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 robust. In particular, we show theoretically and demonstrate via numerical experiments that

(i) the smoothing width of the filter can be alternatively tuned to be proportional to the speed of the slow process or to smaller scales and provide in both cases unbiased results for maximum likelihood parameter estimation. Sharp estimates on the minimal width with respect to the multiscale parameter are provided.

(ii) estimations based on our technique are robust in practice with respect to the parameter of the filter. This is not the case for subsampling, which is strongly influenced by the subsampling frequency.

(iii) 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.

(iv) it is possible to insert the filtering methodology into a continuous-time Bayesian framework by a careful modification of the likelihood function. 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.

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)

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \tag{2.1}$$

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 $V\colon \mathbb{R} \to \mathbb{R}^N$ and $p\colon \mathbb{R} \to \mathbb{R}$ correspond to the slow-scale and the fast-scale confining potentials. In particular, we assume

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

for smooth functions $V_i\colon \mathbb{R} \to \mathbb{R}$, $i = 1, \dots, N$. 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

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t, \tag{2.3}$$

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

$$K = \int_0^L (1 + \Phi'(y))^2\,\mu(\mathrm{d}y), \tag{2.4}$$

with

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

and where the function Φ is the unique solution with zero-mean with respect to the measure μ of the elliptic partial differential equation

$$-p'(y)\Phi'(y) + \sigma\Phi''(y) = p'(y), \quad 0 \le y \le L, \tag{2.5}$$

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

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

We now briefly present the classical methodology for estimating the drift coefficient. Let $T > 0$ and let $X^\varepsilon := (X_t^\varepsilon, 0 \le t \le T)$ 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

$$p(X^\varepsilon \mid A) = \exp\left(-\frac{I(X^\varepsilon \mid A)}{2\Sigma}\right), \tag{2.6}$$

where

$$I(X^\varepsilon \mid A) = \int_0^T A \cdot V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2\,\mathrm{d}t.$$

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

$$\widehat{A}^\varepsilon(T) := \arg\min_{A \in \mathbb{R}^N} I(X^\varepsilon \mid A) = -M^{-1}h, \tag{2.7}$$

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

$$M = \frac{1}{T}\int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t, \quad h = \frac{1}{T}\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon, \tag{2.8}$$

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

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

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

$$\left\|V'(x) - V'(y)\right\|_2 \le C\left|x - y\right|,$$

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

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

$$V^\varepsilon(x) \coloneqq \alpha \cdot V(x) + p\left(\frac{x}{\varepsilon}\right).$$

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 (2.7) is indeed the unique minimizer of the likelihood function, as shown in [27, Theorem 2.4].

Given the convergence of $X_t^\varepsilon \to X_t$ in the space of continuous stochastic processes, one would expect that the MLE (2.7) 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 $T \to \infty$ and $\varepsilon \to 0$, 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*

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

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

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

which is a discretized version of $\widehat{A}^\varepsilon(T)$. 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 $\widehat{A}_\delta^\varepsilon(T)$ is an asymptotically unbiased estimator of A with respect to $T \to \infty$ and $\varepsilon \to 0$, 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

$$\mathrm{d}X_t^\varepsilon = -\sum_{i=1}^N \alpha_i \nabla V_i(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}\nabla p\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t,$$

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 $k\colon \mathbb{R}^+ \to \mathbb{R}$ defined as

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

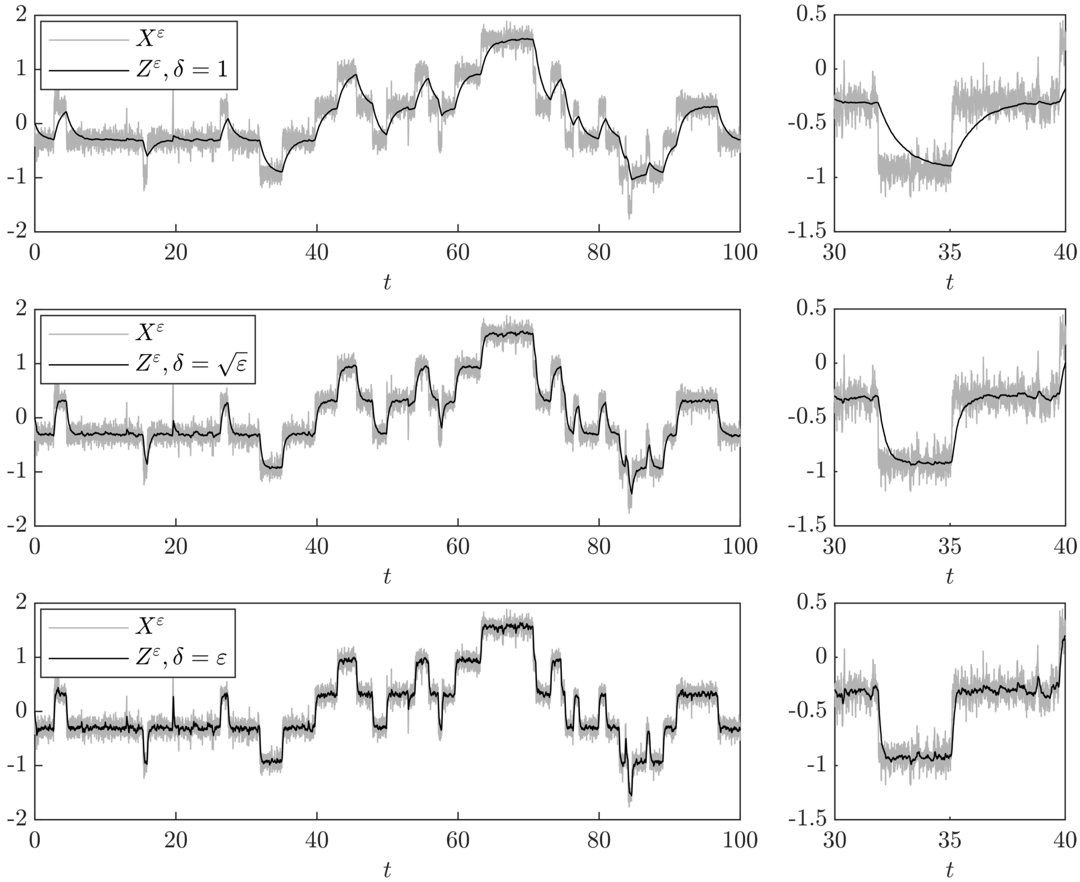


Figure 1: Filtering a trajectory X^ε obtained with $V(x) = x^2/2$, $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 0.5$ and $\varepsilon = 0.1$. The filtering width is $\delta = \{1, \sqrt{\varepsilon}, \varepsilon\}$ from top to bottom, respectively, and $\beta = 1$.

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 $Z^\varepsilon := (Z_t^\varepsilon, 0 \le t \le T)$ defined by the weighted average

$$Z_t^\varepsilon := \int_0^t k(t-s) X_s^\varepsilon \,\mathrm{d}s.$$

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".

Given a trajectory X^ε and the filtered data Z^ε, the estimator of the drift coefficient we propose is given by

$$\widehat{A}_k^\varepsilon(T) = -\widetilde{M}^{-1}\widetilde{h}, \tag{3.2}$$

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

$$\widetilde{M} = \frac{1}{T}\int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t, \qquad \text{and} \qquad \widetilde{h} = \frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon. \tag{3.3}$$

Let us remark that the formula above is obtained from (2.7) 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 $\mathrm{d}X_t^\varepsilon$. Let us furthermore remark that $\widehat{A}_k^\varepsilon(T)$ need not be the minimizer of some filtering-based likelihood function. In fact, if one were to replace Z_t^ε directly in (2.6), the symmetric part of the matrix $\widetilde{M}$ would appear and $\widehat{A}_k^\varepsilon(T)$ would not be the minimizer. Therefore, the estimator $\widehat{A}_k^\varepsilon(T)$ has to be thought of as a perturbation of $\widehat{A}^\varepsilon(T)$, directly at the level of estimators and after the maximization procedure. The only theoretical guarantee which is still needed for the well-posedness of $\widehat{A}_k^\varepsilon(T)$ is for $\widetilde{M}$ to be invertible, which we assume to be true and which we observed to hold in practice.

Let us from now on consider $\beta = 1$. For this value of β, the parameter δ appearing in (3.1) regulates the width of the filtering window. In practice, larger values of δ will lead to trajectories which are smoother and for which fast-scale oscillations are practically canceled. Let us remark that the filtering width resembles the subsampling step employed for the estimator $\widehat{A}_\delta^\varepsilon$ introduced and analyzed in [25]. For subsampling, the choice guaranteeing asymptotically unbiased results is $\delta = \varepsilon^\zeta$ with $\zeta \in (0,1)$, and a similar analysis is due for our filtering technique. For visualization purposes, we depict in Figure 1 the filtered trajectory Z^ε for three different values of δ, namely $\delta = \{1, \sqrt{\varepsilon}, \varepsilon\}$. With $\delta = 1$, all oscillations at the fast scale are canceled and the filtered trajectory Z^ε presents only slow-scale variations. Reducing the value of δ, fast-scale oscillations are progressively taken into account.

In the following, we first focus on the ergodic properties of the process Z^ε when it is coupled with the process X^ε. This analysis is practically independent of the choice of δ, and is therefore presented on its own. Then, we focus on two different cases which depend on the choice of the width δ of the filter. First, we consider δ to be of the magnitude of the slow-scale variations, i.e., $\delta \propto \varepsilon^0$. In this case, an analysis based on multiscale convergence is presented. We then move on to the case $\delta \propto \varepsilon^\zeta$, and show how to choose the exponent ζ in order to obtain unbiased estimators. For this second case, we widely employ techniques and estimates which come from [25].

3.1 Ergodic properties

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

$$\mathrm{d}Z_t^\varepsilon = k(0) X_t^\varepsilon\,\mathrm{d}t + \int_0^t k'(t-s) X_s^\varepsilon\,\mathrm{d}s\,\mathrm{d}t = \frac{1}{\delta}\left(X_t^\varepsilon - Z_t^\varepsilon\right)\mathrm{d}t,$$

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

$$\begin{aligned} \mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\ \mathrm{d}Z_t^\varepsilon &= \frac{1}{\delta}\left(X_t^\varepsilon - Z_t^\varepsilon\right)\mathrm{d}t. \end{aligned} \tag{3.4}$$

The first ingredient for verifying the ergodic properties of the two-dimensional process $(X^\varepsilon, Z^\varepsilon)^\top := ((X_t^\varepsilon, Z_t^\varepsilon)^\top, 0 \le t \le T)$ 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* (3.4) *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* (3.4) *is geometrically ergodic, i.e., there exists $C, \lambda > 0$ such that for all measurable $f \colon \mathbb{R}^2 \to \mathbb{R}$ 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

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

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

$$\sigma \partial_{xx}^2 \rho^\varepsilon(x,z) + \partial_x \left(\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p' \left(\frac{x}{\varepsilon} \right) \right) \rho^\varepsilon(x,z) \right) + \frac{1}{\delta} \partial_z \left((z - x) \rho^\varepsilon(x,z) \right) = 0. \tag{3.5}$$

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 $N > 1$.

In a general case, it is not possible to find an explicit solution to (3.5). 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 3.5. *Under the assumptions of Lemma 3.3, let ρ^ε be the solution of* (3.5) *and let us write*

$$\rho^\varepsilon(x,z) = \varphi^\varepsilon(x) \psi^\varepsilon(z) R^\varepsilon(x,z), \tag{3.6}$$

where φ^ε and ψ^ε are the marginal densities of X^ε and Z^ε respectively, i.e.,

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

Then, it holds

$$\varphi^\varepsilon(x) = \frac{1}{C_{\varphi^\varepsilon}} \exp \left(-\frac{1}{\sigma} \alpha \cdot V(x) - \frac{1}{\sigma} p \left(\frac{x}{\varepsilon} \right) \right), \tag{3.7}$$

where

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

Moreover, it holds

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

3.2 Filtering at the homogenized regime

In this section, we analyze the behavior of the filtering-based estimator $\widehat{A}_k^\varepsilon(T)$ given in (3.2) when the filtering width δ is independent of ε. In particular, we suppose there exist positive constants C_1 and C_2 independent of ε and δ such that $C_1\varepsilon^0 \leq \delta \leq C_2\varepsilon^0$. In the following, we will write more compactly that $\delta = \mathcal{O}(\varepsilon^0)$. The analysis in this case is based on the convergence of the couple $(X^\varepsilon, Z^\varepsilon)^\top$ with respect to the multiscale parameter $\varepsilon \to 0$. 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 guarantees the same kind of convergence for the couple $(X^\varepsilon, Z^\varepsilon)^\top$.

Lemma 3.6. *Under Assumption 2.1, let μ^ε be the invariant measure of the couple $(X^\varepsilon, Z^\varepsilon)^\top$ and let $(X_0^\varepsilon, Z_0^\varepsilon)^\top \sim \mu^\varepsilon$. If $\delta = \mathcal{O}(\varepsilon^0)$, then the measure μ^ε converges weakly to the measure $\mu^0(\mathrm{d}x, \mathrm{d}z) = \rho^0(x,z)\,\mathrm{d}x\,\mathrm{d}z$, whose density ρ^0 is the unique solution of the Fokker–Planck equation*

$$\Sigma\partial_{xx}^2\rho^0(x,z) + \partial_x\left(A \cdot V'(x)\rho^0(x,z)\right) + \frac{1}{\delta}\partial_z\left((z-x)\rho^0(x,z)\right) = 0, \tag{3.8}$$

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

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

$$\begin{aligned}
\mathrm{d}X_t &= -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t, \\
\mathrm{d}Z_t &= \frac{1}{\delta}\left(X_t - Z_t\right)\,\mathrm{d}t,
\end{aligned}$$

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 (3.8). Moreover, standard homogenization theory (see e.g. [8, Chapter 3]) guarantees that $(X^\varepsilon, Z^\varepsilon)^\top \to (X,Z)^\top$ for $\varepsilon \to 0$ 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 $\varepsilon \to 0$. □

Example 3.7. A closed form solution of (3.8) can be obtained in a simple case. Let the dimension of the parameter $N = 1$ and let $V(x) = x^2/2$. Then, the analytical solution is given by

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

where

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

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.

We now present an analogous result to Lemma 3.5 for the limit distribution.

Corollary 3.8. *Let ρ^0 be the solution of* (3.8) *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.,

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

Then, if A and Σ are the coefficients of the homogenized equation (2.3)*, it holds*

$$\varphi^0(x) = \frac{1}{C_{\varphi^0}} \exp\left(-\frac{1}{\Sigma} A \cdot V(x)\right), \quad \textit{where} \quad C_{\varphi^0} = \int_{\mathbb{R}} \exp\left(-\frac{1}{\Sigma} A \cdot V(x)\right) \mathrm{d}x. \tag{3.9}$$

Moreover, it holds

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

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

Let us introduce a notation which will be used throughout the rest of the paper. We denote

$$\widetilde{\mathcal{M}}_\varepsilon := \mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)], \quad \widetilde{\mathcal{M}}_0 := \mathbb{E}^{\rho^0}[V'(Z) \otimes V'(X)], \tag{3.10}$$

i.e., $\widetilde{\mathcal{M}}_\varepsilon$ is the limit for $T \to \infty$ of the matrix $\widetilde{M}$ by the ergodic theorem, and $\widetilde{\mathcal{M}}_0$ is the limit for $\varepsilon \to 0$ of the matrix $\widetilde{\mathcal{M}}_\varepsilon$ due to Lemma 3.6. For completeness, we introduce here the symmetric matrices $\mathcal{M}_\varepsilon$ and $\mathcal{M}_0$ which are defined as

$$\mathcal{M}_\varepsilon := \mathbb{E}^{\rho^\varepsilon}[V'(X^\varepsilon) \otimes V'(X^\varepsilon)], \quad \mathcal{M}_0 := \mathbb{E}^{\rho^0}[V'(X) \otimes V'(X)], \tag{3.11}$$

and which will be employed in the following. We can now introduce the main result, namely the convergence of the filtering-based estimator to the drift coefficient of the homogenized equation.

Theorem 3.9. *Let the assumptions of Lemma 3.3 and Lemma 3.6 hold, and let $\widehat{A}_k^\varepsilon(T)$ be defined in* (3.2) *with $\delta = \mathcal{O}(\varepsilon^0)$. If $\widetilde{M}$ is invertible, then*

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

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

Proof. Replacing the expression of $\mathrm{d}X_t^\varepsilon$ into (3.3), we get for $\widetilde{h}$

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

Therefore, we have

$$\begin{aligned}\widehat{A}_k^\varepsilon(T) &= \alpha + \frac{1}{T}\widetilde{M}^{-1}\int_0^T \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right) V'(Z_t^\varepsilon)\,\mathrm{d}t - \frac{\sqrt{2\sigma}}{T}\widetilde{M}^{-1}\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}W_t \\ &=: \alpha + I_1^\varepsilon(T) - I_2^\varepsilon(T).\end{aligned} \tag{3.12}$$

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

$$\lim_{T\to\infty} I_1^\varepsilon(T) = \widetilde{\mathcal{M}}_\varepsilon^{-1}\,\mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon} p'\left(\frac{X^\varepsilon}{\varepsilon}\right) V'(Z^\varepsilon)\right], \quad \text{a.s.} \tag{3.13}$$

Due to Lemma 3.5 and integrating by parts, we have

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

which implies

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

Replacing the equality above into (3.13), we obtain

$$\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + \widetilde{\mathcal{M}}_\varepsilon^{-1}\sigma\int_{\mathbb{R}}\int_{\mathbb{R}}V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z, \quad \text{a.s.}$$

Due to Lemma 3.5, we therefore have

$$\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta}\widetilde{\mathcal{M}}_\varepsilon^{-1}\,\mathbb{E}^{\rho^\varepsilon}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)], \quad \text{a.s.} \tag{3.14}$$

Since $\delta = \mathcal{O}(\varepsilon^0)$ we can now safely pass to the limit as ε goes to zero and

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta}\widetilde{\mathcal{M}}_0^{-1}\,\mathbb{E}^{\rho^0}[(X-Z)^2V''(Z)], \quad \text{a.s.}, \tag{3.15}$$

where the function ρ^0 is the density of the limit invariant distribution for $\varepsilon \to 0$, as in Lemma 3.6. Due to Corollary 3.8, we have

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

and moreover, an integration by parts yields

$$\begin{aligned}
\frac{1}{\delta}\,\mathbb{E}^{\rho^0}[(X-Z)^2V''(Z)] &= -\Sigma\int_{\mathbb{R}}\int_{\mathbb{R}}V'(z)(\varphi^0)'(x)\psi^0(z)R^0(x,z)\,\mathrm{d}x\,\mathrm{d}z \\
&= -\Sigma\int_{\mathbb{R}}\int_{\mathbb{R}}V'(z)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{C_{\varphi^0}}e^{-\frac{1}{\Sigma}A\cdot V(x)}\right)\psi^0(z)R^0(x,z)\,\mathrm{d}x\,\mathrm{d}z \\
&= \left(\int_{\mathbb{R}}\int_{\mathbb{R}}V'(z)\otimes V'(x)\rho^0(x,z)\,\mathrm{d}x\,\mathrm{d}z\right)A \\
&= \widetilde{\mathcal{M}}_0 A.
\end{aligned}$$

We can therefore conclude that

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + A, \quad \text{a.s.} \tag{3.16}$$

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

$$I_{2,1}^\varepsilon(T) := \left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\otimes V'(X_t^\varepsilon)\,\mathrm{d}t\right)^{-1}\left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\otimes V'(Z_t^\varepsilon)\,\mathrm{d}t\right),$$
$$I_{2,2}^\varepsilon(T) := \left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\otimes V'(Z_t^\varepsilon)\,\mathrm{d}t\right)^{-1}\left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}W_t\right).$$

The ergodic theorem yields

$$\lim_{T\to\infty} I_{2,1}^\varepsilon(T) = \widetilde{\mathcal{M}}_\varepsilon^{-1}\,\mathbb{E}^{\rho^\varepsilon}\left[V'(Z^\varepsilon)\otimes V'(Z^\varepsilon)\right] =: R^\varepsilon,$$

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

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

independently of ε. Therefore

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

which, together with (3.16) and (3.12), proves the desired result. □

Remark 3.10. Let us remark that the condition $\delta = \mathcal{O}(\varepsilon^0)$ is necessary to pass from (3.14) to (3.15) but is not needed before (3.14). Moreover, the term $I_2^\varepsilon(t)$ in the proof vanishes a.s. independently of ε. Therefore, in the analysis of the case $\delta = \mathcal{O}(\varepsilon^\zeta)$ it will be sufficient for unbiasedness to show that

$$\lim_{\varepsilon\to 0}\frac{1}{\delta}\widetilde{\mathcal{M}}_\varepsilon^{-1}\,\mathbb{E}^{\rho^\varepsilon}\left[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)\right] = A,$$

which is a non-trivial limit since $\delta \to 0$ for $\varepsilon \to 0$.

3.3 Filtering at the multiscale regime

Remark for Assyr, Greg, Andrew: This section is all new, therefore it's "red". For easing the reading process, we write it here in black.

We now consider the case of the filtering width $\delta = \mathcal{O}(\varepsilon^\zeta)$, where $\zeta > 0$ will be specified in the following. In this case, the filtered process resembles more the original process X^ε, as can be noted in Figure 1. Moreover, the techniques employed for proving Theorem 3.9 can only be partly exploited, as highlighted by Remark 3.10. In fact, in order to prove unbiasedness it is necessary to characterize precisely the difference between the processes Z^ε and X^ε. A first characterization is given by the following Proposition, whose proof ought to be found in Appendix B.

Proposition 3.11. *Let Assumption 2.1 hold and $\varepsilon, \delta > 0$ be sufficiently small. Then, it holds for every $t > 0$ sufficiently big*

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

where the stochastic process B_t^ε is defined as

$$B_t^\varepsilon := \sqrt{2\sigma}\int_0^t k(t-s)(1+\Phi'(Y_s^\varepsilon))\,\mathrm{d}W_s, \tag{3.17}$$

where Φ is the solution of the cell problem (2.5), W_s is the Brownian motion appearing in (2.1) and $Y_t^\varepsilon = X_t^\varepsilon/\varepsilon$. Moreover, B_t^ε and the remainder $R(\varepsilon,\delta)$ satisfy for every $p \geq 1$ the estimates

$$\left(\mathbb{E}^{\varphi^\varepsilon}\left|B_t^\varepsilon\right|^p\right)^{1/p} \leq C\delta^{-1/2}, \tag{3.18}$$

and

$$\left(\mathbb{E}^{\varphi^\varepsilon}\left|R(\varepsilon,\delta)\right|^p\right)^{1/p} \leq C\left(\delta+\varepsilon\right), \tag{3.19}$$

where C is independent of ε and δ and φ^ε is the density of the invariant measure of X^ε.

It is clear from the Proposition above that understanding the properties of the process B_t^ε is key to understanding the behavior of the difference between X^ε and Z^ε. In particular, we can write the dynamics of B_t^ε with an application of the Itô formula and due to the properties of the kernel $k(t)$ as

$$\mathrm{d}B_t^\varepsilon = -\frac{1}{\delta} B_t^\varepsilon \,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\delta}(1 + \Phi'(Y_t^\varepsilon))\,\mathrm{d}W_t.$$

This equation can be coupled with the dynamics of the processes X_t^ε, Y_t^ε and Z_t^ε, thus describing the evolution of the quadruple $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon, B^\varepsilon)$ together. In particular, it is possible to show that the results of Section 3.1 hold for the quadruple, and the properties of the invariant measure of the quadruple can be exploited to prove the unbiasedness of the estimator in the case $\delta = \mathcal{O}(\varepsilon^\zeta)$ in the same way as in the case $\delta = \mathcal{O}(\varepsilon^0)$. In this context, a further assumption on the potential V is necessary.

Assumption 3.12. The derivatives V', V'' and V''' of the potential $V \colon \mathbb{R} \to \mathbb{R}^N$ are polynomially bounded from above and bounded from below, and the second derivative is Lipschitz, i.e., there exists a constant $L > 0$ such that

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

for all $x, y \in \mathbb{R}$.

In light of Remark 3.10, it is fundamental to understand the behavior of the quantity

$$\frac{1}{\delta}(X_t^\varepsilon - Z_t^\varepsilon)^2 V''(Z_t^\varepsilon),$$

and of its limit for $t \to \infty$ and for $\varepsilon \to 0$. Let us remark that due to Proposition 3.11 we have

$$\frac{1}{\delta}(X_t^\varepsilon - Z_t^\varepsilon)^2 V''(Z_t^\varepsilon) \approx \delta (B_t^\varepsilon)^2 V''(Z_t^\varepsilon),$$

and therefore studying the right hand side of the approximate equality above is the goal of the upcoming discussion. The following result, whose proof is in Appendix C, gives a first characterization.

Lemma 3.13. *Under Assumptions 2.1 and 3.12, let η^ε be the invariant measure of the quadruple $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon, B^\varepsilon)$. Then it holds*

$$\delta\, \mathbb{E}^{\eta^\varepsilon}\left[(B^\varepsilon)^2 V''(Z^\varepsilon)\right] = \sigma\, \mathbb{E}^{\eta^\varepsilon}[(1 + \Phi'(Y^\varepsilon))^2 V''(Z^\varepsilon)] + \widetilde{R}(\varepsilon, \delta),$$

where the remainder $\widetilde{R}(\varepsilon, \delta)$ satisfies

$$\left|\widetilde{R}(\varepsilon, \delta)\right| \leq C\left(\delta^{1/2} + \varepsilon\right).$$

Let us remark that the quantity appearing above hints towards the theory of homogenization. In fact, we recall that the homogenization coefficient K is given by

$$K = \int_0^L \left(1 + \Phi'(y)\right)^2 \mu(\mathrm{d}y),$$

where μ is the marginal measure of the process Y^ε when coupled with X^ε. Therefore, the next step is the homogenization limit, i.e., the limit of vanishing ε, which is considered in the following Lemma, and whose proof is given in Appendix C.

Lemma 3.14. *Let the assumptions of Lemma 3.13 hold, and let $\delta = \varepsilon^\zeta$ with $\zeta > 0$. Then, it holds*

$$\lim_{\varepsilon \to 0} \sigma\, \mathbb{E}^{\eta^\varepsilon}[(1 + \Phi'(Y^\varepsilon))^2 V''(Z^\varepsilon)] = \Sigma\, \mathbb{E}^{\varphi^0}[V''(X)],$$

where Σ is the diffusion coefficient of the homogenized equation (2.3).

Provided with the results presented above, we can prove the following Theorem, stating that the estimator $\widehat{A}_k^\varepsilon(T)$ is asimptotically unbiased even in the case of the filtering width δ vanishing with respect to the multiscale parameter ε.

Theorem 3.15. *Let the assumptions of Lemma 3.3 and Lemma 3.14 hold. Let $\widehat{A}_k^\varepsilon(T)$ be defined in* (3.2) *and $\delta = \varepsilon^\zeta$ with $\zeta \in (0,2)$. If $\widetilde{M}$ is invertible, then*

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{A}_k^\varepsilon(T) = A, \quad \textit{in probability},$$

where A is the drift coefficient of the homogenized equation (2.3).

Proof. 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 $\widetilde{\mathcal{M}}_\varepsilon$ is defined in (3.10). Then following the proof of Theorem 3.9 and in light of Remark 3.10, we only need to show that if $\delta = \varepsilon^\zeta$ with $\zeta \in (0,2)$ we have

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

Using Lemma 3.11 and geometric ergodicity (Lemma 3.3), we have the following equality

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

where $R(\varepsilon,\delta)$ is given in Proposition 3.11, $\mathbb{E}$ denotes the expectation with respect to the Wiener measure and

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

Let us consider the three terms separately. First, by geometric ergodicity and applying Lemma 3.13 and Lemma 3.14 we get

$$\begin{aligned}
\lim_{\varepsilon\to 0}\lim_{t\to\infty} J_1^\varepsilon(t) &= \lim_{\varepsilon\to 0}\delta\,\mathbb{E}^{\eta^\varepsilon}\left[(B^\varepsilon)^2 V''(Z^\varepsilon)\right] \\
&= \lim_{\varepsilon\to 0}\left(\sigma\,\mathbb{E}^{\eta^\varepsilon}[V''(Z^\varepsilon)(1+\Phi'(Y^\varepsilon))^2] + \widetilde{R}(\varepsilon,\delta)\right) \\
&= \Sigma\,\mathbb{E}^{\varphi^0}[V''(X)].
\end{aligned}$$

Let us now consider $J_2^\varepsilon(t)$. Considering Hölder conjugates p, q, r the Hölder inequality yields

$$|J_2^\varepsilon(t)| \le \mathbb{E}[(B_t^\varepsilon)^p]^{1/p}\,\mathbb{E}[R(\varepsilon,\delta)^q]^{1/q}\,\mathbb{E}[V''(Z^\varepsilon)^r]^{1/r}.$$

Now, we can bound the first two terms with (3.18) and (3.19), respectively. The third term is bounded due to Assumption 3.12 and Lemma C.1. Hence, we have

$$|J_2^\varepsilon(t)| \le C\left(\delta^{1/2} + \varepsilon\delta^{-1/2}\right).$$

We consider now $J_3^\varepsilon(t)$. The Hölder inequality yields for conjugates p and q

$$|J_3^\varepsilon(t)| \le \mathbb{E}[R(\varepsilon,\delta)^{2p}]^{1/p}\,\mathbb{E}[V''(Z_t^\varepsilon)^q]^{1/q},$$

which, reasoning as above, yields

$$|J_3^\varepsilon(t)| \leq C\left(\delta + \varepsilon^2\delta^{-1}\right).$$

Therefore, since $\delta = \mathcal{O}(\varepsilon^\zeta)$ for $\zeta \in (0,2)$, the terms $J_2^\varepsilon(t)$ and $J_3^\varepsilon(t)$ vanish in the limit for $t \to \infty$ and $\varepsilon \to 0$. Furthermore, by Lemma C.4 and by weak convergence of the invariant measure μ^ε to μ^0, we have

$$\lim_{\varepsilon \to 0} \widetilde{\mathcal{M}}_\varepsilon = \mathcal{M}_0,$$

where $\mathcal{M}_0$ is defined in (3.11). Therefore

$$\lim_{\varepsilon \to 0} \mathcal{A}^\varepsilon(\delta) = \Sigma \mathcal{M}_0^{-1} \mathbb{E}^{\varphi^0}[V''(X)],$$

and, finally, an integration by parts yields

$$\lim_{\varepsilon \to 0} \mathcal{A}^\varepsilon(\delta) = \Sigma \mathcal{M}_0^{-1} \frac{1}{\Sigma} \mathcal{M}_0 A = A,$$

which implies the desired result. □

We conclude this section with a negative convergence result, i.e., that if $\delta = \varepsilon^\zeta$ with $\zeta > 2$, the filtering estimator converges to the coefficient α of the unhomogenized equation. This result is relevant for two reasons. First, it shows the sharpness of the bound on ζ in the assumptions of Theorem 3.15. Second, it shows an interesting switch with boundary $\zeta = 2$ between two completely different regimes, which happens arbitrarily fast in the limit $\varepsilon \to 0$.

Theorem 3.16. *Let the assumptions of Lemma 3.3 and Assmuption 3.12 hold. Let $\widehat{A}_k^\varepsilon(T)$ be defined in (3.2) and $\delta = \varepsilon^\zeta$ with $\zeta > 2$. If $\widetilde{M}$ is invertible, then*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}_k^\varepsilon(T) = \alpha, \quad \textit{in probability},$$

where α is the drift coefficient of the multiscale equation (2.1).

The proof is given in Appendix C.

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 (2.6). The log-posterior density is therefore given by

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

where M and h are defined in (2.8). 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 $\mu_{T,\varepsilon}(\cdot \mid X^\varepsilon) = \mathcal{N}(m_{T,\varepsilon}, C_{T,\varepsilon})$. In the following theorem, we show that the posterior distribution obtained with no pre-processing of the data contracts asymptotically to the drift coefficient of the unhomogenized equation. We characterize the contraction by verifying that the posterior measure concentrates in arbitrarily small balls. Let us finally remark that the measure $\mu_{T,\varepsilon}$ is a random measure, and therefore contraction has to be considered averaged with respect to the Wiener measure. The choice of the contraction measure and some parts of the proof are taken from [26, Theorem 5.2].

Theorem 4.1. *Under Assumption 2.1, the posterior measure $\mu_{T,\varepsilon}(\cdot \mid X^\varepsilon) = \mathcal{N}(m_{T,\varepsilon}, C_{T,\varepsilon})$ satisfies for all $c > 0$*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \mathbb{E}\left[\mu_{T,\varepsilon}\left(\{a \colon \|a - \alpha\|_2 \geq c\} \mid X^\varepsilon\right)\right] = 0,$$

where $\mathbb{E}$ denotes expectation with respect to the Wiener measure and α is the drift coefficient of the unhomogenized equation (2.1)*.*

Proof. The proof of [26, Theorem 5.2] guarantees that if the trace of $C_{T,\varepsilon}$ tends to zero and if the mean $m_{T,\varepsilon}$ tends to α, then the desired result holds. Indeed, the triangle inequality yields

$$\begin{aligned} \mathbb{E}\left[\mu_{T,\varepsilon}\left(\{a \colon \|a - \alpha\|_2 \geq c\} \mid X^\varepsilon\right)\right] \leq\, & \mathbb{E}\left[\mu_{T,\varepsilon}\left(\left\{a \colon \|a - m_{T,\varepsilon}\|_2 \geq \frac{c}{2}\right\} \mid X^\varepsilon\right)\right] \\ & + \mathbb{P}\left(\|m_{T,\varepsilon} - \alpha\|_2 \geq \frac{c}{2}\right). \end{aligned}$$

If the mean converges in probability, then the second term vanishes. For the first term, Markov's inequality yields

$$\mu_{T,\varepsilon}\left(\left\{a \colon \|a - m_{T,\varepsilon}\|_2 \geq \frac{c}{2}\right\} \mid X^\varepsilon\right) \leq \frac{4}{c^2} \int_{\mathbb{R}^N} \|a - m_{T,\varepsilon}\|_2^2 \, \mu_{T,\varepsilon}(\mathrm{d}a \mid X^\varepsilon),$$

and a change of variable simply gives

$$\int_{\mathbb{R}^N} \|a - m_{T,\varepsilon}\|_2^2 \, \mu_{T,\varepsilon}(\mathrm{d}a \mid X^\varepsilon) = \operatorname{tr}(C_{T\varepsilon}).$$

This proves that we just to have to verify that the covariance matrix vanishes and that the mean tends to the coefficient α. Let us first consider the covariance matrix. An algebraic identity yields

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

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 $T > 0$

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

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

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

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

$$\left\|Q^{-1}\right\|_2 \leq \frac{2\Sigma}{T} \left\|(MC_0M)^{-1}\right\|_2 \leq \frac{2\Sigma}{T} \left\|M^{-1}\right\|_2^2 \left\|C_0^{-1}\right\|_2 = \frac{2\Sigma}{T\bar{\lambda}^2} \left\|C_0^{-1}\right\|_2,$$

which implies

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

and due to the triangle inequality

$$\lim_{T\to\infty} \left\|C_{T,\varepsilon}\right\|_2 = 0. \tag{4.1}$$

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

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

For the second term, Theorem 2.3 implies

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

Let us now consider the first term. Replacing the expression of the maximum likelihood estimator (2.7) and due to the Cauchy–Schwarz and triangle inequalities, we obtain

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

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

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

independently of ε. Finally,

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

which, together with (4.1), implies the desired result. □

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 $T \to \infty$ and for a positive value $\varepsilon > 0$ the Bayesian and the MLE approaches are equivalent. In particular, we have for all $\varepsilon > 0$

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

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

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

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

where

$$\begin{aligned}\widehat{I}(X^\varepsilon \mid A) &= \int_0^T A \cdot V'(Z_t^\varepsilon) \, \mathrm{d}X_t^\varepsilon + \frac{1}{2} \int_0^T \left(A \cdot V'(X_t^\varepsilon)\right) \left(A \cdot V'(Z_t^\varepsilon)\right) \, \mathrm{d}t \\ &= \widetilde{h} \cdot A + \frac{1}{2} A \cdot \widetilde{M} A.\end{aligned}$$

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.,

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

where

$$\begin{aligned}\widetilde{I}(X^\varepsilon \mid A) &= \int_0^T A \cdot V'(Z_t^\varepsilon) \, \mathrm{d}X_t^\varepsilon + \frac{1}{2} \int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2 \, \mathrm{d}t \\ &= \widetilde{h} \cdot A + \frac{1}{2} A \cdot M A.\end{aligned}$$

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 $\widetilde{p}(X^\varepsilon \mid A)$, i.e., the quantity

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

The following theorem guarantees the unbiasedness of this estimator under a condition on the parameter δ of the filter.

Theorem 4.3. *Let the assumptions of Theorem 3.15 hold. Then, if $\delta = \varepsilon^\zeta$, with $\zeta \in (0, 2)$, it holds*

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

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

Proof. 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

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

almost surely, where $\mathcal{M}_\varepsilon$ and $\widetilde{\mathcal{M}}_\varepsilon$ are defined in (3.11) and (3.10), respectively. Therefore, due to Assumption 2.1 which allows controlling the norm of $\mathcal{M}_\varepsilon^{-1}$ and due to Lemma C.4 we have for a constant $C > 0$

$$\lim_{T\to\infty} \left\| \widetilde{A}_k^\varepsilon(T) - \widehat{A}_k^\varepsilon(T) \right\|_2 \leq C \left(\varepsilon + \delta^{1/2} \right). \tag{4.3}$$

Let us remark that $\widehat{A}_k^\varepsilon(T)$ has a bounded norm for ε sufficiently small due to Theorem 3.15. 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 3.15, the inequality (4.3) and since $\delta = \varepsilon^\zeta$, the desired result holds. □

In light of the proof of Theorem 4.1, the result above guarantees that the mean of the posterior distribution $\widetilde{\mu}_{T,\varepsilon}$ converges to the drift coefficient of the homogenized equation. Being the covariance matrix the same for $\mu_{T,\varepsilon}$ and $\widetilde{\mu}_{T,\varepsilon}$, it is possible to prove a positive convergence result for $\widetilde{\mu}_{T,\varepsilon}$, which is given by the following Theorem.

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

$$\lim_{\varepsilon\to 0} \lim_{T\to\infty} \mathbb{E}\left[\widetilde{\mu}_{T,\varepsilon} \left(\{ a \colon \|a - A\|_2 \geq c \} \mid X^\varepsilon \right) \right] = 0,$$

where $\mathbb{E}$ denotes expectation with respect to the Wiener measure and A is the drift coefficient of the homogenized equation (2.3)*.*

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

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

For the first preliminary experiments, 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 the fast potential in the multiscale equation (2.1) as $p(y) = \cos(y)$. In all experiments, data is generated employing the Euler–Maruyama method with a fine time step.

5.1.1 Verification of theoretical results

We first demonstrate numerically the validity of Theorem 3.9, Theorem 3.15 and Theorem 3.16, i.e., the unbiasedness of $\widehat{A}_k^\varepsilon(T)$ for $\delta = \varepsilon^\zeta$ with $\zeta \in [0, 2)$. Let us recall that for $\zeta = 0$ the analysis and the theoretical result are fundamentally different than for $\zeta \in (0, 2)$. We consider $\varepsilon \in \{0.1, 0.05, 0.025\}$, the diffusion coefficient $\sigma = 1$ and generate data X_t^ε for $0 \leq t \leq T$ with $T = 10^3$. Then we filter the data by choosing $\delta = \varepsilon^\zeta$, and $\zeta = 0, 0.1, 0.2, \dots, 3$, and compute $\widehat{A}_k^\varepsilon(T)$. Results are displayed in Figure 2, and show that for $\zeta > 2$, i.e., $\delta = o(\varepsilon^2)$, the estimator tends to the drift coefficient α of the unhomogenized equation. Conversely, as predicted by the theory, for $\zeta \in [0, 2)$ the estimator tends to A, the drift coefficient of the homogenized equation. Therefore, the point $\delta = \varepsilon^2$ acts asymptotically as a switch between two completely different regimes. The transition between these two regimes is theoretically sharp in the limit for $\varepsilon \to 0$. In finite computations the passage between A and α is smoother, and suggests to take, if possible, $\delta = \varepsilon^0$. In the Bayesian case one has to fix $\zeta > 0$, and results indicate that possibly ζ should be close to 0. In particular, all values of $\zeta \in [0, 1]$ seem to output similarly good results.

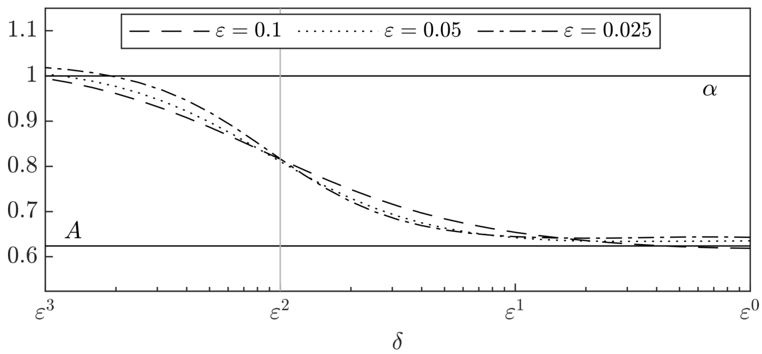


Figure 2: Results for Section 5.1.1. Horizontal lines represent α and A, the drift coefficients of the unhomogenized and homogenized equations. The gray vertical line represents the lower bound for the validity of Theorem 3.15. The curved lines (dashed, dotted and dash-dotted) represents the values of $\widehat{A}_k^\varepsilon(T)$ for $\varepsilon = \{0.1, 0.05, 0.025\}$, respectively. Remark that the δ-axis is in logarithmic scale and is normalized with respect to ε.

5.1.2 Comparison with subsampling

We now compare the results given by the filtering technique with the results given by subsampling the data, i.e., the difference between the estimators $\widehat{A}_k^\varepsilon(T)$ and $\widehat{A}_\delta^\varepsilon(T)$. We fix the diffusion coefficient $\sigma = 0.5$, the multiscale parameter $\varepsilon = 0.1$ and generate data for $0 \leq t \leq T$ with $T = 10^3$. We choose $\delta = \varepsilon^\zeta$ and vary $\zeta \in [0, 1]$, where δ is the filtering and the subsampling width, respectively. Moreover, for the filtering method we consider both $\beta = 1$ and $\beta = 5$. We report in Figure 3 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. This can be observed for a higher value of β but we have no theoretical guarantee in this case.

5.1.3 The influence of β

We finally test the variability of the estimator with respect to β in (3.1). We 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 4, show empirically that the estimator stabilizes fast with respect to β. Nevertheless, there is no theoretical guarantee supporting this empirical observation.

5.2 Multi-dimensional drift coefficient

Let us consider the Chebyshev polynomials of the first kind, i.e., the polynomials $T_i \colon \mathbb{R} \to \mathbb{R}$, $i = 0, 1, \dots$, 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

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

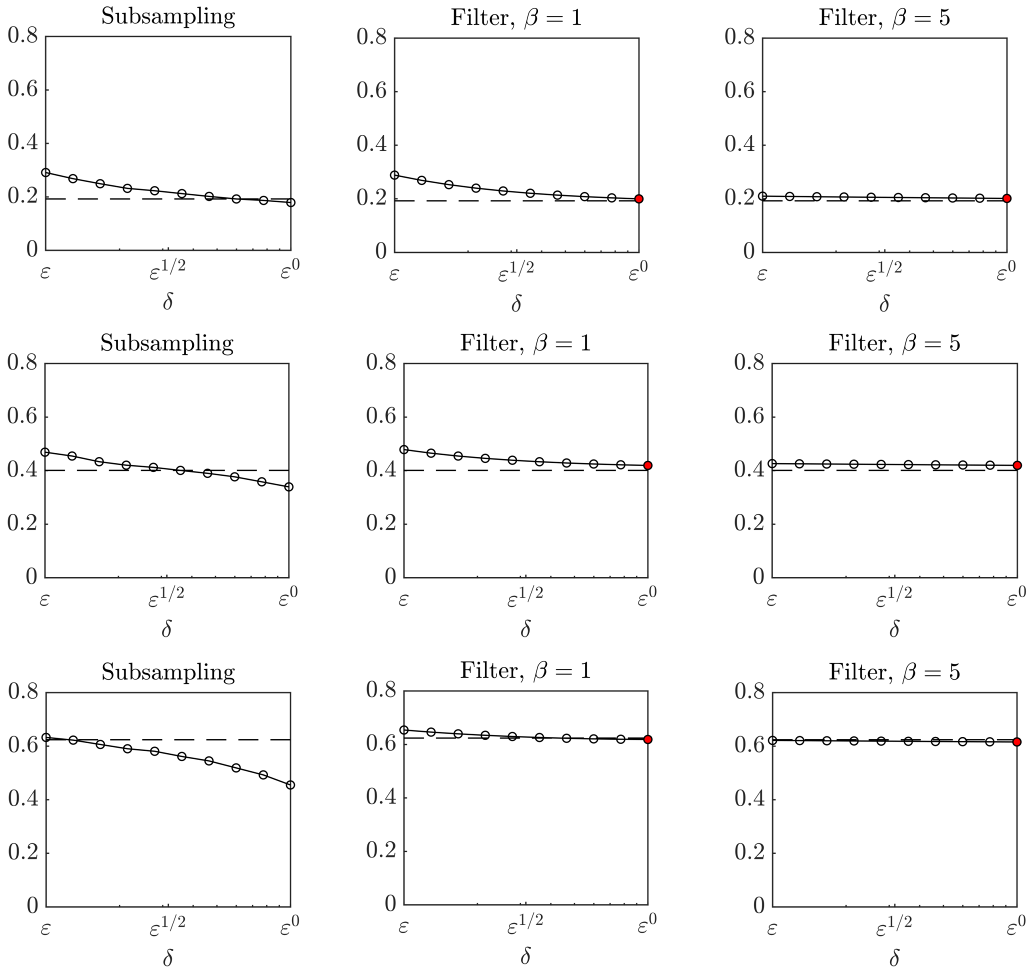


Figure 3: Results for Section 5.1.2. The case of $\delta = \varepsilon^0$ is highlighted as a solid dot for the filtering technique, as the analysis and theoretical result is different in this case. The three rows correspond to $\sigma = 0.5, 0.7, 1.0$ from top to bottom.

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) $\widehat{A}^\varepsilon(T)$ based on the data X^ε itself,

(ii) $\widehat{A}^\varepsilon_\delta(T)$ based on subsampled data with subsampling parameter $\delta = \varepsilon^{2/3}$,

(iii) $\widehat{A}^\varepsilon_k(T)$ 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 5, 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 $\widehat{A}^\varepsilon(T)$ tends to the drift coefficient α of the unhomogenized equation.

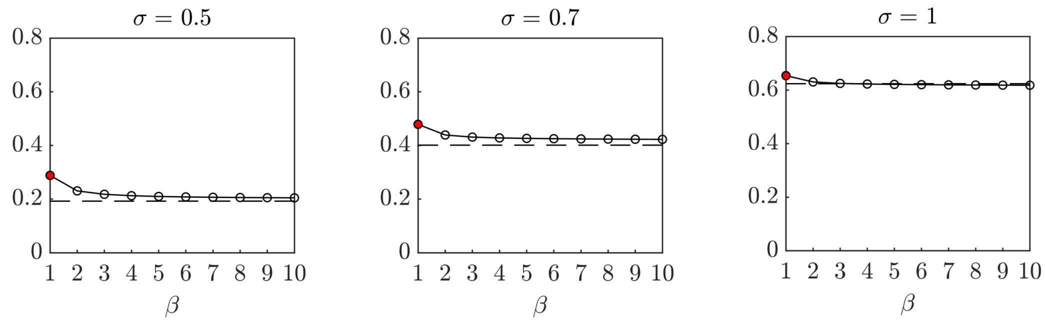
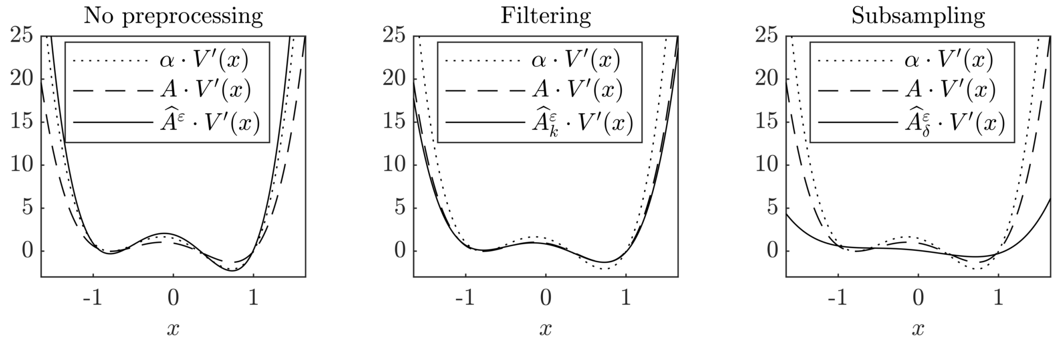


Figure 4: Results for Section 5.1.3. The result for $\beta = 1$, for which there are theoretical guarantees given by Theorem 3.15, is highlighted as a solid dot. From left to right we consider different values of σ.



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

Figure 5: Results for Section 5.2. In the figure, from left to right the potential function estimated with the data itself, the filter, subsampled data. In the table, numerical results for the single components of the true and estimated drift coefficients.

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

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

with coefficients $\alpha_1 = 1$ and $\alpha_2 = 2$. We then consider the multiscale equation with $\sigma = 0.7$, the fast potential $p(y) = \cos(y)$ and $\varepsilon = 0.05$, thus simulating a trajectory X^ε. We adopt here a Bayesian approach and compute the posterior distribution $\widetilde{\mu}_{T,\varepsilon}$ 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 $\widetilde{A}_k^\varepsilon(T)$ with respect to time. In Figure 6, 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.

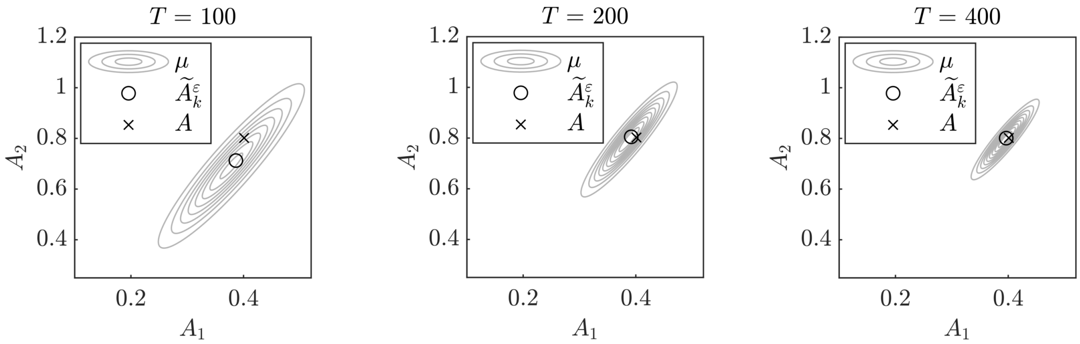


Figure 6: Results for Section 5.3. Posterior distributions over the parameter $A = (A_1, A_2)^\top$ for the bistable potential obtained with the filtering approach. The figures refer to final time $T = 100, 200, 400$ from left to right, respectively. The MLE $\widetilde{A}_k^\varepsilon(t)$ is represented with a circle, while the true value A of the drift coefficient of the homogenized equation is represented with a cross.

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 Proofs of Sections 3.1

Proof of Lemma 3.2. We have to show that the joint process solution to (3.4) is hypo-elliptic. Denoting as $f \colon \mathbb{R} \to \mathbb{R}$ 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,

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

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 (3.4). For geometric ergodicity, let

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

Due to Assumption 2.1(ii), Remark 2.2 and Young's inequality, we then have for all $\gamma > 0$

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

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

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

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

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

and problem (3.4) is dissipative. The result then follows from [19, Theorem 4.4]. $\square$

Proof of Lemma 3.5. Integrating equation (3.5) with respect to z we obtain the stationary Fokker–Planck equation for the process X^ε, i.e.

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

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 (3.7). In view of (3.6) and (A.1), equation (3.5) can be rewritten as

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

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

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

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

$$
\sigma\delta \int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x f(x,z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z = \mathbb{E}^{\rho_\varepsilon}\left[\partial_z f(X^\varepsilon, Z^\varepsilon)(X^\varepsilon - Z^\varepsilon)\right].
$$

Finally, choosing

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

we obtain the desired result. $\square$

Appendix B Proof of Proposition 3.11

B.1 Preliminary estimates

In order to prove the characterization provided by Proposition 3.11, we need to prove two additional results on the filter. 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 $t > 0$, $p \geq 1$ and any function $g \in \mathcal{C}^0([0,t])$ it holds

$$\left| \int_0^t k(t-s) g(s) \, \mathrm{d}s \right|^p \leq \int_0^t k(t-s) \left| g(s) \right|^p \, \mathrm{d}s.$$

Proof. Let us first note that

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

Therefore, the measure $\kappa_t(\mathrm{d}s)$ on $[0,t]$ defined as

$$\kappa_t(\mathrm{d}s) \coloneqq \frac{k(t-s)}{1 - e^{-t/\delta}} \, \mathrm{d}s,$$

is a probability measure. An application of Jensen's inequality therefore yields

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

Finally since $0 < (1 - e^{-t/\delta}) < 1$ and $p \geq 1$, this yields the desired result. $\square$

The following lemma characterizes the action of the filter when it is applied to polynomials in $(t-s)$.

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

$$\int_0^t k(t-s)(t-s)^p \, \mathrm{d}s \leq C\delta^p,$$

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

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

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

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

B.2 Proof of Proposition 3.11

Denoting $Y_t^\varepsilon \coloneqq X_t^\varepsilon/\varepsilon$, we will make use of the decomposition [25, Formula 5.8]

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

which is obtained applying the Itô formula to Φ, the solution of the cell problem (2.5). Recall that by definition of Z_t^ε we have

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

Plugging the decomposition (B.1) 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

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

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

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

where the remainder R_1^ε satisfies

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

Therefore, it holds

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

where we exploited the equality

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

and where

$$\widetilde{R}_1^\varepsilon(t) \coloneqq \int_0^t k(t-s) R_1^\varepsilon(t-s)\,\mathrm{d}s.$$

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

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

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

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

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

$$Q_t^\varepsilon \coloneqq \int_0^t \left(1+\Phi'(Y_r^\varepsilon)\right)\mathrm{d}W_r,$$

and therefore rewrite

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

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

$$\begin{aligned}I_2^\varepsilon(t)&=\sqrt{2\sigma}\left(Q_t^\varepsilon\int_0^t k(t-s)\,\mathrm{d}s-Q_t^\varepsilon+\delta\int_0^t k(t-s)\left(1+\Phi'(Y_s^\varepsilon)\right)\mathrm{d}W_s\right)\\&=\delta B_t^\varepsilon-\sqrt{2\sigma}e^{-t/\delta}Q_t^\varepsilon=:\delta B_t^\varepsilon-R_2^\varepsilon(t).\end{aligned} \tag{B.3}$$

where B_t^ε is defined in (3.17). For the remainder $R_2^\varepsilon(t)$, let us remark that for all $p\geq 1$ it holds

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

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 we have

$$\left(\mathbb{E}^{\varphi^\varepsilon}\left|R_2^\varepsilon(t)\right|^p\right)^{1/p}\leq C\sqrt{t}e^{-t/\delta}. \tag{B.4}$$

In order to obtain the bound (3.18) on B_t^ε, let us remark that from (B.3) it holds for a constant $C>0$ depending only on p

$$\left(\mathbb{E}\left|B_t^\varepsilon\right|^p\right)^{1/p}\leq C\delta^{-1}\left(\mathbb{E}\left|I_2^\varepsilon(t)\right|^p\right)^{1/p}+C\delta^{-1}\left(\mathbb{E}\left|R_2^\varepsilon(t)\right|^p\right)^{1/p}.$$

The second term is bounded exponentially fast with respect to t and δ due to (B.4). For the first term, applying Lemma B.1, the inequality [16, Formula (3.25), p. 163] and Lemma B.2 we obtain for a constant $C>0$ independent of δ and t

$$\begin{aligned}\mathbb{E}\left|I_2^\varepsilon(t)\right|^p&\leq C\int_0^t k(t-s)\,\mathbb{E}\left|Q_t-Q_s\right|^p\mathrm{d}s\\&\leq C\int_0^t k(t-s)(t-s)^{p/2}\,\mathrm{d}s\leq C\delta^{p/2}.\end{aligned}$$

Therefore, it holds

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

which proves the bound (3.18). Let us now consider $I_3^\varepsilon(t)$. Since Φ is bounded, we simply have

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

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

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

which concludes the proof. □

Appendix C Proofs of Section 3.3

C.1 Preliminary estimates

The following lemma shows that Z^ε has bounded moments of all orders.

Lemma C.1. *Under Assumption 2.1, let Z^ε be distributed as the invariant measure μ^ε of the couple $(X^\varepsilon, Z^\varepsilon)^\top$. Then for any $p \geq 1$ there exists a constant $C > 0$ uniform in ε such that*

$$\mathbb{E}^{\rho^\varepsilon} |Z^\varepsilon|^p \leq C.$$

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 and applying Lemma B.1 we have

$$\begin{aligned}\mathbb{E}^{\varphi^\varepsilon} |Z_t^\varepsilon|^p &= \mathbb{E}^{\varphi^\varepsilon} \left| \int_0^t k(t-s) X_s^\varepsilon \,\mathrm{d}s \right|^p \\ &\leq \int_0^t k(t-s)\, \mathbb{E}^{\varphi^\varepsilon} |X_s^\varepsilon|^p \,\mathrm{d}s,\end{aligned}$$

which, together with the definition of k and the fact that X_s^ε has bounded moments of all orders [25, Corollary 5.4], implies

$$\mathbb{E}^{\varphi^\varepsilon} |Z_t^\varepsilon|^p \leq C.$$

In order to conclude, we remark that due to Lemma 3.3 we have for all $t \geq 0$ and for a constant $\widetilde{C}$ depending on the initial condition and p

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

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

Corollary C.2 is a direct consequence of Proposition 3.11 and provides a rough estimate of the difference between the trajectories X_t^ε and Z_t^ε when they are at stationarity.

Corollary C.2. *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 $p \geq 1$, then due to Proposition 3.11 there exists a constant $C > 0$ depending only on p such that

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

Let us now remark that this result holds for X_t^ε being at stationarity and for Z_t^ε being 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 $\widetilde{C}$ depending on the initial condition and p

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

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

The result above can be in some sense rather counter-intuitive. Indeed, for a fixed $\varepsilon > 0$ and for $\delta \to 0$ independently of ε, one expects the filtered trajectory Z^ε to approach X^ε. This is provided by the following Lemma.

Lemma C.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 $p \geq 1$*

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

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

Proof. By equation (2.1) we have for all $0 \leq s < t$

$$X_t^\varepsilon - X_s^\varepsilon = -\alpha \int_s^t V'(X_r^\varepsilon)\,\mathrm{d}r - \frac{1}{\varepsilon}\int_s^t p'\left(\frac{X_r^\varepsilon}{\varepsilon}\right)\mathrm{d}r + \sqrt{2\sigma}(W_t - W_s).$$

Therefore, by Assumption 2.1 and since X_t^ε has bounded moments of all orders at stationarity [25, Corollary 5.4], it holds for any $p \geq 1$ and a constant $C > 0$

$$\mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_s^\varepsilon|^p \leq C\left((t-s)^p + (t-s)^p\varepsilon^{-p} + (t-s)^{p/2}\right), \tag{C.1}$$

where φ^ε is the invariant measure of X^ε. By definition of Z_t^ε we have

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

which, applying Lemma B.1, the inequality (C.1) and Lemma B.2, implies

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

Geometric ergodicity (Lemma 3.3) then implies for ρ^ε the measure of the couple $(X^\varepsilon, Z^\varepsilon)^\top$ and a constant $C > 0$

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

which, for t sufficiently big and since $\delta \leq 1$ yields the desired result. □

Let us conclude with a last preliminary estimate concerning the matrices $\widetilde{\mathcal{M}}_\varepsilon$ and $\mathcal{M}_\varepsilon$ defined in (3.10) and (3.11), respectively.

Lemma C.4. *Let the assumptions of Corollary C.2 hold. Then the matrices $\mathcal{M}_\varepsilon$ and $\widetilde{\mathcal{M}}_\varepsilon$ 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

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

The Lipschitz condition on V' together with the boundedness of the moments of X^ε and Corollary C.2 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. □

C.2 Proof of Lemma 3.13

Let us consider the following system of stochastic differential equations for the processes $X_t^\varepsilon, Z_t^\varepsilon, B_t^\varepsilon, Y_t^\varepsilon$

$$
\begin{aligned}
\mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'(Y_t^\varepsilon)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t,\\
\mathrm{d}Z_t^\varepsilon &= \frac{1}{\delta}\left(X_t^\varepsilon - Z_t^\varepsilon\right)\,\mathrm{d}t,\\
\mathrm{d}B_t^\varepsilon &= -\frac{1}{\delta} B_t^\varepsilon\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\delta}(1+\Phi'(Y_t^\varepsilon))\,\mathrm{d}W_t,\\
\mathrm{d}Y_t^\varepsilon &= -\frac{1}{\varepsilon}\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon^2} p'(Y_t^\varepsilon)\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\varepsilon}\,\mathrm{d}W_t,
\end{aligned}
$$

whose generator $\widetilde{\mathcal{L}}_\varepsilon$ is given by

$$
\begin{aligned}
\widetilde{\mathcal{L}}_\varepsilon = &-\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon}p'(y)\right)\partial_x + \frac{1}{\delta}(x-z)\partial_z - \frac{1}{\delta} b\partial_b - \left(\frac{1}{\varepsilon}\alpha\cdot V'(x) + \frac{1}{\varepsilon^2}p'(y)\right)\partial_y\\
&+ \sigma\left(\partial_{xx}^2 + \frac{2}{\varepsilon}\partial_{xy}^2 + \frac{1}{\varepsilon^2}\partial_{yy}^2 + \frac{2(1+\Phi'(y))}{\delta}\partial_{xb}^2 + \frac{2(1+\Phi'(y))}{\varepsilon\delta}\partial_{yb}^2 + \frac{(1+\Phi'(y))^2}{\delta^2}\partial_{bb}^2\right).
\end{aligned}
$$

Let us denote by $\eta^\varepsilon \colon \mathbb{R}^3 \times [0,L] \to \mathbb{R}$, $\eta^\varepsilon = \eta^\varepsilon(x,z,b,y)$, the invariant measure of the quadruple $(X_t^\varepsilon, Z_t^\varepsilon, B_t^\varepsilon, Y_t^\varepsilon)$. Then η^ε solves the stationary Fokker-Planck equation $\widetilde{\mathcal{L}}_\varepsilon^* \eta^\varepsilon = 0$, i.e., explicitly

$$
\begin{aligned}
&\partial_x\left(\left(\alpha\cdot V'(x) + \frac{1}{\varepsilon}p'(y)\right)\eta^\varepsilon\right) + \frac{1}{\delta}\partial_z\left((z-x)\eta^\varepsilon\right)\\
&+ \frac{1}{\delta}\partial_b(b\eta^\varepsilon) + \partial_y\left(\left(\frac{1}{\varepsilon}\alpha\cdot V'(x) + \frac{1}{\varepsilon^2}p'(y)\right)\eta^\varepsilon\right) + \sigma\left(\partial_{xx}^2\eta^\varepsilon + \frac{2}{\varepsilon}\partial_{xy}^2\eta^\varepsilon + \frac{1}{\varepsilon^2}\partial_{yy}^2\eta^\varepsilon\right)\\
&+ \sigma\left(\frac{2}{\delta}\partial_{xb}^2\left((1+\Phi'(y))\eta^\varepsilon\right) + \frac{2}{\varepsilon\delta}\partial_{yb}^2\left((1+\Phi'(y))\eta^\varepsilon\right) + \frac{1}{\delta^2}\partial_{bb}^2\left((1+\Phi'(y))^2\eta^\varepsilon\right)\right) = 0.
\end{aligned}
\tag{C.2}
$$

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

$$
\frac{1}{\delta}\int_{\mathbb{R}^3\times[0,L]} \partial_z f(z,b)(x-z)\eta^\varepsilon - \frac{1}{\delta}\int_{\mathbb{R}^3\times[0,L]} \partial_b f(z,b) b\eta^\varepsilon + \frac{\sigma}{\delta^2}\int_{\mathbb{R}^3\times[0,L]} \partial_{bb}^2 f(z,b)(1+\Phi'(y))^2\eta^\varepsilon,
$$

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

$$
\delta\,\mathbb{E}^{\eta^\varepsilon}\left[\partial_b f(Z^\varepsilon, B^\varepsilon)B^\varepsilon\right] = \sigma\,\mathbb{E}^{\eta^\varepsilon}\left[\partial_{bb}^2 f(Z^\varepsilon,B^\varepsilon)(1+\Phi'(Y^\varepsilon))\right] + \delta\,\mathbb{E}^{\eta^\varepsilon}\left[\partial_z f(Z^\varepsilon,B^\varepsilon)(X^\varepsilon - Z^\varepsilon)\right].
$$

Choosing

$$
f(z,b) = \frac{1}{2}b^2 V''(z),
$$

we obtain

$$
\begin{aligned}
\delta\,\mathbb{E}^{\eta^\varepsilon}\left[(B^\varepsilon)^2 V''(Z^\varepsilon)\right] &= \sigma\,\mathbb{E}^{\eta^\varepsilon}\left[V''(Z^\varepsilon)(1+\Phi'(Y^\varepsilon))\right] + \frac{\delta}{2}\,\mathbb{E}^{\eta^\varepsilon}\left[(B^\varepsilon)^2 V'''(Z^\varepsilon)(X^\varepsilon - Z^\varepsilon)\right]\\
&=: \sigma\,\mathbb{E}^{\eta^\varepsilon}\left[V''(Z^\varepsilon)(1+\Phi'(Y^\varepsilon))\right] + \widetilde{R}(\varepsilon,\delta).
\end{aligned}
$$

We now consider the remainder and, applying Hölder's inequality, Corollary C.2, Lemma C.1, Assumption 3.12 and (3.18), we get for p,q,r such that $1/p + 1/q + 1/r = 1$

$$
\left|\widetilde{R}(\varepsilon,\delta)\right| \le C\delta\left(\mathbb{E}^{\eta^\varepsilon}\left|B^\varepsilon\right|^{2p}\right)^{1/p}\left(\mathbb{E}^{\eta^\varepsilon}\left|V'''(Z^\varepsilon)\right|^q\right)^{1/q}\left(\mathbb{E}^{\eta^\varepsilon}\left|X^\varepsilon - Z^\varepsilon\right|^r\right)^{1/r} \le C(\delta^{1/2} + \varepsilon),
$$

which completes the proof. □

C.3 Proof of Lemma 3.14

Let us introduce the notation

$$\Delta(\varepsilon) = \left| \sigma \, \mathbb{E}^{\eta^\varepsilon} [V''(Z^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] - \Sigma \, \mathbb{E}^{\varphi^0} [V''(X)] \right|,$$

and note that the aim is to show that $\lim_{\varepsilon \to 0} \Delta(\varepsilon) = 0$. By the triangle inequality we get

$$\begin{aligned}
\Delta(\varepsilon) \leq & \left| \sigma \, \mathbb{E}^{\eta^\varepsilon} [V''(Z^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] - \sigma \, \mathbb{E}^{\eta^\varepsilon} [V''(X^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] \right| \\
& + \left| \sigma \, \mathbb{E}^{\eta^\varepsilon} [V''(X^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] - \Sigma \, \mathbb{E}^{\varphi^0} [V''(X)] \right| \\
=: & \Delta_1(\varepsilon) + \Delta_2(\varepsilon).
\end{aligned}$$

We first study $\Delta_1(\varepsilon)$ and due to the boundedness of Φ', Assumption 3.12 and Lemma C.2 we have

$$\Delta_1(\varepsilon) \leq C \, \mathbb{E}^{\eta^\varepsilon} |X^\varepsilon - Z^\varepsilon| \leq C(\delta^{1/2} + \varepsilon) = C(\varepsilon^{\zeta/2} + \varepsilon),$$

which implies

$$\lim_{\varepsilon \to 0} \Delta_1(\varepsilon) = 0.$$

We now consider $\Delta_2(\varepsilon)$. Integrating equation (C.2) with respect to z and b we obtain the Fokker-Planck equation for the stationary marginal distribution $\lambda \colon \mathbb{R} \times [0, L]$, $\lambda = \lambda(x, y)$, of the couple $(X^\varepsilon, Y^\varepsilon)$

$$\begin{aligned}
\partial_x \left(\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p'(y) \right) \lambda \right) + \partial_y \left(\left(\frac{1}{\varepsilon} \alpha \cdot V'(x) + \frac{1}{\varepsilon^2} p'(y) \right) \lambda \right) & \\
+ \sigma \left(\partial^2_{xx} \lambda + \partial^2_{xy} \left(\frac{2}{\varepsilon} \lambda \right) + \partial^2_{yy} \left(\frac{1}{\varepsilon^2} \lambda \right) \right) & = 0,
\end{aligned}$$

whose solution is given by

$$\lambda(x, y) = \frac{1}{C_\lambda} \exp \left(-\frac{\alpha}{\sigma} V(x) - \frac{1}{\sigma} p(y) \right),$$

where

$$\begin{aligned}
C_\lambda &= \int_{\mathbb{R}} \int_0^L \exp \left(-\frac{\alpha}{\sigma} V(x) - \frac{1}{\sigma} p(y) \right) \mathrm{d}x \, \mathrm{d}y \\
&= \left(\int_{\mathbb{R}} \exp \left(-\frac{\alpha}{\sigma} V(x) \right) \mathrm{d}x \right) \left(\int_0^L \exp \left(-\frac{1}{\sigma} p(y) \right) \mathrm{d}y \right) \\
&=: C_{\lambda_x} C_{\lambda_y}.
\end{aligned}$$

Therefore, since $\Sigma = K\sigma$ and by equations (2.4) and (3.9) we have

$$\begin{aligned}
\sigma \, \mathbb{E}^{\eta^\varepsilon} [V''(X^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] &= \sigma \int_{\mathbb{R}} \int_0^L V''(x)(1 + \Phi'(y))^2 \frac{1}{C_\lambda} \exp \left(-\frac{\alpha}{\sigma} V(x) - \frac{1}{\sigma} p(y) \right) \mathrm{d}x \, \mathrm{d}y \\
&= \sigma \left(\int_{\mathbb{R}} V''(x) \frac{1}{C_{\lambda_x}} \exp \left(-\frac{\alpha}{\sigma} V(x) \right) \mathrm{d}x \right) \\
&\qquad \times \left(\int_0^L (1 + \Phi'(y))^2 \frac{1}{C_{\lambda_y}} \exp \left(-\frac{1}{\sigma} p(y) \right) \mathrm{d}y \right) \\
&= \sigma K \, \mathbb{E}^{\varphi^0} [V''(X)] = \Sigma \, \mathbb{E}^{\varphi^0} [V''(X)],
\end{aligned}$$

which shows that $\Delta_2(\varepsilon) = 0$ and completes the proof. $\square$

C.4 Proof of Theorem 3.16

Let us consider the decomposition (3.12), i.e.,

$$\widehat{A}_k^\varepsilon(T) = \alpha + I_1^\varepsilon(T) - I_2^\varepsilon(T),$$

where $I_1^\varepsilon(T)$ is defined as

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

and, by the proof of Theorem 3.9 we have independently of ε

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

A Taylor expansion of the first order of V' yields

$$V'(Z^\varepsilon) = V'(X^\varepsilon) + V''(\widetilde{X}^\varepsilon)(Z^\varepsilon - X^\varepsilon),$$

where $\widetilde{X}^\varepsilon$ is a random variable which assumes values between X^ε and Z^ε. We can therefore write

$$\begin{aligned}
\mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right] &= \mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(X^\varepsilon)\right] + \mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V''(\widetilde{X}^\varepsilon)(Z^\varepsilon - X^\varepsilon)\right] \\
&=: J_1^\varepsilon + J_2^\varepsilon.
\end{aligned}$$

We now consider the two terms separately and show they vanish for $\varepsilon \to 0$. Integrating by parts in J_1^ε we obtain

$$\begin{aligned}
J_1^\varepsilon &= \int_{\mathbb{R}} \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)V(x)\frac{1}{C_{\rho^\varepsilon}}\exp\left(-\frac{\alpha}{\sigma}V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right)\mathrm{d}x \\
&= \int_{\mathbb{R}} \left(\sigma V''(x) - (V'(x)\otimes V'(x))\,\alpha\right)\frac{1}{C_{\rho^\varepsilon}}\exp\left(-\frac{\alpha}{\sigma}V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right)\mathrm{d}x \\
&= \sigma\,\mathbb{E}^{\rho^\varepsilon}\left[V''(X^\varepsilon)\right] - \mathbb{E}^{\rho^\varepsilon}\left[V'(X^\varepsilon)\otimes V'(X^\varepsilon)\right]\alpha.
\end{aligned}$$

We then pass to the limit as $\varepsilon \to 0$ and integrate by parts again to obtain

$$\lim_{\varepsilon\to 0} J_1^\varepsilon = \sigma\,\mathbb{E}^{\rho^0}\left[V''(X)\right] - \mathbb{E}^{\rho^0}\left[V'(X)\otimes V'(X)\right]\alpha = 0. \tag{C.3}$$

We now turn to J_2^ε. The Hölder's inequality with conjugate exponents p and q and the assumptions on p and V yield

$$|J_2^\varepsilon| \le C\varepsilon^{-1}\left(\mathbb{E}^{\rho^\varepsilon}\left|\widetilde{X}^\varepsilon\right|^q\right)^{1/q}\left(\mathbb{E}^{\rho^\varepsilon}\left|Z^\varepsilon - X^\varepsilon\right|^p\right)^{1/p}.$$

Since $\widetilde{X}^\varepsilon$ assumes values between X^ε and Z^ε, it has bounded moments by [25, Corollary 5.4] and Lemma C.1. Hence, applying Lemma C.3 we have

$$|J_2^\varepsilon| \le C\left(\delta\varepsilon^{-2} + \delta^{1/2}\varepsilon^{-1}\right),$$

which, since $\delta = \varepsilon^\zeta$ with $\zeta > 2$, implies

$$\lim_{\varepsilon\to 0}|J_2^\varepsilon| = 0. \tag{C.4}$$

Finally, Lemma C.4 and the weak convergence of the invariant measure φ^ε to φ^0 imply

$$\lim_{\varepsilon\to 0}\widetilde{\mathcal{M}}_\varepsilon = \mathcal{M}_0,$$

which, together with (C.3), (C.4) implies that $I_1^\varepsilon(T) \to 0$ for $\varepsilon \to 0$, which implies the desired result. $\square$

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. Papavasiliou, 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.