Statistical inference for perturbed multiscale dynamical systems

Siragan Gailus[*], Konstantinos Spiliopoulos

Department of Mathematics & Statistics, Boston University, 111 Cummington Mall, Boston, MA 02215, United States

Abstract

We study statistical inference for small-noise-perturbed multiscale dynamical systems. We prove consistency, asymptotic normality, and convergence of all scaled moments of an appropriately constructed maximum likelihood estimator (MLE) for a parameter of interest, identifying precisely its limiting variance. We allow full dependence of coefficients on both slow and fast processes, which take values in the full Euclidean space; coefficients in the equation for the slow process need not be bounded and there is no assumption of periodic dependence. The results provide a theoretical basis for calibration of small-noise-perturbed multiscale dynamical systems. Data from numerical simulations are presented to illustrate the theory.
© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In many cases, data from physical dynamical systems exhibit multiple characteristic space- or 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. Stochastic noise may be introduced to account for uncertainty or as an essential part of a particular modeling problem. Consequently, multiscale stochastic differential equation (SDE) models are widely deployed in applied fields including physics, chemistry, and biology [5,12,28], neuroscience [13], meteorology [17], and

---

* Corresponding author.
  E-mail addresses: siragan@math.bu.edu (S. Gailus), kspiliop@math.bu.edu (K. Spiliopoulos).

econometrics and mathematical finance [9,27] to describe stochastically perturbed dynamical systems with two or more different space- or time-scales.

In this paper we consider multiscale dynamical systems perturbed by small noise. This is the regime of interest when, for example, one wishes to study rare transition events among equilibrium states of multiscale dynamical systems [6,12,28], small stochastic perturbations of multiscale dynamical systems [10,13], or small-time asymptotics of multiscale models [7,8,24]. Manuscript [15] is devoted to the problem of statistical inference for small-noise-perturbed dynamical systems, although it does not explore multiple scales.

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

$$dX_t^\varepsilon = c_\theta(X_t^\varepsilon, Y_t^\varepsilon)dt + \sqrt{\epsilon}\sigma(X_t^\varepsilon, Y_t^\varepsilon)dW_t \qquad (1)$$

$$dY_t^\varepsilon = \frac{1}{\delta}f(X_t^\varepsilon, Y_t^\varepsilon)dt + \frac{1}{\sqrt{\delta}}\tau_1(X_t^\varepsilon, Y_t^\varepsilon)dW_t + \frac{1}{\sqrt{\delta}}\tau_2(X_t^\varepsilon, Y_t^\varepsilon)dB_t$$

$$X_0^\varepsilon = x_0 \in \mathcal{X} = \mathbb{R}^{\hat{d}}, \qquad Y_0^\varepsilon = y_0 \in \mathcal{Y} = \mathbb{R}^{d-\hat{d}}.$$

Here, W and B 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 $\varepsilon = (\epsilon, \delta) \in \mathbb{R}_+^2$; the notation $\varepsilon \to 0$ should be understood to mean $\epsilon + \delta \to 0$). Conditions on the coefficients are given in Conditions 1 and 2 in Section 3. Note that the driving noises of the slow process X^ε and the fast process Y^ε may exhibit correlation. We will see that (1) may be interpreted as a small-noise multiscale perturbation of a dynamical system described by an ODE; precisely, Theorem 1 establishes that X^ε converges as $\varepsilon \to 0$ to the deterministic solution $\bar{X}$ of the ODE $d\bar{X}_t = \bar{c}_\theta(\bar{X}_t)dt$, where $\bar{c}_\theta(x)$ is an appropriate averaged coefficient.

Statistical inference for diffusions without multiple scales (i.e., $\delta \equiv 1$) is a very well-studied subject in the literature; see for example the classical manuscripts [4,16,23]. In these works $\epsilon \equiv \delta \equiv 1$ and the asymptotic behavior of the MLE is studied in the time horizon limit $T \to \infty$. This is directly analogous to the limit $n \to \infty$ in the classical setting of i.i.d. observations. Apart from the fact that in our case $\delta \not\equiv 1$, we are interested in this work in the regime $\epsilon \to 0$ with *fixed* time horizon T. While there are of course similarities between the two asymptotic regimes $T \to \infty$ and $\epsilon \to 0$, they are not exactly analogous and, as is explained in detail in [15], $\epsilon \to 0$ is the relevant regime when one is interested in small random perturbations of dynamical systems.

Maximum likelihood estimation for multiscale models with noise of order $O(1)$ has been studied in [2,1,14,18,21]. More specifically, the authors of [14] study semiparametric estimation with linear dependence in θ, and the authors of [18,21] prove consistency of the MLE induced by the (nondeterministic) limit of the slow process X^ε in (1) with $\epsilon \equiv 1$ as $\delta \to 0$, assuming that coefficients are bounded and that the fast process Y^ε takes values in a torus. It is important to point out that the regime $\varepsilon \to 0$ which we study in this paper is different in that the diffusion coefficient $\sqrt{\epsilon}\sigma$ vanishes in the limit and, as described precisely by Theorem 1, X^ε converges to the solution of an ODE rather than an SDE; the (deterministic) limit does not induce a well-defined likelihood and consequently we work directly with the likelihood of the multiscale model. Besides [18,21], perhaps most closely related to the present work is [25],

wherein the authors prove consistency and asymptotic normality of the MLE for the special case of (1) in which $Y^\varepsilon = X^\varepsilon/\delta$ with all coefficients bounded and periodic in the fast variable; such assumptions, as we will see, greatly simplify the analysis relative to the present work.

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 consistent (i.e., that it consistently estimates the true value of the parameter), but also that it is asymptotically normal— Theorem 3 establishes a central limit theorem identifying precisely the limiting variance of the estimator (i.e., the Fisher information). Secondly, we allow full dependence of coefficients on both slow and fast processes, which take values in the full Euclidean space; coefficients in the equation for the slow process need not be bounded and there is no assumption of periodic dependence. Essentially, we impose only minimal conditions necessary to guarantee that (1) has a unique strong solution and that averaging is possible in the full Euclidean space as $\delta \to 0$. Thirdly, at a more technical level, we derive in the course of the proofs ergodic-type theorems with explicit rates of convergence, which may be of independent interest (see Theorem 4 and Lemma 10 in the Appendix). To the best of our knowledge, this is the first paper that proves consistency, asymptotic normality, and convergence of all scaled moments of the MLE for small-noise-perturbed multiscale dynamical systems with general coefficients taking values in the full Euclidean space.

Let us conclude the introduction with a bit of methodology. The limiting behavior of the slow process as $\delta \to 0$ is described by the theory of averaging. A key technique in this theory exploits bounds on the solutions of Poisson equations involving the differential operators (infinitesimal generators) associated with the SDEs under consideration. In the classical manuscripts [3,22], these bounds are achieved using assumptions of periodicity or explicit compactness; in this paper, we use the relatively recent results of [19,20] to complete a series of delicate analytic estimates and extend the theory to a fairly general model in the noncompact case.

The rest of this paper is structured as follows. Section 2 discusses the MLE in general terms and introduces some relevant notation. Section 3 specifies basic conditions on the coefficients in our model and describes precisely the limiting behavior of the slow process in Theorem 1, a proof of which may be found in Appendix A.2. Section 4 presents our consistency result in Theorem 2. Section 5 presents our asymptotic normality result in Theorem 3. Section 6 studies an intuitive 'quasi-MLE' obtained by maximizing a simplified 'quasi-likelihood'; the simplified estimator retains the consistency of the MLE, but converges more slowly in numerical simulations to the true value. Section 7 presents data from numerical simulations to supplement and illustrate the theory. Section 8 sketches some possible extensions of our results. A section is reserved for acknowledgments. Finally, an Appendix collects auxiliary theorems and lemmata to which we appeal in the rest of the paper.

2. The maximum likelihood estimator

We suppose throughout that the true value θ_0 of the (unknown) parameter of interest is known to lie in an open, bounded, and convex subset $\Theta \subset \mathbb{R}^D$. In this work, we are interested in studying maximum likelihood estimation of θ_0 based on continuous data; namely, we assume that we observe a continuous trajectory $(x, y)_T = \{(x_t, y_t)\}_{0 \le t \le T}$ of (1).

It is well known that the MLE is defined in the literature as the maximizer of the likelihood. As is common in the literature in diffusion processes, we take as the basic likelihood the Girsanov density of the measure induced by (1) with respect to the measure induced by the same model with $c_\theta \equiv 0$; see for example [15,16]. Denoting these measures respectively by P_θ^ε and P_0^ε we

have by Girsanov's theorem

$$
\epsilon \log \left(\frac{dP_\theta^\varepsilon}{dP_0^\varepsilon} \right) = \sqrt{\epsilon} \int_0^T \langle \sigma^T (\sigma \sigma^T)^{-1} c_\theta, dW_t \rangle (X_t^\varepsilon, Y_t^\varepsilon)
$$

$$
- \sqrt{\epsilon} \int_0^T \langle \tau_2^T (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta, dB_t \rangle (X_t^\varepsilon, Y_t^\varepsilon)
$$

$$
- \frac{1}{2} \int_0^T |\sigma^T (\sigma \sigma^T)^{-1} c_\theta|^2 (X_t^\varepsilon, Y_t^\varepsilon) dt
$$

$$
- \frac{1}{2} \int_0^T |\tau_2^T (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta|^2 (X_t^\varepsilon, Y_t^\varepsilon) dt.
$$

Let us rewrite this in a form that is more convenient for computations. Setting for brevity

$$
\kappa = \begin{pmatrix} \sigma^T (\sigma \sigma^T)^{-1} \\ -\tau_2^T (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} \end{pmatrix},
$$

one sees that $\epsilon \log \left(\frac{dP_\theta^\varepsilon}{dP_0^\varepsilon} \right) = Z_\theta^\varepsilon ((X^\varepsilon, Y^\varepsilon)_T)$ (the equality understood to be in distribution if σ is not a square matrix), where by definition

$$
Z_\theta^\varepsilon ((x, y)_T) = \int_0^T \langle \kappa c_\theta, \kappa \cdot dx_t \rangle (x_t, y_t) - \frac{1}{2} \int_0^T |\kappa c_\theta|^2 (x_t, y_t) dt
$$

$$
+ \sqrt{\epsilon/\delta} \int_0^T \langle (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta, f \rangle (x_t, y_t) dt
$$

$$
- \sqrt{\epsilon \delta} \int_0^T \langle (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta, dY_t^\varepsilon \rangle (x_t, y_t). \tag{2}
$$

For the sake of brevity, we deliberately refer henceforth to Z_θ^ε as 'the likelihood' and to

$$
\hat{\theta}^\varepsilon ((x, y)_T) = \arg \max_{\theta \in \bar{\Theta}} Z_\theta^\varepsilon ((x, y)_T) \tag{3}
$$

as 'the MLE.'

Theoretical analysis of the MLE is complicated on account of the small parameters ϵ and δ in the likelihood; we circumvent this difficulty by using averaging results and related estimates from [19,20] to derive an auxiliary deterministic small-ε limit. Precisely, we establish in Lemma 1 that

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

where $\bar{Z}$ is an appropriately defined limiting function and $\bar{X}$ is the limit of X^ε as per Theorem 1. By comparison with the deterministic limit, we prove in Theorem 2 that the MLE is consistent and in Theorem 3 that it is asymptotically normal with convergent scaled moments.

As mentioned in Section 1, the bounds needed to establish our main result, Theorem 3, are more difficult to obtain in our case due to the fact that we allow (a) unbounded coefficients in the equation for the slow process and (b) for the fast process (as well as the slow) to take values in the full Euclidean space rather than being restricted to a compact space (e.g., a torus). In particular, as will be seen in the course of the proofs, bounds that would otherwise be standard demand delicate estimates exploiting polynomial growth of coefficients and recurrence of the fast dynamics.

We conclude this section by mentioning that the likelihood (2) may appear complicated to evaluate; two points are therefore of interest to note. Firstly, in the case of independent noise ($\tau_1 \equiv 0$) the last two terms in (2) vanish. Secondly, even in the case of dependent noise, as we establish in Section 6, if one is concerned only with consistency of the estimator then one may in fact ignore the last two terms. We refer to the resulting simplified expression as the 'quasi-likelihood' and denote it by $\tilde{Z}_\theta$ to distinguish it from Z_θ^ε (note that it no longer depends *per se* on ε). Our numerical simulations (see Section 7) suggest that the 'quasi-MLE' $\tilde{\theta}$ obtained by maximizing the quasi-likelihood, although still consistent, converges more slowly to the true value than does the MLE $\hat{\theta}^\varepsilon$. This is not, of course, surprising; the likelihoods on which the two are based are themselves after all merely asymptotically equivalent.

3. Preliminaries and assumptions

We work with a canonical probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ satisfying the usual conditions (namely, $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ is right continuous and $\mathcal{F}_0$ contains all P-negligible sets).

To guarantee that (1) is well posed and has a strong solution, and that our limit results are valid, we impose the following regularity and growth conditions:

Condition 1. *Conditions on c_θ*

1. $\exists K > 0, q > 0, r \in [0, 1); \forall \theta \in \Theta, |c_\theta(x, y)| \leq K(1 + |x|^r)(1 + |y|^q)$
2. $\exists K > 0, q > 0; \forall \theta \in \Theta, |\nabla_x c_\theta(x, y)| + |\nabla_x \nabla_x c_\theta(x, y)| \leq K(1 + |y|^q)$
3. $\forall \theta \in \Theta, c_\theta$ *has two continuous derivatives in x, Hölder continuous in y uniformly in x*
4. $\forall \theta \in \Theta, \nabla_y \nabla_y c_\theta(x, y)$ *is jointly continuous in x and y*
5. $c_\theta(x, y)$ *has two locally bounded derivatives in θ with at most polynomial growth in x and y*

   Conditions on σ

1. $\forall N > 0, \exists C(N); \forall x_1, x_2 \in \mathcal{X}, \forall y \in \mathcal{Y}$ *with* $|y| \leq N, |\sigma(x_1, y) - \sigma(x_2, y)| \leq C(N)|x_1 - x_2|$
2. $\exists K > 0, q > 0; |\sigma(x, y)| \leq K(1 + |x|^{1/2})(1 + |y|^q)$
3. $\sigma \sigma^T$ *is uniformly nondegenerate*

   Conditions on f, τ_1, τ_2

1. $f, \tau_1 \tau_1^T$, *and* $\tau_2 \tau_2^T$ *are twice differentiable in x and y, the first and second derivatives in x being bounded, and all partial derivatives up to second order being Hölder continuous in y uniformly in x*
2. $\tau_2 \tau_2^T$ *is uniformly nondegenerate.*

In the limit of infinite scale separation, the slow process appears from the perspective of the fast to become 'frozen.' To guarantee that the fast process has an invariant distribution when the slow process is 'frozen,' we impose the following recurrence condition:

Condition 2.

$$\lim_{|y| \to \infty} \sup_{x \in \mathcal{X}} \left(f(x, y) \cdot y \right) = -\infty.$$

The conditions on f, τ_1, τ_2 in Conditions 1 and 2 guarantee the existence, for each fixed $x \in \mathcal{X}$, of a unique invariant measure μ_x associated with the operator

$$\mathcal{L}_x = f(x, \cdot) \cdot \nabla_y + \frac{1}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, \cdot) : \nabla_y^2$$

(for existence of an invariant measure, see for example [26]; uniqueness is a consequence of nondegenerate diffusion, as for example in [20]).

Conditions 1 and 2 are enough to establish the following averaging theorem, which is essentially the law of large numbers for the slow process X^ε. Incidentally, this justifies the interpretation of (1) as a small-noise multiscale perturbation of a deterministic dynamical system.

Theorem 1. *Assume Conditions 1 and 2. For any fixed $\theta_0 \in \Theta$, initial condition $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, and $0 \leq p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,*

$$E \sup_{0 \leq t \leq T} |X_t^\varepsilon - \bar{X}_t|^p \leq \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p,$$

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

$$\bar{X}_t = x_0 + \int_0^t \bar{c}_{\theta_0}(\bar{X}_s)ds,$$

where $\bar{c}_{\theta_0}(x) = \int_{\mathcal{Y}} c_{\theta_0}(x, y)\mu_x(dy)$.

The proof is deferred to Appendix A.2.

We remark that although convergence of X^ε to $\bar{X}$ is generally expected (see [22,19,20,25]), 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 in the full Euclidean space, specifying an explicit rate of convergence. Of course, we impose Conditions 1 and 2 and exploit the fact that $\bar{X}$ is deterministic. Apart from being an interesting result in its own right, Theorem 1 will play for us a key role in proving that the MLE is consistent (Theorem 2) and asymptotically normal (Theorem 3).

4. Consistency of the MLE

We now identify conditions sufficient for $\hat{\theta}^\varepsilon$ to be an L^p-consistent estimator of the true value of the parameter (Theorem 2). Theoretical analysis is complicated on account of the fact that $Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T)$ is a random variable depending on the small parameters ϵ and δ; we circumvent this difficulty by deriving a deterministic small-ε limit. Recall from Section 2 that

$$\kappa = \begin{pmatrix} \sigma^T(\sigma\sigma^T)^{-1} \\ -\tau_2^T(\tau_2\tau_2^T)^{-1}\tau_1\sigma^T(\sigma\sigma^T)^{-1} \end{pmatrix}.$$

Denoting by θ_0 the true value of the parameter, we have almost surely

$$Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T) = \int_0^T \langle \kappa c_\theta, \kappa \cdot dX_t^\varepsilon \rangle (X_t^\varepsilon, Y_t^\varepsilon) - \frac{1}{2} \int_0^T |\kappa c_\theta|^2 (X_t^\varepsilon, Y_t^\varepsilon)dt$$

$$- \sqrt{\epsilon} \int_0^T \langle (\tau_2\tau_2^T)^{-1}\tau_1\sigma^T(\sigma\sigma^T)^{-1}c_\theta, (\tau_1 dW_t + \tau_2 dB_t) \rangle (X_t^\varepsilon, Y_t^\varepsilon)$$

$$= \int_0^T \langle \kappa c_\theta, \kappa c_{\theta_0} \rangle (X_t^\varepsilon, Y_t^\varepsilon)dt - \frac{1}{2} \int_0^T |\kappa c_\theta|^2 (X_t^\varepsilon, Y_t^\varepsilon)dt$$

$$+ \sqrt{\epsilon} \int_0^T \langle \kappa c_\theta, \kappa \cdot dW_t \rangle (X_t^\varepsilon, Y_t^\varepsilon)$$

$$- \sqrt{\epsilon} \int_0^T \langle (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta, (\tau_1 dW_t + \tau_2 dB_t) \rangle (X_t^\varepsilon, Y_t^\varepsilon). \quad (4)$$

Intuitively speaking, as $\varepsilon = (\epsilon, \delta) \to 0$, three things are happening. Firstly, the last two terms in the above expression tend to zero because of their vanishing prefactor $\sqrt{\epsilon}$. Secondly, as the fast dynamics accelerate relative to the slow, Y_t^ε tends to its invariant distribution at each 'frozen' value X_t^ε. Finally, X^ε itself tends to $\bar{X}$ as per Theorem 1. Thus we are motivated to approximate $Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T)$ in the small-ε limit by $\bar{Z}_{\theta,\theta_0}((\bar{X})_T)$, where the auxiliary function $\bar{Z}_{\theta,\theta_0}$ is defined at a trajectory $(z)_T = \{z_t\}_{0 \le t \le T} \subset \mathcal{X}$ by the formula

$$\bar{Z}_{\theta,\theta_0}((z)_T) = \int_0^T \int_{\mathcal{Y}} \langle \kappa c_\theta, \kappa c_{\theta_0} \rangle (z_t, y) \mu_{z_t}(dy) dt - \frac{1}{2} \int_0^T \int_{\mathcal{Y}} |\kappa c_\theta|^2 (z_t, y) \mu_{z_t}(dy) dt.$$
$$(5)$$

Notice that $\bar{Z}_{\theta,\theta_0}((z)_T)$ attains a maximum at $\theta = \theta_0$, as it is plain to see upon completing the square:

$$\bar{Z}_{\theta,\theta_0}((z)_T) = \frac{1}{2} \int_0^T \int_{\mathcal{Y}} |\kappa c_{\theta_0}|^2 (z_t, y) \mu_{z_t}(dy) dt$$

$$- \frac{1}{2} \int_0^T \int_{\mathcal{Y}} |\kappa (c_\theta - c_{\theta_0})|^2 (z_t, y) \mu_{z_t}(dy) dt.$$

Lemma 1 establishes that if κ is sufficiently smooth, then $Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T)$ is indeed well approximated in the small-ε limit by $\bar{Z}_{\theta,\theta_0}((\bar{X})_T)$. Again, we emphasize that $\bar{Z}_{\theta,\theta_0}$ is an auxiliary function; we use it only as a vehicle to complete our proof of Theorem 2.

Condition 3 (*Smoothness Condition*). *κ has two continuous derivatives in x, Hölder continuous in y uniformly in x.*

Lemma 1. *Let θ_0 be the true value of the parameter. Assume Conditions 1, 2, and 3. For any $0 \le p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,*

$$E \sup_{\theta \in \Theta} \left| Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T) - \bar{Z}_{\theta,\theta_0}((\bar{X})_T) \right|^p \le \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p,$$

where $\bar{X}$ is the limit of X^ε as per Theorem 1.

The proof is deferred to Appendix A.2.

Recall that θ_0 maximizes $\bar{Z}_{\theta,\theta_0}$ whereas $\hat{\theta}^\varepsilon$ (by definition) maximizes Z_θ^ε. Thus, L^p-consistency of $\hat{\theta}^\varepsilon$ can be proved by combining Lemma 1 with an appropriate identifiability condition. Theorem 2 presents the main result of this section.

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

$$\sup_{|u| > \eta} \left(\bar{Z}_{\theta_0+u,\theta_0}((\bar{X})_T) - \bar{Z}_{\theta_0,\theta_0}((\bar{X})_T) \right) \le -\eta.$$

Recalling the form of $\bar{Z}_{\theta,\theta_0}$, we remark that the identifiability condition may be stated equivalently as

$$\inf_{|u|>\eta} \int_0^T \int_{\mathcal{Y}} |\kappa(c_{\theta_0+u} - c_{\theta_0})|^2(\bar{X}_t, y)\mu_{\bar{X}_t}(dy)dt \geq \eta > 0.$$

Theorem 2. *Let θ_0 be the true value of the parameter and let $\hat{\theta}^\varepsilon = argmax_{\theta \in \Theta} Z_{\theta,T}^\varepsilon((X^\varepsilon, Y^\varepsilon)_T)$. Assume Conditions 1, 2, 3, and 4. For any $1 \leq p < \infty$,*

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

Proof. For brevity, denote by Z_θ^ε and $\bar{Z}_\theta$ respectively the quantities $Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T)$ and $\bar{Z}_{\theta,\theta_0}((\bar{X})_T)$. For any $\eta > 0$ and $1 \leq r < \infty$, there is a constant K_r such that for ε sufficiently small,

$$\begin{aligned}
P\left(|\hat{\theta}^\varepsilon - \theta_0| \geq \eta\right) &\leq P\left(\sup_{|u|>\eta}\left(Z_{\theta_0+u}^\varepsilon - Z_{\theta_0}^\varepsilon\right) \geq 0\right) \\
&\leq P\left(\sup_{|u|>\eta}\left((Z_{\theta_0+u}^\varepsilon - Z_{\theta_0}^\varepsilon) - (\bar{Z}_{\theta_0+u} - \bar{Z}_{\theta_0})\right)\right. \\
&\qquad\left. \geq -\sup_{|u|>\eta}\left(\bar{Z}_{\theta_0+u} - \bar{Z}_{\theta_0}\right)\right) \\
&\leq P\left(\sup_{|u|>\eta}\left((Z_{\theta_0+u}^\varepsilon - Z_{\theta_0}^\varepsilon) - (\bar{Z}_{\theta_0+u} - \bar{Z}_{\theta_0})\right) \geq \eta\right) \\
&\leq P\left(\sup_{|u|>\eta}|Z_{\theta_0+u}^\varepsilon - \bar{Z}_{\theta_0+u}| \geq \frac{\eta}{2}\right) + P_{\theta_0}\left(|Z_{\theta_0}^\varepsilon - \bar{Z}_{\theta_0}| \geq \frac{\eta}{2}\right) \\
&\leq \frac{2^{r+1}}{|\eta|^r} E \sup_{\theta \in \Theta}|Z_\theta^\varepsilon - \bar{Z}_\theta|^r \\
&\leq \frac{2^{r+1}}{|\eta|^r} K_r(\sqrt{\epsilon} + \sqrt{\delta})^r,
\end{aligned}$$

where the third inequality follows by Condition 4, the fifth is the Markov inequality, and the last follows by Lemma 1.

Hence, for any $1 \leq r < \infty$ and $\ell > 0$,

$$\begin{aligned}
E|\hat{\theta}^\varepsilon - \theta_0|^p &= E|(\hat{\theta}^\varepsilon - \theta_0) \cdot \chi_{|\hat{\theta}^\varepsilon-\theta_0|<\ell}|^p + E|(\hat{\theta}^\varepsilon - \theta_0) \cdot \chi_{|\hat{\theta}^\varepsilon-\theta_0|\geq\ell}|^p \\
&< \ell^p + \sum_{k=0}^\infty (k+\ell+1)^p \cdot P\left(k+\ell \leq |\hat{\theta}^\varepsilon - \theta_0| < k+\ell+1\right) \\
&\leq \ell^p + \sum_{k=0}^\infty (k+\ell+1)^p \cdot P\left(k+\ell \leq |\hat{\theta}^\varepsilon - \theta_0|\right) \\
&\leq \ell^p + 2^{r+1} K_r(\sqrt{\epsilon} + \sqrt{\delta})^r \sum_{k=0}^\infty \frac{(k+\ell+1)^p}{(k+\ell)^r}.
\end{aligned}$$

Choosing any $r > p+1$, $\sum_{k=0}^\infty \frac{(k+\ell+1)^p}{(k+\ell)^r} < \infty$. Since $\ell > 0$ was arbitrary, we get the desired result as $\varepsilon \to 0$. $\quad\square$

5. Asymptotic normality of the MLE

We now identify conditions sufficient for $\hat{\theta}^\varepsilon$ to be asymptotically normal with convergent scaled moments (Theorem 3). This greatly increases the practical value of the MLE as it allows approximate confidence intervals to be constructed about the point estimate, characterizing its efficiency.

Sufficient conditions for asymptotic normality when $\delta \equiv 1$ (i.e., without multiple scales) appear in [15,16]. In the multiscale setup the only known asymptotic normality results are those of [25]. In [25], the authors study the special case of a multiscale process in which the fast process is simply $Y^\varepsilon = X^\varepsilon/\delta$ and furthermore assume uniformly bounded coefficients that are also periodic in the fast variable. As far as we are aware, the present paper presents the first such result for a small-noise multiscale model in the full Euclidean space with general dependence of coefficients on both slow and fast processes. In particular, we allow general dynamics for the fast process (i.e., we do not restrict ourselves to the case $Y^\varepsilon = X^\varepsilon/\delta$), we allow unbounded coefficients in the equation for the slow process, and we do not restrict the fast process to a compact space (e.g., a torus), but allow it to take values in the full Euclidean space.

We follow the method of Ibragimov–Has'minskii (see Theorem 3.1.1 in [11]; see also Theorem 1.6 in [15] and Theorem 2.6 and Remark 2.7 in [16]). The necessary limits are more difficult to establish for noncompact state spaces; bounds that would otherwise be standard demand delicate estimates exploiting polynomial growth of coefficients and recurrence of the fast dynamics.

Recall from Section 2 that

$$\kappa = \begin{pmatrix} \sigma^T(\sigma\sigma^T)^{-1} \\ -\tau_2^T(\tau_2\tau_2^T)^{-1}\tau_1\sigma^T(\sigma\sigma^T)^{-1} \end{pmatrix}.$$

Definition 1. We define the matrix Q and the Fisher information matrix $I(\theta)$ as

$$Q(x,\theta) = \int_{\mathcal{Y}} \left((\nabla_\theta c_\theta)^T \kappa^T \kappa (\nabla_\theta c_\theta) \right)(x,y)\mu_x(dy)$$

$$I(\theta) = \int_0^T Q(\bar{X}_t,\theta)dt.$$

We shall impose the following continuity and nondegeneracy condition:

Condition 5 (*Continuity of Q and Nondegeneracy of I(θ)*)**.**

1. *The process $Q^{1/2}(X_t^\varepsilon,\theta), t \in [0,T]$ is continuous in probability, uniformly in $L^2[0,T]$ in $\theta \in \Theta$*
2. *$Q^{1/2}(x,\theta_0)$ is continuous in x*
3. *$I(\theta)$ is positive definite uniformly in $\theta \in \Theta$; i.e.,*

$$\exists \eta > 0; \quad \eta \leq \inf_{\theta \in \Theta} \inf_{|\lambda|=1} \lambda^T I(\theta)\lambda.$$

We now state three lemmata that establish Theorem 3, our asymptotic normality theorem. Recall from Section 2 that P_θ^ε is the measure induced by (1). Let us set for brevity

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

and

$$
M_\theta^\varepsilon(u) = \log \frac{dP_{\theta+\phi u}^\varepsilon}{dP_\theta^\varepsilon} = \frac{1}{\sqrt{\epsilon}} \int_0^T \langle \kappa(c_{\theta+\phi u} - c_\theta), d(W, B)_t \rangle (X_t^\varepsilon, Y_t^\varepsilon)
$$
$$
- \frac{1}{2\epsilon} \int_0^T |\kappa(c_{\theta+\phi u} - c_\theta)|^2 (X_t^\varepsilon, Y_t^\varepsilon) dt;
$$

$M_\theta^\varepsilon(u)$ is of course nothing else than the log-likelihood-ratio.

Lemma 2. *Assume Conditions 1, 2, 4, and 5. The family $\{P_\theta^\varepsilon\}_{\theta \in \Theta}$ is uniformly asymptotically normal with normalizing matrix ϕ; that is, for any compact $\tilde{\Theta} \subset \Theta$ and sequences $\{\theta_n\}_{n=1}^\infty \subset \tilde{\Theta}$, $\{\varepsilon_n = (\epsilon_n, \delta_n)\}_{n=1}^\infty \subset \mathbb{R}_+^2$, and $\{u_n\}_{n=1}^\infty \subset \mathbb{R}^p$ with $\varepsilon_n \to 0$, $u_n \to u$, and $\{\theta_n + \phi(\epsilon_n, \theta_n) u_n\}_{n=1}^\infty \subset \Theta$, we have a representation*

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

with $\Delta_n \Rightarrow \mathcal{N}(0, I)$ in $P_{\theta_n}^{\varepsilon_n}$-law and $\lim_{n \to \infty} P_{\theta_n}^{\varepsilon_n}\left(|\psi_{\varepsilon_n}(u_n, \varepsilon_n, \theta_n)| > \eta\right) = 0$ for each $\eta > 0$.

The proof is deferred to Appendix A.3.

Lemma 3. *Assume Conditions 1, 2, 4, and 5. There are constants $m > D$ and $\tilde{K}$ such that for any $\varepsilon \in (0, 1)^2$ and compact $\tilde{\Theta} \subset \Theta$,*

$$
\sup_{\theta \in \tilde{\Theta}} |u_2 - u_1|^{-m} E_\theta^\varepsilon \left| e^{\frac{1}{m} M_\theta^\varepsilon(u_2)} - e^{\frac{1}{m} M_\theta^\varepsilon(u_1)} \right|^m \le \tilde{K}.
$$

The proof of Lemma 5.4 in [25] applies essentially verbatim to establish Lemma 3; we do not repeat it.

Lemma 4. *Assume Conditions 1, 2, 3, 4, and 5. Assume that ϵ does not decay too quickly relative to δ as $\varepsilon = (\epsilon, \delta) \to 0$; that is, suppose there is an $\alpha > 0$ such that we are interested (at least when ε is sufficiently small) only in pairs $\varepsilon = (\epsilon, \delta)$ satisfying $0 < \delta \le \epsilon^\alpha$. For any compact $\tilde{\Theta} \subset \Theta$ and $N > 0$, there are constants $\tilde{K}$ and $\epsilon_0 > 0$ such that*

$$
\sup_{0<\epsilon<\epsilon_0, 0<\delta\le\epsilon^\alpha} \sup_{\theta \in \tilde{\Theta}} \sup_{u;\theta+\phi u \in \tilde{\Theta}} |u|^N E_\theta^\varepsilon e^{\frac{1}{2} M_\theta^\varepsilon(u)} \le \tilde{K}.
$$

The proof is deferred to Appendix A.3.

Theorem 3. *Assume Conditions 1, 2, 3, 4, and 5. Let $\{\varepsilon_n = (\epsilon_n, \delta_n)\}_{n=0}^\infty \subset \mathbb{R}_+^2$ be a sequence with $\varepsilon_n \to 0$. Assume that ϵ_n does not decay too quickly relative to δ_n; that is, suppose that there is an $\alpha > 0$ such that (eventually, at least) $0 < \delta_n \le \epsilon_n^\alpha$. For any compact subset $\tilde{\Theta} \subset \Theta$,*

$$
(I(\theta)/\epsilon_n)^{1/2}(\hat{\theta}^{\varepsilon_n} - \theta) \Rightarrow \mathcal{N}(0, I)
$$

in $P_\theta^{\varepsilon_n}$-distribution uniformly in $\theta \in \tilde{\Theta}$. Moreover, denoting by $\{\mu^p\}_{p=0}^\infty$ the moments of the standard multivariate normal distribution $\mathcal{N}(0, I)$, for all $p > 0$,

$$
\lim_{n \to \infty} \sup_{\theta \in \tilde{\Theta}} \left| E_\theta^{\varepsilon_n} |(I(\theta)/\epsilon_n)^{1/2}(\hat{\theta}^{\varepsilon_n} - \theta)|^p - \mu^p \right| = 0.
$$

Proof. This follows by Theorem 3.1.1 in [11] (see also Theorem 1.6 in [15] and Theorem 2.6 and Remark 2.7 in [16]). Lemmata 2, 3, and 4 establish the necessary conditions. $\square$

6. An alternative simplified likelihood and its properties

As mentioned in Section 2, the likelihood Z_θ^ε given by (2) may appear complicated to evaluate. In the case of independent noise ($\tau_1 \equiv 0$), the last two terms in (2) vanish. Interestingly, it turns out that one may ignore the last two terms even in the case of dependent noise, provided that one is concerned only with consistency of the estimator. We refer to the resulting simplified expression as the 'quasi-likelihood' and denote it by $\tilde{Z}_\theta$ to distinguish it from Z_θ^ε (note that it no longer depends *per se* on ε). We refer to the estimator $\tilde{\theta}$ obtained by maximizing the quasi-likelihood as the 'quasi-MLE.' Precisely,

$$\tilde{\theta}((x, y)_T) = \arg\max_{\theta \in \tilde{\Theta}} \tilde{Z}_\theta((x, y)_T), \tag{6}$$

where

$$\tilde{Z}_\theta((x, y)_T) = \int_0^T \langle \kappa c_\theta, \kappa \cdot dx_t \rangle (x_t, y_t) - \frac{1}{2} \int_0^T |\kappa c_\theta|^2 (x_t, y_t) dt.$$

Recall from Section 4 that almost surely,

$$Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T) = \int_0^T \langle \kappa c_\theta, \kappa c_{\theta_0} \rangle (X_t^\varepsilon, Y_t^\varepsilon) dt - \frac{1}{2} \int_0^T |\kappa c_\theta|^2 (X_t^\varepsilon, Y_t^\varepsilon) dt$$

$$+ \sqrt{\epsilon} \int_0^T \langle \kappa c_\theta, \kappa \cdot dW_t \rangle (X_t^\varepsilon, Y_t^\varepsilon)$$

$$- \sqrt{\epsilon} \int_0^T \langle (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} c_\theta \cdot (\tau_1 dW_t + \tau_2 dB_t) \rangle (X_t^\varepsilon, Y_t^\varepsilon).$$

Recalling the polynomial bounds in Condition 1 and applying Lemma 7 in Appendix A.1, for any $0 \le p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,

$$E \sup_{\theta \in \Theta} \left| Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T) - \tilde{Z}_\theta((X^\varepsilon, Y^\varepsilon)_T) \right|^p \le \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p.$$

Combining this with Lemma 1, we see via the triangle inequality that for any $0 \le p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,

$$E \sup_{\theta \in \Theta} \left| \tilde{Z}_\theta((X^\varepsilon, Y^\varepsilon)_T) - \bar{Z}_{\theta, \theta_0}((\bar{X})_T) \right|^p \le \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p, \tag{7}$$

which is enough for the proof of Theorem 2 to go through with $\tilde{Z}$ in place of Z^ε. We therefore have the following lemma:

Lemma 5. *Let θ_0 be the true value of the parameter. Assume Conditions 1, 2, 3, and 4. For any $1 \le p < \infty$,*

$$\lim_{\varepsilon \to 0} E_{\theta_0} |\tilde{\theta} - \theta_0|^p = 0.$$

We omit the details of the proof of Lemma 5 as, given (7), the proof is identical to that of Theorem 2.

We emphasize that in the case of independent noise (i.e., $\tau_1 \equiv 0$), the MLE and quasi-MLE coincide. On the other hand, as we shall see in Section 7, using the quasi-MLE instead of the MLE when $\tau_1 \not\equiv 0$ comes at a price. Although the estimator is still consistent, our numerical

simulations (see Section 7) suggest that the quasi-MLE converges more slowly to the true value than does the MLE. This is not, of course, surprising; the likelihoods on which the two are based are themselves after all merely asymptotically equivalent.

7. Numerical examples

We now present data from numerical simulations to supplement and illustrate the theory. We begin by considering the system

$$dX_t^\varepsilon = \theta_0(\sin(X_t^\varepsilon))(Y_t^\varepsilon)^2 dt + \sqrt{\epsilon} dW_t \tag{8}$$
$$dY_t^\varepsilon = -\frac{1}{\delta} Y_t^\varepsilon dt + \frac{1}{\sqrt{\delta}} dB_t$$

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

It is easy to see that the limit $\bar{X}$ of the slow process X^ε in (8) 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 is

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

We simulate trajectories using an Euler scheme. Precisely,

$$X_{t_{k+1}}^\varepsilon = \theta_0(\sin(X_{t_k}^\varepsilon))(Y_{t_k}^\varepsilon)^2 (t_{k+1} - t_k) + \sqrt{\epsilon}(W_{t_{k+1}} - W_{t_k})$$
$$Y_{t_{k+1}}^\varepsilon = -\frac{1}{\delta} Y_{t_k}^\varepsilon (t_{k+1} - t_k) + \frac{1}{\sqrt{\delta}}(B_{t_{k+1}} - B_{t_k}),$$

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

Let us fix $\epsilon = 10^{-1}, \delta = 10^{-3}, n = 10^6$, and suppose that the discrete time steps are evenly spaced (i.e. $t_{k+1} - t_k = \Delta t = 10^{-6}$). We remark that the choice of δ influences the error of the Euler approximation. As in [25], one can derive that the error is $O(\Delta t/\delta)$, which implies that with the choice $\delta = 10^{-3}$ and $\Delta t = 10^{-6}$ one has an approximation error on the order of 10^{-3}.

The likelihood is

$$Z_{\theta,1}^\varepsilon((x, y)_1) = \int_0^1 \theta \sin(x_t) y_t^2 dx_t - \frac{1}{2} \int_0^1 \theta^2 \sin^2(x_t) y_t^4 dt.$$

The MLE is therefore

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

Discretizing this we obtain the approximation

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

Evidently, we use a single time-series of 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 1 presents in each case the mean MLE, a normal-based confidence interval using the empirical standard deviation, and the theoretical

Table 1
Estimates of θ_0 with empirical confidence intervals and theoretical standard deviations.

True value of θ_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
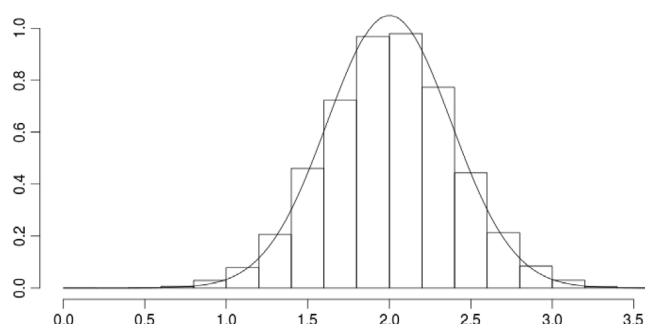
Fig. 1. $\theta_0 = 2$.
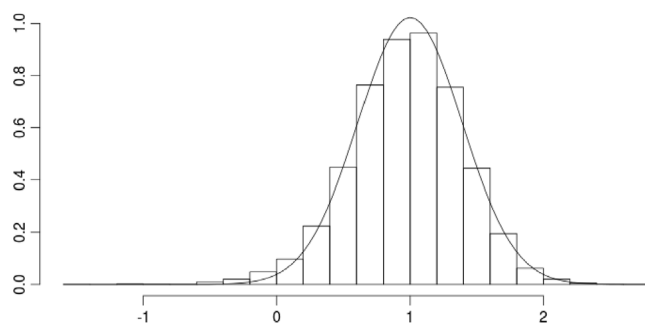
Fig. 2. $\theta_0 = 1$.

standard deviation as per Theorem 3 (i.e., $\sqrt{\epsilon/I(\theta_0)}$). The histograms in Figs. 1–3 compare the empirical distribution of the MLE with the theoretical density curve.

Let us take another example to illustrate the case of dependent noise (i.e., $\tau_1 \neq 0$) and the difference between the true MLE (3) and quasi-MLE (6). We consider the system

$$dX_t^\varepsilon = \theta_0(\sin(X_t^\varepsilon))(Y_t^\varepsilon)^2 dt + \sqrt{\epsilon}\,dW_t$$

$$dY_t^\varepsilon = -\frac{1}{\delta}Y_t^\varepsilon dt + \frac{1}{\sqrt{2\delta}}dW_t + \frac{1}{\sqrt{2\delta}}dB_t$$

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

This time we consider both $\epsilon = 10^{-1}$ and $\epsilon = 10^{-2}$, again fixing $\delta = 10^{-3}$ and simulating trajectories using an Euler scheme with $n = 10^6$ discrete time steps. We simulate the trajectories,
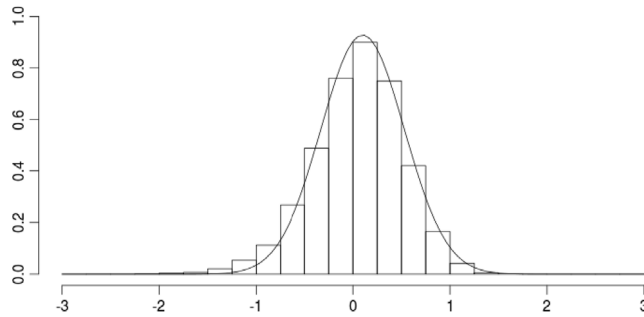
Fig. 3. $\theta_0 = 0.1$.

Table 2
Estimates of $\theta_0 = 1$ with empirical confidence intervals and theoretical standard deviations.

ϵ	MLE	Mean estimator	68% Confidence interval	95% Confidence interval	Theoretical SD ($\hat{\theta}^\varepsilon$)
0.1	$\hat{\theta}^\varepsilon$	0.985	(0.688, 1.282)	(0.391, 1.579)	0.276
0.1	$\tilde{\theta}$	0.972	(0.551, 1.393)	(0.130, 1.814)	–
0.01	$\hat{\theta}^\varepsilon$	0.998	(0.907, 1.090)	(0.815, 1.182)	0.087
0.01	$\tilde{\theta}$	0.996	(0.876, 1.115)	(0.757, 1.235)	–

true MLE $\hat{\theta}^\varepsilon$, and quasi-MLE $\tilde{\theta}$ 10^4 times for $\theta_0 = 1$. Table 2 presents in each case the mean MLE, a normal-based confidence interval using the empirical standard deviation, and, for the true MLE, the theoretical standard deviation as per Theorem 3 (i.e., $\sqrt{\epsilon/I(\theta_0)}$). The histograms in Figs. 4–7 compare the empirical distribution of the true MLE and quasi-MLE with the theoretical density curve for the true MLE.

Remark 1. A few remarks are in order. We notice that both $\hat{\theta}^\varepsilon$ and $\tilde{\theta}$ exhibit smaller variance (and hence tighter confidence bounds) when $\epsilon = 0.01$ than when $\epsilon = 0.1$—this is of course consistent with our asymptotic theory. We also notice that the variance of $\tilde{\theta}$ is larger than the variance of $\hat{\theta}^\varepsilon$, which is to be expected since the former is based on omitting certain terms in the likelihood. Table 2 suggests however that the omission does not matter in the limit, as the gap in the empirical variances of $\hat{\theta}^\varepsilon$ and $\tilde{\theta}$ closes as ϵ gets smaller.

We conclude the section with a final example to illustrate the case of $c_\theta(x, y)$ unbounded in x. We consider the system

$$dX_t^\varepsilon = \theta_0 X_t^\varepsilon (Y_t^\varepsilon)^2 dt + \sqrt{\epsilon}dW_t$$

$$dY_t^\varepsilon = -\frac{1}{\delta}Y_t^\varepsilon dt + \frac{\sqrt{3}}{2}\sqrt{\delta}dW_t + \frac{1}{2\sqrt{\delta}}dB_t$$

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

As in the last example, we consider both $\epsilon = 10^{-1}$ and $\epsilon = 10^{-2}$, fix $\delta = 10^{-3}$, and simulate trajectories using an Euler scheme with $n = 10^6$ discrete time steps. We simulate the trajectories, true MLE $\hat{\theta}^\varepsilon$, and quasi-MLE $\tilde{\theta}$ 10^4 times for $\theta_0 = 1$. Table 3 presents in each case the mean MLE, a normal-based confidence interval using the empirical standard deviation, and, for the true MLE, the theoretical standard deviation as per Theorem 3 (i.e., $\sqrt{\epsilon/I(\theta_0)}$). The histograms
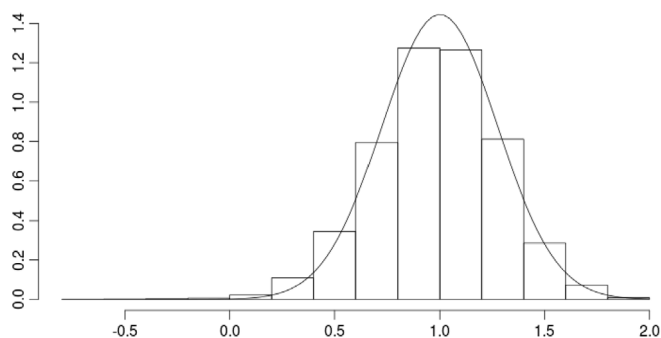
Fig. 4. $\hat{\theta}^{\varepsilon}$, $\epsilon = 0.1$.
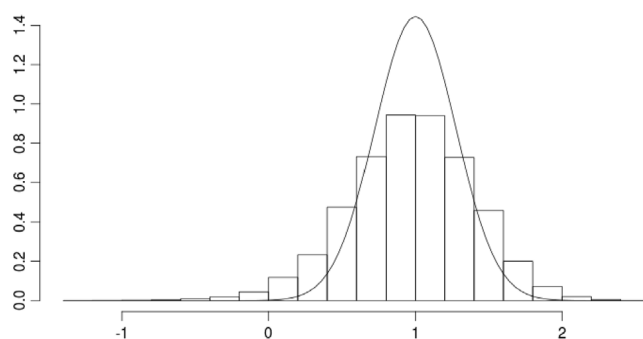
Fig. 5. $\tilde{\theta}$, $\epsilon = 0.1$.
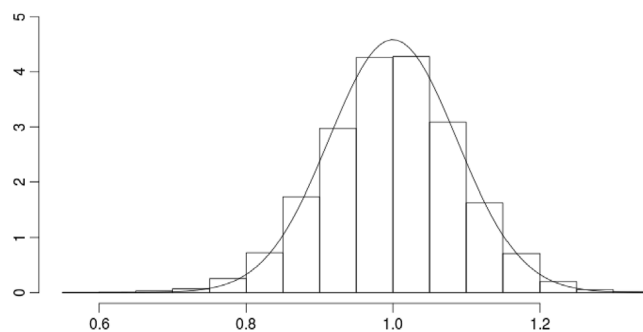
Fig. 6. $\hat{\theta}^{\varepsilon}$, $\epsilon = 0.01$.

that follow in Figs. 8–9 compare the empirical distribution of the true MLE with the theoretical density curve.

8. Possible extensions

In this section we discuss some possible extensions of our results.

Firstly, we sketch a straightforward extension of our results to models with a greater plurality of time-scales. Consider the model

$$dX_t^{\varepsilon} = \left[c_\theta(X_t^{\varepsilon}, Y_t^{\varepsilon}) + h_1(\epsilon, \delta)d_\theta(X_t^{\varepsilon}, Y_t^{\varepsilon}) \right] dt + \sqrt{\epsilon}\sigma(X_t^{\varepsilon}, Y_t^{\varepsilon})dW_t \tag{9}$$
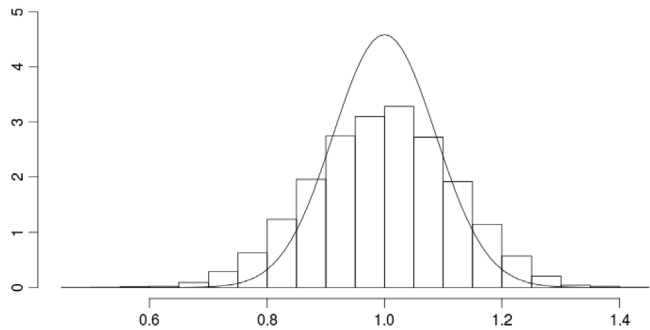
Fig. 7. $\tilde{\theta}$, $\epsilon = 0.01$.

Table 3
Estimates of $\theta_0 = 1$ with empirical confidence intervals and theoretical standard deviations.

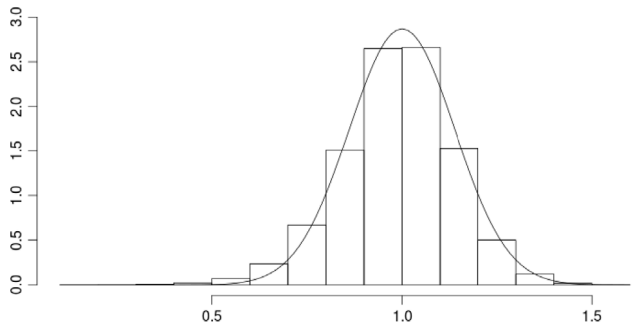| ϵ	MLE	Mean estimator	68% Confidence interval	95% Confidence interval	Theoretical SD ($\hat{\theta}^\epsilon$)
0.1	$\hat{\theta}^\varepsilon$	0.988	(0.839, 1.136)	(0.690, 1.285)	0.139
0.1	$\tilde{\theta}$	0.956	(0.659, 1.253)	(0.361, 1.551)	–
0.01	$\hat{\theta}^\varepsilon$	0.999	(0.955, 1.043)	(0.911,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

Fig. 8. $\hat{\theta}^\varepsilon$, $\epsilon = 0.1$.

$$dY_t^\varepsilon = \left[\frac{1}{\delta} f(X_t^\varepsilon, Y_t^\varepsilon) + \frac{1}{h_2(\epsilon, \delta)} g(X_t^\varepsilon, Y_t^\varepsilon)\right] dt + \frac{1}{\sqrt{\delta}} \tau_1(X_t^\varepsilon, Y_t^\varepsilon) dW_t$$
$$+ \frac{1}{\sqrt{\delta}} \tau_2(X_t^\varepsilon, Y_t^\varepsilon) dB_t$$
$$X_0^\varepsilon = x_0 \in \mathcal{X} = \mathbb{R}^{\hat{d}}, \qquad Y_0^\varepsilon = y_0 \in \mathcal{Y} = \mathbb{R}^{d-\hat{d}},$$

where d_θ satisfies the same conditions as c_θ, g satisfies the same conditions as f, $\lim_{(\epsilon,\delta) \to 0} h_i(\epsilon, \delta) = 0$ for $i \in \{1, 2\}$, and $\delta \leq h_2(\epsilon, \delta)$ for ϵ and δ sufficiently small.

If $\lim_{\varepsilon \to 0} \frac{\delta}{h_2(\epsilon,\delta)} = 0$, then it is relatively easy to see that Theorems 1, 2, 3 hold for (9) as well.
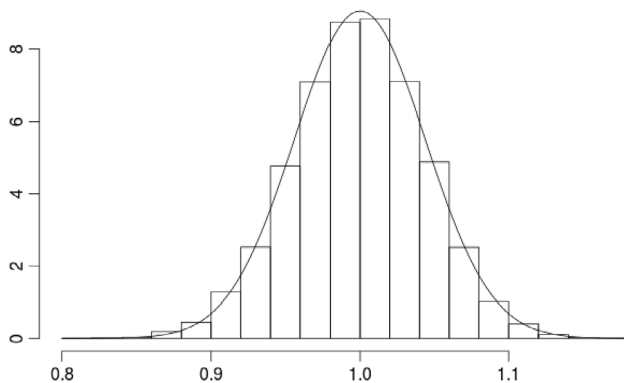
Fig. 9. $\hat{\theta}^{\varepsilon}, \epsilon = 0.01$.

If on the other hand $\lim_{\varepsilon \to 0} \frac{\delta}{h_2(\epsilon, \delta)} = \gamma \in (0, \infty)$, then replacing Condition 2 with

$$\lim_{|y| \to \infty} \sup_{x \in \mathcal{X}} \left((\gamma f + g)(x, y) \cdot y \right) = -\infty.$$

Theorems 1, 2, and 3 hold after using in every instance in place of $\mu_x(dy)$ the invariant measure associated instead with the operator

$$\mathcal{L}_x^{\gamma} = (\gamma f + g)(x, \cdot) \cdot \nabla_y + \frac{1}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, \cdot) : \nabla_y^2.$$

Thus one sees that our theorems and results are valid for the more general class of models (9). This is useful in applications where the models have more than two time-scales; see for example [13].

Secondly, it seems that the methods of this paper can be extended to address estimation of parameters in the drift coefficients of the fast process Y^{ε}. However, an investigation of the likelihood suggests that the limiting behavior of the MLE may then depend upon the limiting behavior of the ratio ϵ / δ, making the treatment somewhat more involved.

Acknowledgments

Appendix

We gather here technical results to which we appeal in the proofs of the main results of this paper. In Appendix A.1 we establish a series of estimates and bounds, including in particular an ergodic-type theorem with an explicit rate of convergence in ϵ and δ. In Appendix A.2 we use the results of Appendix A.1 to prove convergence of the slow process (Theorem 1) and of the likelihood (Lemma 1). In Appendix A.3 we complete our proof of the asymptotic normality of the MLE.

A.1. Auxiliary bounds and an ergodic-type theorem

Recall that $\bar{X}$ is a continuous, deterministic trajectory defined on a compact interval of time; as such, $\sup_{0 \le t \le T} |\bar{X}_t|$ is a finite value. Somewhat less clear is what can be said for the (stochastic) slow process X^ε. We will see in Theorem 1 that, in fact, X^ε converges to $\bar{X}$ in L^p uniformly in $t \in [0, T]$ at the rate of $\left(\sqrt{\epsilon} + \sqrt{\delta} \right)^p$. In order to prove this, however, we must first establish a less ambitious auxiliary bound; namely, that the random variables $\sup_{0 \le t \le T} |X_t^\varepsilon|$ have finite moments uniformly in ε sufficiently small. This auxiliary bound and others derived therefrom will also be used to prove our main statistical results in Theorems 2 and 3.

Lemma 6. *Assume Conditions 1 and 2. For any $p > 0$, there is a constant $\tilde{K}$ such that uniformly in ε sufficiently small,*

$$E \sup_{0 \le t \le T} |X_t^\varepsilon|^p \le \tilde{K},$$

$$E \sup_{0 \le t \le T} |X_t^\varepsilon - \bar{X}_t|^p \le \tilde{K}.$$

Proof. As $\sup_{0 \le t \le T} |\bar{X}_t|$ is a finite value, it is easy to see that the two statements are equivalent. Let us prove the first one. It is enough to prove the lemma for $p \ge 2$. Referring to (1), we recall that by definition

$$X_t^\varepsilon = x_0 + \int_0^t c_\theta(X_s^\varepsilon, Y_s^\varepsilon) ds + \sqrt{\epsilon} \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s.$$

By Condition 1, there are constants $K > 0$, $q > 0$, and $r \in [0, 1)$ such that $|c_\theta(x, y)| \le K(1 + |x|^r)(1 + |y|^q)$ and $|\sigma(x, y)| \le K(1 + |x|^{1/2})(1 + |y|^q)$. By the Burkholder–Davis–Gundy inequality and Young's inequality with conjugate exponents $\frac{1}{r}$ and $\frac{1}{1-r}$, for some constants C_j, for $t \in [0, T]$, and for ε sufficiently small,

$$E \sup_{0 \le s \le t} |X_s^\varepsilon|^p \le C_1 E \left(|x_0|^p + \int_0^t |c_\theta(X_s^\varepsilon, Y_s^\varepsilon)|^p ds + \sup_{0 \le s \le t} \left| \int_0^s \sigma(X_u^\varepsilon, Y_u^\varepsilon) dW_u \right|^p \right)$$

$$\le C_2 \left(|x_0|^p + E \int_0^t (1 + |X_s^\varepsilon|^r)^p (1 + |Y_s^\varepsilon|^q)^p ds \right.$$

$$\left. + E \int_0^t (1 + |X_s^\varepsilon|^{1/2})^p (1 + |Y_s^\varepsilon|^q)^p ds \right)$$

$$\le C_3 \left(|x_0|^p + E \int_0^t |X_s^\varepsilon|^{rp} (1 + |Y_s^\varepsilon|^q)^p ds + E \int_0^t |X_s^\varepsilon|^{p/2} (1 + |Y_s^\varepsilon|^q)^p ds \right.$$

$$\left. + E \int_0^t (1 + |Y_s^\varepsilon|^q)^p ds \right)$$

$$\le C_4 \left(|x_0|^p + E \int_0^t |X_s^\varepsilon|^p ds + E \int_0^t (1 + |Y_s^\varepsilon|^q)^{\frac{p}{1-r}} ds + E \int_0^t (1 + |Y_s^\varepsilon|^q)^{2p} ds \right.$$

$$\left. + E \int_0^t (1 + |Y_s^\varepsilon|^q)^p ds \right)$$

$$\le C_5 \left(1 + \int_0^t E \sup_{0 \le u \le s} |X_u^\varepsilon|^p ds \right),$$

where we appeal to Lemma 1 of [20] in passing to the last inequality. The proof is complete upon applying the Grönwall inequality. $\quad\square$

From here it is easy to extend integrability to functions bounded by polynomials in X^ε and Y^ε. Let us package this observation in a lemma for precision and convenience.

Lemma 7. *Assume Conditions* 1 *and* 2. *Let φ be a function of x and y satisfying $|\varphi(x, y)| \leq K(1 + |x|^r)(1 + |y|^q)$ for some fixed positive constants K, q, r and let V be either of the Wiener processes W, B. For any $\theta \in \Theta$, $\eta > 0$, and $0 \leq p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,*

$$E \int_0^T |\varphi(X_t^\varepsilon, Y_t^\varepsilon)|^p dt \leq \tilde{K}, \tag{10}$$

$$E \sup_{0 \leq t \leq T} \left| \int_0^t \varphi(X_s^\varepsilon, Y_s^\varepsilon) dV_s \right|^p \leq \tilde{K}. \tag{11}$$

Proof.

$$E \int_0^T |\varphi(X_t^\varepsilon, Y_t^\varepsilon)|^p dt \leq E \int_0^T K^p (1 + |X_t^\varepsilon|^r)(1 + |Y_t^\varepsilon|^q)^p dt$$

$$\leq \frac{K^p}{2} \left(E \int_0^T (1 + |X_t^\varepsilon|^r)^{2p} dt + E \int_0^T (1 + |Y_t^\varepsilon|^q)^{2p} dt \right);$$

the terms inside the parentheses are bounded respectively by Lemma 6 and Lemma 1 of [20], giving (10). (11) follows from (10) by the Burkholder–Davis–Gundy inequality:

$$E \sup_{0 \leq t \leq T} \left| \int_0^t \varphi(X_s^\varepsilon, Y_s^\varepsilon) dV_s \right|^p \leq E \int_0^T |\varphi(X_t^\varepsilon, Y_t^\varepsilon)|^p dt. \quad \square$$

Next we prove an ergodic-type theorem that will be used in the proofs of the main results of this paper but which may be of independent interest as well.

Theorem 4. *Assume Conditions* 1 *and* 2. *Let $\varphi(x, y)$ be a function such that $|\varphi(x, y)| \leq K(1 + |x|^r)(1 + |y|^q)$ for some fixed positive constants K, q, r and each derivative up to second order is Hölder continuous in y uniformly in x with absolute value growing at most polynomially in $|y|$ as $y \to \infty$. For any $0 \leq p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,*

$$E \sup_{0 \leq t \leq T} \left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \bar{\varphi}(X_s^\varepsilon) \right) ds \right|^p \leq \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p,$$

where $\bar{\varphi}(x)$ is the averaged function $\int_\mathcal{Y} \varphi(x, y)\mu_x(dy)$.

Proof. It is enough to prove the lemma for $p \geq 2$. By Theorem 3 in [20], the equations

$$\mathcal{L}_x \Phi(x, y) = \varphi(x, y) - \bar{\varphi}(x)$$

$$\int_\mathcal{Y} \Phi(x, y)\mu_x(dy) = 0$$

admit a unique solution Φ in the class of functions that grow at most polynomially in $|y|$ as $y \to \infty$. Applying Itô's lemma, expanding the differential, and rearranging terms we have

$$\int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \bar{\varphi}(X_s^\varepsilon) \right) ds$$

$$= \int_0^t \Big(\delta d\,\Phi - \delta \nabla_x \Phi \cdot c_\theta(X_s^\varepsilon, Y_s^\varepsilon)ds - \frac{\delta \epsilon}{2} \nabla_x^2 \Phi : \sigma \sigma^T (X_s^\varepsilon, Y_s^\varepsilon)ds$$
$$- \sqrt{\delta \epsilon}\nabla_y \nabla_x \Phi : \sigma \tau_1^T (X_s^\varepsilon, Y_s^\varepsilon)ds - \delta\sqrt{\epsilon}\nabla_x \Phi \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon)dW_s$$
$$- \sqrt{\delta}\nabla_y \Phi \cdot \tau_1(X_s^\varepsilon, Y_s^\varepsilon)dW_s - \sqrt{\delta}\nabla_y \Phi \cdot \tau_2(X_s^\varepsilon, Y_s^\varepsilon)dB_s \Big);$$

hence, for ε sufficiently small,

$$E \sup_{0 \le t \le T} \left| \int_0^t \Big(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \bar{\varphi}(X_s^\varepsilon) \Big)ds \right|^p$$

$$\le 7^p (\sqrt{\epsilon} + \sqrt{\delta})^p \Big(E \sup_{0 \le t \le T} \sqrt{\delta}^p \big| \Phi(X_t^\varepsilon, Y_t^\varepsilon) - \Phi(X_0^\varepsilon, Y_0^\varepsilon) \big|^p$$

$$+ E \int_0^T |\nabla_x \Phi \cdot c_\theta(X_s^\varepsilon, Y_s^\varepsilon)|^p ds + E \int_0^T |\nabla_x^2 \Phi : \sigma \sigma^T (X_s^\varepsilon, Y_s^\varepsilon)|^p ds$$

$$+ E \int_0^T |\nabla_y \nabla_x \Phi : \sigma \tau_1^T (X_s^\varepsilon, Y_s^\varepsilon)|^p ds + E \sup_{0 \le t \le T} \left| \int_0^t \nabla_x \Phi \cdot \sigma(X_u^\varepsilon, Y_u^\varepsilon)dW_u \right|^p$$

$$+ E \sup_{0 \le t \le T} \left| \int_0^t \nabla_y \Phi \cdot \tau_1(X_u^\varepsilon, Y_u^\varepsilon)dW_u \right|^p + E \sup_{0 \le t \le T} \left| \int_0^t \nabla_y \Phi \cdot \tau_2(X_u^\varepsilon, Y_u^\varepsilon)dB_u \right|^p \Big).$$

It remains only to show that the expected value terms inside the parentheses are uniformly bounded in ε sufficiently small. For the term $E \sup_{0 \le t \le T} \sqrt{\delta}^p \big| \Phi(X_t^\varepsilon, Y_t^\varepsilon) - \Phi(X_0^\varepsilon, Y_0^\varepsilon) \big|^p$, this follows by the argument of Corollary 1 of [19]. Meanwhile, by the same argument as in the proof of Theorem 3 in [20], all of the derivatives of Φ that appear are continuous in x and y and bounded by expressions of the form $K(1 + |x|^r)(1 + |y|^q)$; that the corresponding terms are bounded therefore follows by Lemma 7. $\square$

Finally, we prove two useful lemmata concerning the invariant measures μ_x.

Lemma 8. *Assume Conditions 1 and 2. For any $q > 0$, there is a constant $\tilde{K}$ such that*

$$\sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} (1 + |y|^q) \mu_x(dy) \le \tilde{K}.$$

Proof. By Theorem 1 in [20], the densities m_x of the measures μ_x admit, for any p, a constant C_p such that $\sup_{x \in \mathcal{X}} |m_x(y)| \le \frac{C_p}{1+|y|^p}$. Choosing p large enough that $\int_{\mathcal{Y}} \frac{1+|y|^q}{1+|y|^p} dy < \infty$,

$$\sup_{x \in \mathcal{X}} \int_{\mathcal{Y}} (1 + |y|^q) \mu_x(dy) \le \int_{\mathcal{Y}} C_p \frac{1 + |y|^q}{1 + |y|^p} dy$$
$$\le \tilde{K}. \quad \square$$

Lemma 9. *Assume Conditions 1 and 2. For any $q > 0$, there is a constant $\tilde{K}$ such that for all $(x_1, x_2) \in \mathcal{X}^2$,*

$$\int_{\mathcal{Y}} (1 + |y|^q)|\mu_{x_1} - \mu_{x_2}|(dy) \le \tilde{K}|x_1 - x_2|.$$

Proof. By Theorem 1 in [20], the densities m_x of the measures μ_x admit, for any p, a constant C_p such that $\sup_{x \in \mathcal{X}} |\nabla_x m_x(y)| \leq \frac{C_p}{1+|y|^p}$. Choosing p large enough that $\int_{\mathcal{Y}} \frac{1+|y|^q}{1+|y|^p} dy < \infty$,

$$\int_{\mathcal{Y}} (1+|y|^q)|m_{x_1} - m_{x_2}|(dy) \leq \int_{\mathcal{Y}} C_p |x_1 - x_2| \frac{1+|y|^q}{1+|y|^p} dy$$
$$\leq \tilde{K} |x_1 - x_2|. \quad \square$$

A.2. Convergence of the slow process and of the likelihood

Let us now prove Theorem 1 and Lemma 1.

Proof of Theorem 1. It is enough to prove the theorem for $p \geq 2$. We begin by noting that

$$
\begin{aligned}
|X_t^\varepsilon - \bar{X}_t|^p &= \left| \int_0^t \sqrt{\epsilon}\sigma(X_s^\varepsilon, Y_s^\varepsilon)dW_s + \int_0^t (c_\theta(X_s^\varepsilon, Y_s^\varepsilon) - \bar{c}_\theta(X_s^\varepsilon))ds \right. \\
&\quad \left. + \int_0^t (\bar{c}_\theta(X_s^\varepsilon) - \bar{c}_\theta(\bar{X}_s))ds \right|^p \\
&\leq 2^p \left(\sup_{0 \leq s \leq t} \left| \int_0^s \sqrt{\epsilon}\sigma(X_u^\varepsilon, Y_u^\varepsilon)dW_u + \int_0^s (c_\theta(X_u^\varepsilon, Y_u^\varepsilon) - \bar{c}_\theta(X_u^\varepsilon))du \right|^p \right. \\
&\quad \left. + \int_0^t |(\bar{c}_\theta(X_s^\varepsilon) - \bar{c}_\theta(\bar{X}_s))|^p ds \right).
\end{aligned}
$$

By Condition 1 and compactness of $\{\bar{X}_t\}_{0 \leq t \leq T}$, the functions $\sup_{x \in \mathcal{X}} |\nabla_x c_\theta(x, y)|$ and $\sup_{0 \leq t \leq T} |c_\theta(\bar{X}_t, y)|$ are bounded by polynomials in $|y|$. Therefore, by Lemmata 8 and 9, there is a constant K such that

$$
\begin{aligned}
|\bar{c}_\theta(X_s^\varepsilon) - \bar{c}_\theta(\bar{X}_s)| &= \left| \int_{\mathcal{Y}} (c_\theta(X_s^\varepsilon) - c_\theta(\bar{X}_s))\mu_{X_s^\varepsilon}(dy) + \int_{\mathcal{Y}} c_\theta(\bar{X}_s)(\mu_{X_s^\varepsilon} - \mu_{\bar{X}_s})(dy) \right| \\
&\leq \int_{\mathcal{Y}} \sup_{x \in \mathcal{X}} |\nabla_x c_\theta(x, y)| \sup_{x \in \mathcal{X}} |m_x(y)| dy \cdot |X_s^\varepsilon - \bar{X}_s| \\
&\quad + \int_{\mathcal{Y}} \sup_{0 \leq t \leq T} |c_\theta(\bar{X}_t, y)||\mu_{X_s^\varepsilon} - \mu_{\bar{X}_s}| dy \\
&\leq K|X_s^\varepsilon - \bar{X}_s|.
\end{aligned}
$$

Therefore, there is a (perhaps larger) constant K such that for $0 \leq t \leq T$,

$$
\begin{aligned}
|X_t^\varepsilon - \bar{X}_t|^p &\leq K \left(\sup_{0 \leq s \leq t} \left| \int_0^s \sqrt{\epsilon}\sigma(X_u^\varepsilon, Y_u^\varepsilon)dW_u + \int_0^s (c_\theta(X_u^\varepsilon, Y_u^\varepsilon) - \bar{c}_\theta(X_u^\varepsilon))du \right|^p \right. \\
&\quad \left. + \int_0^t |X_s^\varepsilon - \bar{X}_s|^p ds \right).
\end{aligned}
$$

Applying the Grönwall inequality, there is a (perhaps larger) constant K such that for $0 \leq t \leq T$,

$$|X_t^\varepsilon - \bar{X}_t|^p \leq K \sup_{0 \leq s \leq t} \left| \int_0^s \sqrt{\epsilon}\sigma(X_u^\varepsilon, Y_u^\varepsilon)dW_u + \int_0^s (c_\theta(X_u^\varepsilon, Y_u^\varepsilon) - \bar{c}_\theta(X_u^\varepsilon))du \right|^p;$$

hence,

$$
E \sup_{0 \le t \le T} |X_t^\varepsilon - \bar{X}_t|^p \le 2^p K \left(\sqrt{\epsilon}^p E \sup_{0 \le t \le T} \left| \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s \right|^p \right.
$$
$$
\left. + E \sup_{0 \le t \le T} \left| \int_0^t (c_\theta(X_s^\varepsilon, Y_s^\varepsilon) - \bar{c}_\theta(X_s^\varepsilon)) ds \right|^p \right).
$$

The conclusion follows by Lemma 7 and Theorem 4. $\square$

Combining Theorems 1 and 4, we have the following lemma:

Lemma 10. *Assume Conditions 1 and 2 and let φ be as in Theorem 4. For any fixed $\theta \in \Theta$ and $0 \le p < \infty$, there is a constant $\tilde{K}$ such that for ε sufficiently small,*

$$
E \sup_{0 \le t \le T} \left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \bar{\varphi}(\bar{X}_s) \right) ds \right|^p \le \tilde{K} (\sqrt{\epsilon} + \sqrt{\delta})^p,
$$

where $\bar{\varphi}(x)$ is the averaged function $\int_{\mathcal{Y}} \varphi(x, y) \mu_x(dy)$.

Proof. By the triangle inequality,

$$
\left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \int_{\mathcal{Y}} \varphi(\bar{X}_s, y) \mu_{\bar{X}_s}(dy) \right) ds \right|
$$
$$
\le \left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \int_{\mathcal{Y}} \varphi(X_s^\varepsilon, y) \mu_{X_s^\varepsilon}(dy) \right) ds \right|
$$
$$
+ \left| \int_0^t \int_{\mathcal{Y}} \varphi(X_s^\varepsilon, y)(\mu_{X_s^\varepsilon} - \mu_{\bar{X}_s})(dy) ds \right|
$$
$$
+ \left| \int_0^t \int_{\mathcal{Y}} (\varphi(X_s^\varepsilon, y) - \varphi(\bar{X}_s, y)) \mu_{\bar{X}_s}(dy) ds \right|.
$$

Hence,

$$
\left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \int_{\mathcal{Y}} \varphi(\bar{X}_s, y) \mu_{\bar{X}_s}(dy) \right) ds \right|^p
$$
$$
\le 3^p \left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \int_{\mathcal{Y}} \varphi(X_s^\varepsilon, y) \mu_{X_s^\varepsilon}(dy) \right) ds \right|^p
$$
$$
+ 3^p \left(\int_0^T \int_{\mathcal{Y}} |\varphi(X_t^\varepsilon, y)| \, |\mu_{X_t^\varepsilon} - \mu_{\bar{X}_t}|(dy) dt \right)^p
$$
$$
+ 3^p \left(\int_0^T \int_{\mathcal{Y}} |\varphi(X_t^\varepsilon, y) - \varphi(\bar{X}_t, y)| \mu_{\bar{X}_t}(dy) dt \right)^p.
$$

It will suffice to bound the expected value of each term on the right-hand side by an expression of the form $K(\sqrt{\epsilon} + \sqrt{\delta})^p$ or else, in light of Theorem 1, an expression of the form $KE \int_0^T |X_t^\varepsilon - \bar{X}_t|^p dt$.

By Theorem 4, there is a constant K such that

$$
3^p E \sup_{0 \le t \le T} \left| \int_0^t \left(\varphi(X_s^\varepsilon, Y_s^\varepsilon) - \int_{\mathcal{Y}} \varphi(X_s^\varepsilon, y) \mu_{X_s^\varepsilon}(dy) \right) ds \right|^p \le 3^p K (\sqrt{\epsilon} + \sqrt{\delta})^p.
$$

By Lemma 9 and the fact that $|\varphi(x, y)|$ is bounded by a polynomial in $|y|$, there is a (perhaps larger) constant K such that

$$3^p E\left(\int_0^T \int_{\mathcal{Y}} |\varphi(X_t^\varepsilon, y)| |\mu_{X_t^\varepsilon} - \mu_{\bar{X}_t}|(dy)dt\right)^p \leq 3^p \left(\int_0^T K|X_t^\varepsilon - \bar{X}_t|dt\right)^p.$$

Similarly, by Lemma 8 and the fact that $|\nabla_x \varphi(x, y)|$ is bounded by a polynomial in $|y|$, there is a (perhaps larger) constant K such that

$$3^p \left(\int_0^T \int_{\mathcal{Y}} |\varphi(X_t^\varepsilon, y) - \varphi(\bar{X}_t, y)|\mu_{\bar{X}_t}(dy)dt\right)^p \leq 3^p \left(\int_0^T K|X_t^\varepsilon - \bar{X}_t|dt\right)^p. \quad \square$$

We are ready to prove convergence of the likelihood (Lemma 1).

Proof of Lemma 1. By Lipschitz dependence in θ and total boundedness of Θ, it is enough to prove the lemma for arbitrary fixed θ.

Using the representations (4) and (5),

$$E\left|Z_\theta^\varepsilon((X^\varepsilon, Y^\varepsilon)_T) - \bar{Z}_{\theta,\theta_0}((\bar{X})_T)\right|^p$$

$$= E\left|\int_0^T \left[\langle \kappa c_\theta, \kappa c_{\theta_0}\rangle(X_t^\varepsilon, Y_t^\varepsilon) - \int_{\mathcal{Y}} \langle \kappa c_\theta, \kappa c_{\theta_0}\rangle(\bar{X}_t, y)\mu_{\bar{X}_t}(dy)\right]dt\right.$$

$$- \frac{1}{2}\int_0^T \left[|\kappa c_\theta|^2(X_t^\varepsilon, Y_t^\varepsilon) - \int_{\mathcal{Y}} |\kappa c_\theta|^2(\bar{X}_t, y)\mu_{\bar{X}_t}(dy)\right]dt$$

$$+ \sqrt{\epsilon}\int_0^T \langle \kappa c_\theta, \kappa \cdot dW_t\rangle(X_t^\varepsilon, Y_t^\varepsilon)$$

$$\left. - \sqrt{\epsilon}\int_0^T \langle (\tau_2\tau_2^T)^{-1}\tau_1\sigma^T(\sigma\sigma^T)^{-1}c_\theta, (\tau_1 dW_t + \tau_2 dB_t)\rangle(X_t^\varepsilon, Y_t^\varepsilon)\right|^p$$

$$\leq 4^p E\left|\int_0^T \left[\langle \kappa c_\theta, \kappa c_{\theta_0}\rangle(X_t^\varepsilon, Y_t^\varepsilon) - \int_{\mathcal{Y}} \langle \kappa c_\theta, \kappa c_{\theta_0}\rangle(\bar{X}_t, y)\mu_{\bar{X}_t}(dy)\right]dt\right|^p$$

$$+ 2^p E\left|\int_0^T \left[|\kappa c_\theta|^2(X_t^\varepsilon, Y_t^\varepsilon) - \int_{\mathcal{Y}} |\kappa c_\theta|^2(\bar{X}_t, y)\mu_{\bar{X}_t}(dy)\right]dt\right|^p$$

$$+ 4^p E\left|\sqrt{\epsilon}\int_0^T \langle \kappa c_\theta, \kappa \cdot dW_t\rangle(X_t^\varepsilon, Y_t^\varepsilon)\right|^p$$

$$+ 4^p E\left|\sqrt{\epsilon}\int_0^T \langle (\tau_2\tau_2^T)^{-1}\tau_1\sigma^T(\sigma\sigma^T)^{-1}c_\theta, (\tau_1 dW_t + \tau_2 dB_t)\rangle(X_t^\varepsilon, Y_t^\varepsilon)\right|^p.$$

As $\varepsilon \to 0$, the first and second terms tend to zero by Lemma 10 while the third and fourth tend to zero by Lemma 7 of Appendix A.1. $\quad \square$

A.3. Lemmata establishing asymptotic normality

In this section we establish Lemmata 2 and 4 to complete the proof of Theorem 3. Recall from Section 5 that these lemmata concern the log-likelihood-ratio

$$M_\theta^\varepsilon(u) = \frac{1}{\sqrt{\epsilon}}\int_0^T \langle \kappa(c_{\theta+\phi u} - c_\theta), d(W, B)_t\rangle(X_t^\varepsilon, Y_t^\varepsilon)$$

$$- \frac{1}{2\epsilon}\int_0^T |\kappa(c_{\theta+\phi u} - c_\theta)|^2(X_t^\varepsilon, Y_t^\varepsilon)dt,$$

where

$$\kappa = \begin{pmatrix} \sigma^T (\sigma \sigma^T)^{-1} \\ -\tau_2^T (\tau_2 \tau_2^T)^{-1} \tau_1 \sigma^T (\sigma \sigma^T)^{-1} \end{pmatrix},$$

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

and $I(\theta)$ is the Fisher information matrix. More concisely, we can write

$$M_\theta^\varepsilon(u) = \int_0^T H(X_t^\varepsilon, Y_t^\varepsilon) d(W, B)_t - \frac{1}{2} \int_0^T |H(X_t^\varepsilon, Y_t^\varepsilon)|^2 dt,$$

where by definition

$$H(x, y) = \frac{1}{\sqrt{\epsilon}} \kappa(x, y)(c_{\theta+\phi u} - c_\theta)(x, y)$$

$$= \kappa(x, y) \int_0^1 \nabla_\theta c_{\theta+h\phi u}(x, y) dh \cdot I^{-1/2}(\theta) u. \tag{12}$$

It is clear that $H(x, y) = H(x, y; u, \epsilon, \theta)$ (i.e., it also depends on the parameters u, ϵ, θ); we nevertheless suppress the additional parameters for the sake of brevity.

Proof of Lemma 2. We must show that for any compact $\tilde{\Theta} \subset \Theta$ and sequences $\{\theta_n\}_{n=1}^\infty \subset \tilde{\Theta}$, $\{\varepsilon_n = (\epsilon_n, \delta_n)\}_{n=1}^\infty \subset \mathbb{R}_+^2$, and $\{u_n\}_{n=1}^\infty \subset \mathbb{R}^p$ with $\varepsilon_n \to 0$, $u_n \to u$, and $\{\theta_n + \phi(\epsilon_n, \theta_n)u_n\}_{n=1}^\infty \subset \Theta$, we have a representation

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

with $\Delta_n \Rightarrow \mathcal{N}(0, I)$ in $P_{\theta_n}^{\varepsilon_n}$-law and $\lim_{n\to\infty} P_{\theta_n}^{\varepsilon_n}\left(|\psi_{\varepsilon_n}(u_n, \varepsilon_n, \theta_n)| > \eta\right) = 0$ for each $\eta > 0$.

Setting $\phi_n = \phi(\epsilon_n, \theta_n)$ and $H_n = \frac{1}{\sqrt{\epsilon_n}}\kappa(c_{\theta_n+\phi_n u_n} - c_{\theta_n})$, $P_{\theta_n}^{\varepsilon_n}$-almost surely,

$$M_{\theta_n}^{\varepsilon_n}(u_n) = \int_0^T H_n(X_t^{\varepsilon_n}, Y_t^{\varepsilon_n}) d(W, B)_t - \frac{1}{2} \int_0^T |H_n(X_t^{\varepsilon_n}, Y_t^{\varepsilon_n})|^2 dt$$

$$= J_n^1 + J_n^2 + J_n^3 + J_n^4,$$

where, writing for brevity $\theta_{\varepsilon,n} = \theta_n + \phi(\epsilon_n, \theta_n)u_n$,

$$J_n^1 = \int_0^T \left\langle \left(H_n - \kappa \nabla_\theta c_{\theta_n+\phi_n u_n} \cdot I^{-1/2}(\theta_n)u_n\right), d(W, B)_t \right\rangle (X_t^{\varepsilon_n}, Y_t^{\varepsilon_n}),$$

$$J_n^2 = \frac{1}{\sqrt{\epsilon_n}} \int_0^T \left\langle \kappa \nabla_\theta c_{\theta_{\varepsilon,n}} \cdot (\theta_{\varepsilon,n} - \theta_n), d(W, B)_t \right\rangle (X_t^{\varepsilon_n}, Y_t^{\varepsilon_n})$$

$$= \left\langle u_n, I^{-1/2}(\theta_n) \int_0^T \left\langle \kappa \nabla_\theta c_{\theta_n+\phi_n u_n}, d(W, B)_t \right\rangle (X_t^{\varepsilon_n}, Y_t^{\varepsilon_n}) \right\rangle,$$

$$J_n^3 = \frac{1}{2}\left(|u_n|^2 - \int_0^T |H_n(X_t^{\varepsilon_n}, Y_t^{\varepsilon_n})|^2 dt\right)$$

$$= \frac{1}{2} \int_0^T \left[\langle I^{-1/2}(\theta_n)u_n, Q^{1/2}(\bar{X}_t, \theta_n)\rangle^2 - |H_n(X_t^{\varepsilon_n}, Y_t^{\varepsilon_n})|\right] dt,$$

$$J_n^4 = -\frac{1}{2}|u_n|^2.$$

J_n^2 converges in distribution to (u, Δ), where $\Delta \sim \mathcal{N}(0, I)$; J_n^4 converges to $-\frac{1}{2}|u|^2$. Let us then show that

$$\sup_{\theta \in \tilde{\Theta}} E\left[|J_n^1|^2 + |J_n^3|^2\right] \to 0.$$

We notice that there is a constant K such that

$$E|J_n^1|^2 = E\left|\int_0^T \left\langle \left(H_n - \kappa \nabla_\theta c_{\theta_n + \phi_n u_n} \cdot I^{-1/2}(\theta_n) u_n\right), d(W, B)_t \right\rangle (X_t^{\varepsilon_n}, Y_t^{\varepsilon_n}) \right|^2$$

$$\leq K \sup_{\theta \in \tilde{\Theta}} \sup_{|v| \leq K\sqrt{\varepsilon_n}} E\left|\int_0^T \left|\kappa \left(\nabla_\theta c_{\theta+v} - \nabla_\theta c_\theta\right) (X_t^{\varepsilon_n}, Y_t^{\varepsilon_n})\right|^2 dt\right|$$

$$\to 0, \tag{13}$$

where the last convergence follows by the uniform continuity of $\nabla_\theta c_\theta$ in $\theta \in \tilde{\Theta}$, tightness of $(X_t^{\varepsilon_n}, Y_t^{\varepsilon_n})$, and the fact that expressions of the form $E\int_0^T \left(1 + |X_t^\varepsilon|^r\right)\left(1 + |Y_t^\varepsilon|^q\right) dt$ are bounded uniformly in ε sufficiently small (recall that $\varepsilon_n \to 0$).

Meanwhile, Theorem 1 and the continuous dependence of the coefficients on the parameter θ together imply that $E|J_n^3|^2 \to 0$ uniformly in $\theta \in \tilde{\Theta}$, completing the proof of the lemma. $\square$

Before giving a proof of Lemma 4, we gather necessary estimates in Lemmata 11 and 12.

Lemma 11. *Assume Conditions 1, 2, 4, and 5. Let $H(x, y)$ be as in (12). For any compact $\tilde{\Theta} \subset \Theta$, there is a constant $\hat{K}$ such that uniformly in $\theta \in \tilde{\Theta}$ and ε sufficiently small,*

$$\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}(dy) dt \geq \hat{K}|u|^2.$$

Proof. Using the inequality $|a|^2 \geq |b|^2 - 2|\langle b, (a - b)\rangle|$ with $a = H(\bar{X}_t, y) = \kappa(\bar{X}_t, y) \int_0^1 \nabla_\theta c_{\theta+h\phi u}(\bar{X}_t, y) dh \cdot I^{-1/2}(\theta) u$ and $b = \kappa(\bar{X}_t, y) \nabla_\theta c_\theta(\bar{X}_t, y) \cdot I^{-1/2}(\theta) u$,

$$\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}(dy) dt \geq I - II,$$

where, recalling the definition of the Fisher information matrix $I(\theta)$,

$$I = \int_0^T \int_{\mathcal{Y}} |\kappa \nabla_\theta c_\theta \cdot I^{-1/2}(\theta) u|^2 (\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt$$

$$= |u|^2$$

and

$$II = 2\int_0^T \int_{\mathcal{Y}} \left|\left\langle \kappa \nabla_\theta c_\theta \cdot I^{-1/2}(\theta) u, \kappa \left(\int_0^1 \nabla_\theta c_{\theta+h\phi u} dh - \nabla_\theta c_\theta\right) \cdot I^{-1/2}(\theta) u \right\rangle\right|$$

$$\times (\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt$$

$$\leq 2\int_0^T \int_{\mathcal{Y}} \left|\left(\nabla_\theta c_\theta \cdot I^{-1/2}(\theta)\right)^T \kappa^T \kappa \left(\int_0^1 \nabla_\theta c_{\theta+h\phi u} dh - \nabla_\theta c_\theta\right) \cdot I^{-1/2}(\theta)\right|$$

$$\times (\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt \cdot |u|^2.$$

By Lemma 8, expressions of the form $\int_{\mathcal{Y}} |y|^p \mu_x(dy)$ are bounded uniformly in x. Using this fact, the polynomial bounds in Condition 1, and the nondegeneracy of the Fisher information matrix in Condition 5,

$$\int_0^T \int_{\mathcal{Y}} \left| \left(\nabla_\theta c_\theta \cdot I^{-1/2}(\theta) \right)^T \kappa^T \kappa \left(\int_0^1 \nabla_\theta c_{\theta+h\phi u} dh - \nabla_\theta c_\theta \right) \cdot I^{-1/2}(\theta) \right|$$
$$\times (\bar{X}_t, y) \mu_{\bar{X}_t}(dy) dt \to 0$$

uniformly in $\theta \in \Theta$ as $\phi = \phi(\epsilon, \theta) = \sqrt{\epsilon} I^{-1/2}(\theta) \to 0$ as $\epsilon \to 0$. The result follows. $\qquad \square$

Lemma 12. *Assume Conditions 1, 2, 4, and 5. Let $H(x, y)$ be defined as in* (12). *Assume that ϵ does not decay too quickly relative to δ as $\varepsilon = (\epsilon, \delta) \to 0$; that is, suppose there is an $\alpha > 0$ such that we are interested (at least when ε is sufficiently small) only in pairs $\varepsilon = (\epsilon, \delta)$ satisfying $0 < \delta \le \epsilon^\alpha$. For every $0 < \gamma, N < \infty$ and for every compact $\tilde{\Theta} \subset \Theta$, there is a constant $\tilde{K}$ such that for ε sufficiently small, uniformly in $\theta \in \tilde{\Theta}$, for every u satisfying $\theta + \phi u \in \tilde{\Theta}$,*

$$P_\theta^\varepsilon \left(\int_0^T \left(\int_{\mathcal{Y}} |H(X_t^\varepsilon, y)|^2 \mu_{X_t^\varepsilon}(dy) - \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}(dy) \right) dt \ge \gamma |u|^2 \right) \le \frac{\tilde{K}}{|u|^{2N}}.$$

Proof. By the triangle inequality,

$$P_\theta^\varepsilon \left(\int_0^T \left(\int_{\mathcal{Y}} |H(X_t^\varepsilon, y)|^2 \mu_{X_t^\varepsilon}(dy) - \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}(dy) \right) dt \ge \gamma |u|^2 \right) \le I + II,$$

where

$$I = P_\theta^\varepsilon \left(\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 (\mu_{\bar{X}_t} - \mu_{X_t^\varepsilon})(dy) dt \ge \frac{\gamma}{2} |u|^2 \right),$$
$$II = P_\theta^\varepsilon \left(\int_0^T \int_{\mathcal{Y}} \left(|H(\bar{X}_t, y)|^2 - |H(X_t^\varepsilon, y)|^2 \right) \mu_{X_t^\varepsilon}(dy) dt \ge \frac{\gamma}{2} |u|^2 \right).$$

Consider the first term, I. By Conditions 1 and 5 and compactness of $\{\bar{X}_t\}_{0 \le t \le T}$, the function $\sup_{0 \le t \le T} |H(\bar{X}_t, y)|^2 / |u|^2$ is bounded by a polynomial in $|y|$. Therefore, by Lemma 9, there is a constant $K < \infty$ such that

$$\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 (\mu_{\bar{X}_t} - \mu_{X_t^\varepsilon})(dy) dt \le K \left(\int_0^T |X_t^\varepsilon - \bar{X}_t| dt \right) \cdot |u|^2.$$

Hence, for any $M < \infty$, by the Markov inequality and Theorem 1, there is a (perhaps larger) constant K such that

$$I = P_\theta^\varepsilon \left(\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 (\mu_{\bar{X}_t} - \mu_{X_t^\varepsilon})(dy) dt \ge \frac{\gamma}{2} |u|^2 \right)$$
$$\le \frac{E_\theta^\varepsilon \left(\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 (\mu_{\bar{X}_t} - \mu_{X_t^\varepsilon})(dy) dt \right)^M}{\left(\frac{\gamma}{2} |u|^2 \right)^M}$$
$$\le K (\sqrt{\epsilon} + \sqrt{\delta})^M.$$

Recall that (by assumption, at least when ε is sufficiently small) there is an $\alpha > 0$ such that $0 < \delta \leq \epsilon^{\alpha}$. Setting $M = \frac{2N}{\alpha \wedge 1}$, a simple calculation verifies that

$$(\sqrt{\epsilon} + \sqrt{\delta})^{\frac{2N}{\alpha \wedge 1}} \leq 2^{2N/(\alpha \wedge 1)} \epsilon^{N};$$

hence,

$$I \leq K \cdot 2^{2N/(\alpha \wedge 1)} \epsilon^{N}.$$

Finally, recall that u must satisfy $\theta + \phi u \in \tilde{\Theta} \subset \Theta$. Since Θ is bounded, there is a constant R for which all admissible u satisfy $\sqrt{\epsilon} u \leq R$. Using this fact, we continue the inequality to obtain

$$I \leq \frac{K \cdot 2^{2N/(\alpha \wedge 1)} \cdot R^{2N}}{|u|^{2N}}.$$

It remains to treat the second term, $II = P_{\theta}^{\varepsilon} \left(\int_0^T \int_{\mathcal{Y}} \left(|H(\bar{X}_t, y)|^2 - |H(X_t^{\varepsilon}, y)|^2 \right) \mu_{X_t^{\varepsilon}} (dy) dt \geq \frac{\gamma}{2} |u|^2 \right)$. By Conditions 1 and 5 and compactness of $\{\bar{X}_t\}_{0 \leq t \leq T}$, there are (new) constants $K, q, r > 0$ such that

$$\left(|H(\bar{X}_t, y)|^2 - |H(X_t^{\varepsilon}, y)|^2 \right) \leq K(|\bar{X}_t|^r + |X_t^{\varepsilon}|^r)(1 + |y|^q)|u|^2 |\bar{X}_t - X_t^{\varepsilon}|.$$

By Lemmata 8 and 6 and Theorem 1, by arguments analogous to those just made for the first term, I, there is a (perhaps larger) constant K such that

$$II = P_{\theta}^{\varepsilon} \left(\int_0^T \int_{\mathcal{Y}} \left(|H(\bar{X}_t, y)|^2 - |H(X_t^{\varepsilon}, y)|^2 \right) \mu_{X_t^{\varepsilon}} (dy) dt \geq \frac{\gamma}{2} |u|^2 \right)$$

$$\leq \frac{E_{\theta}^{\varepsilon} \left(\int_0^T \int_{\mathcal{Y}} (|H(\bar{X}_t, y)|^2 - |H(X_t^{\varepsilon}, y)|^2) \mu_{X_t^{\varepsilon}} (dy) dt \right)^{2N/\alpha}}{\left(\frac{\gamma}{2} |u|^2 \right)^{2N/\alpha}}$$

$$\leq K (\sqrt{\epsilon} + \sqrt{\delta})^{2N/\alpha}$$

$$\leq K \cdot 2^{2N/(\alpha \wedge 1)} \epsilon^{N}$$

$$\leq \frac{K \cdot 2^{2N/(\alpha \wedge 1)} \cdot R^{2N}}{|u|^{2N}}.$$

The proof is complete upon combining the estimates for I and II. □

We now give a proof of Lemma 4.

Proof of Lemma 4. For any $\gamma > 0$, we have by Hölder's inequality that

$$E_{\theta}^{\varepsilon} \left[e^{\frac{1}{2} M_{\theta}^{\varepsilon}(u)} \right] = E_{\theta}^{\varepsilon} \left[e^{\frac{1}{2} M_{\theta}^{\varepsilon}(u)} \cdot \chi_{\{M_{\theta}^{\varepsilon}(u) < -\gamma |u|^2\}} \right] + E_{\theta}^{\varepsilon} \left[e^{\frac{1}{2} M_{\theta}^{\varepsilon}(u)} \cdot \chi_{\{M_{\theta}^{\varepsilon}(u) \geq -\gamma |u|^2\}} \right]$$

$$\leq e^{-\frac{\gamma}{2} |u|^2} + \left(E_{\theta}^{\varepsilon} \left[e^{M_{\theta}^{\varepsilon}(u)} \right] \right)^{1/2} \left(P_{\theta}^{\varepsilon} (M_{\theta}^{\varepsilon}(u) \geq -\gamma |u|^2) \right)^{1/2}$$

$$\leq e^{-\frac{\gamma}{2} |u|^2} + \left(P_{\theta}^{\varepsilon} (M_{\theta}^{\varepsilon}(u) \geq -\gamma |u|^2) \right)^{1/2}.$$

Therefore, it suffices to show that for any $N > 1$ there are constants $\gamma > 0$ and $\tilde{K} > 0$ (which may depend on N) such that

$$\sup_{0 < \epsilon < \epsilon_0, 0 < \delta \le \epsilon^\alpha} \sup_{\theta \in \tilde{\Theta}} P_\theta^\varepsilon(M_\theta^\varepsilon(u) \ge -\gamma|u|^2) \le \frac{\tilde{K}}{|u|^{2N}}.$$

By Lemma 11 there is a positive constant $\hat{K} < \infty$ such that

$$\int_0^T \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}(dy) dt \ge \hat{K}|u|^2$$

uniformly in $\theta \in \Theta$ and ε sufficiently small. Choosing $\gamma = \frac{\hat{K}}{4}$,

$$\begin{aligned}
P_\theta^\varepsilon(M_\theta^\varepsilon(u) \ge -\gamma|u|^2) &= P_\theta^\varepsilon\left(\int_0^T H(X_t^\varepsilon, Y_t^\varepsilon)d(W, B)_t \right.\\
&\qquad \left. -\frac{1}{2}\int_0^T |H(X_t^\varepsilon, Y_t^\varepsilon)|^2 dt \ge -\gamma|u|^2\right)\\
&\le P_\theta^\varepsilon\left(\int_0^T H(X_t^\varepsilon, Y_t^\varepsilon)d(W, B)_t \right.\\
&\qquad \left. +\frac{1}{2}\int_0^T \left(\int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}(dy) - |H(X_t^\varepsilon, Y_t^\varepsilon)|^2\right)dt \ge \gamma|u|^2\right)\\
&\le I + II + II,
\end{aligned}$$

where

$$\begin{aligned}
I &= P_\theta^\varepsilon\left(\int_0^T H(X_t^\varepsilon, Y_t^\varepsilon)d(W, B)_t \ge \frac{\gamma}{4}|u|^2\right),\\
II &= P_\theta^\varepsilon\left(\int_0^T \left(\int_{\mathcal{Y}} |H(X_t^\varepsilon, y)|^2 \mu_{X_t^\varepsilon}(dy) - |H(X_t^\varepsilon, Y_t^\varepsilon)|^2\right)dt \ge \frac{\gamma}{2}|u|^2\right),\\
III &= P_\theta^\varepsilon\left(\int_0^T \left(\int_{\mathcal{Y}} |H(X_t^\varepsilon, y)|^2 \mu_{X_t^\varepsilon} - \int_{\mathcal{Y}} |H(\bar{X}_t, y)|^2 \mu_{\bar{X}_t}\right)dt \ge \frac{\gamma}{2}|u|^2\right).
\end{aligned}$$

The necessary estimate for the third term, III, is given by Lemma 12 with γ replaced by $\hat{K}/8$. It remains to treat the first term, I, and the second term, II.

For I, it is easy to see that by Condition 1 and the nondegeneracy of the Fisher information matrix $I(\theta)$ in Condition 5 there is a (perhaps larger) constant K such that

$$\begin{aligned}
I &= P_\theta^\varepsilon\left(\int_0^T H(X_t^\varepsilon, Y_t^\varepsilon)d(W, B)_t \ge \frac{\gamma}{2}|u|^2\right) \le \frac{E_\theta^\varepsilon\left(\int_0^T H(X_t^\varepsilon, Y_t^\varepsilon)d(W, B)_t\right)^{2N}}{\left(\frac{\gamma}{2}|u|^2\right)^{2N}}\\
&\le \frac{E_\theta^\varepsilon \int_0^T \left(\kappa(X_t^\varepsilon, Y_t^\varepsilon) \int_0^1 \nabla_\theta c_{\theta+h\phi u}(X_t^\varepsilon, Y_t^\varepsilon)dh \cdot I^{-1/2}(\theta)u\right)^{2N} dt}{\left(\frac{\gamma}{2}|u|^2\right)^{2N}}\\
&\le \frac{K}{|u|^{2N}}.
\end{aligned}$$

As for II, by Theorem 4 applied to the function $\varphi(x, y) = |H(x, y)|^2/|u|^2$ there is a (perhaps larger) constant K such that for ϵ sufficiently small and $0 < \delta \leq \epsilon^\alpha$,

$$
\begin{aligned}
II &= P_\theta^\varepsilon\left(\int_0^T\left(\int_{\mathcal{Y}} |H(X_t^\varepsilon, y)|^2 \mu_{X_t^\varepsilon}(dy) - |H(X_t^\varepsilon, Y_t^\varepsilon)|^2\right)dt \geq \frac{\gamma}{2}|u|^2\right)\\
&\leq \frac{E_\theta^\varepsilon\left(\int_0^T\left(\int_{\mathcal{Y}} |H(X_t^\varepsilon, y)|^2 \mu_{X_t^\varepsilon}(dy) - |H(X_t^\varepsilon, Y_t^\varepsilon)|^2\right)dt\right)^{(2N/\alpha\wedge 1)}}{\left(\frac{\gamma}{2}|u|^2\right)^{(2N/\alpha\wedge 1)}}\\
&\leq K(\sqrt{\epsilon} + \sqrt{\delta})^{(2N/\alpha\wedge 1)}\\
&\leq K 2^{2N/(\alpha\wedge 1)}\epsilon^N\\
&\leq \frac{\tilde{K}}{|u|^{2N}},
\end{aligned}
$$

where the last two inequalities follow from arguments analogous to those made in the proof of Lemma 12.

The proof is complete upon combining the estimates for I, II, and III. $\square$

References

[1] R. Azencott, A. Beri, A. Jain, I. Timofeyev, Sub-sampling and parametric estimation for multiscale dynamics, Commun. Math. Sci. 11 (4) (2013) 939–970.

[2] R. Azencott, A. Beri, I. Timofeyev, Adaptive sub-sampling for parametric estimation of Gaussian diffusions, J. Stat. Phys. 139 (6) (2010) 1066–1089.

[3] A. Bensoussan, J.-L. Lions, G. Papanicolaou, Asymptotic methods in periodic structures, in: Studies in Mathematics and its Applications, vol. 5, Elsevier, 1978.

[4] J.P.N. Bishwal, Parameter estimation in stochastic differential equations, in: Lecture Notes in Mathematics, vol. 1923, Springer Science & Business Media, 2008.

[5] A. Chauvière, L. Preziosi, C. Verdier, Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling, CRC Press, 2010.

[6] P. Dupuis, K. Spiliopoulos, H. Wang, Rare event simulation for rough energy landscapes, in: Proceedings of the 2011 Winter Simulation Conference, IEEE, 2011, pp. 504–515.

[7] J. Feng, M. Forde, J.-P. Fouque, Short-maturity asymptotics for a fast mean-reverting heston stochastic volatility model, SIAM J. Financial Math. 1 (1) (2010) 126–141.

[8] J. Feng, J.-P. Fouque, R. Kumar, Small-time asymptotics for fast mean-reverting stochastic volatility models, Ann. Appl. Probab. 22 (4) (2012) 1541–1575.

[9] J.-P. Fouque, G. Papanicolaou, K.R. Sircar, Derivatives in Financial Markets with Stochastic Volatility, Cambridge University Press, 2000.

[10] M.I. Freidlin, A.D. Wentzell, Random perturbations of dynamical systems, second ed., in: A Series of Comprehensive Studies in Mathematics, vol. 260, Springer Science & Business Media, 1998.

[11] I.A. Ibragimov, R.Z. Has'minskii, Statistical estimation: asymptotic theory, in: Applications of Mathematics, vol. 16, Springer Science & Business Media, 1981.

[12] W. Janke, Rugged Free-Energy Landscapes, in: Lecture Notes in Physics, vol. 736, Springer Science & Business Media, 2008.

[13] V.K. Jirsa, W.C. Stacey, P.P. Quilichini, A.I. Ivanov, C. Bernard, On the nature of seizure dynamics, Brain 137 (8) (2014) 2210–2230.

[14] S. Krumscheid, G.A. Pavliotis, S. Kalliadasis, Semiparametric drift and diffusion estimation for multiscale diffusions, Multiscale Model. Simul. 11 (2) (2013) 442–473.

[15] Y. Kutoyants, Identification of Dynamical Systems with Small Noise, Kluwer Academic Publishers, 1994.

[16] Y. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, Springer Science & Business Media, 2004.

[17] A.J. Majda, C. Franzke, B. Khouider, An applied mathematics perspective on stochastic modelling for climate, Phil. Trans. R. Soc. A 366 (1875) (2008) 2429–2455.

[18] A. Papavasiliou, G.A. Pavliotis, A.M. Stuart, Maximum likelihood drift estimation for multiscale diffusions, Stochastic Process. Appl. 119 (10) (2009) 3173–3210.

[19] E. Pardoux, A.Y. Veretennikov, On the Poisson equation and diffusion approximation 1, Ann. Probab. 29 (3) (2001) 1061–1085.

[20] E. Pardoux, A.Y. Veretennikov, On Poisson equation and diffusion approximation 2, Ann. Probab. 31 (3) (2003) 1166–1192.

[21] G.A. Pavliotis, A.M. Stuart, Parameter estimation for multiscale diffusions, J. Stat. Phys. 127 (4) (2007) 741–781.

[22] G.A. Pavliotis, A.M. Stuart, Multiscale methods: averaging and homogenization, in: Texts in Applied Mathematics, vol. 53, Springer Science & Business Media, 2008.

[23] B.L.S.P. Rao, Statistical Inference for Diffusion Type Processes, Arnold, 1999.

[24] K. Spiliopoulos, Fluctuation analysis and short time asymptotics for multiple scales diffusion processes, Stoch. Dyn. 14 (3) (2014) 1350026.

[25] K. Spiliopoulos, A. Chronopoulou, Maximum likelihood estimation for small noise multiscale diffusions, Stat. Inference Stoch. Process. 16 (3) (2013) 237–266.

[26] A.Y. Veretennikov, On polynomial mixing bounds for stochastic differential equations, Stochastic Process. Appl. 70 (1) (1997) 115–127.

[27] L. Zhang, P.A. Mykland, Y. Aït-Sahalia, A tale of two time scales, J. Amer. Statist. Assoc. 100 (472) (2005) 1394–1411.

[28] R. Zwanzig, Diffusion in a rough potential, Proc. Natl. Acad. Sci. USA 85 (7) (1988) 2029–2030.