Thèse n° 10 458

Filtered data and eigenfunction estimators for statistical inference of multiscale and interacting diffusion processes

Présentée le 19 décembre 2022

Faculté des sciences de base
Calcul scientifique et quantification de l'incertitude - Chaire CADMOS
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Andrea ZANONI

Acceptée sur proposition du jury

Prof. C. Hongler, président du jury
Prof. F. Nobile, Prof. G. A. Pavliotis, directeurs de thèse
Prof. S. Reich, rapporteur
Prof. K. Spiliopoulos, rapporteur
Prof. L. Chizat, rapporteur

Mention spéciale
Prof. Assyr Abdulle (1971-2021), ancien directeur de thèse

École
polytechnique
fédérale
de Lausanne

2022

À Assyr

Acknowledgements

I like to consider myself not to be the only author of this thesis, indeed this work would have not been possible without the direct or indirect help of many other people.

First, I would like to thank my three supervisors both for showing me how a good supervisor is, and for being great mathematicians with a large expertise in several areas of math. Assyr, who introduced me to the research world giving me the possibility to begin this beautiful journey. Thank you for teaching me to never stop trying, both in math, when a proof does not work, and in life. I really hope you would be proud of this work. Fabio, who adopted me in his group in the middle of the path. Thank you for your time and precious help, and for letting me discover new adventures that are important for a PhD student: co-supervision of students and conferences. Greg, with whom I collaborated a lot, and who hosted me for one month at Imperial College in London. Thank you for your kindness, and for being a huge source of ideas of interesting problems, which always kept me enthusiastic. If I will ever be a supervisor, I would like to take something from all of you.

I am also thankful to the members of the jury of my private defense, Prof. Hongler for accepting to be the President, and Prof. Chizat, Prof. Reich, and Prof. Spiliopoulos for reading this thesis and providing interesting questions and useful comments.

A big thank to the secretaries Rachel and Virginie, for being always available and fast in answering my emails and solving the bureaucratic issues which I encountered during my PhD.

These years would have not been so great without all the people who I met at EPFL, and I could always count on.

First, I would like to thank all the members of ANMC and CSQI for the nice moments during the group lunches and dinners. It was a pleasure to be in the group together with you. A special thank here goes to Giacomo for being not only the perfect office mate, but also friend, mentor, and coauthor of papers which appear in this thesis. You taught me everything that a PhD student should know.

I now would like to thank all the other members of MATHICSE with whom I shared not only the corridor but also nice activities, from the cookie breaks to the retreats. Some of them deserve a special thank for all the laughs and the wonderful moments we spent together which made these last few years unforgettable. A huge thank to Ale for teaching me orienteering, Bernard for the basketball matches, Eli for the nice working sessions during the confinement, Gianluca for the networking lessons, Niccoletto the best SIAM secretary ever, Nicolino for making me discover "Le Grammont", Ondine the first real and important friend I met in Lausanne, Paride for the chats as my office neighbor, Riccardo because without a P there would not be a VP, and Ali for making me happy every day.

Outside EPFL, when I was at home I was with great friends. Thanks a lot to the members of "Casa SPZ" Leo and Ste for all dinners, hikes, holidays and all the time together. These years would have not been the same without you.

I would also like to thank the friends who, unfortunately, I cannot meet daily. Many thanks to Diego, Fede, Giorgia and Niko for all the beautiful trips together, to Luca for the PACS dinners, to Teo with whom I am lucky to work together again soon and to all my historical friends who I am always happy to meet when I return to Mantova.

Infine, un ringraziamento speciale a tutta la mia famiglia e in particolare a Mamma, Papi e Francy. Anche se sono spesso lontano, vi sento vicino e so che posso sempre contare su di voi. Grazie infinite per la vostra presenza costante e il vostro sostegno, siete e sarete sempre molto importanti per me.

This work is partially supported by the Swiss National Science Foundation, under grant No. 200020_172710.

Abstract

We study the problem of learning unknown parameters of stochastic dynamical models from data. Often, these models are high dimensional and contain several scales and complex structures. One is then interested in obtaining a reduced, coarse-grained description of the dynamics that is valid at macroscopic scales. In this thesis, we consider two stochastic models: multiscale Langevin diffusions and noisy interacting particle systems. In both cases, a simplified description of the model is available through the theory of homogenization and the mean field limit, respectively. Inferring parameters in coarse-grained models using data from the full dynamics is a challenging problem since data are compatible with the surrogate model only at the macroscopic scale.

In the first part of the thesis we consider the framework of overdamped two-scale Langevin equation and aim to fit effective dynamics from continuous observations of the multiscale model. In this setting, estimating parameters of the homogenized equation requires preprocessing of the data, often in the form of subsampling, because traditional maximum likelihood estimators fail. Indeed, they are asymptotically biased in the limit of infinite data and when the multiscale parameter vanishes. We avoid subsampling and work instead with filtered data, found by application of an appropriate kernel of the exponential family and a moving average. We then derive modified maximum likelihood estimators based on the filtered process, and show that they are asymptotically unbiased with respect to the homogenized equation. A series of numerical experiments demonstrate that our new approach allows to successfully infer effective diffusions, and that it is an improvement of traditional methods such as subsampling. In particular, our methodology is more robust, requires less knowledge of the full model, and is easy to implement. We conclude the first part presenting novel theoretical results about multiscale Langevin dynamics and proposing possible developments of the filtering approach.

In the second part of the thesis we consider both multiscale diffusions and interacting particle systems, and we employ a different technique which is suitable for parameter estimation when a sequence of discrete observations is given. In particular, our estimators are defined as the zeros of appropriate martingale estimating functions constructed with the eigenvalues and the eigenfunctions of the generator of the effective dynamics. We first prove homogenization results for the generator of the multiscale Langevin equation and then apply our novel eigenfunction estimators to the two problems under investigation. Moreover, in the case of multiscale diffusions, we combine this strategy with the filtering methodology previously introduced. We prove that our estimators are asymptotically unbiased and present a series of numerical experiments which corroborate our theoretical findings, illustrates the advantages of our approach, and shows that our methodology can be employed to accurately fit simple models from complex phenomena.

Key words: eigenfunction martingale estimator, filtering, Fokker–Planck equation, homogenization, interacting particle systems, Langevin dynamics, maximum likelihood estimator, mean field limit, multiscale diffusion process, statistical inference.

Résumé

Nous étudions le problème d'estimation de paramètres inconnus de modèles dynamiques stochastiques à partir de données. Souvent, ces modèles sont de dimension éleveée, ils contiennent plusieurs échelles et des structures complexes. Nous sommes donc interessés à une description réduite de la dynamique qui soit valable aux échelles macroscopiques. Dans cette thèse, nous considérons deux modèles stochastiques : les processus de diffusion de Langevin multi-échelles et les systèmes de particules en interaction en présence de bruit. Dans les deux cas, une description simplifiée du modèle est disponible grâce à la théorie de l'homogénéisation et de la limite du champ moléculaire, respectivement. Inférer des paramètres dans des modèles réduits en utilisant les données de la dynamique complète est un problème difficile puisque les données ne sont compatibles avec le modèle de substitution qu'à l'échelle macroscopique.

Dans la première partie de la thèse, nous considérons l'équation de Langevin suramortie à deux échelles et nous avons pour but d'ajuster la dynamique effective à partir d'observations continues du modèle multi-échelle. Dans ce cadre, l'estimation des paramètres de l'équation homogénéisée nécessite un prétraitement des données, souvent sous forme de sous-échantillonnage, car les estimateurs traditionnels du maximum de vraisemblance échouent. En effet, ils sont asymptotiquement biaisés dans la limite des données infinies et lorsque le paramètre multi-échelle tend vers zéro. Nous évitons le sous-échantillonnage et nous travaillons avec des données filtrées, obtenues en appliquant un filtre exponentiel et une moyenne mobile. Nous obtenons ensuite des estimateurs de maximum de vraisemblance modifiés basés sur le processus filtré, et nous montrons qu'ils sont asymptotiquement non biaisés par rapport à l'équation homogénéisée. De nombreuses expériences numériques démontrent que notre nouvelle approche permet d'inférer avec succès des processus de diffusion effectifs, et qu'elle constitue une amélioration des méthodes traditionnelles telles que le sous-échantillonnage. En particulier, notre méthodologie est plus robuste, elle nécessite moins de connaissances du modèle complet, et elle est facile à mettre en œuvre. Nous concluons la première partie en présentant de nouveaux résultats théoriques sur la dynamique de Langevin multi-échelle et en proposant des développements possibles de l'approche de filtrage.

Dans la deuxième partie de la thèse, nous considérons à la fois des processus de diffusion multi-échelles et les systèmes de particules en interaction, et nous employons une technique différente qui convient à l'estimation des paramètres lorsque des observations discrètes sont donnée. En particulier, nos estimateurs sont définis comme les zéros de fonctions d'estimation martingales appropriées, construites avec les valeurs propres et les fonctions propres du générateur de la dynamique effective. Nous prouvons d'abord des résultats d'homogénéisation pour le générateur de l'équation de Langevin multi-échelle, puis nous appliquons nos nouveaux estimateurs aux deux problèmes étudiés. De plus, dans le cas des processus de diffusion multi-échelles, nous combinons cette stratégie avec la méthodologie de filtrage introduite précédemment. Nous prouvons que nos estimateurs sont asymptotiquement non biaisés et nous présentons expériences numériques qui valident nos résultats théoriques, qui illustrent les avantages de notre approche et qui montrent que notre méthodologie peut être employée pour inférer des modèles simples de phénomènes complexes.

Mots clés : estimateur martingale avec fonctions propres, filtrage, équation de Fokker–Planck, homogénéisation, systèmes d'interaction de particules, dynamique de Langevin, estimateur du maximum de vraisemblance, théorie du champ moléculaire, processus de diffusion multi-échelle, inférence statistique.

Sommario

Studiamo il problema di stimare parametri incogniti di modelli dinamici stocastici dai dati. Spesso questi modelli sono di dimensione elevata e contengono numerose scale e strutture complesse. Siamo quindi interessati ad ottenere una descrizione ridotta della dinamica che è valida alle scale macroscopiche. In questa tesi consideriamo due modelli stocastici: processi di diffusione di Langevin multiscala e sistemi di interazione di particelle con rumore. In entrambi i casi una descrizione semplificata del modello esiste ed è ottenuta con le teorie dell'omogeneizzazione e del campo medio, rispettivamente. Stimare parametri in modelli ridotti usando i dati della dinamica completa è un problema complesso poiché i dati sono compatibili con il modello surrogato solo alla scala macroscopica.

Nella prima parte della tesi consideriamo l'equazione di Langevin sovrasmorzata con due scale e cerchiamo di adattare una dinamica effettiva alle osservazioni continue del modello multiscala. In questo contesto, stimare i parametri dell'equazione omogeneizzata richiede di preprocessare i dati, spesso attraverso un sottocampionamento, perché i tradizionali stimatori di massima verosimiglianza falliscono. Infatti, essi sono asintoticamente distorti quando il tempo finale di osservazione tende all'infinito e il parametro multiscala tende a zero. Evitiamo il sottocampionamento e lavoriamo invece con dati filtrati, ottenuti applicando un filtro esponenziale e una media mobile. Deriviamo poi nuovi stimatori modificando gli stimatori di massima verosimiglianza e utilizzando il processo filtrato, e dimostriamo che sono asintoticamente non distorti rispetto all'equazione omogeneizzata. Una serie di esperimenti numerici mostra che il nostro approccio permette di inferire correttamente processi di diffusione effettivi ed è un miglioramento dei metodi tradizionali come il sottocampionamento. In particolare, la nostra metodologia è più robusta, richiede una conoscenza minore del modello completo, ed è facile da implementare. Concludiamo la prima parte presentando nuovi risultati teorici sulla dinamica di Langevin multiscala e proponendo possibili sviluppi dell'approccio di filtraggio.

Nella seconda parte della tesi consideriamo sia processi di diffusione multiscala che sistemi di interazione di particelle, e applichiamo una tecnica differente adatta a stimare parametri data una sequenza di osservazioni discrete. In particolare, i nostri stimatori sono definiti come gli zeri di funzioni martingala di stima costruite con gli autovalori e le autofunzioni del generatore della dinamica effettiva. Per prima cosa dimostriamo risultati di omogeneizzazione per il generatore dell'equazione di Langevin multiscala e poi applichiamo i nostri nuovi stimatori ai due problemi in esame. Inoltre, nel caso di processi di diffusione multiscala, combiniamo questa strategia con la metodologia di filtraggio precedentemente introdotta. Dimostriamo che i nostri stimatori sono asintoticamente non distorti e presentiamo numerosi esperimenti numerici che corroborano i nostri risultati teorici, illustrano i vantaggi del nostro approccio, e mostrano che la nostra metodologia può essere impiegata per inferire accuratamente modelli semplici da fenomeni complessi.

Parole chiave: stimatore martingala con autofunzioni, filtraggio, equazione di Fokker–Planck, omogeneizzazione, sistemi di interazione di particelle, dinamica di Langevin, stimatore di massima verosimiglianza, teoria del campo medio, processi di diffusione multiscala, inferenza statistica.

Notation

Sets of numbers

$\mathbb{N}$	set of positive integers
$\mathbb{Z}$	set of integers
$\mathbb{R}$	set of real numbers

Differentials

∇ *or* grad	gradient operator
$\nabla\cdot$ *or* div	divergence operator
Δ	Laplace operator
∇^2	Hessian operator

Functions

Let D be an open domain of $\mathbb{R}^d$, d a positive integer, and consider functions $f\colon D \to R$.

$C^k(D)$	space of k-times continuously differentiable functions
$L^p(D)$	usual Lebesgue space with $p \in [1, \infty]$
χ_D	characteristic function of the set D
$\dot{f}$	partial derivative with respect to time or to a parameter
$\partial_x f$	partial derivative with respect to the variable x

Vectors and matrices

Let $a, b \in R^d$ and $A, B \in \mathbb{R}^{d\times d}$ with d a positive integer.

$\|a\|$	Euclidean norm of a vector
$\|A\|$	2-norm of a matrix
$\langle a; b\rangle$	inner product, i.e., $a^\top b$
$a \otimes b$	outer product, i.e., $ab^\top$
$A : B$	Frobenius inner product, i.e., $\mathrm{tr}(A^\top B)$
$\mathcal{S}(A)$	symmetric part, i.e., $(A + A^\top)/2$

Acronyms

BKE	Backward Kolmogorov Equation
EM	Euler–Maruyama
FPE	Fokker–Planck equation
MLE	Maximum likelihood estimator
OU	Ornstein–Uhlenbeck
PDE	Partial differential equation
SDE	Stochastic Differential Equation

Contents

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 or atmosphere and ocean science, makes it possible to develop physics-informed data driven methodologies for inferring models from data [109, 122, 126]. 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 its macroscopic and microscopic components, then one is usually interested in the derivation of a model for the macroscopic components, while treating the microscopic ones as noise. This often leads to reduced models which are stochastic, thus naturally described by stochastic differential equations (SDEs). The main goal of this thesis is to derive rigorous and systematic methodologies for learning coarse-grained models that accurately describe the dynamics at macroscopic length and time scales from observations of the full microscopic dynamics.

The model which we will mainly consider throughout this thesis is the multiscale overdamped Langevin SDE. This simple equation arises from models of molecular dynamics, and is featured by two fully separated time scales. In particular, it describes the motion of a particle in a confining potential which has slow variations with rapid order-one oscillations superimposed. Hence, it is frequently desirable to infer from data a simpler model which captures effectively large-scale structures, or slow variations, disregarding small-scale fluctuations. In this framework, under appropriate assumptions on the potential, a single-scale surrogate equation in which the fast-scale potential is eliminated is guaranteed to exist due to the theory of homogenization [20, 104]. Given multiscale data in the form of a continuous sample path from this class of model problems of Langevin type, we are therefore interested in determining the drift and diffusion coefficients of the corresponding homogenized equation, i.e., our goal is to obtain effective coarse-grained dynamics from data, in a consistent way with respect to the homogenization theory. The mismatch between the data and their desired slow-scale representation is a typical instance of a problem of model misspecification, which, if ignored or handled incorrectly, can lead to erroneous inference. Indeed, the data, coming from the full dynamics, are compatible with the coarse-grained model only at the time scales at which the effective dynamics is valid. Several methods to take into account model misspecification in multiscale frameworks exist [12, 13, 97, 103]. For diffusion processes, the proposed approaches rely on different sorts of subsampling, which has proved itself effective to some extent in many applications, but which requires nevertheless precise knowledge of how separated the two characteristic scales are. Robustness of this methodology is dubious, too, as inference results tend to be extremely sensitive to the subsampling rate.

We therefore introduce a novel methodology for efficiently estimating the drift and diffusion coefficients of the effective equation given a continuous stream of data from the multiscale model. We propose to modify the maximum likelihood estimator (MLE), which has been proven to fail [103], by using suitably preprocessed data, instead of the raw ones. Rather than subsampling the original trajectory, we smooth the data by applying an appropriate linear time-invariant filter from the exponential family [3]. The estimator we obtain is robust with respect to the parameters of the filter and asymptotically unbiased in the limit of infinite data and of infinitely fast oscillations at the microscale, as long as the filtering width is sufficiently large with respect to the fastest scale. Moreover, we observe numerically that iterating the filtering procedure, i.e., employing smoother data, allows us to obtain reliable estimations of the unknown parameters independently of the filtering width. We also notice from numerical experiments that our estimator appears to be asymptotically normal and this gives a conjecture on its asymptotic variance.

Moreover, instead of employing an exponential filter, we also propose to preprocess the data applying a moving average [51]. This leads to a stable and robust estimator which is even simpler to compute, and for which we prove the same asymptotic unbiasedness property. Difficulties are hidden in the theoretical analysis of the proposed estimator, indeed the main proof is based on the observation that original and filtered data together form a system of stochastic delay differential equations. Furthermore, if the effective diffusion coefficient is assumed to be known, we present how our filtered data methodology can be combined with Bayesian techniques in order to provide a full uncertainty quantification of the inference procedure. Still in the case of known diffusion coefficient, we also show that, instead of employing filtered data, it is possible to use the Stratonovich formulation of the MLE.

The assumption that a continuous path from the solution is observed is however not realistic in most applications. In fact, in all real problems one can only obtain discrete-time measurements of the diffusion process. Hence, we focus on the problem of learning the coarse-grained homogenized model assuming that we are given discrete observations from the microscopic model. We propose a new estimator for learning homogenised SDEs from noisy discrete data that is obtained as the zero of a martingale estimating function which is based on the eigenvalues and the eigenfunctions of the generator of the homogenized process [6]. This technique was origially applied to one-scale SDEs in [73]. We show that this new estimator is asymptotically unbiased only if the distance between two consecutive observations is sufficiently large when compared with the multiscale parameter describing the fastest scale, i.e., if data are 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 introduced in the first part of this thesis. We notice that smoothing the original data makes observations compatible with the homogenized process independently of the rate with which they are sampled. Hence, this second estimator can be employed as a black-box tool for parameter estimation in the case of discrete-time data.

We then employ a similar approach based on martingale estimating functions for inferring unknown parameters in systems of weakly interacting diffusions. For these models the mean field limit exists and is described by a nonlinear diffusion process of McKean type, obtained as the limit when the number of interacting processes goes to infinity. When the number of interacting SDEs is large, the inference problem can become computationally intractable and it is often useful to study the problem of parameter estimation for the limiting mean field SDE. This is related, but distinct, from the previous problem of inference for multiscale diffusions where the objective is to learn the parameters in the homogenized limiting SDE from observations of the full dynamics. Our goal is to show how the inference methodology using martingale estimating functions which is applied to multiscale diffusions can be modified so that it can also be applied to interacting diffusions with a well defined mean field limit. Despite their difference in spirit, it is useful to keep in mind the analogy between homogenization and mean field limits, in the context of parameter estimation.

We consider in particular systems of exchangeable weakly interacting diffusions for which uniform propagation of chaos results are known [18, 19, 39, 84, 93] and for which the mean field SDE has a unique invariant measure. We assume that we are given a sample of discrete-time observations of a single particle. Due to exchangeability, this amount of information should be sufficient to infer parameters in the mean field SDE, in the joint asymptotic limit as the number of observations and the number of particles go to infinity. Our approach consists of constructing martingale estimating functions [21, 73] based on the eigenvalues and the eigenfunctions of the generator of the mean field dynamics. Then, our eigenfunction estimator is the zero of the estimating function [105]. Unlike the finite dimensional case, the mean field SDE is a measure-valued process and the generator is a nonlinear operator, dependent on the law of the process. A direct application of the martingale eigenfunction estimator would require the solution of a nonlinear eigenvalue problem that can be computationally demanding in high dimensions and that would also lead to eigenfunctions depending on time via their dependence on the law of the process. We circumvent this difficulty by replacing the law of the process with the (unique) invariant measure of the mean field dynamics. This leads to a standard Sturm–Liouville type of eigenvalue problem that we can analyze and also solve numerically at a low computational cost. In this thesis we consider the framework where the invariant measure of the mean field SDE is unique. We remark, however, that our numerical experiments show that our methodology applies to McKean SDEs that exhibit phase transitions, i.e., that have multiple stationary measures, as long as we are below the transition point, or the form of the invariant measure is known up to a finite set of parameters, e.g., moments. When the mean field dynamics has a unique invariant measure, we first show the existence of the estimator with high probability when the number of available data and particles is large enough, and then analyze its consistency proving the asymptotic convergence towards the true value of the unknown parameter and providing a rate. Moreover, we prove that the estimator is asymptotically normal under appropriate assumptions on the relationship between the number of observations and particles, in particular the latter must be sufficiently greater than the former.

Finally, on a parallel path, but still related to the first model considered in this thesis, i.e., the Langevin dynamics, we show some additional theoretical results. First, we compute the rate of weak convergence of the invariant measure of the multiscale system to the corresponding homogenized invariant measure, and we present a new proof for the homogenization of the backward Kolmogorov equation (BKE) based on the theory of evolutionary Gamma convergence [87]. Then, we study the homogenization of the Poisson problem and the eigenvalue problem for the generator of the multiscale dynamics, proving convergence results for the eigenpairs and the solution of the Poisson problem. We remark that the analysis is based on the theory of two-scale convergence [9, 10] which we extend to the case of weighted Sobolev spaces.

Outline

The thesis is divided in two parts. Part I is made of Chapters 1 to 4 and mainly concerns the introduction of filtered data to modify MLEs, and Part II is made of Chapters 5 to 8 and is mostly about eigenfunction estimators. We remark that Chapters 2 to 4 and 6 to 8 contain the original contributions of this thesis.

In Chapter 1 we introduce the inference problem for multiscale Langevin dynamics, we give a brief overview of the existing literature on the topic and we present the main contributions of Part I. Chapters 2 and 3, which are based on our papers [3, 51], are about the filtered data methodology based on exponential filter and moving average, respectively, and Chapter 4 is devoted to some additional results and open problems related to the multiscale Langevin model.

In Chapter 5 we introduce interacting particle systems and the technique of martingale estimating functions based on eigenpairs of the generator for parameter estimation. We give a brief overview of the existing literature on the topic and we present the main contributions of Part II. Chapter 6, which is based on our paper [125], is devoted to the study of the homogenization of the generator of the multiscale Langevin dynamics and Chapters 7 and 8, which are based on our papers [6, 105], are about the application of eigenfunction estimators to multiscale and interacting diffusions, respectively.

Part I

Filtered data and modified maximum likelihood estimators: inference from continuous observations

1 Inference for multiscale overdamped Langevin dynamics

In this chapter we consider the multiscale overdamped Langevin equation and its coarse-grained model, which exists due to the theory of homogenization. We then introduce the inference problem we are interested in, and provide an overview of the techniques that have already been employed in the literature. Finally, we summarize the main contributions which will be presented in the remaining chapters of this part, and give the outline of the first part of this thesis.

1.1 Problem setting

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 $b^\varepsilon \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ periodic with respect to its second argument, a multiscale parameter $\varepsilon > 0$, and a diffusion coefficient $\sigma > 0$, we consider the d-dimensional multiscale stochastic differential equation (SDE)

$$\mathrm{d}X^\varepsilon(t) = b^\varepsilon\left(X^\varepsilon(t), \frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t), \tag{1.1}$$

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 to infer a coarse-grained equation, independent of the fastest scale $\mathcal{O}(\varepsilon^{-1})$, which reads

$$\mathrm{d}X^0(t) = b^0\left(X^0(t)\right)\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t), \tag{1.2}$$

where $b^0 \colon \mathbb{R}^d \to \mathbb{R}^d$ 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, one can prove in these cases that $X^\varepsilon \to X^0$ for $\varepsilon \to 0$ in a weak sense (see [104, Chapter 18] or [20, Chapter 3]). In this thesis, 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.

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 thesis, we narrow the scope by considering a gradient structure and a semi-parametric framework, inspired by simple models of molecular dynamics. Let L be a positive integer, and consider a periodic function $p \colon \mathbb{R}^d \to \mathbb{R}$ with period $\mathbb{T}_i$ in the i-th direction in $\mathbb{R}^d$ for $i = 1, \dots, d$, and a potential $\mathcal{V} \colon \mathbb{R}^d \to \mathbb{R}$ defined by

$$\mathcal{V}(x) = \sum_{\ell=1}^{L} \alpha_\ell V_\ell(x),$$

where $\{V_\ell \colon \mathbb{R}^d \to \mathbb{R}\}_{\ell=1}^L$ are smooth functions and $\{\alpha_\ell\}_{i=1}^L$ are scalar drift coefficients. We then let the drift function in the multiscale dynamics (1.1) be given by

$$b^\varepsilon(x,y) = -\nabla\mathcal{V}(x) - \frac{1}{\varepsilon}\nabla p(y) = -\sum_{\ell=1}^{L}\alpha_\ell\nabla V_\ell(x) - \frac{1}{\varepsilon}\nabla p(y),$$

and, with this choice, equation (1.1) reads in the nonparametric form

$$\mathrm{d}X^\varepsilon(t) = -\nabla\mathcal{V}(X^\varepsilon(t))\,\mathrm{d}t - \frac{1}{\varepsilon}\nabla p\left(\frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t), \tag{1.3}$$

and in the semi-parametric form

$$\mathrm{d}X^\varepsilon(t) = -\sum_{\ell=1}^{L}\alpha_\ell\nabla V_\ell(X^\varepsilon(t))\,\mathrm{d}t - \frac{1}{\varepsilon}\nabla p\left(\frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t). \tag{1.4}$$

We remark that the stochastic model we consider is of the overdamped Langevin type. There exists for equation (1.4) a model of the form (1.2) which is effective in the homogenization limit $\varepsilon \to 0$. Let $X^\varepsilon \coloneqq (X^\varepsilon(t), 0 \le t \le T)$ denote the solution of (1.4) for a finite time horizon T. Then, it holds $X^\varepsilon \to X^0$ in law in $C^0([0,T];\mathbb{R}^d)$ for $\varepsilon \to 0$, where $X^0 \coloneqq (X^0(t), 0 \le t \le T)$ is the solution of the overdamped Langevin equation

$$\mathrm{d}X^0(t) = -\sum_{\ell=1}^{L}A_\ell\nabla V_\ell(X^0(t))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t). \tag{1.5}$$

Here, the matrices $A_\ell \coloneqq \alpha_\ell\mathcal{K}$ and $\Sigma \coloneqq \sigma\mathcal{K}$ depend on the symmetric positive semidefinite matrix $\mathcal{K} \in \mathbb{R}^{d\times d}$ defined by

$$\mathcal{K} = \int_{\mathcal{T}}(I + D\Phi(y))(I + D\Phi(y))^\top\,\mathrm{d}\pi(y), \quad \mathcal{T} \coloneqq \bigotimes_{i=1}^{d}[0,\mathbb{T}_i], \tag{1.6}$$

where $D\Phi$ is the Jacobian of the solution $\Phi\colon \mathbb{R}^d \to \mathbb{R}^d$ of the vector-valued partial differential equation (PDE), or cell problem

$$\begin{aligned} &\mathcal{L}_0\Phi = \nabla p, && \text{in } \mathcal{T}, \quad + \text{ periodic b.c. on } \partial\mathcal{T}, \\ &\int_{\mathcal{T}}\Phi(y)\,\mathrm{d}\pi(y) = 0, \end{aligned} \tag{1.7}$$

and 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 π introduced in (1.6) is the probability measure on $\mathcal{T}$ given by $\pi(\mathrm{d}y) = \omega(y)\,\mathrm{d}y$ where the density ω with respect to the Lebesgue measure is defined as

$$\omega(y) = \frac{1}{C_\pi}\exp\left(-\frac{p(y)}{\sigma}\right), \qquad C_\pi = \int_{\mathcal{T}}\exp\left(-\frac{p(y)}{\sigma}\right)\mathrm{d}y, \tag{1.8}$$

which accounts for the fluctuations in the limit as $t \to \infty$ of the fast-scales of the solution of (1.4). We refer the reader to [104, Chapters 11 and 18] for a complete derivation and proof of this homogenization result. Clearly, the homogenized equation corresponding to the non parametric form (1.3) is given by

$$\mathrm{d}X^0(t) = -\mathcal{K}\nabla\mathcal{V}(X^0(t))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t). \tag{1.9}$$

We consider from now on for clarity the case $d = 1$, for which equation (1.4) reads

$$\mathrm{d}X^\varepsilon(t) = -\alpha \cdot V'(X^\varepsilon(t))\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t), \tag{1.10}$$

where $V\colon \mathbb{R} \to \mathbb{R}^L$ is defined as $V(x) = (V_1(x), V_2(x), \dots, V_L(x))^\top$, the derivative V' is computed component-wise, and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_L)^\top$. Let us assume that we are exposed to a continuous-time stream of data X^ε solution of (1.4) 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 $\{\alpha_\ell\}_{\ell=1}^L$, and the diffusion coefficient σ are unknown. Conversely, we assume that the functions $\{V_\ell\}_{\ell=1}^L$ are known. Our goal is then to infer the effective drift coefficient $A = (A_1, A_2, \dots, A_L)^\top \in \mathbb{R}^L$ and diffusion coefficient $\Sigma > 0$ that define the single-scale dynamics

$$\mathrm{d}X^0(t) = -A \cdot V'(X^0(t))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t). \tag{1.11}$$

Let us remark that in this one-dimensional setting it is possible to determine Φ explicitly, and the homogenization coefficient $\mathcal{K}$ is given by

$$\mathcal{K} = \frac{\mathbb{T}^2}{C_\pi \widehat{C}_\pi}, \tag{1.12}$$

where

$$C_\pi = \int_0^{\mathbb{T}} e^{-p(y)/\sigma}\,\mathrm{d}y, \quad \widehat{C}_\pi = \int_0^{\mathbb{T}} e^{p(y)/\sigma}\,\mathrm{d}y.$$

We consider the inferred coefficients to be *asymptotically unbiased* if they converge to the true effective coefficients in the homogenization limit $\varepsilon \to 0$ and for infinite data, i.e., for $T \to \infty$. 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.*

Remark 1.1. For enhancing the clarity of the exposition, throughout most of the thesis, we have chosen to focus on the case of a multi-dimensional parameter in the setting of one-dimensional diffusion processes. In fact, all the theory we present in the following could be generalized to the case of the d-dimensional SDE (1.4). Slight modifications of the proofs demonstrate that analogous results as those presented later on may be obtained in the d-dimensional case.

Remark 1.2. The value of the initial condition X_0^ε in the SDE (1.10) is important neither for the numerical experiments nor for the following analysis and can be chosen arbitrarily. In fact, as shown in [103, Proposition 5.2], the process X_t^ε is geometrically ergodic and therefore it converges to its invariant distribution exponentially fast for any initial condition.

Remark 1.3. We note that our framework may be viewed in the semi-parametric setting as the one of [74]. In particular the functions V_ℓ, $\ell = 1, \dots, L$ can be seen as the known basis functions of a truncated expansion (e.g., Taylor series or Fourier series) for the unknown confining potential $\mathcal{V}$. Numerical examples highlighting the potential of our method in such a setting is given in Sections 2.3.3 and 3.2.2. We also mention that assuming a parametric form for the potential $\mathcal{V}$ is a technique usually employed in the statistics literature in order to regularize the likelihood function and obtain a parametric approximation of the actual MLE of $\mathcal{V}$, which does not exist in general [107].

Let us now state the assumptions which will be employed throughout the rest of our work. In particular, we consider the same dissipative setting as [103, Assumption 3.1].

Assumption 1.4. The potentials p and V satisfy

(i) $p \in \mathcal{C}^\infty(\mathbb{R})$ and is $\mathbb{T}$-periodic for some $\mathbb{T} > 0$;

(ii) $V_\ell \in \mathcal{C}^\infty(\mathbb{R})$ for all $\ell = 1, \dots, L$ is polynomially bounded from above and bounded from below, and there exist $a, b > 0$ such that

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

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

$$\|V'(x) - V'(y)\|_2 \leq C\,|x - y|\,, \qquad \text{and} \qquad \|V''(x) - V''(y)\|_2 \leq C\,|x - y|$$

where $\|\cdot\|$ denotes the Euclidean norm, and the components $V_\ell', V_\ell'', V_\ell'''$ are polynomially bounded for all $\ell = 1, \dots, L$.

Remark 1.5. Assumption 1.4 has the following consequences:

(i) Assumption 1.4(i) allows to employ the theory of periodic homogenization to conclude that (1.11) is effective for equation (1.10).

(ii) Assumption 1.4(ii) implies that the solutions of (1.10) and (1.11) are geometrically ergodic, and thus the existence of unique invariant measures (see [103, Propositions 5.1 and 5.2]).

(iii) Assumption 1.4(iii) is technical, and guarantees sufficient regularity for our main results to hold.

Remark 1.6. In the following, in particular in the proof of Lemma 2.3, we will employ Assumption 1.4(ii) for the whole drift of the SDE (1.1), i.e., the function

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

Since $p \in C^\infty(\mathbb{R})$ and is periodic, all derivatives of p are in $L^\infty(\mathbb{R})$. Therefore, the assumption above is sufficient for $\mathcal{V}^\varepsilon$ to satisfy Assumption 1.4(ii) with different values for a and b. In particular, assume Assumption 1.4(ii) holds for $\mathcal{V}$. Then, we have for all $\gamma > 0$ by Young's inequality

$$\begin{aligned} -(\mathcal{V}^\varepsilon)'(x)x &\leq a - bx^2 - \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)x \\ &\leq \left(a + \frac{1}{2\varepsilon^2\gamma}\|p'\|^2_{L^\infty(\mathbb{R})}\right) - \left(b - \frac{\gamma}{2}\right)x^2. \end{aligned}$$

Hence, Assumption 1.4(ii) holds for $\mathcal{V}^\varepsilon$ with a coefficient b which is arbitrarily close to the coefficient for $\mathcal{V}$, alone. We also remark that the fact that the shift term blows up as $\varepsilon \to 0$ is not an issue because the ergodicity of the multiscale process has to hold for $\varepsilon > 0$ fixed.

1.2 Failure of standard estimators

We now briefly present the classical methodology for estimating the drift coefficient alone. Let $T > 0$ and let $X^0 \coloneqq (X^0(t), 0 \leq t \leq T)$ be a realization of the solution of (1.11) up to final time T. Girsanov's change of measure formula applied to (1.11) allows to write the likelihood of X^0 given a drift coefficient A as

$$\mathfrak{L}_T(X^0 \mid A) = \exp\left(-\frac{I_T(X^0 \mid A)}{2\Sigma}\right), \tag{1.13}$$

where

$$I_T(X^0 \mid A) = \int_0^T A \cdot V'(X^0(t))\,\mathrm{d}X^0(t) + \frac{1}{2}\int_0^T \left(A \cdot V'(X^0(t))\right)^2\,\mathrm{d}t.$$

Minimizing the functional $I_T(X^0 \mid A)$ with respect to A therefore gives the maximum likelihood estimator (MLE) of A, which is independent of the diffusion coefficient Σ and can be formally computed in closed form as

$$\widehat{A}_{\mathrm{MLE}}(X^0,T) \coloneqq \arg\min_{A\in\mathbb{R}^L} I_T(X^0 \mid A) = -M^{-1}(X^0,T)v(X^0,T), \tag{1.14}$$

where $M(X^0,T) \in \mathbb{R}^{L\times L}$ and $v(X^0,T) \in \mathbb{R}^L$ are defined as

$$M(X^0,T) = \frac{1}{T}\int_0^T V'(X^0(t)) \otimes V'(X^0(t))\,\mathrm{d}t, \quad v(X^0,T) = \frac{1}{T}\int_0^T V'(X^0(t))\,\mathrm{d}X^0(t),$$

and where $\otimes$ denotes the outer product in $\mathbb{R}^L$. Under Assumption 1.4, the MLE given in (1.14) is indeed the unique minimizer of the likelihood function, as shown in [107, Theorem 2.4]. Let us consider the modified estimator of the drift coefficient obtained replacing X^0 with $X^\varepsilon \coloneqq (X^\varepsilon(t), 0 \le t \le T)$ solution of (1.10), i.e.,

$$\widehat{A}_{\mathrm{MLE}}(X^\varepsilon,T) \coloneqq \arg\min_{A\in\mathbb{R}^L} I_T(X^\varepsilon \mid A) = -M^{-1}(X^\varepsilon,T)v(X^\varepsilon,T), \tag{1.15}$$

where $I_T(X^\varepsilon \mid A)$, the matrix $M(X^\varepsilon,T)$ and the vector $v(X^\varepsilon,T)$ are obtained replacing each occurrence of X^0 with X^ε. We now introduce the following additional hypothesis.

Assumption 1.7. For all $T > 0$, the symmetric matrices $M(X^0,T)$ and $M(X^\varepsilon,T)$ are positive definite and there exists $\bar\lambda > 0$ such that $\lambda_{\min}(M(X^0,T)) \ge \bar\lambda$ and $\lambda_{\min}(M(X^\varepsilon,T)) \ge \bar\lambda$.

In the following, we simply denote by $M \coloneqq M(X^\varepsilon,T)$ and $v \coloneqq v(X^\varepsilon,T)$ in case of no ambiguity. Given the convergence of $X^\varepsilon \to X^0$ in the space of continuous stochastic processes, one would expect that the MLE (1.15) would be asymptotically unbiased for the drift coefficient A of the homogenized equation (1.11). Instead, it is possible to prove that in the asymptotic limit for $T \to \infty$ and $\varepsilon \to 0$, the MLE tends to the drift coefficient α of the unhomogenized equation (1.10). We report here this result, whose proof can be found for the case $L = 1$ in [103, Theorem 3.4]. We remark that the proof for $L > 1$ follows directly from the one-dimensional case.

Theorem 1.8. *Let Assumption 1.4 hold and let the process X^ε be the solution of* (1.10)*. Then*

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

where α is the drift coefficient of equation (1.10)*.*

Estimating the effective diffusion coefficient Σ of the homogenized SDE (1.11) is as well a relevant problem. Indeed, knowing Σ besides the drift coefficient A gives a complete estimation of the effective model (1.11), which is effective for the multiscale data generated by (1.10) in the sense of homogenization theory. The standard estimator of the diffusion coefficient Σ given the stream of data X^0 is obtained by computing the quadratic variation $\langle X^0\rangle_T$ of the path X^0 and by defining

$$\widehat{\Sigma}_{\mathrm{QV}}(X^0,T) = \frac{\langle X^0\rangle_T}{2T}.$$

We remark that in case the data X^0 would originate from the model (1.11), we would have by definition $\widehat{\Sigma}_{\mathrm{QV}}(X^0,T) = \Sigma$. In [103, Theorem 3.4], the authors show that this approach fails in case the data are not pre-processed, meaning that the estimator $\widehat{\Sigma}_{\mathrm{QV}}(X^\varepsilon,T)$ obtained using the quadratic variation of X^ε equals the diffusion coefficient σ of (1.10), even in the limit for $\varepsilon \to 0$, i.e.,

$$\widehat{\Sigma}_{\mathrm{QV}}(X^\varepsilon,T) = \sigma \neq \Sigma. \tag{1.16}$$

Therefore, we deduce that in the framework of data-driven homogenization both the standard estimators $\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)$ and $\widehat{\Sigma}_{\mathrm{QV}}(X^\varepsilon, T)$ for the effective drift and diffusion coefficients are not asymptotically unbiased and fail. Hence, it is necessary to employ homogenization-informed techniques to infer the effective equation.

Remark 1.9. In case ε is known and due to Theorem 1.8 and (1.16), it would be possible to infer directly the full multiscale model (1.11) 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 (1.10) is possible only by choosing critically small time step for most numerical schemes, which in turn yields dramatically high computational cost.

The main existing tool for obtaining unbiased estimators in the literature is subsampling the data. 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 discretization of the MLE with spacing δ, i.e.,

$$-M^\delta_{\mathrm{sub}}(X^\varepsilon, T)\widehat{A}^\delta_{\mathrm{sub}}(X^\varepsilon, T) = v^\delta_{\mathrm{sub}}(X^\varepsilon, T), \tag{1.17}$$

where

$$M^\delta_{\mathrm{sub}}(X^\varepsilon, T) \coloneqq \frac{\delta}{T}\sum_{i=0}^{n-1} V'(X^\varepsilon(i\delta)) \otimes V'(X^\varepsilon(i\delta)),$$
$$v^\delta_{\mathrm{sub}}(X^\varepsilon, T) \coloneqq \frac{1}{T}\sum_{i=0}^{n-1} V'(X^\varepsilon(i\delta))\left(X^\varepsilon((i+1)\delta) - X^\varepsilon(i\delta)\right).$$

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

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

The subsampling rate which guarantees asymptotically unbiased estimators lays between the time scales of the multiscale and the effective models, i.e., for $\delta = \varepsilon^\zeta$, with $\zeta \in (0,1)$ [103, Theorems 3.5 and 3.6]. The optimal subsampling rate is conjectured in [103] to be $\delta = \varepsilon^{2/3}$. Despite being widely employed in practice, estimators based on subsampling present some drawbacks, such as having a high variance. Other disadvantages of subsampling are mainly two. First, it has been demonstrated numerically [103] 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.

1.3 Literature review

The literature on statistical inference of stochastic models modelled by SDEs is vast. Introductory references on the topic are [17, 23, 77, 108]. 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 [102] for a recent survey, and summarize the approaches which are related to the one presented in this thesis in the following.

Multiscale diffusion processes can be employed for modeling chemical reactions with species reacting at different speeds [79, 80] and for simple models in molecular dynamics, for which the effect of model misspecification was studied in a series of papers [12–14, 97, 103] under the assumption of scale separation. In particular, for Brownian particles moving in two-scale potentials it was shown that, when fitting data from the full dynamics to the homogenized equation, the

MLE is asymptotically biased [103, Theorem 3.4]. To be more precise, in the large sample size limit, the data remains consistent with the multi-scale problem at small scale. Ostensibly this would seem related only to the estimation of the diffusion coefficient. However, because of detail balance, it also has the effect that the MLE, for the drift in a parameter fit of a single-scale model, incorrectly identifies the coefficient of the homogenized equation. The bias of the MLE can be eliminated by subsampling at an appropriate rate, which lies between the two characteristic time scales of the problem.

Similar techniques can be employed in econometrics, in particular for the estimation of the integrated stochastic volatility in the presence of market microstructure noise. In this case, too, the data have to be subsampled at an appropriate rate [8, 94]. The correct subsampling rate can, in some instances, be rather extreme with respect to the frequency of the data itself, resulting in ignoring as much as 99% of the time-series. As the intuition suggests, this increases significantly the variance of the estimator, which is usually taken care of with additional bias corrections and variance reduction procedures. The need of such methodology is accentuated by data being obtained at high-frequency [7, 127].

The problem of extracting large-scale variations from multiscale data is studied in atmosphere and ocean science. In this field, too, subsampling the data is necessary to obtain an accurate coarse-grained model [32, 123].

The necessity to subsample the data can be alleviated by exploiting a martingale property of the likelihood function, and by computing appropriate estimators for conditional expectations, as was done in [68, 74]. This class of estimators can be applied to the case where the noise is multiplicative and also given by a deterministic chaotic system, as opposed to white noise. Estimators of this family have been applied to time series from paleoclimatic data and marine biology and augmented with appropriate model selection methodologies [75]. The estimators proposed in [74] are not well posed on a single trajectory, which is overcome by averaging over a set of short trajectories. For [68], 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 [74]. In [68], theoretical analysis is restricted to the one-dimensional case and when the effective dynamics is of the Ornstein–Uhlenbeck (OU) type.

In their series of works [47, 48, 113], the authors propose estimators for the parameters of a multiscale SDE. The setting is similar to the one studied in this thesis, 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 are due in this framework to the likelihood function induced by the effective dynamics, which is singular. Closely related work is [96], where estimation of multiscale stochastic dynamics is obtained by dimensionality reduction of an appropriate posterior distribution.

Inference of diffusion processes can be naturally performed under a Bayesian perspective. If one focuses on the drift coefficient, the form of the likelihood function guarantees, under a Gaussian prior hypothesis, that the posterior distribution is itself a Gaussian. The versatility of the Bayesian approach in the infinite-dimensional case [37, 115] gives the possibility to extend the study of inferring the drift of a diffusion process to the non-parametric case [106, 107].

The recent work [31] 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.

The issue of model misspecification in inverse problems with a multiscale structure has been treated in the context of PDEs, too. In particular, it has been shown that it is possible to infer a coarse-grained equation from data coming from the full model and to retrieve, in the large data limit, the correct result [92]. A series of papers [1, 2, 4] focuses on retrieving the full model when the multiscale coefficient is endowed with a specific parametrized structure. Since these problems are ill-posed, the latter is achieved via Tikhonov regularization [1, 92], adopting a Bayesian approach [2, 92] or exploiting techniques of Kalman filtering [4]. In [2, 4], the authors highlight the need to account explicitly for the modelling error due to homogenization and apply statistical techniques taken from [26, 27].

1.4 Our main contributions

In the next two chapters, which are based on our research articles [3, 51], we bypass subsampling by designing a methodology based on filtered data.

In Chapter 2 we smooth the time-series data from the multiscale model by application of an appropriate linear time-invariant filter, from the exponential family, and show that doing so allows us to accurately retrieve the drift coefficient of the homogenized model. The methodology we present is straightforward to implement, robust in practice and backed by theory. In particular, we show theoretically and demonstrate via numerical experiments that:

(i) The smoothing width of the filter can be alternatively tuned to be proportional to the speed of the slow process or to smaller scales and provide in both cases unbiased results for maximum likelihood parameter estimation. Sharp estimates on the minimal width with respect to the multiscale parameter are provided. The unbiasedness results are given in Theorems 2.12 and 2.17 for filtered data in the homogenized and in the multiscale regimes, respectively.

(ii) We additionally propose in the multiscale regime an estimator of the effective diffusion coefficient based on filtered data, as shown by Theorem 2.19.

(iii) Estimations based on our technique are robust in practice with respect to the parameter of the filter. This is not the case for subsampling, which is strongly influenced by the subsampling frequency. The robustness of our technique is demonstrated via numerical experiments in Sections 2.3.1 and 2.3.3.

(iv) The entire stream of data is employed, which, in practice, enhances the quality of the filter-based MLE in terms of bias. Moreover, avoiding subsampling and thus discretising the data allows us to employ continuous-time theoretical tools.

(v) It is possible to employ the filtered data approach within a continuous-time Bayesian framework by a careful modification of the likelihood function. Under mild hypotheses on the filter parameters, we are able to show that the posterior distributions obtained with our methodology are asymptotically consistent with respect to the drift parameter of the homogenized equation. Our main theoretical result is given in Theorem 2.24, and a numerical experiment for the combination of the filtered data approach and of Bayesian techniques is presented in Section 2.3.4.

In Chapter 3 we build on the techniques introduced in Chapter 2 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 by testing our method against a complex instance of multi-dimensional Langevin equations. The basic idea underlying the new estimator is similar to the previous one, i.e., in both cases 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 Chapter 3 extend relevantly the work in Chapter 2 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 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.

Finally, in Chapter 4 we present additional theoretical results about the homogenization of multiscale Langevin dynamics and further developments related to the exponential filter introduced in Chapter 2. In particular:

(i) We prove the homogenization of the backward Kolmogorov equation (BKE) corresponding to the multiscale Langevin SDE employing the theory of the evolutionary Gamma convergence.

(ii) We compute a rate of convergence of the expectation with respect to the invariant measure of the multiscale Langevin dynamics towards the expectation with respect to the homogenized invariant measure.

(iii) We propose a different modification of the MLE for the drift estimation of the homogenized Langevin SDE which does not rely on filtered data, but which uses the Stratonovich formulation.

(iv) We propose a new estimator by repeatedly iterating the filtering procedure with the exponential kernel in Chapter 2.

(v) We give a tentative central limit theorem for our estimator with exponentially filtered data of Chapter 2.

We remark that we are still working on points (iv) and (v) and therefore they contain conjectures which we believe are true, but we have not been able to prove rigorously yet.

2 Exponential filter

This chapter, which is based on our research article [3], is devoted to the introduction of filtered data obtained employing an appropriate filter of the exponential family. They are then employed to modify the maximum likelihood estimator (MLE) and give an asymptotic unbiased estimator for the drift coefficient of the homogenized Langevin equation given continuous data from the multiscale system. The chapter is organized as follows. In Section 2.1 we present our filtered data methodology with a particular focus on ergodic properties, on multiscale convergence and, naturally, on the unbiasedness properties of our estimators. In Section 2.2 we introduce the Bayesian framework and show how it can be enhanced employing filtered data, and in Section 2.3 we demonstrate the effectiveness of our methodology via a series of numerical experiments. Section 2.4 collects several technical results which are useful for the proof of the main theorems, and in Section 2.5 we draw our conclusions.

2.1 The filtered data approach

In this section, we introduce and analyse a novel approach based on filtered data to address the issue that the MLE estimator, when confronted with multiscale data, is biased. Let $\beta, \delta > 0$ and let us consider a family of exponential kernel functions $k_{\exp}^{\delta,\beta} \colon \mathbb{R}^+ \to \mathbb{R}$ defined as

$$k_{\exp}^{\delta,\beta}(r) = C_\beta \delta^{-1/\beta} e^{-r^\beta/\delta}, \tag{2.1}$$

where C_β is the normalizing constant given by

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

so that

$$\int_0^\infty k_{\exp}^{\delta,\beta}(r)\,\mathrm{d}r = 1,$$

and where $\Gamma(\cdot)$ is the gamma function. We consider the process $Z_{\exp}^{\delta,\beta,\varepsilon} := (Z_{\exp}^{\delta,\beta,\varepsilon}(t), 0 \le t \le T)$ defined by the weighted average

$$Z_{\exp}^{\delta,\beta,\varepsilon}(t) := \int_0^t k_{\exp}^{\delta,\beta}(t-s) X^\varepsilon(s)\,\mathrm{d}s.$$

The process $Z_{\exp}^{\delta,\beta,\varepsilon}$ can be interpreted as a smoothed version of the original trajectory X^ε. In fact, in the field of signal processing the kernel (2.1) belongs to the class of low-pass linear time-invariant filters, which cut the high frequencies in a signal to highlight its slowest components. In the following, rigorous analysis is conducted only when $\beta = 1$. Nonetheless, numerical experiments show that for higher values of β the performances of estimators computed employing the filter are more robust and qualitatively better.

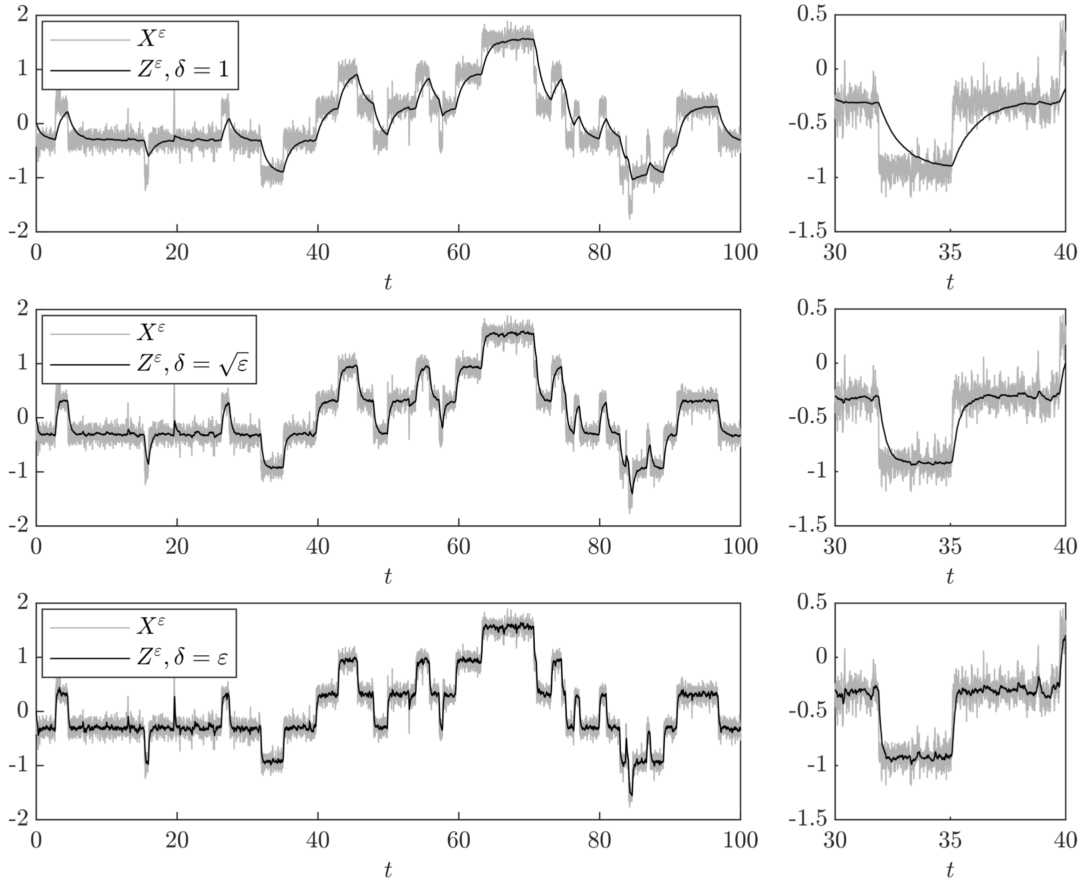


Figure 2.1 – Filtering a trajectory X^ε obtained with $V(x) = x^2/2$, $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 0.5$ and $\varepsilon = 0.1$. The filtering width is $\delta = \{1, \sqrt{\varepsilon}, \varepsilon\}$ from top to bottom, respectively, and $\beta = 1$.

Remark 2.1. Given a trajectory X^ε, it is relatively inexpensive to compute $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$ from a computational standpoint. In particular, the process $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$ is the truncated convolution of the kernel with the process X^ε. Hence, computational tools based on the Fast Fourier Transform (FFT) exist and allow to compute $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$ fast component-wise. Moreover, the process $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$ can be computed, in case $\beta = 1$, in a recursive manner and therefore "online".

Given a trajectory X^ε and the filtered data $Z_{\text{exp}}^{\delta,\beta,\varepsilon}$, the estimator of the drift coefficient we propose is given by

$$\widehat{A}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T) = -(\widetilde{M}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T))^{-1} \widetilde{v}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T), \tag{2.2}$$

where we employ the subscript exp and the superscript δ, β for reference to the filter's kernel in (2.1), and where

$$\begin{aligned} \widetilde{M}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T) &= \frac{1}{T} \int_0^T V'(Z_{\text{exp}}^{\delta,\beta,\varepsilon}(t)) \otimes V'(X^\varepsilon(t)) \, \mathrm{d}t, \\ \widetilde{v}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T) &= \frac{1}{T} \int_0^T V'(Z_{\text{exp}}^{\delta,\beta,\varepsilon}(t)) \, \mathrm{d}X^\varepsilon(t). \end{aligned} \tag{2.3}$$

Let us remark that the formula above is obtained from (1.15) by replacing only one instance of X_t^ε with $Z_{\text{exp}}^{\delta,\beta,\varepsilon}(t)$ in both M and v. In particular, it is fundamental for proving unbiasedness to keep in the definition of v the differential of the original process $\mathrm{d}X^\varepsilon(t)$ (see Remark 2.7). Let us furthermore remark that $\widehat{A}_{\text{exp}}^{\delta,\beta}(X^\varepsilon, T)$ need not be the minimizer of some likelihood function based on filtered data. In fact, if one were to replace $Z_{\exp}^{\delta,\beta,\varepsilon}(t)$ directly in (1.13), the symmetric part of the matrix $\widetilde{M}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ would appear and $\widehat{A}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ would not be the minimizer. Therefore, the estimator $\widehat{A}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ has to be thought of as a perturbation of $\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)$, directly at the level of estimators and after the maximization procedure. The only theoretical guarantee which is still needed for the well-posedness of $\widehat{A}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ is for $\widetilde{M}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ to be invertible, which we assume to be true and which we observed to hold in practice.

We now consider the diffusion coefficient, and propose the estimator for Σ in (1.11) given by

$$\widehat{\Sigma}_{\exp}^{\delta,1}(X^\varepsilon, T) \coloneqq \frac{1}{\delta T} \int_0^T \left(X^\varepsilon(t) - Z_{\exp}^{\delta,\beta,\varepsilon}(t)\right)^2 \,\mathrm{d}t, \tag{2.4}$$

where again we employ the same subscript and superscript for reference to the kernel (2.1) of the filter. As we will show in the following, and in particular in Theorem 2.19, the estimator $\widehat{\Sigma}_{\exp}^{\delta,1}(X^\varepsilon, T)$ is unbiased for the effective diffusion coefficient Σ in case $\beta = 1$ and when we filter data at the multiscale regime, i.e., when δ is a vanishing function of ε.

Let us from now on consider $\beta = 1$. For this value of β, the parameter δ appearing in (2.1) regulates the width of the filtering window. In practice, larger values of δ will lead to trajectories which are smoother and for which fast-scale oscillations are practically canceled. Let us remark that the filtering width resembles the subsampling step employed for the estimator $\widehat{A}_{\mathrm{sub}}^{\delta}(X^\varepsilon, T)$ introduced and analyzed in [103]. For subsampling, the choice guaranteeing asymptotically unbiased results is $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$, and a similar analysis is due for our technique. For visualization purposes, we depict in Figure 2.1 the filtered trajectory $Z_{\exp}^{\delta,\beta,\varepsilon}$ for three different values of δ, namely $\delta = \{1, \sqrt{\varepsilon}, \varepsilon\}$. With $\delta = 1$, all oscillations at the fast scale are canceled and the filtered trajectory $Z_{\exp}^{\delta,\beta,\varepsilon}$ presents only slow-scale variations. Reducing the value of δ, fast-scale oscillations are progressively taken into account.

In the following, we first focus on the ergodic properties of the process $Z_{\exp}^{\delta,\beta,\varepsilon}$ when it is coupled with the process X^ε. This analysis is practically independent of the choice of δ, and is therefore presented on its own. Then, we focus on two different cases which depend on the choice of the width δ of the filter. First, in Section 2.1.2, we consider δ to be independent of ε, and therefore we filter at the speed of the homogenized process. In this case, we are able to prove that our estimator of the drift coefficient of the homogenized equation is asymptotically unbiased almost surely. This result will be presented in Theorem 2.12. We then move on in Section 2.1.3 to the case $\delta \propto \varepsilon^\zeta$, which corresponds to filtering the data at the speed of the multiscale process. In this case, we show that under some conditions on the exponent ζ, we can still obtain estimators which are asymptotically unbiased. This result is proved in Theorem 2.17. For this second case, we widely employ techniques and estimates which come from [103].

Let us finally remark that, for economy of notation, from now on and until the end of this chapter we will simply write X_t^ε and Z_t^ε instead of $X^\varepsilon(t)$ and $Z_{\exp}^{\delta,\beta,\varepsilon}(t)$, respectively, and similarly for all stochastic processes. Moreover, we drop explicit reference to the dependence of $\widetilde{M}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ and $\widetilde{v}_{\exp}^{\delta,\beta}(X^\varepsilon, T)$ on the parameters δ and β and we only write $\widetilde{M}_{\exp}(X^\varepsilon, T)$ and $\widetilde{v}_{\exp}(X^\varepsilon, T)$.

2.1.1 Ergodic properties

Let us consider the filtering kernel (2.1) with $\beta = 1$, i.e.,

$$k_{\exp}^{\delta,1}(r) = \frac{1}{\delta} e^{-r/\delta}. \tag{2.5}$$

In this case, Leibniz integral rule yields the equality

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

which can be interpreted as an ordinary differential equation for Z_t^ε driven by the stochastic signal X^ε. Considering the processes X^ε and Z^ε together, we obtain the system of two one-dimensional stochastic differential equations (SDEs)

$$\begin{aligned}
\mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\
\mathrm{d}Z_t^\varepsilon &= \frac{1}{\delta}\left(X_t^\varepsilon - Z_t^\varepsilon\right)\mathrm{d}t.
\end{aligned} \tag{2.6}$$

The first ingredient for verifying the ergodic properties of the two-dimensional stochastic process $(X^\varepsilon, Z^\varepsilon)^\top := ((X_t^\varepsilon, Z_t^\varepsilon)^\top, 0 \leq t \leq T)$ is verifying that the measure induced by the stochastic process admits a smooth density with respect to the Lebesgue measure. Since noise is present only on the first component, this is a consequence of the theory of hypo-ellipticity, as summarized in the following Lemma, whose proof is given in Section 2.4.1.

Lemma 2.2. *Let $(X^\varepsilon, Z^\varepsilon)^\top$ be the solution of (2.6) and let $\mathfrak{m}_t^\varepsilon$ be the measure induced by the joint process at time t. Then, the measure $\mathfrak{m}_t^\varepsilon$ admits a smooth density with respect to the Lebesgue measure.*

Once it is established that the law of the process admits a smooth density for all times $t > 0$, which satisfies a time-dependent Fokker–Planck equation (FPE), we are interested in the limiting properties of this law. In particular, we know that the process X^ε alone is geometrically ergodic [85, Theorem 4.4], and we wish the couple $(X^\varepsilon, Z^\varepsilon)^\top$ to inherit the same property. The following Lemma guarantees that the couple is indeed geometrically ergodic, and its proof is given in Section 2.4.1.

Lemma 2.3. *Let Assumption 1.4 hold and let $b > 0$ be given in Assumption 1.4(ii). Then, if $\delta > 1/(4b)$, the process $(X^\varepsilon, Z^\varepsilon)^\top$ solution of (2.6) is geometrically ergodic, i.e., there exists $C, \lambda > 0$ such that for all measurable $f\colon \mathbb{R}^2 \to \mathbb{R}$ such that for some integer $q > 0$*

$$f(x,z) \leq 1 + \left\|(x,z)^\top\right\|_2^q,$$

it holds

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

for ρ^ε-a.e. couple $(X_0^\varepsilon, Z_0^\varepsilon)^\top$, where $\mathbb{E}$ denotes expectation with respect to the Wiener measure, and $\mu_{\exp}^\varepsilon$ is the invariant measure of the couple $(X^\varepsilon, Z^\varepsilon)^\top$. Moreover, the density $\rho_{\exp}^\varepsilon$ of $\mu_{\exp}^\varepsilon$ with respect to the Lebesgue measure is the solution to the stationary FPE

$$\sigma\partial_{xx}^2\rho_{\exp}^\varepsilon(x,z) + \partial_x\left(\left(\alpha\cdot V'(x) + \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)\rho_{\exp}^\varepsilon(x,z)\right) + \frac{1}{\delta}\partial_z\left((z-x)\rho_{\exp}^\varepsilon(x,z)\right) = 0. \tag{2.7}$$

Remark 2.4. The condition $\delta > 1/(4b)$ is not very restrictive. Let the parameter dimension $L = 1$ and let $V(x) \propto x^{2r}$ for an integer $r > 1$. Then, Assumption 1.4(ii) holds for an arbitrarily large $b > 0$. Therefore, the parameter of the filter δ can be chosen along the entire positive real axis. A similar argument can be employed for higher dimensions $L > 1$.

In a general case, it is not possible to find an explicit solution to (2.7). Nevertheless, it is possible to show some relevant properties of the solution itself, which are summarized in the following Lemma, whose proof is given in Section 2.4.1.

Lemma 2.5. *Under the assumptions of Lemma 2.3, let $\rho^\varepsilon_{\text{exp}}$ be the solution of* (2.7) *and let us write*

$$\rho^\varepsilon_{\text{exp}}(x,z) = \varphi^\varepsilon(x)\mathfrak{R}^\varepsilon_{\text{exp}}(x,z), \tag{2.8}$$

where φ^ε is the marginal density of the invariant measure ν^ε of X^ε, i.e.,

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

Then, it holds

$$\varphi^\varepsilon(x) = \frac{1}{C_{\nu^\varepsilon}} \exp\left(-\frac{1}{\sigma}\alpha\cdot V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right), \tag{2.9}$$

where

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

Moreover, it holds

$$\sigma\delta\int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)\varphi^\varepsilon(x)\partial_x\mathfrak{R}^\varepsilon_{\text{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z = \mathbb{E}^{\mu^\varepsilon_{\text{exp}}}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)]. \tag{2.10}$$

Remark 2.6. Lemma 2.5, and in particular the equality (2.10), plays a fundamental role in the proof of unbiasedness of the estimator based on filtered data. In particular, this equality allows to bypass the explicit knowledge of the function $\mathfrak{R}_{\text{exp}}(x,z)$, which governs the correlation between the processes X^ε and Z^ε at stationarity, for which a closed-form expression is not available in the general case.

Remark 2.7. Let us return to the definition of $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon,T)$ and replace the differential $\mathrm{d}X^\varepsilon_t$ with $\mathrm{d}Z^\varepsilon_t$ in $\widetilde{v}_{\text{exp}}(X^\varepsilon,T)$. In this case, if $\beta=1$ it holds

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T V'(Z^\varepsilon_t)\,\mathrm{d}Z^\varepsilon_t = \lim_{T\to\infty}\frac{1}{\delta T}\int_0^T V'(Z^\varepsilon_t)(X^\varepsilon_t - Z^\varepsilon_t)\,\mathrm{d}t = \frac{1}{\delta}\,\mathbb{E}^{\mu^\varepsilon_{\text{exp}}}\left[V'(Z^\varepsilon)(X^\varepsilon - Z^\varepsilon)\right] = 0,$$

where the last equality is obtained as in the proof of Lemma 2.5, with the choice $f(x,z)=V(z)$ at the last line. Therefore, we stress again that it is indeed necessary to employ the original differential $\mathrm{d}X^\varepsilon_t$ in the vector $\widetilde{v}_{\text{exp}}(X^\varepsilon,T)$ in the definition (2.2) of $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon,T)$.

Remark 2.8. Let us consider the kernel (2.1) with $\beta>1$. In this case, the steps leading to the system (2.6) do not yield a system of Itô SDEs, but of stochastic delay differential equations. In fact, by the Leibniz integral rule we have

$$\begin{aligned}
\mathrm{d}Z^\varepsilon_t &= k^{\delta,\beta}_{\text{exp}}(0)X^\varepsilon_t\,\mathrm{d}t + \left(\int_0^t (k^{\delta,\beta}_{\text{exp}})'(t-s)X^\varepsilon_s\,\mathrm{d}s\right)\mathrm{d}t \\
&= C_\beta\delta^{-1/\beta}X^\varepsilon_t\,\mathrm{d}t - \left(\frac{\beta}{\delta}\int_0^t (t-s)^{\beta-1}k^{\delta,\beta}_{\text{exp}}(t-s)X^\varepsilon_s\,\mathrm{d}s\right)\mathrm{d}t,
\end{aligned}$$

The analysis of the estimator in case $\beta>1$ is therefore based on different arguments than the one we present here.

2.1.2 Filtered data in the homogenized regime

In this section, we analyze the behavior of the estimator $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon,T)$ based on filtered data given in (2.2) when the filtering width δ is independent of ε. The analysis in this case is based on the convergence of the couple $(X^\varepsilon,Z^\varepsilon)^\top$ with respect to the multiscale parameter $\varepsilon\to 0$. In particular, it is known that the invariant measure of X^ε converges weakly to the invariant measure of X^0, the solution of the homogenized equation (1.11). The following result guarantees the same kind of convergence for the couple $(X^\varepsilon,Z^\varepsilon)^\top$.

Lemma 2.9. *Under Assumption 1.4, let $\mu^\varepsilon_{\mathrm{exp}}$ be the invariant measure of the couple $(X^\varepsilon, Z^\varepsilon)^\top$. If δ is independent of ε, then the measure $\mu^\varepsilon_{\mathrm{exp}}$ converges weakly to the measure $\mu^0_{\mathrm{exp}}(\mathrm{d}x, \mathrm{d}z) = \rho^0_{\mathrm{exp}}(x, z)\,\mathrm{d}x\,\mathrm{d}z$, whose density ρ^0_{exp} is the unique solution of the FPE*

$$\Sigma\partial^2_{xx}\rho^0_{\mathrm{exp}}(x, z) + \partial_x\left(A \cdot V'(x)\rho^0_{\mathrm{exp}}(x, z)\right) + \frac{1}{\delta}\partial_z\left((z - x)\rho^0_{\mathrm{exp}}(x, z)\right) = 0, \tag{2.11}$$

where A and Σ are the coefficients of the homogenized equation (1.11).

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

$$\begin{aligned} \mathrm{d}X^0_t &= -A \cdot V'(X^0_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t, \\ \mathrm{d}Z^0_t &= \frac{1}{\delta}\left(X^0_t - Z^0_t\right)\,\mathrm{d}t, \end{aligned}$$

with $(X^0_0, Z^0_0)^\top \sim \mu^0_{\mathrm{exp}}$. The arguments of Section 2.1.1 can be repeated to conclude that the invariant measure of $(X^0, Z^0)^\top$ admits a smooth density ρ^0_{exp} which satisfies (2.11). Moreover, standard homogenization theory (see e.g. [20, Chapter 3, Theorem 6.4] or [104, Theorem 18.1]) guarantees that $(X^\varepsilon, Z^\varepsilon)^\top \to (X^0, Z^0)^\top$ for $\varepsilon \to 0$ in law as random variables with values in $\mathcal{C}^0([0, T]; \mathbb{R}^2)$, provided that $(X^\varepsilon_0, Z^\varepsilon_0)^\top \sim \mu^\varepsilon_{\mathrm{exp}}$. We remark that traditionally it is assumed that the initial conditions satisfy $(X^\varepsilon_0, Z^\varepsilon_0)^\top = (X^0_0, Z^0_0)^\top$ for the homogenization result to hold, but notice that the proof of e.g. [104, Theorem 18.1] can be shown to hold with a minor modification in case both the multiscale and the homogenized processes are at stationarity. Denoting $E = C^0([0, T], \mathbb{R}^2)$, this means that the measure induced by $(X^\varepsilon, Z^\varepsilon)^\top$ on $(E, \mathcal{B}(E))$ converges weakly to the measure induced by $(X^0, Z^0)^\top$ on the same measurable space (see e.g. [104, Definition 3.24]). Hence, the measure $\mu^\varepsilon_{\mathrm{exp}}$ converges weakly to μ^0_{exp} for $\varepsilon \to 0$. □

Example 2.10. A closed form solution of (2.11) can be obtained in a simple case. Let the dimension of the parameter $L = 1$ and let $V(x) = x^2/2$. Then, the analytical solution is given by

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

where

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

This is the density of a multivariate normal distribution $\mathcal{N}(0, \Gamma)$, where the covariance matrix is given by

$$\Gamma = \frac{\Sigma}{A(1 + A\delta)}\begin{pmatrix} 1 + A\delta & 1 \\ 1 & 1 \end{pmatrix}.$$

Let us remark that this distribution can be obtained from direct computations involving Gaussian processes. In particular, we have that X^0 is in this case an Ornstein–Uhlenbeck (OU) process and it is therefore known that $X^0 \sim \mathcal{GP}(m_t, \mathcal{C}(t, s))$, where at stationarity $m_t = 0$ and

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

The basic properties of Gaussian processes imply that Z^0 is a Gaussian process, and that the couple $(X^0, Z^0)^\top$ is a Gaussian process, too, whose mean and covariance are computable explicitly.

We now present an analogous result to Lemma 2.5 for the limit distribution.

Corollary 2.11. *Let $\rho^0_{\exp}$ be the solution of* (2.11) *and let us write*

$$\rho^0_{\exp}(x,z) = \varphi^0(x)\mathfrak{R}^0_{\exp}(x,z), \tag{2.12}$$

where φ^0 is the marginal density of the invariant measure ν^0 of X^0, i.e.,

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

Then, if A and Σ are the coefficients of the homogenized equation (1.11)*, it holds*

$$\varphi^0(x) = \frac{1}{C_{\nu^0}} \exp\left(-\frac{1}{\Sigma} A \cdot V(x)\right), \quad \text{where} \quad C_{\nu^0} = \int_{\mathbb{R}} \exp\left(-\frac{1}{\Sigma} A \cdot V(x)\right) \mathrm{d}x. \tag{2.13}$$

Moreover, it holds

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

Proof. The proof is directly obtained from Lemma 2.5 setting $p(y) = 0$ and replacing α, σ by A, Σ respectively. □

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

$$\widetilde{\mathcal{M}}^\varepsilon_{\exp} := \mathbb{E}^{\mu^\varepsilon_{\exp}}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)], \quad \widetilde{\mathcal{M}}^0_{\exp} := \mathbb{E}^{\mu^0_{\exp}}[V'(Z^0) \otimes V'(X^0)], \tag{2.14}$$

i.e., $\widetilde{\mathcal{M}}^\varepsilon_{\exp}$ is obtained in the limit for $T \to \infty$ applying the ergodic theorem elementwise to the matrix $\widetilde{M}_{\exp}(X^\varepsilon, T)$, and $\widetilde{\mathcal{M}}^0_{\exp}$ is the limit for $\varepsilon \to 0$ of the matrix $\widetilde{\mathcal{M}}^\varepsilon_{\exp}$ due to Lemma 2.9. For completeness, we introduce here the symmetric matrices $\mathcal{M}^\varepsilon$ and $\mathcal{M}^0$ which are defined as

$$\mathcal{M}^\varepsilon := \mathbb{E}^{\nu^\varepsilon}[V'(X^\varepsilon) \otimes V'(X^\varepsilon)], \quad \mathcal{M}^0 := \mathbb{E}^{\nu^0}[V'(X^0) \otimes V'(X^0)], \tag{2.15}$$

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

Theorem 2.12. *Let the assumptions of Lemma 2.3 and Lemma 2.9 hold, and let $\widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T)$ be defined in* (2.2) *with δ independent of ε and $\beta = 1$. If $\widetilde{M}_{\exp}(X^\varepsilon, T)$ is invertible, then*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T) = A, \quad a.s.,$$

where A is the drift coefficient of the homogenized equation (1.11).

Proof. Replacing the expression of $\mathrm{d}X^\varepsilon_t$ into (2.3), we get for $\widetilde{v}_{\exp}(X^\varepsilon, T)$

$$\widetilde{v}_{\exp}(X^\varepsilon, T) = -\widetilde{M}_{\exp}(X^\varepsilon, T)\alpha - \frac{1}{T}\int_0^T \frac{1}{\varepsilon} p'\left(\frac{X^\varepsilon_t}{\varepsilon}\right) V'(Z^\varepsilon_t)\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{T}\int_0^T V'(Z^\varepsilon_t)\,\mathrm{d}W_t.$$

Therefore, we have

$$\begin{aligned} \widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T) &= \alpha + \frac{1}{T}\widetilde{M}_{\exp}(X^\varepsilon, T)^{-1} \int_0^T \frac{1}{\varepsilon} p'\left(\frac{X^\varepsilon_t}{\varepsilon}\right) V'(Z^\varepsilon_t)\,\mathrm{d}t \\ &\quad - \frac{\sqrt{2\sigma}}{T}\widetilde{M}_{\exp}(X^\varepsilon, T)^{-1} \int_0^T V'(Z^\varepsilon_t)\,\mathrm{d}W_t \\ &=: \alpha + I^\varepsilon_1(T) - I^\varepsilon_2(T). \end{aligned} \tag{2.16}$$

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

$$\lim_{T\to\infty} I_1^\varepsilon(T) = (\widetilde{\mathcal{M}}_{\mathrm{exp}}^\varepsilon)^{-1}\,\mathbb{E}^{\mu_{\mathrm{exp}}^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right], \quad \text{a.s.} \tag{2.17}$$

Replacing the decomposition (2.8), the expression (2.9) of φ^ε and integrating by parts, we have

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

which implies

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

Replacing the equality above into (2.17), we obtain

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

Due to Lemma 2.5, we therefore have

$$\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta}(\widetilde{\mathcal{M}}_{\mathrm{exp}}^\varepsilon)^{-1}\,\mathbb{E}^{\mu_{\mathrm{exp}}^\varepsilon}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)], \quad \text{a.s.} \tag{2.18}$$

Since δ is independent of ε, we can pass to the limit as ε goes to zero and Lemma 2.9 yields

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta}(\widetilde{\mathcal{M}}_{\mathrm{exp}}^0)^{-1}\,\mathbb{E}^{\mu_{\mathrm{exp}}^0}[(X^0 - Z^0)^2 V''(Z^0)], \quad \text{a.s.} \tag{2.19}$$

Due to Corollary 2.11, we have

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

and moreover, an integration by parts yields

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

We can therefore conclude that

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} I_1^\varepsilon(T) = -\alpha + A, \quad \text{a.s.} \tag{2.20}$$

We now consider the second term $I_2^\varepsilon(T)$, and rewrite it as

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

where

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

The ergodic theorem yields

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

where R^ε is bounded uniformly in ε due to the theory of homogenization, Assumption 1.4(iii)-1.7 and Lemma 2.28. Moreover, always due to Lemma 2.28 and Assumption 1.4(iii) we have that $V'(Z^\varepsilon)$ is square integrable, and hence the strong law of large numbers for martingales implies

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

independently of ε. Therefore

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

which, together with (2.20) and (2.16), proves the desired result. □

Remark 2.13. Let us remark that the assumption that δ is independent of ε is necessary to pass from (2.18) to (2.19) but is not needed before (2.18). Moreover, the term $I_2^\varepsilon(t)$ in the proof vanishes a.s. independently of ε. Therefore, in the analysis of the case $\delta = \mathcal{O}(\varepsilon^\zeta)$ it will be sufficient for unbiasedness to show that

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

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

2.1.3 Filtered data in the multiscale regime

We now consider the case of the filtering width $\delta = \mathcal{O}(\varepsilon^\zeta)$, where $\zeta > 0$ will be specified in the following. In this case, the filtered process resembles more the original process X^ε, as can be noted in Figure 2.1. Moreover, the techniques employed for proving Theorem 2.12 can only be partly exploited, as highlighted by Remark 2.13. In fact, in order to prove unbiasedness it is necessary to characterize precisely the difference between the processes Z^ε and X^ε. A first characterization is given by the following Proposition, whose proof is found in Section 2.4.2.

Proposition 2.14. *Let Assumption 1.4 hold and let $\beta = 1$ and $\varepsilon, \delta > 0$ be sufficiently small. Then, it holds for every $t > 0$*

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

where the stochastic process B_t^ε is defined as

$$B_t^\varepsilon \coloneqq \sqrt{2\sigma} \int_0^t k_{\mathrm{exp}}^{\delta,\beta}(t-s)(1 + \Phi'(Y_s^\varepsilon)) \,\mathrm{d}W_s, \tag{2.21}$$

where Φ is the solution of the cell problem (1.7)*, W_s is the Brownian motion appearing in* (1.10) *and $Y_t^\varepsilon = X_t^\varepsilon/\varepsilon$. Moreover, B_t^ε and the remainder $R(\varepsilon,\delta)$ satisfy for every $p \geq 1$ the estimates*

$$\left(\mathbb{E}^{\nu^\varepsilon} |B_t^\varepsilon|^p\right)^{1/p} \leq C\delta^{-1/2}, \tag{2.22}$$

and

$$\left(\mathbb{E}^{\nu^\varepsilon} |R(\varepsilon,\delta)|^p\right)^{1/p} \leq C\left(\delta + \varepsilon + \max\{1,t\}e^{-t/\delta}\right), \tag{2.23}$$

where C is independent of ε, δ and t, and ν^ε is the invariant measure of X^ε.

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

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

This equation can be coupled with the dynamics of the processes X_t^ε, Y_t^ε and Z_t^ε, thus describing the evolution of the quadruple $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon, B^\varepsilon)$ together. In particular, it is possible to show that the results of Section 2.1.1 hold for the quadruple, and the properties of the invariant measure of the quadruple can be exploited to prove the unbiasedness of the estimator in the case $\delta = \mathcal{O}(\varepsilon^\zeta)$ in the same way as in the case δ independent of ε.

In light of Remark 2.13, it is fundamental to understand the behavior of the quantity

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

as well as its limit for $t \to \infty$ and for $\varepsilon \to 0$. Let us remark that due to Proposition 2.14 we have

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

and therefore studying the right hand side of the approximate equality above is the goal of the upcoming discussion. The following result, whose proof is in Section 2.4.3, gives a first characterization.

Lemma 2.15. *Under Assumption 1.4, let η^ε be the invariant measure of the 4-dimensional process $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon, B^\varepsilon)$. Then it holds*

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

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

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

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

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

where π is the marginal measure of the process Y^ε when coupled with X^ε. Therefore, the next step is the homogenization limit, i.e., the limit of vanishing ε, which is considered in the following Lemma, and whose proof is given in Section 2.4.3.

Lemma 2.16. *Let the assumptions of Lemma 2.15 hold, and let $\delta = \varepsilon^\zeta$ with $\zeta > 0$. Then, it holds*

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

where Σ is the diffusion coefficient of the homogenized equation (1.11)*.*

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

Theorem 2.17. *Let the assumptions of Lemma 2.3 and Lemma 2.16 hold. Let $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon, T)$ be defined in* (2.2) *and let $\beta = 1$ and $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$. If $\widetilde{M}_{\text{exp}}(X^\varepsilon, T)$ is invertible, then*

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

where A is the drift coefficient of the homogenized equation (1.11)*.*

Proof. Let us introduce the notation

$$\mathcal{A}^\varepsilon(\delta) \coloneqq \frac{1}{\delta} (\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}})^{-1} \, \mathbb{E}^{\mu^\varepsilon_{\text{exp}}} [(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)],$$

where $\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}}$ is defined in (2.14). Then following the proof of Theorem 2.12 and in light of Remark 2.13, we only need to show that if $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$ we have

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

Using Proposition 2.14 and geometric ergodicity for taking the limit for $t \to \infty$ (Lemma 2.3), we have the following equality

$$\begin{aligned}
\mathcal{A}^\varepsilon(\delta) &= (\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}})^{-1} \frac{1}{\delta} \lim_{t \to \infty} \mathbb{E}[(X^\varepsilon_t - Z^\varepsilon_t)^2 V''(Z^\varepsilon_t)] \\
&= (\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}})^{-1} \frac{1}{\delta} \lim_{t \to \infty} \mathbb{E} \left[(\delta B^\varepsilon_t + R(\varepsilon, \delta))^2 \, V''(Z^\varepsilon_t) \right] \\
&=: (\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}})^{-1} \lim_{t \to \infty} \left(J^\varepsilon_1(t) + J^\varepsilon_2(t) + J^\varepsilon_3(t) \right),
\end{aligned}$$

where $R(\varepsilon, \delta)$ is given in Proposition 2.14, $\mathbb{E}$ denotes the expectation with respect to the Wiener measure and

$$\begin{aligned}
J^\varepsilon_1(t) &= \delta \, \mathbb{E} \left[(B^\varepsilon_t)^2 V''(Z^\varepsilon_t) \right], \\
J^\varepsilon_2(t) &= 2 \, \mathbb{E} \left[B^\varepsilon_t R(\varepsilon, \delta) V''(Z^\varepsilon_t) \right], \\
J^\varepsilon_3(t) &= \frac{1}{\delta} \, \mathbb{E} \left[R(\varepsilon, \delta)^2 V''(Z^\varepsilon_t) \right].
\end{aligned}$$

Let us consider the three terms separately. First, by geometric ergodicity and applying Lemma 2.15 and Lemma 2.16 we get

$$\begin{aligned}
\lim_{\varepsilon \to 0} \lim_{t \to \infty} J^\varepsilon_1(t) &= \lim_{\varepsilon \to 0} \delta \, \mathbb{E}^{\eta^\varepsilon} \left[(B^\varepsilon)^2 V''(Z^\varepsilon) \right] \\
&= \lim_{\varepsilon \to 0} \left(\sigma \, \mathbb{E}^{\eta^\varepsilon} [V''(Z^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] + \widetilde{R}(\varepsilon, \delta) \right) \\
&= \Sigma \, \mathbb{E}^{\nu^0} [V''(X^0)].
\end{aligned}$$

Let us now consider $J^\varepsilon_2(t)$. Considering Hölder conjugates p, q, r the Hölder inequality yields

$$\|J^\varepsilon_2(t)\| \le \mathbb{E}[|B^\varepsilon_t|^p]^{1/p} \, \mathbb{E}[|R(\varepsilon, \delta)|^q]^{1/q} \, \mathbb{E}[\|V''(Z^\varepsilon)\|^r]^{1/r}.$$

Now, we can bound the first two terms with (2.22) and (2.23), respectively. The third term is bounded due to Assumption 1.4 and Lemma 2.28. Hence, we have for t sufficiently large

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

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

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

which, similarly as above, yields for t sufficiently large

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

Therefore, since $\delta = \mathcal{O}(\varepsilon^\zeta)$ for $\zeta \in (0,2)$, the terms $J_2^\varepsilon(t)$ and $J_3^\varepsilon(t)$ vanish in the limit for $t \to \infty$ and $\varepsilon \to 0$. Furthermore, by Lemma 2.31 and by weak convergence of the invariant measure $\mu_{\text{exp}}^\varepsilon$ to μ_{exp}^0, we have

$$\lim_{\varepsilon\to 0} \widetilde{\mathcal{M}}_{\text{exp}}^\varepsilon = \mathcal{M}^0,$$

where $\mathcal{M}^0$ is defined in (2.15). Therefore

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

and, finally, employing (2.13) and (2.15) and integrating by parts yields

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

which implies the desired result. □

We conclude the analysis concerning the estimator $\widehat{A}_{\text{exp}}^{\delta,\beta}(X^\varepsilon,T)$ for the effective drift coefficient with a negative convergence result, i.e., that if $\delta = \varepsilon^\zeta$ with $\zeta > 2$, the estimator based on filtered data converges to the coefficient α of the unhomogenized equation. This result is relevant for two reasons. First, it shows the sharpness of the bound on ζ in the assumptions of Theorem 2.17. Second, it shows an interesting switch between two completely different regimes at $\zeta = 2$, which happens arbitrarily fast in the limit $\varepsilon \to 0$.

Theorem 2.18. *Let the assumptions of Lemma 2.3 hold. Let $\widehat{A}_{\text{exp}}^{\delta,\beta}(X^\varepsilon,T)$ be defined in* (2.2) *and let $\beta = 1$ and $\delta = \varepsilon^\zeta$ with $\zeta > 2$. If $\widetilde{M}_{\text{exp}}(X^\varepsilon,T)$ is invertible, then*

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{A}_{\text{exp}}^{\delta,\beta}(X^\varepsilon,T) = \alpha, \quad \textit{in probability},$$

where α is the drift coefficient of the multiscale equation (1.10).

The proof is given in Section 2.4.3.

We conclude this section by proving a result of asymptotic unbiasedness for the estimator $\widehat{\Sigma}_{\text{exp}}^{\delta,1}(X^\varepsilon,T)$ of the effective diffusion coefficient Σ defined in (2.4). The proof is given in Section 2.4.4.

Theorem 2.19. *Let the Assumptions of Theorem 2.18 hold. Then, if $\delta = \varepsilon^\zeta$, with $\zeta \in (0,2)$, it holds*

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{\Sigma}_{\text{exp}}^{\delta,1}(X^\varepsilon,T) = \Sigma, \quad \textit{in probability},$$

where Σ is the diffusion coefficient of the homogenized equation (1.11).

2.2 The Bayesian setting

In this section we present a Bayesian reinterpretation of the inference procedure, which, given the structure of the problem, allows full uncertainty quantification with little more computational effort than required for the MLE.

Let us fix a Gaussian prior $\pi_0 = \mathcal{N}(A_0, C_0)$ on A, where $A_0 \in \mathbb{R}^L$ and $C_0 \in \mathbb{R}^{L\times L}$ is symmetric positive definite. Then, given a final time $T > 0$, the posterior distribution $\pi_{T,\varepsilon}$ admits a density $p_T(A \mid X^\varepsilon)$ with respect to the Lebesgue measure which satisfies

$$p_T(A \mid X^\varepsilon) = \frac{1}{C_{\pi_{T,\varepsilon}}} \mathfrak{L}_T(X^\varepsilon \mid A)\, p_0(A),$$

where $C_{\pi_{T,\varepsilon}}$ is the normalization constant, p_0 is the density of π_0, and where the likelihood $\mathfrak{L}_T(X^\varepsilon \mid A)$ is given in (1.13). The log-posterior density is therefore given by

$$\log p_T(A \mid X^\varepsilon) = -\log C_{\pi_{T,\varepsilon}} - \frac{T}{2\Sigma} A \cdot v(X^\varepsilon, T) - \frac{T}{4\Sigma} A \cdot M(X^\varepsilon, T) A - \frac{1}{2}(A - A_0) \cdot C_0^{-1}(A - A_0),$$

where $M(X^\varepsilon, T)$ and $v(X^\varepsilon, T)$ are defined in (1.15). Since the log-posterior density is quadratic in A, the posterior is Gaussian, and it is therefore sufficient to determine its mean and covariance to fully characterize it. We denote by $m_{T,\varepsilon}$ and $C_{T,\varepsilon}$ the mean and covariance matrix, respectively. Completing the squares in the log-posterior density, we formally obtain

$$\begin{aligned} C_{T,\varepsilon}^{-1} &= C_0^{-1} + \frac{T}{2\Sigma} M(X^\varepsilon, T), \\ C_{T,\varepsilon}^{-1} m_{T,\varepsilon} &= C_0^{-1} A_0 - \frac{T}{2\Sigma} v(X^\varepsilon, T). \end{aligned} \tag{2.24}$$

Under Assumption 1.4, one can show that the posterior at time $T > 0$ is well defined and given by $\pi_{T,\varepsilon}(\cdot \mid X^\varepsilon) = \mathcal{N}(m_{T,\varepsilon}, C_{T,\varepsilon})$. Let us remark that in order to compute the posterior covariance $C_{T,\varepsilon}$ the value of the diffusion coefficient Σ of the homogenized equation is needed. Although the exact value is in general unknown, it can be estimated employing the subsampling technique presented in [103] or with the estimator $\widehat{\Sigma}^{\delta,1}_{\mathrm{exp}}(X^\varepsilon, T)$ given in (2.4) based on filtered data. In fact, we verified in practice that the estimator of the diffusion coefficient based on subsampling is more robust with respect to the subsampling step than the estimator for the drift coefficient. In the following theorem, we show that the posterior distribution obtained with no pre-processing of the data contracts asymptotically to the drift coefficient of the unhomogenized equation. We characterize the contraction by verifying that the posterior measure concentrates in arbitrarily small balls. Let us finally remark that the measure $\pi_{T,\varepsilon}$ is a random measure, and therefore contraction has to be considered averaged with respect to the Wiener measure. The choice of the contraction measure and some parts of the proof are taken from [106, Theorem 5.2].

Theorem 2.20. *Under Assumption 1.4, the posterior measure $\pi_{T,\varepsilon}(\cdot \mid X^\varepsilon) = \mathcal{N}(m_{T,\varepsilon}, C_{T,\varepsilon})$ satisfies for all $c > 0$*

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

where $\mathbb{E}$ denotes expectation with respect to the Wiener measure and α is the drift coefficient of the unhomogenized equation (1.10)*.*

Remark 2.21. The result above has the same consequences in the Bayesian setting as Theorem 1.8 has for the MLE. In particular, it shows that the posterior distribution obtained when data are not pre-processed concentrates asymptotically on the drift coefficient of the unhomogenized equation (1.10). Moreover, a partial result which can be deduced from the proof is that in the limit for $T \to \infty$ and for a positive value $\varepsilon > 0$ the Bayesian and the MLE approaches are equivalent. In particular, we have for all $\varepsilon > 0$

$$\lim_{T\to\infty} \|C_{T,\varepsilon}\|_2 = 0,$$
$$\lim_{T\to\infty} \left\| m_{T,\varepsilon} - \widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T) \right\|_2 = 0,$$

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

Proof of Theorem 2.20. The proof of [106, Theorem 5.2] guarantees that if the trace of $C_{T,\varepsilon}$ tends to zero and if the mean $m_{T,\varepsilon}$ tends to α, then the desired result holds. Indeed, the triangle inequality yields

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

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

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

and a change of variable simply gives

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

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

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

where

$$Q = M(X^\varepsilon, T) + \frac{T}{2\Sigma} M(X^\varepsilon, T) C_0 M(X^\varepsilon, T).$$

Let us first remark that due to the hypothesis on $M(X^\varepsilon, T)$ (Assumption 1.41.7) and the ergodic theorem it holds for all $T > 0$

$$\left\|M(X^\varepsilon, T)^{-1}\right\|_2 \leq \frac{1}{\bar\lambda},$$

where $\bar\lambda$ is given in Assumption1.41.7. We now have that for generic symmetric positive definite matrices R and S it holds

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

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

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

which implies

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

and due to the triangle inequality

$$\lim_{T\to\infty} \|C_{T,\varepsilon}\|_2 = 0. \tag{2.25}$$

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

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

For the second term, Theorem 1.8 implies

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

Let us now consider the first term. Replacing the expression of the MLE (1.15) and due to the Cauchy–Schwarz and triangle inequalities, we obtain

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

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

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

independently of ε. Finally,

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

which, together with (2.25), implies the desired result. □

2.2.1 The filtered data approach

In this section, we present how to correct the asymptotic biasedness of the posterior highlighted by Theorem 2.20 employing filtered data. In the Bayesian setting, we consider the modified likelihood function

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

where

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

Since M is symmetric positive definite, the function $\widetilde{\mathfrak{L}}_T(X^\varepsilon \mid A)$ is indeed a valid Gaussian likelihood function. We then obtain the modified posterior $\widetilde{\pi}_{T,\varepsilon} = \mathcal{N}(\widetilde{m}_{T,\varepsilon}, C_{T,\varepsilon})$, whose parameters are given by

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

Let us remark that the posterior $\widetilde{\pi}_{T,\varepsilon}$ has the same covariance as $\pi_{T,\varepsilon}$ given in (2.24) and that therefore it is indeed a valid Gaussian posterior distribution. Nevertheless, in order to employ the tool of convergence introduced in Theorem 2.20, we need to study the properties of the MLE based on the likelihood $\widetilde{\mathfrak{L}}_T(X^\varepsilon \mid A)$, i.e., the quantity

$$\widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) = -M(X^\varepsilon, T)^{-1} \widetilde{v}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T). \tag{2.26}$$

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

Theorem 2.22. *Let the assumptions of Theorem 2.17 hold. Then, if $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$ and $\beta = 1$, it holds*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) = A, \quad a.s.,$$

for $\widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ defined in (2.26)*.*

Proof. We first consider the difference between the two estimators $\widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ and $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$. In particular, the ergodic theorem and an algebraic equality imply

$$\begin{aligned}
\lim_{T\to\infty} \left(\widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - \widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) \right) &= \left((\mathcal{M}^\varepsilon)^{-1} - (\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1} \right) \lim_{T\to\infty} \widetilde{v}_{\mathrm{exp}}(X^\varepsilon, T) \\
&= -(\mathcal{M}^\varepsilon)^{-1} \left(\mathcal{M}^\varepsilon - \widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}} \right) (\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1} \lim_{T\to\infty} \widetilde{v}_{\mathrm{exp}}(X^\varepsilon, T) \\
&= (\mathcal{M}^\varepsilon)^{-1} \left(\mathcal{M}^\varepsilon - \widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}} \right) \lim_{T\to\infty} \widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T),
\end{aligned}$$

almost surely, where $\mathcal{M}^\varepsilon$ and $\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}}$ are defined in (2.15) and (2.14), respectively. Therefore, due to Assumption 1.4 which allows controlling the norm of $(\mathcal{M}^\varepsilon)^{-1}$ and due to Lemma 2.31 we have for a constant $C > 0$

$$\lim_{T\to\infty} \left\| \widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - \widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) \right\|_2 \leq C \left(\varepsilon + \delta^{1/2} \right), \tag{2.27}$$

where we remark that $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ has a bounded norm for ε sufficiently small due to Theorem 2.17. Now, the triangle inequality yields

$$\left\| \widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - A \right\|_2 \leq \left\| \widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - \widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) \right\|_2 + \left\| \widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - A \right\|_2 .$$

Therefore, due to Theorem 2.17, the inequality (2.27) and since $\delta = \varepsilon^\zeta$, the desired result holds. $\square$

Remark 2.23. One could argue that we could have carried on the whole analysis for the estimator $\widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ instead of the estimator $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$. Nevertheless, the latter guarantees the strong result of almost sure convergence in case δ is independent of ε, which is false for the former. Conversely, analysing the properties of the estimator $\widetilde{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ is fundamental for the Bayesian setting, in which the matrix $\widetilde{M}_{\mathrm{exp}}(X^\varepsilon, T)$ cannot be employed as its symmetric part is not positive definite in general.

In light of the proof of Theorem 2.20, Theorem 2.22 guarantees that the mean of the posterior distribution $\widetilde{\pi}_{T,\varepsilon}$ converges to the drift coefficient of the homogenized equation. Since the covariance matrix is the same for $\pi_{T,\varepsilon}$ and $\widetilde{\pi}_{T,\varepsilon}$, it is possible to prove a positive convergence result for $\widetilde{\pi}_{T,\varepsilon}$, which is given by the following Theorem.

Theorem 2.24. *Let the Assumptions of Theorem 2.22 hold. Then, the modified posterior measure $\widetilde{\pi}_{T,\varepsilon}(\cdot \mid X^\varepsilon) = \mathcal{N}(\widetilde{m}_{T,\varepsilon}, C_{T,\varepsilon})$ satisfies*

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

where $\mathbb{E}$ denotes expectation with respect to the Wiener measure and A is the drift coefficient of the homogenized equation (1.11)*.*

Proof. The proof follows from the proof of Theorem 2.20 and from Theorem 2.22. □

2.3 Numerical experiments

In this section we show numerical experiments confirming our theoretical findings and showcasing the potential of the filtered data approach to overcome model misspecification arising when multiscale data are used to fit homogenized models.

Remark 2.25. In practice, we consider for numerical experiment the data to be in the form of a high-frequency discrete time series from the solution X^ε of (1.10). Let $\tau > 0$ be the time step at which data are observed, and let $X^\varepsilon \coloneqq (X_0^\varepsilon, X_\tau^\varepsilon, X_{2\tau}^\varepsilon, \dots)$. We then compute the estimator $\widehat{A}^{\delta,\beta}_{\exp,\tau}$ as

$$\widehat{A}^{\delta,\beta}_{\exp,\tau}(X^\varepsilon, T) = -\widetilde{M}^{\delta,\beta}_{\exp,\tau}(X^\varepsilon, T)^{-1} \widetilde{v}^{\delta,\beta}_{\exp,\tau}(X^\varepsilon, T),$$

where

$$\widetilde{M}^{\delta,\beta}_{\exp,\tau}(X^\varepsilon, T) = \frac{\tau}{T} \sum_{j=0}^{n-1} V'(Z^\varepsilon_{j\tau}) \otimes V'(X^\varepsilon_{j\tau}), \quad \widetilde{v}^{\delta,\beta}_{\exp,\tau}(X^\varepsilon, T) = \frac{1}{T} \sum_{j=0}^{n-1} V'(Z^\varepsilon_{j\tau})(X^\varepsilon_{(j+1)\tau} - X^\varepsilon_{j\tau}).$$

We take in all experiments $\tau \ll \varepsilon^2$, so that the discretization of the data has negligible effects and does not compromise the validity of our theoretical results.

2.3.1 Parameters of the filter

For the first preliminary experiments, we consider $L = 1$ and the quadratic potential $V(x) = x^2/2$. In this case, the solution of the homogenized equation is an OU process. Moreover, we set the the fast potential in the multiscale equation (1.10) as $p(y) = \cos(y)$. In all experiments, data are generated employing the Euler–Maruyama (EM) method with a fine time step.

Verification of theoretical results

We first demonstrate numerically the validity of Theorem 2.12, Theorem 2.17 and Theorem 2.18, i.e., the unbiasedness of $\widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T)$ for $\beta = 1$ and $\delta = \varepsilon^\zeta$ with $\zeta \in [0, 2)$ and biasedness for $\zeta > 2$. Let us recall that for $\zeta = 0$ the analysis and the theoretical result are fundamentally different than for $\zeta \in (0, 2)$. We consider $\varepsilon \in \{0.1, 0.05, 0.025\}$, the diffusion coefficient $\sigma = 1$ and generate data X_t^ε for $0 \le t \le T$ with $T = 10^3$. Then we filter the data by choosing $\delta = \varepsilon^\zeta$, and $\zeta = 0, 0.1, 0.2, \dots, 3$, and compute $\widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T)$. Results are displayed in Figure 2.2, and show that for $\zeta > 2$, i.e., $\delta = o(\varepsilon^2)$, the estimator tends to the drift coefficient α of the unhomogenized equation. Conversely, as predicted by the theory, for $\zeta \in [0, 2)$ the estimator tends to A, the drift coefficient of the homogenized equation. Therefore, the point $\delta = \varepsilon^2$ acts asymptotically as a switch between two completely different regimes, which is theoretically sharp in the limit for $T \to \infty$ and $\varepsilon \to 0$. Let us remark that the results displayed in Figure 2.2.(a) demonstrate that the transition occurs more rapidly for the smallest values of ε. Moreover, in Figure 2.2.(b), one can see how with bigger final times T the estimator is closer both to A when $\zeta \in [0, 2]$ and to α when $\zeta > 2$. Still, we observe that in finite computations the switch between A and α is smoother than what we expect from the theory, which suggests to fix, if possible, $\delta = 1$.

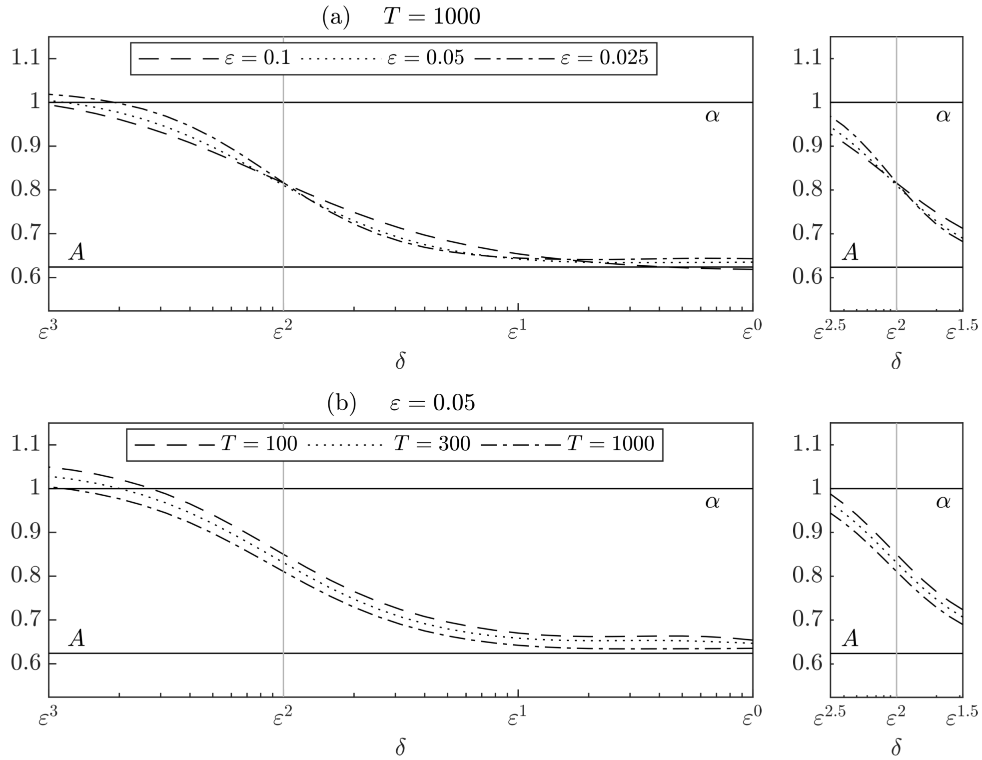


Figure 2.2 – Results for Section 2.3.1. On both figures, horizontal lines represent α and A, the drift coefficients of the unhomogenized and homogenized equations, and the grey vertical line represents the lower bound for the validity of Theorem 2.17. The curved lines (dashed, dotted and dash-dotted) represent on figure (a) the values of $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon, T)$ for $\varepsilon = \{0.1, 0.05, 0.025\}$, respectively, computed with $T = 10^3$. On figure (b), they correspond to the values of $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon, T)$ at $T = \{100, 300, 1000\}$, respectively, computed with $\varepsilon = 0.05$. We plot next to both figures (a) and (b) a zoom on a neighbourhood of ε^2 to show the transition between the two regimes highlighted by the theoretical results. Note that the δ-axis is in logarithmic scale and is normalized with respect to ε.

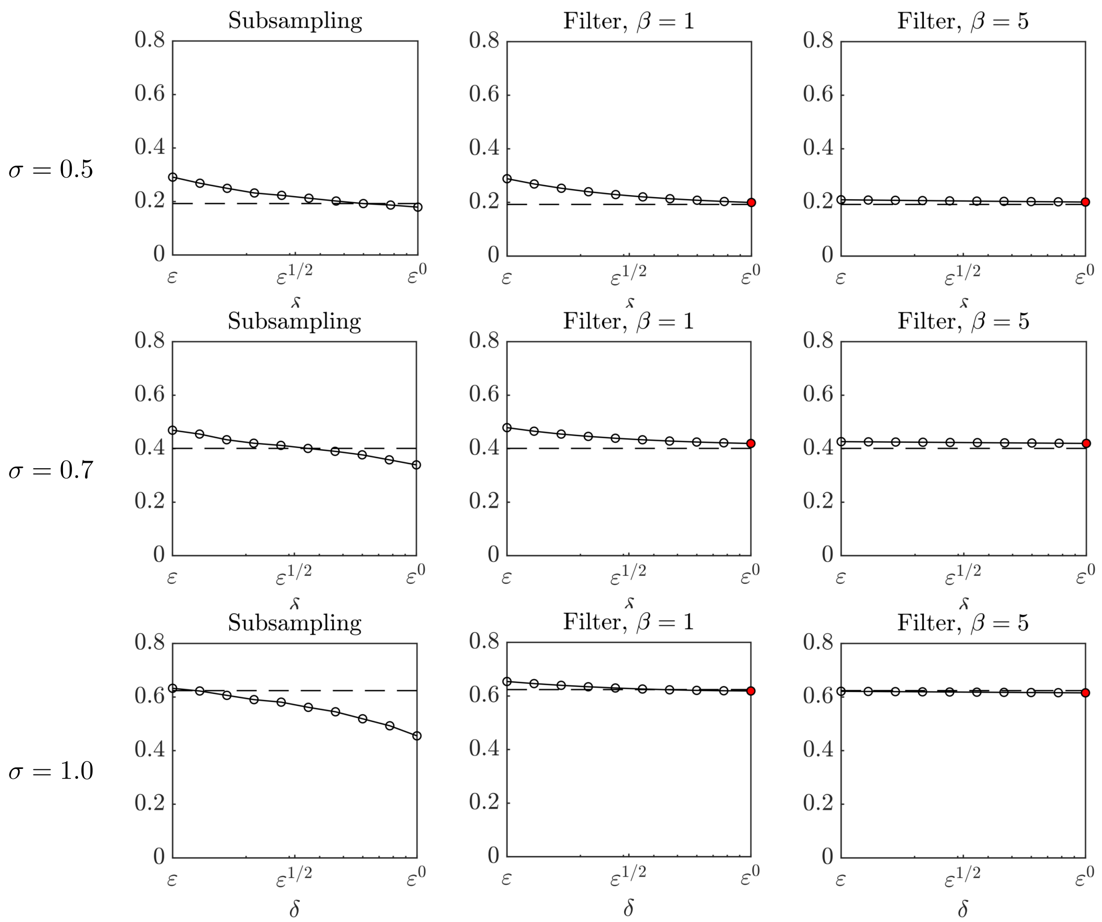


Figure 2.3 – Results for Section 2.3.1. The case of $\delta = 1$ is highlighted as a solid dot for the filtered data technique, as the analysis and theoretical result is different in this case. The three rows correspond to $\sigma = 0.5, 0.7, 1.0$ from top to bottom, and the dashed line corresponds to the true value of A.

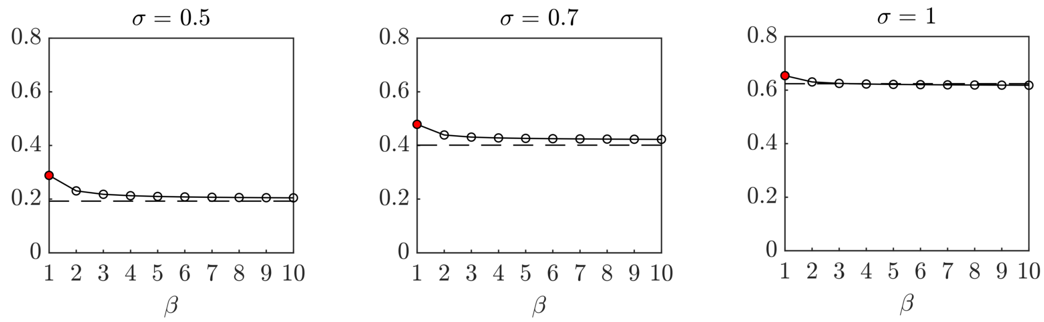


Figure 2.4 – Results for the estimator based on filter data with respect to the parameter β (Section 2.3.1). The result for $\beta = 1$, for which there are theoretical guarantees given by Theorem 2.17, is highlighted as a solid dot. From left to right we consider different values of σ, and the dashed line corresponds to the true value of A.

Comparison with subsampling

We now compare the results given by the filtered data technique with the results given by subsampling the data, i.e., the difference between the estimators $\widehat{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon, T)$ and $\widehat{A}^{\delta}_{\text{sub}}(X^\varepsilon, T)$. We fix the multiscale parameter $\varepsilon = 0.1$ and generate data for $0 \leq t \leq T$ with $T = 10^3$. We choose $\delta = \varepsilon^\zeta$ and vary $\zeta \in [0, 1]$, where δ is the filtering and the subsampling width, respectively. Moreover, for the filtered data approach we consider both $\beta = 1$ and $\beta = 5$. We report in Figure 2.3 the experimental results. Let us remark that:

(i) for $\sigma = 0.5$ the results given by subsampling and by the filter with $\beta = 1$ are similar, while for higher values of σ the filtered data approach seems better than subsampling;

(ii) in general, choosing a higher value of β seems beneficial for the quality of the estimator;

(iii) the dependence on δ of numerical results given by the filter seems relevant only in case $\beta = 1$ and for small values of σ. For $\beta = 1$ and higher values of σ, the estimator is stable with respect to this parameter. This can be observed for a higher value of β but we have no theoretical guarantee in this case.

The influence of β

We finally test the variability of the estimator with respect to β in (2.1). We consider $\delta = \varepsilon$, which corresponds to $\zeta = 1$ and seems to be the worst-case scenario for the filter, at least for $\beta = 1$. We consider again $\sigma = 0.5, 0.7, 1$ and vary $\beta = 1, 2, \dots, 10$. Results, given in Figure 2.4, show empirically that the estimator stabilizes fast with respect to β. Nevertheless, there is no theoretical guarantee supporting this empirical observation.

2.3.2 Variance of the estimators

We now compare the estimators $\widehat{A}^{\delta,\beta}_{\text{exp}}$ based on filtered data and $\widehat{A}^{\delta}_{\text{sub}}$ based on subsampling in terms of variance. We consider for this experiment the SDE (1.10) with $L = 1$, the bistable potential $V(x) = x^4/4 - x^2/2$, the multiscale drift coefficient $\alpha = 1$, the diffusion coefficient $\sigma = 1$ and with $\varepsilon = 0.1$. We then let $X^\varepsilon = (X_t, 0 \leq t \leq T)$ be the solution of (1.10) and generate $N_\text{s} = 500$ i.i.d. samples of X^ε. We then compute the estimators $\widehat{A}^{\delta,\beta}_{\text{exp}}$ and $\widehat{A}^{\delta}_{\text{sub}}$ on each of the realizations of X^ε, thus obtaining N_s replicas $\{\widehat{A}^{\delta,\beta,(i)}_{\text{exp}}\}_{i=1}^{N_\text{s}}$ and $\{\widehat{A}^{\delta,(i)}_{\text{sub}}\}_{i=1}^{N_\text{s}}$. For the estimator

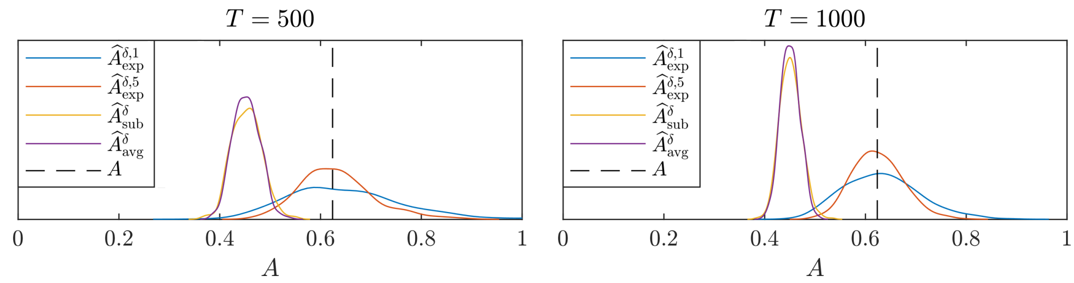


Figure 2.5 – Numerical results for Section 2.3.2. Comparison between the density of the estimator of the drift based on filtered data with $\beta = \{1, 5\}$, the estimator based on subsampling and the estimator based on shift-subsampling and averaging of (2.29). On the left and on the right, the final time is $T = \{500, 1000\}$, respectively.

$\widehat{A}^{\delta,\beta}_{\mathrm{exp}}$, we consider the kernel (2.1) with $\beta = \{1, 5\}$ and with $\delta = 1$. For the estimator $\widehat{A}^{\delta}_{\mathrm{sub}}$, we employ the subsampling width $\delta = \varepsilon^{2/3}$, which is heuristically optimal following [103]. It could be argued that another estimator based on subsampling and shifting could be employed to reduce the variance. In particular, we let $\tau > 0$ be the time step at which the data are observed. Indeed, in practice we work with high-frequency discrete data, and observe $X^\varepsilon := (X^\varepsilon_0, X^\varepsilon_\tau, \dots, X^\varepsilon_{n\tau})$, with $n\tau = T$. We assume for simplicity that the subsampling width δ is a multiple of τ and compute for all $k = 0, 1, \dots, \delta/\tau - 1$

$$\widehat{A}^{\delta,k}_{\mathrm{sub}}(X^\varepsilon, T) = -\frac{\sum_{j=0}^{n-1} V'(X^\varepsilon_{j\delta+k})(X^\varepsilon_{(j+1)\delta+k} - X^\varepsilon_{j\delta+k})}{\delta \sum_{j=0}^{n-1} V'(X^\varepsilon_{j\delta+k})^2}, \tag{2.28}$$

i.e. the subsampling estimator obtained by shifting the origin by $k\tau$. We then average over the index k and obtain the new estimator

$$\widehat{A}^{\delta}_{\mathrm{avg}}(X^\varepsilon, T) = \frac{\tau}{\delta} \sum_{k=0}^{\delta/\tau - 1} \widehat{A}^{\delta,k}_{\mathrm{sub}}(X^\varepsilon, T). \tag{2.29}$$

We include this estimator in the numerical study for completeness, and compute N_{s} replicas of $\widehat{A}^{\delta}_{\mathrm{avg}}$ on all the realizations of X^ε. Results, given in Figure 2.5 for the final times $T = \{500, 1000\}$, show that our novel approach does not outperform subsampling in terms of variance, but clearly does in terms of bias. Moreover, we notice numerically that the shifted-averaged estimator $\widehat{A}^{\delta}_{\mathrm{avg}}$ does not reduce sensibly the variance in this case with respect to $\widehat{A}^{\delta}_{\mathrm{sub}}$. In fact, this is only partly surprising, since the estimators $\widehat{A}^{\delta,k}_{\mathrm{sub}}$ of (2.28) are highly correlated. Finally, we notice that the filtering estimator $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}$ with $\beta = 5$ has a lower variance with respect to the same estimator with $\beta = 1$. This confirms that choosing a higher value of β improves the estimation of the effective drift coefficient.

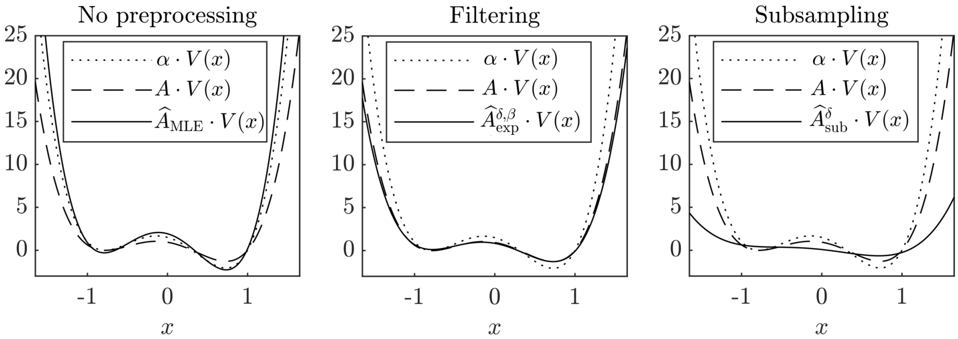
2.3.3 Multidimensional drift coefficient

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

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{i+1}(x) = 2xT_i(x) - T_{i-1}(x).$$

We consider the potential function $V(x)$ with

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



Coefficient	Multiscale α	Homogenized A	No preprocessing $\widehat{A}_{\mathrm{MLE}}$	Filtering $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}$	Subsampling $\widehat{A}^{\delta}_{\mathrm{sub}}$
1	-1	-0.62	-0.92	-0.70	-0.59
2	-0.5	-0.31	-0.70	-0.27	0.05
3	0.5	0.31	0.55	0.31	0.14
4	1	0.62	1.22	0.57	0.13

Figure 2.6 – Results for Section 2.3.3. In the figure, from left to right the potential function estimated with the data itself, the filter, subsampled data. In the table, numerical results for the single components of the true and estimated drift coefficients.

thus considering the semi-parametric framework of Remark 1.3. This potential function satisfies Assumption 1.4 whenever L is even and if the leading coefficient α_L is positive. We set $L = 4$ and the drift coefficient $\alpha = (-1, -1/2, 1/2, 1)$. With this drift coefficient, the potential function is of the bistable kind. Moreover, we set $\varepsilon = 0.05$, the diffusion coefficient $\sigma = 1$, the fast potential $p(y) = \cos(y)$ and simulate a trajectory of X^ε for $0 \le t \le T$ with $T = 10^3$ employing the EM method with time step $\Delta t = \varepsilon^3$. We estimate the drift coefficient $A \in \mathbb{R}^4$ with the estimators:

(i) $\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)$ based on the data X^ε itself;

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

(iii) $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ based on filtered data Z^ε computed with $\beta = 1$ and $\delta = 1$.

In particular, we pick this specific value of δ for the subsampling following the optimality criterion given in [103]. Results, given in Figure 2.6, show that the filter-based estimation captures well the homogenized potential as well as the coefficient A. Moreover, it is possible to remark the negative result given by Theorem 1.8 holds in practice, i.e., with no pre-processing the estimator $\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)$ tends to the drift coefficient α of the unhomogenized equation. Finally, we can observe that the subsampling-based estimator fails to capture the homogenized coefficients. Indeed, the estimator strongly depends on the sampling rate and on the diffusion coefficient, as shown in the numerical experiments of [103]. Even though the authors suggest the choice of $\delta = \varepsilon^{2/3}$, this is just an heuristic and is not guaranteed to be the optimal value in all cases. In the asymptotic limit of $\varepsilon \to 0$ and $T \to \infty$, any valid choice of the subsampling rate is guaranteed theoretically to work, but not in the pre-asymptotic regime. Our estimator, conversely, seems to perform better with no particular tuning of the parameters even in this multi-dimensional case, which demonstrates the robustness of our novel approach.

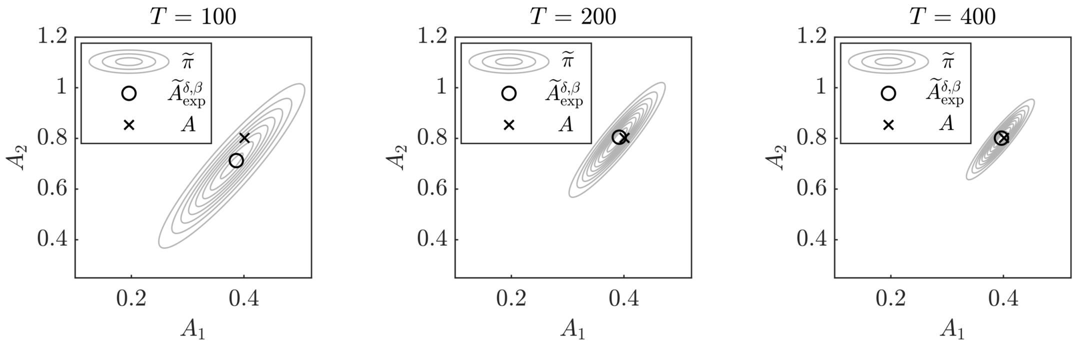


Figure 2.7 – Results for Section 2.3.4. Posterior distributions over the parameter $A = (A_1, A_2)^\top$ for the bistable potential obtained with the filtered data approach. The figures refer to final time $T = 100, 200, 400$ from left to right, respectively. The estimator $\widetilde{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon, t)$ is represented with a circle, while the true value A of the drift coefficient of the homogenized equation is represented with a cross.

2.3.4 The Bayesian approach: bistable potential

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

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

with coefficients $\alpha_1 = 1$ and $\alpha_2 = 2$. We then consider the multiscale equation with $\sigma = 0.7$, the fast potential $p(y) = \cos(y)$ and $\varepsilon = 0.05$, thus simulating a trajectory X^ε. We adopt here a Bayesian approach and compute the posterior distribution $\widetilde{\pi}_{T,\varepsilon}$ obtained with the filtered data approach introduced in Section 2.2.1. The parameters of the filter are set to $\beta = 1$ and $\delta = \varepsilon$ in (2.1). Moreover, we choose the non-informative prior $\pi_0 = \mathcal{N}(0, I)$. Let us remark that in order to compute the posterior covariance the diffusion coefficient Σ of the homogenized equation has to be known. In this case, we pre-compute the value of Σ via the coefficient $\mathcal{K}$ and the theory of homogenization, but notice that Σ could be estimated either employing the subsampling technique of [103] or using the estimator $\widehat{\Sigma}^{\delta,1}_{\text{exp}}$ based on filtered data defined in (2.4). In particular, in this case $\Sigma \approx 0.2807$, and we compute numerically

$$\widehat{\Sigma}^{\delta,1}_{\text{exp}}(X^\varepsilon, 100) = 0.2901, \quad \widehat{\Sigma}^{\delta,1}_{\text{exp}}(X^\varepsilon, 200) = 0.2835, \quad \widehat{\Sigma}^{\delta,1}_{\text{exp}}(X^\varepsilon, 400) = 0.2813,$$

so that employing the estimator $\widehat{\Sigma}^{\delta,1}_{\text{exp}}$ instead of the true value would have negligible effects on the computation of the posterior over the effective drift coefficient. We stop computations at times $T = \{100, 200, 400\}$ in order to observe the shrinkage of the Gaussian posterior towards the estimator $\widetilde{A}^{\delta,\beta}_{\text{exp}}(X^\varepsilon, T)$ with respect to time. In Figure 2.7, we observe that the posterior does indeed shrink towards this estimator, which in turn gets progressively closer to the true value of the drift coefficient A of the homogenized equation.

2.4 Technical results

2.4.1 Proofs of Section 2.1.1

Proof of Lemma 2.2. We have to show that the joint process solution to (2.6) is hypo-elliptic. Denoting as $f\colon \mathbb{R} \to \mathbb{R}$ the function

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

the generator of the process $(X^\varepsilon, Z^\varepsilon)^\top$ is given by

$$\mathcal{L} = f\partial_x + \sigma\partial_{xx}^2 + \frac{1}{\delta}(x-z)\partial_z =: \mathcal{X}_0 + \sigma\mathcal{X}_1^2,$$

where

$$\mathcal{X}_0 = f\partial_x + \frac{1}{\delta}(x-z)\partial_z, \quad \mathcal{X}_1 = \partial_x.$$

The commutator $[\mathcal{X}_0, \mathcal{X}_1]$ applied to a test function v then gives

$$\begin{aligned}[\mathcal{X}_0, \mathcal{X}_1]v &= f\partial_x^2 v + \frac{1}{\delta}(x-z)\partial_x\partial_z v - \partial_x\left(f\partial_x v + \frac{1}{\delta}(x-z)\partial_z v\right) \\ &= -\partial_x f\partial_x v - \frac{1}{\delta}\partial_z v.\end{aligned}$$

Consequently,

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

which spans the tangent space of $\mathbb{R}^2$ at (x,z), denoted $T_{x,z}\mathbb{R}^2$. The desired result then follows from Hörmander's theorem (see e.g. [101, Chapter 6]). □

Proof of Lemma 2.3. Lemma 2.2 guarantees that the FPE can be written directly from the system (2.6). For geometric ergodicity, let

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

Due to Assumption 1.4(ii), Remark 1.6 and Young's inequality, we then have for all $\gamma > 0$

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

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

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

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

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

and problem (2.6) is dissipative. It remains to prove the irreducibility condition [85, Condition 4.3]. We remark that the system (2.6) fits the framework of the example the end of [85, Page 199], and therefore [85, Condition 4.3] is satisfied. The result then follows from [85, Theorem 4.4]. □

Proof of Lemma 2.5. Integrating equation (2.7) with respect to z we obtain the stationary FPE for the process X^ε, i.e.

$$\sigma(\varphi^\varepsilon)''(x) + \frac{\mathrm{d}}{\mathrm{d}x}\left(\left(\alpha\cdot V'(x) + \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)\varphi^\varepsilon(x)\right) = 0, \tag{2.30}$$

whose solution is given by

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

and which proves (2.9). In view of (2.8) and (2.30), equation (2.7) can be rewritten as

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

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

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

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

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

Finally, choosing

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

we obtain the desired result. □

2.4.2 Proof of Proposition 2.14

Preliminary estimates

In order to prove the characterization provided by Proposition 2.14, we need to prove two additional results on the filter. First, we prove a Jensen-like inequality for the kernel of the filter.

Lemma 2.26. *Let $\delta > 0$ and $k^{\delta,1}_{\mathrm{exp}}$ be defined in (2.5). Then, for any $t > 0$, $p \geq 1$ and any function $g \in \mathcal{C}^0([0,t])$ it holds*

$$\left|\int_0^t k^{\delta,1}_{\mathrm{exp}}(t-s)g(s)\,\mathrm{d}s\right|^p \leq \int_0^t k^{\delta,1}_{\mathrm{exp}}(t-s)\left|g(s)\right|^p\,\mathrm{d}s.$$

Proof. Let us first note that

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

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

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

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

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

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

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

Lemma 2.27. *With the notation of Lemma 2.26, it holds for all $p \geq 0$*

$$\int_0^t k_{\exp}^{\delta,1}(t-s)(t-s)^p \,\mathrm{d}s \leq C\delta^p,$$

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

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

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

where γ is the lower incomplete Gamma function and Γ is the complete Gamma function. $\square$

Proof of Proposition 2.14

Denoting $Y_t^\varepsilon \coloneqq X_t^\varepsilon/\varepsilon$, we will make use of the decomposition [103, Formula 5.8]

$$\begin{aligned} X_t^\varepsilon - X_s^\varepsilon = &- \int_s^t (\alpha \cdot V'(X_r^\varepsilon))(1 + \Phi'(Y_r^\varepsilon)) \,\mathrm{d}r \\ &+ \sqrt{2\sigma} \int_s^t (1 + \Phi'(Y_r^\varepsilon)) \,\mathrm{d}W_r - \varepsilon(\Phi(Y_t^\varepsilon) - \Phi(Y_s^\varepsilon)), \end{aligned} \tag{2.31}$$

which is obtained applying the Itô formula to Φ, the solution of the cell problem (1.7). Recall that by definition of Z_t^ε we have

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

Plugging the decomposition (2.31) into the equation above, we obtain

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

where

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

Let us analyze the terms above singularly. For $I_1^\varepsilon(t)$, one can show [103, Proposition 5.8]

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

where the remainder R_1^ε satisfies

$$\left(\mathbb{E}^{\nu^\varepsilon} |R_1^\varepsilon(t-s)|^p\right)^{1/p} \leq C(\varepsilon^2 + \varepsilon(t-s)^{1/2} + (t-s)^{3/2}). \tag{2.32}$$

Therefore, it holds

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

where we exploited the equality

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

and where

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

Now, Lemma 2.26, the inequality (2.32) and Lemma 2.27 yield for all $p \geq 1$

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

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

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

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

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

and therefore rewrite

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

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

$$\begin{aligned}
I_2^\varepsilon(t) &= \sqrt{2\sigma}\left(Q_t^\varepsilon\int_0^t k_{\exp}^{\delta,\beta}(t-s)\,\mathrm{d}s - Q_t^\varepsilon + \delta\int_0^t k_{\exp}^{\delta,\beta}(t-s)\left(1+\Phi'(Y_s^\varepsilon)\right)\,\mathrm{d}W_s\right) \\
&= \delta B_t^\varepsilon - \sqrt{2\sigma}e^{-t/\delta}Q_t^\varepsilon \eqqcolon \delta B_t^\varepsilon - R_2^\varepsilon(t).
\end{aligned} \tag{2.33}$$

where B_t^ε is defined in (2.21). For the remainder $R_2^\varepsilon(t)$, let us remark that for all $p \geq 1$ it holds

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

where we applied Jensen's inequality, an estimate for the moments of stochastic integrals [69, Formula (3.25), p. 163] and the boundedness of Φ. Therefore we have

$$\left(\mathbb{E}^{\nu^\varepsilon}\left|R_2^\varepsilon(t)\right|^p\right)^{1/p} \leq C\sqrt{t}e^{-t/\delta}. \tag{2.34}$$

In order to obtain the bound (2.22) on B_t^ε, let us remark that from (2.33) it holds for a constant $C > 0$ depending only on p

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

The second term is bounded exponentially fast with respect to t and δ due to (2.34). For the first term, applying Lemma 2.26, the inequality [69, Formula (3.25), p. 163] and Lemma 2.27 we obtain for a constant $C > 0$ independent of δ and t

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

Therefore, it holds for δ sufficiently small

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

which proves the bound (2.22). Let us now consider $I_3^\varepsilon(t)$. Since Φ is bounded, we simply have

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

almost surely. Finally, due to [103, Corollary 5.4], we know that X_t^ε has bounded moments of all orders and therefore

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

which concludes the proof. □

2.4.3 Proofs of Section 2.1.3

Preliminary estimates

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

Lemma 2.28. *Under Assumption 1.4, let Z^ε be distributed as the marginal of the invariant measure $\mu_{\mathrm{exp}}^\varepsilon$ of the couple $(X^\varepsilon, Z^\varepsilon)^\top$. Then for any $p \ge 1$ there exists a constant $C > 0$ uniform in ε such that*

$$\mathbb{E}^{\mu_{\mathrm{exp}}^\varepsilon}\,|Z^\varepsilon|^p \le C.$$

Proof. Let X_t^ε be at stationarity with respect to its invariant measure ν^ε. Let Z_t^ε be the corresponding filtered process. By definition of Z_t^ε and applying Lemma 2.26 we have

$$\begin{aligned}\mathbb{E}^{\nu^\varepsilon}\,|Z_t^\varepsilon|^p &= \mathbb{E}^{\nu^\varepsilon}\left|\int_0^t k_{\mathrm{exp}}^{\delta,\beta}(t-s)X_s^\varepsilon\,\mathrm{d}s\right|^p \\ &\le \int_0^t k_{\mathrm{exp}}^{\delta,\beta}(t-s)\,\mathbb{E}^{\nu^\varepsilon}\,|X_s^\varepsilon|^p\,\mathrm{d}s,\end{aligned}$$

which, together with the definition of $k_{\mathrm{exp}}^{\delta,\beta}$ and the fact that X_s^ε has bounded moments of all orders [103, Corollary 5.4], implies for a constant $C > 0$

$$\mathbb{E}^{\nu^\varepsilon}\,|Z_t^\varepsilon|^p \le C.$$

In order to conclude, we remark that due to Lemma 2.3 we have for all $t \ge 0$

$$\mathbb{E}^{\mu_{\mathrm{exp}}^\varepsilon}\,|Z^\varepsilon|^p \le \mathbb{E}^{\nu^\varepsilon}\,|Z_t^\varepsilon|^p + Ce^{-\lambda t},$$

which, for t sufficiently big, yields the desired result. □

Corollary 2.29 is a direct consequence of Proposition 2.14 and provides a rough estimate of the difference between the trajectories X_t^ε and Z_t^ε when they are at stationarity.

Corollary 2.29. *Under Assumption 1.4, let the couple $(X^\varepsilon, Z^\varepsilon)^\top$ be distributed as its invariant measure $\mu_{\mathrm{exp}}^\varepsilon$. Then, if $\delta \le 1$, it holds for any $p \ge 1$*

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

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

Proof. Let $p \geq 1$, then due to Proposition 2.14 there exists a constant $C > 0$ depending only on p such that

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

Let us now remark that this result holds for X_t^ε being at stationarity and for Z_t^ε being its filtered process, and not for a couple $(X^\varepsilon, Z^\varepsilon)^\top \sim \mu_{\text{exp}}^\varepsilon$. In order to conclude, we remark that due to Lemma 2.3 we have for all $t \geq 0$

$$\mathbb{E}^{\mu_{\text{exp}}^\varepsilon} |X^\varepsilon - Z^\varepsilon|^p \leq \mathbb{E}^{\nu^\varepsilon} |X_t^\varepsilon - Z_t^\varepsilon|^p + Ce^{-\lambda t},$$

which, for t sufficiently big, yields the desired result. □

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

Lemma 2.30. *Under Assumption 1.4, let the couple $(X^\varepsilon, Z^\varepsilon)^\top$ be distributed as its invariant measure $\mu_{\text{exp}}^\varepsilon$. Then, if $\delta \leq 1$, it holds for any $p \geq 1$*

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

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

Proof. By equation (1.10) we have for all $0 \leq s < t$

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

Therefore, by Assumption 1.4 and since X_t^ε has bounded moments of all orders at stationarity [103, Corollary 5.4], it holds for any $p \geq 1$ and a constant $C > 0$

$$\mathbb{E}^{\nu^\varepsilon} |X_t^\varepsilon - X_s^\varepsilon|^p \leq C \left((t-s)^p + (t-s)^p \varepsilon^{-p} + (t-s)^{p/2} \right), \tag{2.35}$$

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

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

which, applying Lemma 2.26, the inequality (2.35) and Lemma 2.27, implies

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

Geometric ergodicity (Lemma 2.3) then implies for $\mu_{\text{exp}}^\varepsilon$ the measure of the couple $(X^\varepsilon, Z^\varepsilon)^\top$

$$\mathbb{E}^{\mu_{\text{exp}}^\varepsilon} |X^\varepsilon - Z^\varepsilon|^p \leq \mathbb{E}^{\nu^\varepsilon} |X_t^\varepsilon - Z_t^\varepsilon|^p + Ce^{-\lambda t},$$

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

Let us conclude with a last preliminary estimate concerning the matrices $\widetilde{\mathcal{M}}_{\text{exp}}^\varepsilon$ and $\mathcal{M}^\varepsilon$ defined in (2.14) and (2.15), respectively.

Lemma 2.31. *Let the assumptions of Corollary 2.29 hold. Then the matrices $\mathcal{M}^\varepsilon$ and $\widetilde{\mathcal{M}}^\varepsilon_{\exp}$ satisfy*

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

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

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

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

The Lipschitz condition on V' together with the boundedness of the moments of X^ε and Corollary 2.29 yield for a constant $C > 0$

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

which is the desired result. □

Proof of Lemma 2.15

Let us consider the following system of SDEs for the processes $X^\varepsilon_t, Z^\varepsilon_t, B^\varepsilon_t, Y^\varepsilon_t$

$$\begin{aligned}\mathrm{d}X^\varepsilon_t &= -\alpha \cdot V'(X^\varepsilon_t)\,\mathrm{d}t - \frac{1}{\varepsilon} p'(Y^\varepsilon_t)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\ \mathrm{d}Z^\varepsilon_t &= \frac{1}{\delta}\left(X^\varepsilon_t - Z^\varepsilon_t \right)\mathrm{d}t, \\ \mathrm{d}B^\varepsilon_t &= -\frac{1}{\delta} B^\varepsilon_t\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\delta}(1 + \Phi'(Y^\varepsilon_t))\,\mathrm{d}W_t, \\ \mathrm{d}Y^\varepsilon_t &= -\frac{1}{\varepsilon}\alpha \cdot V'(X^\varepsilon_t)\,\mathrm{d}t - \frac{1}{\varepsilon^2} p'(Y^\varepsilon_t)\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\varepsilon}\,\mathrm{d}W_t,\end{aligned}$$

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

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

Let us denote by $e^\varepsilon \colon \mathbb{R}^3 \times [0, \mathbb{T}] \to \mathbb{R}$, $e^\varepsilon = e^\varepsilon(x, z, b, y)$, the density of the invariant measure η^ε of the quadruple $(X^\varepsilon_t, Z^\varepsilon_t, B^\varepsilon_t, Y^\varepsilon_t)$. Then e^ε solves the stationary FPE $\widetilde{\mathcal{L}}^*_\varepsilon e^\varepsilon = 0$, i.e., explicitly

$$\begin{aligned}&\partial_x \left(\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p'(y) \right) e^\varepsilon \right) + \frac{1}{\delta}\partial_z \left((z - x) e^\varepsilon \right) \\ &+ \frac{1}{\delta}\partial_b (b e^\varepsilon) + \partial_y \left(\left(\frac{1}{\varepsilon}\alpha \cdot V'(x) + \frac{1}{\varepsilon^2} p'(y) \right) e^\varepsilon \right) + \sigma \left(\partial^2_{xx} e^\varepsilon + \frac{2}{\varepsilon}\partial^2_{xy} e^\varepsilon + \frac{1}{\varepsilon^2}\partial^2_{yy} e^\varepsilon \right) \\ &+ \sigma \left(\frac{2}{\delta}\partial^2_{xb} \left((1 + \Phi'(y)) e^\varepsilon \right) + \frac{2}{\varepsilon\delta}\partial^2_{yb} \left((1 + \Phi'(y)) e^\varepsilon \right) + \frac{1}{\delta^2}\partial^2_{bb} \left((1 + \Phi'(y))^2 e^\varepsilon \right) \right) = 0.\end{aligned} \tag{2.36}$$

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

$$\frac{1}{\delta}\int_{\mathbb{R}^3 \times [0,\mathbb{T}]} \partial_z f(z, b)(x - z) e^\varepsilon - \frac{1}{\delta}\int_{\mathbb{R}^3 \times [0,\mathbb{T}]} \partial_b f(z, b) b e^\varepsilon + \frac{\sigma}{\delta^2}\int_{\mathbb{R}^3 \times [0,\mathbb{T}]} \partial^2_{bb} f(z, b)(1 + \Phi'(y))^2 e^\varepsilon,$$

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

$$\delta\,\mathbb{E}^{\eta^\varepsilon}\left[\partial_b f(Z^\varepsilon,B^\varepsilon)B^\varepsilon\right] = \sigma\,\mathbb{E}^{\eta^\varepsilon}\left[\partial^2_{bb} f(Z^\varepsilon,B^\varepsilon)(1+\Phi'(Y^\varepsilon))\right] + \delta\,\mathbb{E}^{\eta^\varepsilon}\left[\partial_z f(Z^\varepsilon,B^\varepsilon)(X^\varepsilon - Z^\varepsilon)\right]. \tag{2.37}$$

Choosing

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

we obtain

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

We now consider the remainder and, applying Hölder's inequality, Corollary 2.29, Lemma 2.28, Assumption 1.4 and (2.22), we get for p,q,r such that $1/p+1/q+1/r=1$

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

which completes the proof. □

Proof of Lemma 2.16

Let us introduce the notation

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

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

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

We first study $\Delta_1(\varepsilon)$ and due to the boundedness of Φ', Assumption 1.4 and Lemma 2.29 we have

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

which implies

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

We now consider $\Delta_2(\varepsilon)$. Integrating equation (2.36) with respect to z and b we obtain the FPE for the stationary marginal distribution $\lambda\colon \mathbb{R}\times[0,\mathbb{T}]$, $\lambda=\lambda(x,y)$, of the couple $(X^\varepsilon,Y^\varepsilon)^\top$

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

whose solution is given by

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

where

$$\begin{aligned} C_\lambda &= \int_{\mathbb{R}} \int_0^{\mathbb{T}} \exp\left(-\frac{\alpha}{\sigma} V(x) - \frac{1}{\sigma} p(y)\right) \mathrm{d}x \, \mathrm{d}y \\ &= \left(\int_{\mathbb{R}} \exp\left(-\frac{\alpha}{\sigma} V(x)\right) \mathrm{d}x\right) \left(\int_0^{\mathbb{T}} \exp\left(-\frac{1}{\sigma} p(y)\right) \mathrm{d}y\right) \\ &= C_{\nu^0} C_\pi. \end{aligned}$$

Therefore, since $\Sigma = \mathcal{K}\sigma$ and by equations (1.6) and (2.13) we have

$$\begin{aligned} \sigma \, \mathbb{E}^{\eta^\varepsilon}[V''(X^\varepsilon)(1 + \Phi'(Y^\varepsilon))^2] &= \sigma \int_{\mathbb{R}} \int_0^{\mathbb{T}} V''(x)(1 + \Phi'(y))^2 \frac{1}{C_\lambda} \exp\left(-\frac{\alpha}{\sigma} V(x) - \frac{1}{\sigma} p(y)\right) \mathrm{d}x \, \mathrm{d}y \\ &= \sigma \left(\int_{\mathbb{R}} V''(x) \frac{1}{C_{\nu^0}} \exp\left(-\frac{\alpha}{\sigma} V(x)\right) \mathrm{d}x\right) \\ &\qquad \times \left(\int_0^{\mathbb{T}} (1 + \Phi'(y))^2 \frac{1}{C_\pi} \exp\left(-\frac{1}{\sigma} p(y)\right) \mathrm{d}y\right) \\ &= \sigma \mathcal{K} \, \mathbb{E}^{\nu^0}[V''(X^0)] = \Sigma \, \mathbb{E}^{\nu^0}[V''(X^0)], \end{aligned}$$

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

Proof of Theorem 2.18

Let us consider the decomposition (2.16), i.e.,

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

where $I_1^\varepsilon(T)$ is defined in (2.16) and satisfies

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

and, by the proof of Theorem 2.12 we have independently of ε

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

A Taylor expansion of the first order of V' yields

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

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

$$\begin{aligned} \lim_{T \to \infty} I_1^\varepsilon(T) &= (\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1} \left(\mathbb{E}^{\nu^\varepsilon} \left[\frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon}{\varepsilon}\right) V'(X^\varepsilon)\right] + \mathbb{E}^{\mu^\varepsilon_{\exp}} \left[\frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon}{\varepsilon}\right) V''(\widetilde{X}^\varepsilon)(Z^\varepsilon - X^\varepsilon)\right]\right) \\ &=: (\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1} \left(J_1^\varepsilon + J_2^\varepsilon\right). \end{aligned}$$

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

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

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

$$\lim_{\varepsilon\to 0} J_1^\varepsilon = \sigma\,\mathbb{E}^{\nu^0}\left[V''(X^0)\right] - \mathbb{E}^{\nu^0}\left[V'(X^0)\otimes V'(X^0)\right]\alpha = 0. \tag{2.38}$$

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

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

Since $\widetilde{X}^\varepsilon$ assumes values between X^ε and Z^ε, it has bounded moments by [103, Corollary 5.4] and Lemma 2.28. Hence, applying Lemma 2.30 we have

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

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

$$\lim_{\varepsilon\to 0}\|J_2^\varepsilon\| = 0. \tag{2.39}$$

Finally, Lemma 2.31 and the weak convergence of the invariant measure ν^ε to ν^0 imply

$$\lim_{\varepsilon\to 0}\widetilde{\mathcal{M}}^\varepsilon_{\exp} = \mathcal{M}^0,$$

which, together with (2.38), (2.39) implies that $I_1^\varepsilon(T) \to 0$ for $T\to\infty$ and $\varepsilon\to 0$, which implies the desired result. □

2.4.4 Proof of Theorem 2.19

First, the ergodic theorem yields

$$\lim_{T\to\infty}\widehat{\Sigma}^{\delta,1}_{\exp} = \frac{1}{\delta}\,\mathbb{E}^{\mu^\varepsilon_{\exp}}\left[(X^\varepsilon - Z^\varepsilon)^2\right],$$

then applying Proposition 2.14 at stationarity we obtain

$$\begin{aligned}\lim_{T\to\infty}\widehat{\Sigma}^{\delta,1}_{\exp} &= \delta\,\mathbb{E}^{\mu^\varepsilon_{\exp}}\left[(B^\varepsilon)^2\right] + 2\,\mathbb{E}^{\mu^\varepsilon_{\exp}}\left[B^\varepsilon R(\varepsilon,\delta)\right] + \frac{1}{\delta}\,\mathbb{E}^{\mu^\varepsilon_{\exp}}\left[R(\varepsilon,\delta)^2\right]\\ &=: I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon,\end{aligned}$$

and due to the Cauchy-Schwarz inequality and estimates (2.22) and (2.23) we have

$$|I_2^\varepsilon| \le C\left(\delta^{1/2} + \varepsilon\delta^{-1/2}\right) \qquad \text{and} \qquad |I_3^\varepsilon| \le C\left(\delta + \varepsilon^2\delta^{-1}\right), \tag{2.40}$$

for a constant $C > 0$ independent of ε and δ. Let us now consider I_1^ε. Employing equation (2.37) with the function $f(z,b) = 1/2b^2$ gives

$$\mathbb{E}^{\eta^\varepsilon}\left[(B^\varepsilon)^2\right] = \frac{\sigma}{\delta}\,\mathbb{E}^{\eta^\varepsilon}\left[1 + \Phi'(Y^\varepsilon)\right] = \frac{\sigma\mathcal{K}}{\delta} = \frac{\Sigma}{\delta},$$

which together with bounds (2.40) and the hypothesis on δ implies

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty}\widehat{\Sigma}^{\delta,1}_{\exp} = \Sigma, \quad \text{in probability},$$

which is the desired result. □

2.5 Conclusion

In this chapter we considered a novel methodology to confront the problem of model misspecification when homogenized models are fit to multiscale data. Our approach is based on using filtered data for the estimation of the drift of the homogenized diffusion process. We proved asymptotic unbiasedness of estimators drawn from our methodology. Moreover, we found a modified Bayesian approach which guarantees robust uncertainty quantification and posterior contraction, based on the same filtered data approach. Numerical experiments demonstrate how the estimator based on filtered data requires less knowledge of the characteristic time-scales of the multiscale equation with respect to subsampling, and how it can be employed as a black-box tool for parameter estimation on a range of academic examples.

3 Moving average

In this chapter, which is based on our research article [51], we use a similar approach to the one proposed in Chapter 2, but instead of employing an exponential filter, we compute filtered data applying a moving average. The chapter is organized as follows. In Section 3.1 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 3.2 corroborating our theoretical findings and showing how to apply our methodology to a complex multi-dimensional scenario. Section 3.3 is dedicated to the proof of unbiasedness for our estimators, and in Section 3.4 we draw our conclusions.

3.1 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

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

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

$$Z_{\mathrm{ma}}^{\delta,\varepsilon}(t) = \int_0^t k_{\mathrm{ma}}^{\delta}(t-s)X^{\varepsilon}(s)\,\mathrm{d}s = \frac{1}{\delta}\int_{t-\delta}^t X^{\varepsilon}(s)\,\mathrm{d}s, \qquad t \geq \delta. \tag{3.1}$$

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

$$Z_{\mathrm{ma}}^{\delta,\varepsilon}(t) = \frac{1}{t}\int_0^t X^{\varepsilon}(s)\,\mathrm{d}s, \qquad 0 \leq t < \delta.$$

We remark that with this choice the process $Z_{\mathrm{ma}}^{\delta,\varepsilon}$ 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 smoothening the data with a low-pass filter has already been introduced in Chapter 2. Instead of a moving average, we consider in Chapter 2 the exponential filtering kernel (2.1). Filtered data $Z_{\mathrm{exp}}^{\delta,\beta,\varepsilon}(t)$ are then obtained similarly to (3.1) by taking a time convolution. We recall that the parameters δ and β in (2.1) have two different roles. In particular, we have that β is a *shape parameter*, and δ is the *filtering width*. The approaches presented here in the previous chapter are closely related. In fact, it is simple to deduce that for almost every $r \geq 0$

$$\lim_{\beta\to\infty} k_{\mathrm{exp}}^{\delta,\beta}(r) = \chi_{[0,1]}(r), \tag{3.2}$$

independently of δ. We remark that the theoretical analysis in Chapter 2 is restricted to the case where the shape parameter $\beta = 1$, despite numerical experiments suggest that choosing $\beta > 1$

yields better estimators. Studying the moving average kernel and due to (3.2) therefore partially fills the theoretical gap of our previous work.

3.1.1 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 (1.15) by replacing one occurrence of the original process X^ε with the filtered process $Z_{\mathrm{ma}}^{\delta,\varepsilon}$ both in M and v. In particular, the drift estimator $\widehat{A}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ is defined as the solution of the linear system

$$-\widetilde{M}_{\mathrm{ma}}^\delta(X^\varepsilon, T)\widehat{A}_{\mathrm{ma}}^\delta(X^\varepsilon, T) = \widetilde{v}_{\mathrm{ma}}^\delta(X^\varepsilon, T), \tag{3.3}$$

where

$$\begin{aligned}\widetilde{M}_{\mathrm{ma}}^\delta(X^\varepsilon, T) &\coloneqq \frac{1}{T}\int_0^T V'(Z_{\mathrm{ma}}^{\delta,\varepsilon}(t)) \otimes V'(X^\varepsilon(t))\,\mathrm{d}t, \\ \widetilde{v}_{\mathrm{ma}}^\delta(X^\varepsilon, T) &\coloneqq \frac{1}{T}\int_0^T V'(Z_{\mathrm{ma}}^{\delta,\varepsilon}(t))\,\mathrm{d}X^\varepsilon(t).\end{aligned}$$

We remark that it is fundamental to keep $X^\varepsilon(t)$ in the differential of the right-hand side $\widetilde{v}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$, as discussed in Remark 2.7. For the well-posedness of the estimator $\widehat{A}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ the matrix $\widetilde{M}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ 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.

3.1.2 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

$$\widehat{\Sigma}_{\mathrm{ma}}^\delta(X^\varepsilon, T) = \frac{1}{\delta T}\int_\delta^T \left(X^\varepsilon(t) - Z_{\mathrm{ma}}^{\delta,\varepsilon}(t)\right)\left(X^\varepsilon(t) - X^\varepsilon(t-\delta)\right)\,\mathrm{d}t, \tag{3.4}$$

which is analogous to the diffusion estimator presented in Section 2.1.

A different methodology for estimating the diffusion coefficient of the homogenized equation can be derived from the particular form of (1.11). 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 (1.16), 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 Theorem 1.8 we know that $\widehat{\alpha} = \widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)$ approximates α, we write

$$\widehat{\alpha}\widehat{\mathcal{K}} = \widehat{A}, \tag{3.5}$$

which is an overdetermined linear system with L 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 (3.5) 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

$$\widetilde{\Sigma}(X^\varepsilon, T) = \frac{\langle X^\varepsilon\rangle_T(\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)^\top\widehat{A}(X^\varepsilon, T))}{2T(\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T)^\top\widehat{A}_{\mathrm{MLE}}(X^\varepsilon, T))}. \tag{3.6}$$

Remark 3.1. 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 $Z_{\mathrm{ma}}^{\delta,\varepsilon}$ is of order $\mathcal{O}(n)$.

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

Theorem 3.2. *Let $\widehat{A}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ be defined in* (3.3) *with δ independent of ε or $\delta = \varepsilon^\zeta$ where $\zeta \in (0, 2)$. Under Assumption 1.4 and if $\widetilde{M}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ is invertible, it holds*

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

where A is the drift coefficient of the homogenized equation (1.11)*.*

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

Theorem 3.3. *Let $\widehat{\Sigma}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ be defined in* (3.4) *with $\delta = \varepsilon^\zeta$ where $\zeta \in (0, 2)$. Under Assumption 1.4, it holds*

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

where Σ is the diffusion coefficient of the homogenized equation (1.11)*.*

We notice from the statement of Theorem 3.3 that the main limitation of $\widehat{\Sigma}_{\mathrm{ma}}^\delta$ 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 3.4. *Let $\widehat{A}$ be an asymptotically unbiased estimator of the effective drift coefficient, i.e.,*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}(X^\varepsilon, T) = A, \quad a.s., \tag{3.7}$$

where A is the drift coefficient of the homogenized equation (1.11)*, and let $\widetilde{\Sigma}(X^\varepsilon, T)$ be defined in* (3.6)*. Under Assumption 1.4, it holds*

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

where Σ is the diffusion coefficient of the homogenized equation (1.11)*.*

The following corollary is a direct consequence of Theorem 3.4.

Corollary 3.5. *Let $\widetilde{\Sigma}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$, $\widetilde{\Sigma}_{\mathrm{exp}}^{\delta,\beta}(X^\varepsilon, T)$, $\widetilde{\Sigma}_{\mathrm{sub}}^\delta(X^\varepsilon, T)$ be defined by* (3.6) *with $\widehat{A}(X^\varepsilon, T)$ replaced by $\widehat{A}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$, $\widehat{A}_{\mathrm{exp}}^{\delta,\beta}(X^\varepsilon, T)$, and $\widehat{A}_{\mathrm{sub}}^\delta(X^\varepsilon, T)$, respectively. If δ is independent of ε or $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$, then*

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

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

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widetilde\Sigma^\delta_{\mathrm{sub}} = \Sigma, \quad a.s.,$$

where Σ *is the diffusion coefficient of the homogenized equation* (1.11).

Two remarks are due.

Remark 3.6. In our approach, the scale-separation parameter ε need not be known. In particular, Theorem 3.2 and Corollary 3.5 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 3.2, 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 3.7. Despite being more accurate in practice, as demonstrated by our numerical experiments, and not requiring knowledge of ε, the estimator $\widetilde\Sigma^\delta_{\mathrm{ma}}$ is computationally more expensive to obtain than $\widehat\Sigma^\delta_{\mathrm{ma}}$. Indeed, computing $\widetilde\Sigma^\delta_{\mathrm{ma}}$ requires estimators for the parameters α and σ of the multiscale equation (1.10), as well as the drift coefficient A of the homogenized model (1.11). Hence, if a very accurate estimate for the whole effective equation is needed, we recommend to employ $\widetilde\Sigma^\delta_{\mathrm{ma}}$, while $\widehat\Sigma^\delta_{\mathrm{ma}}$ can be used in case only the diffusion coefficient is required.

The proof of Theorems 3.2 to 3.4 is obtained by applying techniques similar to the ones employed in [103], and is presented in detail in Section 3.3.

3.1.4 The multidimensional case

In this section we report the expression of the estimators for higher dimensions. Indeed, the choice $d = 1$ is made only for economy of notation, and for clarity. Moreover, the proofs of Theorems 3.2 to 3.4 would be conceptually unchanged by considering $d > 1$, up to tedious technical details. Let $\nabla V\colon \mathbb{R}^d \to \mathbb{R}^{dL}$ be the vector obtained by stacking the gradients of the components V_ℓ, $\ell = 1, \dots, L$, of the slow-scale potential, i.e.,

$$\nabla V(x) = \begin{pmatrix} \nabla V_1(x)^\top & \cdots & \nabla V_L(x)^\top \end{pmatrix}^\top .$$

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

$$A = \begin{pmatrix} A_1^\top & \cdots & A_L^\top \end{pmatrix}^\top ,$$

is the solution of the linear system

$$-\widetilde M^\delta_{\mathrm{ma}}(X^\varepsilon, T)\widehat A^\delta_{\mathrm{ma}}(X^\varepsilon, T) = \widetilde v^\delta_{\mathrm{ma}}(X^\varepsilon, T),$$

where the matrices $\widetilde M^\delta_{\mathrm{ma}}(X^\varepsilon, T) \in \mathbb{R}^{dL\times dL}$ and $\widetilde v^\delta_{\mathrm{ma}}(X^\varepsilon, T) \in \mathbb{R}^{dL\times d}$ are defined as

$$\begin{aligned}
\widetilde M^\delta_{\mathrm{ma}}(X^\varepsilon, T) &:= \frac{1}{T}\int_0^T \nabla V(Z^{\delta,\varepsilon}_{\mathrm{ma}}(t)) \otimes \nabla V(X^\varepsilon(t))\,\mathrm{d}t,\\
\widetilde v^\delta_{\mathrm{ma}}(X^\varepsilon, T) &:= \frac{1}{T}\int_0^T \nabla V(Z^{\delta,\varepsilon}_{\mathrm{ma}}(t)) \otimes \,\mathrm{d}X^\varepsilon(t).
\end{aligned}$$

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 $\widehat\Sigma^\delta_{\mathrm{ma}}(X^\varepsilon, T)$ for the effective diffusion coefficient $\Sigma \in \mathbb{R}^{d\times d}$ as

$$\widehat\Sigma^\delta_{\mathrm{ma}}(X^\varepsilon, T) = \mathcal{S}\left(\frac{1}{\delta T}\int_\delta^T \left(X^\varepsilon(t) - Z^{\delta,\varepsilon}_{\mathrm{ma}}(t)\right) \otimes \left(X^\varepsilon(t) - X^\varepsilon(t-\delta)\right)\,\mathrm{d}t\right),$$

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 $\widehat{\Sigma}^\delta_{\mathrm{ma}}(X^\varepsilon, T)$, but due to asymptotic unbiasedness it is natural to expect that for T large enough and ε small enough the estimator $\widehat{\Sigma}^\delta_{\mathrm{ma}}(X^\varepsilon, T)$ is positive definite as its limit. For the estimators of the form $\widetilde{\Sigma}(X^\varepsilon, T)$, we first estimate the homogenization matrix as

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

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 $A \in \mathbb{R}^{dL\times d}$. It is simple to compute the minimum in practice by imposing $K = BB^\top$, and then minimizing directly over $B \in \mathbb{R}^{d\times d}$. 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 3.2.3, our methodology based on moving average can be naturally and successfully applied to higher-dimensional stochastic differential equations (SDEs).

3.2 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.

3.2.1 Sensitivity analysis with a Ornstein–Uhlenbeck model

We first consider the one-dimensional equation (1.10) with $L = 1$, 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 (OU) equation. 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) $\widehat{A}^\delta_{\mathrm{ma}}$, $\widehat{A}^{\delta,1}_{\mathrm{exp}}$, and $\widehat{A}^\delta_{\mathrm{sub}}$ 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) $\widehat{\Sigma}^\delta_{\mathrm{ma}}$, $\widehat{\Sigma}^{\delta,1}_{\mathrm{exp}}$, and $\widehat{\Sigma}^\delta_{\mathrm{sub}}$ of the effective diffusion coefficient obtained with the same methods, respectively,

(iii) $\widetilde{\Sigma}^\delta_{\mathrm{ma}}$, $\widetilde{\Sigma}^{\delta,1}_{\mathrm{exp}}$, and $\widetilde{\Sigma}^\delta_{\mathrm{sub}}$ of the effective diffusion coefficient based on the corresponding drift estimators given in (i).

We consider $T = 10^4$ and generate data $X^\varepsilon = (X^\varepsilon(t), 0 \le t \le T)$ 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 (EM) method with fixed time step $\Delta_t = \varepsilon^3_{\min}$, where $\varepsilon_{\min} = 0.05$ 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

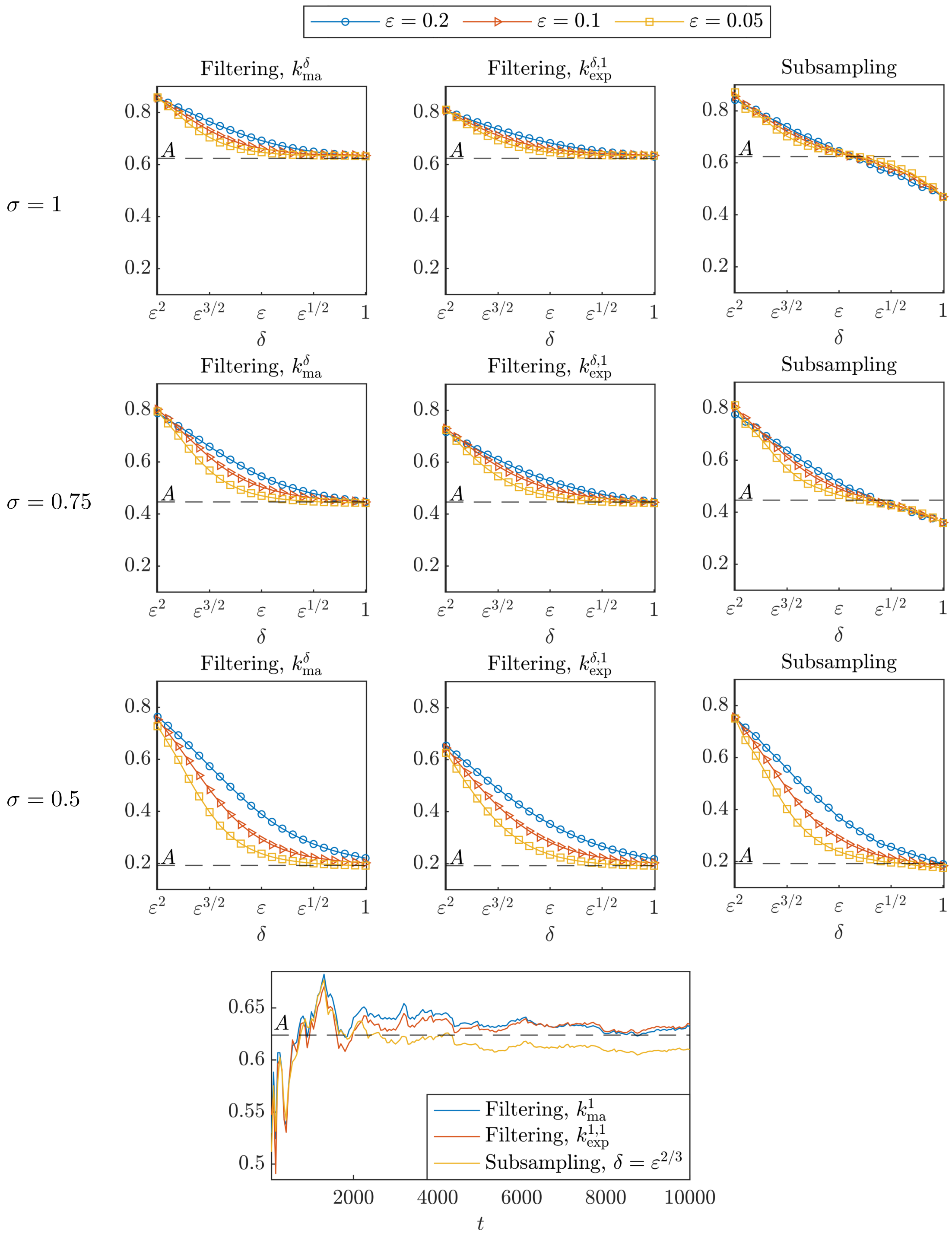


Figure 3.1 – Estimation of the drift coefficient of an effective OU equation. Top: Numerical results at final time $T = 10^4$ for variable $\sigma = 1, 0.75, 0.5$ (per row), for $\varepsilon = 0.2, 0.1, 0.05$ (blue, red, and yellow lines respectively), and for variable filtering/subsampling width δ (horizontal 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.

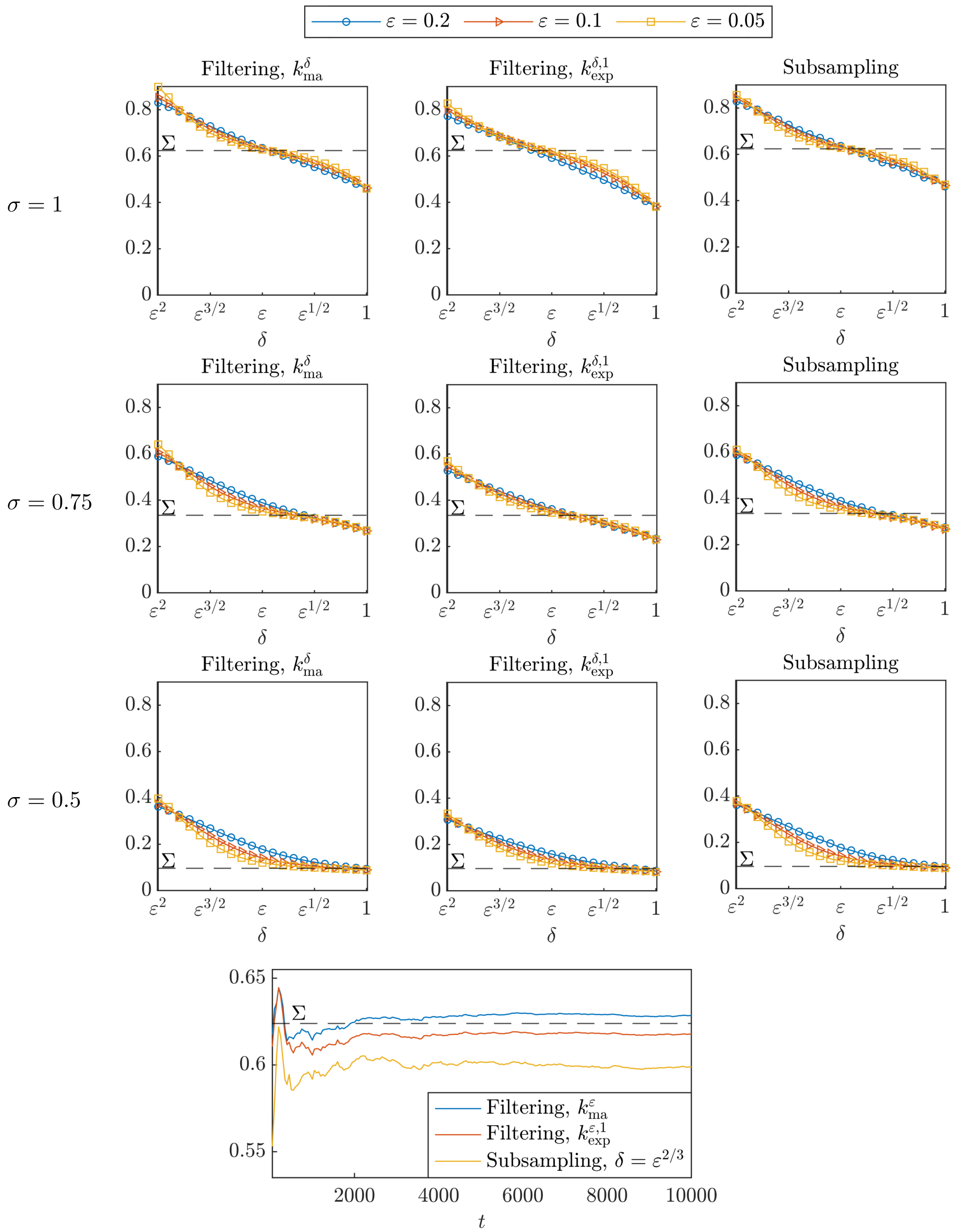


Figure 3.2 – Estimation of the diffusion coefficient of an effective OU equation. Top: Numerical results at final time $T = 10^4$ for variable $\sigma = 1, 0.75, 0.5$ (per row), for $\varepsilon = 0.2, 0.1, 0.05$ (blue, red, and yellow lines respectively), and for variable filtering/subsampling width δ (horizontal axis in all figures). Results for the estimators $\widehat{\Sigma}$: 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 $\delta = \varepsilon$, 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.

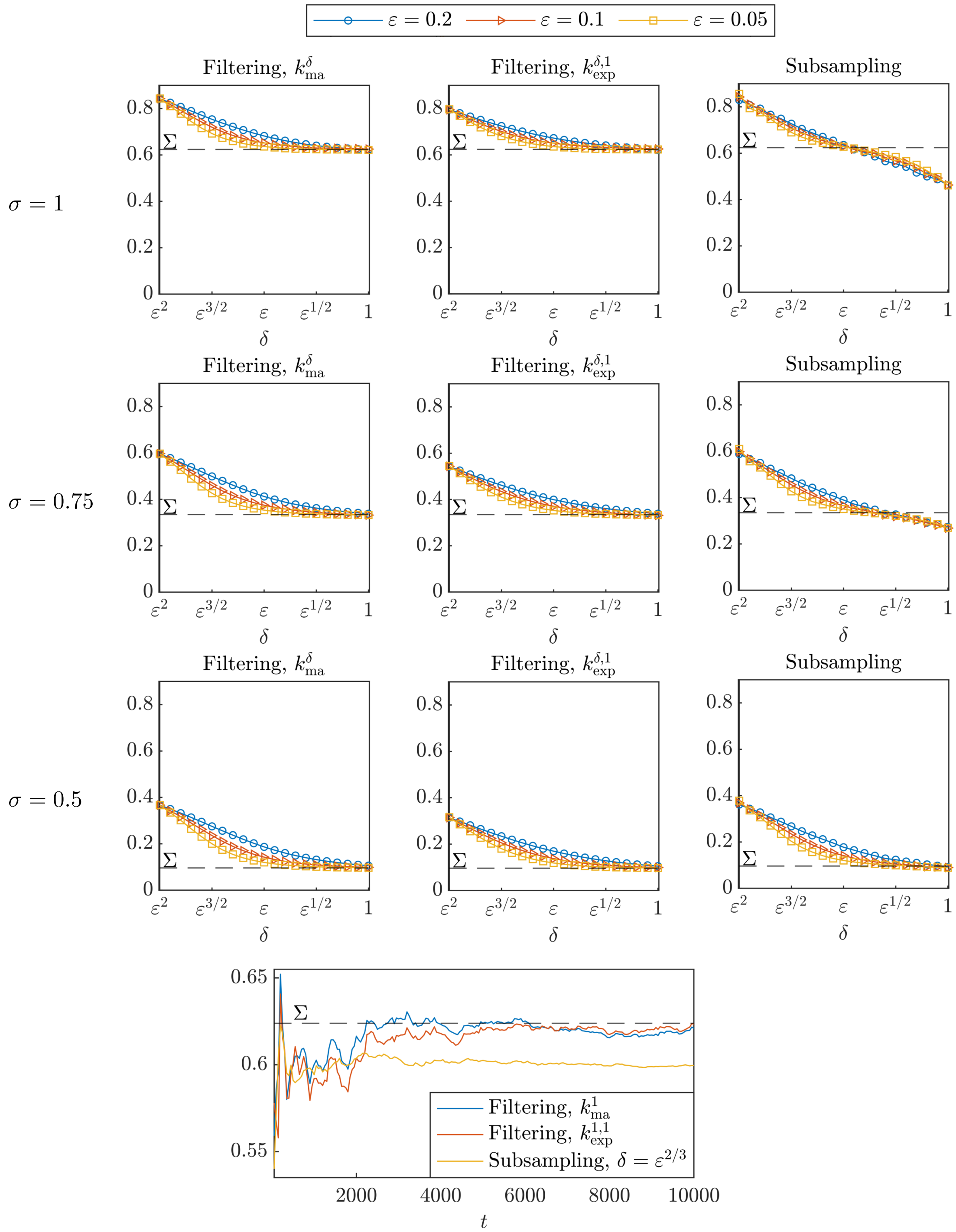


Figure 3.3 – Estimation of the diffusion coefficient of an effective OU equation. Top: Numerical results at final time $T = 10^4$ for variable $\sigma = 1, 0.75, 0.5$ (per row), for $\varepsilon = 0.2, 0.1, 0.05$ (blue, red, and yellow lines respectively), and for variable filtering/subsampling width δ (horizontal axis in all figures). Results for the estimators $\widetilde{\Sigma}$: 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.

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

(i) 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 [103]) for subsampling is similar for the three methods.

(ii) Figure 3.2: The estimators $\widehat{\Sigma}^{\delta}_{\mathrm{ma}}$, $\widehat{\Sigma}^{\delta,1}_{\mathrm{exp}}$, and $\widehat{\Sigma}^{\delta}_{\mathrm{sub}}$ 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 here seems to slightly outperform the concurrent methods.

(iii) Figure 3.3: The estimators $\widetilde{\Sigma}^{\delta}_{\mathrm{ma}}$ and $\widetilde{\Sigma}^{\delta,1}_{\mathrm{exp}}$ show enhanced accuracy with respect to the corresponding estimators of (ii) 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".

3.2.2 The semi-parametric setting

We now consider the semi-parametric setting for a one-dimensional multiscale Langevin equation of the form (1.10). In particular, we consider the number of parameters $L = 6$ and define $V \colon \mathbb{R} \to \mathbb{R}^L$ as

$$V(x) = \begin{pmatrix} \frac{x^6}{6} & \frac{x^5}{5} & \frac{x^4}{4} & \frac{x^3}{3} & \frac{x^2}{2} & x \end{pmatrix}^\top .$$

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

$$\alpha = \begin{pmatrix} 1 & -1 & -5.25 & 4.75 & 5 & -3 \end{pmatrix}^\top .$$

With this choice, the slow-scale potential $\mathcal{V} = \alpha \cdot V$ 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 $X^\varepsilon = (X^\varepsilon(t), 0 \le t \le T)$ with $T = 5 \cdot 10^4$, generated with the EM 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 (1.11). Similarly to Section 3.2.1, we compare the two filtering methodologies (moving average and exponential kernels), and subsampling. Moreover, we compute for all strategies the effective drift estimator $\widehat{A}$, and the effective diffusion estimators $\widehat{\Sigma}$ and $\widetilde{\Sigma}$.

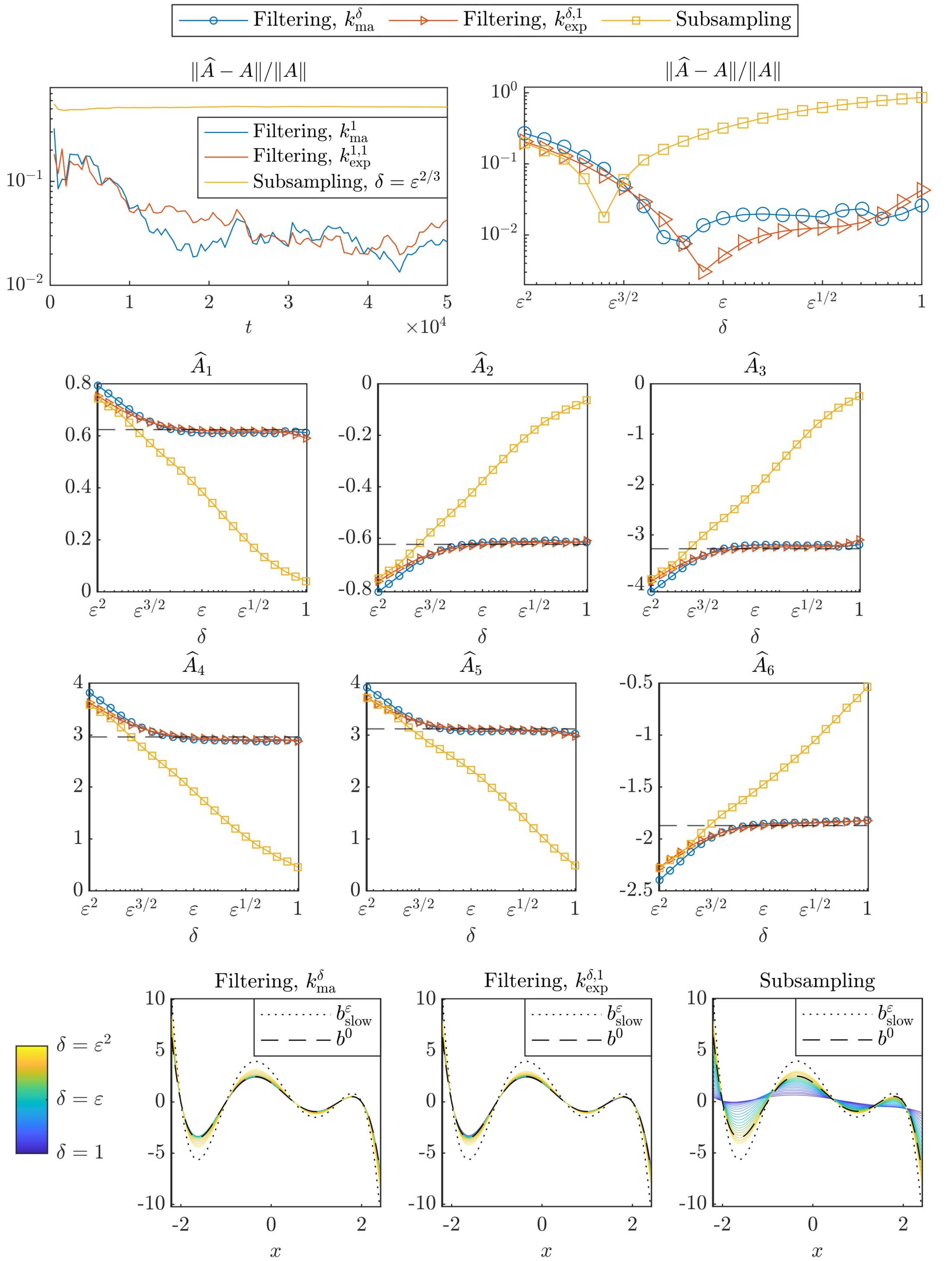


Figure 3.4 – Estimation of the drift coefficient in the one-dimensional semi-parametric setting. First row: on the left, evolution of the relative error with respect to $t \in [0, 5 \cdot 10^4]$, and on the right dependence of the relative error on the filtering/subsampling width $\delta \in [\varepsilon^2, 1]$. 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.

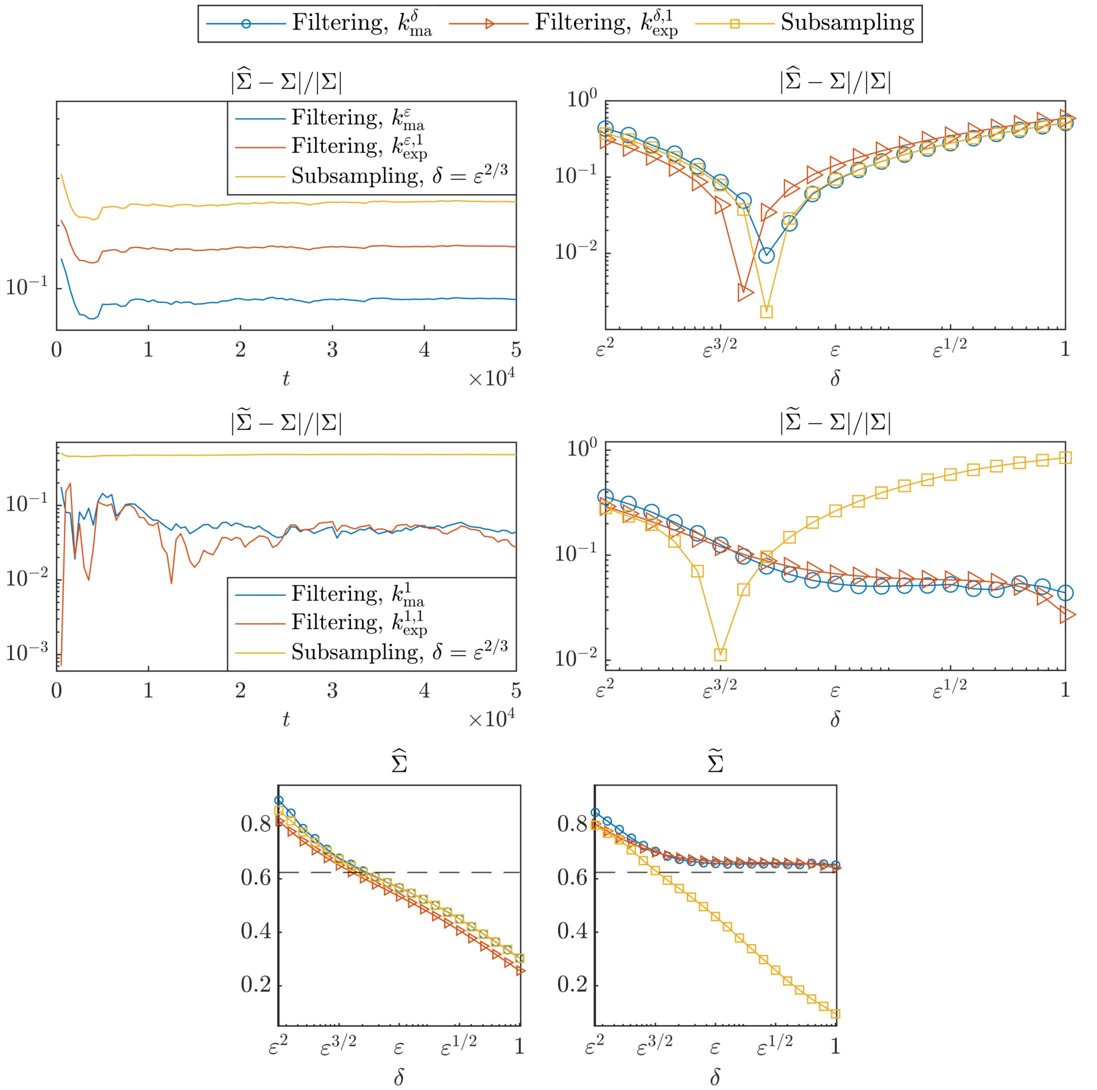


Figure 3.5 – Estimation of the diffusion coefficient in the one-dimensional semi-parametric setting. First ($\widehat{\Sigma}$) and second ($\widetilde{\Sigma}$) row: on the left, we show the evolution of the relative error with respect to $t \in [0, 5 \cdot 10^4]$, and on the right the dependence of 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.

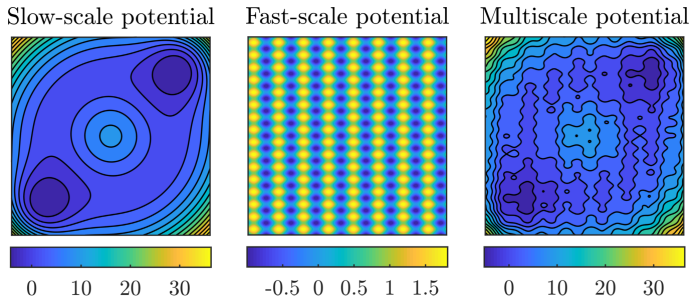


Figure 3.6 – Slow-, fast-, and multiscale potentials for the two-dimensional example of Section 3.2.3, depicted here in the square $(-2.5, 2.5)^2$.

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

(i) 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 that the optimal value for δ appears to be $\delta \approx \varepsilon^{3/2}$, which is surprising in view of the convergence result of [103]. 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.

(ii) Figure 3.5: The diffusion coefficient is estimated more accurately by the estimator $\widetilde{\Sigma}$ than $\widehat{\Sigma}$ when employing filtered data. Indeed, for both filtering kernels the estimator $\widetilde{\Sigma}$ 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 $\widehat{\Sigma}$ strongly depends on the filtering/subsampling width. We note that choosing $\delta = \varepsilon$ the moving average estimator $\widehat{\Sigma}^{\varepsilon}_{\mathrm{ma}}$ outperforms the corresponding estimator $\widehat{\Sigma}^{\varepsilon,1}_{\exp}$ based on exponential filtering. For subsampling, the two estimators are equivalent in terms of accuracy, and are extremely dependent of 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 $\varepsilon^{2/3}$.

3.2.3 A two-dimensional example

As a last numerical example, we consider a two-dimensional SDE ($d = 2$) of the form (1.4). In particular, we let $L = 4$ 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 $p\colon \mathbb{R}^2 \to \mathbb{R}$ as

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

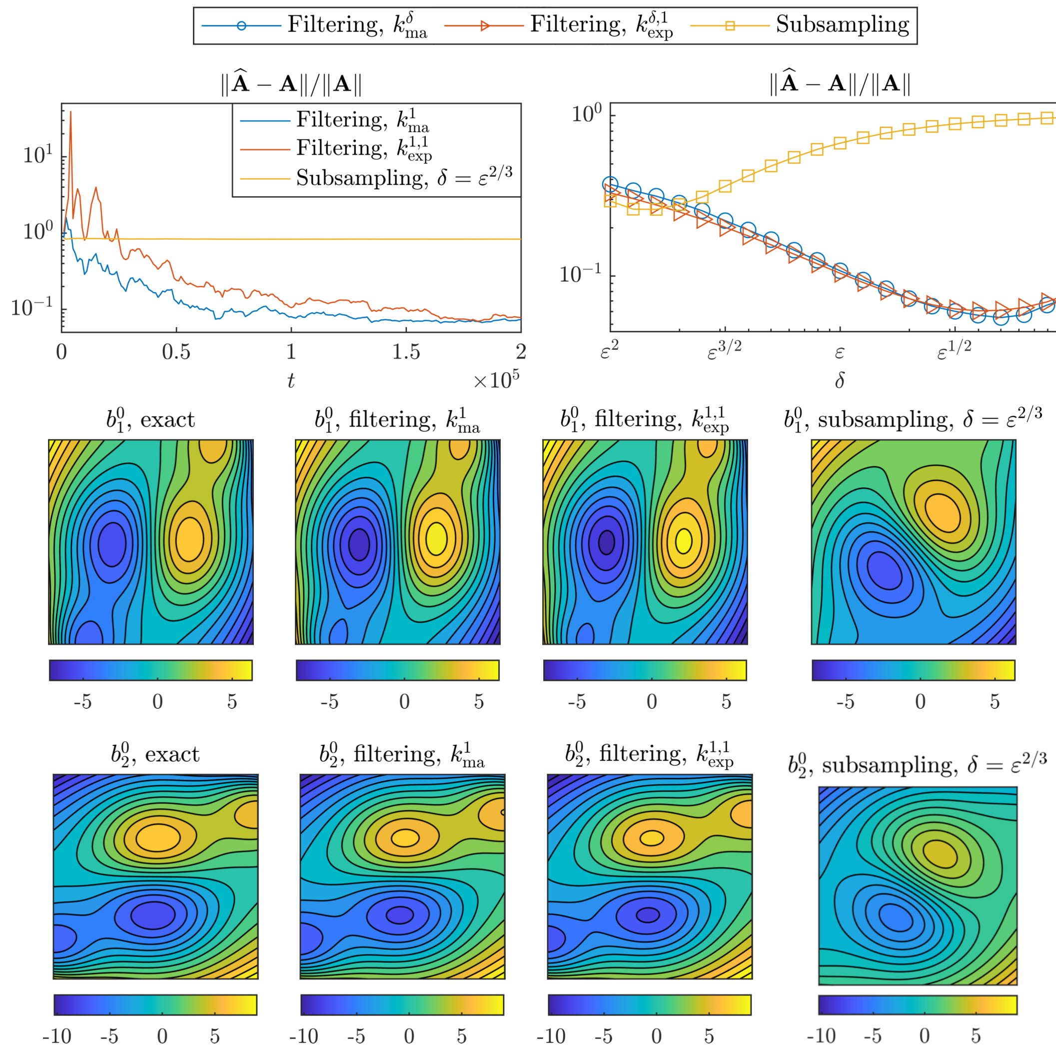


Figure 3.7 – Estimation of the effective drift coefficient for the two-dimensional example of Section 3.2.3. First row: Dependence of the relative error with respect to $t \in [0, T]$ (left) and to the subsampling/filtering width δ (right) for both filtering methods and subsampling. Second and third 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.

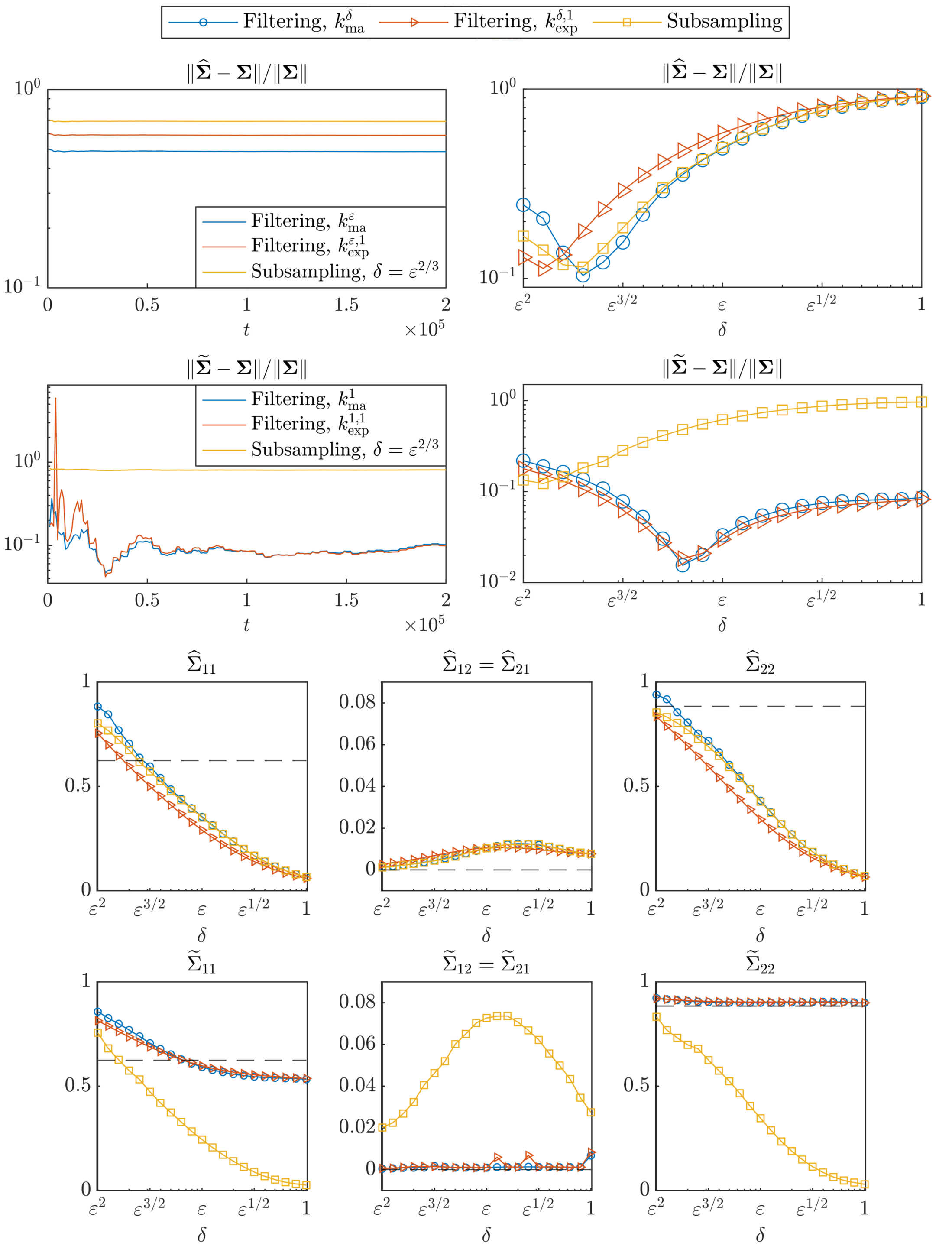


Figure 3.8 – Estimation of the effective diffusion coefficient for the two-dimensional example of Section 3.2.3. First and second row: Dependence of the relative error of the estimators $\widehat{\Sigma}$ (first row) and $\widetilde{\Sigma}$ (second row) with respect to $t \in [0, T]$ (left) and to the 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.

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 (1.12). 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 $X^\varepsilon = (X^\varepsilon(t), 0 \le t \le T)$ by integrating (1.4) with the EM 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

(i) 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 $\widehat{A}^1_{\mathrm{ma}}$ 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 $\widehat{A}^{1,1}_{\exp}$ 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.

(ii) 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 $\widetilde{\Sigma}_{\mathrm{sub}}$, instead, suffers from the lack of accuracy of the drift estimator $\widehat{A}^\delta_{\mathrm{sub}}$, and is not reliable. Always commenting on Figure 3.8, we remark that convergence with respect to $t \in [0, T]$ of the estimators $\widetilde{\Sigma}^1_{\mathrm{ma}}$ and $\widetilde{\Sigma}^{1,1}_{\exp}$ is similar, with the moving average filter seemingly less prone to instabilities for small t.

3.3 Asymptotic unbiasedness of the estimators

In this section we present the proof of Theorems 3.2 to 3.4 and Corollary 3.5, i.e., the results of asymptotic unbiasedness for our filtering-based estimators. In Chapter 2 the proofs of convergence are obtained with the kernel $k_{\mathrm{exp}}^{\delta,1}$ by noticing that the original trajectory X^ε and its filtered version $Z_{\mathrm{exp}}^{\delta,1,\varepsilon}$ are solution of an hypoelliptic system of Itô SDEs. For higher values of $\beta > 1$, the system describing the evolution of X^ε and $Z_{\mathrm{exp}}^{\delta,\beta,\varepsilon}$ is not a Itô system due to the presence of a memory term. In case we consider the moving average kernel k_{ma}^δ which we study here, the memory term simplifies to a constant delay. Hence, the evolution of the filtered trajectory $Z_{\mathrm{ma}}^{\delta,\varepsilon}$ can be coupled with the original trajectory X^ε through the system of stochastic delay differential equations (SDDEs)

$$\begin{aligned} \mathrm{d}X^\varepsilon(t) &= -\alpha \cdot V'(X^\varepsilon(t))\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t), \\ \mathrm{d}Z_{\mathrm{ma}}^{\delta,\varepsilon}(t) &= -\frac{1}{\delta}(X^\varepsilon(t-\delta) - X^\varepsilon(t))\,\mathrm{d}t. \end{aligned} \tag{3.8}$$

To be precise, the system above is a combination of a Itô SDE and a delay ordinary differential equation driven by a stochastic signal. Due to the theory of homogenization (see [20, Chapter 3], or [104, Chapter 18], or the proof of Lemma 2.9), if δ is independent of ε, the solution $(X^\varepsilon, Z_{\mathrm{ma}}^{\delta,\varepsilon})$ converges in law as random variables in $\mathcal{C}^0([0,T],\mathbb{R}^2)$ to the solution $(X^0, Z_{\mathrm{ma}}^{\delta,0})$ of the system

$$\begin{aligned} \mathrm{d}X^0(t) &= -A \cdot V'(X^0(t))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t), \\ \mathrm{d}Z_{\mathrm{ma}}^{\delta,0}(t) &= -\frac{1}{\delta}(X^0(t-\delta) - X^0(t))\,\mathrm{d}t. \end{aligned} \tag{3.9}$$

In the following, we first focus on ergodic properties of the couples $(X^\varepsilon, Z_{\mathrm{ma}}^{\delta,\varepsilon})$ and $(X^0, Z_{\mathrm{ma}}^{\delta,0})$ evolving according to (3.8) and (3.9), respectively. Then, we employ the invariant measures and the Fokker–Planck equations (FPEs) derived through the ergodicity theory to prove asymptotic unbiasedness. We remark that the strategy we adopt is similar to the one of Chapter 2. Still, different techniques need to be employed due to the delay in the second equation of the systems (3.8) and (3.9).

Let us remark that, for economy of notation, from now on in this section we will simply write X_t^ε and Z_t^ε instead of $X^\varepsilon(t)$ and $Z_{\mathrm{ma}}^{\delta,\varepsilon}(t)$, respectively, and similarly for all stochastic processes. Moreover, we drop explicit reference to the dependence of $\widetilde{M}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ and $\widetilde{v}_{\mathrm{ma}}^\delta(X^\varepsilon, T)$ on the parameter δ and we only write $\widetilde{M}_{\mathrm{ma}}(X^\varepsilon, T)$ and $\widetilde{v}_{\mathrm{ma}}(X^\varepsilon, T)$.

3.3.1 Ergodic properties

It is well-known (see, e.g., [103]) that X^ε is geometrically ergodic with invariant measure ν^ε whose density φ^ε takes the Gibbs form

$$\varphi^\varepsilon(x) = \frac{1}{C_{\nu^\varepsilon}} \exp\left(-\frac{\mathcal{V}^\varepsilon(x)}{\sigma}\right), \quad C_{\nu^\varepsilon} = \int_{\mathbb{R}} \exp\left(-\frac{\mathcal{V}^\varepsilon(x)}{\sigma}\right)\mathrm{d}x,$$

where

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

Moreover, an analogous result holds true for the homogenized process X^0, which is geometrically ergodic with invariant measure ν^0 whose density φ^0 is given by

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

We now introduce a similar result of ergodicity for the couples $(X^\varepsilon, Z^\varepsilon)$ and (X^0, Z^0) satisfying (3.8) and (3.9), respectively, i.e., for the multiscale process and its filtered version.

Proposition 3.8. *Under Assumption 1.4, the solution $(X^\varepsilon, Z^\varepsilon)$ of (3.8) is ergodic, and the density $\rho^\varepsilon_{\mathrm{ma}}$ of its invariant measure $\mu^\varepsilon_{\mathrm{ma}}$ on $\mathbb{R}^2$, such that $\mu^\varepsilon_{\mathrm{ma}}(\mathrm{d}x, \mathrm{d}z) = \rho^\varepsilon_{\mathrm{ma}}(x,z)\,\mathrm{d}x\,\mathrm{d}z$, satisfies*

$$\begin{aligned}
&\sigma \partial^2_{xx} \rho^\varepsilon_{\mathrm{ma}}(x,z) + \partial_x \left((\mathcal{V}^\varepsilon)'(x) \rho^\varepsilon_{\mathrm{ma}}(x,z) \right) + \frac{1}{\delta} \partial_z \left(\left(\int_{\mathbb{R}} y \mathfrak{S}^\varepsilon_{\mathrm{ma}}(y \mid x,z)\,\mathrm{d}y - x \right) \rho^\varepsilon_{\mathrm{ma}}(x,z) \right) = 0, \\
&\int_{\mathbb{R}} \int_{\mathbb{R}} \rho^\varepsilon_{\mathrm{ma}}(x,z)\,\mathrm{d}x\,\mathrm{d}z = 1,
\end{aligned} \tag{3.10}$$

where, if $X^\varepsilon_0 \sim \nu^\varepsilon$, it holds

$$\int_{\mathbb{R}} y \mathfrak{S}^\varepsilon_{\mathrm{ma}}(y \mid x,z)\,\mathrm{d}y = \mathbb{E}\left[X^\varepsilon_0 \mid X^\varepsilon_\delta = x, Z^\varepsilon_\delta = z\right],$$

i.e., $\mathfrak{S}^\varepsilon_{\mathrm{ma}}(\cdot \mid x,z)$ is the conditional density of X^ε_0 given $X^\varepsilon_\delta = x$ and $Z^\varepsilon_\delta = z$. Moreover, if δ is independent of ε, the solution (X^0, Z^0) of (3.9) is ergodic, and the density ρ^0_{ma} of its invariant measure μ^0_{ma} on $\mathbb{R}^2$, such that $\mu^0_{\mathrm{ma}}(\mathrm{d}x, \mathrm{d}z) = \rho^0_{\mathrm{ma}}(x,z)\,\mathrm{d}x\,\mathrm{d}z$, satisfies

$$\begin{aligned}
&\Sigma \partial^2_{xx} \rho^0_{\mathrm{ma}}(x,z) + \partial_x \left(A \cdot V'(x) \rho^0_{\mathrm{ma}}(x,z) \right) + \frac{1}{\delta} \partial_z \left(\left(\int_{\mathbb{R}} y \mathfrak{S}^0_{\mathrm{ma}}(y \mid x,z)\,\mathrm{d}y - x \right) \rho^0_{\mathrm{ma}}(x,z) \right) = 0, \\
&\int_{\mathbb{R}} \int_{\mathbb{R}} \rho^0_{\mathrm{ma}}(x,z)\,\mathrm{d}x\,\mathrm{d}z = 1,
\end{aligned} \tag{3.11}$$

where, if $X^0_0 \sim \nu^0$, it holds

$$\int_{\mathbb{R}} y \mathfrak{S}^0_{\mathrm{ma}}(y \mid x,z)\,\mathrm{d}y = \mathbb{E}\left[X^0_0 \mid X^0_\delta = x, Z^0_\delta = z\right],$$

i.e., $\mathfrak{S}^0_{\mathrm{ma}}(\cdot \mid x,z)$ is the conditional density of X^0_0 given $X^0_\delta = x$ and $Z^0_\delta = z$.

Proof. In order to prove that the joint process $(X^\varepsilon, Z^\varepsilon)$ is ergodic, we show that it admits a unique invariant measure. If X^ε_0 is distributed accordingly to its invariant measure ν^ε, which exists due to Assumption 1.4, then the processes $(X^\varepsilon_s, 0 \le s \le \delta)$ and $(X^\varepsilon_s, t-\delta \le s \le t)$ are equally distributed for all $t \ge \delta$. Hence, the two-dimensional random variables $\left(X^\varepsilon_\delta, \frac{1}{\delta}\int_0^\delta X^\varepsilon_s\,\mathrm{d}s\right)$ and $\left(X^\varepsilon_t, \frac{1}{\delta}\int_{t-\delta}^t X^\varepsilon_s\,\mathrm{d}s\right)$ are equal in law for all $t \ge \delta$. Recalling that the joint process $\left(X^\varepsilon_t, \frac{1}{\delta}\int_{t-\delta}^t X^\varepsilon_s\,\mathrm{d}s\right)$ is the solution $(X^\varepsilon_t, Z^\varepsilon_t)$ of the system (3.8), it follows that the invariant measure $\mu^\varepsilon_{\mathrm{ma}}$ on $\mathbb{R}^2$ is the law of the random variable $\left(X^\varepsilon_\delta, \frac{1}{\delta}\int_0^\delta X^\varepsilon_s\,\mathrm{d}s\right)$. The uniqueness of the invariant measure $\mu^\varepsilon_{\mathrm{ma}}$ is then a direct consequence of the uniqueness of the invariant measure ν^ε for the process X^ε since the joint measure $\mu^\varepsilon_{\mathrm{ma}}$ is uniquely determined by its marginal ν^ε. Moreover, the FPE for the one-time PDF related to an SDDE with a single fixed delay is well-known (see, e.g., [46, 61, 82]) and the stationary equation (3.10) for the density $\rho^\varepsilon_{\mathrm{ma}}$ of $\mu^\varepsilon_{\mathrm{ma}}$ is then obtained due to the particular form of the system (3.8). Finally, the results corresponding to the homogenized system (3.9) can be proved analogously. □

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

Lemma 3.9. *Let $\rho^\varepsilon_{\mathrm{ma}}(x,z) = \varphi^\varepsilon(x)\mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x)$, where φ^ε and $\rho^\varepsilon_{\mathrm{ma}}$ are the densities of the invariant measures of X^ε and $(X^\varepsilon, Z^\varepsilon)$, respectively, and where $\mathfrak{R}^\varepsilon_{\mathrm{ma}}$ is the conditional density of Z^ε given X^ε. Then, if $X^\varepsilon_0 \sim \nu^\varepsilon$, it holds*

$$\sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z)\varphi^\varepsilon(x)\partial_x \mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x) \,\mathrm{d}x \,\mathrm{d}z = \frac{1}{\delta}\, \mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}} \left[(X^\varepsilon_\delta - Z^\varepsilon_\delta)\,(X^\varepsilon_\delta - X^\varepsilon_0)\, V''(Z^\varepsilon_\delta) \right]. \tag{3.12}$$

Moreover, if δ is independent of ε and writing $\rho^0_{\mathrm{ma}}(x,z) = \varphi^0(x)\mathfrak{R}^0_{\mathrm{ma}}(z \mid x)$ for the density of the homogenized invariant measure μ^0_{ma} of (X^0, Z^0), it holds

$$\Sigma \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z)\varphi^0(x)\partial_x \mathfrak{R}^0_{\mathrm{ma}}(z \mid x) \,\mathrm{d}x \,\mathrm{d}z = \frac{1}{\delta}\, \mathbb{E}^{\mu^0_{\mathrm{ma}}} \left[\left(X^0_\delta - Z^0_\delta\right) \left(X^0_\delta - X^0_0\right) V''(Z^0_\delta) \right]. \tag{3.13}$$

Proof. We proceed similarly to the proof of Lemma 2.5. Replacing the decomposition $\rho^\varepsilon_{\mathrm{ma}}(x,z) = \varphi^\varepsilon(x)\mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x)$ into the FPE (3.10) gives

$$\partial_x \left(\sigma\varphi^\varepsilon(x)\partial_x \mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x) \right) + \partial_z \left(\frac{1}{\delta} \left(\int_{\mathbb{R}} y \mathfrak{S}^\varepsilon_{\mathrm{ma}}(y \mid x,z) \,\mathrm{d}y - x \right) \varphi^\varepsilon(x)\mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x) \right) = 0.$$

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

$$\sigma \int_{\mathbb{R}} \int_{\mathbb{R}} \partial_x f(x,z)\varphi^\varepsilon(x)\partial_x \mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x) \,\mathrm{d}x \,\mathrm{d}z = \frac{1}{\delta}\, \mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}} \left[\partial_z f(X^\varepsilon_\delta, Z^\varepsilon_\delta)(X^\varepsilon_\delta - X^\varepsilon_0) \right].$$

The choice $f(x,z) = (x-z)V'(z) + V(z)$ gives equation (3.12). Finally, equation (3.13) is obtained analogously employing the FPE of the homogenized SDE (3.11). □

3.3.2 Preliminary results

Let us first introduce the notation

$$\begin{aligned} \widetilde{\mathcal{M}}^\varepsilon_{\mathrm{ma}} &\coloneqq \mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}} \left[V'(Z^\varepsilon) \otimes V'(X^\varepsilon) \right], & \qquad \widetilde{\mathcal{M}}^0_{\mathrm{ma}} &\coloneqq \mathbb{E}^{\mu^0_{\mathrm{ma}}} \left[V'(Z^0) \otimes V'(X^0) \right], \\ \mathcal{M}^\varepsilon &\coloneqq \mathbb{E}^{\nu^\varepsilon} \left[V'(X^\varepsilon) \otimes V'(X^\varepsilon) \right], & \qquad \mathcal{M}^0 &\coloneqq \mathbb{E}^{\nu^0} \left[V'(X^0) \otimes V'(X^0) \right], \end{aligned}$$

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

Lemma 3.10. *Under Assumption 1.4, it holds*

$$X^\varepsilon_\delta - Z^\varepsilon_\delta = \frac{\sqrt{2\sigma}}{\delta} \int_0^\delta t(1 + \Phi'(Y^\varepsilon_t)) \,\mathrm{d}W_t + R(\varepsilon,\delta), \tag{3.14}$$

where Φ is the solution of the cell problem (1.7) and where the remainder $R(\varepsilon,\delta)$ satisfies for all $p \geq 1$ and a constant $C > 0$ independent of ε and δ

$$\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}} \left[|R(\varepsilon,\delta)|^p \right]^{1/p} \leq C\,(\varepsilon + \delta). \tag{3.15}$$

Moreover, if X^ε_0 is stationary, i.e. $X^\varepsilon_0 \sim \varphi^\varepsilon$, it holds

$$\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}} \left[|X^\varepsilon - Z^\varepsilon|^p \right]^{1/p} \leq C \left(\delta^{1/2} + \varepsilon \right), \tag{3.16}$$

$$\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}} \left[|Z^\varepsilon|^p \right]^{1/p} \leq C. \tag{3.17}$$

Proof. Employing the decomposition (5.8) in [103] and due to [103, Lemma 5.5, Propsition 5.8] we have for all $t \in [0, \delta]$

$$X_\delta^\varepsilon = X_t^\varepsilon + \sqrt{2\sigma} \int_t^\delta (1 + \Phi'(Y_s^\varepsilon)) \, \mathrm{d}W_s + R(\varepsilon, \delta), \tag{3.18}$$

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

$$\mathbb{E}^{\mu_{\mathrm{ma}}^\varepsilon} \left[|R(\varepsilon, \delta)|^p \right]^{1/p} \leq C \left(\varepsilon + \delta \right).$$

Therefore, we obtain

$$X_\delta^\varepsilon - Z_\delta^\varepsilon = \frac{1}{\delta} \int_0^\delta (X_\delta^\varepsilon - X_t^\varepsilon) \, \mathrm{d}t = \frac{\sqrt{2\sigma}}{\delta} \int_0^\delta \int_t^\delta (1 + \Phi'(Y_s^\varepsilon)) \, \mathrm{d}W_s \, \mathrm{d}t + R(\varepsilon, \delta),$$

which by Fubini's theorem yields

$$X_\delta^\varepsilon - Z_\delta^\varepsilon = \frac{\sqrt{2\sigma}}{\delta} \int_0^\delta t(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}W_t + R(\varepsilon, \delta),$$

and proves (3.14) and (3.15). By the Itô isometry, it holds

$$\mathbb{E}^{\mu_{\mathrm{ma}}^\varepsilon} \left[\left| \int_0^\delta t(1 + \Phi'(Y_t^\varepsilon)) \, \mathrm{d}W_t \right|^p \right]^{1/p} \leq C\delta^{3/2}, \tag{3.19}$$

which, together with (3.14), (3.15) and the proof of Proposition 3.8, gives (3.16). Finally, (3.17) is proved by applying the triangle inequality and due to (3.16) and [103, Corollary 5.4]. □

Lemma 3.11. *Under Assumption 1.4 and if $\delta = \varepsilon^\zeta$ with $\zeta \in (0, 2)$, then it holds*

$$\lim_{\varepsilon \to 0} \widetilde{\mathcal{M}}_{\mathrm{ma}}^\varepsilon = \mathcal{M}^0.$$

Proof. By the triangle inequality, we have

$$\left\| \widetilde{\mathcal{M}}_{\mathrm{ma}}^\varepsilon - \mathcal{M}^0 \right\| \leq \left\| \widetilde{\mathcal{M}}_{\mathrm{ma}}^\varepsilon - \mathcal{M}^\varepsilon \right\| + \left\| \mathcal{M}^\varepsilon - \mathcal{M}^0 \right\|.$$

The first term vanishes as $\varepsilon \to 0$ due to Lemma 3.10, [103, Corollary 5.4] and since V' is Lipschitz under Assumption 1.4. The second term vanishes due to the theory of homogenization as $\varepsilon \to 0$. □

3.3.3 Proof of the main results

We can now prove our main results, i.e., Theorems 3.2 to 3.4 and Corollary 3.5.

Proof of Theorem 3.2. Following the proof of Theorem 2.12, we have

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

where

$$I_1 = \frac{1}{T} \widetilde{M}_{\mathrm{ma}}(X^\varepsilon, T)^{-1} \int_0^T \frac{1}{\varepsilon} p' \left(\frac{X_t^\varepsilon}{\varepsilon} \right) V'(Z_t^\varepsilon) \, \mathrm{d}t,$$
$$I_2 = \frac{\sqrt{2\sigma}}{T} \widetilde{M}_{\mathrm{ma}}(X^\varepsilon, T)^{-1} \int_0^T V'(Z_t^\varepsilon) \, \mathrm{d}W_t,$$

and where

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

uniformly in ε by Lemma 3.10 and the strong law of large numbers for martingales. Considering I_1, due to Assumption 1.4 the ergodic theorem and an integration by parts yield

$$\lim_{T\to\infty} I_1 = -\alpha + (\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{ma}})^{-1}\sigma \int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)\varphi^\varepsilon(x)\partial_x \mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x)\,\mathrm{d}x\,\mathrm{d}z,$$

where $\mathfrak{R}^\varepsilon_{\mathrm{ma}}(z \mid x)$ is defined in Lemma 3.9, which also implies

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

where

$$\mathcal{A}^\varepsilon(\delta) = \frac{1}{\delta}(\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{ma}})^{-1}\,\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}}\left[(X^\varepsilon_\delta - Z^\varepsilon_\delta)(X^\varepsilon_\delta - X^\varepsilon_0)V''(Z^\varepsilon_\delta)\right]. \tag{3.20}$$

It remains to show that

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

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

$$\lim_{\varepsilon\to 0}\mathcal{A}^\varepsilon(\delta) = \frac{1}{\delta}(\widetilde{\mathcal{M}}^0_{\mathrm{ma}})^{-1}\,\mathbb{E}^{\mu^0_{\mathrm{ma}}}\left[(X^0_\delta - Z^0_\delta)(X^0_\delta - X^0_0)V''(Z^0_\delta)\right],$$

so that applying Lemma 3.9 for the homogenized equation backwards we have

$$\lim_{\varepsilon\to 0}\mathcal{A}^\varepsilon(\delta) = (\widetilde{\mathcal{M}}^0_{\mathrm{ma}})^{-1}\Sigma \int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)\varphi^0(x)\partial_x \mathfrak{R}^0_{\mathrm{ma}}(z \mid x)\,\mathrm{d}x\,\mathrm{d}z.$$

An integration by parts then gives

$$\lim_{\varepsilon\to 0}\mathcal{A}^\varepsilon(\delta) = (\widetilde{\mathcal{M}}^0_{\mathrm{ma}})^{-1}\widetilde{\mathcal{M}}^0_{\mathrm{ma}} A = A,$$

which concludes *Case 1.*

Case 2: $\delta = \varepsilon^\zeta$ with $\zeta \in (0,2)$. Replacing formulas (3.18) with $t = 0$ and (3.14) into (3.20) gives

$$\begin{aligned}\mathcal{A}^\varepsilon(\delta) = {} & \frac{2\sigma}{\delta^2}(\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{ma}})^{-1}\,\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}}\left[\left(\int_0^\delta t(1+\Phi'(Y^\varepsilon_t))\,\mathrm{d}W_t\right)\left(\int_0^\delta (1+\Phi'(Y^\varepsilon_t))\,\mathrm{d}W_t\right)V''(Z^\varepsilon_\delta)\right] \\ & + \widetilde{R}_1(\varepsilon,\delta),\end{aligned}$$

where, due to Lemma 3.10, estimate (3.19) and the fact that by the Itô isometry

$$\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}}\left[\left|\int_t^\delta (1+\Phi'(Y^\varepsilon_s))\,\mathrm{d}W_s\right|^p\right]^{1/p} \le C\delta^{1/2}, \tag{3.21}$$

it follows that the remainder satisfies

$$\left\|\widetilde{R}_1(\varepsilon,\delta)\right\| \le C\left(\delta^{1/2} + \varepsilon\delta^{-1/2} + \varepsilon^2\delta^{-1}\right). \tag{3.22}$$

Moreover, since V'' is Lipschitz under Assumption 1.4 and due to the triangle inequality, equation (3.18), estimates (3.15), (3.21) and Lemma 3.10, it holds for all $t \in [0, \delta]$

$$\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}}\left[\|V''(Z^\varepsilon_\delta) - V''(X^\varepsilon_t)\|^p\right]^{1/p} \le C\left(\varepsilon + \delta^{1/2}\right), \tag{3.23}$$

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

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

where due to (3.22) and (3.23) the remainder satisfies

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

Repeating the last part of the proof of Lemma 2.16, we then obtain

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

Finally, since $\delta = \varepsilon^\zeta$ with $\zeta\in(0,2)$, by (3.22) and due to Lemma 3.11 we obtain

$$\lim_{\varepsilon\to0}\mathcal{A}^\varepsilon(\delta) = \Sigma(\mathcal{M}^0)^{-1}\,\mathbb{E}^{\nu^0}[V''(X^0)],$$

and an integration by parts gives

$$\lim_{\varepsilon\to0}\mathcal{A}^\varepsilon(\delta) = \Sigma(\mathcal{M}^0)^{-1}\frac{1}{\Sigma}\mathcal{M}^0A = A,$$

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

Proof of Theorem 3.3. Due to Assumption 1.4 the ergodic theorem gives

$$\lim_{T\to\infty}\widehat{\Sigma}^\delta_{\mathrm{ma}}(X^\varepsilon,T) = \frac{1}{\delta}\,\mathbb{E}^{\mu^\varepsilon_{\mathrm{ma}}}\left[(X^\varepsilon_\delta - Z^\varepsilon_\delta)(X^\varepsilon_\delta - X^\varepsilon_0)\right]. \tag{3.24}$$

Following step-by-step *Case 2* of the proof of Theorem 3.2 with the value 1 instead of $V''(Z^\varepsilon_\delta)$, and without the pre-multiplication by $(\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{ma}})^{-1}$, we obtain the desired result. □

Remark 3.12. It is clear from the proof of Theorem 3.3 that it is theoretically not possible to choose δ independent of ε in the computation of $\widehat{\Sigma}^\delta_{\mathrm{ma}}(X^\varepsilon,T)$. Let $L=1$ and $V(x)=x^2/2$, so that X^0 is an OU 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

$$\mathcal{C}(t,s) = \frac{\Sigma}{A}e^{-A|t-s|}. \tag{3.25}$$

By (3.24) and (3.25) we can therefore explicitly compute

$$\lim_{\varepsilon\to0}\lim_{T\to\infty}\widehat{\Sigma}^\delta_{\mathrm{ma}}(X^\varepsilon,T) = \frac{1}{\delta}\,\mathbb{E}^{\mu^0_{\mathrm{ma}}}\left[\left(X^0_\delta - Z^0_\delta\right)\left(X^0_\delta - X^0_0\right)\right] = \frac{1-e^{-\delta A}}{\delta A}\Sigma,$$

so that $\widehat{\Sigma}^\delta_{\mathrm{ma}}(X^\varepsilon,T)$ is asymptotically unbiased only if $\delta\to0$.

Proof of Theorem 3.4. Notice that due to (1.16) we have

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

which together with (3.7) and by Theorem 1.8 implies

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

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

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

which is the desired result. □

Proof of Corollary 3.5. The desired results follow directly from Theorems 2.12, 2.17 and 3.2, [103, Theorem 3.5] and Theorem 3.4. We remark that the limit in Theorem 2.17 holds true also a.s. and the proof of [103, Theorem 3.5] can be modified (see Remark 3.13) such that hypothesis (3.7) is satisfied. □

Remark 3.13. The proof of [103, Theorem 3.5] can be modified in order to show that the estimator $\widehat{A}^\delta_{\mathrm{sub}}(X^\varepsilon, T)$ given in (1.17) satisfies

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}^\delta_{\mathrm{sub}}(X^\varepsilon, T) = A, \quad \text{a.s.} \tag{3.26}$$

Due to Assumption 1.4(ii) and by the ergodic theorem we have

$$\lim_{T \to \infty} \widehat{A}^\delta_{\mathrm{sub}}(X^\varepsilon, T) = -\frac{\mathbb{E}^{\nu^\varepsilon}[V'(X_0^\varepsilon)(X_\delta^\varepsilon - X_0^\varepsilon)]}{\mathbb{E}^{\nu^\varepsilon}[V'(X_0^\varepsilon)^2]}.$$

We then employ Lemma 7.17 with f the identity function and $\Delta = \delta$ and we notice that the martingale

$$M_t^\varepsilon := \sqrt{2\sigma} \int_0^t (1 + \Phi'(Y_s^\varepsilon)) \,\mathrm{d}W_s,$$

where Φ is defined in (1.7) and $Y_s^\varepsilon = X_s^\varepsilon / \varepsilon$, is such that $M_0^\varepsilon = 0$. Therefore, we obtain

$$\lim_{T \to \infty} \widehat{A}^\delta_{\mathrm{sub}}(X^\varepsilon, T) = A + \widetilde{R}(\varepsilon, \delta),$$

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

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

Finally, since $\delta = \varepsilon^\zeta$ with $\zeta \in (0,1)$ we deduce the desired result (3.26).

3.4 Conclusion

In this chapter 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 the similar class of estimators that we introduced in Chapter 2. 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.

4 Further results and open problems

In this chapter we present additional theoretical results related to the homogenization of the multiscale Langevin dynamics and potential improvements of the filtering methodology, which appear to work in practice, but are still open problems from the theoretical viewpoint. In particular, a homogenization result for the multiscale Langevin dynamics and the rate of convergence for the expectation of smooth functions with respect to its invariant measure towards the expectation with respect to the homogenized invariant measure are presented in Sections 4.1 and 4.2, respectively. Then, Section 4.3 is devoted to the Stratonovich formulation of the MLE. Finally, we consider the estimator of Chapter 2, and in Section 4.4 we study filtered data obtained by repeatedly applying the exponential filter and in Section 4.5 we analyze its asymptotic normality.

4.1 Homogenization of Langevin dynamics via Γ–convergence

In this section we prove an homogenization result for the backward Kolmogorov equation (BKE) of the multiscale overdamped Langevin stochastic differential equation (SDE), employing the theory of the evolutionary Γ-convergence, which is presented in detail in [87]. This theory provides an abstract framework where it is possible to prove rigorous convergence results for generalized gradient flows. Therefore, it is a powerful tool which can be employed to show homogenization results and it is different from the more classic approaches like multiscale expansion, two-scale convergence and Tartar's method of oscillating test functions. In addition to the result presented here for the Langevin dynamics, we believe that evolutionary Γ-convergence can be useful in more general settings.

4.1.1 Problem setting

We consider the two-scale SDE (1.3) in the nonparametric form in one dimension, i.e.,

$$\mathrm{d}X^\varepsilon(t) = -\mathcal{V}'(X^\varepsilon(t))\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t), \tag{4.1}$$

and whose homogenized counterpart is therefore

$$\mathrm{d}X^0(t) = -\mathcal{K}\mathcal{V}'(X^0(t))\,\mathrm{d}t + \sqrt{2\mathcal{K}\sigma}\,\mathrm{d}W(t), \tag{4.2}$$

We now introduce the assumptions on the slow scale potential $\mathcal{V}$, which will be needed in the following analysis.

Assumption 4.1. The potential $\mathcal{V} \in \mathcal{C}^\infty(\mathbb{R})$ is polynomially bounded from above and bounded from below, and there exist $a, b > 0$ such that

$$-\mathcal{V}'(x)x \leq a - bx^2. \tag{4.3}$$

Moreover, it holds

$$\lim_{|x|\to+\infty}\left(\frac{1}{4}\|\nabla\mathcal{V}(x)\|^2-\frac{1}{2}\Delta\mathcal{V}(x)\right)=+\infty \qquad \text{and} \qquad \lim_{|x|\to+\infty}\|\nabla\mathcal{V}(x)\|=+\infty. \tag{4.4}$$

We recall that, under the dissipative condition (4.3) in Assumption 4.1, the processes $X^\varepsilon(t)$ and $X^0(t)$ are geometrically ergodic with unique invariant measures ν^ε and ν^0, whose densities φ^ε and φ^0 are given by

$$\varphi^\varepsilon(x)=\frac{1}{C_{\nu^\varepsilon}}e^{-\frac{1}{\sigma}\left(\mathcal{V}(x)+p\left(\frac{x}{\varepsilon}\right)\right)}, \quad \text{where} \quad C_{\nu^\varepsilon}=\int_{\mathbb{R}}e^{-\frac{1}{\sigma}\left(\mathcal{V}(x)+p\left(\frac{x}{\varepsilon}\right)\right)}\,\mathrm{d}x, \tag{4.5}$$

and

$$\varphi^0(x)=\frac{1}{C_{\nu^0}}e^{-\frac{1}{\sigma}\mathcal{V}(x)}, \quad \text{where} \quad C_{\nu^0}=\int_{\mathbb{R}}e^{-\frac{1}{\sigma}\mathcal{V}(x)}\,\mathrm{d}x. \tag{4.6}$$

We note that the density φ^ε can be rewritten using φ^0 and the definition of the density ω of the measure π in (1.8) as

$$\varphi^\varepsilon(x)=\frac{C_{\nu^0}C_\pi}{C_{\nu^\varepsilon}}\omega\left(\frac{x}{\varepsilon}\right)\varphi^0(x), \tag{4.7}$$

where

$$\lim_{\varepsilon\to0}C_{\nu^\varepsilon}=\frac{C_{\nu^0}C_\pi}{\mathbb{T}}. \tag{4.8}$$

We now consider the BKE of the multiscale SDE (4.1)

$$\begin{aligned}\frac{\partial u^\varepsilon}{\partial t}(t,x)&=\sigma\frac{\partial^2u^\varepsilon}{\partial x^2}(t,x)-\left(\mathcal{V}'(x)+\frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)\frac{\partial u^\varepsilon}{\partial x}(t,x),\\ u^\varepsilon(0,x)&=u_0(x),\end{aligned} \tag{4.9}$$

for some initial condition u_0, and its homogenized counterpart

$$\begin{aligned}\frac{\partial u^0}{\partial t}(t,x)&=\mathcal{K}\sigma\frac{\partial^2u^0}{\partial x^2}(t,x)-\mathcal{K}\mathcal{V}'(x)\frac{\partial u^0}{\partial x}(t,x),\\ u^0(0,x)&=u_0(x),\end{aligned} \tag{4.10}$$

and we aim to prove an homogenization result for the BKE, i.e., the convergence of the solution u^ε of (4.9) to the solution u^0 of (4.10) in some sense which will be specified later. In particular, we employ the theory of evolutionary Γ-convergence which has been developed in [87]. For the following analysis, it is useful to define the functions $w^\varepsilon(t,x)=e^{\lambda t}u^\varepsilon(t,x)$ and $w^0(t,x)=e^{\lambda t}u^0(t,x)$ for a fixed value of $\lambda>0$. Then, due to equations (4.9) and (4.10), they are the solutions of

$$\begin{aligned}\frac{\partial w^\varepsilon}{\partial t}(t,x)&=\sigma\frac{\partial^2w^\varepsilon}{\partial x^2}(t,x)-\left(\mathcal{V}'(x)+\frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)\frac{\partial w^\varepsilon}{\partial x}(t,x)+\lambda w^\varepsilon(t,x),\\ w^\varepsilon(0,x)&=u_0(x),\end{aligned} \tag{4.11}$$

and

$$\begin{aligned}\frac{\partial w^0}{\partial t}(t,x)&=\mathcal{K}\sigma\frac{\partial^2w^0}{\partial x^2}(t,x)-\mathcal{K}\mathcal{V}'(x)\frac{\partial w^0}{\partial x}(t,x)+\lambda w^0(t,x),\\ w^0(0,x)&=u_0(x),\end{aligned} \tag{4.12}$$

respectively. We remark that once we prove the convergence of w^ε to w^0 then the convergence of u^ε to u^0 follows directly since $u^\varepsilon(t,x)=e^{-\lambda t}w^\varepsilon(t,x)$ and $u^0(t,x)=e^{-\lambda t}w^0(t,x)$.

4.1.2 Gradient flow formulation

In this section we reformulate the homogenization problem of the BKE using the Integrated Evolutionary Variational Estimate (IEVE) presented in [87, Section 3.4.1]. Let us introduce the gradient system $(L^2_{\varphi^0}(\mathbb{R}), \mathcal{E}^\varepsilon, \Psi^\varepsilon)$, where $L^2_{\varphi^0}(\mathbb{R})$ is the Lebesgue space weighted by the invariant measure of the homogenized process and the functionals $\mathcal{E}^\varepsilon, \Psi^\varepsilon \colon L^2_{\varphi^0}(\mathbb{R}) \to [0, +\infty]$ are defined by

$$\mathcal{E}^\varepsilon(u) = \begin{cases} \frac{1}{2}\sigma \int_{\mathbb{R}} (\partial_x u(x))^2 \varphi^\varepsilon(x) \,\mathrm{d}x + \frac{1}{2}\lambda \int_{\mathbb{R}} u(x)^2 \varphi^\varepsilon(x) \,\mathrm{d}x & \text{if } u \in H^1_{\varphi^0}(\mathbb{R}), \\ +\infty & \text{if } u \in L^2_{\varphi^0}(\mathbb{R}) \setminus H^1_{\varphi^0}(\mathbb{R}), \end{cases} \tag{4.13}$$

and

$$\Psi^\varepsilon(v) = \frac{1}{2} \int_{\mathbb{R}} v(x)^2 \varphi^\varepsilon(x) \,\mathrm{d}x, \tag{4.14}$$

where $H^1_{\varphi^0}(\mathbb{R})$ is the Sobolev space weighted by the invariant measure φ^0 of the homogenized process (4.2). Moreover, we introduce the limit functionals $\mathcal{E}^0$ and Ψ^0 given by

$$\mathcal{E}^0(u) = \begin{cases} \frac{1}{2}\mathcal{K}\sigma \int_{\mathbb{R}} (\partial_x u(x))^2 \varphi^0(x) \,\mathrm{d}x + \frac{1}{2}\lambda \int_{\mathbb{R}} u(x)^2 \varphi^0(x) \,\mathrm{d}x & \text{if } u \in H^1_{\varphi^0}(\mathbb{R}), \\ +\infty & \text{if } u \in L^2_{\varphi^0}(\mathbb{R}) \setminus H^1_{\varphi^0}(\mathbb{R}), \end{cases} \tag{4.15}$$

and

$$\Psi^0(v) = \frac{1}{2} \int_{\mathbb{R}} v(x)^2 \varphi^0(x) \,\mathrm{d}x. \tag{4.16}$$

Then, equations (4.11) and (4.12) can be rewritten employing the gradient flow formulation as

$$\mathrm{D}\Psi^\varepsilon(\dot{w}^\varepsilon) = -\mathrm{D}\mathcal{E}^\varepsilon(w^\varepsilon) \qquad \text{and} \qquad \mathrm{D}\Psi^0(\dot{w}^0) = -\mathrm{D}\mathcal{E}^0(w^0), \tag{4.17}$$

where the dot denotes the derivative with respect to the time variable. Let us finally recall the definition of continuous, Gamma and Mosco convergence of functionals $\mathcal{F}^\varepsilon, \mathcal{F}^0 \colon X \to [-\infty, +\infty]$, where X is a reflexive Banach space, which will be employed in the analysis. For further details on the gradient flow formulation and on the different notions of convergence we refer to [87].

Definition 4.2 (Continuous convergence)**.** We say that $\mathcal{F}^\varepsilon$ strongly converges to $\mathcal{F}^0$ and we write $\mathcal{F}^\varepsilon \xrightarrow{\mathrm{C}} \mathcal{F}^0$ if

$$u^\varepsilon \to u^0 \text{ in } X \qquad \Longrightarrow \qquad \mathcal{F}^\varepsilon(u^\varepsilon) \to \mathcal{F}^0(u^0).$$

Moreover, we say that $\mathcal{F}^\varepsilon$ weakly converges to $\mathcal{F}^0$ and we write $\mathcal{F}^\varepsilon \overset{\mathrm{C}}{\rightharpoonup} \mathcal{F}^0$ if

$$u^\varepsilon \rightharpoonup u^0 \text{ in } X \qquad \Longrightarrow \qquad \mathcal{F}^\varepsilon(u^\varepsilon) \to \mathcal{F}^0(u^0).$$

Remark 4.3. Notice that weak continuous convergence implies strong continuous convergence.

Definition 4.4 (Gamma convergence)**.** We say that $\mathcal{F}^\varepsilon$ strongly Gamma converges to $\mathcal{F}^0$ and we write $\mathcal{F}^\varepsilon \xrightarrow{\Gamma} \mathcal{F}^0$ if

$$\begin{aligned} &\text{(G1S)} \qquad u^\varepsilon \to u^0 \text{ in } X \qquad \Longrightarrow \qquad \mathcal{F}^0(u^0) \le \liminf_{\varepsilon \to 0} \mathcal{F}^\varepsilon(u^\varepsilon), \\ &\text{(G2S)} \qquad \forall u^0 \in X \;\exists\, u^\varepsilon \colon u^\varepsilon \to u^0 \text{ in } X \text{ and } \mathcal{F}^0(u^0) = \lim_{\varepsilon \to 0} \mathcal{F}^\varepsilon(u^\varepsilon). \end{aligned}$$

Moreover, we say that $\mathcal{F}^\varepsilon$ weakly Gamma converges to $\mathcal{F}^0$ and we write $\mathcal{F}^\varepsilon \overset{\Gamma}{\rightharpoonup} \mathcal{F}^0$ if

$$\begin{aligned} &\text{(G1W)} \qquad u^\varepsilon \rightharpoonup u^0 \text{ in } X \qquad \Longrightarrow \qquad \mathcal{F}^0(u^0) \le \liminf_{\varepsilon \to 0} \mathcal{F}^\varepsilon(u^\varepsilon), \\ &\text{(G2W)} \qquad \forall u^0 \in X \;\exists\, u^\varepsilon \colon u^\varepsilon \rightharpoonup u^0 \text{ in } X \text{ and } \mathcal{F}^0(u^0) = \lim_{\varepsilon \to 0} \mathcal{F}^\varepsilon(u^\varepsilon). \end{aligned}$$

Remark 4.5. Notice that (G2S) $\implies$ (G2W) and (G1W) $\implies$ (G1S).

Definition 4.6 (Mosco convergence)**.** We say that $\mathcal{F}^\varepsilon$ Mosco converges to $\mathcal{F}^0$ and we write $\mathcal{F}^\varepsilon \xrightarrow{\mathrm{M}} \mathcal{F}^0$ if

$$\mathcal{F}^\varepsilon \xrightarrow{\Gamma} \mathcal{F}^0 \qquad \text{and} \qquad \mathcal{F}^\varepsilon \xrightharpoonup{\Gamma} \mathcal{F}^0.$$

Remark 4.7. Notice that, in view of Remark 4.5, for Mosco convergence it is enough to check conditions (G2S) and (G1W).

4.1.3 Homogenization result

In this section we prove the homogenization of the BKE, but we first need some technical results. In the next two lemmas we prove the convergence of the functionals Ψ^ε and $\mathcal{E}^\varepsilon$ to Ψ^0 and $\mathcal{E}^0$, respectively.

Proposition 4.8. *Let Ψ^ε and Ψ^0 be defined in* (4.14) *and* (4.16) *on the space $L^2_{\varphi^0}(\mathbb{R})$. Then $\Psi^\varepsilon \xrightarrow{\mathrm{C}} \Psi^0$, where $\xrightarrow{\mathrm{C}}$ stands for continuous convergence given in Definition 4.2.*

Proof. Let v^ε be a sequence such that $v^\varepsilon \to v$ in $L^2_{\varphi^0}(\mathbb{R})$ as $\varepsilon \to 0$. Then, by the triangle inequality we have

$$\begin{aligned}\left|\frac{1}{2}\int_\mathbb{R} v^\varepsilon(x)^2\varphi^\varepsilon(x)\,\mathrm{d}x - \frac{1}{2}\int_\mathbb{R} v^0(x)^2\varphi^0(x)\,\mathrm{d}x\right| &\le \frac{1}{2}\left|\int_\mathbb{R} v^\varepsilon(x)^2\varphi^\varepsilon(x)\,\mathrm{d}x - \int_\mathbb{R} v^0(x)^2\varphi^\varepsilon(x)\,\mathrm{d}x\right| \\ &\quad + \frac{1}{2}\left|\int_\mathbb{R} v^0(x)^2\varphi^\varepsilon(x)\,\mathrm{d}x - \int_\mathbb{R} v^0(x)^2\varphi^0(x)\,\mathrm{d}x\right| \\ &=: \frac{1}{2}(I_1^\varepsilon + I_2^\varepsilon),\end{aligned} \tag{4.18}$$

and we now show that the two terms in the right-hand side vanish. First, due to equation (4.7) we obtain

$$I_1^\varepsilon \le \frac{C_{\nu^0}C_\pi}{C_{\nu^\varepsilon}}\int_\mathbb{R}\left|v^\varepsilon(x)^2 - v^0(x)^2\right|\left|\omega\left(\frac{x}{\varepsilon}\right)\right|\varphi^0(x)\,\mathrm{d}x,$$

and, since ω is smooth and periodic and hence bounded, by limit (4.8) and applying the Cauchy–Schwarz inequality we have for a constant $C > 0$ independent of ε

$$I_1^\varepsilon \le C\left\|v^\varepsilon - v^0\right\|_{L^2_{\varphi^0}(\mathbb{R})}\left(\left\|v^\varepsilon\right\|_{L^2_{\varphi^0}(\mathbb{R})} + \left\|v^0\right\|_{L^2_{\varphi^0}(\mathbb{R})}\right), \tag{4.19}$$

which implies that $I_1^\varepsilon \to 0$ as $\varepsilon \to 0$. We now consider I_2^ε and by equations (4.7) and (4.8) and the periodicity of p we get

$$\lim_{\varepsilon\to 0} I_2^\varepsilon = \lim_{\varepsilon\to 0}\left|\frac{C_{\nu^0}}{C_{\nu^\varepsilon}}\int_\mathbb{R} v^0(x)^2 e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)}\varphi^0(x)\,\mathrm{d}x - \int_\mathbb{R} v^0(x)^2\varphi^0(x)\,\mathrm{d}x\right| = 0,$$

which together with (4.18) and (4.19) yields that $\Psi^\varepsilon(v^\varepsilon) \to \Psi^0(v^0)$ as $\varepsilon \to 0$ and concludes the proof. $\square$

Proposition 4.9. *Let $\mathcal{E}^\varepsilon$ and $\mathcal{E}^0$ be defined in* (4.13) *and* (4.15) *on the space $L^2_{\varphi^0}(\mathbb{R})$. Then $\mathcal{E}^\varepsilon \xrightarrow{\mathrm{M}} \mathcal{E}^0$, where $\xrightarrow{\mathrm{M}}$ stands for Mosco convergence given in Definition 4.6.*

Proof. In order to prove Mosco convergence, as highlighted in Remark 4.7 we need to show two conditions, hence we divide the proof in two steps.

Step 1: liminf estimate. Let u^ε be a sequence such that $u^\varepsilon \rightharpoonup u^0$ in $L^2_{\varphi^0}(\mathbb{R})$ as $\varepsilon \to 0$ with $\alpha = \liminf_{\varepsilon \to 0} \mathcal{E}^\varepsilon(u^\varepsilon)$. If $\alpha = +\infty$ then the result is trivial. Otherwise, using equation (4.7) and since ω is such that there exists a constant $\beta > 0$ such that $\omega(y) \geq \beta$ for all $y \in \mathbb{R}$, we have for a constant $C > 0$ independent of ε

$$\|u^\varepsilon\|^2_{H^1_{\varphi^0}(\mathbb{R})} \leq C\mathcal{E}^\varepsilon(u^\varepsilon),$$

which implies that, up to a subsequence, u^ε is bounded in $H^1_{\varphi^0}(\mathbb{R})$. Therefore, there exists a convergent subsequence in $H^1_{\varphi^0}(\mathbb{R})$ and, since $u^\varepsilon \rightharpoonup u^0$ in $L^2_{\varphi^0}(\mathbb{R})$ and by the uniqueness of the limit, we deduce that $u^\varepsilon \rightharpoonup u^0$ in $H^1_{\varphi^0}(\mathbb{R})$ which yields $u^\varepsilon \to u^0$ and $\partial_x u^\varepsilon \rightharpoonup \partial_x u^0$ in $L^2_{\varphi^0}(\mathbb{R})$. Then, we have

$$\begin{aligned}
\mathcal{E}^\varepsilon(u^\varepsilon) &= \frac{1}{2}\sigma \int_{\mathbb{R}} (\partial_x u^\varepsilon(x))^2 \varphi^\varepsilon(x) \,\mathrm{d}x + \frac{1}{2}\lambda \int_{\mathbb{R}} u^\varepsilon(x)^2 \varphi^\varepsilon(x) \,\mathrm{d}x \\
&= \frac{\sigma C_{\nu^0} C_\pi}{2C_{\nu^\varepsilon}} \int_{\mathbb{R}} \left(\partial_x u^\varepsilon(x) - \frac{\mathcal{K}}{\mathbb{T}\omega\left(\frac{x}{\varepsilon}\right)} \partial_x u^0(x) \right)^2 \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x \\
&\quad + \frac{\sigma \mathcal{K} C_{\nu^0} C_\pi}{\mathbb{T} C_{\nu^\varepsilon}} \int_{\mathbb{R}} \partial_x u^\varepsilon(x) \partial_x u^0(x) \varphi^0(x) \,\mathrm{d}x - \frac{\sigma \mathcal{K}^2 C_{\nu^0} C_\pi}{2\mathbb{T}^2 C_{\nu^\varepsilon}} \int_{\mathbb{R}} \frac{1}{\omega\left(\frac{x}{\varepsilon}\right)} \partial_x u^0(x)^2 \varphi^0(x) \,\mathrm{d}x \\
&\quad + \frac{\lambda C_{\nu^0} C_\pi}{2C_{\nu^\varepsilon}} \int_{\mathbb{R}} u^\varepsilon(x)^2 \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x \\
&=: I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon + I_4^\varepsilon,
\end{aligned} \tag{4.20}$$

and we now study the four terms in the right-hand side separately. By equations (1.12) and (4.8), the weak convergence of $\partial_x u^\varepsilon$ to $\partial_x u^0$ in $L^2_{\varphi^0}(\mathbb{R})$ and the periodicity of ω we obtain

$$\begin{aligned}
&I_1^\varepsilon \geq 0, \\
&\lim_{\varepsilon \to 0} I_2^\varepsilon = \mathcal{K}\sigma \int_{\mathbb{R}} (\partial_x u^0(x))^2 \varphi^0(x) \,\mathrm{d}x, \\
&\lim_{\varepsilon \to 0} I_3^\varepsilon = -\frac{1}{2}\mathcal{K}\sigma \int_{\mathbb{R}} (\partial_x u^0(x))^2 \varphi^0(x) \,\mathrm{d}x.
\end{aligned} \tag{4.21}$$

Moreover, by the strong convergence of u^ε to u^0 in $L^2_{\varphi^0}(\mathbb{R})$ and following an argument similar to the proof of Proposition 4.8 we deduce

$$\lim_{\varepsilon \to 0} I_4^\varepsilon = \frac{1}{2}\lambda \int_{\mathbb{R}} u^0(x)^2 \varphi^0(x) \,\mathrm{d}x,$$

which together with (4.20) and (4.21) gives

$$\liminf_{\varepsilon \to 0} \mathcal{E}^\varepsilon(u^\varepsilon) \geq \mathcal{E}^0(u^0),$$

which is the desired liminf estimate.

Step 2: recovery sequence. Let $u^0 \in H^1_{\varphi^0}(\mathbb{R})$ and define the sequence u^ε such that $u^\varepsilon(0) = u^0(0)$ and

$$\partial_x u^\varepsilon(x) = \frac{\mathcal{K}}{\mathbb{T}\omega\left(\frac{x}{\varepsilon}\right)} \partial_x u^0(x).$$

Notice that by the periodicity of ω and equation (1.12) we have $u^\varepsilon \rightharpoonup u^0$ in $H^1_{\varphi^0}(\mathbb{R})$, which implies that $u^\varepsilon \to u^0$ in $L^2_{\varphi^0}(\mathbb{R})$. Then, reasoning similarly to Step 1 we deduce that

$$\lim_{\varepsilon \to 0} \mathcal{E}^\varepsilon(u^\varepsilon) = \mathcal{E}^0(u^0),$$

which shows that u^ε is the desired recovery sequence and completes the proof. □

We are now ready to state and prove the homogenization result.

Theorem 4.10. *Let $u^\varepsilon, u^0 \colon [0,T] \to L^2_{\varphi^0}(\mathbb{R})$ be the solutions of the BKEs* (4.9) *and* (4.10)*, respectively. Then for all $t > 0$ and as $\varepsilon \to 0$ it holds*

$$u^\varepsilon(t) \rightharpoonup u^0(t) \quad \text{in } H^1_{\varphi^0}(\mathbb{R}).$$

Proof. Let us consider the functions $w^\varepsilon, w^0 \colon [0,T] \to L^2_{\varphi^0}(\mathbb{R})$ defined as $w^\varepsilon(t) = e^{\lambda t} u^\varepsilon(t)$ and $w^0(t) = e^{\lambda t} u^0(t)$ which solve equations (4.11) and (4.12) and recall that their gradient flow formulation is given by (4.17) with functionals $\mathcal{E}^\varepsilon$, Ψ^ε, $\mathcal{E}^0$, Ψ^0 defined in (4.13), (4.14), (4.15) and (4.16). Then, by equations (4.7) and (4.8) we deduce that there exist three constants $C_1, C_2, C_3 > 0$ independent of ε such that

$$\mathcal{E}^\varepsilon(u) \geq C_1 \left\| u \right\|^2_{H^1_{\varphi^0}(\mathbb{R})},$$

and

$$C_2 \left\| v \right\|^2_{L^2_{\varphi^0}(\mathbb{R})} \leq \Psi^\varepsilon(v) \leq C_3 \left\| v \right\|^2_{L^2_{\varphi^0}(\mathbb{R})}.$$

Moreover, under conditions (4.4) in Assumption 4.1 and by [5, Propostion A.4] it follows that $H^1_{\varphi^0}(\mathbb{R})$ is compactly embedded into $L^2_{\varphi^0}(\mathbb{R})$. Therefore, due to Propositions 4.8 and 4.9 and applying [87, Theorem 3.4.1] we obtain for all $t > 0$ and as $\varepsilon \to 0$

$$w^\varepsilon(t) \rightharpoonup w^0(t) \quad \text{in } H^1_{\varphi^0}(\mathbb{R}),$$

which, recalling that $u^\varepsilon(t) = e^{-\lambda t} w^\varepsilon(t)$ and $u^0(t) = e^{-\lambda t} w^0(t)$, gives the desired result. □

4.2 Rate of weak convergence of multiscale Langevin dynamics

In this section we aim to compute a rate for the weak convergence of the invariant measure of the multiscale Langevin dynamics towards the corresponding homogenized invariant measure. We first recall that in [103, Proposition 5.2] it is proved that measure ν^ε with density φ^ε in (4.5) weakly converges to the measure ν^0 with density φ^0 in (4.6) by showing that $\varphi^\varepsilon \rightharpoonup \varphi^0$ in $L^1(\mathbb{R})$. Hence, we know that for all bounded and continuous functions $f \in \mathcal{C}^0_b(\mathbb{R})$

$$\lim_{\varepsilon \to 0} \mathbb{E}^{\varphi^\varepsilon}[f(X)] = \mathbb{E}^{\varphi^0}[f(X)]. \tag{4.22}$$

We still work under Assumption 4.1 and we are now interested in studying the rate of convergence of (4.22), i.e., we want to find $r > 0$ and a constant $C > 0$ independent of r such that

$$\left| \mathbb{E}^{\varphi^\varepsilon}[f(X)] - \mathbb{E}^{\varphi^0}[f(X)] \right| \leq C\varepsilon^r.$$

We first need to study some technical results on periodic functions which will be employed later in the main theorem.

4.2.1 Periodic functions

The next lemma provides a condition which guarantees that the primitive of a periodic function is periodic as well. This result will be used to prove Lemma 4.12, which is crucial in computing the rate of weak convergence, which is our main goal.

Lemma 4.11. *Let $u\colon \mathbb{R} \to \mathbb{R}$ be a periodic function with period $\mathbb{T} > 0$. If*

$$\frac{1}{\mathbb{T}} \int_0^{\mathbb{T}} u(y) \, \mathrm{d}y = 0, \tag{4.23}$$

then any primitive $U\colon \mathbb{R} \to \mathbb{R}$ of u is periodic with the same period $\mathbb{T}$. Moreover, U can be chosen such that

$$\frac{1}{\mathbb{T}} \int_0^{\mathbb{T}} U(y) \, \mathrm{d}y = 0. \tag{4.24}$$

Proof. The general form of the primitive U is

$$U(y) = \int_0^y u(t) \, \mathrm{d}t + c, \qquad \text{with } c \in \mathbb{R}.$$

Then, by hypothesis (4.23) and since u is periodic with period $\mathbb{T}$ we have

$$U(y + \mathbb{T}) = \int_0^{y+\mathbb{T}} u(t) \, \mathrm{d}t + c = \int_0^{\mathbb{T}} u(t) \, \mathrm{d}t + \int_{\mathbb{T}}^{\mathbb{T}+y} u(t) \, \mathrm{d}t + c = \int_0^y u(t) \, \mathrm{d}t + c = U(y),$$

which shows that U is periodic with period $\mathbb{T}$. Finally, choosing

$$c = -\frac{1}{\mathbb{T}} \int_0^{\mathbb{T}} \int_0^y u(t) \, \mathrm{d}t \, \mathrm{d}y$$

gives condition (4.24) and completes the proof. □

Lemma 4.12. *Let $g \in \mathcal{C}_b^k(\mathbb{R})$ and $u\colon \mathbb{R} \to \mathbb{R}$ be a periodic function with period $\mathbb{T} > 0$ and such that*

$$\frac{1}{\mathbb{T}} \int_0^{\mathbb{T}} u(y) \, \mathrm{d}y = 0.$$

Then there exists $C > 0$ independent of ε such that

$$\left| \int_{\mathbb{R}} g(x) e^{-\frac{1}{\sigma} \mathcal{V}(x)} u\left(\frac{x}{\varepsilon}\right) \mathrm{d}x \right| \leq C \varepsilon^k.$$

Proof. Integrating by parts and noticing that the boundary term vanishes we have

$$\int_{\mathbb{R}} g(x) e^{-\frac{1}{\sigma} \mathcal{V}(x)} u\left(\frac{x}{\varepsilon}\right) \mathrm{d}x = -\varepsilon \int_{\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}x} \left(g(x) e^{-\frac{1}{\sigma} \mathcal{V}(x)} \right) U\left(\frac{x}{\varepsilon}\right) \mathrm{d}x,$$

where, due to Lemma 4.11, the primitive U is periodic and has zero mean over the period. Moreover, periodicity and continuity give the boundedness of U. Iterating this procedure k times yields

$$\int_{\mathbb{R}} g(x) e^{-\frac{1}{\sigma} \mathcal{V}(x)} u\left(\frac{x}{\varepsilon}\right) \mathrm{d}x = (-1)^k \varepsilon^k \int_{\mathbb{R}} \frac{\mathrm{d}^k}{\mathrm{d}x^k} \left(g(x) e^{-\frac{1}{\sigma} \mathcal{V}(x)} \right) \mathcal{U}_k\left(\frac{x}{\varepsilon}\right) \mathrm{d}x,$$

where $\mathcal{U}_k$ is the k-th antiderivative of u obtained choosing at each step the primitive with zero mean over the period and it is therefore bounded. It now remains to compute the derivative of order k inside the integral and using the general Leibniz rule and the Faà di Bruno's formula we get

$$\begin{aligned}
\frac{\mathrm{d}^k}{\mathrm{d}x^k} \left(g(x) e^{-\frac{1}{\sigma} \mathcal{V}(x)} \right) &= \sum_{n=0}^k \binom{k}{n} g^{(k-n)}(x) \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left(e^{-\frac{1}{\sigma} \mathcal{V}(x)} \right) \\
&= e^{-\frac{1}{\sigma} \mathcal{V}(x)} \sum_{n=0}^k \binom{k}{n} g^{(k-n)}(x) \sum_{i=0}^n \frac{1}{i!} \left(\frac{1}{\sigma} \right)^i \sum_{j=0}^i (-1)^j \binom{i}{j} \mathcal{V}(x)^{i-j} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left(\mathcal{V}(x)^j \right).
\end{aligned}$$

Hence, there exists a constant $\widetilde{C} > 0$ independent of ε such that

$$\left| \int_{\mathbb{R}} g(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} u\left(\frac{x}{\varepsilon}\right) \mathrm{d}x \right| \le \widetilde{C}\varepsilon^k \sum_{n=0}^{k}\sum_{i=0}^{n}\sum_{j=0}^{i} \int_{\mathbb{R}} \left| g^{(k-n)}(x) \right| \left| \mathcal{V}(x) \right|^{i-j} \left| \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left(\mathcal{V}(x)^j\right) \right| e^{-\frac{1}{\sigma}\mathcal{V}(x)} \mathrm{d}x,$$

and defining

$$C = \widetilde{C} \sum_{n=0}^{k}\sum_{i=0}^{n}\sum_{j=0}^{i} \int_{\mathbb{R}} \left| g^{(k-n)}(x) \right| \left| \mathcal{V}(x) \right|^{i-j} \left| \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left(\mathcal{V}(x)^j\right) \right| e^{-\frac{1}{\sigma}\mathcal{V}(x)} \mathrm{d}x$$

gives the desired result. □

Remark 4.13. Notice that similar results can be proven in the multidimensional case, i.e., when $\mathbb{R}$ is replaced by $\mathbb{R}^d$ and the interval $[0, \mathbb{T}]$ is replaced by the hypercube $\mathcal{T} = [0, \mathbb{T}]^d$. In particular, in Lemma 4.11 we consider the vector $U \colon \mathcal{T} \to \mathbb{R}^d$ such that $\operatorname{div}_y U = u$ with components

$$\begin{aligned} U_1(y) &= \int_0^{y_1} u(t, y_2, \dots, y_d) \,\mathrm{d}t - \frac{1}{\mathbb{T}^d} \int_{\mathcal{T}} \int_0^{y_1} u(t, y_2, \dots, y_d) \,\mathrm{d}t \,\mathrm{d}y, \\ U_i(y) &= 0 \qquad \text{for all } i = 2, \dots, d, \end{aligned}$$

and observe that it is periodic with period $\mathbb{T}$ in all directions and it satisfies

$$\frac{1}{\mathbb{T}^d} \int_{\mathcal{T}} U(y) \,\mathrm{d}y = 0.$$

Moreover, in Lemma 4.12 we consider functions $g \in \mathcal{C}_b^k(\mathbb{R}^d)$ and the proof remains the same if all the derivatives with respect to x are replaced by the partial derivatives with respect to the first component x_1, in fact the following integration by parts holds

$$\begin{aligned} \int_{\mathbb{R}^d} g(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} u\left(\frac{x}{\varepsilon}\right) \mathrm{d}x &= \varepsilon \int_{\mathbb{R}^d} g(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} \operatorname{div}_x \left(U\left(\frac{x}{\varepsilon}\right)\right) \mathrm{d}x \\ &= -\varepsilon \int_{\mathbb{R}^d} \nabla \left(g(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} \right) \cdot U\left(\frac{x}{\varepsilon}\right) \mathrm{d}x \\ &= -\varepsilon \int_{\mathbb{R}^d} \frac{\partial}{\partial x_1} \left(g(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} \right) U_1\left(\frac{x}{\varepsilon}\right) \mathrm{d}x, \end{aligned}$$

which finally yields the rate of convergence ε^k.

4.2.2 Main result

We are now ready to state the main result of this section, i.e., the rate of convergence of the limit in (4.22).

Theorem 4.14. *Let $f \in \mathcal{C}_b^k(\mathbb{R})$. Then there exists a constant $C > 0$ independent of ε such that*

$$\left| \mathbb{E}^{\varphi^\varepsilon}[f(X)] - \mathbb{E}^{\varphi^0}[f(X)] \right| \le C\varepsilon^k.$$

Proof. By definition of φ^ε and φ^0 in (4.5) and (4.6) we obtain

$$\begin{aligned} \left| \mathbb{E}^{\varphi^\varepsilon}[f(X)] - \mathbb{E}^{\varphi^0}[f(X)] \right| &= \left| \int_{\mathbb{R}} f(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} \left(\frac{1}{C_{\nu^\varepsilon}} e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)} - \frac{1}{C_{\nu^0}} \right) \mathrm{d}x \right| \\ &\le \frac{1}{C_{\nu^\varepsilon}} \left| \int_{\mathbb{R}} f(x) e^{-\frac{1}{\sigma}\mathcal{V}(x)} \left(e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)} - \frac{C_\pi}{\mathbb{T}} \right) \mathrm{d}x \right| \\ &\quad + \frac{1}{C_{\nu^\varepsilon}} \left| \frac{C_\pi C_{\nu^0}}{\mathbb{T}} - C_{\nu^\varepsilon} \right| \left| \int_{\mathbb{R}} f(x) \frac{1}{C_{\nu^0}} e^{-\frac{1}{\sigma}\mathcal{V}(x)} \mathrm{d}x \right| \\ &=: I_1^\varepsilon + I_2^\varepsilon. \end{aligned} \tag{4.25}$$

We now consider the second term in the right-hand side and we have

$$\left| \frac{C_\pi C_{\nu^0}}{\mathbb{T}} - C_{\nu^\varepsilon} \right| = \left| \int_{\mathbb{R}} e^{-\frac{1}{\sigma}\mathcal{V}(x)} \left(\frac{C_\pi}{\mathbb{T}} - e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)} \right) \mathrm{d}x \right|,$$

then we define the function

$$u(y) = \frac{C_\pi}{\mathbb{T}} - e^{-\frac{1}{\sigma}p(y)},$$

and observe that

$$\frac{1}{\mathbb{T}} \int_0^{\mathbb{T}} u(y) \,\mathrm{d}y = 0.$$

Therefore, applying Lemma 4.12 we deduce the existence of a constant $\widetilde{C}_2 > 0$ independent of ε such that

$$\left| \frac{C_\pi C_{\nu^0}}{\mathbb{T}} - C_{\nu^\varepsilon} \right| \leq \widetilde{C}_2 \varepsilon^k,$$

which together with the boundedness of the sequence $\{C_{\nu^\varepsilon}\}$ due to (4.8) yields

$$|I_2^\varepsilon| \leq C_2 \varepsilon^k, \tag{4.26}$$

for a constant $C_2 > 0$ independent of ε. We now study the term I_1 and, similarly as above, applying Lemma 4.12 and due to the boundedness of the sequence $\{C_{\nu^\varepsilon}\}$ we obtain

$$|I_1^\varepsilon| \leq C_1 \varepsilon^k, \tag{4.27}$$

for a constant $C_1 > 0$ independent of ε. Finally, employing estimates (4.26) and (4.27) together with bound (4.25) and defining $C = \max\{C_1, C_2\}$ give the desired result. □

Remark 4.15. Notice that Theorem 4.14 can be extended due to Remark 4.13 to the multidimensional setting ($f \in \mathcal{C}_b^k(\mathbb{R}^d)$ with $d > 1$). Moreover, the result easily generalizes to the case where the function f takes values in $\mathbb{R}^L$ with $L > 1$. In both cases we obtain the same rate of convergence ε^k.

4.3 Drift estimator employing Stratonovich integral

We consider the one-dimensional multiscale and homogenized SDEs (1.10) and (1.11) with $L = 1$ for simplicity, i.e., $\alpha, A \in \mathbb{R}$. We still aim to correctly estimate the drift coefficient A of the effective equation (1.11) given a trajectory $(X^\varepsilon(t))_{t\in[0,T]}$ of the multiscale dynamics (1.10). In this section we show how one can modify the MLE in order to make it asymptotically unbiased without employing filtered data as done in Chapters 2 and 3. We consider the Stratonovich formulation of the MLE given in [41]

$$\widehat{A}_{\mathrm{Strat}}(X^\varepsilon, T) = \frac{-\int_0^T V'(X^\varepsilon(t)) \circ \mathrm{d}X^\varepsilon(t) + \Sigma \int_0^T V''(X^\varepsilon(t)) \,\mathrm{d}t}{\int_0^T V'(X^\varepsilon(t))^2 \,\mathrm{d}t}, \tag{4.28}$$

and we show that it is asymptotically unbiased in the limit of infinite data. We remark that, even if this technique is simple to implement and does not depend on hyperparameters, it requires the knowledge of the diffusion coefficient Σ of the homogenized equation (1.11). Therefore, in case Σ is known one should employ this estimator rather than using the estimators of Chapters 2 and 3, as this approach is simpler and does not requires the computation of the filtered data. However, this assumption is not realistic in most applications, where the diffusion coefficient is in general unknown and hence our estimators with filtered data should be employed.

4.3.1 Asymptotic unbiasedness

In the following theorem we prove that the estimator $\widehat{A}_{\text{Strat}}(X^\varepsilon, T)$ defined in (4.28) is asymptotically unbiased in the limit of infinite trajectory ($T \to \infty$) and when the multiscale parameter vanishes ($\varepsilon \to 0$). As usual, for economy of notation, we will simply write X_t^ε instead of $X^\varepsilon(t)$.

Theorem 4.16. *Under Assumption 1.4 it holds*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}_{\text{Strat}}(X^\varepsilon, T) = A, \qquad a.s.$$

Proof. Replacing equation (1.10) in the definition (4.28) we have

$$\begin{aligned} \widehat{A}_{\text{Strat}}(X^\varepsilon, T) &= \alpha + \frac{\frac{1}{T}\int_0^T V'(X_t^\varepsilon)\frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\,\mathrm{d}t}{\frac{1}{T}\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t} - \frac{\frac{\sqrt{2\sigma}}{T}\int_0^T V'(X_t^\varepsilon)\circ\mathrm{d}W_t}{\frac{1}{T}\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t} + \frac{\frac{\Sigma}{T}\int_0^T V''(X_t^\varepsilon)\,\mathrm{d}t}{\frac{1}{T}\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t} \\ &=: \alpha + I_1(\varepsilon, T) - I_2(\varepsilon, T) + I_3(\varepsilon, T), \end{aligned} \tag{4.29}$$

and we now consider the three terms in the right-hand side separately. First, following the proof of [103, Theorem 3.4] we obtain

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} I_1(\varepsilon, T) = 0, \qquad a.s. \tag{4.30}$$

Then, converting the Stratonovich integral into the Itô integral we have

$$\begin{aligned} I_2(\varepsilon, T) &= \frac{\frac{\sqrt{2\sigma}}{T}\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}W_t}{\frac{1}{T}\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t} + \frac{\frac{\sigma}{T}\int_0^T V''(X_t^\varepsilon)\,\mathrm{d}t}{\frac{1}{T}\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t} \\ &=: I_2^1(\varepsilon, T) + I_2^2(\varepsilon, T), \end{aligned} \tag{4.31}$$

and following the proof of [103, Theorem 3.4] we obtain

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} I_2^1(\varepsilon, T) = 0, \qquad a.s. \tag{4.32}$$

Then, the ergodic theorem and the homogenization theory give

$$\begin{aligned} \lim_{\varepsilon \to 0} \lim_{T \to \infty} I_2^2(\varepsilon, T) &= \sigma \frac{\mathbb{E}^{\nu^0}[V''(X)]}{\mathbb{E}^{\nu^0}[V'(X)^2]}, \qquad a.s., \\ \lim_{\varepsilon \to 0} \lim_{T \to \infty} I_3(\varepsilon, T) &= \Sigma \frac{\mathbb{E}^{\nu^0}[V''(X)]}{\mathbb{E}^{\nu^0}[V'(X)^2]}, \qquad a.s., \end{aligned}$$

and, due to the invariant measure ν^0 with density φ^0 defined in (2.13), an integration by parts yields

$$\mathbb{E}^{\nu^0}[V''(X)] = \frac{A}{\Sigma}\,\mathbb{E}^{\nu^0}[V'(X)^2] = \frac{\alpha}{\sigma}\,\mathbb{E}^{\nu^0}[V'(X)^2].$$

Therefore, we have

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} I_2^2(\varepsilon, T) = \alpha, \qquad a.s., \qquad \text{and} \qquad \lim_{\varepsilon \to 0} \lim_{T \to \infty} I_3(\varepsilon, T) = A, \qquad a.s.,$$

which together with decompositions (4.29) and (4.31) and limits (4.30) and (4.32) give the desired result. ∎

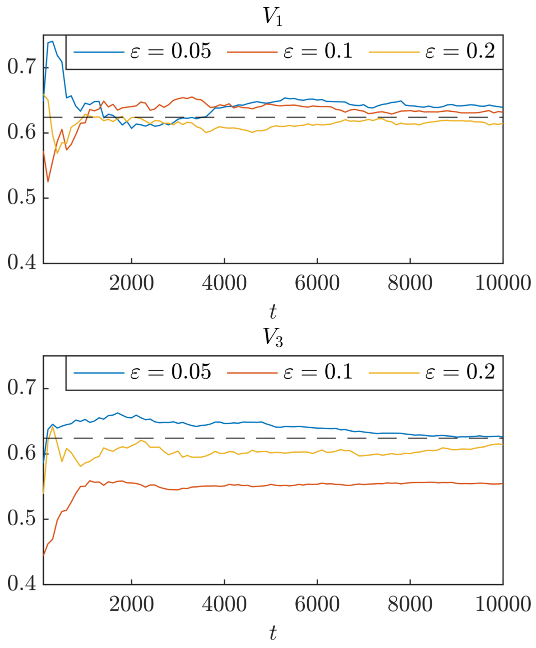


Figure 4.1 – Numerical results for the drift estimator $\widehat{A}_{\text{Strat}}(X^\varepsilon, T)$. The dotted line represents the exact drift coefficient A.

4.3.2 Numerical experiment

We now present numerical experiments which confirm our theoretical result. We consider equation (1.10) and we set the drift coefficient $\alpha = 1$, the diffusion coefficient $\sigma = 1$ and the fast-scale potential $p(y) = \cos(y)$. We then study four different slow-scale potentials

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

and we generate data up to the final time $T = 10^4$ employing the Euler–Maruyama (EM) method with a fine time step, in particular we set $h = \varepsilon^3$. In Figure 4.1 we show how the estimation evolves with respect to time for different values of the multiscale parameter $\varepsilon = 0.05, 0.1, 0.2$. We observe that the results stabilize for large values of the final time T, and that they improve when the multiscale parameter ε is smaller.

4.4 Iterated exponential filter

In this section we present a possible development of the filtering methodology presented in Chapter 2 for inferring the drift coefficient of the homogenized equation (1.11) given continuous observations from the multiscale system (1.10). Since we noticed that using exponentially filtered data was beneficial for the unbiasedness of the estimator, the idea is now to repeatedly apply the filter to the original data multiple times, in order to obtain new smoother data. We remark that this is still a work in progress and rigorous theoretical results have not been proven yet. However, preliminary numerical experiments suggest that this approach could provide estimators which are unbiased independently of the length of the filtering width, and therefore more robust.

Consider the filter (2.1) with $\beta = 1$

$$k_{\exp}^{\delta,1}(r) = \frac{1}{\delta} e^{-r/\delta},$$

and for a positive integer $\mathscr{N}$ let the process $Z_{\mathscr{N}}^{\varepsilon} := (Z_{\mathscr{N}}^{\varepsilon}(t), 0 \le t \le T)$ be defined by

$$Z_{\mathscr{N}}^{\varepsilon}(t) := \int_0^t k(t - s_1) \int_0^{s_1} k(s_1 - s_2) \cdots \int_0^{s_{\mathscr{N}-1}} k(s_{\mathscr{N}-1} - s_{\mathscr{N}}) X_{s_{\mathscr{N}}}^{\varepsilon} \,\mathrm{d}s_{\mathscr{N}} \cdots \mathrm{d}s_2 \,\mathrm{d}s_1. \tag{4.33}$$

The filtered process can be rewritten in different fashions. In particular, noticing that

$$k_{\exp}^{\delta,1}(t-s) k_{\exp}^{\delta,1}(s-r) = \frac{1}{\delta} k_{\exp}^{\delta,1}(t-r),$$

we can write

$$Z_{\mathscr{N}}^{\varepsilon}(t) = \frac{1}{\delta^{\mathscr{N}-1}} \int_0^t \int_0^{s_1} \cdots \int_0^{s_{\mathscr{N}-1}} k_{\exp}^{\delta,1}(t - s_{\mathscr{N}}) X_{s_{\mathscr{N}}}^{\varepsilon} \,\mathrm{d}s_{\mathscr{N}} \cdots \mathrm{d}s_2 \,\mathrm{d}s_1. \tag{4.34}$$

Then, applying Cauchy's iterated integral rule to (4.34), we obtain

$$Z_{\mathscr{N}}^{\varepsilon}(t) = \frac{1}{\delta^{\mathscr{N}-1}(\mathscr{N}-1)!} \int_0^t (t-s)^{\mathscr{N}-1} k_{\exp}^{\delta,1}(t-s) X_s^{\varepsilon} \,\mathrm{d}s.$$

This shows that filtering iteratively the data corresponds to a single application of a different filter, which we call $k_{\exp}^{\delta,\mathscr{N}}$ and which is defined as

$$k_{\exp}^{\delta,\mathscr{N}}(r) = \frac{1}{\delta^{\mathscr{N}}(\mathscr{N}-1)!} r^{\mathscr{N}-1} e^{-r/\delta}. \tag{4.35}$$

Moreover, we showed in Chapter 2 that the original trajectory $X^{\varepsilon}(t)$ and the single-filtered process $Z_1^{\varepsilon}(t)$ satisfy the system of SDEs

$$\begin{aligned}
\mathrm{d}X^{\varepsilon}(t) &= -\alpha \cdot V'(X^{\varepsilon}(t)) \,\mathrm{d}t - \frac{1}{\varepsilon} p' \left(\frac{X^{\varepsilon}(t)}{\varepsilon} \right) \mathrm{d}t + \sqrt{2\sigma} \,\mathrm{d}W(t), \\
\mathrm{d}Z_1^{\varepsilon}(t) &= \frac{1}{\delta} \left(X^{\varepsilon}(t) - Z_1^{\varepsilon}(t) \right) \mathrm{d}t.
\end{aligned}$$

Clearly, with the same argument we have that recursively it holds

$$\begin{aligned}
\mathrm{d}X^{\varepsilon}(t) &= -\alpha \cdot V'(X^{\varepsilon}(t)) \,\mathrm{d}t - \frac{1}{\varepsilon} p' \left(\frac{X^{\varepsilon}(t)}{\varepsilon} \right) \mathrm{d}t + \sqrt{2\sigma} \,\mathrm{d}W(t), \\
\mathrm{d}Z_n^{\varepsilon}(t) &= \frac{1}{\delta} \left(Z_{n-1}^{\varepsilon}(t) - Z_n^{\varepsilon}(t) \right) \mathrm{d}t, \qquad n = 1, \dots, \mathscr{N},
\end{aligned} \tag{4.36}$$

with the initialization $Z_0^{\varepsilon}(t) \equiv X^{\varepsilon}(t)$, and therefore the process $(X^{\varepsilon}(t), Z_1^{\varepsilon}(t), \dots, Z_{\mathscr{N}}^{\varepsilon}(t))$ satisfy a system of $\mathscr{N} + 1$ SDEs. The corresponding stationary Fokker–Planck equation (FPE) for the density $\rho_{\exp}^{\varepsilon,\mathscr{N}}(x, z)$ of the invariant measure $\mu_{\exp}^{\varepsilon,\mathscr{N}}(x, z)$ is given by

$$\begin{aligned}
\sigma \partial_{xx}^2 \rho_{\exp}^{\varepsilon,\mathscr{N}}(x,z) &+ \partial_x \left(\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p' \left(\frac{x}{\varepsilon} \right) \right) \rho_{\exp}^{\varepsilon,\mathscr{N}}(x,z) \right) \\
&+ \frac{1}{\delta} \partial_{z_1} \left((z_1 - x) \rho_{\exp}^{\varepsilon,\mathscr{N}}(x,z) \right) + \frac{1}{\delta} \sum_{n=2}^{N} \partial_{z_n} \left((z_n - z_{n-1}) \rho_{\exp}^{\varepsilon,\mathscr{N}}(x,z) \right) = 0,
\end{aligned} \tag{4.37}$$

where $z = (z_1, \dots, z_N)$. Finally, the new estimator $\widehat{A}_{\exp}^{\delta,\mathscr{N}}(X^{\varepsilon}, T)$ is obtained in the same way as $\widehat{A}_{\exp}^{\delta,1}(X^{\varepsilon}, T)$, by replacing the data Z^{ε} filtered once with the data $Z_{\mathscr{N}}^{\varepsilon}$ filtered $\mathscr{N}$ times. In the following we will simply write X_t^{ε} and $(Z_{\mathscr{N}}^{\varepsilon})_t$ instead of $X^{\varepsilon}(t)$ and $Z_{\mathscr{N}}^{\varepsilon}(t)$, respectively, and similarly for all stochastic processes.

4.4.1 Statement of main results

We first present a negative result which corresponds to Theorem 2.18 for the exponential filter. In particular, we show that even if we apply the filter $\mathscr{N}$ times, if the filtering width is sufficiently small, then the estimator is asymptotically biased and converges to the drift coefficient α of the multiscale equation (1.10). We remark that the proof of Theorem 4.17 requires some technical results, contained in Lemmas 4.28 to 4.30, which for convenience are postponed to Section 4.4.3.

Theorem 4.17. *If $\mathscr{N} = \lfloor \delta^{-\gamma} \rceil$, where $\lfloor \cdot \rceil$ denotes rounding to the nearest integer, with $\gamma \in [0, 1)$ and if $\delta = \varepsilon^\zeta$ with $\zeta > 2/(1-\gamma)$, it holds*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}^{\delta, \mathscr{N}}_{\exp}(X^\varepsilon, T) = \alpha, \quad a.s.,$$

where α is the drift coefficient of the multiscale equation (1.10)*.*

Proof. By the proof of Theorem 2.18, we have that

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

where it holds

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

independently of ε and where $I_1^\varepsilon(T)$ satisfies

$$\lim_{T \to \infty} I_1^\varepsilon(T) = (\widetilde{\mathcal{M}}^{\varepsilon, \mathscr{N}}_{\exp})^{-1} \left(J_1^\varepsilon + J_2^\varepsilon \right),$$

with

$$\begin{aligned} \widetilde{\mathcal{M}}^{\varepsilon, \mathscr{N}}_{\exp} &:= \mathbb{E}^{\mu^{\varepsilon, \mathscr{N}}_{\exp}} \left[V'(X^\varepsilon) \otimes V'(Z^\varepsilon_{\mathscr{N}}) \right], \\ J_1^\varepsilon &:= \mathbb{E}^{\mu^{\varepsilon, \mathscr{N}}_{\exp}} \left[\frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon}{\varepsilon} \right) V'(X^\varepsilon) \right], \\ J_2^\varepsilon &:= \mathbb{E}^{\mu^{\varepsilon, \mathscr{N}}_{\exp}} \left[\frac{1}{\varepsilon} p' \left(\frac{X^\varepsilon}{\varepsilon} \right) V''(\widetilde{X}^\varepsilon)(Z^\varepsilon_{\mathscr{N}} - X^\varepsilon) \right], \end{aligned}$$

where $\widetilde{X}^\varepsilon$ lays between X^ε and $Z^\varepsilon_{\mathscr{N}}$. A similar reasoning as in the proof of Theorem 2.18 allows to conclude that

$$\lim_{\varepsilon \to 0} J_1^\varepsilon = 0, \tag{4.38}$$

and an application of the Hölder's inequality, Lemmas 4.28 and 4.29 below yields

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

Let us remark that since $\mathscr{N} = \lfloor \delta^{-\gamma} \rceil \leq C \delta^{-\gamma}$ with $\gamma \in [0, 1)$ we have that the quantity $(\mathscr{N} - 1)\delta$ is bounded and therefore Lemma 4.29 holds. We then obtain

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

which, since by hypothesis $\delta = \varepsilon^\zeta$ with $\zeta > 2/(1-\gamma)$, gives

$$\lim_{\varepsilon \to 0} \| J_2^\varepsilon \| = 0. \tag{4.39}$$

Finally, Lemma 4.30 together with (4.38) and (4.39) gives that $I_1^\varepsilon(T) \to 0$ for $T \to \infty$ and $\varepsilon \to 0$, which in turn implies the desired result. $\square$

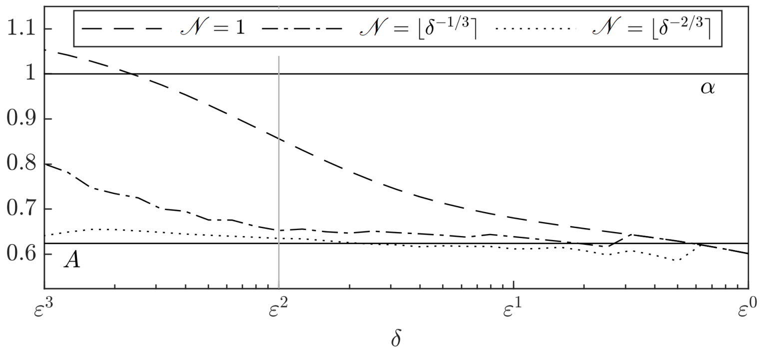


Figure 4.2 – On the x axis, the value δ of the filtering width. The horizontal solid lines represent the drift coefficients α and A of (1.10) and (1.11), respectively. The dashed, dash-dotted and dotted lines represent the value of $\widehat{A}^{\delta,\mathscr{N}}_{\exp}(X^\varepsilon, T)$, with $\mathscr{N} = \lfloor \delta^{-\gamma} \rceil$ and $\gamma = \{0, 1/3, 2/3\}$, respectively.

Inspired by the fast switch between two completely different regimes for the filter in Chapter 2 and based on the previous theorem, we propose the following conjecture. Initial steps which we believe are useful to prove it and which are similar to some results in Chapter 2, are presented in Section 4.4.3. Even if we have not been able to prove the conjecture yet, the numerical experiment in the next section seems to indicate that it holds true.

Conjecture 4.18. *If $\mathscr{N} = \lfloor \delta^{-\gamma} \rceil$, where $\lfloor \cdot \rceil$ denotes rounding to the nearest integer, with $\gamma \in [0, 1)$ and if $\delta = \varepsilon^\zeta$ with $0 \le \zeta < 2/(1-\gamma)$, it holds*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}^{\delta,\mathscr{N}}_{\exp}(X^\varepsilon, T) = A, \quad a.s.,$$

where A is the drift coefficient of the homogenized equation (1.11)*.*

Remark 4.19. Notice that by letting $\gamma \to 1$, then it is possible to take $\zeta \to \infty$, which in turn implies $\delta \to 0$ and $\mathscr{N} \to \infty$. Therefore, the smaller the filtering width is, the more we have to apply the filter to the data. This means that, by iterating the filtering procedure, it would be possible to filter at any regime as long as the number of iterations is sufficiently large.

4.4.2 Numerical experiment

We consider $V(x) = x^2/2$, so that the solution of the homogenized equation (1.11) is a Ornstein–Uhlenbeck (OU) process. We then choose the fluctuating potential as $p(y) = \cos(y)$ and the multiscale parameter $\varepsilon = 0.1$. We generate data $X^\varepsilon = (X^\varepsilon_t, 0 < t \le 10^3)$ from the multiscale equation, sampled at a high frequency. We then compute $Z^\varepsilon_{\mathscr{N}}$, with $\mathscr{N} = \lfloor \delta^{-\gamma} \rceil$ and $\gamma \in \{0, 1/3, 2/3\}$, and the estimator $\widehat{A}^{\delta,\mathscr{N}}_{\exp}(X^\varepsilon, T)$. Let us remark that for $\gamma = 0$ one gets $\mathscr{N} = 1$ and therefore the same framework as Chapter 2. Finally, we fix $\delta = \varepsilon^\zeta$, varying $\zeta \in (0, 3)$. Results, given in Figure 4.2, show the benifits of the iterated approach. Indeed, where for $\mathscr{N} = 1$ (i.e., $\gamma = 0$) results quickly diverge from the drift coefficient A of the homogenized equation towards the coefficient α in (1.11), multiple applications of the filter, well-calibrated with respect to the filtering width δ, give extremely robust results with respect to the latter.

4.4.3 Some partial results toward proving Conjecture 4.18

In this section we introduce some technical results which may be useful to show Conjecture 4.18. Their proofs is similar to the proofs of the corresponding results for the exponential filter in

Chapter 2, but it is in general more complex as it is important to highlight the dependence on the number $\mathscr{N}$ of times we apply the filter, in addition to the parameter δ of the filtering width. In the first part we show some inequalities for the new iterated filter, and in the second part we study the ergodic properties of the corresponding filtered data.

Properties of the filter

Lemma 4.20. *It holds*

$$\int_0^t k_{\exp}^{\delta,\mathscr{N}}(t-s)\,\mathrm{d}s = C_k(t,\mathscr{N},\delta),$$

where $0 < C_k(t,\mathscr{N},\delta) < 1$. *Moreover, if* $t > 6(\mathscr{N}-1)\delta$, *it holds*

$$1 - C_k(t,\mathscr{N},\delta) \leq Ce^{-t/(2\delta)},$$

for a constant $C > 0$ *independent of* δ, $\mathscr{N}$ *and* t.

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

$$\int_0^t k_{\exp}^{\delta,\mathscr{N}}(t-s)\,\mathrm{d}s = \frac{1}{(\mathscr{N}-1)!}\int_0^{t/\delta} u^{\mathscr{N}-1}e^{-u}\,\mathrm{d}s = \frac{1}{(\mathscr{N}-1)!}\gamma\left(\mathscr{N},\frac{t}{\delta}\right) =: C_k(t,\mathscr{N},\delta),$$

where γ is the lower incomplete Gamma function. Now, since $0 < \gamma(\mathscr{N},t/\delta) < \Gamma(\mathscr{N})$, where Γ is the Gamma function, and $\Gamma(\mathscr{N}) = (\mathscr{N}-1)!$, we have $0 < C_k(t,\mathscr{N},\delta) < 1$. Let us remark that it holds

$$1 - C_k(t,\mathscr{N},\delta) = 1 - \frac{1}{(\mathscr{N}-1)!}\gamma\left(\mathscr{N},\frac{t}{\delta}\right) = \frac{1}{(\mathscr{N}-1)!}\Gamma\left(\mathscr{N},\frac{t}{\delta}\right),$$

where Γ is the upper incomplete Gamma function. Let first $\mathscr{N} = 1$. In this case, we have $\Gamma(1,t/\delta) = e^{-t/\delta} \leq e^{-t/(2\delta)}$, so that the desired result holds for all $t \geq 0$. We now consider the case $\mathscr{N} \geq 2$. For any real numbers a, x, the bound

$$\Gamma(a,x) \leq Bx^{a-1}e^{-x},$$

holds for any $B > 1$ and for $x > B(a-1)/(B-1)$ [90]. We take $B = 6/5$ and consider $t > 6(\mathscr{N}-1)\delta$ so that the bound above holds for $\Gamma(\mathscr{N},t/\delta)$. Moreover, by Stirling's formula, we have for $\mathscr{N} \geq 2$ that $(\mathscr{N}-1)! \geq \sqrt{2\pi(\mathscr{N}-1)}(\mathscr{N}-1)^{\mathscr{N}-1}e^{-(\mathscr{N}-1)}$. Then, we obtain

$$1 - C_k(t,\mathscr{N},\delta) \leq C\frac{e^{-t/\delta+\mathscr{N}-1}}{\sqrt{\mathscr{N}-1}}\left(\frac{t}{\delta(\mathscr{N}-1)}\right)^{\mathscr{N}-1},$$

for a constant $C > 0$ and independent of t, δ and $\mathscr{N}$. Then, since $\sqrt{\mathscr{N}-1} \geq 1$ we have

$$\begin{aligned} 1 - C_k(t,\mathscr{N},\delta) &\leq Ce^{-t/\delta+\mathscr{N}-1}\left(\frac{t}{\delta(\mathscr{N}-1)}\right)^{\mathscr{N}-1} \\ &= C\exp\left((\mathscr{N}-1)\left(1 - \frac{t}{\delta(\mathscr{N}-1)} + \log\left(\frac{t}{\delta(\mathscr{N}-1)}\right)\right)\right) \end{aligned}$$

Now, let us remark that if $x > 6$ it holds $1 + \log x < x/2$, so that by choosing $x = t/(\delta(\mathscr{N}-1))$ and since $t > 6(\mathscr{N}-1)\delta$ we obtain

$$1 - C_k(t,\mathscr{N},\delta) \leq Ce^{-t/(2\delta)},$$

which proves the desired result. □

Lemma 4.21. *Let $g \in \mathcal{C}^0((0,T))$, $\mathcal{N}$ be a positive integer and $k_{\exp}^{\delta,\mathcal{N}}$ be defined in* (4.35)*. Then, it holds*

$$\left|\int_0^t k_{\exp}^{\delta,\mathcal{N}}(t-s)g(s)\,\mathrm{d}s\right|^p \leq \int_0^t k_{\exp}^{\delta,\mathcal{N}}(t-s)\,|g(s)|^p\,\mathrm{d}s,$$

for all $p \geq 1$.

Proof. We employ the equivalent formulation of (4.33) for the filter, with a generic continuous function $g(s)$ instead of X_s^ε. Let us define

$$I_1(s_{\mathcal{N}-1}) := \int_0^{s_{\mathcal{N}-1}} k_{\exp}^{\delta,1}(s_{\mathcal{N}-1} - s_{\mathcal{N}})g(s_{\mathcal{N}})\,\mathrm{d}s_{\mathcal{N}},$$

and for $n = 2, \ldots, \mathcal{N}$

$$\begin{aligned} I_n(s_{\mathcal{N}-n}) := \int_0^{s_{\mathcal{N}-n}} & k_{\exp}^{\delta,1}(s_{\mathcal{N}-n} - s_{\mathcal{N}-n+1}) \cdots \\ & \cdots \int_0^{s_{\mathcal{N}-1}} k_{\exp}^{\delta,1}(s_{\mathcal{N}-1} - s_{\mathcal{N}})g(s_{\mathcal{N}})\,\mathrm{d}s_{\mathcal{N}} \cdots \mathrm{d}s_{\mathcal{N}-n+1}, \end{aligned}$$

with $s_0 \equiv t$. We then prove the result by induction on n. For $n = 1$, the result holds due to Lemma 2.26. For the recursion step, assume that it holds

$$\begin{aligned} |I_n(s_{\mathcal{N}-n})|^p \leq \int_0^{s_{\mathcal{N}-n}} & k_{\exp}^{\delta,1}(s_{\mathcal{N}-n} - s_{\mathcal{N}-n+1}) \cdots \\ & \cdots \int_0^{s_{\mathcal{N}-1}} k_{\exp}^{\delta,1}(s_{\mathcal{N}-1} - s_{\mathcal{N}})\,|g(s_{\mathcal{N}})|^p\,\mathrm{d}s_{\mathcal{N}} \cdots \mathrm{d}s_{\mathcal{N}-n+1}. \end{aligned}$$

We can then write

$$I_{n+1}(s_{\mathcal{N}-n-1}) = \int_0^{s_{\mathcal{N}-n-1}} k_{\exp}^{\delta,1}(s_{\mathcal{N}-n-1} - s_{\mathcal{N}-n})I_n(s_{\mathcal{N}-n})\,\mathrm{d}s_{\mathcal{N}-n}.$$

Now $I_n(s_{\mathcal{N}-n})$ is a continuous function of $s_{\mathcal{N}-n}$ and therefore again due to Lemma 2.26 we have

$$|I_{n+1}(s_{\mathcal{N}-n-1})|^p \leq \int_0^{s_{\mathcal{N}-n-1}} k_{\exp}^{\delta,1}(s_{\mathcal{N}-n-1} - s_{\mathcal{N}-n})\,|I_n(s_{\mathcal{N}-n})|^p\,\mathrm{d}s_{\mathcal{N}-n},$$

which, combined with the induction hypothesis, yields the desired result. □

Lemma 4.22. *Let $p \geq 1$. Then, it holds*

$$\int_0^t k_{\exp}^{\delta,\mathcal{N}}(t-s)(t-s)^p\,\mathrm{d}s \leq \Gamma(p+1)\mathcal{N}^p\delta^p.$$

Proof. It holds

$$\int_0^t k_{\exp}^{\delta,\mathcal{N}}(t-s)(t-s)^p\,\mathrm{d}s = \frac{1}{\delta^{\mathcal{N}-1}(\mathcal{N}-1)!}\int_0^t k_{\exp}^{\delta,1}(t-s)(t-s)^{\mathcal{N}+p-1}\,\mathrm{d}s,$$

so that from the proof of Lemma 2.27, we obtain

$$\int_0^t k_{\exp}^{\delta,\mathcal{N}}(t-s)(t-s)^p\,\mathrm{d}s \leq \frac{\Gamma(\mathcal{N}+p)}{(\mathcal{N}-1)!}\delta^p,$$

where Γ denotes the Gamma function. It now remains to show that

$$\frac{\Gamma(\mathscr{N}+p)}{(\mathscr{N}-1)!} \leq \Gamma(p+1)\mathscr{N}^p, \tag{4.40}$$

which we prove by induction. The case $\mathscr{N}=1$ is trivial. Let us now assume that (4.40) holds true, then by the properties of the Gamma function, it holds

$$\frac{\Gamma(\mathscr{N}+1+p)}{\mathscr{N}!} = \frac{(\mathscr{N}+p)\Gamma(\mathscr{N}+p)}{\mathscr{N}(\mathscr{N}-1)!} \leq \frac{\mathscr{N}+p}{\mathscr{N}}\Gamma(p+1)\mathscr{N}^p = \Gamma(p+1)(\mathscr{N}^p + p\mathscr{N}^{p-1}). \tag{4.41}$$

Moreover, notice that for $p \geq 1$

$$(\mathscr{N}+1)^p = \mathscr{N}^p\left(1+\frac{1}{\mathscr{N}}\right)^p \geq \mathscr{N}^p\left(1+\frac{p}{\mathscr{N}}\right) = \mathscr{N}^p + p\mathscr{N}^{p-1},$$

which together with (4.41) gives the desired result. □

Ergodic properties

Lemma 4.23. *Let $(X^\varepsilon, Z_1^\varepsilon, \dots, Z_{\mathscr{N}}^\varepsilon)^\top$ be the solution of* (4.36) *and let $\mathfrak{m}_t^{\varepsilon,\mathscr{N}}$ be the measure induced by the joint process at time t. Then, the measure $\mathfrak{m}_t^{\varepsilon,\mathscr{N}}$ admits a smooth density with respect to the Lebesgue measure.*

Proof. We have to show that the joint process solution to (4.36) is hypo-elliptic. Denoting as $f\colon \mathbb{R}\to\mathbb{R}$ the function

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

the generator of the process $(X^\varepsilon, Z_1^\varepsilon, \dots, Z_{\mathscr{N}}^\varepsilon)^\top$ is given by

$$\mathcal{L} = f\partial_x + \sigma\partial_{xx}^2 + \frac{1}{\delta}(x-z_1)\partial_{z_1} + \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}(z_{n-1}-z_n)\partial_{z_n} =: \mathcal{X}_0 + \sigma\mathcal{X}_1^2,$$

where

$$\mathcal{X}_0 = f\partial_x + \frac{1}{\delta}(x-z_1)\partial_{z_1} + \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}(z_{n-1}-z_n)\partial_{z_n}, \quad \mathcal{X}_1 = \partial_x.$$

Then we have

$$\begin{aligned}
&\mathcal{A}_0 = \mathrm{Lie}(\mathcal{X}_1) = \mathrm{Lie}(\partial_x) \\
&\mathcal{A}_1 = \mathrm{Lie}([\mathcal{X}_0,\mathcal{X}_1]) \supset \mathrm{Lie}\left(-\frac{1}{\delta}\partial_{z_1}\right) \\
&\mathcal{A}_1 = \mathrm{Lie}([\mathcal{X}_0,[\mathcal{X}_0,\mathcal{X}_1]]) \supset \mathrm{Lie}\left(\frac{1}{\delta^2}\partial_{z_2}\right) \\
&\dots \\
&\mathcal{A}_N \supset \mathrm{Lie}\left(\frac{(-1)^{\mathscr{N}}}{\delta^{\mathscr{N}}}\partial_{z_{\mathscr{N}}}\right).
\end{aligned}$$

Consequently,

$$\mathcal{H} = \mathrm{Lie}\left(\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_{\mathscr{N}}\right) \supset \mathrm{Lie}\left(\partial_x, \partial_{z_1}, \dots, \partial_{z_{\mathscr{N}}}\right),$$

which spans the tangent space of $\mathbb{R}^{1+\mathscr{N}}$ at $(x, z_1, \dots, z_{\mathscr{N}})$, denoted $T_{x,z_1,\dots,z_{\mathscr{N}}}\mathbb{R}^{1+\mathscr{N}}$. The desired result then follows from Hörmander's theorem (see e.g. [101, Chapter 6]). □

Lemma 4.24. *Let $\rho^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}$ be the solution of (4.37) and let us write*

$$\rho^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z) = \varphi^\varepsilon(x)\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z), \tag{4.42}$$

where φ^ε is the marginal density of X^ε, i.e.,

$$\varphi^\varepsilon(x) = \int_{\mathbb{R}^{\mathscr{N}}} \rho^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z)\,\mathrm{d}z.$$

Then, it holds

$$\varphi^\varepsilon(x) = \frac{1}{C_{\nu^\varepsilon}} \exp\left(-\frac{1}{\sigma}\alpha\cdot V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right), \tag{4.43}$$

where

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

Moreover, it holds

$$\sigma\delta\int_{\mathbb{R}^{1+\mathscr{N}}} V'(z_{\mathscr{N}})\varphi^\varepsilon(x)\partial_x\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z = \mathbb{E}^{\mu^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}}[(X^\varepsilon - Z^\varepsilon_{\mathscr{N}})(Z^\varepsilon_{\mathscr{N}-1} - Z^\varepsilon_{\mathscr{N}})V''(Z^\varepsilon_{\mathscr{N}})].$$

Proof. Integrating equation (4.37) with respect to $z_1,\dots,z_{\mathscr{N}}$ we obtain the stationary FPE for the process X^ε, i.e.

$$\sigma(\varphi^\varepsilon)''(x) + \frac{\mathrm{d}}{\mathrm{d}x}\left(\left(\alpha\cdot V'(x) + \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)\varphi^\varepsilon(x)\right) = 0, \tag{4.44}$$

whose solution is given by

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

and which proves (4.43). In view of (4.42) and (4.44), equation (4.37) can be rewritten as

$$\partial_x\left(\sigma\varphi^\varepsilon\partial_x\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}\right) + \partial_{z_1}\left(\frac{1}{\delta}(z_1 - x)\varphi^\varepsilon\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}\right) + \sum_{n=2}^{\mathscr{N}}\partial_{z_n}\left(\frac{1}{\delta}(z_n - z_{n-1})\varphi^\varepsilon\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}\right) = 0.$$

We now multiply the equation above by a continuous differentiable function $f\colon \mathbb{R}^{1+\mathscr{N}} \to \mathbb{R}^L$, $f = f(x,z)$, and integrate with respect to x and $z_1,\dots,z_{\mathscr{N}}$. Then an integration by parts yields

$$\begin{aligned}\sigma\int_{\mathbb{R}^{1+\mathscr{N}}}\partial_x f(x,z)\varphi^\varepsilon(x)\partial_x\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z &= \frac{1}{\delta}\int_{\mathbb{R}^{1+\mathscr{N}}}\partial_{z_1}f(x,z)(x - z_1)\varphi^\varepsilon(x)\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z \\ &\quad + \sum_{n=2}^{\mathscr{N}}\frac{1}{\delta}\int_{\mathbb{R}^{1+\mathscr{N}}}\partial_{z_n}f(x,z)(z_{n-1} - z_n)\varphi^\varepsilon(x)\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z,\end{aligned}$$

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

$$\begin{aligned}\sigma\delta\int_{\mathbb{R}^{1+\mathscr{N}}}\partial_x f(x,z)\varphi^\varepsilon(x)\partial_x\mathfrak{R}^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z &= \mathbb{E}^{\mu^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}}\left[\partial_{z_1}f(X^\varepsilon, Z^\varepsilon_1,\dots,Z^\varepsilon_{\mathscr{N}})(X^\varepsilon - Z^\varepsilon_1)\right] \\ &\quad + \sum_{n=2}^{\mathscr{N}}\mathbb{E}^{\mu^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}}\left[\partial_{z_n}f(X^\varepsilon, Z^\varepsilon_1,\dots,Z^\varepsilon_{\mathscr{N}})(Z^\varepsilon_{n-1} - Z^\varepsilon_n)\right].\end{aligned}$$

Finally, choosing

$$f(x,z_1,\dots,z_{\mathscr{N}}) = (x - z_{\mathscr{N}})V'(z_{\mathscr{N}}) + V(z_{\mathscr{N}}),$$

we obtain the desired result. □

Proposition 4.25. *It holds*

$$X_t^\varepsilon - (Z_{\mathscr{N}}^\varepsilon)_t = \delta \sum_{n=1}^{\mathscr{N}} (B_n^\varepsilon)_t + R(\varepsilon, \delta),$$

where the stochastic processes $(B_n^\varepsilon)_t$ are defined for all $n = 1, \dots, \mathscr{N}$ as

$$(B_n^\varepsilon)_t = \sqrt{2\sigma} \int_0^t k_{\exp}^{\delta,n}(t-s)(1 + \Phi'(Y_s^\varepsilon)) \, \mathrm{d}W_s. \tag{4.45}$$

Moreover, the following estimates are satisfied

$$\left(\mathbb{E}^{\nu^\varepsilon} \left| \sum_{n=1}^{\mathscr{N}} (B_n^\varepsilon)_t \right|^p \right)^{1/p} \leq C \mathscr{N}^{1/2} \delta^{-1/2}, \tag{4.46}$$

$$\left(\mathbb{E}^{\nu^\varepsilon} \left| (B_{\mathscr{N}}^\varepsilon)_t \right|^2 \right)^{1/2} \leq C \mathscr{N}^{-1/4} \delta^{-1/2}, \tag{4.47}$$

$$\left(\mathbb{E}^{\nu^\varepsilon} \left| R(\varepsilon, \delta) \right|^p \right)^{1/p} \leq C \left(\varepsilon + \mathscr{N} \delta \right). \tag{4.48}$$

Proof. Denoting $Y_t^\varepsilon \coloneqq X_t^\varepsilon / \varepsilon$, we will use the decomposition [103, Formula 5.8]

$$\begin{aligned} X_t^\varepsilon - X_s^\varepsilon = &- \int_s^t (\alpha \cdot V'(X_r^\varepsilon))(1 + \Phi'(Y_r^\varepsilon)) \, \mathrm{d}r \\ &+ \sqrt{2\sigma} \int_s^t (1 + \Phi'(Y_r^\varepsilon)) \, \mathrm{d}W_r - \varepsilon(\Phi(Y_t^\varepsilon) - \Phi(Y_s^\varepsilon)). \end{aligned} \tag{4.49}$$

Recall that by definition of $(Z_{\mathscr{N}}^\varepsilon)_t$ and by Lemma 4.20 we have

$$X_t^\varepsilon - (Z_{\mathscr{N}}^\varepsilon)_t = \int_0^t k_{\exp}^{\delta,\mathscr{N}}(t-s)(X_t^\varepsilon - X_s^\varepsilon) \, \mathrm{d}s + (1 - C_k(t, \mathscr{N}, \delta)) X_t^\varepsilon.$$

Plugging the decomposition (4.49) into the equation above, we obtain

$$X_t^\varepsilon - (Z_{\mathscr{N}}^\varepsilon)_t = I_1^\varepsilon(t) + I_2^\varepsilon(t) + I_3^\varepsilon(t) + I_4^\varepsilon(t),$$

where

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

Reasoning similarly as in the proof of Proposition 2.14, since Φ and Φ' are bounded and X_t^ε has bounded moments of any order by [103, Corollary 5.4], due to Lemma 4.20, Lemma 4.21 and Lemma 4.22 we deduce for $t > 6(\mathscr{N} - 1)\delta$

$$\begin{aligned} \left(\mathbb{E}^{\nu^\varepsilon} \left| I_1^\varepsilon \right|^p \right)^{1/p} &\leq C \int_0^t k_{\exp}^{\delta,\mathscr{N}}(t-s)(t-s) \, \mathrm{d}s \leq C \mathscr{N} \delta, \\ \left(\mathbb{E}^{\nu^\varepsilon} \left| I_3^\varepsilon \right|^p \right)^{1/p} &\leq C \varepsilon, \\ \left(\mathbb{E}^{\nu^\varepsilon} \left| I_4^\varepsilon \right|^p \right)^{1/p} &\leq C e^{-t/(2\delta)}. \end{aligned} \tag{4.50}$$

Let us now consider $I_2^\varepsilon(t)$. We introduce the notation

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

and therefore we have

$$I_2^\varepsilon(t) = \sqrt{2\sigma} \int_0^t k_{\exp}^{\delta,\mathscr{N}}(t-s)(Q_t^\varepsilon - Q_s^\varepsilon) \, \mathrm{d}s.$$

Applying iteratively the Itô formula to $u_n(s, Q_s^\varepsilon)$ where $u_n(s,x) = k_{\exp}^{\delta,n}(t-s)x$ for all $n = 1, \dots, \mathscr{N}$ yields

$$\begin{aligned} I_2^\varepsilon(t) &= \sqrt{2\sigma} \left((C_k(t, \mathscr{N}, \delta) - 1) Q_t^\varepsilon + \delta \sum_{n=1}^{\mathscr{N}} \int_0^t k_{\exp}^{\delta,n}(t-s) \left(1 + \Phi'(Y_s^\varepsilon)\right) \mathrm{d}W_s \right) \\ &= \delta \sum_{n=1}^{\mathscr{N}} (B_n^\varepsilon)_t - \sqrt{2\sigma}(1 - C_k(t, \mathscr{N}, \delta)) Q_t^\varepsilon \eqqcolon \delta \sum_{n=1}^{\mathscr{N}} (B_n^\varepsilon)_t - \widetilde{I}_2^\varepsilon(t). \end{aligned} \tag{4.51}$$

where the processes B_n^ε, $n = 1, \dots, \mathscr{N}$ are defined in (4.45). For the remainder $\widetilde{I}_2^\varepsilon(t)$, as in the proof of Proposition 2.14, we have

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

which due to Lemma 4.20 implies for all $t > 6(\mathscr{N} - 1)\delta$

$$\left(\mathbb{E}^{\phi^\varepsilon} \left| \widetilde{I}_2^\varepsilon(t) \right|^p \right)^{1/p} \le C \sqrt{t} e^{-t/(2\delta)}, \tag{4.52}$$

which together with (4.50) gives estimate (4.48). Let us now remark that from (4.51) it holds for a constant $C > 0$ depending only on p

$$\left(\mathbb{E} \left| \sum_{n=1}^{\mathscr{N}} (B_n^\varepsilon)_t \right|^p \right)^{1/p} \le C\delta^{-1} \left(\mathbb{E} \left| I_2^\varepsilon(t) \right|^p \right)^{1/p} + C\delta^{-1} \left(\mathbb{E} \left| \widetilde{I}_2^\varepsilon(t) \right|^p \right)^{1/p}.$$

The second term is bounded exponentially fast with respect to t and δ due to (4.52). For the first term, applying Lemma 4.21, the inequality [69, Formula (3.25), p. 163] and Lemma 4.22 we obtain for a constant $C > 0$ independent of δ and t

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

Therefore, it holds

$$\left(\mathbb{E} \left| \sum_{n=1}^{\mathscr{N}} (B_n^\varepsilon)_t \right|^p \right)^{1/p} \le C \mathscr{N}^{1/2} \delta^{-1/2},$$

which proves the bound (4.46). It only remains to prove the estimate (4.47). Since Φ' is bounded and by Itô isometry we have

$$\mathbb{E}^{\nu^\varepsilon} \left| (B_{\mathscr{N}}^\varepsilon)_t \right|^2 \le C \int_0^t (k_{\exp}^{\delta,\mathscr{N}})^2 (t-s) \, \mathrm{d}s = C \frac{(2\mathscr{N} - 2)!}{\delta 2^{2\mathscr{N}-1} ((\mathscr{N} - 1)!)^2}.$$

Moreover, due to Stirling's formula we have for all $\mathscr{N} \geq 2$

$$(\mathscr{N}-1)! \geq \sqrt{2\pi(\mathscr{N}-1)}(\mathscr{N}-1)^{\mathscr{N}-1}e^{-(\mathscr{N}-1)},$$

and

$$(2\mathscr{N}-2)! \leq e\sqrt{(2\mathscr{N}-2)}(2\mathscr{N}-2)^{(2\mathscr{N}-2)}e^{-(2\mathscr{N}-2)},$$

which yield

$$\mathbb{E}^{\nu^\varepsilon}\left|(B^\varepsilon_{\mathscr{N}})_t\right|^2 \leq C\mathscr{N}^{-1/2}\delta^{-1},$$

which implies the desired estimate. □

Lemma 4.26. *It holds for all $n = 1, \dots, \mathscr{N}$*

$$(Z^\varepsilon_{n-1})_t - (Z^\varepsilon_n)_t = \delta(B^\varepsilon_n)_t + R_n(\varepsilon,\delta), \tag{4.53}$$

where $(Z^\varepsilon_0)_t \equiv X^\varepsilon_t$ and the stochastic process $(B^\varepsilon_n)_t$ is defined in (4.45) *and $R_n(\varepsilon,\Delta)$ satisfies*

$$\left(\mathbb{E}^{\nu^\varepsilon}\left|R_n(\varepsilon,\delta)\right|^p\right)^{1/p} \leq C\left(\varepsilon+\delta\right),$$

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

Proof. Let us prove the result by induction. The case $n = 1$ is given by Proposition 2.14. Then, assuming that (4.53) holds true, by definition of the stochastic process $(Z^\varepsilon_n)_t$ we have

$$\begin{aligned}
(Z^\varepsilon_n)_t - (Z^\varepsilon_{n+1})_t &= \int_0^t k^{\delta,1}_{\exp}(t-s)((Z^\varepsilon_{n-1})_s - (Z^\varepsilon_n)_s)\,\mathrm{d}s \\
&= \sqrt{2\sigma}\delta\int_0^t k^{\delta,1}_{\exp}(t-s)\int_0^s k^{\delta,n}_{\exp}(s-r)(1+\Phi'(Y^\varepsilon_r)\,\mathrm{d}W_r\,\mathrm{d}s + \int_0^t k^{\delta,1}_{\exp}(t-s)R_n(\varepsilon,\delta)\,\mathrm{d}s \\
&=: I_{n+1} + R_{n+1}(\varepsilon,\delta).
\end{aligned}$$

Let us first consider I_{n+1}. By exchanging the order of the integrals we obtain

$$\begin{aligned}
I_{n+1} &= \sqrt{2\sigma}\delta\int_0^t \frac{1}{\delta^{n+1}}e^{-\frac{t-r}{\delta}}\left(\int_r^t \frac{(s-r)^{n-1}}{(n-1)!}\,\mathrm{d}s\right)(1+\Phi'(Y^\varepsilon_r))\,\mathrm{d}W_r \\
&= \sqrt{2\sigma}\delta\int_0^t k^{\delta,n+1}_{\exp}(t-r)(1+\Phi'(Y^\varepsilon_r))\,\mathrm{d}W_r.
\end{aligned}$$

Moreover, the remainder $R_{n+1}(\varepsilon,\delta)$ satisfies

$$\left(\mathbb{E}^{\nu^\varepsilon}\left|R_{n+1}(\varepsilon,\delta)\right|^p\right)^{1/p} \leq \left(\int_0^t k^{\delta,1}_{\exp}(t-s)\,\mathrm{d}s\right)^{1/p}\left(\mathbb{E}^{\nu^\varepsilon}\left|R_n(\varepsilon,\delta)\right|^p\right)^{1/p} \leq C\left(\varepsilon+\delta\right),$$

which concludes the proof. □

Lemma 4.27. *Let $\eta^\varepsilon_{\mathscr{N}}$ be the invariant measure of $(X^\varepsilon, Y^\varepsilon, Z^\varepsilon_1, \dots, Z^\varepsilon_{\mathscr{N}}, B^\varepsilon_1, \dots, B^\varepsilon_{\mathscr{N}})$. Then it holds*

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

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

$$\left|\widetilde{R}(\varepsilon,\delta)\right| \leq C\left(\mathscr{N}^{1/4}\delta + \mathscr{N}^{1/2}\delta^{1/2}\varepsilon + \mathscr{N}^{1/2}\delta^{3/2}\right),$$

for a constant $C > 0$ independent of ε, δ and $\mathscr{N}$.

Proof. Let us consider the following system of SDEs for the multidimensional stochastic process $(X_t^\varepsilon, (Z_1^\varepsilon)_t, \dots, (Z_{\mathscr{N}}^\varepsilon)_t, (B_1^\varepsilon)_t, \dots, (B_{\mathscr{N}}^\varepsilon)_t, Y_t^\varepsilon)$

$$\begin{aligned}
\mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'(Y_t^\varepsilon)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\
\mathrm{d}(Z_1^\varepsilon)_t &= -\frac{1}{\delta}\left((Z_1^\varepsilon)_t - X_t^\varepsilon\right)\,\mathrm{d}t, \\
\mathrm{d}(Z_n^\varepsilon)_t &= -\frac{1}{\delta}\left((Z_n^\varepsilon)_t - (Z_{n-1}^\varepsilon)_t\right)\,\mathrm{d}t, \qquad n = 2, \dots, \mathscr{N} \\
\mathrm{d}(B_1^\varepsilon)_t &= -\frac{1}{\delta}(B_1^\varepsilon)_t\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\delta}(1 + \Phi'(Y_t^\varepsilon))\,\mathrm{d}W_t, \\
\mathrm{d}(B_n^\varepsilon)_t &= -\frac{1}{\delta}\left((B_n^\varepsilon)_t - (B_{n-1}^\varepsilon)_t\right)\,\mathrm{d}t, \qquad n = 2, \dots, \mathscr{N} \\
\mathrm{d}Y_t^\varepsilon &= -\frac{1}{\varepsilon}\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon^2}p'(Y_t^\varepsilon)\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{\varepsilon}\,\mathrm{d}W_t,
\end{aligned}$$

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

$$\begin{aligned}
\widetilde{\mathcal{L}}_{\mathscr{N}}^\varepsilon = &- \left(\alpha \cdot V'(x) + \frac{1}{\varepsilon}p'(y)\right)\partial_x - \frac{1}{\delta}(z_1 - x)\partial_{z_1} - \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}(z_n - z_{n-1})\partial_{z_n} \\
&- \frac{1}{\delta}b_1\partial_{b_1} - \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}(b_n - b_{n-1})\partial_{b_n} - \left(\frac{1}{\varepsilon}\alpha \cdot V'(x) + \frac{1}{\varepsilon^2}p'(y)\right)\partial_y \\
&+ \sigma\left(\partial_{xx}^2 + \frac{2}{\varepsilon}\partial_{xy}^2 + \frac{1}{\varepsilon^2}\partial_{yy}^2 + \frac{2(1 + \Phi'(y))}{\delta}\partial_{xb_1}^2 + \frac{2(1 + \Phi'(y))}{\varepsilon\delta}\partial_{yb_1}^2 + \frac{(1 + \Phi'(y))^2}{\delta^2}\partial_{b_1b_1}^2\right).
\end{aligned}$$

Let us denote by $e_{\mathscr{N}}^\varepsilon \colon \mathbb{R}^{1+2\mathscr{N}} \times [0, \mathbb{T}] \to \mathbb{R}$, $e_{\mathscr{N}}^\varepsilon = e_{\mathscr{N}}^\varepsilon(x, z_1, \dots, z_{\mathscr{N}}, b_1, \dots, b_{\mathscr{N}}, y)$, the density of the invariant measure $\eta_{\mathscr{N}}^\varepsilon$ of $(X_t^\varepsilon, (Z_1^\varepsilon)_t, \dots, (Z_{\mathscr{N}}^\varepsilon)_t, (B_1^\varepsilon)_t, \dots, (B_{\mathscr{N}}^\varepsilon)_t, Y_t^\varepsilon)$. Then $e_{\mathscr{N}}^\varepsilon$ solves the stationary FPE $(\widetilde{\mathcal{L}}_{\mathscr{N}}^\varepsilon)^* e_{\mathscr{N}}^\varepsilon = 0$, i.e., explicitly

$$\begin{aligned}
&\partial_x\left(\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon}p'(y)\right)e_{\mathscr{N}}^\varepsilon\right) + \frac{1}{\delta}\partial_{z_1}\left((z_1 - x)e_{\mathscr{N}}^\varepsilon\right) + \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}\partial_{z_n}\left((z_n - z_{n-1})e_{\mathscr{N}}^\varepsilon\right) \\
&+ \frac{1}{\delta}\partial_{b_1}(b_1\eta^\varepsilon) + \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}\partial_{b_n}\left((b_n - b_{n-1})e_{\mathscr{N}}^\varepsilon\right) + \partial_y\left(\left(\frac{1}{\varepsilon}\alpha \cdot V'(x) + \frac{1}{\varepsilon^2}p'(y)\right)e_{\mathscr{N}}^\varepsilon\right) \\
&+ \sigma\left(\partial_{xx}^2 e_{\mathscr{N}}^\varepsilon + \frac{2}{\varepsilon}\partial_{xy}^2 e_{\mathscr{N}}^\varepsilon + \frac{1}{\varepsilon^2}\partial_{yy}^2 e_{\mathscr{N}}^\varepsilon\right) \\
&+ \sigma\left(\frac{2}{\delta}\partial_{xb_1}^2\left((1 + \Phi'(y))e_{\mathscr{N}}^\varepsilon\right) + \frac{2}{\varepsilon\delta}\partial_{yb_1}^2\left((1 + \Phi'(y))e_{\mathscr{N}}^\varepsilon\right) + \frac{1}{\delta^2}\partial_{b_1b_1}^2\left((1 + \Phi'(y))^2 e_{\mathscr{N}}^\varepsilon\right)\right) = 0.
\end{aligned}$$

We now multiply the equation above by a continuous differentiable function $f \colon \mathbb{R}^{2\mathscr{N}} \to \mathbb{R}^L$, $f = f(z_1, \dots, z_{\mathscr{N}}, b_1, \dots, b_{\mathscr{N}})$, and integrate with respect to x, z, b and y. Then an integration by parts yields

$$\begin{aligned}
&\frac{1}{\delta}\int_{\mathbb{R}^{1+2\mathscr{N}} \times [0,\mathbb{T}]} \partial_{z_1} f(z, b)(x - z_1)e_{\mathscr{N}}^\varepsilon + \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}\int_{\mathbb{R}^{1+2\mathscr{N}} \times [0,\mathbb{T}]} \partial_{z_n} f(z, b)(z_{n-1} - z_n)e_{\mathscr{N}}^\varepsilon \\
&\quad - \frac{1}{\delta}\int_{\mathbb{R}^{1+2\mathscr{N}} \times [0,\mathbb{T}]} \partial_{b_1} f(z, b)b_1 e_{\mathscr{N}}^\varepsilon - \frac{1}{\delta}\sum_{n=2}^{\mathscr{N}}\int_{\mathbb{R}^{1+2\mathscr{N}} \times [0,\mathbb{T}]} \partial_{b_n} f(z, b)(b_n - b_{n-1})e_{\mathscr{N}}^\varepsilon \\
&\qquad + \frac{\sigma}{\delta^2}\int_{\mathbb{R}^{1+2\mathscr{N}} \times [0,\mathbb{T}]} \partial_{b_1b_1}^2 f(z, b)(1 + \Phi'(y))^2 e_{\mathscr{N}}^\varepsilon = 0,
\end{aligned}$$

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

$$
\begin{aligned}
\delta\, \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\partial_{b_1} f(Z^\varepsilon, B^\varepsilon) B_1^\varepsilon\right] + \delta \sum_{n=2}^{\mathscr{N}} \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\partial_{b_n} f(Z^\varepsilon, B^\varepsilon)(B_n^\varepsilon - B_{n-1}^\varepsilon)\right] \\
= \sigma\, \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\partial^2_{b_1 b_1} f(Z^\varepsilon, B^\varepsilon)(1+\Phi'(Y^\varepsilon))^2\right] \\
+ \delta\, \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\partial_{z_1} f(Z^\varepsilon, B^\varepsilon)(X^\varepsilon - Z_1^\varepsilon)\right] + \delta \sum_{n=2}^{\mathscr{N}} \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\partial_{z_n} f(Z^\varepsilon, B^\varepsilon)(Z_{n-1}^\varepsilon - Z_n^\varepsilon)\right].
\end{aligned}
$$

Choosing

$$
f(z,b) = \frac{1}{2} \left(\sum_{n=1}^{\mathscr{N}} b_n \right)^2 V''(z_{\mathscr{N}}),
$$

we obtain

$$
\begin{aligned}
\delta\, \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\sum_{n=1}^{\mathscr{N}} B_n^\varepsilon B_{\mathscr{N}}^\varepsilon V''(Z_{\mathscr{N}}^\varepsilon) \right] &= \sigma\, \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} [(1+\Phi'(Y^\varepsilon))^2 V''(Z_{\mathscr{N}}^\varepsilon)] \\
&\quad + \frac{\delta}{2} \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left[\left(\sum_{n=1}^{\mathscr{N}} B_n^\varepsilon \right)^2 V'''(Z_{\mathscr{N}}^\varepsilon)(Z_{\mathscr{N}-1}^\varepsilon - Z_{\mathscr{N}}^\varepsilon) \right] \\
&=: \sigma\, \mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} [(1+\Phi'(Y^\varepsilon))^2 V''(Z_{\mathscr{N}}^\varepsilon)] + \widetilde{R}(\varepsilon,\delta),
\end{aligned}
$$

where, by Hölder's inequality with exponents p, q, r, Lemma 4.26 and bounds (4.46) and (4.47), the remainder satisfies

$$
\begin{aligned}
\left\| \widetilde{R}(\varepsilon,\delta) \right\| &\leq C\delta \left(\mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left| \sum_{n=1}^{\mathscr{N}} B_n^\varepsilon \right|^{2p} \right)^{1/p} \left(\mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \| V'''(Z_{\mathscr{N}}^\varepsilon) \|^q \right)^{1/q} \left(\mathbb{E}^{\eta^{\varepsilon}_{\mathscr{N}}} \left| Z_{\mathscr{N}-1}^\varepsilon - Z_{\mathscr{N}}^\varepsilon \right|^r \right)^{1/r} \\
&\leq C\delta \mathscr{N}^{1/2} \delta^{-1/2} \left(\mathscr{N}^{-1/4} \delta^{1/2} + \varepsilon + \delta \right),
\end{aligned}
$$

which completes the proof. ☐

Lemma 4.28. *It holds*

$$
\mathbb{E}^{\mu^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}} |Z_{\mathscr{N}}^\varepsilon|^p \leq C,
$$

for all $p \geq 1$ and for a constant $C > 0$ independent of ε.

Proof. The proof trivially follows from the proof of Lemma 2.28 and from Lemma 4.21 by replacing the filter $k_{\mathrm{exp}}^{\delta,1}$ with the iterated filter $k_{\mathrm{exp}}^{\delta,\mathscr{N}}$. ☐

Lemma 4.29. *Let $(\mathscr{N}-1)\delta \leq \bar{t} < \infty$. Then, it holds*

$$
\left(\mathbb{E}^{\mu^{\varepsilon,\mathscr{N}}_{\mathrm{exp}}} |X^\varepsilon - Z_{\mathscr{N}}^\varepsilon|^p \right)^{1/p} \leq C \left(\mathscr{N} \delta \varepsilon^{-1} + \mathscr{N}^{1/2} \delta^{1/2} \right),
$$

where $C > 0$ is a constant independent of ε, δ and $\mathscr{N}$.

Proof. By Lemma 4.20, we have

$$
X_t^\varepsilon - (Z_{\mathscr{N}}^\varepsilon)_t = \int_0^t k_{\mathrm{exp}}^{\delta,\mathscr{N}}(t-s) \left(X_t^\varepsilon - X_s^\varepsilon \right) \mathrm{d}s + \left(1 - C_k(t,\mathscr{N},\delta)\right) X_t^\varepsilon.
$$

Moreover, it holds

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

see Lemma 2.30. Therefore, by Lemma 4.21 we obtain

$$\begin{aligned}\mathbb{E}^{\nu^\varepsilon} |X_t^\varepsilon - (Z_{\mathscr{N}}^\varepsilon)_t|^p &\le C \int_0^t k_{\exp}^{\delta,\mathscr{N}}(t-s) \left((t-s)^p \varepsilon^{-p} + (t-s)^{p/2} \right) \mathrm{d}s \\ &\quad + C \left(1 - C_k(t, \mathscr{N}, \delta)\right)^p \mathbb{E}^{\nu^\varepsilon} |X_t^\varepsilon|^p \\ &=: I_1 + I_2,\end{aligned}$$

for a constant C depending only on p. Let us consider the first term. By Lemma 4.22, we have

$$I_1 \le C \mathscr{N}^p \delta^p \varepsilon^{-p} + C \mathscr{N}^{p/2} \delta^{p/2}.$$

For the second term, by Lemma 4.20 and [103, Corollary 5.4] we have for $t > 6(\mathscr{N} - 1)\delta$ the bound

$$I_2 \le C e^{-pt/(2\delta)}.$$

By ergodicity we have that

$$\mathbb{E}^{\mu_{\exp}^{\varepsilon,\mathscr{N}}} |X^\varepsilon - Z_{\mathscr{N}}^\varepsilon|^p = \lim_{t\to\infty} \mathbb{E}^{\varphi^\varepsilon} |X_t^\varepsilon - (Z_{\mathscr{N}}^\varepsilon)_t|^p ,$$

which, for t sufficiently big and since $(\mathscr{N} - 1)\delta \le \bar{t} < \infty$ yields the desired result. □

Lemma 4.30. *Le $\widetilde{\mathcal{M}}_\varepsilon$ and $\mathcal{M}_0$ be defined as*

$$\widetilde{\mathcal{M}}_{\exp}^{\varepsilon,\mathscr{N}} := \mathbb{E}^{\mu_{\exp}^{\varepsilon,\mathscr{N}}} \left[V'(X^\varepsilon) \otimes V'(Z_{\mathscr{N}}^\varepsilon) \right], \qquad \text{and} \qquad \mathcal{M}^0 := \mathbb{E}^{\nu^0} \left[V'(X^0) \otimes V'(X^0) \right].$$

Then, if $\mathscr{N}\delta \to 0$ as $\varepsilon \to 0$ it holds

$$\lim_{\varepsilon\to 0} \widetilde{\mathcal{M}}_{\exp}^{\varepsilon,\mathscr{N}} = \mathcal{M}^0.$$

Proof. Employing triangle inequality and Cauchy–Schwarz inequality we have

$$\begin{aligned}\left\| \widetilde{\mathcal{M}}_{\exp}^{\varepsilon,\mathscr{N}} - \mathcal{M}^0 \right\| &\le \left\| \mathbb{E}^{\mu_{\exp}^{\varepsilon,\mathscr{N}}} \left[V'(X^\varepsilon) \otimes V'(Z_{\mathscr{N}}^\varepsilon) \right] - \mathbb{E}^{\nu^\varepsilon} \left[V'(X^\varepsilon) \otimes V'(X^\varepsilon) \right] \right\| \\ &\quad + \left\| \mathbb{E}^{\nu^\varepsilon} \left[V'(X^\varepsilon) \otimes V'(X^\varepsilon) \right] - \mathbb{E}^{\nu^0} \left[V'(X^0) \otimes V'(X^0) \right] \right\| \\ &\le C \left(\mathbb{E}^{\nu^\varepsilon} \| V'(X^\varepsilon) \|^2 \right)^{1/2} \left(\mathbb{E} |Z_{\mathscr{N}}^\varepsilon - X^\varepsilon|^2 \right)^{1/2} \\ &\quad + \left\| \mathbb{E}^{\nu^\varepsilon} \left[V'(X^\varepsilon) \otimes V'(X^\varepsilon) \right] - \mathbb{E}^{\nu^0} \left[V'(X^0) \otimes V'(X^0) \right] \right\| ,\end{aligned}$$

and due to Proposition 4.25 we obtain

$$\begin{aligned}\left\| \widetilde{\mathcal{M}}_{\exp}^{\varepsilon,\mathscr{N}} - \mathcal{M}^0 \right\| &\le \left\| \mathbb{E}^{\nu^\varepsilon} \left[V'(X^\varepsilon) \otimes V'(X^\varepsilon) \right] - \mathbb{E}^{\nu^0} \left[V'(X^0) \otimes V'(X^0) \right] \right\| \\ &\quad + C \left(\varepsilon + \mathscr{N}^{1/2} \delta^{1/2} + \mathscr{N}\delta \right).\end{aligned}$$

Finally, the desired result follows by homogenization theory and weak convergence of measures. □

We believe that the technical results presented in this section should be useful to show Conjecture 4.18, whose proof should follow the same ideas of Theorem 2.17. The main steps consist of showing that some remainder terms vanish and, in this case, their bound depend not only on ε and δ, but also on $\mathscr{N}$. However, we notice that when we let $\varepsilon \to 0$ some of them blow up. This can be caused by the fact that some bounds which we computed may not be tight enough and should be improved, such as the ones in Proposition 4.25 and consequently in Lemma 4.27.

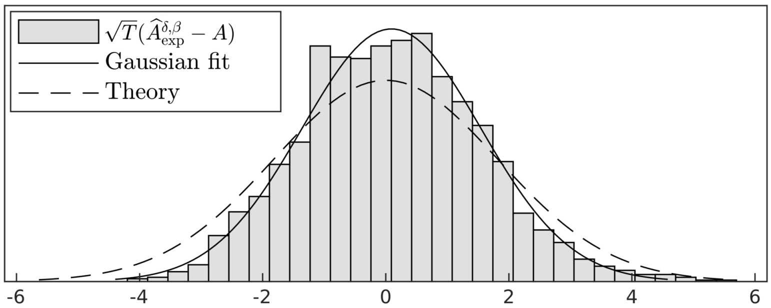


Figure 4.3 – Central limit theorem result. The histogram represents numerical results, the solid curve a Gaussian fit to the latter and the dashed curve the theoretical estimate given in Theorem 4.31.

4.5 Central limit theorem for estimator with exponential filter

The goal of this section is to study the asymptotic normality of the drift estimator proposed in Chapter 2. Simple numerical experiments suggest that it is possible to prove a central limit theorem and preliminary computations led us to the formulation of the following conjecture.

Conjecture 4.31. *Let $T = \varepsilon^{-\gamma}$ with $\gamma > 0$ and consider the exponential filter* (2.1) *with $\beta = 1$ and δ independent of ε. Then, if γ is sufficiently small, it holds*

$$\lim_{\varepsilon \to 0} \sqrt{T} \left(\widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - A \right) = \Lambda, \quad \textit{in law},$$

where $\Lambda \sim \mathcal{N}(0, \Gamma)$ and

$$\Gamma = 2\sigma\, \mathbb{E}^{\mu^0_{\mathrm{exp}}}[V'(Z^0) \otimes V'(X^0)]^{-1}\, \mathbb{E}^{\mu^0_{\mathrm{exp}}}[V'(Z^0) \otimes V'(Z^0)]\, \mathbb{E}^{\mu^0_{\mathrm{exp}}}[V'(X^0) \otimes V'(Z^0)]^{-1}.$$

4.5.1 Numerical experiment

In this experiment we wish to confirm the validity of Conjecture 4.31. We consider the same test equation as for Section 2.3.1, i.e., the quadratic potential $V(x) = x^2/2$ with fluctuating potential $p(y) = \sin(y)$, multiscale parameter $\varepsilon = 0.05$ and diffusion coefficient $\sigma = 1$. The parameters of the filter are set to $\beta = 1$ and $\delta = 1$. We compute the estimator $\widehat{A}^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T)$ with final time $T = 10^3$ on 2000 realizations of the solution and estimate the quantity

$$\Delta^\varepsilon_A(T) \coloneqq \sqrt{T} \left(A^{\delta,\beta}_{\mathrm{exp}}(X^\varepsilon, T) - A \right),$$

where A is the drift coefficient of the homogenized equation (1.11). Results, depicted in Figure 4.3, show that the distribution of $\Delta^\varepsilon_A(T)$ indeed follows a zero-mean Gaussian law, whose covariance however does not exactly agree with the theoretical result.

4.5.2 Ideas behind Conjecture 4.31

In this section we present some theoretical results which guided us to the formulation of Conjecture 4.31 and partially justify it. However, we remark that this is not a rigorous proof and only gives an intuition on why it should be true. We first introduce a technical lemma.

Lemma 4.32. *Let $\mathcal{L}^\varepsilon_{\exp}$ be the generator of the couple $(X^\varepsilon, Z^\varepsilon)^\top$, i.e.,*

$$\mathcal{L}^\varepsilon_{\exp} = -\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p'\left(\frac{x}{\varepsilon}\right)\right)\partial_x + \frac{1}{\delta}(x-z)\partial_z + \sigma\partial^2_{xx}.$$

Moreover, let $\rho^\varepsilon_{\exp}$ be the density of the invariant measure $\mu^\varepsilon_{\exp}$ of $(X^\varepsilon, Z^\varepsilon)^\top$ and $u^\varepsilon\colon \mathbb{R}^2 \to \mathbb{R}^L$ be the solution of

$$-\mathcal{L}^\varepsilon_{\exp} u^\varepsilon = \chi^\varepsilon - \mathbb{E}^{\mu^\varepsilon_{\exp}}[\chi^\varepsilon(X^\varepsilon, Z^\varepsilon)], \tag{4.54}$$

satisfying $\mathbb{E}^{\mu^\varepsilon_{\exp}}[u^\varepsilon(X^\varepsilon, Z^\varepsilon)] = 0$ for $\chi^\varepsilon\colon \mathbb{R}^2 \to \mathbb{R}^L$. Then, it holds

$$\frac{1}{T}\int_0^T \chi^\varepsilon(X^\varepsilon_t, Z^\varepsilon_t)\,\mathrm{d}t = \mathbb{E}^{\mu^\varepsilon_{\exp}}[\chi^\varepsilon(X^\varepsilon, Z^\varepsilon)] - \frac{\mathscr{R}^\varepsilon(T)}{T} + \sqrt{2\sigma}\frac{\mathscr{S}^\varepsilon(T)}{T}, \tag{4.55}$$

where

$$\mathscr{R}^\varepsilon(T) \coloneqq u^\varepsilon(X^\varepsilon_T, Z^\varepsilon_T) - u^\varepsilon(X^\varepsilon_0, Z^\varepsilon_0), \quad \mathscr{S}^\varepsilon(T) \coloneqq \int_0^T \partial_x u^\varepsilon(X^\varepsilon_t, Z^\varepsilon_t)\,\mathrm{d}W_t. \tag{4.56}$$

Moreover, it holds

$$\begin{aligned} 2\sigma\,\mathbb{E}^{\mu^\varepsilon_{\exp}}[\partial_x u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \partial_x u^\varepsilon(X^\varepsilon, Z^\varepsilon)] = \mathbb{E}^{\mu^\varepsilon_{\exp}}\,[&\chi^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes u^\varepsilon(X^\varepsilon, Z^\varepsilon) \\ &+ u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \chi^\varepsilon(X^\varepsilon, Z^\varepsilon)]. \end{aligned} \tag{4.57}$$

Proof. The proof of (4.55) and (4.56) is an application of the Itô formula (see e.g. [104, Remark 6.17]). For (4.57), it is possible to show that since $(\mathcal{L}^\varepsilon_{\exp})^*\rho^\varepsilon_{\exp} = 0$ it holds

$$(\mathcal{L}^\varepsilon_{\exp})^*(u^\varepsilon\rho^\varepsilon_{\exp}) = 2\sigma\rho^\varepsilon_{\exp}\partial^2_{xx}u^\varepsilon - \rho^\varepsilon_{\exp}\mathcal{L}^\varepsilon_{\exp}u^\varepsilon + 2\sigma\partial_x u^\varepsilon\partial_x\rho^\varepsilon_{\exp}.$$

Therefore, an integration by parts yields

$$\begin{aligned} \mathbb{E}^{\mu^\varepsilon_{\exp}}[\mathcal{L}^\varepsilon_{\exp}u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes u^\varepsilon(X^\varepsilon, Z^\varepsilon)] &= \int_\mathbb{R}\int_\mathbb{R} u^\varepsilon \otimes (\mathcal{L}^\varepsilon_{\exp})^*\left(u^\varepsilon\rho^\varepsilon_{\exp}\right)\,\mathrm{d}x\,\mathrm{d}z \\ &= -\int_\mathbb{R}\int_\mathbb{R} u^\varepsilon \otimes \mathcal{L}^\varepsilon_{\exp}u^\varepsilon\rho^\varepsilon_{\exp}\,\mathrm{d}x\,\mathrm{d}z \\ &\quad - 2\sigma\int_\mathbb{R}\int_\mathbb{R} \partial_x u^\varepsilon \otimes \partial_x u^\varepsilon\rho^\varepsilon_{\exp}\,\mathrm{d}x\,\mathrm{d}z \\ &= -\mathbb{E}^{\mu^\varepsilon_{\exp}}[u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \mathcal{L}^\varepsilon_{\exp}u^\varepsilon(X^\varepsilon, Z^\varepsilon)] \\ &\quad - 2\sigma\,\mathbb{E}^{\mu^\varepsilon_{\exp}}[\partial_x u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \partial_x u^\varepsilon(X^\varepsilon, Z^\varepsilon)]. \end{aligned}$$

Finally, since $\mathbb{E}^{\mu^\varepsilon_{\exp}}[u(X^\varepsilon, Z^\varepsilon)] = 0$

$$\begin{aligned} 2\sigma\,\mathbb{E}^{\mu^\varepsilon_{\exp}}[\partial_x u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \partial_x u^\varepsilon(X^\varepsilon, Z^\varepsilon)] &= -\mathbb{E}^{\mu^\varepsilon_{\exp}}\,[\mathcal{L}^\varepsilon_{\exp}u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes u^\varepsilon(X^\varepsilon, Z^\varepsilon) \\ &\qquad + u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \mathcal{L}^\varepsilon_{\exp}u^\varepsilon(X^\varepsilon, Z^\varepsilon)] \\ &= \mathbb{E}^{\mu^\varepsilon_{\exp}}\,[\chi^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes u^\varepsilon(X^\varepsilon, Z^\varepsilon) \\ &\qquad + u^\varepsilon(X^\varepsilon, Z^\varepsilon) \otimes \chi^\varepsilon(X^\varepsilon, Z^\varepsilon)], \end{aligned}$$

which is the desired result. □

The next conjecture, which is crucial to justify Conjecture 4.31, should be rigorously proved using homogenization techniques. We only give an idea based on a multiscale expansion of the solution.

Conjecture 4.33. *Let u^ε be the solution of* (4.54) *with*

$$\chi^\varepsilon(x,z) = \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)V'(z) - V'(z)\otimes V'(x)(\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}})^{-1}\mathfrak{p}^\varepsilon,$$

where

$$\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}} := \mathbb{E}^{\mu^\varepsilon_{\text{exp}}}[V'(Z^\varepsilon)\otimes V'(X^\varepsilon)], \quad \mathfrak{p}^\varepsilon := \mathbb{E}^{\mu^\varepsilon_{\text{exp}}}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right].$$

Then, $u^\varepsilon \to 0$ "in some sense" for $\varepsilon \to 0$.

Idea of proof. We present here a formal proof based on asymptotic expansion with respect to ε. Let us first remark that by definition

$$\mathbb{E}^{\mu^\varepsilon_{\text{exp}}}[\chi^\varepsilon(X^\varepsilon, Z^\varepsilon)] = 0,$$

and therefore problem (4.54) reads

$$-\mathcal{L}^\varepsilon_{\text{exp}}u^\varepsilon = \chi^\varepsilon.$$

Let us now denote $y = x/\varepsilon$ and write

$$u^\varepsilon(x,z) = u_0(x,y,z) + \varepsilon u_1(x,y,z) + \varepsilon^2 u_2(x,y,z) + \dots,$$

which implies that

$$\begin{aligned}\partial_x u^\varepsilon &= \partial_x(u_0(x,y,z) + \varepsilon u_1(x,y,z) + \varepsilon^2 u_2(x,y,z) + \dots) \\ &\quad + \frac{1}{\varepsilon}\partial_y(u_0(x,y,z) + \varepsilon u_1(x,y,z) + \varepsilon^2 u_2(x,y,z) + \dots).\end{aligned}$$

Let us first remark that from the proof of Theorem 2.12 we have that

$$\lim_{\varepsilon\to 0}(\widetilde{\mathcal{M}}^\varepsilon_{\text{exp}})^{-1}\mathfrak{p}^\varepsilon = A - \alpha.$$

Replacing u^ε in (4.54) and grouping the terms of order ε^0, ε^{-1} and ε^{-2} we get the system

$$L_0 u_0 = 0, \tag{4.58}$$

$$L_1 u_0 + L_0 u_1 = -p'(y)V'(z), \tag{4.59}$$

$$L_0 u_2 + L_1 u_1 + L_2 u_0 = (V'(z)\otimes V'(x))(A-\alpha), \tag{4.60}$$

where

$$\begin{aligned} L_0 &= -p'(y)\partial_y + \sigma\partial^2_{yy}, \\ L_1 &= -p'(y)\partial_x - \alpha\cdot V'(x)\partial_y + 2\sigma\partial^2_{xy}, \\ L_2 &= -\alpha\cdot V'(x)\partial_x - \frac{1}{\delta}(z-x)\partial_z + \sigma\partial^2_{xx}. \end{aligned}$$

Let us first remark that equation (4.58) is satisfied for $u_0 = u_0(x,z)$ independent of y. In particular, the kernel of L_0 is made of constants and the kernel of L_0^* is one-dimensional and $\text{Ker}(L_0^*) = \text{Span}\{\omega\}$ where

$$\omega(y) = \frac{1}{C_\pi}e^{-p(y)/\sigma}, \quad \text{where} \quad C_\pi = \int_0^{\mathbb{T}} e^{-p(y)/\sigma}\,\mathrm{d}y, \tag{4.61}$$

where $\mathbb{T}$ is the period of p. Since u_0 is independent of y equation (4.59) reduces to

$$L_0 u_1 = p'(y)\left(\partial_x u_0 - V'(z)\right).$$

Let us remark that the general solution u_1 can be written as

$$u_1(x,y,z) = \Phi(y)\left(\partial_x u_0(x,z) - V'(z)\right),$$

where $\Phi\colon \mathbb{R} \to \mathbb{R}$ satisfies

$$-L_0\Phi = -p'(y).$$

Let us remark that this cell problem is the same as (1.7). We consider now equation (4.60) and remark that we can rewrite it as

$$\begin{aligned} L_0 u_2 &= (V'(z)\otimes V'(x))(A-\alpha) - \left(2\Phi'(y) - \frac{1}{\sigma}p'(y)\Phi(y) + 1\right)\sigma\partial_{xx}^2 u_0 \\ &\quad + (\Phi'(y)+1)\left(\alpha\cdot V'(x)\right)\partial_x u_0 - \left(V'(z)\otimes V'(x)\right)\alpha\Phi'(y) + \frac{1}{\delta}(z-x)\partial_z u_0. \end{aligned}$$

By the Fredholm alternative, in order for (4.60) to have a solution we need the right hand side to have zero-mean with respect to ω in (4.61). Therefore, recalling that the homogenization coefficient $\mathcal{K}$ in (1.6) is given by

$$\mathcal{K} = \int_0^{\mathbb{T}} (1+\Phi'(y))^2\rho(y)\,\mathrm{d}y$$

and remarking that integrations by part allow to rewrite $\mathcal{K}$ as

$$\mathcal{K} = \int_0^{\mathbb{T}} (1+\Phi'(y))\rho(y)\,\mathrm{d}y, \quad \mathcal{K} = \int_0^{\mathbb{T}} \left(2\Phi'(y) - \frac{1}{\sigma}p'(y)\Phi(y) + 1\right)\omega(y)\,\mathrm{d}y,$$

we get

$$0 = (V'(z)\otimes V'(x))(A-\alpha-(\mathcal{K}-1)\alpha) - \mathcal{K}\sigma\partial_{xx}^2 u_0 + \left(\mathcal{K}\alpha\cdot V'(x)\right)\partial_x u_0 + \frac{1}{\delta}(z-x)\partial_z u_0.$$

Moreover, since $A = \mathcal{K}\alpha$ and $\Sigma = \mathcal{K}\sigma$, it implies

$$0 = -\Sigma\partial_{xx}^2 u_0 + \left(A\cdot V'(x)\right)\partial_x u_0 + \frac{1}{\delta}(z-x)\partial_z u_0,$$

which can be written as

$$-\mathcal{L}_0 u_0 = 0, \tag{4.62}$$

where $\mathcal{L}_0$ is the generator of the couple $(X^0, Z^0)^\top$. Finally, the unique solution of (4.62) satisfying

$$\mathbb{E}^{\mu_{\exp}^0}[u_0(X^0, Z^0)] = 0,$$

is given by

$$u_0(x,z) = 0.$$

□

We can finally give an intuition behind the central limit theorem stated at the beginning of the section.

Idea of proof of Conjecture 4.31. Let us introduce the notation

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

and the notation

$$\mathfrak{p}^\varepsilon := \mathbb{E}^{\mu_{\exp}^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right]$$

and let us remark that from the proof of Theorem 2.12 one can deduce the equalities

$$A = \frac{1}{\delta}(\widetilde{\mathcal{M}}^0_{\exp})^{-1}\, \mathbb{E}^{\mu^0_{\exp}}[(X^0 - Z^0)^2 V''(Z^0)], \tag{4.63}$$

and

$$\alpha = (\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1} \left(\frac{1}{\delta}\, \mathbb{E}^{\mu^\varepsilon_{\exp}}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)] - \mathfrak{p}^\varepsilon \right). \tag{4.64}$$

Let us furthermore remark that the decomposition (2.16) yields

$$\sqrt{T}\left(\widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T) - A\right) = \sqrt{T}\left(\alpha - A + I_1^\varepsilon - I_2^\varepsilon\right).$$

Replacing the expression for A and α given in (4.63) and (4.64), we get

$$\begin{aligned} \sqrt{T}\left(\widehat{A}^{\delta,\beta}_{\exp}(X^\varepsilon, T) - A\right) &= \sqrt{T}\left(I_1^\varepsilon(T) - (\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1}\mathfrak{p}^\varepsilon\right) \\ &\quad + \frac{\sqrt{T}}{\delta}(\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1}\, \mathbb{E}^{\mu^\varepsilon_{\exp}}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)] \\ &\quad - \frac{\sqrt{T}}{\delta}(\widetilde{\mathcal{M}}^0_{\exp})^{-1}\, \mathbb{E}^{\mu^0_{\exp}}[(X^0 - Z^0)^2 V''(Z^0)] - \sqrt{T} I_2^\varepsilon(T). \end{aligned}$$

We rewrite the term involving $I_2^\varepsilon(T)$ as

$$\sqrt{T} I_2^\varepsilon(T) = \frac{\sqrt{2\sigma}}{\sqrt{T}} \widetilde{M}_{\exp}(X^\varepsilon, T)^{-1} Q^\varepsilon(T),$$

where, since Z_t^ε is adapted with respect to the natural filtration $\mathcal{F}_t$ of the Wiener process $W := (W_t, t \geq 0)$, the quantity

$$Q^\varepsilon(T) := \int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}W_t,$$

is a martingale whose quadratic variation is given by

$$\langle Q^\varepsilon \rangle_T = \int_0^T V'(Z_t^\varepsilon) \otimes V'(Z_t^\varepsilon)\,\mathrm{d}t.$$

Since the ergodic theorem guarantees that

$$\lim_{T\to\infty} \frac{\langle Q^\varepsilon \rangle_T}{T} = \mathbb{E}^{\mu^\varepsilon_{\exp}}[V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)], \quad \text{in } L^1,$$

the martingale central limit theorem gives

$$\lim_{T\to\infty} \frac{1}{\sqrt{T}} Q^\varepsilon(T) = \Xi^\varepsilon, \quad \text{in law},$$

where $\Xi^\varepsilon \sim \mathcal{N}(0, \mathbb{E}^{\mu^\varepsilon_{\exp}}[V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)])$. Now, by the ergodic theorem

$$\lim_{T\to\infty} \widetilde{M}_{\exp}(X^\varepsilon, T)^{-1} = (\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1}, \quad \text{a.s.}$$

Therefore, we have by Slutsky's theorem that

$$\lim_{T\to\infty} \sqrt{T} I_2^\varepsilon(T) = \Lambda^\varepsilon \sim \mathcal{N}\left(0, \Gamma^\varepsilon\right), \quad \text{in law}.$$

where the covariance matrix Γ^ε is given by

$$\Gamma^\varepsilon = 2\sigma(\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-1}\, \mathbb{E}^{\mu^\varepsilon_{\exp}}[V'(Z^\varepsilon) \otimes V'(Z^\varepsilon)](\widetilde{\mathcal{M}}^\varepsilon_{\exp})^{-\top}.$$

Finally, we denote by Λ the limit for $\varepsilon \to 0$ of the sequence of Gaussian random variables Λ^ε, whose covariance matrix Γ^0 is given by

$$\Gamma^0 = 2\sigma(\widetilde{\mathcal{M}}^0_{\mathrm{exp}})^{-1}\, \mathbb{E}^{\mu^0_{\mathrm{exp}}}[V'(Z^0) \otimes V'(Z^0)](\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-\top}.$$

Let us introduce the notation

$$J_1^\varepsilon(T) \coloneqq \sqrt{T}\left(I_1^\varepsilon(T) - (\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1}\mathfrak{p}^\varepsilon\right),$$

and let us remark that we can rewrite

$$\begin{aligned} J_1^\varepsilon(T) = \sqrt{T}&\left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t\right)^{-1} \\ &\times \left(\frac{1}{T}\int_0^T \left(\frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)V'(Z_t^\varepsilon) - V'(Z_t^\varepsilon)\otimes V'(X_t^\varepsilon)(\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1}\mathfrak{p}^\varepsilon\right)\mathrm{d}t\right). \end{aligned}$$

Let us denote by u^ε the solution of (4.54) with right hand side

$$\chi^\varepsilon(x,z) = \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)V'(z) - V'(z)\otimes V'(x)(\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1}\mathfrak{p}^\varepsilon,$$

and note that $\mathbb{E}^{\mu^\varepsilon_{\mathrm{exp}}}[\chi^\varepsilon(X^\varepsilon, Z^\varepsilon)] = 0$. Then, by Lemma 4.32, we have

$$J_1^\varepsilon(T) = \sqrt{T}\left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\otimes V'(X_t^\varepsilon)\,\mathrm{d}t\right)^{-1}\left(-\frac{\mathscr{R}^\varepsilon(T)}{T} + \sqrt{2\sigma}\frac{\mathscr{S}^\varepsilon(T)}{T}\right),$$

where $\mathscr{R}^\varepsilon(T)$ and $\mathscr{S}^\varepsilon(T)$ are defined in (4.56). Since $\mathscr{R}^\varepsilon(T)$ is bounded independently of ε, we first get by the ergodic theorem

$$\lim_{T\to\infty}\left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\otimes V'(X_t^\varepsilon)\,\mathrm{d}t\right)^{-1}\frac{\mathscr{R}^\varepsilon(T)}{\sqrt{T}} = 0, \quad \text{a.s.}$$

Repeating the same reasoning as for $Q^\varepsilon(T)$ and employing the ergodic theorem and the Slutsky's theorem, we get

$$\lim_{T\to\infty} J_1^\varepsilon(T) = \lim_{T\to\infty}\sqrt{2\sigma}\left(\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\otimes V'(X_t^\varepsilon)\,\mathrm{d}t\right)^{-1}\frac{\mathscr{S}^\varepsilon(T)}{\sqrt{T}} = \Lambda_1^\varepsilon \sim \mathcal{N}(0,\Gamma_1^\varepsilon),$$

where due to (4.57) the covariance is given by

$$\begin{aligned} \Gamma_1^\varepsilon &= 2\sigma(\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1}\,\mathbb{E}^{\mu^\varepsilon_{\mathrm{exp}}}[\partial_x u^\varepsilon(X^\varepsilon,Z^\varepsilon)\otimes\partial_x u^\varepsilon(X^\varepsilon,Z^\varepsilon)](\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-\top} \\ &= (\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1}\,\mathbb{E}^{\mu^\varepsilon_{\mathrm{exp}}}[u^\varepsilon(X^\varepsilon,Z^\varepsilon)\otimes\chi^\varepsilon(X^\varepsilon,Z^\varepsilon) + \chi^\varepsilon(X^\varepsilon,Z^\varepsilon)\otimes u^\varepsilon(X^\varepsilon,Z^\varepsilon)](\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-\top}. \end{aligned}$$

Since due to Conjecture 4.33 the solution $u^\varepsilon \to 0$ in some sense, we expect the covariance Γ_1^ε to behave similarly in the limit $\varepsilon \to 0$. Let us now introduce the notation

$$J_2^\varepsilon(T) = \frac{\sqrt{T}}{\delta}\left((\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1}\,\mathbb{E}^{\mu^\varepsilon_{\mathrm{exp}}}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)] - (\widetilde{\mathcal{M}}^0_{\mathrm{exp}})^{-1}\,\mathbb{E}^{\mu^0_{\mathrm{exp}}}[(X^0 - Z^0)^2 V''(Z^0)]\right).$$

We have the decomposition

$$\begin{aligned} J_2^\varepsilon(T) &= \frac{\sqrt{T}}{\delta}\left((\widetilde{\mathcal{M}}^\varepsilon_{\mathrm{exp}})^{-1} - (\widetilde{\mathcal{M}}^0_{\mathrm{exp}})^{-1}\right)\mathbb{E}^{\mu^\varepsilon_{\mathrm{exp}}}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)] \\ &\quad + \frac{\sqrt{T}}{\delta}(\widetilde{\mathcal{M}}^\varepsilon_0)^{-1}\left(\mathbb{E}^{\mu^\varepsilon_{\mathrm{exp}}}[(X^\varepsilon - Z^\varepsilon)^2 V''(Z^\varepsilon)] - \mathbb{E}^{\mu^0_{\mathrm{exp}}}[(X^0 - Z^0)^2 V''(Z^0)]\right) \\ &\eqqcolon J_{2,1}^\varepsilon(T) + J_{2,2}^\varepsilon(T). \end{aligned}$$

In order to send the terms above to zero, we need the convergence rate of $\mu^{\varepsilon}_{\text{exp}}$ to μ^{0}_{exp} w.r.t. ε, and some arguments in this sense seem to appear in [62]. Then, setting $T = \varepsilon^{-\gamma}$ with an appropriate γ sufficiently small, the quantity $J^{\varepsilon}_{2}(T)$ tends to zero as ε vanishes and implies the expected result. $\square$

Part II

Estimating functions based on eigenpairs of the generator: inference from discrete observations

5 Inference and homogenization problems

This chapter is devoted to the presentation of estimators based on martingale estimating functions constructed with the eigenvalues and eigenfuncions of the generator of the dynamics, for inference problems given a sequence of discrete data. We first introduce the homogenization problem for the eigenpairs of the generator of the multiscale Langevin dynamics and the second model under investigation, i.e., interacting diffusions, and then provide an overview of the literature on this topic. Finally, we present the main contributions and give the outline of the second part of this thesis.

5.1 Problem setting

In the section below we explain in more details and introduce the main notation to analyze the homogenization of the eigenvalue problem for the generator of the multiscale Langevin dynamics. We then want to apply the inference methodologies based on eigenfunction estimators to the simple model of fast/slow stochastic differential equations (SDEs) for which the theory of homogenization exists, that enables us to study the inference problem in a rigorous and systematic manner, and to systems of weakly interacting diffusions for which the mean field limit exists and is described by a nonlinear diffusion process of McKean type, obtained in the limit as the number of interacting processes goes to infinity. Interacting particle systems are therefore outlined below.

5.1.1 Multiscale diffusions and eigenvalue problem for the generator

In Chapters 6 and 7 we study the same class of diffusion processes introduced in Chapter 1, i.e., multiscale Langevin dynamics.

Before focusing on the inference problem, in Chapter 6 we present theoretical results for the generator of the multiscale SDE and, in particular, about the homogenization of its eigenvalues and eigenfunctions. In fact, the estimators which we introduce later are based on the eigenpairs of the generator of the homogenized dynamics, which we verify to be the limit in some sense which will be more clear later of the eigenpairs of the generator of the multiscale dynamics. We consider the multiscale and homogenized Langevin SDEs (1.3) and (1.9)

$$\mathrm{d}X^\varepsilon(t) = -\nabla\mathcal{V}(X^\varepsilon(t))\,\mathrm{d}t - \frac{1}{\varepsilon}\nabla p\left(\frac{X^\varepsilon(t)}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W(t), \tag{5.1}$$

$$\mathrm{d}X^0(t) = -\mathcal{K}\nabla\mathcal{V}(X^0(t))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t). \tag{5.2}$$

We recall that, under Assumption 1.4, it has been shown in [103] that the processes $X^\varepsilon(t)$ and $X^0(t)$ are geometrically ergodic with unique invariant distributions ν^ε and ν^0 whose densities with respect to the Lebesgue measure are given by

$$\varphi^\varepsilon(x) = \frac{1}{C_{\nu^\varepsilon}} e^{-\frac{1}{\sigma}\left(\mathcal{V}(x)+p\left(\frac{x}{\varepsilon}\right)\right)} \qquad \text{with} \qquad C_{\nu^\varepsilon} = \int_{\mathbb{R}^d} e^{-\frac{1}{\sigma}\left(\mathcal{V}(x)+p\left(\frac{x}{\varepsilon}\right)\right)}\,\mathrm{d}x, \tag{5.3}$$

and

$$\varphi^0(x) = \frac{1}{C_{\nu^0}} e^{-\frac{1}{d}\operatorname{tr}(\Sigma^{-1}\mathcal{K})\mathcal{V}(x)} \qquad \text{with} \qquad C_{\nu^0} = \int_{\mathbb{R}^d} e^{-\frac{1}{d}\operatorname{tr}(\Sigma^{-1}\mathcal{K})\mathcal{V}(x)} \,\mathrm{d}x. \tag{5.4}$$

Notice that $\operatorname{tr}(\Sigma^{-1}\mathcal{K})/d = 1/\sigma$ since $\Sigma = \mathcal{K}\sigma$ and that $\varphi^\varepsilon \rightharpoonup \varphi^0$ in $L^1(\mathbb{R}^d)$ by [103, Proposition 5.2]. We then introduce the generators $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$ of the multiscale process (5.1) and (5.2), respectively, which are defined for all $u \in \mathcal{C}^2(\mathbb{R}^d)$ as

$$\mathcal{L}^\varepsilon u(x) = -\left(\nabla\mathcal{V}(x) + \frac{1}{\varepsilon}\nabla p\left(\frac{x}{\varepsilon}\right)\right) \cdot \nabla u(x) + \sigma\Delta u(x), \tag{5.5}$$

and

$$\mathcal{L}^0 = -\mathcal{K}\nabla\mathcal{V}(x) \cdot \nabla u(x) + \Sigma : \nabla^2 u(x), \tag{5.6}$$

where : denotes the Frobenius inner product and ∇^2 the Hessian matrix. Since the process $X^\varepsilon(t)$ is close in a weak sense to the process $X^0(t)$ as $\varepsilon \to 0$, we then expect that also the generators $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$ behave similarly when the multiscale parameter vanishes. In particular, we consider the infinitesimal generator $\mathcal{L}^\varepsilon$ of (5.1) and study first the partial differential equation (PDE)

$$-\mathcal{L}^\varepsilon u^\varepsilon + \eta u^\varepsilon = f, \tag{5.7}$$

for a right-hand side f and where $\eta > 0$, and then the eigenvalue problem

$$-\mathcal{L}^\varepsilon \phi^\varepsilon = \lambda^\varepsilon \phi^\varepsilon. \tag{5.8}$$

We analyze the homogenization of problems (5.7) and (5.8) providing asymptotic results for their solutions in the limit of vanishing ε. In particular, we show that they converge to the solutions of the corresponding problems for the generator $\mathcal{L}^0$ of the homogenized diffusion (5.2). We remark that these equations are defined on the whole space $\mathbb{R}^d$, and this leads us to the introduction of weighted Sobolev spaces where the weight function is the invariant density of the homogenized process (5.2). The proof of the convergence results relies on the theory of two-scale convergence, which we extend to the case of weighted Sobolev spaces in order to make it fit into our framework.

We then study the parameter estimation problem in Chapter 7. We consider $M+1$ uniformly distributed observation times $0 = t_0 < t_1 < t_2 < \dots, < t_M = T$, set $\Delta = t_m - t_{m-1}$ and let $(X^\varepsilon(t))_{t\in[0,T]}$ be a realization of the solution of (1.10). We assume to know a sample $\{\widetilde{X}_m^\varepsilon\}_{m=0}^M$ of the realization where $\widetilde{X}_m^\varepsilon = X^\varepsilon(t_m)$ and we aim to estimate the drift coefficient A of the homogenized equation (1.11). We employ the techinque of eigenfunction martingale estimating functions which is suitable for inference problems where a sequence of discrete observations is given. Moreover, by observing that if the sampling rate 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 Chapter 2, which has been proved to be beneficial for correcting the behavior of the maximum likelihood estimator (MLE) in the setting of continuous observations.

Remark 5.1. For clarity of the presentation, in Chapter 7 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 theory presented can be generalized to the case of multidimensional diffusion processes in $\mathbb{R}^d$, for which we provide further details in Section 7.7 and an example in Section 7.3.5. However, the problem becomes more complex and computationally expensive from a numerical viewpoint and it can be prohibitive if the dimension d is too large, since the methodology proposed requires the solution of the eigenvalue problem for the generator of a d-dimensional diffusion process.

5.1.2 Interacting particle systems

In Chapter 8 we employ eigenfunction estimators for a different model. In particular, we consider a system of interacting particles in one dimension moving in a confining potential $\mathcal{V}$ over the time interval $[0,T]$ whose interaction is governed by an interaction potential $\mathcal{F}$

$$\begin{aligned} \mathrm{d}X_t^{(n)} &= -\mathcal{V}'(X_t^{(n)};\alpha)\,\mathrm{d}t - \frac{1}{N}\sum_{i=1}^N \mathcal{F}'(X_t^{(n)} - X_t^{(i)};\kappa)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t^{(n)}, \qquad n = 1,\dots,N, \\ X_0^{(n)} &\sim \gamma, \qquad n = 1,\dots,N, \end{aligned} \tag{5.9}$$

where N is the number of particles, $\{W_t^{(n)}\}_{n=1}^N$ are standard independent one dimensional Brownian motions, $\mathcal{V}(\cdot;\alpha)$ and $\mathcal{F}(\cdot;\kappa)$ are the confining and interaction potentials, respectively, which depend on some parameters $\alpha \in \mathbb{R}^{L_1}, \kappa \in \mathbb{R}^{L_2}$, and $\sigma > 0$ is the diffusion coefficient. The functions $\mathcal{V}'$ and $\mathcal{F}'$ are then the derivatives of $\mathcal{V}$ and $\mathcal{F}$ with respect to their first argument. We assume chaotic initial conditions, i.e., that the particles are initially distributed according to the same measure γ.

Remark 5.2. We consider the case when the particles move in one dimension for the clarity of exposition. In fact, the proposed method and our rigorous results can be easily generalized to the case of N interacting particles moving in dimension $d > 1$. However in higher dimensions the problem becomes more complex and expensive from the computational point of view.

We place ourselves in the same framework of [84], which is summarized in the following assumption.

Assumption 5.3. The confining and interaction potentials $\mathcal{V}$ and $\mathcal{F}$, respectively, satisfy:

(i) $\mathcal{V}(\cdot;\alpha) \in \mathcal{C}^2(\mathbb{R})$ is uniformly convex and polynomially bounded along with its derivatives uniformly in α;

(ii) $\mathcal{F}(\cdot;\kappa) \in \mathcal{C}^2(\mathbb{R})$ is even, convex and polynomially bounded along with its derivatives uniformly in κ.

It is well-known (see, e.g., [101, Chapter 4]) that under Assumption 5.3 the dynamics described by the system (5.9) is geometrically ergodic with unique invariant measure given by the Gibbs measure $\mu_\theta^N(\mathrm{d}x) = \rho^N(x;\theta)\,\mathrm{d}x$, where

$$\rho^N(x;\theta) = \frac{1}{C_{\mu_\theta^N}} \exp\left\{-\frac{1}{\sigma}\mathcal{E}^N(x;\theta)\right\}, \qquad C_{\mu_\theta^N} = \int_{\mathbb{R}^N} \exp\left\{-\frac{1}{\sigma}\mathcal{E}^N(x;\theta)\right\} \mathrm{d}x,$$

and $\mathcal{E}^N(\cdot;\theta)$ is defined by

$$\mathcal{E}^N(x;\theta) \coloneqq \sum_{n=1}^N \mathcal{V}(x_n;\alpha) + \frac{1}{2N}\sum_{n=1}^N\sum_{i=1}^N \mathcal{F}(x_n - x_i;\kappa).$$

for $\theta = \begin{pmatrix}\alpha^\top & \kappa^\top\end{pmatrix}^\top \in \Theta \subseteq \mathbb{R}^L$ with $L = L_1 + L_2$ and Θ the set of admissible parameters. The main goal is the estimation of the unknown parameter $\theta \in \Theta$, given discrete observations of the path of one single particle. We are interested in applications involving large interacting particle systems, i.e., when $N \gg 1$, hence studying the whole system is not practical and can be computationally unfeasible. Therefore, our approach consists of considering the mean field limit which has already been thoroughly studied (see, e.g., [38, 53]). Letting the number of particles N go to infinity we obtain the nonlinear, in the sense of McKean, SDE

$$\begin{aligned} \mathrm{d}X_t &= -\mathcal{V}'(X_t;\alpha)\,\mathrm{d}t - (\mathcal{F}'(\cdot;\kappa) * u(\cdot,t;\theta))(X_t)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\ X_0 &\sim \gamma, \end{aligned} \tag{5.10}$$

where $u(\cdot, t; \theta)$ is the density with respect to the Lebesgue measure of the law of X_t and the nonlinearity means that the drift of the SDE (5.10) depends on the law of the process. The density u is the solution of the nonlinear Fokker–Planck (McKean–Vlasov) equation

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial}{\partial x}\left(\mathcal{V}'(x;\alpha)u(x,t;\theta) + (\mathcal{F}'(\cdot;\kappa) * u(\cdot,t;\theta))(x,t)u(x,t;\theta) + \sigma\frac{\partial u}{\partial x}(x,t)\right),$$

with initial condition $u(x,0;\theta)\,\mathrm{d}x = \gamma(\mathrm{d}x)$. It is well known that, in contrast to the finite dimensional dynamics, the mean field limit (5.10) can have, in the non-convex case more than one invariant measures $\mu_\theta(\mathrm{d}x) = \rho(x;\theta)\,\mathrm{d}x$ [28, 38]. The density of the stationary state(s) satisfies the stationary Fokker–Planck equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathcal{V}'(x;\alpha)\rho(x;\theta) + (\mathcal{F}'(\cdot;\kappa) * \rho(\cdot;\theta))(x)\rho(x;\theta) + \rho'(x;\theta)\right) = 0,$$

where the second variable θ emphasizes the fact that ρ depends on the parameters α and κ of the potentials $\mathcal{V}$ and $\mathcal{F}$, respectively. However, under Assumption 5.3 it has been proven in [84] that there exists a unique invariant measure which is the solution of

$$\rho(x;\theta) = \frac{1}{C_{\mu_\theta}}\exp\left\{-\frac{1}{\sigma}\left(\mathcal{V}(x;\alpha) + (\mathcal{F}(\cdot;\kappa) * \rho(\cdot;\theta))(x)\right)\right\}, \tag{5.11}$$

where C_{μ_θ} is the normalization constant

$$C_{\mu_\theta} = \int_{\mathbb{R}} \exp\left\{-\frac{1}{\sigma}\left(\mathcal{V}(x;\alpha) + (\mathcal{F}(\cdot;\kappa) * \rho(\cdot;\theta))(x)\right)\right\}\mathrm{d}x.$$

Example 5.4. A particular choice for the interaction potential is the Curie–Weiss quadratic interaction [38], which is also known as harmonic potential. We take $\kappa > 0$ and consider the interaction potential

$$\mathcal{F}(x;\kappa) = \frac{\kappa}{2}x^2.$$

The interacting particle system (5.9) becomes, for all $n = 1, \dots, N$

$$\mathrm{d}X_t^{(n)} = -\mathcal{V}'(X_t^{(n)};\alpha)\,\mathrm{d}t - \kappa\left(X_t^{(n)} - \bar{X}_t^N\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t^{(n)},$$

where $\bar{X}_t^N$ denotes the empirical mean

$$\bar{X}_t^N = \frac{1}{N}\sum_{i=1}^N X_t^{(i)}.$$

This interaction term creates a tendency for the particles to relax toward the center of gravity of the ensemble and the parameter κ measures the strength of the interaction between the agents, hence this model provides a simple example of cooperative interaction.

The mean field limit (5.10) then becomes

$$\mathrm{d}X_t = -\mathcal{V}'(X_t;\alpha)\,\mathrm{d}t - \kappa\left(X_t - m_t\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t,$$

where m_t denotes the expectation of X_t, $m_t = \mathbb{E}[X_t]$, and its unique (when the confining potential $\mathcal{V}$ is convex) invariant measure $\mu_\theta(\mathrm{d}x) = \rho(x;\theta)\,\mathrm{d}x$ is given by

$$\rho(x;\theta) = \frac{1}{C_{\mu_\theta}}\exp\left\{-\frac{1}{\sigma}\left(\mathcal{V}(x;\alpha) + \kappa\left(\frac{1}{2}x^2 - mx\right)\right)\right\}, \tag{5.12}$$

with the constraint for the expectation with respect to the invariant measure

$$m = \int_{\mathbb{R}} x\rho(x;\theta)\,\mathrm{d}x, \tag{5.13}$$

and where

$$C_{\mu_\theta} = \int_{\mathbb{R}} \exp\left\{-\frac{1}{\sigma}\left(\mathcal{V}(x;\alpha) + \kappa\left(\frac{1}{2}x^2 - mx\right)\right)\right\}\mathrm{d}x.$$

Equation (5.13) is the self-consistency equation [38, 46, 59] that enables us to calculate the invariant measure and, then, the stationary state(s). In the case where the confining potential is quadratic, we have a system of linear SDEs and the mean field limit reduces to the mean field Ornstein-Uhlenbeck (OU) SDE. In this case the first moment vanishes, $m = 0$, and the invariant measure is unique (this is the case, of course, of arbitrary strictly convex confining potentials). The inference problem for the linear interacting particle system and for the corresponding mean field limit is easier than that of the general case. We emphasize that, unlike this present work, most earlier papers, e.g., [24, 70], focus on this linear case, i.e., on systems of weakly interacting linear SDEs. The estimator proposed and studied in Chapter 8 can be applied to arbitrary non-quadratic interaction and confining potentials.

In Chapter 8 we therefore present our method for the estimation of the unknown parameter $\theta = (\alpha, \kappa) \in \Theta \subseteq \mathbb{R}^L$, given discrete observation of a single particle of the system (5.9). We consider $M + 1$ equidistant observation times $0 = t_0 < t_1 < \cdots < t_M = T$, let $\Delta = t_m - t_{m-1}$ be the sampling rate and let $(X_t^{(n)})_{t\in[0,T]}$ be a realization of the n-th particle of the solution of the system (5.9) for some $n = 1, \dots, N$. We then aim to estimate the unknown parameter θ given a sample $\{\widetilde{X}_m^{(n)}\}_{m=0}^M$ of the realization where $\widetilde{X}_m^{(n)} = X_{t_m}^{(n)}$ and $t_m = \Delta m$.

5.2 Construction of eigenfunction estimators

In this section we present the eigenfunction estimators which we will employ in Chapters 7 and 8 to estimate unknown parameters in multiscale diffusions and interacting particle systems, respectively, based on discrete-time observations. This methodology has already been studied for one-scale problems without a martingale structure in [73], and our exposition is mainly based on this article. For the clarity of the exposition, we analyze only one-dimensional stochastic processes and parameters and refer to [73] for further details about the multidimensional cases. We consider the class of SDEs

$$\mathrm{d}X_t = b(X_t;\theta)\,\mathrm{d}t + h(X_t;\theta)\,\mathrm{d}W_t, \tag{5.14}$$

where W is a standard one-dimensional Brownian motion, b and h are the drift and diffusion functions, which are assumed to be known, and θ is the unknown parameter which varies in a subset Θ of $\mathbb{R}$. Then, the inference problem consists in estimating the exact value θ_0 given $M + 1$ equidistant discrete observations $X_{t_0}, X_{t_1}, \dots, X_{t_M}$, where $t_m - t_{m-1} = \Delta$. In order to construct the estimator we need the eigenvalues and eigenfunctions of the generator of (5.14) which solve the eigenvalue problem

$$-\mathcal{L}_\theta\phi(x;\theta) = -\lambda(\theta)\phi(x;\theta),$$

where $\mathcal{L}_\theta$ is defined for all twice continuously differentiable functions u as

$$\mathcal{L}_\theta u(x) = b(x;\theta)u'(x) + \frac{1}{2}h^2(x;\theta)u''(x).$$

The spectral theory of diffusion processes guarantees that the spectrum of $\mathcal{L}_\theta$ is discrete, the eigenvalues satisfy $0 \le \lambda_0(\theta) < \lambda_1(\theta) < \cdots < \lambda_j(\theta) \uparrow \infty$ and the eigenfunctions form an orthonormal basis of the weighted space $L^2_{\mu_\theta}(\mathbb{R}^d)$, where μ_θ is the invariant measure of the process X_t in (5.14). Moreover, under weak regularity conditions it can be shown that the following formula holds

$$\mathbb{E}\left[\phi(X_{t_m};\theta) \mid X_{t_{m-1}}\right] = e^{-\lambda(\theta)\Delta}\phi(X_{t_{m-1}};\theta),$$

and it will be fundamental in the analysis. We can therefore select the first J eigenpairs $\{(\phi_j,\lambda_j)\}_{j=1}^J$, which are the most relevant since the dependence on the past is mainly determined by the small eigenvalues, and define the martingale estimating function as

$$G_M^J(\theta) = \sum_{m=1}^M \sum_{j=1}^J \psi_j(X_{t_{m-1}};\theta)\left(\phi_j(X_{t_m};\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(X_{t_{m-1}};\theta)\right),$$

where $\{\psi_j(\cdot;\theta)\}_{j=1}^J$ is a set of smooth functions possibly dependent on the parameter θ. The eigenfunction estimator $\widehat{\theta}_M^J$ is finally defined as the zero of the function G_M^J, i.e, $G_M^J(\widehat{\theta}_M^J) = 0$. We now introduce some assumptions which are important to prove the asymptotic unbiasedness and normality of the proposed estimator.

Assumption 5.5. Let us define the following quantities:

$$\begin{aligned}
s(x;\theta) &= \exp\left(-2\int_0^x \frac{b(y;\theta)}{h^2(y;\theta)}\,\mathrm{d}y\right),\\
g_j(x,y,z;\theta) &= \psi_j(z;\theta)\left(\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)\right),\\
\bar{g}_j(x,y;\theta) &\coloneqq g_j(x,y,x;\theta)\\
Q_\Delta^\theta(x,y) &= \mu_\theta(x)\Pi_\Delta(y,x;\theta),
\end{aligned}$$

where μ_θ and Π_Δ are the invariant and transition densities, respectively. Then:

(i) the following holds for all $\theta \in \Theta$

$$\int_0^{+\infty} s(x;\theta)\,\mathrm{d}x = \int_{-\infty}^0 s(x;\theta)\,\mathrm{d}x = \infty \qquad \text{and} \qquad \int_{-\infty}^{+\infty} \frac{1}{s(x;\theta)h^2(x;\theta)}\,\mathrm{d}x = A(\theta) < \infty,$$

(ii) the functions $\bar{g}_j(x,y;\theta)$, $j = 1,\dots,J$, are in $L^2(Q_\Delta^{\theta_0})$ and continuously differentiable with respect to θ for all x and y,

(iii) the functions $\partial_\theta \bar{g}_j(x,y;\theta)$, $j = 1,\dots,J$, are locally dominated square integrable with respect to $Q_\Delta^{\theta_0}$ and it holds

$$\mathfrak{f}(\theta_0) = \sum_{j=1}^J \int \partial_\theta \bar{g}_j(x,y;\theta_0)\,\mathrm{d}Q_\Delta^{\theta_0}(x,y) \neq 0.$$

We can now state the main convergence result in [73], which will be employed in our analysis in Chapters 7 and 8.

Theorem 5.6. *Under Assumption 5.5 an estimator $\widehat{\theta}_M^J$, which solves the equation $G_M^J(\widehat{\theta}_M^J) = 0$, exists with a probability tending to one as $M \to \infty$. Moreover, it holds*

$$\lim_{M\to\infty} \widehat{\theta}_M^J = \theta_0, \qquad \textit{in probability},$$

and

$$\lim_{M\to\infty} \sqrt{M}(\widehat{\theta}_M^J - \theta_0) = \mathcal{N}\left(0, \frac{\mathfrak{v}(\theta_0)}{\mathfrak{f}^2(\theta_0)}\right), \qquad \textit{in distribution},$$

where

$$\mathfrak{v}(\theta_0) = \sum_{j=1}^{J} \sum_{k=1}^{J} \int \psi_j(x;\theta_0)\psi_k(x;\theta_0)a_{jk}(x;\theta_0)\mu_{\theta_0}(x)\,\mathrm{d}x,$$

with

$$a_{jk}(x;\theta_0) = \int \left(\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)\right)\left(\phi_k(y;\theta) - e^{-\lambda_k(\theta)\Delta}\phi_k(x;\theta)\right)\Pi_\Delta(y,x;\theta)\,\mathrm{d}y.$$

5.3 Literature review

Multiscale diffusion processes are a powerful tool in many applications, and in all these scenarios it is relevant to extract single-scale surrogates, which are effective for modeling the slowest component of the system, which often governs its macroscopic behavior. The literature about inference of unknown parameters in stochastic systems from continuous observations has been already analyzed in detail in Section 1.3. We add here that, in case the data consist of discrete observations instead of continuous time series, it is possible to employ estimators based on martingale properties and on a spectral decomposition of the generator of the stochastic process. In addition to the work of [73], other methodologies of this kind have been applied successfully to inference problems for single-scale SDEs [34, 35], as well as more recently for multiscale diffusions [36]. We also remark that a general theory for martingale estimating functions exists and is thoroughly outlined in [21]. They appear to be appropriate for multiscale problems due to their robustness properties.

We now focus on the study of the homogenization of the Poisson problem and the eigenvalue problem for the generator of the multiscale Langevin dynamics. We remark that the Poisson problem for elliptic operators corresponding to infinitesimal generators of diffusion processes has been thoroughly investigated in [98–100], where more probabilistic approaches and the method of corrector are employed. In particular, the authors prove the existence and uniqueness of the solution in suitable weighted Sobolev spaces and its continuity with respect to parameters in the equations. Moreover, the Poisson problem for an extended generator defined in terms of an appropriate version of the Dynkin formula is analyzed in [120]. Regarding the study of the homogenization of the eigenvalue problem for elliptic operators, several results exist in the context of bounded domains [11, 71, 72], and additional first-order corrections for the eigenvalues of the homogenized generator are provided in [89]. Our theoretical analysis is based on the notion of two-scale convergence, which was initially introduced in [91] and then studied in greater detail in [9, 10]. Our contribution to this field consists in the extension of this theory from Lebesgue spaces in bounded domains to the more general case of weighted Sobolev spaces in unbounded domains. An advantage of the two-scale convergence approach with respect to other techniques presented in the literature is the fact that it gives a mathematical justification to the formal asymptotic multiscale expansion which is usually employed to derive the homogenized equation.

In Chapter 8 we then consider interacting particle systems and, more generally interacting multiagent models, which appear frequently in the natural and social sciences. In addition to the well known applications, e.g., plasma physics [58] and stellar dynamics [22], new applications include, e.g., the modeling of chemotaxis [116], pedestrian dynamics [60, 83], crowd dynamics [86], urban modeling [43], models for opinion formation [52, 57], collective behavior [38], and models for systemic risk [54]. In many of these applications, the phenomenological models involve unknown parameters that need to be estimated from data. This is particularly the case for multiagent models used in the social sciences and in economics, where no physics-informed choices of parameters are available. Learning parameters or even models, in a nonparametric setting, from data is becoming an increasingly important aspect of the overall mathematical modeling strategy. This is particularly the case in view of the huge quantity of available data in different areas, which allows the development of accurate data-driven techniques for learning parameters from data.

Inference for large interacting systems has attracted considerable attention, starting from the work of Kasonga [70], in which the MLE was considered. In particular, it was proved that the MLE for estimating parameters in the drift, when the drift is linearly dependent on the parameters, given continuous time observations of *all* the particles of the N-particle system, is consistent and asymptotically normal in the limit as $N \to \infty$. In this setting, it is possible to test whether the particles are interacting or not, at least in the linear case, i.e., for a system of interacting OU processes. Consistency and asymptotic normality of the sieve estimator and an approximate MLE estimator, i.e., when discrete observations of all the particles are given, was studied in [24] in the same framework of linear dependence on the parameters for the drift and known diffusion coefficient. Moreover, MLE inference of the mean field OU SDE was also considered. Properties of the MLE for the McKean SDE, when a continuous path of the SDE is observed, were studied in [121]. Consistency of the MLE was proved and an application to a model for ionic diffusion was presented. The MLE estimator for the McKean SDE was also considered in [81] and numerical experiments for the mean field OU process were presented. The combined large particle and long time asymptotics, $N \to \infty$ and $T \to \infty$, of the MLE for the case of a quadratic interaction, i.e., for interacting OU processes, was studied in [29]. Unlike the previous works mentioned in this literature review, the case where only a single particle trajectory is observed was considered in this work. It was shown that the parameters in the drift can be estimated with optimal rate of convergence simultaneously in mean-field limit and in long-time dynamics. Offline and online inference for the McKean SDE was studied in [112]. Consistency and asymptotic normality of the offline MLE for the interacting particle system in the limit as the number of particles $N \to \infty$ was shown. In addition, an online parameter estimator for the mean field SDE was proposed, which evolves according to a continuous-time stochastic gradient descent algorithm on the asymptotic log-likelihood of the interacting particle system.

5.4 Our main contributions

In Chapter 6, which is based on our research article [125], we prove homogenization results using two-scale convergence for two different problems involving the generator of the multiscale Langevin dynamics. In particular, the main contribution of this chapter, in addition to the extension of the theory of two-scale convergence to weighted Sobolev spaces in unbounded domains, is the homogenization of the Poisson equation with a reaction term (5.7) and of the eigenvalue problem (5.8) for the generator of multiscale Langevin dynamics. We work with weighted Sobolev spaces in the whole space $\mathbb{R}^d$ and we show:

(i) strong convergence in L^2 sense and weak convergence in H^1 sense of the solution of the multiscale equation (5.7) to the solution of the corresponding homogenized problem,

(ii) convergence of the eigenvalues of the multiscale generator to the corresponding eigenvalues of the homogenized generator;

(iii) strong convergence in L^2 sense and weak convergence in H^1 sense of the eigenvectors of the multiscale generator to the corresponding eigenvectors of the homogenized generator.

On the other hand, the main goal of Chapters 7 and 8, which are based on our research articles [6, 105], is to propose new robust algorithms based on martingale estimating functions for learning parameters in coarse-grained models from observations originated by more complex phenomena. In Chapter 7 we focus on multiscale diffusions and combine two main ideas:

(i) 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,

(ii) the filtering methodology for smoothing the data in order to make them compatible with the homogenized model, which was introduced in Chapter 2.

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

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

(ii) 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. Hence, this second estimator is not sensitive with respect to the sampling rate.

In Chapter 8 we adopt a similar approach for interacting particle systems and our main contributions are summarized below.

(i) We propose a new methodology for estimating parameters in the drift of large interacting particle systems when a sequence of discrete observations of a single particle is given. Our proposed estimator is based on the eigenvalues and eigenfunctions of the generator of the mean field SDE at the steady state.

(ii) We show theoretically that our estimator is asymptotically unbiased and asymptotically normal in the limit as the number of observations and the number of particles go to infinity and we compute the rate of convergence.

(iii) We demonstrate numerically that our proposed estimator is reliable and robust with respect to the sampling rate.

6 Homogenization of the generator of multiscale Langevin dynamics

In this chapter, which is based on our research article [125], we study the homogenization of the Poisson problem and the eigenvalue problem for the generator of the multiscale Langevin dynamics employing weighted Sobolev spaces and the theory of two-scale convergence. The chapter is organized as follows. In Section 6.1 we introduce the weighted Sobolev spaces which are employed in the analysis and we present some preliminary results. Then, in Sections 6.2 and 6.3 we study the homogenization of the Poisson problem with a reaction term and of the eigenvalue problem for the generator, respectively, and in Section 6.4 we show numerical examples which confirm our theoretical findings. In Section 6.5 we summarize the main results of the chapter.

We first introduce the functional spaces which will be employed throughout the rest of the chapter.

Notation

Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a probability density function (either φ^ε or φ^0 defined in (5.3) and (5.4)) and $\mathcal{T}$ be the cell defined in (1.6).

- $L^2_\varphi(\mathbb{R}^d)$ is the space of measurable functions $u\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$\|u\|_{L^2_\varphi(\mathbb{R}^d)} \coloneqq \left(\int_{\mathbb{R}^d} u(x)^2 \varphi(x) \,\mathrm{d}x \right)^{1/2} < \infty.$$

- $L^2_\varphi(\mathbb{R}^d \times \mathcal{T})$ is the space of measurable functions $u\colon \mathbb{R}^d \times \mathcal{T} \to \mathbb{R}$ such that

$$\|u\|_{L^2_\varphi(\mathbb{R}^d \times \mathcal{T})} \coloneqq \left(\int_{\mathbb{R}^d} \int_{\mathcal{T}} u(x,y)^2 \varphi(x) \,\mathrm{d}y \,\mathrm{d}x \right)^{1/2} < \infty.$$

- $H^1_\varphi(\mathbb{R}^d)$ is the space of measurable weakly differentiable functions $u\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$\|u\|_{H^1_\varphi(\mathbb{R}^d)} \coloneqq \left(\int_{\mathbb{R}^d} u(x)^2 \varphi(x) \,\mathrm{d}x + \int_{\mathbb{R}^d} \|\nabla u(x)\|^2 \varphi(x) \,\mathrm{d}x \right)^{1/2} < \infty.$$

- $C^k_{\mathrm{per}}(\mathcal{T})$ with $k \in \mathbb{N}$ is the subspace of $C^k(\mathbb{R}^d)$ of $\mathcal{T}-$periodic functions.
- $H^1_{\mathrm{per}}(\mathcal{T})$ is the closure of $C^\infty_{\mathrm{per}}(\mathcal{T})$ with respect to the norm in $H^1(\mathcal{T})$.
- $\mathcal{W}_{\mathrm{per}}(\mathcal{T})$ is the quotient space $H^1_{\mathrm{per}}(\mathcal{T})/\mathbb{R}$ and it is endowed with the norm

$$\|u\|_{\mathcal{W}_{\mathrm{per}}(\mathcal{T})} = \|\nabla u\|_{L^2(\mathcal{T})}.$$

- $L^2_\varphi(\mathbb{R}^d; C^0_{\mathrm{per}}(\mathcal{T}))$ is the space of measurable functions $u\colon x \mapsto u(x) \in C^0_{\mathrm{per}}(\mathcal{T})$ such that $\|u(x)\|_{L^\infty(\mathcal{T})} \in L^2_\varphi(\mathbb{R}^d)$ and it is endowed with the norm

$$\|u\|_{L^2_\varphi(\mathbb{R}^d; C^0_{\mathrm{per}}(\mathcal{T}))} = \left(\int_{\mathbb{R}^d} \sup_{y \in \mathcal{T}} |u(x,y)|^2 \varphi(x) \,\mathrm{d}x \right)^{1/2}.$$

- $L^2_\varphi(\mathbb{R}^d; \mathcal{W}_{\mathrm{per}}(\mathcal{T}))$ is the space of measurable functions $u\colon x \mapsto u(x) \in \mathcal{W}_{\mathrm{per}}(\mathcal{T})$ such that $\|u(x)\|_{\mathcal{W}_{\mathrm{per}}(\mathcal{T})} \in L^2_\varphi(\mathbb{R}^d)$ and it is endowed with the norm

$$\|u\|_{L^2_\varphi(\mathbb{R}^d;\mathcal{W}_{\mathrm{per}}(\mathcal{T}))} = \left(\int_{\mathbb{R}^d} \int_{\mathcal{T}} \|\nabla_y u(x,y)\|^2 \varphi(x) \,\mathrm{d}y \,\mathrm{d}x \right)^{1/2}.$$

6.1 Preliminary results

In this section we introduce the main functional spaces which will be employed in the following analysis and we study their relations. Let us consider the weighted Sobolev spaces $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$, $L^2_{\varphi^0}(\mathbb{R}^d)$, $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and $H^1_{\varphi^0}(\mathbb{R}^d)$, where the weight functions are the densities of invariant measures ν^ε and ν^0 defined in (5.3) and (5.4), where we recall that p is $\mathcal{T}$-periodic. First, we show that the weighted Lebesgue spaces $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$ and $L^2_{\varphi^0}(\mathbb{R}^d)$ describe the same space of functions but they are endowed with different norms.

Lemma 6.1. *Under Assumption 1.4(i), there exist two constants $C_{\mathrm{low}}, C_{\mathrm{up}} > 0$ independent of ε such that*

$$C_{\mathrm{low}} \|u\|_{L^2_{\varphi^0}(\mathbb{R}^d)} \le \|u\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \le C_{\mathrm{up}} \|u\|_{L^2_{\varphi^0}(\mathbb{R}^d)}.$$

In particular, the injections $\mathcal{I}_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^0}(\mathbb{R}^d)}$ and $\mathcal{I}_{L^2_{\varphi^0}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}$ are continuous.

Proof. Since $p \in C^\infty(\mathbb{R}^d)$ is $\mathcal{T}$-periodic, then there exists a constant $M > 0$ such that $|p(y)| \le M$ for all $y \in \mathbb{R}^d$. Therefore, we have

$$0 < e^{-\frac{M}{\sigma}} \le e^{-\frac{1}{\sigma} p\left(\frac{x}{\varepsilon}\right)} \le e^{\frac{M}{\sigma}},$$

which implies

$$e^{-\frac{M}{\sigma}} \|u\|_{L^2_{\varphi^0}(\mathbb{R}^d)} \le \|u\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \le e^{\frac{M}{\sigma}} \|u\|_{L^2_{\varphi^0}(\mathbb{R}^d)}.$$

Finally, defining $C_{\mathrm{low}} \coloneqq e^{-\frac{M}{\sigma}}$ and $C_{\mathrm{up}} \coloneqq e^{\frac{M}{\sigma}}$ we obtain the desired result. □

An analogous result holds true also for the weighted Sobolev spaces $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and $H^1_{\varphi^0}(\mathbb{R}^d)$ and follows directly from Lemma 6.1.

Corollary 6.2. *Under Assumption 1.4(i), there exist two constants $C_{\mathrm{low}}, C_{\mathrm{up}} > 0$ independent of ε such that*

$$C_{\mathrm{low}} \|u\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \le \|u\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \le C_{\mathrm{up}} \|u\|_{H^1_{\varphi^0}(\mathbb{R}^d)}.$$

In particular, the injections $\mathcal{I}_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \hookrightarrow H^1_{\varphi^0}(\mathbb{R}^d)}$ and $\mathcal{I}_{H^1_{\varphi^0}(\mathbb{R}^d) \hookrightarrow H^1_{\varphi^\varepsilon}(\mathbb{R}^d)}$ are continuous.

Let us now consider the injections $\mathcal{I}_{H^1_{\varphi^0}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^0}(\mathbb{R}^d)}$ and $\mathcal{I}_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}$, which are continuous since by definition we have

$$\|u\|_{L^2_{\varphi^0}(\mathbb{R}^d)} \le \|u\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \qquad \text{and} \qquad \|u\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \le \|u\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)}.$$

We remark that these injections are not compact in general, differently from classical non weighted Sobolev spaces in bounded and regular domains, where the compactness is always guaranteed by the Rellich–Kondrachov theorem [44, Theorem 5.7.1]. Hence, in order to ensure the compactness of the injections $\mathcal{I}_{H^1_{\varphi^0}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^0}(\mathbb{R}^d)}$ and $\mathcal{I}_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}$ we make the following additional assumption.

Assumption 6.3. There exists a constant $\beta > 2$ such that the slow-scale potential $\mathcal{V}$ satisfies

$$\lim_{\|x\|\to+\infty} \left(\frac{1}{\beta} \|\nabla \mathcal{V}(x)\|^2 - \Delta \mathcal{V}(x) \right) = +\infty \qquad \text{and} \qquad \lim_{\|x\|\to+\infty} \|\nabla \mathcal{V}(x)\| = +\infty.$$

Then, Assumption 6.3 implies condition (82) in [5], and due to [5, Proposition A.4] it follows that the injection $\mathcal{I}_{H^1_{\varphi^0}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^0}(\mathbb{R}^d)}$ is compact and the measure φ^0 satisfies the Poincaré inequality for all $u \in H^1_{\varphi^0}(\mathbb{R}^d)$ and for a constant $C_P > 0$

$$\int_{\mathbb{R}^d} (u(x) - \bar{u}^0)^2 \varphi^0(x) \,\mathrm{d}x \le C_P \int_{\mathbb{R}^d} \|\nabla u(x)\|^2 \varphi^0(x) \,\mathrm{d}x, \tag{6.1}$$

where $\bar{u}^0 = \int_{\mathbb{R}^d} u(x)\varphi^0(x)\,\mathrm{d}x$.

Remark 6.4. Assumption 6.3 is satisfied, e.g., by quadratic potentials in $\mathbb{R}^d$ of the form $\mathcal{V}(x) = \frac{1}{2}x^\top Dx$, where $D \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, and by the bistable potential $\mathcal{V}(x) = x^4/4 - x^2/2$ in $\mathbb{R}$. Moreover, Assumption 6.3 is not the only sufficient condition to ensure the compactness of the injection $\mathcal{I}_{H^1_{\varphi^0}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^0}(\mathbb{R}^d)}$. Two other necessary and sufficient conditions are presented in Proposition 1.3 and Lemma 2.2 in [49]. In particular, it is required that the potential $\mathcal{V}$ is such that either the Schrödinger operator

$$\mathcal{S} = -\Delta + \frac{1}{4} \left\|\nabla \varphi^0\right\|^2 - \frac{1}{2}\Delta\varphi^0,$$

or the operator

$$\mathcal{P} = -\Delta + \nabla\varphi^0 \cdot \nabla,$$

has compact resolvent. Moreover, another sufficient condition is given in [66, Theorem 3.1], where it is proved that the potentials of the form $V = \|x\|^{2p}$ with p integer greater than zero satisfy the condition.

Given Assumption 6.3 and using [5, Proposition A.4], we can now prove that the same compactness result holds true also for the spaces $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$.

Lemma 6.5. *Under Assumptions 1.4 and 6.3, the injection $\mathcal{I}_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \hookrightarrow L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}$ is a compact operator and the measure φ^ε satisfies the Poincaré inequality for all $u \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and for a constant $\widetilde{C}^\varepsilon_P > 0$*

$$\int_{\mathbb{R}^d} (u(x) - \bar{u}^\varepsilon)^2 \varphi^\varepsilon(x) \,\mathrm{d}x \le \widetilde{C}^\varepsilon_P \int_{\mathbb{R}^d} \|\nabla u(x)\|^2 \varphi^\varepsilon(x) \,\mathrm{d}x, \tag{6.2}$$

where $\bar{u}^\varepsilon = \int_{\mathbb{R}^d} u(x)\varphi^\varepsilon(x)\,\mathrm{d}x$.

Proof. Let $\mathcal{V}^\varepsilon$ be defined as

$$\mathcal{V}^\varepsilon(x) = \mathcal{V}(x) + p\left(\frac{x}{\varepsilon}\right).$$

Then, due to Assumption 1.4(i) there exists a constant $M > 0$ such that $|p(y)| \le M$, $\|\nabla p(y)\| \le M$ and $|\Delta p(y)| \le M$ for all $y \in \mathcal{T}$. We now show that for any $\varepsilon > 0$ the potential $\mathcal{V}^\varepsilon$ satisfies condition (82) in [5]. In fact, by the triangle inequality we first have

$$\|\nabla\mathcal{V}^\varepsilon(x)\| = \left\|\nabla\mathcal{V}(x) + \frac{1}{\varepsilon}\nabla p\left(\frac{x}{\varepsilon}\right)\right\| \ge \|\nabla\mathcal{V}(x)\| - \frac{1}{\varepsilon}\left\|\nabla p\left(\frac{x}{\varepsilon}\right)\right\| \ge \|\nabla\mathcal{V}(x)\| - \frac{M}{\varepsilon},$$

which due to Assumption 6.3 implies that

$$\lim_{\|x\|\to+\infty} \|\nabla \mathcal{V}^\varepsilon(x)\| \geq \lim_{\|x\|\to+\infty} \|\nabla \mathcal{V}(x)\| - \frac{M}{\varepsilon} = +\infty.$$

Moreover, let β be given in Assumption 6.3, and notice that by Young's inequality we get

$$\begin{aligned} \|\nabla \mathcal{V}^\varepsilon(x)\|^2 &\geq \|\nabla \mathcal{V}(x)\|^2 + \frac{1}{\varepsilon^2} \left\| \nabla p\left(\frac{x}{\varepsilon}\right)\right\|^2 - \frac{2}{\varepsilon} \|\nabla \mathcal{V}(x)\| \left\| \nabla p\left(\frac{x}{\varepsilon}\right)\right\| \\ &\geq \|\nabla \mathcal{V}(x)\|^2 - \frac{\beta-2}{\beta} \|\nabla \mathcal{V}(x)\|^2 - \frac{\beta}{(\beta-2)\varepsilon^2} \left\| \nabla p\left(\frac{x}{\varepsilon}\right)\right\|^2 \\ &\geq \frac{2}{\beta} \|\nabla \mathcal{V}(x)\|^2 - \frac{\beta M^2}{(\beta-2)\varepsilon^2}, \end{aligned}$$

which due to Assumption 6.3 yields

$$\begin{aligned} \lim_{\|x\|\to+\infty} \left(\frac{1}{4} \|\nabla \mathcal{V}^\varepsilon(x)\|^2 - \frac{1}{2}\Delta \mathcal{V}^\varepsilon(x)\right) &\geq \frac{1}{2} \lim_{\|x\|\to+\infty} \left(\frac{1}{\beta} \|\nabla \mathcal{V}(x)\|^2 - \Delta \mathcal{V}(x)\right) - \frac{\beta M^2}{4(\beta-2)\varepsilon^2} - \frac{M}{2\varepsilon^2} \\ &= +\infty. \end{aligned}$$

Therefore, condition (82) in [5] is satisfied, and following the same argument of [5, Proposition A.4] we obtain the desired result. □

6.2 Poisson equation with a reaction term

Let us recall that the generators $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$ of the multiscale and homogenized diffusions are defined in equations (5.5) and (5.6), respectively. In this section we study the problem for the multiscale generator

$$-\mathcal{L}^\varepsilon u^\varepsilon + \eta u^\varepsilon = f, \tag{6.3}$$

with $f \in L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$ and where the reaction term with coefficient $\eta > 0$ is added in order to ensure the well-posedness of the problem, and we analyze its homogenization. In particular, we show that the solution u^ε converges in some sense which will be specified later to the solution u^0 of the Poisson problem for the homogenized generator with a reaction term

$$-\mathcal{L}^0 u^0 + \eta u^0 = f, \tag{6.4}$$

where, in view of Lemma 6.1, f is now seen as a function of $L^2_{\varphi^0}(\mathbb{R}^d)$.

Remark 6.6. We decided to study the Poisson equation with a reaction term with coefficient $\eta > 0$ so that, as we will see later, the bilinear form of the corresponding weak formulation is coercive. This guarantees the well-posedness of the problem without additional conditions on the solution and on the right-hand side, which would be otherwise needed if the bilinear form was only weakly coercive as in the case $\eta = 0$. Moreover, this partial differential equation (PDE) will be useful in the study of the homogenization of the eigenvalue problem for the generator, which is the focus of Section 6.3 and the main purpose of this chapter. We finally remark that the solutions to the equations (6.3) and (6.4) depend not only on ε, but also on the choice of the parameter η. However, in order to simplify the notation, we decided not to include η as a subscript of the solutions u^ε and u^0.

6.2.1 Weak formulation

We first write the weak formulation of problems (6.3) and (6.4) and, applying the Lax–Milgram lemma, we prove that they admit a unique solution respectively in the spaces $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and

$H^1_{\varphi^0}(\mathbb{R}^d)$. Since the proof is analogous for both the cases, we present the details only in the multiscale setting. Letting $\psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ be a test function, multiplying equation (6.3) by $\psi(x)\varphi^\varepsilon(x)$ and integrating over $\mathbb{R}^d$ and by parts we obtain

$$\sigma \int_{\mathbb{R}^d} \nabla u^\varepsilon(x) \cdot \nabla \psi(x) \varphi^\varepsilon(x) \,\mathrm{d}x + \eta \int_{\mathbb{R}^d} u^\varepsilon(x)\psi(x)\varphi^\varepsilon(x) \,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)\psi(x)\varphi^\varepsilon(x) \,\mathrm{d}x.$$

Therefore, the weak formulation of problem (6.3) reads:

$$\text{find } u^\varepsilon \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \text{ such that } B^\varepsilon(u^\varepsilon, \psi) = F^\varepsilon(\psi) \text{ for all } \psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d), \tag{6.5}$$

where $B^\varepsilon \colon H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \times H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \to \mathbb{R}$ and $F^\varepsilon \colon H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \to \mathbb{R}$ are defined as

$$\begin{aligned} B^\varepsilon(v, \psi) &= \sigma \int_{\mathbb{R}^d} \nabla v(x) \cdot \nabla \psi(x) \varphi^\varepsilon(x) \,\mathrm{d}x + \eta \int_{\mathbb{R}^d} v(x)\psi(x)\varphi^\varepsilon(x) \,\mathrm{d}x, \\ F^\varepsilon(\psi) &= \int_{\mathbb{R}^d} f(x)\psi(x)\varphi^\varepsilon(x) \,\mathrm{d}x. \end{aligned} \tag{6.6}$$

Similarly, the weak formulation of problem (6.4) reads:

$$\text{find } u^0 \in H^1_{\varphi^0}(\mathbb{R}^d) \text{ such that } B^0(u^0, \psi) = F^0(\psi) \text{ for all } \psi \in H^1_{\varphi^0}(\mathbb{R}^d), \tag{6.7}$$

where $B^0 \colon H^1_{\varphi^0}(\mathbb{R}^d) \times H^1_{\varphi^0}(\mathbb{R}^d) \to \mathbb{R}$ and $F^0 \colon H^1_{\varphi^0}(\mathbb{R}^d) \to \mathbb{R}$ are defined as

$$\begin{aligned} B^0(v, \psi) &= \int_{\mathbb{R}^d} \Sigma \nabla v(x) \cdot \nabla \psi(x) \varphi^0(x) \,\mathrm{d}x + \eta \int_{\mathbb{R}^d} v(x)\psi(x)\varphi^0(x) \,\mathrm{d}x, \\ F^0(\psi) &= \int_{\mathbb{R}^d} f(x)\psi(x)\varphi^0(x) \,\mathrm{d}x. \end{aligned} \tag{6.8}$$

Then, the well-posedness of the two problems is given by the following lemmas.

Lemma 6.7. *Problem* (6.5) *has a unique solution* $u^\varepsilon \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ *which satisfies*

$$\left\| u^\varepsilon \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \le \frac{1}{\min\{\sigma, \eta\}} \left\| f \right\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}. \tag{6.9}$$

Proof. The existence and uniqueness of the solution follow from the Lax–Milgram lemma once we show the continuity and coercivity of B^ε and the continuity of F^ε defined in (6.6). Applying the Cauchy–Schwarz inequality we obtain

$$\left| B^\varepsilon(v, \psi) \right| \le 2 \max\{\sigma, \eta\} \left\| v \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \left\| \psi \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)},$$

and

$$\left| F^\varepsilon(\psi) \right| \le \left\| f \right\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \left\| \psi \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)}, \tag{6.10}$$

which prove the continuity of B^ε and F^ε. Moreover, we also have

$$B^\varepsilon(\psi, \psi) \ge \min\{\sigma, \eta\} \left\| \psi \right\|^2_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)}, \tag{6.11}$$

which gives the coercivity of B^ε. Finally, due to inequalities (6.10) and (6.11) we deduce

$$\min\{\sigma, \eta\} \left\| u^\varepsilon \right\|^2_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \le B^\varepsilon(u^\varepsilon, u^\varepsilon) = F^\varepsilon(u^\varepsilon) \le \left\| f \right\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \left\| u^\varepsilon \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)}$$

which implies estimate (6.9) and concludes the proof. $\square$

Lemma 6.8. *Problem* (6.7) *has a unique solution* $u^0 \in H^1_{\varphi^0}(\mathbb{R}^d)$ *which satisfies*

$$\left\| u^0 \right\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \le \frac{1}{\min\{\lambda_{\min}(\Sigma), \eta\}} \left\| f \right\|_{L^2_{\varphi^0}(\mathbb{R}^d)},$$

where $\lambda_{\min}(\Sigma) > 0$ *is the smallest eigenvalue of the matrix* Σ*.*

We omit the proof of Lemma 6.8 since it follows the same argument of Lemma 6.7.

6.2.2 Two-scale convergence

We now focus on the homogenization of problem (6.5) and our strategy is based on the two-scale convergence method outlined in [30, Chapter 9]. We remark that we extend this theory to the case of weighted Sobolev spaces in unbounded domains, hence also the definition of two-scale convergence has to be adapted and it is given in Definition 6.9. We first introduce some preliminary results, and in the last part of this section we prove the main convergence theorem.

Definition 6.9. A sequence of functions $\{v^\varepsilon\}$ in $L^2_{\varphi^0}(\mathbb{R}^d)$ is said to *two-scale converge* to the limit $v^0 \in L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T})$ if for any function $\psi \in L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ it holds

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} v^\varepsilon(x) \psi\left(x, \frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x = \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} v^0(x,y) \psi(x,y) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x.$$

We then write $v^\varepsilon \rightsquigarrow v^0$.

Remark 6.10. From Definition 6.9 it follows that two-scale convergence implies weak convergence. In fact, choosing ψ independent of y we obtain

$$v^\varepsilon \rightharpoonup \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} v^0(\cdot, y) \,\mathrm{d}y \qquad \text{in } L^2_{\varphi^0}(\mathbb{R}^d),$$

and if also the two-scale limit is independent of y then $v^\varepsilon \rightharpoonup v^0$ in $L^2_{\varphi^0}(\mathbb{R}^d)$.

The following lemmas are technical results which will be useful in the proof of next theorems. The former studies the properties of the space $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ and the latter is a convergence result for two-scale functions in the same space.

Lemma 6.11. *The space* $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ *is separable and dense in* $L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T})$*.*

Proof. Since the space $C^0_{\text{per}}(\mathcal{T})$ is separable, then by [30, Proposition 3.55] it follows that the space $L^2(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ is separable. Moreover, $L^2(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ is isomorphic to $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ through the isomorphism

$$\mathscr{I} \colon L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T})) \to L^2(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T})), \qquad u \mapsto \mathscr{I}(u) = \sqrt{\varphi^0} u,$$

and thus the space $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ is separable as well. Concerning the density result, since $\mathcal{D}(\mathcal{T})$ is dense in $L^2(\mathcal{T})$, then $L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{D}(\mathcal{T}))$ is dense in $L^2_{\varphi^0}(\mathbb{R}^d; L^2(\mathcal{T}))$. Finally, the property that $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ is dense in $L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T})$ follows from the inclusion $L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{D}(\mathcal{T})) \subset L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$ and the fact that $L^2_{\varphi^0}(\mathbb{R}^d; L^2(\mathcal{T})) = L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T})$. $\square$

Lemma 6.12. *Let* $\psi \in L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$*. Then*

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} \psi\left(x, \frac{x}{\varepsilon}\right)^2 \varphi^0(x) \,\mathrm{d}x = \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} \psi(x,y)^2 \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x. \tag{6.12}$$

Proof. The proof follows the same steps of the proof of Lemma 5.2 in [9], where the spaces $L^1(\Omega)$ and $L^1(\Omega; C^0_{\text{per}}(Y))$ are replaced by $L^2_{\varphi^0}(\mathbb{R}^d)$ and $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$, respectively. Accordingly, the integrals $\int_\Omega v(x)\,\mathrm{d}x$ for a function $v = v(x)$ are replaced by $\int_{\mathbb{R}^d} v(x)\varphi^0(x)\,\mathrm{d}x$. $\square$

The following propositions are compactness results in the spaces $L^2_{\varphi^0}(\mathbb{R}^d)$ and $H^1_{\varphi^0}(\mathbb{R}^d)$, respectively, which highlight the importance of the notion of two-scale convergence and thus justify the introduction of Definition 6.9. The proof of Proposition 6.14 is based on the proof of Theorem 9.9 in [30].

Proposition 6.13. *Let $\{v^\varepsilon\}$ be a bounded sequence in $L^2_{\varphi^0}(\mathbb{R}^d)$. Then, there exist a subsequence $\{v^{\varepsilon'}\}$ and a function $v^0 \in L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T})$ such that*

$$v^{\varepsilon'} \rightsquigarrow v^0.$$

Proof. The proof follows the same steps of the proof of Theorem 9.7 in [30], where Proposition 3.61, equation (9.2) and the spaces $L^2(\Omega)$, $L^2(\Omega \times Y)$, $L^2(\Omega; C^0_{\text{per}}(Y))$ are replaced by Lemma 6.11, equation (6.12) and $L^2_{\varphi^0}(\mathbb{R}^d)$, $L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T})$, $L^2_{\varphi^0}(\mathbb{R}^d; C^0_{\text{per}}(\mathcal{T}))$, respectively. Accordingly, the integrals $\int_\Omega v(x)\,\mathrm{d}x$ for a function $v = v(x)$ are replaced by $\int_{\mathbb{R}^d} v(x)\varphi^0(x)\,\mathrm{d}x$. $\square$

Proposition 6.14. *Let $\{v^\varepsilon\}$ be a sequence of functions in $H^1_{\varphi^0}(\mathbb{R}^d)$ such that*

$$v^\varepsilon \rightharpoonup v^0 \qquad \text{in } H^1_{\varphi^0}(\mathbb{R}^d). \tag{6.13}$$

Then, $v^\varepsilon \rightsquigarrow v^0$ and there exist a subsequence $\{v^{\varepsilon'}\}$ and $v_1 \in L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{W}_{\text{per}}(\mathcal{T}))$ such that

$$\nabla v^{\varepsilon'} \rightsquigarrow \nabla v^0 + \nabla_y v_1.$$

Proof. By Proposition 6.13, there exists a subsequence (still denoted by ε) such that

$$v^\varepsilon \rightsquigarrow v \in L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T}) \qquad \text{and} \qquad \nabla v^\varepsilon \rightsquigarrow \Xi \in (L^2_{\varphi^0}(\mathbb{R}^d \times \mathcal{T}))^d. \tag{6.14}$$

Letting $\psi \in (\mathcal{D}(\mathbb{R}^d; C^\infty_{\text{per}}(\mathcal{T})))^d$ and integrating by parts we have

$$\begin{aligned}\int_{\mathbb{R}^d} \nabla v^\varepsilon(x) \cdot \psi\left(x, \frac{x}{\varepsilon}\right) \varphi^0(x)\,\mathrm{d}x = &- \int_{\mathbb{R}^d} v^\varepsilon(x) \left[\operatorname{div}_x \psi\left(x, \frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon} \operatorname{div}_y \psi\left(x, \frac{x}{\varepsilon}\right)\right] \varphi^0(x)\,\mathrm{d}x \\ &+ \frac{1}{\sigma} \int_{\mathbb{R}^d} v^\varepsilon(x) \psi\left(x, \frac{x}{\varepsilon}\right) \cdot \nabla \mathcal{V}(x) \varphi^0(x)\,\mathrm{d}x,\end{aligned}$$

which implies

$$\begin{aligned}\int_{\mathbb{R}^d} v^\varepsilon(x) \operatorname{div}_y \psi\left(x, \frac{x}{\varepsilon}\right) \varphi^0(x)\,\mathrm{d}x = &\ \varepsilon \int_{\mathbb{R}^d} v^\varepsilon(x) \left[\frac{1}{\sigma} \psi\left(x, \frac{x}{\varepsilon}\right) \cdot \nabla \mathcal{V}(x) - \operatorname{div}_x \psi\left(x, \frac{x}{\varepsilon}\right)\right] \varphi^0(x)\,\mathrm{d}x \\ &- \varepsilon \int_{\mathbb{R}^d} \nabla v^\varepsilon(x) \cdot \psi\left(x, \frac{x}{\varepsilon}\right) \varphi^0(x)\,\mathrm{d}x.\end{aligned}$$

Passing to the limit as $\varepsilon \to 0$ and due to (6.14) we obtain

$$\frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} v(x, y) \operatorname{div}_y \psi(x, y) \varphi^0(x)\,\mathrm{d}y\,\mathrm{d}x = 0,$$

which yields for all $\psi \in (\mathcal{D}(\mathbb{R}^d \times \mathcal{T}))^d$

$$\int_{\mathbb{R}^d} \int_{\mathcal{T}} \nabla_y v(x, y) \cdot \psi(x, y) \varphi^0(x)\,\mathrm{d}y\,\mathrm{d}x = 0.$$

Hence, by [30, Theorem 1.44] and since $\varphi^0(x) > 0$ for all $x \in \mathbb{R}^d$ we get

$$\nabla_y v = 0 \qquad a.e. \text{ on } \mathbb{R}^d \times \mathcal{T}.$$

Therefore, from [30, Proposition 3.38] with Ω replaced by $\mathcal{T}$ and x fixed we deduce that v does not depend on y and due to Remark 6.10 and hypothesis (6.13) this implies that $v = v^0 \in H^1_{\varphi^0}(\mathbb{R}^d)$. Let now $\psi \in (\mathcal{D}(\mathbb{R}^d; C^\infty_{\text{per}}(\mathcal{T})))^d$ such that $\operatorname{div}_y \psi = 0$. Integrating by parts and by (6.14) we obtain

$$\begin{aligned}
\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} \nabla v^\varepsilon(x) \cdot \psi\left(x, \frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x \\
&= \lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} v^\varepsilon(x) \left[\frac{1}{\sigma}\psi\left(x, \frac{x}{\varepsilon}\right) \cdot \nabla \mathcal{V}(x) - \operatorname{div}_x \psi\left(x, \frac{x}{\varepsilon}\right)\right] \varphi^0(x) \,\mathrm{d}x \\
&= \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} v^0(x) \left[\frac{1}{\sigma}\psi(x,y) \cdot \nabla \mathcal{V}(x) - \operatorname{div}_x \psi(x,y)\right] \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x \\
&= \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} \nabla v^0(x) \cdot \psi(x,y) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x.
\end{aligned}$$

Due to (6.14) we also have

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} \nabla v^\varepsilon(x) \cdot \psi\left(x, \frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x = \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} \Xi(x,y) \cdot \psi(x,y) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x,$$

and defining $\widetilde{\psi}(x,y) = \sqrt{\varphi^0(x)}\psi(x,y)$ it follows that

$$\int_{\mathbb{R}^d} \int_{\mathcal{T}} \sqrt{\varphi^0(x)} \left[\Xi(x,y) - \nabla v^0(x)\right] \cdot \widetilde{\psi}(x,y) \,\mathrm{d}y \,\mathrm{d}x = 0,$$

for all $\widetilde{\psi} \in (\mathcal{D}(\mathbb{R}^d; C^\infty_{\text{per}}(\mathcal{T})))^d$ such that $\operatorname{div}_y \psi = 0$. Therefore, by a classical result (see, e.g., [55, 118]) there exists a unique function $\widetilde{v}_1 \in L^2(\mathbb{R}^d; \mathcal{W}_{\text{per}}(\mathcal{T}))$ such that

$$\left(\Xi(x,y) - \nabla v^0(x)\right) \sqrt{\varphi^0(x)} = \nabla_y \widetilde{v}_1(x,y).$$

Finally, defining $v_1 \in L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{W}_{\text{per}}(\mathcal{T}))$ as $v_1(x,y) = \widetilde{v}_1(x,y)/\sqrt{\varphi^0(x)}$ gives the desired result. □

6.2.3 Homogenization

We are now ready to state and prove the homogenization of problem (6.3) employing the two-scale convergence methodology introduced in the previous section. The proof of next theorem is inspired by [30, Section 9.3].

Theorem 6.15. *Let u^ε and u^0 be respectively the unique solutions of problems* (6.5) *and* (6.7)*. Then, under Assumptions 1.4 and 6.3 and as $\varepsilon \to 0$*

(i) $u^\varepsilon \to u^0$ in $L^2_{\varphi^0}(\mathbb{R}^d)$,

(ii) $u^\varepsilon \rightharpoonup u^0$ in $H^1_{\varphi^0}(\mathbb{R}^d)$.

Proof. By Lemmas 6.1 and 6.7 and Corollary 6.2 we have

$$\|u^\varepsilon\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \le \frac{1}{C_{\text{low}}} \|u\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \le \frac{1}{C_{\text{low}} \min\{\sigma, \eta\}} \|f\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \le \frac{C_{\text{up}}}{C_{\text{low}} \min\{\sigma, \eta\}} \|f\|_{L^2_{\varphi^0}(\mathbb{R}^d)},$$

which implies that the sequence $\{u^\varepsilon\}$ is bounded in $H^1_{\varphi^0}(\mathbb{R}^d)$. Then, there exist $\widetilde{u} \in H^1_{\varphi^0}(\mathbb{R}^d)$ and a subsequence (still denoted by ε) such that

$$u^\varepsilon \rightharpoonup \widetilde{u} \quad \text{in } H^1_{\varphi^0}(\mathbb{R}^d) \qquad \text{and} \qquad u^\varepsilon \to \widetilde{u} \quad \text{in } L^2_{\varphi^0}(\mathbb{R}^d).$$

Due to Proposition 6.14 there exists $u_1 \in L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{W}_{\mathrm{per}}(\mathcal{T}))$ such that, up to a subsequence

$$u^\varepsilon \rightsquigarrow \widetilde{u} \qquad \text{and} \qquad \nabla u^\varepsilon \rightsquigarrow \nabla \widetilde{u} + \nabla_y u_1.$$

We now want to prove that $\widetilde{u}$ is the unique solution of problem (6.7), i.e., $\widetilde{u} = u^0$. Let $\psi_0 \in \mathcal{D}(\mathbb{R}^d)$ and $\psi_1 \in \mathcal{D}(\mathbb{R}^d; C^\infty_{\mathrm{per}}(\mathcal{T}))$ and note that $\psi_0(\cdot) + \varepsilon\psi_1\left(\cdot, \frac{\cdot}{\varepsilon}\right) \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and thus it can be chosen as a test function in (6.5). We then have

$$\begin{aligned} \sigma \int_{\mathbb{R}^d} \nabla u^\varepsilon(x) \cdot \left(\nabla\psi_0(x) + \varepsilon \nabla_x \psi_1\left(x, \frac{x}{\varepsilon}\right) + \nabla_y \psi_1\left(x, \frac{x}{\varepsilon}\right)\right) \varphi^\varepsilon(x) \,\mathrm{d}x \\ + \eta \int_{\mathbb{R}^d} u^\varepsilon(x) \left(\psi_0(x) + \varepsilon\psi_1\left(x, \frac{x}{\varepsilon}\right)\right) \varphi^\varepsilon(x) \,\mathrm{d}x = \int_{\mathbb{R}^d} f(x) \left(\psi_0(x) + \varepsilon\psi_1\left(x, \frac{x}{\varepsilon}\right)\right) \varphi^\varepsilon(x) \,\mathrm{d}x, \end{aligned} \tag{6.15}$$

and noting that

$$\varphi^\varepsilon(x) = \frac{C_\pi C_{\nu^0}}{C_{\nu^\varepsilon}} \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x), \tag{6.16}$$

where π is defined in (1.8), equation (6.15) can be rewritten as

$$I^\varepsilon_{1,1} + I^\varepsilon_{1,2} + \varepsilon\left(I^\varepsilon_{2,1} + I^\varepsilon_{2,2}\right) = J^\varepsilon_1 + \varepsilon J^\varepsilon_2, \tag{6.17}$$

where

$$\begin{aligned} I^\varepsilon_{1,1} &\coloneqq \sigma \int_{\mathbb{R}^d} \nabla u^\varepsilon(x) \cdot \left(\nabla\psi_0(x) + \nabla_y\psi_1\left(x, \frac{x}{\varepsilon}\right)\right) \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x, \\ I^\varepsilon_{1,2} &\coloneqq \eta \int_{\mathbb{R}^d} u^\varepsilon(x)\psi_0(x)\omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x, \\ I^\varepsilon_{2,1} &\coloneqq \sigma \int_{\mathbb{R}^d} \nabla u^\varepsilon(x) \cdot \nabla_x\psi_1\left(x, \frac{x}{\varepsilon}\right) \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x, \\ I^\varepsilon_{2,2} &\coloneqq \eta \int_{\mathbb{R}^d} u^\varepsilon(x)\psi_1\left(x, \frac{x}{\varepsilon}\right) \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x, \\ J^\varepsilon_1 &\coloneqq \int_{\mathbb{R}^d} f(x)\psi_0(x)\omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x, \\ J^\varepsilon_2 &\coloneqq \int_{\mathbb{R}^d} f(x)\psi_1\left(x, \frac{x}{\varepsilon}\right) \omega\left(\frac{x}{\varepsilon}\right) \varphi^0(x) \,\mathrm{d}x. \end{aligned}$$

Passing to the limit as $\varepsilon \to 0$ in equation (6.17) and by two-scale convergence we obtain

$$\begin{aligned} \lim_{\varepsilon\to 0} I^\varepsilon_{1,1} &= \frac{\sigma}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} (\nabla\widetilde{u}(x) + \nabla_y u_1(x,y)) \cdot (\nabla\psi_0(x) + \nabla_y\psi_1(x,y))\, \omega(y)\varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x, \\ \lim_{\varepsilon\to 0} I^\varepsilon_{1,2} &= \frac{\eta}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} \widetilde{u}(x)\psi_0(x)\omega(y)\varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x = \frac{\eta}{|\mathcal{T}|} \int_{\mathbb{R}^d} \widetilde{u}(x)\psi_0(x)\varphi^0(x) \,\mathrm{d}x, \\ \lim_{\varepsilon\to 0} I^\varepsilon_{2,1} &= \frac{\sigma}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} (\nabla\widetilde{u}(x) + \nabla_y u_1(x,y)) \cdot \nabla_x\psi_1(x,y)\omega(y)\varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x, \\ \lim_{\varepsilon\to 0} I^\varepsilon_{2,2} &= \frac{\eta}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} \widetilde{u}(x)\psi_1(x,y)\omega(y)\varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x, \\ \lim_{\varepsilon\to 0} J^\varepsilon_1 &= \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} f(x)\psi_0(x)\omega(y)\varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x = \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} f(x)\psi_0(x)\varphi^0(x) \,\mathrm{d}x, \\ \lim_{\varepsilon\to 0} J^\varepsilon_2 &= \frac{1}{|\mathcal{T}|} \int_{\mathbb{R}^d} \int_{\mathcal{T}} f(x)\psi_1(x,y)\omega(y)\varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x, \end{aligned}$$

which yield

$$
\begin{aligned}
\sigma \int_{\mathbb{R}^d} \int_{\mathcal{T}} (\nabla \widetilde{u}(x) + \nabla_y u_1(x,y)) \cdot (\nabla \psi_0(x) + \nabla_y \psi_1(x,y))\, \omega(y) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x& \\
+ \eta \int_{\mathbb{R}^d} \widetilde{u}(x) \psi_0(x) \varphi^0(x) \,\mathrm{d}x = \int_{\mathbb{R}^d} f(x) \psi_0(x) \varphi^0(x) \,\mathrm{d}x.&
\end{aligned}
\tag{6.18}
$$

We now show that equation (6.18) is a variational equation in the functional space

$$
\mathcal{H} = H^1_{\varphi^0}(\mathbb{R}^d) \times L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{W}_{\mathrm{per}}(\mathcal{T})),
$$

endowed with the norm

$$
\|\Psi\|_{\mathcal{H}} = \left(\|\psi_0\|^2_{H^1_{\varphi^0}(\mathbb{R}^d)} + \|\psi_1\|^2_{L^2_{\varphi^0}(\mathbb{R}^d;\mathcal{W}_{\mathrm{per}}(\mathcal{T}))} \right)^{1/2}, \qquad \text{for all } \Psi = (\psi_0, \psi_1) \in \mathcal{H},
$$

and that the hypotheses of the Lax–Milgram lemma are satisfied. Let $a \colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be the bilinear form defined for any $\Xi = (\xi_0, \xi_1) \in \mathcal{H}$ and $\Psi = (\psi_0, \psi_1) \in \mathcal{H}$ by

$$
\begin{aligned}
a(\Xi, \Psi) = \sigma &\int_{\mathbb{R}^d} \int_{\mathcal{T}} (\nabla \xi_0(x) + \nabla_y \xi_1(x,y)) \cdot (\nabla \psi_0(x) + \nabla_y \psi_1(x,y))\, \omega(y) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x \\
&+ \eta \int_{\mathbb{R}^d} \xi_0(x) \psi_0(x) \varphi^0(x) \,\mathrm{d}x,
\end{aligned}
$$

and let $F \colon \mathcal{H} \to \mathbb{R}$ be the linear functional defined by

$$
F(\Psi) = \int_{\mathbb{R}^d} f(x) \psi_0(x) \varphi^0(x) \,\mathrm{d}x.
$$

Notice that due to the definition of ω in (1.8) and the hypotheses on p in Assumption 1.4(i) there exist two constants $C_1, C_2 > 0$ such that $0 < C_1 \leq |\omega(y)| \leq C_2$ for all $y \in Y$. It follows that a and F are continuous, in fact applying the Cauchy–Schwarz inequality we get

$$
|a(\Xi, \Psi)| \leq (2\sigma(1 + C_2) + \eta) \|\Xi\|_{\mathcal{H}} \|\Psi\|_{\mathcal{H}},
$$

and

$$
|F(\Psi)| \leq \|f\|_{L^2_{\varphi^0}(\mathbb{R}^d)} \|\Psi\|_{\mathcal{H}}.
$$

Moreover, we also have

$$
\begin{aligned}
a(\Psi, \Psi) &\geq C_1 \sigma \int_{\mathbb{R}^d} \int_{\mathcal{T}} \|\nabla \psi_0(x) + \nabla_y \psi_1(x,y)\|^2 \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x + \eta \int_{\mathbb{R}^d} \psi_0(x)^2 \varphi^0(x) \,\mathrm{d}x \\
&= C_1 \sigma |\mathcal{T}| \int_{\mathbb{R}^d} \|\nabla \psi_0(x)\|^2 \varphi^0(x) \,\mathrm{d}x + \eta \int_{\mathbb{R}^d} \psi_0(x)^2 \varphi^0(x) \,\mathrm{d}x \\
&\quad + C_1 \sigma \int_{\mathbb{R}^d} \int_{\mathcal{T}} \|\nabla_y \psi_1(x,y)\|^2 \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x \\
&\geq \min\{C_1 \sigma |\mathcal{T}|, \eta, C_1 \sigma\} \|\Psi\|_{\mathcal{H}},
\end{aligned}
$$

which shows that a is coercive and where we used the fact that due to the periodicity of $\psi_1(x, \cdot)$ in $\mathcal{T}$ for all $x \in \mathbb{R}^d$

$$
\begin{aligned}
\int_{\mathbb{R}^d} \int_{\mathcal{T}} \nabla \psi_0(x) \cdot \nabla_y \psi_1(x,y) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x &= \int_{\mathbb{R}^d} \int_{\mathcal{T}} \operatorname{div}_y (\nabla \psi_0(x) \psi_1(x,y)) \varphi^0(x) \,\mathrm{d}y \,\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \int_{\partial Y} \psi_1(x,y) \nabla \psi_0(x) \cdot \mathbf{n_y} \varphi^0(x) \,\mathrm{d}\gamma_y \,\mathrm{d}x = 0,
\end{aligned}
$$

where $\mathbf{n}_y$ denotes the outward unit normal vector to $\partial\mathcal{T}$. Therefore, the Lax-Milgram lemma gives the existence and uniqueness of the solution $U = (\widetilde{u}, u_1) \in \mathcal{H}$ of equation (6.18) for any $\Psi = (\psi_0, \psi_1) \in \mathcal{H}$. Then, notice that the components of the unique solution U must satisfy

$$\widetilde{u}(x) = u^0(x) \qquad \text{and} \qquad \nabla_y u_1(x,y) = (\nabla\Phi(y))^\top \nabla u^0(x),$$

where u^0 is the unique solution of problem (6.7) and Φ solves equation (1.7). In fact, replacing U into (6.18) we obtain

$$\begin{aligned}\left(\int_{\mathbb{R}^d} \sigma \left(\int_{\mathcal{T}} (I + \nabla\Phi(y)^\top)\omega(y)\,\mathrm{d}y\right) \nabla u^0(x) \cdot \nabla\psi_0(x)\varphi^0(x)\,\mathrm{d}x\right) + \eta \int_{\mathbb{R}^d} u^0(x)\psi_0(x)\varphi^0(x)\,\mathrm{d}x \\ + \sigma \int_{\mathbb{R}^d}\int_{\mathcal{T}} \left(I + \nabla\Phi(y)^\top\right) \nabla u^0(x) \cdot \nabla_y \psi_1(x,y)\omega(y)\varphi^0(x)\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)\psi_0(x)\varphi^0(x)\,\mathrm{d}x,\end{aligned} \tag{6.19}$$

and, due to definition (1.6) and problem (6.7), equation (6.19) holds true for all $\Psi = (\psi_0, \psi_1) \in \mathcal{H}$ if we show that for any $\psi_1 \in L^2_{\varphi^0}(\mathbb{R}^d; \mathcal{W}_{\mathrm{per}}(\mathcal{T}))$

$$I \coloneqq \sigma \int_{\mathbb{R}^d}\int_{\mathcal{T}} \left(I + \nabla\Phi(y)^\top\right) \nabla u^0(x) \cdot \nabla_y \psi_1(x,y)\omega(y)\varphi^0(x)\,\mathrm{d}y\,\mathrm{d}x = 0.$$

Integrating by parts and by definition of ω in (1.8) we indeed have

$$\begin{aligned} I &= \sigma \int_{\mathbb{R}^d}\int_{\partial\mathcal{T}} \left(I + \nabla\Phi(y)^\top\right) \nabla u^0(x)\psi_1(x,y)\omega(y)\varphi^0(x) \cdot \mathbf{n}_y\,\mathrm{d}\gamma_y\,\mathrm{d}x \\ &\quad - \int_{\mathbb{R}^d}\int_{\mathcal{T}} (\sigma\Delta\Phi(y) - \nabla\Phi(y)\nabla p(y) - \nabla p(y)) \cdot \nabla u^0(x)\psi_1(x,y)\omega(y)\varphi^0(x)\,\mathrm{d}y\,\mathrm{d}x \\ &= 0,\end{aligned}$$

where the last equality is given by (1.7) and the periodicity of the functions $\Phi, \psi_1(x,\cdot)$ and ω in $\mathcal{T}$. We have thus proved that the only admissible limit for the subsequence of $\{u^\varepsilon\}$ is the solution u^0 of problem (6.7), which implies that the whole sequence $\{u^\varepsilon\}$ converges to u^0 and completes the proof. □

The previous result can be generalized to the case where also the right-hand side depends on the multiscale parameter ε.

Corollary 6.16. *Let $\{f^\varepsilon\}$ be a sequence in $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$ such that $f^\varepsilon \to f^0$ in $L^2_{\varphi^0}(\mathbb{R}^d)$ and let u^ε be the unique solution of problem*

$$B^\varepsilon(u^\varepsilon, \psi) = \langle f^\varepsilon; \psi\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}, \qquad \textit{for all } \psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d), \tag{6.20}$$

where $\langle\cdot;\cdot\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}$ denotes the inner product in $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$. Then, under Assumptions 1.4 and 6.3 and as $\varepsilon \to 0$

$$u^\varepsilon \rightharpoonup u^0 \ \textit{in } H^1_{\varphi^0}(\mathbb{R}^d) \qquad \textit{and} \qquad u^\varepsilon \to u^0 \ \textit{in } L^2_{\varphi^0}(\mathbb{R}^d),$$

where u^0 is the unique solution of the problem

$$B^0(u^0, \psi) = \left\langle f^0; \psi\right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)}, \qquad \textit{for all } \psi \in H^1_{\varphi^0}(\mathbb{R}^d),$$

where $\langle\cdot;\cdot\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)}$ denotes the inner product in $L^2_{\varphi^0}(\mathbb{R}^d)$.

Proof. Let $\widetilde{u}^\varepsilon$ be the solution of problem

$$B^\varepsilon(\widetilde{u}^\varepsilon, \psi) = \left\langle f^0; \psi\right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}, \qquad \text{for all } \psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d), \tag{6.21}$$

and notice that by Theorem 6.15 and as $\varepsilon \to 0$

$$\widetilde{u}^\varepsilon \rightharpoonup u^0 \text{ in } H^1_{\varphi^0}(\mathbb{R}^d) \qquad \text{and} \qquad \widetilde{u}^\varepsilon \to u^0 \text{ in } L^2_{\varphi^0}(\mathbb{R}^d). \tag{6.22}$$

Consider now the difference between problems (6.20) and (6.21)

$$B^\varepsilon(u^\varepsilon - \widetilde{u}^\varepsilon, \psi) = \left\langle f^\varepsilon - f^0; \psi \right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)},$$

and choose $\psi = u^\varepsilon - \widetilde{u}^\varepsilon$. Since B^ε is coercive by (6.11) and using the Cauchy–Schwarz inequality we have

$$\begin{aligned}
\min\{\sigma, \eta\} \left\| u^\varepsilon - \widetilde{u}^\varepsilon \right\|^2_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} &\leq B^\varepsilon(u^\varepsilon - \widetilde{u}^\varepsilon, u^\varepsilon - \widetilde{u}^\varepsilon) \\
&= \left\langle f^\varepsilon - f^0; u^\varepsilon - \widetilde{u}^\varepsilon \right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \\
&\leq \left\| f^\varepsilon - f^0 \right\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \left\| u^\varepsilon - \widetilde{u}^\varepsilon \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)},
\end{aligned}$$

which implies

$$\left\| u^\varepsilon - \widetilde{u}^\varepsilon \right\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \leq \frac{1}{\min\{\sigma, \eta\}} \left\| f^\varepsilon - f^0 \right\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)},$$

and employing Lemma 6.1 and Corollary 6.2 we obtain

$$\left\| u^\varepsilon - \widetilde{u}^\varepsilon \right\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \leq \frac{C_{\mathrm{up}}}{C_{\mathrm{low}} \min\{\sigma, \eta\}} \left\| f^\varepsilon - f^0 \right\|_{L^2_{\varphi^0}(\mathbb{R}^d)}.$$

Therefore, since $f^\varepsilon \to f^0$ in $L^2_{\varphi^0}(\mathbb{R}^d)$ we deduce that $u^\varepsilon - \widetilde{u}^\varepsilon \to 0$ in $H^1_{\varphi^0}(\mathbb{R}^d)$, which together with the limits in (6.22) gives the desired result. □

Finally, the next theorem is a corrector result which justifies the two first term in the asymptotic expansion of the solution u^ε of (6.5)

$$u^\varepsilon(x) = u^0(x) + \varepsilon u_1\left(x, \frac{x}{\varepsilon}\right) + \varepsilon^2 u_2\left(x, \frac{x}{\varepsilon}\right) + \dots,$$

which is usually employed in homogenization theory.

Theorem 6.17. *Let u^ε and u^0 be respectively the unique solutions of problems* (6.5) *and* (6.7)*. Then, under Assumptions 1.4 and 6.3*

$$\lim_{\varepsilon \to 0} \left\| u^\varepsilon - u^0 - \varepsilon u_1\left(\cdot, \frac{\cdot}{\varepsilon}\right) \right\|_{H^1_{\varphi^0}(\mathbb{R}^d)} = 0,$$

where $u_1(x,y) = \Phi(y) \cdot \nabla u^0(x)$ and Φ is the solution of (1.7)*.*

Proof. Let us first recall that from the proof of Theorem 6.15 we know that as $\varepsilon \to 0$

$$u^\varepsilon \rightsquigarrow u^0 \qquad \text{and} \qquad \nabla u^\varepsilon \rightsquigarrow \nabla u^0 + \nabla_y u_1. \tag{6.23}$$

Let z^ε be defined as

$$z^\varepsilon(x) \coloneqq u^\varepsilon(x) - u^0(x) - \varepsilon u_1\left(x, \frac{x}{\varepsilon}\right),$$

and let $\bar{z}^\varepsilon$ be its mean with respect to the invariant distribution φ^0, i.e.,

$$\bar{z}^\varepsilon \coloneqq \int_{\mathbb{R}^d} z^\varepsilon(x) \varphi^0(x) \,\mathrm{d}x.$$

Then, applying the Poincaré inequality (6.1) we obtain

$$\begin{aligned} \|z^\varepsilon\|^2_{H^1_{\varphi^0}(\mathbb{R}^d)} &= \|z^\varepsilon\|^2_{L^2_{\varphi^0}(\mathbb{R}^d)} + \|\nabla z^\varepsilon\|^2_{(L^2_{\varphi^0}(\mathbb{R}^d))^d} \\ &= \|z^\varepsilon - \bar z^\varepsilon\|^2_{L^2_{\varphi^0}(\mathbb{R}^d)} + (\bar z^\varepsilon)^2 + \|\nabla z^\varepsilon\|^2_{(L^2_{\varphi^0}(\mathbb{R}^d))^d} \\ &\le (\bar z^\varepsilon)^2 + (C_P+1)\,\|\nabla z^\varepsilon\|^2_{(L^2_{\varphi^0}(\mathbb{R}^d))^d}\,, \end{aligned} \tag{6.24}$$

and we now study the two terms in the right-hand side separately. First, by the two-scale convergence (6.23) and the fact that Φ is bounded by [103, Lemma 5.5] we have

$$\lim_{\varepsilon\to 0} \bar z^\varepsilon = \lim_{\varepsilon\to0}\left(\int_{\mathbb{R}^d} u^\varepsilon(x)\varphi^0(x)\,\mathrm{d}x - \int_0 u^0(x)\varphi^0(x)\,\mathrm{d}x - \varepsilon\int_{\mathbb{R}^d}\Phi\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^0(x)\varphi^0(x)\,\mathrm{d}x\right) = 0. \tag{6.25}$$

We then consider the second term in the right-hand side of (6.24) and using Lemma 6.1 we have

$$\begin{aligned} \|\nabla z^\varepsilon\|^2_{(L^2_{\varphi^0}(\mathbb{R}^d))^d} &\le \frac{1}{C^2_{\mathrm{low}}}\,\|\nabla z^\varepsilon\|^2_{(L^2_{\varphi^\varepsilon}(\mathbb{R}^d))^d} \\ &= \frac{1}{C^2_{\mathrm{low}}}\int_{\mathbb{R}^d}\left\|\nabla u^\varepsilon(x) - \left(I+\nabla\Phi\left(\frac{x}{\varepsilon}\right)^\top\right)\nabla u^0(x) - \varepsilon\nabla^2u^0(x)\Phi\left(\frac{x}{\varepsilon}\right)\right\|^2\varphi^\varepsilon(x)\,\mathrm{d}x \\ &\le \frac{2}{C^2_{\mathrm{low}}}(I_1^\varepsilon + I_2^\varepsilon), \end{aligned}$$

where

$$\begin{aligned} I_1^\varepsilon &\coloneqq \varepsilon^2\int_{\mathbb{R}^d}\left\|\nabla^2u^0(x)\Phi\left(\frac{x}{\varepsilon}\right)\right\|^2\varphi^\varepsilon(x)\,\mathrm{d}x \\ I_2^\varepsilon &\coloneqq \int_{\mathbb{R}^d}\left\|\nabla u^\varepsilon(x) - \left(I+\nabla\Phi\left(\frac{x}{\varepsilon}\right)^\top\right)\nabla u^0(x)\right\|^2\varphi^\varepsilon(x)\,\mathrm{d}x. \end{aligned}$$

Since Φ and ω are bounded, due to equation (6.16) and noting that

$$\lim_{\varepsilon\to0}\frac{C_\pi C_{\nu^0}}{C_{\nu^\varepsilon}} = |\mathcal{T}|\,, \tag{6.26}$$

we obtain

$$\lim_{\varepsilon\to0} I_1^\varepsilon = \lim_{\varepsilon\to0}\varepsilon^2\frac{C_\pi C_{\nu^0}}{C_{\nu^\varepsilon}}\int_{\mathbb{R}^d}\left\|\nabla^2u^0(x)\Phi\left(\frac{x}{\varepsilon}\right)\right\|^2\omega\left(\frac{x}{\varepsilon}\right)\varphi^0(x)\,\mathrm{d}x = 0. \tag{6.27}$$

Moreover, since u^ε solves problem (6.5) we have

$$\begin{aligned} \sigma I_2^\varepsilon &= \int_{\mathbb{R}^d} f(x)u^\varepsilon(x)\varphi^\varepsilon(x)\,\mathrm{d}x + \sigma\int_{\mathbb{R}^d}\left\|\left(I+\nabla\Phi\left(\frac{x}{\varepsilon}\right)^\top\right)\nabla u^0(x)\right\|^2\varphi^\varepsilon(x)\,\mathrm{d}x \\ &\quad - 2\sigma\int_{\mathbb{R}^d}\left(I+\nabla\Phi\left(\frac{x}{\varepsilon}\right)^\top\right)\nabla u^0(x)\cdot\nabla u^\varepsilon(x)\varphi^\varepsilon(x)\,\mathrm{d}x - \eta\int_{\mathbb{R}^d}u^\varepsilon(x)^2\varphi^\varepsilon(x)\,\mathrm{d}x, \end{aligned}$$

which by equation (6.16) yields

$$\begin{aligned} \frac{\sigma C_{\nu^\varepsilon}}{C_\pi C_{\nu^0}} I_2^\varepsilon &= \int_{\mathbb{R}^d} f(x)u^\varepsilon(x)\omega\left(\frac{x}{\varepsilon}\right)\varphi^0(x)\,\mathrm{d}x + \sigma\int_{\mathbb{R}^d}\left\|\left(I+\nabla\Phi\left(\frac{x}{\varepsilon}\right)^\top\right)\nabla u^0(x)\right\|^2\omega\left(\frac{x}{\varepsilon}\right)\varphi^0(x)\,\mathrm{d}x \\ &\quad - 2\sigma\int_{\mathbb{R}^d}\left(I+\nabla\Phi\left(\frac{x}{\varepsilon}\right)^\top\right)\nabla u^0(x)\cdot\nabla u^\varepsilon(x)\omega\left(\frac{x}{\varepsilon}\right)\varphi^0(x)\,\mathrm{d}x \\ &\quad - \eta\int_{\mathbb{R}^d}u^\varepsilon(x)^2\omega\left(\frac{x}{\varepsilon}\right)\varphi^0(x)\,\mathrm{d}x. \end{aligned}$$

Passing to the limit as $\varepsilon \to 0$, due to the two-scale converge (6.23), equation (6.26) and the definition of $\mathcal{K}$ in (1.6) we have

$$\lim_{\varepsilon\to 0} \sigma I_2^\varepsilon = \int_{\mathbb{R}^d} f(x)u^0(x)\varphi^0(x) - \int_{\mathbb{R}^d} \Sigma\nabla u^0(x)\cdot\nabla u^0(x)\varphi^0(x)\,\mathrm{d}x - \eta\int_{\mathbb{R}^d} u^0(x)^2 = 0,$$

where the last equality follows from the fact that u^0 is the solution of problem (6.7), and which together with (6.27) implies

$$\lim_{\varepsilon\to 0} \|\nabla z^\varepsilon\|^2_{(L^2_{\varphi^0}(\mathbb{R}^d))^d} = 0. \tag{6.28}$$

Finally, bound (6.24) and limits (6.25) and (6.28) imply the desired result. □

6.3 Eigenvalue problem

In this section we study the homogenization of the eigenvalue problem for the multiscale generator $\mathcal{L}^\varepsilon$. Let $(\lambda^\varepsilon, \phi^\varepsilon)$ be an eigenpair of $\mathcal{L}^\varepsilon$ which solves

$$-\mathcal{L}^\varepsilon\phi^\varepsilon = \lambda^\varepsilon\phi^\varepsilon, \tag{6.29}$$

and let (λ^0, ϕ^0) be an eigenpair of $\mathcal{L}^0$ which solves

$$-\mathcal{L}^0\phi^0 = \lambda^0\phi^0. \tag{6.30}$$

We first show that the spectra of the generators $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$ are discrete and afterwards we prove the convergence of the eigenvalues and the eigenfunctions of the former to the eigenvalues and the eigenfunctions of the latter as the multiscale parameter ε vanishes.

Lemma 6.18. *Let $\mathcal{L}^\varepsilon$ be the generator defined in* (5.5)*. Under Assumptions 1.4 and 6.3, there exists a sequence of couples eigenvalue-eigenvector $\{(\lambda_n^\varepsilon, \phi_n^\varepsilon)\}_{n\in\mathbb{N}}$ which solve* (6.29)*. Moreover, the eigenvalues satisfy*

$$0 = \lambda_0^\varepsilon < \lambda_1^\varepsilon < \lambda_2^\varepsilon < \cdots < \lambda_n^\varepsilon < \cdots \nearrow +\infty,$$

and the eigenfunctions belong to $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ with $\phi_0^\varepsilon \equiv 1$ and

$$\|\phi_n^\varepsilon\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} = \sqrt{1 + \frac{\lambda_n^\varepsilon}{\sigma}}, \tag{6.31}$$

and form an orthonormal basis of $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$.

Proof. By Lemma 6.5 and in particular the Poincaré inequality (6.2), the generator $\mathcal{L}^\varepsilon$ has a spectral gap. Therefore, by [101, Section 4.7] $-\mathcal{L}^\varepsilon$ is a non-negative self-adjoint operator in $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$ with discrete spectrum. Hence, the eigenvalues are real, non-negative, simple and can be ordered as

$$0 = \lambda_0^\varepsilon < \lambda_1^\varepsilon < \lambda_2^\varepsilon < \cdots < \lambda_n^\varepsilon < \cdots \nearrow +\infty.$$

Notice that $\lambda_0^\varepsilon = 0$ and $\phi_0^\varepsilon \equiv 1$. Moreover, using the the unitary transformation which maps the generator to a Schrödinger operator, it follows that the eigenfunctions $\{\phi_n^\varepsilon\}_{n=0}^\infty$ span $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$ and can be normalized such that they form an orthonormal basis (see, e.g., [65, 110]). It now only remains to show that the eigenfunctions belong to $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ and the equality (6.31). Let us consider problem (6.5), which has a unique solution due to Lemma 6.7 and let us denote by $\mathcal{S}_\eta^\varepsilon \colon L^2_{\varphi^\varepsilon}(\mathbb{R}^d) \to H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ the operator which maps the right-hand side f to the solution u^ε, i.e., $\mathcal{S}_\eta^\varepsilon f = u^\varepsilon$. A couple $(\lambda_n^\varepsilon, \phi_n^\varepsilon)$ satisfies for all $\psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$

$$B^\varepsilon(\phi_n^\varepsilon, \psi) = \langle(\lambda_n^\varepsilon + \eta)\phi_n^\varepsilon; \psi\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}\,, \tag{6.32}$$

where B^ε is defined in (6.6) and $\langle \cdot ; \cdot \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}$ denotes the inner product in $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$, and hence

$$\mathcal{S}^\varepsilon_\eta \phi^\varepsilon_n = \frac{1}{\lambda^\varepsilon_n + \eta} \phi^\varepsilon_n,$$

which shows that ϕ^ε_n is also an eigenfunction of $\mathcal{S}^\varepsilon_\eta$ with corresponding eigenvalue $1/(\lambda^\varepsilon_n + \eta)$ and therefore $\phi^\varepsilon_n \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$. Finally, choosing $\psi = \phi^\varepsilon_n$ in (6.32) and since $\|\phi^\varepsilon_n\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} = 1$ we deduce that

$$\|\nabla \phi^\varepsilon_n\|^2_{(L^2_{\varphi^\varepsilon}(\mathbb{R}^d))^d} = \frac{\lambda^\varepsilon_n}{\sigma},$$

which yields equation (6.31) and concludes the proof. □

An analogous results holds true also for the homogenized generator $\mathcal{L}^0$, for which we omit the details since the proof is similar to proof of the previous lemma.

Lemma 6.19. *Let $\mathcal{L}^0$ be the generator defined in* (5.6)*. Under Assumptions 1.4 and 6.3, there exists a sequence of couples eigenvalue-eigenvector $\{(\lambda^0_n, \phi^0_n)\}_{n\in\mathbb{N}}$ which solve* (6.30)*. Moreover, the eigenvalues satisfy*

$$0 = \lambda^0_0 < \lambda^0_1 < \lambda^0_2 < \cdots < \lambda^0_n < \cdots \nearrow +\infty,$$

and the eigenfunctions belong to $H^1_{\varphi^0}(\mathbb{R}^d)$ with $\phi^0_0 \equiv 1$ and

$$\sqrt{1 + \frac{\lambda^0_n}{\lambda_{\max}(\Sigma)}} \leq \left\|\phi^0_n\right\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \leq \sqrt{1 + \frac{\lambda^0_n}{\lambda_{\min}(\Sigma)}},$$

and form an orthonormal basis of $L^2_{\varphi^0}(\mathbb{R}^d)$.

We remark that the eigenvalues and the eigenfunctions of the generators $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$ can be computed employing the Rayleigh quotients R^ε and R^0, respectively, which are defined as

$$\begin{aligned} R^\varepsilon(\psi) &= \sigma \frac{\|\nabla \psi\|^2_{(L^2_{\varphi^\varepsilon}(\mathbb{R}^d))^d}}{\|\psi\|^2_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}} \qquad && \text{for all } \psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d), \quad \psi \neq 0, \\ R^0(\psi) &= \frac{\langle \Sigma \nabla \psi ; \nabla \psi \rangle_{(L^2_{\varphi^0}(\mathbb{R}^d))^d}}{\|\psi\|^2_{L^2_{\varphi^0}(\mathbb{R}^d)}} \qquad && \text{for all } \psi \in H^1_{\varphi^0}(\mathbb{R}^d), \quad \psi \neq 0. \end{aligned} \tag{6.33}$$

Let E^ε_n be the finite dimensional subspace of $H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$ spanned by the first n eigenfunctions $\{\phi^\varepsilon_0, \phi^\varepsilon_1, \dots, \phi^\varepsilon_n\}$ and let E^0_n be the finite dimensional subspace of $H^1_{\varphi^0}(\mathbb{R}^d)$ spanned by the first n eigenfunctions $\{\phi^0_0, \phi^0_1, \dots, \phi^0_n\}$. Then, the "minimax principle" (see, e.g., [33, 114]) gives the characterization for the n-th eigenvalue

$$\begin{aligned} \lambda^\varepsilon_n &= R^\varepsilon(\phi^\varepsilon_n) = \max_{\psi \in E^\varepsilon_n} R^\varepsilon(\psi) = \min_{\psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d), \psi \perp E^\varepsilon_{n-1}} R^\varepsilon(\psi) = \min_{W \in D^\varepsilon_n} \max_{\psi \in W} R^\varepsilon(\psi), \\ \lambda^0_n &= R^0(\phi^0_n) = \max_{\psi \in E^0_n} R^0(\psi) = \min_{\psi \in H^1_{\varphi^0}(\mathbb{R}^d), \psi \perp E^0_{n-1}} R^0(\psi) = \min_{W \in D^0_n} \max_{\psi \in W} R^0(\psi), \end{aligned} \tag{6.34}$$

where

$$\begin{aligned} D^\varepsilon_n &= \{W \subset H^1_{\varphi^\varepsilon}(\mathbb{R}^d) \colon \dim W = n\}, \\ D^0_n &= \{W \subset H^1_{\varphi^0}(\mathbb{R}^d) \colon \dim W = n\}. \end{aligned}$$

We can now state and prove the homogenization of the spectrum of the multiscale generator, whose proof is inspired by the proof of Theorem 2.1 in [71].

Theorem 6.20. *Let $(\lambda_n^\varepsilon, \phi_n^\varepsilon)$ and (λ_n^0, ϕ_n^0) be ordered couples eigenvalue-eigenfunction of the generators $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$, respectively, with $\|\phi_n^\varepsilon\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} = 1$ and $\left\|\phi_n^0\right\|_{L^2_{\varphi^0}(\mathbb{R}^d)} = 1$. Then, under Assumptions 1.4 and 6.3 and choosing the sign of ϕ_n^ε such that $\left\langle \phi_n^\varepsilon; \phi_n^0 \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)} > 0$, it holds for all $n \in \mathbb{N}$ and as $\varepsilon \to 0$*

(i) $\lambda_n^\varepsilon \to \lambda_n^0$,
(ii) $\phi_n^\varepsilon \to \phi_n^0$ *in* $L^2_{\varphi^0}(\mathbb{R}^d)$,
(iii) $\phi_n^\varepsilon \rightharpoonup \phi_n^0$ *in* $H^1_{\varphi^0}(\mathbb{R}^d)$.

Proof. The proof is divided into several steps.
Step 1: *Boundedness of eigenvalues and eigenfunctions.*
Let $\psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d)$, which due to Corollary 6.2 belongs to $H^1_{\varphi^0}(\mathbb{R}^d)$ as well. Employing Lemma 6.1 we have

$$\frac{C_{\text{low}} \|\nabla \psi\|_{(L^2_{\varphi^0}(\mathbb{R}^d))^d}}{C_{\text{up}} \|\psi\|_{L^2_{\varphi^0}(\mathbb{R}^d)}} \leq \frac{\|\nabla \psi\|_{(L^2_{\varphi^\varepsilon}(\mathbb{R}^d))^d}}{\|\psi\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}} \leq \frac{C_{\text{up}} \|\nabla \psi\|_{(L^2_{\varphi^0}(\mathbb{R}^d))^d}}{C_{\text{low}} \|\psi\|_{L^2_{\varphi^0}(\mathbb{R}^d)}},$$

which by the definitions of the Rayleigh quotients in (6.33) implies

$$\frac{C_{\text{low}}^2}{\lambda_{\max}(\mathcal{K}) C_{\text{up}}^2} R^0(\psi) \leq R^\varepsilon(\psi) \leq \frac{C_{\text{up}}^2}{\lambda_{\min}(\mathcal{K}) C_{\text{low}}^2} R^0(\psi),$$

where $\mathcal{K}$ is defined in (1.6). Then, applying the “minimax principle” in (6.34) we obtain for all $n \in \mathbb{N}$

$$\frac{C_{\text{low}}^2}{\lambda_{\max}(\mathcal{K}) C_{\text{up}}^2} \lambda_n^0 \leq \lambda_n^\varepsilon \leq \frac{C_{\text{up}}^2}{\lambda_{\min}(\mathcal{K}) C_{\text{low}}^2} \lambda_n^0,$$

which shows that the sequence of eigenvalues $\{\lambda_n^\varepsilon\}$ is bounded for all $n \in \mathbb{N}$. Moreover, due to equation (6.31) and Corollary 6.2 we deduce that also the sequence of eigenfunctions $\{\phi_n^\varepsilon\}$ is bounded in $H^1_{\varphi^0}(\mathbb{R}^d)$, in fact we have

$$\|\phi_n^\varepsilon\|_{H^1_{\varphi^0}(\mathbb{R}^d)} \leq \frac{1}{C_{\text{low}}} \|\phi_n^\varepsilon\|_{H^1_{\varphi^\varepsilon}(\mathbb{R}^d)} \leq \frac{1}{C_{\text{low}}} \sqrt{1 + \frac{C_{\text{up}}^2}{\lambda_{\min}(\Sigma) C_{\text{low}}^2} \lambda_n^0}.$$

Step 2: *Extraction of a subsequence.*
Due to Step 1 we can extract a subsequence ε' of ε such that $\{\lambda_0^{\varepsilon'}\}$ is convergent and $\{\phi_0^{\varepsilon'}\}$ is weakly convergent in $H^1_{\varphi^0}(\mathbb{R}^d)$ and strongly convergent in $L^2_{\varphi^0}(\mathbb{R}^d)$ and a further subsequence ε'' of ε' such that $\{\lambda_0^{\varepsilon''}\}$ and $\{\lambda_1^{\varepsilon''}\}$ are convergent and $\{\phi_0^{\varepsilon''}\}$ and $\{\phi_1^{\varepsilon''}\}$ are weakly convergent in $H^1_{\varphi^0}(\mathbb{R}^d)$ and strongly convergent in $L^2_{\varphi^0}(\mathbb{R}^d)$. Repeating this procedure for all $n \in \mathbb{N}$ and choosing the standard diagonal subsequence we can find a subsequence, which is still denoted by ε, such that for all $n \in \mathbb{N}$

$$\lambda_n^\varepsilon \to \widetilde{\lambda}_n, \qquad \phi_n^\varepsilon \rightharpoonup \widetilde{\phi}_n \text{ in } H^1_{\varphi^0}(\mathbb{R}^d), \qquad \phi_n^\varepsilon \to \widetilde{\phi}_n \text{ in } L^2_{\varphi^0}(\mathbb{R}^d),$$

where $\widetilde{\lambda}_n \in \mathbb{R}$ and $\widetilde{\phi}_n \in H^1_{\varphi^0}(\mathbb{R}^d)$. From now on we will always consider this final subsequence, if not stated differently.
Step 3: *Identification of the limits.*
A couple eigenvalue-eigenfunction $(\lambda_n^\varepsilon, \phi_n^\varepsilon)$ of the multiscale generator $\mathcal{L}^\varepsilon$ solves the problem

$$B^\varepsilon(\phi_n^\varepsilon, \psi) = \langle (\lambda_n^\varepsilon + \eta) \phi_n^\varepsilon; \psi \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}, \qquad \text{for all } \psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d),$$

where B^ε is defined in (6.6) and by Step 2

$$(\lambda_n^\varepsilon + \eta) \phi_n^\varepsilon \to (\widetilde{\lambda}_n + \eta) \widetilde{\phi}_n \text{ in } L^2_{\varphi^0}(\mathbb{R}^d).$$

Hence, by Corollary 6.16 and the uniqueness of the limit it follows that the couple $(\widetilde{\lambda}_n, \widetilde{\phi}_n)$ solves the problem

$$B^0(\widetilde{\phi}_n, \psi) = \left\langle (\widetilde{\lambda}_n + \eta)\widetilde{\phi}_n; \psi \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)}, \qquad \text{for all } \psi \in H^1_{\varphi^0}(\mathbb{R}^d),$$

where B^0 is defined in (6.8) and therefore it is a couple eigenvalue-eigenfunction of the homogenized generator $\mathcal{L}^0$.
Step 4: *Ordering of the limits.*
We now show that the sequence of limits $\{\widetilde{\lambda}_n\}_{n\in\mathbb{N}}$ is such that $\widetilde{\lambda}_0 < \widetilde{\lambda}_1 < \widetilde{\lambda}_2 < \cdots < \widetilde{\lambda}_n < \cdots$. First, due to Lemma 6.18 we know that $\lambda_0^\varepsilon < \lambda_1^\varepsilon < \lambda_2^\varepsilon < \cdots < \lambda_n^\varepsilon < \cdots$, hence their limits must satisfy $\widetilde{\lambda}_0 \le \widetilde{\lambda}_1 \le \widetilde{\lambda}_2 \le \cdots \le \widetilde{\lambda}_n \le \cdots$. Let us now assume by contradiction that there exist $l, m \in \mathbb{N}$ such that $\widetilde{\lambda}_l = \widetilde{\lambda}_m =: \widetilde{\lambda}$. Since the eigenfunctions ϕ_l^ε and ϕ_m^ε corresponding to the eigenvalues λ_l^ε and λ_m^ε are orthogonal in $L^2_{\varphi^\varepsilon}(\mathbb{R}^d)$, then

$$\langle \phi_l^\varepsilon; \phi_m^\varepsilon \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} = 0,$$

and passing to the limit as ε vanishes we obtain

$$\left\langle \widetilde{\phi}_l; \widetilde{\phi}_m \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)} = 0. \tag{6.35}$$

In fact, we have

$$\begin{aligned} &\left| \left\langle \widetilde{\phi}_l; \widetilde{\phi}_m \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)} - \langle \phi_l^\varepsilon; \phi_m^\varepsilon \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \right| \le \\ &\qquad \left| \left\langle \widetilde{\phi}_l; \widetilde{\phi}_m \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)} - \left\langle \widetilde{\phi}_l; \widetilde{\phi}_m \right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \right| + \left| \left\langle \widetilde{\phi}_l; \widetilde{\phi}_m \right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} - \langle \phi_l^\varepsilon; \phi_m^\varepsilon \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \right|, \end{aligned} \tag{6.36}$$

where the first term in the right hand side vanishes due to the convergence of the measure ν^ε with density φ^ε towards the measure ν^0 with density φ^0 and the second term tends to zero due to the convergence of the eigenvectors and because by Cauchy–Schwarz inequality and Lemma 6.1 we have

$$\left| \left\langle \widetilde{\phi}_l; \widetilde{\phi}_m \right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} - \langle \phi_l^\varepsilon; \phi_m^\varepsilon \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \right| \le \left\| \widetilde{\phi}_l \widetilde{\phi}_m - \phi_l^\varepsilon \phi_m^\varepsilon \right\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \le C_{\mathrm{up}} \left\| \widetilde{\phi}_l \widetilde{\phi}_m - \phi_l^\varepsilon \phi_m^\varepsilon \right\|_{L^2_{\varphi^0}(\mathbb{R}^d)}. \tag{6.37}$$

Therefore, equality (6.35) implies that the eigenvectors $\widetilde{\phi}_l$ and $\widetilde{\phi}_m$ corresponding to the eigenvalue $\widetilde{\lambda}$ are linearly independent and hence $\widetilde{\lambda}$ is not a simple eigenvalue, which is impossible due to Lemma 6.19.
Step 5: *Entire spectrum.*
We now prove that there is no eigenvalue of the homogenized generator $\mathcal{L}^0$ other than those in the sequence $\{\widetilde{\lambda}_n\}_{n\in\mathbb{N}}$. Let us assume by contradiction that $\{\widetilde{\lambda}_n\}_{n\in\mathbb{N}}$ is a subsequence of $\{\lambda_n^0\}_{n\in\mathbb{N}}$, i.e., that there exists an eigenvalue $\lambda \in \mathbb{R}$ of the homogenized generator $\mathcal{L}^0$ such that $\lambda \neq \widetilde{\lambda}_n$ for all $n \in \mathbb{N}$ and let $\phi \in H^1_{\varphi^0}(\mathbb{R}^d)$ be its corresponding normalized eigenfunction, which due to Lemma 6.19 satisfies

$$\left\langle \phi; \widetilde{\phi}_n \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)} = 0, \qquad \text{for all } n \in \mathbb{N}.$$

Then, there exists $m \in \mathbb{N}$ such that $\lambda < \widetilde{\lambda}_{m+1}$. Let v^ε be the solution of the problem

$$B^\varepsilon(v^\varepsilon, \psi) = (\lambda + \eta) \langle \phi; \psi \rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}, \qquad \text{for all } \psi \in H^1_{\varphi^\varepsilon}(\mathbb{R}^d), \tag{6.38}$$

and notice that due to Theorem 6.15

$$v^\varepsilon \rightharpoonup \phi \text{ in } H^1_{\varphi^0}(\mathbb{R}^d) \qquad \text{and} \qquad v^\varepsilon \to \phi \text{ in } L^2_{\varphi^0}(\mathbb{R}^d).$$

Choosing $\psi = v^\varepsilon$ in (6.38) we then have

$$\lim_{\varepsilon\to 0} R^\varepsilon(v^\varepsilon) = \lim_{\varepsilon\to 0} \sigma \frac{\|\nabla v^\varepsilon\|_{(L^2_{\varphi^\varepsilon}(\mathbb{R}^d))^d}}{\|v^\varepsilon\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}} = \lim_{\varepsilon\to 0} \frac{(\lambda+\eta)\left\langle \phi; v^\varepsilon\right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}}{\|v^\varepsilon\|_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)}} - \eta = \lambda, \tag{6.39}$$

where the last equality is justified by an argument similar to (6.36) and (6.37). Let now ξ^ε be defined as

$$\xi^\varepsilon \coloneqq v^\varepsilon - \sum_{n=0}^m \left\langle v^\varepsilon; \phi^\varepsilon_n\right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} \phi^\varepsilon_n, \tag{6.40}$$

which has the same limit as v^ε, i.e.,

$$\xi^\varepsilon \rightharpoonup \phi \text{ in } H^1_{\varphi^0}(\mathbb{R}^d) \qquad \text{and} \qquad \xi^\varepsilon \to \phi \text{ in } L^2_{\varphi^0}(\mathbb{R}^d),$$

since a similar computation to (6.36) and (6.37) yields

$$\lim_{\varepsilon\to 0} \left\langle v^\varepsilon; \phi^\varepsilon_n\right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} = \left\langle \phi; \widetilde{\phi}_n \right\rangle_{L^2_{\varphi^0}(\mathbb{R}^d)} = 0. \tag{6.41}$$

Moreover, due to (6.41) also its Rayleigh quotient has the same limit as (6.39), i.e.,

$$\lim_{\varepsilon\to 0} R^\varepsilon(\xi^\varepsilon) = \lambda,$$

and by definition (6.40) it follows for all $n = 1,\dots,m$

$$\left\langle \xi^\varepsilon; \phi^\varepsilon_n\right\rangle_{L^2_{\varphi^\varepsilon}(\mathbb{R}^d)} = 0.$$

Therefore, by the "minimax principle" (6.34), $\lambda^\varepsilon_{m+1} \le R^\varepsilon(\xi^\varepsilon)$ and passing to the limit as $\varepsilon \to 0$ we deduce that $\widetilde{\lambda}_{m+1} \le \lambda$ which contradicts the fact that m is such that $\lambda < \widetilde{\lambda}_{m+1}$.

Step 6: *Convergence to the homogenized spectrum.*

From Steps 3,4,5 and by Lemma 6.19 it follows that the sequence of limits $\{\widetilde{\lambda}_n\}_{n\in\mathbb{N}}$ is the same as the sequence of eigenvalues $\{\lambda^0_n\}_{n\in\mathbb{N}}$ of the homogenized generator $\mathcal{L}^0$, hence we have $\widetilde{\lambda}_n = \lambda^0_n$ for all $n \in \mathbb{N}$. Moreover, since the eigenfunctions are normalized, then the limit $\widetilde{\phi}_n$ can be either ϕ^0_n or $-\phi^0_n$. The hypothesis that the sign of ϕ^ε_n is chosen such that $(\phi^\varepsilon_n, \phi^0_n)_{L^2_{\varphi^0}(\mathbb{R}^d)} > 0$ implies that the positive sign is the right one, i.e., $\widetilde{\phi}_n = \phi^0_n$ for all $n \in \mathbb{N}$.

Step 7: *Convergence of the whole sequence.*

For all $n \in \mathbb{N}$ the fact that the only admissible limit for the subsequence $\{\lambda^\varepsilon_n\}$ is λ^0_n implies that the whole sequence converges to λ^0_n. Indeed, assuming by contradiction that $\{\lambda^\varepsilon_n\}$ does not converge to λ^0_n gives the existence of a subsequence $\{\lambda^{\varepsilon'}_n\}$ and $\delta > 0$ such that

$$\left|\lambda^{\varepsilon'}_n - \lambda^0_n\right| > \delta. \tag{6.42}$$

However, repeating all the previous steps we can extract a subsequence $\{\lambda^{\varepsilon''}_n\}$ from $\{\lambda^{\varepsilon'}_n\}$ such that

$$\lim_{\varepsilon\to 0} \lambda^{\varepsilon''}_n = \lambda^0_n,$$

which contradicts (6.42). Finally, a similar argument shows the convergence of the whole sequence of eigenfunctions $\{\phi^\varepsilon_n\}$ to ϕ^0_n and concludes the proof. □

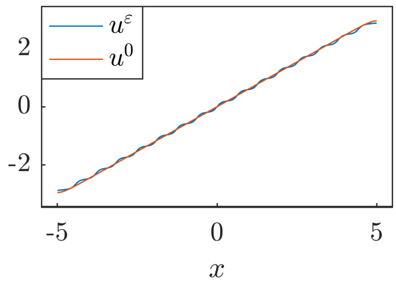


Figure 6.1 – Multiscale and homogenized solution of the Poisson problem with a reaction term setting $\varepsilon = 0.1$.

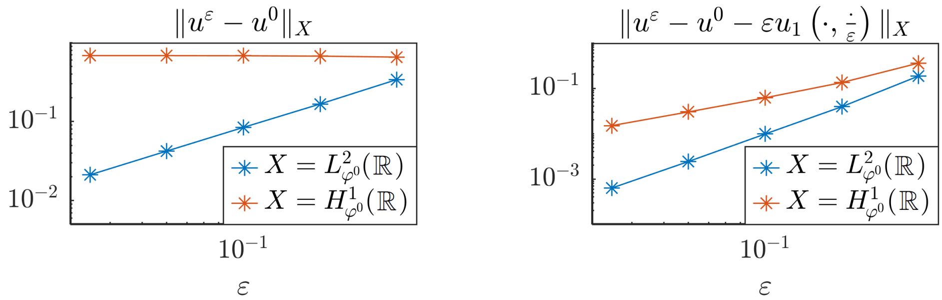


Figure 6.2 – Poisson problem with a reaction term varying ε. Left: distance between the multiscale and homogenized solution. Right: distance between the multiscale solution and its first order expansion.

6.4 Numerical illustration

In this section we present an example complementing our theoretical results. We consider the one-dimensional ($d = 1$) multiscale Ornstein–Uhlenbeck (OU) process with slow-scale potential $\mathcal{V}(x) = x^2/2$, fast-scale potential $p(y) = \cos(y)$ and diffusion coefficient $\sigma = 1$. The numerical results are obtained setting the discretization size $h = \varepsilon^2$ and replacing the real line $\mathbb{R}$ with a truncated domain $D = [-R, R]$ with $R = 5$. The error introduced by this approximation is negligible since the invariant measures φ^0 and φ^ε, which appear as weight functions in the integrals, decay exponentially fast for $\|x\| \to \infty$. In particular, the problems are discretized employing the finite element method with continuous piecewise linear functions as basis functions. We also remark that, due to the fast decay of the invariant measures, we did not need to impose any boundary condition on the solution, but we only assumed the functions φ^0 and φ^ε to be zero on the boundary of the truncated domain D.

6.4.1 Poisson problem with a reaction term

We consider the Poisson problems (6.3) and (6.4) with reaction coefficient $\eta = 1$ and right-hand side $f(x) = x$. In this particular case the homogenized equation (6.4) admits the analytical solution

$$u^0(x) = \frac{x}{\mathcal{K} + \eta}.$$

In Figure 6.1 we plot the numerical solutions u^ε and u^0 setting $\varepsilon = 0.1$, and we observe that the multiscale solution oscillates around the homogenized one. We then solve equation (6.3) for different values of the mutiscale parameter $\varepsilon = 0.025, 0.05, 0.1, 0.2, 0.4$, and we compute the distance between u^ε and u^0 both in $L^2_{\varphi^0}(\mathbb{R})$ and $H^1_{\varphi^0}(\mathbb{R})$. On the left of Figure 6.2 we observe that the theoretical results given by Theorem 6.15 are confirmed in practice. In particular, $\left\| u^\varepsilon - u^0 \right\|_{L^2_{\varphi^0}(\mathbb{R})}$ decreases as ε vanishes, while $\left\| u^\varepsilon - u^0 \right\|_{H^1_{\varphi^0}(\mathbb{R})}$ remains constant. Indeed, the solution u^ε converges to u^0 strongly in $L^2_{\varphi^0}(\mathbb{R})$ but only weakly in $H^1_{\varphi^0}(\mathbb{R})$. We now consider a better approximation of the multiscale solution u^ε, which is given by the first order expansion

$$\widetilde{u}^\varepsilon(x) = u^0(x) + \varepsilon u_1\left(x, \frac{x}{\varepsilon}\right),$$

where

$$u_1(x,y) = (u^0)'(x)\Phi(y).$$

The analytical solution Φ of equation (1.7), which is periodic in $\mathcal{T} = [0, \mathbb{T}]$ and has zero-mean with respect to ω, is

$$\Phi(y) = C_\Phi - y + \frac{\mathbb{T}}{\widehat{C}_\pi} \int_0^y e^{\frac{1}{\sigma}p(z)} \,\mathrm{d}z,$$

where

$$\widehat{C}_\pi = \int_0^{\mathbb{T}} e^{\frac{1}{\sigma}p(y)} \,\mathrm{d}y,$$

and

$$C_\Phi = \frac{1}{C_\pi} \int_0^{\mathbb{T}} y e^{-\frac{1}{\sigma}p(y)} \,\mathrm{d}y - \frac{\mathbb{T}}{C_\pi \widehat{C}_\pi} \int_0^{\mathbb{T}} \int_0^y e^{\frac{1}{\sigma}(p(z)-p(y))} \,\mathrm{d}z \,\mathrm{d}y.$$

On the right of Figure 6.2 we plot the distance between u^ε and its first order approximation $\widetilde{u}^\varepsilon$ both in $L^2_{\varphi^0}(\mathbb{R})$ and $H^1_{\varphi^0}(\mathbb{R})$, and we observe that we now also have strong convergence in $H^1_{\varphi^0}(\mathbb{R})$ as shown by Theorem 6.17.

6.4.2 Eigenvalue problem

We now consider the homogenization of the eigenvalue problem for the multiscale generator. First, in Figure 6.3 we set $\varepsilon = 0.1$ and plot the first four eigenvalues and eigenfunctions of both $\mathcal{L}^\varepsilon$ and $\mathcal{L}^0$. We observe that the eigenvalues λ_n^ε are close to the eigenvalues λ_n^0 and that the mismatch increases for n bigger, i.e., for eigenvalues with greater magnitude. Moreover, the eigenfunctions behave similarly to the solution of the Poisson problem, in the sense that ϕ_n^ε oscillates around ϕ_n^0. We remark that in the particular case of the OU process the eigenvalue problem for the homogenized generator $\mathcal{L}^0$ can be solved analytically and the eigenfunctions are given by the normalized Hermite polynomials [101, Section 4.4]. In particular, we have for all $n \in \mathbb{N}$ that $\lambda_n^0 = \mathcal{K} n$ and

$$\phi_n^0(x) = \frac{1}{\sqrt{n!}} H_n\left(\sqrt{\frac{\mathcal{K}}{\Sigma}} x\right),$$

where

$$H_n(z) = (-1)^n e^{\frac{z^2}{2}} \frac{\mathrm{d}^n}{\mathrm{d}z^n}\left(e^{-\frac{z^2}{2}}\right).$$

We then solve the eigenvalue problem for different values of the multiscale parameter $\varepsilon = 0.025, 0.05, 0.1, 0.2, 0.4$, and we compute the distance between the multiscale and homogenized eigenvalues and eigenfunctions. Figure 6.4 demonstrates numerically what we proved theoretically in Theorem 6.20, i.e., that we have convergence of the eigenvalues and strong convergence in $L^2_{\varphi^0}(\mathbb{R})$, but only weak in $H^1_{\varphi^0}(\mathbb{R})$, of the eigenfunctions.

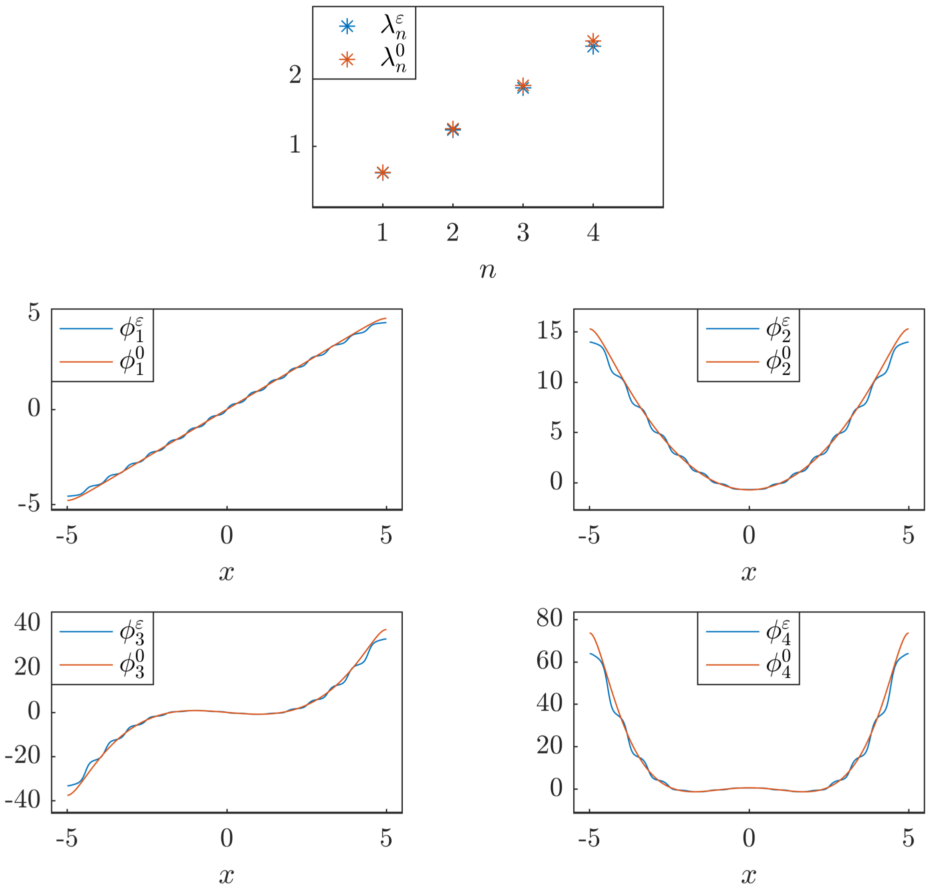


Figure 6.3 – First four eigenvalues and eigenfunctions of the multiscale and homogenized generator setting $\varepsilon = 0.1$.

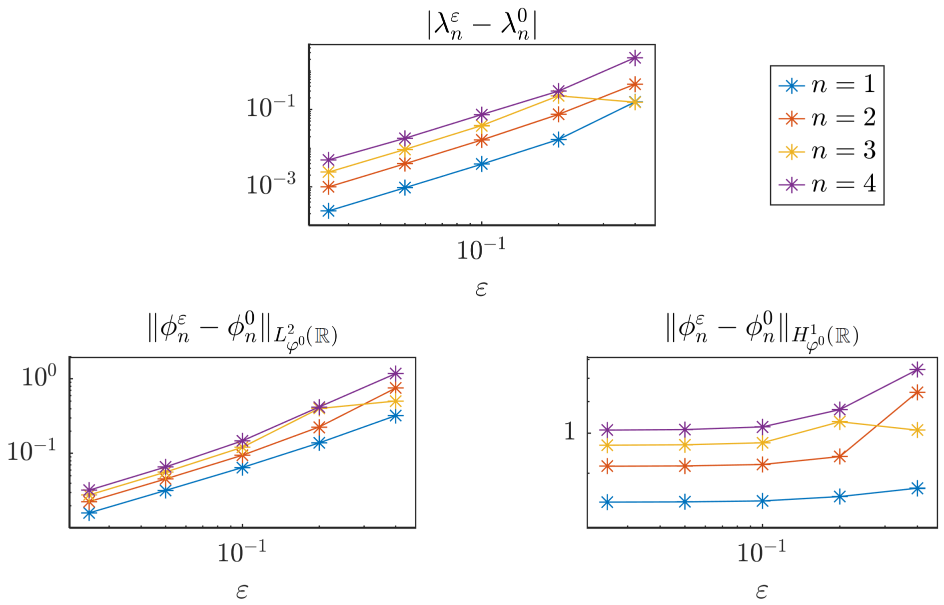


Figure 6.4 – Distance between the first four eigenvalues and eigenfunctions of the multiscale and homogenized generator varying ε.

6.5 Conclusion

We presented the homogenization of two problems involving the infinitesimal generator of the multiscale overdamped Langevin stochastic differential equation (SDE). We first considered the Poisson problem with a reaction term and, after introducing appropriate weighted Sobolev spaces and extending the theory of two-scale convergence, we proved in Theorem 6.15 the strong convergence in L^2 sense and the weak convergence in H^1 sense of the multiscale solution to the solution of the same problem where the multiscale generator is replaced by its homogenized surrogate. In Theorem 6.17 we also provided a corrector result which justifies the two first terms in the usual asymptotic expansion in homogenization theory. We then analyzed the eigenvalue problem and in Theorem 6.20 we showed homogenization results for the eigenvalues and the eigenfunctions of the multiscale generator. In particular, we demonstrated the convergence of the eigenvalues and the strong convergence in L^2 sense and the weak convergence in H^1 sense of the eigenvectors towards the corresponding eigenvalues and eigenfunctions of the generator of the coarse-grained dynamics. Finally, we verified numerically our theoretical results simulating the multiscale one-dimensional OU process.

7 Eigenfunction estimators for multiscale diffusions

In this chapter, which is based on our research article [6], we apply eigenfunction estimators to infer the parameters of the homogenized Langevin dynamics given discrete observations from the multiscale system. The chapter is organized as follows. In Section 7.1 we introduce the proposed estimators and in Section 7.2 we present the main theoretical results. Then, in Section 7.3 we perform numerical experiments which validate the efficacy of our method and Section 7.4 is devoted to the proofs of the main theorems. Moreover, in Section 7.5 we show some technical results which are employed in the analysis and in Sections 7.6 and 7.7 we explain some details about the implementation of the proposed methodology and the extension to the multidimensional case, respectively. Finally, in Section 7.8 we draw our conclusions.

7.1 Martingale estimating functions based on eigenfunctions

In this section we develop martingale estimating functions based on the eigenfunctions of the generator of the process, since the theory of the eigenvalue problem for elliptic differential operators and the multiscale analysis of this eigenvalue problem are well developed. Let $\mathcal{A} \subset \mathbb{R}^L$ be the set of admissible drift coefficients for which Assumption 1.4(ii) is satisfied. To describe our methodology we consider the solution $X_t(a)$ of the homogenized process (1.11) with a generic parameter $a \in \mathcal{A}$ instead of the exact drift coefficient A:

$$\mathrm{d}X_t(a) = -a \cdot V'(X_t(a))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t, \tag{7.1}$$

which has invariant measure

$$\varphi_a(x) = \frac{1}{C_{\nu_a}} \exp\left(-\frac{1}{\Sigma} a \cdot V(x)\right), \quad \text{where} \quad C_{\nu_a} = \int_{\mathbb{R}} \exp\left(-\frac{1}{\Sigma} a \cdot V(x)\right) \mathrm{d}x. \tag{7.2}$$

The generator $\mathcal{L}_a$ of (7.1) is defined for all $u \in C^2(\mathbb{R})$ as

$$\mathcal{L}_a u(x) = -a \cdot V'(x) u'(x) + \Sigma u''(x), \tag{7.3}$$

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. [63]). 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

$$\mathcal{L}_a \phi_j(x; a) = -\lambda_j(a) \phi_j(x; a), \tag{7.4}$$

which, due to (7.3), is equivalent to

$$\Sigma \phi_j''(x; a) - a \cdot V'(x) \phi_j'(x; a) + \lambda_j(a) \phi_j(x; a) = 0, \tag{7.5}$$

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 $L^2_{\varphi_a}(\mathbb{R})$. 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 [101, Chapter 4]. We now state a formula, which has been proved in [73] and will be fundamental in the rest of the work

$$\mathbb{E}\left[\phi_j(X_{t_m}(a);a)|X_{t_{m-1}}(a)=x\right]=e^{-\lambda_j(a)\Delta}\phi_j(x;a), \tag{7.6}$$

where $\Delta = t_m - t_{m-1}$ 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 $\{\psi_j(\cdot;a)\}_{j=1}^J$ be J arbitrary functions $\psi_j(\cdot;a)\colon \mathbb{R}\to\mathbb{R}^L$ possibly dependent on the parameter a, which satisfy Assumption 7.2(i)(ii) stated below, and define for $x,y,z\in\mathbb{R}$ the martingale estimating function

$$g_j(x,y,z;a)=\psi_j(z;a)\left(\phi_j(y;a)-e^{-\lambda_j(a)\Delta}\phi_j(x;a)\right). \tag{7.7}$$

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

$$\widehat{G}_{M,\varepsilon}^J(a)=\frac{1}{\Delta}\sum_{m=0}^{M-1}\sum_{j=1}^J g_j(\widetilde{X}_m^\varepsilon,\widetilde{X}_{m+1}^\varepsilon,\widetilde{X}_m^\varepsilon;a). \tag{7.8}$$

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. The full derivation of a martingale estimating function as an approximation of the true score function is given in detail in [21, Section 2]. The first step is a discretization of the gradient of the continuous-time log-likelihood, which yields a biased estimating function. Hence, the next step is adjusting this function by adding its compensator in order to obtain a zero-mean martingale. Moreover, by using the eigenfunctions of the generator, it is shown in [73] that this approach is suitable for scalar diffusion processes with no multiscale structure, i.e., processes with a single characteristic length/time scale. In fact, by a classical result for ergodic diffusion processes [101, 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.

Remark 7.1. In the construction of the martingale estimating function $\widehat{G}_{M,\varepsilon}^J(a)$ we omitted the first index $j=0$ because, for ergodic diffusion processes, the first eigenvalue is zero, $\lambda_0(a)=0$, and its corresponding eigenfunction is constant, $\phi_0(a)=1$, and hence they would give $g_0(x,y,z;a)=0$ independently of the function $\psi_0(z;a)$. Therefore, it would not provide us with any information about the unknown parameters in the drift.

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

$$\widehat{G}_{M,\varepsilon}^J(a)=0. \tag{7.9}$$

An intuition on why $\widehat{G}_{M,\varepsilon}^J$ is a good score function is given by the following result. Let $\widehat{G}_{M,0}^J$ be the score function where the observations of the slow variable of the multiscale process are replaced by the homogenized ones, then due to equation (7.6)

$$\mathbb{E}\left[\widehat{G}_{M,0}^J(A)\right]=0,$$

which means that the zero of the expectation of the score function with homogenized observations is exactly the drift coefficient of the effective equation. In Algorithm 1 we summarize the main steps for computing the estimator $\widehat{A}_{M,\varepsilon}^J$ and further details about the implementation can be found in Section 7.6. We finally introduce the following technical assumption which will be employed in the analysis.

Assumption 7.2. The following hold for all $a \in \mathcal{A}$ and for all $j = 1, \dots, J$:

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

(ii) all components of $\psi_j(\cdot; a)$, $\psi_j'(\cdot; a)$, $\dot\psi_j(\cdot; a)$, $\dot\psi_j'(\cdot; a)$ 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 7.3. In [73] the authors propose a method to choose the functions $\{\psi_j(\cdot; a)\}_{j=1}^J$ in order to obtain optimality in the sense of [56]: this optimal set of functions can be seen as the projection of the score function onto the set of martingale estimating functions obtained by varying the function $\{\psi_j(\cdot; a)\}_{j=1}^J$. 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 [64]. For further details on how to choose these functions we refer to [73], and we remark that their calculation requires additional computational cost. Nevertheless, the theory we develop is valid for all functions which satisfy Assumptions 7.2(i) and 7.2(ii) and we observed in practice that choosing simple functions independent of the unknown parameter, e.g. monomials of the form $\psi_j(z; a) = z^k$ 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 polynomials as eigenfunctions [15, Section 2.7]. Partial results in this directions also exist in higher dimensions.

Algorithm 1: Estimation of A without filtered data

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

Output: Estimation $\widehat A_{M,\varepsilon}^J$ 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 $\{\lambda_j(a)\}_{j=1}^J$ and eigenfunctions $\{\phi_j(\cdot; a)\}_{j=1}^J$.
3: Construct the function $g_j(x, y, z; a) = \psi_j(z; a)\left(\phi_j(y; a) - e^{-\lambda_j(a)\Delta}\phi_j(x; a)\right)$.
4: Construct the score function $\widehat G_{M,\varepsilon}^J(a) = \frac{1}{\Delta}\sum_{m=0}^{M-1}\sum_{j=1}^J g_j(\widetilde X_m^\varepsilon, \widetilde X_{m+1}^\varepsilon, \widetilde X_m^\varepsilon; a)$.
5: Let $\widehat A_{M,\varepsilon}^J$ be the solution of the nonlinear system $\widehat G_{M,\varepsilon}^J(a) = 0$.

7.1.1 The filtering approach

We now go back to our multiscale stochastic differential equation (SDE) (1.10) and, inspired by Chapter 2, we propose a second estimator for the homogenized drift coefficient by filtering the data. In particular, we modify $\widehat A_{M,\varepsilon}^J$ by filtering the observations and inserting the new data into the score function $\widehat G_{M,\varepsilon}^J$ 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

$$k_{\exp}^{1,1}(r) = e^{-r},$$

for which a rigorous theory has been developed in Chapter 2. We remark that this exponential kernel is a low-pass filter, which cuts the high frequencies and highlights the slowest components. We then define the filtered observations $\{\widetilde{Z}_m^\varepsilon\}_{m=0}^M$ choosing $\widetilde{Z}_0^\varepsilon = 0$ and computing the weighted average for all $m = 1, \dots, M$

$$\widetilde{Z}_m^\varepsilon = \Delta \sum_{k=0}^{m-1} k_{\exp}^{1,1}(\Delta(m-k))\widetilde{X}_k^\varepsilon, \tag{7.10}$$

where the fast-scale component of the original multiscale trajectory is eliminated, and we define the new score function as a modification of (7.8), i.e.,

$$\widetilde{G}_{M,\varepsilon}^J(a) = \frac{1}{\Delta} \sum_{m=0}^{M-1} \sum_{j=1}^J g_j(\widetilde{X}_m^\varepsilon, \widetilde{X}_{m+1}^\varepsilon, \widetilde{Z}_m^\varepsilon; a). \tag{7.11}$$

Remark 7.4. Notice that the filtered data only partially replace the original data in the definition of the score function. This idea is inspired by Chapter 2 where the same approach is used with the maximum likelihood estimator (MLE). The importance of keeping also the original observations becomes apparent in the proofs of the main results. However, a simple intuition is provided by equation (7.6). This equation is essential in order to obtain the unbiasedness of the estimators when the sampling rate Δ is independent of the multiscale parameter ε, but it is not valid for the filtered process.

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

$$\widetilde{G}_{M,\varepsilon}^J(a) = 0. \tag{7.12}$$

The main steps to compute the estimator $\widetilde{A}_{M,\varepsilon}^J$ are highlighted in Algorithm 2 and additional details about the implementation can be found in Section 7.6. Note that (7.10) can be rewritten as

$$\widetilde{Z}_m^\varepsilon = \Delta \sum_{k=0}^{m-1} e^{-\Delta(m-k)} \widetilde{X}_k^\varepsilon. \tag{7.13}$$

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

$$Z_t^\varepsilon = \int_0^t e^{-(t-s)} X_s^\varepsilon \,\mathrm{d}s. \tag{7.14}$$

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

$$\begin{aligned} \mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \\ \mathrm{d}Z_t^\varepsilon &= (X_t^\varepsilon - Z_t^\varepsilon)\,\mathrm{d}t, \end{aligned} \tag{7.15}$$

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 Sections 2.1.1 and 2.1.2 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 $\mu_{\text{exp}}^\varepsilon$ and μ_{exp}^0 with respect to the Lebesgue measure denoted respectively by $\rho_{\text{exp}}^\varepsilon = \rho_{\text{exp}}^\varepsilon(x,z)$ and $\rho_{\text{exp}}^0 = \rho_{\text{exp}}^0(x,z)$. Let us finally remark that given discrete observations $\widetilde X_m^\varepsilon$ we can only compute $\widetilde Z_m^\varepsilon$, 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 7.5. The only difference in the construction of the estimators $\widehat A_{M,\varepsilon}^J$ and $\widetilde A_{M,\varepsilon}^J$ 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 $\left\{\widetilde X_m^\varepsilon\right\}_{m=0}^M$.
Distance between two consecutive observations Δ.
Number of eigenvalues and eigenfunctions J.
Functions $\{\psi_j(z;a)\}_{j=1}^J$.
Slow-scale potential V.
Diffusion coefficient Σ.

Output: Estimation $\widetilde A_{M,\varepsilon}^J$ 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 $\{\lambda_j(a)\}_{j=1}^J$ and eigenfunctions $\{\phi_j(\cdot;a)\}_{j=1}^J$.
3: *Compute the filtered data* $\left\{\widetilde Z_m^\varepsilon\right\}_{m=0}^M$ *as* $\widetilde Z_0^\varepsilon = 0$ *and* $\widetilde Z_m^\varepsilon = \Delta \sum_{k=0}^{m-1} e^{-\Delta(m-k)} \widetilde X_k^\varepsilon$.
4: Construct the function $g_j(x,y,z;a) = \psi_j(z;a)\left(\phi_j(y;a) - e^{-\lambda_j(a)\Delta}\phi_j(x;a)\right)$.
5: Construct the score function $\widetilde G_{M,\varepsilon}^J(a) = \frac{1}{\Delta}\sum_{m=0}^{M-1}\sum_{j=1}^J g_j(\widetilde X_m^\varepsilon, \widetilde X_{m+1}^\varepsilon, \widetilde Z_m^\varepsilon; a)$.
6: Let $\widetilde A_{M,\varepsilon}^J$ be the solution of the nonlinear system $\widetilde G_{M,\varepsilon}^J(a) = 0$.

7.2 Main results

In this section we present the main results of this chapter, i.e., the asymptotic unbiasedness of the proposed estimators. We first need to introduce the following technical assumption, which is a nondegeneracy hypothesis related to the use of the implicit function theorem for the functions (7.8) and (7.11) in the limit as $M \to \infty$.

Assumption 7.6. Let A be the homogenized drift coefficient of equation (1.11). Then the following hold

(i) $\det\left(\sum_{j=1}^J \mathbb{E}^{\widetilde\mu_{\text{exp}}^0}\left[\left(\psi_j(\widetilde Z_0^0;A)\otimes\nabla_a X_\Delta(A)\right)\phi_j'(X_\Delta^0;A)\right]\right) \neq 0$,

(ii) $\det\left(\sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\left(\psi_j(X_0^0;A)\otimes\nabla_a X_\Delta(A)\right)\phi_j'(X_\Delta^0;A)\right]\right) \neq 0$,

(iii) $\det\left(\sum_{j=1}^J \mathbb{E}^{\mu_{\text{exp}}^0}\left[(\psi_j(Z_0^0;A)\otimes V'(X_0^0))\phi_j'(X_0^0;A)\right]\right) \neq 0$,

(iv) $\det\left(\sum_{j=1}^J \mathbb{E}^{\nu^0}\left[(\psi_j(X_0^0;A)\otimes V'(X_0^0))\phi_j'(X_0^0;A)\right]\right) \neq 0$,

where $\widetilde\mu_{\text{exp}}^0$ is the invariant measure of the couple $(\widetilde X_m^0, \widetilde Z_m^0)$ with density $\widetilde\rho_{\text{exp}}^0$, whose existence is guaranteed by Lemma 7.28, and $\nabla_a X_t(a)$ is the gradient of the stochastic process $X_t(a)$ in (7.1) with respect to the drift coefficient a.

Remark 7.7. The nondegeneracy Assumption 7.6, which is analogous to Condition 4.2(a) in [73], holds true in all nonpathological examples and does not constitute an essential limitation on the range of validity of the results proved in this work. Further details about the necessity of this assumption for the analysis of the proposed estimator will be given in Section 7.4.2.

The proofs of the following two main theorems are the focus of Section 7.4.

Theorem 7.8. *Let J be a positive integer. Under Assumptions 1.4, 7.2 and 7.6 and if Δ is independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta \in (0,1)$, there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, an estimator $\widehat{A}^J_{M,\varepsilon}$ which solves the system $\widehat{G}^J_{M,\varepsilon}(\widehat{A}^J_{M,\varepsilon}) = 0$ exists with probability tending to one as $M \to \infty$. Moreover*

$$\lim_{\varepsilon\to 0} \lim_{M\to\infty} \widehat{A}^J_{M,\varepsilon} = A, \qquad \textit{in probability},$$

where A is the homogenized drift coefficient of equation (1.11).

Theorem 7.9. *Let J be a positive integer. Under Assumptions 1.4, 7.2 and 7.6 and if Δ is independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta > 0$ and $\zeta \neq 1$, $\zeta \neq 2$, there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ an estimator $\widetilde{A}^J_{M,\varepsilon}$ which solves the system $\widetilde{G}^J_{M,\varepsilon}(\widetilde{A}^J_{M,\varepsilon}) = 0$ exists with probability tending to one as $M \to \infty$. Moreover*

$$\lim_{\varepsilon\to 0} \lim_{M\to\infty} \widetilde{A}^J_{M,\varepsilon} = A, \qquad \textit{in probability},$$

where A is the homogenized drift coefficient of equation (1.11).

Remark 7.10. Notice that in both Theorem 7.8 and Theorem 7.9 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 M tends to infinity, and then we let the multiscale parameter ε vanish. Moreover, in Theorem 7.9 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 (1.10) converge to the homogenized drift coefficient A of (1.11). 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 7.11. Since the main goal of this chapter 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 $\theta = (a, s) \in \mathbb{R}^{L+1}$ where a stands for the drift and s stands for the diffusion, yielding nonlinear systems of dimension $L + 1$ corresponding to (7.9) and (7.12). The proofs of the asymptotic unbiasedness of the new estimators $\widehat{\theta}^J_{M,\varepsilon}$ and $\widetilde{\theta}^J_{M,\varepsilon}$ can be adjusted analogously. For completeness, we provide a more detailed explanation and a numerical experiment illustrating this approach in Section 7.3.6.

7.2.1 A particular case

Before analysing the general framework, let us consider the simple case of the Ornstein-Uhlenbeck (OU) process, i.e. let the dimension of the parameter $L = 1$ and let $V(x) = x^2/2$. Then the multiscale SDE (1.10) becomes

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

and its homogenized version is

$$\mathrm{d}X_t^0 = -AX_t^0 \,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t.$$

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 [101, 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 $\psi_1(z; a) = z$, which implies

$$g_1(x,y,z;a) = z\left(y - e^{-a\Delta}x\right).$$

Then the score functions (7.8) and (7.11) become

$$\begin{aligned}
\widehat{G}^1_{M,\varepsilon} &= \frac{1}{\Delta}\sum_{m=0}^{M-1} \widetilde{X}_m^\varepsilon \left(\widetilde{X}_{m+1}^\varepsilon - e^{-a\Delta}\widetilde{X}_m^\varepsilon\right),\\
\widetilde{G}^1_{M,\varepsilon}(a) &= \frac{1}{\Delta}\sum_{m=0}^{M-1} \widetilde{Z}_m^\varepsilon \left(\widetilde{X}_{m+1}^\varepsilon - e^{-a\Delta}\widetilde{X}_m^\varepsilon\right).
\end{aligned}$$

The solutions of the equations $\widehat{G}^1_{M,\varepsilon}(a) = 0$ and $\widetilde{G}^1_{M,\varepsilon}(a) = 0$ can be computed analytically and are given by

$$\widehat{A}^1_{M,\varepsilon} = -\frac{1}{\Delta}\log\left(\frac{\sum_{m=0}^{M-1}\widetilde{X}_m^\varepsilon\widetilde{X}_{m+1}^\varepsilon}{\sum_{m=0}^{M-1}(\widetilde{X}_m^\varepsilon)^2}\right), \tag{7.16}$$

and

$$\widetilde{A}^1_{M,\varepsilon} = -\frac{1}{\Delta}\log\left(\frac{\sum_{m=0}^{M-1}\widetilde{Z}_m^\varepsilon\widetilde{X}_{m+1}^\varepsilon}{\sum_{m=0}^{M-1}\widetilde{Z}_m^\varepsilon\widetilde{X}_m^\varepsilon}\right). \tag{7.17}$$

Comparing these estimators with the discrete MLE (defined in the same way as the subsampling estimator (1.17)) without filtered data

$$\widehat{\mathrm{MLE}}^\Delta_{M,\varepsilon} = -\frac{\sum_{m=0}^{M-1}\widetilde{X}_m^\varepsilon(\widetilde{X}_{m+1}^\varepsilon - \widetilde{X}_m^\varepsilon)}{\Delta\sum_{m=0}^{M-1}(\widetilde{X}_m^\varepsilon)^2},$$

and the discrete MLE with filtered data

$$\widetilde{\mathrm{MLE}}^\Delta_{M,\varepsilon} = -\frac{\sum_{m=0}^{M-1}\widetilde{Z}_m^\varepsilon(\widetilde{X}_{m+1}^\varepsilon - \widetilde{X}_m^\varepsilon)}{\Delta\sum_{m=0}^{M-1}\widetilde{Z}_m^\varepsilon\widetilde{X}_m^\varepsilon},$$

we notice that they coincide in the limit as Δ vanishes. We remark that we are comparing our estimator with the discrete MLE instead of the analytical formula for the MLE in continuous time since we assume that we are observing our process at discrete times. Therefore, the continuous time MLE has to be approximated using the available discrete data [101, Section 5.3]. In the following theorems we show the asymptotic limit of the estimators. We do not provide a proof for these results since Theorem 7.12 and Theorem 7.14 are particular cases of Theorem 7.8 and Theorem 7.9 respectively, and Theorem 7.13 follows from the proof of Theorem 7.8 as highlighted in Remark 7.26.

Theorem 7.12. *Let Δ be independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta \in (0, 1)$. Then, under Assumption 1.4, the estimator* (7.16) *satisfies*

$$\lim_{\varepsilon \to 0} \lim_{M \to \infty} \widehat{A}^1_{M,\varepsilon} = A, \quad \textit{in probability},$$

where A is the drift coefficient of the homogenized equation (1.11)*.*

Theorem 7.13. *Let Δ be independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta > 2$. Then, under Assumption 1.4, the estimator* (7.16) *satisfies*

$$\lim_{\varepsilon \to 0} \lim_{M \to \infty} \widehat{A}^1_{M,\varepsilon} = \alpha, \quad \textit{in probability},$$

where α is the drift coefficient of the homogenized equation (1.10)*.*

Theorem 7.14. *Let Δ be independent of ε or $\Delta = \varepsilon^\zeta$ with $\zeta \neq 1$, $\zeta \neq 2$. Then, under Assumption 1.4, the estimator* (7.17) *satisfies*

$$\lim_{\varepsilon \to 0} \lim_{M \to \infty} \widetilde{A}^1_{M,\varepsilon} = A, \quad \textit{in probability},$$

where A is the drift coefficient of the homogenized equation (1.11)*.*

Remark 7.15. Notice that it is possible to write different proofs for Theorems 7.12, 7.13 and 7.14, 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.

7.3 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 M of observations and the number J of eigenvalues and eigenfunctions taken into account. In the experiments that we present data are generated employing the Euler–Maruyama (EM) 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 $\{\widetilde{X}^\varepsilon_m\}_{m=0}^M$, where $M = T/\Delta$ and $\widetilde{X}^\varepsilon_m = X^\varepsilon_{t_m}$ with $t_m = m\Delta$. In view of Remark 1.2 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 $\{\widetilde{Z}^\varepsilon_m\}_{m=1}^M$ we employ equation (7.13).

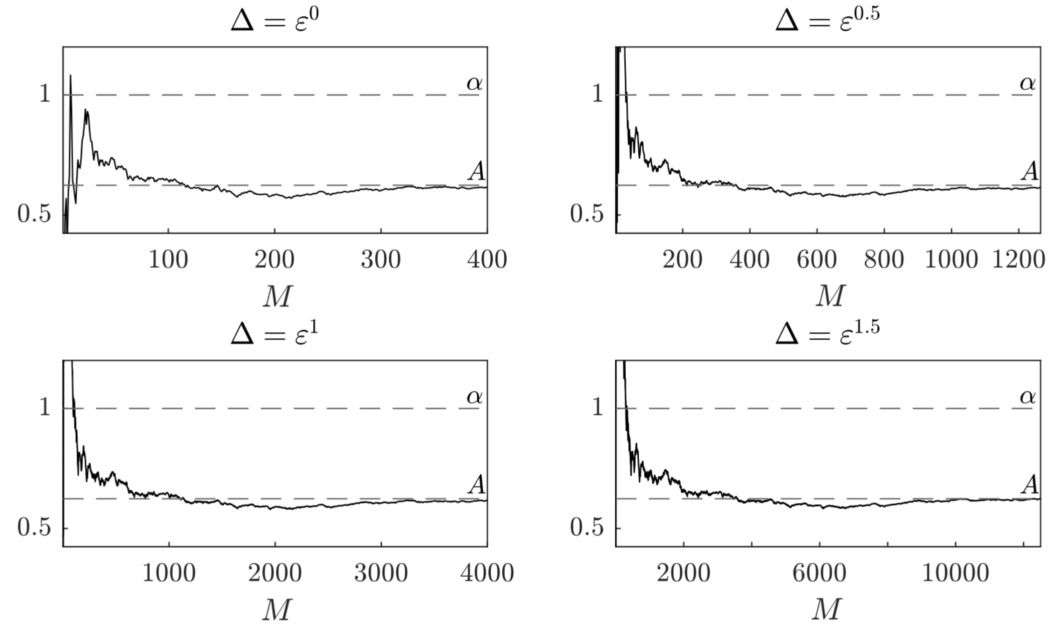


Figure 7.1 – Sensitivity analysis with respect to the number M of observations for different values of $\Delta \leq 1$, for the estimator $\widetilde{A}^J_{M,\varepsilon}$ with $J = 1$.

We repeat this procedure for 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 $\mathcal{K}$ given by the theory of homogenization, but we also remark that its value could be estimated employing the subsampling technique presented in [103] or modifying the estimating function as explained in Remark 7.11.

7.3.1 Sensitivity analysis with respect to the number of observations

We consider the multiscale OU process, i.e. equation (1.10) 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 $\sigma = 1$. Notice that for this choice of the slow-scale potential the technical assumptions required in the main Theorems 7.8, 7.9 can be easily checked. We plot the results computed by the estimator $\widetilde{A}^J_{M,\varepsilon}$ with $J = 1$ and $\psi_1(x; a) = x$ 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 7.1 we plot the results of the estimator as a function of the number of observations M. 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 M 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 $T = 2^{15}$.

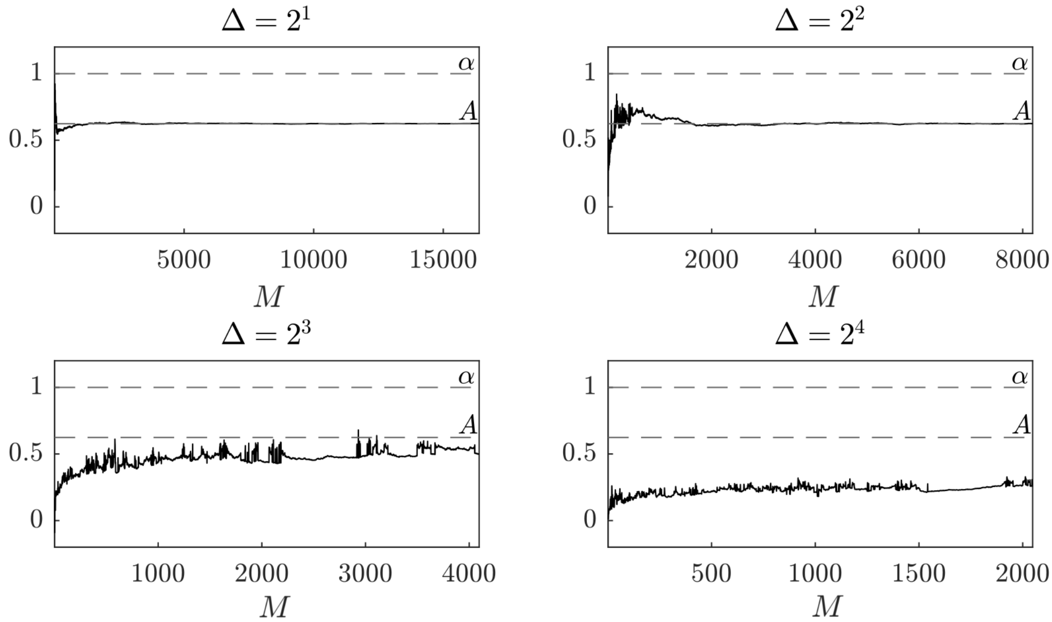


Figure 7.2 – Sensitivity analysis with respect to the number M of observations for different values of $\Delta > 1$, for the estimator $\widetilde{A}^J_{M,\varepsilon}$ with $J = 1$.

Figure 7.2 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 are 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.

7.3.2 Sensitivity analysis with respect to the number of eigenpairs

Let us now consider equation (1.10) with four different slow-scale potentials

$$V_1(x) = \frac{x^2}{2}, \qquad V_2(x) = \frac{x^4}{4}, \qquad V_3(x) = \frac{x^6}{6}, \qquad V_4(x) = \frac{x^4}{4} - \frac{x^2}{2}. \tag{7.18}$$

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 $J = 1, \dots, 10$. The functions $\{\psi_j\}_{j=1}^{10}$ appearing in the estimating function are given by $\psi_j(x; a) = x$ for all $j = 1, \dots, J$.

In Figure 7.3, where we plot the values computed by $\widehat{A}^J_{M,\varepsilon}$ and $\widetilde{A}^J_{M,\varepsilon}$, 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, 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 grow and, therefore, how quickly the corresponding exponential terms 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.

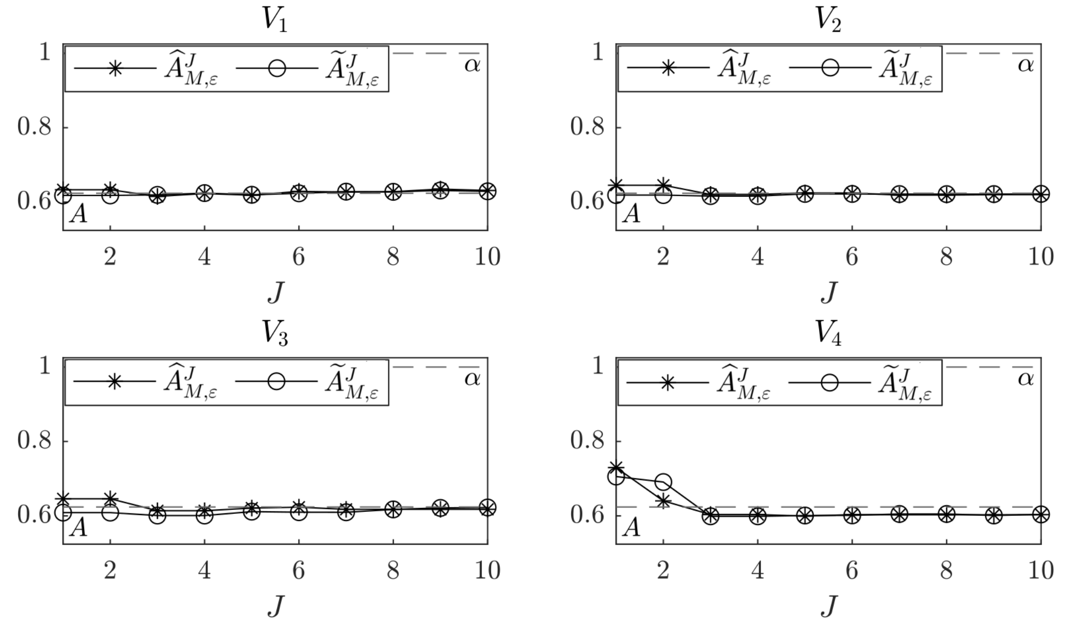


Figure 7.3 – Sensitivity analysis with respect to the number J of eigenvalues and eigenfunctions for different slow-scale potentials, for the estimators $\widehat{A}^J_{M,\varepsilon}$ and $\widetilde{A}^J_{M,\varepsilon}$.

7.3.3 Verification of the theoretical results

We consider the same setting as in the previous subsection, i.e. equation (1.10) with slow-scale potentials given by (7.18) and $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 1$ and $\varepsilon = 0.1$. Moreover, we set $J = 1$, $\psi_1(x; a) = x$ 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 7.4 we compare our martingale estimator $\widehat{A}^J_{M,\varepsilon}$ without filtered data with the discrete maximum likelihood estimator denoted $\widehat{\mathrm{MLE}}^{\Delta}_{M,\varepsilon}$. The MLE does not provide good results for two reasons:

(i) 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;

(ii) 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 investigated in greater detail in [103], there exists an optimal value of Δ such that $\widehat{\mathrm{MLE}}^{\Delta}_{M,\varepsilon}$ works well, but this value is not known a priori and is strongly dependent on the problem, hence this technique is not robust. Figure 7.4 shows that the second issue, i.e., when Δ is relatively big, can be solved employing $\widehat{A}^J_{M,\varepsilon}$, an estimator for discrete observations, and that filtering the data is not needed as proved in Theorem 7.8.

Then, in order to solve also the first problem, in Figure 7.5 we compare $\widehat{A}^J_{M,\varepsilon}$ with our martingale

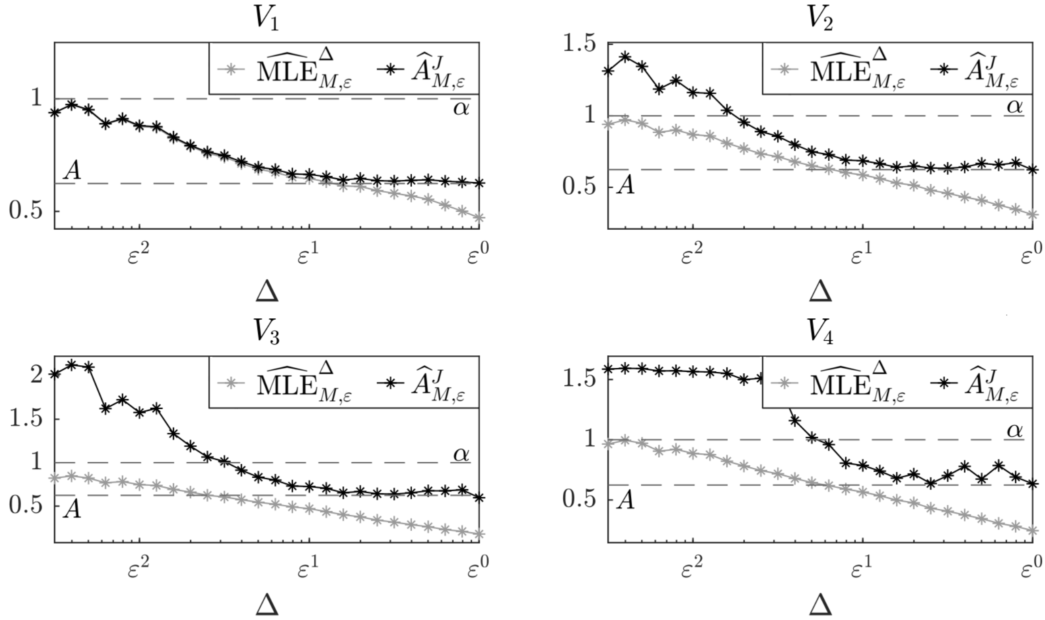


Figure 7.4 – Comparison between the discrete maximum likelihood estimator $\widehat{\mathrm{MLE}}^{\Delta}_{M,\varepsilon}$ presented in [103] and our estimator $\widehat{A}^J_{M,\varepsilon}$ with $J = 1$ without filtered data as a function of the distance Δ between two successive observations for different slow-scale potentials.

estimator $\widetilde{A}^J_{M,\varepsilon}$ with filtered data. We observe that inserting filtered data in the estimator allows us 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 7.9. 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 7.26, we observe that the limiting value of the estimator $\widehat{A}^J_{M,\varepsilon}$ as the number of observations M goes to infinity and the multiscale parameter ε vanishes is strongly dependent on the problem and cannot be computed theoretically. However, if we consider the slow-scale potential $V_1(x) = x^2/2$, i.e. the multiscale OU process, then the limit, as proved in Theorem 7.13, is the drift coefficient α of the multiscale equation.

7.3.4 Multidimensional drift coefficient

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

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

and the fast-scale potential $p(y) = \cos(y)$. We choose the exact drift coefficient of the multiscale equation (1.10) to be $\alpha = \begin{pmatrix} 1.2 & 0.7 \end{pmatrix}^\top$ and the diffusion coefficient to be $\sigma = 0.7$. We also set the number of eigenfunctions $J = 1$, the function $\psi_1(x; a) = \begin{pmatrix} x^3 & x \end{pmatrix}^\top$, the distance between two consecutive observations $\Delta = 1$ and the final time $T = 1000$. We then compute the estimator $\widehat{A}^J_{M,\varepsilon}$ after $M = 100, 200, \dots, 1000$ observations and in Figure 7.6 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 $\widehat{A}^J_{M,\varepsilon}$ which is computationally less expensive to compute.

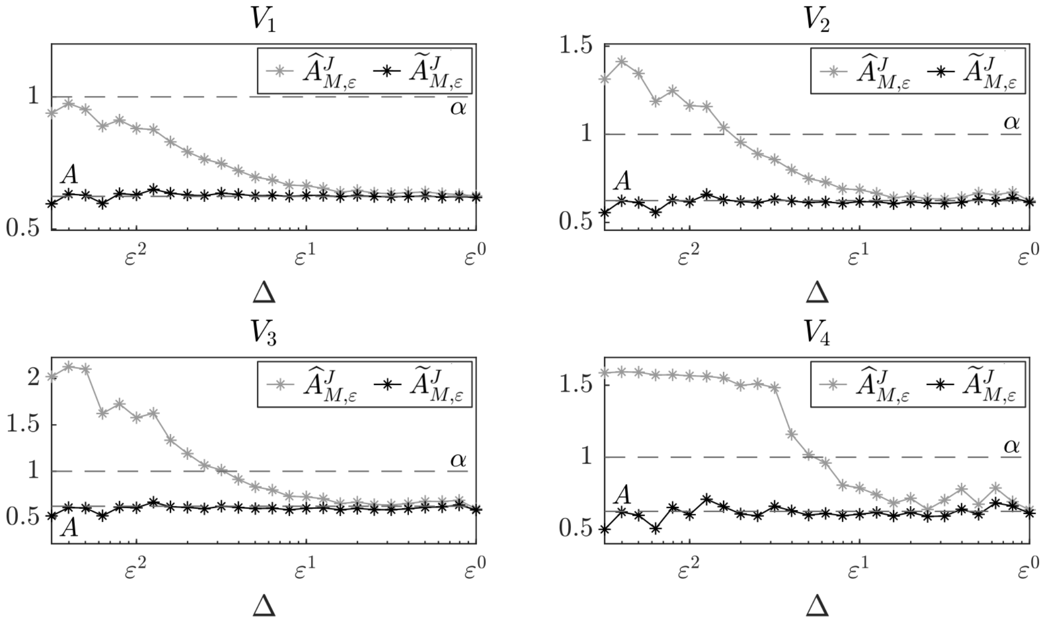


Figure 7.5 – Comparison between our two estimators $\widehat{A}^J_{M,\varepsilon}$ without filtered data and $\widetilde{A}^J_{M,\varepsilon}$ with filtered data with $J = 1$ as a function of the distance Δ between two successive observations for different slow-scale potentials.

M	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 7.1 – Absolute error $\widehat{e}^J_{M,\varepsilon}$ defined in (7.19) between the homogenized drift coefficient A and the estimator $\widehat{A}^J_{M,\varepsilon}$ with $J = 1$ for a two-dimensional drift coefficient.

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 the true value $A = \begin{pmatrix} 0.48 & 0.28 \end{pmatrix}^\top$. 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 7.1 we report the absolute error $\widehat{e}^J_{M,\varepsilon}$ defined as

$$\widehat{e}^J_{M,\varepsilon} = \left\| A - \widehat{A}^J_{M,\varepsilon} \right\|_2, \tag{7.19}$$

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

7.3.5 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 [59]. For $t \in [0, T]$ and for all

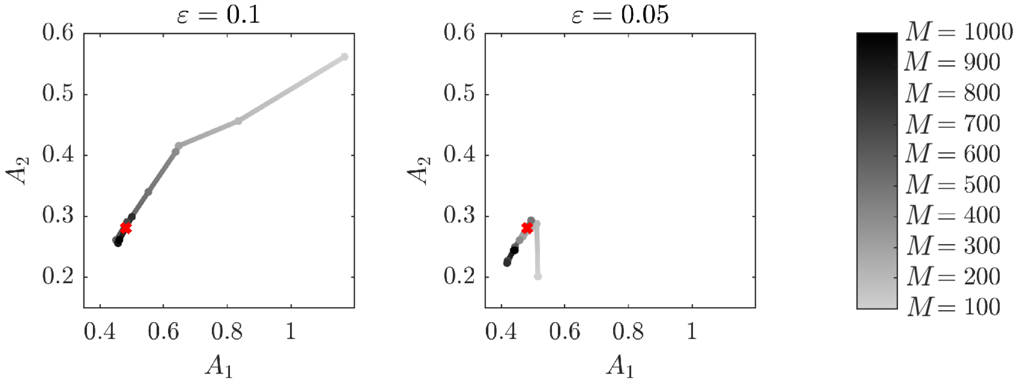


Figure 7.6 – Evolution in time of the estimator $\widehat{A}^J_{M,\varepsilon}$ with $J = 1$ for a two-dimensional drift coefficient.

$i = 1, \dots, d$, consider the system of SDEs

$$\mathrm{d}X_i^\varepsilon(t) = -\alpha X_i^\varepsilon(t)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_i^\varepsilon(t)}{\varepsilon}\right) - \frac{\theta}{d}\sum_{j=1}^d \left(X_i^\varepsilon(t) - X_j^\varepsilon(t)\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_i(t). \tag{7.20}$$

In this section 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 in Chapter 8. It can be shown (see e.g. [59, Section 2.1] and [40,42]) that $(X_1^\varepsilon, \dots X_d^\varepsilon)$ converges in law as ε goes to zero to the solution $(X_1^0, \dots, X_d^0)$ of the homogenized system

$$\mathrm{d}X_i^0(t) = -AX_i^0(t)\,\mathrm{d}t - \frac{\Theta}{d}\sum_{j=1}^d \left(X_i^0(t) - X_j^0(t)\right)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_i(t). \tag{7.21}$$

where $\Theta = \mathcal{K}\theta$ and $\mathcal{K}$ is defined in (1.6). Moreover, the first eigenvalue and eigenfunction of the generator of the homogenized system can be computed explicitly and they are given respectively by

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

Hence, letting $\Delta > 0$ independent of ε, given a sequence of observations $((\widetilde{X}_1^\varepsilon)_m, \dots (\widetilde{X}_d^\varepsilon)_m)_{m=0}^M$, we can express the estimators analytically

$$\widehat{A}^1_{M,\varepsilon} = -\frac{1}{\Delta}\log\left(\frac{\sum_{m=0}^{M-1}\left(\sum_{i=1}^d (\widetilde{X}_i^\varepsilon)_m\right)\left(\sum_{i=1}^d (\widetilde{X}_i^\varepsilon)_{m+1}\right)}{\sum_{m=0}^{M-1}\left(\sum_{i=1}^d (\widetilde{X}_i^\varepsilon)_m\right)^2}\right),$$

$$\widetilde{A}^1_{M,\varepsilon} = -\frac{1}{\Delta}\log\left(\frac{\sum_{m=0}^{M-1}\left(\sum_{i=1}^d (\widetilde{Z}_i^\varepsilon)_m\right)\left(\sum_{i=1}^d (\widetilde{X}_i^\varepsilon)_{m+1}\right)}{\sum_{m=0}^{M-1}\left(\sum_{i=1}^d (\widetilde{Z}_i^\varepsilon)_m\right)\left(\sum_{i=1}^d (\widetilde{X}_i^\varepsilon)_m\right)}\right).$$

Let us now set $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 1$ and $\theta = 1$. We then simulate system (7.20) for different final times $T = 100, 200, \dots, 1000$ and approximate the drift coefficient A of the homogenized system (7.21) for $d = 2$ and $d = 5$. In Figure 7.7 and Figure 7.8 we plot the results respectively of the estimators $\widehat{A}^J_{M,\varepsilon}$ with $\Delta = 1$ and $\widetilde{A}^J_{M,\varepsilon}$ 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 $T = 500$.

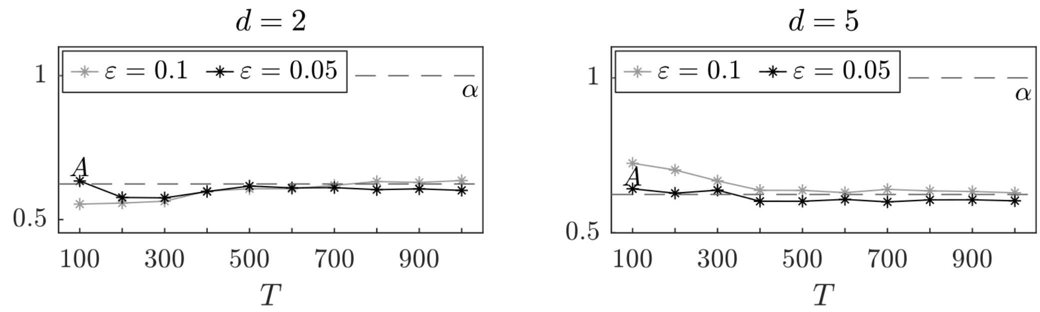


Figure 7.7 – Evolution in time of the estimator $\widehat{A}^J_{M,\varepsilon}$ with $J = 1$ for a d-dimensional system of interacting particles with sampling rate $\Delta = 1$.

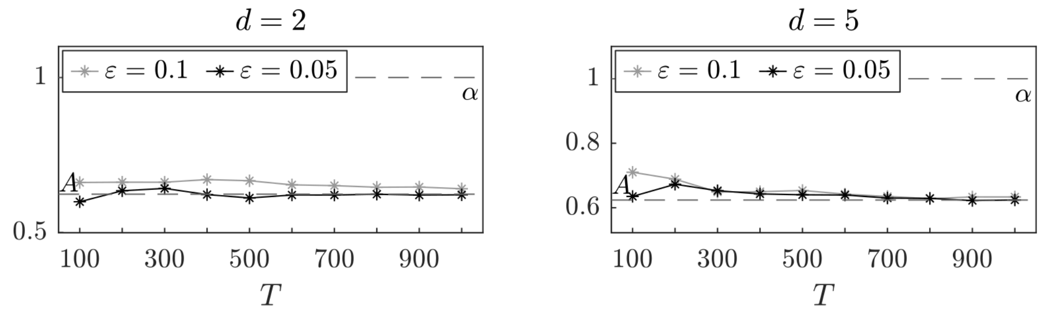


Figure 7.8 – Evolution in time of the estimator $\widetilde{A}^J_{M,\varepsilon}$ with $J = 1$ for a d-dimensional system of interacting particles with sampling rate $\Delta = \varepsilon$.

7.3.6 Simultaneous inference of drift and diffusion coefficients

As highlighted by Remark 7.11, a small modification of our methodology allows us to estimate the diffusion coefficient, in addition to drift coefficients. Define the parameter $\theta = \begin{pmatrix} a^\top & s \end{pmatrix}^\top \in \mathbb{R}^{L+1}$, whose exact value is given by $\theta_0 = \begin{pmatrix} A^\top & \Sigma \end{pmatrix}^\top \in \mathbb{R}^{L+1}$, where A and Σ are the drift and diffusion coefficients of the homogenized equation, respectively. Then, the eigenvalue problem reads for all $j \in \mathbb{N}$

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

where the eigenvalues and eigenfunctions are now dependent on the new parameter θ. Accordingly, also the functions $\{\psi_j\}_{j=1}^J$ can be chosen dependent on both the drift and diffusion coefficients and, moreover, they have to take values in $\mathbb{R}^{L+1}$, i.e., $\psi_j(\cdot;\theta)\colon \mathbb{R} \to \mathbb{R}^{L+1}$. Therefore, the new score functions $\widehat{G}^J_{M,\varepsilon}$ and $\widetilde{G}^J_{M,\varepsilon}$ are defined from $\Theta = \mathcal{A} \times \mathcal{S} \subset \mathbb{R}^{L+1}$, which is the set of admissible parameters θ, to $\mathbb{R}^{L+1}$ and thus give nonlinear systems of dimension $L + 1$. Finally, the solutions $\widehat{\theta}^1_{M,\varepsilon}$ and $\widetilde{\theta}^1_{M,\varepsilon}$ of the systems are the estimators of both the drift and diffusion coefficients of the homogenized equation. In fact, small modifications in the proofs of the main results, in particular in the notation, yield the asymptotic unbiasedness of the estimators under the same conditions, i.e.,

$$\lim_{\varepsilon\to 0}\lim_{M\to\infty} \widehat{\theta}^1_{M,\varepsilon} = \lim_{\varepsilon\to 0}\lim_{M\to\infty} \widetilde{\theta}^1_{M,\varepsilon} = \theta_0 = \begin{pmatrix} A^\top & \Sigma \end{pmatrix}^\top, \qquad \text{in probability.}$$

Consider now the same setting of Section 7.3.1, i.e., the multiscale OU potential with $V(x) = x^2/2$, $p(y) = \cos(y)$, $\alpha = 1$, $\sigma = 1$ and let us assume that both the drift and diffusion coefficients are unknown. We remark that in this case we have $L = 1$. Then, set the final time $T = 1000$, the sampling rate $\Delta = 1$ and the number of eigenfunctions and eigenvalues $J = 2$. Moreover, we

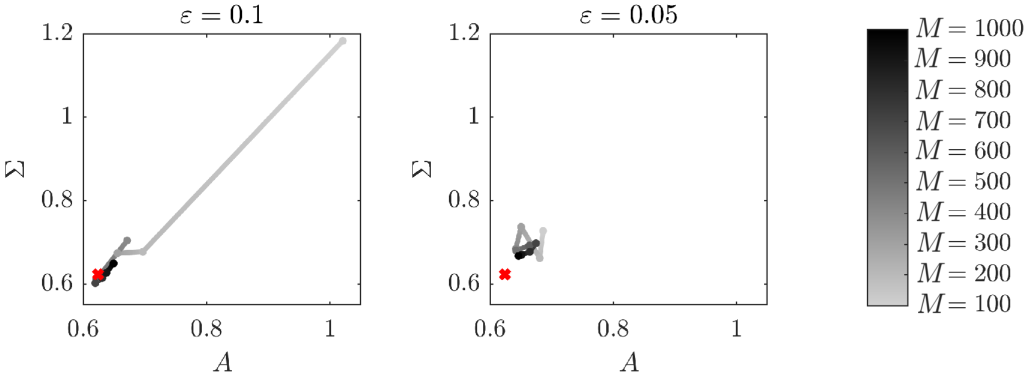


Figure 7.9 – Simultaneous inference of drift and diffusion coefficient for the estimator $\widehat{A}^J_{M,\varepsilon}$ with $J = 2$.

choose the functions $\psi_1(x;\theta) = \psi_2(x;\theta) = \begin{pmatrix} x^2 & x \end{pmatrix}^\top$. Since the distance between two consecutive observations is independent of the multiscale parameter ε, we consider the estimator $\widehat{A}^J_{M,\varepsilon}$ without filtered data. In Figure 7.9 we plot the evolution of our estimator varying the number of observations M for two different values of ε, in particular $\varepsilon = 0.1$ and $\varepsilon = 0.05$. We observe that if the multiscale parameter is smaller, then the number of observations needed to obtain a reliable approximation of the unknown parameters is lower.

7.4 Asymptotic unbiasedness

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

(i) we first study the limiting behaviour of the score functions $\widehat{G}^J_{M,\varepsilon}$ and $\widetilde{G}^J_{M,\varepsilon}$ defined in (7.8) and (7.11) as the number of observations M goes to infinity, i.e., as the final time T tends to infinity;

(ii) 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 7.4.1);

(iii) we finally prove our main results, i.e., the asymptotic unbiasedness of the drift estimators (Section 7.4.2).

We first define the Jacobian matrix of the function g_j introduced in (7.7) with respect to a:

$$\begin{aligned} h_j(x,y,z;a) &= \dot{\psi}_j(z;a)\left(\phi_j(y,a) - e^{-\lambda_j(a)\Delta}\phi_j(x;a)\right) \\ &\quad + \psi_j(z;a) \otimes \left(\dot{\phi}_j(y;a) - e^{-\lambda_j(a)\Delta}\left(\dot{\phi}_j(x;a) - \Delta\dot{\lambda}_j(a)\phi_j(x,a)\right)\right), \end{aligned} \tag{7.22}$$

which will be employed in the following and where $\otimes$ denotes the outer product in $\mathbb{R}^L$ 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 1.4, due to ergodicity and stationarity and by [21, Lemma 3.1] we have

$$\lim_{M\to\infty} \frac{1}{M}\widehat{G}^J_{M,\varepsilon}(a) = \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\nu^\varepsilon}\left[g_j\left(X_0^\varepsilon, X_\Delta^\varepsilon, X_0^\varepsilon; a\right)\right] =: \widehat{\mathcal{G}}_J(\varepsilon,a),$$

and

$$\lim_{M\to\infty} \frac{1}{M}\widetilde{G}^J_{M,\varepsilon}(a) = \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}}\left[g_j\left(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a\right)\right] =: \widetilde{\mathcal{G}}_J(\varepsilon,a), \tag{7.23}$$

where $\mathbb{E}^{\nu^\varepsilon}$ and $\mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}$ denotes respectively that X_0^ε and $(X_0^\varepsilon, \widetilde{Z}_0^\varepsilon)$ are distributed according to their invariant distribution. We remark that the invariant distribution $\widetilde{\mu}^\varepsilon_{\exp}$ exists due to Lemma 7.28. By equation (7.22) the Jacobian matrices of $\widehat{\mathcal{G}}_J(\varepsilon, a)$ and $\widetilde{\mathcal{G}}_J(\varepsilon, a)$ with respect to a are given by

$$\widehat{\mathcal{H}}_J(\varepsilon, a) \coloneqq \frac{\partial}{\partial a}\widehat{\mathcal{G}}_J(\varepsilon, a) = \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\nu^\varepsilon}\left[h_j\left(X_0^\varepsilon, X_\Delta^\varepsilon, X_0^\varepsilon; a\right)\right],$$

and

$$\widetilde{\mathcal{H}}_J(\varepsilon, a) \coloneqq \frac{\partial}{\partial a}\widetilde{\mathcal{G}}_J(\varepsilon, a) = \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}\left[h_j\left(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a\right)\right]. \tag{7.24}$$

7.4.1 Continuity of the limit of the score function

In this section, we first prove the continuity of the functions $\widehat{\mathcal{G}}_J, \widetilde{\mathcal{G}}_J \colon (0,\infty)\times\mathcal{A} \to \mathbb{R}^L$ and $\widehat{\mathcal{H}}_J, \widetilde{\mathcal{H}}_J, \colon (0,\infty)\times\mathcal{A} \to \mathbb{R}^{L\times L}$. 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 Sections 7.5.1 and 7.5.3.

Lemma 7.16. *Let $\widetilde{Z}^\varepsilon$ be defined in* (7.10) *and distributed according to the invariant measure $\widetilde{\mu}^\varepsilon_{\exp}$ of the process $(\widetilde{X}_m, \widetilde{Z}_m)$. Then for any $p \geq 1$ there exists a constant $C > 0$ uniform in ε such that*

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

Lemma 7.17. *Let $f\colon \mathbb{R}\to\mathbb{R}$ 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

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

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

Proposition 7.18. *Under Assumption 7.2, the functions $\widetilde{\mathcal{G}}_J\colon (0,\infty)\times\mathcal{A}\to\mathbb{R}^L$ and its derivative $\widetilde{\mathcal{H}}_J, \colon (0,\infty)\times\mathcal{A}\to\mathbb{R}^{L\times L}$ defined in* (7.23) *and* (7.24)*, 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 ψ_j and ϕ_j are continuously differentiable with respect to a for all $j = 1, \dots, J$ respectively due to Assumption 7.2 and Lemma 7.30, then also g_j is continuously differentiable with respect to a. Therefore, by the mean value theorem for vector-valued functions we have

$$\begin{aligned} Q_1(\varepsilon, a) &\leq \frac{1}{\Delta} \sum_{j=1}^J \left\| \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a) \right] - \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*) \right] \right\| \\ &= \frac{1}{\Delta} \sum_{j=1}^J \left\| \int_0^1 \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[h_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^* + t(a - a^*)) \right] \mathrm{d}t \; (a - a^*) \right\|. \end{aligned}$$

Then, letting $C > 0$ be a constant independent of ε, since ψ_j and ϕ_j are polynomially bounded still by Assumption 7.2 and X_0^ε, X_Δ^ε and $\widetilde{Z}_0^\varepsilon$ have bounded moments of any order by [103, Corollary 5.4] and Lemma 7.16, we obtain

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

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

$$\begin{aligned} \lim_{\varepsilon \to \varepsilon^*} Q_2(\varepsilon) &= \lim_{\varepsilon \to \varepsilon^*} \left\| \frac{1}{\Delta} \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*) \right] - \frac{1}{\Delta} \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^{\varepsilon^*}_{\mathrm{exp}}} \left[g_j(X_0^{\varepsilon^*}, X_\Delta^{\varepsilon^*}, \widetilde{Z}_0^{\varepsilon^*}; a^*) \right] \right\| \\ &\leq \lim_{\varepsilon \to \varepsilon^*} \frac{1}{\Delta} \sum_{j=1}^J \left\| \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*) \right] - \mathbb{E}^{\widetilde{\mu}^{\varepsilon^*}_{\mathrm{exp}}} \left[g_j(X_0^{\varepsilon^*}, X_\Delta^{\varepsilon^*}, \widetilde{Z}_0^{\varepsilon^*}; a^*) \right] \right\|, \end{aligned}$$

and the right hand side vanishes due to the continuity of g_j for all $j = 1, \dots, J$ and the continuity of the solution of an SDE with respect to a parameter (see [76, 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 $j = 1, \dots, J$

$$\begin{aligned} \phi_j(X_\Delta^\varepsilon; a^*) &= \phi_j(X_{\Delta^*}^\varepsilon; a^*) - \alpha \cdot \int_{\Delta^*}^\Delta V'(X_t^\varepsilon) \phi_j'(X_t^\varepsilon; a^*) \,\mathrm{d}t - \frac{1}{\varepsilon} \int_{\Delta^*}^\Delta \phi_j'(X_t^\varepsilon; a^*) p' \left(\frac{X_t^\varepsilon}{\varepsilon} \right) \mathrm{d}t \\ &\quad + \sigma \int_{\Delta^*}^\Delta \phi_j''(X_t^\varepsilon; a^*) \,\mathrm{d}t + \sqrt{2\sigma} \int_{\Delta^*}^\Delta \phi_j'(X_t^\varepsilon; a^*) \,\mathrm{d}W_t, \end{aligned}$$

then we can write

$$\begin{aligned} \widetilde{\mathcal{G}}_J(\varepsilon, a^*) = \frac{1}{\Delta} \sum_{j=1}^J & \left(\mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[\psi_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j(X_{\Delta^*}^\varepsilon; a^*) \right] - e^{-\lambda(a^*)\Delta} \, \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\mathrm{exp}}} \left[\psi_j(\widetilde{Z}_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) \right] \right) \\ & + R(\varepsilon), \end{aligned}$$

where $R(\varepsilon)$ is given by

$$
\begin{aligned}
R(\varepsilon) = &-\frac{1}{\Delta}\sum_{j=1}^{J}\int_{\Delta^*}^{\Delta}\mathbb{E}^{\widetilde{\mu}^{\varepsilon}_{\mathrm{exp}}}\left[\psi_j(\widetilde{Z}^{\varepsilon}_0;a^*)\phi_j'(X^{\varepsilon}_t;a^*)\alpha\cdot V'(X^{\varepsilon}_t)\right]\,\mathrm{d}t \\
&-\frac{1}{\varepsilon\Delta}\sum_{j=1}^{J}\int_{\Delta^*}^{\Delta}\mathbb{E}^{\widetilde{\mu}^{\varepsilon}_{\mathrm{exp}}}\left[\psi_j(\widetilde{Z}^{\varepsilon}_0;a^*)\phi_j'(X^{\varepsilon}_t;a^*)p'\left(\frac{X^{\varepsilon}_t}{\varepsilon}\right)\right]\,\mathrm{d}t \\
&+\frac{\sigma}{\Delta}\int_{\Delta^*}^{\Delta}\mathbb{E}^{\widetilde{\mu}^{\varepsilon}_{\mathrm{exp}}}\left[\psi_j(\widetilde{Z}^{\varepsilon}_0;a^*)\phi_j''(X^{\varepsilon}_t;a^*)\right]\,\mathrm{d}t+\frac{\sqrt{2\sigma}}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\mu}^{\varepsilon}_{\mathrm{exp}}}\left[\int_{\Delta^*}^{\Delta}\psi_j(\widetilde{Z}^{\varepsilon}_0;a^*)\phi_j'(X^{\varepsilon}_t;a^*)\,\mathrm{d}W_t\right].
\end{aligned}
$$

Let $C>0$ be independent of ε and notice that since p' is bounded, $\psi_j,\phi_j',\phi_j'',V'$ are polynomially bounded and X^{ε}_t and $\widetilde{Z}^{\varepsilon}_0$ have bounded moments of any order by [103, Corollary 5.4] and Lemma 7.16, applying Hölder's inequality we obtain

$$
\|R(\varepsilon)\| \le \frac{C}{\Delta}\left(\|\alpha\|+\sigma+\frac{1}{\varepsilon}\right)(\Delta-\Delta^*)+\frac{C}{\Delta}\sqrt{2\sigma}(\Delta-\Delta^*)^{1/2}. \tag{7.25}
$$

Therefore, by the continuity of the solution of an SDE with respect to a parameter (see [88]) and due to the bound (7.25), we deduce that

$$
\begin{aligned}
\lim_{\varepsilon\to\varepsilon^*}\widetilde{\mathcal{G}}_J(\varepsilon,a^*) &= \frac{1}{\Delta^*}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\mu}^{\varepsilon^*}_{\mathrm{exp}}}\left[\psi_j(\widetilde{Z}^{\varepsilon^*}_0;a^*)\left(\phi_j(X^{\varepsilon^*}_{\Delta^*};a^*)-e^{-\lambda(a^*)\Delta^*}\phi_j(X^{\varepsilon^*}_0;a^*)\right)\right] \\
&= \widetilde{\mathcal{G}}_J(\varepsilon^*,a^*),
\end{aligned}
$$

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

Remark 7.19. Notice that the proof of Proposition 7.18 can be repeated analogously for the functions $\widehat{\mathcal{G}}_J\colon(0,\infty)\times\mathcal{A}\to\mathbb{R}^L$ and $\widehat{\mathcal{H}}_J\colon(0,\infty)\times\mathcal{A}\to\mathbb{R}^{L\times L}$ 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 7.20) data are sampled at the homogenized regime ignoring the fact that the they are generated by a multiscale model, while in the second case (Proposition 7.22) the distance between two consecutive observations is proportional to the multiscale parameter and thus data are sampled at the multiscale regime preserving the multiscale structure of the full path.

Proposition 7.20. *Let the functions $\widetilde{\mathcal{G}}_J\colon(0,\infty)\times\mathcal{A}\to\mathbb{R}^L$ and $\widetilde{\mathcal{H}}_J,\colon(0,\infty)\times\mathcal{A}\to\mathbb{R}^{L\times L}$ be defined in* (7.23) *and* (7.24) *and let Δ be independent of ε. Under Assumption 7.2 and for any $a^*\in\mathcal{A}$ we have*

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

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

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

$$
\left\|\widetilde{\mathcal{G}}_J(\varepsilon,a)-\frac{1}{\Delta}\sum_{j=1}^{J}\mathbb{E}^{\widetilde{\mu}^{0}_{\mathrm{exp}}}\left[g_j\left(X^0_0,X^0_\Delta,\widetilde{Z}^0_0;a^*\right)\right]\right\| \le Q_1(\varepsilon,a)+Q_2(\varepsilon),
$$

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 7.18 and

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

Let us remark that the convergence in law of the joint process $\{(\widetilde{X}_m^\varepsilon, \widetilde{Z}_m^\varepsilon)\}_{m=0}^M$ to the joint process $\{(\widetilde{X}_m^0, \widetilde{Z}_m^0)\}_{m=0}^M$ by Lemma 7.28 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

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

which implies the desired result. □

Remark 7.21. Similar results to Proposition 7.18 and Proposition 7.20 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

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

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

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

Proposition 7.22. *Let the functions $\widetilde{\mathcal{G}}_J \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^L$ and $\widetilde{\mathcal{H}}_J, \colon (0, \infty) \times \mathcal{A} \to \mathbb{R}^{L \times L}$ be defined in* (7.23) *and* (7.24) *and let $\Delta = \varepsilon^\zeta$ with $\zeta > 0$ and $\zeta \neq 1$, $\zeta \neq 2$. Under Assumption 7.2 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*

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

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

$$\begin{aligned} \widetilde{\mathfrak{h}}_J^0(a) \coloneqq & \sum_{j=1}^J \mathbb{E}^{\mu^0_{\exp}} \left[\dot{\psi}_j(Z_0^0; a) \left(\mathcal{L}_A \phi_j(X_0^0; a) + \lambda_j(a) \phi_j(X_0^0; a) \right) \right] \\ & + \sum_{j=1}^J \mathbb{E}^{\mu^0_{\exp}} \left[\psi_j(Z_0^0; a) \otimes \left(\mathcal{L}_A \dot{\phi}_j(X_0^0; a) + \lambda_j(a) \dot{\phi}_j(X_0^0; a) \right) \right] \\ & + \sum_{j=1}^J \mathbb{E}^{\mu^0_{\exp}} \left[\psi_j(Z_0^0; a) \phi_j(X_0^0; a) \right] \otimes \dot{\lambda}_j(a), \end{aligned}$$

where the generator $\mathcal{L}_A$ is defined in (7.3).

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

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

then we need to show that the two terms vanish and we distinguish two cases.
Case 1: $\zeta \in (0,1)$.
Applying Lemma 7.17 to the functions $\phi_j(\cdot;a^*)$ for all $j=1,\dots,J$ and noting that

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

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

$$\begin{aligned}
\widetilde{\mathcal{G}}_J(\varepsilon,a^*) &= \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}\left[\psi_j(\widetilde{Z}_0^\varepsilon;a^*)\left(\phi_j(X_\Delta^\varepsilon;a^*)-e^{-\lambda_j(a^*)\Delta}\phi_j(X_0^\varepsilon;a^*)\right)\right]\\
&= \frac{1-e^{-\lambda_j(a^*)\Delta}}{\Delta}\sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}\left[\psi_j(\widetilde{Z}_0^\varepsilon;a^*)\phi_j(X_0^\varepsilon;a^*)\right] + \sum_{j=1}^J \frac{1}{\Delta}\,\mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}\left[\psi_j(\widetilde{Z}_0^\varepsilon;a^*)R(\varepsilon,\Delta)\right]\\
&\quad + \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}\left[\psi_j(\widetilde{Z}_0^\varepsilon;a^*)\left(\Sigma\phi_j''(X_0^\varepsilon;a^*)-A\cdot V'(X_0^\varepsilon)\phi_j'(X_0^\varepsilon;a^*)\right)\right]\\
&=: I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon,
\end{aligned}$$

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

$$\left(\mathbb{E}^{\widetilde{\mu}^\varepsilon_{\exp}}\left|R(\varepsilon,\Delta)\right|^p\right)^{1/p} \leq C(\varepsilon+\Delta^{3/2}). \tag{7.26}$$

We now study the three terms separately. First, by Cauchy-Schwarz inequality, since $\psi_j(\cdot;a^*)$ is polynomially bounded, $\widetilde{Z}_0^\varepsilon$ has bounded moments of any order by Lemma 7.16 and due to (7.26) we obtain

$$\|I_2^\varepsilon\| \leq C\left(\varepsilon\Delta^{-1}+\Delta^{1/2}\right). \tag{7.27}$$

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

$$\begin{aligned}
I_1^\varepsilon &= \frac{1-e^{-\lambda_j(a^*)\Delta}}{\Delta}\sum_{j=1}^J \mathbb{E}^{\mu^\varepsilon_{\exp}}\left[\psi_j(Z_0^\varepsilon;a^*)\phi_j(X_0^\varepsilon;a^*)\right]\\
&\quad + \frac{1-e^{-\lambda_j(a^*)\Delta}}{\Delta}\sum_{j=1}^J \mathbb{E}\left[\left(\psi_j(\widetilde{Z}_0^\varepsilon;a^*)-\psi_j(Z_0^\varepsilon;a^*)\right)\phi_j(X_0^\varepsilon;a^*)\right],
\end{aligned}$$

where Z_0^ε is distributed according to the invariant measure $\mu^\varepsilon_{\exp}$ of the continuous process $(X_t^\varepsilon, Z_t^\varepsilon)$ and

$$\lim_{\varepsilon\to 0}\frac{1-e^{-\lambda_j(a^*)\Delta}}{\Delta} = \lambda_j(a^*). \tag{7.28}$$

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

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

and since $\psi_j'(\cdot;a^*), \phi_j(\cdot;a^*)$ are polynomially bounded, X_0^ε, Z_0^ε, $\widetilde Z_0^\varepsilon$ have bounded moments of any order respectively by [103, Corollary 5.4], Lemma 2.28 and Lemma 7.16 and applying Hölder's inequality and Corollary 7.29 we obtain

$$\left\| \mathbb{E}\left[\left(\psi_j(\widetilde Z_0^\varepsilon;a^*) - \psi_j(Z_0^\varepsilon;a^*) \right) \phi_j(X_0^\varepsilon;a^*) \right] \right\| \le C \left(\Delta^{1/2} + \varepsilon \right). \tag{7.29}$$

Moreover, notice that by homogenization theory (see Section 2.1.2) the joint process $(X_0^\varepsilon, Z_0^\varepsilon)$ converges in law to the joint process (X_0^0, Z_0^0) and therefore

$$\lim_{\varepsilon\to 0} \mathbb{E}^{\mu^\varepsilon_{\exp}} \left[\psi_j(Z_0^\varepsilon;a^*)\phi_j(X_0^\varepsilon;a^*) \right] = \mathbb{E}^{\mu^0_{\exp}} \left[\psi_j(Z_0^0;a^*)\phi_j(X_0^0;a^*) \right],$$

which together with (7.28) and (7.29) yields

$$\lim_{\varepsilon\to 0} I_1^\varepsilon = \sum_{j=1}^J \lambda_j(a^*)\, \mathbb{E}^{\mu^0_{\exp}} \left[\psi_j(Z_0^0;a^*)\phi_j(X_0^0;a^*) \right]. \tag{7.30}$$

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

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

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

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

Therefore, we obtain

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

which together with (7.27) and (7.30) implies

$$\begin{aligned} &\lim_{\varepsilon\to 0} \widetilde{\mathcal{G}}_J(\varepsilon,a^*) \\ &\quad = \sum_{j=1}^J \mathbb{E}^{\mu^0_{\exp}} \left[\psi_j(Z_0^0;a) \left(\Sigma\phi_j''(X_0^0;a^*) - A\cdot V'(X_0^0)\phi_j'(X_0^0;a^*) + \lambda_j(a^*)\phi_j(X_0^0;a^*) \right) \right], \end{aligned} \tag{7.31}$$

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 7.18 we have

$$\begin{aligned} Q_1(\varepsilon,a) &\le \frac{1}{\Delta} \sum_{j=1}^J \left\| \mathbb{E}^{\widetilde\mu^\varepsilon_{\exp}} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde Z_0^\varepsilon; a) \right] - \mathbb{E}^{\widetilde\mu^\varepsilon_{\exp}} \left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde Z_0^\varepsilon; a^*) \right] \right\| \\ &\le \sum_{j=1}^J \left\| \frac{1}{\Delta}\, \mathbb{E}^{\widetilde\mu^\varepsilon_{\exp}} \left[h_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde Z_0^\varepsilon; \widetilde a) \right] \right\| \left\| (a-a^*) \right\|, \end{aligned}$$

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

$$Q_1(\varepsilon,a) \le C \left\| a - a^* \right\|,$$

which together with (7.31) gives the desired result.
Case 2: $\zeta \in (1,2) \cup (2,\infty)$.
Let Z_0^ε be distributed according to the invariant measure $\mu_{\text{exp}}^\varepsilon$ of the continuous process $(X_t^\varepsilon, Z_t^\varepsilon)$ and define

$$\begin{aligned}\widetilde{R}(\varepsilon,\Delta) &\coloneqq \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}_{\text{exp}}^\varepsilon}\left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, \widetilde{Z}_0^\varepsilon; a^*)\right] - \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\mu_{\text{exp}}^\varepsilon}\left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, Z_0^\varepsilon; a^*)\right] \\ &= \frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}\left[\left(\psi_j(\widetilde{Z}_0^\varepsilon; a^*) - \psi_j(Z_0^\varepsilon; a^*)\right)\left(\phi_j(X_\Delta^\varepsilon; a^*) - e^{-\lambda_j(a^*)\Delta}\phi_j(X_0^\varepsilon; a^*)\right)\right].\end{aligned}$$

Then we have

$$\widetilde{\mathcal{G}}_J(\varepsilon, a^*) = \sum_{j=1}^J \frac{1}{\Delta}\mathbb{E}^{\mu_{\text{exp}}^\varepsilon}\left[g_j(X_0^\varepsilon, X_\Delta^\varepsilon, Z_0^\varepsilon; a^*)\right] + \widetilde{R}(\varepsilon,\Delta) =: \sum_{j=1}^J Q_j^\varepsilon + \widetilde{R}(\varepsilon,\Delta), \tag{7.32}$$

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

$$\begin{aligned}\phi_j(X_\Delta^\varepsilon; a^*) = \phi_j(X_0^\varepsilon; a^*) &- \int_0^\Delta \alpha \cdot V'(X_t^\varepsilon)\phi_j'(X_t^\varepsilon; a^*)\,\mathrm{d}t - \int_0^\Delta \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\phi_j'(X_t^\varepsilon; a^*)\,\mathrm{d}t \\ &+ \sigma\int_0^\Delta \phi_j''(X_t^\varepsilon; a^*)\,\mathrm{d}t + \sqrt{2\sigma}\int_0^\Delta \phi_j'(X_t^\varepsilon; a^*)\,\mathrm{d}W_t,\end{aligned} \tag{7.33}$$

and observing that

$$\mathbb{E}\left[\left(\psi_j(\widetilde{Z}_0^\varepsilon; a^*) - \psi_j(Z_0^\varepsilon; a^*)\right)\int_0^\Delta \phi_j'(X_t^\varepsilon; a^*)\,\mathrm{d}W_t\right] = 0, \tag{7.34}$$

since

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

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

$$\begin{aligned}\widetilde{R}(\varepsilon,\Delta) = &\sum_{j=1}^J \frac{1 - e^{-\lambda_j(a^*)\Delta}}{\Delta}\mathbb{E}\left[\left(\psi_j(\widetilde{Z}_0^\varepsilon; a^*) - \psi_j(Z_0^\varepsilon; a^*)\right)\phi_j(X_0^\varepsilon; a^*)\right] \\ &+ \sum_{j=1}^J \frac{1}{\Delta}\int_0^\Delta \mathbb{E}\left[\left(\psi_j(\widetilde{Z}_0^\varepsilon; a^*) - \psi_j(Z_0^\varepsilon; a^*)\right)\left(\sigma\phi_j''(X_t^\varepsilon; a^*) - \alpha\cdot V'(X_t^\varepsilon)\phi_j'(X_t^\varepsilon; a^*)\right)\right]\mathrm{d}t \\ &- \sum_{j=1}^J \frac{1}{\varepsilon\Delta}\int_0^\Delta \mathbb{E}\left[\left(\psi_j(\widetilde{Z}_0^\varepsilon; a^*) - \psi_j(Z_0^\varepsilon; a^*)\right)p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\phi_j'(X_t^\varepsilon; a^*)\right]\mathrm{d}t \\ =:\ &\widetilde{R}_1(\varepsilon,\Delta) + \widetilde{R}_2(\varepsilon,\Delta) + \widetilde{R}_3(\varepsilon,\Delta).\end{aligned}$$

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

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

and since $\psi_j'(\cdot; a^*), \phi_j(\cdot; a^*)$ are polynomially bounded, X_0^ε, Z_0^ε, $\widetilde Z_0^\varepsilon$ have bounded moments of any order respectively by [103, Corollary 5.4], Lemma 2.28 and Lemma 7.16 and applying Hölder's inequality we obtain

$$\left\| \widetilde R_1(\varepsilon, \Delta) \right\| \le C \left(\mathbb{E} \left| \widetilde Z_0^\varepsilon - Z_0^\varepsilon \right|^2 \right)^{1/2}, \tag{7.35}$$

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

$$\left\| \widetilde R_2(\varepsilon, \Delta) \right\| \le C \left(\mathbb{E} \left| \widetilde Z_0^\varepsilon - Z_0^\varepsilon \right|^2 \right)^{1/2} \quad \text{and} \quad \left\| \widetilde R_3(\varepsilon, \Delta) \right\| \le C\varepsilon^{-1} \left(\mathbb{E} \left| \widetilde Z_0^\varepsilon - Z_0^\varepsilon \right|^2 \right)^{1/2},$$

which together with (7.35) yield

$$\left\| \widetilde R(\varepsilon, \Delta) \right\| \le C \left(\mathbb{E} \left| \widetilde Z_0^\varepsilon - Z_0^\varepsilon \right|^2 \right)^{1/2} \left(1 + \varepsilon^{-1} \right). \tag{7.36}$$

Moreover, applying Lemma 7.17 and proceeding similarly to the first part of the first case of the proof we have

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

which together with (7.36) and Corollary 7.29 implies

$$\begin{aligned} \left\| \widetilde R(\varepsilon, \Delta) \right\| &\le C \left(\mathbb{E} \left| \widetilde Z_0^\varepsilon - Z_0^\varepsilon \right|^2 \right)^{1/2} \left(1 + \min\{\varepsilon^{-1}, \varepsilon\Delta^{-1} + \Delta^{1/2}\} \right) \\ &\le C \left(\Delta^{1/2} + \min\{\varepsilon, \varepsilon^{-1}\Delta\} \right) \left(1 + \min\{\varepsilon^{-1}, \varepsilon\Delta^{-1} + \Delta^{1/2}\} \right). \end{aligned} \tag{7.37}$$

Let us now consider Q_j^ε. Replacing equation (7.33) into the definition of Q_j^ε in (7.32) and observing that similarly to (7.34) it holds

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

we obtain

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

We rewrite $\psi_j(Z_0^\varepsilon; a^*)$ inside the integrals employing equation (7.15) and Itô's lemma

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

hence due to stationarity we have

$$Q_j^\varepsilon = Q_{j,1}^\varepsilon + Q_{j,2}^\varepsilon, \tag{7.38}$$

where

$$\begin{aligned} Q_{j,1}^\varepsilon = \frac{1 - e^{-\lambda_j(a^*)}}{\Delta} & \mathbb{E}^{\mu^\varepsilon_{\exp}} \left[\psi_j(Z_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*) \right] - \mathbb{E}^{\mu^\varepsilon_{\exp}} \left[\left(\psi_j(Z_0^\varepsilon; a^*) \otimes V'(X_0^\varepsilon) \right) \phi_j'(X_0^\varepsilon; a^*) \right] \alpha \\ & - \mathbb{E}^{\mu^\varepsilon_{\exp}} \left[\psi_j(Z_0^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_0^\varepsilon}{\varepsilon} \right) \phi_j'(X_0^\varepsilon; a^*) \right] + \sigma \, \mathbb{E}^{\mu^\varepsilon_{\exp}} \left[\psi_j(Z_0^\varepsilon; a^*) \phi_j''(X_0^\varepsilon; a^*) \right] \end{aligned}$$

and

$$
\begin{aligned}
Q_{j,2}^\varepsilon = \frac{1}{\Delta} &\left(\int_0^\Delta \int_0^t \mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[(\psi_j'(Z_s^\varepsilon; a^*) \otimes V'(X_t^\varepsilon)) \phi_j'(X_t^\varepsilon; a^*)(X_s^\varepsilon - Z_s^\varepsilon) \right] \mathrm{d}s \, \mathrm{d}t \right) \alpha \\
&+ \frac{1}{\Delta} \int_0^\Delta \int_0^t \mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[\psi_j'(Z_s^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_t^\varepsilon}{\varepsilon} \right) \phi_j'(X_t^\varepsilon; a^*)(X_s^\varepsilon - Z_s^\varepsilon) \right] \mathrm{d}s \, \mathrm{d}t \\
&- \frac{\sigma}{\Delta} \int_0^\Delta \int_0^t \mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[\psi_j'(Z_s^\varepsilon; a^*) \phi_j''(X_t^\varepsilon; a^*)(X_s^\varepsilon - Z_s^\varepsilon) \right] \mathrm{d}s \, \mathrm{d}t.
\end{aligned}
$$

Since $\phi_j'(\cdot; a^*), \phi_j''(\cdot; a^*)$ and $\psi_j'(\cdot; a^*)$ are polynomially bounded, p' is bounded and X_t^ε and Z_t^ε have bounded moments of any order respectively by [103, Corollary 5.4] and Lemma 2.28, $Q_{j,2}^\varepsilon$ is bounded by

$$
\left\| Q_{j,2}^\varepsilon \right\| \leq C \left(\Delta + \varepsilon^{-1} \Delta \right). \tag{7.39}
$$

Let us now move to $Q_{j,1}^\varepsilon$ and let us recall the functions defined in (2.8) and (2.12)

$$
\mathfrak{R}_{\text{exp}}^\varepsilon(x, z) := \frac{\rho_{\text{exp}}^\varepsilon(x, z)}{\varphi^\varepsilon(x)} \qquad \text{and} \qquad \mathfrak{R}_{\text{exp}}^0(x, z) := \frac{\rho_{\text{exp}}^0(x, z)}{\varphi^0(x)},
$$

where $\rho_{\text{exp}}^\varepsilon$ and ρ_{exp}^0 are respectively the densities with respect to the Lebesgue measure of the invariant distributions $\mu_{\text{exp}}^\varepsilon$ and μ_{exp}^0 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

$$
\begin{aligned}
\mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[\psi_j(Z_0^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_0^\varepsilon}{\varepsilon} \right) \phi_j'(X_0^\varepsilon; a^*) \right] &= \int_\mathbb{R} \int_\mathbb{R} \psi_j(z; a^*) \frac{1}{\varepsilon} p' \left(\frac{x}{\varepsilon} \right) \phi_j'(x; a^*) \rho_{\text{exp}}^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z \\
&= -\sigma \int_\mathbb{R} \int_\mathbb{R} \frac{1}{C_{\varphi^\varepsilon}} \psi_j(z; a^*) \frac{\mathrm{d}}{\mathrm{d}x} \left(e^{-\frac{1}{\sigma} p\left(\frac{x}{\varepsilon}\right)} \right) \phi_j'(x; a^*) e^{-\frac{1}{\sigma} \alpha \cdot V(x)} \mathfrak{R}_{\text{exp}}^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z \\
&= \sigma \int_\mathbb{R} \int_\mathbb{R} \frac{1}{C_{\varphi^\varepsilon}} \psi_j(z; a^*) \frac{\partial}{\partial x} \left(\phi_j'(x; a^*) e^{-\frac{1}{\sigma} \alpha \cdot V(x)} \mathfrak{R}_{\text{exp}}^\varepsilon(x, z) \right) e^{-\frac{1}{\sigma} p\left(\frac{x}{\varepsilon}\right)} \, \mathrm{d}x \, \mathrm{d}z,
\end{aligned}
$$

which implies

$$
\begin{aligned}
\mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[\psi_j(Z_0^\varepsilon; a^*) \frac{1}{\varepsilon} p' \left(\frac{X_0^\varepsilon}{\varepsilon} \right) \phi_j'(X_0^\varepsilon; a^*) \right] &= \sigma \, \mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[\psi_j(Z_0^\varepsilon; a^*) \phi_j''(X_0^\varepsilon; a^*) \right] \\
&\quad - \mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[(\psi_j(Z_0^\varepsilon; a^*) \otimes V(X_0^\varepsilon)) \phi_j'(X_0^\varepsilon; a^*) \right] \alpha \\
&\quad + \sigma \int_\mathbb{R} \int_\mathbb{R} \psi_j(z; a^*) \phi_j'(x; a^*) \varphi^\varepsilon(x) \frac{\partial}{\partial x} \mathfrak{R}_{\text{exp}}^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z.
\end{aligned}
$$

Employing the last equation in the proof of Lemma 2.5 with $f(x, z) = \psi_j(z; a^*) \phi_j'(x; a^*)$ and $\delta = 1$ we have

$$
\sigma \int_\mathbb{R} \int_\mathbb{R} \psi_j(z; a^*) \phi_j'(x; a^*) \varphi^\varepsilon(x) \frac{\partial}{\partial x} \mathfrak{R}_{\text{exp}}^\varepsilon(x, z) \, \mathrm{d}x \, \mathrm{d}z = \mathbb{E}^{\mu_{\text{exp}}^\varepsilon} \left[\psi_j'(Z_0^\varepsilon; a^*) \phi_j(X_0^\varepsilon; a^*)(X_0^\varepsilon - Z_0^\varepsilon) \right], \tag{7.40}
$$

and thus we obtain

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

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

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

then applying formula (7.40) 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

$$\begin{aligned}\mathbb{E}^{\mu^0_{\mathrm{exp}}}\left[\psi_j'(Z_0^0;a^*)\phi_j(X_0^0;a^*)(X_0^0-Z_0^0)\right] &= \Sigma\int_{\mathbb{R}}\int_{\mathbb{R}}\psi_j(z;a^*)\phi_j'(x;a^*)\varphi^0(x)\frac{\partial}{\partial x}\mathfrak{R}^0_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z \\ &= -\Sigma\int_{\mathbb{R}}\int_{\mathbb{R}}\psi_j(z;a^*)\frac{\mathrm{d}}{\mathrm{d}x}\left(\phi_j'(x;a^*)\varphi^0(x)\right)\mathfrak{R}^0_{\mathrm{exp}}(x,z)\,\mathrm{d}x\,\mathrm{d}z \\ &= \mathbb{E}^{\mu^0_{\mathrm{exp}}}\left[\psi_j(Z_0^0;a^*)\left(\Sigma\phi_j''(X_0^0;a^*) - A\cdot V'(X_0^0)\phi_j'(X_0^0;a^*)\right)\right].\end{aligned}$$

Therefore, we obtain

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

which together with (7.32), (7.38) and bounds (7.37) and (7.39) 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. □

Remark 7.23. A similar result to Proposition 7.22 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

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

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

$$\begin{aligned}\widehat{\mathfrak{h}}_J^0(a) \coloneqq &\sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\dot\psi_j(X_0^0;a)\left(\mathcal{L}_A\phi_j(X_0^0;a) + \lambda_j(a)\phi_j(X_0^0;a)\right)\right] \\ &+ \sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\psi_j(X_0^0;a)\otimes\left(\mathcal{L}_A\dot\phi_j(X_0^0;a) + \lambda_j(a)\dot\phi_j(X_0^0;a)\right)\right] \\ &+ \mathbb{E}^{\nu^0}\left[\psi_j(X_0^0;a)\phi_j(X_0^0;a)\right]\otimes\dot\lambda_j(a),\end{aligned}$$

where the generator $\mathcal{L}_A$ is defined in (7.3). 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

$$\mathfrak{g}_J^0(a) \coloneqq \sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\psi_j(X_0^0;a)\left(\sigma\phi_j''(X_0^0;a) - \alpha\cdot V'(X_0^0)\phi_j'(X_0^0;a) + \lambda_j(a)\phi_j(X_0^0;a)\right)\right], \tag{7.41}$$

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

$$\begin{aligned}\mathfrak{h}_J^0(a) \coloneqq &\sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\dot\psi_j(X_0^0;a)\left(\sigma\phi_j''(X_0^0;a) - \alpha\cdot V'(X_0^0)\phi_j'(X_0^0;a) + \lambda_j(a)\phi_j(X_0^0;a)\right)\right] \\ &+ \sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\psi_j(X_0^0;a)\otimes\left(\sigma\dot\phi_j''(X_0^0;a) - \alpha\cdot V'(X_0^0)\dot\phi_j'(X_0^0;a) + \lambda_j(a)\dot\phi_j(X_0^0;a)\right)\right] \\ &+ \sum_{j=1}^J \mathbb{E}^{\nu^0}\left[\psi_j(X_0^0;a)\phi_j(X_0^0;a)\right]\otimes\dot\lambda_j(a).\end{aligned}$$

The proof is omitted since it is similar to the second case of the proof of Proposition 7.22.

7.4.2 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 M tends to infinity and the multiscale parameter ε vanishes. Therefore, we study the limit of the score functions and their Jacobian matrices as $M \to \infty$ 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 7.20 and Remark 7.21 evaluated in $a^* = A$. Then due to equation (7.6) we get

$$\begin{aligned} \frac{1}{\Delta} \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^0_{\mathrm{exp}}} & \left[g_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; A \right) \right] \\ &= \frac{1}{\Delta} \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^0_{\mathrm{exp}}} \left[\psi_j(\widetilde{Z}_0^0; A) \left(\phi_j(X_\Delta^0; A) - e^{-\lambda_j(A)\Delta} \phi_j(X_0^0; A) \right) \right] \\ &= \frac{1}{\Delta} \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^0_{\mathrm{exp}}} \left[\psi_j(\widetilde{Z}_0^0; A) \left(\mathbb{E} \left[\phi_j(X_\Delta^0; A) \middle| (X_0^0, \widetilde{Z}_0^0) \right] - e^{-\lambda_j(A)\Delta} \phi_j(X_0^0; A) \right) \right] \\ &= 0, \end{aligned} \tag{7.42}$$

and similarly we obtain

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

On the other hand, if Δ is a power of ε we study the limit of Proposition 7.22 and Remark 7.23 evaluated in $a^* = A$ and by (7.4) we have

$$\widetilde{\mathfrak{g}}_J^0(A) = 0 \qquad \text{and} \qquad \widehat{\mathfrak{g}}_J^0(A) = 0. \tag{7.43}$$

Moreover, differentiating equation (7.6) with respect to a, we get

$$\begin{aligned} \mathbb{E} \left[\dot{\phi}_j(X_{t_m}(a); a) | X_{t_{m-1}}(a) = x \right] &= e^{-\lambda_j(a)\Delta} \dot{\phi}_j(x; a) - \dot{\lambda}_j(a) \Delta e^{-\lambda_j(a)\Delta} \phi_j(x; a) \\ &\quad - \mathbb{E} \left[\phi_j'(X_{t_m}(a); a) \nabla_a X_{t_m}(a) | X_{t_{m-1}}(a) = x \right], \end{aligned} \tag{7.44}$$

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 (7.6) and (7.44) we have

$$\frac{1}{\Delta} \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^0_{\mathrm{exp}}} \left[h_j \left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; A \right) \right] = - \sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}^0_{\mathrm{exp}}} \left[\left(\psi_j(\widetilde{Z}_0^0; A) \otimes \nabla_a X_\Delta(A) \right) \phi_j'(X_\Delta^0; A) \right], \tag{7.45}$$

and

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

Then due to Lemma 7.30 we can differentiate the eigenvalue problem (7.5) 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 (7.5) implies

$$\widetilde{\mathfrak{h}}_J^0(A) = \sum_{j=1}^J \mathbb{E}^{\mu_{\exp}^0}\left[(\psi_j(Z_0^0;A) \otimes V'(X_0^0))\phi_j'(X_0^0;A)\right], \tag{7.46}$$

and

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

Before showing the main results, we need two auxiliary lemmas, which in turn rely on the technical Assumption 7.6, which can now be rewritten as

(i) $\det\left(\frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\widetilde{\mu}_{\exp}^0}\left[h_j\left(X_0^0, X_\Delta^0, \widetilde{Z}_0^0; A\right)\right]\right) \neq 0,$

(ii) $\det\left(\frac{1}{\Delta}\sum_{j=1}^J \mathbb{E}^{\nu^0}\left[h_j\left(X_0^0, X_\Delta^0, X_0^0; A\right)\right]\right) \neq 0,$

(iii) $\det\left(\widetilde{\mathfrak{h}}_J^0(A)\right) \neq 0,$

(iv) $\det\left(\widehat{\mathfrak{h}}_J^0(A)\right) \neq 0.$

Since the proofs of the two lemmas are similar we only write the details of the first one.

Lemma 7.24. *Under Assumption 7.2 and Assumption 7.6 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 $\zeta \neq 1$, $\zeta \neq 2$*

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

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 7.20 and Proposition 7.22 depending on Δ and note that due to (7.42) if Δ is independent of ε and (7.43) otherwise, we have

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

Moreover, by (7.45), (7.46) and Assumption 7.6, 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 7.18, the implicit function theorem (see [67, Theorem 2]) gives the desired result. □

Lemma 7.25. *Under Assumption 7.2 and Assumption 7.6 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 $\zeta \in (0, 1)$*

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

Moreover

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

We are now ready to prove the asymptotic unbiasedess of the estimators, i.e., Theorem 7.8 and Theorem 7.9. We only prove Theorem 7.9 for the estimator $\widetilde{A}^J_{M,\varepsilon}$ with filtered data. The proof of Theorem 7.8 for the estimator $\widehat{A}^J_{M,\varepsilon}$ without filtered data is analogous and is omitted here.

Proof of Theorem 7.9. We need to show for a fixed $0 < \varepsilon < \varepsilon_0$:

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

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

We first note that by Lemma 7.24 we have

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

We then follow the steps of the proof of [21, Theorem 3.2]. Due to [16, Theorem A.1], claims (i) and (ii) hold true if we verify that

$$\lim_{M\to\infty} \sup_{a\in B^\varepsilon_{C,M}} \left\| \frac{1}{M} \dot{\widetilde{G}}^J_{M,\varepsilon}(a) - \widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right\| = 0, \qquad \text{in probability}, \tag{7.47}$$

and as $M \to \infty$

$$\frac{1}{\sqrt{N}} \widetilde{G}^J_{M,\varepsilon}(A + \widetilde{\gamma}(\varepsilon)) \to \mathcal{N}\left(0, \Lambda^\varepsilon\right), \qquad \text{in law}, \tag{7.48}$$

where Λ^ε is a positive definite covariance matrix and

$$B^\varepsilon_{C,M} = \left\{ a \in \mathcal{A} \colon\ \|a - (A + \widetilde{\gamma}(\varepsilon))\| \le \frac{C}{\sqrt{M}} \right\},$$

for $C > 0$ small enough such that $B_{C,1} \subset \mathcal{A}$. Result (7.48) is a consequence of [45, Theorem 1]. We then have

$$\begin{aligned}
&\sup_{a\in B^\varepsilon_{C,M}} \left\| \frac{1}{M} \dot{\widetilde{G}}^\varepsilon_{M,J}(a) - \widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right\| \\
&\qquad\qquad \le \sup_{a\in B^\varepsilon_{C,1}} \left\| \frac{1}{M\Delta} \sum_{m=0}^{M-1} \sum_{j=1}^{J} h_j(\widetilde{X}^\varepsilon_m, \widetilde{X}^\varepsilon_{m+1}, \widetilde{Z}^\varepsilon_m; a) - \widetilde{\mathcal{H}}_J(\varepsilon, a) \right\| \\
&\qquad\qquad\quad + \sup_{a\in B^\varepsilon_{C,M}} \left\| \widetilde{\mathcal{H}}_J(\varepsilon, a) - \widetilde{\mathcal{H}}_J(\varepsilon, A + \widetilde{\gamma}(\varepsilon)) \right\|,
\end{aligned}$$

where the right-hand side vanishes by [21, Lemma 3.3] and the continuity of $\widetilde{\mathcal{H}}$ (Proposition 7.18), implying result (7.47). Hence, we proved (i) and (ii), which conclude the proof of the theorem. $\square$

Remark 7.26. Notice that if $\Delta = \varepsilon^\zeta$ with $\zeta > 2$ and we do not employ the filter, in view of (7.41) and following the same proof of Theorem 7.9, we could compute the asymptotic limit of $\widehat{A}^J_{M,\varepsilon}$ as M goes to infinity and ε vanishes if we knew a^* such that

$$\sum_{j=1}^{J} \mathbb{E}^{\nu^0} \left[\psi_j(X^0_0; a^*) \left(\sigma \phi_j''(X^0_0; a^*) - \alpha \cdot V'(X^0_0) \phi_j'(X^0_0; a^*) + \lambda_j(a^*) \phi_j(X^0_0; a^*) \right) \right] = 0.$$

The value of a^* cannot be found analytically since it is, in general, different from the drift coefficients α and A of the multiscale and homogenized equations (1.10) and (1.11). Nevertheless, we observe that in the simple scale of the multiscale OU process we have $a^* = \alpha$.

7.5 Technical results

In this section we prove some technical results which are used to show the unbiasedness of the estimators $\widehat{A}^J_{M,\varepsilon}$ and $\widetilde{A}^J_{M,\varepsilon}$. 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.

7.5.1 Application of the filter to discrete data

The following result quantifies the expected distance among the continuous process Z^ε_t and the filtered observations $\widetilde{Z}^\varepsilon_m$.

Lemma 7.27. *Let $0 < \Delta < 1$, M be a positive integer and let $\widetilde{Z}^\varepsilon_m$ and Z^ε_t be defined respectively in (7.10) and (7.14) with $\widetilde{X}^\varepsilon_0 = X^\varepsilon_0$ distributed according to its invariant measure ν^ε. Then there exists a constant $C > 0$ independent of ε, Δ and M such that for all $m = 0, \dots, M$ and for all $p \geq 1$*

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

where $\mathbb{E}^{\nu^\varepsilon}$ denotes the expectation with respect to the Wiener measure and the fact that X^ε_0 is distributed according to ν^ε.

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

$$E := \mathbb{E}^{\nu^\varepsilon} \left| Z^\varepsilon_{m\Delta} - \widetilde{Z}^\varepsilon_m \right|^p,$$

which is equivalent to

$$E = \mathbb{E}^{\nu^\varepsilon} \left| \sum_{k=0}^{m-1} \int_{k\Delta}^{(k+1)\Delta} \left(e^{-(m\Delta - s)} X^\varepsilon_s - e^{-\Delta(m-k)} \widetilde{X}^\varepsilon_k \right) \mathrm{d}s \right|^p.$$

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

$$\begin{aligned} E &\leq 2^{p-1}\, \mathbb{E}^{\nu^\varepsilon} \left(\sum_{k=0}^{m-1} \int_{k\Delta}^{(k+1)\Delta} e^{-(m\Delta - s)} \left| X^\varepsilon_s - X^\varepsilon_{k\Delta} \right| \mathrm{d}s \right)^p \\ &\quad + 2^{p-1}\, \mathbb{E}^{\nu^\varepsilon} \left(\sum_{k=0}^{m-1} \int_{k\Delta}^{(k+1)\Delta} \left(e^{-(m\Delta - s)} - e^{-\Delta(m-k)} \right) \mathrm{d}s \left| \widetilde{X}^\varepsilon_k \right| \right)^p \\ &=: 2^{p-1} \left(E_1 + E_2 \right). \end{aligned} \tag{7.49}$$

We now study the two terms separately. Applying Lemma 2.26 we first get

$$\begin{aligned} E_1 &= \mathbb{E}^{\nu^\varepsilon} \left(\int_0^{m\Delta} e^{-(m\Delta - s)} \left| X^\varepsilon_s - \sum_{k=0}^{m-1} X^\varepsilon_{k\Delta} \chi_{[k\Delta,(k+1)\Delta)}(s) \right| \mathrm{d}s \right)^p \\ &\leq \int_0^{m\Delta} e^{-(m\Delta - s)}\, \mathbb{E}^{\nu^\varepsilon} \left| X^\varepsilon_s - \sum_{k=0}^{m-1} X^\varepsilon_{k\Delta} \chi_{[k\Delta,(k+1)\Delta)}(s) \right|^p \mathrm{d}s, \end{aligned} \tag{7.50}$$

and, in order to bound the term inside the integral, we can follow two different procedures. Either we employ [103, Lemma 6.1], which gives

$$\mathbb{E}^{\nu^\varepsilon} \left| X^\varepsilon_s - \sum_{k=0}^{m-1} X^\varepsilon_{k\Delta} \chi_{[k\Delta,(k+1)\Delta)}(s) \right|^p \leq C \left(\Delta^p + \Delta^{p/2} + \varepsilon^p \right), \tag{7.51}$$

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 [103, Corollary 5.4] and p is bounded, it holds for all $s \in [k\Delta, (k+1)\Delta)$

$$\begin{aligned}\mathbb{E}^{\nu^\varepsilon} \left| X_s^\varepsilon - X_{k\Delta}^\varepsilon \right|^p &= \mathbb{E}^{\nu^\varepsilon} \left| -\alpha \int_{k\Delta}^s X_r^\varepsilon \,\mathrm{d}r - \frac{1}{\varepsilon} \int_{k\Delta}^s p'\left(\frac{X_r^\varepsilon}{\varepsilon}\right) \mathrm{d}r + \sqrt{2\sigma} W_s \right|^p \\ &\leq C\left(\Delta^p + \Delta^p \varepsilon^{-p} + \Delta^{p/2}\right).\end{aligned} \tag{7.52}$$

Therefore, due to (7.50), (7.51) and (7.52), we obtain

$$E_1 \leq C\left(\Delta^{p/2} + \min\left\{\varepsilon^p, \Delta^p \varepsilon^{-p}\right\}\right). \tag{7.53}$$

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

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

and note that

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

Therefore, applying Jensen's inequality and due to the fact that $\widetilde{X}_k^\varepsilon$ has bounded moments of any order by [103, Corollary 5.4] we have

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

which, together with (7.49) and (7.53), gives the desired result. □

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

Lemma 7.28. *Let $\Delta > 0$ and let Assumption 1.4 hold. Then the couple $(\widetilde{X}_m^\varepsilon, \widetilde{Z}_m^\varepsilon)$, where $\widetilde{X}_m^\varepsilon$ is a sample from the continuous process* (1.10) *and $\widetilde{Z}_m^\varepsilon$ is defined in* (7.10)*, admits a unique invariant measure $\widetilde{\mu}_{\mathrm{exp}}^\varepsilon$ with density with respect to the Lebesgue measure denoted by $\widetilde{\rho}_{\mathrm{exp}}^\varepsilon = \widetilde{\rho}_{\mathrm{exp}}^\varepsilon(x,z)$. Moreover, if Δ is independent of ε, it converges in law to the two-dimensional process $(\widetilde{X}_m^0, \widetilde{Z}_m^0)$ with $\widetilde{\rho}_{\mathrm{exp}}^0 = \widetilde{\rho}_{\mathrm{exp}}^0(x,z)$ as density of the invariant measure $\widetilde{\mu}_{\mathrm{exp}}^0$.*

Proof. By definition (7.13) we obtain the following SDE

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

where $\widetilde{X}_m^\varepsilon$ is a stationary and ergodic sequence. Observing that $\log e^{-\Delta} = -\Delta < 0$, applying Theorem 1 and in view of Remark 1.3 in [25] we deduce the existence of a unique invariant measure for the couple $(\widetilde{X}_m^\varepsilon, \widetilde{Z}_m^\varepsilon)$. Let us notice that in the theorem the sequence $\widetilde{X}_m^\varepsilon$ must be defined for all $m \in \mathbb{Z}$ while in our framework $m \in \mathbb{N}$, 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 $(\widetilde{X}_m^0, \widetilde{Z}_m^0)$. Finally, standard homogenization theory implies the weak convergence of $\widetilde{\rho}_{\mathrm{exp}}^\varepsilon$ to $\widetilde{\rho}_{\mathrm{exp}}^0$, which concludes the proof. □

Corollary 7.29. *Let Z^ε and $\widetilde{Z}^\varepsilon$ be at stationarity. 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 7.27 by letting m go to infinity, noting that the constant C is independent of n and employing ergodicity given by Lemma 7.28. □

It directly follows that $\widetilde{Z}_m^\varepsilon$ has bounded moments of all order and, in particular, we can prove Lemma 7.16.

Proof of Lemma 7.16. Applying Jensen's inequality to the function $x \mapsto |x|^p$, we have

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

then bounding the two terms in the right-hand side respectively with Lemma 2.28 and Corollary 7.29 gives the desired result. □

7.5.2 Properties of eigenfunctions and eigenvalues of the generator

Let us now consider the eigenvalue and the eigenfunctions of the generator of SDE (7.1).

Lemma 7.30. *Let $\{(\lambda_j(a), \phi_j(\cdot; a))\}_{j=0}^\infty$ be the solutions of the eigenvalue problem (7.4). 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 [111]. Let us remark that the fact that the spectrum is discrete and non-degenerate is guaranteed by [101, Section 4.7]. Finally, the second result in the statement is a direct consequence of the elliptic regularity theory. □

7.5.3 Approximation of the Itô formula

In this section we prove Lemma 7.17, 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 $S_0^\varepsilon = X_0^\varepsilon$

$$\mathrm{d}S_t^\varepsilon = -\alpha V'(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))\mathrm{d}t + \sqrt{2\sigma}(1 + \Phi'(Y_t^\varepsilon))\mathrm{d}W_t, \tag{7.54}$$

where $Y_t^\varepsilon = X_t^\varepsilon/\varepsilon$ and Φ is the cell function which solves equation (1.7), 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 [103] we have

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

and, since Φ is bounded by [103, Lemma 5.5], we get for a constant $C > 0$ independent of Δ and ε

$$\left|X_\Delta^\varepsilon - S_\Delta^\varepsilon\right| \leq C\varepsilon. \tag{7.55}$$

Before showing the main formula, we need two preliminary estimates which will be employed later in the analysis. The proofs of Lemma 7.31 and Lemma 7.32 are inspired by the proof of Proposition 5.8 in [103].

Lemma 7.31. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that f, f' are polynomially bounded. Then*

$$\int_0^\Delta \alpha \cdot V'(X_t^\varepsilon) f(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \,\mathrm{d}t = A \cdot V'(X_0^\varepsilon) f(X_0^\varepsilon)\Delta + R_1(\varepsilon, \Delta), \tag{7.56}$$

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

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

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 (7.56). Applying Jensen's inequality to the function $z \mapsto |z|^p$ we have

$$\begin{aligned}
\mathbb{E}^{\nu^\varepsilon} |R_1(\varepsilon, \Delta)|^p \leq 3^{p-1}\, \mathbb{E}^{\nu^\varepsilon} & \left| \int_0^\Delta \alpha \cdot (V'(X_t^\varepsilon) - V'(X_0^\varepsilon))\, f(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon)) \,\mathrm{d}t \right|^p \\
& + 3^{p-1}\, \mathbb{E}^{\nu^\varepsilon} \left| \alpha \cdot V'(X_0^\varepsilon) \int_0^\Delta (f(X_t^\varepsilon) - f(X_0^\varepsilon))\,(1 + \Phi'(Y_t^\varepsilon)) \,\mathrm{d}t \right|^p \\
& + 3^{p-1}\, \mathbb{E}^{\nu^\varepsilon} \left| f(X_0^\varepsilon) V'(X_0^\varepsilon) \cdot \int_0^\Delta (\alpha(1 + \Phi'(Y_t^\varepsilon)) - A) \,\mathrm{d}t \right|^p \\
& =: I_1(\varepsilon, \Delta) + I_2(\varepsilon, \Delta) + I_3(\varepsilon, \Delta).
\end{aligned} \tag{7.57}$$

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 [103, Corollary 5.4], we have

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

then applying [103, Lemma 6.1] we obtain

$$I_1(\varepsilon, \Delta) \leq C\left(\Delta^{2p} + \Delta^{3p/2} + \varepsilon^p \Delta^p\right). \tag{7.58}$$

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

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

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

$$I_2(\varepsilon, \Delta) \leq C\left(\Delta^{2p} + \Delta^{3p/2} + \varepsilon^p \Delta^p\right). \tag{7.59}$$

We now consider the function

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

which by definition of A and due to (1.6) has zero mean with respect to π defined in (1.8). Therefore, since f and V' are polynomially bounded and X_0^ε has bounded moments of any order by [103, Corollary 5.4], applying [103, Lemma 5.6] we obtain

$$I_3(\varepsilon, \Delta) \leq C \left(\varepsilon^{2p} + \varepsilon^p \Delta^p + \varepsilon^p \Delta^{p/2} \right). \tag{7.60}$$

Finally, for ε and Δ sufficiently small, the desired result follows from (7.57) and from estimates (7.58), (7.59) and (7.60). □

Lemma 7.32. *Let $f \colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that f, f' are polynomially bounded. Then*

$$\int_0^\Delta \sigma f(X_t^\varepsilon)(1 + \Phi'(Y_t^\varepsilon))^2 \,\mathrm{d}t = \Sigma f(X_0^\varepsilon)\Delta + R_2(\varepsilon, \Delta), \tag{7.61}$$

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

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

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 (7.61). Applying Jensen's inequality to the function $z \mapsto |z|^p$ we have

$$\begin{aligned} \mathbb{E}^{\nu^\varepsilon} \left| R_2(\varepsilon, \Delta) \right|^p &\leq 2^{p-1} \, \mathbb{E}^{\nu^\varepsilon} \left| \int_0^\Delta \sigma \left(f(X_t^\varepsilon) - f(X_0^\varepsilon) \right) (1 + \Phi'(Y_t^\varepsilon)^2) \,\mathrm{d}t \right|^p \\ &\quad + 2^{p-1} \, \mathbb{E}^{\nu^\varepsilon} \left| f(X_0^\varepsilon) \int_0^\Delta \left(\sigma(1 + \Phi'(Y_t^\varepsilon))^2 - \Sigma \right) \,\mathrm{d}t \right|^p \\ &=: I_1(\varepsilon, \Delta) + I_2(\varepsilon, \Delta). \end{aligned} \tag{7.62}$$

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

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

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 [103, Corollary 5.4], we have

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

and applying [103, Lemma 6.1] we obtain

$$I_1(\varepsilon, \Delta) \leq C \left(\Delta^{2p} + \Delta^{3p/2} + \varepsilon^p \Delta^p \right). \tag{7.63}$$

We now consider the function

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

which by definition of Σ and due to (1.6) has zero mean with respect to μ defined in (1.8). Therefore, since f is polynomially bounded and X_0^ε has bounded moments of any order by [103, Corollary 5.4], applying [103, Lemma 5.6] we obtain

$$I_2(\varepsilon, \Delta) \leq C \left(\varepsilon^{2p} + \varepsilon^p \Delta^p + \varepsilon^p \Delta^{p/2} \right). \tag{7.64}$$

Finally, for ε and Δ sufficiently small, the desired result follows from (7.62) and from estimates (7.63) and (7.64). □

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 7.17. Applying Itô's lemma to the process S_t^ε defined in (7.54) 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 7.31 and Lemma 7.32 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 (7.55), the hypotheses on f and the fact that X_t^ε has bounded moments of any order by [103, Corollary 5.4], we obtain

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

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. ☐

7.6 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 (7.5)

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

and the solution of the non-linear system (7.9) or (7.12) 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 (7.5), i.e., we multiply it by $v\varphi_a$, where v is a test function and φ_a is the invariant distribution defined in (7.2), 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 $\epsilon > 0$ there exists $r > 0$ such that

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

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

$$\Sigma \int_{-R}^{+R} \phi_j'(x; a) v'(x) \varphi_a(x) \,\mathrm{d}x = \lambda_j(a) \int_{-R}^{+R} \phi_j(x; a) v(x) \varphi_a(x) \,\mathrm{d}x. \tag{7.65}$$

Notice that R must be chosen big enough and such that

$$R \geq \max_{m=0,\dots,M} \max \left\{ \left| \widetilde{X}_m^\varepsilon \right|, \left| \widetilde{Z}_m^\varepsilon \right| \right\} =: \bar{R},$$

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

$$\phi_j(R; a) > 0 \qquad \text{and} \qquad \int_{-R}^{+R} \phi_j(x; a)^2 \varphi_a(x) \,\mathrm{d}x = 1. \tag{7.66}$$

We then introduce a partition $\mathcal{P}_h$ 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

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

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 $\{\beta_k\}_{k=0}^{N_h}$ of X_h^1 characterized by the following property

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

where δ_{ik} is the Kronecker delta. We want to find $\phi_j(\cdot; a) \in X_h^1$ such that equation (7.65) holds true for all $v \in X_h^1$. Therefore, in equation (7.65) we substitute

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

and we obtain the discrete formulation

$$S\Theta_j(a) = \lambda_j(a) M \Theta_j(a), \tag{7.67}$$

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

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

where the integrals are approximated through the composite Simpson's quadrature rule. Equation (7.67) 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 (7.66), 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 $\widehat{G}_{M,\varepsilon}^J(a)$ in (7.8) or $\widetilde{G}_{M,\varepsilon}^J(a)$ in (7.11) with filtered data. Hence, it only remains to solve systems (7.9) or (7.12), i.e.,

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

To solve these equations we can follow two approaches:

(i) find the zero of $\widehat{G}^J_{M,\varepsilon}(a)$ or $\widetilde{G}^J_{M,\varepsilon}(a)$;

(ii) find the minimum of $\left\|\widehat{G}^J_{M,\varepsilon}(a)\right\|$ or $\left\|\widetilde{G}^J_{M,\varepsilon}(a)\right\|$.

In practice, for the first approach we can use the function `fsolve` in MATLAB or the function `scipy.optimize.fsolve` in PYTHON, 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.

7.7 Multidimensional diffusion processes

In this section we present how our methodology for estimating the drift coefficient of the homogenized equation can be extended to the case of the d-dimensional multiscale diffusion process (1.4). Using the tensor notation, we can then define the drift coefficient $A \in \mathbb{R}^{L\times d\times d}$, which collects together the L matrices A_ℓ for $\ell = 1, \dots, L$ of the homogenized equation (1.5). Our goal is now to estimate the tensor A and thus we need to define the score functions. First, the d-dimensional eigenvalue problem for $j = 1, \dots, J$ corresponding to (7.5) is

$$\Sigma : \nabla^2 \phi_j(x;a) - \left(\sum_{\ell=1}^L a_\ell \nabla V_\ell(x) \right) \cdot \nabla \phi_j(x;a) + \lambda_j(a)\phi_j(x;a) = 0,$$

where : denotes the Frobenius inner product, ∇^2 the Hessian matrix and the parameter $a \in \mathbb{R}^{L\times d\times d}$ collects together the L matrices a_ℓ for $\ell = 1, \dots, L$. Then, in order to define the martingale estimating functions g_j for $j = 1, \dots, J$, we take a collection $\{\psi_j\}_{j=1}^J$ of functions $\psi_j(\cdot;a)\colon \mathbb{R}^d \to \mathbb{R}^{L\times d\times d}$ and we use equation (7.7). Finally, we construct the score functions $\widehat{G}^J_{M,\varepsilon}$ and $\widetilde{G}^J_{M,\varepsilon}$ in the same way as we did in the one dimensional case, i.e., employing equations (7.8) and (7.11). We remark that the filtered data are obtained as in equation (7.10) by applying the filter component-wise. We can now compute the estimators $\widehat{A}^J_{M,\varepsilon}$ and $\widetilde{A}^J_{M,\varepsilon}$ by solving the nonlinear systems

$$\widehat{G}^J_{M,\varepsilon}(a) = 0 \qquad \text{and} \qquad \widetilde{G}^J_{M,\varepsilon}(a) = 0,$$

which have dimension Ld^2. From a theoretical point of view, slight modifications of the proofs allow to conclude that analogous results to the main theorems hold true, i.e., that the estimators are asymptotically unbiased in the limit of infinite observations and when the multiscale parameter vanishes. However, the problem becomes more complex and computationally expensive from a numerical viewpoint, in particular when the dimension d is large. In fact, the final nonlinear system, which has to be solved, has dimension Ld^2 instead of L and, most importantly, it is required to solve the eigenvalue problem for the generator of a diffusion process in d dimensions.

7.8 Conclusion

In this chapter 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 7.8 and Theorem 7.9 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.

8 Eigenfunction estimators for interacting particle systems

In this chapter, which is based on our research article [105], we apply eigenfunction estimators to infer the parameters of the mean field limit of a system of interacting particles given discrete observations of one single particle. The chapter is organized as follows. In Section 8.1 we construct the proposed estimator and in Section 8.2 we present the main theoretical results. Then, in Section 8.3 we show several numerical experiments illustrating the potentiality of our approach and in Section 8.4 we present the proofs of the main results. Finally, in Section 8.5 we draw our conclusions.

8.1 Parameter estimation problem

We want to construct martingale estimating functions based on the eigenfunctions and the eigenvalues of the generator of the dynamics, the same technique which was initially proposed in [73] for single-scale stochastic differential equations (SDEs) and which we applied to multiscale SDEs in Chapter 7. In principle, the methodology developed in [73] could be applied to the N-particle system. However, this would require solving the eigenvalue problem for the generator of an N-dimensional diffusion process, which is computationally expensive. Moreover, our fundamental assumption is that we are observing a single particle and thus we do not have a complete knowledge of the system. Therefore, we construct the martingale estimating functions employing the generator of the mean field dynamics, which is a good approximation of the path of a single particle when the number N of particles is large [117]. Let $\mathcal{L}_t$ be the generator of the mean field limit SDE (5.10)

$$\mathcal{L}_t = -\left(\mathcal{V}'(\cdot;\alpha) + (\mathcal{F}'(\cdot;\kappa) * u(\cdot,t;\theta))\right)\frac{\mathrm{d}}{\mathrm{d}x} + \sigma\frac{\mathrm{d}^2}{\mathrm{d}x^2},$$

and let $\mathcal{L}$ be the generator obtained replacing the density $u(\cdot,t;\theta)$ with the density $\rho(\cdot;\theta)$ of the invariant measure μ_θ

$$\mathcal{L} = -\left(\mathcal{V}'(\cdot;\alpha) + (\mathcal{F}'(\cdot;\kappa) * \rho(\cdot;\theta))\right)\frac{\mathrm{d}}{\mathrm{d}x} + \sigma\frac{\mathrm{d}^2}{\mathrm{d}x^2}.$$

We remark that now the generator $\mathcal{L}$ is time-independent. We then consider the eigenvalue problem $-\mathcal{L}\phi(\cdot;\theta) = \lambda(\theta)\phi(\cdot;\theta)$, which reads

$$\sigma\phi''(x;\theta) - \left(\mathcal{V}'(x;\alpha) + (\mathcal{F}'(\cdot;\kappa) * \rho(\cdot;\theta))(x)\right)\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0, \tag{8.1}$$

and from the well-known spectral theory of diffusion processes (see, e.g., [63]) we deduce the existence of a countable set of eigenvalues $0 = \lambda_0(\theta) < \lambda_1(\theta) < \cdots < \lambda_j(\theta) \uparrow \infty$ whose corresponding eigenfunctions $\{\phi_j(\cdot;\theta)\}_{j=0}^\infty$ form an orthonormal basis of the weighted space $L^2_{\rho(\cdot;\theta)}(\mathbb{R})$. In fact, even if the SDE (5.10) is nonlinear, when $X_0 \sim \mu_\theta$ then the solution X_t behaves like a classic diffusion process with drift function $-\mathcal{V}'(\cdot;\alpha) - \mathcal{F}'(\cdot;\kappa) * \rho(\cdot;\theta)$, hence the spectral theory for diffusion processes still holds. We also state here the variational formulation of the eigenvalue problem, which will be employed to implement numerically the proposed methodology. Let v be a test function and multiply equation (8.1) by $v\rho(\cdot;\theta)$, where the density $\rho(\cdot;\theta)$ of the invariant measure μ_θ is defined in (5.11). Then, integrating over $\mathbb{R}$ and by parts we obtain

$$\sigma \int_{\mathbb{R}} \phi'(x;\theta)v'(x)\rho(x;\theta)\,\mathrm{d}x = \lambda(\theta)\int_{\mathbb{R}} \phi(x;\theta)v(x)\rho(x;\theta)\,\mathrm{d}x.$$

We are now ready to present how to employ the eigenvalue problem in the construction of the martingale estimation function and afterwords in the definition of our estimator. Let J be a positive integer and let $\psi_j(\cdot;\theta)\colon \mathbb{R}\to\mathbb{R}^L$ for $j=1,\dots,J$ be arbitrary functions dependent on the parameter θ which satisfy Assumption 8.2 below, and define the martingale estimating function $G^J_{M,N}\colon \Theta\to\mathbb{R}^L$ as

$$G^J_{M,N}(\theta) \coloneqq \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J} \bar g_j(\widetilde X^{(n)}_m, \widetilde X^{(n)}_{m+1};\theta),$$

where

$$\bar g_j(x,y;\theta) \coloneqq \psi_j(x;\theta)\left(\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)\right), \tag{8.2}$$

and $\{\widetilde X^{(n)}_m\}_{m=0}^M$ is the set of observations of the n-th particle from the system with N particles. The estimator we propose is then given by the solution $\widehat\theta^J_{M,N}$ of the p-dimensional nonlinear system

$$G^J_{M,N}(\theta) = 0, \tag{8.3}$$

where in this case 0 denotes the vector with all components equal to zero. An intuition on why considering the solution of equation (2.8) as a good estimator is the following and will be more clear later. Let $\mathbb{G}^J_M$ defined in (8.19) be the estimating function where the observations from the inetracting particle system have been replaced by the observations from the corresponding mean field limit. Then, employing formula (8.18) we have

$$\mathbb{E}^{\mu_{\theta_0}}\left[\mathbb{G}^J_M(\theta_0)\right] = 0,$$

which means that the zero of the expectation of the estimating function with observations from the mean field limit is exactly the true unknown coefficient. The main steps needed to obtain the estimator $\widehat\theta^J_{M,N}$ are summarized in Algorithm 3. For further details about the implementation and for discussions about the choice of the arbitrary functions $\{\psi_j(\cdot;\theta)\}_{j=1}^J$ we refer to Section 7.6 and Remark 7.3.

Remark 8.1. The main limitation of our approach is that the knowledge of the invariant measure is required in order to construct the martingale estimating function (step 1 in Algorithm 3). However, it is often the case that the invariant measure is known up to a set of parameters, such as moments, i.e., only the functional form of the invariant measure is known. These parameters (moments) are obtained by solving appropriate self-consistency equations [46, Section 2.3]. When such a situation arises, it is possible to first learn these parameters using the available data, e.g., estimate the moments that appear in the invariant measure by employing the law of large numbers. Then, we are in the setting where our technique applies and we can proceed in the same way, as shown in the numerical experiments in Sections 8.3.5 and 8.3.6. In summary, it is sufficient to replace step 1 in Algorithm 3 with "estimate the moments in the invariant measure $\rho(\cdot;\theta)$".

We finally introduce a technical hypothesis which will be needed for the proofs of our main results.

Assumption 8.2. Let $\Theta\subseteq\mathbb{R}^L$ be a compact set. Then the following hold for all $\theta\in\Theta$ and for all $j=1,\dots,J$:

(i) $\psi_j(x;\theta)$ is continuously differentiable with respect to θ for all $x\in\mathbb{R}$;

(ii) all components of $\psi_j(\cdot;\theta)$, $\psi_j'(\cdot;\theta)$, $\dot\psi_j(\cdot;\theta)$, $\dot\psi_j'(\cdot;\theta)$ are polynomially bounded uniformly in θ;

(iii) the potentials $\mathcal{V}$ and $\mathcal{F}$ are such that $\phi_j(\cdot;\theta)$, $\phi_j'(\cdot;\theta)$ and all components of $\dot\phi_j(\cdot;\theta)$, $\dot\phi_j'(\cdot;\theta)$ are polynomially bounded uniformly in θ;

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

Remark 8.3. Assumption 8.2(i) together with [111, Sections 2 and 6] gives the continuous differentiability of the vector-valued function $G_{M,N}^J(\theta)$ with respect to the unknown parameter θ.

Algorithm 3: Estimation of $\theta \in \Theta$

Input: Observations $\{\widetilde X_m^{(n)}\}_{m=0}^M$.
Distance between two consecutive observations Δ.
Number of eigenvalues and eigenfunctions J.
Functions $\{\psi_j(x;\theta)\}_{j=1}^J$.
Confining potential $\mathcal{V}$ and interaction potential $\mathcal{F}$.
Diffusion coefficient σ.

Output: Estimation $\widehat\theta_{M,N}^J$ of θ.

1: Find the invariant measure $\rho(\cdot;\theta)$.

2: Consider the equation
$\sigma\phi''(x;\theta) - (\mathcal{V}'(x;\alpha) + (\mathcal{F}'(\cdot;\kappa) * \rho(\cdot;\theta))(x))\,\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0.$

3: Compute the first J eigenvalues $\{\lambda_j(\theta)\}_{j=1}^J$ and eigenfunctions $\{\phi_j(\cdot;\theta)\}_{j=1}^J$.

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

5: Construct the score function $G_{M,N}^J(\theta) = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^J \bar g_j(\widetilde X_m^{(n)}, \widetilde X_{m+1}^{(n)};\theta)$.

6: Let $\widehat\theta_{M,N}^J$ be the solution of the nonlinear system $G_{M,N}^J(\theta) = 0$.

Remark 8.4. In this chapter we always assume that the diffusion coefficient σ in (5.9) is known. We remark that this is not an essential limitation of our methodology; in fact, if the diffusion coefficient is also unknown, we can consider the parameter set to be estimated to be $\widetilde\theta = (\theta,\sigma) = (\alpha,\kappa,\sigma) \in \mathbb{R}^{L+1}$ and repeat the same procedure. The estimator is then obtained as the solution of the nonlinear system of dimension $L+1$ corresponding to (8.3). A numerical experiment illustrating this procedure is given in Section 8.3.3. Moreover, our main theoretical results remain valid and the proofs do not need any major changes. Alternatively, if the sampling rate is sufficiently small, it is possible to first estimate the diffusion coefficient using the quadratic variation and then proceed with the methodology proposed in this chapter.

Example 8.5. Let us consider the Curie–Weiss quadratic interaction introduced in Example 5.4 as well as a quadratic Ornstein–Uhlenbeck (OU) confining potential $V(x;\alpha) = \frac{1}{2}x^2$. In this case the only unknown parameter is κ and the eigenvalue problem (8.1) reads

$$\sigma\phi''(x;\theta) - (1+\kappa)x\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0, \tag{8.4}$$

so that the eigenvalue and eigenfunctions can be computed analytically (see Section 7.2.1). In particular, the first eigenvalue and eigenfunction are given by $\lambda_1(\theta) = 1+\kappa$ and $\phi_1(x;\theta) = x$, respectively. Therefore, letting $\psi_1(x;\theta) = x$ we have an explicit expression for our estimator

$$\widehat{\theta}^1_{M,N} = -1 - \frac{1}{\Delta} \log \left(\frac{\sum_{m=0}^{M-1} \widetilde{X}_m^{(n)} \widetilde{X}_{m+1}^{(n)}}{\sum_{m=0}^{M-1} (\widetilde{X}_m^{(n)})^2} \right). \tag{8.5}$$

For additional details regarding the eigenvalue problem (8.4) we refer to Section 7.2.1. We also remark that when the drift coefficient of the OU process is unknown, i.e., if we consider the confining potential $V(x;\alpha) = \frac{\alpha}{2}x^2$, then the eigenvalue problem reads

$$\sigma\phi''(x;\theta) - (\alpha+\kappa)x\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0,$$

which only depends on the sum $\alpha + \kappa$ and not on the single parameters alone. Therefore, in this case it is not possible to estimate the unknown coefficients α and κ, but we can only estimate their sum. This is in contrast with the set up in [70], where *all* the particles are observed in continuous time. When this amount of information is available, it is possible to check whether or not the particles are interacting, i.e., to check whether $\kappa = 0$ or not (see [70, Section 4]).

8.2 Main results

In this section we present the main theoretical results of this chapter. In particular, we prove that our estimator $\widehat{\theta}^J_{M,N}$ is asymptotically unbiased (consistent) and asymptotically normal as the number of observations M and particles N go to infinity and we compute the rate of convergence towards the true value of the parameter, which we denote by θ_0. Part of the proof of the consistency of the estimator, which will be presented in detail in Section 8.4, is inspired by Section 7.4. In that chapter we studied the asymptotic properties of a similar estimator for multiscale SDEs letting the number of observations go to infinity and the multiscale parameter vanish. The proofs or our results in the present chapter also requires us to perform a rigorous asymptotic analysis with respect to two parameters, the number of observations and the number of particles.

We first define the Jacobian matrix of the function $\bar g_j$ introduced in (8.2) with respect to the parameter θ, with $\otimes$ denoting the outer product in $\mathbb{R}^L$,

$$\begin{aligned} h_j(x,y;\theta) &:= \dot{\bar g}_j(x,y;\theta) \\ &= \dot\psi_j(x;\theta)\left(\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)\right) \\ &\quad + \psi_j(x;\theta) \otimes \left(\dot\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\left(\dot\phi_j(x;\theta) - \Delta\dot\lambda_j(\theta)\phi_j(x,\theta)\right)\right), \end{aligned}$$

as well as the following quantity

$$\ell_{j,k}(x,y;\theta) := (\psi_j(x;\theta)\otimes\psi_k(x;\theta))\left(\phi_j(y;\theta)\phi_k(y;\theta) - e^{-(\lambda_j(\theta)+\lambda_k(\theta))\Delta}\phi_j(x;\theta)\phi_k(x;\theta)\right).$$

We remark that whenever we write $\mathbb{E}^{\mu_\theta}$ we mean that $X_0 \sim \mu_\theta$ and similarly for the other probability measures.

We now present our main results. In Theorem 8.6 we prove that our estimator is consistent.

Theorem 8.6. *Let J be a positive integer and let $\{\widetilde{X}_m^{(n)}\}_{m=1}^M$ be a set of observations obtained by system* (5.9) *with true parameter θ_0. Under Assumptions 5.3 and 8.2 and if*

$$\det\left(\sum_{j=1}^J \mathbb{E}^{\mu_{\theta_0}}\left[h_j(X_0, X_\Delta;\theta_0)\right]\right) \neq 0, \tag{8.6}$$

there exists $N_0 > 0$ such that for all $N > N_0$ an estimator $\widehat{\theta}^J_{M,N}$, which solves the system $G^J_{M,N}(\theta) = 0$, exists with probability tending to one as M goes to infinity. Moreover, the estimator $\widehat{\theta}^J_{M,N}$ is asymptotically unbiased, i.e.,

$$\lim_{N\to\infty}\lim_{M\to\infty} \widehat{\theta}^J_{M,N} = \theta_0, \qquad \textit{in probability,} \tag{8.7}$$

$$\lim_{M\to\infty}\lim_{N\to\infty} \widehat{\theta}^J_{M,N} = \theta_0, \qquad \textit{in probability,} \tag{8.8}$$

and if $M = o(N)$

$$\lim_{M,N\to\infty} \widehat{\theta}^J_{M,N} = \theta_0, \qquad \textit{in probability.} \tag{8.9}$$

Then, in Theorem 8.7 we provide a rate of convergence for our estimator.

Theorem 8.7. *Let the assumptions of Theorem 8.6 hold, and let us introduce the notation*

$$\Xi^J_{M,N} \coloneqq \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)^{-1} \left\|\widehat{\theta}^J_{M,N} - \theta_0\right\|.$$

Then, for all $\varepsilon > 0$ there exists $K_\varepsilon > 0$ such that

$$\lim_{N\to\infty}\lim_{M\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon, \tag{8.10}$$

$$\lim_{M\to\infty}\lim_{N\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon, \tag{8.11}$$

and if $M = o(\sqrt{N})$

$$\lim_{M,N\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon. \tag{8.12}$$

Finally, in Theorem 8.8 we show that our estimator is asymptotically normal.

Theorem 8.8. *Let the assumptions of Theorem 8.6 hold with $M = o(\sqrt{N})$. Then, the estimator $\widehat{\theta}^J_{M,N}$ is asymptotically normal, i.e.,*

$$\lim_{M,N\to\infty} \sqrt{M}\left(\widehat{\theta}^J_{M,N} - \theta_0\right) = \Lambda^J \sim \mathcal{N}(0, \Gamma^J_0), \qquad \textit{in distribution,}$$

where

$$\begin{aligned} \Gamma^J_0 = \left(\sum_{j=1}^J \mathbb{E}^{\mu_{\theta_0}}\left[h_j(X_0, X_\Delta; \theta_0)\right]\right)^{-1} &\left(\sum_{j=1}^J\sum_{k=1}^J \mathbb{E}^{\mu_{\theta_0}}\left[\ell_{j,k}(X_0, X_\Delta; \theta_0)\right]\right) \\ &\times \left(\sum_{j=1}^J \mathbb{E}^{\mu_{\theta_0}}\left[h_j(X_0, X_\Delta; \theta_0)\right]\right)^{-\top}. \end{aligned} \tag{8.13}$$

Remark 8.9. We note that the technical assumption (8.6) is not a serious limitation of the validity of the theorem; in fact, it is a nondegeneracy hypothesis which holds true in all nonpathological cases and is equivalent to [73, Condition 4.2(a)] and Assumption 7.6. Moreover, it is not necessary to assume that the matrix Γ^J_0 in Theorem 8.8 is indeed a covariance matrix because, due to the particular form of the estimating function, this follows directly from the central limit theorem as explained in [73].

Remark 8.10. For the proof of the main results, we need to assume that, roughly speaking, the number of particles goes to infinity faster than the number of observations. It is not clear whether this assumption is strictly necessary. We expect that noncommutativity issues between the different distinguished limits may arise in the case where the mean field dynamics exhibits phase transitions, i.e., when the stationary state is not unique, see [40]. We will study the consequences of this noncommutativity due to phase transitions to the performance of our estimator and, more generally, to the inference problem in future work.

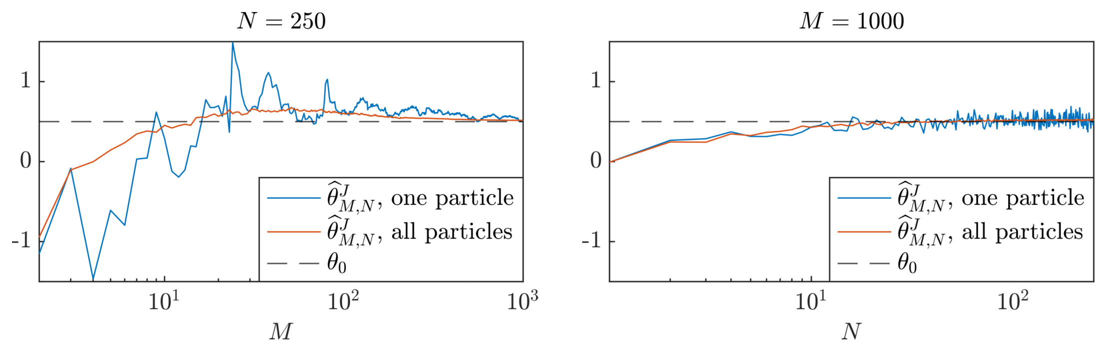


Figure 8.1 – Sensitivity analysis for the OU potential with respect to the number M of observations and N of particles, for the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$.

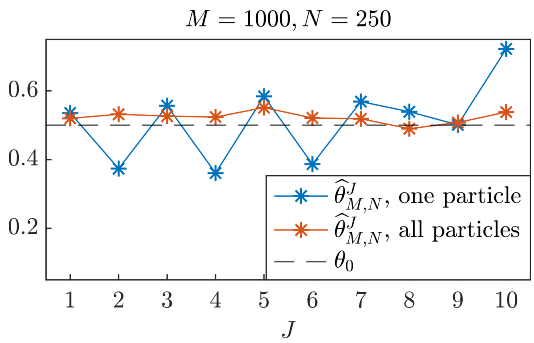


Figure 8.2 – Sensitivity analysis for the OU potential with respect to the number J of eigenvalues and eigenfunctions, for the estimator $\widehat{\theta}^J_{M,N}$.

8.3 Numerical experiments

In this section we present a series of numerical experiments to validate our theoretical results and demonstrate the effectiveness of our estimator in estimate unknown drift parameters of interacting particle systems. In order to generate synthetic data we employ the Euler–Maruyama (EM) method with a time step $h = 0.01$ to solve numerically system (5.9) and obtain $(X_t^{(n)})_{t\in[0,T]}$ for all $n = 1, \dots, N$. Notice that in order to preserve the exchangeability property of the system it is important to set the same initial condition for all the particles, hence we take $X_0^{(n)} = 0$ for all $n = 1, \dots, N$. We then randomly choose a value $n^* \in \{1, \dots, N\}$ and we assume to know a sample $\{X_m^{(n^*)}\}_{m=0}^M$ of observations obtained from the n^*-th particle with sampling rate Δ. We remark that the parameters h and Δ are not related to each other, in fact the former is only used to generate the data, while the latter is the actual distance between two consecutive observations. We repeat the same procedure for 5 different realizations of the Brownian motions and then we compute the average of the values obtained employing our estimator $\widehat{\theta}^J_{M,N}$. In the following, we first perform a sensitivity analysis with respect to the number of observations M, particles N and eigenvalues and eigenfunctions employed in the estimation J, then we confirm our theoretical results given in Theorems 8.6 to 8.8 and finally we test our technique with more challenging academic examples which do not exactly fit into the theory.

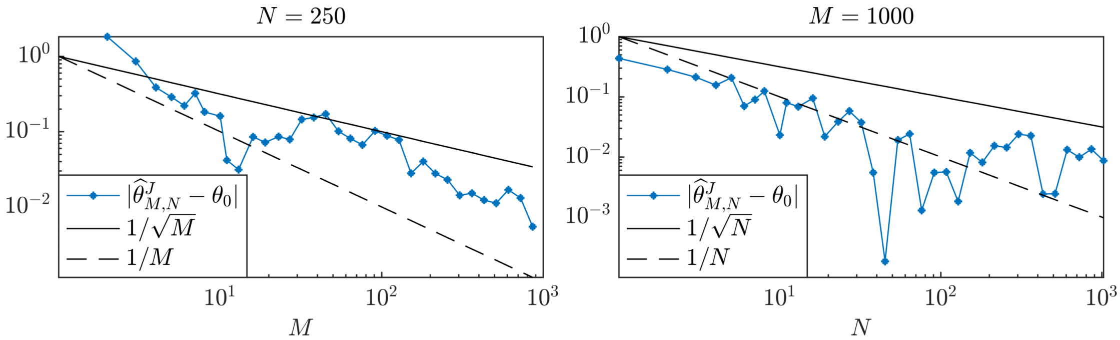


Figure 8.3 – Rates of convergence for the OU potential with respect to the number M of observations and N of particles, for the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$.

8.3.1 Sensitivity analysis and rate of convergence

We consider the setting of Example 8.5 choosing $\sigma = 1$, i.e., the interacting particle system reads

$$\mathrm{d}X_t^{(n)} = -X_t^{(n)}\,\mathrm{d}t - \kappa\left(X_t^{(n)} - \bar{X}_t^N\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{(n)}, \qquad n = 1, \dots, N, \tag{8.14}$$

and we aim to estimate the interaction parameter κ, so we write $\theta = \kappa$. We set $\kappa = 0.5$ and the number of eigenvalues and eigenfunctions $J = 1$ with $\psi_1(x;\theta) = x$, so that we can employ the analytical expression of our estimator given in (8.5). In Figure 8.1 we perform a sensitivity analysis for the estimator $\widehat{\theta}^1_{M,N}$ fixing $\Delta = 1$, varying the number M of observations and N of particles and choosing as other parameter respectively $N = 250$ and $M = 1000$, for which convergence has been reached. The blue line is the estimation given by one single particle while the red line is obtained by averaging the estimations computed employing all the different particles. We notice that convergence is reached when both N and M are large enough and, as expected, the estimation computed by averaging over all the particles stabilizes faster. Moreover, in Figure 8.2 we fix $M = 1000$ and $N = 250$ and we compare the results for different numbers J of eigenvalues and eigenfunctions employed in the construction of the estimating function. We observe that increasing the value of J does not significantly improves the results, hence it seems preferable to always choose $J = 1$ in order to reduce the computational cost. Finally, in Figure 8.3 we verify that the rates of convergence of the estimator $\widehat{\theta}^1_{M,N}$ towards the exact value θ_0 with respect to the number of observations M and particles N are consistent with the theoretical results given in Theorem 8.7. In particular, we observe that approximately it holds

$$\left|\widehat{\theta}^1_{M,N} - \theta_0\right| \simeq \mathcal{O}\left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right).$$

8.3.2 Comparison with the maximum likelihood estimator

We keep the same setting of Section 8.3.1 and we compare the results of our estimator with a maximum likelihood estimator (MLE). In particular, in [70] MLE for the interacting particle system with continuous observations is rigorously derived. Since for large values of N all the particles are approximately independent and identically distributed and we are assuming to observe only one particle, we replace the sample mean with the expectation with respect to the invariant measure, i.e., $\bar{X}_t^N = 0$, and we ignore the sum over all the particles. We then discretize the integrals in the formulation obtaining a modified MLE

$$\widetilde{\theta}^{\mathrm{MLE}}_{M,N} = -1 - \frac{\sum_{m=0}^{M-1} \widetilde{X}_m^{(n)}(\widetilde{X}_{m+1}^{(n)} - \widetilde{X}_m^{(n)})}{\Delta \sum_{m=0}^{M-1} (\widetilde{X}_m^{(n)})^2}. \tag{8.15}$$

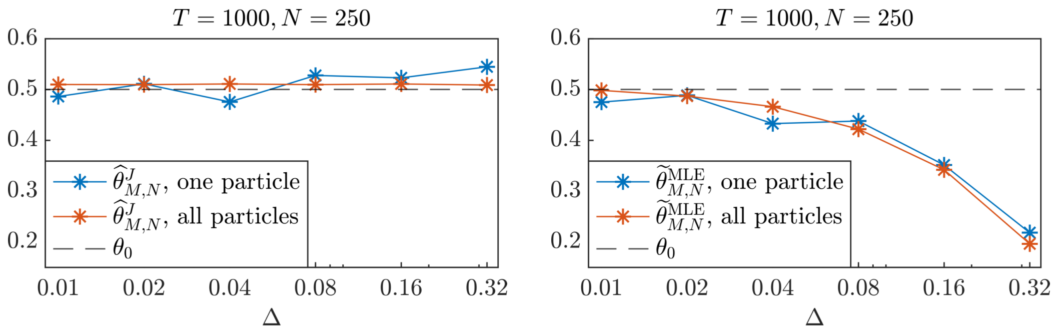


Figure 8.4 – Comparison between the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$ (left) and the maximum likelihood estimator $\widetilde{\theta}^{\mathrm{MLE}}_{M,N}$ (right) varying the distance Δ between two consecutive observations for the OU potential.

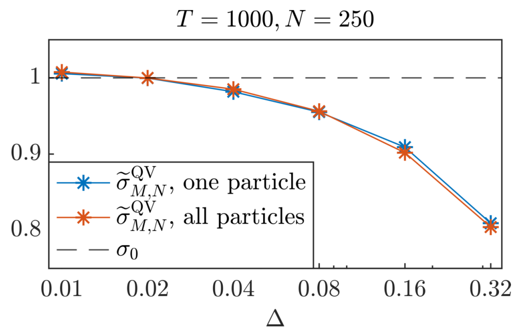


Figure 8.5 – Inference of the diffusion coefficient based on the quadratic variation varying the distance Δ between two consecutive observations for the OU potential.

In Figure 8.4 we fix the final time $T = 1000$ and we repeat the estimation for different values of $\Delta = 0.01 \cdot 2^i$ with $i = 0, \dots, 5$. We observe that, differently from our estimator, the MLE is unbiased only for small values of the sampling rate Δ, i.e., when the discrete observations approximate well the continuous trajectory. Notice also that, as highlighted by the numerical experiments, our estimator $\widehat{\theta}^1_{M,N}$ and the MLE $\widetilde{\theta}^{\mathrm{MLE}}_{M,N}$ defined respectively in (8.5) and (8.15) coincide in the limit of vanishing Δ. In fact, we can rewrite equation (8.5) as

$$\widehat{\theta}^1_{M,N} = -1 - \frac{1}{\Delta} \log \left(1 + \frac{\sum_{m=0}^{M-1} \widetilde{X}^{(n)}_m (\widetilde{X}^{(n)}_{m+1} - \widetilde{X}^{(n)}_m)}{\sum_{m=0}^{M-1} (\widetilde{X}^{(n)}_m)^2} \right),$$

observe that the fraction in the argument of the logarithm is $\mathcal{O}(\Delta)$ and employ the asymptotic expansion $\log(1 + x) \sim x$ for $x = o(1)$.

8.3.3 Diffusion coefficient

We still consider the setting of Example 8.5, but, differently from Section 8.3.1, we now assume the diffusion coefficient to be unknown and we aim to retrieve the correct values of the interaction parameter and the diffusion coefficient, which are given by $\kappa = 0.5$ and $\sigma = 1$, respectively. We set the number of particles $N = 250$ and the number of observations $M = 1000$. A first approach consists in first estimating the diffusion coefficient alone employing the quadratic variation and then infer the interaction parameter as in the previous numerical experiments. In particular, the

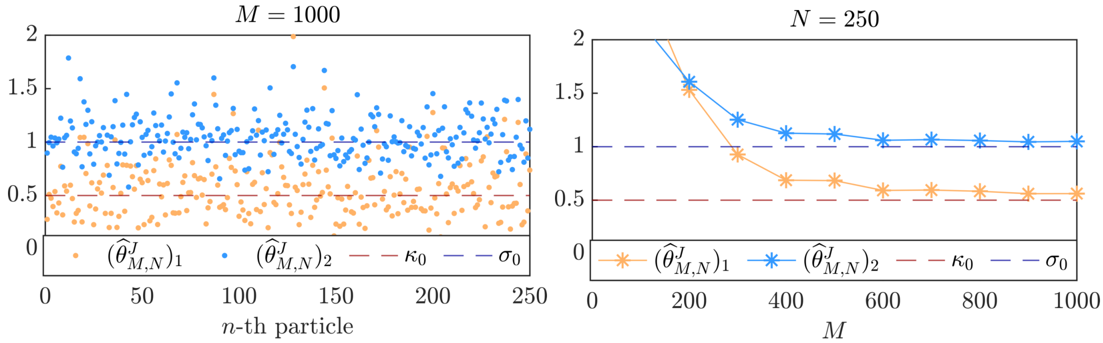


Figure 8.6 – Simultaneous inference of the interaction and diffusion coefficients for the OU potential. Left: estimation $\widehat{\theta}^J_{M,N}$ obtained from each particle with $J = 2$. Right: average of the estimations varying the number of observations.

diffusion coefficient can be approximated as

$$\widetilde{\sigma}^{\mathrm{QV}}_{M,N} = \frac{1}{2\Delta M} \sum_{m=0}^{M-1} (\widetilde{X}^{(n)}_{m+1} - \widetilde{X}^{(n)}_m)^2.$$

However, this estimator is asymptotically unbiased only in the limit of Δ vanishing and is therefore reliable only if the sampling rate is sufficiently small. In fact, one can prove that

$$\lim_{N\to\infty} \lim_{M\to\infty} \widetilde{\sigma}^{\mathrm{QV}}_{M,N} = \lim_{N\to\infty} \frac{1}{2\Delta} \mathbb{E}\left[(X^{(n)}_\Delta - X^{(n)}_0)^2\right] = \frac{1}{2\Delta} \mathbb{E}\left[(X_\Delta - X_0)^2\right],$$

which, due to the fact that in the framework of Example 8.5 X_t at stationarity is a Gaussian process with zero mean and covariance function

$$\mathcal{C}(t,s) = \frac{\sigma}{1+\kappa} e^{-(1+\kappa)|t-s|},$$

implies

$$\lim_{N\to\infty} \lim_{M\to\infty} \widetilde{\sigma}^{\mathrm{QV}}_{M,N} = \sigma \frac{1 - e^{-(1+\kappa)\Delta}}{(1+\kappa)\Delta},$$

where the right-hand side converges to σ if Δ goes to zero. This is also shown in Figure 8.5 where we estimate the diffusion coefficient for different values of the sampling rate $\Delta = 0.01 \cdot 2^i$ with $i = 0, \dots, 5$. Hence, if Δ is far from its vanishing limit we have to follow a different procedure. We now fix $\Delta = 1$ and aim to simultaneously infer the diffusion coefficient and the interaction parameter using our eigenfunction martingale estimators. We then write $\theta = \begin{pmatrix} \kappa & \sigma \end{pmatrix}^\top$ and in order to construct the estimating functions we employ $J = 2$ eigenvalues and eigenfunctions with functions $\psi_1(x;\theta) = \psi_2(x;\theta) = \begin{pmatrix} x^2 & x \end{pmatrix}^\top$. We remark that in the particular case of the OU process it is possible to express the eigenvalues and eigenfunctions analytically and the first two are given by

$$\begin{aligned} \lambda_1 &= 1 + \kappa, & \phi_1(x;\theta) &= x, \\ \lambda_2 &= 2(1+\kappa), & \phi_2(x;\theta) &= x^2 - \frac{\sigma}{1+\kappa}. \end{aligned}$$

Note that the first eigenvalue and eigenfunction do not depend on the diffusion coefficient σ and therefore they alone do not provide enough information, hence it is important to choose at least $J = 2$. In Figure 8.6 we show the numerical results. On the left and we plot the estimation computed employing one single particle for all the N particles and we observe that the estimators are concentrated around the exact values. On the other hand, on the right, we average all the estimations previously computed and we plot the results varying the number of observations M. We notice that the estimations stabilize fast near the correct coefficients.

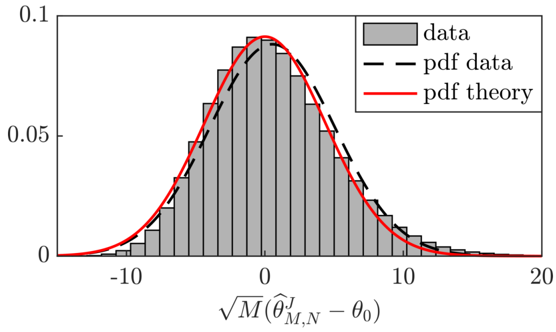


Figure 8.7 – Central limit theorems for the OU potential, for the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$.

8.3.4 Central limit theorem

We keep the same setting of Section 8.3.1 and we validate numerically the central limit theorem which we proved theoretically in Theorem 8.8. In this particular case, the asymptotic variance Γ_0^J can be computed analytically. In fact, the mean field limit of (8.14) at stationarity is

$$\mathrm{d}X_t = -(1+\kappa)X_t \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

and its solution $(X_t)_{t\in[0,T]}$ is a Gaussian process, i.e., $X \sim \mathcal{GP}(m(t), \mathcal{C}(t,s))$, where $m(t) = 0$ and

$$\mathcal{C}(t,s) = \frac{1}{1+\kappa} e^{-(1+\kappa)|t-s|}.$$

Moreover, we have

$$h_1(x,y;\theta) = \Delta e^{-(1+\kappa)\Delta} x^2 \qquad \text{and} \qquad \ell_{1,1}(x,y;\theta) = x^2 \left(y^2 - e^{-2(1+\kappa)\Delta} x^2 \right),$$

and therefore we obtain

$$\Gamma_0^J = \frac{e^{2(1+\kappa)\Delta} - 1}{\Delta^2}.$$

We then fix the number of particles $N = 1500$, the number of observations $M = 1000$ and the sampling rate $\Delta = 1$. In Figure 8.7 we plot the quantity $\sqrt{M}(\widehat{\theta}^J_{M,N} - \theta_0)$ for any particle $n = 1, \dots, N$ and for 500 realizations of the Brownian motion and we observe that it is approximately distributed as $\mathcal{N}(0, \Gamma_0^J)$ accordingly to the theoretical result.

8.3.5 Double well potential

We consider the setting of Example 5.4 and we analyse the double well potential, i.e., we let the confining potential $\mathcal{V}(\cdot;\alpha)$ be

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

with $\alpha = \begin{pmatrix} 1 & 2 \end{pmatrix}^\top$, which is the parameter that we aim to estimate, so we write $\theta = \alpha$. Moreover, we set the interaction term $\kappa = 0.5$ and the number of observations $M = 2000$ with sampling rate $\Delta = 0.5$. Finally, to construct the estimating functions we use $J = 1$ eigenfunctions and eigenvalues and we employ the function $\psi_1(x;\theta) = \begin{pmatrix} x & x^3 \end{pmatrix}^\top$. We remark that this example does not fit in Assumption 5.3, but if the diffusion coefficient σ is chosen sufficiently large, then we are below the phase transition and the mean field limit admits a unique invariant measure [38], so the theory applies. However, when the diffusion coefficient σ is below the critical noise strength,

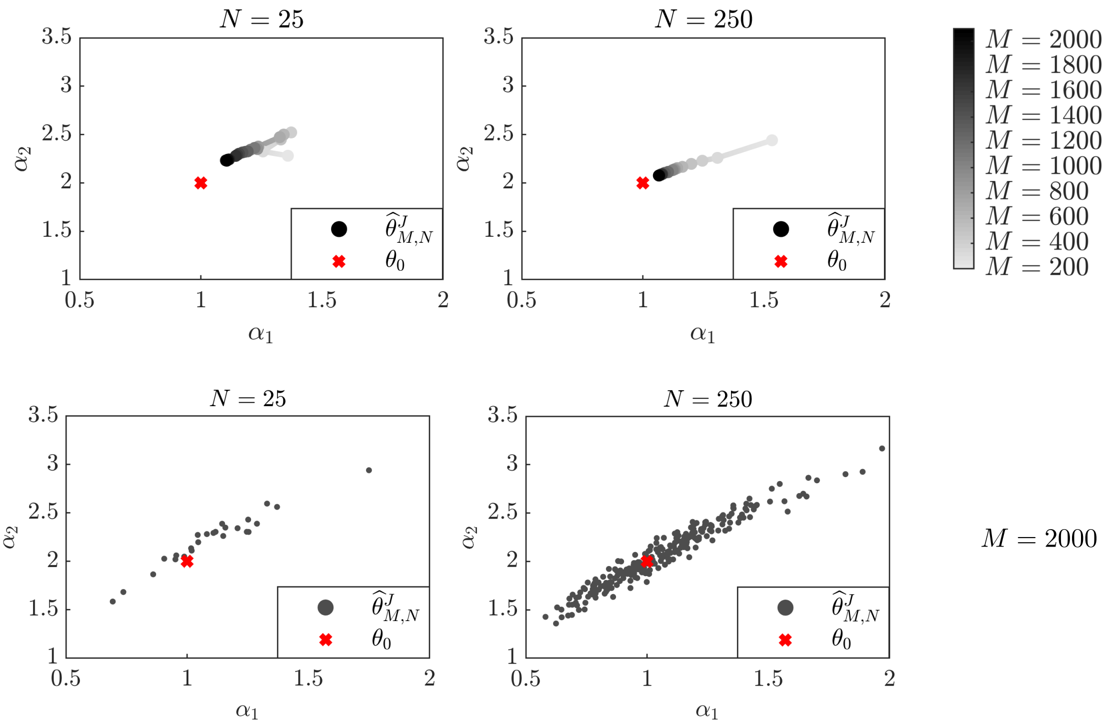


Figure 8.8 – Inference of the two-dimensional drift coefficient of the double well potential below the phase transition. Top: average of the estimations $\widehat{\theta}^J_{M,N}$ with $J = 1$ varying the number of observations. Bottom: scatter plot of the estimations obtained from each particle.

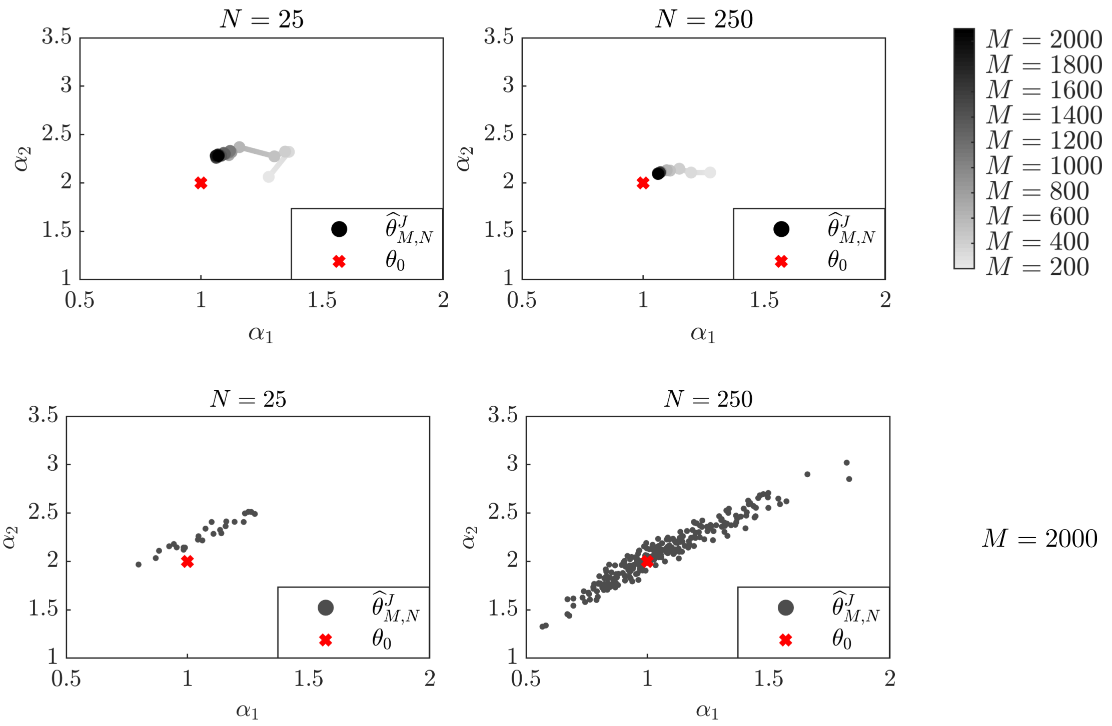


Figure 8.9 – Inference of the two-dimensional drift coefficient of the double well potential above the phase transition. Top: average of the estimations $\widehat{\theta}^J_{M,N}$ with $J = 1$ varying the number of observations. Bottom: scatter plot of the estimations obtained from each particle.

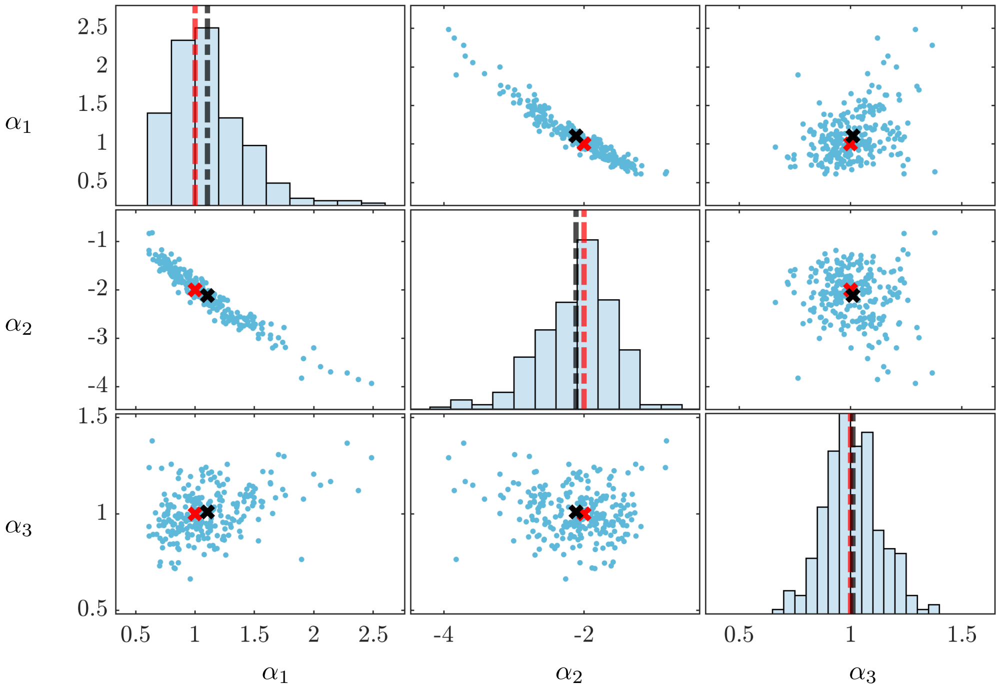


Figure 8.10 – Inference of the three-dimensional drift coefficient of a nonsymmteric potential for the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$. Diagonal: histogram of the estimations of each component obtained from all particles. Off-diagonal: scatter plot of the estimations obtained from all particles for two components at a time. Black and red stars/lines represent the average of the estimations and the exact value, respectively.

then a continuous phase transition occurs and two stationary states exist [59]. In particular, the transition point occurs at $\sigma \simeq 0.6$ with these data. We therefore perform two numerical experiments, one below and one above the phase transition, setting $\sigma = 0.75$ and $\sigma = 0.5$. In the former we have a unique invariant measure, so we can follow the usual approach, while in the latter we do not know in which state the data are converging. Nevertheless, the invariant distribution is known up to the first moment by equation (5.12), so we first estimate the expectation using the law of large numbers with the available observations and then repeat the same procedure as in the previous case. In Figures 8.8 and 8.9 we plot the results of these two experiments. On the top of the figures we plot the evolution of our estimator varying the number of observations M for two different values of the number of particles, in particular $N = 25$ and $N = 250$. We observe that the estimator approaches the correct drift coefficient α as the number of observations M increases and, as expected, the final approximation is better when the number of particles is sufficiently large. Moreover, on the bottom of the same figures we show the scatter plot of the estimations obtained from each particle with $M = 2000$ observations and we can see that they are concentrated around the exact drift coefficient α. We finally remark that we do not notice significant differences between two cases, yielding that the initial estimation of the first moment of the invariant measure does not affect the final results and thus that our methodology can be employed even when multiple stationary states exist.

8.3.6 Nonsymmetric confining potential

We still consider the same setting of Example 5.4 and we now study the case of a nonsymmetric potential. In particular, we let the confining potential $\mathcal{V}(\cdot;\alpha)$ be

$$\mathcal{V}(x;\alpha) = \alpha \cdot \begin{pmatrix} \frac{x^4}{4} & \frac{x^2}{2} & x \end{pmatrix}^\top,$$

with $\alpha = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}^\top$, which is the unknown parameter that we want to infer, hence we set $\theta = \alpha$. Notice that the confining potential is given by the sum of the double well potential and a linear term which breaks the symmetry. This type of potentials of the form $\mathcal{V}(x) = \sum_{\ell=1}^{\mathscr{L}} a_{2\ell} s^{2\ell} + a_1 s$, where $\mathscr{L} \geq 2$, $a_1, a_2 \in \mathbb{R}$, $a_4, \dots, a_{2(\mathscr{L}-1)} \geq 0$ and $a_{2\mathscr{L}} > 0$, which is used in the study of metastability and phase transitions and may have arbitrarily deep double wells, has been analyzed in [119, 124]. Similarly to the experiment in Section 8.3.5, this example does not satisfy Assumption 5.3 and more stationary states can exist. In particular, in [119] it has been proved the existence of an invariant measure around each critical point of the potential. We therefore adopt the same strategy as in the second part of Section 8.3.5 and, since the invariant measure is known up to the first moment by equation (5.12), we first approximate the expectation using the sample mean of the available observations, and then proceed with the following steps of the algorithm. We further set the interaction term $\kappa = 0.5$, the diffusion coefficient $\sigma = 1.5$, the number of particles $N = 250$ and the number of observations $M = 2000$ with sampling rate $\Delta = 0.5$. Moreover, to construct the estimating functions we use $J = 1$ eigenfunctions and eigenvalues and we employ the function $\psi_1(x;\theta) = \begin{pmatrix} x & x^2 & x^3 \end{pmatrix}^\top$. In Figure 8.10 we plot the results of the inference procedure considering two components of the three-dimensional drift coefficient at a time and the single components alone. We observe that the majority of the estimations obtained from all particles are concentrated around the exact values and that their average provides a reliable approximation of the true unknown. A peculiarity of this numerical experiment is the relationship between the first and second components of the estimated drift coefficient, in fact one increases when the other decreases and vice-versa, meaning that the two approximations appear to be correlated.

8.4 Proof of the main results

In this section we present the proof of Theorems 8.6 to 8.8, which are the main results of this chapter. We first recall that due to [50, Lemma 2.3.1] the solution of the interacting particle system $X_t^{(n)}$ and of its mean field limit X_t have bounded moments of any order, in particular there exists a constant $C > 0$ independent of N such that for all $t \in [0,T]$, $n = 1, \dots, N$ and $q \geq 1$

$$\mathbb{E}\left[\left|X_t^{(n)}\right|^q\right]^{1/q} \leq C \qquad \text{and} \qquad \mathbb{E}\left[|X_t|^q\right]^{1/q} \leq C. \tag{8.16}$$

Moreover, in [84, Theorem 3.3] it is shown that each particle converges to the solution of the mean field limit with the same Brownian motion in L^2, i.e, that

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left|X_t^{(n)} - X_t\right|^2\right]^{1/2} \leq \frac{C}{\sqrt{N}}, \tag{8.17}$$

where the constant C is also independent of the final time T. We also state here a formula which has been proved in [73] and will be crucial in the last part of the proof

$$\mathbb{E}^{\mu_{\theta_0}}[\phi_j(X_\Delta;\theta_0) \mid X_0 = x] = e^{-\lambda_j(\theta_0)\Delta}\phi_j(x;\theta_0), \qquad \text{for all } j = 1, \dots, J, \tag{8.18}$$

where θ_0 is the true parameter which generates the path $(X_t)_{t\in[0,T]}$ and $\mathbb{E}^{\mu_{\theta_0}}$ denotes the fact that $X_0 \sim \mu_{\theta_0}$. Before entering the main part of the proof, we introduce some notation and technical results which will be used later. We finally remark that all the constants will be denoted by C and their value can change from line to line.

8.4.1 Limits of the estimating function and its derivative

Let us first define the following vector-valued functions $\mathbb{G}_M^J(\theta), \mathcal{G}_N^J(\theta), \mathscr{G}^J(\theta) \colon \mathbb{R}^L \to \mathbb{R}^L$ and matrix-valued functions $\mathbb{H}_M^J(\theta), \mathcal{H}_N^J(\theta), \mathscr{H}^J(\theta) \colon \mathbb{R}^L \to \mathbb{R}^{L\times L}$

$$
\begin{aligned}
\mathbb{G}_M^J(\theta) &\coloneqq \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J} \bar{g}_j(\widetilde{X}_m, \widetilde{X}_{m+1};\theta), & \mathbb{H}_M^J(\theta) &\coloneqq \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J} h_j(\widetilde{X}_m, \widetilde{X}_{m+1};\theta), \\
\mathcal{G}_N^J(\theta) &\coloneqq \sum_{j=1}^{J} \mathbb{E}^{\mu_{\theta_0}^N}\left[\bar{g}_j(X_0^{(n)}, X_\Delta^{(n)};\theta)\right], & \mathcal{H}_N^J(\theta) &\coloneqq \sum_{j=1}^{J} \mathbb{E}^{\mu_{\theta_0}^N}\left[h_j(X_0^{(n)}, X_\Delta^{(n)};\theta)\right], \\
\mathscr{G}^J(\theta) &\coloneqq \sum_{j=1}^{J} \mathbb{E}^{\mu_{\theta_0}}\left[\bar{g}_j(X_0, X_\Delta;\theta)\right], & \mathscr{H}^J(\theta) &\coloneqq \sum_{j=1}^{J} \mathbb{E}^{\mu_{\theta_0}}\left[h_j(X_0, X_\Delta;\theta)\right].
\end{aligned}
\tag{8.19}
$$

The following lemma then shows that these quantities are bounded in a suitable norm and thus well defined.

Lemma 8.11. *Under Assumptions 5.3 and 8.2 there exists a constant $C > 0$ independent of M, N such that for all $q \geq 1$*

$$
(i)\ \mathbb{E}\left[\left\|G_{M,N}^J(\theta)\right\|^q\right] \leq C, \qquad (ii)\ \mathbb{E}\left[\left\|\mathbb{G}_M^J(\theta)\right\|^q\right] \leq C,
$$
$$
(iii)\ \left\|\mathcal{G}_N^J(\theta)\right\| \leq C, \qquad (iv)\ \left\|\mathscr{G}^J(\theta)\right\| \leq C.
$$

Proof. Since the argument is similar for the four cases, we only write the details of (i). Using the triangle inequality we have

$$
\mathbb{E}\left[\left\|G_{M,N}^J(\theta)\right\|^q\right] \leq \frac{2^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^q\left(\left|\phi_j(\widetilde{X}_{m+1}^{(n)};\theta)\right|^q + \left|\phi_j(\widetilde{X}_m^{(n)};\theta)\right|^q\right)\right],
$$

and due to the Cauchy–Schwarz inequality we obtain

$$
\begin{aligned}
\mathbb{E}\left[\left\|G_{M,N}^J(\theta)\right\|\right] &\leq \frac{2^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^{2q}\right]^{1/2}\mathbb{E}\left[\left|\phi_j(\widetilde{X}_{m+1}^{(n)};\theta)\right|^{2q}\right]^{1/2} \\
&\quad + \frac{2^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^{2q}\right]^{1/2}\mathbb{E}\left[\left|\phi_j(\widetilde{X}_m^{(n)};\theta)\right|^{2q}\right]^{1/2}.
\end{aligned}
$$

Finally, bound (8.16) together with the fact that ψ_j and ϕ_j are polynomially bounded for all $j = 1, \dots, J$ by Assumption 8.2 gives the desired result. □

In the next proposition we study the behaviour of the estimating function $G_{M,N}^J$ as the number of observations M and particles N go to infinity.

Proposition 8.12. *Under Assumptions 5.3 and 8.2 it holds for all $1 \leq q < 2$*

$$
(i)\ \lim_{N\to\infty} G_{M,N}^J(\theta) = \mathbb{G}_M^J(\theta), \quad \text{in } L^q, \qquad (ii)\ \lim_{M\to\infty} \mathbb{G}_M^J(\theta) = \mathscr{G}^J(\theta), \quad \text{in } L^2,
$$
$$
(iii)\ \lim_{M\to\infty} G_{M,N}^J(\theta) = \mathcal{G}_N^J(\theta), \quad \text{in } L^2, \qquad (iv)\ \lim_{N\to\infty} \mathcal{G}_N^J(\theta) = \mathscr{G}^J(\theta).
$$

Moreover, there exists a constant $C > 0$ independent of M, N and θ such that

$$
(i)'\ \mathbb{E}\left[\left\|G_{M,N}^J(\theta) - \mathbb{G}_M^J(\theta)\right\|^q\right]^{1/q} \leq \frac{C}{\sqrt{N}}, \qquad (iv)'\ \left\|\mathcal{G}_N^J(\theta) - \mathscr{G}^J(\theta)\right\| \leq \frac{C}{\sqrt{N}}.
$$

Proof. Results (ii) and (iii) are direct consequences of [21, Lemma 3.1] and of the ergodicity of the processes $(X_t^{(n)})_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$ given by [59, Section 1] and [84, Theorem 3.16], respectively. Let us now consider cases (i) and $(i)'$. Using the triangle inequality we have

$$\mathbb{E}\left[\left\|G_{M,N}^J(\theta)-\mathbb{G}_M^J(\theta)\right\|^q\right] \le \frac{4^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\left(Q_{m,j}^{(1)}+Q_{m,j}^{(2)}+Q_{m,j}^{(3)}+Q_{m,j}^{(4)}\right),$$

where

$$\begin{aligned}
Q_{m,j}^{(1)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde X_m^{(n)};\theta)\right\|^q\left|\phi_j(\widetilde X_{m+1}^{(n)};\theta)-\phi_j(\widetilde X_{m+1};\theta)\right|^q\right],\\
Q_{m,j}^{(2)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde X_m^{(n)};\theta)\right\|^q\left|\phi_j(\widetilde X_{m}^{(n)};\theta)-\phi_j(\widetilde X_{m};\theta)\right|^q\right],\\
Q_{m,j}^{(3)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde X_m^{(n)};\theta)-\psi_j(\widetilde X_m;\theta)\right\|^q\left|\phi_j(\widetilde X_{m+1};\theta)\right|^q\right],\\
Q_{m,j}^{(4)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde X_m^{(n)};\theta)-\psi_j(\widetilde X_m;\theta)\right\|^q\left|\phi_j(\widetilde X_{m};\theta)\right|^q\right],
\end{aligned}$$

and applying the mean value theorem we obtain

$$\begin{aligned}
Q_{m,j}^{(1)} &\le \mathbb{E}\left[\left\|\psi_j(\widetilde X_m^{(n)};\theta)\right\|^q\left|\int_0^1\phi_j'(\widetilde X_{m+1}+s(\widetilde X_{m+1}^{(n)}-\widetilde X_{m+1});\theta)\,\mathrm{d}s\right|^q\left|\widetilde X_{m+1}^{(n)}-\widetilde X_{m+1}\right|^q\right],\\
Q_{m,j}^{(2)} &\le \mathbb{E}\left[\left\|\psi_j(\widetilde X_m^{(n)};\theta)\right\|^q\left|\int_0^1\phi_j'(\widetilde X_{m}+s(\widetilde X_{m}^{(n)}-\widetilde X_{m});\theta)\,\mathrm{d}s\right|^q\left|\widetilde X_{m}^{(n)}-\widetilde X_{m}\right|^q\right],\\
Q_{m,j}^{(3)} &\le \mathbb{E}\left[\left\|\int_0^1\psi_j'(\widetilde X_{m}+s(\widetilde X_{m}^{(n)}-\widetilde X_{m});\theta)\,\mathrm{d}s\right\|^q\left|\widetilde X_{m}^{(n)}-\widetilde X_{m}\right|^q\left|\phi_j(\widetilde X_{m+1};\theta)\right|^q\right],\\
Q_{m,j}^{(4)} &\le \mathbb{E}\left[\left\|\int_0^1\psi_j'(\widetilde X_{m}+s(\widetilde X_{m}^{(n)}-\widetilde X_{m});\theta)\,\mathrm{d}s\right\|^q\left|\widetilde X_{m}^{(n)}-\widetilde X_{m}\right|^q\left|\phi_j(\widetilde X_{m};\theta)\right|^q\right].
\end{aligned}$$

Then, employing the Hölder inequality with exponents $4/(2-q), 4/(2-q), 2/q$ and since the functions $\phi_j, \phi_j', \psi_j, \psi_j'$ are polynomially bounded by Assumption 8.2 and $\widetilde X_m^{(n)}, \widetilde X_m$ have bounded moments of any order by (8.16) we deduce

$$\mathbb{E}\left[\left\|G_{M,N}^J(\theta)-\mathbb{G}_M^J(\theta)\right\|^q\right] \le \frac{C}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\left(\mathbb{E}\left[(\widetilde X_m^{(n)}-\widetilde X_m)^2\right]^{\frac{q}{2}}+\mathbb{E}\left[(\widetilde X_{m+1}^{(n)}-\widetilde X_{m+1})^2\right]^{\frac{q}{2}}\right),$$

which due to (8.17) proves $(i)'$, which directly implies (i). Finally, the proofs of results (iv) and $(iv)'$ are similar to cases (i) and $(i)'$, respectively, and are omitted here. □

Corollary 8.13. *Under Assumptions 5.3 and 8.2 it holds for all* $1 \le q < 2$

$$\lim_{M,N\to\infty} G_{M,N}^J(\theta) = \mathscr{G}^J(\theta), \quad \text{in } L^q.$$

Proof. Employing the triangle inequality we have

$$\mathbb{E}\left[\left\|G_{M,N}^J(\theta)-\mathscr{G}^J(\theta)\right\|^q\right] \le 2^{q-1}\left(\mathbb{E}\left[\left\|G_{M,N}^J(\theta)-\mathbb{G}_M^J(\theta)\right\|^q\right]+\mathbb{E}\left[\left\|\mathbb{G}_M^J(\theta)-\mathscr{G}^J(\theta)\right\|^q\right]\right),$$

where the right-hand side vanishes by $(i)'$ and (ii) in Proposition 8.12, yielding the desires result. □

The limits considered in Proposition 8.12 are summarized schematically in the following diagram

$$
\begin{tikzcd}
& \mathbb{G}^J_M(\theta) \arrow[dr, "\text{in } L^2", "M\to\infty"'] & \\
G^J_{M,N}(\theta) \arrow[ur, "\text{in } L^q", "N\to\infty"'] \arrow[dr, "M\to\infty", "\text{in } L^2"'] & & \mathscr{G}^J(\theta) \\
& \mathcal{G}^J_N(\theta) \arrow[ur, "N\to\infty"] &
\end{tikzcd}
$$

where $q \in [1, 2)$.

Remark 8.14. Notice that all the results in this section hold true also for the derivatives $\mathbb{H}^J_M(\theta)$, $\mathcal{H}^J_N(\theta)$, $\mathscr{H}^J(\theta)$ with respect to the parameter θ defined in (8.19). Since the arguments are analogous we omit the details here.

8.4.2 Zeros of the limits of the estimating function

The goal of this section is to show that the limits of the estimating functions previously defined admit zeros and to study their asymptotic limit. We already know by (8.18) that $\mathscr{G}^J(\theta_0) = 0$, where θ_0 is the true parameter. Then, in the following lemma we consider the zero of the function $\mathcal{G}^J_N(\theta)$ and its limit as $N \to \infty$.

Lemma 8.15. *Under Assumptions 5.3 and 8.2 and if* $\det(\mathscr{H}^J(\theta_0)) \neq 0$ *there exists* $N_0 > 0$ *such that for all* $N > N_0$ *there exists* $\vartheta^J_N \in \Theta$ *which solves the system* $\mathcal{G}^J_N(\theta) = 0$ *and satisfies* $\det(\mathcal{H}^J_N(\vartheta^J_N)) \neq 0$. *Moreover, there exists a constant* $C > 0$ *independent of* N *such that*

$$\left\|\vartheta^J_N - \theta_0\right\| \le \frac{C}{\sqrt{N}}. \tag{8.20}$$

Proof. We first remark that by (8.18) we have $\mathscr{G}^J(\theta_0) = 0$ and, without loss of generality, we can assume that $\det(\mathscr{H}^J(\theta_0)) > 0$. Let $\delta > 0$ sufficiently small, by point $(iv)'$ in Proposition 8.12 and Remark 8.14 we know that $\mathcal{H}^J_N(\theta)$ converges to $\mathscr{H}^J(\theta)$ uniformly in θ and therefore there exist $N_1 > 0$ and $\varepsilon > 0$ such that for all $N > N_1$ and for all $\theta \in B_\varepsilon(\theta_0)$

$$0 < \det(\mathscr{H}^J(\theta_0)) - \delta \le \det(\mathcal{H}^J_N(\theta)) \le \det(\mathscr{H}^J(\theta_0)) + \delta, \tag{8.21}$$

$$0 < \left\|\mathscr{H}^J(\theta_0)^{-1}\right\| - \delta \le \left\|\mathcal{H}^J_N(\theta)^{-1}\right\| \le \left\|\mathscr{H}^J(\theta_0)^{-1}\right\| + \delta. \tag{8.22}$$

Hence, due to equation (8.21) and applying the inverse function theorem we deduce the existence of $\eta > 0$ such that

$$B_\eta(\mathcal{G}^J_N(\theta_0)) \subseteq \mathcal{G}^J_N(B_\varepsilon(\theta_0)).$$

Notice that the radius $\eta > 0$ can be chosen independently of $N > N_1$. In fact, by the proof of [95, Theorem 2.3] and [78, Lemma 1.3] we observe that η is dependent on the radius ε of the ball $B_\varepsilon(\theta_0)$ and the quantity $\left\|\mathcal{H}^J_N(\theta_0)^{-1}\right\|$, which can be bounded independently of $N > N_1$ due to estimate (8.22). Moreover, since

$$\lim_{N\to\infty} \mathcal{G}^J_N(\theta_0) = \mathscr{G}^J(\theta_0) = 0,$$

then there exists $N_2 > 0$ such that for all $N > N_2$ we have $0 \in B_\eta(\mathcal{G}^J_N(\theta_0))$. Therefore, setting $N_0 = \max\{N_1, N_2\}$ for all $N > N_0$ there exists $\vartheta^J_N \in B_\varepsilon(\theta_0)$ such that $\mathcal{G}^J_N(\vartheta^J_N) = 0$, which proves the existence. Furthermore, equation (8.21) gives $\det(\mathcal{H}^J_N(\vartheta^J_N)) \neq 0$. It now remains to show estimate (8.20). Since the set $\overline{B_\varepsilon(\theta_0)}$ is compact, there exist $\widetilde{\vartheta}^J \in \overline{B_\varepsilon(\theta_0)}$ and a subsequence $\vartheta^J_{N_k}$ such that

$$\lim_{k\to\infty} \vartheta^J_{N_k} = \widetilde{\vartheta}^J.$$

By point $(iv)'$ in Proposition 8.12 the function $\mathcal{G}_N^J(\theta)$ converges to $\mathscr{G}^J(\theta)$ uniformly in θ, thus we have

$$0 = \lim_{k\to\infty} \mathcal{G}_{N_k}^J(\vartheta_{N_k}^J) = \lim_{k\to\infty} \left[\mathcal{G}_{N_k}^J(\vartheta_{N_k}^J) - \mathscr{G}^J(\vartheta_{N_k}^J) + \mathscr{G}^J(\vartheta_{N_k}^J)\right] = \mathscr{G}^J(\widetilde{\vartheta}^J),$$

which yields $\widetilde{\vartheta}^J = \theta_0$. This is guaranteed by the fact that ε can be previously chosen sufficiently small such that θ_0 is the only zero of the function $\mathscr{G}^J(\theta)$ in $B_\varepsilon(\theta_0)$. Since θ_0 is the unique limit point for the subsequence $\vartheta_{N_k}^J$, it follows that the whole sequence converges. Then, applying the mean value theorem we obtain

$$\mathscr{G}^J(\vartheta_N^J) - \mathcal{G}_N^J(\vartheta_N^J) = \mathscr{G}^J(\vartheta_N^J) - \mathscr{G}^J(\theta_0) = \left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)(\vartheta_N^J - \theta_0),$$

which implies

$$\left\|\vartheta_N^J - \theta_0\right\| \le \left\|\left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)^{-1}\right\| \left\|\mathscr{G}^J(\vartheta_N^J) - \mathcal{G}_N^J(\vartheta_N^J)\right\|.$$

Since ϑ_N^J converges to θ_0 as N goes to infinity, then

$$\lim_{N\to\infty}\left\|\left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)^{-1}\right\| = \left\|\mathscr{H}^J(\theta_0)^{-1}\right\|,$$

where the right-hand side is well defined because $\det(\mathscr{H}^J(\theta_0)) \neq 0$. Therefore, if N is sufficiently large there exists a constant $C > 0$ independent of N such that

$$\left\|\left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)^{-1}\right\| \le C,$$

which together with point $(iv)'$ in Proposition 8.12 yields estimate (8.20) and concludes the proof. $\square$

In the next lemma we study the zero of the random function $\mathbb{G}_M^J(\theta)$ and its limit as $M \to \infty$. This result is almost the same as [73, Theorem 4.3].

Lemma 8.16. *Let the assumptions of Lemma 8.15 hold. Then, an estimator $\widehat{\vartheta}_M^J$, which solves the equation $\mathbb{G}_M^J(\theta) = 0$ and is such that $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0$, exists with a probability tending to one as $M \to \infty$. Moreover,*

$$\lim_{M\to\infty} \widehat{\vartheta}_M^J = \theta_0, \qquad \textit{in probability},$$

and

$$\lim_{M\to\infty} \sqrt{M}\left(\widehat{\vartheta}_M^J - \theta_0\right) = \Lambda^J \sim \mathcal{N}(0, \Gamma_0^J), \qquad \textit{in distribution},$$

where Γ_0^J is defined in (8.13).

Proof. The existence of the estimator $\widehat{\vartheta}_M^J$ which solves the equation $\mathbb{G}_M^J(\theta) = 0$ with a probability tending to one as $M \to \infty$ and its asymptotic unbiasedness and normality is given by [73, Theorem 4.3], whose prove can be found in [21, Theorem 3.2] and is based on [16, Theorem A.1]. Moreover, by the last line of the proof of [21, Theorem 3.2] or by (A.5) in [73, Theorem 4.3] we have

$$\lim_{M\to\infty} \mathbb{H}_M^J(\widehat{\vartheta}_M^J) = \mathscr{H}^J(\theta_0), \qquad \text{in probability}, \tag{8.23}$$

where $\det(\mathscr{H}^J(\theta_0)) \neq 0$ by assumption. Hence, there exists $\delta > 0$ such that if

$$\left\| \mathbb{H}_M^J(\widehat{\vartheta}_M^J) - \mathscr{H}^J(\theta_0) \right\| \leq \delta,$$

then $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0$. Moreover, for M large enough it holds

$$\mathbb{P}\left(\left\| \mathbb{H}_M^J(\widehat{\vartheta}_M^J) - \mathscr{H}^J(\theta_0) \right\| \leq \delta \right) \geq 1 - \varepsilon_M,$$

where $\varepsilon_M \to 0$ as $M \to \infty$. Let us now define the events

$$A_M \coloneqq \left\{ \exists\, \widehat{\vartheta}_M^J \colon \mathbb{G}_M^J(\widehat{\vartheta}_M^J) \right\} \qquad \text{and} \qquad B_M \coloneqq \left\{ \left\| \mathbb{H}_M^J(\widehat{\vartheta}_M^J) - \mathscr{H}^J(\theta_0) \right\| \leq \delta \right\},$$

and notice that by the first part of the proof we have $\mathbb{P}(A_M) = p_M$ where $p_M \to 1$ as $M \to \infty$. Then, using the basic properties of probability measures we obtain

$$\mathbb{P}\left(A_M \cap \{\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0\} \right) \geq \mathbb{P}\left(A_M \cap B_M \right) \geq \mathbb{P}(A_M) + \mathbb{P}(B_M) - 1 \geq p_M - \varepsilon_M,$$

where the last term tends to one as $M \to \infty$, and which gives the desired result. $\square$

We now consider the zero of the actual estimating function $G_{M,N}^J(\theta)$ and we first analyze its limit as $M \to \infty$.

Lemma 8.17. *Let the assumptions of Theorem 8.6 hold. Then, there exists $N_0 > 0$ such that for all $N > N_0$ an estimator $\widehat{\theta}_{M,N}^J$, which solves the system $G_{M,N}^J(\theta) = 0$, exists with a probability tending to one as M goes to infinity. Moreover, there exist ϑ_N^J solving $\mathcal{G}_N^J(\theta) = 0$ such that*

$$\lim_{M \to \infty} \widehat{\theta}_{M,N}^J = \vartheta_N^J, \qquad \textit{in probability,}$$

and

$$\lim_{M \to \infty} \sqrt{M} \left(\widehat{\theta}_{M,N}^J - \vartheta_N^J \right) = \Lambda_N^J \sim \mathcal{N}(0, \Gamma_N^J), \qquad \textit{in distribution,}$$

where Γ_N^J is a positive definite covariance matrix such that $\lim_{N \to \infty} \Gamma_N^J = \Gamma_0^J$ where Γ_0^J is defined in (8.13).

Proof. First, by Lemma 8.15 there exists $N_0 > 0$ such that for all $N > N_0$ there exists ϑ_N^J such that

$$\mathcal{G}_N^J(\vartheta_N^J) = 0 \qquad \text{and} \qquad \det(\mathcal{H}_N^J(\vartheta_N^J)) \neq 0.$$

Then, the results are equivalent to Lemma 8.16 and therefore the argument follows the same steps of its proof, which is given in detail in [21, Theorem 3.2] and is based on [16, Theorem A.1]. Finally, the convergence of the covariance matrix Γ_N^J is implied by (8.17). $\square$

We then study the limit of the zero of $G_{M,N}^J(\theta)$ as $N \to \infty$.

Lemma 8.18. *Let the assumptions of Lemma 8.17 hold and let $M \ll N$. Then, the estimator $\widehat{\theta}_{M,N}^J$ satisfies for some $\widehat{\vartheta}_M^J$ solving $\mathbb{G}_M^J(\theta) = 0$ and for a constant $C > 0$ independent of M and N*

$$\mathbb{E}\left[\left\| \widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J \right\| \right] \leq C \sqrt{\frac{M}{N}}.$$

Proof. The existence of the estimators $\widehat{\vartheta}^J_M$, such that $\mathbb{G}^J_M(\widehat{\vartheta}^J_M) = 0$ and $\det(\mathbb{H}^J_M(\widehat{\vartheta}^J_M)) \neq 0$, and $\widehat{\theta}^J_{M,N}$, such that $G^J_{M,N}(\widehat{\theta}^J_{M,N}) = 0$, with a probability tending to one as M goes to infinity is guaranteed by Lemmas 8.16 and 8.17, respectively. Then, all the following events are considered as conditioned on the existence of $\widehat{\vartheta}^J_M$ and $\widehat{\theta}^J_{M,N}$ and the fact that $\det(\mathbb{H}^J_M(\widehat{\vartheta}^J_M)) \neq 0$. Let us now define the function $f\colon \mathbb{R}^L \times \mathbb{R}^{M+1} \to \mathbb{R}^L$ as

$$f(\theta, x) = \frac{1}{M} \sum_{m=0}^{M-1} \sum_{j=1}^{J} \bar{g}_j(x_m, x_{m+1}; \theta),$$

where x_m denotes the m-th component of the vector $x \in \mathbb{R}^{m+1}$, and the vectors $\mathbb{X}^{(n)}$ and $\mathbb{X}$ whose m-th components for $m = 0, \dots, M$ are given by

$$\mathbb{X}^{(n)}_m = \widetilde{X}^{(n)}_m \qquad \text{and} \qquad \mathbb{X}_m = \widetilde{X}_m,$$

where $\{\widetilde{X}^{(n)}_m\}_{m=0}^M$ is the set of observations and $\{\widetilde{X}_m\}_{m=0}^M$ are the corresponding realizations of the mean field limit. Notice that $f \in C^1(\Theta \times \mathbb{R}^{M+1})$ due to Assumption 8.2 and Remark 8.3 and by definition we have

$$f(\widehat{\vartheta}^J_M, \mathbb{X}) = 0 \qquad \text{and} \qquad \det\left(\frac{\partial f}{\partial \theta}(\widehat{\vartheta}^J_M, \mathbb{X})\right) \neq 0.$$

Therefore, applying the implicit function theorem there exist $\varepsilon, \delta > 0$ and a continuously differentiable function $F\colon B_\varepsilon(\mathbb{X}) \to B_\delta(\widehat{\vartheta}^J_M)$ such that $f(F(x), x) = 0$ for all $x \in B_\varepsilon(\mathbb{X})$. Hence, if $\mathbb{X}^{(n)}$ is close enough to $\mathbb{X}$ then there must be one $\widehat{\theta}^J_{M,N} \in B_\delta(\widehat{\vartheta}^J_M)$ such that $F(\mathbb{X}^{(n)}) = \widehat{\theta}^J_{M,N}$. Then, employing Jensen's inequality and by estimate (8.17) we have

$$\mathbb{E}\left[\left\|\mathbb{X}^{(n)} - \mathbb{X}\right\|\right] = \mathbb{E}\left[\left(\sum_{m=0}^{M} \left|\widetilde{X}^{(n)}_m - \widetilde{X}_m\right|^2\right)^{1/2}\right] \leq \left(\sum_{m=0}^{M} \mathbb{E}\left[\left|\widetilde{X}^{(n)}_m - \widetilde{X}_m\right|^2\right]\right)^{1/2} \leq C\sqrt{\frac{M}{N}},$$

where the constant C is independent of M and N. Therefore, letting $\varepsilon > 0$ and applying Markov's inequality we obtain

$$\mathbb{P}\left(\left\|\mathbb{X}^{(n)} - \mathbb{X}\right\| \geq \varepsilon\right) \leq \frac{1}{\varepsilon}\,\mathbb{E}\left[\left\|\mathbb{X}^{(n)} - \mathbb{X}\right\|\right] \leq \frac{C}{\varepsilon}\sqrt{\frac{M}{N}}. \tag{8.24}$$

Defining the event $A = \{\|\mathbb{X}^{(n)} - \mathbb{X}\| < \varepsilon\}$ and using the law of total expectation conditioning on A we deduce

$$\mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\|\right] = \mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\| |A\right] \mathbb{P}(A) + \mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\| |A^{\mathsf{C}}\right] \mathbb{P}(A^{\mathsf{C}}),$$

which since $\widehat{\theta}^J_{M,N}, \widehat{\vartheta}^J_M \in \Theta$, a compact set, and due to estimate (8.24) implies

$$\mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\|\right] \leq \mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\| |A\right] + C\sqrt{\frac{M}{N}}. \tag{8.25}$$

It now remains to study the first term in the right-hand side. Applying the mean value theorem we obtain

$$\begin{aligned}
\mathbb{G}^J_M(\widehat{\theta}^J_{M,N}) - G^J_{M,N}(\widehat{\theta}^J_{M,N}) &= \mathbb{G}^J_M(\widehat{\theta}^J_{M,N}) - \mathbb{G}^J_M(\widehat{\vartheta}^J_M) \\
&= \left(\int_0^1 \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M))\,\mathrm{d}t\right)(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M),
\end{aligned}$$

which implies

$$\left\| \widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M \right\| \le \left\| \left(\int_0^1 \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \,\mathrm{d}t \right)^{-1} \right\| \left\| \mathbb{G}^J_M(\widehat{\theta}^J_{M,N}) - G^J_{M,N}(\widehat{\theta}^J_{M,N}) \right\|.$$

Using Hölder inequality with exponents $q \in (1,2)$ and its conjugate q' such that $1/q + 1/q' = 1$ we have

$$\mathbb{E}\left[\left\| \widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M \right\| |A \right] \le Q\,\mathbb{E}\left[\left\| \mathbb{G}^J_M(\widehat{\theta}^J_{M,N}) - G^J_{M,N}(\widehat{\theta}^J_{M,N}) \right\|^q |A \right]^{1/q}, \tag{8.26}$$

where

$$Q = \mathbb{E}\left[\left\| \left(\int_0^1 \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \,\mathrm{d}t \right)^{-1} \right\|^{q'} |A \right]^{1/q'}.$$

Employing the inequality $\mathbb{E}[Y|A] \le \mathbb{E}[Y]/\mathbb{P}(A)$, which holds for any positive random variable Y, point $(i)'$ in Proposition 8.12 and estimate (8.24), the second term in the right-hand side can be bounded by

$$\mathbb{E}\left[\left\| \mathbb{G}^J_M(\widehat{\theta}^J_{M,N}) - G^J_{M,N}(\widehat{\theta}^J_{M,N}) \right\|^q |A \right]^{1/q} \le \frac{C}{\sqrt{N}} \left(\frac{1}{1 - C\sqrt{\frac{M}{N}}} \right)^{1/q} \le \frac{C}{\sqrt{N}}, \tag{8.27}$$

where the last inequality is justified by the fact that $M \ll N$ and by changing the value of the constant C. We now have to bound the first term Q in the right-hand side of equation (8.26). Employing the inequality $\left\| M^{-1} \right\| \le \left\| M \right\|^{p-1} / \left| \det(M) \right|$, which holds for any square nonsingular matrix $M \in \mathbb{R}^{L \times L}$, we have

$$Q \le \mathbb{E}\left[\frac{\left\| \int_0^1 \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \,\mathrm{d}t \right\|^{q'(p-1)}}{\left| \det\left(\int_0^1 \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \,\mathrm{d}t \right) \right|^{q'}} |A \right].$$

Since we are conditioning on the event A, by the first part of the proof, we know that

$$\left\| \widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M \right\| \le \delta,$$

and, by taking ε sufficiently small, we can always find δ small enough, but still finite, such that the absolute value of the determinant in the denominator is lower bounded by a constant independent of M and N because $\det(\mathbb{H}^J_M(\widehat{\vartheta}^J_M)) \neq 0$ and by (8.23) it converges in probability to $\det(\mathscr{H}^J(\theta_0))$, which is invertible. Hence, applying Jensen's inequality we obtain

$$\begin{aligned} Q &\le C\,\mathbb{E}\left[\left\| \int_0^1 \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \,\mathrm{d}t \right\|^{q'(p-1)} |A \right] \\ &\le C\,\mathbb{E}\left[\int_0^1 \left\| \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \right\|^{q'(p-1)} \mathrm{d}t |A \right], \end{aligned}$$

which due to Lemma 8.11, Remark 8.14, the property $\mathbb{E}[Y|A] \le \mathbb{E}[Y]/\mathbb{P}(A)$, which holds for any positive random variable Y, and estimate (8.24) yields

$$Q \le \frac{C}{\mathbb{P}(A)} \int_0^1 \mathbb{E}\left[\left\| \mathbb{H}^J_M(\widehat{\vartheta}^J_M + t(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M)) \right\|^{q'(p-1)} \right] \mathrm{d}t \le C,$$

which together with equations (8.25), (8.26) and (8.27) gives the desired result. □

The results of this section are summarized in the following diagram

$$
\begin{tikzcd}
 & \widehat{\vartheta}^J_M \arrow[dr, "in\ \mathbb{P}", "M \to \infty"'] & \\
\widehat{\theta}^J_{M,N} \arrow[ur, "in\ L^1", "N \to \infty"'] \arrow[dr, "M \to \infty", "in\ \mathbb{P}"'] & & \theta_0 \\
 & \vartheta^J_N \arrow[ur, "N \to \infty"] &
\end{tikzcd}
$$

where $\mathbb{P}$ stands for convergence in probability.

Remark 8.19. All the previous results only prove the existence of such estimators with high probability and do not guarantee their uniqueness. However, as we will see in the next section, any of these estimators converge to the exact value of the unknown.

8.4.3 Proof of the main theorems

In this section we finally present the proofs of the main results of this chapter, i.e., Theorems 8.6 to 8.8.

Proof of Theorem 8.6. First, by Lemma 8.17 we deduce the existence of $N_0 > 0$ such that for all $N > N_0$ the estimator $\widehat{\theta}^J_{M,N}$ exists with a probability tending to one as M goes to infinity. Then, we prove separately equations (8.7), (8.8) and (8.9).

Proof of (8.7)**.** By Lemmas 8.15 and 8.17 we have

$$\lim_{N\to\infty} \lim_{M\to\infty} \widehat{\theta}^J_{M,N} = \lim_{N\to\infty} \vartheta^J_N = \theta_0, \qquad \text{in probability,}$$

which proves (8.7).

Proof of (8.8)**.** By Lemma 8.18 the estimator $\widehat{\theta}^J_{M,N}$ converges to $\widehat{\vartheta}^J_M$ in L^1 as N goes to infinity and hence in probability. Therefore, applying Lemma 8.16 we obtain

$$\lim_{M\to\infty} \lim_{N\to\infty} \widehat{\theta}^J_{M,N} = \lim_{M\to\infty} \widehat{\vartheta}^J_M = \theta_0, \qquad \text{in probability,}$$

which shows (8.8).

Proof of (8.9)**.** We introduce the following decomposition

$$\widehat{\theta}^J_{M,N} - \theta_0 = (\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M) + (\widehat{\vartheta}^J_M - \theta_0) =: Q_1 + Q_2,$$

where $\widehat{\vartheta}^J_M$ is defined in Lemma 8.16 and due to Lemma 8.18 the first quantity satisfies

$$\mathbb{E}\left[\|Q_1\|\right] \leq C\sqrt{\frac{M}{N}}, \tag{8.28}$$

with the constant C independent of M and N. Therefore, since $M = o(N)$, estimate (8.28) together with Lemma 8.16 and the fact that convergence in L^1 implies convergence in probability gives the desired result (8.9) and ends the proof. □

Proof of Theorem 8.7. The existence of the estimator $\widehat{\theta}^J_{M,N}$ is given by Theorem 8.6. Then, we prove separately equations (8.10), (8.11) and (8.12).

Proof of (8.10)**.** Let ϑ_N be defined in Lemma 8.15. Using basic properties of probability measures we have

$$\begin{aligned}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) &= \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \theta_0\right\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right) K_\varepsilon\right) \\ &\le \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| + \|\vartheta_N - \theta_0\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right) K_\varepsilon\right),\end{aligned} \tag{8.29}$$

which implies

$$\begin{aligned}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) &\le \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right) \frac{K_\varepsilon}{2}\right) \\ &\quad + \mathbb{P}\left(\|\vartheta_N - \theta_0\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right) \frac{K_\varepsilon}{2}\right) \\ &\le \mathbb{P}\left(\sqrt{M}\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| > \frac{K_\varepsilon}{2}\right) + \mathbb{P}\left(\|\vartheta_N - \theta_0\| > \frac{K_\varepsilon}{2\sqrt{N}}\right),\end{aligned}$$

and we now study two terms in the right-hand side separately. First, letting M and N go to infinity by Lemma 8.17 we obtain

$$\lim_{N\to\infty}\lim_{M\to\infty} \mathbb{P}\left(\sqrt{M}\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| > \frac{K_\varepsilon}{2}\right) = \mathbb{P}\left(\left\|\Lambda^J\right\| > \frac{K_\varepsilon}{2}\right),$$

where the right-hand side can be made arbitrarily small by taking $K_\varepsilon > 0$ sufficiently large. Moreover, we have

$$\mathbb{P}\left(\|\vartheta_N - \theta_0\| > \frac{K_\varepsilon}{2\sqrt{N}}\right) = \mathbb{E}\left[\mathbb{1}_{\left\{\|\vartheta_N - \theta_0\| > \frac{K_\varepsilon}{2\sqrt{N}}\right\}}\right],$$

where the right-hand side is identically equal to zero if we set $K_\varepsilon > 2C$, where the constant C is given by Lemma 8.15. Hence, for all $\varepsilon > 0$ we can take $K_\varepsilon > 0$ sufficiently large such that

$$\lim_{N\to\infty}\lim_{M\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon,$$

which proves (8.10).

Proof of (8.11)**.** Let $\widehat{\vartheta}_M$ be defined in Lemma 8.16. Repeating a procedure similar to (8.29) and applying Markov's inequality we get

$$\begin{aligned}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) &\le \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}_M\right\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right) \frac{K_\varepsilon}{2}\right) + \mathbb{P}\left(\sqrt{M}\left\|\widehat{\vartheta}_M - \theta_0\right\| > \frac{K_\varepsilon}{2}\right) \\ &\le \frac{2\sqrt{MN}}{K_\varepsilon(\sqrt{M} + \sqrt{N})} \mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}_M\right\|\right] + \mathbb{P}\left(\sqrt{M}\left\|\widehat{\vartheta}_M - \theta_0\right\| > \frac{K_\varepsilon}{2}\right),\end{aligned}$$

and we now study two terms in the right-hand side separately. First, by Lemma 8.16 we have

$$\lim_{M\to\infty} \mathbb{P}\left(\sqrt{M}\left\|\widehat{\vartheta}_M - \theta_0\right\| > \frac{K_\varepsilon}{2}\right) = \mathbb{P}\left(\left\|\Lambda^J\right\| > \frac{K_\varepsilon}{2}\right),$$

where the right-hand side can be made arbitrarily small by taking $K_\varepsilon > 0$ sufficiently large. Moreover, by Lemma 8.18 we have

$$\frac{2\sqrt{MN}}{K_\varepsilon(\sqrt{M} + \sqrt{N})} \mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}_M\right\|\right] \le \frac{2CM}{K_\varepsilon(\sqrt{M} + \sqrt{N})}, \tag{8.30}$$

where the constant C is independent of M and N. Hence, for all $\varepsilon > 0$ we can take $K_\varepsilon > 0$ sufficiently large such that

$$\lim_{M\to\infty} \lim_{N\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon,$$

which shows (8.11).

Proof of (8.12)**.** Equation (8.12) is obtained following verbatim the proof of (8.11) in the previous step and using the fact that $M = o(\sqrt{N})$ to show that the right-hand side in equation (8.30) vanishes. □

Proof of Theorem 8.8. The existence of the estimator $\widehat{\theta}^J_{M,N}$ is given by Theorem 8.6. Then, let us introduce the following decomposition

$$\sqrt{M}\left(\widehat{\theta}^J_{M,N} - \theta_0\right) = \sqrt{M}\left(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right) + \sqrt{M}\left(\widehat{\vartheta}^J_M - \theta_0\right),$$

where $\widehat{\vartheta}^J_M$ is defined in Lemma 8.16. We now study two terms in the right-hand side separately. By Lemma 8.18 we have

$$\sqrt{M}\,\mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\|\right] \le C\frac{M}{\sqrt{N}},$$

where the constant C is independent of M and N, hence since $M = o(\sqrt{N})$ by hypothesis we obtain

$$\lim_{M,N\to\infty} \sqrt{M}\left(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right) = 0, \qquad \text{in probability.} \tag{8.31}$$

Moreover, by Lemma 8.16 we know that

$$\lim_{M\to\infty} \sqrt{M}\left(\widehat{\vartheta}^J_M - \theta_0\right) = \Lambda^J \sim \mathcal{N}(0, \Gamma^J_0), \qquad \text{in distribution,} \tag{8.32}$$

where the covariance matrix Γ^J_0 is defined in (8.13). Finally, limits (8.31) and (8.32) together with Slutsky's theorem imply the desired result. □

8.5 Conclusion

In this chapter we considered inference problems for large systems of exchangeable interacting particles. When the number of particles is large, then the path of a single particle is well approximated by its mean field limit. The limiting mean field SDE is on the one hand more complex because it is a nonlinear SDE (in the sense of McKean), but on the other hand more tractable from a computational viewpoint as it reduces an N-dimensional SDE to a one dimensional one. Our aim was to infer unknown parameters of the dynamics, in particular of the confining and interaction potentials, from a set of discrete observations of a single particle. We propose a novel estimator which is obtained by computing the zero of a martingale estimating function based on the eigenvalues and the eigenfunctions of the generator of the mean field limit, where the law of the process is replaced by the (unique) invariant measure of the mean field dynamics. We showed both theoretically and numerically the asymptotic unbiasedness and normality of our estimator in the limit of infinite data and particles, providing also a rate of convergence towards the true value of the unknown parameter. In particular, we observed that these properties hold true if the number of particles is much larger than the number of observations. Even though our theoretical results require uniqueness of the steady state for the mean field dynamics, our numerical experiments suggest that our method works well even when phase transitions are present, i.e., when there are more than one stationary states. Moreover, we compared our estimator with the MLE, demonstrating that our approach is more robust with respect to small values of the sampling rate. We believe, therefore, that the inference methodology proposed and analyzed in this chapter can be very efficient when learning parameters in mean field SDE models from data.

Future perspectives

In this last section we present possible directions of future research related to the problems studied in this thesis. In the first part, and in particular in Chapters 2 and 3, we introduced filtered data to tackle the problem of model misspecification when we aim to fit homogenized dynamics from multiscale continuous-time data. We believe that this approach opens the way to several further developments. In particular, we think it would be relevant to:

(i) Analyse the exponential filter for $\beta > 1$ in (2.1), which seems to provide more robust results in practice, and study different kinds of filters which do not belong to the exponential family or are not moving averages.

(ii) Consider multiscale stochastic differential equations for which the homogenized equations present non-constant diffusion terms (e.g., multiplicative noise), or drift functions which do not depend linearly on the parameters.

(iii) Extend the analysis to the non-parametric framework most likely by means of Bayesian regularization techniques, thus allowing to recover whole effective functions.

(iv) Derive asymptotically unbiased estimators for the diffusion coefficient which are robust in practice and do not rely on the drift estimator.

(v) Apply similar methodologies to correct faulty behavior of other methods.

In the second part of the thesis we considered eigenfunction estimators for inferring unknown parameters in effective models given discrete-time observations from multiscale and interacting diffusions. We first remark that Chapter 6 provides rigorous homogenization results for the eigenpairs of the generator in the setting of the Langevin dynamics, but we believe that similar theorems can be proved for more general classes of multiscale diffusion processes. Regarding the application of eigenfunction estimators to multiscale diffusions in Chapter 7, in order to be able to assess the accuracy, it would be interesting to analyse its rate of convergence with respect to both the number of observations and the fastest scale. This is a highly nontrivial problem since it first requires the development of a fully quantitative periodic homogenization theory. On the other hand, the work presented in Chapter 8 about the application of eigenfunction estimators to interacting particle systems can be extended in several other interesting directions. First, the main limitation of our methodology is the fact that in order to construct the martingale estimating function we have to know the functional form of the invariant measure of the mean field limit, possibly parameterized in terms of a finite number of moments. There are many examples of mean field partial differential equations where the self-consistency equation cannot be solved analytically or, at least, its solution depends on the unknown parameters in the model. Therefore, it would be interesting to lift this assumption by first learning the invariant measure from data and then applying our approach. This leads naturally to our second objective, namely the extension of our methodology to a nonparametric setting, i.e., when the functional form of the confining and interaction potentials are unknown. Thirdly, we want to obtain more detailed information on the computational complexity of the proposed algorithm, in particular when more eigenfunctions are needed for our martingale estimator and when we are in higher dimensions in space. Finally, it would be interesting to mix the two models under investigation and study the problem of parameter estimation for multiscale interacting particle systems for which we have to combine the homogenization and mean field limits.

Bibliography

[1] A. Abdulle and A. Di Blasio, *Numerical homogenization and model order reduction for multiscale inverse problems*, Multiscale Model. Simul., 17 (2019), pp. 399–433.

[2] 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.

[3] A. Abdulle, G. Garegnani, G. A. Pavliotis, A. M. Stuart, and A. Zanoni, *Drift estimation of multiscale diffusions based on filtered data*, Found. Comput. Math., (2021).

[4] A. Abdulle, G. Garegnani, and A. Zanoni, *Ensemble Kalman Filter for Multiscale Inverse Problems*, Multiscale Model. Simul., 18 (2020), pp. 1565–1594.

[5] A. Abdulle, G. A. Pavliotis, and U. Vaes, *Spectral methods for multiscale stochastic differential equations*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 720–761.

[6] A. Abdulle, G. A. Pavliotis, and A. Zanoni, *Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions*, Stat. Comput. (to appear), (2022).

[7] Y. Aït-Sahalia and J. Jacod, *High-frequency financial econometrics*, Princeton University Press, 2014.

[8] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang, *How often to sample a continuous-time process in the presence of market microstructure noise*, Rev. Financ. Stud, 18 (2005), pp. 351–416.

[9] G. Allaire, *Homogenization and two-scale convergence*, SIAM J. Math. Anal., 23 (1992), pp. 1482–1518.

[10] G. Allaire, *Two-scale convergence: a new method in periodic homogenization*, in Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XII (Paris, 1991–1993), vol. 302 of Pitman Res. Notes Math. Ser., Longman Sci. Tech., Harlow, 1994, pp. 1–14.

[11] G. Allaire, Y. Capdeboscq, A. Piatnitski, V. Siess, and M. Vanninathan, *Homogenization of periodic systems with large potentials*, Arch. Ration. Mech. Anal., 174 (2004), pp. 179–220.

[12] R. Azencott, A. Beri, A. Jain, and I. Timofeyev, *Sub-sampling and parametric estimation for multiscale dynamics*, Commun. Math. Sci., 11 (2013), pp. 939–970.

[13] R. Azencott, A. Beri, and I. Timofeyev, *Adaptive sub-sampling for parametric estimation of Gaussian diffusions*, J. Stat. Phys., 139 (2010), pp. 1066–1089.

[14] R. Azencott, A. Beri, and I. Timofeyev, *Parametric estimation of stationary stochastic processes under indirect observability*, J. Stat. Phys., 144 (2011), pp. 150–170.

[15] 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.

[16] 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.

[17] I. V. Basawa and B. L. S. Prakasa Rao, *Statistical inference for stochastic processes*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1980. Probability and Mathematical Statistics.

[18] S. Benachour, B. Roynette, D. Talay, and P. Vallois, *Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos*, Stochastic Process. Appl., 75 (1998), pp. 173–201.

[19] S. Benachour, B. Roynette, and P. Vallois, *Nonlinear self-stabilizing processes. II. Convergence to invariant probability*, Stochastic Process. Appl., 75 (1998), pp. 203–224.

[20] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam, 1978.

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

[22] J. Binney and S. Tremaine, *Galactic Dynamics*, Princeton University Press, Princeton, second ed., 2008.

[23] J. P. N. Bishwal, *Parameter estimation in stochastic differential equations*, vol. 1923 of Lecture Notes in Mathematics, Springer, Berlin, 2008.

[24] J. P. N. Bishwal, *Estimation in interacting diffusions: continuous and discrete sampling*, Appl. Math. (Irvine), 2 (2011), pp. 1154–1158.

[25] 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.

[26] D. Calvetti, M. Dunlop, E. Somersalo, and A. M. Stuart, *Iterative updating of model error for Bayesian inversion*, Inverse Problems, 34 (2018), pp. 025008, 38.

[27] D. Calvetti, O. Ernst, and E. Somersalo, *Dynamic updating of numerical model discrepancy using sequential sampling*, Inverse Problems, 30 (2014), pp. 114019, 19.

[28] J. A. Carrillo, R. S. Gvalani, G. A. Pavliotis, and A. Schlichting, *Long-time behaviour and phase transitions for the Mckean-Vlasov equation on the torus*, Arch. Ration. Mech. Anal., 235 (2020), pp. 635–690.

[29] X. Chen, *Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data*, 2021.

[30] D. Cioranescu and P. Donato, *An introduction to homogenization*, vol. 17 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, New York, 1999.

[31] M. Coghi, T. Nilssen, N. Nüsken, and S. Reich, *Rough Mckean–Vlasov dynamics for robust ensemble Kalman filtering.* arXiv preprint arXiv:2107.06621, 2022.

[32] C. J. Cotter and G. A. Pavliotis, *Estimating eddy diffusivities from noisy Lagrangian observations*, Commun. Math. Sci., 7 (2009), pp. 805–838.

[33] R. Courant and D. Hilbert, *Methods of mathematical physics. Vol. II: Partial differential equations*, (Vol. II by R. Courant.), Interscience Publishers (a division of John Wiley & Sons), New York-Lon don, 1962.

[34] 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.

[35] D. Crommelin and E. Vanden-Eijnden, *Reconstruction of diffusions using spectral data from timeseries*, Commun. Math. Sci., 4 (2006), pp. 651–668.

[36] D. Crommelin and E. Vanden-Eijnden, *Diffusion estimation from multiscale data by operator eigenpairs*, Multiscale Model. Simul., 9 (2011), pp. 1588–1623.

[37] M. Dashti and A. M. Stuart, *The Bayesian Approach to Inverse Problems*, in Handbook of Uncertainty Quantification, Springer, 2016, pp. 1–118.

[38] D. A. Dawson, *Critical dynamics and fluctuations for a mean-field model of cooperative behavior*, J. Statist. Phys., 31 (1983), pp. 29–85.

[39] F. Delarue and A. Tse, *Uniform in time weak propagation of chaos on the torus*, 2021.

[40] M. G. Delgadino, R. S. Gvalani, and G. A. Pavliotis, *On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions*, Arch. Ration. Mech. Anal., 241 (2021), pp. 91–148.

[41] J. Diehl, P. Friz, and H. Mai, *Pathwise stability of likelihood estimators for diffusions via rough paths*, Ann. Appl. Probab., 26 (2016), pp. 2169–2192.

[42] A. B. Duncan and G. A. Pavliotis, *Brownian motion in an n-scale periodic potential.* preprint arXiv:1605.05854, 2016.

[43] L. Ellam, M. Girolami, G. A. Pavliotis, and A. Wilson, *Stochastic modelling of urban structure*, Proc. A., 474 (2018), pp. 20170700, 20.

[44] L. C. Evans, *Partial differential equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 1998.

[45] D. Florens-Zmirou, *Approximate discrete-time schemes for statistics of diffusion processes*, Statistics, 20 (1989), pp. 547–557.

[46] T. D. Frank, *Delay Fokker–Planck equations, Novikov's theorem, and Boltzmann distributions as small delay approximations*, Phys. Rev. E, 72 (2005), p. 011112.

[47] S. Gailus and K. Spiliopoulos, *Statistical inference for perturbed multiscale dynamical systems*, Stochastic Process. Appl., 127 (2017), pp. 419–448.

[48] S. Gailus and K. Spiliopoulos, *Discrete-time statistical inference for multiscale diffusions*, Multiscale Model. Simul., 16 (2018), pp. 1824–1858.

[49] K. Gansberger, *An idea on proving weighted sobolev embeddings*, Preprint arXiv:1007.3525, (2010).

[50] A. Ganz Bustos, *Approximations des distributions d'équilibre de certains systèmes stochastiques avec interactions McKean-Vlasov*, PhD thesis, Nice, 2008.

[51] G. Garegnani and A. Zanoni, *Robust estimation of effective diffusions from multiscale data*, Commun. Math. Sci. (to appear), (2022).

[52] J. Garnier, G. Papanicolaou, and T.-W. Yang, *Consensus convergence with stochastic effects*, Vietnam J. Math., 45 (2017), pp. 51–75.

[53] J. Gärtner, *On the McKean-Vlasov limit for interacting diffusions*, Math. Nachr., 137 (1988), pp. 197–248.

[54] K. Giesecke, G. Schwenkler, and J. A. Sirignano, *Inference for large financial systems*, Math. Finance, 30 (2020), pp. 3–46.

[55] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations, Theory and Algorithms*, vol. 5 of Springer Series in Computational Mathematics, Springer-Verlag, 1986.

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

[57] B. D. Goddard, B. Gooding, H. Short, and G. A. Pavliotis, *Noisy bounded confidence models for opinion dynamics: the effect of boundary conditions on phase transitions*, IMA Journal of Applied Mathematics, (2021).

[58] F. Golse, *On the dynamics of large particle systems in the mean field limit*, in Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, vol. 3 of Lect. Notes Appl. Math. Mech., Springer, [Cham], 2016, pp. 1–144.

[59] 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.

[60] S. N. Gomes, A. M. Stuart, and M.-T. Wolfram, *Parameter estimation for macroscopic pedestrian dynamics models from microscopic data*, SIAM J. Appl. Math., 79 (2019), pp. 1475–1500.

[61] S. Guillouzic, I. L'Heureux, and A. Longtin, *Small delay approximation of stochastic delay differential equations*, Phys. Rev. E, 59 (1999), pp. 3970–3982.

[62] M. Hairer and G. A. Pavliotis, *Periodic homogenization for hypoelliptic diffusions*, J. Statist. Phys., 117 (2004), pp. 261–279.

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

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

[65] P. D. Hislop and I. M. Sigal, *Introduction to spectral theory*, vol. 113 of Applied Mathematical Sciences, Springer-Verlag, New York, 1996. With applications to Schrödinger operators.

[66] J. G. Hooton, *Compact sobolev imbeddings on finite measure spaces*, Journal of Mathematical Analysis and Applications, 83 (1981), pp. 570–581.

[67] L. Hurwicz and M. K. Richter, *Implicit functions and diffeomorphisms without C^1*, in Advances in mathematical economics. Vol. 5, vol. 5 of Adv. Math. Econ., Springer, Tokyo, 2003, pp. 65–96.

[68] 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.

[69] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.

[70] R. A. Kasonga, *Maximum likelihood theory for large interacting systems*, SIAM J. Appl. Math., 50 (1990), pp. 865–875.

[71] S. Kesavan, *Homogenization of elliptic eigenvalue problems. I*, Appl. Math. Optim., 5 (1979), pp. 153–167.

[72] S. Kesavan, *Homogenization of elliptic eigenvalue problems. II*, Appl. Math. Optim., 5 (1979), pp. 197–216.

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

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

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

[76] 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.

[77] Y. A. Kutoyants, *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London, Ltd., London, 2004.

[78] S. Lang, *Real and functional analysis*, vol. 142 of Graduate Texts in Mathematics, Springer-Verlag, New York, third ed., 1993.

[79] F. Legoll and T. Lelièvre, *Some remarks on free energy and coarse-graining*, in Numerical analysis of multiscale computations, vol. 82 of Lect. Notes Comput. Sci. Eng., Springer, Heidelberg, 2012, pp. 279–329.

[80] T. Lelièvre and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numer., 25 (2016), pp. 681–880.

[81] M. Liu and H. Qiao, *Parameter estimation of path-dependent mckean-vlasov stochastic differential equations*, 2020.

[82] S. A. M. Loos and S. H. L. Klapp, *Fokker–Planck equations for time-delayed systems via Markovian embedding*, J. Stat. Phys., 177 (2019), pp. 95–118.

[83] N. K. Mahato, A. Klar, and S. Tiwari, *Particle methods for multi-group pedestrian flow*, Appl. Math. Model., 53 (2018), pp. 447–461.

[84] F. Malrieu, *Logarithmic Sobolev inequalities for some nonlinear PDE's*, Stochastic Process. Appl., 95 (2001), pp. 109–132.

[85] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl., 101 (2002), pp. 185–232.

[86] B. Maury and S. Faure, *Crowds in equations*, Advanced Textbooks in Mathematics, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2019. An introduction to the microscopic modeling of crowds, With a foreword by Laure Saint-Raymond.

[87] A. Mielke, *On evolutionary Γ-convergence for gradient systems*, in Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, vol. 3 of Lect. Notes Appl. Math. Mech., Springer, [Cham], 2016, pp. 187–249.

[88] 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.

[89] S. Moskow and M. Vogelius, *First-order corrections to the homogenised eigenvalues of a periodic composite medium. a convergence proof*, Proc. Roy. Soc. Edinburgh, 127A (1997), pp. 1263–1299.

[90] P. Natalini and B. Palumbo, *Inequalities for the incomplete gamma function*, Math. Inequal. Appl., 3 (2000), pp. 69–77.

[91] G. Nguetseng, *A general convergence result for a functional related to the theory of homogenization*, SIAM J. Math. Anal., 20 (1989), pp. 608–623.

[92] J. Nolen, G. A. Pavliotis, and A. M. Stuart, *Multiscale modeling and inverse problems*, in Numerical analysis of multiscale problems, vol. 83 of Lect. Notes Comput. Sci. Eng., Springer, Heidelberg, 2012, pp. 1–34.

[93] K. Oelschläger, *A martingale approach to the law of large numbers for weakly interacting stochastic processes*, Ann. Probab., 12 (1984), pp. 458–479.

[94] S. C. Olhede, A. M. Sykulski, and G. A. Pavliotis, *Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise*, Multiscale Model. Simul., 8 (2010), pp. 393–427.

[95] C. D. Pagani and S. Salsa, *Analisi matematica - Volume 1*, Masson, 1995.

[96] A. Papanicolaou and K. Spiliopoulos, *Filtering the maximum likelihood for multiscale problems*, Multiscale Model. Simul., 12 (2014), pp. 1193–1229.

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

[98] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. I*, Ann. Probab., 29 (2001), pp. 1061–1085.

[99] E. Pardoux and A. Y. Veretennikov, *On Poisson equation and diffusion approximation. II*, Ann. Probab., 31 (2003), pp. 1166–1192.

[100] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. III*, Ann. Probab., 33 (2005), pp. 1111–1133.

[101] 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.

[102] G. A. Pavliotis, Y. Pokern, and A. M. Stuart, *Parameter estimation for multiscale diffusions: an overview*, in Statistical methods for stochastic differential equations, vol. 124 of Monogr. Statist. Appl. Probab., CRC Press, Boca Raton, FL, 2012, pp. 429–472.

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

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

[105] G. A. Pavliotis and A. Zanoni, *Eigenfunction martingale estimators for interacting particle systems and their mean field limit*, SIAM J. Appl. Dyn. Syst. (to appear), (2022).

[106] Y. Pokern, A. M. Stuart, and J. H. van Zanten, *Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs*, Stochastic Process. Appl., 123 (2013), pp. 603–628.

[107] Y. Pokern, A. M. Stuart, and E. Vanden-Eijnden, *Remarks on drift estimation for diffusion processes*, Multiscale Model. Simul., 8 (2009), pp. 69–95.

[108] B. L. S. Prakasa Rao, *Statistical inference for diffusion type processes*, vol. 8 of Kendall's Library of Statistics, Edward Arnold, London; Oxford University Press, New York, 1999.

[109] 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.

[110] M. Reed and B. Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[111] 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.

[112] L. Sharrock, N. Kantas, P. Parpas, and G. A. Pavliotis, *Parameter estimation for the mckean-vlasov stochastic differential equation*, 2021.

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

[114] G. Strang and G. J. Fix, *An analysis of the finite element method*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1973. Prentice-Hall Series in Automatic Computation.

[115] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[116] T. Suzuki, *Free energy and self-interacting particles*, vol. 62 of Progress in Nonlinear Differential Equations and their Applications, Birkhäuser Boston, Inc., Boston, MA, 2005.

[117] A.-S. Sznitman, *Topics in propagation of chaos*, in École d'Été de Probabilités de Saint-Flour XIX—1989, vol. 1464 of Lecture Notes in Math., Springer, Berlin, 1991, pp. 165–251.

[118] R. Temam, *Navier-Stokes equations*, vol. 2 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam-New York, revised ed., 1979. Theory and numerical analysis, With an appendix by F. Thomasset.

[119] J. Tugaut, *Phase transitions of McKean-Vlasov processes in double-wells landscape*, Stochastics, 86 (2014), pp. 257–284.

[120] A. Y. Veretennikov and A. M. Kulik, *The extended Poisson equation for weakly ergodic Markov processes*, Teor. Ĭmovīr. Mat. Stat., (2011), pp. 22–38.

[121] J. Wen, X. Wang, S. Mao, and X. Xiao, *Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application*, Appl. Math. Comput., 274 (2016), pp. 237–246.

[122] 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.

[123] Y. Ying, J. Maddison, and J. Vanneste, *Bayesian inference of ocean diffusivity from Lagrangian trajectory data*, Ocean Model., 140 (2019).

[124] N. Yoshida, *Phase transition from the viewpoint of relaxation phenomena*, Rev. Math. Phys., 15 (2003), pp. 765–788.

[125] A. Zanoni, *Homogenization results for the generator of multiscale Langevin dynamics in weighted Sobolev spaces.* arXiv preprint arXiv:2112.04921, 2021.

[126] 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.

[127] 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.

Curriculum Vitae

Personal data

Name	Andrea Zanoni
Date of birth	March 21, 1994
Nationality	Italian

Education

2019 - 2022	**PhD in Mathematics** École Polytechnique Fédérale de Lausanne, Switzerland Thesis advisors: Professors A. Abdulle and F. Nobile Thesis co-advisor: Professor G. A. Pavliotis
2016 - 2019	**MSc in Computational Science and Engineering** École Polytechnique Fédérale de Lausanne, Switzerland **MSc in Mathematical Engineering** Politecnico di Milano, Italy Thesis: Ensemble Kalman filter for multiscale inverse problems Advisors: Professors A. Abdulle and S. Salsa
2013 - 2016	**BSc in Mathematical Engineering** Politecnico di Milano, Italy Thesis: Il principio di concentrazione compattezza di P. L. Lions Advisor: Professor F. Gazzola
2008 - 2013	**Diploma** Liceo Scientifico Belfiore Matova, Italy

Work Experience

2018	**Internship in Machine Learning and Data Science** Basque Center of Applied Mathematics – Bilbao, Spain Supervisor: Doctor S. Mazuelas

PhD Publications

- G. A. Pavliotis and AZ, *Eigenfunction martingale estimators for interacting particle systems and their mean field limit*, SIAM J. Appl. Dyn. Syst. (to appear), (2022).

- G. Garegnani and AZ, *Robust estimation of effective diffusions from multiscale data*, Commun. Math. Sci. (to appear), (2022).

- A. Abdulle, G. A. Pavliotis, and AZ, *Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions*, Stat. Comput., 32 (2022), p. Paper No. 34.
- AZ, *Homogenization results for the generator of multiscale Langevin dynamics in weighted Sobolev spaces*, preprint arXiv:2112.04921, 2021.
- A. Abdulle, G. Garegnani, G. A. Pavliotis, A. M. Stuart, and AZ, *Drift estimation of multiscale diffusions based on filtered data*, Found. Comput. Math., (2021), pp. 1–52.

Other Publications

- R. Levi, F. Carli, A. R. Arévalo, Y. Altinel, D. J. Stein, M. M. Naldini, F. Grassi, AZ, S. Finkelstein, S. M. Vieira, J. Sousa, R. Barbieri, and L. A. Celi, *Artificial intelligence-based prediction of transfusion in the intensive care unit in patients with gastrointestinal bleeding*, BMJ Health & Care Informatics, 28 (2021).
- A. Abdulle, G. Garegnani, and AZ, *Ensemble Kalman filter for multiscale inverse problems*, Multiscale Model. Simul., 18 (2020), pp. 1565–1594.
- S. Mazuelas, AZ, and A. Pérez, *Minimax classification with 0-1 loss and performance guarantees*, in Advances in Neural Information Processing Systems, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds., vol. 33, Curran Associates, Inc., 2020, pp. 302–312.

Conferences and schools

- *Inference for systems of interacting particles and their mean field limit* (poster), Swiss Numerics Day 2022, Zurich (Switzerland), 12 September 2022.
- Summer School on *Mathematical modeling for epidemiology: analysis, simulation and forecasting*, Cetraro (Italy), 05 - 09 September 2022.
- *Eigenfunction martingale estimators for multiscale and interacting diffusions*, International Conference on Scientific Computation and Differential Equations (SciCADE), Reykjavík (Iceland), 25 - 29 July 2022.
- *Homogénéisation guidée par les données de l'équation de Langevin multi-échelle*, Congrès National d'Analyse Numérique (CANUM), Évian-les-Bains (France), 13 - 17 June 2022.
- *Data-driven homogenization of multiscale Langevin dynamics*, Hybrid SIAM Conference on Uncertainty Quantification (UQ), Atlanta (USA), 12 - 15 April 2022.
- *Inference of effective diffusions from multiscale data*, Swiss Numerics Day 2021, Lausanne (Switzerland), 13 September 2021.
- *Solution of multiscale inverse problems through filtering techniques and numerical homogenization*, 14th WCCM & ECCOMAS Congress 2020, Virtual Congress, 11 - 15 January 2021.

Academic visits

- Imperial College London, UK, February-March 2022
  Visit of one month to Professor Grigorios A. Pavliotis in the Department of Mathematics

Distinctions

- Vice President of the EPFL chapter of SIAM (2021 - 2022)
- Best project of *Milan Critical Care Datathon and ESICM's Big Datatalk 2019* organized by MIT, Humanitas Research Hospital and Politecnico di Milano
- Best freshmen of the academic year 2013/2014 at Politecnico di Milano
- Honorable mention at the Italian final phase of the Mathematical Olympiad 2012

Teaching

Co-supervised semester projects:

- Konstantin Medyanikov, *Generalized method of moments for learning unknown parameters in SDEs*, September-December 2022
- Max Hirsch, *Continuous time stochastic gradient descent for parameter identification in multiscale diffusion*, September-December 2022

Teaching Assistant for the courses:

- Analysis I, BSc at EPFL, September-December 2022
- Numerical Integration of Stochastic Differential Equations, MSc at EPFL, February-June 2022
- Numerical Integration of Dynamic Systems, MSc at EPFL, September-December 2021
- Analysis IV, BSc at EPFL, February-June 2021
- Numerical Integration of Dynamic Systems, MSc at EPFL, September-December 2020
- Analysis IV, BSc at EPFL, February-June 2020
- Analysis I, BSc at EPFL, September-December 2019
- Numerical Analysis, BSc at EPFL, February-June 2019
- Numerical Analysis, BSc at EPFL, February-June 2018
- Analysis I, BSc at EPFL, September-December 2017