

ROBUST ESTIMATION OF EFFECTIVE DIFFUSIONS FROM MULTISCALE DATA*

GIACOMO GAREGNANI[†] AND ANDREA ZANONI[‡]

Abstract. We present a novel methodology based on filtered data and moving averages for estimating effective dynamics from observations of multiscale systems. We show in a semi-parametric framework of the Langevin type that our approach is asymptotically unbiased with respect to the theory of homogenization. Moreover, we demonstrate on a range of challenging numerical experiments that our method is accurate in extracting coarse-grained dynamics from multiscale data. In particular, the estimators we propose are more robust and require less knowledge of the full model than the standard technique of subsampling, which is widely employed in practice in this setting.

AMS subject classifications. 60J60; 62F12; 62M05; 62M20; 65C30.

Keywords. Parameter inference; diffusion processes; data-driven homogenization; filtering; Langevin equation.

1. Introduction

Inferring simple effective models from observations of complex phenomena characterized by multiple scales is a problem of interest in several fields. Examples range from chemical models of molecular dynamics [28, 29, 36, 39], where reactions may occur at widely separated time scales, to the modelling of financial markets characterized by market-microstructure noise [6, 7, 34, 45]. Extracting from data a simple surrogate of multiscale models is of the utmost relevance also in oceanography, meteorology, and marine biology [15, 26].

In this paper, we are interested in inferring coarse-grained equations from observations of diffusion processes evolving on multiple time scales. Given a positive integer d , a drift function $b^\varepsilon: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ periodic with respect to its second argument, a multiscale parameter $\varepsilon > 0$, and a diffusion coefficient $\sigma > 0$, we consider the d -dimensional multiscale SDE

$$dX^\varepsilon(t) = b^\varepsilon\left(X^\varepsilon(t), \frac{X^\varepsilon(t)}{\varepsilon}\right) dt + \sqrt{2\sigma} dW(t), \quad (1.1)$$

where $W := (W(t), t \geq 0)$ is a standard d -dimensional Brownian motion, and where $X^\varepsilon(0)$ is a given initial condition. Assuming that continuous-time data $X^\varepsilon := (X^\varepsilon(t), 0 \leq t \leq T)$ are provided, with T a finite time horizon, our goal is then inferring a coarse-grained equation, independent of the fastest scale $\mathcal{O}(\varepsilon^{-1})$, which reads

$$dX^0(t) = b^0(X^0(t)) dt + \sqrt{2\Sigma} dW(t), \quad (1.2)$$

where $b^0: \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ are the effective drift function and diffusion matrix, respectively. Knowledge of the full model (1.1) yields, in specific instances, a single-scale model (1.2) which is effective in the sense of the theory of homogenization. In particular,

*Received: September 08, 2021; Accepted (in revised form): May 29, 2022. Communicated by Tony Lelièvre.

[†]ANMC, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (giacomo.garegnani@epfl.ch).

[‡]ANMC, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (andrea.zanoni@epfl.ch).

The authors are partially supported by the Swiss National Science Foundation, under grant No. 200020_172710.

one can prove in these cases that $X^\varepsilon \rightarrow X^0$ for $\varepsilon \rightarrow 0$ in a weak sense (see [40, Chapter 18] or [12, Chapter 3]). In this work, we consider b^ε and σ to be unknown and wish to infer the parameters b^0 and Σ of (1.2) from multiscale data. Hence, the problem we consider here could be framed into the setting of data-driven homogenization.

1.1. Setup: multiscale overdamped Langevin dynamics. The class of multiscale SDEs which can be written as (1.1) is vast, and can be employed for modeling a wide range of physical and social phenomena. In this work, we narrow the scope by considering a semi-parametric framework and a gradient structure, inspired by simple models of molecular dynamics. Let N be a positive integer, consider smooth functions $\{V_i: \mathbb{R}^d \rightarrow \mathbb{R}\}_{i=1}^N$, and a periodic function $p: \mathbb{R}^d \rightarrow \mathbb{R}$ with period L_i in the i -th direction in \mathbb{R}^d for $i=1, \dots, d$. We then let the drift function in the multiscale dynamics (1.1) be given by

$$b^\varepsilon(x, y) = - \sum_{i=1}^N \alpha_i \nabla V_i(x) - \frac{1}{\varepsilon} \nabla p(y),$$

where $\{\alpha_i\}_{i=1}^N$ are scalar drift coefficients. With this choice, Equation (1.1) reads

$$dX^\varepsilon(t) = - \sum_{i=1}^N \alpha_i \nabla V_i(X^\varepsilon(t)) dt - \frac{1}{\varepsilon} \nabla p\left(\frac{X^\varepsilon(t)}{\varepsilon}\right) dt + \sqrt{2\sigma} dW(t), \tag{1.3}$$

and the stochastic model we consider is of the overdamped Langevin type. There exists for Equation (1.3) a model of the form (1.2) which is effective in the homogenization limit $\varepsilon \rightarrow 0$. Let $X^\varepsilon := (X^\varepsilon(t), 0 \leq t \leq T)$ denote the solution of (1.3) for a finite time horizon T . Then, it holds $X^\varepsilon \rightarrow X^0$ in law in $C^0([0, T]; \mathbb{R}^d)$ for $\varepsilon \rightarrow 0$, where $X^0 := (X^0(t), 0 \leq t \leq T)$ is the solution of the overdamped Langevin equation

$$dX^0(t) = - \sum_{i=1}^N A_i \nabla V_i(X^0(t)) dt + \sqrt{2\Sigma} dW(t).$$

Here, the matrices $A_i := \alpha_i \mathcal{K}$ and $\Sigma := \sigma \mathcal{K}$, where $\mathcal{K} \in \mathbb{R}^{d \times d}$ is the symmetric positive semidefinite matrix defined by

$$\mathcal{K} = \int_{\mathbb{L}} (I + D\Phi(y))(I + D\Phi(y))^\top d\nu(y), \quad \mathbb{L} := \bigotimes_{i=1}^d [0, L_i], \tag{1.4}$$

where $D\Phi$ is the Jacobian of the solution $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ of the vector-valued PDE, or cell problem

$$\begin{aligned} \mathcal{L}_0 \Phi &= \nabla p, & \text{in } \mathbb{L}, & \quad + \text{ periodic b.c. on } \partial\mathbb{L}, \\ \int_{\mathbb{L}} \Phi(y) d\nu(y) &= 0, \end{aligned} \tag{1.5}$$

and where the differential operator \mathcal{L}_0 , applied component-wise to Φ , is defined as

$$\mathcal{L}_0 = -\nabla p \cdot \nabla + \sigma \Delta.$$

The measure ν introduced in (1.4) is the probability measure on \mathbb{L} given by

$$\nu(dy) = \frac{1}{C_\nu} \exp\left(-\frac{p(y)}{\sigma}\right) dy, \quad C_\nu = \int_{\mathbb{L}} \exp\left(-\frac{p(y)}{\sigma}\right) dy,$$

which accounts for the fluctuations at infinity of the fast-scales of the solution of (1.3). We refer the reader to [40, Chapters 11 and 18] for a complete derivation and proof of this homogenization result.

1.2. Failure of standard estimators. We consider from now on for clarity the case $d=1$, for which Equation (1.3) reads

$$dX^\varepsilon(t) = -\alpha \cdot V'(X^\varepsilon(t))dt - \frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon(t)}{\varepsilon} \right) dt + \sqrt{2\sigma} dW(t), \tag{1.6}$$

where $V: \mathbb{R} \rightarrow \mathbb{R}^N$ is defined as $V(x) = (V_1(x), V_2(x), \dots, V_N(x))^\top$, the derivative V' is computed component-wise, and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^\top$. Let us assume that we are exposed to a continuous-time stream of data X^ε solution of (1.3) for a finite time horizon T . Moreover, let us assume that the periodic function p , as well as the scale-separation parameter ε , the drift coefficients $\{\alpha_i\}_{i=1}^N$, and the diffusion coefficient σ are unknown. Conversely, we assume that the functions $\{V_i\}_{i=1}^N$ are known. In the semi-parametric framework, indeed, these functions often coincide with the first N elements of the basis of some appropriate function space. Our goal is then to infer the effective drift coefficient $A = (A_1, A_2, \dots, A_N)^\top \in \mathbb{R}^N$ and diffusion coefficient $\Sigma > 0$ that define the single-scale dynamics

$$dX^0(t) = -A \cdot V'(X^0(t))dt + \sqrt{2\Sigma} dW(t). \tag{1.7}$$

We consider the inferred coefficients to be *asymptotically unbiased* if they converge to the true effective coefficients in the homogenization limit $\varepsilon \rightarrow 0$ and for infinite data, i.e., for $T \rightarrow \infty$. Since the coarse-grained dynamics are inferred from data instead of being computed using the homogenization formulas, we are in the setting of *data-driven homogenization*.

Let us first consider the drift coefficient. Applying Girsanov change of measure formula to the effective dynamics (see e.g. [30, 31, 37, 42]) yields the likelihood function

$$L_T(X | A) = \exp \left(-\frac{I_T(X | A)}{2\Sigma} \right),$$

where $X = (X(t), 0 \leq t \leq T)$ is a continuous-time stream of data and where

$$I_T(X | A) = \int_0^T A \cdot V'(X(t)) dX(t) + \frac{1}{2} \int_0^T (A \cdot V'(X(t)))^2 dt.$$

Minimizing the function I_T with respect to A yields the maximum likelihood estimator (MLE) $\hat{A}(X, T)$, which is defined as the solution of the linear system

$$-M(X, T) \hat{A}_{\text{MLE}}(X, T) = v(X, T), \tag{1.8}$$

where

$$M(X, T) := \frac{1}{T} \int_0^T V'(X(t)) \otimes V'(X(t)) dt, \quad v(X, T) := \frac{1}{T} \int_0^T V'(X(t)) dX(t).$$

The standard estimator of the diffusion coefficient Σ , given the stream of data X , is obtained by computing the quadratic variation $\langle X \rangle_T$ of the path X and by defining

$$\hat{\Sigma}(X, T) = \frac{\langle X \rangle_T}{2T}.$$

We remark that in case the data X would originate from the model (1.7), we would have by definition $\hat{\Sigma}(X, T) = \Sigma$.

Let us now consider the stream of data $X = X^\varepsilon$, i.e., the framework of data-driven homogenization. In this case, both the estimators $\widehat{A}(X^\varepsilon, T)$ and $\widehat{\Sigma}(X^\varepsilon, T)$ for the effective drift and diffusion coefficients are not asymptotically unbiased. In particular, it holds by [39, Theorem 3.4]

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widehat{A}_{\text{MLE}}(X^\varepsilon, T) &= \alpha, \quad \text{a.s.}, \\ \widehat{\Sigma}(X^\varepsilon, T) &= \sigma, \end{aligned} \tag{1.9}$$

where α and σ are the coefficients of the multiscale dynamics (1.6). In this setting, the standard estimators fail and it is necessary to employ homogenization-informed techniques to infer the effective equation.

REMARK 1.1. In case ε is known and due to (1.9), it would be possible to infer directly the full multiscale model (1.7) employing a periodic parametrisation of the function p . We argue that this would be less useful, at least for predictive purposes, than estimating directly the effective model. Indeed, numerical integration of (1.6) is possible only by choosing critically small time step for most numerical schemes, which in turn yields dramatically high computational cost.

1.3. Literature review. The literature on statistical inference of stochastic models modelled by SDEs is vast. Introductory references on the topic are [11, 13, 27, 43]. The articles [41, 42] give a complete overview on recent frequentist and Bayesian perspectives on the topic. A series of methods have been proposed in recent years for multiscale models, in different settings and with different purposes. We refer the reader to [38] for a recent survey, and summarize the approaches which are related to the one presented in this article in the following.

The focus of the early papers [36, 39] are simple multiscale models of molecular dynamics. In these works, the authors show that multiscale continuous-time data should be subsampled in order to infer single-scale equations which are effective in the sense of the theory of homogenization. In particular, let $\delta > 0$ be the subsampling rate, chosen for simplicity such that $T = n\delta$ for a positive integer n . The subsampled drift estimator is computed with a Euler–Maruyama-type discretisation of the MLE with spacing δ , i.e.,

$$-M_{\text{sub}}^\delta(X^\varepsilon, T) \widehat{A}_{\text{sub}}^\delta(X^\varepsilon, T) = v_{\text{sub}}^\delta(X^\varepsilon, T), \tag{1.10}$$

where

$$\begin{aligned} M_{\text{sub}}^\delta(X^\varepsilon, T) &:= \frac{\delta}{T} \sum_{i=0}^{n-1} V'(X^\varepsilon(i\delta)) \otimes V'(X^\varepsilon(i\delta)), \\ v_{\text{sub}}^\delta(X^\varepsilon, T) &:= \frac{1}{T} \sum_{i=0}^{n-1} V'(X^\varepsilon(i\delta)) (X^\varepsilon((i+1)\delta) - X^\varepsilon(i\delta)). \end{aligned}$$

For the diffusion coefficient, the same subsampling and discretization procedure yields the estimator

$$\widehat{\Sigma}_{\text{sub}}^\delta = \frac{1}{2T} \sum_{i=0}^{n-1} (X^\varepsilon((i+1)\delta) - X^\varepsilon(i\delta))^2.$$

The subsampling rate, which guarantees asymptotically unbiased estimators, lies between the time scales of the multiscale and the effective models, i.e., for $\delta = \varepsilon^\zeta$, with

$\zeta \in (0,1)$ [39, Theorems 3.5 and 3.6]. The optimal subsampling rate is conjectured in [39] to be $\delta = \varepsilon^{2/3}$. The necessity of subsampling the data have been further observed in [8–10] for unobservable multiscale processes. In econometrics, the works [6, 7, 34, 45] show that subsampling the data is indispensable in order to infer integrated volatility in the presence of market-microstructure noise. The disadvantages of subsampling are mainly two. First, it has been demonstrated numerically [3, 39] that inference results based on subsampling highly depend on δ for $\varepsilon > 0$ and finite T . Second, knowledge of the scale-separation parameter ε is necessary to build asymptotically unbiased estimators, which in practice could be a severe limitation.

The methodologies proposed in [23, 25] and further applied in [26] for model selection with paleoclimatic and biological data bypass subsampling by exploiting a martingale property of the likelihood function, and by computing appropriate estimators for conditional expectations. The estimators proposed in [25] are not well posed on a single trajectory, which is overcome by averaging over a set of short trajectories. For [23], estimators are obtained through a computationally expensive procedure employed to approximate conditional expectations via Nadarya–Watson techniques. Let us remark that, to our knowledge, unbiasedness for these estimators of the effective dynamics is not theoretically justified and is just conjectured in [25]. In [23], theoretical analysis is restricted to the one-dimensional case and when the effective dynamics is of the Ornstein–Uhlenbeck type.

Another alternative to subsampling has been proposed in [3]. In this work, maximum likelihood and Bayesian estimators for the effective dynamics are modified by smoothing the continuous-time data with a low-pass exponential filter. Under the assumptions of clear separation between the fast and the slow time scales, and of periodicity for the fast-scale drift function, asymptotic unbiasedness of the estimators is shown. An enhanced robustness with respect to subsampling is demonstrated via numerical experiments. The setting presented in our current paper closely relates to the work [3], which we simplify here without any loss of accuracy.

In their series of works [20, 21, 44], the authors propose estimators for the parameters of a multiscale SDE. The setting is similar to the one studied in this paper, with the difference that the stochastic dynamics are driven by a small noise, which vanishes in the homogenization limit. Hence, the effective equation is in this case a deterministic dynamical system. Theoretical difficulties in this framework are due to the likelihood function induced by the effective dynamics, which is singular. Closely related work is [35], where estimation of multiscale stochastic dynamics is obtained by dimensionality reduction of an appropriate posterior distribution.

In the case of discrete-time data, it is possible to employ information based on the eigenpairs of the generator of the dynamics to obtain asymptotically unbiased estimators. In [24] and in the series of works [16, 17], the authors develop the theory of this class of spectral estimators for single-scale SDEs, which is further applied to the multiscale case in [18]. Recently, the technique based on filtered data of [3] has been combined with eigenfunction martingale estimating functions for inferring effective dynamics from discrete-time data from the full model in [5].

The recent work [14] deals with inference of a similar multiscale equation with Kalman filtering methodologies. In this work, though, the authors focus on retrieving the coefficients of the multiscale dynamics given misspecified data from the reduced model, while we are interested in the opposite direction.

In the related field of multiscale partial differential equations (PDEs), the issue of model misspecification due to the mismatch of multiscale data and homogenized models

has been studied in [33] and in the series of works [1, 2, 4].

1.4. Our contributions. In this paper, we build on the techniques introduced in our previous work [3] and design a new class of estimators for effective diffusions based on moving averages. The methodologies we introduce here are easy to implement, computationally cheap, robust, and unbiased with respect to the theory of homogenization. Furthermore, we complete the numerical analysis of [3] by testing our method against a complex instance of multi-dimensional Langevin equations.

The basic idea underlying both our previous work [3] and our manuscript is similar: in both works we propose to smoothen the data through some filtering kernel, and to compute modified estimators which employ – in some appropriate fashion – the filtered data. Nevertheless, the theoretical and numerical analysis presented in this paper extend relevantly our previous work and have their own originality. In particular:

- (i) Numerical experiments show that the straightforward application of a moving average to the data yields extremely robust inference of effective dynamics when presented with multiscale data. The accuracy of the inference procedure is remarkably higher than previously existing techniques.
- (ii) We propose a novel straightforward estimator for the effective diffusion coefficient, which was not previously analysed in our previous work, or elsewhere in the literature.
- (iii) Implementing our methodology does not require prior knowledge of the scale-separation parameter, and is simple and efficient even for data and parameters in multiple dimensions.
- (iv) We present an original analysis of asymptotic unbiasedness based on the ergodic properties of an appropriate system of stochastic delay differential equations (SDDEs). To our knowledge, this analysis is a novel theoretical contribution to the literature of statistical inference for SDEs.

1.5. Outline. The remainder of this paper is organized as follows. In Section 2 we present our methodology and introduce the main results of unbiasedness for the estimators based on filtered data. We then present numerical experiments in Section 3 corroborating our theoretical findings and showing how to apply our methodology to a complex multi-dimensional scenario. Section 4 is dedicated to the proof of unbiasedness for our estimators, and in Section 5 we draw our conclusions and present some possible future directions of research.

2. The filtered data approach

In this work, we propose an alternative to subsampling for preprocessing the data and infer effective dynamics. In particular, we smoothen the data with a continuous moving average, identified by the kernel

$$k_{\text{ma}}^\delta(r) = \frac{1}{\delta} \chi_{[0,\delta]}(r),$$

where $\chi_{[a,b]}$ denotes the indicator function of the interval $[a,b]$ for real numbers $a < b$, and where $\delta > 0$ is the size of the moving average window. We then take a time convolution of the kernel and the data and obtain the filtered trajectory

$$Z_{\text{ma}}^{\delta,\varepsilon}(t) = \int_0^t k_{\text{ma}}^\delta(t-s) X^\varepsilon(s) ds = \frac{1}{\delta} \int_{t-\delta}^t X^\varepsilon(s) ds, \quad t \geq \delta. \quad (2.1)$$

The process $Z_{\text{ma}}^{\delta,\varepsilon}$ is fully defined on the interval $t \in [0, T]$ by defining for small times $t \leq \delta$

$$Z_{\text{ma}}^{\delta,\varepsilon}(t) = \frac{1}{t} \int_0^t X^\varepsilon(s) \, ds, \quad 0 \leq t < \delta.$$

We remark that with this choice the process $Z_{\text{ma}}^{\delta,\varepsilon}$ is continuous on $[0, T]$, and can be seen as a smoothed version of the original trajectory where the fast oscillations are damped.

The idea of smoothing the data with a low-pass filter has already been introduced in our previous work [3]. Instead of a moving average, we consider in [3] an exponential filtering kernel of the form

$$k_{\text{exp}}^{\delta,\beta}(r) = \frac{C_\beta}{\delta^{1/\beta}} \exp\left(-\frac{r^\beta}{\delta}\right), \quad C_\beta = \frac{\beta}{\Gamma(1/\beta)}, \tag{2.2}$$

where the constant C_β normalizes the kernel in the sense

$$\int_0^\infty k_{\text{exp}}^{\delta,\beta}(r) \, dr = 1.$$

Filtered data $Z_{\text{exp}}^{\delta,\beta,\varepsilon}(t)$ are then obtained similarly to (2.1) by taking a time convolution. We remark that the parameters δ and β in (2.2) have two different roles. In particular, we have that β is a *shape parameter*, and δ is the *filtering width*. The approaches presented here and in our previous work are closely related. In fact, it is simple to deduce that for almost every $r \geq 0$

$$\lim_{\beta \rightarrow \infty} k_{\text{exp}}^{\delta,\beta}(r) = \chi_{[0,1]}(r), \tag{2.3}$$

independently of δ . We remark that the theoretical analysis in [3] is restricted to the case where the shape parameter $\beta = 1$, despite numerical experiments suggesting that choosing $\beta > 1$ yields better estimators. Studying the moving average kernel and (2.3) therefore partially fills the theoretical gap of our previous work.

2.1. Estimating the drift coefficient. We now present how in practice one employs filtered data to obtain asymptotically unbiased estimators of the effective drift coefficient. The main idea is modifying the classical MLE (1.8) by replacing one occurrence of the original process X^ε with the filtered process $Z_{\text{ma}}^{\delta,\varepsilon}$ both in M and v . In particular, the drift estimator $\hat{A}_{\text{ma}}^\delta(X^\varepsilon, T)$ is defined as the solution of the linear system

$$-M_{\text{ma}}^\delta(X^\varepsilon, T) \hat{A}_{\text{ma}}^\delta(X^\varepsilon, T) = v_{\text{ma}}^\delta(X^\varepsilon, T), \tag{2.4}$$

where

$$M_{\text{ma}}^\delta(X^\varepsilon, T) := \frac{1}{T} \int_0^T V'(Z_{\text{ma}}^{\delta,\varepsilon}(t)) \otimes V'(X^\varepsilon(t)) \, dt,$$

$$v_{\text{ma}}^\delta(X^\varepsilon, T) := \frac{1}{T} \int_0^T V'(Z_{\text{ma}}^{\delta,\varepsilon}(t)) \, dX^\varepsilon(t).$$

An asymptotically unbiased estimator $\hat{A}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T)$ of the drift coefficient based on exponential filtering kernels is obtained in the same way, by replacing $Z_{\text{ma}}^{\delta,\varepsilon}$ with $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$, as shown in [3]. We remark that it is fundamental to keep $X^\varepsilon(t)$ in the differential of the right-hand side $v(X^\varepsilon, T)$, as discussed in [3, Remark 3.7]. For the well-posedness of the estimator $\hat{A}_{\text{ma}}^\delta(X^\varepsilon, T)$ the matrix $M_{\text{ma}}^\delta(X^\varepsilon, T)$ needs to be invertible. We take this property as an assumption in the theoretical analysis but we observe that it holds in practice in all the numerical experiments.

2.2. Estimating the diffusion coefficient. We now focus on inferring the effective diffusion coefficient and we propose two different estimators. The first one is given by

$$\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T) = \frac{1}{\delta T} \int_\delta^T (X^\varepsilon(t) - Z_{\text{ma}}^{\delta, \varepsilon}(t)) (X^\varepsilon(t) - X^\varepsilon(t - \delta)) dt, \quad (2.5)$$

which is analogous to the diffusion estimator presented in [3, Section 3].

A different methodology for estimating the diffusion coefficient of the homogenized equation can be derived from the particular form of (1.7). We know that $\Sigma = \sigma \mathcal{K}$ and therefore an estimation of Σ can be obtained by first estimating σ and \mathcal{K} . The former can be computed exactly due to (1.9), indeed we have

$$\sigma = \frac{\langle X^\varepsilon \rangle_T}{2T},$$

while for the latter we use the fact that $A = \alpha \mathcal{K}$. Given an estimator \widehat{A} for the effective drift coefficient and since by (1.9) we know that $\widehat{\alpha} = \widehat{A}_{\text{MLE}}(X^\varepsilon, T)$ approximates α , we write

$$\widehat{\alpha} \widehat{\mathcal{K}} = \widehat{A}, \quad (2.6)$$

which is an overdetermined linear system with N equations and one unknown, and where $\widehat{\mathcal{K}}$ denotes the estimator of the coefficient \mathcal{K} which we aim to infer. The least squares solution to (2.6) is given by

$$\widehat{\mathcal{K}} = \frac{\widehat{\alpha}^\top \widehat{A}}{\widehat{\alpha}^\top \widehat{\alpha}}.$$

Assuming that an estimator $\widehat{A}(X^\varepsilon, T)$ of the effective drift coefficient has already been computed, the effective diffusion coefficient can then be estimated as

$$\widetilde{\Sigma}(X^\varepsilon, T) = \frac{\langle X^\varepsilon \rangle_T (\widehat{A}_{\text{MLE}}(X^\varepsilon, T)^\top \widehat{A}(X^\varepsilon, T))}{2T (\widehat{A}_{\text{MLE}}(X^\varepsilon, T)^\top \widehat{A}_{\text{MLE}}(X^\varepsilon, T))}. \quad (2.7)$$

REMARK 2.1. In practice, the stream X^ε consists of high-frequency discrete data, and not of a continuous process. Hence, the filtered data and all our estimators are computed in practice with the usual appropriate discretizations. We notice that if X^ε consists of n data points, the time complexity needed to compute the filtered trajectory $Z_{\text{ma}}^{\delta, \varepsilon}$ is of order $\mathcal{O}(n)$.

2.3. Statement of asymptotic unbiasedness results. In this section we present the main theoretical results of this work, i.e., the asymptotic unbiasedness of the proposed estimators. We first introduce the assumptions which will be employed in the analysis. In particular, we restrict the analysis to the same dissipative framework as [3, 39].

ASSUMPTION 2.1. *The functions p and V satisfy:*

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

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

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

$$\|V'(x) - V'(y)\| \leq C|x - y|, \quad \text{and} \quad \|V''(x) - V''(y)\| \leq C|x - y|,$$

where $\|\cdot\|$ denotes the Euclidean norm, and the components V'_i and V''_i are polynomially bounded for all $i = 1, \dots, N$.

REMARK 2.2. Assumption 2.1 has the following consequences:

- (i) Assumption 2.1 (i) allows to employ the theory of periodic homogenization to conclude that (1.7) is effective for Equation (1.6).
- (ii) Assumption 2.1 (ii) implies that the solutions of (1.6) and (1.7) are geometrically ergodic, and thus the existence of unique invariant measures (see [39, Propositions 5.1 and 5.2]).
- (iii) Assumption 2.1 (iii) is technical, and guarantees sufficient regularity for our main results to hold.

We first consider the drift estimator $\widehat{A}_{\text{ma}}^\delta(X^\varepsilon, T)$, which is asymptotically unbiased due to the following result.

THEOREM 2.1. Let $\widehat{A}_{\text{ma}}^\delta(X^\varepsilon, T)$ be defined in (2.4) with δ independent of ε or $\delta = \varepsilon^\zeta$ where $\zeta \in (0, 2)$. Under Assumption 2.1 and if $M_{\text{ma}}^\delta(X^\varepsilon, T)$ is invertible, it holds

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

where A is the drift coefficient of the homogenized Equation (1.7).

The following result of asymptotic unbiasedness holds for the estimator $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$ of the effective diffusion.

THEOREM 2.2. Let $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$ be defined in (2.5) with $\delta = \varepsilon^\zeta$ where $\zeta \in (0, 2)$. Under Assumption 2.1, it holds

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T) = \Sigma, \quad \text{a.s.},$$

where Σ is the diffusion coefficient of the homogenized Equation (1.7).

We notice from the statement of Theorem 2.2 that the main limitation of $\widehat{\Sigma}_{\text{ma}}^\delta$ is that knowledge of the scale-separation parameter ε is necessary for unbiasedness. Conversely, the estimator $\widetilde{\Sigma}(X^\varepsilon, T)$ of the effective diffusion can be asymptotically unbiased even if δ is independent of ε . Indeed, unbiasedness of $\widetilde{\Sigma}(X^\varepsilon, T)$ solely relies on the accuracy of the corresponding drift estimator, as shown by the following result.

THEOREM 2.3. Let \widehat{A} be an asymptotically unbiased estimator of the effective drift coefficient, i.e.,

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widehat{A}(X^\varepsilon, T) = A, \quad \text{a.s.}, \tag{2.8}$$

where A is the drift coefficient of the homogenized Equation (1.7), and let $\widetilde{\Sigma}(X^\varepsilon, T)$ be defined in (2.7). Under Assumption 2.1, it holds

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

where Σ is the diffusion coefficient of the homogenized Equation (1.7).

The following corollary is a direct consequence of Theorem 2.3.

COROLLARY 2.1. *Let $\widetilde{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$, $\widetilde{\Sigma}_{\text{exp}}^{\delta, \beta}(X^\varepsilon, T)$, $\widetilde{\Sigma}_{\text{sub}}^\delta(X^\varepsilon, T)$ be defined by (2.7) with $\widehat{A}(X^\varepsilon, T)$ replaced by $\widehat{A}_{\text{ma}}^\delta(X^\varepsilon, T)$, $\widehat{A}_{\text{exp}}^{\delta, \beta}(X^\varepsilon, T)$, and $\widehat{A}_{\text{sub}}^\delta(X^\varepsilon, T)$, respectively. If δ is independent of ε or $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$, then*

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widetilde{\Sigma}_{\text{ma}}^\delta = \Sigma, \quad a.s., \quad \lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widetilde{\Sigma}_{\text{exp}}^{\delta, \beta} = \Sigma, \quad a.s.,$$

and if $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widetilde{\Sigma}_{\text{sub}}^\delta = \Sigma, \quad a.s.,$$

where Σ is the diffusion coefficient of the homogenized Equation (1.7).

Two remarks are due.

REMARK 2.3. In our approach, the scale-separation parameter ε need not be known. In particular, Theorem 2.1 and Corollary 2.1 show that our estimators are asymptotically unbiased if δ is independent of ε , i.e., the filtering width equals the speed of the homogenized process. Since ε is in general unknown in practice, this constitutes an advantage with respect to, e.g., subsampling, for which knowledge of ε is necessary. Moreover, as we demonstrate numerically in Section 3, modifying the filtering width has a weak impact on the inference results as long as $\delta \in [\varepsilon, 1]$ when $T < \infty$ and $\varepsilon > 0$.

REMARK 2.4. Despite being more accurate in practice, as demonstrated by our numerical experiments, and not requiring knowledge of ε , the estimator $\widetilde{\Sigma}_{\text{ma}}^\delta$ is computationally more expensive to obtain than $\widehat{\Sigma}_{\text{ma}}^\delta$. Indeed, computing $\widetilde{\Sigma}_{\text{ma}}^\delta$ requires estimators for the parameters α and σ of the multiscale Equation (1.6), as well as the drift coefficient A of the homogenized model (1.7). Hence, if a very accurate estimate for the whole effective equation is needed, we recommend to employ $\widetilde{\Sigma}_{\text{ma}}^\delta$, while $\widehat{\Sigma}_{\text{ma}}^\delta$ can be used in case only the diffusion coefficient is required.

The proof of Theorems 2.1 to 2.3 is obtained by applying techniques similar to the ones employed in [3, 5, 39], and is presented in detail in Section 4.

2.4. The Multi-dimensional case. In this section we report the expression of the estimators for higher dimensions. Indeed, the choice $d=1$ in the remainder of the paper is made only for economy of notation, and for clarity. Moreover, the proofs of Theorems 2.1 to 2.3 would be conceptually unchanged by considering $d > 1$, up to tedious technical details. Let $\nabla V: \mathbb{R}^d \rightarrow \mathbb{R}^{dN}$ be the vector obtained by stacking the gradients of the components V_i , $i = 1, \dots, N$, of the slow-scale potential, i.e.,

$$\nabla V(x) = (\nabla V_1(x)^\top \cdots \nabla V_N(x)^\top)^\top.$$

Then, the drift estimator for the block matrix $A \in \mathbb{R}^{dN \times d}$ given by

$$A = (A_1^\top \cdots A_N^\top)^\top,$$

is the solution of the linear system

$$-M_{\text{ma}}^\delta(X^\varepsilon, T) \widehat{A}_{\text{ma}}^\delta(X^\varepsilon, T) = v_{\text{ma}}^\delta(X^\varepsilon, T),$$

where the matrices $M_{\text{ma}}^\delta(X^\varepsilon, T) \in \mathbb{R}^{dN \times dN}$ and $v_{\text{ma}}^\delta(X^\varepsilon, T) \in \mathbb{R}^{dN \times d}$ are defined as

$$M_{\text{ma}}^\delta(X^\varepsilon, T) := \frac{1}{T} \int_0^T \nabla V(Z_{\text{ma}}^{\delta, \varepsilon}(t)) \otimes \nabla V(X^\varepsilon(t)) dt,$$

$$v_{\text{ma}}^\delta(X^\varepsilon, T) := \frac{1}{T} \int_0^T \nabla V(Z_{\text{ma}}^{\delta, \varepsilon}(t)) \otimes dX^\varepsilon(t).$$

Concerning the diffusion coefficient, it is natural to impose that it is symmetric and positive definite, so that the square root $\sqrt{2\Sigma}$ is well defined. For this reason, we define the estimator $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$ for the effective diffusion coefficient $\Sigma \in \mathbb{R}^{d \times d}$ as

$$\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T) = \mathcal{S} \left(\frac{1}{\delta T} \int_\delta^T (X^\varepsilon(t) - Z_{\text{ma}}^{\delta, \varepsilon}(t)) \otimes (X^\varepsilon(t) - X^\varepsilon(t - \delta)) dt \right),$$

where $\mathcal{S}(B)$ denotes the symmetric part of a matrix $B \in \mathbb{R}^{d \times d}$, i.e., $\mathcal{S}(B) = (B + B^\top)/2$. Remark that positive definiteness is not guaranteed for $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$, but due to asymptotic unbiasedness it is natural to expect that for T large enough and ε small enough the estimator $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$ is positive definite as its limit. For the estimators of the form $\widetilde{\Sigma}(X^\varepsilon, T)$, we first estimate the homogenization matrix as

$$\widehat{\mathcal{K}} = \arg \min_{\mathcal{K} \in \text{Sym}_d^+} \left\| \widehat{A}_{\text{MLE}}(X^\varepsilon, T) \mathcal{K} - \widehat{A}(X^\varepsilon, T) \right\|_F,$$

where Sym_d^+ is the space of symmetric positive definite matrices of size $d \times d$, $\|\cdot\|_F$ is the Frobenius norm, and $\widehat{A}(X^\varepsilon, T)$ is any estimator of the effective drift coefficient $A \in \mathbb{R}^{dN \times d}$. It is simple to compute the minimum in practice by imposing $\mathcal{K} = LL^\top$, and then minimizing directly over $L \in \mathbb{R}^{d \times d}$. Then, we define

$$\widetilde{\Sigma}(X^\varepsilon, T) = \frac{\langle X^\varepsilon \rangle_T}{2T} \widehat{\mathcal{K}}.$$

In this case, the estimator is symmetric and positive definite by construction.

As we present in the numerical experiment of Section 3.3, our methodology based on moving average can be naturally and successfully applied to higher-dimensional SDEs.

3. Numerical experiments

In this section, we present a series of numerical experiments which have the twofold goal of validating our theoretical analysis, and of showcasing the effectiveness of our technique on challenging academic examples.

3.1. Sensitivity analysis with a Ornstein–Uhlenbeck model.

We first consider the one-dimensional Equation (1.6) with $N = 1$, with the slow scale potential $V(x) = x^2/2$, and with the fluctuating potential $p(y) = \sin(y)$. In this case, the effective model is an Ornstein–Uhlenbeck equation, and it is simple to verify that the homogenization coefficient $\mathcal{K} > 0$ is given by

$$\mathcal{K} = \frac{L^2}{C_\eta^+ C_\eta^-}, \quad C_\eta^\pm = \int_0^L \exp\left(\pm \frac{p(y)}{\sigma}\right) dy, \tag{3.1}$$

where $L = 2\pi$ is the period of p . In this scenario, we can compute exact values for the effective drift and diffusion coefficients to assess the accuracy of our inference method. We then compare numerically the accuracy of the estimators

- (i) $\widehat{A}_{\text{ma}}^\delta$, $\widehat{A}_{\text{exp}}^{\delta,1}$, and $\widehat{A}_{\text{sub}}^\delta$ of the effective drift coefficient obtained by employing data preprocessed with the moving average filter of width δ , the exponential filter with δ and shape parameter $\beta = 1$, and subsampling with period δ ,
- (ii) $\widehat{\Sigma}_{\text{ma}}^\delta$, $\widehat{\Sigma}_{\text{exp}}^{\delta,1}$, and $\widehat{\Sigma}_{\text{sub}}^\delta$ of the effective diffusion coefficient obtained with the same methods, respectively,
- (iii) $\widetilde{\Sigma}_{\text{ma}}^\delta$, $\widetilde{\Sigma}_{\text{exp}}^{\delta,1}$, and $\widetilde{\Sigma}_{\text{sub}}^\delta$ of the effective diffusion coefficient based on the corresponding drift estimators given in (i).

We consider $T = 10^4$ and generate data $X^\varepsilon = (X^\varepsilon(t), 0 \leq t \leq T)$ from the multiscale model with drift coefficient $\alpha = 1$ and for a variable diffusion coefficient $\sigma = 0.5, 0.75, 1$, so that the homogenization coefficient $\mathcal{K} \approx 0.19, 0.45, 0.62$ respectively. Moreover, we consider scale-separation parameters $\varepsilon = 0.2, 0.1, 0.05$ to observe convergence with respect to the homogenization limit. Data is generated with the Euler–Maruyama method with fixed time step $\Delta_t = \varepsilon_{\text{min}}^3$, where $\varepsilon_{\text{min}} = 0.05$ is the smallest value we employ for the scale-separation parameter. With this choice, we have the twofold advantage of introducing negligible numerical errors which do not compromise the validity of our results, and of capturing well the fast-scale oscillations. The filtering/subsampling widths are set to $\delta = \varepsilon^\zeta$ for $\zeta = i/10$ for $i = 0, 1, \dots, 20$ to observe robustness with respect to preprocessing. Let us remark that subsampling-based estimators are asymptotically unbiased only for $\zeta \in (0, 1)$, and that the theory for both filtering kernels is different in case $\zeta = 0$, i.e., when the filtering width is independent of ε .

Numerical results, given in Figures 3.1 to 3.3, demonstrate that

- (a) Figure 3.1: The two filtering techniques yield estimators of the drift coefficient of comparable accuracy across all parameters σ , ε and δ , and they are both more robust than subsampling when varying the parameters σ and δ . We observe that robustness with respect to δ is particularly improved for higher values of σ . For the two filtering methodologies asymptotic unbiasedness with respect to ε seems to hold in practice. Finally, convergence with respect to $t \in [0, T]$, verified with $\sigma = 1$, $\delta = 1$ for the filtering methods and $\delta = \varepsilon^{2/3}$ (conjectured optimal in [39]) for subsampling is similar for the three methods.
- (b) Figure 3.2: The estimators $\widehat{\Sigma}_{\text{ma}}^\delta$, $\widehat{\Sigma}_{\text{exp}}^{\delta,1}$, and $\widehat{\Sigma}_{\text{sub}}^\delta$ of Σ have similar accuracy across all values of ε , δ , and σ . Again, convergence with respect to ε seems to be in practice respected. Let us remark that for these estimators $\delta = 1$ is not a viable choice (for neither the two filtering methods, nor subsampling). Convergence in time is therefore demonstrated for $\sigma = 1$, for $\delta = \varepsilon$ for the two filtering methods, and with $\delta = \varepsilon^{2/3}$ for subsampling. With these choices, the moving average filter introduced in this paper seems to slightly outperform the concurrent methods.
- (c) Figure 3.3: The estimators $\widetilde{\Sigma}_{\text{ma}}^\delta$ and $\widetilde{\Sigma}_{\text{exp}}^{\delta,1}$ show enhanced accuracy with respect to the corresponding estimators of Item b across all values of ε , σ , and δ . Convergence with respect to ε seems to be respected for all methods. Convergence in time is demonstrated for $\sigma = 1$, for $\delta = 1$ for the two filtering methods, and with $\delta = \varepsilon^{2/3}$ for subsampling. With these choices, the moving average filter introduced here seems to show a faster time transient towards the effective diffusion coefficient with respect to the concurrent methods. We remark that the diffusion estimators identified by the “hat” seem to converge faster with respect to t than the ones identified with the “tilde”.

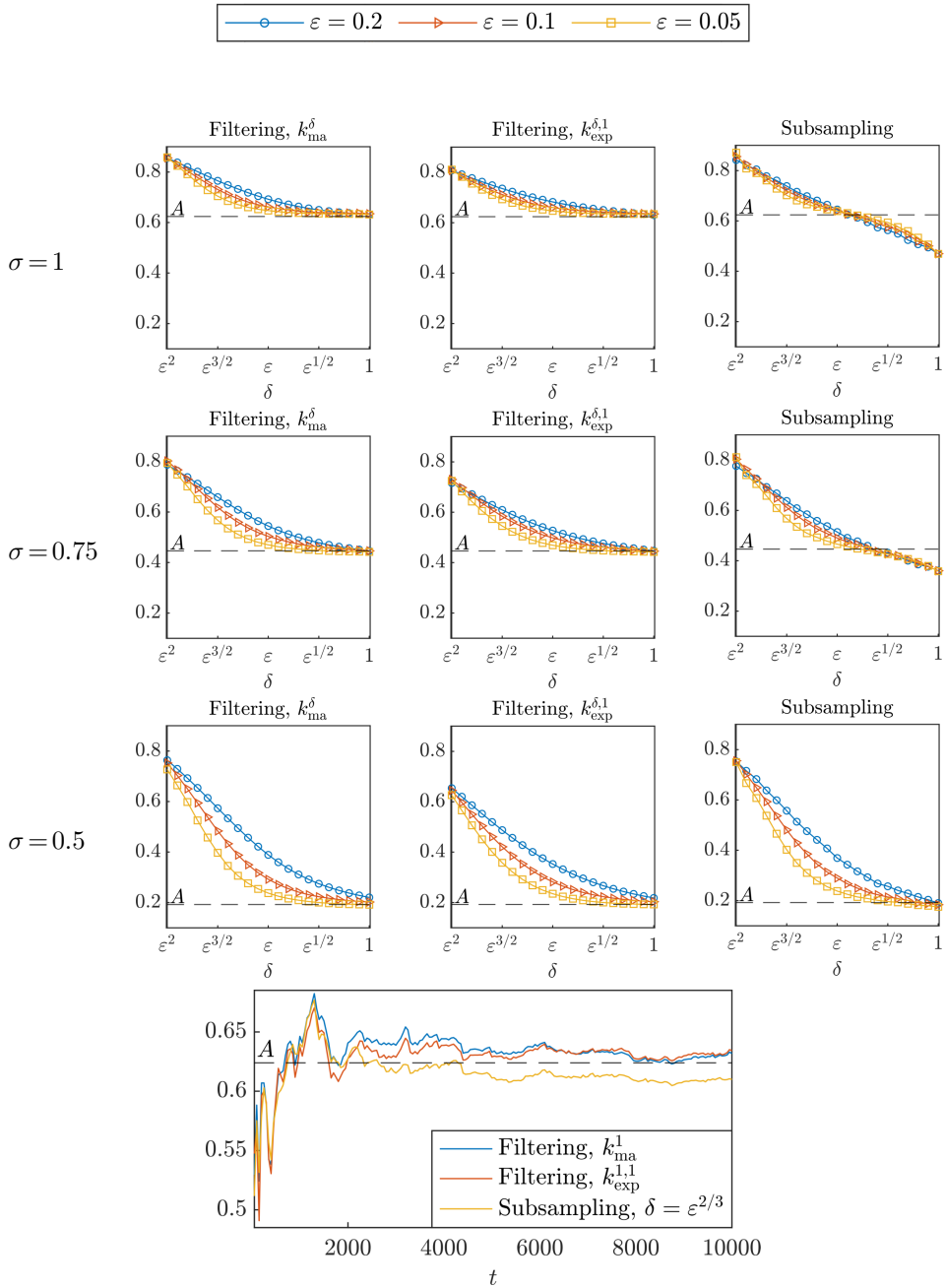


Fig. 3.1: Estimation of the drift coefficient of an effective Ornstein–Uhlenbeck equation. Top: Numerical results at final time $T = 10^4$ for variable $\sigma = 1, 0.75, 0.5$ (per row), for $\varepsilon = 0.2, 0.1, 0.05$ (blue, red, and yellow lines respectively), and for variable filtering/subsampling width δ (horizontal axis in all figures). Comparison between filtering with moving average and exponential filters (first two columns) and subsampling (last column). Bottom: Convergence with respect to $t \in [0, 10^4]$ of the two estimators based on filtered data with $\delta = 1$, and of the subsampling estimator with $\delta = \varepsilon^{2/3}$, for fixed $\sigma = 1$ and $\varepsilon = 0.05$. Remark: The legend on top is valid for all plots, except the last row.

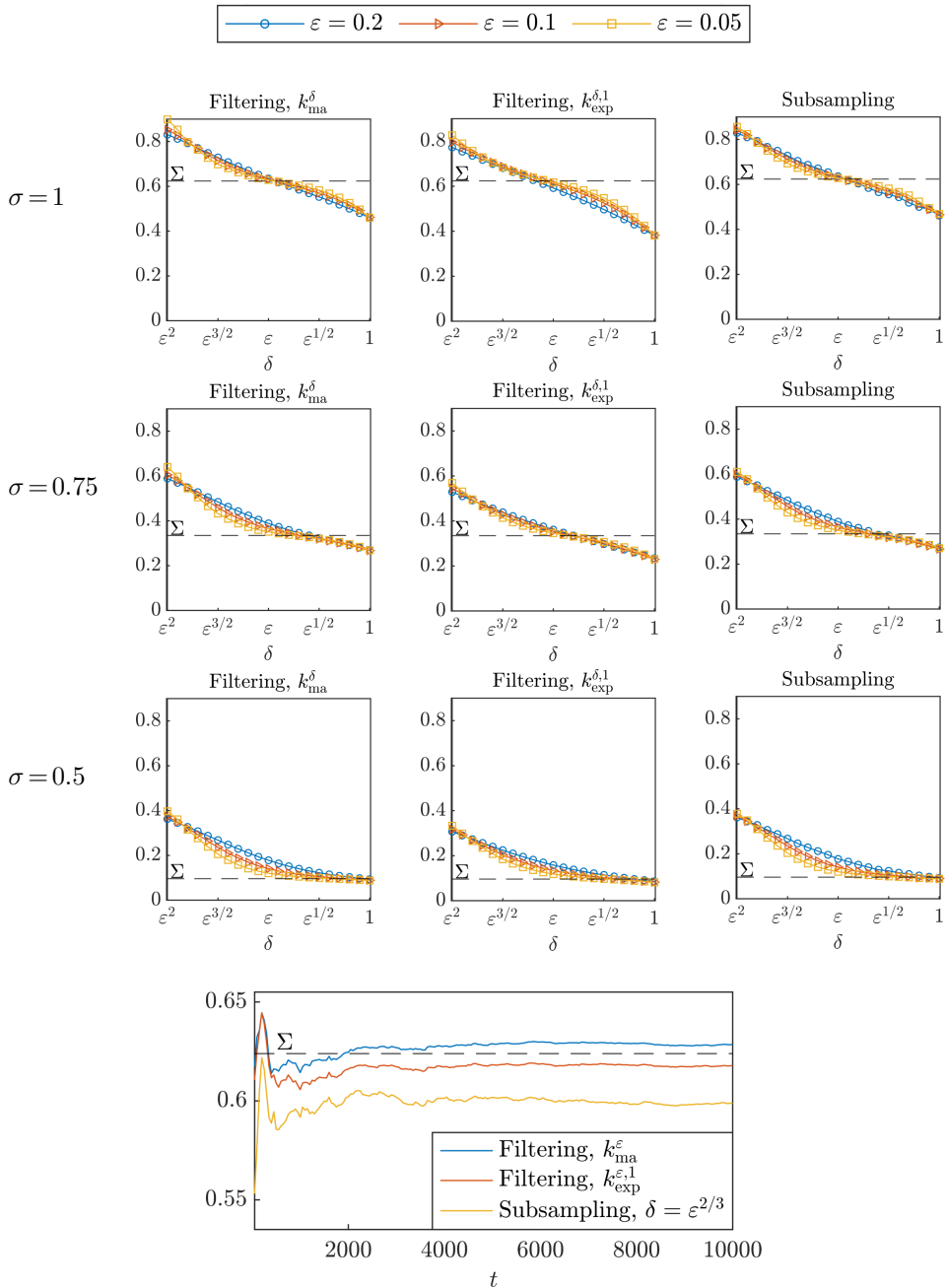


Fig. 3.2: Estimation of the diffusion coefficient of an effective Ornstein–Uhlenbeck equation. Top: Numerical results at final time $T = 10^4$ for variable $\sigma = 1, 0.75, 0.5$ (per row), for $\epsilon = 0.2, 0.1, 0.05$ (blue, red, and yellow lines respectively), and for variable filtering/subsampling width δ (horizontal axis in all figures). Results for the estimators $\hat{\Sigma}$: Comparison between filtering with moving average and exponential filters (first two columns) and subsampling (last column). Bottom: Convergence with respect to $t \in [0, 10^4]$ of the two estimators based on filtered data with $\delta = \epsilon$, and of the subsampling estimator with $\delta = \epsilon^{2/3}$, for fixed $\sigma = 1$ and $\epsilon = 0.05$. Remark: The legend on top is valid for all plots, except the last row.

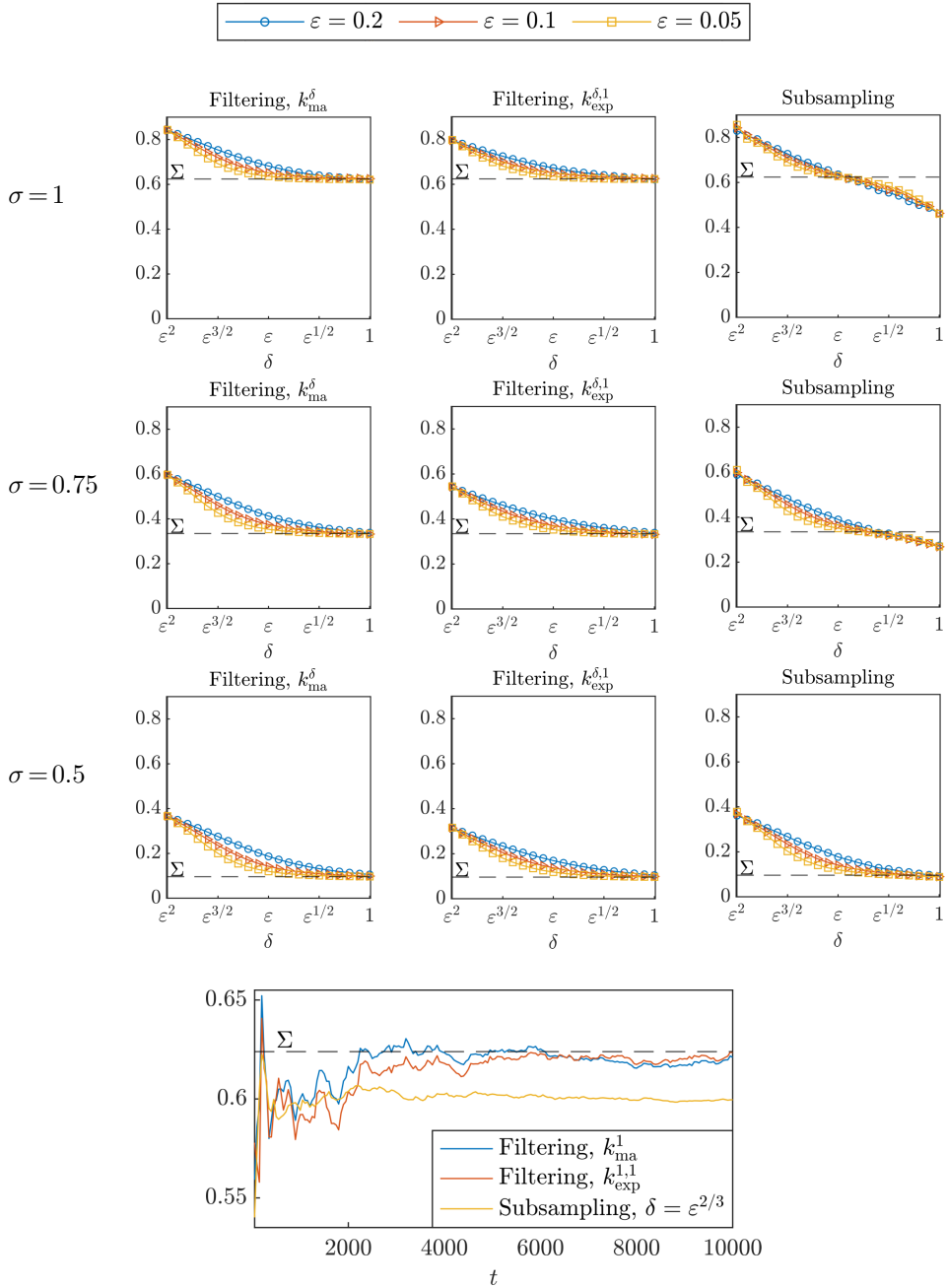


Fig. 3.3: Estimation of the diffusion coefficient of an effective Ornstein–Uhlenbeck equation. Top: Numerical results at final time $T = 10^4$ for variable $\sigma = 1, 0.75, 0.5$ (per row), for $\varepsilon = 0.2, 0.1, 0.05$ (blue, red, and yellow lines respectively), and for variable filtering/subsampling width δ (horizontal axis in all figures). Results for the estimators $\tilde{\Sigma}$: Comparison between filtering with moving average and exponential filters (first two columns) and subsampling (last column). Bottom: Convergence with respect to $t \in [0, 10^4]$ of the two estimators based on filtered data with $\delta = 1$, and of the subsampling estimator with $\delta = \varepsilon^{2/3}$, for fixed $\sigma = 1$ and $\varepsilon = 0.05$. Remark: The legend on top is valid for all plots, except the last row.

3.2. The Semi-parametric setting. We now consider the semi-parametric setting for a one-dimensional multiscale Langevin equation of the form (1.6). In particular, we consider the number of parameters $N=6$ and define $V: \mathbb{R} \rightarrow \mathbb{R}^N$ as

$$V(x) = \left(\frac{x^6}{6} \quad \frac{x^5}{5} \quad \frac{x^4}{4} \quad \frac{x^3}{3} \quad \frac{x^2}{2} \quad x \right)^\top.$$

The slow-scale potential is premultiplied by the six-dimensional drift coefficient $\alpha \in \mathbb{R}^6$

$$\alpha = (1 \quad -1 \quad -5.25 \quad 4.75 \quad 5 \quad -3)^\top.$$

With this choice, the slow-scale potential $\alpha \cdot V$ has three stable points. Moreover, we choose the fast-scale potential as $p = \sin(y)$, the diffusion coefficient $\sigma = 1$, and the multiscale parameter $\varepsilon = 0.05$. We then wish to infer the effective drift and diffusion coefficients $A \in \mathbb{R}^6$ and $\Sigma > 0$ from synthetic data $X^\varepsilon = (X^\varepsilon(t), 0 \leq t \leq T)$ with $T = 5 \cdot 10^4$, generated with the Euler–Maruyama method with time step $\Delta_t = \varepsilon^3$. In this case, the homogenization coefficient $\mathcal{K} \approx 0.62$. We then infer the effective drift and diffusion coefficients $A \in \mathbb{R}^6$ and $\Sigma > 0$ which define the homogenized Equation (1.7). Similarly to Section 3.1, we compare the two filtering methodologies (moving average and exponential kernels), and subsampling. Moreover, we compute for all strategies the effective drift estimator \hat{A} , and the effective diffusion estimators $\hat{\Sigma}$ and $\tilde{\Sigma}$.

Numerical results, given in Figures 3.4 and 3.5, demonstrate that

- (a) Figure 3.4: The six-dimensional effective drift coefficient is estimated accurately by both filtering-based methodologies, which yield comparable results both in terms of time convergence and of robustness with respect to the filtering width δ . At final time, both estimators have relative errors of magnitude 10^{-2} , and all six components of the drift coefficient are accurately retrieved. We note that for $\delta = 1$, i.e., the go-to implementation when ε is unknown, the moving average estimator appears to be slightly better than the one obtained with the exponential filter. Subsampling, conversely, does not enable to retrieve the drift coefficient accurately and strongly depends on the subsampling width δ . We remark that the optimal value for δ appears to be $\delta \approx \varepsilon^{3/2}$, which is surprising in view of the convergence result of [39]. Finally, we notice that for all values of δ the estimated drift function is visually almost identical to the effective drift, and is clearly differentiated from the slow component of the multiscale drift.
- (b) Figure 3.5: The diffusion coefficient is estimated more accurately by the estimator $\tilde{\Sigma}$ than $\hat{\Sigma}$ when employing filtered data. Indeed, for both filtering kernels the estimator $\tilde{\Sigma}$ is very robust with respect to the filtering width δ , and results are very accurate in case $\delta = 1$, the go-to implementation when the scale-separation parameter ε is unknown. Conversely, the estimator $\hat{\Sigma}$ strongly depends on the filtering/subsampling width. We note that choosing $\delta = \varepsilon$ the moving average estimator $\hat{\Sigma}_{\text{ma}}^\varepsilon$ outperforms the corresponding estimator $\hat{\Sigma}_{\text{exp}}^{\varepsilon,1}$ based on exponential filtering. For subsampling, the two estimators are equivalent in terms of accuracy, and are extremely dependent on the subsampling width δ . Equivalently to the drift estimator, the best inference results seem to be given by $\delta \approx \varepsilon^{3/2}$, and not at the conjectured optimal value $\varepsilon^{2/3}$.

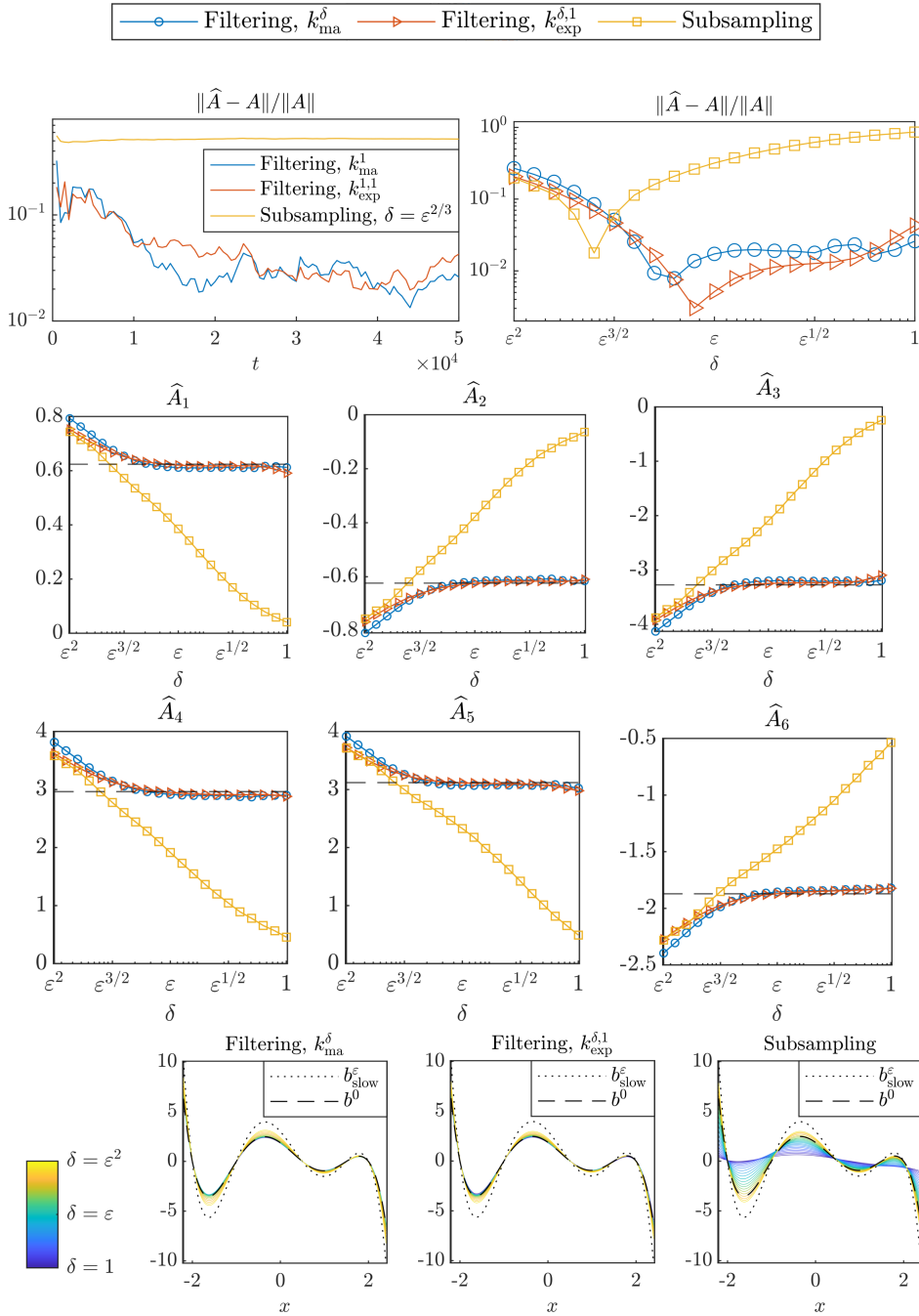


Fig. 3.4: Estimation of the drift coefficient in the one-dimensional semi-parametric setting. First row: on the left, we show the evolution of the relative error with respect to $t \in [0, 5 \cdot 10^4]$, and on the right the dependence of the relative error on the filtering/subsampling width $\delta \in [\varepsilon^2, 1]$. Second and third rows: Dependence on δ of the estimators for the components A_i , $i = 1, \dots, 6$ of the effective drift coefficient obtained with filtered data with both kernels, and with subsampling. Fourth row: Dependence on the filtering/subsampling width δ of the estimated drift function with the same three methodologies. Remark: The legend on top is valid for all plots, except the last row.

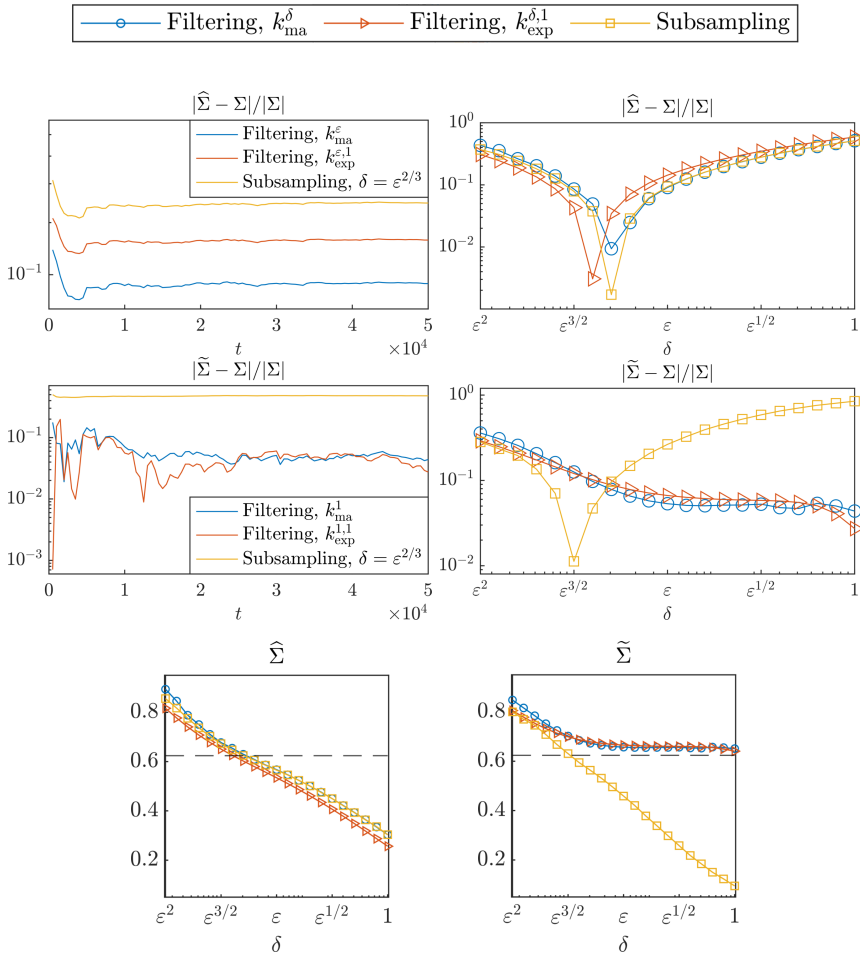


Fig. 3.5: Estimation of the diffusion coefficient in the one-dimensional semi-parametric setting. First ($\widehat{\Sigma}$) and second ($\widetilde{\Sigma}$) row: on the left, we show the evolution of the relative error with respect to $t \in [0, 5 \cdot 10^4]$, and on the right the dependence of the relative error with respect to the filtering/subsampling width $\delta \in [\varepsilon^2, 1]$. Third row: Dependence on δ of the estimators $\widehat{\Sigma}$ and $\widetilde{\Sigma}$ of the effective diffusion coefficient obtained with filtered data with both kernels, and with subsampling.

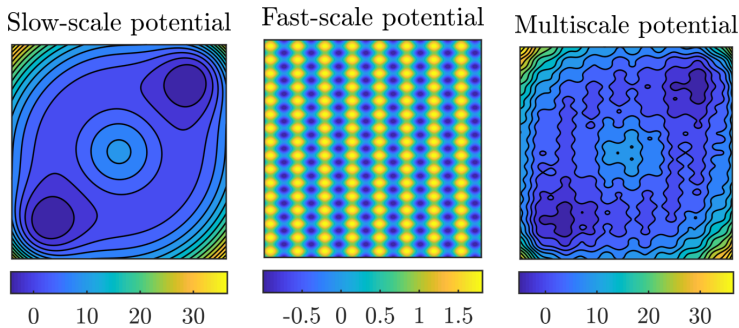


Fig. 3.6: Slow-, fast-, and multiscale potentials for the two-dimensional example of Section 3.3, depicted here in the square $(-2.5, 2.5)^2$.

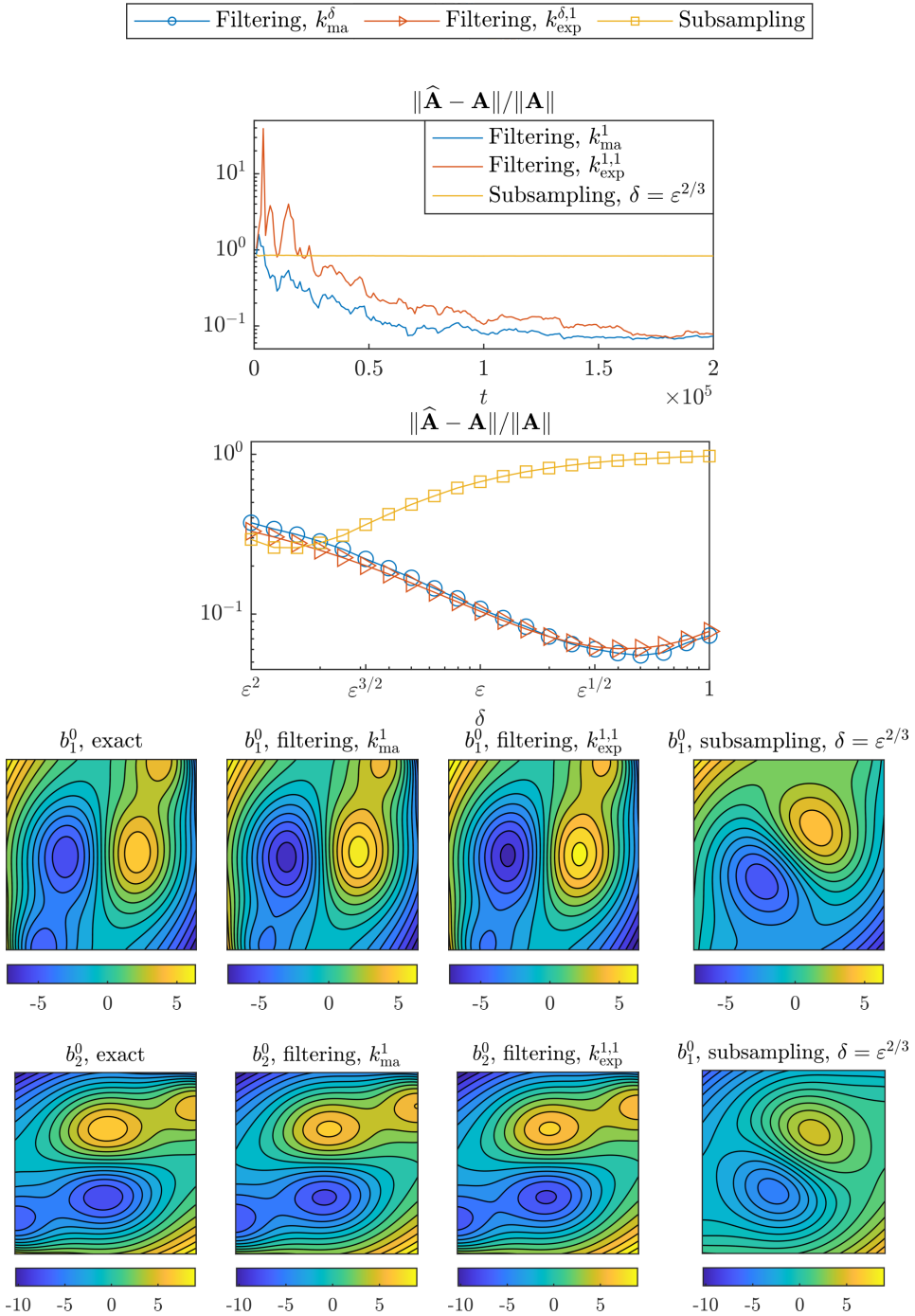


Fig. 3.7: Estimation of the effective drift coefficient for the two-dimensional example of Section 3.3. First row: Dependence of the relative error with respect to $t \in [0, T]$ (left) and to the sub-sampling/filtering width δ (right) for both filtering methods and subsampling. Second and third row: Graphical representation in the square $(-2.5, 2.5)^2$ of the estimated effective drift function at final time for the same three methods.

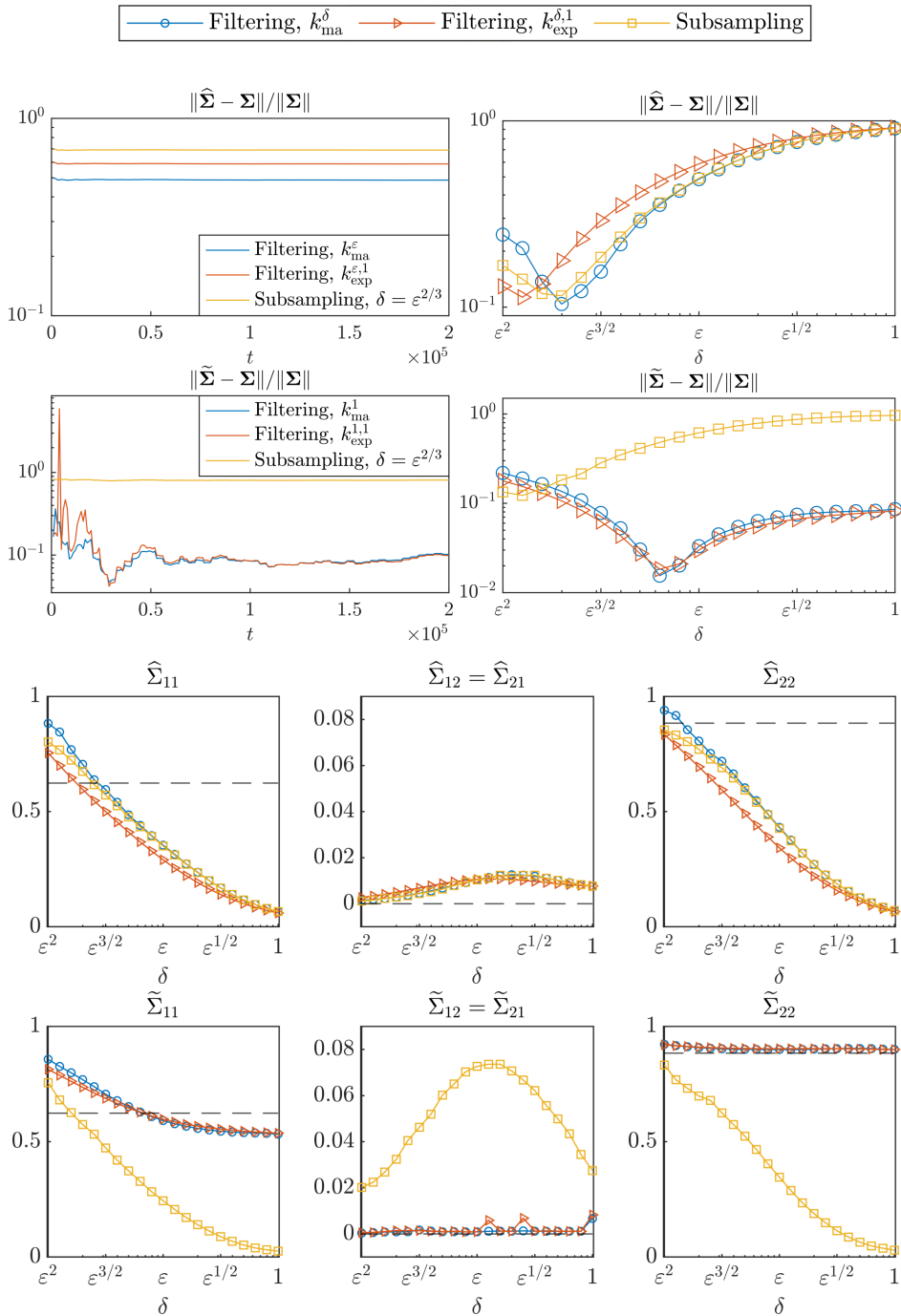


Fig. 3.8: Estimation of the effective diffusion coefficient for the two-dimensional example of Section 3.3. First and second row: Dependence of the relative error of the estimators $\hat{\Sigma}$ (first row) and $\tilde{\Sigma}$ (second row) with respect to $t \in [0, T]$ (left) and to the subsampling/filtering width δ (right) for both filtering methods and subsampling. Third and fourth row: Sensitivity of the estimators of the entries of the diffusion matrix Σ with respect to δ , with both “hat” (third row) and “tilde” (fourth row) estimators, and for the same three methodologies as above.

3.3. A Two-dimensional example. As a last numerical example, we consider a two-dimensional SDE ($d=2$) of the form (1.3). In particular, we let $N=4$ and define

$$\begin{aligned} V_1(x) &= \exp\left(-\|x-x_1\|^2\right), & V_2(x) &= \exp\left(-\|x-x_2\|^2\right), \\ V_3(x) &= \exp\left(-\|x\|^2\right), & V_4(x) &= \frac{1}{4}\|x\|^4, \end{aligned}$$

where $x_1=(2,2)^\top, x_2=(-2,-2)^\top$. The exact drift coefficient in the multiscale dynamics is defined by $\alpha_1=\alpha_2=-15, \alpha_3=10$ and $\alpha_4=1$. We choose the fast-scale periodic potential $p: \mathbb{R}^2 \rightarrow \mathbb{R}$ as

$$p(y) = \sin(y_1) + \sin^2(y_2),$$

and let the diffusion coefficient $\sigma=1$ and the scale-separation parameter $\varepsilon=0.1$. Since the fast-scale potential can be decomposed as $p(y)=p_1(y_1)+p_2(y_2)$, the homogenization coefficient \mathcal{K} is diagonal and its diagonal components can be computed employing the one-dimensional formula (3.1). In particular, we have

$$\mathcal{K} \approx \begin{pmatrix} 0.62 & 0 \\ 0 & 0.88 \end{pmatrix}.$$

We remark that choosing larger values for the diffusion coefficient σ makes the diagonal elements of the matrix \mathcal{K} close to 1. Hence, in this case the homogenized potential and the slow-scale component of the multiscale potential would be close, which in turn may lead to a misinterpretation of the numerical results in case $\delta=\varepsilon^\zeta$ with $\zeta>1$. The slow, fast, and multiscale potential functions are represented in Figure 3.6. The slow-scale potential presents two wells around the points x_1 and x_2 , a local maximum in the origin, and diverges outside any ball large enough and centered in the origin. The superposition of the slow and fast-scale potentials (evaluated in $y=x/\varepsilon$) perturbs the slow-scale potential and is responsible for an infinity of non-negligible local minima. We note that due to the local minima, the local maximum in the origin, and the two-dimensional setup, transitions between the potential wells are rare. This compromises the accuracy of the inference results, especially for the drift coefficient, unless final time is taken large enough.

We set $T=2 \cdot 10^5$ and generate synthetic observations $X^\varepsilon=(X^\varepsilon(t), 0 \leq t \leq T)$ by integrating (1.3) with the Euler–Maruyama method with time step $\Delta_t=\varepsilon^3$. The necessity to perform experiments over long time horizons is due to the bistable nature of our setting, and to the low probability of transitioning between the potential wells due to the components V_1 and V_2 of the potential. It would have been also possible to increase the probability of such transitions by increasing the value of the diffusion coefficient σ , which however would have been problematic, as explained above. We then estimate the effective drift coefficients $\{A_i \in \mathbb{R}^{2 \times 2}\}_{i=1}^4$ and the effective diffusion matrix $\Sigma \in \mathbb{R}^{2 \times 2}$ employing data filtered with the moving average and exponential kernels, and with subsampling for a comparison. For the drift coefficient, we measure accuracy by computing the relative error on the 16-dimensional vector obtained by stacking all coefficients of the four 2×2 effective drift matrices. Numerical results, given in Figures 3.7 and 3.8, demonstrate that

- (a) Figure 3.7: The drift estimator obtained with both filtering methodologies is extremely accurate, given the complexity of the setting and the high-dimensionality of the coefficient. In particular, for all values of $\delta \in [\varepsilon, 1]$ we obtain relative errors

below 10%. Moreover, implementing both filtering methodologies with $\delta=1$, i.e., when the scale-separation parameter is unknown, yields quasi-optimal results. We remark that the estimator $\widehat{A}_{\text{ma}}^1$ obtained with the moving average kernel and $\delta=1$ appears to converge sensibly faster with respect to $t \in [0, T]$ than the corresponding estimator $\widehat{A}_{\text{exp}}^1$ obtained with the exponential kernel. Conversely, the relative error for subsampling is dramatically higher, and subsampling should not in our opinion be employed in this high-dimensional setting. Always commenting on Figure 3.7, we note that the drift function estimated with both filtering methods at final time is visually almost indistinguishable from the exact effective drift function.

- (b) Figure 3.8: Likewise the numerical experiments of the previous sections, the diffusion estimators $\widehat{\Sigma}$ obtained with both filtering kernels and subsampling is not robust with respect to the filtering width, with the moving average kernel that seems to perform slightly better than the exponential kernel for $\delta=\varepsilon$, and than subsampling when $\delta=\varepsilon^{3/2}$. The estimators $\widetilde{\Sigma}$ obtained with the two filtering methods are instead extremely accurate at identifying both the diagonal components – especially Σ_{22} , and the zero off-diagonal elements. The subsampling-based estimator $\widetilde{\Sigma}_{\text{sub}}$, instead, suffers from the lack of accuracy of the drift estimator $\widehat{A}_{\text{sub}}^\delta$, and is not reliable. Always commenting on Figure 3.8, we remark that convergence with respect to $t \in [0, T]$ of the estimators $\widetilde{\Sigma}_{\text{ma}}^1$ and $\widetilde{\Sigma}_{\text{exp}}^{1,1}$ is similar, with the moving average filter seemingly less prone to instabilities for small t .

4. Asymptotic unbiasedness of the estimators

In this section we present the proof of Theorems 2.1 to 2.3 and Corollary 2.1, i.e., the results of asymptotic unbiasedness for our filtering-based estimators. In [3] the proofs of convergence are obtained with the kernel $k_{\text{exp}}^{\delta,1}$ by noticing that the original trajectory X^ε and its filtered version $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$ are solution of an hypoelliptic system of Itô SDEs. For higher values of $\beta > 1$, the system describing the evolution of X^ε and $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$ is not an Itô system due to the presence of a memory term. In case we consider the moving average kernel k_{ma}^δ which we study in this paper, the memory term simplifies to a constant delay. Hence, the evolution of the filtered trajectory $Z_{\text{ma}}^{\delta,\varepsilon}$ can be coupled with the original trajectory X^ε through the system of stochastic delay differential equations (SDDEs)

$$\begin{aligned} dX^\varepsilon(t) &= -\alpha \cdot V'(X^\varepsilon(t))dt - \frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon(t)}{\varepsilon} \right) dt + \sqrt{2\sigma} dW(t), \\ dZ_{\text{ma}}^{\delta,\varepsilon}(t) &= -\frac{1}{\delta} (X^\varepsilon(t-\delta) - X^\varepsilon(t)) dt. \end{aligned} \tag{4.1}$$

To be precise, the system above is a combination of an Itô SDE and a delay ordinary differential equation driven by a stochastic signal. Due to the theory of homogenization (see [12, Chapter 3], or [40, Chapter 18], or the proof of [3, Lemma 3.9]), if δ is independent of ε , the solution $(X^\varepsilon, Z_{\text{ma}}^{\delta,\varepsilon})$ converges in law as random variables in $C^0([0, T], \mathbb{R}^2)$ to the solution $(X^0, Z_{\text{ma}}^{\delta,0})$ of the system

$$\begin{aligned} dX^0(t) &= -A \cdot V'(X^0(t))dt + \sqrt{2\Sigma} dW(t), \\ dZ_{\text{ma}}^{\delta,0}(t) &= -\frac{1}{\delta} (X^0(t-\delta) - X^0(t))dt. \end{aligned} \tag{4.2}$$

In the following, we first focus on ergodic properties of the couples $(X^\varepsilon, Z_{\text{ma}}^{\delta,\varepsilon})$ and $(X^0, Z_{\text{ma}}^{\delta,0})$ evolving according to (4.1) and (4.2), respectively. Then, we employ the

invariant measures and the Fokker–Planck equations derived through the ergodicity theory to prove asymptotic unbiasedness. We remark that the strategy we adopt is similar to the one of [3]. Still, different techniques need to be employed due to the delay in the second equation of the systems (4.1) and (4.2).

4.1. Ergodic properties. It is well-known (see, e.g., [3, 39]) that X^ε is geometrically ergodic with invariant measure μ^ε on \mathbb{R} , whose density ρ^ε satisfying $\mu^\varepsilon(dx) = \rho^\varepsilon(x)dx$ takes the Gibbs form

$$\rho^\varepsilon(x) = \frac{1}{C_\mu^\varepsilon} \exp\left(-\frac{V_\varepsilon(x)}{\sigma}\right), \quad C_\mu^\varepsilon = \int_{\mathbb{R}} \exp\left(-\frac{V_\varepsilon(x)}{\sigma}\right) dx,$$

where

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

Moreover, an analogous result holds true for the homogenized process X^0 , which is geometrically ergodic with invariant measure μ^0 on \mathbb{R} , whose density ρ^0 satisfying $\mu^0(dx) = \rho^0(x)dx$ is given by

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

We now introduce a similar result of ergodicity for the couples $(X^\varepsilon, Z_{\text{ma}}^{\delta, \varepsilon})$ and $(X^0, Z_{\text{ma}}^{\delta, 0})$ satisfying (4.1) and (4.2), respectively, i.e., for the multiscale process and its filtered version.

PROPOSITION 4.1. *Under Assumption 2.1, the solution $(X^\varepsilon, Z_{\text{ma}}^{\delta, \varepsilon})$ of (4.1) is ergodic, and the density $\tilde{\rho}^\varepsilon$ of its invariant measure $\tilde{\mu}^\varepsilon$ on \mathbb{R}^2 , such that $\tilde{\mu}^\varepsilon(dx, dz) = \tilde{\rho}^\varepsilon(x, z)dx dz$, satisfies*

$$\begin{aligned} \sigma \partial_{xx}^2 \tilde{\rho}^\varepsilon(x, z) + \partial_x (V'_\varepsilon(x) \tilde{\rho}^\varepsilon(x, z)) + \frac{1}{\delta} \partial_z \left(\left(\int_{\mathbb{R}} y \psi^\varepsilon(y | x, z) dy - x \right) \tilde{\rho}^\varepsilon(x, z) \right) &= 0, \\ \int_{\mathbb{R}} \int_{\mathbb{R}} \tilde{\rho}^\varepsilon(x, z) dx dz &= 1, \end{aligned} \tag{4.3}$$

where, if $X^\varepsilon(0) \sim \mu^\varepsilon$, it holds

$$\int_{\mathbb{R}} y \psi^\varepsilon(y | x, z) dy = \mathbb{E} [X^\varepsilon(0) | X^\varepsilon(\delta) = x, Z_{\text{ma}}^{\delta, \varepsilon}(\delta) = z],$$

i.e., $\psi^\varepsilon(\cdot | x, z)$ is the conditional density of $X^\varepsilon(0)$ given $X^\varepsilon(\delta) = x$ and $Z_{\text{ma}}^{\delta, \varepsilon}(\delta) = z$. Moreover, if δ is independent of ε , the solution $(X^0, Z_{\text{ma}}^{\delta, 0})$ of (4.2) is ergodic, and the density $\tilde{\rho}^0$ of its invariant measure $\tilde{\mu}^0$ on \mathbb{R}^2 , such that $\tilde{\mu}^0(dx, dz) = \tilde{\rho}^0(x, z)dx dz$, satisfies

$$\begin{aligned} \Sigma \partial_{xx}^2 \tilde{\rho}^0(x, z) + \partial_x (A \cdot V'(x) \tilde{\rho}^0(x, z)) + \frac{1}{\delta} \partial_z \left(\left(\int_{\mathbb{R}} y \psi^0(y | x, z) dy - x \right) \tilde{\rho}^0(x, z) \right) &= 0, \\ \int_{\mathbb{R}} \int_{\mathbb{R}} \tilde{\rho}^0(x, z) dx dz &= 1, \end{aligned} \tag{4.4}$$

where, if $X^0(0) \sim \mu^0$, it holds

$$\int_{\mathbb{R}} y \psi^0(y | x, z) dy = \mathbb{E} [X^0(0) | X^0(\delta) = x, Z_{\text{ma}}^{\delta, 0}(\delta) = z],$$

i.e., $\psi^0(\cdot|x,z)$ is the conditional density of $X^0(0)$ given $X^0(\delta) = x$ and $Z_{\text{ma}}^{\delta,0}(\delta) = z$.

Proof. In order to prove that the joint process $(X^\varepsilon, Z_{\text{ma}}^{\delta,\varepsilon})$ is ergodic, we show that it admits a unique invariant measure. If $X^\varepsilon(0)$ is distributed accordingly to its invariant measure μ^ε , which exists due to Assumption 2.1, then the processes $(X^\varepsilon(s), 0 \leq s \leq \delta)$ and $(X^\varepsilon(s), t - \delta \leq s \leq t)$ are equally distributed for all $t \geq \delta$. Hence, the two-dimensional random variables $(X^\varepsilon(\delta), \frac{1}{\delta} \int_0^\delta X^\varepsilon(s) ds)$ and $(X^\varepsilon(t), \frac{1}{\delta} \int_{t-\delta}^t X^\varepsilon(s) ds)$ are equal in law for all $t \geq \delta$. Recalling that the joint process $(X^\varepsilon(t), \frac{1}{\delta} \int_{t-\delta}^t X^\varepsilon(s) ds)$ is the solution $(X^\varepsilon(t), Z_{\text{ma}}^{\delta,\varepsilon}(t))$ of the system (4.1), it follows that the invariant measure $\tilde{\mu}^\varepsilon$ on \mathbb{R}^2 is the law of the random variable $(X^\varepsilon(\delta), \frac{1}{\delta} \int_0^\delta X^\varepsilon(s) ds)$. The uniqueness of the invariant measure $\tilde{\mu}^\varepsilon$ is then a direct consequence of the uniqueness of the invariant measure μ^ε for the process X^ε since the joint measure $\tilde{\mu}^\varepsilon$ is uniquely determined by its marginal μ^ε . Moreover, the Fokker–Planck equation for the one-time PDF related to an SDDE with a single fixed delay is well-known (see, e.g., [19, 22, 32]) and the stationary Equation (4.3) for the density $\tilde{\rho}^\varepsilon$ of $\tilde{\mu}^\varepsilon$ is then obtained due to the particular form of the system (4.1). Finally, the results corresponding to the homogenized system (4.2) can be proved analogously. \square

The following formulas, which will be employed in the proof of the main results, are then direct consequences of the Fokker–Planck equations obtained above.

LEMMA 4.1. *Let $\tilde{\rho}^\varepsilon(x,z) = \rho^\varepsilon(x)\varphi^\varepsilon(z|x)$, where ρ^ε and $\tilde{\rho}^\varepsilon$ are the densities of the invariant measures μ^ε and $\tilde{\mu}^\varepsilon$ of X^ε and $(X^\varepsilon, Z_{\text{ma}}^{\delta,\varepsilon})$, respectively, and where φ^ε is the conditional density of $Z_{\text{ma}}^{\delta,\varepsilon}$ given X^ε . Then, if $X^\varepsilon(0) \sim \mu^\varepsilon$, it holds*

$$\begin{aligned} & \sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \rho^\varepsilon(x) \partial_x \varphi^\varepsilon(z|x) dx dz \\ &= \frac{1}{\delta} \mathbb{E}^{\tilde{\mu}^\varepsilon} [(X^\varepsilon(\delta) - Z_{\text{ma}}^{\delta,\varepsilon}(\delta)) (X^\varepsilon(\delta) - X^\varepsilon(0)) V''(Z_{\text{ma}}^{\delta,\varepsilon}(\delta))]. \end{aligned} \tag{4.5}$$

Moreover, if δ is independent of ε and writing $\tilde{\rho}^0(x,z) = \rho^0(x)\varphi^0(z|x)$ for the density of the homogenized invariant measure $\tilde{\mu}^0$ of $(X^0, Z_{\text{ma}}^{\delta,0})$, it holds

$$\begin{aligned} & \Sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \rho^0(x) \partial_x \varphi^0(z|x) dx dz \\ &= \frac{1}{\delta} \mathbb{E}^{\tilde{\mu}^0} [(X^0(\delta) - Z_{\text{ma}}^{\delta,0}(\delta)) (X^0(\delta) - X^0(0)) V''(Z_{\text{ma}}^{\delta,0}(\delta))]. \end{aligned} \tag{4.6}$$

Proof. We proceed similarly to the proof of [3, Lemma 3.5]. Replacing the decomposition $\tilde{\rho}^\varepsilon(x,z) = \rho^\varepsilon(x)\varphi^\varepsilon(z|x)$ into the Fokker–Planck Equation (4.3) gives

$$\partial_x (\sigma \rho^\varepsilon(x) \partial_x \varphi^\varepsilon(z|x)) + \partial_z \left(\frac{1}{\delta} \left(\int_{\mathbb{R}} y \psi^\varepsilon(y|x,z) dy - x \right) \rho^\varepsilon(x) \varphi^\varepsilon(z|x) \right) = 0.$$

We then multiply the equation above by a smooth function $f: \mathbb{R}^2 \rightarrow \mathbb{R}^N$, $f = f(x,z)$, and integrate first with respect to x and z and then by parts, obtaining

$$\sigma \int_{\mathbb{R}} \int_{\mathbb{R}} \partial_x f(x,z) \rho^\varepsilon(x) \partial_x \varphi^\varepsilon(z|x) dx dz = \frac{1}{\delta} \mathbb{E}^{\tilde{\mu}^\varepsilon} [\partial_z f(X^\varepsilon(\delta), Z_{\text{ma}}^{\delta,\varepsilon}(\delta)) (X^\varepsilon(\delta) - X^\varepsilon(0))].$$

The choice $f(x,z) = (x-z)V'(z) + V(z)$ gives Equation (4.5). Finally, Equation (4.6) is obtained analogously employing the Fokker–Planck equation of the homogenized SDE (4.4). \square

4.2. Preliminary results. Let us first introduce the notation

$$\begin{aligned} \widetilde{\mathcal{M}}_\varepsilon &:= \mathbb{E}^{\widetilde{\mu}^\varepsilon} [V'(Z_{\text{ma}}^{\delta,\varepsilon}) \otimes V'(X^\varepsilon)], & \widetilde{\mathcal{M}}_0 &:= \mathbb{E}^{\widetilde{\mu}^0} [V'(Z_{\text{ma}}^{\delta,0}) \otimes V'(X^0)], \\ \mathcal{M}_\varepsilon &:= \mathbb{E}^{\mu^\varepsilon} [V'(X^\varepsilon) \otimes V'(X^\varepsilon)], & \mathcal{M}_0 &:= \mathbb{E}^{\mu^0} [V'(X^0) \otimes V'(X^0)], \end{aligned}$$

which is repeatedly employed below. Before presenting the main proofs, we introduce two auxiliary lemmas.

LEMMA 4.2. *Under Assumption 2.1, it holds*

$$X^\varepsilon(\delta) - Z_{\text{ma}}^{\delta,\varepsilon}(\delta) = \frac{\sqrt{2\sigma}}{\delta} \int_0^\delta t(1 + \Phi'(Y^\varepsilon(t))) dW(t) + R(\varepsilon, \delta), \tag{4.7}$$

where Φ is the solution of the cell problem (1.5) and where the remainder $R(\varepsilon, \delta)$ satisfies for all $p \geq 1$ and a constant $C > 0$ independent of ε and δ

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} [|R(\varepsilon, \delta)|^p]^{1/p} \leq C(\varepsilon + \delta). \tag{4.8}$$

Moreover, if $X^\varepsilon(0)$ is stationary, i.e. $X^\varepsilon(0) \sim \mu^\varepsilon$, it holds

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} [|X^\varepsilon - Z_{\text{ma}}^{\delta,\varepsilon}|^p]^{1/p} \leq C(\delta^{1/2} + \varepsilon), \tag{4.9}$$

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} [|Z_{\text{ma}}^{\delta,\varepsilon}|^p]^{1/p} \leq C. \tag{4.10}$$

Proof. Employing the decomposition (5.8) in [39] and due to [39, Lemma 5.5, Proposition 5.8] we have for all $t \in [0, \delta]$

$$X^\varepsilon(\delta) = X^\varepsilon(t) + \sqrt{2\sigma} \int_t^\delta (1 + \Phi'(Y^\varepsilon(s))) dW(s) + R(\varepsilon, \delta), \tag{4.11}$$

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

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} [|R(\varepsilon, \delta)|^p]^{1/p} \leq C(\varepsilon + \delta).$$

Therefore, we obtain

$$\begin{aligned} X^\varepsilon(\delta) - Z_{\text{ma}}^{\delta,\varepsilon}(\delta) &= \frac{1}{\delta} \int_0^\delta (X^\varepsilon(\delta) - X^\varepsilon(t)) dt = \frac{\sqrt{2\sigma}}{\delta} \int_0^\delta \int_t^\delta (1 + \Phi'(Y^\varepsilon(s))) dW(s) dt + R(\varepsilon, \delta), \end{aligned}$$

which by Fubini's theorem yields

$$X^\varepsilon(\delta) - Z_{\text{ma}}^{\delta,\varepsilon}(\delta) = \frac{\sqrt{2\sigma}}{\delta} \int_0^\delta t(1 + \Phi'(Y^\varepsilon(t))) dW(t) + R(\varepsilon, \delta),$$

and proves (4.7) and (4.8). By the Itô isometry, it holds

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} \left[\left| \int_0^\delta t(1 + \Phi'(Y^\varepsilon(t))) dW(t) \right|^p \right]^{1/p} \leq C\delta^{3/2}, \tag{4.12}$$

which, together with (4.7), (4.8) and the proof of Proposition 4.1, gives (4.9). Finally, (4.10) is proved by applying the triangle inequality and due to (4.9) and [39, Corollary 5.4]. \square

LEMMA 4.3. *Under Assumption 2.1 and if $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$, then it holds*

$$\lim_{\varepsilon \rightarrow 0} \widetilde{\mathcal{M}}_\varepsilon = \mathcal{M}_0.$$

Proof. By the triangle inequality, we have

$$\left\| \widetilde{\mathcal{M}}_\varepsilon - \mathcal{M}_0 \right\| \leq \left\| \widetilde{\mathcal{M}}_\varepsilon - \mathcal{M}_\varepsilon \right\| + \left\| \mathcal{M}_\varepsilon - \mathcal{M}_0 \right\|.$$

The first term vanishes as $\varepsilon \rightarrow 0$ due to Lemma 4.2, [39, Corollary 5.4] and since V' is Lipschitz under Assumption 2.1. The second term vanishes due to the theory of homogenization as $\varepsilon \rightarrow 0$. \square

4.3. Proof of the main results. We can now prove our main results, i.e., Theorems 2.1 to 2.3 and Corollary 2.1.

Proof. (Proof of Theorem 2.1.) Following the proof of [3, Theorem 3.12], we have

$$\widehat{A}_{\text{ma}}^\delta(X^\varepsilon, T) = \alpha + I_1 - I_2,$$

where

$$I_1 = \frac{1}{T} M_{\text{ma}}^\delta(X^\varepsilon, T)^{-1} \int_0^T \frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon(t)}{\varepsilon} \right) V'(Z_{\text{ma}}^{\delta, \varepsilon}(t)) dt,$$

$$I_2 = \frac{\sqrt{2\sigma}}{T} M_{\text{ma}}^\delta(X^\varepsilon, T)^{-1} \int_0^T V'(Z_{\text{ma}}^{\delta, \varepsilon}(t)) dW(t),$$

and where

$$\lim_{T \rightarrow \infty} I_2 = 0,$$

uniformly in ε by Lemma 4.2 and the strong law of large numbers for martingales. Considering I_1 , due to Assumption 2.1 the ergodic theorem and an integration by parts yield

$$\lim_{T \rightarrow \infty} I_1 = -\alpha + \widetilde{\mathcal{M}}_\varepsilon^{-1} \sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \rho^\varepsilon(x) \partial_x \varphi^\varepsilon(z|x) dx dz,$$

where $\varphi^\varepsilon(z|x)$ is defined in Lemma 4.1, which also implies

$$\lim_{T \rightarrow \infty} I_1 = -\alpha + \mathcal{A}^\varepsilon(\delta),$$

where

$$\mathcal{A}^\varepsilon(\delta) = \frac{1}{\delta} \widetilde{\mathcal{M}}_\varepsilon^{-1} \mathbb{E}^{\widetilde{\mu}^\varepsilon} [(X^\varepsilon(\delta) - Z_{\text{ma}}^{\delta, \varepsilon}(\delta))(X^\varepsilon(\delta) - X^\varepsilon(0)) V''(Z_{\text{ma}}^{\delta, \varepsilon}(\delta))]. \tag{4.13}$$

It remains to show that

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = A,$$

for which we consider two cases, corresponding to δ independent of ε and $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$, respectively.

Case 1: δ independent of ε . In this case, the theory of homogenization yields

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = \frac{1}{\delta} \widetilde{\mathcal{M}}_0^{-1} \mathbb{E}^{\widetilde{\mu}^0} [(X^0(\delta) - Z_{\text{ma}}^{\delta,0}(\delta))(X^0(\delta) - X^0(0))V''(Z_{\text{ma}}^{\delta,0}(\delta))],$$

so that applying Lemma 4.1 for the homogenized equation backwards we have

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = \widetilde{\mathcal{M}}_0^{-1} \Sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z) \rho^0(x) \partial_x \varphi^0(z|x) dx dz.$$

An integration by parts then gives

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\delta) = \widetilde{\mathcal{M}}_0^{-1} \widetilde{\mathcal{M}}_0 A = A,$$

which concludes *Case 1*.

Case 2: $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$. Replacing formulas (4.11) with $t=0$ and (4.7) into (4.13) gives

$$\begin{aligned} & \mathcal{A}^\varepsilon(\delta) \\ &= \frac{2\sigma}{\delta^2} \widetilde{\mathcal{M}}_\varepsilon^{-1} \mathbb{E}^{\widetilde{\mu}^\varepsilon} \left[\left(\int_0^\delta t(1 + \Phi'(Y^\varepsilon(t))) dW(t) \right) \left(\int_0^\delta (1 + \Phi'(Y^\varepsilon(t))) dW(t) \right) V''(Z_{\text{ma}}^{\delta,\varepsilon}(\delta)) \right] \\ & \quad + \widetilde{R}_1(\varepsilon, \delta), \end{aligned}$$

where, due to Lemma 4.2, estimate (4.12) and the fact that by the Itô isometry

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} \left[\left| \int_t^\delta (1 + \Phi'(Y^\varepsilon(s))) dW(s) \right|^p \right]^{1/p} \leq C \delta^{1/2}, \tag{4.14}$$

it follows that the remainder satisfies

$$\|\widetilde{R}_1(\varepsilon, \delta)\| \leq C (\delta^{1/2} + \varepsilon \delta^{-1/2} + \varepsilon^2 \delta^{-1}). \tag{4.15}$$

Moreover, since V'' is Lipschitz under Assumption 2.1 and due to the triangle inequality, Equation (4.11), estimates (4.8), (4.14) and Lemma 4.2, it holds for all $t \in [0, \delta]$

$$\mathbb{E}^{\widetilde{\mu}^\varepsilon} \left[\|V''(Z_{\text{ma}}^{\delta,\varepsilon}(\delta)) - V''(X^\varepsilon(t))\|^p \right]^{1/p} \leq C (\varepsilon + \delta^{1/2}), \tag{4.16}$$

which for ε and δ sufficiently small is at most of order $\mathcal{O}(\|\widetilde{R}_1(\varepsilon, \delta)\|)$. Hence, by the Itô isometry

$$\begin{aligned} & \mathcal{A}^\varepsilon(\delta) \\ &= \frac{2\sigma}{\delta^2} \widetilde{\mathcal{M}}_\varepsilon^{-1} \mathbb{E}^{\widetilde{\mu}^\varepsilon} \left[\left(\int_0^\delta t(1 + \Phi'(Y^\varepsilon(t))) dW(t) \right) \left(\int_0^\delta (1 + \Phi'(Y^\varepsilon(t))) V''(X^\varepsilon(t)) dW(t) \right) \right] \\ & \quad + \widetilde{R}_2(\varepsilon, \delta) \\ &= \frac{2\sigma}{\delta^2} \widetilde{\mathcal{M}}_\varepsilon^{-1} \int_0^\delta t \mathbb{E}^{\widetilde{\mu}^\varepsilon} [(1 + \Phi'(Y^\varepsilon(t)))^2 V''(X^\varepsilon(t))] dt + \widetilde{R}_2(\varepsilon, \delta), \end{aligned}$$

where due to (4.15) and (4.16) the remainder satisfies

$$\left\| \tilde{R}_2(\varepsilon, \delta) \right\| \leq C \left(\delta^{1/2} + \varepsilon \delta^{-1/2} + \varepsilon^2 \delta^{-1} \right).$$

Repeating the last part of the proof of [3, Lemma 3.17], we then obtain

$$\begin{aligned} \mathcal{A}^\varepsilon(\delta) &= \frac{2\sigma\mathcal{K}}{\delta^2} \widetilde{\mathcal{M}}_\varepsilon^{-1} \mathbb{E}^{\mu^0} [V''(X^0)] \int_0^\delta t dt + \tilde{R}_2(\varepsilon, \delta) \\ &= \Sigma \widetilde{\mathcal{M}}_\varepsilon^{-1} \mathbb{E}^{\mu^0} [V''(X^0)] + \tilde{R}_2(\varepsilon, \delta). \end{aligned}$$

Finally, since $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$, by (4.15) and due to Lemma 4.3 we obtain

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

and an integration by parts gives

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

which proves *Case 2* and therefore concludes the proof. □

Proof. (Proof of Theorem 2.2.) Due to Assumption 2.1 the ergodic theorem gives

$$\lim_{T \rightarrow \infty} \widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T) = \frac{1}{\delta} \mathbb{E}^{\tilde{\mu}^\varepsilon} \left[(X^\varepsilon(\delta) - Z_{\text{ma}}^{\delta, \varepsilon}(\delta)) (X^\varepsilon(\delta) - X^\varepsilon(0)) \right]. \quad (4.17)$$

Following step-by-step *Case 2* of the proof of Theorem 2.1 with the value 1 instead of $V''(Z_{\text{ma}}^{\delta, \varepsilon}(\delta))$, and without the pre-multiplication by $\widetilde{\mathcal{M}}_\varepsilon^{-1}$, we obtain the desired result. □

REMARK 4.1. It is clear from the proof of Theorem 2.2 that it is theoretically not possible to choose δ independent of ε in the computation of $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$. Let $N = 1$ and $V(x) = x^2/2$, so that X^0 is an Ornstein-Uhlenbeck process. In this case, the process X^0 is a Gaussian process such that at stationarity $X^0 \sim \mathcal{GP}(0, \mathcal{C}(t, s))$ where

$$\mathcal{C}(t, s) = \frac{\Sigma}{A} e^{-A|t-s|}. \quad (4.18)$$

By (4.17) and (4.18) we can therefore explicitly compute

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T) = \frac{1}{\delta} \mathbb{E}^{\tilde{\mu}^0} \left[(X^0(\delta) - Z_{\text{ma}}^{\delta, 0}(\delta)) (X^0(\delta) - X^0(0)) \right] = \frac{1 - e^{-\delta A}}{\delta A} \Sigma,$$

so that $\widehat{\Sigma}_{\text{ma}}^\delta(X^\varepsilon, T)$ is asymptotically unbiased only if $\delta \rightarrow 0$.

Proof. (Proof of Theorem 2.3.) Notice that due to (1.9) we have

$$\frac{\langle X^\varepsilon \rangle_T}{2T} = \sigma,$$

which together with (2.8) and by (1.9) implies

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \widetilde{\Sigma}(X^\varepsilon, T) = \frac{\alpha^\top A}{(\alpha^\top \alpha)} \sigma.$$

Finally, since $A = \mathcal{K}\alpha$ and $\Sigma = \mathcal{K}\sigma$ we obtain

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \tilde{\Sigma}(X^\varepsilon, T) = \mathcal{K}\sigma = \Sigma,$$

which is the desired result. □

Proof. (Proof of Corollary 2.1.) The desired results follow directly from Theorem 2.1, [3, Theorems 3.12 and 3.18], [39, Theorem 3.5] and Theorem 2.3. We remark that the limit in [3, Theorem 3.18] holds true also a.s. and the proof of [39, Theorem 3.5] can be modified (see Remark 4.2) such that hypothesis (2.8) is satisfied. □

REMARK 4.2. The proof of [39, Theorem 3.5] can be modified in order to show that the estimator $\hat{A}_{\text{sub}}(X^\varepsilon, T)$ given in (1.10) satisfies

$$\lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \hat{A}_{\text{sub}}(X^\varepsilon, T) = A, \quad \text{a.s.} \tag{4.19}$$

Due to Assumption 2.1 Item ii and by the ergodic theorem we have

$$\lim_{T \rightarrow \infty} \hat{A}_{\text{sub}}(X^\varepsilon, T) = - \frac{\mathbb{E}^{\mu^\varepsilon} [V'(X^\varepsilon(0))(X^\varepsilon(\delta) - X^\varepsilon(0))]}{\mathbb{E}^{\mu^\varepsilon} [V'(X^\varepsilon(0))^2]}.$$

We then employ [5, Lemma 5.2] with f the identity function and $\Delta = \delta$ and we notice that the martingale

$$M^\varepsilon(t) := \sqrt{2\sigma} \int_0^t (1 + \Phi'(Y^\varepsilon(s))) dW(s),$$

where Φ is defined in (1.5) and $Y^\varepsilon(s) = X^\varepsilon(s)$, is such that $M^\varepsilon(0) = 0$. Therefore, we obtain

$$\lim_{T \rightarrow \infty} \hat{A}_{\text{sub}}(X^\varepsilon, T) = A + \tilde{R}(\varepsilon, \delta),$$

where the remainder satisfies for a constant $C > 0$ independent of ε and δ

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

Finally, since $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$ we deduce the desired result (4.19).

5. Conclusion

In this work, we introduced a novel methodology for inferring effective diffusions from observations of multiscale dynamics based on filtering the data with moving averages. Asymptotic unbiasedness is rigorously proved by originally exploiting an ergodicity result for SDDEs. Our method is robust, easy to implement, computationally uninvolved, and outperforms the standard technique of subsampling on a range of test cases. Moreover, the performances are comparable to a similar class of estimators that we introduced in our previous work [3]. The accuracy of our methodology in the multiscale, multi-dimensional, and highly-parametrised case is surprisingly high in view of its simplicity and low computational involvement.

We believe that future developments could go in the direction of:

- (i) Deriving asymptotically unbiased estimators for the diffusion coefficient which are robust in practice and do not rely on the drift estimator,
- (ii) Extending the filtered data methodology to multiscale SDEs with non-constant diffusion terms, or for drift functions which do not depend linearly on the parameters,
- (iii) Extending the analysis to the non-parametric case.

Acknowledgements. We thank the anonymous reviewer whose comments helped improve and clarify this manuscript. We are grateful to Assyr Abdulle and Grigorios A. Pavliotis for invaluable support and advice over the years.

REFERENCES

- [1] A. Abdulle and A. Di Blasio, *Numerical homogenization and model order reduction for multiscale inverse problems*, *Multiscale Model. Simul.*, **17**:399–433, 2019. 1.3
- [2] A. Abdulle and A. Di Blasio, *A Bayesian numerical homogenization method for elliptic multiscale inverse problems*, *SIAM/ASA J. Uncertain. Quantif.*, **8**:414–450, 2020. 1.3
- [3] A. Abdulle, G. Garegnani, G.A. Pavliotis, A.M. Stuart, and A. Zaroni, *Drift estimation of multiscale diffusions based on filtered data*, *Found. Comput. Math.*, **2021**. 1.3, 1.4, 2, 2, 2.1, 2.2, 2.3, 2.3, 4, 4, 4, 4.1, 4.1, 4.3, 4.3, 4.3, 5
- [4] A. Abdulle, G. Garegnani, and A. Zaroni, *Ensemble Kalman filter for multiscale inverse problems*, *Multiscale Model. Simul.*, **18**:1565–1594, 2020. 1.3
- [5] A. Abdulle, G.A. Pavliotis, and A. Zaroni, *Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions*, *Stat. Comput.*, **32**:34, 2022. 1.3, 2.3, 4.2
- [6] Y. Ait-Sahalia and J. Jacod, *High-Frequency Financial Econometrics*, Princeton University Press, 2014. 1, 1.3
- [7] Y. Ait-Sahalia, P.A. Mykland, and L. Zhang, *How often to sample a continuous-time process in the presence of market microstructure noise*, *Rev. Financ. Stud.*, **18**:351–416, 2005. 1, 1.3
- [8] R. Azencott, A. Beri, A. Jain, and I. Timofeyev, *Sub-sampling and parametric estimation for multiscale dynamics*, *Commun. Math. Sci.*, **11**:939–970, 2013. 1.3
- [9] R. Azencott, A. Beri, and I. Timofeyev, *Adaptive sub-sampling for parametric estimation of Gaussian diffusions*, *J. Stat. Phys.*, **139**:1066–1089, 2010. 1.3
- [10] R. Azencott, A. Beri, and I. Timofeyev, *Parametric estimation of stationary stochastic processes under indirect observability*, *J. Stat. Phys.*, **144**:150–170, 2011. 1.3
- [11] I.V. Basawa and B.L.S. Prakasa Rao, *Statistical Inference for Stochastic Processes*, Academic Press, London-New York, 1980. 1.3
- [12] A. Bensoussan, J.L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, North-Holland Publishing Co., Amsterdam, 1978. 1, 4
- [13] J.P.N. Bishwal, *Parameter Estimation in Stochastic Differential Equations*, Lecture Notes in Mathematics, Springer, Berlin, 2008. 1.3
- [14] M. Coghi, T. Nilssen, N. Nüsken, and S. Reich, *Rough Mckean–Vlasov dynamics for robust ensemble Kalman filtering*, arXiv preprint, [arXiv:2107.06621](https://arxiv.org/abs/2107.06621), 2022. 1.3
- [15] C.J. Cotter and G.A. Pavliotis, *Estimating eddy diffusivities from noisy Lagrangian observations*, *Commun. Math. Sci.*, **7**:805–838, 2009. 1
- [16] D. Crommelin and E. Vanden-Eijnden, *Fitting timeseries by continuous-time Markov chains: a quadratic programming approach*, *J. Comput. Phys.*, **217**:782–805, 2006. 1.3
- [17] D. Crommelin and E. Vanden-Eijnden, *Reconstruction of diffusions using spectral data from timeseries*, *Commun. Math. Sci.*, **4**:651–668, 2006. 1.3
- [18] D. Crommelin and E. Vanden-Eijnden, *Diffusion estimation from multiscale data by operator eigenpairs*, *Multiscale Model. Simul.*, **9**(4):1588–1623, 2011. 1.3
- [19] T.D. Frank, *Delay Fokker–Planck equations, Novikov’s theorem, and Boltzmann distributions as small delay approximations*, *Phys. Rev. E*, **72**:011112, 2005. 4.1
- [20] S. Gailus and K. Spiliopoulos, *Statistical inference for perturbed multiscale dynamical systems*, *Stochast. Process. Appl.*, **127**:419–448, 2017. 1.3
- [21] S. Gailus and K. Spiliopoulos, *Discrete-time statistical inference for multiscale diffusions*, *Multiscale Model. Simul.*, **16**:1824–1858, 2018. 1.3
- [22] S. Guillouzic, I. L’Heureux, and A. Longtin, *Small delay approximation of stochastic delay differential equations*, *Phys. Rev. E*, **59**:3970–3982, 1999. 4.1
- [23] 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**:314–328, 2015. 1.3
- [24] M. Kessler and M. Sørensen, *Estimating equations based on eigenfunctions for a discretely observed diffusion process*, *Bernoulli*, **5**:299–314, 1999. 1.3
- [25] S. Krumscheid, G.A. Pavliotis, and S. Kalliadasis, *Semiparametric drift and diffusion estimation for multiscale diffusions*, *Multiscale Model. Simul.*, **11**:442–473, 2013. 1.3
- [26] 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**:042139, 2015. 1, 1.3
- [27] Y.A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*, Springer Series in Statistics,

- Springer-Verlag, London, 2004. 1.3
- [28] F. Legoll and T. Lelièvre, *Some remarks on free energy and coarse-graining*, in B. Engquist, O. Runborg, and Y.H. Tsai (eds.), *Numerical Analysis of Multiscale Computations*, Springer, Heidelberg, 279–329, 2012. 1
- [29] T. Lelièvre and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, *Acta Numer.*, 25:681–880, 2016. 1
- [30] R.S. Liptser and A.N. Shiryaev, *Statistics of Random Processes. I. General Theory*, Translated from the 1974 Russian original by A.B. Aries, Springer-Verlag, 5, 2001. 1.2
- [31] R.S. Liptser and A.N. Shiryaev, *Statistics of Random Processes. II. Applications*, Translated from the 1974 Russian original by A.B. Aries, Springer-Verlag, 6, 2001. 1.2
- [32] S.A.M. Loos and S.H.L. Klapp, *Fokker–Planck equations for time-delayed systems via Markovian embedding*, *J. Stat. Phys.*, 177:95–118, 2019. 4.1
- [33] J. Nolen, G.A. Pavliotis, and A.M. Stuart, *Multiscale modeling and inverse problems*, in I. Graham, T. Hou, O. Lakkis, and R. Scheichl (eds.), *Numerical Analysis of Multiscale Problems*, Springer, Heidelberg, 1–34, 2012. 1.3
- [34] S.C. Olhede, A.M. Sykulski, and G.A. Pavliotis, *Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise*, *Multiscale Model. Simul.*, 8:393–427, 2010. 1, 1.3
- [35] A. Papanicolaou and K. Spiliopoulos, *Filtering the maximum likelihood for multiscale problems*, *Multiscale Model. Simul.*, 12:1193–1229, 2014. 1.3
- [36] A. Papavasiliou, G.A. Pavliotis, and A.M. Stuart, *Maximum likelihood drift estimation for multiscale diffusions*, *Stochast. Process. Appl.*, 119:3173–3210, 2009. 1, 1.3
- [37] G.A. Pavliotis, *Stochastic Processes and Applications*, Texts in Applied Mathematics, Springer, New York, 60, 2014. 1.2
- [38] G.A. Pavliotis, Y. Pokern, and A.M. Stuart, *Parameter estimation for multiscale diffusions: an overview*, in M. Kessler, A. Lindner, and M. Sorensen (eds.), *Statistical Methods for Stochastic Differential Equations*, CRC Press, Boca Raton, FL, 124:429–472, 2012. 1.3
- [39] G.A. Pavliotis and A.M. Stuart, *Parameter estimation for multiscale diffusions*, *J. Stat. Phys.*, 127:741–781, 2007. 1, 1.2, 1.3, 1.3, 2.3, ii, 2.3, a, a, 4.1, 4.2, 4.2, 4.2, 4.3, 4.2
- [40] G.A. Pavliotis and A.M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts in Applied Mathematics, Springer, New York, 53, 2008. 1, 1.1, 4
- [41] Y. Pokern, A.M. Stuart, and J.H. van Zanten, *Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs*, *Stochast. Process. Appl.*, 123:603–628, 2013. 1.3
- [42] Y. Pokern, A.M. Stuart, and E. Vanden-Eijnden, *Remarks on drift estimation for diffusion processes*, *Multiscale Model. Simul.*, 8:69–95, 2009. 1.2, 1.3
- [43] B.L.S. Prakasa Rao, *Statistical Inference for Diffusion Type Processes: Kendall’s Library of Statistics 8*, Edward Arnold, London; Oxford University Press, New York, 8, 1999. 1.3
- [44] K. Spiliopoulos and A. Chronopoulou, *Maximum likelihood estimation for small noise multiscale diffusions*, *Stat. Inference Stoch. Process.*, 16:237–266, 2013. 1.3
- [45] L. Zhang, P.A. Mykland, and Y. Ait-Sahalia, *A tale of two time scales: determining integrated volatility with noisy high-frequency data*, *J. Amer. Statist. Assoc.*, 100:1394–1411, 2005. 1, 1.3