

# Drift Estimation of Multiscale Diffusion via Filtering

Assyr Abdulle <sup>\*</sup>      Giacomo Garegnani <sup>\*</sup>      Grigorios A. Pavliotis <sup>†</sup>  
Andrew M. Stuart <sup>‡</sup>      Andrea Zanoni <sup>\*</sup>

**Abstract.** We study the problem of drift estimation for two-scale continuous time series. We set ourselves in the framework of overdamped Langevin equations, for which a single-scale surrogate homogenized equation exists. In this setting, estimating the drift coefficient of the homogenized equation requires pre-processing of the data, often in the form of subsampling. We avoid subsampling by filtering the data with an appropriate kernel function and compute maximum likelihood estimators based on the filtered process. We show that the estimators we propose are asymptotically unbiased and demonstrate numerically the advantages of our method with respect to subsampling. Finally, we show how our filtering methodology can be combined with Bayesian techniques and provide a full uncertainty quantification of the inference procedure.

**AMS subject classifications.** 62F15, 65C30, 62M05, 74Q10.

**Keywords.** Parameter inference, diffusion process, data-driven homogenization, filtering, Bayesian inference, Langevin equation.

## 1 Introduction

Efficient parameter estimation for stochastic models is essential in a wide range of applications in natural and social sciences. In several areas, the data originates from phenomena which vary continuously in time and which are endowed with a multiscale structure. This is the case, for example, in molecular dynamics, oceanography and atmosphere science or in econometrics. Frequently, it is desirable in this areas to infer from data a simpler model which captures effectively large-scale structures, or slow variations, disregarding small-scale fluctuations or treating them as a source of noise. The mismatch between the data and their desired slow-scale representation is a typical instance of a problem of model misspecification, which, if ignored or mistreated, can lead to wrong solutions. Indeed, the data, coming from the full dynamics, are compatible with the coarse-grained model only at the time scales at which the effective dynamics is valid.

In this paper we consider a simple multiscale setting arising from models of molecular dynamics, with a complete separation between a fast and a slow scale. In particular, we consider diffusion processes for which the confining potential has slow variations, and whose motion is perturbed by a fast-scale potential with rapid periodic and bounded oscillations. Given this simple class of model problems, we are interested in determining the drift coefficient of an equation of the overdamped Langevin type in which the fast-scale potential is eliminated. The theory of homogenization guarantees that such a single-scale equation can be uniquely determined, and our goal is therefore to obtain effective coarse-grained dynamics from data consistently with respect to the homogenization result.

Several methods to take into account model misspecification in multiscale frameworks as above exist. For diffusion processes, the proposed approaches rely in different measures to subsampling, which has proved itself to some extent effective in many applications, but which requires nevertheless precise knowledge of how separated the two characteristic time scales are. Robustness of this

---

<sup>\*</sup>Institute of Mathematics, École Polytechnique Fédérale de Lausanne

<sup>†</sup>Department of Mathematics, Imperial College London

<sup>‡</sup>Department of Computing and Mathematical Sciences, Caltech

methodology is dubious, too, as inference results tend to be extremely sensitive to the subsampling rate.

In the rest of the introduction, we first give a brief overview of the existing literature on the topic of deterministic and stochastic multiscale inference problems, then introduce our novel methodology and its favourable properties and conclude with an outline of this paper.

## 1.1 Literature Review

For simple models in molecular dynamics, the effect of model misspecification was studied in a series of papers [11, 12, 19, 21, 22] under the assumption of scale separation. In particular, for Brownian particles moving in two-scale potentials it was shown that, when fitting data from the full dynamics to the homogenized equation, the maximum likelihood estimator (MLE) is asymptotically biased [22, Theorem 3.4]. To be more precise, in the large sample size limit the MLE converges to the coefficients of the unhomogenized equation, rather than to those of the homogenized one. The bias of the MLE can be eliminated by subsampling at an appropriate rate, which lies between the two characteristic time scales of the problem [22, Theorems 3.5 and 3.6].

Similar techniques can be employed in econometrics, in particular for the estimation of the integrated stochastic volatility in the presence of market micro-structure noise. In this case, too, the data have to be subsampled at an appropriate rate [5, 18]. The correct subsampling rate can be in some instances rather extreme with respect to the frequency of the data and lead to get rid of more than 99% of the data. As the intuition suggests, this increases significantly the bias of the estimator, which is usually taken care of with additional bias corrections and variance reduction procedures. The need of such methodology is accentuated by data being obtained at high-frequency [4, 28].

The problem of extracting large-scale variations from multiscale data is studied in atmosphere and ocean science. In this field, too, subsampling the data is necessary to obtain an accurate coarse-grained model [9, 27].

The necessity to subsample the data can be alleviated by using appropriate martingale estimators, as was done in [13, 14]. This class of estimators can be applied to the case where the noise is multiplicative and also given by a deterministic chaotic system, as opposed to white noise. Estimators of this family have been applied to time series from paleoclimatic data and marine biology and augmented with appropriate model selection methodologies [15].

Inference of diffusion processes can be naturally performed under a Bayesian perspective. If one focuses on the drift coefficient, the form of the likelihood function guarantees, under a Gaussian prior hypothesis, that the posterior distribution is itself a Gaussian. The versatility of the Bayesian approach in the infinite-dimensional case [10, 26] gives the possibility to extend the problem of inferring the drift of a diffusion process to the non-parametric case [24, 25].

The issue of model misspecification in inverse problems with a multiscale structure has been treated in the context of partial differential equations, too. In particular, it has been shown that it is possible to infer a coarse-grained equation from data coming from the full model and to retrieve asymptotically the correct result [17]. A series of papers [1–3] focuses on retrieving the full model when the multiscale coefficient is endowed with a specific parametrized structure. Being these problems ill-posed, the latter is achieved via Tikhonov regularization [2, 17], adopting a Bayesian approach [1, 17] or exploiting techniques of Kalman filtering [3]. In [1, 3], the authors highlight the need of accounting explicitly for the modelling error due to homogenization and apply statistical techniques taken from [7, 8].

## 1.2 Our contributions

In this paper, we bypass the issue of subsampling by implementing an appropriate filtering methodology. In particular, we show that smoothing the data coming from the multiscale model with an appropriate linear time-invariant filter of the exponential family allows to retrieve the drift coefficient

of the homogenized equation. The methodology we present is not involved computationally, easy to implement in practice and presents two main advantages:

- (i) the filter's kernel depends on two parameters which can be tuned to obtain more robust results and which can be interpreted as analogous to the subsampling rate. Nevertheless, while the MLE is extremely sensitive with respect to the latter, our filtering methodology is robust with respect to its parameters, and can therefore be applied as a black-box tool for parameter estimation,
- (ii) the entire stream of data is employed, which, in practice, enhances the quality of the filter-based MLE in terms of bias. Moreover, avoiding subsampling and thus discretising the data allows us to employ continuous-time theoretical tools.

After having deduced unbiased estimators, we focus on how to insert the filtering methodology into a Bayesian approach. In particular, we analyse the effects of modifying the likelihood function if the smoothed trajectory is inserted in an appropriate manner. Under mild hypotheses on the filter parameters, we are able to show that the posterior distributions obtained with our methodology are asymptotically consistent with respect to the drift parameter of the homogenized equation. To our knowledge, this is the first instance of a successful Bayesian approach applied to multiscale diffusion processes.

### 1.3 Outline

The rest of the paper is organised as follows. In Section 2 we introduce the problem and lay the basis of our analysis setting the main assumptions and notation. In Section 3 we present our filtering methodology, with a particular focus on ergodic properties, on multiscale convergence and, naturally, on the properties of our estimators. In Section 4 we introduce the Bayesian framework and show how it can be enhanced with filtering techniques. Finally, in Section 5 we demonstrate the effectiveness of our methodology via a series of numerical experiments.

## 2 Problem setting

In this section, we introduce the class of diffusion processes which we treat in this paper and the classical methodology employed for the estimation of the drift. Let  $\varepsilon > 0$  and let us consider the one-dimensional multiscale stochastic differential equation (SDE)

$$dX_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon) dt - \frac{1}{\varepsilon} p' \left( \frac{X_t^\varepsilon}{\varepsilon} \right) dt + \sqrt{2\sigma} dW_t, \quad (2.1)$$

where, given a positive integer  $N$ , we have that  $\alpha \in \mathbb{R}^N$  and  $\sigma > 0$  are the drift and diffusion coefficients respectively and  $W_t$  is a standard one-dimensional Brownian motion. The functions  $V: \mathbb{R} \rightarrow \mathbb{R}^N$  and  $p: \mathbb{R} \rightarrow \mathbb{R}$  correspond to the slow-scale and the fast-scale confining potentials. In particular, we assume

$$V(x) = (V_1(x) \quad V_2(x) \quad \cdots \quad V_N(x))^\top, \quad (2.2)$$

for smooth functions  $V_i: \mathbb{R} \rightarrow \mathbb{R}$ ,  $i = 1, \dots, N$ . Moreover, we assume  $p$  to be smooth and periodic of period  $L$ . Theory of homogenization [6, Chapter 3] guarantees the existence of an SDE of the form

$$dX_t = -A \cdot V'(X_t) dt + \sqrt{2\Sigma} dW_t, \quad (2.3)$$

where  $W_t$  is the same Brownian motion as in (2.1), such that  $X_t^\varepsilon \rightarrow X_t$  for  $\varepsilon \rightarrow 0$  in law as random variables in  $C^0([0, T]; \mathbb{R})$ . In particular, we have  $A = K\alpha$  and  $\Sigma = K\sigma$ , where the coefficient  $0 < K < 1$  is given by the formula

$$K = \int_0^L (1 + \Phi'(y))^2 \mu(dy),$$

with

$$\mu(dy) = \frac{1}{Z} e^{-p(y)/\sigma} dy, \quad \text{where} \quad Z = \int_0^L e^{-p(y)/\sigma} dy,$$

and where the function  $\Phi$  is the unique solution with zero-mean with respect to the measure  $\mu$  of the elliptic partial differential equation

$$-p'(y)\Phi'(y) + \sigma\Phi''(y) = p'(y), \quad 0 \leq y \leq L,$$

endowed with periodic boundary conditions. **Let us remark that in this one-dimensional setting it is possible to determine  $\Phi$  explicitly, and the homogenization coefficient  $K$  is given by**

$$K = \frac{L^2}{Z\widehat{Z}},$$

**where**

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

We now briefly present the classical methodology for estimating the drift coefficient. Let  $T > 0$  and let  $X^\varepsilon := (X_t^\varepsilon, 0 \leq t \leq T)$  be a realization of the solution of (2.1) up to final time. Girsanov's change of measure formula applied to (2.3) allows to write the likelihood of  $X^\varepsilon$  given a drift coefficient  $A$  as

$$p(X^\varepsilon | A) = \exp\left(-\frac{I(X^\varepsilon | A)}{2\Sigma}\right), \quad (2.4)$$

where

$$I(X^\varepsilon | A) = \int_0^T A \cdot V'(X_t^\varepsilon) dX_t^\varepsilon + \frac{1}{2} \int_0^T (A \cdot V'(X_t^\varepsilon))^2 dt.$$

Minimizing the functional  $I(X^\varepsilon | A)$  with respect to  $A$  therefore gives the maximum likelihood estimator (MLE) of  $A$ , which can be formally computed in closed form as

$$\widehat{A}^\varepsilon(T) := \arg \min_{A \in \mathbb{R}^N} I(X^\varepsilon | A) = M^{-1}h, \quad (2.5)$$

where  $M \in \mathbb{R}^{N \times N}$  and  $h \in \mathbb{R}^N$  are defined as

$$M = \frac{1}{T} \int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon) dt, \quad h = \frac{1}{T} \int_0^T V'(X_t^\varepsilon) dX_t^\varepsilon, \quad (2.6)$$

and where  $\otimes$  denotes the outer product in  $\mathbb{R}^N$ . Let us now state the assumptions which will be employed throughout the rest of our work. In particular, we consider the same dissipative setting as [22, Assumption 3.1].

*Assumption 2.1.* The potentials  $p$  and  $V$  satisfy

- (i)  $p \in C^\infty(\mathbb{R}) \cap L^\infty(\mathbb{R})$  and is  $L$ -periodic for some  $L > 0$ ,
- (ii)  $V_i \in C^\infty(\mathbb{R})$  for all  $i = 1, \dots, N$  is polynomially bounded from above and bounded from below, and there exist  $a, b > 0$  such that

$$-\sum_{i=1}^N V'_i(x)x \leq a - bx^2.$$

- (iii)  $V'$  is Lipschitz continuous, i.e. there exists a constant  $C > 0$  such that

$$\|V'(x) - V'(y)\|_2 \leq C|x - y|,$$

- (iv) for all  $T > 0$ , the symmetric matrix  $M$  is positive definite and there exists  $\bar{\lambda} > 0$  such that  $\lambda_{\min}(M) \geq \bar{\lambda}$ .

Under these assumptions, the MLE given in (2.5) is indeed the unique minimizer of the likelihood function, as shown in [25, Theorem 2.4].

Given the convergence of  $X_t^\varepsilon \rightarrow X_t$  in the space of continuous stochastic processes, one would expect that the MLE (2.5) would be asymptotically unbiased for the drift coefficient  $A$  of the homogenized equation (2.3). Instead, it is possible to prove that in the asymptotic limit for  $T \rightarrow \infty$  and  $\varepsilon \rightarrow 0$ , the MLE tends to the drift coefficient  $\alpha$  of the unhomogenized equation (2.1). We report here this result, whose proof can be found for the case  $N = 1$  in [22, Theorem 3.4]. Let us remark that the proof for  $N > 1$  follows directly from the one-dimensional case.

**Theorem 2.2.** *Let Assumption 2.1 hold and let  $X_0^\varepsilon$  be distributed according to the invariant measure of the process  $X^\varepsilon$  solution of (2.1). Then*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \hat{A}^\varepsilon(T) = \alpha, \quad a.s.,$$

where  $\alpha$  is the drift coefficient of equation (2.1).

As anticipated in the introduction, the main tool for obtaining unbiased estimators in the literature is subsampling the data. In particular, let the dimension of the parameter  $N = 1$ , let  $\delta > 0$  and let  $T = n\delta$  with  $n$  a positive integer. Then, a subsampled estimator for  $A$  is given by

$$\hat{A}_\delta^\varepsilon(T) = - \frac{\sum_{j=0}^{n-1} V'(X_{j\delta}^\varepsilon) (X_{(j+1)\delta}^\varepsilon - X_{j\delta}^\varepsilon)}{\delta \sum_{j=0}^{n-1} V'(X_{j\delta}^\varepsilon)^2},$$

which is a discretized version of  $\hat{A}^\varepsilon(T)$ . It is possible to show [22, Theorem 3.5] that choosing  $\delta = \varepsilon^\zeta$  with  $\zeta \in (0, 1)$  and if  $T$  is sufficiently big, then  $\hat{A}_\delta^\varepsilon(T)$  is an asymptotically unbiased estimator of  $A$  with respect to  $T \rightarrow \infty$  and  $\varepsilon \rightarrow 0$ , in probability. Despite being widely employed in practice, estimators based on subsampling present some drawbacks, as discussed in the introduction. In the following, we will introduce and analyse a novel approach for the drift estimation.

*Remark 2.3.* Let us remark that for enhancing the clarity of the exposition, in this article we chose to focus on the case of a multi-dimensional parameter in the setting of one-dimensional diffusion processes. In fact, all the theory we present in the following could be generalized to the case of  $d$ -dimensional SDEs equivalent to (2.1), which can be written as

$$dX_t^\varepsilon = - \sum_{i=1}^N \alpha_i \nabla V_i(X_t^\varepsilon) dt - \frac{1}{\varepsilon} \nabla p \left( \frac{X_t^\varepsilon}{\varepsilon} \right) dt + \sqrt{2\sigma} dW_t,$$

where  $W_t$  is a standard  $d$ -dimensional Brownian motion. The proof of all results below should be slightly modified, but we verified that all arguments still hold true in the  $d$ -dimensional case.

### 3 The filtering approach

In this section, we introduce and analyse a novel filtering approach to solve the biasedness issue highlighted by Theorem 2.2. Let  $\beta, \delta > 0$  and let us consider a family of exponential kernel functions  $k: \mathbb{R}^+ \rightarrow \mathbb{R}$  defined as

$$k(r) = C_\beta \delta^{-1/\beta} e^{-r^\beta/\delta}, \quad (3.1)$$

where  $C_\beta$  is a normalizing constant given by

$$C_\beta = \beta \Gamma(1/\beta)^{-1},$$

and where  $\Gamma(\cdot)$  is the gamma function. We consider the process  $Z^\varepsilon := (Z_t^\varepsilon, 0 \leq t \leq T)$  defined by the weighted average

$$Z_t^\varepsilon := \int_0^t k(t-s) X_s^\varepsilon ds.$$

The process  $Z^\varepsilon$  can be interpreted as a smoothed version of the original trajectory  $X^\varepsilon$ . In fact, in the field of signal processing the kernel (3.1) belongs to the class of low-pass linear time-invariant filters, which cut the high frequencies in a signal to highlight its slowest components. In the following, only in case  $\beta = 1$  a rigorous analysis is carried on. Nonetheless, numerical experiments show that for higher values of  $\beta$  the performances of estimators computed employing the filter are more robust and qualitatively better.

*Remark 3.1.* Given a trajectory  $X^\varepsilon$ , it is relatively inexpensive to compute  $Z^\varepsilon$  from a computational standpoint. In particular, the process  $Z^\varepsilon$  is the truncated convolution of the kernel with the process  $X^\varepsilon$ . Hence, computational tools based on the Fast Fourier Transform (FFT) exist and allow to compute  $Z^\varepsilon$  fast component-wise. Moreover, the process  $Z^\varepsilon$  can be computed, in case  $\beta = 1$ , in a recursive manner and therefore “online”.

In the rest of this section we will focus on the properties of the process  $Z^\varepsilon$  when it is considered together with the original process  $X^\varepsilon$ . In particular, we focus on ergodic properties and multiscale convergence. Finally, we present and analyse unbiased estimators obtained with the filtering process.

### 3.1 Ergodic properties of the filter

Let us consider the filtering kernel (3.1) with  $\beta = 1$ , i.e.,

$$k(r) = \frac{1}{\delta} e^{-r/\delta}.$$

In this case, Leibniz integral rule yields the equality

$$dZ_t^\varepsilon = k(0)X_t^\varepsilon dt + \int_0^t k'(t-s)X_s^\varepsilon ds dt = \frac{1}{\delta}(X_t^\varepsilon - Z_t^\varepsilon) dt,$$

which can be interpreted as an ordinary differential equation for  $Z_t^\varepsilon$  driven by the stochastic signal  $X^\varepsilon$ . Considering the processes  $X^\varepsilon$  and  $Z^\varepsilon$  together, we obtain the system of two one-dimensional SDEs

$$\begin{aligned} dX_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon) dt - \frac{1}{\varepsilon} p' \left( \frac{X_t^\varepsilon}{\varepsilon} \right) dt + \sqrt{2\sigma} dW_t, \\ dZ_t^\varepsilon &= \frac{1}{\delta} (X_t^\varepsilon - Z_t^\varepsilon) dt. \end{aligned} \tag{3.2}$$

The first ingredient for verifying the ergodic properties of the two-dimensional process  $(X^\varepsilon, Z^\varepsilon)^\top := ((X_t^\varepsilon, Z_t^\varepsilon)^\top, 0 \leq t \leq T)$  is verifying that the measure induced by the stochastic process admits a smooth density with respect to the Lebesgue measure. Since noise is present only on the first component, this is a consequence of the theory of hypo-ellipticity, as summarized in the following Lemma, *whose proof is given in Appendix A*.

**Lemma 3.2.** *Let  $(X^\varepsilon, Z^\varepsilon)^\top$  be the solution of (3.2) and let  $\mu_t^\varepsilon$  be the measure induced by the couple at time  $t$ . Then, the measure  $\mu_t^\varepsilon$  admits a smooth density  $\rho_t^\varepsilon$  with respect to the Lebesgue measure.*

Once it is established that the law of the process admits a smooth density for all times  $t > 0$ , which satisfies a time-dependent Fokker–Planck equation, we are interested in the limiting properties of this law. In particular, we know that the process  $X^\varepsilon$  alone is geometrically ergodic [16, Theorem 4.4], and we wish the couple  $(X^\varepsilon, Z^\varepsilon)^\top$  to be endowed with the same property. The following Lemma guarantees that the couple is indeed geometrically ergodic, *and its proof is given in Appendix A*.

**Lemma 3.3.** *Let Assumption 2.1 hold and let  $b > 0$  be given in Assumption 2.1(ii). Then, if  $\delta > 1/(4b)$ , the process  $(X^\varepsilon, Z^\varepsilon)^\top$  solution of (3.2) is geometrically ergodic, i.e., there exists  $C, \lambda > 0$  such that for all measurable  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  such that for some integer  $q > 0$*

$$f(x, z) \leq 1 + \left\| \begin{pmatrix} x & z \end{pmatrix}^\top \right\|_2^q,$$

it holds

$$\left| \mathbb{E} f(X_t^\varepsilon, Z_t^\varepsilon) - \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, z) \rho^\varepsilon(x, z) dx dz \right| \leq C \left( 1 + \left\| (X_0^\varepsilon \quad Z_0^\varepsilon)^\top \right\|_2^q \right) e^{-\lambda t},$$

where  $\mathbb{E}$  denotes expectation with respect to the Wiener measure and  $\rho^\varepsilon$  is the solution to the stationary Fokker–Planck equation

$$\sigma \partial_{xx}^2 \rho^\varepsilon(x, z) + \partial_x \left( \left( \alpha \cdot V'(x) + \frac{1}{\varepsilon} p' \left( \frac{x}{\varepsilon} \right) \right) \rho^\varepsilon(x, z) \right) + \frac{1}{\delta} \partial_z ((z - x) \rho^\varepsilon(x, z)) = 0. \quad (3.3)$$

*Remark 3.4.* Let us remark that for the multiscale equation (2.1) one should check the condition above for the potential  $V(x) + p(x/\varepsilon)$ . Nevertheless, for functions  $p$  satisfying Assumption 2.1, the result holds trivially.

*Remark 3.5.* The condition  $\delta > 1/(4b)$  is not very restrictive. Let the parameter dimension  $N = 1$  and let  $V(x) \propto x^{2p}$  for an integer  $p > 1$ . Then, Assumption 2.1(ii) holds for an arbitrarily large  $b > 0$ . Therefore, the parameter of the filter  $\delta$  can be chosen along the entire positive real axis. A similar argument can be employed for higher dimensions  $N > 1$ .

*Example 3.6.* A closed form solution of (3.3) can be obtained in a simple homogenized case with the dimension of the parameter  $N = 1$ . Let  $p(y) = 0$  and the parameters  $\alpha, \sigma$  be replaced respectively by  $A$  and  $\Sigma$ . Then, if  $V(x) = x^2/2$ , equation (3.3) has the analytical solution

$$\rho^0(x, z) = \frac{1}{C_{\rho^0}} \exp \left( -\frac{A}{\Sigma} \frac{x^2}{2} - \frac{1}{\delta \Sigma} \frac{(x - (1 + A\delta)z)^2}{2} \right),$$

where

$$C_{\rho^0} = \int_{\mathbb{R}} \int_{\mathbb{R}} \exp \left( -\frac{A}{\Sigma} \frac{x^2}{2} - \frac{1}{\delta \Sigma} \frac{(x - (1 + A\delta)z)^2}{2} \right) dx dz.$$

This is the density of a multivariate normal distribution  $\mathcal{N}(0, \Gamma)$ , where the covariance matrix is given by

$$\Gamma = \frac{\Sigma}{A(1 + A\delta)} \begin{pmatrix} 1 + A\delta & 1 \\ 1 & 1 \end{pmatrix}.$$

Let us remark that this distribution can be obtained from direct computations involving Gaussian processes. In particular, it is known that  $X \sim \mathcal{GP}(\mu_t, \mathcal{C}(t, s))$ , where at stationarity  $\mu_t = 0$  and

$$\mathcal{C}(t, s) = \frac{\Sigma}{A} e^{-A|t-s|}.$$

The basic properties of Gaussian processes imply that  $Z$  is a Gaussian process, and that the couple  $(X, Z)^\top$  is a Gaussian process, too, whose mean and covariance are computable explicitly.

In a general case, it is not possible to find an explicit solution to (3.3). Nevertheless, it is possible to show some relevant properties of the solution itself, which are summarized in the following Lemma, **whose proof is given in Appendix A**.

**Lemma 3.7.** *Under the assumptions of Lemma 3.3, let  $\rho^\varepsilon$  be the solution of (3.3) and let us write*

$$\rho^\varepsilon(x, z) = \varphi^\varepsilon(x) \psi^\varepsilon(z) R^\varepsilon(x, z), \quad (3.4)$$

where  $\varphi^\varepsilon$  and  $\psi^\varepsilon$  are the marginal densities of  $X^\varepsilon$  and  $Z^\varepsilon$  respectively, i.e.,

$$\varphi^\varepsilon(x) = \int_{\mathbb{R}} \rho^\varepsilon(x, z) dz, \quad \psi^\varepsilon(z) = \int_{\mathbb{R}} \rho^\varepsilon(x, z) dx.$$

Then, it holds

$$\varphi^\varepsilon(x) = \frac{1}{C_{\varphi^\varepsilon}} \exp \left( -\frac{1}{\sigma} \alpha \cdot V(x) - \frac{1}{\sigma} p \left( \frac{x}{\varepsilon} \right) \right), \quad (3.5)$$

where

$$C_{\varphi^\varepsilon} = \int_{\mathbb{R}} \exp \left( -\frac{1}{\sigma} \alpha \cdot V(x) - \frac{1}{\sigma} p \left( \frac{x}{\varepsilon} \right) \right) dx.$$

Moreover, it holds

$$\sigma \delta \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^\varepsilon(x) \psi^\varepsilon(z) \partial_x R^\varepsilon(x, z) dx dz = \mathbb{E}^{\rho^\varepsilon} [((X^\varepsilon)^2 - (Z^\varepsilon)^2) V''(Z^\varepsilon)].$$

### 3.2 Multiscale convergence

We now investigate the convergence of the couple  $(X^\varepsilon, Z^\varepsilon)^\top$  with respect to the multiscale parameter  $\varepsilon \rightarrow 0$ . In particular, it is known that the invariant measure of  $X^\varepsilon$  converges weakly to the invariant measure of  $X$  solution of the homogenized equation (2.3). The following result, **whose proof is given in Appendix A**, guarantees the same kind of convergence for the couple  $(X^\varepsilon, Z^\varepsilon)^\top$ .

**Lemma 3.8.** *Under Assumption 2.1, let  $\mu^\varepsilon$  be the invariant measure of the couple  $(X^\varepsilon, Z^\varepsilon)^\top$  and let  $(X_0^\varepsilon, Z_0^\varepsilon)^\top \sim \mu^\varepsilon$ . Then, the measure  $\mu^\varepsilon$  converges weakly to the measure  $\mu^0(dx, dz) = \rho^0(x, z) dx dz$ , whose density  $\rho^0$  is the unique solution of the Fokker–Planck equation*

$$\Sigma \partial_{xx}^2 \rho^0(x, z) + \partial_x (A \cdot V'(x) \rho^0(x, z)) + \frac{1}{\delta} \partial_z ((z - x) \rho^0(x, z)) = 0, \quad (3.6)$$

where  $A$  and  $\Sigma$  are the coefficients of the homogenized equation (2.3).

We conclude this section presenting an analogous result to Lemma 3.7 for the limit distribution.

**Corollary 3.9.** *Let  $\rho^0$  be the solution of (3.6) and let us write*

$$\rho^0(x, z) = \varphi^0(x) \psi^0(z) R^0(x, z),$$

where  $\varphi^0$  and  $\psi^0$  are the marginal densities, i.e.,

$$\varphi^0(x) = \int_{\mathbb{R}} \rho^0(x, z) dz, \quad \psi^0(z) = \int_{\mathbb{R}} \rho^0(x, z) dx.$$

Then, if  $A$  and  $\Sigma$  are the coefficients of the homogenized equation (2.3), it holds

$$\varphi^0(x) = \frac{1}{C_{\varphi^0}} \exp \left( -\frac{1}{\Sigma} A \cdot V(x) \right), \quad \text{where} \quad C_{\varphi^0} = \int_{\mathbb{R}} \exp \left( -\frac{1}{\Sigma} A \cdot V(x) \right) dx.$$

Moreover, it holds

$$\Sigma \delta \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^0(x) \psi^0(z) \partial_x R^0(x, z) dx dz = \mathbb{E}^{\rho^0} [(X^2 - Z^2) V''(Z)].$$

*Proof.* The proof is directly obtained from Lemma 3.7 replacing  $p(y) = 0$  and  $\alpha, \sigma$  by  $A, \Sigma$  respectively.  $\square$

### 3.3 The filtering-based estimator

We now consider the inference problem and propose our filtering-based estimator of the drift, which is formally given by the formula

$$\hat{A}_k^\varepsilon(T) = -\widetilde{M}^{-1} \widetilde{h}, \quad (3.7)$$

where we employ the subscript  $k$  for reference to the filter's kernel in (3.1), and where

$$\widetilde{M} = \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon) dt, \quad \text{and} \quad \widetilde{h} = \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) dX_t^\varepsilon. \quad (3.8)$$

Let us remark that the formula above is obtained from (2.5) by replacing only one instance of  $X_t^\varepsilon$  with  $Z_t^\varepsilon$  in both  $M$  and  $h$ . In particular, it is fundamental for proving unbiasedness to keep in the definition of  $h$  the differential of the original process  $dX_t^\varepsilon$ . Let us furthermore remark that  $\hat{A}_k^\varepsilon(T)$  need not be the minimizer of some filtering-based likelihood function. In fact, if one were to replace  $Z_t^\varepsilon$  directly in (2.4), the symmetric part of the matrix  $\widetilde{M}$  would appear and  $\hat{A}_k^\varepsilon(T)$  would not be the minimizer. Therefore, the estimator  $\hat{A}_k^\varepsilon(T)$  has to be thought of as a perturbation of  $\hat{A}^\varepsilon(T)$ , directly at the level of estimators and after the maximization procedure. The only theoretical guarantee which is still needed for the well-posedness of  $\hat{A}_k^\varepsilon(T)$  is for  $\widetilde{M}$  to be invertible, which we assume to be true and which we observed to hold in practice.

We are now able to present the main result of unbiasedness of the filtering-based estimator.



**Theorem 3.10.** *Let the assumptions of Lemma 3.3 and Lemma 3.8 hold, and let  $\widehat{A}_k^\varepsilon(T)$  be defined in (3.7). If  $\widetilde{M}$  is invertible, then*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widehat{A}_k^\varepsilon(T) = A, \quad \text{a.s.},$$

where  $A$  is the drift coefficient of the homogenized equation (2.1).

*Proof.* Replacing the expression of  $dX_t^\varepsilon$  into (3.8), we get for  $\widetilde{h}$

$$\widetilde{h} = -\widetilde{M}\alpha - \frac{1}{T} \int_0^T \frac{1}{\varepsilon} p' \left( \frac{X_t^\varepsilon}{\varepsilon} \right) V'(Z_t^\varepsilon) dt + \frac{\sqrt{2\sigma}}{T} \int_0^T V'(Z_t^\varepsilon) dW_t.$$

Therefore, we have

$$\begin{aligned} \widehat{A}_k^\varepsilon(T) &= \alpha + \frac{1}{T} \widetilde{M}^{-1} \int_0^T \frac{1}{\varepsilon} p' \left( \frac{X_t^\varepsilon}{\varepsilon} \right) V'(Z_t^\varepsilon) dt - \frac{\sqrt{2\sigma}}{T} \widetilde{M}^{-1} \int_0^T V'(Z_t^\varepsilon) dW_t \\ &=: \alpha + I_1^\varepsilon(T) - I_2^\varepsilon(T). \end{aligned} \quad (3.9)$$

We study the terms  $I_1^\varepsilon(T)$  and  $I_2^\varepsilon(T)$  separately. The ergodic theorem applied to  $I_1^\varepsilon(T)$  yields

$$\lim_{T \rightarrow \infty} I_1^\varepsilon(T) = \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \mathbb{E}^{\rho^\varepsilon} \left[ \frac{1}{\varepsilon} p' \left( \frac{X^\varepsilon}{\varepsilon} \right) V'(Z^\varepsilon) \right], \quad \text{a.s.} \quad (3.10)$$

Due to Lemma 3.7 and integrating by parts, we have

$$\begin{aligned} \mathbb{E}^{\rho^\varepsilon} \left[ \frac{1}{\varepsilon} p' \left( \frac{X^\varepsilon}{\varepsilon} \right) V'(Z^\varepsilon) \right] &= \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \frac{1}{\varepsilon} p' \left( \frac{x}{\varepsilon} \right) \frac{1}{C_{\varphi^\varepsilon}} e^{-\frac{1}{\sigma} \alpha \cdot V(x)} e^{-\frac{1}{\sigma} p(\frac{x}{\varepsilon})} \psi^\varepsilon(z) R^\varepsilon(x, z) dx dz \\ &= -\sigma \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{d}{dx} \left( e^{-\frac{1}{\sigma} p(\frac{x}{\varepsilon})} \right) \frac{1}{C_{\varphi^\varepsilon}} e^{-\frac{1}{\sigma} \alpha \cdot V(x)} V'(z) \psi^\varepsilon(z) R^\varepsilon(x, z) dx dz \\ &= \sigma \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{C_{\varphi^\varepsilon}} e^{-\frac{1}{\sigma} p(\frac{x}{\varepsilon})} \partial_x \left( e^{-\frac{1}{\sigma} \alpha \cdot V(x)} R^\varepsilon(x, z) \right) V'(z) \psi^\varepsilon(z) dx dz, \end{aligned}$$

which implies

$$\begin{aligned} \mathbb{E}^{\rho^\varepsilon} \left[ \frac{1}{\varepsilon} p' \left( \frac{X^\varepsilon}{\varepsilon} \right) V'(Z^\varepsilon) \right] &= - \left( \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \otimes V'(x) \rho^\varepsilon(x, z) dx dz \right) \alpha \\ &\quad + \sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^\varepsilon(x) \psi^\varepsilon(z) \partial_x R^\varepsilon(x, z) dx dz \\ &= -\mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)] \alpha + \sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^\varepsilon(x) \psi^\varepsilon(z) \partial_x R^\varepsilon(x, z) dx dz. \end{aligned}$$

Replacing the equality above into (3.10), we obtain

$$\lim_{T \rightarrow \infty} I_1^\varepsilon(T) = -\alpha + \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^\varepsilon(x) \psi^\varepsilon(z) \partial_x R^\varepsilon(x, z) dx dz, \quad \text{a.s.}$$

Due to Lemma 3.7, we therefore have

$$\lim_{T \rightarrow \infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta} \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \mathbb{E}^{\rho^\varepsilon} [((X^\varepsilon)^2 - (Z^\varepsilon)^2) V''(Z^\varepsilon)], \quad \text{a.s.}$$

We now pass to the limit as  $\varepsilon$  goes to zero and

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta} \mathbb{E}^{\rho^0} [V'(Z) \otimes V'(X)]^{-1} \mathbb{E}^{\rho^0} [(X^2 - Z^2) V''(Z)], \quad \text{a.s.},$$

where the function  $\rho^0$  is the density of the limit invariant distribution for  $\varepsilon \rightarrow 0$ , as in Lemma 3.8. Due to Corollary 3.9, we have

$$\frac{1}{\delta} \mathbb{E}^{\rho^0} [(X^2 - Z^2) V''(Z)] = \Sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^0(x) \psi^0(z) \partial_x R^0(x, z) dx dz,$$

and moreover, an integration by parts yields

$$\begin{aligned}
\frac{1}{\delta} \mathbb{E}^{\rho^0}[(X^2 - Z^2)V''(Z)] &= -\Sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z)(\varphi^0)'(x)\psi^0(z)R^0(x, z) \, dx \, dz \\
&= -\Sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \frac{d}{dx} \left( \frac{1}{C_{\varphi^0}} e^{-\frac{1}{\Sigma} A \cdot V(x)} \right) \psi^0(z) R^0(x, z) \, dx \, dz \\
&= \left( \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \otimes V'(x) \rho^0(x, z) \, dx \, dz \right) A \\
&= \mathbb{E}^{\rho^0}[V'(Z) \otimes V'(X)]A.
\end{aligned}$$

We can therefore conclude that

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} I_1^\varepsilon(T) = -\alpha + A, \quad \text{a.s.} \quad (3.11)$$

We now consider the second term  $I_2^\varepsilon(T)$ , and rewrite it as

$$I_2^\varepsilon(T) = \sqrt{2\sigma} I_{2,1}^\varepsilon(T) I_{2,2}^\varepsilon(T),$$

where

$$\begin{aligned}
I_{2,1}^\varepsilon(T) &:= \left( \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon) \, dt \right)^{-1} \left( \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) \otimes V'(Z_t^\varepsilon) \, dt \right), \\
I_{2,2}^\varepsilon(T) &:= \left( \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) \otimes V'(Z_t^\varepsilon) \, dt \right)^{-1} \left( \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) \, dW_t \right).
\end{aligned}$$

The ergodic theorem yields

$$\lim_{T \rightarrow \infty} I_{2,1}^\varepsilon(T) = \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)] =: R^\varepsilon,$$

where  $R^\varepsilon$  is bounded uniformly in  $\varepsilon$ . Moreover, the strong law of large numbers for martingales implies

$$\lim_{T \rightarrow \infty} I_{2,2}^\varepsilon(T) = 0, \quad \text{a.s.},$$

independently of  $\varepsilon$ . Therefore

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} I_2^\varepsilon(T) = 0, \quad \text{a.s.},$$

which, together with (3.11) and (3.9), proves the desired result.  $\square$

Central limit theorem text.

**Theorem 3.11.** *Central limit theorem statement:*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \sqrt{T} \left( \widehat{A}_k^\varepsilon(T) - A \right) = \Lambda, \quad \text{in law,}$$

where  $\Lambda \sim \mathcal{N}(0, \Gamma)$  and

$$\Gamma = 2\sigma \mathbb{E}^{\rho^0} [V'(Z) \otimes V'(X)]^{-1} \mathbb{E}^{\rho^0} [V'(Z) \otimes V'(Z)] \mathbb{E}^{\rho^0} [V'(X) \otimes V'(Z)]^{-1}.$$

*Proof.* Let us consider the decomposition (3.9), which yields

$$\sqrt{T} \left( \widehat{A}_k^\varepsilon(T) - A \right) = \sqrt{T} (\alpha - A + I_1^\varepsilon(T)) - \sqrt{T} I_2^\varepsilon(T).$$

**Second term is OK:** We rewrite the term involving  $I_2^\varepsilon(T)$  as

$$\sqrt{T} I_2^\varepsilon(T) = \frac{\sqrt{2\sigma}}{\sqrt{T}} \widetilde{M}^{-1} Q_T^\varepsilon,$$

where, since  $Z_t^\varepsilon$  is adapted with respect to the natural filtration  $\mathcal{F}_t$  of the Wiener process  $W := (W_t, t \geq 0)$ , the quantity

$$Q_T^\varepsilon := \int_0^T V'(Z_t^\varepsilon) dW_t,$$

is a martingale whose quadratic variation is given by

$$\langle Q^\varepsilon \rangle_T = \int_0^T V'(Z_t^\varepsilon) \otimes V'(Z_t^\varepsilon) dt.$$

Since the ergodic theorem guarantees that

$$\lim_{T \rightarrow \infty} \frac{\langle Q^\varepsilon \rangle_T}{T} = \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)], \quad \text{in } L^1(\mu^\varepsilon),$$

the martingale central limit theorem gives

$$\lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} Q_T^\varepsilon = \Xi^\varepsilon, \quad \text{in law},$$

where  $\Xi^\varepsilon \sim \mathcal{N}(0, \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)])$ . Now, by the ergodic theorem

$$\lim_{T \rightarrow \infty} \widetilde{M}^{-1} = \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1}, \quad \text{a.s.}$$

Therefore, we have by Slutsky's theorem that

$$\lim_{T \rightarrow \infty} \sqrt{T} I_2^\varepsilon(T) = \Lambda^\varepsilon \sim \mathcal{N}(0, \Gamma^\varepsilon), \quad \text{in law.}$$

where the covariance matrix  $\Gamma^\varepsilon$  is given by

$$\Gamma^\varepsilon = 2\sigma \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)] \mathbb{E}^{\rho^\varepsilon} [V'(X^\varepsilon) \otimes V'(Z^\varepsilon)]^{-1}.$$

Finally, we denote by  $\Lambda$  the limit for  $\varepsilon \rightarrow 0$  of the sequence of Gaussian random variables  $\Lambda^\varepsilon$ .

**Idea for the first term:** Let  $\mathcal{L}^\varepsilon$  be the generator of the joint process  $(X^\varepsilon, Z^\varepsilon)$ , i.e.,

$$\mathcal{L}^\varepsilon = \begin{pmatrix} -V'(x) - \frac{1}{\varepsilon} p' \left( \frac{x}{\varepsilon} \right) \\ \frac{1}{\delta} (x - z) \end{pmatrix} \cdot \nabla_{x,z} + \sigma \partial_{xx}^2.$$

Then, let  $\Phi^\varepsilon$  the solution of the Poisson equation

$$\mathcal{L}^\varepsilon \Phi^\varepsilon(x, z) = V'(z) \frac{1}{\varepsilon} p' \left( \frac{x}{\varepsilon} \right) - \mathbb{E}^{\rho^\varepsilon} \left[ V'(Z^\varepsilon) \frac{1}{\varepsilon} p' \left( \frac{X^\varepsilon}{\varepsilon} \right) \right].$$

An application of Itô formula [23, Remark 6.17] yields

$$\frac{1}{T} \int_0^T \frac{1}{\varepsilon} p' \left( \frac{X_t^\varepsilon}{\varepsilon} \right) V'(Z_t^\varepsilon) dt = \mathbb{E}^{\rho^\varepsilon} \left[ V'(Z^\varepsilon) \frac{1}{\varepsilon} p' \left( \frac{X^\varepsilon}{\varepsilon} \right) \right] + \frac{R_T^\varepsilon}{T} - \sqrt{2\sigma} \frac{S_T^\varepsilon}{T},$$

where the remainder  $R_T^\varepsilon$  is given by

$$R_T^\varepsilon := \Phi^\varepsilon(X_T^\varepsilon, Z_T^\varepsilon) - \Phi^\varepsilon(X_0^\varepsilon, Z_0^\varepsilon),$$

and  $S_T^\varepsilon$  is the martingale defined as

$$S_T^\varepsilon := \int_0^T \partial_x \Phi^\varepsilon(X_t^\varepsilon, Z_t^\varepsilon) dt.$$

Therefore, we have

$$\begin{aligned} \sqrt{T} (\alpha - A + I_1^\varepsilon(T)) &= \sqrt{T} \left( \alpha - A + \widetilde{M}^{-1} \mathbb{E}^{\rho^\varepsilon} \left[ V'(Z^\varepsilon) \frac{1}{\varepsilon} p' \left( \frac{X^\varepsilon}{\varepsilon} \right) \right] \right) \\ &\quad + \frac{\widetilde{M}^{-1} R_T^\varepsilon}{\sqrt{T}} - \sqrt{2\sigma} \frac{\widetilde{M}^{-1} S_T^\varepsilon}{\sqrt{T}}. \end{aligned}$$

Since  $R_T^\varepsilon$  is bounded independently of  $\varepsilon$ , we first get by the ergodic theorem that

$$\lim_{T \rightarrow \infty} \frac{\widetilde{M}^{-1} R_T^\varepsilon}{\sqrt{T}} = 0, \quad \text{a.s.}$$

Repeating the same reasoning as for  $Q_T^\varepsilon$ , we get

$$\lim_{T \rightarrow \infty} \sqrt{2\sigma} \frac{\widetilde{M}^{-1} S_T^\varepsilon}{\sqrt{T}} = \Lambda_1^\varepsilon \sim \mathcal{N}(0, \Gamma_1^\varepsilon),$$

where the covariance is given by

$$\Gamma_1^\varepsilon = 2\sigma \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \mathbb{E}^{\rho^\varepsilon} [\partial_x \Phi^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \partial_x \Phi^\varepsilon(X^\varepsilon, Z^\varepsilon)] \mathbb{E}^{\rho^\varepsilon} [V'(X^\varepsilon) \otimes V'(Z^\varepsilon)]^{-1}.$$

□

## 4 The Bayesian setting

In this section we present a Bayesian reinterpretation of the inference procedure, which, given the structure of the problem, allows to get at a full uncertainty quantification with a low computational effort.

Let us fix a Gaussian prior  $\mu_0 = \mathcal{N}(A_0, C_0)$  on  $A$ , where  $A_0 \in \mathbb{R}^N$  and  $C_0 \in \mathbb{R}^{N \times N}$  is symmetric positive definite. Then, given a final time  $T > 0$ , the posterior distribution  $\mu_{T,\varepsilon}$  admits a density  $p(A | X^\varepsilon)$  with respect to the Lebesgue measure which satisfies

$$p(A | X^\varepsilon) = \frac{1}{Z^\varepsilon} p(X^\varepsilon | A) p_0(A),$$

where  $Z^\varepsilon$  is the normalization constant,  $p_0$  is the density of  $\mu_0$ , and where the likelihood  $p(X^\varepsilon | A)$  is given in (2.4). The log-posterior density is therefore given by

$$\log p(A | X^\varepsilon) = -\log Z^\varepsilon - \frac{T}{2\Sigma} A \cdot h - \frac{T}{4\Sigma} A \cdot M A - \frac{1}{2} (A - A_0) \cdot C_0^{-1} (A - A_0),$$

where  $M$  and  $h$  are defined in (2.6). Since the log-posterior density is quadratic in  $A$ , the posterior is Gaussian, and it is therefore sufficient to determine its mean and covariance to fully characterize it. We denote by  $m_{T,\varepsilon}$  and  $C_{T,\varepsilon}$  the mean and covariance matrix, respectively. Completing the squares in the log-posterior density, we formally obtain

$$\begin{aligned} C_{T,\varepsilon}^{-1} &= C_0^{-1} + \frac{T}{2\Sigma} M, \\ C_{T,\varepsilon}^{-1} m_{T,\varepsilon} &= C_0^{-1} A_0 - \frac{T}{2\Sigma} h. \end{aligned}$$

Under Assumption 2.1, one can show that the posterior at time  $T > 0$  is indeed given by  $\mu_{T,\varepsilon} = \mathcal{N}(m_{T,\varepsilon}, C_{T,\varepsilon})$ . We are now interested in the asymptotic limit of the posterior distribution for  $T \rightarrow \infty$  and for  $\varepsilon \rightarrow 0$ .

We can now state the main result for asymptotic convergence of the posterior distribution.

**Theorem 4.1.** *Under Assumption 2.1, the posterior measure  $\mu_{T,\varepsilon} = \mathcal{N}(m_{T,\varepsilon}, C_{T,\varepsilon})$  satisfies*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \mu_{T,\varepsilon}(\cdot) = \delta(\cdot - \alpha), \quad \text{a.s.},$$

*weakly in the space of probability measures on  $\mathbb{R}^N$ , where  $\delta$  is the Dirac measure on  $\mathbb{R}^N$ .*

*Proof.* Let us first consider the covariance matrix. Hua's identity yields

$$C_{T,\varepsilon} = \frac{2\Sigma}{T} (M^{-1} - Q^{-1}),$$

where

$$Q = M + \frac{T}{2\Sigma} MC_0 M.$$

Let us first remark that due to the hypothesis on  $M$  and the ergodic theorem it holds for all  $T > 0$

$$\|M^{-1}\|_2 \leq \frac{1}{\lambda}.$$

We now have that for generic symmetric positive definite matrices  $R$  and  $S$  it holds

$$\|(R + S)^{-1}\|_2 \leq \|S^{-1}\|_2.$$

Applying this inequality to  $Q^{-1}$ , we obtain

$$\|Q^{-1}\|_2 \leq \frac{2\Sigma}{T} \|(MC_0 M)^{-1}\|_2 \leq \frac{2\Sigma}{T} \|M^{-1}\|_2^2 \|C_0^{-1}\|_2 = \frac{2\Sigma}{T\lambda^2} \|C_0^{-1}\|_2,$$

which implies

$$\lim_{T \rightarrow \infty} \|Q^{-1}\|_2 = 0,$$

and due to the triangle inequality

$$\lim_{T \rightarrow \infty} \|C_{T,\varepsilon}\|_2 = 0. \quad (4.1)$$

We proved that in the limit for  $T \rightarrow \infty$  the covariance shrinks to zero independently of  $\varepsilon$ . We now consider the mean. First, we remark that the triangle inequality yields

$$\|m_{T,\varepsilon} - \alpha\|_2 \leq \|m_{T,\varepsilon} - \hat{A}^\varepsilon(T)\|_2 + \|\hat{A}^\varepsilon(T) - \alpha\|_2.$$

For the second term, Theorem 2.2 implies

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \|\hat{A}^\varepsilon(T) - \alpha\|_2 = 0, \quad \text{a.s.}$$

Let us now consider the first term. Replacing the expression of the maximum likelihood estimator (2.5) and due to the Cauchy–Schwarz and triangle inequalities, we obtain

$$\begin{aligned} \|m_{T,\varepsilon} - \hat{A}^\varepsilon(T)\|_2 &= \frac{2\Sigma}{T} \left\| M^{-1} C_0^{-1} A_0 - Q^{-1} \left( C_0^{-1} A_0 - \frac{T}{2\Sigma} h \right) \right\|_2 \\ &\leq \frac{2\Sigma}{T\lambda} \|C_0^{-1}\|_2 \left( \|A_0\|_2 + \frac{1}{\lambda} \|h\|_2 + \frac{2\Sigma}{T\lambda} \|C_0^{-1}\|_2 \|A_0\|_2 \right). \end{aligned}$$

Moreover, the ergodic theorem and the strong law of large numbers for martingales guarantee that  $\|h\|_2$  is bounded a.s. for  $T \rightarrow \infty$ . Therefore

$$\lim_{T \rightarrow \infty} \|m_{T,\varepsilon} - \hat{A}^\varepsilon(T)\|_2 = 0, \quad \text{a.s.},$$

independently of  $\varepsilon$ . Finally,

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \|m_{T,\varepsilon} - \alpha\|_2 = 0, \quad \text{a.s.},$$

which, together with (4.1), implies the desired result.  $\square$

*Remark 4.2.* The result above has the same consequences in the Bayesian setting as Theorem 2.2 has for the MLE. In particular, it shows that the posterior distribution obtained when data is not pre-processed concentrates asymptotically on the drift coefficient of the unhomogenized equation (2.1). Moreover, a partial result which can be deduced from the proof is that in the limit for  $T \rightarrow \infty$  and for a positive value  $\varepsilon > 0$  the Bayesian and the MLE approaches are equivalent. In particular, we have for all  $\varepsilon > 0$

$$\begin{aligned}\lim_{T \rightarrow \infty} \|C_{T,\varepsilon}\|_2 &= 0, \\ \lim_{T \rightarrow \infty} \|m_{T,\varepsilon} - \hat{A}^\varepsilon(T)\|_2 &= 0,\end{aligned}$$

i.e., the weak limit of the posterior  $\mu_{T,\varepsilon}$  for  $T \rightarrow \infty$  is the Dirac delta concentrated on the limit of  $\hat{A}^\varepsilon(T)$  for  $T \rightarrow \infty$ .

## 4.1 The filtering approach

In this section, we present how to correct the faulty behaviour highlighted by Theorem 4.1 with filtering techniques. Employing the filter in the same way as for the MLE, we can modify the likelihood function and obtain

$$\hat{p}(X^\varepsilon | A) = \exp\left(-\frac{\hat{I}(X^\varepsilon | A)}{2\Sigma}\right),$$

where

$$\begin{aligned}\hat{I}(X^\varepsilon | A) &= \int_0^T A \cdot V'(Z_t^\varepsilon) dX_t^\varepsilon + \frac{1}{2} \int_0^T (A \cdot V'(X_t^\varepsilon)) (A \cdot V'(Z_t^\varepsilon)) dt \\ &= \tilde{h} \cdot A + \frac{1}{2} A \cdot \widetilde{M} A.\end{aligned}$$

Let us remark that  $\widetilde{M}$  is not symmetric. Nevertheless, we have that

$$A \cdot \widetilde{M} A = A \cdot \widetilde{M}_S A,$$

where  $\widetilde{M}_S$  is the symmetric part of  $\widetilde{M}$ . Repeating the same reasoning as above, one would therefore obtain a modified posterior  $\tilde{\mu}_{T,\varepsilon} = \mathcal{N}(\tilde{m}_{T,\varepsilon}, \hat{C}_{T,\varepsilon})$ , whose covariance matrix and mean would be given by

$$\begin{aligned}\hat{C}_{T,\varepsilon}^{-1} &= C_0^{-1} + \frac{T}{2\Sigma} \widetilde{M}_S, \\ \hat{C}_{T,\varepsilon}^{-1} \tilde{m}_{T,\varepsilon} &= C_0^{-1} A_0 - \frac{T}{2\Sigma} \tilde{h}.\end{aligned}$$

There is no guarantee under Assumption 2.1 that  $\hat{C}_{T,\varepsilon}$  is positive definite and therefore a covariance matrix. Hence, in this Bayesian setting, we consider a different modification of the likelihood where we perturb only the vector  $h$ , i.e.,

$$\tilde{p}(X^\varepsilon | A) = \exp\left(-\frac{\tilde{I}(X^\varepsilon | A)}{2\Sigma}\right),$$

where

$$\begin{aligned}\tilde{I}(X^\varepsilon | A) &= \int_0^T A \cdot V'(Z_t^\varepsilon) dX_t^\varepsilon + \frac{1}{2} \int_0^T (A \cdot V'(X_t^\varepsilon))^2 dt \\ &= \tilde{h} \cdot A + \frac{1}{2} A \cdot M A.\end{aligned}$$

so that we obtain the modified posterior  $\tilde{\mu}_{T,\varepsilon} = \mathcal{N}(\tilde{m}_{T,\varepsilon}, C_{T,\varepsilon})$ , whose parameters are given by

$$\begin{aligned}C_{T,\varepsilon}^{-1} &= C_0^{-1} + \frac{T}{2\Sigma} M, \\ C_{T,\varepsilon}^{-1} \tilde{m}_{T,\varepsilon} &= C_0^{-1} A_0 - \frac{T}{2\Sigma} \tilde{h}.\end{aligned}$$

Let us remark that the posterior  $\tilde{\mu}_{T,\varepsilon}$  has the same covariance as  $\mu_{T,\varepsilon}$  and that therefore it is indeed a valid Gaussian posterior distribution. Nevertheless, in order to employ the tool of convergence introduced in Theorem 4.1, we need to study the properties of the MLE based on the likelihood  $\tilde{p}(X^\varepsilon | A)$ , i.e., the quantity

$$\tilde{A}_k^\varepsilon(T) = -M^{-1}\tilde{h}. \quad (4.2)$$

The following theorem guarantees the unbiasedness of this estimator under a condition on the parameter  $\delta$  of the filter, and we refer to Appendix B for its proof.

**Theorem 4.3.** *Let the assumptions of Theorem 3.10 hold. Then, if  $\delta$  in (3.1) satisfies  $\delta = \delta(\varepsilon)$  with  $\delta(\varepsilon) \rightarrow 0$  for  $\varepsilon \rightarrow 0$ , it holds*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \tilde{A}_k^\varepsilon(T) = A, \quad a.s.,$$

for  $\tilde{A}_k^\varepsilon(T)$  defined in (4.2).

The following result gives the convergence of the modified posterior distribution  $\tilde{\mu}_{T,\varepsilon}$  to the drift coefficient of the homogenized equation.

**Theorem 4.4.** *Let the Assumptions of Theorem 4.3 hold. Then, the modified posterior measure  $\tilde{\mu}_{T,\varepsilon} = \mathcal{N}(\tilde{m}_{T,\varepsilon}, C_{T,\varepsilon})$  satisfies*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \tilde{\mu}_{T,\varepsilon}(\cdot) = \delta(\cdot - A), \quad a.s.,$$

weakly in the space of probability measures on  $\mathbb{R}^N$ , where  $\delta$  is the Dirac measure on  $\mathbb{R}^N$ .

*Proof.* The proof follows from the proof of Theorem 4.1 and from Theorem 4.3.  $\square$

## 5 Numerical experiments

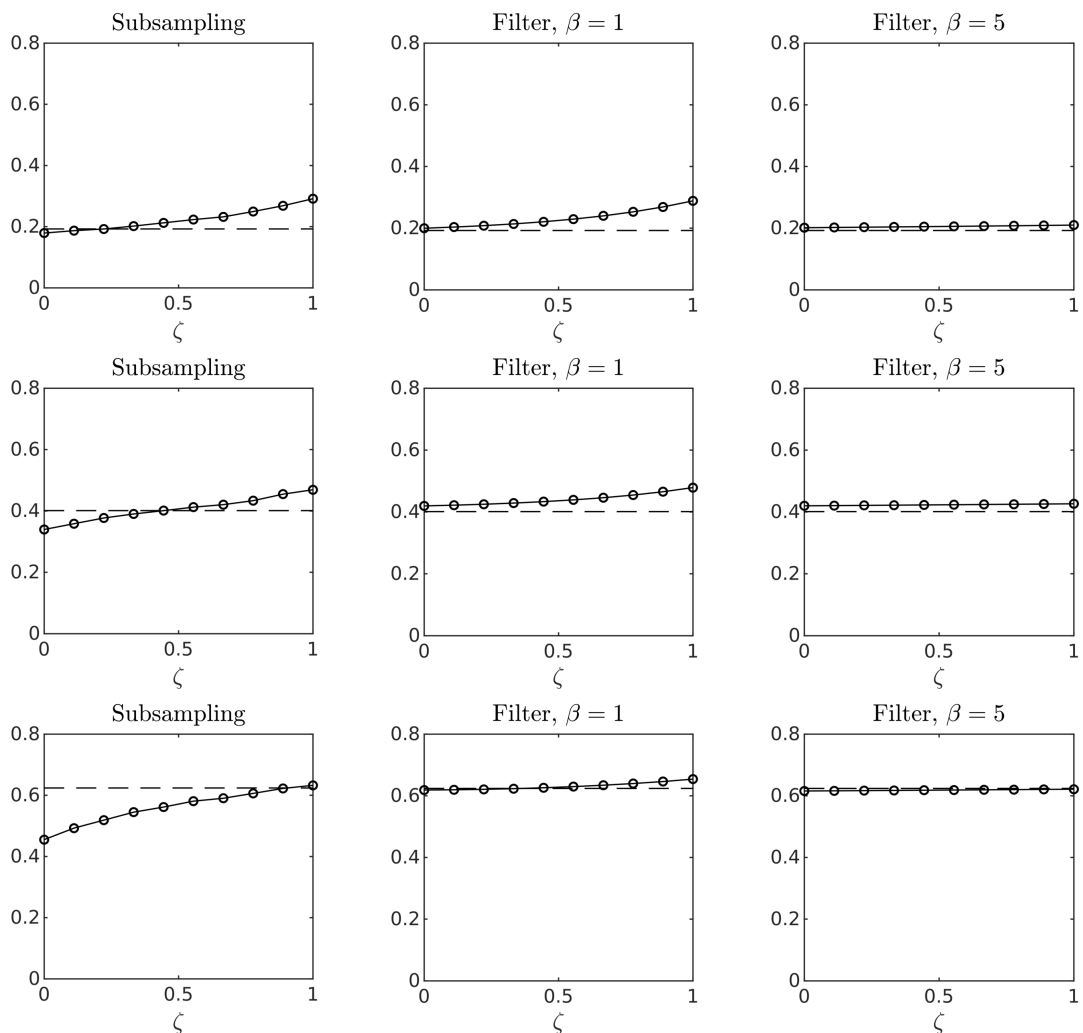
In this section we show numerical experiments confirming our theoretical findings and showing the potentiality of the filtering approach.

### 5.1 Parameters of the filter

We consider  $N = 1$  and the quadratic potential  $V(x) = x^2/2$ . In this case, the solution of the homogenized equation is an Ornstein–Uhlenbeck process. Moreover, we set the multiscale equation (2.1) with  $\varepsilon = 0.1$  and the fast potential  $p(y) = \cos(y)$ . We generate data  $X^\varepsilon$  for  $0 \leq t \leq T$  and  $T = 10^3$  employing the Euler–Maruyama method with time step  $\Delta_t = \varepsilon^3$ .

As a first experiment, we consider the effect of setting different values of the parameter  $\delta$  of the filter (3.1) on the estimator. As a comparison and for similarity with the subsampling approach, we consider  $\delta = \varepsilon^\zeta$  and vary  $\zeta \in [0, 1]$ . We employ the same values of  $\delta$  for the subsampling, too, as the theoretical results guarantee convergence only for these values. Moreover, we repeat the experiments fixing  $\beta = 1, 5$  and for three different values of the diffusion, i.e., we choose  $\sigma = 0.5, 0.7, 1$ . We report in Figure 1 the experimental results. Let us remark that

- (i) for  $\sigma = 0.5$  the results given by subsampling and by the filter with  $\beta = 1$  are similar, while for higher values of  $\sigma$  the filtering approach seems better than subsampling,
- (ii) in general, choosing a higher value of  $\beta$  seems beneficial for the quality of the estimator,
- (iii) the dependence on  $\delta$  of numerical results given by the filter seems relevant only in case  $\beta = 1$  and for small values of  $\sigma$ . For  $\beta = 1$  and higher values of  $\sigma$ , the estimator is stable with respect to this parameter, as predicted in Theorem 3.10. This can be observed for a higher value of  $\beta$  but we have no theoretical guarantee in this case.



**Figure 1:** Results for the Ornstein–Uhlenbeck process. The axis of abscissae  $\zeta$  represents the value of  $\delta = \varepsilon^\zeta$  for both subsampling and the filter. The three rows correspond to  $\sigma = 0.5, 0.7, 1.0$  from top to bottom.

As a second experiment, we therefore test the variability of the estimator with respect to  $\beta$  in (3.1). We repeat the numerical experiment in Figure 1 and we choose to consider  $\delta = \varepsilon$ , which corresponds to  $\zeta = 1$  and seems to be the worst-case scenario for the filter, at least for  $\beta = 1$ . We consider again  $\sigma = 0.5, 0.7, 1$  and vary  $\beta = 1, 2, \dots, 10$ . Results, given in Figure 2, show empirically that the estimator is fast stable with respect to  $\beta$ .

## 5.2 Multi-dimensional drift coefficient

Let us consider the Chebyshev polynomials of the first kind, i.e., the polynomials  $T_i: \mathbb{R} \rightarrow \mathbb{R}$ ,  $i = 0, 1, \dots$ , defined by the recurrence relation

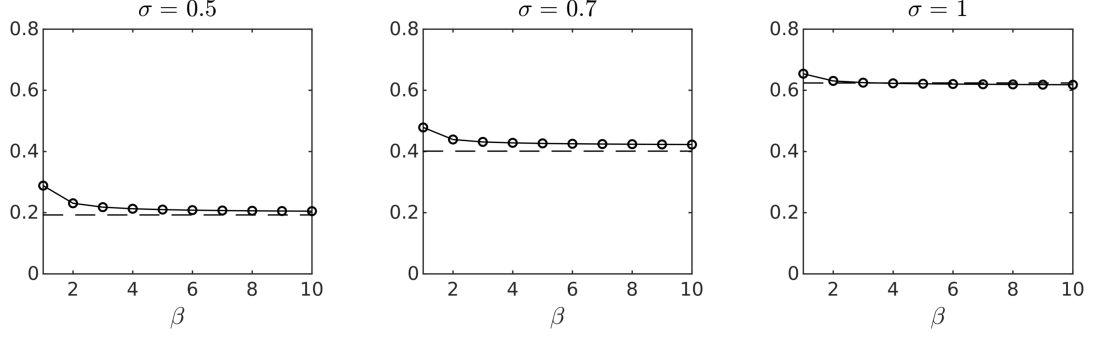
$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{i+1}(x) = 2xT_i(x) - T_{i-1}(x).$$

We consider the potential function  $V(x)$  as in (2.2) with

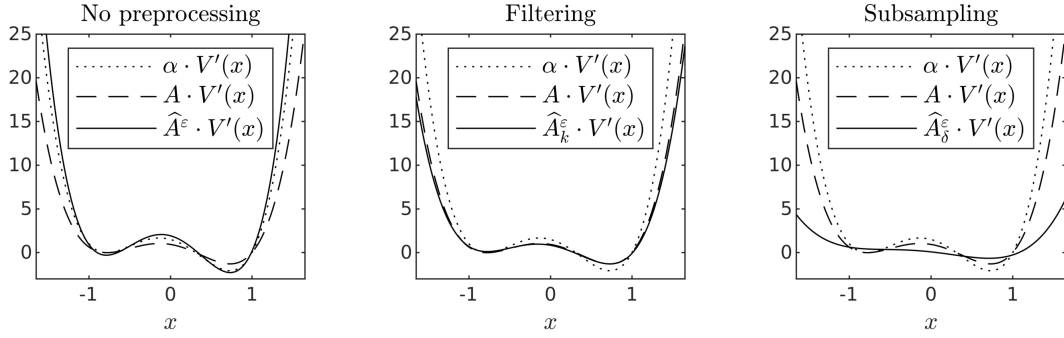
$$V_i(x) = T_i(x), \quad i = 1, \dots, 4.$$

This potential function satisfies Assumption 2.1 whenever  $N$  is even and if the leading coefficient  $\alpha_N$  is positive. We set  $N = 4$  and the drift coefficient  $\alpha = (-1, -1/2, 1/2, 1)$ . With this drift





**Figure 2:** Dependence of the estimator given by the filter for the Ornstein–Uhlenbeck process on the parameter  $\beta$  in (3.1). From left to right we consider different values of  $\sigma$ .



**Figure 3:** Results for the four-dimensional parameter. From left to right the potential function estimated with the data itself, the filter, subsampled data.

coefficient, the potential function is of the bistable kind. Moreover, we set  $\varepsilon = 0.05$ , the diffusion coefficient  $\sigma = 1$ , the fast potential  $p(y) = \cos(y)$  and simulate a trajectory of  $X^\varepsilon$  for  $0 \leq t \leq T$  with  $T = 10^3$  employing the Euler–Maruyama method with time step  $\Delta_t = \varepsilon^3$ . We estimate the drift coefficient  $A \in \mathbb{R}^4$  with the estimators

- (i)  $\hat{A}^\varepsilon(T)$  based on the data  $X^\varepsilon$  itself,
- (ii)  $\hat{A}_\delta^\varepsilon(T)$  based on subsampled data with subsampling parameter  $\delta = \varepsilon^{2/3}$ ,
- (iii)  $\hat{A}_k^\varepsilon(T)$  based on filtered data  $Z^\varepsilon$  computed with  $\beta = 1$  and  $\delta = 1$ .

In particular, we pick this specific value of  $\delta$  for the subsampling following the optimality criterion given in [22]. Results, given in Figure 3 and Table 1, show that the filter-based estimation captures well the homogenized potential as well as the coefficient  $A$ . Moreover, it is possible to remark the negative result given by Theorem 2.2 holds in practice, i.e., with no pre-processing the estimator  $\hat{A}^\varepsilon(T)$  tends to the drift coefficient  $\alpha$  of the unhomogenized equation.

### 5.3 Central limit theorem

### 5.4 The Bayesian approach: bistable potential

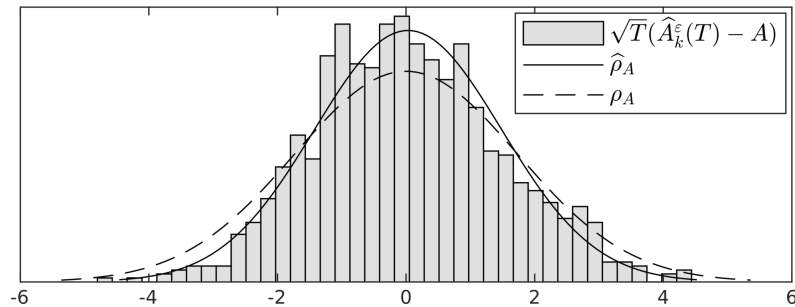
In this numerical experiment we consider  $N = 2$  and the bistable potential, i.e., the function  $V$  defined as

$$V(x) = \begin{pmatrix} \frac{x^4}{4} & -\frac{x^2}{2} \end{pmatrix}^\top,$$

with coefficients  $\alpha_1 = 1$  and  $\alpha_2 = 2$ . We then consider the multiscale equation with  $\sigma = 0.7$ , the fast potential  $p(y) = \cos(y)$  and  $\varepsilon = 0.05$ , thus simulating a trajectory  $X^\varepsilon$ . We adopt here a

Coefficient	Multiscale	Homogenized	No preprocessing	Filtering	Subsampling
	$\alpha$	$A$	$\hat{A}^\varepsilon$	$\hat{A}_k^\varepsilon$	$\hat{A}_\delta^\varepsilon$
1	-1	-0.62	-0.92	-0.70	-0.59
2	-0.5	-0.31	-0.70	-0.27	0.05
3	0.5	0.31	0.55	0.31	0.14
4	1	0.62	1.22	0.57	0.13

**Table 1:** Numerical results for the four-dimensional parameter.



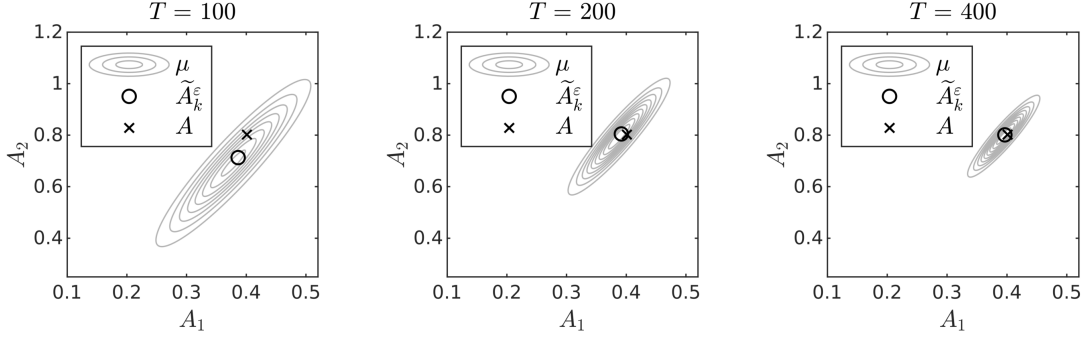
**Figure 4:** Central limit theorem result. The histogram represents numerical results, the solid curve a Gaussian fit to the latter and the dashed curve  $\rho_A$  represents the theoretical estimate given in Theorem 3.11.

Bayesian approach and compute the posterior distribution  $\tilde{\mu}_{T,\varepsilon}$  obtained with the filtering approach introduced in Section 4.1. The parameters of the filter are set to  $\beta = 1$  and  $\delta = \varepsilon$  in (3.1). Let us remark that in order to compute the posterior covariance the diffusion coefficient  $\Sigma$  of the homogenized equation has to be known. In this case, we pre-compute the value of  $\Sigma$  via the coefficient  $K$  and the theory of homogenization, but let us remark that  $\Sigma$  could be estimated employing the subsampling technique of [22]. We stop computations at times  $T = 100, 200, 400$  in order to observe the shrinkage of the Gaussian posterior towards the MLE  $\tilde{A}_k^\varepsilon(T)$  with respect to time. In Figure 5, we observe that the posterior does indeed shrink towards the MLE, which in turn gets progressively closer to the true value of the drift coefficient  $A$  of the homogenized equation.

## 6 Conclusion

In this work we considered a novel methodology based on filtering for the estimation of the drift of multiscale diffusion processes. We proved results of ergodicity, convergence and, most importantly, unbiasedness of estimators drawn from our methodology. Moreover, we combined a Bayesian approach and our new technique to guarantee a robust uncertainty quantification of the inference procedure. Numerical experiments demonstrate how the filtering-based estimator requires less knowledge of the characteristic time-scales of the multiscale equation and how it can be employed as a black-box tool for parameter estimation on a range of academic examples. We believe this work gives way to several further developments. In particular, we believe it would be relevant to

- (i) analyse the filtering methodology for  $\beta > 1$  in (3.1), which seems to give more robust results in practice,
- (ii) extend the analysis to the non-parametric framework most likely by means of Bayesian regularization techniques,
- (iii) consider multiscale models for which the homogenized equation presents multiplicative noise,
- (iv) test the filtering methodology against real-world data,



**Figure 5:** Posterior distributions over the parameter  $A = (A_1, A_2)^\top$  for the bistable potential obtained with the filtering approach. The figures refer to final time  $T = 100, 200, 400$  from left to right, respectively. The MLE  $\tilde{A}_k^\varepsilon(t)$  is represented with a circle, while the true value  $A$  of the drift coefficient of the homogenized equation is represented with a cross.

- (v) apply similar filtering methodologies to correct faulty behaviour of non likelihood-based estimators.

## Appendix A Proof of Lemmas

*Proof of Lemma 3.2.* We have to show that the joint process solution to (3.2) is hypo-elliptic. Denoting as  $f: \mathbb{R} \rightarrow \mathbb{R}$  the function

$$f(x) = -\alpha \cdot V'(x) - \frac{1}{\varepsilon} p' \left( \frac{x}{\varepsilon} \right),$$

the generator of the process  $(X^\varepsilon, Z^\varepsilon)^\top$  is given by

$$\mathcal{L} = f \partial_x + \sigma \partial_{xx}^2 + \frac{1}{\delta} (x - z) \partial_z =: \mathcal{X}_0 + \sigma \mathcal{X}_1^2,$$

where

$$\mathcal{X}_0 = f \partial_x + \frac{1}{\delta} (x - z) \partial_z, \quad \mathcal{X}_1 = \partial_x.$$

The commutator  $[\mathcal{X}_0, \mathcal{X}_1]$  applied to a test function  $v$  then gives

$$\begin{aligned} [\mathcal{X}_0, \mathcal{X}_1]v &= f \partial_x^2 v + \frac{1}{\delta} (x - z) \partial_x \partial_z v - \partial_x \left( f \partial_x v + \frac{1}{\delta} (x - z) \partial_z v \right) \\ &= -\partial_x f \partial_x v - \frac{1}{\delta} \partial_z v. \end{aligned}$$

Consequently,

$$\text{Lie}(\mathcal{X}_1, [\mathcal{X}_0, \mathcal{X}_1]) = \text{Lie} \left( \partial_x, -\partial_x f \partial_x - \frac{1}{\delta} \partial_z \right),$$

which spans  $T_x \mathbb{R}^2$ . The desired result then follows from Hörmander's theorem (see e.g. [20, Chapter 6]).  $\square$

*Proof of Lemma 3.3.* Lemma 3.2 guarantees that the Fokker–Planck equation can be written directly from the system (3.2). For geometric ergodicity, let

$$\mathcal{S}(x, z) := \left( -\sum_{i=1}^N V'_i(x) \right) \cdot \begin{pmatrix} x \\ z \end{pmatrix} = -\sum_{i=1}^N V'_i(x) x + \frac{1}{\delta} (xz - z^2).$$

Due to Assumption 2.1(ii) and Young's inequality, we then have for all  $\gamma > 0$

$$\mathcal{S}(x, z) \leq a + \left( \frac{1}{2\gamma\delta} - b \right) x^2 + \frac{1}{\delta} \left( \frac{\gamma}{2} - 1 \right) z^2.$$

We choose  $\gamma = \gamma^* := 1 - b\delta + \sqrt{1 + (1 - b\delta)^2} > 0$  so that

$$C(\gamma^*) := -\frac{1}{2\gamma^*\delta} + b = -\frac{1}{\delta} \left( \frac{\gamma^*}{2} - 1 \right),$$

and we notice that  $C(\gamma^*) > 0$  if  $\delta > 1/(4b)$ . In this case, we have

$$\mathcal{S}(x, z) \leq a - C(\gamma^*) \left\| \begin{pmatrix} x & z \end{pmatrix}^\top \right\|^2,$$

and problem (3.2) is dissipative. The result then follows from [16, Theorem 4.4].  $\square$

*Proof of Lemma 3.7.* Integrating equation (3.3) with respect to  $z$  we obtain the stationary Fokker-Planck equation for the process  $X^\varepsilon$ , i.e.

$$\sigma(\varphi^\varepsilon)''(x) + \frac{d}{dx} \left( \left( \alpha \cdot V'(x) + \frac{1}{\varepsilon} p' \left( \frac{x}{\varepsilon} \right) \right) \varphi^\varepsilon(x) \right) = 0, \quad (\text{A.1})$$

whose solution is given by

$$\varphi^\varepsilon(x) = \frac{1}{C_{\varphi^\varepsilon}} \exp \left( -\frac{1}{\sigma} \alpha \cdot V(x) - \frac{1}{\sigma} p \left( \frac{x}{\varepsilon} \right) \right),$$

and which proves (3.5). By integrating equation (3.3) with respect to  $x$  and integrating by parts we obtain

$$\psi^\varepsilon(z) + z(\psi^\varepsilon)'(z) = \frac{d}{dz} \int_{\mathbb{R}} x \rho^\varepsilon(x, z) dx,$$

which can be written as

$$\frac{d}{dz} (z \psi^\varepsilon(z)) = \frac{d}{dz} \int_{\mathbb{R}} x \rho^\varepsilon(x, z) dx.$$

Now, since  $\psi^\varepsilon$  is the density of a probability distribution, this implies that

$$z \psi^\varepsilon(z) = \int_{\mathbb{R}} x \rho^\varepsilon(x, z) dx,$$

which, replacing the decomposition (3.4), yields

$$z = \int_{\mathbb{R}} x \varphi^\varepsilon(x) R^\varepsilon(x, z) dx. \quad (\text{A.2})$$

In view of (3.4) and (A.1), equation (3.3) can be rewritten as

$$\sigma \varphi^\varepsilon \psi^\varepsilon \partial_{xx}^2 R^\varepsilon + \frac{1}{\delta} \varphi^\varepsilon \psi^\varepsilon R^\varepsilon + \sigma (\varphi^\varepsilon)' \psi^\varepsilon \partial_x R^\varepsilon + \frac{1}{\delta} (z - x) \varphi^\varepsilon ((\psi^\varepsilon)' R + \psi^\varepsilon \partial_z R^\varepsilon) = 0. \quad (\text{A.3})$$

We now multiply the equation above by  $x$  and integrate with respect to  $x$ . Let us consider some simplifications explicitly. First, an integration by parts yields

$$\sigma \psi^\varepsilon \int_{\mathbb{R}} x \varphi^\varepsilon \partial_{xx}^2 R^\varepsilon dx + \sigma \psi^\varepsilon \int_{\mathbb{R}} x (\varphi^\varepsilon)' \partial_x R^\varepsilon dx = -\sigma \psi^\varepsilon \int_{\mathbb{R}} \varphi^\varepsilon \partial_x R^\varepsilon dx. \quad (\text{A.4})$$

Then, applying (A.2), we have

$$\frac{1}{\delta} \psi^\varepsilon \int_{\mathbb{R}} x \varphi^\varepsilon R^\varepsilon dx = \frac{1}{\delta} z \psi^\varepsilon. \quad (\text{A.5})$$

Moreover, again applying (A.2), we can compute

$$\frac{1}{\delta}(\psi^\varepsilon)' \int_{\mathbb{R}} (z-x)x\varphi^\varepsilon R \, dx = \frac{1}{\delta}(\psi^\varepsilon)' \left( z^2 - \int_{\mathbb{R}} x^2 \varphi^\varepsilon R^\varepsilon \, dx \right). \quad (\text{A.6})$$

Finally, we compute the last term always applying (A.2) obtaining

$$\frac{1}{\delta}\psi^\varepsilon \int_{\mathbb{R}} (z-x)x\varphi^\varepsilon \partial_z R \, dx = \frac{1}{\delta}\psi^\varepsilon \left( z - \int_{\mathbb{R}} x^2 \varphi^\varepsilon \partial_z R^\varepsilon \, dx \right). \quad (\text{A.7})$$

Replacing the equalities (A.4), (A.5), (A.6) and (A.7) into (A.3) we obtain

$$(\psi^\varepsilon)' \left( z^2 - \int_{\mathbb{R}} x^2 \varphi^\varepsilon R^\varepsilon \, dx \right) + \psi^\varepsilon \left( 2z - \int_{\mathbb{R}} x^2 \varphi^\varepsilon \partial_z R^\varepsilon \, dx - \delta \sigma \int_{\mathbb{R}} \varphi^\varepsilon \partial_x R^\varepsilon \, dx \right) = 0.$$

We rewrite the equality above as

$$\begin{aligned} \delta \sigma \psi^\varepsilon(z) \int_{\mathbb{R}} \varphi^\varepsilon(x) \partial_x R^\varepsilon(x, z) \, dx &= (\psi^\varepsilon)'(z) z^2 - (\psi^\varepsilon)'(z) \int_{\mathbb{R}} x^2 \varphi^\varepsilon(x) R^\varepsilon(x, z) \, dx \\ &\quad + 2\psi^\varepsilon(x) z - \psi^\varepsilon \int_{\mathbb{R}} x^2 \varphi^\varepsilon(x) \partial_z R^\varepsilon(x, z) \, dx. \end{aligned}$$

Multiplying by  $V'(z)$ , integrating with respect to  $z$  and integrating by parts, we obtain the following identity in  $\mathbb{R}^N$

$$\begin{aligned} \sigma \delta \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \varphi^\varepsilon(x) \psi^\varepsilon(z) \partial_x R^\varepsilon(x, z) \, dx \, dz &= \int_{\mathbb{R}} \int_{\mathbb{R}} (x^2 - z^2) V''(z) \rho_\varepsilon(x, z) \, dx \, dz \\ &= \mathbb{E}^{\rho^\varepsilon} [(X^\varepsilon)^2 - (Z^\varepsilon)^2] V''(Z^\varepsilon), \end{aligned}$$

which is the desired result.  $\square$

*Proof of Lemma 3.8.* Let  $(X, Z)^\top := ((X_t, Z_t)^\top, 0 \leq t \leq T)$  be the solution of

$$\begin{aligned} dX_t &= -A \cdot V'(X_t) \, dt + \sqrt{2\Sigma} \, dW_t, \\ dZ_t &= \frac{1}{\delta} (X_t - Z_t) \, dt, \end{aligned}$$

with  $(X_0, Z_0)^\top \sim \mu^0$ . The arguments of Section 3.1 can be repeated to conclude that the invariant measure of  $(X, Z)^\top$  admits a smooth density  $\rho^0$  which satisfies (3.6). Moreover, standard homogenization theory (see e.g. [6, Chapter 3]) guarantees that  $(X^\varepsilon, Z^\varepsilon)^\top \rightarrow (X, Z)^\top$  for  $\varepsilon \rightarrow 0$  in law as random variables with values in  $\mathcal{C}^0([0, T]; \mathbb{R}^2)$ . The Portmanteau theorem can be employed to conclude that the measure  $\mu^\varepsilon$  converges weakly to  $\mu^0$  for  $\varepsilon \rightarrow 0$ .  $\square$

## Appendix B Proof of Theorem 4.3

Let us consider the estimator defined in (4.2), that is

$$\tilde{A}_k^\varepsilon(T) = -M^{-1} \tilde{h},$$

where the matrix  $M$  is taken from (2.6) and the vector  $\tilde{h}$  is taken from (3.8). We prove here Theorem 4.3, i.e., the unbiasedness of this estimator, for which it is necessary to prove two technical lemmas. Let us start with an estimate concerning the distance between  $X^\varepsilon$  and  $Z^\varepsilon$ .

**Lemma B.1.** *Under Assumption 2.1, let the couple  $(X^\varepsilon, Z^\varepsilon)^\top$  be distributed as its invariant measure  $\mu^\varepsilon$ . Then, if  $\delta \leq 1$ , it holds for any integer  $p \geq 1$*

$$\mathbb{E}^{\rho^\varepsilon} |X^\varepsilon - Z^\varepsilon|^{2p} \leq C (\varepsilon^{2p} + \delta^p),$$

for a constant  $C > 0$  independent of  $\varepsilon$  and  $\delta$ .

*Proof.* Let us first remark that the filtering kernel satisfies

$$\int_0^t k(t-s) ds = 1 - e^{-t/\delta}, \quad (\text{B.1})$$

which implies that the measure  $\kappa_t(ds)$  on  $(0, t)$  defined as

$$\kappa_t(ds) := \frac{k(t-s)}{1 - e^{-t/\delta}} ds,$$

is a probability measure. Let  $X_t^\varepsilon$  be at stationarity with respect to its invariant measure, which we recall having density denoted as  $\varphi^\varepsilon$ . Let  $Z_t^\varepsilon$  be the corresponding filtered process. By definition of  $Z_t^\varepsilon$  and due (B.1) we now have

$$\begin{aligned} X_t^\varepsilon - Z_t^\varepsilon &= X_t^\varepsilon - \int_0^t k(t-s) X_s^\varepsilon ds \\ &= \int_0^t k(t-s) (X_t^\varepsilon - X_s^\varepsilon) ds + e^{-t/\delta} X_t^\varepsilon \\ &= (1 - e^{-t/\delta}) \int_0^t (X_t^\varepsilon - X_s^\varepsilon) \kappa_t(ds) + e^{-t/\delta} X_t^\varepsilon. \end{aligned}$$

Therefore, Jensen's inequalities yields for a constant  $C > 0$  depending only on  $p$

$$\begin{aligned} \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - Z_t^\varepsilon|^{2p} &\leq C(1 - e^{-t/\delta})^{2p} \int_0^t \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_s^\varepsilon|^{2p} \kappa_t(ds) + C e^{-2pt/\delta} \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon|^{2p} \\ &\leq C \int_0^t \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_s^\varepsilon|^{2p} k(t-s) ds + C e^{-2pt/\delta} \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon|^{2p}, \end{aligned}$$

where we replaced the definition of  $\kappa(ds)$  and exploited the fact that  $1 - e^{-t/\delta} \leq 1$  for all  $t \geq 0$ . Due to [22, Corollary 5.4], we know that  $X_t^\varepsilon$  has bounded moments of all orders with respect to its invariant measure. Moreover, the estimate

$$\mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_s^\varepsilon|^{2p} \leq C (\varepsilon^{2p} + (t-s)^{2p} + (t-s)^p),$$

holds for all  $t > s$  due to [22, Lemma 6.1]. Finally, it is possible to verify by induction on  $r$  that for all integers  $r$  it holds

$$\int_0^t k(t-s)(t-s)^r ds \leq r! \delta^r,$$

which implies that there exists a constant  $C > 0$  depending only on  $p$  such that

$$\mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - Z_t^\varepsilon|^{2p} \leq C (\varepsilon^{2p} + \delta^p + e^{-2pt/\delta}).$$

Let us now remark that this result holds for  $X_t^\varepsilon$  being at stationarity and for  $Z_t^\varepsilon$  be its filtered process, and not for a couple  $(X^\varepsilon, Z^\varepsilon)^\top \sim \mu^\varepsilon$ . In order to conclude, we remark that due to Lemma 3.3 we have for all  $t \geq 0$  and for a constant  $C$  depending on the initial condition and  $p$

$$\mathbb{E}^{\rho^\varepsilon} |X^\varepsilon - Z^\varepsilon|^{2p} \leq \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - Z_t^\varepsilon|^{2p} + C e^{-\lambda t},$$

which, for  $t$  sufficiently big, yields

$$\mathbb{E}^{\rho^\varepsilon} |X^\varepsilon - Z^\varepsilon|^{2p} \leq C (\varepsilon^{2p} + \delta^p),$$

which is the desired result.  $\square$

Let us consider a couple  $(X^\varepsilon, Z^\varepsilon)^\top$ . In the following, we will adopt the notation

$$\mathcal{M} = \mathbb{E}^{\rho^\varepsilon} [V'(X^\varepsilon) \otimes V'(X^\varepsilon)] \quad \text{and} \quad \widetilde{\mathcal{M}} = \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)], \quad (\text{B.2})$$

which due to the ergodic theorem satisfy

$$\lim_{T \rightarrow \infty} M = \mathcal{M}, \quad \text{and} \quad \lim_{T \rightarrow \infty} \widetilde{M} = \widetilde{\mathcal{M}},$$

almost surely. In the following Lemma, we consider the difference between the two matrices  $\mathcal{M}$  and  $\widetilde{\mathcal{M}}$ .

**Lemma B.2.** *Let the assumptions of Lemma B.1 hold. Then the matrices  $\mathcal{M}$  and  $\widetilde{\mathcal{M}}$  defined in (B.2) satisfy*

$$\left\| \mathcal{M} - \widetilde{\mathcal{M}} \right\|_2 \leq C \left( \varepsilon + \delta^{1/2} \right),$$

for a constant  $C > 0$  independent of  $\varepsilon$  and  $\delta$ .

*Proof.* Applying Jensen's and Cauchy–Schwarz inequalities we have

$$\begin{aligned} \left\| \mathcal{M} - \widetilde{\mathcal{M}} \right\|_2 &\leq \mathbb{E}^{\rho^\varepsilon} \left\| (V'(Z^\varepsilon) - V'(X^\varepsilon)) \otimes V'(X^\varepsilon) \right\|_2 \\ &\leq \left( \mathbb{E}^{\rho^\varepsilon} \|V'(Z^\varepsilon) - V'(X^\varepsilon)\|_2^2 \right)^{1/2} \left( \mathbb{E}^{\rho^\varepsilon} \|V'(X^\varepsilon)\|_2^2 \right)^{1/2}. \end{aligned}$$

The Lipschitz condition on  $V$  together with the boundedness of the moments of  $X^\varepsilon$  and Lemma B.1 yield for a constant  $C > 0$

$$\left\| \mathcal{M} - \widetilde{\mathcal{M}} \right\|_2 \leq C \left( \mathbb{E}^{\rho^\varepsilon} |Z^\varepsilon - X^\varepsilon|^2 \right)^{1/2} \leq C \left( \varepsilon + \delta^{1/2} \right),$$

which is the desired result.  $\square$

We can finally prove the unbiasedness of the estimator  $\widetilde{A}_k^\varepsilon(T)$ .

*Proof of Theorem 4.3.* We first consider the difference between the two estimators  $\widetilde{A}_k^\varepsilon(T)$  and  $\widehat{A}_k^\varepsilon(T)$ . In particular, the ergodic theorem and an algebraic equality imply

$$\begin{aligned} \lim_{T \rightarrow \infty} \left( \widetilde{A}_k^\varepsilon(T) - \widehat{A}_k^\varepsilon(T) \right) &= \left( \mathcal{M}^{-1} - \widetilde{\mathcal{M}}^{-1} \right) \lim_{T \rightarrow \infty} \widetilde{h} \\ &= \mathcal{M}^{-1} \left( \mathcal{M} - \widetilde{\mathcal{M}} \right) \widetilde{\mathcal{M}}^{-1} \lim_{T \rightarrow \infty} \widetilde{h} \\ &= \mathcal{M}^{-1} \left( \mathcal{M} - \widetilde{\mathcal{M}} \right) \lim_{T \rightarrow \infty} \widehat{A}_k^\varepsilon(T), \end{aligned}$$

almost surely. Therefore, due to Assumption 2.1 which allows controlling the norm of  $\mathcal{M}^{-1}$  and due to Lemma B.2 we have for a constant  $C > 0$

$$\lim_{T \rightarrow \infty} \left\| \widetilde{A}_k^\varepsilon(T) - \widehat{A}_k^\varepsilon(T) \right\|_2 \leq C \left( \varepsilon + \delta^{1/2} \right). \quad (\text{B.3})$$

Let us remark that  $\widehat{A}_k^\varepsilon(T)$  has a bounded norm for  $\varepsilon$  sufficiently small due to Theorem 3.10. Now, the triangle inequality yields

$$\left\| \widetilde{A}_k^\varepsilon(T) - A \right\|_2 \leq \left\| \widetilde{A}_k^\varepsilon(T) - \widehat{A}_k^\varepsilon(T) \right\|_2 + \left\| \widehat{A}_k^\varepsilon(T) - A \right\|_2.$$

Therefore, due to Theorem 3.10, the inequality (B.3) and since  $\delta \rightarrow 0$  for  $\varepsilon \rightarrow 0$ , the desired result holds.  $\square$

## References

- [1] A. ABDULLE AND A. DI BLASIO, *A Bayesian numerical homogenization method for elliptic multiscale inverse problems*. Accepted for publication in SIAM-ASA J Uncertain, 2018.
- [2] ———, *Numerical homogenization and model order reduction for multiscale inverse problems*, Multiscale Model. Simul., 17 (2019), pp. 399–433.
- [3] A. ABDULLE, G. GAREGNANI, AND A. ZANONI, *Ensemble Kalman filter for multiscale inverse problems*. arXiv preprint arXiv:1908.05495, 2019.
- [4] Y. AÏT-SAHALIA AND J. JACOD, *High-frequency financial econometrics*, Princeton University Press, 2014.
- [5] Y. AÏT-SAHALIA, P. A. MYKLAND, AND L. ZHANG, *How often to sample a continuous-time process in the presence of market microstructure noise*, Rev. Financ. Stud., 18 (2005), pp. 351–416.
- [6] A. BENSOUSSAN, J.-L. LIONS, AND G. PAPANICOLAOU, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam, 1978.
- [7] D. CALVETTI, M. DUNLOP, E. SOMERSALO, AND A. STUART, *Iterative updating of model error for Bayesian inversion*, Inverse Problems, 34 (2018), pp. 025008, 38.
- [8] D. CALVETTI, O. ERNST, AND E. SOMERSALO, *Dynamic updating of numerical model discrepancy using sequential sampling*, Inverse Problems, 30 (2014), pp. 114019, 19.
- [9] C. J. COTTER AND G. A. PAVLIOTIS, *Estimating eddy diffusivities from noisy Lagrangian observations*, Commun. Math. Sci., 7 (2009), pp. 805–838.
- [10] M. DASHTI AND A. M. STUART, *The Bayesian Approach to Inverse Problems*, in Handbook of Uncertainty Quantification, Springer, 2016, pp. 1–118.
- [11] S. GAILUS AND K. SPILIOPOULOS, *Statistical inference for perturbed multiscale dynamical systems*, Stochastic Process. Appl., 127 (2017), pp. 419–448.
- [12] ———, *Discrete-time statistical inference for multiscale diffusions*, Multiscale Model. Simul., 16 (2018), pp. 1824–1858.
- [13] S. KALLIADASIS, S. KRUMSCHEID, AND G. A. PAVLIOTIS, *A new framework for extracting coarse-grained models from time series with multiscale structure*, J. Comput. Phys., 296 (2015), pp. 314–328.
- [14] S. KRUMSCHEID, G. A. PAVLIOTIS, AND S. KALLIADASIS, *Semiparametric drift and diffusion estimation for multiscale diffusions*, Multiscale Model. Simul., 11 (2013), pp. 442–473.
- [15] S. KRUMSCHEID, M. PRADAS, G. A. PAVLIOTIS, AND S. KALLIADASIS, *Data-driven coarse graining in action: Modeling and prediction of complex systems*, Physical Review E, 92 (2015), p. 042139.
- [16] J. C. MATTINGLY, A. M. STUART, AND D. J. HIGHAM, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic processes and their applications, 101 (2002), pp. 185–232.
- [17] J. NOLEN, G. A. PAVLIOTIS, AND A. M. STUART, *Multiscale modeling and inverse problems*, in Numerical analysis of multiscale problems, vol. 83 of Lect. Notes Comput. Sci. Eng., Springer, Heidelberg, 2012, pp. 1–34.
- [18] S. C. OLHEDE, A. M. SYKULSKI, AND G. A. PAVLIOTIS, *Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise*, Multiscale Model. Simul., 8 (2010), pp. 393–427.



- [19] A. PAPAVALIOU, G. A. PAVLIOTIS, AND A. M. STUART, *Maximum likelihood drift estimation for multiscale diffusions*, Stochastic Process. Appl., 119 (2009), pp. 3173–3210.
- [20] G. A. PAVLIOTIS, *Stochastic processes and applications*, vol. 60 of Texts in Applied Mathematics, Springer, New York, 2014. Diffusion processes, the Fokker-Planck and Langevin equations.
- [21] G. A. PAVLIOTIS, Y. POKERN, AND A. M. STUART, *Parameter estimation for multiscale diffusions: an overview*, in Statistical methods for stochastic differential equations, vol. 124 of Monogr. Statist. Appl. Probab., CRC Press, Boca Raton, FL, 2012, pp. 429–472.
- [22] G. A. PAVLIOTIS AND A. M. STUART, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
- [23] G. A. PAVLIOTIS AND A. M. STUART, *Multiscale methods: averaging and homogenization*, vol. 53 of Texts in Applied Mathematics, Springer, New York, 2008.
- [24] Y. POKERN, A. M. STUART, AND J. H. VAN ZANTEN, *Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs*, Stochastic Process. Appl., 123 (2013), pp. 603–628.
- [25] Y. POKERN, A. M. STUART, AND E. VANDEN-ELJNDEN, *Remarks on drift estimation for diffusion processes*, Multiscale Model. Simul., 8 (2009), pp. 69–95.
- [26] A. M. STUART, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
- [27] Y. YING, J. MADDISON, AND J. VANNESTE, *Bayesian inference of ocean diffusivity from Lagrangian trajectory data*, Ocean Model., 140 (2019).
- [28] L. ZHANG, P. A. MYKLAND, AND Y. AÏT-SAHALIA, *A tale of two time scales: determining integrated volatility with noisy high-frequency data*, J. Amer. Statist. Assoc., 100 (2005), pp. 1394–1411.