

Drift Estimation of Multiscale Diffusion via Filtering

Assyr Abdulle ^{*} Giacomo Garegnani ^{*} Grigorios A. Pavliotis [†]
Andrew M. Stuart [‡] Andrea Zanoni ^{*}

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 originate 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

^{*}Institute of Mathematics, École Polytechnique Fédérale de Lausanne

[†]Department of Mathematics, Imperial College London

[‡]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 [6, 7, 13, 14, 22, 24, 25] 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 [25, 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 [25, 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, 21]. 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, 30].

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 [11, 29].

The necessity to subsample the data can be alleviated by using appropriate martingale estimators, as was done in [15, 17]. 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 [18].

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 [12, 28] gives the possibility to extend the problem of inferring the drift of a diffusion process to the non-parametric case [26, 27].

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 [20]. 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, 20], adopting a Bayesian approach [1, 20] 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 [9, 10].

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

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

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 [8, 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 Φ is the unique solution with zero-mean with respect to the measure μ of the elliptic partial differential equation

$$-p'(y)\Phi'(y) + \sigma\Phi''(y) = p'(y), \quad 0 \leq y \leq L, \quad (2.4)$$

endowed with periodic boundary conditions. **Let us remark that in this one-dimensional setting it is possible to determine Φ 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^ε given a drift coefficient A as

$$p(X^\varepsilon | A) = \exp\left(-\frac{I(X^\varepsilon | A)}{2\Sigma}\right), \quad (2.5)$$

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

$$\widehat{A}^\varepsilon(T) := \arg \min_{A \in \mathbb{R}^N} I(X^\varepsilon | A) = M^{-1}h, \quad (2.6)$$

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

$$M = \frac{1}{T} \int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon) dt, \quad h = \frac{1}{T} \int_0^T V'(X_t^\varepsilon) dX_t^\varepsilon, \quad (2.7)$$

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 [25, Assumption 3.1].

Assumption 2.1. The potentials p and V satisfy

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

$$-\alpha \cdot V'(x)x \leq a - bx^2.$$

- (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 $\lambda_{\min}(M) \geq \bar{\lambda}$.

Remark 2.2. In the following, in particular in the proof of Lemma 3.3, we will employ Assumption 2.1(ii) for the whole drift of the SDE (2.1), i.e., the function

$$V^\varepsilon(x) := \alpha \cdot V(x) - p\left(\frac{x}{\varepsilon}\right).$$

Since $p \in L^\infty$, the assumption above is sufficient for V^ε to satisfy Assumption 2.1(ii) with different values for a and b . In particular, if the condition holds for $\alpha \cdot V$ for a given $b > 0$, then one can choose a new value for the same coefficient which is arbitrarily close to the original one for V^ε .

Under these assumptions, the MLE given in (2.6) is indeed the unique minimizer of the likelihood function, as shown in [27, 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 (2.6) 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 α of the unhomogenized equation (2.1). We report here this result, whose proof can be found for the case $N = 1$ in [25, Theorem 3.4]. Let us remark that the proof for $N > 1$ follows directly from the one-dimensional case.

Theorem 2.3. *Let Assumption 2.1 hold and let X_0^ε be distributed according to the invariant measure of the process X^ε solution of (2.1). Then*

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

where α 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 [25, 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 2.4. 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 2.3. 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_β 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^ε can be interpreted as a smoothed version of the original trajectory X^ε . 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 β the performances of estimators computed employing the filter are more robust and qualitatively better.

Remark 3.1. Given a trajectory X^ε , it is relatively inexpensive to compute Z^ε from a computational standpoint. In particular, the process Z^ε is the truncated convolution of the kernel with the process X^ε . Hence, computational tools based on the Fast Fourier Transform (FFT) exist and allow to compute Z^ε fast component-wise. Moreover, the process Z^ε 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^ε when it is considered together with the original process X^ε . 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^ε driven by the stochastic signal X^ε . Considering the processes X^ε and Z^ε 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 μ_t^ε be the measure induced by the couple at time t . Then, the measure μ_t^ε admits a smooth density ρ_t^ε 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^ε alone is geometrically ergodic [19, 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

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

where \mathbb{E} denotes expectation with respect to the Wiener measure and ρ^ε 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. 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 δ can be chosen along the entire positive real axis. A similar argument can be employed for higher dimensions $N > 1$.

Example 3.5. 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 α, σ be replaced respectively by A and Σ . 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 3.6. *Under the assumptions of Lemma 3.3, let ρ^ε 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 φ^ε and ψ^ε are the marginal densities of X^ε and Z^ε 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

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

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^ε 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 3.7. *Under Assumption 2.1, let μ^ε 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 μ^ε converges weakly to the measure $\mu^0(dx, dz) = \rho^0(x, z) dx dz$, whose density ρ^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 Σ are the coefficients of the homogenized equation (2.3).

We conclude this section presenting an analogous result to Lemma 3.6 for the limit distribution.

Corollary 3.8. *Let ρ^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 φ^0 and ψ^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 Σ 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

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

Proof. The proof is directly obtained from Lemma 3.6 replacing $p(y) = 0$ and α, σ by A, Σ 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

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

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 (2.6) by replacing only one instance of X_t^ε with Z_t^ε 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^ε . 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^ε directly in (2.5), the symmetric part of the matrix \tilde{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 \tilde{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 3.9. *Let the assumptions of Lemma 3.3 and Lemma 3.7 hold, and let $\hat{A}_k^\varepsilon(T)$ be defined in (3.7). If \tilde{M} is invertible, then*

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

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

Proof. Replacing the expression of dX_t^ε into (3.8), we get for \tilde{h}

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

Therefore, we have

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

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 3.6 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 3.6, we therefore have

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

We now pass to the limit as ε goes to zero and

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

where the function ρ^0 is the density of the limit invariant distribution for $\varepsilon \rightarrow 0$, as in Lemma 3.7. Due to Corollary 3.8, we have

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

and moreover, an integration by parts yields

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

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^ε is bounded uniformly in ε . 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 ε . 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

CASE $\delta = \varepsilon^\zeta$ (PROOF IN APPENDIX C).

Theorem 3.10. *Let the assumptions of Lemma 3.3 and Lemma 3.7 hold, let $\hat{A}_k^\varepsilon(T)$ be defined in (3.7) and let $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$. If \tilde{M} is invertible, then*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \hat{A}_k^\varepsilon(T) = A, \quad \text{in law,}$$

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

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 $p(A \mid X^\varepsilon)$ with respect to the Lebesgue measure which satisfies

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

where Z^ε is the normalization constant, p_0 is the density of μ_0 , and where the likelihood $p(X^\varepsilon \mid A)$ is given in (2.5). The log-posterior density is therefore given by

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

where M and h are defined in (2.7). 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*

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

weakly in the space of probability measures on \mathbb{R}^N , where δ 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

$$Q = M + \frac{T}{2\Sigma} M C_0 M.$$

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

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

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 ε . 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 2.3 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 (2.6) 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 ε . 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 2.3 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

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

where

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

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 $\widehat{\mu}_{T,\varepsilon} = \mathcal{N}(\widehat{m}_{T,\varepsilon}, \widehat{C}_{T,\varepsilon})$, whose covariance matrix and mean would be given by

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

There is no guarantee under Assumption 2.1 that $\widehat{C}_{T,\varepsilon}$ 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.,

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

where

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

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

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

Let us remark that the posterior $\widetilde{\mu}_{T,\varepsilon}$ 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 $\widetilde{p}(X^\varepsilon | A)$, i.e., the quantity

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

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

Theorem 4.3. *Let the assumptions of Theorem 3.9 hold. Then, if δ in (3.1) satisfies $\delta = \delta(\varepsilon)$ with $\delta(\varepsilon) \rightarrow 0$ for $\varepsilon \rightarrow 0$, it holds*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widetilde{A}_k^\varepsilon(T) = A, \quad \text{in law,}$$

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

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

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

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widetilde{\mu}_{T,\varepsilon}(\cdot) = \delta(\cdot - A), \quad \text{in law,}$$

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

Proof. The proof follows from the proof of Theorem 4.1 and from Theorem 4.3. □

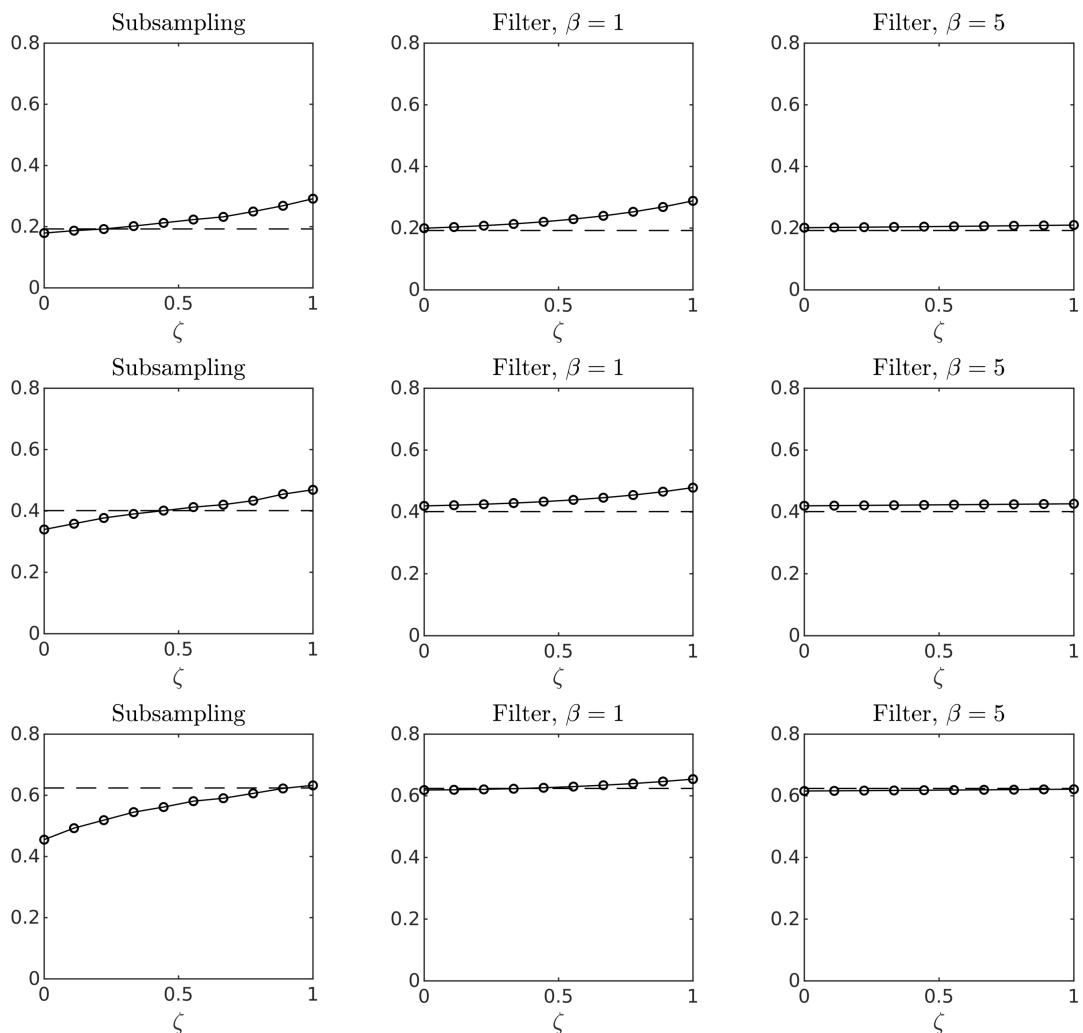


Figure 1: Results for the Ornstein–Uhlenbeck process. The axis of abscissae ζ represents the value of $\delta = \varepsilon^\zeta$ for both subsampling and the filter. The three rows correspond to $\sigma = 0.5, 0.7, 1.0$ from top to bottom.

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^ε 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 δ 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 δ 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

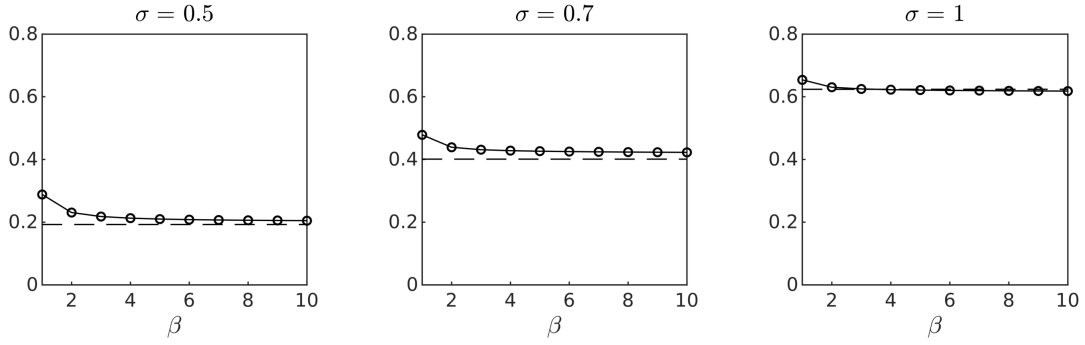


Figure 2: Dependence of the estimator given by the filter for the Ornstein–Uhlenbeck process on the parameter β in (3.1). From left to right we consider different values of σ .

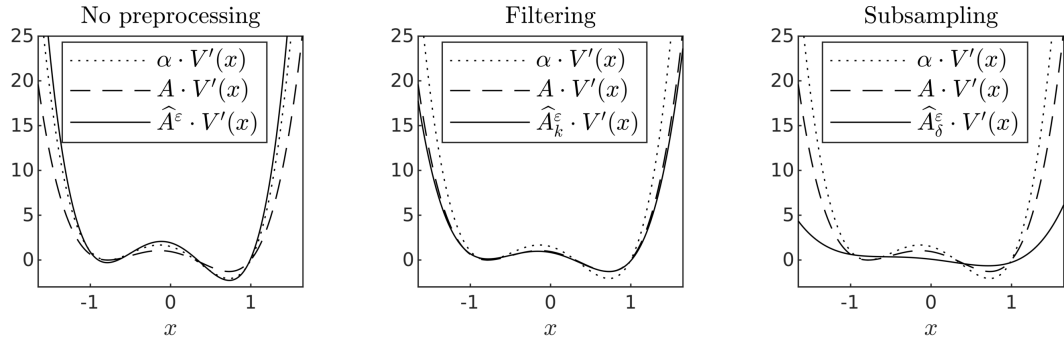


Figure 3: Results for the four-dimensional parameter. From left to right the potential function estimated with the data itself, the filter, subsampled data.

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 σ the filtering approach seems better than subsampling,
- (ii) in general, choosing a higher value of β seems beneficial for the quality of the estimator,
- (iii) the dependence on δ of numerical results given by the filter seems relevant only in case $\beta = 1$ and for small values of σ . For $\beta = 1$ and higher values of σ , the estimator is stable with respect to this parameter, as predicted in Theorem 3.9. This can be observed for a higher value of β but we have no theoretical guarantee in this case.

As a second experiment, we therefore test the variability of the estimator with respect to β 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 β .

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

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

Coefficient	Multiscale α	Homogenized A	No preprocessing \hat{A}^ε	Filtering \hat{A}_k^ε	Subsampling $\hat{A}_\delta^\varepsilon$
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.

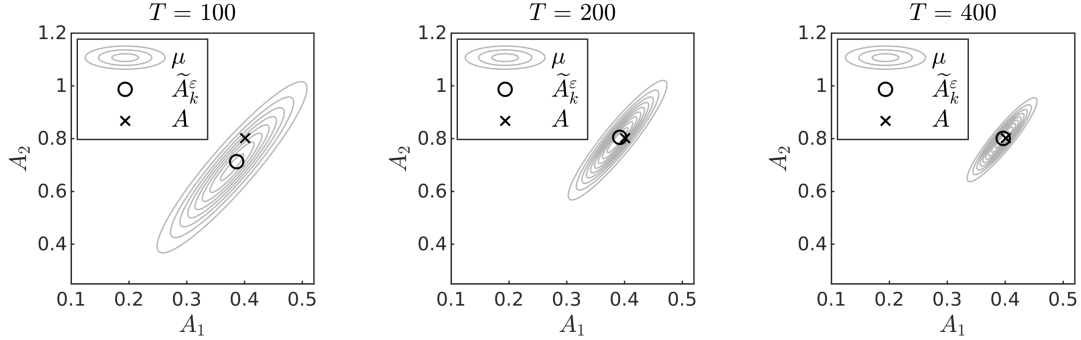


Figure 4: Posterior distributions over the parameter $A = (A_1, A_2)^\top$ for the bistable potential obtained with the filtering approach. The figures refer to final time $T = 100, 200, 400$ from left to right, respectively. The MLE $\hat{A}_k^\varepsilon(t)$ is represented with a circle, while the true value A of the drift coefficient of the homogenized equation is represented with a cross.

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

In particular, we pick this specific value of δ for the subsampling following the optimality criterion given in [25]. 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 2.3 holds in practice, i.e., with no pre-processing the estimator $\hat{A}^\varepsilon(T)$ tends to the drift coefficient α of the unhomogenized equation.

5.3 The Bayesian approach: bistable potential

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

$$V(x) = \left(\frac{x^4}{4} \quad -\frac{x^2}{2} \right)^\top,$$

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^ε . We adopt here a 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 Σ of the homogenized equation has to be known. In this case, we pre-compute the value of Σ via the coefficient K and the theory of homogenization, but let us remark that Σ could be estimated employing the subsampling technique of [25]. 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 4, 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,
- (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 the tangent space of \mathbb{R}^2 at (x, z) , denoted $T_{x,z}\mathbb{R}^2$. The desired result then follows from Hörmander's theorem (see e.g. [23, 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

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

Due to Assumption 2.1(ii), Remark 2.2 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 [19, Theorem 4.4]. \square

Proof of Lemma 3.6. Integrating equation (3.3) with respect to z we obtain the stationary Fokker–Planck equation for the process X^ε , i.e.

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

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

$$\partial_x (\sigma \varphi^\varepsilon \psi^\varepsilon \partial_x R^\varepsilon) + \partial_z \left(\frac{1}{\delta} (z - x) \varphi^\varepsilon \psi^\varepsilon R^\varepsilon \right) = 0.$$

We now multiply the equation above by a continuous differentiable function $f: \mathbb{R}^2 \rightarrow \mathbb{R}^N$, $f = f(x, z)$, and integrate with respect to x and z . Then integration by parts yields

$$\sigma \int_{\mathbb{R}} \int_{\mathbb{R}} \partial_x f(x, z) \varphi^\varepsilon(x) \psi^\varepsilon(z) \partial_x R^\varepsilon(x, z) dx dz = \frac{1}{\delta} \int_{\mathbb{R}} \int_{\mathbb{R}} \partial_z f(x, z) (x - z) \varphi^\varepsilon(x) \psi^\varepsilon(z) R^\varepsilon(x, z) dx dz,$$

which implies the following identity in \mathbb{R}^N

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

Finally, choosing

$$f(x, z) = (x - z)V'(z) + V(z),$$

we obtain the desired result. \square

Proof of Lemma 3.7. 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 ρ^0 which satisfies (3.6). Moreover, standard homogenization theory (see e.g. [8, 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 μ^ε converges weakly to μ^0 for $\varepsilon \rightarrow 0$. \square

Appendix B Estimates on the filtered solution

In this section, we prove a series of results which quantify the distance between the multiscale solution X_t^ε and the filtered trajectory Z_t^ε . First, we prove a Jensen-like inequality for the kernel of the filter.

Lemma B.1. *Let $\delta > 0$ and $k(r)$ be defined as*

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

Then, for any $t > 0$, $p \geq 0$ and any function $g \in \mathcal{C}^0(0, t)$ it holds

$$\left| \int_0^t k(t-s) g(s) ds \right|^p \leq (1 - e^{-t/\delta})^{p-1} \int_0^t k(t-s) |g(s)|^p ds.$$

Moreover, if $p \geq 1$ and since $0 < (1 - e^{-t/\delta}) < 1$ for all $t > 0$, it holds

$$\left| \int_0^t k(t-s) g(s) ds \right|^p \leq \int_0^t k(t-s) |g(s)|^p ds.$$

Proof. Let us first note that

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

Therefore, 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. An application of Jensen's inequality therefore yields

$$\begin{aligned} \left| \int_0^t k(t-s) g(s) ds \right|^p &\leq (1 - e^{-t/\delta})^p \int_0^t |g(s)|^p \kappa_t(ds) \\ &= (1 - e^{-t/\delta})^{p-1} \int_0^t k(t-s) |g(s)|^p ds, \end{aligned}$$

which is the desired result. \square

The following lemma characterizes the action of the filter when it is applied to polynomials.

Lemma B.2. *With the notation of Lemma B.1, it holds for all $p \geq 0$*

$$\int_0^t k(t-s) (t-s)^p ds \leq C \delta^p,$$

where $C > 0$ is a positive constant independent of δ .

Proof. The change of variable $u = (t - s)/\delta$ yields

$$\int_0^t k(t-s)(t-s)^p ds = \delta^p \int_0^{t/\delta} u^p e^{-u} du = \delta^p \gamma\left(p+1, \frac{t}{\delta}\right),$$

where γ is the lower incomplete Gamma function, which is bounded by the complete Gamma function $\Gamma(p+1)$ independently of the second argument. \square

We can now prove a rough estimate of the difference between the trajectories X_t^ε and Z_t^ε when they are at stationarity.

Lemma B.3. *Under Assumption 2.1, let the couple $(X^\varepsilon, Z^\varepsilon)^\top$ be distributed as its invariant measure μ^ε . Then, if $\delta \leq 1$, it holds for any $p \geq 1$*

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

for a constant $C > 0$ independent of ε and δ .

Proof. Let X_t^ε be at stationarity with respect to its invariant measure, which we recall having density denoted as φ^ε . Let Z_t^ε be the corresponding filtered process. By definition of Z_t^ε we have

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

Therefore, Lemma B.1 yields for a constant $C > 0$ depending only on p

$$\mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - Z_t^\varepsilon|^p \leq C \int_0^t \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_s^\varepsilon|^p k(t-s) ds + C e^{-pt/\delta} \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon|^p.$$

Moreover, the estimate

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

holds for all $t > s$ due to [25, Lemma 6.1]. Finally, since the moments of X_t^ε are bounded [25, Corollary 5.4] and due to Lemma B.2 there exists a constant $C > 0$ depending only on p such that

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

Let us now remark that this result holds for X_t^ε being at stationarity and for Z_t^ε 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

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

which, for t sufficiently big, yields the desired result. \square

It is relevant to study the quantities

$$\mathcal{M}_\varepsilon := \mathbb{E}^{\rho^\varepsilon} [V'(X^\varepsilon) \otimes V'(X^\varepsilon)] \quad \text{and} \quad \widetilde{\mathcal{M}}_\varepsilon := \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)], \quad (\text{B.1})$$

which are by the ergodic theorem the time limit of the matrices appearing in the estimators $\widehat{A}^\varepsilon(T)$ and $\widehat{A}_k^\varepsilon(T)$, respectively. In particular, the distance between the matrices \mathcal{M}_ε and $\widetilde{\mathcal{M}}_\varepsilon$ can be bounded, which is summarized in the following lemma.

Lemma B.4. *Let the assumptions of Lemma B.3 hold. Then the matrices \mathcal{M} and $\widetilde{\mathcal{M}}_\varepsilon$ defined in (B.1) satisfy*

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

for a constant $C > 0$ independent of ε and δ .

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

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

The Lipschitz condition on V' together with the boundedness of the moments of X^ε and Lemma B.3 yield for a constant $C > 0$

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

which is the desired result. \square

The final result presented in this section is a sharper quantification of the difference between X_t^ε and Z_t^ε than Lemma B.3. In particular, denoting $Y_t^\varepsilon := X_t^\varepsilon/\varepsilon$, we will make use of the decomposition [25, Formula 5.8]

$$\begin{aligned} X_t^\varepsilon - X_s^\varepsilon &= - \int_s^t (\alpha \cdot V'(X_r^\varepsilon))(1 + \Phi'(Y_r^\varepsilon)) \, dr \\ &\quad + \sqrt{2\sigma} \int_s^t (1 + \Phi'(Y_r^\varepsilon)) \, dW_r - \varepsilon(\Phi(Y_t^\varepsilon) - \Phi(Y_s^\varepsilon)), \end{aligned} \tag{B.2}$$

which is obtained applying the Itô formula to Φ , the solution of the cell problem (2.4).

Lemma B.5. *Let Assumption 2.1 hold and $\varepsilon, \delta > 0$ be sufficiently small. Then, it holds for every $t > 0$ sufficiently big*

$$X_t^\varepsilon - Z_t^\varepsilon = \delta \beta_t^\varepsilon + R(\varepsilon, \delta),$$

where the stochastic process β_t^ε is defined as

$$\beta_t^\varepsilon := \sqrt{2\sigma} \int_0^t k(t-s)(1 + \Phi'(Y_s^\varepsilon)) \, dW_s, \tag{B.3}$$

where Φ is the solution of the cell problem (2.4), W_s is the Brownian motion appearing in (2.1) and $Y_t^\varepsilon = X_t^\varepsilon/\varepsilon$. Moreover, the remainder $R(\varepsilon, \delta)$ satisfies for every $p > 0$ the estimate

$$(\mathbb{E}^{\rho^\varepsilon} |R(\varepsilon, \delta)|^p)^{1/p} \leq C \left(\delta + \varepsilon^2 + \varepsilon \delta^{1/2} \right), \tag{B.4}$$

where C is independent of ε and δ .

Proof. Recall that by definition of Z_t^ε we have

$$X_t^\varepsilon - Z_t^\varepsilon = \int_0^t k(t-s)(X_t^\varepsilon - X_s^\varepsilon) \, ds + e^{-t/\delta} X_t^\varepsilon.$$

Plugging the decomposition (B.2) into the equation above, we obtain

$$X_t^\varepsilon - Z_t^\varepsilon = I_1^\varepsilon(t) + I_2^\varepsilon(t) + I_3^\varepsilon(t) + I_4^\varepsilon(t),$$

where

$$\begin{aligned}
I_1^\varepsilon(t) &:= - \int_0^t k(t-s) \int_s^t (\alpha \cdot V'(X_r^\varepsilon))(1 + \Phi'(Y_r^\varepsilon)) \, dr \, ds, \\
I_2^\varepsilon(t) &:= \sqrt{2\sigma} \int_0^t k(t-s) \int_s^t (1 + \Phi'(Y_r^\varepsilon)) \, dW_r \, ds, \\
I_3^\varepsilon(t) &:= -\varepsilon \int_0^t k(t-s) (\Phi(Y_t^\varepsilon) - \Phi(Y_s^\varepsilon)) \, ds, \\
I_4^\varepsilon(t) &= e^{-t/\delta} X_t.
\end{aligned}$$

Let us analyze the terms above singularly. For $I_1^\varepsilon(t)$, one can show [25, Proposition 5.8]

$$\int_s^t (\alpha \cdot V'(X_r^\varepsilon))(1 + \Phi'(Y_r^\varepsilon)) \, dr = (t-s)(A \cdot V'(X_t^\varepsilon)) + R_1^\varepsilon(t-s),$$

in law, where the remainder R_1^ε satisfies

$$\left(\mathbb{E}^{\varphi^\varepsilon} |R_1^\varepsilon(t-s)|^p \right)^{1/p} \leq C(\varepsilon^2 + \varepsilon(t-s)^{1/2} + (t-s)^{3/2}). \quad (\text{B.5})$$

Therefore, it holds

$$\begin{aligned}
I_1^\varepsilon(t) &= -(A \cdot V'(X_t^\varepsilon)) \int_0^t k(t-s)(t-s) \, ds + \int_0^t k(t-s) R_1^\varepsilon(t-s) \, ds \\
&= -\delta(A \cdot V'(X_t^\varepsilon)) + e^{-t/\delta}(t+\delta)(A \cdot V'(X_t^\varepsilon)) + \tilde{R}_1^\varepsilon(t),
\end{aligned}$$

where we exploited the equality

$$\int_0^t k(t-s)(t-s) \, ds = \delta - e^{-t/\delta}(t+\delta),$$

and where

$$\tilde{R}_1^\varepsilon(t) := \int_0^t k(t-s) R_1^\varepsilon(t-s) \, ds.$$

Now, Lemma B.1, the inequality (B.5) and Lemma B.2 yield for all $p \geq 1$

$$\begin{aligned}
\mathbb{E}^{\varphi^\varepsilon} \left| \tilde{R}_1^\varepsilon(t) \right|^p &\leq C \int_0^t k(t-s) \mathbb{E}^{\varphi^\varepsilon} |R_1^\varepsilon(t-s)|^p \, ds \\
&\leq C \int_0^t k(t-s) (\varepsilon^{2p} + \varepsilon^p(t-s)^{p/2} + (t-s)^{3p/2}) \, ds \\
&\leq C \left(\varepsilon^{2p} + \varepsilon^p \delta^{p/2} + \delta^{3p/2} \right),
\end{aligned}$$

where C is a positive constant independent of ε and δ . Therefore, for t sufficiently big and δ sufficiently small, we get

$$\left(\mathbb{E}^{\varphi^\varepsilon} |I_1^\varepsilon(t)|^p \right)^{1/p} \leq C \left(\delta + \varepsilon^2 + \varepsilon \delta^{1/2} \right).$$

We now consider the second term. Let us introduce the notation

$$Q_t := \int_0^t (1 + \Phi'(Y_r^\varepsilon)) \, dW_r,$$

and therefore rewrite

$$I_2^\varepsilon(t) = \sqrt{2\sigma} Q_t \int_0^t k(t-s) (Q_t - Q_s) \, ds.$$

An application of the Itô formula to $u(s, Q_s)$ where $u(s, x) = k(t-s)x$ yields

$$\begin{aligned}
I_2^\varepsilon(t) &= \sqrt{2\sigma} \left(Q_t \int_0^t k(t-s) \, ds - Q_t + \delta \int_0^t k(t-s) (1 + \Phi'(Y_s^\varepsilon)) \, dW_s \right) \\
&= \delta \beta_t^\varepsilon - \sqrt{2\sigma} e^{-t/\delta} Q_t =: \delta \beta_t^\varepsilon - R_2^\varepsilon(t).
\end{aligned}$$

where β_t^ε is defined in (B.3). For the remainder $R_2^\varepsilon(t)$, let us remark that for all $p \geq 1$ it holds

$$(\mathbb{E}|Q_t|^p)^2 \leq \mathbb{E}|Q_t|^{2p} \leq Ct^{p-1} \int_0^t \mathbb{E}|1 + \Phi'(Y_r^\varepsilon)|^{2p} dr \leq Ct$$

where we applied Jensen's inequality, an estimate for the moments of stochastic integrals [16, Formula (3.25), p. 163] and the boundedness of Φ . Therefore, for t large enough, we have

$$\mathbb{E}|R_2^\varepsilon(t)|^p \leq C\sqrt{t}e^{-pt/\delta}.$$

Let us now consider $I_3^\varepsilon(t)$. Since Φ is bounded, we simply have

$$|I_3^\varepsilon(t)| \leq C\varepsilon,$$

almost surely. Finally, due to [25, Corollary 5.4], we know that X_t^ε has bounded moments of all orders and therefore

$$\left(\mathbb{E}^{\varphi^\varepsilon}|I_4^\varepsilon(t)|^p\right)^{1/p} \leq Ce^{-t/\delta},$$

which concludes the proof. \square

Remark B.6. Let us remark that one can show for all $p \geq 1$ and a constant $C > 0$ independent of δ

$$(\mathbb{E}|\beta_t^\varepsilon|^p)^{1/p} \leq C\delta^{-1/2}.$$

Therefore, the estimate presented in Lemma B.5 is consistent with the rough estimate of Lemma B.3, and the term highlighted in the statement of Lemma B.5 is indeed the dominating term in terms of powers of δ .

Appendix C Proof of Theorem 3.10

Proof of Theorem 3.10. Let us introduce the notation

$$\mathcal{A}^\varepsilon(\delta) = \frac{1}{\delta} \widetilde{\mathcal{M}}_\varepsilon^{-1} \mathbb{E}^{\rho^\varepsilon}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)],$$

where

$$\widetilde{\mathcal{M}}_\varepsilon = \mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)].$$

Then following the proof of Theorem 3.9 we only need to show that if $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$ we have

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = A, \quad \text{in law.}$$

Using Lemma B.5 and geometric ergodicity (Lemma 3.3), we have the following equality in law

$$\begin{aligned} \mathcal{A}^\varepsilon(\delta) &= \widetilde{\mathcal{M}}_\varepsilon^{-1} \frac{1}{\delta} \lim_{t \rightarrow \infty} \mathbb{E}[(X_t^\varepsilon - Z_t^\varepsilon)^2 V''(Z_t^\varepsilon)] \\ &= \widetilde{\mathcal{M}}_\varepsilon^{-1} \frac{1}{\delta} \lim_{t \rightarrow \infty} \mathbb{E} \left[\left(\sqrt{2\Sigma} \delta \int_0^t k(t-s) dB_s + R(\varepsilon, \delta) \right)^2 V''(Z_t^\varepsilon) \right] \\ &=: \widetilde{\mathcal{M}}_\varepsilon^{-1} \lim_{t \rightarrow \infty} (J_1^\varepsilon(t) + J_2^\varepsilon(t) + J_3^\varepsilon(t)), \end{aligned}$$

where

$$\begin{aligned} J_1^\varepsilon(t) &= 2\Sigma\delta \mathbb{E} \left[\left(\int_0^t k(t-s) dB_s \right)^2 V''(Z_t^\varepsilon) \right], \\ J_2^\varepsilon(t) &= 2\sqrt{2\Sigma} \mathbb{E} \left[\left(\int_0^t k(t-s) dB_s \right) R(\varepsilon, \delta) V''(Z_t^\varepsilon) \right], \\ J_3^\varepsilon(t) &= \frac{1}{\delta} \mathbb{E} [R(\varepsilon, \delta)^2 V''(Z_t^\varepsilon)]. \end{aligned}$$

Let us consider the three terms separately. First, since the Brownian motion B_s is independent of the process Z_t^ε , applying Itô isometry we get

$$J_1^\varepsilon(t) = 2\Sigma\delta \left(\int_0^t k(t-s)^2 ds \right) \mathbb{E}[V''(Z_t^\varepsilon)] = \Sigma \mathbb{E}[V''(Z_t^\varepsilon)],$$

and, by Lemma B.3 and weak convergence of the invariant measure, we obtain

$$\lim_{\varepsilon \rightarrow 0} \lim_{t \rightarrow \infty} J_1^\varepsilon(t) = \Sigma \lim_{\varepsilon \rightarrow 0} \mathbb{E}^{\rho^\varepsilon}[V''(Z_t^\varepsilon)] = \Sigma \mathbb{E}^{\rho^0}[V''(X)].$$

Moreover, due to (B.4), we have the following bounds for $J_2^\varepsilon(t)$ and $J_3^\varepsilon(t)$

$$\begin{aligned} |J_2^\varepsilon(t)| &\leq C \left(\delta^{1/2} + \varepsilon^\iota \delta^{-1/2} + \varepsilon \right), \\ |J_3^\varepsilon(t)| &\leq C \left(\delta + \varepsilon^{2\iota} \delta^{-1} + \varepsilon^2 \right). \end{aligned}$$

Furthermore, by Lemma B.4 and weak convergence of the invariant measure, we have

$$\lim_{\varepsilon \rightarrow 0} \widetilde{\mathcal{M}}_\varepsilon^{-1} = \mathbb{E}^{\rho^0}[V'(X) \otimes V'(X)] =: \mathcal{M}_0.$$

Therefore

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = \Sigma \mathbb{E}^{\rho^0}[V'(X) \otimes V'(X)]^{-1} \mathbb{E}^{\rho^0}[V''(X)],$$

and, finally, an integration by parts yields

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = \Sigma \mathbb{E}^{\rho^0}[V'(X) \otimes V'(X)]^{-1} \frac{1}{\Sigma} \mathbb{E}^{\rho^0}[V'(X) \otimes V'(X)] A = A,$$

which implies the desired result. \square

Appendix D Proof of Theorem 4.3

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

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

where the matrix M is taken from (2.7) and the vector \widetilde{h} is taken from (3.8). We prove here Theorem 4.3, i.e., the unbiasedness of this estimator. Let us consider a couple $(X^\varepsilon, Z^\varepsilon)^\top$.

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

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

almost surely. Therefore, due to Assumption 2.1 which allows controlling the norm of $\mathcal{M}_\varepsilon^{-1}$ and due to Lemma B.4 we have for a constant $C > 0$

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

Let us remark that $\widehat{A}_k^\varepsilon(T)$ has a bounded norm for ε sufficiently small due to Theorem 3.9. 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 (D.1) 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] R. AZENCOTT, A. BERI, A. JAIN, AND I. TIMOFEYEV, *Sub-sampling and parametric estimation for multiscale dynamics*, Commun. Math. Sci., 11 (2013), pp. 939–970.
- [7] R. AZENCOTT, A. BERI, AND I. TIMOFEYEV, *Adaptive sub-sampling for parametric estimation of Gaussian diffusions*, J. Stat. Phys., 139 (2010), pp. 1066–1089.
- [8] A. BENSOUSSAN, J.-L. LIONS, AND G. PAPANICOLAOU, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam, 1978.
- [9] D. CALVETTI, M. DUNLOP, E. SOMERSALO, AND A. STUART, *Iterative updating of model error for Bayesian inversion*, Inverse Problems, 34 (2018), pp. 025008, 38.
- [10] D. CALVETTI, O. ERNST, AND E. SOMERSALO, *Dynamic updating of numerical model discrepancy using sequential sampling*, Inverse Problems, 30 (2014), pp. 114019, 19.
- [11] C. J. COTTER AND G. A. PAVLIOTIS, *Estimating eddy diffusivities from noisy Lagrangian observations*, Commun. Math. Sci., 7 (2009), pp. 805–838.
- [12] M. DASHTI AND A. M. STUART, *The Bayesian Approach to Inverse Problems*, in Handbook of Uncertainty Quantification, Springer, 2016, pp. 1–118.
- [13] S. GAILUS AND K. SPILIOPOULOS, *Statistical inference for perturbed multiscale dynamical systems*, Stochastic Process. Appl., 127 (2017), pp. 419–448.
- [14] ———, *Discrete-time statistical inference for multiscale diffusions*, Multiscale Model. Simul., 16 (2018), pp. 1824–1858.
- [15] 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.
- [16] I. KARATZAS AND S. E. SHREVE, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.
- [17] S. KRUMSCHEID, G. A. PAVLIOTIS, AND S. KALLIADASIS, *Semiparametric drift and diffusion estimation for multiscale diffusions*, Multiscale Model. Simul., 11 (2013), pp. 442–473.
- [18] 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.
- [19] 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.

- [20] 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.
- [21] 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.
- [22] A. PAPAVALIOU, G. A. PAVLIOTIS, AND A. M. STUART, *Maximum likelihood drift estimation for multiscale diffusions*, Stochastic Process. Appl., 119 (2009), pp. 3173–3210.
- [23] 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.
- [24] 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.
- [25] G. A. PAVLIOTIS AND A. M. STUART, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
- [26] 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.
- [27] Y. POKERN, A. M. STUART, AND E. VANDEN-EIJNDEN, *Remarks on drift estimation for diffusion processes*, Multiscale Model. Simul., 8 (2009), pp. 69–95.
- [28] A. M. STUART, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
- [29] Y. YING, J. MADDISON, AND J. VANNESTE, *Bayesian inference of ocean diffusivity from Lagrangian trajectory data*, Ocean Model., 140 (2019).
- [30] 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.