Parameter Estimation for Multiscale Diffusions

G. A. Pavliotis[1] and A. M. Stuart[2]

We study the problem of parameter estimation for time-series possessing two, widely separated, characteristic time scales. The aim is to understand situations where it is desirable to fit a homogenized single-scale model to such multiscale data. We demonstrate, numerically and analytically, that if the data is sampled too finely then the parameter fit will fail, in that the correct parameters in the homogenized model are not identified. We also show, numerically and analytically, that if the data is subsampled at an appropriate rate then it is possible to estimate the coefficients of the homogenized model correctly.

The ideas are studied in the context of thermally activated motion in a two-scale potential. However the ideas may be expected to transfer to other situations where it is desirable to fit an averaged or homogenized equation to multiscale data.

1. INTRODUCTION

Parameter estimation for continuous time stochastic models is an increasingly important part of the overall modeling strategy in a wide variety of applications. It is quite often the case that the data to be fitted to a diffusion process has a multiscale character. One example is the field of molecular dynamics, where it is desirable to find effective models for low dimensional phenomena (such as conformational dynamics, vacancy diffusion and so forth) which are embedded within higher dimensional time-series. Another example is the ocean-atmosphere sciences where it is desirable to find effective models for large-scale structures, whilst representing the small-scales stochastically. The multiscale structure of the

---

[1] Department of Mathematics, Imperial College London, London SW7 2AZ, UK; e-mail: g.pavliotis@ imperial.ac.uk.
[2] Mathematics Institute, Warwick University, Coventry CV4 7AL, UK; e-mail: stuart@maths. warwick.ac.uk.

data in these problems renders the problem of parameter estimation very subtle, and great care has to be taken in order to estimate the coefficients correctly. The aim of the paper is to shed light on this estimation problem through the study of a simple class of model problems, typical of those arising in molecular dynamics. In particular, we will work under the assumption of scale separation.

Suppose that we are presented with data (either real or from a computational model) from a physical system which possesses many characteristic length and time scales. There are two important questions which then arise naturally:

- *Question 1.* Is the data compatible with the assumption of scale separation; if yes, what is the size of the small parameter ϵ whose inverse measures this scale separation?
- *Question 2.* Given knowledge of ϵ, how should the data be used to facilitate parameter estimation in a single-scale, coarse grained model which does not contain the small scale separation parameter ϵ?

Question 1 is a hard and challenging problem which we leave for future study. In this paper we study Question 2. To generate the data we employ multiscale diffusion processes. We use this data to fit a single-scale homogenized equation, investigating how to use the data to obtain the correct effective or single-scale model. The *key issue* is that the data is only compatible with the effective model if it is subsampled at an appropriate rate. On fine scales the underlying multiscale character of the data dominates.

There are two forms of multiscale diffusions which are of particular interest in the context of parameter estimation. The first gives rise to *averaging* for SDEs, and the second to *homogenization* for SDEs. For averaging one has, for $\epsilon \ll 1$,

$$dx^{\epsilon}(t) = f(x^{\epsilon}(t), y^{\epsilon}(t)) \, dt + \alpha \, (x^{\epsilon}(t), y^{\epsilon}(t)) \, dU(t), \tag{1.1a}$$

$$dy^{\epsilon}(t) = \frac{1}{\epsilon} g(x^{\epsilon}(t), y^{\epsilon}(t)) \, dt + \frac{1}{\sqrt{\epsilon}} \beta(x^{\epsilon}(t), y^{\epsilon}(t)) \, dV(t), \tag{1.1b}$$

with U, V standard Brownian motions. Averaging f and $\alpha \alpha^{T}$ over the invariant measure of the y^{ϵ} equation, with x^{ϵ} viewed as fixed, gives an averaged SDE for x. The fast process y, with timescale ϵ, is eliminated. For homogenization one has

$$dx^{\epsilon}(t) = \left(\frac{1}{\epsilon} f_0(x^{\epsilon}(t), y^{\epsilon}(t)) + f_1(x^{\epsilon}(t), y^{\epsilon}(t)) \right) dt$$
$$+ \alpha(x^{\epsilon}(t), y^{\epsilon}(t)) \, dU(t), \tag{1.2a}$$

$$dy^{\epsilon}(t) = \frac{1}{\epsilon^2} g(x^{\epsilon}(t), y^{\epsilon}(t)) \, dt + \frac{1}{\epsilon} \beta(x^{\epsilon}(t), y^{\epsilon}(t)) \, dV(t), \tag{1.2b}$$

where it is assumed that f_0 averages to zero against the invariant measure of the fast process y^{ϵ} with x^{ϵ} fixed. Now y^{ϵ} has time-scale ϵ^2 and is eliminated. The

fluctuations in f_0, suitably amplified by ϵ^{-1}, induce $\mathcal{O}(1)$ effects in the homogenized equation for x^ϵ. In both cases (1.1) and (1.2) it is possible to show [5] that the process $x^\epsilon(t)$ converges in law, as $\epsilon \to 0$, to the solution of an effective SDE of the form

$$dx(t) = F(x(t)) \, dt + A(x(t)) \, dU(t). \tag{1.3}$$

Explicit formulae can be derived for the effective coefficients $F(x)$ and $A(x)$ in the above equation.[5,25] A natural question that arises then is how to fit an SDE of the form (1.3) to data generated by a multiscale stochastic equation of the form (1.1) or (1.2), under the assumption of scale separation, i.e. when $\epsilon \ll 1$. This paper is a first attempt towards the study of this interesting problem, for a specific class of SDEs of the form (1.2).

Our basic model will be the first order Langevin equation

$$dx^\epsilon(t) = -\nabla V \left(x^\epsilon(t), \frac{x^\epsilon(t)}{\epsilon}; \alpha \right) dt + \sqrt{2\sigma} \, d\beta(t), \tag{1.4}$$

where $\beta(t)$ denotes standard Brownian motion on $\mathbb{R}^d$ and σ is a positive constant. The two-scale potential $V^\epsilon(x, y; \alpha)$ is assumed to consist of a large-scale and a fluctuating part:

$$V(x, y; \alpha) = \alpha V(x) + p(y). \tag{1.5}$$

As we show explicitly in (5.4) this set-up puts us in the framework of homogenization for SDEs.

Under (1.5), the SDE (1.4) becomes

$$dx^\epsilon(t) = -\alpha \nabla V(x^\epsilon(t)) \, dt - \frac{1}{\epsilon} \nabla p \left(\frac{x^\epsilon(t)}{\epsilon} \right) dt + \sqrt{2\sigma} \, d\beta(t). \tag{1.6}$$

If p is a sufficiently smooth periodic function, which, for simplicity, we take to have the same period L in all directions, then it is well known (see e.g. Refs. 5, 24) that, as $\epsilon \to 0$, the solution $x^\epsilon(t)$ of (1.4) converges in law to the solution of the SDE

$$dx(t) = -\alpha K \nabla V(x(t)) \, dt + \sqrt{2\sigma K} \, d\beta(t), \tag{1.7}$$

with

$$K = \int_{[0,L]^d} (I + \nabla_y \phi(y)) \, (I + \nabla_y \phi(y))^T \, \mu(dy) \tag{1.8}$$

and

$$\mu(dy) = \rho(y) \, dy = \frac{1}{Z} e^{-p(y)/\sigma} \, dy, \quad Z = \int_{[0,L]^d} e^{-p(y)/\sigma} \, dy. \tag{1.9}$$
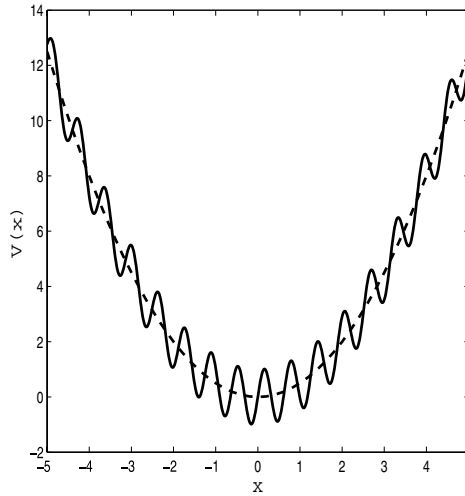
Fig. 1. $V^\epsilon(x, x/\epsilon) = \frac{1}{2}x^2 + \sin(\frac{x}{\epsilon})$ with $\epsilon = 0.1$ and averaged potential $V(x) = \frac{1}{2}x^2$.

The field $\phi(y)$ is the solution of the Poisson equation

$$-\mathcal{L}_0\phi(y) = -\nabla_y p(y), \quad \mathcal{L}_0 := -\nabla_y p(y) \cdot \nabla_y + \sigma\Delta_y, \tag{1.10}$$

on $[0, L]^d$ with periodic boundary conditions. The function $\rho(y)$ spans the null-space of $\mathcal{L}_0^*$, the L^2-adjoint of $\mathcal{L}_0$. The effective diffusion tensor K is positive definite and the diffusivity is always depleted.[23] Physically this occurs because the homogenized process must represent the cost of traversing the many small energy barriers present in the original multiscale problem but which are not explicitly captured in the homogenized potential. In Fig. 1 we plot the potential $V^\epsilon(x, x/\epsilon)$, as well as the average potential $V(x)$, illustrating this phenomenon. In fact, the effective diffusivity $\Sigma = \sigma K$ decays exponentially fast in σ as $\sigma \to 0$. See Ref. 7 and the references therein. Thus the original and homogenized diffusivities are exponentially different at small temperatures.

To illustrate these facts explicitly, consider the problem in one dimension, $d = 1$. In this case the limiting equation takes the form

$$dx(t) = -AV'(x(t))\,dt + \sqrt{2\Sigma}d\beta(t). \tag{1.11}$$

The effective coefficients are

$$A = \frac{\alpha L^2}{Z\hat{Z}} \quad \text{and} \quad \Sigma = \frac{\sigma L^2}{Z\hat{Z}}, \tag{1.12}$$

where

$$\hat{Z} = \int_0^L e^{p(y)/\sigma} dy, \quad Z = \int_0^L e^{-p(y)/\sigma} dy. \tag{1.13}$$

Notice that $L^2 \leq Z\hat{Z}$ by the Cauchy-Schwarz inequality. This explicitly shows that the homogenized equation in one dimension comprises motion in the average potential $V(x)$, at a new slower time-scale contracted by A/α.

The main results of the paper can be summarized as follows. Assume that we are given a path $\{x^\epsilon(t)\}_{t \in [0,T]}$ of Eq. (1.6) and that we want to fit an SDE of the form (1.11) to the given data, estimating the parameters A, Σ as $\hat{A}, \hat{\Sigma}$. Then the following is a loose statement of our main results; these will be formulated precisely, and proved, below.

Theorem 1.1. *If we do not subsample, then the estimators $\hat{A}$ and $\hat{\Sigma}$ are asymptotically biased – they converge to α, σ.*

Theorem 1.2. *If the sampling rate is between the two characteristic time scales of the SDE (1.4) then the estimators $\hat{A}$ and $\hat{\Sigma}$ are asymptotically unbiased – they converge to A, Σ.*

In econometrics and finance, the problem of estimating parameters for continuous time diffusion processes in the presence of small scale fluctuations (market microstructure noise) has been considered by Aït-Sahalia and collaborators[1,2] and more recently in Ref. 3. In that work the microscale is input as an independent white observational noise that is superimposed on-top of a single-scale diffusion process. We have a somewhat different framework: we work in the context of coupled systems of diffusions exhibiting multiple scales. Our aim is to fit a single-scale homogenized diffusion to data. Models similar to the ones considered in this paper have been studied extensively in finance, see Ref. 12 and the references therein. In that book there is discussion of parameter estimation for multiscale diffusions, with emphasis on the estimation of the rate of mean reversion of volatility from historical asset price data; see Ref. 12, Chap. 4. Despite differences from the framework used in Refs. 1–3 to study problems arising in econometrics and finance, similarities with our work remain. Indeed, as we will see later on, trying to fit the models on the basis of data sampled at too high a frequency leads to incorrect parameter inference; furthermore, there is an optimal subsampling rate for the data to obtain correct inference.

Various numerical algorithms for diffusions with multiple scales have been developed[17,27] and analyzed.[10,14] Those papers are finely honed to optimize the fitting of the homogenized diffusion in situations where the multiscale model is known explicitly. In contrast, in this paper we introduce multiscale diffusions

primarily as a device to generate multiscale data; we do not assume that the multiscale model is available to us when doing parameter estimation. This enables us to gain understanding of parameter estimation in situations where the multiscale data is given to us from experiments, or comes from a model where the scale-separation is not explicit. Two recent papers contain numerical experiments relating to the extraction of averaged or homogenized diffusions from data generated by a multiscale diffusion; see Refs. 6, 9.

The rest of the paper is organized as follows. In Sec. 2 we present the estimators that we will use. In Sec. 3 we present the main results of this paper. In Sec. 4 we present various numerical experiments illustrating the theoretical results presented in the previous section. Section 5 contains some technical results that are needed for the proofs of our theorems. Section 6 contains the proofs of two central propositions concerning the behavior of the multiscale diffusion when observed on time-scales long compared with the fast time-scales of the process, but small compared with the slow time-scales of the process. Section 7 is devoted to the proofs of our theorems. Finally, in Sec. 8 we present some concluding remarks.

2. THE ESTIMATORS

In this section we describe various estimators for the parameters arising in Eq. (1.7). We use $\langle \cdot, \cdot \rangle$ to denote the standard inner-product on $\mathbb{R}^d$ and $| \cdot |$ the induced Euclidean norm.

We assume that we are given a path $x = \{x(t)\}_{t \in [0,T]}$, or samples from such a path, $x = \{x_n\}_{n=0}^N$, with $x_n = x(n\delta)$. For simplicity we aim to fit the equation in the form

$$dx(t) = -A \nabla V(x(t)) \, dt + \sqrt{2\Sigma} d\beta(t), \tag{2.1}$$

where A and Σ are scalars. In one dimension this reduces to the form (1.11). Note that in general this is only the correct form for the homogenized equation in one dimension since, typically, the average potential has a matrix as a pre-factor, as in (1.7). However it suffices to exemplify the main ideas in this work, and simplifies the presentation.

The standard way to estimate the diffusion coefficient is via the quadratic variation of the path:

$$\hat{\Sigma}_{N,\delta}(x) = \frac{1}{2N\delta d} \sum_{n=0}^{N-1} |x_{n+1} - x_n|^2. \tag{2.2}$$

The standard way to estimate drift coefficients is via the path-space likelihood of (2.1) with respect to a pure diffusion with no drift, namely (see, for example, Refs. 4, 20)

$$\mathbb{L}(x) \propto \exp\{-I(x)/4\Sigma\}$$

where

$$I(x) = \int_0^T (|A\nabla(x(t))|^2 dt + 2A\langle \nabla V(x(t)), dx(t)\rangle).$$

Maximizing the log-likelihood then gives the estimate $\hat{A}$ of A given by

$$\hat{A}(x) = -\frac{\int_0^T \langle \nabla V(x(t)), dx(t)\rangle}{\int_0^T |\nabla V(x(t))|^2 dt}. \tag{2.3}$$

If the data is given in discrete but finely spaced increments, as often happens in practice, then this estimator can be approximated to yield

$$\hat{A}_{N,\delta}(x) = -\frac{\sum_{n=0}^{N-1} \langle \nabla V(x_n), (x_{n+1} - x_n)\rangle}{\sum_{n=0}^{N-1} |\nabla V(x_n)|^2 \delta}. \tag{2.4}$$

A key issue in this paper is to understand how to chose δ as a function of ϵ to ensure that data generated by (1.4) can be effectively fit to obtain the correct homogenized drift coefficients in equations such as (2.1), via the estimators (2.2) and (2.4).

The gradient structure of the SDE (2.1) can be used to obtain a second estimator for the drift coefficients. This second estimator, which we now derive, is of interest for two different reasons: firstly it may be useful in practice as it may lead to smaller variance in estimators; secondly it highlights the fact that working out how to sample the data to obtain the correct estimation of the diffusion coefficient alone will lead to correct estimation of the drift parameters, at least for the class of gradient-structure SDEs that we consider in this paper; see Theorem 3.8. The second estimator requires the input of an estimator $\hat{\Sigma}$ for the diffusion coefficient and is

$$\tilde{A}(x) = \hat{\Sigma} \frac{\frac{1}{T}\int_0^T \Delta V(x(t))\, dt}{\frac{1}{T}\int_0^T |\nabla V(x(t))|^2 dt}. \tag{2.5}$$

Approximating to allow for the input of discrete-time data gives

$$\tilde{A}_{N,\delta}(x) = \hat{\Sigma} \frac{\sum_{n=0}^{N-1} \Delta V(x_n)\delta}{\sum_{n=0}^{N-1} |\nabla V(x_n)|^2 \delta}. \tag{2.6}$$

In Proposition 3.3 below we show that, under appropriate assumptions on the potential $V(x)$, $\tilde{A}(x)$ is a natural approximation to $\hat{A}(x)$. This is because $\tilde{A}$ follows from $\hat{A}$ by an application of the Itô formula to the numerator of (2.3).

3. STATEMENT OF MAIN RESULTS

In this section we present the main results of our paper. Throughout the paper we make the following standing assumptions on the drift vector fields:

Assumptions 3.1. *The potentials p and V satisfy:*

- $p(y) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$;
- $V(x) \in C^\infty(\mathbb{R}^d, \mathbb{R})$;
- $|\nabla V(x_1) - \nabla V(x_2)| \le C_{\text{Lip}}|x_1 - x_2| \quad \forall x_1, x_2 \in \mathbb{R}^d$;
- $\exists a, b > 0 : \langle -\nabla V(x), x \rangle \le a - b|x|^2 \quad \forall x \in \mathbb{R}^d$;
- $e^{-\frac{\alpha}{\sigma} V(x)} \in L^1(\mathbb{R}^d, \mathbb{R}^+)$.

Notice that we have taken the field $p(y)$ to have period 1 in all directions. This assumption is made only for notational simplicity. In the numerical experiments reported in Sec. 4 we actually use 2π-periodic functions.

Remark 3.2. The third assumption will be used primarily to deduce that, by choice of origin for V,

$$|\nabla V(x)| \le C_{\text{Lip}}|x|. \tag{3.1}$$

This assumption could be relaxed and replaced by a polynomial growth bound; however this complicates the analysis without adding new insight. Similarly it is not necessary, of course, that V and p are C^∞. The fourth condition, however, is essential: it drives the ergodicity of the process which we use in a fundamental way in the analysis of the drift parameter estimators; it would not, however, be fundamental for estimation of diffusion coefficients alone. The fourth condition implies the fifth, which is simply the requirement that the invariant measure is indeed a probability measure; we state the two conditions separately for clarity of exposition.

The first result shows that the second estimator for the drift coefficient, defined in (2.6), is asymptotically equivalent to the maximum likelihood estimator.

Proposition 3.3. *Let $x = \{x(t)\}_{t \in [0,T]}$ satisfy (2.1). If $\hat{\Sigma} = \Sigma$ then the estimator $\tilde{A}(x)$ defined in (2.6) is asymptotically equivalent to the maximum likelihood estimator $\hat{A}$:*

$$\lim_{T \to \infty} \tilde{A}(x) = \hat{A}(x), \quad a.s.$$

The next result shows that, without subsampling, the parameter estimators for the homogenized model will be asymptotically biased: they recover the parameters from the unhomogenized equations.

Theorem 3.4. *Let $x^\epsilon(t)$ be the solution of (1.6) with $x^\epsilon(0)$ distributed according to the invariant measure of the process. Then the estimator (2.3) satisfies*

$$\lim_{\epsilon \to 0} \lim_{T \to \infty} \hat{A}(x^\epsilon) = \alpha \quad a.s. \tag{3.2}$$

Fix $T = N\delta$ in (2.2). Then for every $\epsilon > 0$ we have

$$\lim_{N \to \infty} \hat{\Sigma}_{N,\delta}(x^\epsilon) = \sigma \quad a.s. \tag{3.3}$$

Now consider the one dimensional problem

$$dx^\epsilon(t) = -\alpha V'(x^\epsilon(t))\,dt - \frac{1}{\epsilon}p'\left(\frac{x^\epsilon(t)}{\epsilon}\right)dt + \sqrt{2\sigma}\,d\beta(t). \tag{3.4}$$

The next two results show that, with appropriate subsampling, the estimators recover the correct drift and diffusion coefficients for the homogenized model (1.11) when taking data from the unhomogenized Eq. (3.4).

Theorem 3.5. *Let $x^\epsilon(t)$ be the solution of (3.4) with $x^\epsilon(0)$ distributed according to the invariant measure of the process. Further, let $\delta = \epsilon^\zeta$, $\zeta \in (0, 1)$ and $N = [\epsilon^{-\gamma}]$, $\gamma > \zeta$, where $[\cdot]$ denotes the integer part of a number. Then*

$$\lim_{\epsilon \to \infty} \hat{A}_{N,\delta}(x^\epsilon) = A \quad \text{in law}, \tag{3.5}$$

where A is given by (1.12).

Theorem 3.6. *Let $x^\epsilon(t)$ be the solution of (3.4) with $x^\epsilon(0)$ distributed according to the invariant measure of the process. Fix $T = N\delta$ with $\delta = \epsilon^\zeta$ and $\zeta \in (0, 1)$. Then*

$$\lim_{\epsilon \to 0} \hat{\Sigma}_{N,\delta}(x^\epsilon) = \Sigma \quad \text{in law}, \tag{3.6}$$

where Σ is given by (1.12).

Remarks 3.7.

- The two previous results require $\epsilon/\delta \to 0$ as $\epsilon \to 0$. In view of the fact that the fast time-scale is $\mathcal{O}(\epsilon^2)$ (see Eq. (5.4b)) it might be expected that this could be relaxed to $\epsilon^2/\delta \to 0$ as $\epsilon \to 0$. However we have not been able to prove this, and serious obstructions appear to prevent this.
- The value of ζ which optimizes the error in Theorem 3.5 is $\zeta = 2/3$. This appears to be a tight estimate, if the Euler approximation (2.4) to the estimator (2.3) is used to estimate the data.

- The value of ζ which optimizes the error in Theorem 3.6 is $\zeta = 1/2$. This may not be a tight estimate. Indeed if a conjectured improvement can be made on one of our estimates, then $\zeta = 2/3$ becomes the optimal choice.

See Remark 5.9 for further discussion of all of these points.

The final result concerns the second drift estimator and again concerns input of data from the unhomogenized Eq. (3.4) into the parameter estimator for the homogenized Eq. (1.11). It requires an estimate of the diffusion coefficient, $\hat{\Sigma}$. If $\hat{\Sigma} = \sigma$, then we estimate the drift coefficient incorrectly with $\tilde{A}(x^\epsilon)$; on the other hand, if $\hat{\Sigma} = \Sigma$, then the estimator $\tilde{A}(x^\epsilon)$ gives the drift of the homogenized equation. (To see the last result recall that $A/\Sigma = \alpha/\sigma$, see (1.12)). Consequently, for multi-scale gradient systems, it is sufficient only to subsample in a fashion which leads to the correct diffusion coefficient. This offers a clear computational advantage: the rate of subsampling does not need to be considered for drift parameter estimation, provided the second estimator $\tilde{A}$ is used.

Theorem 3.8.   *Let $x^\epsilon(t)$ be the solution of (3.4) with $x^\epsilon(0)$ distributed according to the invariant measure of the process. Assume that the diffusion coefficient has been estimated to be $\hat{\Sigma}$. Then*

$$\lim_{\epsilon \to 0} \lim_{T \to \infty} \tilde{A}(x^\epsilon) = \frac{\hat{\Sigma}}{\sigma}\alpha \quad a.s.$$

The proofs of our theorems are based on a combination of limit theorems for continuous time ergodic Markov processes together with results from homogenization theory. We quite often employ the gradient structure of the SDE, in particular the fact that the invariant measure is a Gibbs measure. It is not immediately clear how to extend the results to non-gradient SDE. For the proof of Theorems 3.5 and 3.6 we need to perform a careful asymptotic analysis of solutions to (1.6) for two small parameters ϵ and δ. We accomplish this by using quantitative versions of the ergodic theorem and of the martingale central limit theorem; see Propositions 5.7 and 5.8.

4. NUMERICAL RESULTS

In all cases we solve the multiscale SDE (1.4) using the Euler-Marayama scheme[19] for a single realization of the noise, with a time-step Δt sufficiently small so that the error due to the discretization is negligible; this requires that the time-step is small compared with ϵ^2, the fastest scale in the problem. We also employ a sufficiently long time interval so that the invariant measure is well sampled by the single path. Since the convergence to the invariant measure is

uniform in $\epsilon \to 0$, this is not prohibitive. We then use the data generated from the multiscale process as input to the estimators for the homogenized diffusion (1.7). We present numerical results for three model problems: a one dimensional polynomial potential of even degree, a one dimensional bistable potential and a two dimensional quadratic potential. In all three cases we perturb the large-scale part of the potential V by small-scale fast oscillations, usually in the form of a cosine potential p.

We present two types of numerical results. Note that δ, the time interval between two consecutive observations, is the inverse sampling rate. In the first we use $\delta = \Delta t$ as the time interval between two consecutive observations in the estimators. In the second we subsample the data, using $\delta > \Delta t$ and study how the estimated coefficients behave as a function of the subsampling. We use the data generated from our simulation in the estimators (2.4) and (2.6) to estimate the drift coefficient and in (2.2) to estimate the diffusion coefficient of (1.11). For the most part we work in one dimension and fit a single drift and diffusion parameter so that (1.7) becomes (1.11). When we work in more than one dimension, or estimate more than just a single drift or diffusion parameter, we use natural generalizations of the estimators defined in the previous section.

Let us summarize the main conclusions that can be drawn from the numerical experiments; recall that $\Delta t \ll \epsilon^2$. First, if we choose $\delta = \Delta t$, that is, if we don't subsample, then the resulting estimators do not generate the correct estimates of the homogenized coefficients. If, on the other hand, we subsample with $\epsilon^2 \ll \delta \ll \mathcal{O}(1)$, then the estimators generate the values of the parameters of the homogenized equation. Furthermore, there is an optimal sampling rate: there exists a δ^* which minimizes the distance between the homogenized value of the parameter and the value generated by the estimator. The optimal sampling rate depends sensitively on σ. It is also of interest that, in higher dimensions, the optimal sampling rate can be different for different parameters.

The above observations appear to hold independently of the detailed form of the large-scale part of the potential V (provided, of course, that it satisfies appropriate convexity conditions to ensure ergodicity). In addition, the performance of the estimators seems to be the same irrespective of the dimension of the problem.

Another interesting observation is that the second estimator for the drift coefficient (2.6) performs at least as well as the maximum likelihood estimator (2.4), and in some instances outperforms it.

4.1. Failure without Subsampling

In this section we study the estimators $\hat{A}$ and $\hat{\Sigma}$ when the data is given from the solution of Eq. (1.6) with $\epsilon \ll 1$ and $\Delta t = \delta$–no subsampling is used. We use
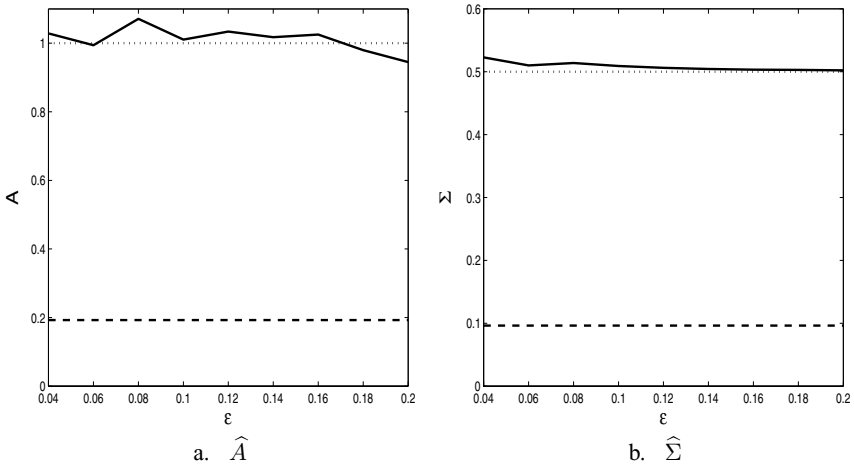
Fig. 2. Estimation of the drift and diffusion coefficients vs ϵ for the potential (4.1). *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.

the potential

$$V(x) = \frac{1}{2}\alpha x^2 \tag{4.1}$$

The small-scale part of the potential is

$$p(y) = \cos(y). \tag{4.2}$$

In Fig. 2 we plot the estimators $\hat{A}$ and $\hat{\Sigma}$ for various values of ϵ. For comparison we also plot the homogenized coefficients A and Σ and the unhomogenized coefficients α and σ. We observe that the estimators always give us the coefficients α and σ of the original SDE (1.6). In particular, the performance of the estimators does not improve as $\epsilon \to 0$. In Fig. 3 we plot the estimators for various values of the diffusion coefficient σ. We notice that the estimators give the values of the coefficients α and σ, for all values of σ. Since the homogenized coefficients decay to 0 exponentially fast in σ, the results of Fig. 3 indicate that the estimators give exponentially wrong results when $\sigma \ll 1$.

These results indicate the need to subsample – i.e. to choose δ appropriately as a function of ϵ.

4.2. Success with Subsampling

Now, rather than using all the data that were generated from the solution of Eq. (1.4) we use only a fraction of them. We choose δ in the estimators (2.2), (2.4)
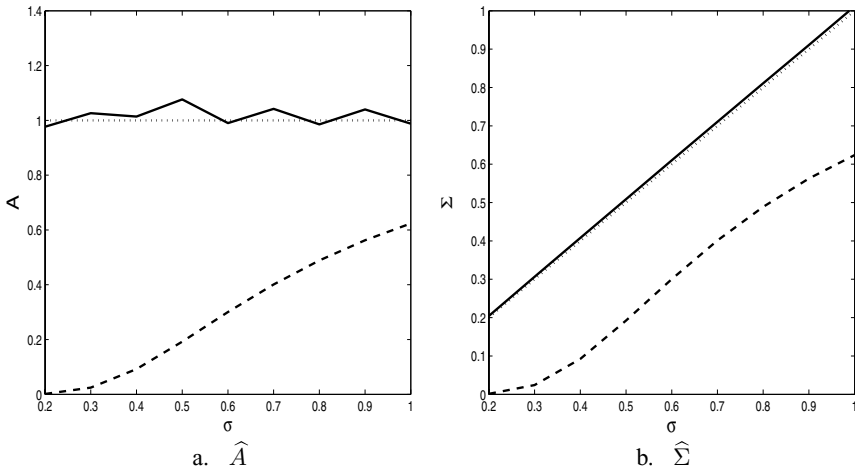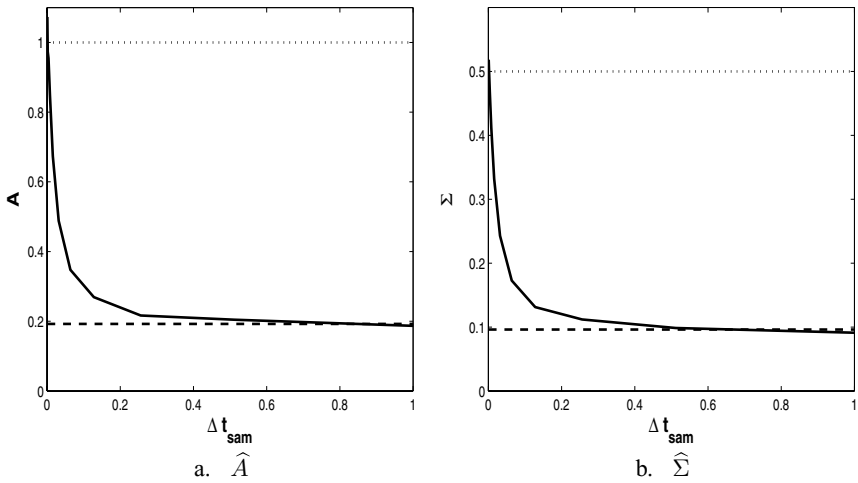
Fig. 3. Estimation of the drift and diffusion coefficients vs σ for the potential (4.1) with $\epsilon = 0.1$. *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.

and (2.6) as follows:

$$\Delta t_{\mathrm{sam}} = \delta = 2^k \Delta t, \quad k = 0, 1, 2, \ldots,$$

and we study the performance of the estimators as a function of the sampling rate. We investigate this issue for three different model problems.

4.2.1. OU Processes in 1D

We study the problem in one dimension with the large-scale part of the potential given by (4.1) and with the fluctuating part being the cosine potential (4.2). The two estimators $\hat{A}$ and $\tilde{A}$ for the drift coefficient produce almost identical results and we only present results for the maximum likelihood estimator $\hat{A}$. In Fig. 4 we present the estimated values of the drift and diffusion coefficients as a function of the inverse sampling rate $\delta = \Delta t_{\mathrm{sam}}$ when $\epsilon = 0.1, \alpha = 1.0, \sigma = 0.5$. We observe that, provided that we subsample at an appropriate rate, we are able to estimate the parameters of the homogenized equation correctly. Notice also that the estimators for the drift and the diffusion coefficient show very similar dependence on the sampling rate. This is in accordance with our theoretical results; see Theorem 3.8. In Fig. 5 we plot $\hat{\Sigma}$ as a function of the sampling rate for two different values of σ. We observe that the estimator of the diffusion coefficient is a decreasing function of the sampling rate, as expected. In addition to this, there is a well defined optimal sampling rate, which depends sensitively on σ. In particular the optimal δ is a decreasing function of σ. This is to be expected, since when

Fig. 4. Estimation of the drift and diffusion coefficients vs Δt_{sam} for the potential (4.1) with $\epsilon = 0.1$. *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.

$\sigma \gg 1$ the process $x^\epsilon(t)$ loses its multiscale character and becomes effectively a standard Brownian motion. Consequently, when σ is sufficiently large, the optimal δ becomes Δt, the integration time step. Notice furthermore that the slope of the $\hat{\Sigma} - \delta$ curve depends on σ. In Fig. 6 we plot the estimators of the drift and diffusion
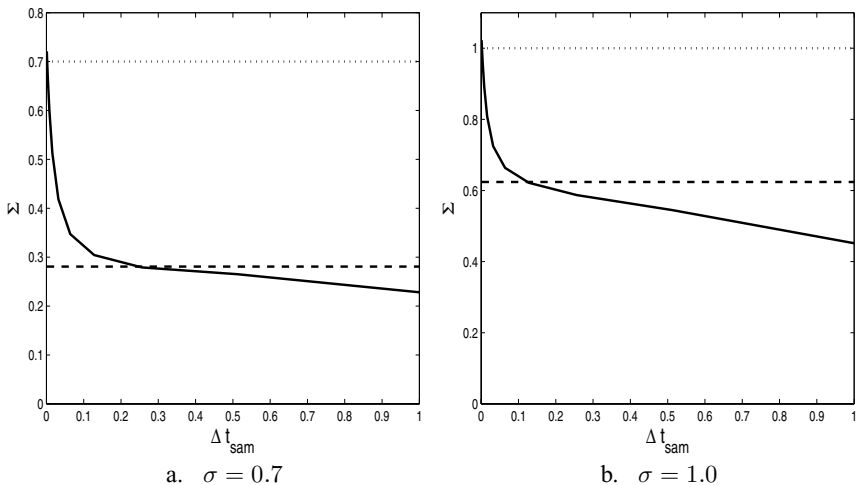

Fig. 5. Estimation of the diffusion coefficient vs Δt_{sam} for the potential (4.1) with $\epsilon = 0.1$, for two different values of σ. *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.
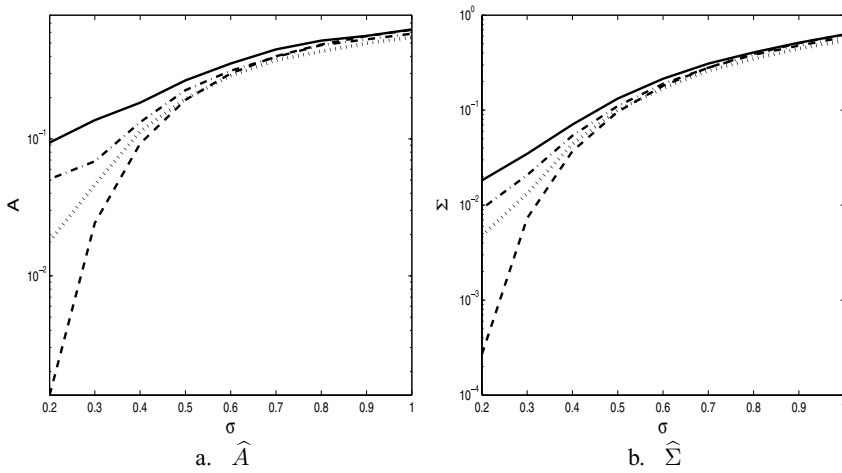
Fig. 6. Estimation of the drift and diffusion coefficient vs σ for the potential (4.1) with $\epsilon = 0.1$, $\alpha = 1.0$, for three different sampling rates. *Solid line:* $\Delta t_{\text{sam}} = 0.128$. *Dash-dotted line:* $\Delta t_{\text{sam}} = 0.256$. *Dotted line:* $\Delta t_{\text{sam}} = 0.512$. *Dashed line:* homogenized coefficient.

coefficients versus σ, for three different sampling rates. For comparison we also plot the homogenized coefficients. We observe that all three sampling rates lead to reasonably accurate estimates for A and Σ, when σ is not too small. On the other hand, the estimators become less accurate as $\sigma \to 0$. This is also to be expected: when $\sigma \ll 1$, the accurate simulation of (1.4) requires a very small time step; moreover, the equation has to be solved over a very long time interval in order for the invariant measure of the process to be well represented. Hence, our hypothesis that the errors due to discretization and finite time of integration are small, is not valid. In addition, as σ tends to 0, the optimal sampling rate increases, and becomes much larger than the coarser sampling rate that we use in the simulations. In Fig. 7 we plot the estimators versus ϵ, for three different values of the sampling rate. As expected, the deviation of the estimated values of the drift and diffusion coefficients from the homogenized values is an increasing function of ϵ. On the other hand, the optimal sampling rate does not appear to depend sensitively on ϵ: when it is small, a range of sampling rates give similar parameter estimates.

4.2.2. A Bistable Potential

We consider Eq. (1.4) in one dimension with a mean potential of the bistable form

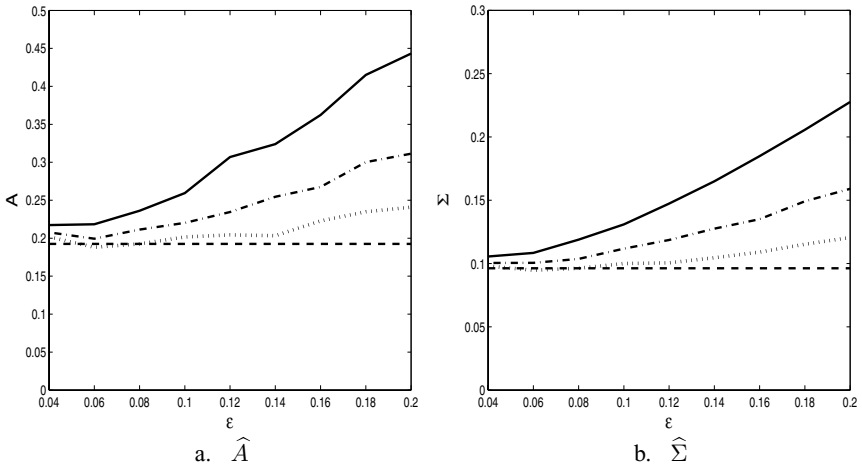$$V(x; \alpha, \beta) = -\frac{1}{2}\alpha x^2 + \frac{1}{4}\beta x^4. \tag{4.3}$$

Fig. 7. Estimation of the drift and diffusion coefficient vs ϵ for the potential (4.1) with $\alpha = 1.0$, $\sigma = 0.5$, for three different sampling rates. *Solid line:* $\Delta t_{\mathrm{sam}} = 0.128$. *Dash-dotted line:* $\Delta t_{\mathrm{sam}} = 0.256$. *Dotted line:* $\Delta t_{\mathrm{sam}} = 0.512$. *Dashed line:* homogenized coefficient.

The fluctuating part of the potential is given by (4.2). The homogenized equation is

$$dX(t) = (AX(t) - BX(t)^3)\,dt + \sqrt{2\Sigma}\,d\beta(t), \qquad (4.4)$$

where the homogenized coefficients are given by

$$A = \alpha K, \quad B = \beta K, \quad \Sigma = \sigma K, \quad K = \frac{4\pi^2}{Z\hat{Z}},$$

where Z and $\hat{Z}$ are given by (1.13) with $L = 2\pi$ and $p(y) = \cos(y)$. We estimate the diffusion coefficient using formula (2.2) with $d = 1$. For the two parameters of the drift we use generalizations of the maximum likelihood estimator $\hat{A}$.

In Figs. 8 and 9 we present the estimators for the two drift coefficients versus the sampling rate, for two different values of σ. We observe that the performance of the estimators is qualitatively similar to the OU case. Notice also that the optimal sampling rate is approximately the same for both coefficients.

In Fig. 10 we plot the estimator for the diffusion coefficient versus the sampling rate, for two different values of σ. The conclusions reached from the numerical study of $\hat{\Sigma}$ for the one dimensional OU process carry over almost verbatim to this case.
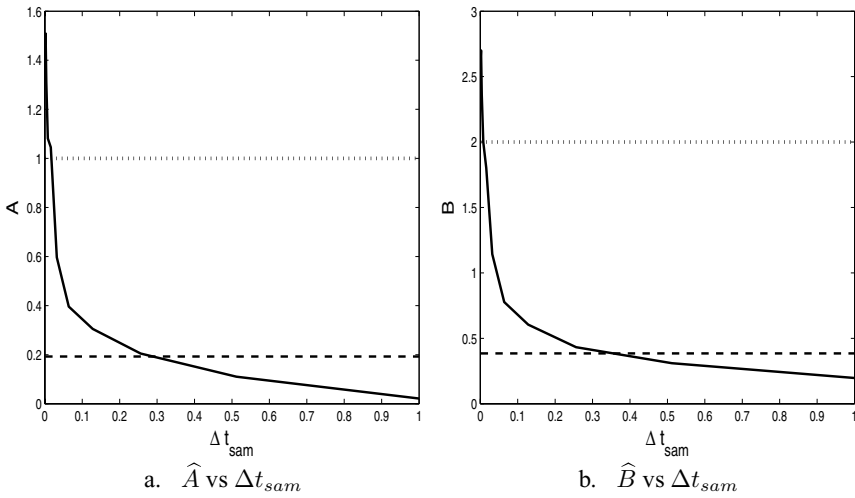
Fig. 8. Estimation of the parameters of the bistable potential (4.3) as a function of the sampling rate for $\sigma = 0.5$, $\epsilon = 0.1$. *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.

Fig. 9. Estimation of the parameters of the bistable potential (4.3) as a function of the sampling rate for $\sigma = 0.7$, $\epsilon = 0.1$. *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.
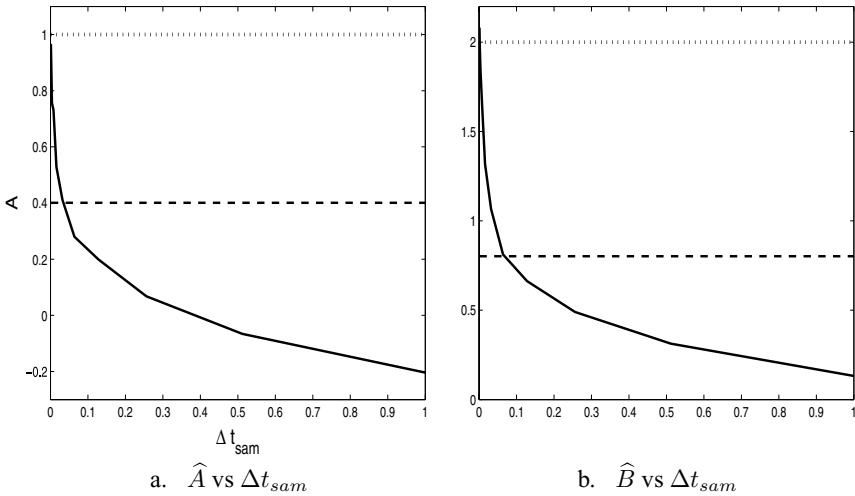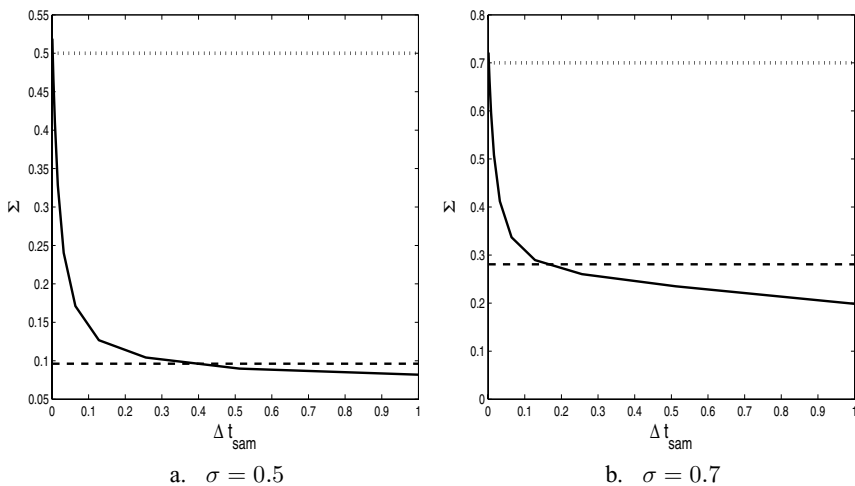
Fig. 10. Estimation of the diffusion coefficient for the bistable potential (4.3) as a function of the sampling rate for $\alpha = 1.0$, $\beta = 2.0$, $\epsilon = 0.1$. *Solid line:* estimated coefficient. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.

4.2.3. A Quadratic Potential in 2D

We consider now (1.4) in two dimensions with a separable fast potential $p(y)$:

$$dx^{\epsilon}(t) = -\nabla V(x^{\epsilon}(t), B) \, dt - \frac{1}{\epsilon} \nabla p_1 \left(\frac{x_1^{\epsilon}(t)}{\epsilon} \right)$$

$$-\frac{1}{\epsilon} \nabla p_2 \left(\frac{x_2^{\epsilon}(t)}{\epsilon} \right) dt + \sqrt{2\sigma} \, d\beta(t), \tag{4.5}$$

where B is the set of the drift parameters that we wish to estimate. The homogenized equation reads

$$dX(t) = -K\nabla V(X(t), B) \, dt + \sqrt{2\sigma K} \, d\beta(t), \tag{4.6}$$

where

$$K = \begin{pmatrix} \frac{L^2}{Z_1 \hat{Z}_1} & 0 \\ 0 & \frac{L^2}{Z_2 \hat{Z}_2} \end{pmatrix} \tag{4.7}$$

and

$$Z_i = \int_0^L e^{-\frac{p_i(y_i)}{\sigma}} \, dy_i, \quad \hat{Z}_i = \int_0^L e^{\frac{p_i(y_i)}{\sigma}} \, dy_i, \quad i = 1, 2.$$

In the above L denotes the period of $p(y)$.

We will consider the case of a general quadratic potential in two dimensions:

$$V(x, B_0) = \frac{1}{2}x^T B_0 x, \qquad (4.8)$$

with B symmetric and positive-definite. For the fluctuations we will use a simple two-dimensional extension of the cosine potential (4.2):

$$p_1(y_1) = \cos(y_1), \quad p_2(y_2) = \frac{1}{2}\cos(y_2).$$

With these choices for the large and small scale parts of the potential the homogenized Eq. (4.6) can be written in the form

$$dX(t) = -\nabla V_h(X(t))\,dt + \sqrt{2\Sigma}d\beta(t),$$

with

$$V_h(x) = \frac{1}{2}x^T B_h x, \quad B_h = K B_o \quad \text{and} \quad \Sigma = \sigma K.$$

Our goal is to estimate the diffusion tensor and the components of the matrix B_h. We will estimate the diffusion tensor through the quadratic variation:

$$\hat{\Sigma}_{N,\delta}(x(t)) = \frac{1}{2N\delta}\sum_{n=0}^{N-1}(x_{n+1} - x_n) \otimes (x_{n+1} - x_n), \qquad (4.9)$$

where $\otimes$ stands for the tensor product. For simplicity we will assume that the diffusion tensor in our model is diagonal. This is consistent with the homogenized diffusion tensor, see Eq. (4.7). We will use generalizations of the maximum likelihood estimator $\hat{A}$ in order to estimate the parameters of the quadratic potential. In the experiments we use the matrix

$$B_o = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}. \qquad (4.10)$$

In Fig. 11 we present the estimated values of the two non-zero components of the diffusion tensor versus the sampling rate.[3] The performance of the estimator for the diffusion tensor is, qualitatively at least, similar to its performance in the one dimensional problems considered in the previous two subsections. Notice, however, that the optimal sampling rate is quite different for the two non-zero components of the diffusion tensor.

In Fig. 12 we present the estimated values of the four drift coefficients. The results are in accordance with the one dimensional theory developed in this paper, as well as with the numerical experiments shown in one dimension. We remark

---

[3] The estimated value of the off-diagonal elements is almost 0 for all values of the sampling rate, in accordance with the theoretical result (4.7).
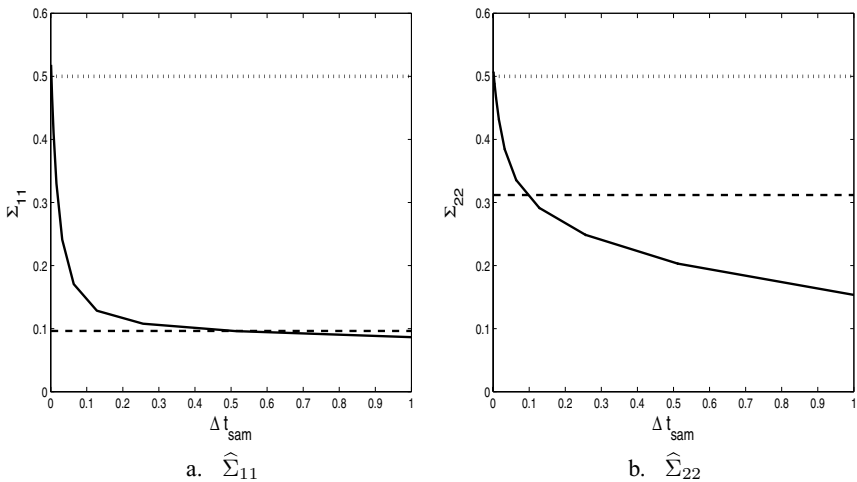
a. $\widehat{\Sigma}_{11}$ b. $\widehat{\Sigma}_{22}$

Fig. 11. Estimation of the non-zero elements of the diffusion tensor for the 2d quadratic potential (4.8) as a function of the sampling rate for B_o given by (4.10) and $\sigma = 0.5$, $\epsilon = 0.1$. *Solid line:* estimated homogenized coefficient. *Dashed line:* true homogenized coefficient. *Dotted line:* unhomogenized coefficient.

that the estimators capture successfully the fact that the homogenized matrix B_h, is not symmetric. Notice furthermore that, as for the diffusion matrix, the optimal sampling rate is different for different components of B_h.

Thus, in this simple two dimensional multiscale model, the optimal sampling rate is different in different directions. This suggests that extreme care has to be taken when estimating parameters for multidimensional, multiscale stochastic processes.

4.3. The Second Estimator for the Drift Coefficient

In this section we compare between the performances of the two estimators for the drift coefficient, namely $\hat{A}$ and $\tilde{A}$ given by Eqs. (2.4) and (2.6) respectively. In (2.6) we use the quadratic variation estimator for the diffusion coefficient, Eq. (2.2).

We estimate the drift parameter of (1.4) in one dimension for a quartic and a sixth-degree large-scale potential $V(x)$:

$$V(x) = \frac{1}{4}\alpha x^4 \tag{4.11}$$

and

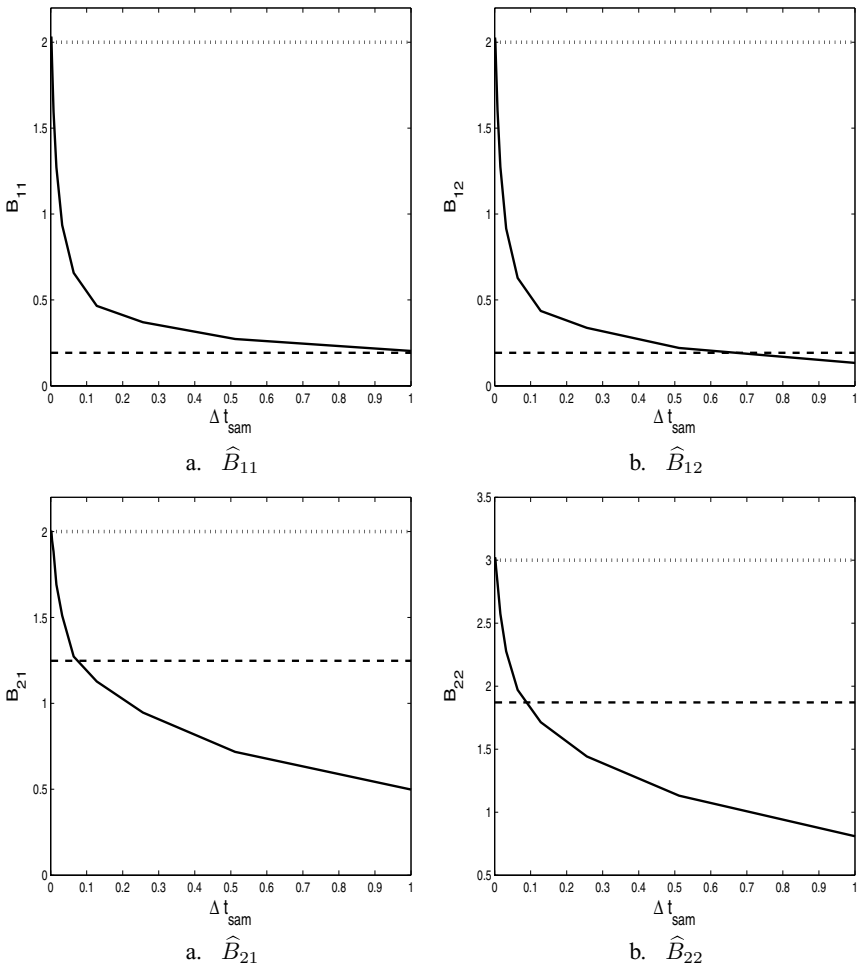$$V(x) = \frac{1}{6}\alpha x^6. \tag{4.12}$$

Fig. 12. Estimation of the parameters of the 2d quadratic potential (4.8) as a function of the sampling rate for B_o given by (4.10) and $\sigma = 0.5$, $\epsilon = 0.1$. *Solid line:* estimated homogenized coefficient. *Dashed line:* true homogenized coefficient. *Dotted line:* unhomogenized coefficient.

In both cases the small scale fluctuations are represented by the cosine potential (4.2) In Fig. 13 we present the estimated values of the drift coefficient as a function of the sampling rate for two different σ for the quartic potential (4.11). We also plot the effective and the unhomogenized values of the drift coefficient. Similar results for the sixth-degree potential (4.12) are presented in Fig. 14. In both cases we observe that the alternative estimator $\tilde{A}$ performs better than $\hat{A}$ in this situation where the data is subsampled.
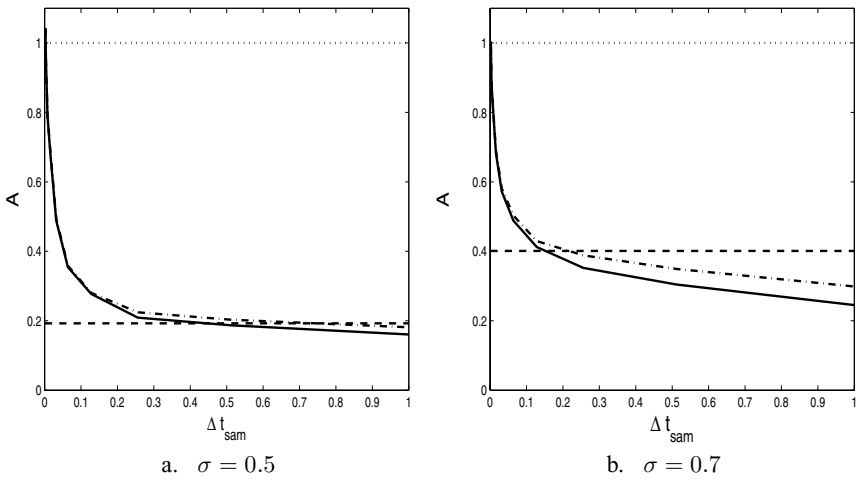
Fig. 13. Estimation of the drift coefficients for the quartic potential (4.11) as a function of the sampling rate for $\epsilon = 0.1$. *Solid line:* $\hat{A}$. *Dash-dot line:* $\tilde{A}$. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.
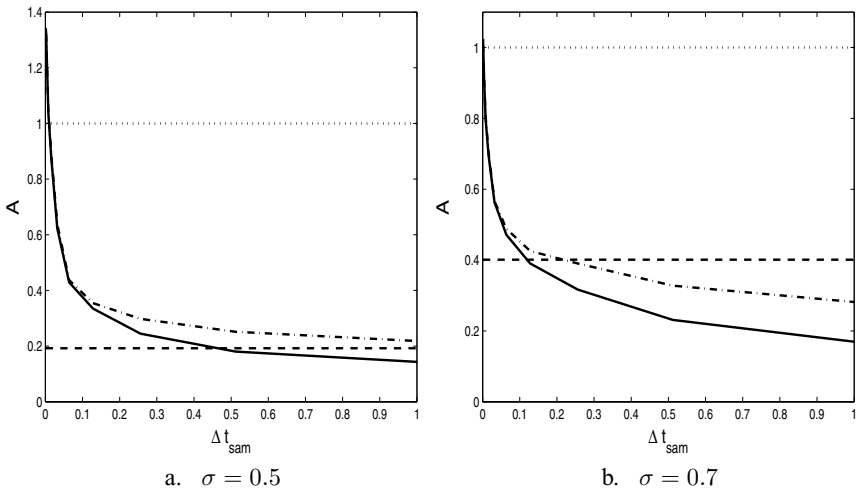

Fig. 14. Estimation of the drift coefficients for the sixth-degree potential (4.12) as a function of the sampling rate for $\epsilon = 0.1$. *Solid line:* $\hat{A}$. *Dash-dotted line:* $\tilde{A}$. *Dashed line:* homogenized coefficient. *Dotted line:* unhomogenized coefficient.

5. PRELIMINARY RESULTS

The rest of the paper is devoted to the proofs of our main results. For the convenience of the reader we have broken the material into three sections. In this section we collect various results that will be used in the proofs. We start by investigating some of the properties of the invariant measures of the unhomogenized and of the homogenized equation. We then introduce some tools useful in the study of homogenization for SDEs.

Proposition 5.1. *The invariant measure of the homogenized Eq. (1.7) is the Gibbs measure*

$$\mu(dx) = \rho(x)\,dx = \frac{1}{Z}e^{-\alpha V(x)/\sigma}\,dx, \quad Z = \int_{\mathbb{R}^d} e^{-\alpha V(x)/\sigma}\,dx. \quad (5.1)$$

The Markov process $x(t)$ given by (1.7) is geometrically ergodic: there are C, $\lambda > 0$ such that, for every measurable $f(x)$ satisfying

$$|f(x)| \leq 1 + |x|^p,$$

for some integer $p > 0$, we have, for μ-a.e. $x(0)$,

$$\left|\mathbb{E}f(x(t)) - \int_{\mathbb{R}^d} f(x)\rho(x)\,dx\right| \leq C(1 + |x(0)^p|)e^{-\lambda t},$$

where $\mathbb{E}$ denotes expectation with respect to Wiener measure.

Proof: Assumptions 3.1, together with the formulae for the effective drift and the effective diffusion coefficient, Eq. (1.8), imply that the solution $x(t)$ of the homogenized Eq. (1.7) has a unique invariant measure with smooth density. The Gibbs measure (5.1) satisfies

$$\alpha \nabla V \rho + \sigma \nabla \rho = 0$$

and hence

$$K(\alpha \nabla V \rho + \sigma \nabla \rho) = 0.$$

Because K is constant we deduce that

$$\alpha K \nabla V \rho + \nabla \cdot (\sigma K \rho) = 0.$$

Thus

$$\nabla \cdot (\alpha K \nabla V \rho + \nabla \cdot (\sigma K \rho)) = 0.$$

This is the stationary Fokker-Planck equation for (1.7) showing that the Gibbs measure p is indeed an invariant measure. For the geometric ergodicity we use [Ref. 22, Theorem 5.3]. □

Proposition 5.2. *The invariant measure of the unhomogenized Eq. (1.6) is the Gibbs measure*

$$\mu^\epsilon(dx) = \rho^\epsilon(x)\,dx = \frac{1}{Z^\epsilon}e^{-\frac{\alpha}{\sigma}V(x)-\frac{1}{\sigma}p(\frac{x}{\epsilon})}, \quad Z^\epsilon = \int_{\mathbb{R}^d}e^{-\frac{\alpha}{\sigma}V(x)-\frac{1}{\sigma}p(\frac{x}{\epsilon})}\,dx. \quad (5.2)$$

For every $\epsilon > 0$ the Markov process (1.6) is geometrically ergodic: there are C, $\lambda > 0$ such that, for every measurable $f(x)$ satisfying

$$|f(x)| \le 1 + |x|^p,$$

for some integer $p > 0$ we have, for μ^ϵ-a.e. $x^\epsilon(0)$,

$$\left|\mathbb{E}f(x^\epsilon(t)) - \int_{\mathbb{R}}f(x)\rho^\epsilon(x)\,dx\right| \le C(1 + |x^\epsilon(0)^p|)e^{-\lambda t},$$

where $\mathbb{E}$ denotes expectation with respect to Wiener measure.

  Furthermore, the measure μ^ϵ converges weakly to the invariant measure of the homogenized dynamics μ given by (5.1).

Proof: Assumptions 3.1 imply that $x^\epsilon(t)$ is an ergodic Markov process. Direct calculation with the Fokker-Planck equation shows that the unique invariant measure of the process is the Gibbs measure

$$\rho^\epsilon(x)\,dx = \frac{1}{Z^\epsilon}e^{-\frac{1}{\sigma}V(x,\frac{x}{\epsilon}\alpha)}dx$$

$$= \frac{1}{Z^\epsilon}e^{-\frac{\alpha}{\sigma}V(x)-\frac{1}{\sigma}p(\frac{x}{\epsilon})}dx,$$

with Z^ϵ given by (5.2). For the geometric ergodicity we use [Ref. 22, Theorem 5.3].

  Now let

$$u(x,y) := e^{-\frac{\alpha}{\sigma}V(x)-\frac{1}{\sigma}p(y)}.$$

Since $u(x,y) \in L^1(\mathbb{R}^d; C_{\text{per}}(\mathbb{T}^d))$, by [Ref. 8, Lemma 9.1] we have that

$$u\left(\cdot,\frac{\cdot}{\epsilon}\right) \rightharpoonup \int_{\mathbb{T}^d}u(\cdot,y)\,dy, \quad \text{weakly in } L^1(\mathbb{R}^d).$$

In particular, since $1 \in L^\infty(\mathbb{R}^d)$,

$$\lim_{\epsilon\to 0}Z^\epsilon = \int_{\mathbb{R}^d}\int_{\mathbb{T}^d}e^{-\frac{\alpha}{\sigma}V(x)-\frac{1}{\sigma}p(y)}dy.$$

We combine the above two results to conclude that

$$\rho^\epsilon(x) \rightharpoonup \frac{1}{Z}e^{-\frac{\alpha}{\sigma}V(x)}, \quad \text{weakly in } L^1(\mathbb{R}^d), \quad (5.3)$$

where Z is given by (5.1). The weak convergence of the densities in $L^1(\mathbb{R}^d)$ implies the weak convergence of the corresponding probability measures.　　　□

Remark 5.3.　　The assumption of stationarity of the process $x^\epsilon(t)$ is not necessary for the proof of the theorems stated in the previous section is made only for simplicity. This follows because $x^\epsilon(t)$ is geometrically ergodic and consequently it converges to its invariant distribution exponentially fast for arbitrary initial conditions. Furthermore, the fact that the invariant measure of the process $x^\epsilon(t)$ converges weakly, as $\epsilon \to 0$, to the invariant measure of the homogenized process is important for us as many of our results will be deduced by taking expectations with respect to the invariant measure $\mu^\epsilon(dx)$ of the multiscale dynamics (1.6). The weak convergence alluded to demonstrates that the measure μ^ϵ behaves uniformly in $\epsilon \to 0$.

An immediate corollary of the above proposition is that $x^\epsilon(t)$ has bounded moments of all orders. We will use the notation $\mathbb{E}^{\mu^\epsilon}$ to denote expectation with respect to the stationary measure of (3.1) on path space, when initial data is distributed according to the Gibbs measure (5.2).

Corollary 5.4.　　*Let $x^\epsilon(t)$ be the solution of (1.4) with the potential given by (1.5) and assume that conditions (3.1) are satisfied. Assume furthermore that $x^\epsilon(0)$ is distributed according to μ^ϵ. Then, for all $p \geq 1$, there is a constant $C = C(p)$, uniform in $\epsilon \to 0$, such that*

$$\mathbb{E}^{\mu^\epsilon}|x^\epsilon(t)|^p \leq C \quad \forall t \geq 0.$$

It is convenient for the subsequent analysis to introduce the auxiliary variable

$$y^\epsilon(t) = \frac{x^\epsilon(t)}{\epsilon}.$$

We can then write Eq. (1.6) in the form

$$dx^\epsilon(t) = -\alpha \nabla V(x^\epsilon(t))\, dt - \frac{1}{\epsilon}\nabla p(y^\epsilon(t))\, dt + \sqrt{2\sigma}\, d\beta(t), \qquad (5.4a)$$

$$dy^\epsilon(t) = -\frac{1}{\epsilon}\alpha \nabla V(x^\epsilon(t))\, dt - \frac{1}{\epsilon^2}\nabla p(y^\epsilon(t))\, dt + \sqrt{\frac{2\sigma}{\epsilon^2}}\, d\beta(t). \qquad (5.4b)$$

Notice that both processes $x^\epsilon(t)$ and $y^\epsilon(t)$ are driven by the same Brownian motion. Written in this fashion it is clear that we are in a situation where homogenization applies. The homogenized equation is found by eliminating $y^\epsilon(t)$ from the scale separated system for $\{x^\epsilon(t),\, y^\epsilon(t)\}$. Note that $\mathcal{L}_0$ defined in (1.10) is the generator

of the process

$$dy(t) = -\nabla p(y(t)) \, dt + \sqrt{2\sigma} \, d\beta(t),$$

on the unit torus, which governs the dynamics of $y^\epsilon(t)$ to leading order in ϵ. The generator of the joint process $\{x^\epsilon(t), y^\epsilon(t)\}$ reads

$$\mathcal{L}^\epsilon = \frac{1}{\epsilon^2} \mathcal{L}_0 + \frac{1}{\epsilon} \mathcal{L}_1 + \mathcal{L}_2,$$

where

$$\mathcal{L}_0 = -\nabla_y p(y) \cdot \nabla_y + \sigma \nabla_y,$$

$$\mathcal{L}_1 = -\nabla_y p(y) \cdot \nabla_x - \alpha \nabla_x V(x) \cdot \nabla_y + 2\sigma \nabla_x \cdot \nabla_y,$$

$$\mathcal{L}_2 = -\alpha \nabla_x V(x) \cdot \nabla_x + \sigma \Delta_x.$$

The following result can be found in, e.g. [Ref. 5, Chap. 3].

Lemma 5.5. *Assume that $p(y) \in C^\infty_{per}(\mathbb{T}^d, \mathbb{R})$ and that $H(y) \in C^\infty_{per}(\mathbb{T}^d, \mathbb{R}^d)$. Let $\mu(dy)$ be the Gibbs measure (1.9) and assume that $H(y)$ is centered with respect to $\mu(dy)$:*

$$\int_{\mathbb{T}^d} H(y)\mu(dy) = 0. \tag{5.5}$$

Then the Poisson equation

$$-\mathcal{L}_0 \chi = H(y), \tag{5.6}$$

has a unique mean-zero solution in $L^2_{per}(\mathbb{T}^d, \mu(dy); \mathbb{R}^d)$. This solution, together with all its derivatives, is bounded.

We will need an estimate on integrals whose integrand is centered with respect to the invariant measure $\mu(dy)$.

Lemma 5.6. *Let $H(y) \in C^\infty_{per}(\mathbb{T}^d; \mathbb{R}^d)$ satisfy condition (5.5). Assume that $x^\epsilon(0)$ is distributed according to (5.2). Then the following estimate holds for any $p > 1$ and $T > 0$:*

$$\mathbb{E}^{\mu^\epsilon} \left| \int_0^T H(y^\epsilon(s)) \, ds \right|^p \leq C\left(\epsilon^{2p} + \epsilon^p T^p + \epsilon^p T^{\frac{p}{2}}\right).$$

Proof: Consider the Poisson Eq. (5.6) with periodic boundary conditions. Since $H(y)$ satisfies (5.5), Lemma 5.5 applies and we have that $\chi(y)$ is smooth and bounded, together with all its derivatives. We now apply the Itô formula to $\chi(y^\epsilon(t))$,

where $y^\epsilon(t)$ is the solution of (5.4b), and use (5.6) to obtain

$$\int_0^T H(y^\epsilon(s))\,ds = -\epsilon^2(\chi(y^\epsilon(T)) - \chi(y^\epsilon(0)))$$

$$+ \epsilon\sqrt{2\sigma}\int_0^T \langle \nabla_y\chi(y^\epsilon(s)), d\beta(s)\rangle$$

$$- \alpha\epsilon\int_0^T \langle \nabla V(x^\epsilon(s)), \nabla_y\chi(y^\epsilon(s))\rangle\,ds.$$

Now, using the boundedness of χ and its derivatives, we have, for

$$I(T) := E^{\mu^\epsilon}\left|\int_0^T H(y^\epsilon(s))\,ds,\right|^p,$$

$$I(T) \le C\left(\epsilon^{2p} + \epsilon^p E^{\mu^\epsilon}\left|\int_0^T |\nabla V(x^\epsilon(s))|ds\right|^p + \epsilon^p E^{\mu^\epsilon}\left|\int_0^T \langle \nabla_y\chi(y^\epsilon(s)), d\beta(s)\rangle\right|^p\right)$$

$$\le C\left(\epsilon^{2p} + \epsilon^p T^{p-1}\int_0^T |x^\epsilon(s)|^p ds + \epsilon^p T^{\frac{p}{2}-1}\int_0^T E^{\mu^\epsilon}|\nabla_y\chi(y^\epsilon(s))|^p ds\right)$$

$$\le C\left(\epsilon^{2p} + \epsilon^p T^p + \epsilon^p T^{\frac{p}{2}}\right),$$

from which the desired estimate follows. In deriving the above we used Corollary 5.4 and the estimate [Ref. 18, Eq. (3.25), p. 163] on moments of stochastic integrals. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

For the rest of this section we will restrict ourselves to the one dimensional case. If we apply the Itô formula to $\phi(y^\epsilon(s))$, the solution of the Poisson Eq. (1.10), then we obtain

$$x_{n+1}^\epsilon - x_n^\epsilon = -\alpha\int_{n\delta}^{(n+1)\delta} V'(x^\epsilon(s))(1 + \partial_y\phi(y^\epsilon(s)))\,ds \qquad (5.7)$$

$$+ \sqrt{2\sigma}\int_{n\delta}^{(n+1)\delta} (1 + \partial_y\phi(y^\epsilon(s)))\,d\beta(s)$$

$$- \epsilon(\phi(y^\epsilon((n+1)\delta)) - \phi(y^\epsilon(n\delta))). \qquad (5.8)$$

The proofs of Theorems 3.5 and 3.6 is based on careful asymptotic analysis of the behavior of $x_{n+1}^\epsilon - x_n^\epsilon$ given by this formula when both ϵ and δ are small. Specifically we will use the following two propositions. They show how the effective homogenized behavior is manifest in the time-δ Markov chain induced by sampling the path $x^\epsilon(t)$ from (1.6).

Proposition 5.7. *For ϵ, $\delta > 0$ sufficiently small and $n \in \mathbb{N}$ there exists an i.i.d. sequence of random variables $\xi_n \in \mathcal{N}(0, 1)$ such that*

$$\sqrt{2\sigma} \int_{n\delta}^{(n+1)\delta} (1 + \partial_y \phi(y^\epsilon(s))) \, d\beta(s) = \sqrt{2\Sigma\delta} \xi_n + R_1(\delta, \epsilon) \qquad (5.9)$$

in law. The remainder $R_1(\delta, \epsilon)$ satisfies, for every $\iota \in (0, \frac{1}{2})$ and $p > 0$, the estimate

$$(\mathbb{E}^{\mu^\epsilon} |R_1(\epsilon, \delta)|^p)^{1/p} \le C(\epsilon^{2\iota} + \epsilon^\iota), \qquad (5.10)$$

where C is independent of ϵ and δ.

Proposition 5.8. *For ϵ, $\delta > 0$ sufficiently small and $n \in \mathbb{N}$ we have that*

$$\alpha \int_{n\delta}^{(n+1)\delta} V'(x^\epsilon(s))(1 + \partial_y \phi(y^\epsilon(s))) \, ds = A\delta V'(x_n^\epsilon) + R_2(\epsilon, \delta) \qquad (5.11)$$

in law. The remainder $R_2(\delta, \epsilon)$ satisfies, for every $p > 0$, the estimate

$$(\mathbb{E}^{\mu^\epsilon} |R_2(\epsilon, \delta)|^p)^{1/p} \le C\left(\epsilon^2 + \delta^{\frac{1}{2}} + \delta^{3/2}\right), \qquad (5.12)$$

where C independent of ϵ and δ.

Remark 5.9.

- Estimate (5.10) is almost certainly not optimal. Indeed, informal calculations lead us to expect the estimate

$$(\mathbb{E}^{\mu^\epsilon} |R_1(\epsilon, \delta)|^p)^{1/p} \le C\left(\epsilon^{2\iota} + \epsilon^\iota \delta^\iota + \epsilon^\iota \delta^{\frac{1}{2}}\right).$$

   However, we have not been able to prove this.
- When estimating the diffusion coefficient in Theorem 3.6 and using estimate (5.10) the leading error contributions are $\mathcal{O}(\delta)$ and $\mathcal{O}(\epsilon^{\frac{1}{2}})$ suggesting that $\zeta = \frac{1}{2}$ is optimal. However if the preceding improved estimate could be proved then the optimal ζ in Theorem 3.6 is $\zeta = 2/3$.
- When estimating the drift coefficient in Theorem 3.5 the leading error contributions are $\mathcal{O}(\delta^{3/2})$ and $\mathcal{O}(\epsilon)$ suggesting that $\zeta = \frac{2}{3}$ is optimal. The analysis leading to this result appears tight.
- Note that it is the presence of the terms of size $\epsilon\phi(y^\epsilon(s))$, where $\phi(y)$ is the solution of (1.10), in (5.8) which necessitate taking steps δ which are large compared to ϵ and not large compared with ϵ^2, the fast scale in the problem. It is hard to see how this issue can be avoided.

6. PROOF OF PROPOSITIONS 5.7 AND 5.8

In this section we prove Propositions 5.7 and 5.8. These are central to the proof of the two theorems concerning the behavior of the estimators with subsampled data. We start with a rough estimate on $x^\epsilon_{n+1} - x^\epsilon_n$ that we will need for the proofs of the propositions.

6.1. A Rough Estimate

Lemma 6.1.   *Let Assumptions 3.1 hold and assume that $x^\epsilon(t)$, the solution of (3.4), is stationary. Then there exists a constant C, independent of δ and ϵ, such that*

$$\mathbb{E}^{\mu^\epsilon} |x^\epsilon(s) - x^\epsilon_{n\delta}|^p \leq C\big(\delta^p + \delta^{\frac{p}{2}} + \epsilon^p\big), \tag{6.1}$$

for every $s \in [n\delta, (n+1)\delta]$ and every $p \geq 1$.
Proof: Using the same derivation that leads to (5.8), but with $(n+1)\delta$ replaced by s, we have:

$$
\begin{aligned}
x^\epsilon(s) - x^\epsilon_n &= -\alpha \int_{n\delta}^s V'(x^\epsilon(s))(1 + \partial_y \phi(y^\epsilon(s)))\,ds \\
&\quad + \sqrt{2\sigma} \int_{n\delta}^s (1 + \partial_y \phi(y^\epsilon(s)))\,d\beta(s) - \epsilon(\phi(y^\epsilon(s)) - \phi(y^\epsilon(n\delta))) \\
&=: I^1_{n,\delta} + I^2_{n,\delta} + I^3_{n,\delta}.
\end{aligned}
\tag{6.2}
$$

We need to estimate the terms in (6.2). We start with $I^3_{n,\delta}$. By Lemma 5.5 we have

$$\| \phi(y) \|_{L^\infty} \leq C.$$

Consequently

$$\mathbb{E}^{\mu^\epsilon} |I^3_{n,\delta}|^p \leq C^p_\epsilon.$$

To estimate $I^1_{n,\delta}$ we use again Lemma 5.5 to conclude that

$$\|1 + \partial_y \phi\|_{L^\infty}. \tag{6.3}$$

The above estimate, together with Assumptions 3.1, Corollary 5.4 and the stationarity of the process $x^\epsilon(t)$, give

$$
\begin{aligned}
\mathbb{E}^{\mu^\epsilon} |I^1_{n,\delta}|^p &\leq C\delta^{p-1} \int_{n\delta}^{(n+1)\delta} \mathbb{E}^{\mu^\epsilon} |V'(x^\epsilon(s))|^p ds \\
&\leq C\delta^{p-1} \int_{n\delta}^{(n+1)\delta} \mathbb{E}^{\mu^\epsilon} |x^\epsilon(s)|^p ds \\
&\leq C\delta^p.
\end{aligned}
$$

Estimate [Ref. 18, Eq. (3.25), p. 163] on moments of stochastic integrals, together with Eq. (6.3), enable us to conclude that

$$E^{\mu^\epsilon}\left|I_{n,\delta}^2\right|^p \leq C\delta^{\frac{p}{2}-1}\int_{n\delta}^{(n+1)\delta}E^{\mu^\epsilon}|1+\partial_y\phi(y^\epsilon(s))|^p ds$$

$$\leq C\delta^{\frac{p}{2}}.$$

We combine the above estimates to obtain (6.1).   □

6.2. Proof of Proposition 5.7

From Theorem [Ref. 13, Sec. 1.3], [Ref. 18, Theorem. 3.4.6] we know that the martingale

$$M(t) := \sqrt{2\sigma}\int_0^t \left(1+\partial_y\phi(y_s^\epsilon)\right)d\beta(s)$$

is equal in law to a time-changed Brownian motion $\hat{\beta}$:,

$$M(t) = \hat{\beta}\left(2\sigma\int_0^t (1+\partial_y\phi(y^\epsilon(s)))^2 ds\right).$$

Also the quadratic variation satisfies

$$\langle M\rangle_t = 2\sigma\int_0^t (1+\partial_y\phi(y^\epsilon(s)))^2 ds \approx 2\Sigma t.$$

Indeed

$$\mathbb{E}^{\mu^\epsilon}\langle M\rangle_t = 2\sigma\mathbb{E}^{\mu^\epsilon}\int_0^t (1+\partial\phi(y^\epsilon(s)))^2 ds$$

$$= 2\Sigma t, 11$$

where the last equality follows from Eq. (1.8) for $d=1$. Using these observations we write

$$J_n := \sqrt{2\sigma}\int_{n\delta}^{(n+1)\delta} (1+\partial_y\phi(y^\epsilon(s)))\,d\beta(s)$$

$$= \sqrt{2\sigma}\int_{n\delta}^{(n+1)\delta} (1+\partial_y\phi(y^\epsilon(s)))\,d\beta(s) - \sqrt{2\sigma}\int_0^{n\delta} (1+\partial_y\phi(y^\epsilon(s)))\,d\beta(s)$$

$$= \hat{\beta}(2\Sigma(n+1)\delta) - \hat{\beta}(2\Sigma n\delta) + r_{n+1} - r_n$$

$$= \sqrt{2\Sigma\delta}\xi_n + r_{n+1} - r_n,$$

where the ξ_n are i.i.d unit Gaussian random variables and

$$r_n = \hat{\beta}(\langle M \rangle_{n\delta}) - \hat{\beta}(2\Sigma n\delta).$$

To estimate this difference we follow the proof of [Ref. 16, Theorem. 2.1]. We start by employing the Hölder continuity of Brownian motion, together with Hölder inequality, to estimate:

$$\mathbb{E}^{\mu^\epsilon} |\hat{\beta}(\langle M \rangle_{n\delta}) - \hat{\beta}(\mathbb{E}^{\mu^\epsilon} \langle M \rangle_{n\delta})|^p \leq \mathbb{E}^{\mu^\epsilon} |\text{Höl}_\iota(\hat{\beta})(\langle M \rangle_{n\delta} - \mathbb{E}^{\mu^\epsilon} \langle M \rangle_{n\delta})^\iota|^p$$

$$\leq \mathbb{E}^{\mu^\epsilon} |\text{Höl}_\iota(\hat{\beta})|^p \left(E^{\mu^\epsilon} |\langle M \rangle_{n\delta} - E^{\mu^\epsilon} \langle M \rangle_{n\delta}|^{\iota q} \right)^{\frac{p}{q}}$$

$$\leq C \left(\mathbb{E}^{\mu^\epsilon} \left| \int_0^{n\delta} H(y^\epsilon(s)) \, dz \right|^{\iota q} \right)^{\frac{p}{q}},$$

with $\iota \in (0, \frac{1}{2})$. We have used the notation

$$H(y) := 2\sigma(1 + \partial_y \phi(y))^2 - 2\Sigma.$$

We have also used the fact that, for every $\beta \in (0, \frac{1}{2})$ and every bounded time interval, the ι-Hölder exponent of Brownian motion is uniformly bounded with probability one. We have that

$$\int_{\mathbb{T}} H(y)\mu(dy) = 0,$$

where $\mu(dy)$ is defined in (1.9). Since $n\delta \leq T$, Lemma 5.6 applies and we have that, for q sufficiently large and for ϵ sufficiently small,

$$\mathbb{E}^{\mu^\epsilon} |J_n - \sqrt{2\Sigma\delta}\xi_n|^p \leq C(\epsilon^{2q\iota} + \epsilon^{q\iota})^{\frac{p}{q}}$$

$$\leq C(\epsilon^{2p\iota} + \epsilon^{p\iota}).$$

This completes the proof of the proposition. $\qquad\square$

6.3. Proof of Proposition 5.8

We have

$$\mathbb{E}^{\mu^\epsilon} |R_2(\epsilon, \delta)|^p = \mathbb{E}^{\mu^\epsilon} \left| \int_{n\delta}^{(n+1)\delta} \alpha V'(x^\epsilon(s))(1 + \partial_y \phi(y^\epsilon(s))) \, ds - \delta A V'(x_{n\delta}^\epsilon) \right|^p$$

$$= \mathbb{E}^{\mu^\epsilon} \left| \int_{n\delta}^{(n+1)\delta} \alpha V'(x_{n\delta}^\epsilon)(1 + \partial_y \phi(y^\epsilon(s))) \, ds - A \int_{n\delta}^{(n+1)\delta} V'(x_{n\delta}^\epsilon) \, ds \right.$$

$$+ \alpha \int_{n\delta}^{(n+1)\delta} V'\big(x^\epsilon(s)\big) - V'\big(x_{n\delta}^\epsilon\big)\big)(1 + \partial_y \phi(y^\epsilon(s))) \, ds \bigg|^p$$

$$\leq C \mathbb{E}^{\mu^\epsilon} \bigg| V'\big(x_{n\delta}^\epsilon\big) \int_{n\delta}^{(n+1)\delta} (\alpha(1 + \partial_y \phi(y^\epsilon(s))) - A) \, ds \bigg|^p$$

$$+ \alpha^p C \mathbb{E}^{\mu^\epsilon} \bigg| \int_{n\delta}^{(n+1)\delta} \big(V'(x^\epsilon(s)) - V'\big(x_{n\delta}^\epsilon\big)\big)(1 + \partial_y \phi(y^\epsilon(s))) \, d\beta \bigg|^p$$

$$=: I_{\epsilon,\delta}^1 + I_{\epsilon,\delta}^2,$$

where the constant C depends only on p. We use the Hölder inequality, Assumptions 3.1, Lemma 6.1 and the uniform bound on $\partial_y \phi(y)$ to obtain, for ϵ, δ sufficiently small,

$$I_{\epsilon,\delta}^2 \leq C \delta^{p-1} \int_{n\delta}^{(n+1)\delta} \mathbb{E}^{\mu^\epsilon} \big| x^\epsilon(s) - x_{n\delta}^\epsilon \big|^p \, ds$$

$$\leq C \delta^{p-1} \int_{n\delta}^{(n+1)\delta} \big(\delta^{\frac{p}{2}} + \epsilon^p \big) \, ds$$

$$\leq C \big(\delta^{\frac{3p}{2}} + \delta^p \epsilon^p \big).$$

Consequently

$$\big(\mathbb{E}^{\mu^\epsilon} \big| I_{\epsilon,\delta}^2 \big| \big)^{1/p} \leq C(\delta^{3/2} + \delta\epsilon). \tag{6.4}$$

Consider now the function

$$H(y) := \alpha(1 + \partial_y \phi(y)) - A,$$

From the definition of A we get that

$$\int_{\mathbb{T}} (\alpha(1 + \partial_y \phi(y)) - A) \mu(dy) = 0.$$

Hence, Lemma 5.6 applies and we get

$$\mathbb{E}^{\mu^\epsilon} \bigg| \int_{n\delta}^{(n+1)\delta} (\alpha(1 + \partial_y \phi(y^\epsilon(s))) - A) \, ds \bigg|^p \leq C(\epsilon^{2p} + \epsilon^p \delta^p + \epsilon^p \delta^{p/2}).$$

We combine the above estimate with (3.1) and Corollary 5.4 to obtain,

$$\big(\mathbb{E}^{\mu^\epsilon} \big| I_{\epsilon,\delta}^2 \big|^p \big)^{1/p} \leq C(\epsilon^2 + \epsilon \delta^{1/2}), \tag{6.5}$$

for ϵ, δ sufficiently small. The proof of the proposition follows from estimates (6.4) and (6.5). □

7. PROOF OF MAIN THEOREMS

Here we combine the results from the preceding two sections to complete the proofs of the main theorems and of Proposition 3.3.

7.1. Proof of Proposition 3.3

We apply the Itô formula to $V(x(t))$ for $x(t)$ solving (2.1) and use formula (2.3) to obtain

$$\hat{A}(x) = \frac{V(x(0)) - V(x(T)) + \Sigma \int_0^T \Delta V(x(t)) \, dt}{\int_0^T |\nabla V(x)|^2 dt}$$

$$= \frac{(V(x(0)) - V(x(T)))}{\int_0^T |\nabla V(x)|^2 dt} + \frac{\frac{1}{T} \Sigma \int_0^T \Delta V(x(t)) dt}{\frac{1}{T} \int_0^T |\nabla V(x)|^2 dt}$$

$$= \frac{\frac{1}{T}(V(x(0)) - V(x(T)))}{\frac{1}{T} \int_0^T |\nabla V(x)|^2 dt} + \tilde{A}(x).$$

Under the Assumptions 3.1 it follows from[21] that

$$\lim_{T \to 0} \frac{1}{T}(V(x(0)) - V(x(T))) = 0, \quad \text{a.s.}$$

and hence, by ergodicity applied to deduce almost sure convergence of the denominator,

$$\lim_{T \to 0} \frac{\frac{1}{T}(V(x(0)) - V(x(T)))}{\int_0^T \frac{1}{T}|\nabla V(x(t))|^2 dt} = 0, \quad \text{a.s.}$$

The result follows. □

7.2. Proof of Theorem 3.4

We combine Eqs. (2.3) and (1.6) to calculate

$$\hat{A}(x^\epsilon) = \frac{\int_0^T -\langle \Delta V(x^\epsilon(t)), dx^\epsilon(t) \rangle}{\int_0^T |\nabla V(x^\epsilon(t))|^2 dt}$$

$$
= \frac{\int_0^T \left\langle -\Delta V(x^\epsilon(t)), -\alpha \nabla V(x^\epsilon(t)) \, dt - \frac{1}{\epsilon} \nabla p\left(\frac{x^\epsilon(t)}{\epsilon}\right) dt + \sqrt{2\sigma} \, d\beta(t) \right\rangle}{\int_0^T |\nabla V(x^\epsilon(t))|^2 dt}
$$

$$
= \alpha + \frac{\frac{1}{\epsilon} \int_0^T \left\langle \nabla V(x^\epsilon(t)), \nabla p\left(\frac{x^\epsilon(t)}{\epsilon}\right)\right\rangle dt}{\int_0^T |\nabla V(x^\epsilon(t))|^2 dt} - \sqrt{2\sigma} \frac{\int_0^T \left\langle \nabla V(x^\epsilon(t)), d\beta(t)\right\rangle}{\int_0^T |\nabla V(x^\epsilon(t))|^2 dt}
$$

$$
=: \alpha + I_1(T, \epsilon) - I_2(T, \epsilon).
$$

We will treat the terms $I_1(T, \epsilon)$ and $I_2(T, \epsilon)$ separately. We start with $I_2(t, \epsilon)$. Since the stochastic integral

$$
M_T := \int_0^T \left\langle \nabla V(x^\epsilon(t)), d\beta(t)\right\rangle
$$

is a continuous martingale which is null at 0, the strong law of large numbers for martingales [Ref. 26, p. 187] applies and we have that

$$
\lim_{T \to +\infty} \frac{M_T}{\langle M \rangle_T} = 0 \quad \text{a.s.}
$$

Consequently

$$
\lim_{T \to +\infty} I_2(T, \epsilon) = 0 \quad \text{a.s.} \tag{7.1}
$$

Let us consider now the term $I_1(T, \epsilon)$. We use the ergodic theorem to deduce that

$$
\lim_{T \to \infty} I_1(T, \epsilon) = \lim_{T \to \infty} \frac{\frac{1}{\epsilon T} \int_0^T \left\langle \nabla V(x^\epsilon(t)), \nabla p\left(\frac{x^\epsilon(t)}{\epsilon}\right)\right\rangle ds}{\frac{1}{T} \int_0^T |\nabla V(x^\epsilon(t))|^2 dt}
$$

$$
= \frac{\mathbb{E}^{\mu^\epsilon}\left(\left\langle \nabla V(x), \frac{1}{\epsilon} \nabla p\left(\frac{x}{\epsilon}\right)\right\rangle\right)}{\mathbb{E}^{\mu^\epsilon} |\nabla V(x)|^2} \quad \text{a.s.}
$$

Now we use Proposition 5.2 to compute

$$
\frac{\mathbb{E}^{\mu^\epsilon}\left(\left\langle \nabla V(x), \frac{1}{\epsilon} \nabla p\left(\frac{x}{\epsilon}\right)\right\rangle\right)}{\mathbb{E}^{\mu^\epsilon} |\nabla V(x)|^2} = \frac{\int_{\mathbb{R}^d} \left\langle \nabla V(x), \frac{1}{\epsilon} \nabla p\left(\frac{x}{\epsilon}\right)\right\rangle \rho^\epsilon(x) \, dx}{\mathbb{E}^{\mu^\epsilon} |\nabla V(x)|^2}
$$

$$
= \frac{-\sigma \frac{1}{Z^\epsilon} \int_{\mathbb{R}^d} \left\langle \nabla V(x) e^{-\frac{\alpha}{\sigma} V(x)}, \nabla \left(e^{-\frac{1}{\sigma} p(x/\epsilon)}\right)\right\rangle dx}{\mathbb{E}^{\mu^\epsilon} |\nabla V(x)|^2}
$$

$$
= \sigma \frac{\mathbb{E}^{\mu^\epsilon}(\nabla V(x))}{\mathbb{E}^{\mu^\epsilon} |\nabla V(x)|^2} - \alpha.
$$

In deriving the penultimate line we used an integration by parts. The weak convergence of μ^ϵ to μ (second part of Proposition 5.2), formula (5.1), together with another integration by parts give

$$\lim_{\epsilon \to \infty} \frac{\mathbb{E}^{\mu^\epsilon}(\nabla V(x))}{\mathbb{E}^{\mu^\epsilon}(|\nabla V(x)|^2)} = \frac{\mathbb{E}^{\mu^\epsilon}(\nabla V(x))}{\mathbb{E}^{\mu^\epsilon}(|\nabla V(x)|^2)}$$

$$= \frac{\mathbb{E}^\mu(\nabla V(x))}{-\frac{\sigma}{\alpha} \frac{1}{Z} \int_{\mathbb{R}^d} \langle \nabla V(x), \nabla(e^{-\frac{\alpha}{\sigma} V(x)}) \rangle \, dx}$$

$$= \frac{\alpha}{\sigma}.$$

We combine the above calculations to conclude that

$$\lim_{\epsilon \to \infty} \lim_{T \to \infty} I_1(T, \epsilon) = 0 \quad \text{a.s.} \tag{7.2}$$

The proof of the convergence of the maximum likelihood estimator, Eq. (3.2) now follows from Eqs. (7.2) and (7.1).

The proof of the convergence of the estimator for the diffusion coefficient, Eq. (3.3), follows from the definition of the quadratic variation, see e.g. Ref. 4. □

Remark 7.1. *An immediate corollary of the proof of the above theorem is that*

$$\lim_{T \to \infty} \hat{A}(x^\epsilon) = \frac{\mathbb{E}^{\mu^\epsilon}(\Delta V(x))}{\mathbb{E}^{\mu^\epsilon}|\nabla V(x)|^2} \quad \text{a.s.}$$

7.3. Proof of Theorem 3.5

We combine Proposition 5.8 and (5.8) to conclude that

$$x_{n+1}^\epsilon - x_n^\epsilon = J_n - A\delta V'(x_n^\epsilon) + R(\epsilon, \delta),$$

where J_n is as defined in the proof of Proposition 5.7 and, for ϵ, δ sufficiently small and $\zeta \in (0, 1)$,

$$(\mathbb{E}^{\mu^\epsilon}|R(\epsilon, \delta)|^p)^{1/p} \leq C(\delta^{3/2} + \epsilon). \tag{7.3}$$

Notice that

$$\mathbb{E}^{\mu^\epsilon}|J_n|^2 = \mathcal{O}(\delta).$$

We combine this with formula (2.4) to obtain

$$\hat{A}_{N,\delta}(x^\epsilon) = A - \frac{\sum_{n=0}^{N-1} V'(x_n^\epsilon) J_n}{\sum_{n=1}^{N-1} |V'(x_n^\epsilon)|^2 \delta} - \frac{\sum_{n=0}^{N-1} V'(x_n^\epsilon) R(\epsilon, \delta)}{\sum_{n=0}^{N-1} |V'(x_n^\epsilon)|^2 \delta}$$

$$:= A - I_1 - I_2, \tag{7.4}$$

We need to control the terms I_1 and I_2. We start with I_1, which we rewrite in the form

$$I_1 = \epsilon^{\frac{\gamma-\zeta}{2}} \frac{\frac{1}{\sqrt{(N\delta)}} \sum_{n=0}^{N-1} V'(x_n^\epsilon) J_n}{\frac{1}{N} \sum_{n=0}^{N-1} |V'(x_n^\epsilon)|^2}.$$

The central limit theorem for (discrete) martingales implies that

$$\lim_{N \to +\infty} \frac{1}{\sqrt{(N\delta)}} \sum_{n=0}^{N-1} V'(x_n^\epsilon) J_n = \frac{1}{\sqrt{\delta}} \mathcal{N}(0, \mathbb{E}^{\mu^\epsilon}(|V'(x^\epsilon(0))|^2 |J_0|^2))$$

$$= \frac{1}{\sqrt{\delta}} \mathcal{N}(0, c\delta) = c\mathcal{N}(0, 1) \quad \text{in law,}$$

for some c uniform in $\epsilon \to 0$. In the above we have used the fact that $\mathbb{E}^{\mu^\epsilon} |J_0|^2 = 2\Sigma\delta$.

On the other hand, the ergodic theorem implies that

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{n=0}^{N-1} |V'(x_n^\epsilon)|^2 = \mathbb{E}^{\mu^\epsilon} |V(x)|^2, \quad \text{a.s.} \tag{7.5}$$

Hence, by Slutsky's theorem, and remembering that $N = [\epsilon^{-\gamma}]$, we have that

$$\lim_{\epsilon \to 0} I_1 = 0 \quad \text{in law.} \tag{7.6}$$

Consider now the term I_2. It can be written as

$$I_2 = \frac{\epsilon^{\gamma-\zeta} \sum_{n=0}^{N-1} V'(x_n^\epsilon) R(\epsilon, \delta)}{\frac{1}{N} \sum_{n=0}^{N-1} |V'(x_n^\epsilon)|^2}.$$

The ergodic theorem implies that the denominator in the above expression converges a.s. to a finite value. To study the numerator of the above expression we use estimate (7.3), together with Hölder inequality to estimate

$$\mathbb{E}^{\mu^\epsilon} \left| \epsilon^{\gamma-\zeta} \sum_{n=0}^{N-1} V'(x_n^\epsilon) R(\epsilon, \delta) \right| \le \epsilon^{\gamma-\zeta} \sum_{n=0}^{N-1} (\mathbb{E}^{\mu^\epsilon} |V'(x_n^\epsilon)|^q)^{1/q} (\mathbb{E}^{\mu^\epsilon} |R(\epsilon, \delta)|^p)^{1/p}$$

$$\le C\epsilon^{\gamma-\zeta} \sum_{n=0}^{N-1} (\mathbb{E}^{\mu^\epsilon} |R(\epsilon, \delta)|^p)^{1/p}$$

$$\le C(\epsilon^{\zeta/2} + \epsilon^{1-\zeta}).$$

In the above we have used Corollary 5.4, together with Assumptions 3.1. The above calculation shows that numerator of I_2 converges to 0 in L^1, and hence

in law. This, together with the a.s. convergence of the denominator and Slutsky's theorem gives

$$\lim_{\epsilon \to 0} I_2 = 0 \quad \text{in law.} \tag{7.7}$$

Combining (7.4), (7.6) and (7.7) completes the proof of the theorem. $\quad\square$

7.4. Proof of Theorem 3.6

We combine Proposition 5.7 with (5.8) to write the difference $x_{n+1}^\epsilon - x_n^\epsilon$ in the form

$$x_{n+1}^\epsilon - x_n^\epsilon = \sqrt{2\Sigma\delta}\xi_n + \hat{R}(\delta, \epsilon) \tag{7.8}$$

in law, where, for ϵ, δ sufficiently small,

$$(\mathbb{E}^{\mu^\epsilon} |\hat{R}(\epsilon, \delta)|^p)^{1/p} \leq C(\delta + \epsilon^\beta). \tag{7.9}$$

We substitute (7.8) into the formula for the estimator (2.2) with $d = 1$ to obtain

$$\hat{\Sigma}_{N,\delta}(x^\epsilon) = \Sigma \frac{1}{N} \sum_{n=0}^{N-1} \xi_n^2 + \frac{1}{2N\delta} \sum_{n=0}^{N-1} (\hat{R}(\delta, \epsilon))^2 + \frac{1}{N\delta} \sum_{n=0}^{N-1} \sqrt{2\Sigma\delta}\xi_n \hat{R}(\delta, \epsilon)$$

$$=: \Sigma \frac{1}{N} \sum_{n=0}^{N-1} \xi_n^2 + I_1 + I_2.$$

By the law of large numbers the first term tends almost surely to Σ as $\epsilon \to 0$ (which implies $N \to \infty$.) Thus it suffices to show that the remaining terms tend to zero in law. We do this by showing that they tend to zero in L^1.

Note that

$$\mathbb{E}^{\mu^\epsilon} |I_1| \leq C \sum_{n=0}^{N-1} \mathbb{E}^{\mu^\epsilon} (\hat{R}(\delta, +\epsilon))^2$$

$$= CN(\delta + \epsilon^\beta)^2$$

$$\leq C(\delta + \epsilon^{2\beta}\delta^{-1})$$

$$= C(\xi^\zeta + \epsilon^{2\beta-\zeta})$$

$$= o(1),$$

for $\zeta \in (0, 1)$, since β can be chosen arbitrarily close to $\frac{1}{2}$.

Similarly

$$\mathbb{E}^{\mu^\epsilon}|I_2| \le C \sum_{n=0}^{N-1} \delta^{\frac{1}{2}}(\delta, \epsilon^\beta)$$

$$\le C\left(\delta^{\frac{1}{2}} + \epsilon^\beta \delta^{-\frac{1}{2}}\right)$$

$$= C\left(\epsilon^{\frac{\zeta}{2}} + \epsilon^{\beta - \frac{\zeta}{2}}\right)$$

$$= o(1),$$

For $\zeta \in (0, 1)$, since β can be chosen arbitrarily close to $\frac{1}{2}$. This completes the proof.   □

7.5. Proof of Theorem 3.8

Taking the limit $T \to +\infty$ in (2.5) gives

$$\lim_{T\to\infty} \tilde{A}(x^\epsilon) = \hat{\Sigma} \frac{\mathbb{E}^{\mu^\epsilon}(\Delta V(x))}{\mathbb{E}^{\mu^\epsilon}|\nabla V(x)|^2} \quad \text{a.s.}$$

Proposition 5.2, Eq. (5.3) in particular, implies that

$$\lim_{\epsilon\to\infty} \hat{\Sigma} \frac{\mathbb{E}^{\mu^\epsilon}(\Delta V(x))}{\mathbb{E}^{\mu^\epsilon}|\nabla V(x)|^2} = \hat{\Sigma} \frac{\mathbb{E}^{\mu}(\Delta V(x))}{\mathbb{E}^{\mu}|\nabla V(x)|^2},$$

where $\mathbb{E}^\mu$ denotes expectation with respect to the invariant distribution $\rho(x)$ of the homogenized process, given by formula (5.1). An integration by parts now gives that

$$\mathbb{E}^\mu|\nabla V(x)|^2 = \frac{\sigma}{\alpha}\mathbb{E}^\mu(\Delta V(x)).$$

Thus, the final result of our considerations is that

$$\lim_{\epsilon\to0}\lim_{T\to\infty} \tilde{A}(X^\epsilon) = \frac{\hat{\Sigma}}{\sigma}\alpha \quad \text{a.s.}$$

□

8. CONCLUSIONS AND FUTURE WORK

The problem of parameter estimation for continuous time diffusion processes is studied in this paper. In particular our goal is to accurately fit a homogenized equation from data which has a multiscale character. In order to obtain theorems and numerical insight we generate this data from continuous time diffusion processes with two, widely separated, characteristic time scales. Our main conclusions are as follows:

- In order to estimate the drift and diffusion coefficients accurately it is necessary to subsample.
- There is an optimal subsampling rate, between the two characteristic time-scales of the multiscale data.
- The optimal subsampling rate may differ for different parameters.
- For gradient multiscale systems it is only necessary to estimate the diffusion coefficient correctly, provided one uses the second estimator for the drift-$\tilde{A}$, defined in Eqs. (2.5) and (2.6).

Both analysis and numerics are given to substantiate these claims. Many open questions remain; we list those which seem important to us.

- Question 1 from the introduction: how can multiscale data be diagnosed as such, and how can the parameter ϵ be estimated from the data in this case. In this paper we have generated simulated multiscale data by using a multiscale diffusion process. However this was done to provide a convenient analytical framework. In applications it is of interest to develop tools for characterizing the multiscale structure of a time-series-to estimate characteristic time-scales. Related work has been done in Ref. 11. Further study would be of interest.
- Rough heuristics indicate that any subsampling rate which is between the two characteristic time scales of the processes, namely $\mathcal{O}(\epsilon^2)$ and $\mathcal{O}(1)$, should enable accurate estimation of the drift and diffusion coefficients. However our analysis works only in the case where the subsampling is between $\mathcal{O}(\epsilon)$ and $\mathcal{O}(1)$. We suspect that this analysis is tight, but it would be of interest to establish this rigorously.
- Analyze other parameter estimation problems for multiscale diffusions, not necessarily of gradient form. In particular study both averaging and homogenization set-ups, as outlined in the introductory section.
- Determine precisely the range of subsamplings which will give accurate parameter estimates and optimize the subsampling rate for accuracy.
- Optimize the algorithm by combining estimates based on shifts of the subsampled data so that information is not thrown away; this is done in the context of econometrics and finance in Refs. 1, 2. More generally use kernels to facilitate optimal use of data.
- Analyze questions analogous to those raised here for multidimensional multiscale processes.
- Analyze questions analogous to those raised here for hypoelliptic multiscale diffusions; in particular the case where the homogenized equation is a fully elliptic first order Langevin equation which is derived from an overdamped second-order Langevin equation.
- Study whether there is any advantage in using random subsampling rates.

- Study drift that depends non-linearly on the parameters to be estimated, or where the noise of the model to be fitted is state-dependent.
- Parameter estimation for deterministic multiscale problems where the fast process is a strongly mixing chaotic deterministic process.
- Our work is highly dependent on having strict scale separation. It would be interesting to understand related issues for problems without such clear scale separation, such as Kac-Zwanzig heat baths.[15] The approach in Ref. 9 is potentially interesting in this context, although limited to low dimensional state space. We anticipate that the ideas and methods from our work will transfer to arbitrary dimensions.

ACKNOWLEDGMENTS

REFERENCES

1. Y. Ait-Sahalia, P. A. Mykland and L. Zhang, How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financ. Studies* **18**:351–416 (2005).
2. Y. Ait-Sahalia, P. A. Mykland and L. Zhang, A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Stat. Assoc.* **100**:1394–1411 (2005).
3. O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde and N. Shephard, Designing realised kernels to measure the ex-post variation of equity in the presence of noise. Preprint (2006).
4. I. V. Basawa and B. L. S. Prakasa Rao, *Statistical inference for stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers] (London, 1980).
5. A. Bensoussan, J. L. Lions and G. Papanicolaou, *Asymptotic analysis of periodic structures*. (North-Holland, Amsterdam, 1978).
6. C. P. Calderon, Fitting effective diffusion models to data associated with a glassy potential: Estimation, classical inference procedures and some heuristics. *SIAM Multiscale Modeling and Simulation*, to appear.
7. F. Campillo and A. Piatnitski, Effective diffusion in vanishing viscosity. In *Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XIV* (Paris, 1997/1998), volume 31 of *Stud. Math. Appl.* (pp. 133–145, North-Holland, Amsterdam, 2002).
8. D. Cioranescu and P. Donato, *An Introduction to Homogenization*. (Oxford University Press, New York, 1999).
9. D. T. Crommelin and E. Vanden-Eijnden, Reconstruction of diffusions using spectral data from timeseries. *Commun. Math. Sci.* **4**(3):651–668 (2006).
10. W. E. D. Liu and E. Vanden-Eijnden, Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* **58**(11):1544–1585 (2005).
11. J.-P. Fouque, G. Papanicolaou R. Sircar, and K. Solna, Short time scale in S and P volatility. *J. Comp. Finance* **6**(4):1–23 (2003).
12. J.-P. Fouque, G. C. Papanicolaou, and R. K. Sircar, *Derivatives in financial markets with stochastic volatility*. (Cambridge University Press, Cambridge, 2000).

13. M. Freidlin, *Functional integration and partial differential equations*, volume 109 of *Annals of Mathematics Studies*. (Princeton University Press, Princeton, NJ, 1985).
14. D. Givon, I. G. Kevrekidis and R. Kupferman, Strong convergence schemes of projective integration schemes for singularly perturbed stochastic differential equations. *Comm. Math. Sci.* **4**(4):707–729 (2006).
15. D. Givon, R. Kupferman and A. M. Stuart, Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity* **17**(6):R55–R127 (2004).
16. M. Hairer and G. A. Pavliotis, Periodic homogenization for hypoelliptic diffusions. *J. Statist. Phys.* **117**(1–2):261–279 (2004).
17. G. Humer and I. G. Kevrekidis, Coarse molecular dynamics of a peptide fragment: Free energy, kinetics and long time dynamics computations. *J. Chem. Phys.* **118**(23):10762–10773 (2003).
18. I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. (Springer-Verlag, New York, second edition 1991).
19. P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, volume 23 of Applications of Mathematics (New York). (Springer-Verlag, Berlin, 1992).
20. R. S. Liptser and A. N. Shiryaev, *Statistics of random processes. I*, volume 5 of Applications of Mathematics (New York). (Springer-Verlag, Berlin, 2001).
21. X. Mao, *Stochastic differential equations and their applications*. (Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997).
22. J. C. Mattingly, A. M. Stuart and D. J. Higham, Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101**(2):185–232 (2002).
23. S. Olla, *Homogenization of diffusion processes in random fields*. (Lecture Notes, 1994).
24. E. Pardoux, Homogenization of linear and semilinear second order parabolic PDEs with periodic coefficients: A probabilistic approach. *J. Funct. Anal.* **167**:498–520 (1999).
25. G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization* (Lecture Notes, 2006).
26. D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, volume 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. (Springer-Verlag, Berlin, third edition, 1999).
27. E. Vanden-Eijnden, Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* **1**(2):385–391 (2003).