

Drift Estimation of Multiscale Diffusion via Filtering

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

Abstract. Copy-paste of SIAM UQ abstract, modify. We present a novel technique for estimating the drift function of a diffusion process possessing two separated time scales. Our aim is fitting a homogenized diffusion model to a continuous sample path coming from the full multiscale process, thus dealing with an issue of model misspecification. We consider a Bayesian framework and study the asymptotic limit of posterior distributions over the drift function. In this setting, we show on the one hand that if the continuous multiscale data are not pre-processed, then the posterior distribution concentrates asymptotically on the wrong value of the drift function. On the other hand, we show that data can be treated ahead of the inference procedure in order to obtain the desired posterior. In particular, we prove that there exists a family of transformations which are linear on the space of continuous sample paths and which, when applied to multiscale data, allow the posterior distribution to be asymptotically correct. We present a series of numerical examples on test cases which corroborate our theoretical findings.

AMS subject classifications.

Keywords.

1 Introduction

Data-driven approaches to the derivation of effective stochastic coarse-grained dynamics has become a standard tool in different areas such as molecular dynamics (CITE) or atmosphere and ocean science [9, 13, 26]. When fitting data to a low dimensional coarse-grained model, one is faced with the problem of model misspecification; indeed, the data, coming from the full dynamics, is compatible with the coarse-grained model only at the time scales at which the effective dynamics is valid. This is related to the issue of high frequency data that arises in econometrics [4, 5, 18, 27].

The issue of model misspecification in inverse problems with a multiscale structure has been treated in the context of partial differential equations (PDE). 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 homogenisation and apply statistical techniques taken from [7, 8].

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 homogenised 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

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

[†]Department of Mathematics, Imperial College London

[‡]Department of Computing and Mathematical Sciences, Caltech

the coefficients of the unhomogenised equation, rather than to those of the homogenised one. The bias of the MLE estimator 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]. This is similar to the estimation of the integrated stochastic volatility in the presence of market micro-structure noise, where the data has to be subsampled at an appropriate rate [5, 27]. 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 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].

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, changing the filter parameters does not have a strong influence on the result, therefore allowing our technique to 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.

It is natural to reinterpret the problem of inferring the drift parameter of multiscale diffusion processes under a Bayesian perspective [23, 24]. In particular, the form of the likelihood function guarantees, under a Gaussian prior hypothesis, that the posterior distribution is itself a Gaussian. Therefore, one can simply compute the mean and covariance of the posterior distribution to obtain a full quantification of the inference process. The Bayesian approach has been adopted in several works concerning single and multiscale PDEs, too, and has proved itself to be versatile and efficient in this framework (see, e.g., [1, 3]). Moreover, infinite-dimensional inverse problems are intrinsically ill-posed, and interpreting them in a Bayesian manner is an efficient tool for regularizing them and guaranteeing their well-posedness [10, 25]. The regularization property of a Bayesian approach has been recently demonstrated for inverse problems involving multiscale PDE [1]. In this work, we investigate the effects of inserting our filtering methodology in a Bayesian framework. In particular, we analyse the effects of modifying the likelihood function inserting the filtered trajectory, showing how this corrects the faulty asymptotic behaviour which would be present in case no pre-processing would be applied.

1.1 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, to the properties of our estimator. 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

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 $\alpha_i \in \mathbb{R}$, $i = 1, \dots, N$ 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 $p: \mathbb{R} \rightarrow \mathbb{R}$ are slow and fast potentials driving the dynamics of the solution X_t^ε . Given $N \in \mathbb{N}_{>0}$, we assume the slow potential to be of the form

$$V(x) = \begin{pmatrix} V_1(x) & V_2(x) & \cdots & V_N(x) \end{pmatrix}^\top, \quad (2.2)$$

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] guarantees the existence of an SDE of the form

$$dX_t^0 = -A \cdot V'(X_t^0) dt + \sqrt{2\Sigma} dW_t, \quad (2.3)$$

where W_t is the same Brownian motion and where the fast dynamics have been eliminated, such that $X_t^\varepsilon \rightarrow X_t^0$ for $\varepsilon \rightarrow 0$ in law as random variables in $\mathcal{C}^0((0, T), \mathbb{R})$. In particular, we have $A_i = K\alpha_i$ and $\Sigma = K\sigma$, where the coefficient K is given by the formula

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

with

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

and where the function Φ is the solution 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 $T > 0$ 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. In order to estimate the drift coefficient A of the homogenized equation we maximize the likelihood function

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

where Girsanov formula allows to write explicitly the log-likelihood as

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

Therefore, the maximum likelihood estimator can be formally computed in closed form as

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

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

$$M = \frac{1}{2\Sigma T} \int_0^T V'(X_t) \otimes V'(X_t) dt, \quad h = \frac{1}{2\Sigma T} \int_0^T V'(X_t) dX_t,$$

and where \otimes denotes the tensor 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 dissipative setting of [22].

Assumption 2.1. The potentials p and V satisfy

$$(i) \quad p \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}) \cap L^\infty(\mathbb{R}, \mathbb{R}),$$

(ii) the potential $V \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ and there exist $a, b > 0$ such that

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

Moreover, for all $T > 0$, the symmetric matrix M is positive definite and its minimum eigenvalue satisfies $\lambda_{\min}(M) \geq \bar{\lambda} > 0$.

Under these assumptions, the estimator $\hat{A}^\varepsilon(T)$ is indeed well-posed. Moreover, these assumptions guarantee that the process X^ε is geometrically ergodic, and

3 The filtering approach

In the previous section, we have shown that posterior distributions over the drift coefficient of the homogenized equation are asymptotically incorrect if multiscale data are replaced into the expression of the likelihood. Hence, the need of pre-processing the data is highlighted. In particular, we consider a filtering approach. Let $k: \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}$ be a kernel function and consider the process $Z^\varepsilon := (Z_t^{\varepsilon, k}, 0 \leq t \leq T)$ defined by the weighted average

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

Let $\beta, \delta > 0$ and let us consider a family of exponential kernel functions defined as

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

where C_β is a normalizing constant given by

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

and where $\Gamma(\cdot)$ is the gamma function. 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.

Given a trajectory $X^\varepsilon := (X_t^\varepsilon, 0 \leq t \leq T)$, it is relatively inexpensive to compute Z^ε from a computational standpoint. In particular, the process Z^ε is the 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.

3.1 Ergodic properties of the filter

In the following, we consider the filter with $\beta = 1$, i.e.,

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

since in this case it holds

$$dZ_t^\varepsilon = k(t, t) X_t^\varepsilon dt + \int_0^t \partial_t k(t, s) X_s^\varepsilon ds dt = \frac{1}{\delta} (X_t^\varepsilon - Z_t^\varepsilon) dt.$$

Considering the processes X^ε and Z^ε together, we obtain a system of stochastic differential equations

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

The following result guarantees that the measure induced by the joint process $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ admits a smooth density.

Lemma 3.1. *Let $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ be the solution of (3.2). Then the measure μ^ε of the process admits a smooth density ρ^ε with respect to the Lebesgue measure.*

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

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

Then, the generator of the process $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ 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

Having established that the law of the joint process admits a smooth density with respect to the Lebesgue measure, we now show in the following Lemma a property of ergodicity.

Lemma 3.2. *Let Assumption 2.1 hold and let $b > 0$ be given in Assumption 2.1(ii). Then, if $\delta > 1/(4b)$, the process $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ solution of (3.2) is geometrically ergodic, i.e., there exists $C, \lambda > 0$ such that for all measurable f such that for some integer $p > 0$*

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

it holds

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

where ρ^ε is the solution to the stationary Fokker–Planck equation

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

Proof. Lemma 3.1 guarantees that the Fokker–Planck equation can be written directly from the system (3.2). For geometric ergodicity, let

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

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 $\gamma = \gamma^* := 1 - b\delta + \sqrt{2 - 2b\delta + b^2\delta^2} > 0$ 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

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

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

Remark 3.3. Let us remark that for the multiscale equation (2.1) one should check the condition above for the drift function $V(x) + \varepsilon^{-1}p(x/\varepsilon)$. Nevertheless, being p bounded, the result holds trivially.

Example 3.4. 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 $X_t \sim \mathcal{GP}(\mu_t, \mathcal{C}(t, s))$, where

$$\mu_t = X_0 e^{-At}, \quad \mathcal{C}(t, s) = \frac{\Sigma}{A} \left(e^{-A|t-s|} - e^{-A(t+s)} \right).$$

The basic properties of Gaussian processes imply that Z_t is a Gaussian process, and that the couple $(X_t, Z_t)^\top$ 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.

Lemma 3.5. *Let ρ^ε be the solution of (3.3) and let us write*

$$\rho^\varepsilon(x, z) = \varphi^\varepsilon(x) \psi^\varepsilon(z) R^\varepsilon(x, z), \tag{3.4}$$

where φ^ε and ψ^ε are the marginal densities, 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

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

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

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

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

$$\sigma(\varphi^\varepsilon)''(x) + \left(\alpha \cdot V''(x) + \frac{1}{\varepsilon^2} p''\left(\frac{x}{\varepsilon}\right) \right) \varphi^\varepsilon(x) + \left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p'\left(\frac{x}{\varepsilon}\right) \right) (\varphi^\varepsilon)'(x) = 0, \quad (3.6)$$

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 ψ^ε 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

$$z = \int_{\mathbb{R}} x \varphi^\varepsilon(x) R^\varepsilon(x, z) dx. \quad (3.7)$$

Replacing the decomposition (3.4) in equation (3.3) and using equation (3.6) we obtain the following equality

$$\sigma \varphi^\varepsilon \psi^\varepsilon \partial_{xx}^2 R^\varepsilon + \frac{1}{\delta} \varphi^\varepsilon \psi^\varepsilon R^\varepsilon + \sigma (\varphi^\varepsilon)' \psi^\varepsilon \partial_x R^\varepsilon + \frac{1}{\delta} (z - x) \varphi^\varepsilon ((\psi^\varepsilon)' R + \psi^\varepsilon \partial_z R^\varepsilon) = 0. \quad (3.8)$$

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

$$\sigma \psi^\varepsilon \int_{\mathbb{R}} x \varphi^\varepsilon \partial_{xx}^2 R^\varepsilon dx + \sigma \psi^\varepsilon \int_{\mathbb{R}} x (\varphi^\varepsilon)' \partial_x R^\varepsilon dx = -\sigma \psi^\varepsilon \int_{\mathbb{R}} \varphi^\varepsilon \partial_x R^\varepsilon dx. \quad (3.9)$$

Then, applying (3.7), we have

$$\frac{1}{\delta} \psi^\varepsilon \int_{\mathbb{R}} x \varphi^\varepsilon R^\varepsilon dx = \frac{1}{\delta} z \psi^\varepsilon. \quad (3.10)$$

Moreover, again applying (3.7), we can compute

$$\frac{1}{\delta} (\psi^\varepsilon)' \int_{\mathbb{R}} (z - x) x \varphi^\varepsilon R^\varepsilon dx = \frac{1}{\delta} (\psi^\varepsilon)' \left(z^2 - \int_{\mathbb{R}} x^2 \varphi^\varepsilon R^\varepsilon dx \right). \quad (3.11)$$

Finally, we compute the last term always applying (3.7) obtaining

$$\frac{1}{\delta} \psi^\varepsilon \int_{\mathbb{R}} (z - x) x \varphi^\varepsilon \partial_z R^\varepsilon dx = \frac{1}{\delta} \psi^\varepsilon \left(z - \int_{\mathbb{R}} x^2 \varphi^\varepsilon \partial_z R^\varepsilon dx \right). \quad (3.12)$$

Replacing the equalities (3.9), (3.10), (3.11) and (3.12) into (3.8) 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

Remark 3.6. It could be possible to find a closed-form solution to (3.3) in a more general framework than Example 3.4. Nonetheless, an explicit solution is not necessary for our argument.

3.2 Multiscale convergence

We now investigate the convergence of the couple $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ with respect to the multiscale parameter $\varepsilon \rightarrow 0$. **Rewrite the following two lemmas with considerations from discussion with Greg: maybe one lemma is enough and clearer.**

Lemma 3.7. *Let $T > 0$. Then, the couple $(X^\varepsilon, Z^\varepsilon)^\top := ((X_t^\varepsilon, Z_t^\varepsilon)^\top, 0 \leq t \leq T)$ converges in law for $\varepsilon \rightarrow 0$ in $\mathcal{C}^0((0, T), \mathbb{R}^2)$ to the couple $(X, Z)^\top := ((X_t, Z_t)^\top, 0 \leq t \leq T)$ solution to*

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

where $A \in \mathbb{R}^N$ and $\Sigma > 0$ are the coefficients of (2.3).

Proof. Standard homogenization theory (see e.g. [6, Chapter 3]) can be employed, and therefore the proof is omitted here. \square

It is known that the invariant distribution of the process X^ε converges weakly to the invariant distribution of the process X^0 solution to (2.3). We now consider the random variable $(X^\varepsilon, Z^\varepsilon)^\top$ and present an analogous convergence result for the joint process.

Lemma 3.8. *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*

$$\Sigma \partial_{xx}^2 \rho^0(x, z) + \left(A \cdot V''(x) + \frac{1}{\delta} \right) \rho^0(x, z) + A \cdot V'(x) \partial_x \rho^0(x, z) + \frac{1}{\delta} (z - x) \partial_z \rho^0(x, z) = 0, \quad (3.13)$$

where A and Σ are the coefficients of the homogenized equation (2.3).

Proof. Lemma 3.7, together with the assumption that the joint process has an initial condition distributed as the invariant measures, guarantees via the Portmanteau theorem the weak convergence of μ^ε to μ^0 . \square

We conclude this section presenting an analogous result to Lemma 3.5 for the limit distribution.

Corollary 3.9. *Let ρ^0 be the solution of (3.13) and let us write*

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

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

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

Proof. The proof is directly obtained from Lemma 3.5 replacing $p(y) = 0$ and α, σ by A, Σ respectively. \square

3.3 The modified estimator

The estimator of the homogenized drift coefficient A we analyse is given by the formula

$$\widehat{A}^\varepsilon(T) = -\widetilde{M}^{-1} \widetilde{h}, \quad (3.14)$$

where

$$\widetilde{M} = \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon) dt \quad \text{and} \quad \widetilde{h} = \frac{1}{T} \int_0^T V'(Z_t^\varepsilon) dX_t^\varepsilon. \quad (3.15)$$

In practice, one evaluation of $V'(\cdot)$ is computed with the filter, while the differential and the second evaluation of $V'(\cdot)$ are maintained with the original process X^ε .

Remark 3.10. Let us remark that in the one-dimensional case $N = 1$, the estimator can be written as

$$\widehat{A}^\varepsilon(T) = -\frac{\int_0^T V'(Z_t^\varepsilon) dX_t^\varepsilon}{\int_0^T V'(Z_t^\varepsilon) V'(X_t^\varepsilon) dt}.$$

In the one-dimensional case, it is clear that the maximum likelihood estimator based on X^ε alone, i.e., the quantity

$$\widehat{A}^\varepsilon(T) = -\frac{\int_0^T V'(X_t^\varepsilon) dX_t^\varepsilon}{\int_0^T V'(X_t^\varepsilon)^2 dt}.$$

is well defined for $V(x) \neq 0$. Equivalently, for a N -dimensional parameter, a sufficient assumption for the well-posedness of the estimator is to require the symmetric matrix M defined as

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

to be positive definite. In the following lemma, we ensure that the estimator based on the filtering method is well-defined.

Lemma 3.11. *The matrix \widetilde{M} defined in (3.15) is invertible almost surely.*

Proof. **TO DO** \square

Theorem 3.12. *Let $\widehat{A}^\varepsilon(T)$ be defined in (3.14). Then*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widehat{A}^\varepsilon(T) = A, \quad \text{a.s.}, \quad (3.16)$$

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

Proof. Replacing the expression of dX_t^ε into (3.14), 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

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

We study the terms I_1^ε and I_2^ε separately. The ergodic theorem applied to I_1^ε yields

$$\lim_{T \rightarrow \infty} I_1^\varepsilon(T) = \mathbb{E}^{\rho^\varepsilon} [V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1} \mathbb{E}^{\rho^\varepsilon} \left[\frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon}{\varepsilon} \right) V'(Z^\varepsilon) \right], \quad \text{a.s.} \quad (3.17)$$

Due to Lemma 3.5 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

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

Replacing the equality above into (3.17), 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 3.5, 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 ε goes to zero and

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

where the function ρ^0 is the density of the limit invariant distribution for $\varepsilon \rightarrow 0$, as in Lemma 3.8.

Due to Lemma 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

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

We can therefore conclude that

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} I_1^\varepsilon(T) = -\alpha + A, \quad \text{a.s.} \quad (3.18)$$

We now consider the second term I_2^ε , and rewrite it as

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

where

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

The ergodic theorem yields

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

where M^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 (3.18), proves (3.16). \square

4 The Bayesian setting

In a Bayesian setting, the goal is to determine the posterior distribution $\mu(A \mid X)$ given a continuous trajectory $X := (X_t, 0 \leq t \leq T)$. Choosing 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, the posterior distribution admits a density $p(A \mid X)$ with respect to the Lebesgue measure which satisfies

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

where Z is the normalization constant, p_0 is the density of μ_0 , and where Girsanov's formula allows to write the likelihood as

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

with

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

The log-posterior density is then given by

$$\begin{aligned} \log p(A \mid X) &= -\log Z - \frac{1}{2\Sigma} \int_0^T A \cdot V'(X_t) dX_t - \frac{1}{4\Sigma} \int_0^T (A \cdot V'(X_t))^2 dt \\ &\quad - \frac{1}{2} \left\| C_0^{-1/2} (A - A_0) \right\|_2^2. \end{aligned}$$

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 and

C_T the mean and covariance matrix, respectively. Let us consider the matrix-valued function $M: \mathbb{R} \rightarrow \mathbb{R}^{N \times N}$ whose entries are given by

$$M_{ij} = \frac{1}{2\Sigma T} \int_0^T V_i'(X_t) V_j'(X_t) dt, \quad i, j = 1, \dots, N,$$

and the vector-valued function $h: \mathbb{R} \rightarrow \mathbb{R}^N$ defined by

$$h_i = \frac{1}{2\Sigma T} \int_0^T V_i'(X_t) dX_t, \quad i = 1, \dots, N.$$

Let us remark that we can rewrite M and h in vector notation as

$$M = \frac{1}{2\Sigma T} \int_0^T V'(X_t) \otimes V'(X_t) dt, \quad h = \frac{1}{2\Sigma T} \int_0^T V'(X_t) dX_t,$$

where \otimes denotes the tensor product in \mathbb{R}^N . Therefore, we can rewrite the log-posterior density as

$$\log p(A | X) = -\log Z - TA \cdot h - \frac{T}{2} A \cdot MA - \frac{1}{2} (A - A_0) \cdot C_0^{-1} (A - A_0).$$

Completing the squares in the log-posterior density, it is possible to show that the posterior is a Gaussian $\mu(A | X) = \mathcal{N}(m_T, C_T)$, whose precision matrix and mean are formally given by

$$\begin{aligned} C_T^{-1} &= C_0^{-1} + TM, \\ C_T^{-1} m_T &= C_0^{-1} A_0 - Th. \end{aligned}$$

We are now interested in the limit of the posterior distribution for $T \rightarrow \infty$. Let us first define the maximum likelihood estimator (MLE) $\hat{A}(X)$ of A , which is obtained by maximizing the log-likelihood function $\log p(X | A)$, and which is hence formally given by

$$\hat{A}(X) = -M^{-1}h.$$

We now introduce regularity assumptions on the SDE.

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

Proposition 4.1. *Given a continuous stochastic process $X := (X_t, 0 \leq t \leq T)$ which is ergodic with invariant measure μ^∞ such that $\mathbb{E}^{\mu^\infty}(V_i'(\cdot)) \leq C$ for some constant $C > 0$ and for all $i = 1, \dots, N$, the posterior $\mu(A | X)$ contracts to the limit of $\hat{A}(X)$ for $T \rightarrow \infty$.*

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

$$C_T = T^{-1} (M^{-1} - Q^{-1}),$$

where

$$Q = M + TMC_0M.$$

The eigenvalues of Q satisfy for all $i = 1, \dots, N$

$$\lambda_i(Q) = \lambda_i(M) + T\lambda_i(MC_0M) \geq \bar{\lambda} + T\lambda_{\min}(C_0),$$

and therefore we obtain

$$\lambda_{\max}(Q^{-1}) \leq \frac{1}{\bar{\lambda} + T\lambda_{\min}(C_0)}. \quad (4.1)$$

Similarly, we have a bound for the maximum eigenvalue of M^{-1} given by

$$\lambda_{\max}(M^{-1}) \leq \frac{1}{\bar{\lambda}}, \quad (4.2)$$

which implies that

$$\text{tr } C_T \leq \frac{N}{T} \left(\frac{1}{\bar{\lambda}} + \frac{1}{\bar{\lambda} + T\lambda_{\min}(C_0)} \right),$$

and which therefore allows us to conclude that

$$\lim_{T \rightarrow \infty} \|C_T\|_F = 0.$$

We now consider the mean. Replacing the expression of the maximum likelihood estimator, we obtain

$$\|m_T - \hat{A}(X)\|_2 = T^{-1} \|M^{-1}C_0^{-1}A_0 - Q^{-1}(C_0^{-1}A_0 - Th)\|_2.$$

Let us remark that (4.1) implies $\|Q^{-1}\|_F \rightarrow 0$ for $T \rightarrow \infty$ and (4.2) implies $\|M^{-1}\|_F \leq C$ for some $C > 0$ independently of T . Moreover, the ergodic theorem guarantees that $\|h\|_2 \leq C$ for T sufficiently big, and therefore the Cauchy–Schwarz and the triangle inequalities imply

$$\lim_{T \rightarrow \infty} \|m_T - \hat{A}(X)\|_2 = 0,$$

which proves the desired result. \square

Remark 4.2. Proposition 4.1 guarantees that in the asymptotic limit of $T \rightarrow \infty$, it is equivalent to consider the Bayesian approach and the maximum likelihood approach, and can therefore be interpreted as a consistency result for both approaches. Nonetheless, in this Gaussian framework the Bayesian approach provides richer information on the inference result with a negligible additional cost.

4.1 The filtering approach

In this section we show how the filtering approach can be employed in a Bayesian framework. As in Section 4 let us consider a Gaussian prior $\mu_0 = \mathcal{N}(A_0, C_0)$. Then, the density $p^\varepsilon(A | X)$ of the posterior distribution $\mu^\varepsilon(A | X^\varepsilon)$ given by multiscale data satisfies

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

where Z^ε is the normalization term and where the likelihood function is given by

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

with

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

Let us consider the likelihood function $\tilde{p}^\varepsilon(X | A)$ modified to include the filtered data Z^ε given by

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

with

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

Moreover, let us consider the modified posterior distribution $\tilde{\mu}^\varepsilon(A | X)$ with density

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

Let us remark that employing (3.15) we can write

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

With respect to Section 4, the matrix \widetilde{M} is not symmetric. Nonetheless, we remark that

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

where \widetilde{M}_S is the symmetric part of \widetilde{M} , i.e.,

$$\widetilde{M}_S = \frac{1}{2} (\widetilde{M} + \widetilde{M}^\top).$$

Hence, we can rewrite

$$\log \widetilde{p}^\varepsilon(A | X) = -\log \widetilde{Z}^\varepsilon - TA \cdot \widetilde{h} - \frac{T}{2} A \cdot \widetilde{M}_S A - \frac{1}{2} (A - A_0) \cdot C_0^{-1} (A - A_0).$$

This shows that the posterior $\widetilde{\mu}^\varepsilon$ is a Gaussian $\mathcal{N}(\widetilde{m}_T, \widetilde{C}_T)$ where the covariance and mean can be obtained as

$$\begin{aligned} \widetilde{C}_T^{-1} &= C_0^{-1} + T\widetilde{M}, \\ \widetilde{C}_T^{-1}\widetilde{m}_T &= C_0^{-1}A_0 - T\widetilde{h}. \end{aligned}$$

In the following theorem we show that the posterior we obtain

Theorem 4.3. *Under*

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 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 3.12. 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 .

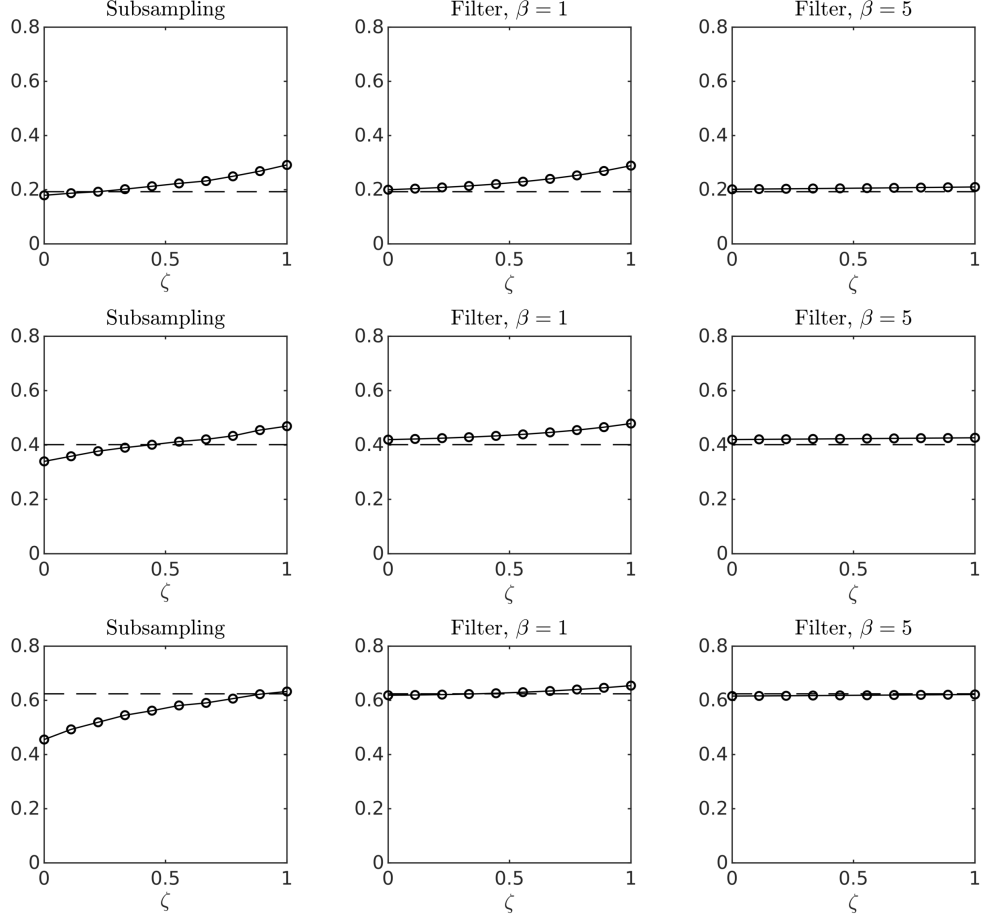


Figure 1: Results for the Ornstein–Uhlenbeck process. The axis of abscissae ζ represents the value of $\delta = \varepsilon^\zeta$ for both subsampling and the filter. The three rows correspond to $\sigma = 0.5, 0.7, 1.0$ from top to bottom.

5.2 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 $\sigma = 0.8$, the fast potential $p(y) = \cos(y)$ and $\varepsilon = 0.05$, thus simulating a trajectory X^ε for time $0 \leq t \leq T$, with $T = 800$. We adopt here a Bayesian approach and compute the posterior distribution $\mu(A \mid X^\varepsilon)$ obtained with the filtering approach. The filter’s parameters are set to $\beta = 1$ and $\delta = \varepsilon^{1/2}$ 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 [22]. We stop computations at times $t = 200, 400, 800$ in order to observe the shrinkage of the Gaussian posterior towards the MLE with respect to time. In Figure 3, 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.

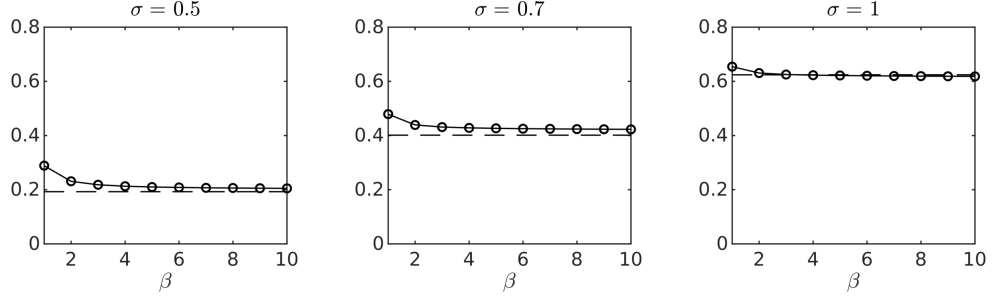


Figure 2: Dependence of the estimator given by the filter for the Ornstein–Uhlenbeck process on the parameter β in (3.1). From left to right we consider different values of σ .

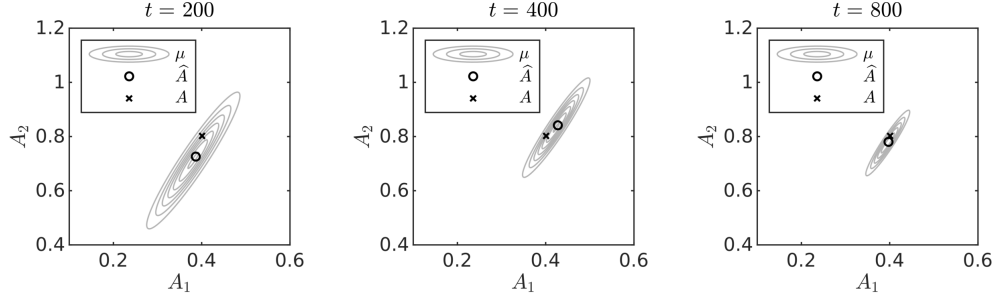


Figure 3: Posterior distributions over the parameter $A = (A_1, A_2)^\top$ for the bistable potential obtained with the filtering approach. The figures refer to time $t = 200, 400, 800$ from left to right, respectively. The MLE \hat{A} is represented with a circle, while the true value A of the drift coefficient of the homogenized equation is represented with a cross.

5.3 Multi-dimensional parameter

We consider $N = 4$ and the potential function $V(x)$ as in (2.2) with

$$V_i(x) = \frac{x^{2i}}{2i}, \quad i = 1, \dots, 4,$$

so that $\alpha \cdot V'(x) = \alpha_1 x + \alpha_2 x^3 + \alpha_3 x^5 + \alpha_4 x^7$, with $\alpha = (-2, -1, 1, 1/2)$. We consider $\varepsilon = 0.05$, 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 compute the MLE of the drift coefficient $A \in \mathbb{R}^4$ of the homogenized equation (2.3) employing

- (i) the data X^ε itself,
- (ii) subsampled data with subsampling parameter $\delta = \varepsilon^{2/3}$,
- (iii) filtered data Z^ε computed with $\beta = 1$ and $\delta = 0.2$.

In particular, we pick this specific value of δ for the subsampling following the optimality criterion given in [22]. Results, given in Figure 4, show that the filter-based estimation captures well the homogenized potential.

References

- [1] A. ABDULLE AND A. DI BLASIO, *A Bayesian numerical homogenization method for elliptic multiscale inverse problems*. Accepted in SIAM-ASA J Uncertain, 2019.
- [2] ———, *Numerical homogenization and model order reduction for multiscale inverse problems*, Multiscale Model. Simul., 17 (2019), pp. 399–433.

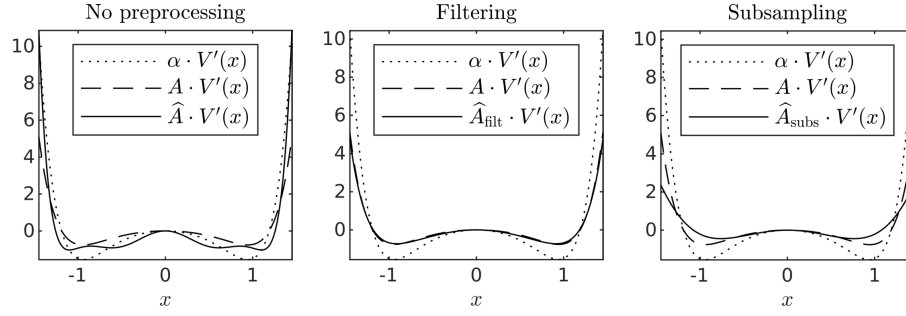


Figure 4: Results for the four-dimensional parameters. From left to right the potential function estimated with (i) the data itself, (ii) the filter, (iii) subsampled data.

- [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. AIT-SAHALIA, P. A. MYKLAND, AND L. ZHANG, *How often to sample a continuous-time process in the presence of market microstructure noise*, Rev. Financ. 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] 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.
- [24] Y. POKERN, A. M. STUART, AND E. VANDEN-EIJNDEN, *Remarks on drift estimation for diffusion processes*, Multiscale Model. Simul., 8 (2009), pp. 69–95.
- [25] A. M. STUART, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
- [26] Y. YING, J. MADDISON, AND J. VANNESTE, *Bayesian inference of ocean diffusivity from Lagrangian trajectory data*, Ocean Model., 140 (2019).
- [27] 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.