof of Theorem 4. It is enough to prove the theorem for \( p \geq 2 \). By Theorem 3 in [21], the equations
\[
\mathcal{L}_x \Phi(x,y) = c_0(x,y) - \bar{c}_0(x) \tag{21}
\]
\[
\int_{\mathcal{Y}} \Phi(x,y) \mu_x(dy) = 0 \tag{22}
\]
admit a unique solution \( \Phi \) in the class of functions that grow at most polynomially in \( |x| \) as \( x \to \infty \), and moreover, the unique solution is continuous in \( x \) and bounded by an expression of the form \( K(1 + |y|^q) \). Applying this solution to the multiscale process and expanding the Itô differential, we have
\[ |X_t^x - X_s^x|_p = \left| \int_0^t \left[ (c_\theta(X_s^x, Y_s^x) - \bar{c}_\theta(X_s^x))ds + (\bar{c}_\theta(X_s^x) - \bar{c}_\theta(X_s))ds + \sqrt{c}(X_s^x, Y_s^x)dW_s \right] \right|^p \]

\[
= \int_0^t \left[ \delta d\Phi - \delta \nabla x \Phi \cdot c_\theta(X_s^x, Y_s^x)ds - \delta \nabla x \Phi : \sigma \sigma^T(X_s^x, Y_s^x)ds \right. \\
- \sqrt{\delta \nabla y \nabla x \Phi : \sigma \sigma^T(X_s^x, Y_s^x)ds - \delta \nabla y \Phi \sigma(X_s^x, Y_s^x)dW_s - \sqrt{\delta \nabla y \Phi \tau_1(X_s^x, Y_s^x)dW_s} \right|^p \\
- \sqrt{\delta \nabla y \Phi \tau_2(X_s^x, Y_s^x)dB_s + (\bar{c}_\theta(X_s^x) - \bar{c}_\theta(X_s))ds + \sqrt{c}(X_s^x, Y_s^x)dW_s} \bigg]^p \\
\leq C_1 \left( \delta^p \int_0^t d\Phi \right|^p + \delta^p \int_0^t \nabla x \Phi \cdot c_\theta(X_s^x, Y_s^x)ds \bigg|^p + (\delta \sqrt{c})^{p \sup_{0 \leq s \leq t} \sup} \int_0^s \nabla x \Phi \sigma(X_u^x, Y_u^x)dW_u \bigg|^p \\
+ \sqrt{\delta \epsilon}^p \sup_{0 \leq s \leq t} \int_0^t \nabla y \Phi \tau_1(X_u^x, Y_u^x)dW_u \bigg|^p + \sqrt{\delta \epsilon}^p \sup_{0 \leq s \leq t} \int_0^t \nabla y \Phi \tau_2(X_u^x, Y_u^x)dB_u \bigg|^p \\
+ \sqrt{\delta \epsilon}^p \sup_{0 \leq s \leq t} \int_0^t \sigma(X_u^x, Y_u^x)dW_u \bigg|^p + \left( \frac{\delta \epsilon}{2} \right)^p \int_0^t \nabla x \Phi : \sigma \sigma^T(X_s^x, Y_s^x)ds \bigg|^p \bigg)
\]

where \( C_1 \) is a constant that does not depend on \( \epsilon \) or \( \delta \).

In the following, let \( K \) denote a generic positive constant and \( q \) a generic positive integer.

Let us focus on the Riemann integrals for a moment. Looking at the first two terms, each is bounded for \( \epsilon \) sufficiently small by an expression \( \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)ds \). Now consider the last three terms. We see that similarly, for \( \epsilon \) sufficiently small,

\[ \sqrt{\delta \epsilon} \int_0^t \nabla y \nabla x \Phi : \sigma \sigma^T(X_s^x, Y_s^x)ds \bigg|^p \leq \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)(1 + |X_s^x|^{1/2})^pds; \]

therefore, with new constants as necessary,

\[ \sqrt{\delta \epsilon} \int_0^t \nabla y \nabla x \Phi : \sigma \sigma^T(X_s^x, Y_s^x)ds \bigg|^p \leq \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)ds + \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)|X_s^x - \bar{X}_s|^pds. \]

Next, it is clear that also

\[ \left( \frac{\delta \epsilon}{2} \right)^p \int_0^t \nabla x \Phi : \sigma \sigma^T(X_s^x, Y_s^x)ds \bigg|^p \leq \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)ds \leq \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)|X_s^x - \bar{X}_s|^pds. \]

Finally, because \( \bar{c}_\theta \) satisfies the hypotheses of Lemmas 6 and 7,

\[ \left| \int_0^t (\bar{c}_\theta(X_s^x) - \bar{c}_\theta(\bar{X}_s))ds \right|^p \leq K \int_0^t |X_s^x - \bar{X}_s|^pds. \]

Taking all of this together, for \( \epsilon \) sufficiently small, perhaps with a new value of \( K \),

\[ |X_t^x - \bar{X}_t|^p \leq C_1 \left( \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s^x|^q)ds + \sqrt{\delta \epsilon} K \int_0^t (1 + |Y_s|^q)|X_s^x - \bar{X}_s|^pds + K \int_0^t |X_s^x - \bar{X}_s|^pds \right. \\
+ \left( \frac{\delta \sqrt{c}}{2} \right)^p \sup_{0 \leq s \leq t} \int_0^t \nabla x \Phi \sigma(X_u^x, Y_u^x)dW_u \bigg|^p + \sqrt{\delta \epsilon} \sup_{0 \leq s \leq t} \int_0^t \nabla y \Phi \tau_1(X_u^x, Y_u^x)dW_u \bigg|^p \\
+ \sqrt{\delta \epsilon} \sup_{0 \leq s \leq t} \int_0^t \nabla y \Phi \tau_2(X_u^x, Y_u^x)dW_u \bigg|^p + \left( \frac{\delta \epsilon}{2} \right)^p \int_0^t \sigma(X_u^x, Y_u^x)dW_u \bigg)^p \bigg)\].

20
Applying the Gronwall inequality to the full inequality and using the fact that the trajectory of $\bar{X}$ is compact, we obtain that for $\varepsilon$ sufficiently small,

$$|X^\varepsilon_t - \bar{X}_t|^p \leq M_t(\varepsilon)e^{\sqrt{C^p}K \int_0^T (1+|Y^\varepsilon_t|^q) \, ds},$$

where by definition

$$M_t(\varepsilon) = \left( \sqrt{\varepsilon C} K \int_0^t (1 + |Y^\varepsilon_s|^q) + \left( \sqrt{\varepsilon} \right)^p \sup_{0 \leq s \leq t} \left| \int_0^s \nabla_x \Phi(x_s, Y^\varepsilon_s) \, dW_u \right|^p \right)^{1/2}

+ \sup_{0 \leq s \leq t} \left| \int_0^s \nabla_y \Phi \tau_1(x_s, Y^\varepsilon_s) \, dW_u \right|^p + \sqrt{C} \sup_{0 \leq s \leq t} \left| \int_0^s \nabla_x \Phi \tau_2(x_s, Y^\varepsilon_s) \, dB_u \right|^p

+ \varepsilon t \left( \epsilon \right)$$

Therefore, for conjugate exponents $p_1, p_2$, uniformly in $\varepsilon$ sufficiently small,

$$E \sup_{0 \leq t \leq T} |X^\varepsilon_t - \bar{X}_t|^p \leq EM_T(\varepsilon)e^{\sqrt{C^p}K \int_0^T (1+|Y^\varepsilon_t|^q) \, ds}

\leq \left( E(M_T(\varepsilon))^{p_1} \right)^{1/p_1} \left( Ee^{\sqrt{C^p}K \int_0^T (1+|Y^\varepsilon_t|^q) \, ds} \right)^{1/p_2}

\leq K_1 \left( E(M_T(\varepsilon))^{p_1} \right)^{1/p_1},$$

for some constant $K_1$ by Lemma 4.

Let us examine the $p_1$ power of second factor on the right-hand side. The idea is to bound the power of the sum by the sum of the powers and apply the Burkholder-Davis-Gundy inequality. Precisely, for conjugate exponents $p_3, p_4$ and with new $C(\varepsilon) \to 0$ as $\varepsilon \to 0$ in each line as necessary,

$$E(M_T(\varepsilon))^{p_1} \leq C(\varepsilon) \left( K_2 + K_3 E \int_0^T (1 + |Y^\varepsilon_t|^q)(1 + |X^\varepsilon_t - \bar{X}_t|^\frac{p_2}{p_4} \, dt \right)

\leq C(\varepsilon) \left( K_4 + K_3 E \int_0^T (1 + |Y^\varepsilon_t|^q)|X^\varepsilon_t - \bar{X}_t|^\frac{p_2}{p_4} \, dt \right)

\leq C(\varepsilon) \left( 1 + \left( E \int_0^T (1 + |Y^\varepsilon_t|^q) \, dt \right)^{1/p_3} \left( E \int_0^T |X^\varepsilon_t - \bar{X}_t|^\frac{p_2p_4}{p_4} \, dt \right)^{1/p_4} \right)

\leq C(\varepsilon) \left( 1 + \left( \sup_{0 \leq t \leq T} |X^\varepsilon_t - \bar{X}_t|^\frac{p_2p_4}{p_4} \right)^{1/p_4} \right)

\leq C(\varepsilon),$$

where we appeal to the bound in Lemma 1 of [21] in passing from line 3 to line 4, and where the final inequality follows from the fact that $E \sup_{0 \leq t \leq T} |X^\varepsilon_t - \bar{X}_t|^\frac{p_2p_4}{p_4}$ is bounded uniformly as $\varepsilon \to 0$ by Lemma 5.

Immediately, we get

$$E \sup_{0 \leq t \leq T} |X^\varepsilon_t - \bar{X}_t|^p \leq K_1 \left( C(\varepsilon) \right)^{1/p_1},$$

hence $\lim_{\varepsilon \to 0} E \sup_{0 \leq t \leq T} |X^\varepsilon_t - \bar{X}_t|^p = 0.$
9.2 Lemmas for Asymptotic Consistency of the MLE

The Lemmas we prove in this section are used to tie the rescaled log-likelihood to its limiting value (see Theorem [2]), which is a stepping stone to establishing asymptotic consistency of the MLE.

**Lemma 8.** Let \( V_t \) be either of the Wiener processes \( W_t, B_t \). Assume Conditions 1 and 2. Let \( \phi \) be a function of \( x, y \) satisfying \( |\phi(x, y)| \leq K(1 + |y|^q)(1 + |x|^r) \) for some fixed positive constants \( K, q, r \). Then for any fixed \( \theta \in \Theta, \eta > 0, \) and \( 1 \leq p < \infty \), there is a constant \( \tilde{K} \) such that for \( \varepsilon \) sufficiently small,

\[
E \sup_{0 \leq t \leq T} \left| \int_0^t \phi(X^\varepsilon_s, Y^\varepsilon_s) dV_s \right|^p \leq \tilde{K};
\]

in particular, if \( C(\varepsilon) \) is any function such that \( \lim_{\varepsilon \to 0} C(\varepsilon) = 0 \), then

\[
\lim_{\varepsilon \to 0} E \sup_{0 \leq t \leq T} \left| C(\varepsilon) \int_0^t \phi(X^\varepsilon_s, Y^\varepsilon_s) dV_s \right|^p = 0.
\]

**Proof.** This is an application of the Burkholder-Davis-Gundy inequality. For some constants \( C_1, C_2, \)

\[
E \sup_{0 \leq t \leq T} \left| \int_0^t \phi(X^\varepsilon_s, Y^\varepsilon_s) dV_s \right|^p \leq C_p E \left( \int_0^T |\phi(X^\varepsilon_t, Y^\varepsilon_t)|^2 dt \right)^{p/2}
\]

\[
\leq C_p K E \left( \int_0^T (1 + |Y^\varepsilon|^q)^2 (1 + |X^\varepsilon|^r)^2 dt \right)^{p/2}
\]

\[
\leq C_p K \sqrt{E \int_0^T (1 + |Y^\varepsilon|^q)^2 (1 + |X^\varepsilon|^r)^2 dt}^p
\]

\[
\leq C_p C_2 \sqrt{T^{p-1} E \int_0^T (1 + |Y^\varepsilon|^q)^2 (1 + |X^\varepsilon| - \bar{X}_t|^r)^2 dt}^p
\]

\[
\leq \tilde{K}
\]

for \( \varepsilon \) sufficiently small, where we appeal to the Hölder inequality, the bound in Lemma 1 of [21], and Theorem 1.

**Lemma 9.** Assume Conditions 1 and 2. Let \( \phi \) be a function of \( x, y \) such that

1. \( \exists K, q > 0; \sum_{j=0}^2 |\frac{\partial^j}{\partial x^j} \phi(x, y)| \leq K(1 + |y|^q) \)

2. \( \phi \in C^{2,\alpha}(\mathcal{X}, \mathcal{Y}); \) namely, \( \phi \) has two continuous derivatives in \( x \), Hölder \( \alpha \) in \( y \) uniformly in \( x \).

Then for any fixed \( \theta \in \Theta \) and any \( 1 \leq p < \infty \),

\[
\lim_{\varepsilon \to 0} E \left| \int_0^T \phi(X^\varepsilon_t, Y^\varepsilon_t) - \int_{\mathcal{Y}} \phi(X^\varepsilon_t, \tilde{y}) \mu_{X^\varepsilon_t}(dy) dt \right|^p = 0.
\]

**Proof.** By Theorem 3 in [21], the equations

\[
\mathcal{L}_x \Phi(x, y) = \varphi(x, y) - \int_{\mathcal{Y}} \varphi(x, \tilde{y}) \mu_x(dy)
\]

\[
\int_{\mathcal{Y}} \Phi(x, y) \mu_x(dy) = 0
\]
admit a unique solution $\Phi$ in the class of functions that grow at most polynomially in $|x|$ as $x \to \infty$, and moreover, the unique solution is continuous in $x$ and bounded by an expression of the form $K(1 + |y|^q)$. Applying this solution to the multiscale process and expanding the Itô differential, we have, suppressing the arguments $(X^\varepsilon_t, Y^\varepsilon_t)$,

$$
\mathcal{L}_{X^\varepsilon_t} \Phi dt = \delta d\Phi - \delta \nabla_x \Phi \cdot c_\theta dt - \frac{\delta}{2} \nabla^2_x \Phi : \sigma \sigma^T \phi \tau dt - \sqrt{\delta} \nabla_y \Phi \tau \phi dW_t - \sqrt{\delta} \nabla_y \Phi \tau_1 dB_t,
$$

In any case, by the same argument as in the proof of Theorem 3 in [21], all of the derivatives of $\Phi$ that appear are continuous in $x$ and bounded by expressions of the form $K(1 + |y|^q)$. Hence, the general term on the right hand side of (22) is either

$$
C(\varepsilon) \psi(X^\varepsilon_t, Y^\varepsilon_t) dt
$$

or else

$$
C(\varepsilon) \psi(X^\varepsilon_t, Y^\varepsilon_t) dV_t,
$$

where $|\psi(X^\varepsilon_t, Y^\varepsilon_t)| \leq K(1 + |Y^\varepsilon_t|^q)(1 + |X^\varepsilon_t|)$ for some $K, q > 0$, $V_t$ is one of the Wiener processes $W_t, B_t$, and $C(\varepsilon) \to 0$ as $\varepsilon \to 0$.

For (23), the inequalities in the proof of Lemma 8 show that

$$
\lim_{\varepsilon \to 0} E \left| \int_0^T C(\varepsilon) \psi(X^\varepsilon_t, Y^\varepsilon_t) dt \right|^p \leq \lim_{\varepsilon \to 0} E \int_0^T |C(\varepsilon) \psi(X^\varepsilon_t, Y^\varepsilon_t)|^p dt,
$$

while for (24), Lemma 8 itself shows that

$$
\lim_{\varepsilon \to 0} E \left| \int_0^T C(\varepsilon) \psi(X^\varepsilon_t, Y^\varepsilon_t) dV_t \right|^p = 0.
$$

The proof is complete upon noting that the $p$th absolute power of a sum is bounded by a constant times the sum of the $p$th absolute powers of the terms. \hfill \Box

**Corollary 1.** Assume Conditions 1 and 2. Let $\varphi$ be a function of $x$ and $y$ such that

1. $\exists K, q > 0; \sum_{j=0}^2 |\frac{\partial^j}{\partial x^j} \varphi(x, y)| \leq K(1 + |y|^q)$

2. $\varphi \in C^{2,\alpha}(X, \mathcal{Y})$; namely, $\varphi$ has two continuous derivatives in $x$, Hölder $\alpha$ in $y$ uniformly in $x$.

Then for any fixed $\theta \in \Theta$ and any $1 \leq p \leq \infty$,

$$
\lim_{\varepsilon \to 0} E \left| \int_0^T \varphi(X^\varepsilon_t, Y^\varepsilon_t) - \int_y \varphi(\bar{X}_t, y) \mu(\bar{X}_t) dy dt \right|^p = 0.
$$
Proof. By the triangle inequality,

\[
\left| \int_0^T \varphi(X^\varepsilon_t, Y^\varepsilon_t) - \int_Y \varphi(\bar{X}_t, y) \mu_{X^\varepsilon_t}(dy) dt \right| \leq \left| \int_0^T \varphi(X^\varepsilon_t, Y^\varepsilon_t) - \int_Y \varphi(X^\varepsilon_t, y) \mu_{X^\varepsilon_t}(dy) dt \right| \\
+ \left| \int_0^T \int_Y \varphi(X^\varepsilon_t, y) (\mu_{X^\varepsilon_t} - \mu_{\bar{X}_t})(dy) dt \right| \\
+ \left| \int_0^T \int_Y (\varphi(X^\varepsilon_t, y) - \varphi(\bar{X}_t, y)) \mu_{\bar{X}_t}(dy) dt \right|.
\]

Hence, for some constant \(C_p\),

\[
\left| \int_0^T \varphi(X^\varepsilon_t, Y^\varepsilon_t) - \int_Y \varphi(X^\varepsilon_t, y) \mu_{X^\varepsilon_t}(dy) dt \right|^{p} \leq C_p \left| \int_0^T \varphi(X^\varepsilon_t, Y^\varepsilon_t) - \int_Y \varphi(X^\varepsilon_t, y) \mu_{X^\varepsilon_t}(dy) dt \right|^{p} \\
+ C_p \left( \int_0^T \int_Y |\varphi(X^\varepsilon_t, y)| |\mu_{X^\varepsilon_t} - \mu_{\bar{X}_t}|(dy) dt \right)^{p} \\
+ C_p \left( \int_0^T \int_Y |\varphi(X^\varepsilon_t, y) - \varphi(\bar{X}_t, y)| \mu_{\bar{X}_t}(dy) dt \right)^{p}.
\]

It will suffice to show that the expectation of each term on the right tends to 0.

By Lemma 6,

\[
\lim_{\varepsilon \to 0} E \left( C_p \left| \int_0^T \varphi(X^\varepsilon_t, Y^\varepsilon_t) - \int_Y \varphi(X^\varepsilon_t, y) \mu_{X^\varepsilon_t}(dy) dt \right|^{p} \right) = 0.
\]

By Lemma 6, there is a constant \(\tilde{K}\) such that

\[
C_p \left( \int_0^T \int_Y |\varphi(X^\varepsilon_t, y)| |\mu_{X^\varepsilon_t} - \mu_{\bar{X}_t}|(dy) dt \right)^{p} \leq C_p \left( \int_0^T \tilde{K} \left| X^\varepsilon_t - \bar{X}_t \right| dt \right)^{p}
\]

whence the expectation tends to 0 by the convergence of Theorem 1.

Similarly, by Lemma 7, it is possible to choose \(\tilde{K}\) such that

\[
C_p \left( \int_0^T \int_Y |\varphi(X^\varepsilon_t, y) - \varphi(\bar{X}_t, y)| \mu_{\bar{X}_t}(dy) dt \right)^{p} \leq C_p \left( \int_0^T \tilde{K} \left| X^\varepsilon_t - \bar{X}_t \right| dt \right)^{p},
\]

whence the expectation tends to 0 by the convergence of Theorem 1.

\[\square\]

9.3 Exponential Bound for the Slow Motion and Lemmas for Asymptotic Normality of the MLE

The first main result of this section is the existence of a bound, uniform in \(\varepsilon\) sufficiently small, on the expectation of a random variable of the form \(e^{\tilde{K} \int_0^T |X^\varepsilon_t|^2 dt}\) (see Lemma 11). Expressions such as this one arise in the proof of Lemma 6. We remark that if \(c_\theta\) and \(\sigma\) are bounded then such a uniform bound is trivial. Dropping the assumption of bounded coefficients makes the uniform bound much harder to obtain.
The second main result of this section is the proof of Lemmas 11, 12, and 13 which establish the conditions of Theorem 1.6 in [14], which we use to prove Theorem 4.

Before we prove Lemma 11, let us establish a preliminary version in Lemma 10 below.

**Lemma 10. (Preliminary Exponential Bound for \( X \))** Assume Conditions 1 and 2. Then for any \( \theta \in \Theta \) and \( K > 0 \), there exists a positive constant \( \tilde{K} \) such that for all \( \varepsilon \) sufficiently small,

\[
E e^{K \int_0^T |X_t^\varepsilon| dt} \leq \tilde{K}.
\]

**Proof.** Fix \( R > 0 \) and let us define \( \tau = \tau(R) = T \wedge \inf \{ t \in [0, T] : |X_t^\varepsilon| \geq R \} \). Because \( \tau \to T \) almost surely as \( R \to \infty \), it will suffice to find a bound for \( E e^{K \int_0^\tau |X_t^\varepsilon| dt} \) that is uniform in \( R > 0 \) and \( \varepsilon \) sufficiently small.

Letting \( \nu \) denote the distribution of \( \tau \) and \( E^\nu[\cdot] \) the conditional expectation \( E[\cdot | \tau = y] \), we condition on the value of the stopping time to obtain

\[
E e^{K \int_0^\tau |X_t^\varepsilon| dt} \leq E^\nu e^{K \int_0^\tau |X_t^\varepsilon| dt} \nu(dy) \leq \int_0^\tau \left( \frac{1}{y} \int_0^y E^\nu e^{K \int_0^\tau |X_t^\varepsilon| dt} dy \right) \nu(dy),
\]

which motivates us to take a closer look at the integrand \( E^\nu e^{K y |X_t^\varepsilon|} \).

Since

\[
|X_t^\varepsilon| = |x_0| + \int_0^t X_s^\varepsilon \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon) ds + \varepsilon \int_0^t A_s : \sigma^T(X_s^\varepsilon, Y_s^\varepsilon) ds + \varepsilon \int_0^t \sqrt{\varepsilon} \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s,
\]

where \( A_s \) denotes the random matrix \( A_s = (a_{i,j})_s = \frac{1}{|X_s^\varepsilon|^2} (\delta_{ij} |X_s^\varepsilon|^2 - X_s^\varepsilon X_s^\varepsilon) \), we get the bound

\[
E^\nu e^{K y |X_t^\varepsilon|} \leq \frac{1}{2} E^\nu e^{2KyI} + \frac{1}{2} E^\nu \left[ e^{3/2Ky} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s \right],
\]

where

\[
I = |x_0| + \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\varepsilon}{2} \int_0^t A_s : \sigma^T(X_s^\varepsilon, Y_s^\varepsilon) ds.
\]

Let us focus on expectation in the right-most term in (26). Letting \( \alpha = e^{3/2K^2}y^2 \) and suppressing the arguments in \( \sigma \), for \( \nu \)-a.e. \( y \),

\[
E^\nu \left[ e^{3/2Ky} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] \leq E^\nu \left[ e^{3/2Ky} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] + E^\nu \left[ e^{-3/2Ky} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] \]

\[
\leq \frac{1}{2} E^\nu \left[ e^{2\sqrt{\alpha}} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s - 2\alpha \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] + \frac{1}{2} E^\nu \left[ e^{2\alpha} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] \]

\[
+ \frac{1}{2} E^\nu \left[ e^{-2\sqrt{\alpha}} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s - 2\alpha \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] + \frac{1}{2} E^\nu \left[ e^{2\alpha} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right] \]

\[
\leq 1 + E^\nu \left[ e^{2\alpha} \int_0^t \frac{X_s^\varepsilon}{|X_s^\varepsilon|^2} \sigma dW_s \right].
\]
Furthermore, for any pair conjugate exponents \( p_1, p_2 \) and some constant \( C \),
\[
E^y[e^{2\alpha C \int_0^T |X_s^y|^q} \sigma^2] ds \leq E^y[e^{2\alpha C \int_0^T (1 + |y_s|^q)(1 + |X_s^y|^{2r})} ds]
\leq E^y[e^{2\alpha C \int_0^T (1 + |y_s|^q) + \frac{1}{p_1}(1 + |y_s|^q) p_1 + \frac{1}{p_2} |X_s^y|^{2r}p_2} ds]
\leq \frac{1}{2} E^y[e^{4\alpha C \int_0^T 1} ds] + \frac{1}{2} E^y[e^{4\alpha C \int_0^T |X_s^y|^{2r}p_2} ds],
\]
where
\[
II = (1 + |y_s|^q) + \frac{1}{p_1}(1 + |y_s|^q) p_1;
\]
choosing \( p_2 = 1/2r \), (27) reads
\[
E^y[e^{2\alpha C \int_0^T |X_s^y|^q} \sigma^2] ds \leq \frac{1}{2} E^y[e^{4\alpha C \int_0^T 1} ds] + \frac{1}{2} E^y[e^{4\alpha C \int_0^T 2r|X_s^y|} ds].
\]
Recall that \( \alpha = 4\epsilon K^2 y^2 = O(\epsilon^3) \).

Before putting everything together, we note that the lemma we are trying to prove is trivial over intervals of time for which the process remains in a given bounded set, say \( |X_s^y| < 1 \), whereas on an interval where \( \inf_t |X_t^y| \geq 1 \), we can replace \( A_s : \sigma \sigma^T \) with some polynomial \( C_1(1 + |Y_s^y|^q) \). For this reason it suffices to prove the lemma after replacing \( A_s : \sigma \sigma^T \) with such a polynomial. Returning to (26), for some constants \( \tilde{C}, \tilde{K}, \tilde{q} > 0 \), for \( \nu \)-a.e. \( y \), and for \( \epsilon \) sufficiently small,
\[
E^y[e^{\epsilon \tilde{K} y |X_s^y|}] \leq \frac{\tilde{C}}{4} E^y[e^{\tilde{K} \int_0^T (1 + |y_s|^q)} ds] + \frac{1}{4} E^y[e^{\epsilon \tilde{K} \int_0^T |X_s^y|} ds]
\leq \frac{\tilde{C}}{2} E^y[e^{\tilde{K} \int_0^T (1 + |y_s|^q)} dt] + \frac{1}{4} E^y[e^{\epsilon \tilde{K} \int_0^T |X_s^y|} dt],
\]
whence
\[
E^y[e^{\epsilon \tilde{K} \int_0^T |X_s^y|}] \leq \frac{1}{4} \int_0^y E^y[e^{\epsilon y |X_s^y|}] dt
\]
implies
\[
E^y[e^{\epsilon \tilde{K} \int_0^T |X_s^y|}] \leq \frac{4\tilde{C}}{3} E^y[e^{\epsilon \tilde{K} \int_0^T (1 + |y_s|^q)} dt].
\]
Finally, returning to (25), for \( \epsilon \) sufficiently small,
\[
E e^{\epsilon \tilde{K} \int_0^T |X_s^y|} dt \leq \int_0^T E^y[e^{\epsilon \tilde{K} \int_0^T |X_s^y|}] \nu(dy)
\leq \frac{4\tilde{C}}{3} \int_0^T E^y[e^{\epsilon \tilde{K} \int_0^T (1 + |y_s|^q)} dt] \nu(dy),
\]
which is bounded uniformly by Lemma 4.

\textbf{Lemma 11. (Exponential Bound for }X\text{)} \text{Let } V_t \text{ be either of the Wiener processes } W_t, B_t \text{. Assume Conditions } 7 \text{ and } 3 \text{. Then for any } \theta \in \Theta \text{ and } K > 0, \text{ there exists a positive constant } \tilde{K} \text{ such that for all } \epsilon \text{ sufficiently small and all } t \in [0, T],
\[
E_0 e^{K |X_t^y|} dt \leq \tilde{K};
\]
it follows that it is possible to choose $\tilde{K}$ such that

\[
E_\theta \int_0^T e^{K|X_t^\epsilon|} dt \leq \tilde{K}, \tag{29}
\]

\[
E_\theta e^{K \int_0^T |X_t^\epsilon| dt} \leq \tilde{K}, \tag{30}
\]

\[
E_\theta e^{K \int_0^T |X_t^\epsilon|^{1/2} d\mathcal{V}_t} \leq \tilde{K}. \tag{31}
\]

**Proof.** With Lemma 10 in hand, (28) follows by a simpler version of the same argument:

\[
E e^{K|X_t^\epsilon|} dt \leq \frac{1}{2} E[e^{2Kt}] + \frac{1}{2} E[e^{e^{1/22K}\int_0^t \frac{X_s^\epsilon}{|X_s^\epsilon|^2} \sigma(X_s^\epsilon, Y_s^\epsilon) dW_s}],
\]

where as in Lemma 10

\[
A_s = (a_{i,j})_s = \frac{1}{|X_s^\epsilon|^2} (\delta_{ij}|X_s^\epsilon|^2 - X_{s,i}^\epsilon X_{s,j}^\epsilon)
\]

and

\[
I = |x_0| + \left| \int_0^t X_s^\epsilon \cdot c_\theta(X_s^\epsilon, Y_s^\epsilon) ds + \frac{\epsilon}{2} \int_0^t A_s : \sigma \sigma^T(X_s^\epsilon, Y_s^\epsilon) ds \right|.
\]

As in Lemma 10, it suffices to prove the lemma after replacing $A_s : \sigma \sigma^T$ with some polynomial $C_1(1 + |Y_s^\epsilon|^9)$, which makes $E[e^{2Kt}]$ uniformly bounded; meanwhile, with $\alpha = 4K^2$, the inequalities

\[
E[e^{\alpha} |f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} \sigma dW_s|] \leq E[e^{\alpha} |f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} \sigma dW_s|]
\]

\[
\leq \frac{1}{2} E[e^{2\alpha} f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} \sigma dW_s, -2\alpha f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} |\sigma|^2 ds] + \frac{1}{2} E[e^{2\alpha} f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} |\sigma|^2 ds]
\]

\[
= 1 + E[e^{2\alpha} f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} |\sigma|^2 ds]
\]

and an appeal to Lemma 10 establish that $E[e^{\alpha} |f^1_0 \frac{X_s^\epsilon}{|X_s^\epsilon|^2} \sigma(X_s^\epsilon, Y_s^\epsilon) dW_s|]$ also is uniformly bounded. This proves (28), (29) follows by Tonelli’s theorem and the uniformity in $t$ of our bound. (30) and (31) reduce to (29) as (10) and (11) reduce to (9) in Lemma 4.

**Proof of Lemma 7** The proof follows similarly to that of Lemma 5.3 in [24]. We present here the main steps in order to illuminate the provenance of the normalizing matrix.

Recall that

\[
\phi = \phi(\epsilon, \theta) = \sqrt{\epsilon} I^{-1/2}(\theta).
\]

We must show that for any compact $\tilde{\Theta} \subset \Theta$ and arbitrary sequences $\theta_n \in \tilde{\Theta}, \epsilon_n \to 0 \in \mathbb{R}^+$, and $u_n \to u \in \mathbb{R}^p$ such that $\theta_n + \phi(\epsilon_n, \theta_n) u_n \in \Theta$, we have a representation

\[
M_{\epsilon_n}(\theta_n, u_n) = (\Delta_n, u) - \frac{1}{2} |u|^2 + \psi_{\epsilon_n}(u_n, u, \theta_n)
\]

with

\[
\mathcal{L}_{\theta_n}(\Delta_n) = \mathcal{N}(0, I)
\]

and

\[
\lim_{n \to \infty} P_{\theta_n}(\|\psi_{\epsilon_n}(u_n, u, \theta_n)\| > \delta) = 0.
\]
Let us write for brevity

\[ \theta_{\epsilon,n} = \theta_n + \phi(\epsilon_n, \theta_n)u_n \]

and

\[ \Delta c_{\theta_n, \epsilon_n} = \frac{c_{\theta_n} - c_{\theta}}{\sqrt{\epsilon}}. \]

Recall that

\[ \kappa^T \kappa = (\sigma \sigma^T)^{-1} \sigma (I + \tau_1^T (\tau_2 \tau_2^T)^{-1} \tau_1) \sigma^T \sigma^{-1}. \] (32)

and that \( \mu_\epsilon \) denotes the invariant measure. One may for example define

\[ \kappa = \left( -\tau_2^T (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T \sigma^{-1} \right). \]

Almost surely with respect to \( P_{\theta_n} \),

\[ M_{\epsilon,n}(\theta_n, u_n) = \int_0^T \langle \kappa \Delta c_{\theta_n, \epsilon_n}, d(W, B)_t \rangle(x_t, y_t) - \frac{1}{2} \int_0^T ||\kappa \Delta c_{\theta_n, \epsilon_n}||^2(x_t, y_t) dt \]

\[ = J_1^n + J_2^n + J_3^n + J_4^n, \]

where

\[ J_1^n = \frac{1}{\sqrt{\epsilon}} \int_0^T \langle \kappa \left[ (c_{\theta_{\epsilon,n}} - c_{\theta_n}) - \nabla_{\theta} c_{\theta_{\epsilon,n}} \cdot (\theta_{\epsilon,n} - \theta_n) \right], d(W, B)_t \rangle(x_t, y_t), \]

\[ J_2^n = \frac{1}{\sqrt{\epsilon}} \int_0^T \langle \kappa \nabla_{\theta} c_{\theta_{\epsilon,n}} \cdot (\theta_{\epsilon,n} - \theta_n), d(W, B)_t \rangle(x_t, y_t) \]

\[ = \left\langle I^{-1/2}(\theta_{\epsilon,n})u_n, \int_0^T \langle \kappa \nabla_{\theta} c_{\theta_{\epsilon,n}}, d(W, B)_t \rangle(x_t, y_t) \right\rangle, \]

\[ J_3^n = -\frac{1}{2} \int_0^T \left[ ||\kappa \Delta c_{\theta_n, \epsilon_n}||^2(x_t, y_t) - \langle u_n, u_n \rangle \right] dt \]

\[ = -\frac{1}{2} \int_0^T \left[ ||\kappa \Delta c_{\theta_n, \epsilon_n}||^2(x_t, y_t) - \langle I^{-1/2}(\theta_n)u_n, q^{1/2}(X_t, \theta_n) \rangle \right] dt, \]

\[ J_4^n = -\frac{1}{2} \langle u_n, u_n \rangle. \]

\( J_2^n \) converges in distribution to \((\Phi, u)\), where \( \Phi \sim \mathcal{N}(0, I) \), and clearly \( J_4^n \) converges to \(-\frac{1}{2}||u||^2\).

Meanwhile, as in [24], \( J_1^n \) and \( J_3^n \) converge to zero in probability.

Proof of Lemma 2. The proof of Lemma 5.4 in [24] applies essentially verbatim to establish this lemma. We do not repeat it here.

Proof of Lemma 3. This is the most involved lemma to prove; the exponential bounds of Lemmas 4 and 11 play a crucial role in the necessary estimates. Recall that we have defined

\[ \phi = \phi(\epsilon, \theta) = \sqrt{\epsilon}I^{-1/2}(\theta) \]

and

\[ \Delta c_{\theta} = \frac{c_{\theta + \phi u} - c_{\theta}}{\sqrt{\epsilon}}. \]

28
For any $q$ and conjugate exponents $p_1, p_2$, \
\[ E_\theta e^{pM_\varepsilon(x, \theta)} = E_\theta \left[ e^{p_1 \int_0^\varepsilon (\kappa \Delta c_\theta, d(W, B)_t) (X^\varepsilon_t, Y^\varepsilon_t) - \frac{2}{p_1} \int_0^\varepsilon ||\kappa \Delta c_\theta||^2 (X^\varepsilon_t, Y^\varepsilon_t) dt} \right] \]
\[ \leq \left( E_\theta \left[ e^{-p_1 \frac{p_2}{2} \int_0^\varepsilon ||\kappa \Delta c_\theta||^2 (X^\varepsilon_t, Y^\varepsilon_t) dt} \right] \right)^{1/p_1} \times \left( E_\theta \left[ e^{p_2 \int_0^\varepsilon (\kappa \Delta c_\theta, d(W, B)_t) (X^\varepsilon_t, Y^\varepsilon_t) - \frac{2p_2}{p_1} \int_0^\varepsilon ||\kappa \Delta c_\theta||^2 (X^\varepsilon_t, Y^\varepsilon_t) dt} \right] \right)^{1/p_2}. \]

Choose $q$ so that $p^2 < q < p$; letting $p_2 = q/p^2 > 1$, the base in the second factor in the product is 
\[ E_\theta \left[ e^{p_2 \int_0^\varepsilon (\kappa \Delta c_\theta, d(W, B)_t) (X^\varepsilon_t, Y^\varepsilon_t) - \frac{2p_2}{p_1} \int_0^\varepsilon ||\kappa \Delta c_\theta||^2 (X^\varepsilon_t, Y^\varepsilon_t) dt} \right] \leq 1, \]
hence, with $\gamma = \frac{p^2 - q}{2(q - p^2)} > 0$, we have 
\[ E_\theta e^{pM_\varepsilon(x, \theta)} \leq \left( E_\theta \left[ e^{-\gamma \int_0^\varepsilon ||\kappa \Delta c_\theta||^2 (X^\varepsilon_t, Y^\varepsilon_t) dt} \right] \right)^{(q - p^2)/q}. \]

We wish to replace the multiscale-dependent integrand $||\kappa \Delta c_\theta||^2 (X^\varepsilon_t, Y^\varepsilon_t)$ with an averaged integrand that depends only on $X^\varepsilon_t$. Fixing $x$, we find 
\[ \int_Y ||\kappa \Delta c_\theta(x, y)||^2 \mu_x(dy) = \int_Y \left( \int_0^1 ||I^{-1/2}(\theta)u, \nabla \theta c_{\theta+\kappa \phi}(\varepsilon, \theta)_u||^2 \mu(y) \right) \mu_x(dy) \]
\[ = \int_Y \left( I^{-1/2}(\theta)u, \int_0^1 \kappa \nabla \theta c_{\theta+\kappa \phi}(\varepsilon, \theta)_u(x, y) \right)^2 \mu_x(dy) \]
\[ = \left( I^{-1/2}(\theta)u, q^{1/2}(\varepsilon, \theta)_u(x, \theta) \right)^2, \]
where by definition 
\[ q_{u}(x, \theta) = \int_Y \left( \int_0^T S(\theta + \varepsilon h, x, y) dh \right) \left( \int_0^T S(\theta + \varepsilon h, x, y) dh \right) \mu_x(dy). \]

Just as in Lemma 9, we can use a Poisson equation technique to obtain a convenient expression for the error in our approximation.

By Theorem 3 in [21], the equations
\[ \mathcal{L}_x \Phi(x, y) = ||\kappa \Delta c_\theta(x, y)||^2 - \langle I^{-1/2}(\theta)u, q^{1/2}(\varepsilon, \theta)_u(x, \theta) \rangle^2 \]
\[ \int_Y \Phi(x, y) \mu_x(dy) = 0 \]
admire a unique solution $\Phi$ in the class of functions that grow at most polynomially in $|x|$ as $x \to \infty$, and moreover, the unique solution is continuous in $x$ and bounded by an expression of the form $K(1 + |y|^q)$. Applying this solution to the multiscale process and expanding the Itô differential, we have, suppressing the arguments $(X^\varepsilon_t, Y^\varepsilon_t)$,
\[ \begin{align*} 
\mathcal{L}_{X^\varepsilon} \Phi dt &= \delta d\Phi - \delta \nabla_x \Phi \cdot c_{\theta} dt - \delta \frac{1}{2} \nabla_x^2 \Phi : \sigma \sigma^T dt - \sqrt{\delta} \nabla_y \nabla_x \Phi : \sigma \tau_1^T dt \\
&\quad - \frac{1}{2} \sqrt{\delta} \nabla_y \Phi \sigma dW_t - \sqrt{\delta} \nabla_y \Phi \tau_1 dW_t - \sqrt{\delta} \nabla_y \Phi \tau_2 dB_t. 
\end{align*} \]
By the same argument as in the proof of Theorem 3 in [21], all of the derivatives of \( \Phi \) that appear are continuous in \( x \) and bounded by expressions of the form \( K(1 + |y|^q) \). In light of Lemmas 4 and 11 for any constant \( C \) there is a constant \( \bar{K} \) such that for all \( \varepsilon \) sufficiently small,

\[
E \theta e^{C \int_0^T \mathcal{L}_{X^*_t}(X^*_t, Y^*_t) \, dt} \leq \bar{K}.
\]  

(35)

Returning to (33), we continue the estimate; for conjugate exponents \( p_3, p_4, \)

\[
E \theta e^{-\gamma \int_0^T \|\kappa \Delta c_{\theta}\| \|X_t^*, Y^*_t\|^2 \, dt} \leq \left( E \theta e^{-\gamma p_3 \int_0^T (1-1/2)(\theta) u, q_{\theta(\varepsilon, \theta)}(X_t^*, \theta)^2 \, dt} \right)^{1/p_3} \times \left( E \theta e^{-\gamma p_4 \int_0^T \|\kappa \Delta c_{\theta}(X_t^*, Y_t^*)\|^2 - (1-1/2)(\theta) u, q_{\theta(\varepsilon, \theta)}(X_t^*, \theta)^2 \, dt} \right)^{1/p_4};
\]

the second factor is bounded by a constant by the definition (34) and the inequality (35), so it remains to find an appropriate bound for \( E \theta e^{-\gamma p_3 \int_0^T (1-1/2)(\theta) u, q_{\theta(\varepsilon, \theta)}(X_t^*, \theta)^2 \, dt} \).

Denoting the quantity \( \sqrt{m_{\varepsilon}(y) \kappa(x, y) c_{\theta}(x, y)} \) by \( d_{\theta}(x, y) \) and using the inequality \( a^2 \geq b^2 - 2|a-b| \),

\[
\langle I^{-1/2}(\theta) u, q_{\theta(\varepsilon, \theta)}(X_t^*, \theta) \rangle^2 = \frac{1}{\varepsilon} \int \|d_{\theta+\phi(\varepsilon, \theta)} u - d_{\theta}\|^2(X_t^*, y) dy \\
\geq \frac{1}{\varepsilon} \int \|d_{\theta+\phi(\varepsilon, \theta)} u - d_{\theta}\|^2(X_t^*, y) dy \\
- \frac{2}{\varepsilon} \int \left[ \|d_{\theta+\phi(\varepsilon, \theta)} u(X_t^*, y) - d_{\theta+\phi(\varepsilon, \theta)} u(X_t, y)\| + \|d_{\theta}(X_t^*, y) - d_{\theta}(X_t, y)\| \right] \\
\times \|d_{\theta+\phi(\varepsilon, \theta)} u - d_{\theta}\|(X_t^*, y) \right] dy;
\]

(36)

to find the desired bound for \( E \theta e^{-\gamma p_3 \int_0^T (1-1/2)(\theta) u, q_{\theta(\varepsilon, \theta)}(X_t^*, \theta)^2 \, dt} \), we examine more closely the time integrals of the terms on the right hand side of (36).

As \( \varepsilon \to 0, \phi(\varepsilon, \theta) u \to 0 \) and

\[
\frac{1}{\varepsilon} \int_0^T \int_Y \|d_{\theta+\phi(\varepsilon, \theta)} u - d_{\theta}\|^2(X_t^*, y) dy dt = \int_0^T \int_Y \|\kappa \Delta c_{\theta}(X_t, y)\|^2 \mu_{X_t}(dy) dt \\
= \int_0^T \int_Y \|\kappa \int_0^1 \langle I^{-1/2}(\theta) u, \nabla_\theta c_{\theta+h\phi(\varepsilon, \theta)} \rangle(X_t^*, y) \rangle^T dh \|^2 \mu_{X_t}(dy) dt \\
= \int_0^T \int_Y \|\kappa \int_0^1 \langle I^{-1/2}(\theta) u, \nabla_\theta c_{\theta}(X_t^*, y) \rangle^T dh \|^2 \mu_{X_t}(dy) dt \\
+ \int_0^T \int_Y \|\kappa \int_0^1 \langle I^{-1/2}(\theta) u, (\nabla_\theta c_{\theta+h\phi(\varepsilon, \theta)} - \nabla_\theta c_{\theta})(X_t, y) \rangle^T dh \|^2 \mu_{X_t}(dy) dt \\
+ o(||I^{-1/2}(\theta) u||^2) \\
= \int_0^T \langle I^{-1/2}(\theta) u, q_{\theta(\varepsilon, \theta)}(X_t^*, \theta) \rangle^2 dt + o(||I^{-1/2}(\theta) u||^2) \\
= ||u||^2 + o(||I^{-1/2}(\theta) u||^2) \\
\geq ||I^{-1/2}(\theta) u||^2 (C_1 + o(1)),
\]

(37)

where the uniform positive definiteness of \( I(\theta) \) supplies the positive constant \( C_1 \). From here, the uniform boundedness of \( \kappa \nabla c_{\theta} \) in \( \theta \) and the lower bound (37) imply that we can find positive constants \( C_2, C_3, \) perhaps
depending on $\hat{\Theta}$, such that for all $\varepsilon$ sufficiently small,

$$C_2||I^{-1/2}(\theta)u||^2 \leq \frac{1}{\varepsilon} \int_0^T \int_Y ||d_{\theta+\phi(\varepsilon,\theta)}u - d_\theta||^2(\tilde{X}_t, y)dydt \leq C_3||I^{-1/2}(\theta)u||^2.$$  

Moving on to the second integral in (36), we have for $\tilde{\Theta} = \theta + \phi(\varepsilon, \theta)u$,

$$\frac{1}{\varepsilon} \int_0^T \int_Y (||d_{\theta+\phi(\varepsilon,\theta)}u(X_t, y) - d_\theta(X_t, y)|| ||d_{\tilde{\Theta}}(X_t^\varepsilon, y) - d_{\tilde{\Theta}}(X_t, y)||)dydt$$

$$\leq \frac{1}{\varepsilon} \left( \int_0^T \int_Y ||d_{\theta+\phi(\varepsilon,\theta)}u(X_t, y) - d_\theta(X_t, y)||^2dydt \right)^{1/2} \left( \int_0^T \int_Y ||d_{\tilde{\Theta}}(X_t^\varepsilon, y) - d_{\tilde{\Theta}}(X_t, y)||^2dydt \right)^{1/2}$$

$$= ||I^{-1/2}(\theta)u|| \left( \int_0^T \int_Y ||d_{\tilde{\Theta}}(X_t^\varepsilon, y) - d_{\tilde{\Theta}}(X_t, y)||^2dydt \right)^{1/2}. \quad (38)$$

By Theorem 1 in [21], the densities $m_x$ of the measures $\mu_x$ admit, for any $p$, a constant $C_p$ such that $|m_x(y)| \leq C_p^{1+|y|^q}$. Recalling that by definition $d_\theta(x, y) = \sqrt{m_x(y)\kappa^{-1/2}(x, y)c_\theta(x, y)}$, we see that for some constant $C_4$, we can bound the integrand in the term on the right hand of (38) as

$$||d_{\tilde{\Theta}}(X_t^\varepsilon, y) - d_{\tilde{\Theta}}(X_t, y)||^2 \leq (||d_{\theta}(X_t^\varepsilon, y)|| + ||d_{\tilde{\Theta}}(X_t, y)||)^2$$

$$\leq C_4C_p^{1+|y|^q};$$

choosing $p$ large enough that $\int_Y \frac{1+|y|^q}{1+|y|^p}dy < \infty$, it follows that there is a constant $C_5$ such that

$$\left( \int_0^T \int_Y ||d_{\tilde{\Theta}}(X_t^\varepsilon, y) - d_{\tilde{\Theta}}(X_t, y)||^2dydt \right)^{1/2} \leq C_5.$$ 

The second integral in (36) may therefore be bounded as

$$\frac{2}{\varepsilon} \int_Y \left( ||d_{\theta+\phi(\varepsilon,\theta)}u(X_t^\varepsilon, y) - d_{\theta+\phi(\varepsilon,\theta)}u(X_t, y)|| + ||d_{\theta}(X_t^\varepsilon, y) - d_{\theta}(X_t, y)|| \right)$$

$$\times ||d_{\theta+\phi(\varepsilon,\theta)}u - d_{\theta}||(\tilde{X}_t, y)dy$$

$$\leq 4C_5||I^{-1/2}(\theta)u||.$$ 

Putting everything together, for some constant $C_6$, for $\varepsilon$ sufficiently small,

$$E_{\theta}e^{\tilde{M}_t(\theta, u)} \leq \left( E_{\theta}\gamma \int_0^T ||\kappa\Delta u||^2(X_t^\varepsilon, Y_t^\varepsilon)dt \right)^{(q-p^2)/q}$$

$$\leq C_6 \left( E_{\theta}\gamma \int_0^T ||I^{-1/2}(\theta)u||^{p_3/2}d_{\theta+\phi(\varepsilon,\theta)}(X_t^\varepsilon, \theta)||^2dt \right)^{(q-p^2)/qp_3}$$

$$\leq C_6 \left( E_{\theta}\gamma \int_0^T ||I^{-1/2}(\theta)u||||d_{\theta+\phi(\varepsilon,\theta)}(X_t^\varepsilon, \theta)||^2d_{\theta}(X_t, \theta)||^{p_3} dt \right)^{(q-p^2)/qp_3}$$

$$= C_6 \left( \gamma \int_0^T ||I^{-1/2}(\theta)u||^2 + \gamma \int_0^T ||I^{-1/2}(\theta)u||^{p_3} dt \right)^{(q-p^2)/qp_3}.$$ 

This gives the function $g(||u||)$ required to prove the lemma. \qed
[21] Pardoux, E., & Veretennikov, A. Y. (2003). On Poisson equation and diffusion approximation 2. The Annals of Probability, 31(3), pp. 1166-1192.

[22] Rao, B. P. (1999). Statistical inference for diffusion type processes. Arnold, Oxford University Press.

[23] Spiliopoulos, K. (2014). Fluctuation analysis and short time asymptotics for multiple scales diffusion processes. Stochastics and Dynamics, 14(03), pp. 1350026.

[24] Spiliopoulos, K., & Chronopoulou, A. (2013). Maximum likelihood estimation for small noise multiscale diffusions. Statistical Inference for Stochastic Processes, 16(3), pp. 237-266.

[25] Veretennikov, A. Y. (1988). Bounds for the mixing rate in the theory of stochastic equations. Theory of Probability & Its Applications, 32(2), pp. 273-281.

[26] Zhang, L., Mykland, P. A., & At-Sahalia, Y. (2005). A tale of two time scales. Journal of the American Statistical Association, 100(472), pp. 1394-1411.

[27] Zwanzig, R. (1988). Diffusion in a rough potential. Proceedings of the National Academy of Sciences, 85(7), pp. 2029-2030.