Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions

Assyr Abdulle [*] Grigorios A. Pavliotis [†] Andrea Zanoni [*]

Abstract

We propose a novel method for drift estimation of multiscale diffusion processes when a sequence of discrete observations is given. For the Langevin dynamics in a two-scale potential, our approach relies on the eigenvalues and the eigenfunctions of the homogenized dynamics. Our first estimator is derived from a martingale estimating function of the generator of the homogenized diffusion process. However, the unbiasedness of the estimator depends on the rate with which the observations are sampled. We therefore introduce a second estimator which relies also on filtering the data and we prove that it is asymptotically unbiased independently of the sampling rate. A series of numerical experiments illustrate the reliability and efficiency of our different estimators.

AMS subject classifications. 62F15, 65C30, 62M05, 74Q10, 35B27, 60J60, 62M05, 76M50.

Key words. Langevin dynamics, diffusion process, homogenization, parameter estimation, discrete observations, eigenvalue problem, filtering, martingale estimators.

1 Introduction

Learning models from data is a problem of fundamental importance in modern applied mathematics. The abundance of data in many application areas such as molecular dynamics, atmosphere/ocean science make it possible to develop physics-informed data driven methodologies for deriving models from data [27, 30, 31]. Naturally, most problems of interest are characterised by a very high dimensional state space and by the presence of many characteristic length and time scales. When it is possible to decompose the state space into the resolved and unresolved degrees of freedom, then one is usually interested in the derivation of a model for the resolved degrees of freedom, while treating the unresolved scales as noise. Clearly, these reduced models are stochastic, often described by stochastic differential equations (SDEs). The goal of this paper is to derive rigorous and systematic methodologies for learning coarse-grained models that describes accurately the dynamics at macroscopic length and time scales from noisy observations of the full, unresolved dynamics. We apply the proposed methodologies to simple models of fast/slow SDEs for which the theory of homogenization exists, that enables us to study the inference problem in a rigorous and systematic manner.

In many applications the available data is noisy, not equidistant and certainly not compatible with the coarse-grained model. The presence of observation noise and of the model-data mismatch renders the problem of learning macroscopic models from microscopic data highly ill-posed. Several examples from econometrics (market microstructure noise) [4] and molecular dynamics show that standard algorithms, e.g. maximum likelihood or quadratic variation for the diffusion coefficient,

---

[*]ANMC, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland, assyr.abdulle@epfl.ch, andrea.zanoni@epfl.ch

[†]Department of Mathematics, Imperial College London, London SW7 2AZ, UK, g.pavliotis@imperial.ac.uk

are *asymptotically biased* and they fail to estimate correctly the parameters in the coarse-grained model. We mention in particular the pioneering work on estimating the integrated stochastic volatility in the presence of market microstructure noise [4, 32]. In a series of earlier works this problem was studied using maximum likelihood techniques with subsampled data [23, 25], methodologies based on the method of moments [17, 19, 20], quadratic programming approaches [10] as well as Bayesian approaches [1, 3].

The main observation in [23, 25] is that when the maximum likelihood estimator (MLE) of the fast/slow system is evaluated at the full data, then the MLE becomes asymptotically biased; in fact, the original data is not compatible with the homogenized equation, and therefore data need to be preprocessed, for instance under the form of subsampling. On the other hand, when the MLE is evaluated at appropriately subsampled data, then it becomes asymptotically unbiased. Although this is an interesting theoretical observation (see also later developments in [29]), it does not lead to an efficient algorithm. The reason for this is that the performance of the estimator depends very sensitively on the choice of the sampling rate. In addition, the optimal sampling rate is not known and is strongly dependent on the problem into account. Furthermore, subsampling naturally leads to an increase in the variance, unless appropriate variance reduction methodologies are used.

In a recent work [2] we addressed the problem of lack of robustness of the MLE with subsampling algorithm by introducing an appropriate filtering methodology that leads to a stable and robust algorithm. In particular, rather than subsampling the original trajectory, we smoothed the data by applying an appropriate linear time-invariant filter from the exponential family and we modified the MLE by inserting the new filtered data. This new estimator was thus independent of the subsampling rate and also asymptotically unbiased and robust with respect to the parameters of the filter.

However, the assumption that the full path of the solution is observed is not realistic in most applications. In fact, in all real problems one can only obtain discrete measurements of the diffusion process. Hence, in this paper we focus on the problem of learning the coarse-grained homogenized model assuming that we are given discrete observations from the microscopic model. In this paper we use the martingale estimating functions that were introduced in [8], where the authors study drift estimation for discrete observations of one-scale processes and show that estimators based on the discretized continuous-version likelihood function can be strongly biased. They therefore propose martingale estimating functions obtained by adjusting the discretized continuous-version score function by its compensator which lead to unbiased estimators. Moreover, in [18] a different type of martingale estimating function, which is dependent on the eigenvalues and eigenfunctions of the generator of the stochastic process, is introduced and asymptotic unbiasedness and normality are proved. Furthermore, another inference methodology that uses spectral information is proposed in [11]. Their approach consists of inferring the drift and diffusion functions of a diffusion process by minimizing an objective function which measures how close the generator is to having a reference spectrum which is obtained from the time series through the construction of a discrete-time Markov chain. This idea has been further expanded in several directions in [12].

In this paper, we propose a new estimator for learning homogenised SDEs from noisy discrete data that is based on the martingale estimators that were introduced in [18]. The main idea is to consider the eigenvalues and eigenfunctions of the generator of the homogenized process. This new estimator is asymptotically unbiased only if the distance between two consecutive observations is not too small compared with the multiscale parameter describing the fastest scale, i.e., if data is compatible with the homogenized model. Therefore, in order to obtain unbiased approximations independently of the sampling rate with which the observations are obtained, we propose a second estimator which, in addition to the original observations, relies also on filtered data obtained following the filtering methodology presented in [2]. We observe that smoothing the original data makes observations compatible with the homogenized process independently of the rate with which they are sampled and hence this second estimator gives a black-box tool for parameter estimation.

1.1 Our main contributions

The main goal of this paper is to propose new algorithms based on martingale estimating functions and filtered data for which we can prove rigorously that they are asymptotically unbiased and not sensitive with respect to e.g. the sampling rate and the observation error. In particular, we combine two main ideas:

- the use of martingale estimating functions for discretely observed diffusion processes based on the eigenvalues and the eigenfunctions of the generator of the homogenized process, which was originally presented for one-scale problems in [18];

- the filtering methodology for smoothing the data in order to make them compatible with the homogenized model, which was introduced in [2].

We prove theoretically and observe numerically that the estimator without filtered data is asymptotically unbiased if:

- the observations are taken at the homogenized regime, i.e., the sampling rate is independent of the parameter measuring scale separation;

- the observations are taken at the multiscale regime, i.e., the sampling rate is dependent on the fastest scale, and the sampling rate is bigger than the multiscale parameter.

Moreover, we show that the estimator with filtered data corrects the bias caused by a sampling rate smaller than the multiscale parameter and therefore it is asymptotically unbiased independently of the sampling rate.

Outline. The rest of the paper is organized as follows. In Section 2 we present the Langevin dynamics and its corresponding homogenized equation and we introduce the two proposed estimators based on eigenvalues and eigenfunctions of the generator with and without filtered data. In Section 3 we present the main results of this work, i.e., the asymptotic unbiasedness of the two estimators, and in Section 4 we perform numerical experiments which validate the efficacy of our methods. Section 5 is devoted to the proof of the main results which are presented in Section 3. Finally, in the Appendix we show some technical results which are employed in the analysis and we explain some details about the implementation of the proposed methodology.

2 Problem setting

In this work we study the following class of multiscale diffusion processes. Consider the following two-scale SDE, observed over the time interval $[0, T]$

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

where $\varepsilon > 0$ describes the fast scale, $\alpha \in \mathbb{R}^M$ and $\sigma > 0$ are respectively the drift and diffusion coefficients and W_t is a standard one-dimensional Brownian motion. The functions $V \colon \mathbb{R} \to \mathbb{R}^M$ and $p \colon \mathbb{R} \to \mathbb{R}$ are the slow-scale and fast-scale parts of the potential and they are assumed to be smooth. Moreover, we also assume p to be periodic with period $L > 0$.

Remark 2.1. For clarity of the presentation, we focus our analysis on scalar multiscale diffusions with a finite number of parameters in the drift that have to be learned from data. Nevertheless, we remark that all the following theory can be easily generalized to the case of multidimensional diffusion processes in $\mathbb{R}^d$, for which we provide an example in Section 4.5.

The theory of homogenization (see e.g. [7, Chapter 3] or [26, Chapter 18]) guarantees the existence of the following homogenized SDE whose solution X_t^0 is the limit in law of the solutions X_t^ε of (2.1) as random variables in $\mathcal{C}^0([0,T];\mathbb{R})$

$$\mathrm{d}X_t^0 = -A \cdot V'(X_t^0)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t, \tag{2.2}$$

where $A = K\alpha$, $\Sigma = K\sigma$. The coefficient $0 < K < 1$ has the explicit formula

$$K = \int_0^L (1 + \Phi'(y))^2\,\mu(\mathrm{d}y) = \int_0^L (1 + \Phi'(y))\,\mu(\mathrm{d}y), \tag{2.3}$$

with

$$\mu(\mathrm{d}y) = \frac{1}{C_\sigma}e^{-p(y)/\sigma}\,\mathrm{d}y, \quad \text{where} \quad C_\sigma = \int_0^L e^{-p(y)/\sigma}\,\mathrm{d}y, \tag{2.4}$$

and where the function Φ is the unique solution with zero-mean with respect to the measure μ of the 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. In particular, for one-dimensional diffusion processes we have

$$\Phi'(y) = \frac{L}{\widehat{C}_\sigma}e^{p(y)/\sigma} - 1, \quad \text{where} \quad \widehat{C}_\sigma = \int_0^L e^{p(y)/\sigma}\,\mathrm{d}y,$$

which implies

$$K = \frac{L^2}{C_\sigma \widehat{C}_\sigma}.$$

Our goal is to derive estimators for the homogenized drift coefficient A based on multiscale data originating from (2.1). In this work we consider the same setting as [2], which is summarized by the following assumption.

Assumption 2.2. 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) $V \in \mathcal{C}^\infty(\mathbb{R};\mathbb{R}^N)$ and each component is polynomially bounded from above and bounded from below, and there exist $b_1, b_2 > 0$ such that

$$-b_1 + b_2 x^2 \le \alpha \cdot V'(x)x,$$

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

$$\|V'(x) - V'(y)\| \le C\,|x - y|.$$

Let us remark that, under Assumption 2.2, it has been proved in [25] that both processes (2.1) and (2.2) are geometrically ergodic and their invariant measure has a density with respect to the Lebesgue measure. In particular, let us denote by φ^ε and φ^0 the densities of the invariant measures of X_t^ε and X_t^0, respectively defined by

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

and

$$\varphi^0(x) = \frac{1}{C_{\varphi^0}}\exp\left(-\frac{1}{\Sigma}A \cdot V(x)\right), \quad \text{where} \quad C_{\varphi^0} = \int_\mathbb{R} \exp\left(-\frac{1}{\Sigma}A \cdot V(x)\right)\,\mathrm{d}x. \tag{2.6}$$

Remark 2.3. The value of the initial condition X_0^ε in the SDE (2.1) is important neither for the numerical experiments nor for the following analysis and can be chosen arbitrarily. In fact, the process X_t^ε is geometrically ergodic and therefore it converges to its invariant distribution with density φ^ε exponentially fast for any initial condition.

Drift estimation problem. Consider $N + 1$ uniformly distributed observations $0 = t_0 < t_1 < t_2 < \ldots, < t_N = T$, set $\Delta = t_n - t_{n-1}$ and let $(X_t^\varepsilon)_{t \in [0,T]}$ be a realization of the solution of (2.1). We then assume to know a sample $\{\widetilde{X}_n^\varepsilon\}_{n=0}^N$ of the realization where $\widetilde{X}_n^\varepsilon = X_{t_n}^\varepsilon$ and we aim to estimate the drift coefficient A of the homogenized equation (2.2). First, since we deal with discrete observations of stochastic processes, we employ martingale estimating functions based on eigenfunctions, which have already been studied for problems without a martingale structure in [18]. Second, by observing that if the time-step Δ is too small with respect to the multiscale parameter ε, then the data could be compatible with the full dynamics rather than with the coarse-grained model, we also adopt the filtering methodology presented in [2], which has been proved to be beneficial for correcting the behavior of the maximum likelihood estimator (MLE) in the setting of continuous observations.

2.1 Martingale estimating functions based on eigenfunctions

Let $\mathcal{A} \subset \mathbb{R}^M$ be the set of admissible drift coefficients for which Assumption 2.2(ii) is satisfied. To describe our methodology we consider the solution $X_t(a)$ of the homogenized process (2.2) with a generic parameter $a \in \mathcal{A}$ instead of the exact drift coefficient A:

$$\mathrm{d}X_t(a) = -a \cdot V'(X_t(a)) \, \mathrm{d}t + \sqrt{2\Sigma} \, \mathrm{d}W_t, \tag{2.7}$$

which, according to (2.6), has invariant measure

$$\varphi_a(x) = \frac{1}{C_{\varphi_a}} \exp\left(-\frac{1}{\Sigma} a \cdot V(x)\right), \quad \text{where} \quad C_{\varphi_a} = \int_{\mathbb{R}} \exp\left(-\frac{1}{\Sigma} a \cdot V(x)\right) \, \mathrm{d}x. \tag{2.8}$$

The generator $\mathcal{L}_a$ of (2.7) is defined for all $u \in C^2(\mathbb{R})$ as

$$\mathcal{L}_a u(x) = -a \cdot V'(x) u'(x) + \Sigma u''(x), \tag{2.9}$$

where the subscript denotes the dependence of the generator on the unknown drift coefficient a. From the well-known spectral theory of diffusion processes and under our assumptions on the potential V we deduce that $\mathcal{L}_a$ has a countable set of eigenvalues (see e.g. [15]). In particular, let $\{(\lambda_j(a), \phi_j(\cdot; a))\}_{j=0}^\infty$ be the sequence of eigenvalue-eigenfunction couples of the generator which solve the eigenvalue problem

$$\mathcal{L}_a \phi_j(x; a) = -\lambda_j(a) \phi_j(x; a), \tag{2.10}$$

which, due to (2.9), is equivalent to

$$\Sigma \phi_j''(x; a) - a \cdot V'(x) \phi_j'(x; a) + \lambda_j(a) \phi_j(x; a) = 0, \tag{2.11}$$

and where the eigenvalues satisfy $0 = \lambda_0(a) < \lambda_1(a) < \cdots < \lambda_j(a) \uparrow \infty$ and the eigenfunctions form an orthonormal basis for the weighted space $L^2(\varphi_a^0)$. We mention in passing that, by making a unitary transformation, the eigenvalue problem for the generator of the Langevin dynamics can be transformed to the standard Sturm-Liouville problem for Schrödinger operators [24, Chapter 4]. We now state a formula, which has been proved in [18] and will be fundamental in the rest of the paper

$$\mathbb{E}\left[\phi_j(X_{t_n}(a); a) | X_{t_{n-1}}(a) = x\right] = e^{-\lambda_j(a)\Delta} \phi_j(x; a), \tag{2.12}$$

where $\Delta = t_n - t_{n-1}$ is the constant distance between two consecutive observations. We now discuss how this eigenvalue problem can be used for parameter estimation. Let J be a positive integer and let $\{\beta_j(\cdot; a)\}_{j=1}^J$ be J arbitrary functions $\beta_j(\cdot; a) \colon \mathbb{R} \to \mathbb{R}^M$ possibly dependent on the parameter a, which satisfy Assumption 2.4(i)(ii) stated below, and define for $x, y, z \in \mathbb{R}$ the martingale estimating function

$$g_j(x, y, z; a) = \beta_j(z; a) \left(\phi_j(y; a) - e^{-\lambda_j(a)\Delta} \phi_j(x; a)\right). \tag{2.13}$$

Then, given a set of observations $\{\widetilde{X}_n^\varepsilon\}_{n=0}^N$, we consider the score function $\widehat{G}_{N,J}^\varepsilon \colon \mathcal{A} \to \mathbb{R}^M$ defined by

$$\widehat{G}_{N,J}^\varepsilon(a) = \frac{1}{\Delta} \sum_{n=0}^{N-1} \sum_{j=1}^J g_j(\widetilde{X}_n^\varepsilon, \widetilde{X}_{n+1}^\varepsilon, \widetilde{X}_n^\varepsilon; a). \tag{2.14}$$

5

This function can be seen as an approximation in terms of eigenfunctions of the true score function, i.e., the gradient of the log-likelihood function with respect to the unknown parameter. In fact, by a classical result for ergodic diffusion processes [24, Section 4.7], any function in the L^2 space weighted by the invariant measure can be written as an infinite linear combination of the eigenfunctions of the generator of the diffusion process.

The estimator $\widehat{A}_{N,J}^{\varepsilon}$. The first estimator we propose for the homogenized drift coefficient A is given by the solution $\widehat{A}_{N,J}^{\varepsilon}$ of the M-dimensional nonlinear system

$$\widehat{G}_{N,J}^{\varepsilon}(a) = 0. \tag{2.15}$$

In Algorithm 1 we summarize the main steps for computing the estimator $\widehat{A}_{N,J}^{\varepsilon}$ and further details about the implementation can be found in Appendix B. We finally introduce the following technical assumption which will be employed in the analysis.

Assumption 2.4. The following hold for all $a \in \mathcal{A}$ and for all $j = 1, \dots, J$:

(i) $\beta_j(z; a)$ is continuously differentiable with respect to a for all $z \in \mathbb{R}$;

(ii) all components of $\beta_j(\cdot; a)$, $\beta_j'(\cdot; a)$, $\dot{\beta}_j(\cdot; a)$, $\dot{\beta}_j'(\cdot; a)$ are polynomially bounded;

(iii) the slow-scale potential V is such that $\phi_j(\cdot; a)$, $\phi_j'(\cdot; a)$, $\phi_j''(\cdot; a)$, and all components of $\dot{\phi}_j(\cdot; a)$, $\dot{\phi}_j'(\cdot; a)$, $\dot{\phi}_j''(\cdot; a)$ are polynomially bounded;

where the dot denotes either the Jacobian matrix or the gradient with respect to a.

Remark 2.5. In [18] the authors propose a method to choose the functions $\{\beta_j(\cdot; a)\}_{j=1}^{J}$ in order to obtain optimality in the sense of [13]: the optimal set of functions in this sense can be seen as the projection in a suitable sense of the score function onto the set of martingale estimating functions obtained by varying the function $\{\beta_j(\cdot; a)\}_{j=1}^{J}$. For the class of diffusion processes for which the eigenfunctions are polynomials, the optimal estimating functions can be computed analytically. In fact, they are related to the moments of the transition density, which can be computed explicitly. Moreover, another procedure is to choose functions which depend only on the unknown parameter and which minimize the asymptotic variance. This approach is strongly related to the asymptotic optimality criterion considered by [16]. For further details on how to compute these functions we refer to [18]. Nevertheless, the theory we develop is valid for all functions which satisfy Assumption 2.4(i)(ii) and we observed in practice that choosing simple functions independent of the unknown parameter, e.g. $\beta_j(z) = z^k$ with $k \in \mathbb{N}$, is sufficient to obtain satisfactory estimations. We also remark that in one dimension we can characterize completely all diffusion processes whose generator has orthogonal polynomial eigenfunctions [5, Section 2.7].

2.2 The filtering approach

We now go back to our multiscale SDE (2.1) and, inspired by [2], we propose a second estimator for the homogenized drift coefficient by filtering the data. In particular, we modify $\widehat{A}_{N,J}^{\varepsilon}$ by filtering the observations and inserting the new data into the score function $\widehat{G}_{N,J}^{\varepsilon}$ in order to take into account the case when the step size Δ is too small with respect to the multiscale parameter ε. Let us consider the exponential kernel $k \colon \mathbb{R}^+ \to \mathbb{R}$ defined as

$$k(r) = e^{-r},$$

for which a rigorous theory has been developed in [2]. We then define the filtered observations $\{\widetilde{Z}_n^{\varepsilon}\}_{n=0}^{N}$ choosing $\widetilde{Z}_0^{\varepsilon} = 0$ and computing the weighted average for all $n = 1, \dots, N$

$$\widetilde{Z}_n^{\varepsilon} = \Delta \sum_{k=0}^{n-1} k(\Delta(n-k)) \widetilde{X}_k^{\varepsilon}, \tag{2.16}$$

Algorithm 1: Estimation of A without filtered data

Input: Observations $\left\{\widetilde{X}_n^\varepsilon\right\}_{n=0}^N$.
Distance between two consecutive observations Δ.
Number of eigenvalues and eigenfunctions J.
Functions $\{\beta_j(z;a)\}_{j=1}^J$.
Slow-scale potential V.
Diffusion coefficient Σ.

Output: Estimation $\widehat{A}_{N,J}^\varepsilon$ of A.

1: Consider the eigenvalue problem $\Sigma\phi_j''(x;a) - a \cdot V'(x)\phi_j'(x;a) + \lambda_j(a)\phi_j(x;a) = 0$.
2: Compute the first J eigenvalues $\{\lambda_j(a)\}_{j=1}^J$ and eigenfunctions $\{\phi_j(\cdot;a)\}_{j=1}^J$.
3: Construct the function $g_j(x,y,z;a) = \beta_j(z;a)\left(\phi_j(y;a) - e^{-\lambda_j(a)\Delta}\phi_j(x;a)\right)$.
4: Construct the score function $\widehat{G}_{N,J}^\varepsilon(a) = \frac{1}{\Delta}\sum_{n=0}^{N-1}\sum_{j=1}^J g_j(\widetilde{X}_n^\varepsilon, \widetilde{X}_{n+1}^\varepsilon, \widetilde{X}_n^\varepsilon; a)$.
5: Let $\widehat{A}_{N,J}^\varepsilon$ be the solution of the nonlinear system $\widehat{G}_{N,J}^\varepsilon(a) = 0$.

---

and we define the new score function as a modification of (2.14), i.e.,

$$\widetilde{G}_{N,J}^\varepsilon(a) = \frac{1}{\Delta}\sum_{n=0}^{N-1}\sum_{j=1}^J g_j(\widetilde{X}_n^\varepsilon, \widetilde{X}_{n+1}^\varepsilon, \widetilde{Z}_n^\varepsilon; a). \tag{2.17}$$

The estimator $\widetilde{A}_{N,J}^\varepsilon$. The second estimator $\widetilde{A}_{N,J}^\varepsilon$ is given by the solution of the M-dimensional nonlinear system

$$\widetilde{G}_{N,J}^\varepsilon(a) = 0. \tag{2.18}$$

The main steps to compute the estimator $\widetilde{A}_{N,J}^\varepsilon$ are highlighted in Algorithm 2 and additional details about the implementation can be found in Appendix B. Note that (2.16) can be rewritten as

$$\widetilde{Z}_n^\varepsilon = \Delta\sum_{k=0}^{n-1} e^{-\Delta(n-k)}\widetilde{X}_k^\varepsilon. \tag{2.19}$$

We introduce its continuous version Z_t^ε which will be employed in the analysis

$$Z_t^\varepsilon = \int_0^t e^{-(t-s)}X_s^\varepsilon \, \mathrm{d}s. \tag{2.20}$$

We remark that the joint process $(X_t^\varepsilon, Z_t^\varepsilon)$ satisfies the system of multiscale SDEs

$$\begin{aligned}
\mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\
\mathrm{d}Z_t^\varepsilon &= (X_t^\varepsilon - Z_t^\varepsilon)\,\mathrm{d}t,
\end{aligned} \tag{2.21}$$

and, using the theory of homogenization, when ε goes to zero it converges in law as a random variable in $\mathcal{C}^0([0,T];\mathbb{R}^2)$ to the two-dimensional process (X_t^0, Z_t^0), which solves

$$\begin{aligned}
\mathrm{d}X_t^0 &= -A \cdot V'(X_t^0)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t, \\
\mathrm{d}Z_t^0 &= \left(X_t^0 - Z_t^0\right)\,\mathrm{d}t.
\end{aligned}$$

Moreover, it has been proved in [2] that the two-dimensional processes $(X_t^\varepsilon, Z_t^\varepsilon)$ and (X_t^0, Z_t^0) are geometrically ergodic and their respective invariant measures have densities with respect to the Lebesgue measure denoted respectively by $\rho^\varepsilon = \rho^\varepsilon(x,z)$ and $\rho^0 = \rho^0(x,z)$. Let us finally remark that given discrete observations $\widetilde{X}_n^\varepsilon$ we can only compute $\widetilde{Z}_n^\varepsilon$, but the theory, which has to be employed for proving the convergence results, has been studied for the continuous-time process Z_t^ε.

Remark 2.6. The only difference in the construction of the estimators $\widehat{A}^\varepsilon_{N,J}$ and $\widetilde{A}^\varepsilon_{N,J}$ is the fact that the latter requires filtered data, which are obtained from discrete observations, and thus it is computationally more expensive. Therefore, when it is possible to use the estimator without filtered data, it is preferable to employ it.

Algorithm 2: Estimation of A with filtered data

    Input: Observations $\left\{ \widetilde{X}^\varepsilon_n \right\}^N_{n=0}$.

               Distance between two consecutive observations Δ.

               Number of eigenvalues and eigenfunctions J.

               Functions $\{\beta_j(z; a)\}^J_{j=1}$.

               Slow-scale potential V.

               Diffusion coefficient Σ.

    Output: Estimation $\widetilde{A}^\varepsilon_{N,J}$ of A.

1: Consider the eigenvalue problem $\Sigma\phi''_j(x; a) - a \cdot V'(x)\phi'_j(x; a) + \lambda_j(a)\phi_j(x; a) = 0$.

2: Compute the first J eigenvalues $\{\lambda_j(a)\}^J_{j=1}$ and eigenfunctions $\{\phi_j(\cdot; a)\}^J_{j=1}$.

3: *Compute the filtered data* $\left\{ \widetilde{Z}^\varepsilon_n \right\}^N_{n=0}$ *as* $\widetilde{Z}^\varepsilon_0 = 0$ *and* $\widetilde{Z}^\varepsilon_n = \Delta \sum^{n-1}_{k=0} e^{-\Delta(n-k)}\widetilde{X}^\varepsilon_k$.

4: Construct the function $g_j(x, y, z; a) = \beta_j(z; a)\left(\phi_j(y; a) - e^{-\lambda_j(a)\Delta}\phi_j(x; a)\right)$.

5: Construct the score function $\widetilde{G}^\varepsilon_{N,J}(a) = \frac{1}{\Delta}\sum^{N-1}_{n=0}\sum^J_{j=1} g_j(\widetilde{X}^\varepsilon_n, \widetilde{X}^\varepsilon_{n+1}, \widetilde{Z}^\varepsilon_n; a)$.

6: Let $\widetilde{A}^\varepsilon_{N,J}$ be the solution of the nonlinear system $\widetilde{G}^\varepsilon_{N,J}(a) = 0$.

---

3   Main results

In this section we present the main results of this work, i.e., the asymptotic unbiasedness of the proposed estimators. We first introduce a technical assumption which is a nondegeneracy hypothesis and which holds true in all nonpathological examples. We remark that this assumption is analogous to Condition 4.2(a) in [18].

Assumption 3.1. Let A be the homogenized drift coefficient of equation (2.2). Then the following hold

  (i) $\det\left(\sum^J_{j=1}\mathbb{E}^{\widetilde{\rho}^0}\left[\left(\beta_j(\widetilde{Z}^0_0; A) \otimes \nabla_a X_\Delta(A)\right)\phi'_j(X^0_\Delta; A)\right]\right) \neq 0$,

  (ii) $\det\left(\sum^J_{j=1}\mathbb{E}^{\varphi^0}\left[\left(\beta_j(X^0_0; A) \otimes \nabla_a X_\Delta(A)\right)\phi'_j(X^0_\Delta; A)\right]\right) \neq 0$,

  (iii) $\det\left(\sum^J_{j=1}\mathbb{E}^{\rho^0}\left[(\beta_j(Z^0_0; A) \otimes V'(X^0_0))\phi'_j(X^0_0; A)\right]\right) \neq 0$,

  (iv) $\det\left(\sum^J_{j=1}\mathbb{E}^{\varphi^0}\left[(\beta_j(X^0_0; A) \otimes V'(X^0_0))\phi'_j(X^0_0; A)\right]\right) \neq 0$.

We first state the two main theorems, whose proofs are the focus of Section 5.

Theorem 3.2. *Let J be a positive integer. Then there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ under Assumptions 2.2, 2.4, 3.1 and if Δ is independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$, an estimator $\widehat{A}^\varepsilon_{N,J}$ which solves the system $\widehat{G}^\varepsilon_{N,J}(\widehat{A}^\varepsilon_{N,J}) = 0$ exists with probability tending to one as $N \to \infty$. Moreover*

$$\lim_{\varepsilon \to 0}\lim_{N \to \infty}\widehat{A}^\varepsilon_{N,J} = A, \qquad \text{in probability,}$$

where A is the homogenized drift coefficient of equation (2.2).

Theorem 3.3. *Let J be a positive integer. Then there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ under Assumptions 2.2, 2.4, 3.1 and if Δ is independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta > 0$ and $\zeta \neq 1$, $\zeta \neq 2$, an estimator $\widetilde{A}^\varepsilon_{N,J}$ which solves the system $\widetilde{G}^\varepsilon_{N,J}(\widetilde{A}^\varepsilon_{N,J}) = 0$ exists with probability tending to one as $N \to \infty$. Moreover*

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} \widetilde{A}^\varepsilon_{N,J} = A, \qquad \text{in probability,}$$

where A is the homogenized drift coefficient of equation (2.2).

Remark 3.4. Notice that in both Theorem 3.2 and Theorem 3.3 the order of the limits is important and they cannot be interchanged. In fact, we first consider the large data limit, i.e., the number of observations N tends to infinity, and then we let the multiscale parameter ε vanish. Moreover, in Theorem 3.3 the values $\zeta = 1$ and $\zeta = 2$ are not allowed because of technicalities in the proof, but we observe numerically that the estimator works well also in these two particular cases.

These two theorems show that both estimators based on the multiscale data from (2.1) converge to the homogenized drift coefficient A of (2.2). Since the analysis is similar for the two cases, we will mainly focus on the second score function with filtered observations and at the end of each step we will state the differences with respect to the estimator without pre-processed data.

Remark 3.5. Since the main goal of this work is the estimation of the effective drift coefficient A, in the numerical experiments and in the following analysis we will always assume the effective diffusion coefficient Σ to be known. Nevertheless, we remark that our methodology can be slightly modified in order to take into account the estimation of the effective diffusion coefficient too. In fact, the parameter a can be replaced by the parameter $\theta = (a, s) \in \mathbb{R}^{M+1}$ where a stands for the drift and s stands for the diffusion, yielding nonlinear systems of dimension $M + 1$ corresponding to (2.15) and (2.18). The proofs of the asymptotic unbiasedness of the new estimators $\widehat{\theta}^\varepsilon_{N,J}$ and $\widetilde{\theta}^\varepsilon_{N,J}$ can be adjusted analogously.

3.1 A particular case

Before analysing the general framework, let us consider the simple case of the Ornstein-Uhlenbeck process, i.e. let the dimension of the parameter $N = 1$ and let $V(x) = x^2/2$. Then multiscale SDE (2.1) becomes

$$dX^\varepsilon_t = -\alpha X^\varepsilon_t \, dt - \frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon_t}{\varepsilon} \right) dt + \sqrt{2\sigma} \, dW_t,$$

and its homogenized version is

$$dX^0_t = -A X^0_t \, dt + \sqrt{2\Sigma} \, dW_t.$$

Letting $a \in \mathcal{A}$, then the eigenfunctions $\phi_j(\cdot; a)$ and the eigenvalues $\lambda_j(a)$ satisfy

$$\phi''_j(x; a) - \frac{a}{\Sigma} x \phi'(x) + \frac{\lambda(a)}{\Sigma} \phi(\cdot; a) = 0.$$

The solution of the eigenvalue problem can be computed explicitly (see [24, Section 4.4]); we have

$$\lambda_j(a) = ja,$$

and $\phi_j(\cdot; a)$ satisfies the recurrence relation

$$\phi_{j+1}(x; a) = x \phi_j(x; a) - j \frac{\Sigma}{a} \phi_{j-1}(x; a),$$

with $\phi_0(x; a) = 1$ and $\phi_1(x; a) = x$. It is also possible to prove by induction that

$$\phi'_j(x; a) = j \phi_{j-1}(x).$$

Let us consider the simplest case with only one eigenfunction, i.e. $J = 1$, and $\beta_1(z; a) = z$, which implies
$$g_1(x, y, z; a) = z\left(y - e^{-a\Delta}x\right).$$
Then the score functions (2.14) and (2.17) become
$$\widehat{G}_{N,1}^{\varepsilon}(a) = \frac{1}{\Delta}\sum_{n=0}^{N-1}\widetilde{X}_n^{\varepsilon}\left(\widetilde{X}_{n+1}^{\varepsilon} - e^{-\lambda(a)\Delta}\widetilde{X}_n^{\varepsilon}\right),$$
$$\widetilde{G}_{N,1}^{\varepsilon}(a) = \frac{1}{\Delta}\sum_{n=0}^{N-1}\widetilde{Z}_n^{\varepsilon}\left(\widetilde{X}_{n+1}^{\varepsilon} - e^{-\lambda(a)\Delta}\widetilde{X}_n^{\varepsilon}\right).$$

The solutions of the equations $\widehat{G}_{N,1}^{\varepsilon}(a) = 0$ and $\widetilde{G}_{N,1}^{\varepsilon}(a) = 0$ can be computed analytically and are given by
$$\widehat{A}_{N,1}^{\varepsilon} = -\frac{1}{\Delta}\log\left(\frac{\sum_{n=0}^{N-1}\widetilde{X}_n^{\varepsilon}\widetilde{X}_{n+1}^{\varepsilon}}{\sum_{n=0}^{N-1}(\widetilde{X}_n^{\varepsilon})^2}\right), \tag{3.1}$$
and
$$\widetilde{A}_{N,1}^{\varepsilon} = -\frac{1}{\Delta}\log\left(\frac{\sum_{n=0}^{N-1}\widetilde{Z}_n^{\varepsilon}\widetilde{X}_{n+1}^{\varepsilon}}{\sum_{n=0}^{N-1}\widetilde{Z}_n^{\varepsilon}\widetilde{X}_n^{\varepsilon}}\right). \tag{3.2}$$
Comparing these estimators with the discrete MLE defined in [25] without filtered data as
$$\widehat{\mathrm{MLE}}_{N,\Delta}^{\varepsilon} = -\frac{\sum_{n=0}^{N-1}\widetilde{X}_n^{\varepsilon}(\widetilde{X}_{n+1}^{\varepsilon} - \widetilde{X}_n^{\varepsilon})}{\Delta\sum_{n=0}^{N-1}(\widetilde{X}_n^{\varepsilon})^2},$$
and the discrete MLE with filtered data
$$\widetilde{\mathrm{MLE}}_{N,\Delta}^{\varepsilon} = -\frac{\sum_{n=0}^{N-1}\widetilde{Z}_n^{\varepsilon}(\widetilde{X}_{n+1}^{\varepsilon} - \widetilde{X}_n^{\varepsilon})}{\Delta\sum_{n=0}^{N-1}\widetilde{Z}_n^{\varepsilon}\widetilde{X}_n^{\varepsilon}},$$
we notice that they coincide in the limit as Δ vanishes. In the following theorems we show the asymptotic limit of the estimators. We do not provide a proof for these results since Theorem 3.6 and Theorem 3.8 are particular cases of Theorem 3.2 and Theorem 3.3 respectively, and Theorem 3.7 follows from the proof of Theorem 3.2 as highlighted in Remark 5.11.

Theorem 3.6. *Let Δ be independent of ε or $\Delta = \varepsilon^{\zeta}$ with $\zeta \in (0,1)$. Then, under Assumption 2.2, the estimator* (3.1) *satisfies*
$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\widehat{A}_{N,1}^{\varepsilon} = A, \quad \text{in probability},$$
where A is the drift coefficient of the homogenized equation (2.2).

Theorem 3.7. *Let Δ be independent of ε or $\Delta = \varepsilon^{\zeta}$ with $\zeta > 2$. Then, under Assumption 2.2, the estimator* (3.1) *satisfies*
$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\widehat{A}_{N,1}^{\varepsilon} = \alpha, \quad \text{in probability},$$
*where α is the drift coefficient of the homogenized equation* (2.1).

Theorem 3.8. *Let Δ be independent of ε or $\Delta = \varepsilon^{\zeta}$ with $\zeta \neq 1$, $\zeta \neq 2$. Then, under Assumption 2.2, the estimator* (3.2) *satisfies*
$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\widetilde{A}_{N,1}^{\varepsilon} = A, \quad \text{in probability},$$
where A is the drift coefficient of the homogenized equation (2.2).

Remark 3.9. Notice that it is possible to write different proofs for Theorems 3.6, 3.7 and 3.8, which take into account the specific form of the estimators, and thus show stronger results. In fact, if the distance Δ between two consecutive observations is independent of the multiscale parameter ε, then the convergences in the statements do not only hold in probability, but also almost surely. We expect that almost sure convergence can be proved for a larger class of equations, but are neither aware of related literature showing such a stronger result, nor have been able to prove it.
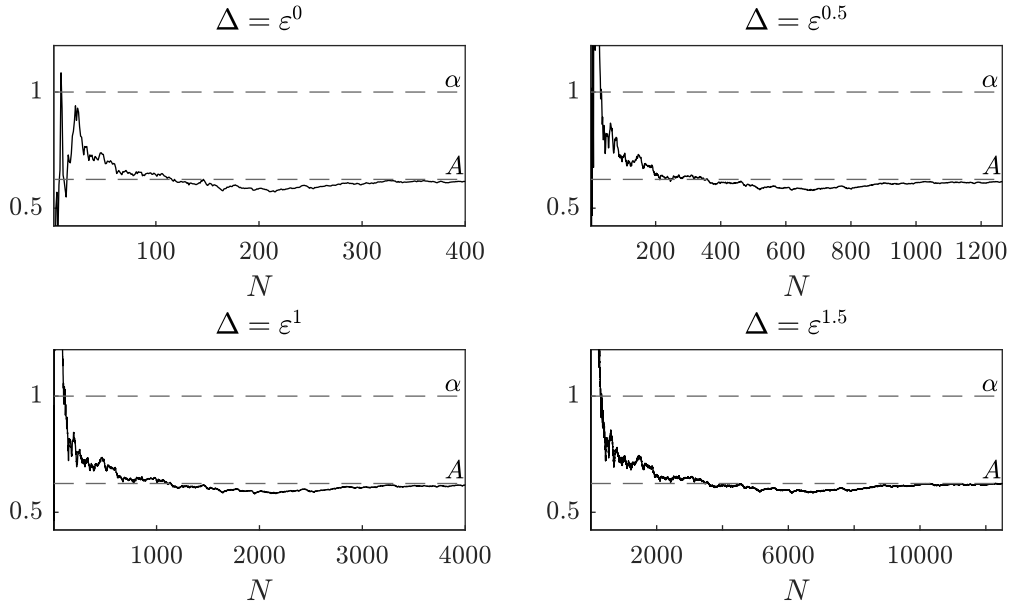
Figure 1: Sensitivity analysis with respect to the number N of observations for different values of $\Delta \leq 1$, for the estimator $\widetilde{A}^\varepsilon_{N,J}$ with $J = 1$.

4 Numerical experiments

In this section we present numerical experiments which confirm our theoretical results and show the power of the martingale estimating functions based on eigenfunctions and filtered data to correct the unbiasedness caused by discretization and the fact that we are using multiscale data to fit homogenized models. Moreover, we present a sensitivity analysis with respect to the number N of observations and the number J of eigenvalues and eigenfunctions taken into account. In the experiments that we present data is generated employing the Euler–Maruyama method with a fine time step h, in particular we set $h = \varepsilon^3$. Letting $\Delta, T > 0$, we generate data X^ε_t for $0 \leq t \leq T$ and we select a sequence of observations $\{\widetilde{X}^\varepsilon_n\}^N_{n=0}$, where $N = T/\Delta$ and $\widetilde{X}^\varepsilon_n = X^\varepsilon_{t_n}$ with $t_n = n\Delta$. In view of Remark 2.3 we do not require stationarity of the multiscale dynamics, hence we always set the initial condition to be $X^\varepsilon_0 = 0$. Notice that the time step h is only used to generate numerically the original data and has to be chosen sufficiently small in order to have a reliable approximation of the continuous path. However, the distance between two consecutive observations Δ is the rate at which we sample the data, which we assume to know, from the original trajectory. In order to compute the filtered data $\{\widetilde{Z}^\varepsilon_n\}^N_{n=1}$ we employ equation (2.19). We repeat this procedure for $M = 15$ different realizations of Brownian motion and we plot the average of the drift coefficients computed by the estimators. We finally remark that in order to compute our estimators we need the value of the diffusion coefficient Σ of the homogenized equation. In all the numerical experiments we compute it exactly using the formula for the coefficient K given by the theory of homogenization, but we also remark that its value could be estimated employing the subsampling technique presented in [25] or modifying the estimating function as explained in Remark 3.5.

4.1 Sensitivity analysis with respect to the number of observations

We consider the multiscale Ornstein–Uhlenbeck process, i.e. equation (2.1) with $V(x) = x^2/2$, and we take $p(y) = \cos(y)$, the multiscale parameter $\varepsilon = 0.1$, the drift coefficient $\alpha = 1$ and the diffusion coefficient $\sigma = 1$. We plot the results computed by the estimator $\widetilde{A}^\varepsilon_{N,J}$ with $J = 1$ and $\beta(x; a) = x$ and we then divide the analysis in two cases: Δ "small" and Δ "big".

Let us first consider Δ "small", i.e. $\Delta = \varepsilon^\zeta$ with $\zeta = 0, 0.5, 1, 1.5$, and take $T = 400$. In Figure
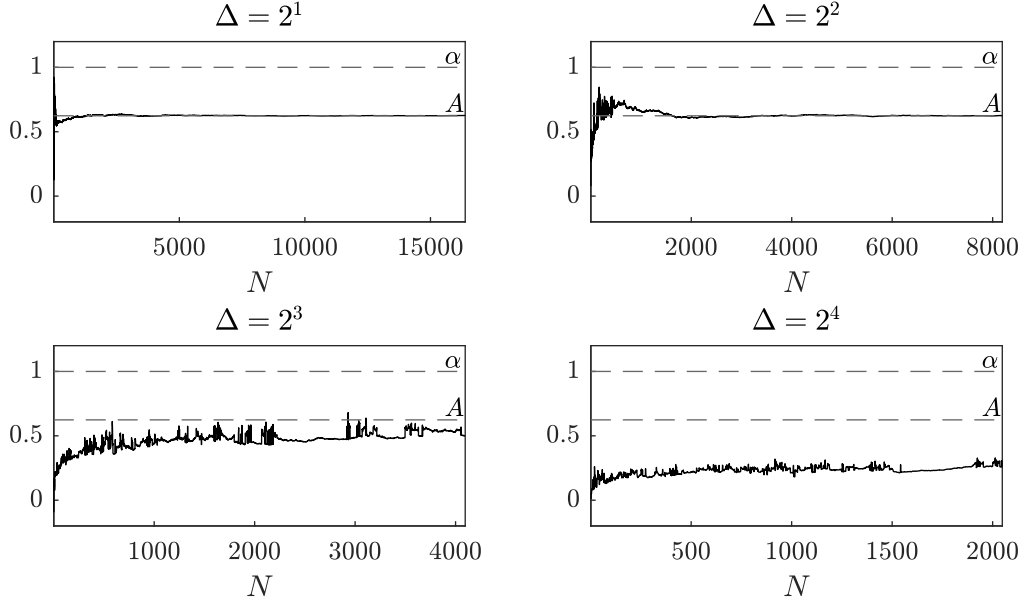
Figure 2: Sensitivity analysis with respect to the number N of observations for different values of $\Delta > 1$, for the estimator $\widetilde{A}^{\varepsilon}_{N,J}$ with $J = 1$.

1 we plot the results of the estimator as a function of the number of observations N. We remark that in this case the number of observations needed to reach convergence is strongly dependent and inversely proportional to the distance Δ between two consecutive observations. This means that in order to reach convergence we need the final time T to be sufficiently large independently of Δ. In fact, when the distance Δ is small, the discrete observations are a good approximation of the continuous trajectory and therefore what matters most is the length T of the original path rather than the number N of observations.

In order to study the case Δ "big", i.e. $\Delta > 1$, we set $\Delta = 2^{\zeta}$ with $\zeta = 1, 2, 3, 4$, and take $T = 2^{15}$. Figure 2 shows that in this case the number of observations needed to reach convergence is an increasing function of Δ. Therefore, in order to have a reliable approximation of the drift coefficient of the homogenized equation, the final time T has to be chosen depending on Δ. This is justified by the fact that, differently from the previous case, the discrete data is less correlated and therefore they do not well approximate the continuous trajectory. In particular, when the distance Δ between two consecutive observations is very large, then in practice we need a huge amount of data because a good approximation of the unknown coefficient is obtained only if the final time T is very large.
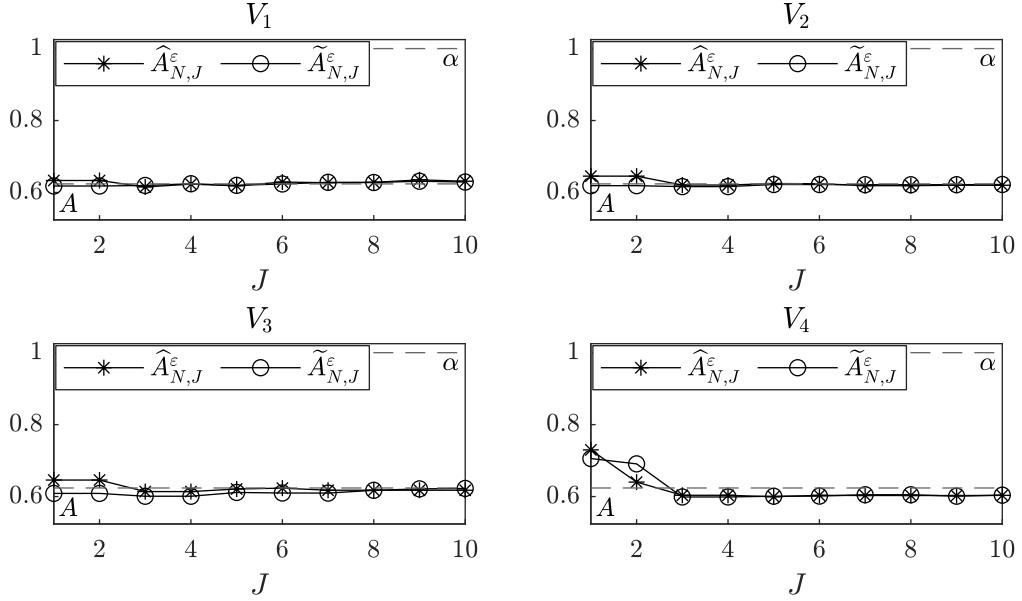
4.2 Sensitivity analysis with respect to the number of eigenvalues and eigenfunctions

Let us now consider equation (2.1) with four different slow-scale potentials

$$V_1(x) = \frac{x^2}{2}, \qquad V_2(x) = \frac{x^4}{4}, \qquad V_3(x) = \frac{x^6}{6}, \qquad V_4(x) = \frac{x^4}{4} - \frac{x^2}{2}. \qquad (4.1)$$

The other functions and parameters of the SDE are chosen as in the previous subsection, i.e. $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 1$ and $\varepsilon = 0.1$. Moreover, we set $\Delta = \varepsilon$ and $T = 500$ and we vary $J = 1, \ldots, 10$. The functions $\{\beta_j\}_{j=1}^{10}$ appearing in the estimating function are given by $\beta_j(x; a) = x$ for all $j = 1, \ldots, J$.

In Figure 3, where we plot the values computed by $\widehat{A}^{\varepsilon}_{N,J}$ and $\widetilde{A}^{\varepsilon}_{N,J}$, we observe that the number J of eigenvalues and eigenfunctions slightly improve the results, in particular for the fourth potential, but the estimation stabilizes when the number of eigenvalues J is still small, e.g. $J = 3$. Therefore,

Figure 3: Sensitivity analysis with respect to the number J of eigenvalues and eigenfunctions for different slow-scale potentials, for the estimators $\widehat{A}^\varepsilon_{N,J}$ and $\widetilde{A}^\varepsilon_{N,J}$.

in order to reduce the computational cost, it seems to be preferable not to take large values of J. This is related to how quickly the eigenvalues decay. The rigorous study of the accuracy of the spectral estimators as a function of the number of eigenvalues and eigenfunctions that we take into account will be investigated elsewhere.

4.3   Verification of the theoretical results

We consider the same setting as in the previous subsection, i.e. equation (2.1) with slow-scale potentials given by (4.1) and $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 1$ and $\varepsilon = 0.1$. Moreover, we set $J = 1$, $\beta(x; a) = x$ and $T = 500$ and we choose the distance between two successive observations to be $\Delta = \varepsilon^\zeta$ with $\zeta = 0, 0.1, 0.2, \dots, 2.5$.

In Figure 4 we compare our martingale estimator $\widehat{A}^\varepsilon_{N,J}$ without filtered data with the discrete maximum likelihood estimator denoted $\widehat{\mathrm{MLE}}^\varepsilon_{N,\Delta}$. The MLE does not provide good results for two reasons:

- if Δ is small, more precisely if $\Delta = \varepsilon^\zeta$ with $\zeta > 1$, sampling the data does not completely eliminate the fast-scale components of the original trajectory, therefore, since we are employing data generated by the multiscale model, the estimator is trying to approximate the drift coefficient α of the multiscale equation, rather than the one of the homogenized equation;

- if Δ is relatively big, in particular if $\Delta = \varepsilon^\zeta$ with $\zeta \in [0, 1)$, then we are taking into account only the slow-scale components of the original trajectory, but a bias is still introduced because we are discretizing an estimator which is usually used for continuous data.

Nevertheless, as observed in these numerical experiments and deeply investigated in [25], there exists an optimal value of Δ such that $\widehat{\mathrm{MLE}}^\varepsilon_{N,\Delta}$ works well, but this value is not known a priori and is strongly dependent on the problem, hence this technique is not robust. Figure 4 shows that the second issue, i.e., when Δ is relatively big, can be solved employing $\widehat{A}^\varepsilon_{N,J}$, an estimator for discrete observations, and that filtering the data is not needed as proved in Theorem 3.2.
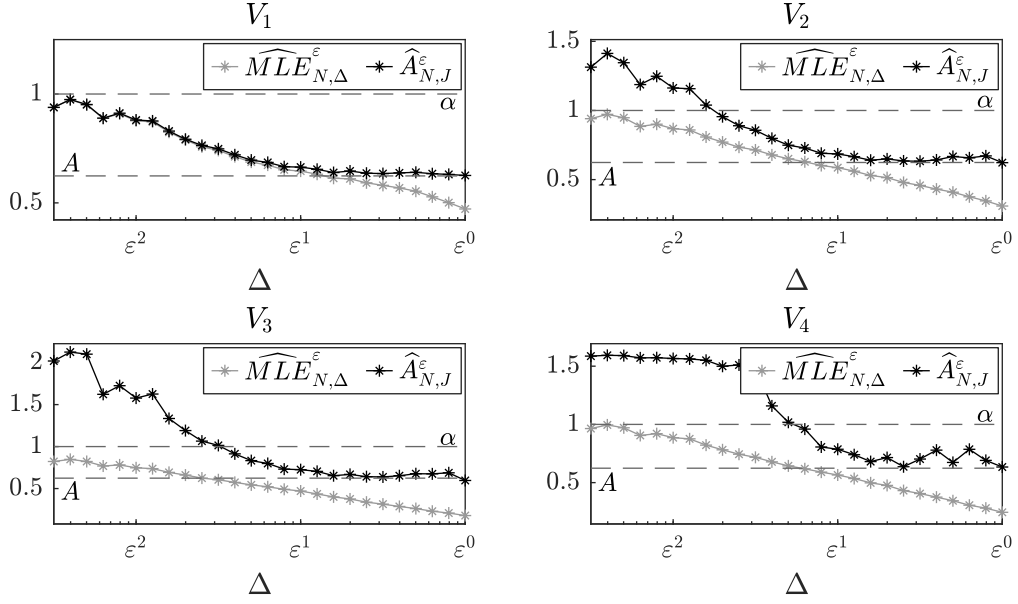
Figure 4: Comparison between the discrete maximum likelihood estimator $\widehat{\mathrm{MLE}}^\varepsilon_{N,\Delta}$ presented in [25] and our estimator $\widehat{A}^\varepsilon_{N,J}$ with $J = 1$ without filtered data as a function of the distance Δ between two successive observations for different slow-scale potentials.

Then, in order to solve also the first problem, in Figure 5 we compare $\widehat{A}^\varepsilon_{N,J}$ with our martingale estimator $\widetilde{A}^\varepsilon_{N,J}$ with filtered data. We observe that inserting filtered data in the estimator allows to disregard the fast-scale components of the original trajectory and to obtain good approximations of the drift coefficient A of the homogenized equation independently of Δ, as already shown in Theorem 3.3. In particular, we notice that the results still improve even for big values of Δ if we employ the estimator based on filtered data. Finally, as highlighted in Remark 5.11, we observe that the limiting value of the estimator $\widehat{A}^\varepsilon_{N,J}$ as the number of observations N goes to infinity and the multiscale parameter ε vanishes is strongly dependent on the problem and can not be computed theoretically. However, if we consider the slow-scale potential $V_1(x) = x^2/2$, i.e. the multiscale Ornstein–Uhlenbeck process, then the limit, as proved in Theorem 3.7, is the drift coefficient α of the multiscale equation.

4.4 Multidimensional drift coefficient

In this experiment we consider a multidimensional drift coefficient, in particular we set $N = 2$. We then consider the bistable potential, i.e.,

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

and the fast-scale potential $p(y) = \cos(y)$. We choose the exact drift coefficient of the multiscale equation (2.1) to be $\alpha = \left(1.2 \quad 0.7 \right)^\top$ and the diffusion coefficient to be $\sigma = 0.7$. We also set the number of eigenfunctions $J = 1$, the function $\beta(x;a) = \left(x^3 \quad x \right)^\top$, the distance between two consecutive observations $\Delta = 1$ and the final time $T = 1000$. We then compute the estimator $\widehat{A}^\varepsilon_{N,J}$ after $N = 100, 200, \ldots, 1000$ observations and in Figure 6 we plot the result of the experiment for the cases $\varepsilon = 0.1$ and $\varepsilon = 0.05$. Since we are analysing the case Δ independent of ε, filtering the data is not necessary and therefore we consider the estimator $\widehat{A}^\varepsilon_{N,J}$ which is computationally less expensive to compute.

We observe that the estimation is approaching the exact value A of the drift coefficient of the homogenized equation as the number of observations increases, until it starts oscillating around
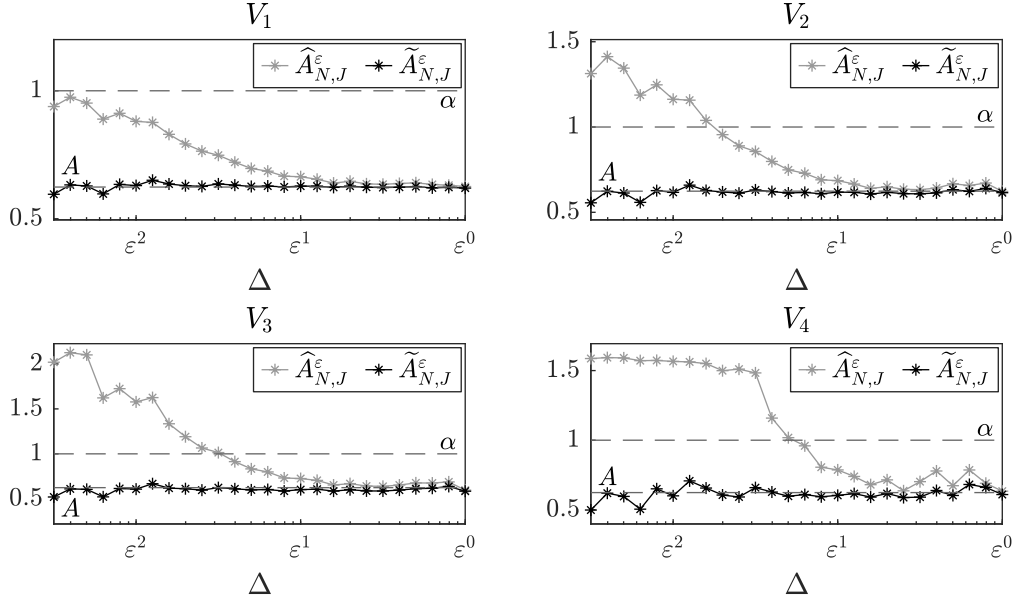
Figure 5: Comparison between our two estimators $\widehat{A}^\varepsilon_{N,J}$ without filtered data and $\widetilde{A}^\varepsilon_{N,J}$ with filtered data with $J = 1$ as a function of the distance Δ between two successive observations for different slow-scale potentials.

N	100	200	300	400	500	600	700	800	900	1000
$\varepsilon = 0.1$	0.742	0.395	0.215	0.201	0.093	0.036	0.011	0.027	0.034	0.028
$\varepsilon = 0.05$	0.086	0.031	0.019	0.031	0.018	0.049	0.081	0.085	0.055	0.053

Table 1: Absolute error $\widehat{e}^\varepsilon_N$ defined in (4.2) between the homogenized drift coefficient A and the estimator $\widehat{A}^\varepsilon_{N,J}$ with $J = 1$ for problem in Section 4.4.

the true value A. Moreover, we notice that the time needed to reach a neighborhood of A is smaller when the multiscale parameter ε is closer to its vanishing limit. In Table 1 we report the absolute error $\widehat{e}^\varepsilon_N$ defined as

$$\widehat{e}^\varepsilon_N = \left\| A - \widehat{A}^\varepsilon_{N,J} \right\|_2, \tag{4.2}$$

where $\|\cdot\|_2$ denotes the euclidean norm, varying the number of observations N for the two values of the multiscale parameter.

## 4.5 Multidimensional stochastic process: interacting particles

In this section we consider a system of d interacting particles in a two-scale potential, a problem with a wide range of applications which has been studied in [14]. For $t \in [0, T]$ and for all $i = 1, \ldots, d$, consider the system of SDEs

$$\mathrm{d}X_i^\varepsilon(t) = -\alpha X_i^\varepsilon(t)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_i^\varepsilon(t)}{\varepsilon}\right) - \frac{\theta}{d}\sum_{j=1}^d \left(X_i^\varepsilon(t) - X_j^\varepsilon(t)\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_i(t). \tag{4.3}$$

In this paper we fix the number of particles and study the performance of our estimators as ε vanishes. The very interesting problem of inference for mean field SDEs, obtained in the limit as $d \to \infty$, will be investigated elsewhere. It can be shown that $(X_1^\varepsilon, \ldots X_d^\varepsilon)$ converges in law as ε goes to zero to the solution $(X_1^0, \ldots, X_d^0)$ of the homogenized system

$$\mathrm{d}X_i^0(t) = -A X_i^0(t)\,\mathrm{d}t - \frac{\Theta}{d}\sum_{j=1}^d \left(X_i^0(t) - X_j^0(t)\right)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_i(t). \tag{4.4}$$
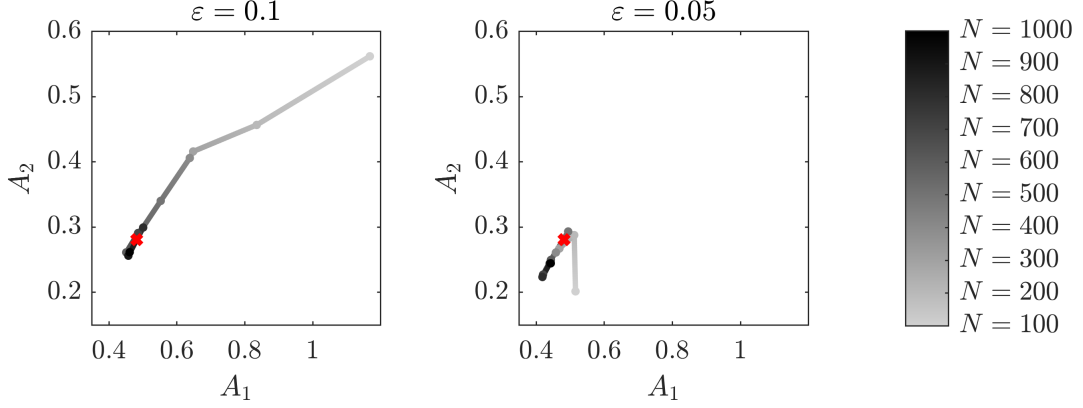
Figure 6: Evolution in time of the estimator $\widehat{A}^\varepsilon_{N,J}$ with $J = 1$ for a two-dimensional drift coefficient.
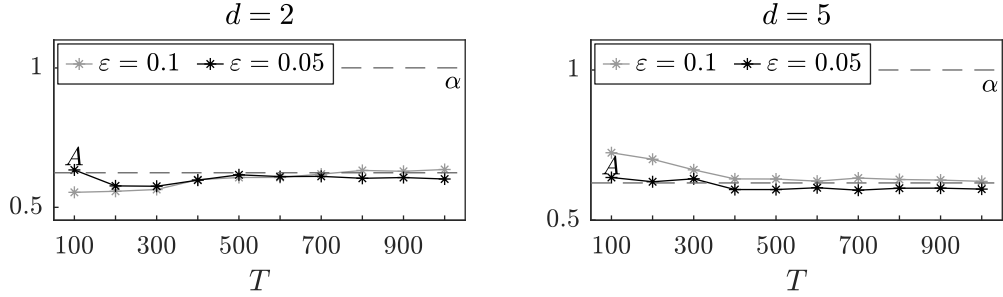
Figure 7: Evolution in time of the estimator $\widehat{A}^\varepsilon_{N,1}$ for a d-dimensional system of interacting particles.

where $\Theta = K\theta$ and K is defined in (2.3). Moreover, the first eigenvalue and eigenfunction of the generator of the homogenized system can be computed explicitly and they are given respectively by

$$\phi_1(x_1,\ldots,x_d) = \sum_{i=1}^d x_i \qquad \text{and} \qquad \lambda_1 = A.$$

Hence, letting $\Delta > 0$ independent of ε, given a sequence of observations $((\widetilde{X}^\varepsilon_1)_n, \ldots (\widetilde{X}^\varepsilon_d)_n)_{n=0}^N$, we can express the estimators analytically

$$\widehat{A}^\varepsilon_{N,1} = -\frac{1}{\Delta} \log\left(\frac{\sum_{n=0}^{N-1} \left(\sum_{i=1}^d (\widetilde{X}^\varepsilon_i)_n\right)\left(\sum_{i=1}^d (\widetilde{X}^\varepsilon_i)_{n+1}\right)}{\sum_{n=0}^{N-1} \left(\sum_{i=1}^d (\widetilde{X}^\varepsilon_i)_n\right)^2} \right),$$

$$\widetilde{A}^\varepsilon_{N,1} = -\frac{1}{\Delta} \log\left(\frac{\sum_{n=0}^{N-1} \left(\sum_{i=1}^d (\widetilde{Z}^\varepsilon_i)_n\right)\left(\sum_{i=1}^d (\widetilde{X}^\varepsilon_i)_{n+1}\right)}{\sum_{n=0}^{N-1} \left(\sum_{i=1}^d (\widetilde{Z}^\varepsilon_i)_n\right)\left(\sum_{i=1}^d (\widetilde{X}^\varepsilon_i)_n\right)} \right).$$

Let us now set $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 1$ and $\theta = 1$. We then simulate system (4.3) for different final times $T = 100, 200, \ldots, 1000$ and approximate the drift coefficient A of the homogenized system (4.4) for $d = 2$ and $d = 5$. In Figure 7 and Figure 8 we plot the results respectively of the estimators $\widehat{A}^\varepsilon_{N,J}$ with $\Delta = 1$ and $\widetilde{A}^\varepsilon_{N,J}$ with $\Delta = \varepsilon$ for two different values of $\varepsilon = 0.1, 0.05$. As expected, we observe that our estimator provides a better approximation of the unknown coefficient A when the time T increases and that this value stabilizes after approximately $T = 500$.

16
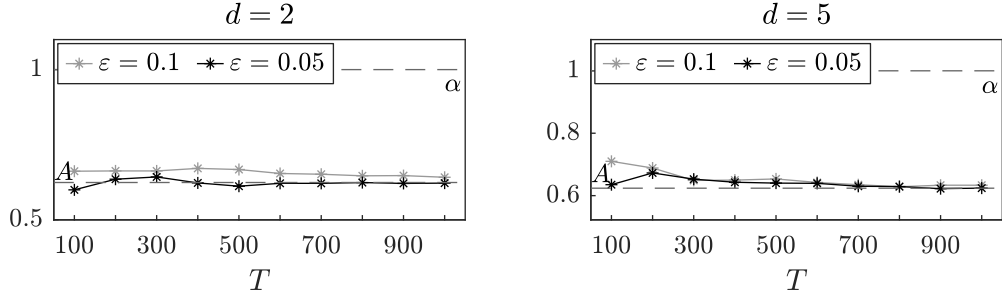
Figure 8: Evolution in time of the estimator $\widetilde{A}^{\varepsilon}_{N,1}$ for a d-dimensional system of interacting particles.

5 Asymptotic unbiasedness

In this section we prove our main results. The plan of the proof is the following:

- we first study the limiting behaviour of the score functions $\widehat{G}^{\varepsilon}_{N,J}$ and $\widetilde{G}^{\varepsilon}_{N,J}$ defined in (2.14) and (2.17) as the number of observations N goes to infinity, i.e., as the final time T tends to infinity;

- we then show the continuity of the limit of the score functions obtained in the previous step and we compute their limits as the multiscale parameter ε vanishes (Section 5.1);

- we finally prove our main results, i.e., the asymptotic unbiasedness of the drift estimators (Section 5.2).

We first define the Jacobian matrix of the function g_j introduced in (2.13) with respect to a:

$$
\begin{aligned}
h_j(x,y,z;a) = \dot{\beta}_j(z;a)\left(\phi_j(y,a) - e^{-\lambda_j(a)\Delta}\phi_j(x;a)\right) \\
+ \beta_j(z;a) \otimes \left(\dot{\phi}_j(y;a) - e^{-\lambda_j(a)\Delta}\left(\dot{\phi}_j(x;a) - \Delta\dot{\lambda}_j(a)\phi_j(x,a)\right)\right),
\end{aligned}
\tag{5.1}
$$

which will be employed in the following and where $\otimes$ denotes the outer product in $\mathbb{R}^M$ and the dot denotes either the Jacobian matrix or the gradient with respect to a, e.g. $h_j = \dot{g}_j$. Then note that, under Assumption 2.2, due to ergodicity and stationarity and by [8, Lemma 3.1] we have

$$
\lim_{N\to\infty}\frac{1}{N}\widehat{G}^{\varepsilon}_{N,J}(a) = \frac{1}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\varphi^{\varepsilon}}\left[g_j\left(X^{\varepsilon}_0, X^{\varepsilon}_{\Delta}, X^{\varepsilon}_0; a\right)\right] =: \widehat{\mathcal{G}}_J(\varepsilon, a),
$$

and

$$
\lim_{N\to\infty}\frac{1}{N}\widetilde{G}^{\varepsilon}_{N,J}(a) = \frac{1}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\rho}^{\varepsilon}}\left[g_j\left(X^{\varepsilon}_0, X^{\varepsilon}_{\Delta}, \widetilde{Z}^{\varepsilon}_0; a\right)\right] =: \widetilde{\mathcal{G}}_J(\varepsilon, a),
\tag{5.2}
$$

where $\mathbb{E}^{\varphi^{\varepsilon}}$ and $\mathbb{E}^{\widetilde{\rho}^{\varepsilon}}$ denotes respectively that X^{ε}_0 and $(X^{\varepsilon}_0, \widetilde{Z}^{\varepsilon}_0)$ are distributed according to their invariant distribution. We remark that the invariant distribution $\widetilde{\rho}^{\varepsilon}$ exists due to Lemma A.2. By equation (5.1) the Jacobian matrices of $\widehat{\mathcal{G}}_J(\varepsilon, a)$ and $\widetilde{\mathcal{G}}_J(\varepsilon, a)$ with respect to a are given by

$$
\widehat{\mathcal{H}}_J(\varepsilon, a) := \frac{\partial}{\partial a}\widehat{\mathcal{G}}_J(\varepsilon, a) = \frac{1}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\varphi^{\varepsilon}}\left[h_j\left(X^{\varepsilon}_0, X^{\varepsilon}_{\Delta}, X^{\varepsilon}_0; a\right)\right],
$$

and

$$
\widetilde{\mathcal{H}}_J(\varepsilon, a) := \frac{\partial}{\partial a}\widetilde{\mathcal{G}}_J(\varepsilon, a) = \frac{1}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\rho}^{\varepsilon}}\left[h_j\left(X^{\varepsilon}_0, X^{\varepsilon}_{\Delta}, \widetilde{Z}^{\varepsilon}_0; a\right)\right].
\tag{5.3}
$$

17

5.1 Continuity of the limit of the score function

In this section, we first prove the continuity of the functions $\widehat{\mathcal{G}}_J, \widetilde{\mathcal{G}}_J : (0,\infty) \times \mathcal{A} \to \mathbb{R}^M$ and $\widehat{\mathcal{H}}_J, \widetilde{\mathcal{H}}_J, : (0,\infty) \times \mathcal{A} \to \mathbb{R}^{M \times M}$. We then study the limit of these functions for $\varepsilon \to 0$. As the proof for the filtered and the non-filtered are similar, we will concentrate on the filtered case and comment on the non-filtered case. Before entering into the proof, we give two preliminary technical lemmas which will be used repeatedly and whose proof can be found respectively in Appendix A.1 and Appendix A.3.

Lemma 5.1. *Let* $\widetilde{Z}^\varepsilon$ *be defined in* (2.16) *and distributed according to the invariant measure* $\widetilde{\rho}^\varepsilon$ *of the process* $(\widetilde{X}_n, \widetilde{Z}_n)$. *Then for any* $p \geq 1$ *there exists a constant* $C > 0$ *uniform in* ε *such that*

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

Lemma 5.2. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a* $\mathcal{C}^\infty(\mathbb{R})$ *function which is polynomially bounded along with all its derivatives. Then*

$$f(X_\Delta^\varepsilon) = f(X_0^\varepsilon) - A \cdot V'(X_0^\varepsilon) f'(X_0^\varepsilon) \Delta + \Sigma f''(X_0^\varepsilon) \Delta + \sqrt{2\sigma} \int_0^\Delta f'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}W_t + R(\varepsilon, \Delta),$$

where $R(\varepsilon, \Delta)$ *satisfies for all* $p \geq 1$ *and for a constant* $C > 0$ *independent of* Δ *and* ε

$$\left(\mathbb{E}^{\varphi^\varepsilon} |R(\varepsilon, \Delta)|^p \right)^{1/p} \leq C(\varepsilon + \Delta^{3/2}).$$

We start here with a continuity result for the score function and its Jacobian matrix with respect to the unknown parameter.

Proposition 5.3. *Under Assumption 2.4, the functions* $\widetilde{\mathcal{G}}_J : (0,\infty) \times \mathcal{A} \to \mathbb{R}^M$ *and* $\widetilde{\mathcal{H}}_J, : (0,\infty) \times \mathcal{A} \to \mathbb{R}^{M \times M}$ *defined in* (5.2) *and* (5.3), *where* Δ *can be either independent of* ε *or* $\Delta = \varepsilon^\zeta$ *with* $\zeta > 0$, *are continuous.*

Proof. We only prove the statement for $\widetilde{\mathcal{G}}_J$, then the argument is similar for $\widetilde{\mathcal{H}}_J$. Letting $\varepsilon^* \in (0,\infty)$ and $a^* \in \mathcal{A}$, we want to show that

$$\lim_{(\varepsilon, a) \to (\varepsilon^*, a^*)} \left\| \widetilde{\mathcal{G}}_J(\varepsilon, a) - \widetilde{\mathcal{G}}_J(\varepsilon^*, a^*) \right\| = 0.$$

By the triangle inequality we have

$$\left\| \widetilde{\mathcal{G}}_J(\varepsilon, a) - \widetilde{\mathcal{G}}_J(\varepsilon^*, a^*) \right\| \leq \left\| \widetilde{\mathcal{G}}_J(\varepsilon, a) - \widetilde{\mathcal{G}}_J(\varepsilon, a^*) \right\| + \left\| \widetilde{\mathcal{G}}_J(\varepsilon, a^*) - \widetilde{\mathcal{G}}_J(\varepsilon^*, a^*) \right\| =: Q_1(\varepsilon, a) + Q_2(\varepsilon),$$

then we divide the proof in two steps and we show that the two terms vanish.
Step 1: $Q_1(\varepsilon, a) \to 0$ as $(\varepsilon, a) \to (\varepsilon^*, a^*)$.
Since β_j and ϕ_j are continuously differentiable with respect to a for all $j = 1, \ldots, J$ respectively due to Assumption 2.4 and Lemma A.4, then also g_j is continuously differentiable with respect to a. Therefore, by the mean value theorem for vector-valued functions we have

$$Q_1(\varepsilon, a) \leq \frac{1}{\Delta} \sum_{j=1}^J \left\| \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a) \right] - \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*) \right] \right\|$$

$$= \frac{1}{\Delta} \sum_{j=1}^J \left\| \int_0^1 \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[h_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^* + t(a - a^*)) \right] \mathrm{d}t \, (a - a^*) \right\|.$$

Then, letting $C > 0$ be a constant independent of ε, since β_j and ϕ_j are polynomially bounded still by Assumption 2.4 and X_0^ε, X_Δ^ε and $\widetilde{Z}_0^\varepsilon$ have bounded moments of any order by [25, Corollary 5.4] and Lemma 5.1, we obtain

$$Q_1(\varepsilon, a) \leq \frac{C}{\Delta} \left\| a - a^* \right\|,$$

18

which implies that $Q_1(\varepsilon, a)$ vanishes as (ε, a) goes to (ε^*, a^*) both if Δ is independent of ε and if $\Delta = \varepsilon^\xi$.

Step 2: $Q_2(\varepsilon) \to 0$ as $\varepsilon \to \varepsilon^*$.

If Δ is independent of ε, then we have

$$\lim_{\varepsilon \to \varepsilon^*} Q_2(\varepsilon) = \lim_{\varepsilon \to \varepsilon^*} \left\| \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*) \right] - \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^{\varepsilon^*}} \left[g_j(X_0^{\varepsilon^*}, X_\Delta^{\varepsilon^*}, \widetilde{Z}_0^{\varepsilon^*}; a^*) \right] \right\|$$

$$\leq \lim_{\varepsilon \to \varepsilon^*} \frac{1}{\Delta} \sum_{j=1}^{J} \left\| \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*) \right] - \mathbb{E}^{\widetilde{\rho}^{\varepsilon^*}} \left[g_j(X_0^{\varepsilon^*}, X_\Delta^{\varepsilon^*}, \widetilde{Z}_0^{\varepsilon^*}; a^*) \right] \right\|,$$

and the right hand side vanishes due to the continuity of g_j for all $j = 1, \ldots, J$ and the continuity of the solution of a stochastic differential equation with respect to a parameter (see [21, Theorem 2.8.1]). Let us now consider the case $\Delta = \varepsilon^\zeta$ with $\zeta > 0$ and let us assume, without loss of generality, that $\varepsilon > \varepsilon^*$. Denoting $\Delta^* = (\varepsilon^*)^\zeta$ and applying Itô's lemma we have for all $j = 1, \ldots, J$

$$\phi_j(X_\Delta^\varepsilon; a^*) = \phi_j(X_{\Delta^*}^\varepsilon; a^*) - \alpha \cdot \int_{\Delta^*}^{\Delta} V'(X_t^\varepsilon) \phi_j'(X_t^\varepsilon; a^*) \, \mathrm{d}t - \frac{1}{\varepsilon} \int_{\Delta^*}^{\Delta} \phi_j'(X_t^\varepsilon; a^*) p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right) \mathrm{d}t$$

$$+ \sigma \int_{\Delta^*}^{\Delta} \phi_j''(X_t^\varepsilon; a^*) \, \mathrm{d}t + \sqrt{2\sigma} \int_{\Delta^*}^{\Delta} \phi_j'(X_t^\varepsilon; a^*) \, \mathrm{d}W_t,$$

then we can write

$$\widetilde{\mathcal{G}}_J(\varepsilon, a^*) = \frac{1}{\Delta} \sum_{j=1}^{J} \left(\mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[\beta_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j(X_{\Delta^*}^\varepsilon; a^*) \right] - e^{-\lambda(a^*)\Delta} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[\beta_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) \right] \right) + R(\varepsilon),$$

where $R(\varepsilon)$ is given by

$$R(\varepsilon) = -\frac{1}{\Delta} \sum_{j=1}^{J} \int_{\Delta^*}^{\Delta} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[\beta_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j'(X_t^\varepsilon; a^*) \alpha \cdot V'(X_t^\varepsilon) \right] \mathrm{d}t$$

$$- \frac{1}{\varepsilon \Delta} \sum_{j=1}^{J} \int_{\Delta^*}^{\Delta} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[\beta_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j'(X_t^\varepsilon; a^*) p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right) \right] \mathrm{d}t$$

$$+ \frac{\sigma}{\Delta} \int_{\Delta^*}^{\Delta} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[\beta_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j''(X_t^\varepsilon; a^*) \right] \mathrm{d}t + \frac{\sqrt{2\sigma}}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[\int_{\Delta^*}^{\Delta} \beta_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j'(X_t^\varepsilon; a^*) \, \mathrm{d}W_t \right].$$

Let $C > 0$ be independent of ε and notice that since p' is bounded, $\beta_j, \phi_j', \phi_j'', V'$ are polynomially bounded and X_t^ε and $\widetilde{Z}_0^\varepsilon$ have bounded moments of any order by [25, Corollary 5.4] and Lemma 5.1, applying Hölder's inequality we obtain

$$|R(\varepsilon)| \leq \frac{C}{\Delta} \left(\|\alpha\| + \sigma + \frac{1}{\varepsilon} \right) (\Delta - \Delta^*) + \frac{C}{\Delta} \sqrt{2\sigma} (\Delta - \Delta^*)^{1/2}. \tag{5.4}$$

Therefore, by the continuity of the solution of a stochastic differential equation with respect to a parameter (see [22]) and due to the bound (5.4), we deduce that

$$\lim_{\varepsilon \to \varepsilon^*} \widetilde{\mathcal{G}}_J(\varepsilon, a^*) = \frac{1}{\Delta^*} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^{\varepsilon^*}} \left[\beta_j(\widetilde{Z}_0^{\varepsilon^*}; a^*) \left(\phi_j(X_{\Delta^*}^{\varepsilon^*}; a^*) - e^{-\lambda(a^*)\Delta^*} \phi_j(X_0^{\varepsilon^*}; a^*) \right) \right] = \widetilde{\mathcal{G}}_J(\varepsilon^*, a^*),$$

which implies that $Q_2(\varepsilon)$ vanishes as ε goes to ε^* and concludes the proof. $\qquad\square$

Remark 5.4. Notice that the proof of Proposition 5.3 can be repeated analogously for the functions $\widehat{\mathcal{G}}_J \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^M$ and $\widehat{\mathcal{H}}_J \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^{M \times M}$ without filtered data in order to prove their continuity.

Next we study the limit as ε vanishes and we divide the analysis in two cases. In particular, we consider Δ independent of ε and $\Delta = \varepsilon^\zeta$ with $\zeta > 0$. In the first case (Proposition 5.5) data is sampled at the homogenized regime ignoring the fact that the they are generated by a multiscale model, while in the second case (Proposition 5.7) the distance between two consecutive observations is proportional to the multiscale parameter and thus data is sampled at the multiscale regime preserving the multiscale structure of the full path.

Proposition 5.5. *Let the functions $\widetilde{\mathcal{G}}_J \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^M$ and $\widetilde{\mathcal{H}}_J, \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^{M \times M}$ be defined in (5.2) and (5.3) and let Δ be independent of ε. Under Assumption 2.4 and for any $a^* \in \mathcal{A}$ we have*

$$(i) \lim_{(\varepsilon, a) \to (0, a^*)} \widetilde{\mathcal{G}}_J(\varepsilon, a) = \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[g_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; a^* \right) \right],$$

$$(ii) \lim_{(\varepsilon, a) \to (0, a^*)} \widetilde{\mathcal{H}}_J(\varepsilon, a) = \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[h_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; a^* \right) \right].$$

Proof. We only prove the statement for $\widetilde{\mathcal{G}}_J$, then the argument is similar for $\widetilde{\mathcal{H}}_J$. By the triangle inequality we have

$$\left\| \widetilde{\mathcal{G}}_J(\varepsilon, a) - \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[g_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; a^* \right) \right] \right\| \leq Q_1(\varepsilon, a) + Q_2(\varepsilon),$$

where

$$Q_1(\varepsilon, a) = \left\| \widetilde{\mathcal{G}}_J(\varepsilon, a) - \widetilde{\mathcal{G}}_J(\varepsilon, a^*) \right\|,$$

which vanishes due to the first step of the proof of Proposition 5.3 and

$$Q_2(\varepsilon) = \left\| \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[g_j \left(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^* \right) \right] - \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[g_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; a^* \right) \right] \right\|.$$

Let us remark that the convergence in law of the joint process $\{(\widetilde{X}_n^\varepsilon, \widetilde{Z}_n^\varepsilon)\}_{n=0}^N$ to the joint process $\{(\widetilde{X}_n^0, \widetilde{Z}_n^0)\}_{n=0}^N$ by Lemma A.2 implies the convergence in law of the triple $(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon)$ to the triple $(X_0^0, X_\Delta^0, \widetilde{Z}_0^0)$ since $\widetilde{X}_0^\varepsilon = X_0^\varepsilon$, $\widetilde{X}_1^\varepsilon = X_\Delta^\varepsilon$ and $\widetilde{X}_0^0 = X_0^0$, $\widetilde{X}_1^0 = X_\Delta^0$. Therefore we have

$$\lim_{\varepsilon \to 0} Q_2(\varepsilon) \leq \lim_{\varepsilon \to 0} \frac{1}{\Delta} \sum_{j=1}^{J} \left\| \mathbb{E}^{\widetilde{\rho}^\varepsilon} \left[g_j \left(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^* \right) \right] - \mathbb{E}^{\widetilde{\rho}^0} \left[g_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; a^* \right) \right] \right\| = 0,$$

which implies the desired result. $\square$

Remark 5.6. Similar results to Proposition 5.3 and Proposition 5.5 can be shown for the estimator without filtered data. In particular we have that $\widehat{\mathcal{G}}_J(\varepsilon, a)$ and $\widehat{\mathcal{H}}_J(\varepsilon, a)$ are continuous in $(0, \infty) \times \mathcal{A}$ and

$$(i) \lim_{(\varepsilon, a) \to (0, a^*)} \widehat{\mathcal{G}}_J(\varepsilon, a) = \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[g_j \left(X_0^0, X_\Delta^0, X_0^0; a^* \right) \right],$$

$$(ii) \lim_{(\varepsilon, a) \to (0, a^*)} \widehat{\mathcal{H}}_J(\varepsilon, a) = \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[h_j \left(X_0^0, X_\Delta^0, X_0^0; a^* \right) \right].$$

Since the proof is analogous, we do not report here the details.

Proposition 5.7. *Let the functions $\widetilde{\mathcal{G}}_J \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^M$ and $\widetilde{\mathcal{H}}_J, \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^{M \times M}$ be defined in (5.2) and (5.3) and let $\Delta = \varepsilon^\zeta$ with $\zeta > 0$ and $\zeta \neq 1$, $\zeta \neq 2$. Under Assumption 2.4 and for any $a^* \in \mathcal{A}$ we have*

(i) $\lim_{(\varepsilon,a)\to(0,a^*)} \widetilde{\mathcal{G}}_J(\varepsilon,a) = \widetilde{\mathfrak{g}}_J^0(a^*)$, where

$$\widetilde{\mathfrak{g}}_J^0(a) := \sum_{j=1}^{J} \mathbb{E}^{\rho^0}\left[\beta_j(Z_0^0;a)\left(\mathcal{L}_A\phi_j(X_0^0;a) + \lambda_j(a)\phi_j(X_0^0;a)\right)\right],$$

(ii) $\lim_{(\varepsilon,a)\to(0,a^*)} \widetilde{\mathcal{H}}_J(\varepsilon,a) = \widetilde{\mathfrak{h}}_J^0(a^*)$, where

$$\widetilde{\mathfrak{h}}_J^0(a) := \sum_{j=1}^{J} \mathbb{E}^{\rho^0}\left[\dot{\beta}_j(Z_0^0;a)\left(\mathcal{L}_A\phi_j(X_0^0;a) + \lambda_j(a)\phi_j(X_0^0;a)\right)\right]$$

$$+ \sum_{j=1}^{J} \mathbb{E}^{\rho^0}\left[\beta_j(Z_0^0;a) \otimes \left(\mathcal{L}_A\dot{\phi}_j(X_0^0;a) + \lambda_j(a)\dot{\phi}_j(X_0^0;a)\right)\right]$$

$$+ \sum_{j=1}^{J} \mathbb{E}^{\rho^0}\left[\beta_j(Z_0^0;a)\phi_j(X_0^0;a)\right] \otimes \dot{\lambda}_j(a),$$

where the generator $\mathcal{L}_A$ is defined in (2.9).

Proof. We only prove the statement for $\widetilde{\mathcal{G}}_J$, then the argument is similar for $\widetilde{\mathcal{H}}_J$. By the triangle inequality we have

$$\left\|\widetilde{\mathcal{G}}_J(\varepsilon,a) - \widetilde{\mathfrak{g}}_J^0(a^*)\right\| \le \left\|\widetilde{\mathcal{G}}_J(\varepsilon,a) - \widetilde{\mathcal{G}}_J(\varepsilon,a^*)\right\| + \left\|\widetilde{\mathcal{G}}_J(\varepsilon,a^*) - \widetilde{\mathfrak{g}}_J^0(a^*)\right\| =: Q_1(\varepsilon,a) + Q_2(\varepsilon),$$

then we need to show that the two terms vanish and we distinguish two cases.
Case 1: $\zeta \in (0,1)$.
Applying Lemma 5.2 to the functions $\phi_j(\cdot;a^*)$ for all $j = 1,\dots,J$ and noting that

$$\mathbb{E}^{\widetilde{\rho}^\varepsilon}\left[\beta_j(\widetilde{Z}_0^\varepsilon;a^*)\int_0^\Delta \phi_j'(X_t^\varepsilon;a^*)(1+\Phi'(Y_t^\varepsilon))\,\mathrm{d}W_t\right] = 0,$$

since

$$M_s := \int_0^s \phi_j'(X_t^\varepsilon;a^*)(1+\Phi'(Y_t^\varepsilon))\,\mathrm{d}W_t$$

is a martingale with $M_0 = 0$, we have

$$\widetilde{\mathcal{G}}_J(\varepsilon,a^*) = \frac{1}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\rho}^\varepsilon}\left[\beta_j(\widetilde{Z}_0^\varepsilon;a^*)\left(\phi_j(X_\Delta^\varepsilon;a^*) - e^{-\lambda_j(a^*)\Delta}\phi_j(X_0^\varepsilon;a^*)\right)\right]$$

$$= \frac{1 - e^{-\lambda_j(a^*)\Delta}}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\rho}^\varepsilon}\left[\beta_j(\widetilde{Z}_0^\varepsilon;a^*)\phi_j(X_0^\varepsilon;a^*)\right] + \sum_{j=1}^{J}\frac{1}{\Delta}\mathbb{E}^{\widetilde{\rho}^\varepsilon}\left[\beta_j(\widetilde{Z}_0^\varepsilon;a^*)R(\varepsilon,\Delta)\right]$$

$$+ \sum_{j=1}^{J}\mathbb{E}^{\widetilde{\rho}^\varepsilon}\left[\beta_j(\widetilde{Z}_0^\varepsilon;a^*)\left(\Sigma\phi_j''(X_0^\varepsilon;a^*) - A\cdot V'(X_0^\varepsilon)\phi_j'(X_0^\varepsilon;a^*)\right)\right]$$

$$=: I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon,$$

where $R(\varepsilon,\Delta)$ satisfies for a constant $C > 0$ independent of ε and Δ and for all $p \ge 1$

$$\left(\mathbb{E}^{\widetilde{\rho}^\varepsilon}|R(\varepsilon,\Delta)|^p\right)^{1/p} \le C(\varepsilon + \Delta^{3/2}). \tag{5.5}$$

We now study the three terms separately. First, by Cauchy-Schwarz inequality, since $\beta_j(\cdot;a^*)$ is polynomially bounded, $\widetilde{Z}_0^\varepsilon$ has bounded moments of any order by Lemma 5.1 and due to (5.5) we obtain

$$\|I_2^\varepsilon\| \le C\left(\varepsilon\Delta^{-1} + \Delta^{1/2}\right). \tag{5.6}$$

Let us now focus on I_1^ε for which we have

$$I_1^\varepsilon = \frac{1 - e^{-\lambda_j(a^*)\Delta}}{\Delta} \sum_{j=1}^{J} \left(\mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*)\phi_j(X_0^\varepsilon; a^*) \right] + \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^\varepsilon; a^*) - \beta_j(Z_0^\varepsilon; a^*) \right) \phi_j(X_0^\varepsilon; a^*) \right] \right),$$

where Z_0^ε is distributed according to the invariant measure ρ^ε of the continuous process $(X_t^\varepsilon, Z_t^\varepsilon)$ and

$$\lim_{\varepsilon \to 0} \frac{1 - e^{-\lambda_j(a^*)\Delta}}{\Delta} = \lambda_j(a^*). \tag{5.7}$$

By the mean value theorem for vector-valued functions we have

$$\mathbb{E} \left[(\beta_j(\widetilde{Z}_0^\varepsilon; a^*) - \beta_j(Z_0^\varepsilon; a^*))\phi_j(X_0^\varepsilon; a^*) \right] = \mathbb{E} \left[\int_0^1 \beta_j'(Z_0^\varepsilon + t(\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon); a^*) \, dt \, (\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon)\phi_j(X_0^\varepsilon; a^*) \right],$$

and since $\beta_j'(\cdot; a^*), \phi_j(\cdot; a^*)$ are polynomially bounded, X_0^ε, Z_0^ε, $\widetilde{Z}_0^\varepsilon$ have bounded moments of any order respectively by [25, Corollary 5.4], [2, Lemma C.1] and Lemma 5.1 and applying Hölder's inequality and Corollary A.3 we obtain

$$\left\| \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^\varepsilon; a^*) - \beta_j(Z_0^\varepsilon; a^*) \right) \phi_j(X_0^\varepsilon; a^*) \right] \right\| \leq C \left(\Delta^{1/2} + \varepsilon \right). \tag{5.8}$$

Moreover, notice that by homogenization theory (see [2, Section 3.2]) the joint process $(X_0^\varepsilon, Z_0^\varepsilon)$ converges in law to the joint process (X_0^0, Z_0^0) and therefore

$$\lim_{\varepsilon \to 0} \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*)\phi_j(X_0^\varepsilon; a^*) \right] = \mathbb{E}^{\rho^0} \left[\beta_j(Z_0^0; a^*)\phi_j(X_0^0; a^*) \right],$$

which together with (5.7) and (5.8) yields

$$\lim_{\varepsilon \to 0} I_1^\varepsilon = \sum_{j=1}^{J} \lambda_j(a^*) \, \mathbb{E}^{\rho^0} \left[\beta_j(Z_0^0; a^*)\phi_j(X_0^0; a^*) \right]. \tag{5.9}$$

We now consider I_3^ε and we follow an argument similar to I_2^ε. We first have

$$\begin{aligned}
I_3^\varepsilon &= \sum_{j=1}^{J} \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \left(\Sigma\phi_j''(X_0^\varepsilon; a^*) - A \cdot V'(X_0^\varepsilon)\phi_j'(X_0^\varepsilon; a^*) \right) \right] \\
&\quad + \sum_{j=1}^{J} \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^\varepsilon; a^*) - \beta_j(Z_0^\varepsilon; a^*) \right) \left(\Sigma\phi_j''(X_0^\varepsilon; a^*) - A \cdot V'(X_0^\varepsilon)\phi_j'(X_0^\varepsilon; a^*) \right) \right] \\
&=: I_{3,1}^\varepsilon + I_{3,2}^\varepsilon,
\end{aligned}$$

where the first term in the right-hand side converges due to homogenization theory and the second one is bounded by

$$\left\| I_{3,2}^\varepsilon \right\| \leq C \left(\Delta^{1/2} + \varepsilon \right).$$

Therefore, we obtain

$$\lim_{\varepsilon \to 0} I_3^\varepsilon = \sum_{j=1}^{J} \mathbb{E}^{\rho^0} \left[\beta_j(Z_0^0; a^*) \left(\Sigma\phi_j''(X_0^0; a^*) - A \cdot V'(X_0^0)\phi_j'(X_0^0; a^*) \right) \right],$$

which together with (5.6) and (5.9) implies

$$\lim_{\varepsilon \to 0} \widetilde{\mathcal{G}}_J(\varepsilon, a^*) = \sum_{j=1}^{J} \mathbb{E}^{\rho^0} \left[\beta_j(Z_0^0; a) \left(\Sigma\phi_j''(X_0^0; a^*) - A \cdot V'(X_0^0)\phi_j'(X_0^0; a^*) + \lambda_j(a^*)\phi_j(X_0^0; a^*) \right) \right],$$

$$\tag{5.10}$$

which shows that $Q_2(\varepsilon)$ vanishes as ε goes to zero. Let us now consider $Q_1(\varepsilon, a)$. Following the first step of the proof of Proposition 5.3 we have

$$Q_1(\varepsilon, a) \leq \frac{1}{\Delta} \sum_{j=1}^{J} \left\| \mathbb{E}^{\widetilde{\rho}^{\varepsilon}} \left[g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, \widetilde{Z}_0^{\varepsilon}; a) \right] - \mathbb{E}^{\widetilde{\rho}^{\varepsilon}} \left[g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, \widetilde{Z}_0^{\varepsilon}; a^*) \right] \right\|$$

$$\leq \sum_{j=1}^{J} \left\| \frac{1}{\Delta} \mathbb{E}^{\widetilde{\rho}^{\varepsilon}} \left[h_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, \widetilde{Z}_0^{\varepsilon}; \widetilde{a}) \right] \right\| \|(a - a^*)\|,$$

where $\widetilde{a}$ assumes values in the line connecting a and a^*, and repeating the same computation as above we obtain

$$Q_1(\varepsilon, a) \leq C \|a - a^*\|,$$

which together with (5.10) gives the desired result.

Case 2: $\zeta \in (1, 2) \cup (2, \infty)$.

Let Z_0^{ε} be distributed according to the invariant measure ρ^{ε} of the continuous process $(X_t^{\varepsilon}, Z_t^{\varepsilon})$ and define

$$\widetilde{R}(\varepsilon, \Delta) := \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^{\varepsilon}} \left[g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, \widetilde{Z}_0^{\varepsilon}; a^*) \right] - \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\rho^{\varepsilon}} \left[g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, Z_0^{\varepsilon}; a^*) \right]$$

$$= \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^{\varepsilon}; a^*) - \beta_j(Z_0^{\varepsilon}; a^*) \right) \left(\phi_j(X_{\Delta}^{\varepsilon}; a^*) - e^{-\lambda_j(a^*)\Delta} \phi_j(X_0^{\varepsilon}; a^*) \right) \right].$$

Then we have

$$\widetilde{\mathcal{G}}_J(\varepsilon, a^*) = \sum_{j=1}^{J} \frac{1}{\Delta} \mathbb{E}^{\rho^{\varepsilon}} \left[g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, Z_0^{\varepsilon}; a^*) \right] + \widetilde{R}(\varepsilon, \Delta) =: \sum_{j=1}^{J} Q_j^{\varepsilon} + \widetilde{R}(\varepsilon, \Delta), \qquad (5.11)$$

and we first bound the remainder $\widetilde{R}(\varepsilon, \Delta)$. Applying Itô's lemma to the process X_t^{ε} with the functions $\phi_j(\cdot; a^*)$ for each $j = 1, \dots, J$ we have

$$\phi_j(X_{\Delta}^{\varepsilon}; a^*) = \phi_j(X_0^{\varepsilon}; a^*) - \int_0^{\Delta} \alpha \cdot V'(X_t^{\varepsilon}) \phi_j'(X_t^{\varepsilon}; a^*) \, dt - \int_0^{\Delta} \frac{1}{\varepsilon} p' \left(\frac{X_t^{\varepsilon}}{\varepsilon} \right) \phi_j'(X_t^{\varepsilon}; a^*) \, dt$$

$$+ \sigma \int_0^{\Delta} \phi_j''(X_t^{\varepsilon}; a^*) \, dt + \sqrt{2\sigma} \int_0^{\Delta} \phi_j'(X_t^{\varepsilon}; a^*) \, dW_t, \qquad (5.12)$$

and observing that

$$\mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^{\varepsilon}; a^*) - \beta_j(Z_0^{\varepsilon}; a^*) \right) \int_0^{\Delta} \phi_j'(X_t^{\varepsilon}; a^*) \, dW_t \right] = 0, \qquad (5.13)$$

since

$$M_s = \int_0^s \phi_j'(X_t^{\varepsilon}; a^*) \, dW_t$$

is a martingale with $M_0 = 0$, we obtain

$$\widetilde{R}(\varepsilon, \Delta) = \sum_{j=1}^{J} \frac{1 - e^{-\lambda_j(a^*)\Delta}}{\Delta} \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^{\varepsilon}; a^*) - \beta_j(Z_0^{\varepsilon}; a^*) \right) \phi_j(X_0^{\varepsilon}; a^*) \right]$$

$$+ \sum_{j=1}^{J} \frac{1}{\Delta} \int_0^{\Delta} \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^{\varepsilon}; a^*) - \beta_j(Z_0^{\varepsilon}; a^*) \right) \left(\sigma \phi_j''(X_t^{\varepsilon}; a^*) - \alpha \cdot V'(X_t^{\varepsilon}) \phi_j'(X_t^{\varepsilon}; a^*) \right) \right] dt$$

$$- \sum_{j=1}^{J} \frac{1}{\varepsilon \Delta} \int_0^{\Delta} \mathbb{E} \left[\left(\beta_j(\widetilde{Z}_0^{\varepsilon}; a^*) - \beta_j(Z_0^{\varepsilon}; a^*) \right) p' \left(\frac{X_t^{\varepsilon}}{\varepsilon} \right) \phi_j'(X_t^{\varepsilon}; a^*) \right] dt$$

$$=: \widetilde{R}_1(\varepsilon, \Delta) + \widetilde{R}_2(\varepsilon, \Delta) + \widetilde{R}_3(\varepsilon, \Delta).$$

By the mean value theorem for vector-valued functions we have

$$\mathbb{E}\left[(\beta_j(\widetilde{Z}_0^\varepsilon; a^*) - \beta_j(Z_0^\varepsilon; a^*))\phi_j(X_0^\varepsilon; a^*)\right] = \mathbb{E}\left[\int_0^1 \beta_j'(Z_0^\varepsilon + t(\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon); a^*)\,\mathrm{d}t\,(\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon)\phi_j(X_0^\varepsilon; a^*)\right],$$

and since $\beta_j'(\cdot; a^*), \phi_j(\cdot; a^*)$ are polynomially bounded, $X_0^\varepsilon, Z_0^\varepsilon, \widetilde{Z}_0^\varepsilon$ have bounded moments of any order respectively by [25, Corollary 5.4], [2, Lemma C.1] and Lemma 5.1 and applying Hölder's inequality we obtain

$$\left\|\widetilde{R}_1(\varepsilon, \Delta)\right\| \leq C\left(\mathbb{E}\left|\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon\right|^2\right)^{1/2}, \tag{5.14}$$

for a constant $C > 0$ independent of ε and Δ. We repeat a similar argument for $\widetilde{R}_2(\varepsilon, \Delta)$ and $\widetilde{R}_3(\varepsilon, \Delta)$ to get

$$\left\|\widetilde{R}_2(\varepsilon, \Delta)\right\| \leq C\left(\mathbb{E}\left|\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon\right|^2\right)^{1/2} \quad \text{and} \quad \left\|\widetilde{R}_3(\varepsilon, \Delta)\right\| \leq C\varepsilon^{-1}\left(\mathbb{E}\left|\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon\right|^2\right)^{1/2},$$

which together with (5.14) yield

$$\left\|\widetilde{R}(\varepsilon, \Delta)\right\| \leq C\left(\mathbb{E}\left|\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon\right|^2\right)^{1/2}\left(1 + \varepsilon^{-1}\right). \tag{5.15}$$

Moreover, applying Lemma 5.2 and proceeding similarly to the first part of the first case of the proof we have

$$\left\|\widetilde{R}(\varepsilon, \Delta)\right\| \leq C\left(\mathbb{E}\left|\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon\right|^2\right)^{1/2}\left(1 + \varepsilon\Delta^{-1} + \Delta^{1/2}\right),$$

which together with (5.15) and Corollary A.3 implies

$$\begin{aligned}
\left\|\widetilde{R}(\varepsilon, \Delta)\right\| &\leq C\left(\mathbb{E}\left|\widetilde{Z}_0^\varepsilon - Z_0^\varepsilon\right|^2\right)^{1/2}\left(1 + \min\{\varepsilon^{-1}, \varepsilon\Delta^{-1} + \Delta^{1/2}\}\right) \\
&\leq C\left(\Delta^{1/2} + \min\{\varepsilon, \varepsilon^{-1}\Delta\}\right)\left(1 + \min\{\varepsilon^{-1}, \varepsilon\Delta^{-1} + \Delta^{1/2}\}\right).
\end{aligned} \tag{5.16}$$

Let us now consider Q_j^ε. Replacing equation (5.12) into the definition of Q_j^ε in (5.11) and observing that similarly to (5.13) it holds

$$\mathbb{E}^{\rho^\varepsilon}\left[\beta_j(Z_0^\varepsilon; a^*)\int_0^\Delta \phi_j'(X_t^\varepsilon; a^*)\,\mathrm{d}W_t\right] = 0,$$

we obtain

$$\begin{aligned}
Q_j^\varepsilon ={}& \frac{1 - e^{-\lambda_j(a^*)}}{\Delta}\mathbb{E}^{\rho^\varepsilon}\left[\beta_j(Z_0^\varepsilon; a^*)\phi_j(X_0^\varepsilon; a^*)\right] \\
&- \frac{1}{\Delta}\left(\int_0^\Delta \mathbb{E}^{\rho^\varepsilon}\left[(\beta_j(Z_0^\varepsilon; a^*) \otimes V'(X_t^\varepsilon))\phi_j'(X_t^\varepsilon; a^*)\right]\,\mathrm{d}t\right)\alpha \\
&- \frac{1}{\Delta}\int_0^\Delta \mathbb{E}^{\rho^\varepsilon}\left[\beta_j(Z_0^\varepsilon; a^*)\frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\phi_j'(X_t^\varepsilon; a^*)\right]\,\mathrm{d}t + \frac{\sigma}{\Delta}\int_0^\Delta \mathbb{E}^{\rho^\varepsilon}\left[\beta_j(Z_0^\varepsilon; a^*)\phi_j''(X_t^\varepsilon; a^*)\right]\,\mathrm{d}t.
\end{aligned}$$

We rewrite $\beta_j(Z_0^\varepsilon; a^*)$ inside the integrals employing equation (2.21) and Itô's lemma

$$\beta_j(Z_0^\varepsilon; a^*) = \beta_j(Z_t^\varepsilon; a^*) - \int_0^t \beta_j'(Z_s^\varepsilon; a^*)(X_s^\varepsilon - Z_s^\varepsilon)\,\mathrm{d}s,$$

hence due to stationarity we have

$$Q_j^\varepsilon = Q_{j,1}^\varepsilon + Q_{j,2}^\varepsilon, \tag{5.17}$$

24

where

$$Q_{j,1}^\varepsilon = \frac{1 - e^{-\lambda_j(a^*)}}{\Delta} \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) \right] - \mathbb{E}^{\rho^\varepsilon} \left[(\beta_j(Z_0^\varepsilon; a^*) \otimes V'(X_0^\varepsilon)) \phi_j'(X_0^\varepsilon; a^*) \right] \alpha$$

$$- \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_0^\varepsilon}{\varepsilon} \right) \phi_j'(X_0^\varepsilon; a^*) \right] + \sigma \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \phi_j''(X_0^\varepsilon; a^*) \right]$$

and

$$Q_{j,2}^\varepsilon = \frac{1}{\Delta} \left(\int_0^\Delta \int_0^t \mathbb{E}^{\rho^\varepsilon} \left[(\beta_j'(Z_s^\varepsilon; a^*) \otimes V'(X_t^\varepsilon)) \phi_j'(X_t^\varepsilon; a^*) (X_s^\varepsilon - Z_s^\varepsilon) \right] \, \mathrm{d}s \, \mathrm{d}t \right) \alpha$$

$$+ \frac{1}{\Delta} \int_0^\Delta \int_0^t \mathbb{E}^{\rho^\varepsilon} \left[\beta_j'(Z_s^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_t^\varepsilon}{\varepsilon} \right) \phi_j'(X_t^\varepsilon; a^*) (X_s^\varepsilon - Z_s^\varepsilon) \right] \, \mathrm{d}s \, \mathrm{d}t$$

$$- \frac{\sigma}{\Delta} \int_0^\Delta \int_0^t \mathbb{E}^{\rho^\varepsilon} \left[\beta_j'(Z_s^\varepsilon; a^*) \phi_j''(X_t^\varepsilon; a^*) (X_s^\varepsilon - Z_s^\varepsilon) \right] \, \mathrm{d}s \, \mathrm{d}t.$$

Since $\phi_j'(\cdot; a^*), \phi_j''(\cdot; a^*)$ and $\beta_j'(\cdot; a^*)$ are polynomially bounded, p' is bounded and X_t^ε and Z_t^ε have bounded moments of any order respectively by [25, Corollary 5.4] and [2, Lemma C.1], $Q_{j,2}^\varepsilon$ is bounded by

$$\left\| Q_{j,2}^\varepsilon \right\| \le C \left(\Delta + \varepsilon^{-1} \Delta \right). \tag{5.18}$$

Let us now move to $Q_{j,1}^\varepsilon$ and let us define the functions

$$\eta^\varepsilon(x, z) := \frac{\rho^\varepsilon(x, z)}{\varphi^\varepsilon(x)} \qquad \text{and} \qquad \eta^0(x, z) := \frac{\rho^0(x, z)}{\varphi^0(x)},$$

where ρ^ε and ρ^0 are respectively the densities with respect to the Lebesgue measure of the invariant distributions of the joint processes $(X_t^\varepsilon, Z_t^\varepsilon)$ and (X_t^0, Z_t^0) and φ^ε and φ^0 are their marginals with respect to the first component. Integrating by parts we have

$$\mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_0^\varepsilon}{\varepsilon} \right) \phi_j'(X_0^\varepsilon; a^*) \right] = \int_\mathbb{R} \int_\mathbb{R} \beta_j(z; a^*) \frac{1}{\varepsilon} p' \left(\frac{x}{\varepsilon} \right) \phi_j'(x; a^*) \rho^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z$$

$$= -\sigma \int_\mathbb{R} \int_\mathbb{R} \frac{1}{C_{\varphi^\varepsilon}} \beta_j(z; a^*) \frac{\mathrm{d}}{\mathrm{d}x} \left(e^{-\frac{1}{\sigma} p \left(\frac{x}{\varepsilon} \right)} \right) \phi_j'(x; a^*) e^{-\frac{1}{\sigma} \alpha \cdot V(x)} \eta^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z$$

$$= \sigma \int_\mathbb{R} \int_\mathbb{R} \frac{1}{C_{\varphi^\varepsilon}} \beta_j(z; a^*) \frac{\partial}{\partial x} \left(\phi_j'(x; a^*) e^{-\frac{1}{\sigma} \alpha \cdot V(x)} \eta^\varepsilon(x, z) \right) e^{-\frac{1}{\sigma} p \left(\frac{x}{\varepsilon} \right)} \, \mathrm{d}x \, \mathrm{d}z,$$

which implies

$$\mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_0^\varepsilon}{\varepsilon} \right) \phi_j'(X_0^\varepsilon; a^*) \right] = \sigma \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \phi_j''(X_0^\varepsilon; a^*) \right]$$

$$- \mathbb{E}^{\rho^\varepsilon} \left[(\beta_j(Z_0^\varepsilon; a^*) \otimes V(X_0^\varepsilon)) \phi_j'(X_0^\varepsilon; a^*) \right] \alpha$$

$$+ \sigma \int_\mathbb{R} \int_\mathbb{R} \beta_j(z; a^*) \phi_j'(x; a^*) \varphi^\varepsilon(x) \frac{\partial}{\partial x} \eta^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z.$$

Employing the last equation in the proof of Lemma 3.5 in [2] with $\delta = 1$ and $f(x, z) = \beta_j(z; a^*) \phi_j'(x; a^*)$ we have

$$\sigma \int_\mathbb{R} \int_\mathbb{R} \beta_j(z; a^*) \phi_j'(x; a^*) \varphi^\varepsilon(x) \frac{\partial}{\partial x} \eta^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z = \mathbb{E}^{\rho^\varepsilon} \left[\beta_j'(Z_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) (X_0^\varepsilon - Z_0^\varepsilon) \right], \tag{5.19}$$

and thus we obtain

$$Q_{j,1}^\varepsilon = \frac{1 - e^{-\lambda_j(a^*)}}{\Delta} \mathbb{E}^{\rho^\varepsilon} \left[\beta_j(Z_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) \right] - \mathbb{E}^{\rho^\varepsilon} \left[\beta_j'(Z_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) (X_0^\varepsilon - Z_0^\varepsilon) \right].$$

Letting ε go to zero and due to homogenization theory, it follows

$$\lim_{\varepsilon \to 0} Q_{j,1}^\varepsilon = \lambda_j(a^*) \mathbb{E}^{\rho^0} \left[\beta_j(Z_0^0; a^*) \phi_j(X_0^0; a^*) \right] - \mathbb{E}^{\rho^0} \left[\beta_j'(Z_0^0; a^*) \phi_j(X_0^0; a^*) (X_0^0 - Z_0^0) \right],$$

25

then applying formula (5.19) for the homogenized equation, i.e. with $p(y) = 0$ and α and σ replaced by A and Σ, and integrating by parts we have

$$\mathbb{E}^{\rho^0}\left[\beta_j'(Z_0^0; a^*)\phi_j(X_0^0; a^*)(X_0^0 - Z_0^0)\right] = \Sigma \int_{\mathbb{R}}\int_{\mathbb{R}} \beta_j(z; a^*)\phi_j'(x; a^*)\varphi^0(x)\frac{\partial}{\partial x}\eta^0(x, z)\,\mathrm{d}x\,\mathrm{d}z$$

$$= -\Sigma \int_{\mathbb{R}}\int_{\mathbb{R}} \beta_j(z; a^*)\frac{\mathrm{d}}{\mathrm{d}x}\left(\phi_j'(x; a^*)\varphi^0(x)\right)\eta^0(x, z)\,\mathrm{d}x\,\mathrm{d}z$$

$$= \mathbb{E}^{\rho^0}\left[\beta_j(Z_0^0; a^*)\left(\Sigma\phi_j''(X_0^0; a^*) - A \cdot V'(X_0^0)\phi_j'(X_0^0; a^*)\right)\right].$$

Therefore, we obtain

$$\lim_{\varepsilon \to 0} Q_{j,1}^\varepsilon = \mathbb{E}^{\rho^0}\left[\beta_j(Z_0^0; a^*)\left(\Sigma\phi_j''(X_0^0; a^*) - A \cdot V'(X_0^0)\phi_j'(X_0^0; a^*) + \lambda_j(a^*)\phi_j(X_0^0; a^*)\right)\right],$$

which together with (5.11), (5.17) and bounds (5.16) and (5.18) implies that $Q_2(\varepsilon)$ vanishes as ε goes to zero. Finally, analogously to the first case we can show that also $Q_1(\varepsilon, a)$ vanishes, concluding the proof. $\qquad\square$

Remark 5.8. A similar result to Proposition 5.7 can be shown for the estimator without filtered data only if $\zeta \in (0, 1)$, i.e. the first case in the proof. In particular, we have

(i) $\lim_{(\varepsilon, a) \to (0, a^*)} \widehat{\mathcal{G}}_J(\varepsilon, a) = \widehat{\mathfrak{g}}_J^0(a^*)$, where

$$\widehat{\mathfrak{g}}_J^0(a) := \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\beta_j(X_0^0; a)\left(\mathcal{L}_A\phi_j(X_0^0; a) + \lambda_j(a)\phi_j(X_0^0; a)\right)\right],$$

(ii) $\lim_{(\varepsilon, a) \to (0, a^*)} \widehat{\mathcal{H}}_J(\varepsilon, a) = \widehat{\mathfrak{h}}_J^0(a^*)$, where

$$\widehat{\mathfrak{h}}_J^0(a) := \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\dot{\beta}_j(X_0^0; a)\left(\mathcal{L}_A\phi_j(X_0^0; a) + \lambda_j(a)\phi_j(X_0^0; a)\right)\right]$$

$$+ \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\beta_j(X_0^0; a) \otimes \left(\mathcal{L}_A\dot{\phi}_j(X_0^0; a) + \lambda_j(a)\dot{\phi}_j(X_0^0; a)\right)\right]$$

$$+ \mathbb{E}^{\varphi^0}\left[\beta_j(X_0^0; a)\phi_j(X_0^0; a)\right] \otimes \dot{\lambda}_j(a),$$

where the generator $\mathcal{L}_A$ is defined in (2.9). Since the proof is analogous, we do not report here the details. On the other hand, if $\zeta > 2$ we can show that

(i) $\lim_{(\varepsilon, a) \to (0, a^*)} \widehat{\mathcal{G}}_J(\varepsilon, a) = \mathfrak{g}_J^0(a^*)$, where

$$\mathfrak{g}_J^0(a) := \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\beta_j(X_0^0; a)\left(\sigma\phi_j''(X_0^0; a) - \alpha \cdot V'(X_0^0)\phi_j'(X_0^0; a) + \lambda_j(a)\phi_j(X_0^0; a)\right)\right], \quad (5.20)$$

(ii) $\lim_{(\varepsilon, a) \to (0, a^*)} \widehat{\mathcal{H}}_J(\varepsilon, a) = \mathfrak{h}_J^0(a^*)$, where

$$\mathfrak{h}_J^0(a) := \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\dot{\beta}_j(X_0^0; a)\left(\sigma\phi_j''(X_0^0; a) - \alpha \cdot V'(X_0^0)\phi_j'(X_0^0; a) + \lambda_j(a)\phi_j(X_0^0; a)\right)\right]$$

$$+ \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\beta_j(X_0^0; a) \otimes \left(\sigma\dot{\phi}_j''(X_0^0; a) - \alpha \cdot V'(X_0^0)\dot{\phi}_j'(X_0^0; a) + \lambda_j(a)\dot{\phi}_j(X_0^0; a)\right)\right]$$

$$+ \sum_{j=1}^J \mathbb{E}^{\varphi^0}\left[\beta_j(X_0^0; a)\phi_j(X_0^0; a)\right] \otimes \dot{\lambda}_j(a).$$

The proof is omitted since it is similar to the second case of the proof of Proposition 5.7.

5.2 Proof of the main results

Let us remark that we aim to prove the asymptotic unbiasedness of the proposed estimators, i.e., their convergence to the homogenized drift coefficient A as the the number of observations N tends to infinity and the multiscale parameter ε vanishes. Therefore, we study the limit of the score functions and their Jacobian matrices as $N \to \infty$ and $\varepsilon \to 0$ evaluated in the desired limit point A.

We first analyse the case Δ independent of ε and we consider the limit of Proposition 5.5 and Remark 5.6 evaluated in $a^* = A$. Then due to equation (2.12) we get

$$
\begin{aligned}
\frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[g_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; A \right) \right] &= \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[\beta_j(\widetilde{Z}_0^0; A) \left(\phi_j(X_\Delta^0; A) - e^{-\lambda_j(A)\Delta} \phi_j(X_0^0; A) \right) \right] \\
&= \frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[\beta_j(\widetilde{Z}_0^0; A) \left(\mathbb{E} \left[\phi_j(X_\Delta^0; A) \middle| (X_0^0, \widetilde{Z}_0^0) \right] - e^{-\lambda_j(A)\Delta} \phi_j(X_0^0; A) \right) \right] \\
&= 0,
\end{aligned}
\tag{5.21}
$$

and similarly we obtain

$$
\frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[g_j \left(X_0^0, X_\Delta^0, X_0^0; A \right) \right] = 0.
$$

On the other hand, if Δ is a power of ε we study the limit of Proposition 5.7 and Remark 5.8 evaluated in $a^* = A$ and by (2.10) we have

$$
\widetilde{\mathfrak{g}}_J^0(A) = 0 \qquad \text{and} \qquad \widehat{\mathfrak{g}}_J^0(A) = 0.
\tag{5.22}
$$

Moreover, differentiating equation (2.12) with respect to a, we get

$$
\begin{aligned}
\mathbb{E} \left[\dot{\phi}_j(X_{t_n}(a); a) \middle| X_{t_{n-1}}(a) = x \right] &= e^{-\lambda_j(a)\Delta} \dot{\phi}_j(x; a) - \dot{\lambda}_j(a)\Delta e^{-\lambda_j(a)\Delta} \phi_j(x; a) \\
&\quad - \mathbb{E} \left[\phi_j'(X_{t_n}(a); a) \nabla_a X_{t_n}(a) \middle| X_{t_{n-1}}(a) = x \right],
\end{aligned}
\tag{5.23}
$$

where the process $\nabla_a X_t(a)$ satisfies

$$
\mathrm{d} \left(\nabla_a X_t(a) \right) = -V'(X_t) \, \mathrm{d}t - a \cdot V''(X_t) \nabla_a X_t(a) \, \mathrm{d}t.
$$

Therefore, due to (2.12) and (5.23) we have

$$
\frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[h_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; A \right) \right] = - \sum_{j=1}^{J} \mathbb{E}^{\widetilde{\rho}^0} \left[\left(\beta_j(\widetilde{Z}_0^0; A) \otimes \nabla_a X_\Delta(A) \right) \phi_j'(X_\Delta^0; A) \right],
\tag{5.24}
$$

and

$$
\frac{1}{\Delta} \sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[h_j \left(X_0^0, X_\Delta^0, X_0^0; A \right) \right] = - \sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[\left(\beta_j(X_0^0; A) \otimes \nabla_a X_\Delta(A) \right) \phi_j'(X_\Delta^0; A) \right].
$$

Then due to Lemma A.4 we can differentiate the eigenvalue problem (2.11) with respect to a and deduce that

$$
\Sigma \dot{\phi}_j''(x; a) - a \cdot V'(x) \dot{\phi}_j'(x; a) + \lambda_j(a) \dot{\phi}_j(x; a) = V'(x) \phi_j'(x; a) - \dot{\lambda}_j \phi_j(x; a),
$$

where the dot denotes the gradient with respect to a and which together with (2.11) implies

$$
\widetilde{\mathfrak{h}}_J^0(A) = \sum_{j=1}^{J} \mathbb{E}^{\rho^0} \left[(\beta_j(Z_0^0; A) \otimes V'(X_0^0)) \phi_j'(X_0^0; A) \right],
\tag{5.25}
$$

and

$$\widehat{\mathfrak{h}}_J^0(A) = \sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[(\beta_j(X_0^0; A) \otimes V'(X_0^0)) \phi_j'(X_0^0; A) \right].$$

Before showing the main results, we need two auxiliary lemmas and since the proofs are similar we only write the details of the first one.

Lemma 5.9. *Under Assumption 2.4 and Assumption 3.1 there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists $\widetilde{\gamma} = \widetilde{\gamma}(\varepsilon)$ such that if Δ is independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta > 0$ and $\zeta \neq 1$, $\zeta \neq 2$*

$$\widetilde{\mathcal{G}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) = 0 \qquad and \qquad \det\left(\widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right) \neq 0.$$

Moreover

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

Proof. Let us first extend the functions $\widetilde{\mathcal{G}}_J$ and $\widetilde{\mathcal{H}}_J$ by continuity in $\varepsilon = 0$ with their limit given by Proposition 5.5 and Proposition 5.7 depending on Δ and note that due to (5.21) if Δ is independent of ε and (5.22) otherwise, we have

$$\widetilde{\mathcal{G}}_J(0, A) = 0.$$

Moreover, by (5.24), (5.25) and Assumption 3.1, we know that

$$\det\left(\widetilde{\mathcal{H}}_J(0, A) \right) \neq 0.$$

Therefore, since the functions $\widetilde{\mathcal{G}}_J$ and $\widetilde{\mathcal{H}}_J$ are continuous by Proposition 5.3, the implicit function theorem gives the desired result. □

Lemma 5.10. *Under Assumption 2.4 and Assumption 3.1 there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists $\widehat{\gamma} = \widehat{\gamma}(\varepsilon)$ such that if Δ is independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$*

$$\widehat{\mathcal{G}}_J(\varepsilon, A + \widehat{\gamma}(\varepsilon)) = 0 \qquad and \qquad \det\left(\widehat{\mathcal{H}}_J(\varepsilon, A + \widehat{\gamma}(\varepsilon)) \right) \neq 0.$$

Moreover

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

We are now ready to prove the asymptotic unbiasedess of the estimators, i.e., Theorem 3.2 and Theorem 3.3. We only prove Theorem 3.3 for the estimator $\widetilde{A}_{N,J}^\varepsilon$ with filtered data. The proof of Theorem 3.2 for the estimator $\widehat{A}_{N,J}^\varepsilon$ without filtered data is analogous and is omitted here.

Proof of Theorem 3.3. We need to show for a fixed $0 < \varepsilon < \varepsilon_0$:

  (i) the existence of the solution $\widetilde{A}_{N,J}^\varepsilon$ of the system $\widetilde{G}_{N,J}^\varepsilon(a) = 0$ with probability tending to one as $N \to \infty$;

  (ii) $\lim_{N \to \infty} \widetilde{A}_{N,J}^\varepsilon = A + \widetilde{\gamma}(\varepsilon)$ in probability with $\lim_{\varepsilon \to 0} \widetilde{\gamma}(\varepsilon) = 0$.

We first note that by Lemma 5.9 we have

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

We then follow the steps of the proof of [8, Theorem 3.2]. Due to [6, Theorem A.1], claims (i) and (ii) hold true if we verify that

$$\lim_{N \to \infty} \sup_{a \in B_{C,N}^\varepsilon} \left\| \frac{1}{N} \dot{\widetilde{G}}_{N,J}^\varepsilon(a) - \widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right\| = 0, \qquad \text{in probability,} \qquad (5.26)$$

and as $N \to \infty$

$$\frac{1}{\sqrt{N}} \widetilde{G}_{N,J}^{\varepsilon}(A + \widetilde{\gamma}(\varepsilon)) \to \mathcal{N}\left(0, \Lambda^{\varepsilon}\right), \qquad \text{in law}, \tag{5.27}$$

where

$$B_{C,N}^{\varepsilon} = \left\{ a \in \mathcal{A} \colon \|a - (A + \widetilde{\gamma}(\varepsilon))\| \leq \frac{C}{\sqrt{N}} \right\},$$

for $C > 0$ small enough such that $B_{C,1} \subset \mathcal{A}$ and

$$\Lambda^{\varepsilon} = \frac{1}{\Delta^2} \mathbb{E}^{\widetilde{\rho}^{\varepsilon}} \left[\left(\sum_{j=1}^{J} g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, \widetilde{Z}_0^{\varepsilon}; A + \widetilde{\gamma}(\varepsilon)) \right) \otimes \left(\sum_{j=1}^{J} g_j(X_0^{\varepsilon}, X_{\Delta}^{\varepsilon}, \widetilde{Z}_0^{\varepsilon}; A + \widetilde{\gamma}(\varepsilon)) \right) \right].$$

Result (5.27) is a consequence of [8, Lemma 3.1]. We then have

$$\sup_{a \in B_{C,N}^{\varepsilon}} \left\| \frac{1}{N} \dot{\widetilde{G}}_{N,J}^{\varepsilon}(a) - \widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right\| \leq \sup_{a \in B_{C,1}^{\varepsilon}} \left\| \frac{1}{N\Delta} \sum_{i=0}^{N-1} \sum_{j=1}^{J} h_j(\widetilde{X}_n^{\varepsilon}, \widetilde{X}_{n+1}^{\varepsilon}, \widetilde{Z}_n^{\varepsilon}; a) - \widetilde{\mathcal{H}}_J(\varepsilon, a) \right\|$$
$$+ \sup_{a \in B_{C,N}^{\varepsilon}} \left\| \widetilde{\mathcal{H}}_J(\varepsilon, a) - \widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right\|,$$

where the right-hand side vanishes by [8, Lemma 3.3] and the continuity of $\widetilde{\mathcal{H}}$ (Proposition 5.3), implying result (5.26). Hence, we proved (i) and (ii), which conclude the proof of the theorem. $\square$

Remark 5.11. Notice that if $\Delta = \varepsilon^{\zeta}$ with $\zeta > 2$ and we do not employ the filter, in view of (5.20) and following the same proof of Theorem 3.3, we could compute the asymptotic limit of $\widehat{A}_{N,J}^{\varepsilon}$ as N goes to infinity and ε vanishes if we knew a^* such that

$$\sum_{j=1}^{J} \mathbb{E}^{\varphi^0} \left[\beta_j(X_0^0; a^*) \left(\sigma \phi_j''(X_0^0; a^*) - \alpha \cdot V'(X_0^0) \phi_j'(X_0^0; a^*) + \lambda_j(a^*) \phi_j(X_0^0; a^*) \right) \right] = 0.$$

The value of a^* can not be found analytically since it is, in general, different from the drift coefficients α and A of the multiscale and homogenized equations (2.1) and (2.2). Nevertheless, we observe that in the simple scale of the multiscale Ornstein-Uhlenbeck process we have $a^* = \alpha$.

6 Conclusion

In this work we presented new estimators for learning the effective drift coefficient of the homogenized Langevin dynamics when we are given discrete observations from the original multiscale diffusion process. Our approach relies on a martingale estimating function based on the eigenvalues and eigenfunctions of the generator of the coarse-grained model and on a linear time-invariant filter from the exponential family, which is employed to smooth the original data. We studied theoretically the convergence properties of our estimators when the sample size goes to infinity and the multiscale parameter describing the fastest scale vanishes. In Theorem 3.2 and Theorem 3.3 we proved respectively the asymptotic unbiasedness of the estimators with and without filtered data. We remark that the former is not robust with respect to the sampling rate at finite multiscale parameter while the estimator with filtered data is robust independently of the sampling rate. We analysed numerically the dependence of our estimators on the number of observations and the number of eigenfunctions employed in the estimating function noticing that the first eigenvalues in magnitude are sufficient to approximate the drift coefficient. Moreover, we performed several numerical experiments, which highlighted the effectiveness of our approach and confirmed our theoretical results. We believe that eigenfunction estimators can be very useful in applications, for example to multiparticle systems and their mean field limit [14], since the eigenvalue problem for the generator of a reversible Markov process is a very well studied problem. This means, in particular, that it is possible to study rigorously the proposed estimators and to prove asymptotic unbiasedness and asymptotic normality. We will return to this problem in future work. Finally, we think that it would be interesting to extend our estimators to the non-parametric framework and consider more general multiscale models.

Acknowledgements

Appendix A   Technical results

In this section we prove some technical results which are used to show the unbiasedness of the estimators $\widehat{A}_{N,J}^\varepsilon$ and $\widetilde{A}_{N,J}^\varepsilon$. We first study the properties of the filter applied to discrete data and then we focus on the regularity of the eigenfunctions and eigenvalues of the generator. We finally prove a formula which can be interpreted as an approximation of the Itô's lemma.

A.1   Application of the filter to discrete data

The following result quantifies the expected distance among the continuous process Z_t^ε and the filtered observations $\widetilde{Z}_n^\varepsilon$.

Lemma A.1. *Let $0 < \Delta < 1$, N be a positive integer and let $\widetilde{Z}_n^\varepsilon$ and Z_t^ε be defined respectively in (2.16) and (2.20) with $\widetilde{X}_0^\varepsilon = X_0^\varepsilon$ distributed according to its invariant measure φ^ε. Then there exists a constant $C > 0$ independent of ε, Δ and N such that for all $n = 0, \dots, N$ and for all $p \geq 1$*

$$\left(\mathbb{E}^{\varphi^\varepsilon} \left| Z_{n\Delta}^\varepsilon - \widetilde{Z}_n^\varepsilon \right|^p \right)^{1/p} \leq C \left(\Delta^{1/2} + \min\left\{ \varepsilon, \Delta \varepsilon^{-1} \right\} \right),$$

where $\mathbb{E}^{\varphi^\varepsilon}$ denotes the expectation with respect to the Wiener measure and the fact that X_0^ε is distributed according to φ^ε.

Proof. In order to simplify the notation, let us define the quantity

$$E := \mathbb{E}^{\varphi^\varepsilon} \left| Z_{n\Delta}^\varepsilon - \widetilde{Z}_n^\varepsilon \right|^p.$$

Then by Jensen's inequality applied to the convex function $y \mapsto |y|^p$ and since $X_{k\Delta}^\varepsilon = \widetilde{X}_k^\varepsilon$ we have

$$
\begin{aligned}
E &= \mathbb{E}^{\varphi^\varepsilon} \left| \sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} \left(e^{-(n\Delta-s)} X_s^\varepsilon - e^{-\Delta(n-k)} \widetilde{X}_k^\varepsilon \right) \mathrm{d}s \right|^p \\
&\leq 2^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left(\sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} e^{-(n\Delta-s)} \left| X_s^\varepsilon - X_{k\Delta}^\varepsilon \right| \mathrm{d}s \right)^p \\
&\quad + 2^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left(\sum_{k=0}^{n-1} \int_{k\Delta}^{(k+1)\Delta} \left(e^{-(n\Delta-s)} - e^{-\Delta(n-k)} \right) \mathrm{d}s \left| \widetilde{X}_k^\varepsilon \right| \right)^p \\
&=: 2^{p-1} \left(E_1 + E_2 \right).
\end{aligned}
\tag{A.1}
$$

We now study the two terms separately. Applying [2, Lemma B.1] we first get

$$
\begin{aligned}
E_1 &= \mathbb{E}^{\varphi^\varepsilon} \left(\int_0^{n\Delta} e^{-(n\Delta-s)} \left| X_s^\varepsilon - \sum_{k=0}^{n-1} X_{k\Delta}^\varepsilon \chi_{[k\Delta,(k+1)\Delta)}(s) \right| \mathrm{d}s \right)^p \\
&\leq \int_0^{n\Delta} e^{-(n\Delta-s)} \mathbb{E}^{\varphi^\varepsilon} \left| X_s^\varepsilon - \sum_{k=0}^{n-1} X_{k\Delta}^\varepsilon \chi_{[k\Delta,(k+1)\Delta)}(s) \right|^p \mathrm{d}s,
\end{aligned}
\tag{A.2}
$$

and, in order to bound the term inside the integral, we can follow two different procedures. Either we employ [25, Lemma 6.1], which gives

$$\mathbb{E}^{\varphi^\varepsilon} \left| X_s^\varepsilon - \sum_{k=0}^{n-1} X_{k\Delta}^\varepsilon \chi_{[k\Delta,(k+1)\Delta)}(s) \right|^p \le C \left(\Delta^p + \Delta^{p/2} + \varepsilon^p \right), \qquad (A.3)$$

where $C > 0$ is a constant independent of ε and Δ or we notice that, since X_t^ε has bounded moments of any order by [25, Corollary 5.4] and p is bounded, it holds for all $s \in [k\Delta, (k+1)\Delta)$

$$\mathbb{E}^{\varphi^\varepsilon} |X_s^\varepsilon - X_{k\Delta}^\varepsilon|^p = \mathbb{E}^{\varphi^\varepsilon} \left| -\alpha \int_{k\Delta}^s X_r^\varepsilon \, dr - \frac{1}{\varepsilon} \int_{k\Delta}^s p' \left(\frac{X_r^\varepsilon}{\varepsilon} \right) \, dr + \sqrt{2\sigma} W_s \right|^p$$
$$\le C \left(\Delta^p + \Delta^p \varepsilon^{-p} + \Delta^{p/2} \right). \qquad (A.4)$$

Therefore, due to (A.2), (A.3) and (A.4), we obtain

$$E_1 \le C \left(\Delta^{p/2} + \min \left\{ \varepsilon^p, \Delta^p \varepsilon^{-p} \right\} \right). \qquad (A.5)$$

Let us now consider E_2, which can be first bounded by

$$E_2 \le \Delta^p \, \mathbb{E}^{\varphi^\varepsilon} \left(\sum_{k=0}^{n-1} \left(e^{-\Delta(n-1-k)} - e^{-\Delta(n-k)} \right) \left| \widetilde{X}_k^\varepsilon \right| \right)^p,$$

and note that

$$\sum_{k=0}^{n-1} \left(e^{-\Delta(n-1-k)} - e^{-\Delta(n-k)} \right) = \sum_{k=0}^{n-1} \left(e^{-\Delta k} - e^{-\Delta(k+1)} \right) = 1 - e^{-\Delta n}.$$

Therefore, applying Jensen's inequality and due to the fact that $\widetilde{X}_k^\varepsilon$ has bounded moments of any order by [25, Corollary 5.4] we have

$$E_2 \le \Delta^p (1 - e^{-\Delta n})^{p-1} \sum_{k=0}^{n-1} \left(e^{-\Delta(n-1-k)} - e^{-\Delta(n-k)} \right) \mathbb{E}^{\varphi^\varepsilon} \left| \widetilde{X}_k^\varepsilon \right|^p \le C \Delta^p,$$

which, together with (A.1) and (A.5), gives the desired result. $\qquad \square$

We now show the ergodicity of the process $(\widetilde{X}_n^\varepsilon, \widetilde{Z}_n^\varepsilon)$, where the first component is a sample from the continuous-time process, i.e. $\widetilde{X}_n^\varepsilon = X_{n\Delta}^\varepsilon$, while the second component is computed starting from the discrete observations $\widetilde{X}_n^\varepsilon$.

Lemma A.2. *Let $\Delta > 0$ and let Assumption 2.2 hold. Then the couple $(\widetilde{X}_n^\varepsilon, \widetilde{Z}_n^\varepsilon)$, where $\widetilde{X}_n^\varepsilon$ is a sample from the continuous process* (2.1) *and $\widetilde{Z}_n^\varepsilon$ is defined in* (2.16)*, admits a unique invariant measure with density with respect to the Lebesgue measure denoted by $\widetilde{\rho}^\varepsilon = \widetilde{\rho}^\varepsilon(x, z)$. Moreover, if Δ is independent of ε, it converges in law to the two-dimensional process $(\widetilde{X}_n^0, \widetilde{Z}_n^0)$ with $\widetilde{\rho}^0 = \widetilde{\rho}^0(x, z)$ as density of the invariant measure.*

Proof. By definition (2.19) we obtain the following stochastic difference equation

$$\widetilde{Z}_{n+1}^\varepsilon = e^{-\Delta} \widetilde{Z}_n^\varepsilon + \Delta e^{-\Delta} \widetilde{X}_n^\varepsilon,$$

where $\widetilde{X}_n^\varepsilon$ is a stationary and ergodic sequence. Observing that $\log e^{-\Delta} = -\Delta < 0$, applying Theorem 1 and in view of Remark 1.3 in [9] we deduce the existence of a unique invariant measure for the couple $(\widetilde{X}_n^\varepsilon, \widetilde{Z}_n^\varepsilon)$. Let us notice that in the theorem the sequence $\widetilde{X}_n^\varepsilon$ must be defined for all $n \in \mathbb{Z}$ while in our framework $n \in \mathbb{N}$, but let us also remark that any stationary process indexed by $\mathbb{N}$ can be extended to one indexed by $\mathbb{Z}$ in an essentially unique way. Moreover, if Δ is independent of ε, the same reasoning can be repeated to get the existence of a unique invariant measure for the couple $(\widetilde{X}_n^0, \widetilde{Z}_n^0)$. Finally, standard homogenization theory implies the weak convergence of $\widetilde{\rho}^\varepsilon$ to $\widetilde{\rho}^0$, which concludes the proof. $\qquad \square$

Let us now denote the marginal invariant distributions of $Z^\varepsilon, \widetilde{Z}^\varepsilon$ and $Z^0, \widetilde{Z}^0$ respectively by $\psi^\varepsilon = \psi^\varepsilon(z), \widetilde{\psi}^\varepsilon = \widetilde{\psi}^\varepsilon(z)$ and $\psi^0 = \psi^0(z), \widetilde{\psi}^0 = \widetilde{\psi}^0(z)$.

Corollary A.3. *Let Z^ε and $\widetilde{Z}^\varepsilon$ be distributed respectively according to ψ^ε and $\widetilde{\psi}^\varepsilon$. Then there exists a constant $C > 0$ independent of ε and Δ such that*

$$\left(\mathbb{E} \left| Z^\varepsilon - \widetilde{Z}^\varepsilon \right|^p \right)^{1/p} \leq C \left(\Delta^{1/2} + \min \left\{ \varepsilon, \Delta \varepsilon^{-1} \right\} \right).$$

Proof. The result follows directly from Lemma A.1 by letting n go to infinity, noting that the constant C is independent of n and employing ergodicity given by Lemma A.2. $\quad\square$

It directly follows that $\widetilde{Z}_n^\varepsilon$ has bounded moments of all order and, in particular, we can prove Lemma 5.1.

Proof of Lemma 5.1. Applying Jensen's inequality to the function $x \mapsto |x|^p$, we have

$$\mathbb{E}^{\widetilde{\rho}^\varepsilon} \left| \widetilde{Z}^\varepsilon \right|^p \leq 2^{p-1} \mathbb{E}^{\rho^\varepsilon} |Z^\varepsilon|^p + 2^{p-1} \mathbb{E} \left| \widetilde{Z}^\varepsilon - Z^\varepsilon \right|^p,$$

then bounding the two terms in the right-hand side respectively with [2, Lemma C.1] and Corollary A.3 gives the desired result. $\quad\square$

A.2 Properties of eigenfunctions and eigenvalues of the generator

Let us now consider the eigenvalue and the eigenfunctions of the generator of SDE (2.7).

Lemma A.4. *Let $\{(\lambda_j(a), \phi_j(\cdot; a))\}_{j=0}^\infty$ be the solutions of the eigenvalue problem (2.10). Then $\phi_j(x; a)$ and $\lambda_j(a)$ are continuously differentiable with respect to a for all $x \in \mathbb{R}$ and for all $j \in \mathbb{N}$. Moreover, $\phi_j(\cdot; a)$ and $\dot{\phi}_j(\cdot; a)$ belong to $\mathcal{C}^\infty(\mathbb{R})$.*

Proof. The first result follows from Section 2 and Section 6 in [28]. Let us remark that the fact that the spectrum is discrete and non-degenerate is guaranteed by [24, Section 4.7]. Finally, the second result in the statement is a direct consequence of the elliptic regularity theory. $\quad\square$

A.3 Approximation of the Itô formula

In this section we prove Lemma 5.2, which is an approximation of the Itô's lemma applied to the stochastic process X_t^ε. Let us introduce the process S_t^ε defined by the following SDE with initial condition $S_0^\varepsilon = X_0^\varepsilon$

$$\mathrm{d}S_t^\varepsilon = -\alpha V'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))\mathrm{d}t + \sqrt{2\sigma}(1 + \Phi'(Y_t^\varepsilon))\mathrm{d}W_t, \tag{A.6}$$

where $Y_t^\varepsilon = X_t^\varepsilon/\varepsilon$ and Φ is the cell function which solves equation (2.5), and notice that

$$S_\Delta^\varepsilon = X_0^\varepsilon - \alpha \int_0^\Delta V'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}t + \sqrt{2\sigma} \int_0^\Delta (1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}W_t.$$

Therefore, due to equation (5.7) in [25] we have

$$|X_\Delta^\varepsilon - S_\Delta^\varepsilon| = \varepsilon \left| \Phi(Y_\Delta^\varepsilon) - \Phi(Y_0^\varepsilon) \right|,$$

and, since Φ is bounded by [25, Lemma 5.5], we get for a constant $C > 0$ independent of Δ and ε

$$|X_\Delta^\varepsilon - S_\Delta^\varepsilon| \leq C\varepsilon. \tag{A.7}$$

Before showing the main formula, we need two preliminary estimates which will be employed later in the analysis. The proofs of Lemma A.5 and Lemma A.6 are inspired by the proof of Proposition 5.8 in [25].

Lemma A.5. *Let $f \colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that f, f' are polynomially bounded. Then*

$$\int_0^\Delta \alpha \cdot V'(X_t^\varepsilon) f(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}t = A \cdot V'(X_0^\varepsilon) f(X_0^\varepsilon)\Delta + R_1(\varepsilon, \Delta), \qquad (\text{A.8})$$

where the remainder satisfies for all $p \geq 1$ and for a constant $C > 0$ independent of Δ and ε

$$\left(\mathbb{E}^{\varphi^\varepsilon} |R_1(\varepsilon, \Delta)|^p \right)^{1/p} \leq C(\varepsilon^2 + \Delta^{1/2}\varepsilon + \Delta^{3/2}).$$

Proof. To obtain the remainder $R_1(\varepsilon, \Delta)$ we decompose suitably the difference between the left-hand side and the right-hand side of (A.8). Applying Jensen's inequality to the function $z \mapsto |z|^p$ we have

$$\begin{aligned}
\mathbb{E}^{\varphi^\varepsilon} |R_1(\varepsilon, \Delta)|^p \leq{}& 3^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| \int_0^\Delta \alpha \cdot \left(V'(X_t^\varepsilon) - V'(X_0^\varepsilon) \right) f(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}t \right|^p \\
&+ 3^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| \alpha \cdot V'(X_0^\varepsilon) \int_0^\Delta \left(f(X_t^\varepsilon) - f(X_0^\varepsilon) \right)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}t \right|^p \\
&+ 3^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| f(X_0^\varepsilon) V'(X_0^\varepsilon) \cdot \int_0^\Delta \left(\alpha(1 + \Phi'(Y_t^\varepsilon)) - A \right) \mathrm{d}t \right|^p \\
=:{}& I_1(\varepsilon, \Delta) + I_2(\varepsilon, \Delta) + I_3(\varepsilon, \Delta).
\end{aligned} \qquad (\text{A.9})$$

Letting $C > 0$ be a constant independent of ε and δ, we now bound the three terms separately. First, applying Hölder inequality and since V' is Lipschitz, Φ' is bounded, f is polynomially bounded and X_t^ε has bounded moments of any order by [25, Corollary 5.4], we have

$$I_1(\varepsilon, \Delta) \leq C\Delta^{p-1} \int_0^\Delta \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_0^\varepsilon|^p \, \mathrm{d}t,$$

then applying [25, Lemma 6.1] we obtain

$$I_1(\varepsilon, \Delta) \leq C \left(\Delta^{2p} + \Delta^{3p/2} + \varepsilon^p \Delta^p \right). \qquad (\text{A.10})$$

We then rewrite $I_2(\varepsilon, \Delta)$ employing the mean value theorem

$$I_2(\varepsilon, \Delta) = 3^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| \alpha \cdot V'(X_0^\varepsilon) \int_0^\Delta f'(\widetilde{X}_t^\varepsilon)(X_t^\varepsilon - X_0^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}t \right|^p,$$

where $\widetilde{X}_t^\varepsilon$ assumes values between X_0^ε and X_t^ε, and we repeat the same reasoning as for $I_1(\varepsilon, \Delta)$ to get

$$I_2(\varepsilon, \Delta) \leq C \left(\Delta^{2p} + \Delta^{3p/2} + \varepsilon^p \Delta^p \right). \qquad (\text{A.11})$$

We now consider the function

$$H(y) =: \alpha(1 + \Phi'(y)) - A,$$

which by definition of A and due to (2.3) has zero mean with respect to μ defined in (2.4). Therefore, since f and V' are polynomially bounded and X_0^ε has bounded moments of any order by [25, Corollary 5.4], applying [25, Lemma 5.6] we obtain

$$I_3(\varepsilon, \Delta) \leq C \left(\varepsilon^{2p} + \varepsilon^p \Delta^p + \varepsilon^p \Delta^{p/2} \right). \qquad (\text{A.12})$$

Finally, for ε and Δ sufficiently small, the desired result follows from (A.9) and from estimates (A.10), (A.11) and (A.12). $\qquad \square$

Lemma A.6. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that f, f' are polynomially bounded. Then*

$$\int_0^\Delta \sigma f(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))^2 \, \mathrm{d}t = \Sigma f(X_0^\varepsilon)\Delta + R_2(\varepsilon, \Delta), \tag{A.13}$$

where the remainder satisfies for all $p \geq 1$ and for a constant $C > 0$ independent of Δ and ε

$$\left(\mathbb{E}^{\varphi^\varepsilon} |R_2(\varepsilon, \Delta)|^p\right)^{1/p} \leq C(\varepsilon^2 + \Delta^{1/2}\varepsilon + \Delta^{3/2}).$$

Proof. To obtain the remainder $R_2(\varepsilon, \Delta)$ we decompose suitably the difference between the left-hand side and the right-hand side of (A.13). Applying Jensen's inequality to the function $z \mapsto |z|^p$ we have

$$\mathbb{E}^{\varphi^\varepsilon} |R_2(\varepsilon, \Delta)|^p \leq 2^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| \int_0^\Delta \sigma \left(f(X_t^\varepsilon) - f(X_0^\varepsilon)\right) (1 + \Phi'(Y_t^\varepsilon)^2) \, \mathrm{d}t \right|^p$$

$$+ 2^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| f(X_0^\varepsilon) \int_0^\Delta \left(\sigma(1 + \Phi'(Y_t^\varepsilon))^2 - \Sigma\right) \mathrm{d}t \right|^p \tag{A.14}$$

$$=: I_1(\varepsilon, \Delta) + I_2(\varepsilon, \Delta).$$

Letting $C > 0$ be a constant independent of ε and Δ, we now bound the two terms separately. First, we rewrite $I_1(\varepsilon, \Delta)$ employing the mean value theorem

$$I_1(\varepsilon, \Delta) = 2^{p-1} \mathbb{E}^{\varphi^\varepsilon} \left| \int_0^\Delta \sigma f'(\widetilde{X}_t^\varepsilon)(X_t^\varepsilon - X_0^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))^2 \, \mathrm{d}t \right|^p,$$

where $\widetilde{X}_t^\varepsilon$ assumes values between X_0^ε and X_t^ε, then applying Hölder inequality and since Φ' is bounded, f' is polynomially bounded and X_t^ε has bounded moments of any order by [25, Corollary 5.4], we have

$$I_1(\varepsilon, \Delta) \leq C\Delta^{p-1} \int_0^\Delta \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - X_0^\varepsilon|^p \, \mathrm{d}t,$$

and applying [25, Lemma 6.1] we obtain

$$I_1(\varepsilon, \Delta) \leq C \left(\Delta^{2p} + \Delta^{3p/2} + \varepsilon^p \Delta^p\right). \tag{A.15}$$

We now consider the function

$$H(y) =: \sigma(1 + \Phi'(y))^2 - \Sigma,$$

which by definition of Σ and due to (2.3) has zero mean with respect to μ defined in (2.4). Therefore, since f is polynomially bounded and X_0^ε has bounded moments of any order by [25, Corollary 5.4], applying [25, Lemma 5.6] we obtain

$$I_2(\varepsilon, \Delta) \leq C \left(\varepsilon^{2p} + \varepsilon^p \Delta^p + \varepsilon^p \Delta^{p/2}\right). \tag{A.16}$$

Finally, for ε and Δ sufficiently small, the desired result follows from (A.14) and from estimates (A.15) and (A.16). $\qquad \square$

We can now prove the main formula, which is employed repeatedly in the proof of the asymptotic unbiasedness of the drift estimators.

Proof of Lemma 5.2. Applying Itô's lemma to the process S_t^ε defined in (A.6) with the function f we have

$$f(S_\Delta^\varepsilon) = f(X_0^\varepsilon) - \int_0^\Delta \alpha \cdot V'(X_t^\varepsilon)f'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}t + \int_0^\Delta \sigma f''(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))^2 \, \mathrm{d}t$$

$$+ \sqrt{2\sigma} \int_0^\Delta f'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}W_t,$$

and due to Lemma A.5 and Lemma A.6 we obtain

$$f(S_\Delta^\varepsilon) = f(X_0^\varepsilon) - A \cdot V'(X_0^\varepsilon)f'(X_0^\varepsilon)\Delta + \Sigma f''(X_0^\varepsilon)\Delta + \sqrt{2\sigma} \int_0^\Delta f'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))\,\mathrm{d}W_t$$
$$- R_1(\varepsilon, \Delta) + R_2(\varepsilon, \Delta).$$

Then we write

$$f(X_\Delta^\varepsilon) = f(S_\Delta^\varepsilon) + [f(X_\Delta^\varepsilon) - f(S_\Delta^\varepsilon)] =: f(S_\Delta^\varepsilon) + R_3(\varepsilon, \Delta),$$

and, in order to conclude, it only remains to bound the expectation of $R_3(\varepsilon, \Delta)$. Applying the mean value theorem and the Cauchy-Schwarz inequality and due to (A.7), the hypotheses on f and the fact that X_t^ε has bounded moments of any order by [25, Corollary 5.4], we obtain

$$\mathbb{E}^{\varphi^\varepsilon}\,|R_3(\varepsilon, \Delta)|^p \leq \left(\mathbb{E}^{\varphi^\varepsilon}\left|f'(\widetilde{X})\right|^{2p}\right)^{1/2}\left(\mathbb{E}^{\varphi^\varepsilon}\,|X_\Delta^\varepsilon - S_\Delta^\varepsilon|^{2p}\right)^{1/2} \leq C\varepsilon^p,$$

where $\widetilde{X}$ takes values between X_Δ^ε and S_Δ^ε, and which together with the estimates for R_1 and R_2 implies the desired result. $\qquad\square$

Appendix B   Implementation details

In this section we present the main techniques that we employed in the implementation of the proposed method. The most important steps in the algorithm are the computation of the eigenvalues and eigenfunctions of the eigenvalue problem (2.11)

$$\Sigma\phi_j''(x;a) - a \cdot V'(x)\phi_j'(x;a) + \lambda_j(a)\phi_j(x;a) = 0,$$

and the solution of the non-linear system (2.15) or (2.18) with filtered data. Let us first focus on the eigenvalue problem. We note that the domain of the eigenfunctions is the whole real line $\mathbb{R}$ and need to be truncated for numerical computations. We first consider the variational formulation of equation (2.11), i.e., we multiply it by $v\varphi_a$, where v is a test function and φ_a is the invariant distribution defined in (2.8), and integrating by parts we obtain for all $j \in \mathbb{N}$ the following eigenvalue problem

$$\Sigma \int_\mathbb{R} \phi_j'(x;a)v'(x)\varphi_a(x)\,\mathrm{d}x = \lambda_j(a)\int_\mathbb{R} \phi_j(x;a)v(x)\varphi_a(x)\,\mathrm{d}x.$$

Since φ_a decays to zero exponentially fast, for all $\delta > 0$ there exists $r > 0$ such that

$$|\varphi_a(x)| < \delta \qquad \text{for all } x \notin [-r, r].$$

Hence, letting $R > 0$ we assume that $\varphi_a(\pm R) \simeq 0$ and we solve the truncated problem

$$\Sigma \int_{-R}^{+R} \phi_j'(x;a)v'(x)\varphi_a(x)\,\mathrm{d}x = \lambda_j(a)\int_{-R}^{+R} \phi_j(x;a)v(x)\varphi_a(x)\,\mathrm{d}x. \tag{B.1}$$

Notice that R must be chosen big enough and such that

$$R \geq \max_{n=0,\dots,N} \max\left\{\left|\widetilde{X}_n^\varepsilon\right|, \left|\widetilde{Z}_n^\varepsilon\right|\right\} =: \bar{R},$$

and we take $R = \max\{\bar{R}+0.1, 1.7\}$. Moreover, in order to have a unique solution for the eigenvector $\phi_j(\cdot;a)$ we impose the additional conditions

$$\phi_j(R;a) > 0 \qquad \text{and} \qquad \int_{-R}^{+R} \phi_j(x;a)^2\varphi_a(x)\,\mathrm{d}x = 1. \tag{B.2}$$

We then introduce a partition $\mathcal{T}_h$ of $[-R, R]$ in N_h subintervals $K_i = [x_{i-1}, x_i]$ with

$$-R = x_0 < x_1 < \cdots < x_{N_h} < x_{N_h} = +R,$$

and $h = 2R/N_h$, and we construct the discrete space

$$X_h^1 = \left\{ v_h \in C^0([-R, +R]) \colon v_h|_{K_i} \in \mathbb{P}^1 \ \forall \ K_i \in \mathcal{T}_h \right\},$$

which is constituted by continuous piecewise linear functions. Note that the discretization parameter h is chosen to be $h = 0.1$ or $h = 0.05$. We pick the characteristic Lagrangian basis $\{\psi_k\}_{k=0}^{N_h}$ of X_h^1 characterized by the following property

$$\psi_k(x_i) = \delta_{ik} \qquad \text{for all } i, k = 0, \ldots, N_h,$$

where δ_{ik} is the Kronecker delta. We want to find $\phi_j(\cdot; a) \in X_h^1$ such that equation (B.1) holds true for all $v \in X_h^1$. Therefore, in equation (B.1) we substitute

$$\phi_j(x; a) = \sum_{k=0}^{N_h} \theta_j^{(k)}(a) \psi_k(x) \qquad \text{and} \qquad v(x) = \psi_i(x) \text{ for all } i = 0, \ldots, N_h,$$

and we obtain the discrete formulation

$$S\Theta_j(a) = \lambda_j(a) M \Theta_j(a), \tag{B.3}$$

where $\Theta_j(a) \in \mathbb{R}^{N_h+1}$ is such that $(\Theta_j(a))_k = \theta_j^{(k-1)}(a)$ and the components of the matrices $S, M \in \mathbb{R}^{N_h+1 \times N_h+1}$ are given by

$$S_{ik} = \Sigma \int_{-R}^{+R} \psi_{i-1}'(x) \psi_{k-1}'(x) \varphi_a(x) \, \mathrm{d}x, \qquad \text{and} \qquad M_{ik} = \int_{-R}^{+R} \psi_{i-1}(x) \psi_{k-1}(x) \varphi_a(x) \, \mathrm{d}x,$$

where the integrals are approximated through the composite Simpson's quadrature rule. Equation (B.3) is a generalized eigenvalue problem which can be solved in MATLAB using the function `eigs` or in PHYTON using the function `scipy.sparse.linalg.eigsh`. Then we normalize $\Theta_j(a)$ or change its sign in order to impose the conditions (B.2), which can be rewritten as

$$\theta_j^{(N_h)}(a) > 0 \qquad \text{and} \qquad \Theta_j(a)^\top M \Theta_j(a) = 1.$$

Once we compute $\lambda_j(a)$ and $\Theta_j(a)$ we have an approximation of the eigenvalues and eigenfunctions and we can construct the function $\widehat{G}_{N,J}^\varepsilon(a)$ in (2.14) or $\widetilde{G}_{N,J}^\varepsilon(a)$ in (2.17) with filtered data. Hence, it only remains to solve systems (2.15) or (2.18), i.e.,

$$\widehat{G}_{N,J}^\varepsilon(a) = 0, \qquad \text{or} \qquad \widetilde{G}_{N,J}^\varepsilon(a) = 0.$$

To solve these equations we can follow two approaches:

- find the zero of $\widehat{G}_{N,J}^\varepsilon(a)$ or $\widetilde{G}_{N,J}^\varepsilon(a)$;

- find the minimum of $\left\| \widehat{G}_{N,J}^\varepsilon(a) \right\|$ or $\left\| \widetilde{G}_{N,J}^\varepsilon(a) \right\|$.

In practice, for the first approach the function `fsolve` in MATLAB or the function `scipy.optimize.fsolve` in PYTHON can be used, while for the second one the function `fmincon` in MATLAB or the function `scipy.optimize.minimize` in PYTHON can be used. Finally, note that the functions implemented in MATLAB or PYTHON have been employed with their default parameters.

References

[1] A. ABDULLE AND A. DI BLASIO, *A Bayesian numerical homogenization method for elliptic multiscale inverse problems*, SIAM/ASA J. Uncertain. Quantif., 8 (2020), pp. 414–450.

[2] A. Abdulle, G. Garegnani, G. A. Pavliotis, A. M. Stuart, and A. Zanoni, *Drift estimation of multiscale diffusions based on filtered data.* arXiv preprint arXiv:2009.13457[math.NA], submitted to Found. Comput. Math., 2020.

[3] A. Abdulle, G. Garegnani, and A. Zanoni, *Ensemble Kalman filter for multiscale inverse problems*, Multiscale Model. Simul., 18 (2020), pp. 1565–1594.

[4] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang, *How often to sample a continuous-time process in the presence of market microstructure noise*, in Stochastic Finance, A. N. Shiryaev, M. R. Grossinho, P. E. Oliveira, and M. L. Esquível, eds., Boston, MA, 2006, Springer US, pp. 3–72.

[5] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, vol. 348 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer, Cham, 2014.

[6] O. E. Barndorff-Nielsen and M. Sørensen, *A review of some aspects of asymptotic likelihood theory for stochastic processes*, International Statistical Review / Revue Internationale de Statistique, 62 (1994), pp. 133–165.

[7] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, AMS Chelsea Publishing, Providence, RI, 2011. Corrected reprint of the 1978 original [MR0503330].

[8] B. M. Bibby and M. Sørensen, *Martingale estimation functions for discretely observed diffusion processes*, Bernoulli, 1 (1995), pp. 17–39.

[9] A. Brandt, *The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients*, Adv. in Appl. Probab., 18 (1986), pp. 211–220.

[10] D. Crommelin and E. Vanden-Eijnden, *Fitting timeseries by continuous-time Markov chains: a quadratic programming approach*, J. Comput. Phys., 217 (2006), pp. 782–805.

[11] ——, *Reconstruction of diffusions using spectral data from timeseries*, Commun. Math. Sci., 4 (2006), pp. 651–668.

[12] ——, *Diffusion estimation from multiscale data by operator eigenpairs*, Multiscale Model. Simul., 9 (2011), pp. 1588–1623.

[13] V. P. Godambe and C. C. Heyde, *Quasi-likelihood and optimal estimation*, Internat. Statist. Rev., 55 (1987), pp. 231–244.

[14] S. N. Gomes and G. A. Pavliotis, *Mean field limits for interacting diffusions in a two-scale potential*, J. Nonlinear Sci., 28 (2018), pp. 905–941.

[15] L. P. Hansen, J. A. Scheinkman, and N. Touzi, *Spectral methods for identifying scalar diffusions*, J. Econometrics, 86 (1998), pp. 1–32.

[16] C. C. Heyde and R. Gay, *On asymptotic quasi-likelihood estimation*, Stochastic Process. Appl., 31 (1989), pp. 223–236.

[17] 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.

[18] M. Kessler and M. Sørensen, *Estimating equations based on eigenfunctions for a discretely observed diffusion process*, Bernoulli, 5 (1999), pp. 299–314.

[19] S. Krumscheid, G. A. Pavliotis, and S. Kalliadasis, *Semiparametric drift and diffusion estimation for multiscale diffusions*, Multiscale Model. Simul., 11 (2013), pp. 442–473.

[20] S. Krumscheid, M. Pradas, G. A. Pavliotis, and S. Kalliadasis, *Data-driven coarse graining in action: Modeling and prediction of complex systems*, Phys. Rev. E, 92 (2015), p. 042139.

[21] N. V. Krylov, *Controlled diffusion processes*, vol. 14 of Stochastic Modelling and Applied Probability, Springer-Verlag, Berlin, 2009. Translated from the 1977 Russian original by A. B. Aries, Reprint of the 1980 edition.

[22] Y. S. Mīshura, S. V. Posashkova, and S. V. Posashkov, *Continuous dependence of solutions of stochastic differential equations controlled by standard and fractional Brownian motions on a parameter*, Teor. Ĭmovīr. Mat. Stat., (2010), pp. 92–105.

[23] A. Papavasiliou, G. A. Pavliotis, and A. M. Stuart, *Maximum likelihood drift estimation for multiscale diffusions*, Stochastic Process. Appl., 119 (2009), pp. 3173–3210.

[24] 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.

[25] G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.

[26] G. A. Pavliotis and A. M. Stuart, *Multiscale methods*, vol. 53 of Texts in Applied Mathematics, Springer, New York, 2008. Averaging and homogenization.

[27] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys., 378 (2019), pp. 686–707.

[28] T. B. Scheffler, *Analyticity of the eigenvalues and eigenfunctions of an ordinary differential operator with respect to a parameter*, Proc. Roy. Soc. London Ser. A, 336 (1974), pp. 475–486.

[29] K. Spiliopoulos and A. Chronopoulou, *Maximum likelihood estimation for small noise multiscale diffusions*, Stat. Inference Stoch. Process., 16 (2013), pp. 237–266.

[30] L. Yang, X. Meng, and G. E. Karniadakis, *B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data*, J. Comput. Phys., 425 (2021), pp. 109913, 23.

[31] D. Zhang, L. Lu, L. Guo, and G. E. Karniadakis, *Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems*, J. Comput. Phys., 397 (2019), pp. 108850, 19.

[32] 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.