Limit Theorems for Stochastic Differential Equations and Their Applications to Statistical Inference under Multiscale Observations

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

Dissertation

von

Jaroslav I. Borodavka

Tag der mündlichen Prüfung: 08.07.2026

Erster Referent: Prof. Dr. Sebastian Krumscheid
Zweiter Referent: Prof. Dr. Mathias Trabs
Dritter Referent: Prof. Dr. Carsten Hartmann

Manuscripts don't burn.

–

Mikhail A. Bulgakov, *The Master and Margarita*

Acknowledgments

Amid writing a thesis, it is easy to get carried away and start to think that it is an endeavor of a single individual, but, as many things in life, it is after all an effort of many people working together, whether directly or indirectly.

First of all, I would like to thank my academic supervisor, Sebastian Krumscheid, for his supportive guidance and ceaseless optimism towards my research. You were always very enthusiastic about discussing new research ideas with me or solving a tricky proof together. Although our meetings became less frequent over the years due to time constraints, you still somehow found time to talk about anything research-related and I personally have benefited immensely from your insights and your multifaceted view on mathematical research. You made me ask questions I have not asked before. I would also like to thank you for supporting my every so often desire to work on my weak points and at the same time polish my existing strong points. I always had the feeling that you are on my side no matter what and you managed to do that without losing your scientific objectivity. For all these reasons, I consider myself privileged to have had a supervisor like you, so thank you.

I am grateful for receiving financial support from the German Research Foundation (DFG) through the CRC 1481 "Sparsity and Singular Structures", which made this research possible. I further thank Mathias Trabs and Carsten Hartmann for agreeing to review this thesis and be referees on my PhD committee.

I would also like to thank all my colleagues from the UQ team and all the researchers from the CRC 1481 for making the PhD journey so much more enjoyable. The countless team-building events, relaxing lunches and dinners, the communication of research through presentations, seminars, and occasional meetings in each other's offices, and the many laughs we have shared over this time were truly enriching. Special thanks go to Maximilian Kruse for being the funniest office mate at KIT and from whom I have learned a lot about good software practices. I would also like to thank Benjamin Berkels for his mentoring and his various insights surrounding my whole PhD. Thanks are also due to my other collaborators, Bruno Ebner, Grigorios Pavliotis, Max Hirsch, and Andrea Zanoni. You made scientific collaboration easy and I learned a lot from your expertise.

Last but certainly not least, I would like to express my deepest and most sincere gratitude to my family and friends. My parents like to think that my life's accomplishments have all been due to me, but the truth is that I would have never come this far without them and their everlasting support. On the same note, I thank my brothers for being the best brothers a youngest brother can wish for; thank you, Ivan, for always showing genuine interest towards my research and thank you, Dennis, for always making me laugh with your sarcastic humor. Truly heartfelt thanks go to my fiancée, Lily, for always believing in me. Your emotional support has given me the mental fortitude to push through and persevere even in moments of utter self-doubt and self-reproach. Finally, I would like to thank my friends, John and Shara, for being exceptional friends with whom I shared many fun moments, giving me an ideal balance to the busy PhD studies. From the bottom of my heart, I thank all of you.

Abstract

We study the problem of statistical estimation under model misspecification, that is, when there exists a discrepancy between the data-generating model and the actual model of interest. Our framework is set in the mathematical theory of multiscale systems and homogenization for stochastic differential equations and is substantially driven by ergodic properties of the underlying stochastic processes. The discrepancy is of a weak kind: the stochastic process of the data-generating model, the multiscale model, converges weakly, in the sense of probability measures, to the stochastic process of the model of interest, the homogenized model. Despite this weak model misspecification, many commonly used statistical methodologies fail to obtain reasonable estimates for quantities of the homogenized model when they are confronted with multiscale structure in the data. In this thesis, we adopt a new perspective on classical limit theorems for stochastic differential equations, provide numerous extensions of those, including but not limited to mean ergodic theorems and central limit theorems, and propose novel statistical estimators in parametric and nonparametric estimation problems under weak misspecification. We not only demonstrate the effectiveness of the limit theorems in establishing various asymptotic properties of our proposed estimators, but we also substantiate many of the theoretical findings through extensive numerical simulations.

Zusammenfassung

Wir untersuchen das Problem der statistischen Schätzung unter dem Aspekt der Modellfehlspezifikation, das heißt, wenn eine Diskrepanz zwischen dem datenerzeugenden Modell und dem tatsächlich interessierenden Modell besteht. Unser Ansatz ist in der mathematischen Theorie multiskaliger Systeme sowie der Homogenisierung stochastischer Differentialgleichungen verankert und wird maßgeblich durch ergodische Eigenschaften der zugrundeliegenden stochastischen Prozesse bestimmt. Die betrachtete Diskrepanz ist von einer schwachen Art: Der stochastische Prozess des datenerzeugenden Multiskalensystems konvergiert schwach, im Sinne von Wahrscheinlichkeitsmaßen, gegen den stochastischen Prozess des interessierenden homogenisierten Modells. Trotz dieser schwachen Modellfehlspezifikation versagen viele bewährte statistische Verfahren, wenn sie auf Daten mit multiskaliger Struktur angewendet werden, und liefern keine zuverlässigen Schätzwerte für Größen des homogenisierten Modells. In dieser Arbeit nehmen wir eine neue Perspektive bezüglich klassischer Grenzwertsätze für stochastische Differentialgleichungen ein, geben zahlreiche Verallgemeinerungen an, darunter insbesondere Ergodensätze sowie zentrale Grenzwertsätze, und schlagen neuartige statistische Schätzer in parametrischen und nichtparametrischen Schätzproblemen unter schwacher Modellfehlspezifikation vor. Wir zeigen nicht nur, dass sich die Grenzwertsätze effektiv zur Herleitung verschiedener asymptotischer Eigenschaften der vorgeschlagenen Schätzer einsetzen lassen, sondern untermauern viele der theoretischen Resultate auch durch umfangreiche numerische Simulationen.

Аннотация

Мы исследуем задачу статистического оценивания в условиях некорректной спецификации модели, то есть, если существует расхождение между моделью, порождающей данные, и в самом деле интересующей моделью. Наш подход основан на математической теории многомасштабных систем и гомогенизации для стохастических дифференциальных уравнений и значительно опирается на эргодические свойства соответствующих стохастических процессов. Рассматриваемое расхождение носит слабый характер: стохастический процесс многомасштабной модели, порождающей данные, сходится слабо, в смысле мер вероятности, к стохастическому процессу гомогенизированной модели интереса. Несмотря на такую слабую некорректную спецификацию модели, многие общепринятые статистические методы не позволяют получать удовлетворительные оценки величин гомогенизированной модели при наличии в данных многомасштабной структуры. В данной работе мы представляем новый взгляд на классические предельные теоремы для стохастических дифференциальных уравнений, приводим многочисленные их обобщения, включая, помимо прочего, эргодические теоремы среднего рода и центральные предельные теоремы, а также разрабатываем новые статистические оценки в параметрических и непараметрических задачах оценивания при слабой некорректной спецификации модели. Мы не только демонстрируем эффективность предельных теорем для установления различных асимптотических свойств предложенных оценок, но и подтверждаем многие теоретические результаты с помощью обширной численной симуляции.

초록

본 논문은 데이터 생성 모델과 실제 관심 모델 사이에 불일치가 존재하는 경우, 즉 모델 오지정(model misspecification) 하에서의 통계적 추정 문제를 연구한다. 본 연구의 틀은 확률 미분 방정식에 대한 다중 규모 시스템(multiscale systems) 및 균질화(homogenization)의 수학적 이론 위에 설정되며, 그 근저에 있는 확률 과정의 에르고딕(ergodic) 성질에 의해 본질적으로 구동된다. 이 불일치는 약한 의미에서의 것으로, 데이터 생성 모델인 다중 규모 모델의 확률 과정이 확률 측도의 의미에서 약하게(weakly), 즉 약수렴(weak convergence)의 의미로, 관심 모델인 균질화 모델의 확률 과정으로 수렴한다. 이러한 약한 모델 오지정에도 불구하고, 널리 사용되는 많은 통계적 방법론들은 데이터에 다중 규모 구조가 내재되어 있을 때 균질화 모델의 모수에 대한 합리적인 추정값을 얻는 데 실패한다. 본 학위논문에서는 확률 미분 방정식에 대한 고전적 극한 정리를 새로운 관점에서 재조명하고, 평균 에르고딕 정리(mean ergodic theorem) 및 중심 극한 정리(central limit theorem)를 포함하는 다양한 확장을 제시하며, 약한 오지정 하에서의 모수적(parametric) 및 비모수적(nonparametric) 추정 문제에 대한 새로운 통계적 추정량(estimator)을 제안한다. 제안된 추정량의 다양한 점근적(asymptotic) 성질을 확립하는 데 있어 극한 정리의 유효성을 입증할 뿐만 아니라, 광범위한 수치 시뮬레이션을 통해 많은 이론적 결과를 실증적으로 뒷받침한다.

Notation

This is a list of frequently appearing symbols, conventions, and abbreviations, without any claim to completeness.

Numbers, vectors and matrices

$\mathbb{R}^d$	d-dimensional Euclidean space with vectors understood as column vectors
I_d	identity matrix in $\mathbb{R}^{d\times d}$
$\top$	transpose of a vector or matrix
$\lvert\cdot\rvert_2$	Euclidean norm on $\mathbb{R}^d$ or $\mathbb{C}^d$
$\lvert\cdot\rvert$	absolute value on $\mathbb{R}$ or $\mathbb{C}$
$a \wedge b$	$\min\{a,b\}$
$a \vee b$	$\max\{a,b\}$

Functions

∇f	gradient of f understood as a row vector
$\nabla_x f(x,y)$	denotes $(\partial_{x_1} f(x,y), \ldots, \partial_{x_n} f(x,y))$
∂^β	higher-order partial derivative for a multi-index $\beta \in \mathbb{N}_0$
$C(X;Y)$	space of continuous functions defined on metric space X with values in metric space Y
$C(X)$	space of real-valued continuous functions defined on metric space X
$C^\infty(X)$	space of infinitely differentiable real-valued functions defined on $X \in \{\mathbb{R}^d, \mathbb{T}^d\}$
$\mathscr{S}(\mathbb{R}^d)$	space of real-valued Schwartz functions defined on $\mathbb{R}^d$
$W^{k,1}(\mathbb{R}^d)$	Sobolev space of real-valued functions defined on $\mathbb{R}^d$ whose weak derivatives are integrable up to order k
$\mathscr{F}$	Fourier transform operator $(\mathscr{F} g)(\xi) = \int_{\mathbb{R}} g(x) e^{-2\pi i \xi x}\, dx$

Probability

i.i.d.	independent and identically distributed
$\sim$	distributional equivalence
$\mathcal{N}_d(m, \Sigma)$	d-dimensional normal distribution with mean $m \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$
$L^p(\Omega, \mathcal{F}, \mathbb{P})$	space of p-integrable random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$
$\mathscr{M}_2^{\mathrm{c}}$	space of continuous square-integrable martingales starting in zero
$\mathscr{M}^{\mathrm{c,\,loc}}$	space of continuous local martingales starting in zero
$\langle X \rangle$	quadratic variation process of a continuous local martingale $X \in \mathscr{M}^{\mathrm{c,\,loc}}$
$\xrightarrow{L^2}$	convergence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$
$\xrightarrow{\mathbb{P}}$	convergence in probability
$\xrightarrow{\mathcal{D}}$	weak convergence

Metric spaces

Let A, B be subsets of a metric space X with metric d and $x \in X$.

A°	interior of A
$\overline{A}$	closure of A
∂A	boundary of A
$\mathrm{dist}(x, A)$	$\inf_{y \in A} d(x, y)$
$\mathrm{dist}(A, B)$	$\inf_{x \in A, y \in B} d(x, y)$
$\mathrm{diam}(A)$	$\sup_{x, y \in A} d(x, y)$
$B_\delta(x)$	open ball around x with radius $\delta > 0$
$\mathscr{B}(X)$	Borel σ-algebra of X

Landau symbols

$f = \mathcal{O}(g)$	f grows at most as fast as g
$f = \Omega(g)$	f grows at least as fast as g
$f = o(g)$	f grows slower than g
$X_n = o_{\mathbb{P}}(1)$	$X_n \xrightarrow{\mathbb{P}} 0$

Miscellaneous

sgn	left-continuous version of sign function
supp	support of a function or a measure
δ_{ij}	Kronecker delta

Acronyms

C	coefficients
CLT	central limit theorem
FEM	finite element method
L	Langevin
LHS	left hand side
MDE	minimum distance estimator
MET	mean ergodic theorem
MLE	maximum likelihood estimator
MSE	mean squared error
ODE	ordinary differential equation
ONB	orthonormal basis
PDE	partial differential equation
QET	quantitative ergodic theorem
R	recurrence
RHS	right hand side
SDE	stochastic differential equation

Table of Contents

Acknowledgments **i**

Abstract (English/Deutsch/Русский/한국어) **iii**

Notation **xi**

List of Figures **xv**

List of Tables **xvii**

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Homogenization for SDEs . . . 3
- 1.3 Related Work . . . 7
- 1.4 Contributions and Outline of the Thesis . . . 9

2 Limit Theorems **11**
- 2.1 Ergodic Theorems . . . 12
  - 2.1.1 A Qualitative ε-Dependent Mean Ergodic Theorem . . . 14
  - 2.1.2 A Quantitative ε-Dependent Mean Ergodic Theorem . . . 36
- 2.2 Asymptotic Vanishing of Tail Probabilities . . . 47
- 2.3 An ε-Dependent Central Limit Theorem . . . 54
- 2.4 Multiscale Overdamped Langevin Equation . . . 58
  - 2.4.1 Multidimensional Case . . . 58
  - 2.4.2 One-Dimensional Case . . . 65

3 Parametric Estimation **77**
- 3.1 A Primer on Parametric Inference under Weak Misspecification . . . 77
  - 3.1.1 Example – Maximum Likelihood Estimation . . . 79
  - 3.1.2 Example – Minimum Distance Estimation . . . 81
- 3.2 A General Minimum Distance Estimator Approach . . . 84
  - 3.2.1 Problem Setting and Approach . . . 84
  - 3.2.2 Asymptotic Properties . . . 88
  - 3.2.3 Multiscale Overdamped Langevin Equation . . . 97
  - 3.2.4 Numerical Experiments . . . 102

4 Nonparametric Estimation **115**
- 4.1 The Density of Interest . . . 115
- 4.2 Density Estimation by Hermite Expansion . . . 116
- 4.3 Estimator Robustness and Asymptotic Normality . . . 118
- 4.4 Numerical Illustration . . . 133
  - 4.4.1 General Implementation Details . . . 133
  - 4.4.2 Robustness . . . 134

4.4.3 Asymptotic Normality 136
4.4.4 Inference of the Scale Separation Parameter 141

5 Concluding Remarks and Outlook 145

A Stochastic Calculus 149
A.1 Martingales, Quadratic Variation, and Local Time 149
A.2 Markov Processes and Strong SDE Solutions 152

B Weak Convergence of Measures 157
B.1 General Metric Spaces 157
B.2 Continuous Functions on $\mathbb{R}^d$ 159
B.3 Separable Hilbert Spaces 160

C Interior Estimates of the Schauder Type 161

Bibliography 167

List of Figures

1.1 Sample paths of a multiscale SDE system with slow-scale and fast-scale variable and their corresponding homogenized limit SDE process 4

3.1 Estimation results of the MLE under model specification and model misspecification 80
3.2 Robustness results of the MDE for a multiscale overdamped Langevin equation with quadratic potential and sine oscillation 105
3.3 Robustness results of the MDE for a multiscale overdamped Langevin equation with quartic potential and sine oscillation 106
3.4 Asymptotic normality results of the MDE for a multiscale overdamped Langevin equation with quadratic potential and sine oscillation 107
3.5 Robustness results of the MDE for a 2-dimensional multiscale overdamped Langevin equation with quadratic potential and additive sine oscillations . 108
3.6 Robustness results of the MDE for a bilinearly coupled fast-slow SDE system with Ornstein–Uhlenbeck limit 113

4.1 Robustness results of the projection-based Hermite density estimator across varying values of the observation time and number of Fourier coefficients for a 1-dimensional multiscale overdamped Langevin equation . . . 135
4.2 Robustness results of the projection-based Hermite density estimator extended to a 2-dimensional setting for a 2-dimensional multiscale overdamped Langevin equation 136
4.3 Sample paths of the centered and scaled projection-based Hermite density estimator and its centered Gaussian limit element in $L^2(\mathbb{R})$ 137
4.4 Robustness results of the projection-based Hermite density estimator for 1-dimensional multiscale overdamped Langevin equations using multiple estimates 138
4.5 Asymptotic normality results of the Fourier coefficients of the centered and scaled projection-based Hermite density estimator for a 1-dimensional multiscale overdamped Langevin equation I 139
4.6 Asymptotic normality results of the Fourier coefficients of the centered and scaled projection-based Hermite density estimator for a 1-dimensional multiscale overdamped Langevin equation II 140
4.7 Estimation results of a Fourier-based inference method for the scale separation parameter 142

List of Tables

3.1 Robustness results of the MDE for a chaotic ODE with a fast component that solves the Lorenz equation . . . 111

C.1 Tracking of a constant in a priori interior estimates of the Schauder type for a general parabolic PDE . . . 164

Chapter 1

Introduction

1.1 Motivation

In the classical statistics literature, the topic of estimation is one of the main building blocks of statistical inference for stochastic processes. It is a recurring theme to study such inference tasks in a specified model setting, that is, when the data is assumed to come from one particular family of distributions, see, for example, [IK81; Kut04; LS00; Pra99; Kut94]. This setting undoubtedly holds up relevant, but is often violated in applications where a mismatch between data and model is common. One prominent example of such a mismatch exists in the realm of multiscale modeling where complicated dynamical systems, characterized by processes evolving on multiple time scales, are approximated by a surrogate model capturing the effective dynamics of the full system. This approximation not only achieves a substantial reduction in complexity and dimensionality but also, in certain contexts that we will introduce later, a gain in structure. When using a reduction method, such as scale separation, to derive a surrogate model, then this is commonly known as coarse-graining; the surrogate model is thus called coarse-grained model, and the rigorous, mathematical multiscale analysis is often termed averaging or homogenization, see [BLP78; FW12; PS08; RX21].

Stochastic systems with multiscale structure arise naturally in a wide range of applications across the physical, biological, financial, and engineering sciences. It is therefore not surprising that the method of deriving a surrogate model is an established practice in such fields of study; for example, in econometrics when dealing with high-frequency financial data and market microstructure [Tsa10], in network traffic data [Abr+02], in molecular dynamics when identifying reaction coordinates and collective variables [Aze+13; Sch10], or in the study of intracellular biochemical reaction networks involving multiple reactions and chemical species [Bal+06; KRK17]. A comprehensive overview and additional examples of applications can be found in the books [Kue15; PS08].

However, even after having derived a coarse-grained model, it is usually not possible to give analytical closed-form expressions for the relevant model quantities. One promising way to still obtain coarse-grained model quantities in such a situation is to resort to data, which, however, is compatible with the coarse-grained model only at the appropriate macroscopic length and time scales. This is a particular setting of misspecification between data and model, cf. [Kut04, Section 2.6.1]. Apart from the mathematical interest, misspecification can be practically relevant for two main reasons. One reason is the imperfect knowledge of the data-generating model. After all, a coarse-grained model may very well approximate several different data-generating models, which are unknown in general, so that there is an identifiability issue from the coarse-grained model to the possible set of data-generating models. The second reason is that, even if there is perfect knowledge of the data-generating model, then the computational complexity for simulation or estimation can potentially be so high that it becomes infeasible or virtually impossible to work with the data-generating model. Hence, from a practitioner's point of view, it may be desirable to work with a simpler model.

In this work, we explore statistical estimation for misspecified models where a sequence of stochastic processes, assumed to come from the data-generating model and characterized by a small-scale parameter, converges weakly, in the sense of probability measures, to a stochastic process that describes the coarse-grained model as the small-scale parameter goes to zero. The basic idea in this setting is usually the same. One constructs an estimator, often through quantities of the coarse-grained model, e.g., likelihood, plugs in observations not from the coarse-grained model process but from the weakly converging data-generating model process, and then analyzes how this estimator performs in the limit for vanishing small-scale parameter. In this context of weak convergence, we will occasionally refer to the model misspecification as weak misspecification and to the observations as weakly perturbed observations or data. Although many of the theoretical results and proposed estimators in this thesis may be applicable in and extendable to more general settings, we will mostly confine ourselves to a setting where the stochastic processes are given by solutions to stochastic differential equations (SDEs) and the preferred method of coarse-graining is homogenization for SDEs. Since homogenization models naturally arise from coarse-graining multiscale systems, we will refer to the coarse-grained model as homogenization model or homogenized equation and to the data-generating model as multiscale model or multiscale system. A review of homogenization for SDEs with relevant references will be given in Section 1.2. The whole thesis is built on the underlying assumption that we have continuous observations of the SDE paths, which is, of course, not realistic, but facilitates theoretical proofs and makes the definition of estimators more elegant with less burdensome notation. However, for the numerical implementations we tacitly use properly discretized versions of these estimators and, by implication, plug in high-frequency observations during simulation.

This kind of estimation problem with a weak misspecification between the data-generating model and the coarse-grained model appears to be difficult to solve robustly with standard statistical estimators, especially the ones that depend on path increments, e.g., quadratic variation estimators for constant diffusion coefficients in SDE models, which is less surprising, since weak convergence only quantifies closeness of expectations against bounded continuous functions. As a consequence, considerable effort has been put into solving these robustness issues of estimators, as will be recapped in Section 1.3. As for the thesis, its line of thought stems not only from this shared desire to develop new robust estimators, but also, more originally, from the personal wish to provide mathematical tools, first and foremost, limit theorems for certain functionals of the weakly perturbed observations that allow one to prove asymptotic properties of a variety of estimators rigorously. As we will see in all subsequent chapters, powerful limit theorems are one of the missing links for the statistical asymptotic analysis of estimators when models are misspecified.

1.2 Homogenization for SDEs

Homogenization for SDEs is a type of diffusion approximation for a system of SDEs involving two widely separated time scales such that the method of scale separation allows for the derivation of a coarse-grained model. There are many multiscale SDE systems with varying levels of generality and with different approaches to their homogenization, so the following exposition is far from exhaustive, but, in order to illustrate the main ideas, let us consider the following fully coupled multiscale SDE system in $\mathbb{R}^{d_x+d_y}$, $d_x, d_y \in \mathbb{N}$,

$$dX_t^\varepsilon = \left[\frac{1}{\varepsilon}H(X_t^\varepsilon, Y_t^\varepsilon) + F(X_t^\varepsilon, Y_t^\varepsilon)\right]dt + \Sigma(X_t^\varepsilon, Y_t^\varepsilon)\, dW_t^{(x)}, \quad X_0^\varepsilon = x_0 \in \mathbb{R}^{d_x}, \tag{1.2.1a}$$

$$dY_t^\varepsilon = \left[\frac{1}{\varepsilon^2}f_0(X_t^\varepsilon, Y_t^\varepsilon) + \frac{1}{\varepsilon}f_1(X_t^\varepsilon, Y_t^\varepsilon)\right]dt + \frac{1}{\varepsilon}\sigma(X_t^\varepsilon, Y_t^\varepsilon)\, dW_t^{(y)}, \quad Y_0^\varepsilon = y_0 \in \mathbb{R}^{d_y}, \tag{1.2.1b}$$

where $W^{(x)}$, $W^{(y)}$ are independent standard Brownian motions of dimension d_x and d_y, respectively, and $H, F \colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}^{d_x}$, $\Sigma \colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}^{d_x \times d_x}$, as well as $f_0, f_1 \colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}^{d_y}$, and $\sigma \colon \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}^{d_y \times d_y}$ are the coefficient functions. Here, $0 < \varepsilon \ll 1$ denotes a small scale parameter controlling the time scale separation. Under suitable assumptions on these coefficient functions, one can show with homogenization techniques, see the references later in this section, that the process X^ε converges weakly in $C([0,T]; \mathbb{R}^{d_x})$ as $\varepsilon \to 0$ to the process X which solves the so-called homogenized SDE

$$dX_t = \bar{F}(X_t)\, dt + \bar{\Sigma}(X_t)\, d\bar{W}_t, \quad X_0 = x_0, \tag{1.2.2}$$

with certain limit coefficient functions $\bar{F} \colon \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$, $\bar{\Sigma} \colon \mathbb{R}^{d_x} \to \mathbb{R}^{d_x \times d_x}$, and a potentially new standard Brownian motion $\bar{W}$ of dimension d_x. Figure 1.1 shows exemplary sample paths of a multiscale SDE system with corresponding homogenized limit process.

Before we specify the limit coefficient functions, let us first identify from this abstract presentation what homogenization achieves. In homogenization theory, it is generally understood that X^ε evolves on a time scale that is much slower than the time scale of Y^ε when $0 < \varepsilon \ll 1$; this is reflected in the equations (1.2.1) through terms of different orders of powers of $1/\varepsilon$. Hence, it is common terminology in this context to call X^ε the slow or slow-scale variable and Y^ε the fast or fast-scale variable. Intuitively speaking, homogenization is a procedure in which the explicit evolution of the fast variable Y^ε is eliminated through scale separation, that is, letting $\varepsilon \to 0$, in order to obtain a simplified equation that approximates the slow variable X^ε and captures the effective slow-scale dynamics of the multiscale system. Of course, the information from the fast variable is not lost in the procedure, but rather embedded in an elegant way into the single homogenized equation (1.2.2). Besides the immediate reduction in complexity and dimensionality, a less obvious gain concerns structure. While the solution of the homogenized equation is a Markov process, the slow variable itself is only a continuous semimartingale if it is explicitly coupled to Y^ε. As such, homogenization shows that the slow variable decouples after scale separation from the fast variable in a way that semimartingale structure is upgraded to Markovian structure.

The limit coefficient functions are given in terms of the solution to an auxiliary Poisson equation involving the drift term H in (1.2.1a). For the solvability of that Poisson

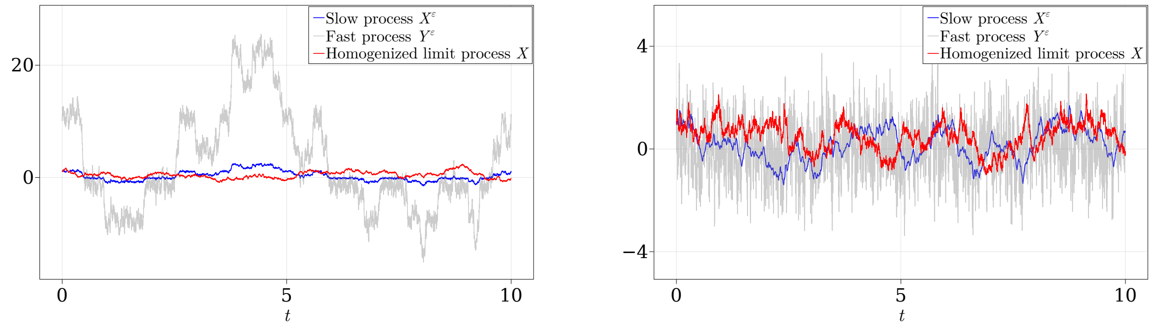


Figure 1.1: Sample paths of a multiscale SDE system with slow-scale and fast-scale variable and their corresponding homogenized limit SDE process.

equation, it is necessary to assume the following centering condition for H

$$\int_{\mathbb{R}^{d_y}} H(x,y)\, d\mu^x(y) = 0, \quad x \in \mathbb{R}^{d_x}, \tag{1.2.3}$$

where μ^x is the invariant measure of the so-called frozen SDE

$$dY_t^x = f_0(x, Y_t^x)\, dt + \sigma(x, Y_t^x)\, dW_t^{(y)}, \quad Y_0^x = y_0. \tag{1.2.4}$$

Here, $x \in \mathbb{R}^{d_x}$ acts as a parameter of the equation. The invariant measure μ^x for Y^x exists under a weak recurrence assumption on the drift f_0 and uniform ellipticity of $a := \sigma\sigma^\top$. The reason (1.2.4) is called the frozen SDE is that on time scales long compared to ε^2 and short compared to 1, the process X^ε appears frozen while the process Y^ε traverses its invariant measure on such time scales because it is evolving more quickly. Of course, X^ε is not frozen, but since it evolves on a slower time scale than Y^ε, intuition based on freezing X^ε and considering (1.2.4) helps in understanding how homogenization arises. Coming back to theory, if we assume the centering condition for H, weak recurrence for f_0, uniform ellipticity for a, and enough regularity of the coefficient functions, then there exists a unique solution χ satisfying the centering condition (1.2.3), i.e.,

$$\int_{\mathbb{R}^{d_y}} \chi(x,y)\, d\mu^x(y) = 0, \quad x \in \mathbb{R}^{d_x}, \tag{1.2.5}$$

and solving the following Poisson equation, which is called the cell problem in homogenization,

$$-\mathcal{L}_0(x,y)\chi(x,y) = H(x,y), \qquad x \in \mathbb{R}^{d_x}, \quad y \in \mathbb{R}^{d_y}. \tag{1.2.6}$$

Here, the second-order differential operator $\mathcal{L}_0$ is given by

$$\mathcal{L}_0(x,y) := \frac{1}{2}\sum_{i,j=1}^{d_y} a_{ij}(x,y)\frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{d_y} f_0^{(i)}(x,y)\frac{\partial}{\partial y_i}, \tag{1.2.7}$$

with component functions $f_0^{(i)}$ of f_0 and matrix entries a_{ij} of a for $i,j = 1,\dots,d_y$. Note that this is a differential operator in the variable y only while x is just a parameter. Using the solution to the cell problem, we can write down the homogenized drift coefficient $\bar{F}$

and diffusion coefficient $\bar{\Sigma}$, namely, for $x \in \mathbb{R}^{d_x}$,

$$\bar{F}(x) := \int_{\mathbb{R}^{d_y}} \left[F(x,y) + \nabla_x \chi(x,y) H(x,y) + \nabla_y \chi(x,y) f_1(x,y)\right] d\mu^x(y), \tag{1.2.8}$$

$$\bar{\Sigma}(x)\bar{\Sigma}(x)^\top := \int_{\mathbb{R}^{d_y}} \left[\Sigma\Sigma^\top(x,y) + [\nabla_y \chi(x,y)\sigma(x,y)]\,[\nabla_y \chi(x,y)\sigma(x,y)]^\top\right] d\mu^x(y). \tag{1.2.9}$$

Since the RHS of formula (1.2.9) is positive semi-definite, there indeed exists a solution $\bar{\Sigma}$ to that equation. Although this solution is not unique in general, this poses no problem in a setting of weak convergence of measures as the relevant information about the law of the SDE processes is encoded through the matrix $\bar{\Sigma}\bar{\Sigma}^\top$ and, of course, the drift vector $\bar{F}$; compare with the various discussions in [SV79, Chapter 5].

As we can see from these equations for the limit coefficients, the $\mathcal{O}(\varepsilon^{-1})$ terms H and f_1 in combination with the solution χ to the cell problem (1.2.6), which is itself based on the $\mathcal{O}(\varepsilon^{-2})$ terms f_0 and σ, lead to subtle homogenization effects, so that the limit coefficients $\bar{F}$ and $\bar{\Sigma}$ are not only the $\mathcal{O}(1)$ terms F and Σ integrated with respect to the invariant measure μ^x. In fact, if H and f_1 were not present, i.e., $H \equiv f_1 \equiv 0$, then the averaging principle pioneered by Khas'minskii [Kha66b; Kha66a] would give rise to exactly these averaged quantities, that is, the terms F and Σ integrated with respect to the invariant measure μ^x. Later on, Papanicolaou, Stroock, and Varadhan [PSV77; Pap76] employed the powerful martingale approach, initiated by Stroock and Varadhan [SV79], to identify the limiting equation and prove weak convergence of the slow variable to the homogenized process. Their proof relies on the previously stated assumptions and sufficient regularity and boundedness conditions of the coefficient functions in (1.2.1). They claim in several remarks that these regularity and boundedness assumptions can be substantially relaxed due to the way limit theorems are obtained through the martingale approach, cf. [SV79], but they do not provide an explicit proof. Similar results can be found in [BLP78; Par99], where a multitude of other homogenization problems with periodic structures and compact state spaces, that is, the space in which the processes obtain their values, are thoroughly analyzed. In the series of papers [PV01; PV03; PV05], the authors studied the regularity properties of the solution to the Poisson equation (1.2.6) and provided probabilistic as well as analytic proofs for the homogenization of SDEs when the coefficients belong to Hölder function classes and are unbounded not only in the fast variable but also in the slow variable. The method they use to obtain weak convergence is again the martingale approach. An alternative proof idea follows a corrector approach [PS07; PPS09], where the so-called corrected process defined through $\hat{X}^\varepsilon := X^\varepsilon + \varepsilon\chi(X^\varepsilon, Y^\varepsilon)$ is analyzed in the limit for $\varepsilon \to 0$. This approach is especially intuitive as many of the homogenization terms already appear in the equations for $\hat{X}^\varepsilon$ through Itô's formula. Yet again, an entirely different proof method – neither relying on the martingale nor on the corrector approach – with similar assumptions to the aforementioned series of papers [PV01; PV03; PV05], has been considered in the more recent contribution [RX21], where a transfer formula and specific fluctuation estimates are derived to establish weak convergence. Though their approach does not retrieve weak convergence at the process level, they do obtain explicit weak convergence rates of X_t^ε to X_t for fixed $t \in [0,T]$ in terms of ε.

As we can see from this short exposition, the theory of homogenization for SDEs can be approached from different angles and with different sets of assumptions, but they all share similar core principles, e.g., the centering condition, the frozen SDE, and the

Poisson equation. A modern overview of homogenization for SDEs, including more general multiscale systems and other aspects of multiscale analysis, is given in [PS08].

Accompanying example. Although many of our results will be stated and proved in more generality, we will consistently study one particular example of homogenization throughout the manuscript. To be precise, consider the following SDE in $\mathbb{R}^d$

$$dX_t^\varepsilon = \left[-\alpha\nabla V(X_t^\varepsilon) - \frac{1}{\varepsilon}\nabla p\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\right]^\top dt + \sqrt{2\sigma}\, dW_t, \quad X_0^\varepsilon = x_0, \tag{1.2.10}$$

where $V\colon \mathbb{R}^d \to \mathbb{R}$ is a large-scale potential, $p\colon \mathbb{R}^d \to \mathbb{R}$ a periodic function with period $\mathbb{L} > 0$ in all directions, and $\alpha, \sigma, \varepsilon > 0$. We assume for the introduction of this example that all quantities are well-defined; precise conditions will be stated in the subsequent chapters. If we define $Y^\varepsilon := X^\varepsilon/\varepsilon$, then this puts us into a similar setting[1] as given by the multiscale system (1.2.1)

$$dX_t^\varepsilon = \left[-\alpha\nabla V(X_t^\varepsilon) - \frac{1}{\varepsilon}\nabla p\left(Y_t^\varepsilon\right)\right]^\top dt + \sqrt{2\sigma}\, dW_t, \quad X_0^\varepsilon = x_0 \in \mathbb{R}^d \tag{1.2.11a}$$

$$dY_t^\varepsilon = \left[-\frac{1}{\varepsilon^2}\nabla p\left(Y_t^\varepsilon\right) - \frac{\alpha}{\varepsilon}\nabla V(X_t^\varepsilon)\right]^\top dt + \frac{\sqrt{2\sigma}}{\varepsilon}\, dW_t, \quad Y_0^\varepsilon = x_0/\varepsilon \in \mathbb{R}^d. \tag{1.2.11b}$$

We will call equation (1.2.10) the multiscale overdamped Langevin equation. As a special case of a more general setting, it has been proved in [DDP23] that X^ε converges weakly in $C([0,T];\mathbb{R}^d)$ to the solution X of the homogenized overdamped Langevin equation

$$dX_t = \left[-\alpha\mathcal{K}\nabla V(X_t)\right]^\top dt + \sqrt{2\sigma\mathcal{K}}\, d\bar{W}_t, \quad X_0 = x_0, \tag{1.2.12}$$

with $\mathcal{K}$, a positive homogenization factor that emerges from the cell problem, given by

$$\mathcal{K} := \int_{\mathbb{T}_{\mathbb{L}}^d} (I_d + \nabla_y\chi(y))(I_d + \nabla_y\chi(y))^\top d\mu^x(y), \quad \mathbb{T}_{\mathbb{L}}^d := \mathbb{R}^d/(\mathbb{L}\mathbb{Z})^d, \tag{1.2.13}$$

invariant measure μ^x of the frozen SDE given through the Lebesgue-density

$$d\mu^x(y) := \frac{1}{Z_p}\exp\left(-p(y)/\sigma\right) dy, \quad Z_p := \int_{\mathbb{T}_{\mathbb{L}}^d} \exp\left(-p(y)/\sigma\right) dy, \tag{1.2.14}$$

and the solution $\chi\colon \mathbb{R}^d \to \mathbb{R}^d$ to the cell problem with periodic boundary conditions

$$\mathcal{L}_0(y)\chi(y) = \nabla p(y), \quad y \in [0,\mathbb{L}]^d, \quad \mathcal{L}_0(y) = \sigma\sum_{i,j=1}^d \frac{\partial^2}{\partial y_i \partial y_j} - \sum_{i=1}^d \partial_{y_i} p(y)\frac{\partial}{\partial y_i}. \tag{1.2.15}$$

In the 1-dimensional case $d = 1$, the solution to the cell problem can be explicitly determined, and the homogenization factor $\mathcal{K}$ is then given by

$$\mathcal{K} = \frac{1}{\Pi^+\Pi^-}, \quad \Pi^\pm := \frac{1}{\mathbb{L}}\int_0^{\mathbb{L}} \exp\left(\pm p(y)/\sigma\right) dy. \tag{1.2.16}$$

[1] System (1.2.1) only involves independent Brownian motions, but it is perfectly possible to include a dependent Brownian motion and derive similar homogenized equations, see, e.g., [PS08; PPS09; PV01].

With the actual equations at hand, we can now clarify what we mean by statistical estimation under weak misspecification. We want to construct estimators for quantities of the homogenized equation (1.2.2), e.g., for the drift coefficient, diffusion coefficient, invariant measures, etc., and plug in observations solely from the slow variable X^ε of the multiscale system (1.2.1) into the constructed estimator, pretending as if we do not or cannot observe the fast variable Y^ε. We then want to analyze the estimator's asymptotic behavior when the small-scale parameter ε goes to zero. For instance, in the setting of the previously presented accompanying example, we will analyze parametric estimators for the parameter $\alpha\mathcal{K}$ in the drift coefficient of (1.2.12) and a nonparametric estimator for the invariant density of the solution process X to (1.2.12). In both cases, the estimators are built based on the homogenized overdamped Langevin equation (1.2.12), but their statistical asymptotic properties will be analyzed under continuous observation paths of the solution process X^ε to the multiscale overdamped Langevin equation (1.2.10). This type of model misspecification is a delicate problem as we will review in Section 1.3.

1.3 Related Work

In recent decades, a lot of effort has been put into the investigation of estimation problems with the aforementioned weak misspecification. The most fundamental among these efforts is arguably the seminal paper [PS07] where the authors considered a class of multiscale overdamped Langevin equations as their data-generating model with a well-defined homogenized limit process that is again given by a Langevin equation with new coefficients damped by a homogenization factor. Essentially, they considered the accompanying example of the previous section. They then analyzed maximum likelihood estimation built on basis of the homogenized limit process while subject to weakly perturbed observations coming from the slow variable of the multiscale system. Their main theoretical and numerical findings are that the maximum likelihood estimator (MLE) is asymptotically biased in the large-time and small-scale limit and recovers parameters from the unhomogenized multiscale system instead of the homogenized limit equation. To solve this inconsistency of the MLE, they proposed a subsampling approach in which a minimal subsampling interval between observation points must be respected in order to retrieve the correct homogenized parameters. The estimation procedure depends sensitively on the subsampling rate, which itself depends on the unknown scale parameter of the multiscale system. Above all, they showcase that different unknown parameters in multidimensional homogenized models require different subsampling rates, advocating for extreme care when handling multidimensional estimation problems. In the subsequent work [PPS09], the authors expanded on the previous results for the MLE and proved on a very rigorous and general level that the MLE is asymptotically biased under weak misspecification but can be repaired with appropriate subsampling. Chapter 7 of [PPS12] gives a unifying summary of the preceding two papers and provides additional material regarding estimation problems for multiscale models.

The idea of subsampling data has already appeared prior to these publications, for example, in the context of market microstructure and high-frequency financial data [AMZ05; ZMA05]. In the years after, similar subsampling strategies have been investigated, as well, see [Aze+13; OSP10]. Despite the apparent success of subsampling in multiscale SDE settings, the authors in [FR11], for example, discuss limitations and accuracy issues of the subsampling procedure in the context of estimation for diffusion coefficients in homogenized models. They instead propose a diffusion coefficient estimator with a preprocessing step involving the filtering of drift contributions, although one must note that they require prior knowledge of the homogenized drift coefficient, which potentially mitigates the practicality of their approach.

The keyword filtering already fell in the preceding paragraph and recently brought a multitude of estimation methods applied not only to multiscale systems but also to other stochastic dynamical systems. For instance, the authors of [Abd+23; GZ23] introduced a technique of filtering multiscale observations through a low-pass exponential filter to then use these filtered observations, in addition to the original observations, in a modified MLE and diffusion coefficient estimator. Under certain assumptions on the multiscale system and the homogenized equation, they proved the asymptotic unbiasedness of their estimators. Motivated by these results, the filtering approach has then been successfully combined with martingale estimating functions based on eigenfunctions of the generator [APZ22] and, more recently, with stochastic gradient descent in continuous time [HZ24] to infer homogenized quantities. In addition to multiscale systems, the same filtered data approach has been applied to stochastic processes driven by colored noise [PRZ25]. A nuanced analysis of this whole approach is given in the PhD thesis [Zan22].

Another popular source for estimators are martingale estimating equations. To give an example, the series of works [KPK13; KKP15; Kru18], which was also applied in [Kru+15] to paleoclimatic and biological data, is based on such martingale estimating equations in a semiparametric framework under weak misspecification. Instead of a single long multiscale observation, they assume to have multiple short trajectories as observations, which makes it possible to approximate certain conditional expectations necessary for the approach to work. This approximation is accomplished through specific Nadaraya-Watson techniques. The effectiveness of this approach has been shown in several numerical studies of these publications, but theoretical justification for robustness are, thus far, still limited.

Other noteworthy contributions include the series of works [GS17; GS18; SC13], where they use appropriately adjusted maximum likelihood estimation for homogenized models that emerge from a multiscale system with vanishing noise in the limit, so that the homogenized equation is actually deterministic, rather than an SDE. In [CV11], the authors take advantage of the asymptotic closedness of the eigenvalue problems between the generator of the multiscale SDE system and the generator of the homogenized SDE to perform nonparametric drift and diffusion estimation for the homogenized model by estimating operator eigenpairs through Galerkin methods. The effectiveness of their method has been numerically tested, although statistical properties of their proposed estimators have not been investigated in that paper. Regarding diffusion coefficient estimation without the usage of hyperparameter-dependent subsampling or filtering, there exists the paper [MP19] where the authors propose a so-called extrema quadratic variation estimator for the diffusion coefficient of the homogenized model in the case when the slow variable is of bounded variation. Their estimator is inspired by the total quadratic variation from rough path theory and employs a natural subsampling. The result is a purely data-driven estimation method without the need of adjusting any hyperparameters, e.g., subsampling rates or filtering widths.

This brief presentation of existing methodologies for the estimation of homogenized quantities from multiscale observations shows that there are quite some commonalities between the different methods and the way they are analyzed. To reiterate, there are many approaches that alter the observations through subsampling or filtering and there are approaches that focus on certain objects or quantities which already behave well under multiscale observations, such as martingale estimating equations or spectral quantities of the generators. The estimators that we will analyze in this thesis follow a similar spirit as the latter approaches. More specifically, we aim to analyze quantities that already behave well under weak misspecification. In our case, these will be time averages of the observations and integrals with respect to invariant measures. Ultimately, our goal is two-fold. On the one hand, we want to develop estimators that can be numerically validated in their asymptotic properties and, on the other hand, we want to rigorously justify not only their robustness but also asymptotic normality, where proofs for the latter property are a fairly rare sight in the field of estimation under multiscale observations.

1.4 Contributions and Outline of the Thesis

Although we will provide more detailed and motivational introductory material for each chapter separately, let us at least give a crude overview of the thesis and present some of the most relevant contributions here.

In Chapter 2, we motivate, present, and discuss a plethora of limit theorems for certain functionals of continuous observation paths coming from the solution process X^ε of a weakly perturbed SDE model. More specifically, we will develop a general mean ergodic theorem (MET) of the kind

$$\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\,dt \xrightarrow{L^2} \int_{\mathbb{R}} h(x)\mu(x)\,dx, \quad \varepsilon \to 0,$$

where h_ε and h are functions from an admissible class such that $h_\varepsilon \to h$ as $\varepsilon \to 0$, $T_\varepsilon \to \infty$ as $\varepsilon \to 0$, and μ is the invariant density of a limit SDE. We provide a qualitative convergence result and a quantitative convergence result with explicit convergence rates with respect to ε. A bigger theoretical obstacle that we successfully overcome in our analysis is the one of a nonstationary initialization of the SDE process; in other words, we do not assume that we start the process with the invariant distribution but rather with an arbitrary deterministic initial state. Building on these results, we will then prove a central limit theorem (CLT) of the type

$$\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\,dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\tau^2), \quad \varepsilon \to 0,$$

where h_ε is an appropriately centered function and $\mathcal{N}_1(0,\tau^2)$ is a normal distribution with mean zero and variance $\tau^2 > 0$. These results are novel in the sense that we actually establish well-defined long-time behavior of quantities for which homogenization results thus far only exist with respect to finite time intervals. To be more precise, in Section 1.2, we have mentioned that the slow process X^ε converges weakly to X in $C([0,T];\mathbb{R}^d)$ as $\varepsilon \to 0$ for fixed $T > 0$, but here we consider a time horizon that diverges to infinity. The reasons for coupling the time horizon with the small scale parameter ε will be motivated in Chapter 2. However, a shortcoming of the acquired results is that they are formulated and proved only for 1-dimensional processes and when the slow variable is decoupled from the fast variable of the multiscale system. At least some words of solace with respect to this matter are that a lot of these results are stated in a more general setting than what is enforced through the diffusion approximation via homogenization. Additionally, the 1-dimensional case is a useful didactic tool for the multidimensional case, for which we will give comments and proof ideas, as well. The work of this chapter is mainly based on the submitted paper [BK26] and in parts on the results in [Bor+25; BKP25].

With these limit theorems at our disposal, we are then in a position enabling us to analyze the specific estimation problems we are interested in. We start this endeavor in Chapter 3 by considering parametric estimation problems. We define a general statistical framework for parametric inference under weak misspecification, introduce different notions of robustness, and investigate the asymptotic properties, such as robustness and asymptotic normality, of several parametric estimators. For instance, we will meticulously analyze the following minimum distance estimator (MDE) that is defined through a minimization problem

$$\hat{\vartheta}_{T_\varepsilon}(X^\varepsilon) \in \operatorname*{arg\,inf}_{\vartheta\in\Theta_0} \int_{\mathbb{R}^d} \left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp\left(iu^\top X_t^\varepsilon\right) \, dt - \int_{\mathbb{R}^d} \exp\left(iu^\top x\right) \, d\mu(\vartheta, x) \right|^2 d\nu(u),$$

where ν is a selectable weight measure and $\mu(\vartheta, \cdot)$ is the invariant measure of a limit process, parametrized by the parameter $\vartheta \in \Theta_0$ which we are trying to estimate with the given observations of X^ε. Although minimum distance estimators are well-studied in the existing statistics literature [Wol57; Kut94; Vaa98; Pra99], available results on asymptotic properties of such estimators under weak misspecification and, especially, results concerning asymptotic normality in a setting where observations come from a multiscale SDE model are limited. In this thesis, we provide such a thorough theoretical and empirical analysis of our proposed minimum distance estimator. The work of this chapter is primarily based on the submitted papers [BK26; BKP25].

Chapter 4 is dedicated to a topic of nonparametric estimation where we employ the following projection approach based on Hermite expansions for the nonparametric estimation of homogenized invariant densities, i.e., we analyze the estimator

$$\widehat{\mu}^\varepsilon_{N_\varepsilon,T_\varepsilon} = \sum_{n=0}^{N_\varepsilon-1} \widehat{\alpha}^\varepsilon_{n,T_\varepsilon} \, \psi_n, \qquad \widehat{\alpha}^\varepsilon_{n,T_\varepsilon} = \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \psi_n(X_t^\varepsilon) \, dt,$$

where $(\psi_n)_{n\in\mathbb{N}_0}$ is the Hermite orthonormal basis of $L^2(\mathbb{R})$. Under a certain set of assumptions and specific scaling conditions on the number of Fourier coefficients N_ε and the observation time T_ε, we establish, theoretically and numerically, that this estimator recovers the invariant density $\mu \in L^2(\mathbb{R})$ of the homogenized overdamped Langevin equation (1.2.12) when using observations of X^ε of the multiscale overdamped Langevin equation (1.2.10). We prove this robustness, a functional CLT in $L^2(\mathbb{R})$ and we remark possible extensions to other spaces and multiple dimensions. The novelty of our results lies again in the rigorous mathematical analysis of a nonparametric estimator under the aspect of weak misspecification. While the presented results build upon the submitted work in [Bor+25], they are substantially extended here.

Finally, in the Appendices A, B, and C, we collect somewhat advanced topics from stochastic calculus, basic concepts from the theory of weak convergence of measures on different spaces, and technical derivations of a priori interior estimates of the Schauder type for parabolic partial differential equations.

Chapter 2

Limit Theorems

Limit theorems are some of the main technical pillars for the rigorous analysis of asymptotic properties of statistical estimators. Without a good set of general limit theorems, it becomes tedious to prove asymptotic results for every new estimator over and over again. In particular, having a type of law of large numbers and central limit theorem builds the backbone of any serious asymptotic analysis of estimators. Hence, this whole chapter is dedicated to the derivation of qualitative and quantitative limit theorems of certain functionals of the observation X^ε. Although the limit theorems that we are about to present are very well-studied in the case where no ε-dependence is present, there appears to be a theoretical gap for general limit theorems in a setting of weak misspecification. Thus, we aim to partially fill this gap by proving a number of limit theorems in such a setting. As already mentioned in the introduction, we are, thus far, able to prove these limit theorems in the 1-dimensional case only. This is due to some technical limitations that will become apparent soon. Nevertheless, we do provide several hints and ideas for the multidimensional case.

In the first section, which is itself divided into two subsections, we motivate the type of limit theorems we are interested in. The application of these limit theorems to specific estimators occurs in the later chapters only. The first subsection is devoted to a qualitative limit theorem without explicit convergence rates but fairly mild assumptions. Several interesting tools from stochastic calculus, e.g., the Meyer-Tanaka formula and occupation density formula, are employed in this subsection. In the second subsection, we then impose stronger assumptions which allow us to prove a quantitative limit theorem with explicit convergence rates. The second section deals with certain tail probabilities of the process X^ε and their convergence to zero as $\varepsilon \to 0$, a result which is especially useful for the proof of the CLT, which is obtained in the third section through the well-known technique of the Poisson equation. Finally, in the last section, we verify that the multiscale overdamped Langevin equation (1.2.10) fulfills the main assumptions of the limit theorems and we prove several other related asymptotic results in this case. As a side note, we supplement this chapter with remarks on multidimensional extensions of the limit theorems at the end of certain sections. These remarks are devoid of the utmost mathematical rigor and act more as an illustration of challenges and prospective research ideas.

This chapter is fairly technical, with some of the proofs even deferred to Appendix C, but the usefulness of our main results will eventually manifest in Chapter 3 and 4. Also, Appendix A is a useful reference aid for this chapter and we will frequently refer to the results stated there.

2.1 Ergodic Theorems

Since the details of this chapter may obscure the idea, we first outline a thorough high-level reasoning behind our analysis. For this, we recall the multiscale overdamped Langevin equation (1.2.10) for $d = 1$

$$dX_t^\varepsilon = -\alpha V'(X_t^\varepsilon) - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right) dt + \sqrt{2\sigma}\, dW_t, \quad t \in [0, T], \tag{2.1.1}$$

and its homogenized limit equation

$$dX_t = -\alpha \mathcal{K} V'(X_t)\, dt + \sqrt{2\sigma\mathcal{K}}\, d\bar{W}_t, \quad t \in [0, T]. \tag{2.1.2}$$

In the paper [PS07], the authors analyzed maximum likelihood estimation built on basis of the homogenized model equation (2.1.2) in order to estimate the homogenized drift parameter $\alpha\mathcal{K}$ in (2.1.2). The MLE is given by the formula

$$\hat{A}_T(X) := -\frac{\int_0^T V'(X_t)\, dX_t}{\int_0^T |V'(X_t)|^2\, dt}. \tag{2.1.3}$$

They then proceed to show that this estimator is asymptotically biased in the sense that it estimates the unhomogenized drift parameter α in (2.1.1)

$$\lim_{\varepsilon\to 0} \lim_{T\to\infty} \hat{A}_T(X^\varepsilon) = \alpha, \quad \text{almost surely}, \tag{2.1.4}$$

and propose, in addition to their estimators with subsampling, the following drift estimator

$$\tilde{A}_T(X) = \hat{\Sigma}\, \frac{\int_0^T V''(X_t)\, dX_t}{\int_0^T |V'(X_t)|^2\, dt}, \tag{2.1.5}$$

which requires an estimate $\hat{\Sigma}$ for the coefficient $\sigma\mathcal{K}$ in (2.1.2). They establish the following result

$$\lim_{\varepsilon\to 0} \lim_{T\to\infty} \tilde{A}_T(X^\varepsilon) = \frac{\hat{\Sigma}}{\sigma}\alpha, \quad \text{almost surely}, \tag{2.1.6}$$

showing that their estimator is asymptotically unbiased in the sense that the homogenized drift parameter is correctly estimated if $\hat{\Sigma} = \sigma\mathcal{K}$.

As we can see, their established limit results are sequential in nature, that is, they first let the time horizon $T \to \infty$ and then $\varepsilon \to 0$. One of the main theoretical tools applied in that paper are limit theorems for SDEs. More precisely, if we consider the solution X of the general 1-dimensional SDE

$$dX_t = \bar{F}(X_t)\, dt + \bar{\Sigma}(X_t)\, d\bar{W}_t, \quad t \in [0, T], \tag{2.1.7}$$

and suppose that X has an invariant density μ on $\mathbb{R}$, then, under various assumptions, there exist well-established limit theorems in the literature, cf. [Kut04; DZ96; MT59; KMN12], of the type

$$\frac{1}{T}\int_0^T h(X_t)\, dt \to \int_{\mathbb{R}} h(x)\mu(x)\, dx, \quad T \to \infty, \tag{2.1.8}$$

for a suitable test function h, with the convergence understood either in the mean or almost sure sense, or as convergence in probability. In their most general form, Birkhoff-Khinchin's ergodic theorem or Von Neumann's mean ergodic theorem [DZ96] are used to derive these results, and, as such, they then act as a limit theorem for solutions of SDEs in the same way as the law of large numbers acts as a limit theorem for random variables. Another closely related limit result is the CLT

$$\frac{1}{\sqrt{T}}\int_0^T h(X_t)\,dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\tau^2), \quad T\to\infty, \tag{2.1.9}$$

for a function h centered with respect to μ. We want to extend the preceding two types of limit theorems in the following way for a particular class of SDEs, including (2.1.1), given by the solution X^ε of the 1-dimensional SDE

$$dX_t^\varepsilon = F_\varepsilon(X_t^\varepsilon)\,dt + \Sigma(X_t^\varepsilon)\,dW_t, \quad t\in[0,T], \tag{2.1.10}$$

where F_ε depends on $\varepsilon>0$ in a way that we will soon specify. Assuming this general framework, our aim is to prove the following two limit theorems. Firstly, an MET of the type

$$\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\,dt \xrightarrow{L^2} \int_{\mathbb{R}} h(x)\mu(x)\,dx, \quad \varepsilon\to 0, \tag{2.1.11}$$

for a suitable sequence of functions h_ε depending on ε such that $h_\varepsilon \to h$, and exploding time horizon such that T_ε explicitly depends on ε with $T_\varepsilon\to\infty$ as $\varepsilon\to 0$. Secondly, we want to prove a CLT of the type

$$\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\,dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\tau^2), \quad \varepsilon\to 0. \tag{2.1.12}$$

In the aforementioned reference [PS07], as well as in [Kru18; Abd+23; HZ24; GZ23; PRZ25], the authors utilize sequential limits to prove asymptotic properties of their estimators, that is, they first let $T\to\infty$ and then $\varepsilon\to 0$, or vice versa. As far as we know, the case of taking coupled limits as in (2.1.11) and (2.1.12) where almost every relevant quantity depends on ε has not been considered yet. At first glance this coupling of observation time with the scale parameter might seem unmotivated, but, in fact, it is particularly important when rigorously proving asymptotic normality of estimators for misspecified models. To be more precise, in statistical estimation problems under weak misspecification, one often has to deal with quantities like

$$\frac{1}{\sqrt{T}}\int_0^T h(X_t^\varepsilon) - \langle h\rangle_\mu\,dt, \quad \langle h\rangle_\mu := \int_{\mathbb{R}} h(x)\mu(x)\,dx,$$

but, in order to prove a CLT for this time integral, the function h needs to be centered with respect to the correct invariant density, which is μ_ε here, the one associated with the process X^ε. Doing so, we obtain the simple splitting

$$\frac{1}{\sqrt{T}}\int_0^T h(X_t^\varepsilon) - \langle h\rangle_\mu\,dt = \frac{1}{\sqrt{T}}\int_0^T h(X_t^\varepsilon) - \langle h\rangle_{\mu_\varepsilon}\,dt + \sqrt{T}\left(\langle h\rangle_{\mu_\varepsilon} - \langle h\rangle_\mu\right).$$

But here comes the crux. Due to the second summand, we cannot rigorously justify a sequential limit as in the previously mentioned references, i.e., letting $T\to\infty$ and then

$\varepsilon \to 0$. We let T depend on ε so that we can leverage the two factors in the second term and obtain a rigorous and logically sound CLT. Some authors, such as the ones of the paper [Abd+23], derived a CLT for their estimator formally, but not on the systematic mathematical level as we aim for.

At this point, one might argue that exchanging the order in the sequential limit, i.e., letting $\varepsilon \to 0$ first and then $T \to \infty$, fixes the problem. And, indeed, given certain natural assumptions (e.g., weak convergence of X^ε, ergodic theorem for X), it is almost trivial to obtain, for example, the sequential limit

$$\lim_{T\to\infty} \lim_{\varepsilon\to 0} \frac{1}{T} \int_0^T h(X_t^\varepsilon)\, dt \overset{\mathbb{P}}{=} \int_{\mathbb{R}} h(x)\mu(x)\, dx, \tag{2.1.13}$$

for a sufficiently regular test function h such that the continuous mapping theorem for weakly convergent sequences holds. However, there is a practical reason for why one does not want to consider the limit order in (2.1.13). If we let $\varepsilon \to 0$ first, then this is as if we were to assume that we already have data from the limit process X, a setting we are not considering here. Additionally, we need to emphasize that we generally have no control over the scale separation parameter ε. What we do have control over is the time horizon T and what many other authors do assume is that ε is small but not vanishing. Hence, the sequential limits

$$\lim_{\varepsilon\to 0} \lim_{T\to\infty} \frac{1}{T} \int_0^T h_\varepsilon(X_t^\varepsilon)\, dt \overset{\mathbb{P}}{=} \int_{\mathbb{R}} h(x)\mu(x)\, dx, \tag{2.1.14}$$

are more appropriate to consider and this can be seen numerous times in the literature, cf. [Abd+23; MP19; PPS09; PRZ25; PS07].

The avoidance of analyzing coupled limits can possibly be attributed to the technical difficulty that many relevant quantities and bounds explode for $\varepsilon \to 0$, so that classically established results in the literature for SDE processes cannot be applied directly. In addition to that, many results concerned with the homogenization for SDEs are obtained for fixed, finite time horizons T and $\varepsilon \to 0$, so that existing bounds and estimates in that field of study will be rather crude and insufficient in terms of T. Hence, an important objective in the following sections will be to identify quantities that actually behave well and we will see many instances of those.

2.1.1 A Qualitative ε-Dependent Mean Ergodic Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$ that satisfies the usual conditions, cf. Appendix A, and consider a 1-dimensional standard Brownian motion $W := (W_t)_{t\in[0,\infty)}$ on that probability space.

Fix $x_0 \in \mathbb{R}$ once and for all as the initial condition for the following 1-dimensional SDE in $\mathbb{R}$, depending on a small parameter $\varepsilon > 0$,

$$dX_t^\varepsilon = F_\varepsilon(X_t^\varepsilon)\, dt + \Sigma(X_t^\varepsilon)\, dW_t, \quad t \in [0, T], \quad X_0^\varepsilon = x_0, \tag{2.1.15}$$

with suitable Borel-measurable functions $F_\varepsilon \colon \mathbb{R} \to \mathbb{R}$, $\Sigma \colon \mathbb{R} \to \mathbb{R}$. Assume, for the moment, that there exists a unique strong solution to this SDE on the given probability space. We suppose a setting in which X^ε converges weakly in $C([0,T];\mathbb{R})$ to X as $\varepsilon \to 0$, where X is the unique strong solution to the SDE

$$dX_t = \bar{F}(X_t)\,dt + \bar{\Sigma}(X_t)\,d\bar{W}_t, \quad t \in [0,T], \quad X_0 = x_0, \tag{2.1.16}$$

with Borel-measurable functions $\bar{F}\colon \mathbb{R} \to \mathbb{R}$ and $\bar{\Sigma}\colon \mathbb{R} \to \mathbb{R}$ and a standard Brownian motion $\bar{W} := (\bar{W}_t)_{t\in[0,\infty)}$ on a potentially new probability space. Under Assumptions (C) and (MET), which will be introduced soon, the following scale functions exist and are twice differentiable on $\mathbb{R}$ with strictly positive first derivative

$$f_\varepsilon(x) := \int_0^x \exp\left(-\int_0^y \frac{2F_\varepsilon(z)}{\Sigma(z)^2}\,dz\right)dy, \quad f(x) := \int_0^x \exp\left(-\int_0^y \frac{2\bar{F}(z)}{\bar{\Sigma}(z)^2}\,dz\right)dy. \tag{2.1.17}$$

These functions are harmonic with respect to the differential operators

$$\mathcal{L}_\varepsilon := F_\varepsilon\,\partial_x + \frac{1}{2}\Sigma^2\,\partial_{xx}, \quad \mathcal{L} := \bar{F}\,\partial_x + \frac{1}{2}\bar{\Sigma}^2\,\partial_{xx}, \tag{2.1.18}$$

respectively, that is, $\mathcal{L}_\varepsilon f_\varepsilon = 0$ and $\mathcal{L}f = 0$. Setting $\xi^\varepsilon := f_\varepsilon \circ X^\varepsilon$, $\xi := f \circ X$, and applying Itô's formula yields the transformed SDEs

$$d\xi_t^\varepsilon = \frac{1}{\rho_\varepsilon(\xi_t^\varepsilon)}\,dW_t, \quad t \in [0,T], \quad \xi_0^\varepsilon = f_\varepsilon(x_0), \tag{2.1.19}$$

$$d\xi_t = \frac{1}{\rho(\xi_t)}\,d\bar{W}_t \quad t \in [0,T], \quad \xi_0 = f(x_0). \tag{2.1.20}$$

Here, the functions ρ_ε and ρ are given by

$$\rho_\varepsilon(x) := \frac{1}{\Sigma(g_\varepsilon(x))f_\varepsilon'(g_\varepsilon(x))}, \quad \rho(x) := \frac{1}{\bar{\Sigma}(g(x))f'(g(x))}, \quad x \in \mathbb{R}, \tag{2.1.21}$$

where g and g_ε are the inverse functions to f and f_ε, respectively. The transformed processes ξ^ε and ξ are in the so-called natural scale because they satisfy an SDE without drift and thus have the identity as their scale function, respectively. Note also that f_ε' and f' are strictly positive implying that g_ε' and g' are strictly positive, as well.

In the following, we want to derive an "ε-dependent" MET, i.e., under appropriate assumptions and for arbitrary initial conditions, we want to prove the validity of

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^\varepsilon)\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \to 0, \quad \varepsilon \to 0, \tag{2.1.22}$$

where $\varphi_\varepsilon\colon \mathbb{R} \to \mathbb{R}$ is a sequence of Borel-measurable, real-valued functions, $C_{\rho_\varepsilon} := \int_{\mathbb{R}} \rho_\varepsilon(x)^2\,dx$, and $T_\varepsilon \to \infty$ as $\varepsilon \to 0$. If we further assume the convergence

$$\frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy \to \frac{1}{C_\rho}\int_{\mathbb{R}} \varphi(y)\rho(y)^2\,dy, \quad \varepsilon \to 0, \tag{2.1.23}$$

for some function $\varphi\colon \mathbb{R} \to \mathbb{R}$ and $C_\rho := \int_{\mathbb{R}} \rho(x)^2\,dx$, then

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^\varepsilon)\,dt - \frac{1}{C_\rho}\int_{\mathbb{R}} \varphi(y)\rho(y)^2\,dy\right|^2 \to 0, \quad \varepsilon \to 0, \tag{2.1.24}$$

follows. To prove (2.1.22) we will follow the material by Gikhman and Skorokhod [GS68] on the one hand and by Sundar [Sun89] on the other hand. They proved the MET for the case where no ε-dependence is present and the limit is established for $T \to \infty$. We will carefully adapt and combine the separate approaches whenever necessary.

In this section, we will constrict ourselves to the following assumptions:

Assumptions (C).

i) The functions $\bar{F}$ and $\bar{\Sigma}$ are locally Lipschitz-continuous on $\mathbb{R}$, $\bar{\Sigma}^2$ is strictly positive, and $\bar{F}/\bar{\Sigma}^2$ is locally integrable, that is, for all $x \in \mathbb{R}$, there exists a $\delta > 0$ such that

$$\int_{x-\delta}^{x+\delta} \frac{|\bar{F}(y)|}{\bar{\Sigma}(y)^2}\, dy < \infty. \tag{2.1.25}$$

ii) The function Σ is locally Lipschitz-continuous on $\mathbb{R}$, Σ^2 is strictly positive, and F_ε/Σ^2 is locally integrable for each fixed $\varepsilon > 0$. Moreover, for any $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists a constant $L_N > 0$ such that for all $x, y \in [-N, N]$

$$|F_\varepsilon(x) - F_\varepsilon(y)| \leq \frac{L_N}{\varepsilon^2}|x - y|.$$

Remark 2.1.1. *Most homogenization results for SDEs are, to the best of our knowledge, established under global Lipschitz conditions, cf. [BLP78; Par99; PV01; PS08; RX21; DDP23], but, according to [PSV77], it is possible to relax the assumptions to local conditions as long as the solution to* (2.1.16) *has almost surely infinite explosion time.*

Assumptions (MET).

i) $\lim_{x\to\pm\infty} f(x) = \pm\infty$.

ii) $C_\rho = \int_{\mathbb{R}} \rho(x)^2\, dx < \infty$.

iii) There exist constants $c_f, c_\Sigma, C_f, C_\Sigma > 0$ such that for all $\varepsilon > 0$ and on $\mathbb{R}$,

$$c_f f' \leq f'_\varepsilon \leq C_f f', \qquad c_\Sigma \bar{\Sigma}^2 \leq \Sigma^2 \leq C_\Sigma \bar{\Sigma}^2.$$

We are not suggesting that these assumptions are minimal requirements for the following results to holds. On the contrary, some of them, e.g., the pointwise bounds in (MET) iii), can be rather restrictive and for certain proofs even superfluous. Indeed, the same results can be obtained if, for example, in (MET) iii) there exist two integrable functions k_1, k_2 such that $c_f k_1 \leq f'_\varepsilon \leq C_f k_2$ on $\mathbb{R}$. But in order to keep the exposition clear without repeating similar arguments and altering assumptions over and over again, we will stick to them throughout this work. Note, however, that we do not impose global Lipschitz conditions on the drift coefficients, as it was done in [GS68] and [Sun89]. We also need to emphasize the circumstance that the local Lipschitz constant in (C) ii) explodes as $\varepsilon \to 0$ and we need to let the test functions in the integrals to depend on ε. This makes the analysis substantially more difficult compared to the case where there is no ε-dependence in the coefficients and in the time horizon.

Remark 2.1.2. *As it has been established in [GS68, §18] and [Sun89] but will also become apparent here, the invariant density of a stochastic process X defined through an SDE such as* (2.1.16) *exists whenever its corresponding scale function satisfies $f(x) \to \pm\infty$ as $x \to \pm\infty$, and the function ρ^2 is integrable. In this case, the invariant density of X is given by*

$$\mu(x) := \frac{1}{Z\bar{\Sigma}(x)^2} \exp\left(\int_0^x \frac{2\bar{F}(y)}{\bar{\Sigma}(y)^2}\, dy\right), \quad Z := \int_{\mathbb{R}} \bar{\Sigma}(x)^{-2} \exp\left(\int_0^x \frac{2\bar{F}(y)}{\bar{\Sigma}(y)^2}\, dy\right) dx. \tag{2.1.26}$$

For fixed $\varepsilon > 0$ and under analog assumptions, a similar formula holds for the invariant density of X^ε, namely

$$\mu_\varepsilon(x) := \frac{1}{Z_\varepsilon\Sigma(x)^2} \exp\left(\int_0^x \frac{2F_\varepsilon(y)}{\Sigma(y)^2}\, dy\right), \quad Z_\varepsilon := \int_{\mathbb{R}} \Sigma(y)^{-2} \exp\left(\int_0^x \frac{2F_\varepsilon(y)}{\Sigma(y)^2}\, dy\right) dx. \tag{2.1.27}$$

Observe that an ordinary integral substitution gives

$$\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(X_t^\varepsilon)\, dt - \int_{\mathbb{R}} \varphi_\varepsilon(x)\mu_\varepsilon(x)\, dx = \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(g_\varepsilon(\xi_t^\varepsilon))\, dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(g_\varepsilon(x))\rho_\varepsilon(x)^2\, dx,$$

indicating that it is indeed sufficient to prove the mean ergodic theorem for the process in the natural scale (2.1.19). We start with a lemma that summarizes some properties, following from Assumptions (C) and (MET). These simple facts will be used throughout what follows.

Lemma 2.1.3. *Assume (C) and (MET). Then the following hold:*

i) For each of the SDEs defined in (2.1.15), (2.1.16), (2.1.19), (2.1.20) *with their given initial conditions, there exists a unique strong solution. Furthermore, for $\varepsilon > 0$, we have the moment inequality*

$$\mathbb{E}\,|\xi_t^\varepsilon| \leq |f_\varepsilon(x_0)| \exp(t), \quad t \in [0, \infty). \tag{2.1.28}$$

ii) There exist constants $d_\rho, D_\rho > 0$ such that $d_\rho \leq C_{\rho_\varepsilon} \leq D_\rho$ for all $\varepsilon > 0$.

iii) For every $\delta > 0$, there exists an $M > 0$ such that for all $\varepsilon > 0$,

$$\int_{|y|>M} \rho_\varepsilon(y)^2\, dy < \delta.$$

Proof. We will prove each of the statements separately.

i) Strong uniqueness for (2.1.19), (2.1.20), and inequality (2.1.28) are a consequence of Theorem A.2.3. Indeed, the generator of ξ^ε, whose general definition is given in Appendix A, equals

$$\frac{1}{2}\rho_\varepsilon^{-2}\partial_{xx},$$

so that the function $v(x) := |x|$, $x \in \mathbb{R}$, satisfies (A.2.13) and (A.2.12) for $R = c = 1$. The same ideas apply to ξ. Strong uniqueness for (2.1.15), (2.1.16) now follows from [KS91, Proposition 5.5.13], because the existence of a strong solution to an SDE transformed by the scale function is equivalent to the existence of a strong solution to the original equation. Note here that, according to Feller's test for explosions [KS91, Problem 5.5.27, Theorem 5.5.29], the solution processes do not explode in finite time due to (MET) i).

ii) This follows by a simple integral substitution and (MET) iii), namely

$$C_{\rho_\varepsilon} = \int_{\mathbb{R}} \frac{dy}{\Sigma(g_\varepsilon(y))^2 f_\varepsilon'(g_\varepsilon(y))^2} = \int_{\mathbb{R}} \frac{dy}{\Sigma(y)^2 f_\varepsilon'(y)} \leq \frac{1}{c_\Sigma c_f} \int_{\mathbb{R}} \rho(y)^2 \, dy =: D_\rho,$$

and, similarly, for the lower bound.

iii) First, (MET) iii) implies that

$$c_f f \leq f_\varepsilon \leq C_f f, \quad \text{on } \mathbb{R}, \tag{2.1.29}$$

for any $\varepsilon > 0$. Now let $\delta > 0$ be arbitrary. Due to (MET) ii), we can choose $M > 0$ such that

$$\int_{|y|>M/(c_f \vee C_f)} \rho(y)^2 \, dy < \delta c_\Sigma c_f,$$

where $c_f \vee C_f = \max\{c_f, C_f\}$. Using this inequality, (2.1.29), and the strict monotonicity of f, f_ε, g, and g_ε, we can therefore estimate

$$\begin{aligned} \int_{|y|>M} \rho_\varepsilon(y)^2 \, dy &= \int_{|f_\varepsilon(y)|>M} \frac{dy}{\Sigma(y)^2 f_\varepsilon'(y)} \leq \frac{1}{c_\Sigma c_f} \int_{|f(y)|>M/(c_f \vee C_f)} \frac{dy}{\bar{\Sigma}(y)^2 f'(y)} \\ &= \frac{1}{c_\Sigma c_f} \int_{|y|>M/(c_f \vee C_f)} \rho(y)^2 \, dy < \delta \end{aligned}$$

for all $\varepsilon > 0$. □

Under Assumptions (C) and (MET), we define, for any $x, y \in \mathbb{R}$ and $\varepsilon > 0$, the stopping time $\tau^y(\xi^{\varepsilon,x}) := \inf\{s > 0 \,|\, \xi_s^{\varepsilon,x} = y\}$, that is, the time that the process $\xi^{\varepsilon,x}$ solving (2.1.19) with $\xi_0^{\varepsilon,x} = x$ requires to get to the point y for the first time. By [GS68, §16, Theorem 1], we have

$$\mathbb{P}\left(\sup_{t>0} \xi_t^{\varepsilon,x} = \infty\right) = \mathbb{P}\left(\inf_{t>0} \xi_t^{\varepsilon,x} = -\infty\right) = 1,$$

so that $\mathbb{P}(\tau^y(\xi^{\varepsilon,x}) < \infty) = 1$, that is to say, $\xi^{\varepsilon,x}$ is a recurrent process, cf. [Dur96, Section 6.3]. Set also $\tau^{(a,b)}(\xi^{\varepsilon,x}) := \inf\{s > 0 \,|\, \xi_s^{\varepsilon,x} \notin (a,b)\}$ for any interval $(a,b) \subset \mathbb{R}$ and $x \in (a,b)$. By [GS68, §15, Corollary 15] and some rearrangements of terms, we get the formula

$$\mathbb{E}\, \tau^{(a,b)}(\xi^{\varepsilon,x}) = \frac{x-a}{b-a} \int_x^b 2(b-z)\rho_\varepsilon(z)^2 \, dz + \frac{b-x}{b-a} \int_a^x 2(z-a)\rho_\varepsilon(z)^2 \, dz.$$

The following lemma gives explicit formulas for mean hitting times of the process $\xi^{\varepsilon,x}$ in terms of its invariant density. The proof for it is found under [GS68, §18, Lemma 1]. As $\varepsilon > 0$ is fixed in this case, nothing needs to be modified.

Lemma 2.1.4. *Under Assumptions (C), (MET), for fixed $\varepsilon > 0$, and for any $x, y \in \mathbb{R}$ with $y < x$, we have*

$$\mathbb{E}\, \tau^y\left(\xi^{\varepsilon,x}\right) = 2(x-y) \int_x^\infty \rho_\varepsilon(z)^2\, dz + 2 \int_y^x (z-y)\rho_\varepsilon(z)^2\, dz, \tag{2.1.30}$$

while for $x < y$, we have

$$\mathbb{E}\, \tau^y\left(\xi^{\varepsilon,x}\right) = 2(y-x) \int_{-\infty}^x \rho_\varepsilon(z)^2\, dz + 2 \int_x^y (y-z)\rho_\varepsilon(z)^2\, dz. \tag{2.1.31}$$

The next proposition essentially shows that time averages of the process forget their initial condition as the time horizon grows.

Proposition 2.1.5. *Let $(\varphi_\varepsilon)_{\varepsilon>0}$ be a uniformly bounded sequence of Borel-measurable, real-valued functions on $\mathbb{R}$. Under Assumptions (C), (MET), and for any $x, y \in \mathbb{R}$ and $\varepsilon > 0$, we have*

$$\left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \mathbb{E}\, \varphi_\varepsilon(\xi_t^{\varepsilon,x})\, dt - \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \mathbb{E}\, \varphi_\varepsilon(\xi_t^{\varepsilon,y})\, dt \right| \le \frac{4 \sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty D_\rho |x-y|}{T_\varepsilon}. \tag{2.1.32}$$

Proof. The proof is basically analog to the original one, see [GS68, §18, Remark 1, Lemma 2], with the only difference being the estimate

$$\mathbb{E}\, \tau^y\left(\xi^{\varepsilon,x}\right) \le 2D_\rho |x-y|,$$

which is an immediate consequence of Lemma 2.1.4 and Lemma 2.1.3 ii). □

Before we proceed, we have to introduce a sequence of nondecreasing stopping times that will aid us several times in the subsequent proofs. The introduction of this stopping time is mostly a technical inconvenience due to the local Lipschitz conditions on the coefficients but, nonetheless, required to make the arguments that follow rigorous. In the subsequent material, a lot of the notation and the results are explained in Appendix A, so that we do not always introduce everything. To carry out the localization procedure through the stopping times, fix $\varepsilon > 0$, $x \in \mathbb{R}$, and a bounded, Borel-measurable function $\varphi_\varepsilon \colon \mathbb{R} \to \mathbb{R}$. We define the following processes

$$K_t^\varepsilon := \int_0^t \varphi_\varepsilon(\xi_s^{\varepsilon,x})\, ds, \quad t \in [0,\infty), \tag{2.1.33}$$

$$M_t^{\varepsilon,z} := \int_0^t \operatorname{sgn}(\xi_s^{\varepsilon,x} - z)\rho_\varepsilon^{-1}(\xi_s^{\varepsilon,x})\, dW_s, \quad t \in [0,\infty), \quad z \in \mathbb{R}, \tag{2.1.34}$$

where sgn is defined to be left-continuous, i.e., $\operatorname{sgn} = \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0]}$. The appearing integrands, thus the integrals, are measurable and adapted to the filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$. For any $t \ge 0$, observe that we have

$$\int_0^t \rho_\varepsilon^{-2}(\xi_s^{\varepsilon,x})\, ds \le t \sup_{0\le s\le t} \rho_\varepsilon^{-2}(\xi_s^{\varepsilon,x}) < \infty, \quad \mathbb{P}\text{-a.s.},$$

by continuity of ρ_ε^{-1} and $\xi^{\varepsilon,x}$, so that $(M_t^{\varepsilon,z})_{t\in[0,\infty)} \in \mathscr{M}^{\mathrm{c,\,loc}}$ for each $z \in \mathbb{R}$ and $\varepsilon > 0$, where $\mathscr{M}^{\mathrm{c,\,loc}}$ is the space of continuous local martingales starting in zero. Given the boundedness of φ_ε and the fact that the quadratic variation process $\langle M^{\varepsilon,z}\rangle$ satisfies $d\langle M^{\varepsilon,z}\rangle_s = \rho_\varepsilon^{-2}(\xi_s^{\varepsilon,x})\,ds$, a similar argument shows that

$$\Big(I_t^{\varepsilon,z} := \int_0^t K_s^\varepsilon \, dM_s^{\varepsilon,z}\Big)_{t\in[0,\infty)} \in \mathscr{M}^{\mathrm{c,\,loc}}$$

for each $z\in\mathbb{R}$. Hence, by definition of $\mathscr{M}^{\mathrm{c,\,loc}}$, there exists a nondecreasing sequence of stopping times $(\tau_n^{\varphi_\varepsilon,x})_{n\in\mathbb{N}}$ of $(\mathcal{F}_t)_{t\in[0,\infty)}$ such that $\tau_n^{\varphi_\varepsilon,x}\to\infty$, $\mathbb{P}$-a.s., as $n\to\infty$ and

$$\Big(M^{\varepsilon,z}_{t\wedge\tau_n^{\varphi_\varepsilon,x}}\Big)_{t\in[0,\infty)}, \quad \Big(I^{\varepsilon,z}_{t\wedge\tau_n^{\varphi_\varepsilon,x}}\Big)_{t\in[0,\infty)} \in \mathscr{M}_2^{\mathrm{c}}, \tag{2.1.35}$$

for each $n\in\mathbb{N}$, where $t\wedge\tau_n^{\varphi_\varepsilon,x} = \min\{t,\tau_n^{\varphi_\varepsilon,x}\}$ and $\mathscr{M}_2^{\mathrm{c}}$ is the space of continuous, square-integrable martingales starting in zero. Note that the quadratic variation process of $I^{\varepsilon,z}$ is given by

$$\langle I^{\varepsilon,z}\rangle_t = \int_0^t (K_s^\varepsilon)^2\,\rho_\varepsilon^{-2}(\xi_s^{\varepsilon,x})\,ds, \quad t\in[0,\infty).$$

In particular, it does not depend on $z\in\mathbb{R}$. Therefore, and because $\langle I^{\varepsilon,z}\rangle$ and $\xi^{\varepsilon,x}$ are continuous, we can additionally choose the localizing sequence $\tau_n^{\varphi_\varepsilon,x}$ in such a way that

$$\langle I^{\varepsilon,z}\rangle_{\cdot\wedge\tau_n^{\varphi_\varepsilon,x}} \le n, \quad \text{and} \quad |\xi^{\varepsilon,x}_{\cdot\wedge\tau_n^{\varphi_\varepsilon,x}}| \le n, \quad \mathbb{P}\text{-a.s.}, \tag{2.1.36}$$

for each $n\in\mathbb{N}$ and all $z\in\mathbb{R}$. Indeed, all we have to do is replace $\tau_n^{\varphi_\varepsilon,x}$ by $s_n\wedge\tau_n^{\varphi_\varepsilon,x}$ where $s_n := \inf\{t\ge 0 \,|\, \langle I^{\varepsilon,z}\rangle_t \ge n \text{ or } |\xi_t^{\varepsilon,x}|\ge n\}$. Here, we used the notation $\xi^{\varepsilon,x}_{\cdot\wedge\tau_n^{\varphi_\varepsilon,x}}$ for the map $[0,\infty)\ni t\mapsto \xi^{\varepsilon,x}_{t\wedge\tau_n^{\varphi_\varepsilon,x}}$ defined elementwise on Ω, and similarly for $\langle I^{\varepsilon,z}\rangle_{\cdot\wedge\tau_n^{\varphi_\varepsilon,x}}$.

In addition to the previous localization procedure, recall the Meyer-Tanaka formula (A.1.10) for a continuous local martingale, say $Y_t = Y_0 + M_t$ with $(M_t)_{t\in[0,\infty)}\in\mathscr{M}^{\mathrm{c,\,loc}}$, that is, for $z\in\mathbb{R}$ and $t\in[0,\infty)$, the identity

$$|Y_t - z| = |Y_0 - z| + \int_0^t \operatorname{sgn}(Y_s - z)\,dM_s + L_t^z(Y), \quad \mathbb{P}\text{-a.s.}, \tag{2.1.37}$$

holds. Here, $L_t^z(Y)$ is the local time process for Y, cf. Theorem A.1.9. Recall also the occupation density formula (A.1.9) that gives us the identity

$$\int_0^t k(Y_s)\,d\langle M\rangle_s = \int_{\mathbb{R}} k(z)L_t^z(Y)\,dz, \quad t\in[0,\infty), \quad \mathbb{P}\text{-a.s.}, \tag{2.1.38}$$

for every Borel-measurable function $k\colon\mathbb{R}\to[0,\infty)$.

We will now utilize these formulas and the introduced sequence of stopping times to prove a result similar to Proposition 2.1.5. The proof idea is borrowed from [Sun89], but note that we do not assume global Lipschitz conditions here, and we need an explicit estimate rather than just a qualitative convergence statement.

Proposition 2.1.6. *Let $(\varphi_\varepsilon)_{\varepsilon>0}$ be a uniformly bounded sequence of Borel-measurable, nonnegative functions on $\mathbb{R}$ with $\bigcup_{\varepsilon>0}\operatorname{supp}(\varphi_\varepsilon)\subset[-R,R]$ for some $R>0$. Under Assumptions (C) and (MET), and for any $x,y\in\mathbb{R}$ and $\varepsilon>0$, we have*

$$\left|\frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} t\,\mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x})\,dt - \frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} t\,\mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt\right| \le \frac{8\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty D_\rho|x-y|}{T_\varepsilon}. \tag{2.1.39}$$

Proof. Assume without loss of generality that $x \geq y$. Using Fubini's theorem, we first have

$$
\begin{aligned}
&\frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} t\,\mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x})\,dt - \frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} t\,\mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt \\
=&\frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon}\int_s^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x}) - \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt\,ds \qquad (2.1.40)\\
=&\frac{2}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x}) - \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt - \frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon}\mathbb{E}\int_0^s \varphi_\varepsilon(\xi_t^{\varepsilon,x}) - \varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt\,ds.
\end{aligned}
$$

Define $\tau_n := \tau_n^{\varphi_\varepsilon,x} \wedge \tau_n^{\varphi_\varepsilon,y}$ through the previously constructed stopping times. With the occupation density formula (2.1.38), we can write

$$
\mathbb{E}\int_0^{s\wedge\tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x}) - \varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt = \mathbb{E}\int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\left[L_{s\wedge\tau_n}^z(\xi^{\varepsilon,x}) - L_{s\wedge\tau_n}^z(\xi^{\varepsilon,y})\right]dz \qquad (2.1.41)
$$

for any $s \in [0,\infty)$. The Meyer-Tanaka formula (2.1.37) gives

$$
L_{s\wedge\tau_n}^z(\xi^{\varepsilon,x}) = |\xi_{s\wedge\tau_n}^{\varepsilon,x} - z| - |x-z| - \int_0^{s\wedge\tau_n} \operatorname{sgn}(\xi_r^{\varepsilon,x} - z)\rho_\varepsilon^{-1}(\xi_r^{\varepsilon,x})\,dW_r, \quad \mathbb{P}\text{-a.s.},
$$

and similarly for $L_{s\wedge\tau_n}^z(\xi^{\varepsilon,y})$. The respective stochastic integrals in these formulas are martingales, cf. (2.1.35), starting in zero and, thus, have expectation zero. Hence, it follows that

$$
\mathbb{E}\left[L_{s\wedge\tau_n}^z(\xi^{\varepsilon,x}) - L_{s\wedge\tau_n}^z(\xi^{\varepsilon,y})\right] = \mathbb{E}\,|\xi_{s\wedge\tau_n}^{\varepsilon,x} - z| - |x-z| - \mathbb{E}\,|\xi_{s\wedge\tau_n}^{\varepsilon,y} - z| + |y-z|. \qquad (2.1.42)
$$

With the uniqueness of the solution to the SDE (2.1.19) and $x \geq y$, a stopping time argument shows that $\xi_r^{\varepsilon,x} \geq \xi_r^{\varepsilon,y}$ for all $r \in [0,\infty)$, $\mathbb{P}$-a.s. Indeed, if τ is the first time that $\xi^{\varepsilon,x}$ and $\xi^{\varepsilon,y}$ coincide, then, by uniqueness, they must do so for all times afterwards as well. Using this fact, a simple case distinction and the triangle inequality show that

$$
\begin{aligned}
\mathbb{E}\,|\xi_{s\wedge\tau_n}^{\varepsilon,x} - z| - \mathbb{E}\,|\xi_{s\wedge\tau_n}^{\varepsilon,y} - z| &\leq \mathbb{E}\left[\xi_{s\wedge\tau_n}^{\varepsilon,x} - \xi_{s\wedge\tau_n}^{\varepsilon,y}\right] = x - y,\\
|y-z| - |x-z| &\leq x - y,
\end{aligned}
$$

where we also used the martingale property in the first line. Therefore,

$$
\mathbb{E}\left[L_{s\wedge\tau_n}^z(\xi^{\varepsilon,x}) - L_{s\wedge\tau_n}^z(\xi^{\varepsilon,y})\right] \leq 2(x-y). \qquad (2.1.43)
$$

Observe that by (2.1.36), the compact support of φ_ε, and the Burkholder-Davis-Gundy inequalities (A.1.8), we can upper bound and obtain

$$
\begin{aligned}
&\mathbb{E}\int_{\mathbb{R}} |\varphi_\varepsilon(z)|\rho_\varepsilon(z)^2 L_{s\wedge\tau_n}^z(\xi^{\varepsilon,x})\,dz\\
\leq&\,\mathbb{E}\int_{\mathbb{R}} |\varphi_\varepsilon(z)|\rho_\varepsilon(z)^2\left[|\xi_{s\wedge\tau_n}^{\varepsilon,x} - z| + |x-z| + \left|\int_0^{s\wedge\tau_n} \operatorname{sgn}(\xi_r^{\varepsilon,x} - z)\rho_\varepsilon^{-1}(\xi_r^{\varepsilon,x})\,dW_r\right|\right]dz\\
\leq&\,\|\varphi_\varepsilon\|_\infty \int_{[-R,R]} \rho_\varepsilon(z)^2\left[\mathbb{E}\,|\xi_{s\wedge\tau_n}^{\varepsilon,x}| + |x| + 2|z| + \mathbb{E}\left|\int_0^{s\wedge\tau_n} \operatorname{sgn}(\xi_r^{\varepsilon,x} - z)\rho_\varepsilon^{-1}(\xi_r^{\varepsilon,x})\,dW_r\right|\right]dz
\end{aligned}
$$

$$\leq \|\varphi_\varepsilon\|_\infty D_\rho \left[n + |x| + 2R + B_{1/2}\mathbb{E}\,\langle I^{\varepsilon,z}\rangle_{s\wedge\tau_n}^{1/2}\right] \leq \|\varphi_\varepsilon\|_\infty D_\rho \left[n + |x| + 2R + B_{1/2}n^{1/2}\right] < \infty,$$

where $B_{1/2} > 0$ is the universal constant coming from the Burkholder-Davis-Gundy inequalities for $m = 1/2$. A completely analog estimate is obtained if $L^z_{s\wedge\tau_n}(\xi^{\varepsilon,x})$ is replaced with $L^z_{s\wedge\tau_n}(\xi^{\varepsilon,y})$. Hence, we can use Fubini's theorem together with (2.1.43) to obtain the estimate

$$\mathbb{E}\int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2 \left[L^z_{s\wedge\tau_n}(\xi^{\varepsilon,x}) - L^z_{s\wedge\tau_n}(\xi^{\varepsilon,y})\right] dz \leq 2(x-y)\int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz.$$

Letting $n \to \infty$, the dominated convergence theorem implies through (2.1.41) and the preceding inequality

$$\mathbb{E}\int_0^s \varphi_\varepsilon(\xi_t^{\varepsilon,x}) - \varphi_\varepsilon(\xi_t^{\varepsilon,y})\, dt \leq 2(x-y)\int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz \leq 2(x-y)\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty D_\rho$$

for any $s \in [0,\infty)$. Therefore, we conclude from (2.1.40) and Proposition 2.1.5

$$\left|\frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} t\mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x})\, dt - \frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} \mathbb{E}\, t\varphi_\varepsilon(\xi_t^{\varepsilon,y})\, dt\right| \leq \frac{8\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty D_\rho(x-y)}{T_\varepsilon}. \qquad \square$$

Since $\xi^{\varepsilon,x}$ is a Markov process, cf. Appendix A and Theorem A.2.3, we have the following stationarity result, whose proof can be obtained, once again, by the same arguments as in the original material in [GS68, §18] since ε is fixed here.

Corollary 2.1.7. *Let $\varepsilon > 0$. Under Assumptions (C), (MET), for any $t \in [0,\infty)$ and any Borel-measurable function $\varphi_\varepsilon \colon \mathbb{R} \to \mathbb{R}$ with $\int_{\mathbb{R}} |\varphi_\varepsilon(x)|\rho_\varepsilon(x)^2\, dx < \infty$, we have*

$$\int_{\mathbb{R}} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\rho_\varepsilon(y)^2\, dy = \int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\, dy. \tag{2.1.44}$$

The following lemma gives us another key ingredient for our sought-after main result. If we have a convergence like in (2.1.23), then this results shows that time averages of the process $\xi^{\varepsilon,x}$ converge in expectation to the space integrals with respect to the invariant measure as the time horizon grows.

Lemma 2.1.8. *Let $(\varphi_\varepsilon)_{\varepsilon>0}$ be a uniformly bounded sequence of Borel-measurable, non-negative functions on $\mathbb{R}$. Under Assumptions (C) and (MET), we have*

$$\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\, dy\right| \to 0, \quad \varepsilon \to 0, \tag{2.1.45}$$

for $x_\varepsilon := f_\varepsilon(x_0)$. Similarly, if $\bigcup_{\varepsilon>0}\operatorname{supp}(\varphi_\varepsilon) \subset [-R,R]$ for some $R > 0$, then

$$\left|\frac{2}{T_\varepsilon^2}\int_0^{T_\varepsilon} t\,\mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\, dy\right| \to 0, \quad \varepsilon \to 0. \tag{2.1.46}$$

Proof. We will only prove the first claim (2.1.45) as the proof of the second claim works the same. For that second claim we would only have to replace estimate (2.1.32) with (2.1.39) in the following.

To prove claim (2.1.45), we first obtain from equation (2.1.44)

$$\begin{aligned}
&\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right| \\
=&\frac{1}{C_{\rho_\varepsilon}}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\ \int_{\mathbb{R}} \rho_\varepsilon(y)^2\,dy - \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\int_{\mathbb{R}} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\rho_\varepsilon(y)^2\,dy\,dt\right| \\
=&\frac{1}{C_{\rho_\varepsilon}}\left|\int_{\mathbb{R}}\left(\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) - \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt\right)\rho_\varepsilon(y)^2\,dy\right| \\
\le&\frac{1}{d_\rho}\Bigg[2\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty\int_{|x_\varepsilon-y|>T_\varepsilon}\rho_\varepsilon(y)^2\,dy \\
&\qquad+\int_{|x_\varepsilon-y|\le T_\varepsilon}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) - \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,y})\,dt\right|\rho_\varepsilon(y)^2\,dy\Bigg].
\end{aligned}$$

For the second term inside the brackets of the last line, we use estimate (2.1.32) to get

$$\begin{aligned}
&\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right| \\
\le&\frac{2\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty}{d_\rho}\left[\int_{|x_\varepsilon-y|>T_\varepsilon}\rho_\varepsilon(y)^2\,dy + \frac{2D_\rho}{T_\varepsilon}\int_{|x_\varepsilon-y|\le T_\varepsilon}|x_\varepsilon-y|\rho_\varepsilon(y)^2\,dy\right].
\end{aligned} \tag{2.1.47}$$

Now let $\delta>0$ be arbitrary. By Lemma 2.1.3 iii) we can choose $M>0$ such that for all $\varepsilon>0$,

$$\int_{|y|>M}\rho_\varepsilon(y)^2\,dy<\delta.$$

Therefore, for sufficiently small $\varepsilon>0$, we have with (2.1.29)

$$\int_{|x_\varepsilon-y|>T_\varepsilon}\rho_\varepsilon(y)^2\,dy \le \int_{|y|>T_\varepsilon-|x_\varepsilon|}\rho_\varepsilon(y)^2\,dy \le \int_{|y|>T_\varepsilon-(c_f\vee C_f)|f(x_0)|}\rho_\varepsilon(y)^2\,dy<\delta,$$

and

$$\begin{aligned}
&\frac{1}{T_\varepsilon}\int_{|x_\varepsilon-y|\le T_\varepsilon}|x_\varepsilon-y|\rho_\varepsilon(y)^2\,dy \\
=&\frac{1}{T_\varepsilon}\left[\int_{|x_\varepsilon-y|\le T_\varepsilon,\,|y|\le M}|x_\varepsilon-y|\rho_\varepsilon(y)^2\,dy + \int_{|x_\varepsilon-y|\le T_\varepsilon,\,|y|>M}|x_\varepsilon-y|\rho_\varepsilon(y)^2\,dy\right] \\
\le&\frac{[(c_f\vee C_f)|f(x_0)|+M]\,D_\rho}{T_\varepsilon} + \int_{|y|>M}\rho_\varepsilon(y)^2\,dy \\
\le&\frac{[(c_f\vee C_f)|f(x_0)|+M]\,D_\rho}{T_\varepsilon} + \delta.
\end{aligned}$$

Thus, it follows

$$\limsup_{\varepsilon\to0}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{E}\,\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right| \le \frac{2\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty(1+2D_\rho)}{d_\rho}\delta.$$

Letting $\delta\to0$ yields the claim. □

The next lemma will be used twice in the proof of the MET. Compared to the results in [Sun89], we again resort to localization arguments with the previously introduced sequence of stopping times to account for the lack of global Lipschitz conditions; recall the paragraphs before Proposition 2.1.6 for the definition of those stopping times.

Lemma 2.1.9. *Let $\varphi_\varepsilon, \psi_\varepsilon$ be bounded, Borel-measurable, nonnegative functions such that* $\operatorname{supp}(\varphi_\varepsilon) \subset [-R_1, R_1]$ *and* $\operatorname{supp}(\psi_\varepsilon) \subset [-R_2, R_2]$ *for some $R_1, R_2 > 0$. Set $\tau_n := \tau_n^{\varphi_\varepsilon, x_\varepsilon}$. Under Assumptions (C) and (MET), and for all $\varepsilon > 0$ and $n \in \mathbb{N}$, we have*

$$\begin{aligned} &\mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \int_t^{T_\varepsilon \wedge \tau_n} \psi_\varepsilon(\xi_s^{\varepsilon, x_\varepsilon})\, ds\, dt \\ &= \mathbb{E} \int_{\mathbb{R}} \psi_\varepsilon(z) \rho_\varepsilon(z)^2 \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \Big[|\xi_{T_\varepsilon \wedge \tau_n}^{\varepsilon, x_\varepsilon} - z| - |\xi_t^{\varepsilon, x_\varepsilon} - z| \Big]\, dt\, dz. \end{aligned} \tag{2.1.48}$$

Proof. Except where expectations occur, the following equations should be understood $\mathbb{P}$-almost surely. Applying the occupation density formula (2.1.38) gives

$$\begin{aligned} &\int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \int_t^{T_\varepsilon \wedge \tau_n} \psi_\varepsilon(\xi_s^{\varepsilon, x_\varepsilon})\, ds\, dt \\ &= \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \int_{\mathbb{R}} \psi_\varepsilon(z) \rho_\varepsilon(z)^2 \Big[L_{T_\varepsilon \wedge \tau_n}^z(\xi^{\varepsilon, x_\varepsilon}) - L_t^z(\xi^{\varepsilon, x_\varepsilon}) \Big]\, dz\, dt \\ &= \int_{\mathbb{R}} \psi_\varepsilon(z) \rho_\varepsilon(z)^2 \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \Big[L_{T_\varepsilon \wedge \tau_n}^z(\xi^{\varepsilon, x_\varepsilon}) - L_t^z(\xi^{\varepsilon, x_\varepsilon}) \Big]\, dt\, dz, \end{aligned}$$

where we used Tonelli's theorem in the last equality since $L_t^z(\xi^{\varepsilon, x_\varepsilon})$ is nondecreasing in t, cf. Theorem A.1.9. The Meyer-Tanaka formula (2.1.37) yields

$$\begin{aligned} L_{T_\varepsilon \wedge \tau_n}^z(\xi^{\varepsilon, x_\varepsilon}) - L_t^z(\xi^{\varepsilon, x_\varepsilon}) &= |\xi_{T_\varepsilon \wedge \tau_n}^{\varepsilon, x_\varepsilon} - z| - |\xi_t^{\varepsilon, x_\varepsilon} - z| - \int_t^{T_\varepsilon \wedge \tau_n} \operatorname{sgn}(\xi_s^{\varepsilon, x_\varepsilon} - z) \rho_\varepsilon^{-1}(\xi_s^{\varepsilon, x_\varepsilon})\, dW_s \\ &= |\xi_{T_\varepsilon \wedge \tau_n}^{\varepsilon, x_\varepsilon} - z| - |\xi_t^{\varepsilon, x_\varepsilon} - z| - \Big[M_{T_\varepsilon \wedge \tau_n}^{\varepsilon, z} - M_t^{\varepsilon, z} \Big], \end{aligned}$$

so that

$$\begin{aligned} &\int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \int_t^{T_\varepsilon \wedge \tau_n} \psi_\varepsilon(\xi_s^{\varepsilon, x_\varepsilon})\, ds\, dt \\ &= \int_{\mathbb{R}} \psi_\varepsilon(z) \rho_\varepsilon(z)^2 \Bigg[\int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \Big[|\xi_{T_\varepsilon \wedge \tau_n}^{\varepsilon, x_\varepsilon} - z| - |\xi_t^{\varepsilon, x_\varepsilon} - z| \Big]\, dt \\ &\qquad - \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon}) \Big[M_{T_\varepsilon \wedge \tau_n}^{\varepsilon, z} - M_t^{\varepsilon, z} \Big]\, dt \Bigg] dz, \end{aligned}$$

since $M^{\varepsilon, z}$, $\xi^{\varepsilon, x_\varepsilon}$ are continuous processes, $\mathbb{P}$-a.s., and φ_ε is bounded with compact support. Stochastic integration by parts, see Lemma A.1.7, implies

$$\int_0^{T_\varepsilon \wedge \tau_n} K_t^\varepsilon\, dM_t^{\varepsilon, z} = K_{T_\varepsilon \wedge \tau_n}^\varepsilon M_{T_\varepsilon \wedge \tau_n}^{\varepsilon, z} - \int_0^{T_\varepsilon \wedge \tau_n} M_t^{\varepsilon, z} \varphi_\varepsilon(\xi_t^{\varepsilon, x_\varepsilon})\, dt$$

$$= \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) \left[M^{\varepsilon,z}_{T_\varepsilon \wedge \tau_n} - M_t^{\varepsilon,z} \right] dt,$$

or equivalently

$$\int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) \int_t^{T_\varepsilon \wedge \tau_n} \psi_\varepsilon(\xi_s^{\varepsilon,x_\varepsilon})\, ds\, dt$$

$$= \int_{\mathbb{R}} \psi_\varepsilon(z)\rho_\varepsilon(z)^2 \Bigg[\int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) \left[|\xi^{\varepsilon,x_\varepsilon}_{T_\varepsilon \wedge \tau_n} - z| - |\xi_t^{\varepsilon,x_\varepsilon} - z| \right] dt$$

$$- \int_0^{T_\varepsilon \wedge \tau_n} K_t^\varepsilon \, dM_t^{\varepsilon,z} \Bigg] dz.$$

In what follows, we require integrability in order to apply linearity and Fubini's theorem with respect to $\mathbb{E}$. For this, observe that, by (2.1.28) and (2.1.36), we can estimate

$$\mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) \int_{\mathbb{R}} \psi_\varepsilon(z)\rho_\varepsilon(z)^2 \left[|\xi^{\varepsilon,x_\varepsilon}_{T_\varepsilon \wedge \tau_n} - z| + |\xi_t^{\varepsilon,x_\varepsilon} - z| \right] dz\, dt$$

$$\leq T_\varepsilon \|\varphi_\varepsilon\|_\infty \|\psi_\varepsilon\|_\infty D_\rho \left(\mathbb{E}\, |\xi^{\varepsilon,x_\varepsilon}_{T_\varepsilon \wedge \tau_n}| + 2R_2 + \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \mathbb{E}\, |\xi_t^{\varepsilon,x_\varepsilon}|\, dt \right)$$

$$\leq T_\varepsilon \|\varphi_\varepsilon\|_\infty \|\psi_\varepsilon\|_\infty D_\rho \left(n + 2R_2 + \frac{|f_\varepsilon(x_0)|}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(t)\, dt \right) < \infty,$$

and, using the Burkholder-Davis-Gundy inequalities (A.1.8) and (2.1.36), we can upper bound

$$\mathbb{E} \int_{\mathbb{R}} \psi_\varepsilon(z)\rho_\varepsilon(z)^2 \left| \int_0^{T_\varepsilon \wedge \tau_n} K_t^\varepsilon \, dM_t^{\varepsilon,z} \right| dz \leq B_{1/2} \int_{\mathbb{R}} \psi_\varepsilon(z)\rho_\varepsilon(z)^2\, \mathbb{E}\, \langle I^{\varepsilon,z} \rangle^{1/2}_{T_\varepsilon \wedge \tau_n} dz$$

$$\leq B_{1/2} \|\psi_\varepsilon\|_\infty D_\rho n^{1/2} < \infty.$$

Hence, it follows

$$\mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) \int_t^{T_\varepsilon \wedge \tau_n} \psi_\varepsilon(\xi_s^{\varepsilon,x_\varepsilon})\, ds\, dt$$

$$= \int_{\mathbb{R}} \psi_\varepsilon(z)\rho_\varepsilon(z)^2 \Bigg[\mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon}) \left[|\xi^{\varepsilon,x_\varepsilon}_{T_\varepsilon \wedge \tau_n} - z| - |\xi_t^{\varepsilon,x_\varepsilon} - z| \right] dt$$

$$- \mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} K_t^\varepsilon \, dM_t^{\varepsilon,z} \Bigg] dz.$$

To reach the claim we finally use the fact that the last displayed expectation is zero due to (2.1.35). □

Using these preparatory results, we are now in a position to state and prove the MET.

Theorem 2.1.10. *Let $(\varphi_\varepsilon)_{\varepsilon>0}$ be a uniformly bounded sequence of Borel-measurable, real-valued functions on $\mathbb{R}$. Under Assumptions (C) and (MET), and for $x_\varepsilon = f_\varepsilon(x_0)$ and $T_\varepsilon \to \infty$ as $\varepsilon \to 0$, we have*

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \to 0, \quad \varepsilon \to 0. \tag{2.1.49}$$

Remark 2.1.11. *In terms of the original quantities, i.e., with respect to the process X^ε that solves* (2.1.15) *and the invariant density μ_ε in* (2.1.27)*, the limit result* (2.1.49) *reads as*

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi_\varepsilon(y)\mu_\varepsilon(y)\,dy\right|^2 \to 0, \quad \varepsilon \to 0. \tag{2.1.50}$$

If we further assume the convergence

$$\int_{\mathbb{R}} \varphi_\varepsilon(y)\mu_\varepsilon(y)\,dy \to \int_{\mathbb{R}} \varphi(y)\mu(y)\,dy, \quad \varepsilon \to 0, \tag{2.1.51}$$

for a bounded, Borel-measurable function $\varphi\colon \mathbb{R} \to \mathbb{R}$, then we obtain the final MET

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi(y)\mu(y)\,dy\right|^2 \to 0, \quad \varepsilon \to 0. \tag{2.1.52}$$

Proof. We will first prove the result for the case where the functions φ_ε are nonnegative with $\bigcup_{\varepsilon>0}\operatorname{supp}(\varphi_\varepsilon) \subset [-R, R]$ for some $R > 0$. Note that we can write

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \\
&= \mathbb{E}\left[\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2\right] - \left(\frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right)^2 \\
&\quad - \mathbb{E}\left(\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right)\frac{2}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy.
\end{aligned}$$

It is enough to prove

$$\limsup_{\varepsilon\to 0}\mathbb{E}\left[\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2\right] - \left(\frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right)^2 \le 0 \tag{2.1.53}$$

because (2.1.49) then follows from the assumptions and Lemma 2.1.8. For this, let $\delta > 0$ be arbitrary. Choose $M > 0$ such that for all $\varepsilon > 0$,

$$\int_{|y|>M} \rho_\varepsilon(y)^2\,dy < \delta.$$

With the sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ from Lemma 2.1.9, for all $\varepsilon > 0$, and any $n \in \mathbb{N}$, we have

$$\begin{aligned}
&\mathbb{E}\int_0^{T_\varepsilon\wedge\tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\int_t^{T_\varepsilon\wedge\tau_n} \varphi_\varepsilon(\xi_s^{\varepsilon,x_\varepsilon})\,ds\,dt \\
&= \mathbb{E}\int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\int_0^{T_\varepsilon\wedge\tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\left[|\xi_{T_\varepsilon\wedge\tau_n}^{\varepsilon,x_\varepsilon} - z| - |\xi_t^{\varepsilon,x_\varepsilon} - z|\right]dt\,dz,
\end{aligned} \tag{2.1.54}$$

and

$$\begin{aligned}
&\mathbb{E}\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\int_t^{T_\varepsilon\wedge\tau_n}\mathbb{1}_{[-M,M]}(\xi_s^{\varepsilon,x_\varepsilon})\,ds\,dt \\
&=\mathbb{E}\int_{|y|\le M}\rho_\varepsilon(y)^2\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\left[|\xi_{T_\varepsilon\wedge\tau_n}^{\varepsilon,x_\varepsilon}-y|-|\xi_t^{\varepsilon,x_\varepsilon}-y|\right]dt\,dy.
\end{aligned} \tag{2.1.55}$$

First, observe that

$$\begin{aligned}
&\int_{\mathbb{R}}\rho_\varepsilon(y)^2\,dy\,\mathbb{E}\left[\left|\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2\right] \\
&\le\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty\right)^2\mathbb{E}\,(T_\varepsilon\wedge\tau_n)^2\delta+\int_{|y|\le M}\rho_\varepsilon(y)^2\,dy\,\mathbb{E}\left[\left|\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2\right] \\
&\le\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty T_\varepsilon\right)^2\delta+\int_{|y|\le M}\rho_\varepsilon(y)^2\,dy\,\mathbb{E}\left[\left|\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2\right].
\end{aligned}$$

Let us analyze the second term in detail. Fubini's theorem, equation (2.1.54), and the triangle inequality allow us to estimate

$$\begin{aligned}
&\int_{|y|\le M}\rho_\varepsilon(y)^2\,dy\,\mathbb{E}\left[\left|\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2\right] \\
&=2\int_{|y|\le M}\rho_\varepsilon(y)^2\,dy\,\mathbb{E}\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\int_t^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_s^{\varepsilon,x_\varepsilon})\,ds\,dt \\
&=2\,\mathbb{E}\int_{|y|\le M}\rho_\varepsilon(y)^2\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\left[|\xi_{T_\varepsilon\wedge\tau_n}^{\varepsilon,x_\varepsilon}-z|-|\xi_t^{\varepsilon,x_\varepsilon}-z|\right]dt\,dz\,dy \\
&\le2\,\mathbb{E}\int_{|y|\le M}\rho_\varepsilon(y)^2\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\left[|\xi_{T_\varepsilon\wedge\tau_n}^{\varepsilon,x_\varepsilon}-y|-|\xi_t^{\varepsilon,x_\varepsilon}-y|\right]dt\,dz\,dy \\
&\quad+4\,\mathbb{E}\int_{|y|\le M}\rho_\varepsilon(y)^2\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})|y-z|\,dt\,dz\,dy.
\end{aligned}$$

For the first term of the last inequality, equation (2.1.55) implies that

$$\begin{aligned}
&\mathbb{E}\int_{|y|\le M}\rho_\varepsilon(y)^2\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\left[|\xi_{T_\varepsilon\wedge\tau_n}^{\varepsilon,x_\varepsilon}-y|-|\xi_t^{\varepsilon,x_\varepsilon}-y|\right]dt\,dz\,dy \\
&=\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\,dz\,\mathbb{E}\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\int_t^{T_\varepsilon\wedge\tau_n}\mathbb{1}_{[-M,M]}(\xi_s^{\varepsilon,x_\varepsilon})\,ds\,dt \\
&\le\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\,dz\,\mathbb{E}\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})(T_\varepsilon\wedge\tau_n-t)\,dt.
\end{aligned}$$

For the second term, our assumptions yield

$$\left|\mathbb{E}\int_{|y|\le M}\rho_\varepsilon(y)^2\int_{\mathbb{R}}\varphi_\varepsilon(z)\rho_\varepsilon(z)^2\int_0^{T_\varepsilon\wedge\tau_n}\varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})|y-z|\,dt\,dz\,dy\right|$$

$$\leq \mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt \int_{|y|\leq M} \rho_\varepsilon(y)^2 \int_{|z|\leq R} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2 |y-z|\, dz\, dy$$

$$\leq \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty D_\rho \right)^2 (M+R)T_\varepsilon.$$

Therefore, combining all these estimates yields

$$\int_{\mathbb{R}} \rho_\varepsilon(y)^2\, dy\, \mathbb{E} \left[\left| \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt \right|^2 \right]$$

$$\leq \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty T_\varepsilon \right)^2 \delta + 4 \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty D_\rho \right)^2 (M+R)T_\varepsilon$$

$$+ 2 \int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz\, \mathbb{E} \int_0^{T_\varepsilon \wedge \tau_n} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})(T_\varepsilon \wedge \tau_n - t)\, dt.$$

Letting $n \to \infty$ and using the dominated convergence theorem gives the estimate

$$\int_{\mathbb{R}} \rho_\varepsilon(y)^2\, dy\, \mathbb{E} \left[\left| \int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt \right|^2 \right] \leq \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty T_\varepsilon \right)^2 \delta + 4 \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty D_\rho \right)^2 (M+R)T_\varepsilon$$

$$+ 2 \int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz\, \mathbb{E} \int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})(T_\varepsilon - t)\, dt.$$

Hence, it follows

$$\mathbb{E} \left[\left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt \right|^2 \right] - \left(\frac{1}{C_{\rho_\varepsilon}} \int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz \right)^2$$

$$\leq \frac{1}{C_{\rho_\varepsilon}} \int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz \left[\frac{2}{T_\varepsilon^2} \mathbb{E} \int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})(T_\varepsilon - t)\, dt - \frac{1}{C_{\rho_\varepsilon}} \int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz \right]$$

$$+ \frac{\delta}{d_\rho} \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty \right)^2 + \frac{4}{C_{\rho_\varepsilon} T_\varepsilon} \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty D_\rho \right)^2 (M+R).$$

The RHS converges by Lemma 2.1.8 to $(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty)^2\, \delta/d_\rho$ as $\varepsilon \to 0$, so that

$$\limsup_{\varepsilon\to 0} \left(\mathbb{E} \left[\left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\, dt \right|^2 \right] - \left(\frac{1}{C_{\rho_\varepsilon}} \int_{\mathbb{R}} \varphi_\varepsilon(z)\rho_\varepsilon(z)^2\, dz \right)^2 \right) \leq \frac{\delta}{d_\rho} \left(\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty \right)^2. \tag{2.1.56}$$

Letting $\delta \to 0$ yields the claim (2.1.53). Thus, the convergence result (2.1.49) is established for Borel-measurable, nonnegative functions φ_ε with $\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty < \infty$ and $\bigcup_{\varepsilon>0} \operatorname{supp}(\varphi_\varepsilon) \subset [-R, R]$ for some $R > 0$. We now extend the convergence result to Borel-measurable, nonnegative functions φ_ε with $\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty < \infty$. For this, let $\delta > 0$ be arbitrary and choose again $M > 0$ such that for all $\varepsilon > 0$,

$$\int_{|y|>M} \rho_\varepsilon(y)^2\, dy < \delta.$$

Define $\psi_\varepsilon^{(m)} := \varphi_\varepsilon \mathbb{1}_{[-m^*,m^*]}$ with $m^* := m^*(m) := m \vee M$, $m \in \mathbb{N}$. We can upper bound

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \\
\leq{}& 8\Bigg[\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \psi_\varepsilon^{(m)}(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2 \\
&+ \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \psi_\varepsilon^{(m)}(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \\
&+ \left|\frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \left(\psi_\varepsilon^{(m)}(y) - \varphi_\varepsilon(y)\right)\rho_\varepsilon(y)^2\,dy\right|^2\Bigg].
\end{aligned}$$

We estimate the first term as follows

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \psi_\varepsilon^{(m)}(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2 \\
={}& \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\mathbb{1}_{[-m^*,m^*]^c}(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2 \\
\leq{}& \left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty\right)^2 \mathbb{E}\left|1 - \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{1}_{[-m^*,m^*]}(\xi_t^{\varepsilon,x_\varepsilon})\,dt\right|^2 \\
\leq{}& 4\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty\right)^2 \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{1}_{[-m^*,m^*]}(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{|y|\leq m^*} \rho_\varepsilon(y)^2\,dy\right|^2 \\
&+ 4\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty\right)^2 \left(\frac{1}{C_{\rho_\varepsilon}}\int_{|y|>m^*} \rho_\varepsilon(y)^2\,dy\right)^2 \\
\leq{}& 4\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty\right)^2 \left[\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \mathbb{1}_{[-m^*,m^*]}(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{|y|\leq m^*} \rho_\varepsilon(y)^2\,dy\right|^2 + \left(\frac{\delta}{d_\rho}\right)^2\right],
\end{aligned}$$

and the last upper bound converges to $4\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty \delta/d_\rho\right)^2$ as $\varepsilon \to 0$. By our previously established result in this proof, we also have convergence of the second term for any $m \in \mathbb{N}$ as $\varepsilon \to 0$. The third term satisfies

$$\begin{aligned}
\left|\frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \left(\psi_\varepsilon^{(m)}(y) - \varphi_\varepsilon(y)\right)\rho_\varepsilon(y)^2\,dy\right|^2 &= \left|\frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\mathbb{1}_{[-m^*,m^*]^c}(y)\rho_\varepsilon(y)^2\,dy\right|^2 \\
&\leq \left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty \frac{\delta}{d_\rho}\right)^2.
\end{aligned}$$

In summary, we have

$$\limsup_{\varepsilon\to 0}\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \leq 40\left(\sup_{\varepsilon>0}\|\varphi_\varepsilon\|_\infty \frac{\delta}{d_\rho}\right)^2,$$

which implies, by letting $\delta \to 0$,

$$\limsup_{\varepsilon\to 0} \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi_\varepsilon(\xi_t^{\varepsilon,x_\varepsilon})\,dt - \frac{1}{C_{\rho_\varepsilon}}\int_{\mathbb{R}} \varphi_\varepsilon(y)\rho_\varepsilon(y)^2\,dy\right|^2 \leq 0.$$

Hence, the convergence result holds for a uniformly bounded sequence $(\varphi_\varepsilon)_{\varepsilon>0}$ of Borel-measurable, nonnegative functions on $\mathbb{R}$, too. The asserted full result is now straightforward; it can be accomplished by splitting the test functions into positive and negative part and applying the preceding result to these nonnegative functions. □

Remarks on multidimensional extensions I.
We somehow simplified the proofs in this section by using stationarity with respect to the invariant measure as stated in Corollary 2.1.7 and the fact that we have a Markov process X^ε already at the ε-level. But the remarkable insight of Sundar's work [Sun89] is that ergodicity is not a Markovian property; it is a semimartingale property. If we briefly recall the multiscale system (1.2.1), then the best we can say about X^ε in a system where slow and fast variable are fully coupled is that it is a semimartingale. It is these two clues that might strike a path for a multidimensional mean ergodic theorem. We want to illustrate this in the following.

We will keep the exposition of the ideas formal and systematic without any guarantee for utmost rigor, because, while many of the claims would be natural in our setting, they may not exist yet in the current literature. In particular, we will not specify concrete regularity assumptions, but instead assume sufficient regularity and boundedness such that the main steps are illuminating and lead to a plausible result. This line of thought shall hold for any of the subsequent remarks on multidimensional limit theorems, too.

Let us consider a d-dimensional multiscale system as in (1.2.1) with 1-dimensional slow variable X^ε, $(d-1)$-dimensional fast variable Y^ε, and 1-dimensional homogenized limit process X as in (1.2.2). We assume that all appearing diffusion coefficients are globally nondegenerate, bounded on $\mathbb{R}$, and that the limit diffusion coefficient $\bar{\Sigma}^2$ is Lipschitz-continuous. The reason for considering a 1-dimensional slow variable is that in the preceding section we utilized the local time process, which exists for semimartingales in a meaningful sense only in dimension 1. Another reason is that the homogenization procedure is a particularly favorable case of reduction in dimensionality here, because it essentially eliminates the potentially high-dimensional fast variable. Our aim is to derive an MET of the form

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi(y)\mu(y)\,dy\right|^2 \to 0, \quad \varepsilon \to 0, \tag{2.1.57}$$

for a Lipschitz-continuous, bounded, nonnegative test function φ with compact support, where μ is the invariant density of X. Results for more general test functions, say continuous and bounded, may be accomplished through sensible approximations and decompositions. If that invariant density exists, then it is given by the formula from Remark 2.1.2

$$\mu(x) = \frac{1}{Z\bar{\Sigma}(x)^2}\exp\left(\int_0^x \frac{2\bar{F}(y)}{\bar{\Sigma}(y)^2}\,dy\right) = \frac{1}{Z\bar{\Sigma}(x)^2 f'(x)},$$

where f is the scale function of X, recall (2.1.17). In contrast to X, it does not make sense to define a scale function for X^ε as there is no generator for X^ε alone. Regardless of this, we can make the ideas of the preceding section still work and one of the main tools for this is the solution χ to the cell problem (1.2.6). In [RX21], the authors have thoroughly analyzed the regularity properties of χ in both arguments, cf. [RX21, Theorem 2.1], essentially giving pointwise estimates for all relevant derivatives of χ, with boundedness in the slow variable and at most polynomial growth in the fast variable. They also used the following uniform bounds

$$\sup_{t\in[0,\infty)}\sup_{\varepsilon\in(0,1]}\mathbb{E}\,|Y_t^\varepsilon|_2^m \le C_0(1+|y_0|_2^m), \tag{2.1.58}$$

for any $m>0$ and some constant $C_0>0$, which was originally proved in [Ver97], see also [PV01; Ver00]. Using these two facts, one can therefore get estimates like

$$\mathbb{E}\int_0^t |\chi(X_s^\varepsilon, Y_s^\varepsilon)|\,ds + \mathbb{E}\int_0^t |\nabla_y\chi(X_s^\varepsilon, Y_s^\varepsilon)|_2\,ds = \mathcal{O}\left(t\right), \tag{2.1.59}$$

and similar ones for higher derivatives as long as they exist. Apart from such estimates, we also need the following quantitative bounds, which were proved in [RX21] for various regimes,

$$\mathbb{E}\int_0^t h(X_s^\varepsilon, Y_s^\varepsilon)\,ds \le C_t\varepsilon, \tag{2.1.60}$$

$$\sup_{t\in[0,T]}|\mathbb{E}\,\varphi(X_t^\varepsilon) - \mathbb{E}\,\varphi(X_t)| \le C_T\varepsilon, \tag{2.1.61}$$

for any function h satisfying the centering condition

$$\int_{\mathbb{R}^{d-1}} h(x,y)\,d\mu^x(y) = 0, \quad x\in\mathbb{R}. \tag{2.1.62}$$

These were proved in [RX21, Lemma 4.2] and [RX21, Theorem 2.3], respectively, and their ”Regime 4” corresponds to our multiscale system (1.2.1) with scale parameter $\alpha_\varepsilon = \varepsilon$. Tracking the dependence on $t>0$, one finds $C_t = \mathcal{O}\left(t\right)$. One must remark at this point that their results are built upon the assumption that the coefficient functions in the multiscale system are bounded in the slow variable; an assumption that is violated for ergodic diffusion processes with an invariant density on the whole space. However, one might still get similar results in the unbounded case through a smooth cut-off of the coefficients and a standard exit stopping time argument. This would most certainly require uniform control of exit time probabilities for X^ε and X. For X, this can be accomplished through a Lyapunov function and Markov's inequality, whereas for X^ε, one should use a corrected version of said Lyapunov function with the help of the corrector term χ. Such a construction was, for example, carried out in [BP78, Section 5]. Despite these for now unresolved issues, we will pretend like we actually have the preceding results and estimates so that we can continue with the derivation of the MET.

We start by applying the Itô formula to χ and using the Poisson equation (1.2.6) to obtain

$$d\chi(X_t^\varepsilon, Y_t^\varepsilon) = \frac{1}{\varepsilon^2}\mathcal{L}_0\chi(X_t^\varepsilon, Y_t^\varepsilon)\,dt + \frac{1}{\varepsilon}\mathcal{L}_1\chi(X_t^\varepsilon, Y_t^\varepsilon) + \mathcal{L}_2\chi(X_t^\varepsilon, Y_t^\varepsilon) +$$

$$
\begin{aligned}
&\frac{1}{\varepsilon}\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(y)}+\partial_x\chi(X_t^\varepsilon,Y_t^\varepsilon)\Sigma(X_t^\varepsilon,Y_t^\varepsilon)dW_t^{(x)}\\
=&-\frac{1}{\varepsilon^2}H(X_t^\varepsilon,Y_t^\varepsilon)\,dt+\frac{1}{\varepsilon}\mathcal{L}_1\chi(X_t^\varepsilon,Y_t^\varepsilon)+\mathcal{L}_2\chi(X_t^\varepsilon,Y_t^\varepsilon)+\\
&\frac{1}{\varepsilon}\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(y)}+\partial_x\chi(X_t^\varepsilon,Y_t^\varepsilon)\Sigma(X_t^\varepsilon,Y_t^\varepsilon)dW_t^{(x)},
\end{aligned}
$$

where for $x\in\mathbb{R}$, $y\in\mathbb{R}^{d-1}$,

$$
\mathcal{L}_0:=\mathcal{L}_0(x,y)=\frac{1}{2}\sum_{i,j=1}^{d-1}a_{ij}(x,y)\frac{\partial^2}{\partial y_i\partial y_j}+\sum_{i=1}^{d-1}f_0^{(i)}(x,y)\frac{\partial}{\partial y_i}, \tag{2.1.63}
$$

$$
\mathcal{L}_1:=\mathcal{L}_1(x,y):=H(x,y)\frac{\partial}{\partial x}+\sum_{i=1}^{d-1}f_1^{(i)}(x,y)\frac{\partial}{\partial y_i}, \tag{2.1.64}
$$

$$
\mathcal{L}_2:=\mathcal{L}_2(x,y):=\frac{1}{2}\Sigma(x,y)^2\frac{\partial^2}{\partial x^2}+F(x,y)\frac{\partial}{\partial x}. \tag{2.1.65}
$$

We now define a corrected process $\hat{X}^\varepsilon:=X^\varepsilon+\varepsilon\chi(X^\varepsilon,Y^\varepsilon)$ for which we have, based on the previous calculations,

$$
\begin{aligned}
d\hat{X}_t^\varepsilon&=[F(X_t^\varepsilon,Y_t^\varepsilon)+\mathcal{L}_1\chi(X_t^\varepsilon,Y_t^\varepsilon)]\,dt+\Sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(x)}\\
&\quad+\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(y)}+\mathcal{O}(\varepsilon),
\end{aligned}
\tag{2.1.66}
$$

where we assumed that some of the remaining terms are of order $\mathcal{O}(\varepsilon)$. Observe here that the worrisome $\mathcal{O}(\varepsilon^{-1})$ term in the slow variable is canceled out perfectly through the correction $\varepsilon\chi(X^\varepsilon,Y^\varepsilon)$, demonstrating the usefulness of the Poisson equation and correction at path-level. Another application of the Itô formula to the scale function f, while neglecting the $\mathcal{O}(\varepsilon)$ terms of the previous equation, yields

$$
\begin{aligned}
df(\hat{X}_t^\varepsilon)&=\left[f'(\hat{X}_t^\varepsilon)\hat{F}_t+\frac{1}{2}f''(\hat{X}_t^\varepsilon)\hat{\Sigma}_t^2\right]dt+f'(\hat{X}_t^\varepsilon)\Sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(x)}\\
&\quad+f'(\hat{X}_t^\varepsilon)\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(y)},
\end{aligned}
\tag{2.1.67}
$$

where

$$
\hat{F}_t:=F(X_t^\varepsilon,Y_t^\varepsilon)+\mathcal{L}_1\chi(X_t^\varepsilon,Y_t^\varepsilon), \tag{2.1.68}
$$

$$
\hat{\Sigma}_t:=\Sigma(X_t^\varepsilon,Y_t^\varepsilon)^2+[\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)][\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)]^\top. \tag{2.1.69}
$$

Here comes a neat trick. Since f is the scale function of X, we know that $\mathcal{L}f=0$ on $\mathbb{R}$, where $\mathcal{L}$ is the generator of X defined in (2.1.18), allowing us to write

$$
\begin{aligned}
df(\hat{X}_t^\varepsilon)&=\left[f'(\hat{X}_t^\varepsilon)(\hat{F}_t-\bar{F}(\hat{X}_t^\varepsilon))+\frac{1}{2}f''(\hat{X}_t^\varepsilon)(\hat{\Sigma}_t^2-\bar{\Sigma}(\hat{X}_t^\varepsilon)^2)\right]dt\\
&\quad+f'(\hat{X}_t^\varepsilon)\Sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(x)}+f'(\hat{X}_t^\varepsilon)\nabla_y\chi(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(y)}.
\end{aligned}
\tag{2.1.70}
$$

Let us assume that φ is nonnegative, bounded, globally Lipschitz with Lipschitz constant $L_\varphi > 0$, and has compact support. We can start just like in the proof of Theorem 2.1.10 by writing

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt - \int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right|^2 \\
&= \mathbb{E}\left[\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt\right|^2\right] - \left(\int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right)^2 \\
&\quad - 2\mathbb{E}\left(\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt - \int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right)\int_{\mathbb{R}}\varphi(y)\mu(y)\,dy.
\end{aligned}$$

For the second term, we have by (2.1.59) and (2.1.61)

$$\begin{aligned}
&\left|\mathbb{E}\,\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt - \int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right| \\
&\leq \left|\mathbb{E}\,\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt - \mathbb{E}\,\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(X_t^\varepsilon)\,dt\right| + \left|\mathbb{E}\,\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right| \\
&\leq \mathbb{E}\,\frac{L_\varphi\varepsilon}{T_\varepsilon}\int_0^{T_\varepsilon}|\chi(X_t^\varepsilon, Y_t^\varepsilon)|\,dt + \left|\mathbb{E}\,\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right| \\
&\leq \mathcal{O}(\varepsilon) + \mathcal{O}(T_\varepsilon\varepsilon) + \left|\mathbb{E}\,\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(X_t)\,dt - \int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right|,
\end{aligned}$$

so that the mean ergodic theorem for X, which is essentially just a special case of Theorem 2.1.10, shows that this second term converges to zero as $\varepsilon \to 0$ as long as $T_\varepsilon\varepsilon \to 0$. This last restriction comes from (2.1.61) and could be removed if one were to have uniform-in-time convergence rates, that is, bounds where the constant C_T in (2.1.61) does not depend on time. Estimates of this kind have been established in [Cri+26] for averaging systems, that is, when $H \equiv f_1 \equiv 0$ in (1.2.1a), and could potentially follow for homogenization systems in the future. Coming back to the above, we want to emphasize here that subtracting and adding quantities of the homogenized process does not work on the L^2-level above, because we only have weak convergence. The processes X^ε and X are in general not even defined on the same probability space. Having dealt with this term, we now see that it is enough to prove

$$\limsup_{\varepsilon\to 0}\mathbb{E}\left[\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt\right|^2\right] \leq \left(\int_{\mathbb{R}}\varphi(y)\mu(y)\,dy\right)^2. \tag{2.1.71}$$

Assuming this holds, then the MET (2.1.57) follows easily from Lipschitz-continuity of φ and (2.1.59). To prove (2.1.71), we proceed as in the previous section and write

$$\mathbb{E}\left[\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\,dt\right|^2\right] = \frac{2}{T_\varepsilon^2}\mathbb{E}\int_0^{T_\varepsilon}\varphi(\hat{X}_t^\varepsilon)\int_t^{T_\varepsilon}\varphi(\hat{X}_s^\varepsilon)\,ds\,dt.$$

The inner integral can be written as

$$\int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\,ds = \int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\left(1 - \frac{\hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds + \int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\hat{\Sigma}_s^2\left(\frac{1}{\bar{\Sigma}(X_s^\varepsilon)^2} - \frac{1}{\bar{\Sigma}(\hat{X}_s^\varepsilon)^2}\right) ds$$
$$+ \int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\frac{\hat{\Sigma}_s^2}{\bar{\Sigma}(\hat{X}_s^\varepsilon)^2}\,ds =: I_\varepsilon^1 + I_\varepsilon^2 + I_\varepsilon^3.$$

For the first integral I_ε^1, we have by Lipschitz-continuity, nondegeneracy and boundedness of the diffusion coefficients, and by (2.1.59)

$$\begin{aligned}
\mathbb{E}\, I_\varepsilon^1 = \mathbb{E}\int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\left(1 - \frac{\hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds &= \mathbb{E}\int_t^{T_\varepsilon} (\varphi(\hat{X}_s^\varepsilon) - \varphi(X_s^\varepsilon))\left(1 - \frac{\hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds \\
&\quad + \mathbb{E}\int_t^{T_\varepsilon} \varphi(X_s^\varepsilon)\left(\frac{\bar{\Sigma}(X_s^\varepsilon)^2 - \hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds \\
&\leq L_\varphi \varepsilon \mathbb{E}\int_0^{T_\varepsilon} |\chi(X_s^\varepsilon, Y_s^\varepsilon)|\left(1 + \|\hat{\Sigma}^2\|_\infty \|\bar{\Sigma}^{-2}\|_\infty\right) ds \\
&\quad + \mathbb{E}\int_t^{T_\varepsilon} \varphi(X_s^\varepsilon)\left(\frac{\bar{\Sigma}(X_s^\varepsilon)^2 - \hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds \\
&\leq \mathcal{O}\left(T_\varepsilon \varepsilon\right) + \mathbb{E}\int_t^{T_\varepsilon} \varphi(X_s^\varepsilon)\left(\frac{\bar{\Sigma}(X_s^\varepsilon)^2 - \hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds.
\end{aligned}$$

For the second term in that last line, we note that the integrand satisfies the centering condition (2.1.62) by definition of $\bar{\Sigma}$. We can see this from the following short calculation

$$\int_{\mathbb{R}^{d-1}} \varphi(x)\left(\frac{\bar{\Sigma}(x)^2 - \Sigma(x,y)^2 + [\nabla_y\chi(x,y)\sigma(x,y)][\nabla_y\chi(x,y)\sigma(x,y)]^\top}{\bar{\Sigma}(x)^2}\right) d\mu^x(y)$$
$$= \frac{\varphi(x)}{\bar{\Sigma}(x)^2}\left[\bar{\Sigma}(x)^2 - \int_{\mathbb{R}^{d-1}} \Sigma(x,y)^2 + [\nabla_y\chi(x,y)\sigma(x,y)][\nabla_y\chi(x,y)\sigma(x,y)]^\top d\mu^x(y)\right] = 0.$$

Hence, we can use (2.1.60) to obtain

$$\mathbb{E}\int_t^{T_\varepsilon} \varphi(X_s^\varepsilon)\left(\frac{\bar{\Sigma}(X_s^\varepsilon)^2 - \hat{\Sigma}_s^2}{\bar{\Sigma}(X_s^\varepsilon)^2}\right) ds = \mathcal{O}\left(T_\varepsilon \varepsilon\right).$$

For the second integral I_ε^2, we use Lipschitz-continuity and nondegeneracy of $\bar{\Sigma}$, boundedness of φ and the functions defining $\hat{\Sigma}$, and (2.1.59) to obtain

$$\mathbb{E}\, I_\varepsilon^2 = \mathbb{E}\int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\hat{\Sigma}_s^2\left(\frac{\bar{\Sigma}(\hat{X}_s^\varepsilon)^2 - \bar{\Sigma}(X_s^\varepsilon)^2}{\bar{\Sigma}(\hat{X}_s^\varepsilon)^2}\right) ds = \mathcal{O}\left(T_\varepsilon \varepsilon\right)$$

For the third integral I_ε^3, we finally use the occupation density formula for semimartingales (A.1.9). Since $\hat{\xi}^\varepsilon := f(\hat{X}^\varepsilon)$ is, by (2.1.67), a 1-dimensional continuous semimartingale, we can express that integral in terms of the local time of $\hat{\xi}^\varepsilon$ as follows

$$\int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\frac{\hat{\Sigma}_s^2}{\bar{\Sigma}(\hat{X}_s^\varepsilon)^2}\,ds = \int_t^{T_\varepsilon} \frac{\varphi(\hat{X}_s^\varepsilon)}{f'(\hat{X}_s^\varepsilon)\bar{\Sigma}(\hat{X}_s^\varepsilon)^2} f'(\hat{X}_s^\varepsilon)\hat{\Sigma}_s^2\,ds = \int_{\mathbb{R}} \frac{\psi(z)}{\tilde{\sigma}(z)^2}\left[L_{T_\varepsilon}^z(\hat{\xi}^\varepsilon) - L_t^z(\hat{\xi}^\varepsilon)\right]dz,$$

where

$$\psi(z) := \varphi \circ f^{-1}(z), \qquad \tilde{\sigma}(z) := (f'\bar{\Sigma}) \circ f^{-1}(z), \quad z \in \mathbb{R}.$$

Using the Meyer-Tanaka formula for semimartingales (A.1.10), it follows

$$L_{T_\varepsilon}^z(\hat{\xi}^\varepsilon) - L_t^z(\hat{\xi}^\varepsilon) = |\hat{\xi}_{T_\varepsilon}^\varepsilon - z| - |\hat{\xi}_t^\varepsilon - z| - \int_t^{T_\varepsilon} \operatorname{sgn}(\hat{\xi}_s^\varepsilon - z)\,d\hat{\xi}_s^\varepsilon,$$

so that after inserting the differential of $\hat{\xi}^\varepsilon$, see (2.1.70), and eliminating the martingale terms through expectation, we obtain

$$\begin{aligned}
\mathbb{E}\, L_{T_\varepsilon}^z(\hat{\xi}^\varepsilon) - L_t^z(\hat{\xi}^\varepsilon) = &\,\mathbb{E}\left(|\hat{\xi}_{T_\varepsilon}^\varepsilon - z| - |\hat{\xi}_t^\varepsilon - z|\right) \\
&- \mathbb{E}\int_t^{T_\varepsilon} \operatorname{sgn}(\hat{\xi}_s^\varepsilon - z) f'(\hat{X}_t^\varepsilon)(\hat{F}_t - \bar{F}(\hat{X}_t^\varepsilon))\,dt \\
&- \frac{1}{2}\mathbb{E}\int_t^{T_\varepsilon} \operatorname{sgn}(\hat{\xi}_s^\varepsilon - z) f''(\hat{X}_t^\varepsilon)(\hat{\Sigma}_t^2 - \bar{\Sigma}(\hat{X}_t^\varepsilon)^2)\,dt.
\end{aligned}$$

This is the point where we hypothesize that the last displayed expectations are of order $\mathcal{O}(T_\varepsilon\varepsilon)$, which might be provable through suitable centering as we have done before and an exit stopping time argument to handle the growth of f' and f'', but the sign function definitely poses an additional problem. In any case, if we were to have this result, then, in a similar spirit as Lemma 2.1.9, we would obtain the following estimate from all the preceding considerations

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \frac{2}{T_\varepsilon^2}\mathbb{E}\int_0^{T_\varepsilon} \varphi(\hat{X}_t^\varepsilon)\int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\,ds\,dt \\
&\quad\le \limsup_{\varepsilon\to 0} \mathbb{E}\int_{\mathbb{R}} \frac{\psi(z)}{\tilde{\sigma}(z)^2}\int_0^{T_\varepsilon} \psi(\hat{\xi}_t^\varepsilon)\left[|\hat{\xi}_{T_\varepsilon}^\varepsilon - z| - |\hat{\xi}_t^\varepsilon - z|\right]dt\,dz.
\end{aligned} \tag{2.1.72}$$

The rest is fairly straightforward. One multiplies both sides of the last inequality by $\int_{[-R,R]} \tilde{\sigma}(y)^{-2}\,dy$, $R > 0$, and uses the compact support of φ to eventually obtain

$$\begin{aligned}
&\int_{[-R,R]} \tilde{\sigma}(y)^{-2}\,dy\left(\limsup_{\varepsilon\to 0} \frac{2}{T_\varepsilon^2}\mathbb{E}\int_0^{T_\varepsilon} \varphi(\hat{X}_t^\varepsilon)\int_t^{T_\varepsilon} \varphi(\hat{X}_s^\varepsilon)\,ds\,dt\right) \\
&\qquad\le \int_{\mathbb{R}} \frac{\psi(z)}{\tilde{\sigma}(z)^2}\,dz\left(\limsup_{\varepsilon\to 0} \frac{2}{T_\varepsilon^2}\mathbb{E}\int_0^{T_\varepsilon} \psi(\hat{\xi}_t^\varepsilon)(T_\varepsilon - t)\,dt\right) \\
&\qquad= \left(\int_{\mathbb{R}} \frac{1}{\tilde{\sigma}(y)^2}\,dy\right)^{-1}\left(\int_{\mathbb{R}} \frac{\psi(z)}{\tilde{\sigma}(z)^2}\,dz\right)^2.
\end{aligned}$$

The exact details of the last argument can be found in the proof of Theorem 2.1.10 or, if one prefers, in [Sun89]. Letting $R \to \infty$ and performing the substitution $dz = f'(y)\,dy$, we finally arrive at (2.1.71). As previously mentioned, this yields the final MET (2.1.57).

As we can see, given the fact that we used several deep structural properties of the underlying equations, the preceding formal derivation of the MET seems quite plausible, but in the current state it is not rigorous enough to call it a proof. Additionally, many nontrivial, although technical, details persist, but they could be circumvented with the right theoretical tools, so that we leave this for potential future research.

2.1.2 A Quantitative ε-Dependent Mean Ergodic Theorem

In this subsection, we want to derive quantitative versions of the preceding results in terms of convergence rates. These quantitative estimates will become especially important for our nonparametric estimator in Chapter 4. For this subsection, we concretize our setting slightly and consider the 1-dimensional SDE (2.1.15) with drift $F_\varepsilon := F + H_\varepsilon$ and $\varepsilon \in (0, 1]$, where H_ε may be understood as a general additive perturbation. For $q \in \mathbb{N}_0$, $r > 0$, and $s \in \mathbb{R}$, we define the following function classes

$$\mathscr{P}(q) := \{f : \mathbb{R} \to \mathbb{R} \text{ continuous} \,|\, \exists C_{q,f} > 0 : |f(x)| \le C_{q,f}(1 + |x|^q),\ \forall x \in \mathbb{R}\}\,, \tag{2.1.73}$$

$$\mathscr{R}(r) := \left\{(f, g) \,|\, f, g : \mathbb{R} \to \mathbb{R} \text{ measurable},\ g \neq 0, \limsup_{|x|\to\infty} \frac{\operatorname{sgn}(x) f(x)}{g(x)} \le -r\right\}, \tag{2.1.74}$$

$$\mathscr{I}(s) := \left\{(f, g) \,|\, f, g : \mathbb{R} \to \mathbb{R} \text{ measurable},\ g \neq 0, \limsup_{|x|\to\infty} \sup_{\substack{|v|\ge 1,\\ \operatorname{sgn}(x)v>0}} \frac{1}{v}\int_{x\wedge x+v}^{x\vee x+v} \frac{f(u)}{g(u)}\,du \le s\right\}, \tag{2.1.75}$$

$$\mathscr{P} := \bigcup_{q\in\mathbb{N}_0} \mathscr{P}(q), \quad \mathscr{R} := \bigcup_{r>0} \mathscr{R}(r). \tag{2.1.76}$$

The first and fourth class represent classes of functions with polynomial majorants. The second and last class will be used to define recurrence conditions and the third one will be a convenient class to work with when dealing with ε-dependence in the drift coefficient and in the invariant density of X^ε.

We will work under the following assumptions.

Assumptions (QET).

i) The functions F, $\bar{F}$ and $\Sigma, \bar{\Sigma}$ are locally Lipschitz-continuous on $\mathbb{R}$, $\Sigma^2, \bar{\Sigma}^2 > 0$ on $\mathbb{R}$, $\Sigma^{-2}, \bar{\Sigma}^{-2} \in \mathscr{P}$, and $(F + H_\varepsilon, \Sigma^2), (\bar{F}, \bar{\Sigma}^2) \in \mathscr{R}$.

ii) For $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists a constant $L_N > 0$ such that for all $x, y \in [-N, N]$,

$$|H_\varepsilon(x) - H_\varepsilon(y)| \le \frac{L_N}{\varepsilon^2}|x - y|.$$

iii) There exist $r > 0$ and $s \in \mathbb{R}$ with $r > s$ such that $(F, \Sigma^2) \in \mathscr{R}(r)$ and $(H_\varepsilon, \Sigma^2) \in \mathscr{I}(s)$.

iv) For every $N \in \mathbb{N}$, there exists a bounded, Borel-measurable function $b_N\colon [-N,N] \to \mathbb{R}$ such that for all $\varepsilon > 0$,

$$\left| \int_0^x \frac{H_\varepsilon(y)}{\Sigma(y)^2}\, dy \right| \le b_N(x), \quad x \in [-N,N].$$

Under Assumptions (QET) i) and ii), there exists a unique strong solution to the SDEs (2.1.15) and (2.1.16) by Lemma 2.1.3 i). Moreover, by $(\bar{F}, \bar{\Sigma}^2) \in \mathscr{R}$, there exist $r, A > 0$ such that for all $|x| > A$,

$$\frac{\operatorname{sgn}(x)\bar{F}(x)}{\bar{\Sigma}(x)^2} \le -\frac{r}{2}. \tag{2.1.77}$$

Hence, we obtain for, say $y > A$, the following estimate

$$\int_A^y \frac{\bar{F}(z)}{\bar{\Sigma}(z)^2}\, dz \le -r(y-A), \tag{2.1.78}$$

so that the scale function of X can be bounded, for $x > A$, by

$$\begin{aligned} f(x) = \int_0^x \exp\left(-\int_0^y \frac{\bar{F}(z)}{\bar{\Sigma}(z)^2}\, dz \right) dy &= f(A) + f'(A) \int_A^x \exp\left(-\int_A^y \frac{\bar{F}(z)}{\bar{\Sigma}(z)^2}\, dz \right) dy \\ &\ge f(A) + \frac{f'(A)}{r} \left(\exp(rx) - \exp(rA) \right), \end{aligned}$$

yielding $\lim_{x\to\infty} f(x) \to \infty$. Similar conclusions hold for $x < A$ due to the symmetry of the inequality (2.1.77). Thus, we just proved condition (MET) i). For condition (MET) ii), we use $\bar{\Sigma}^{-2} \in \mathscr{P}$ and (2.1.78) again to obtain

$$\begin{aligned} C_\rho &= \int_{\mathbb{R}} \rho(x)^2\, dx = \int_{\mathbb{R}} \frac{dx}{\bar{\Sigma}(x)^2 f'(x)} \\ &\le C_{q,\bar{\Sigma}^{-2}} \left[\int_{|x|\le A} \frac{(1+|x|^q)}{f'(x)}\, dx + \frac{\exp(rA)}{f'(A)} \int_{|x|>A} (1+|x|^q) \exp(-r|x|)\, dx \right] < \infty. \end{aligned}$$

Analog ideas and conclusions apply to the scale function f_ε of X^ε, too. Hence, if conditions (QET) i) and ii) hold, then the invariant densities for X^ε and X exist by Remark 2.1.2 and are given by (2.1.27) and (2.1.26), respectively. Due to condition (QET) iv), we also have the following uniform lower bound on the normalization constant of the invariant density μ_ε of X^ε

$$\inf_{\varepsilon>0} Z_\varepsilon \ge \exp\left(-2 \sup_{x\in[-N,N]} b_N(x) \right) \int_{-N}^N \Sigma(x)^{-2} \exp\left(2 \int_0^x \frac{F(y)}{\Sigma(y)^2}\, dy \right) dx > 0. \tag{2.1.79}$$

Our aim in this subsection is to prove, for any deterministic initial condition $x_0 \in \mathbb{R}$,

$$\mathbb{E} \left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\, dt - \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\, dy \right|^2 = \mathcal{O}\left(r(\varepsilon) \right), \quad \varepsilon \to 0, \tag{2.1.80}$$

where $\varphi \in \mathscr{P}$, $T_\varepsilon = \varepsilon^{-\alpha}$ with $\alpha > 0$, and $r(\varepsilon)$ is a convergence rate such that $r(\varepsilon) \to 0$ as $\varepsilon \to 0$. If we further assume the convergence of

$$\left| \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\, dy - \int_{\mathbb{R}} \varphi(y)\mu(y)\, dy \right|^2 = \mathcal{O}(s(\varepsilon)), \quad \varepsilon \to 0,$$

with a certain explicit convergence rate $s(\varepsilon) \to 0$ as $\varepsilon \to 0$, see Lemma 2.4.7 for an example, then we get the full mean ergodic theorem with explicit convergence rates and limit quantities coming from the SDE (2.1.16), i.e.,

$$\mathbb{E} \left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\, dt - \int_{\mathbb{R}} \varphi(y)\mu(y)\, dy \right|^p = \mathcal{O}(r(\varepsilon) \wedge s(\varepsilon)), \quad \varepsilon \to 0. \tag{2.1.81}$$

The proof idea for this result is not entirely new as it goes back to [Kut04, Section 1.2], which is why we only give a sketch of the main proof steps and more detailed explanations for the parts which actually require modifications. The main assumptions that come into play in our setting are (QET) iii) and iv). But before we present that proof, we would like to shed some light on the perturbation H_ε and on the possibly obscure assumption that $(H_\varepsilon, \Sigma^2) \in \mathscr{I}(s)$ for some $s \in \mathbb{R}$. We will provide some examples with additional assumptions on the coefficients yielding the membership to that function class. Bear in mind that the presentation of these examples is nowhere near comprehensive, so that other assumptions may yield similar conclusions.

Example 2.1.12. If we consider a somewhat natural perturbation of the form $H_\varepsilon(x) = \varepsilon H(x)$ with a locally Lipschitz-continuous function $H\colon \mathbb{R} \to \mathbb{R}$, then the assumption $(H_\varepsilon, \Sigma^2) \in \mathscr{I}(s)$ for some $s \in \mathbb{R}$ becomes, for $x > 0$ and $v \geq 1$,

$$\int_x^{x+v} \frac{H_\varepsilon(y)}{\Sigma(y)^2}\, dy \leq \left| \int_x^{x+v} \frac{H(y)}{\Sigma(y)^2}\, dy \right| = |I(x+v) - I(x)|, \quad I(x) := \int_0^x \frac{H(y)}{\Sigma(y)^2}\, dy,$$

where we used $0 < \varepsilon \leq 1$. One viable assumption to have now is "asymptotic" global Lipschitz-continuity of I with asymptotic Lipschitz constant $L_I > 0$, that is, there exists $L_I > 0$ such that

$$\limsup_{x \to \infty} \sup_{v \geq 1} \frac{|I(x+v) - F(x)|}{v} \leq L_I,$$

so that $(H_\varepsilon, \Sigma^2) \in \mathscr{I}(L_I)$. Otherwise, one could also assume boundedness of $H\Sigma^{-2}$ and obtain $(H_\varepsilon, \Sigma^2) \in \mathscr{I}(\|H\Sigma^{-2}\|_\infty)$. Similar ideas apply to the case $x < 0$.

Example 2.1.13. A slightly more delicate example is given by $H_\varepsilon(x) = H(x/\varepsilon)/\varepsilon$ with a locally Lipschitz-continuous function $H\colon \mathbb{R} \to \mathbb{R}$ having an antiderivative I such that I/Σ^2 is bounded on $\mathbb{R}$. Further assuming that the diffusion coefficient Σ is continuously differentiable, integration by parts gives, for $x > 0$ and $v \geq 1$,

$$\begin{aligned} \int_x^{x+v} \frac{H_\varepsilon(y)}{\Sigma(y)^2}\, dy &\leq \left| \left[\frac{I(y/\varepsilon)}{\Sigma(y)^2} \right]_x^{x+v} + 2 \int_x^{x+v} \Sigma'(y) \frac{I(y/\varepsilon)}{\Sigma(y)^3}\, dy \right| \\ &\leq 2\|I/\Sigma^2\|_\infty \left(v + \log \frac{\Sigma(x+v)}{\Sigma(x)} \right). \end{aligned}$$

Assuming global Lipschitz-continuity of Σ with Lipschitz constant $L_\Sigma > 0$ eventually gives $(H_\varepsilon, \Sigma^2) \in \mathscr{I}(2(1+L_\Sigma)\|I\Sigma^{-2}\|_\infty)$. Similar ideas apply to the case $x < 0$. Note that the multiscale overdamped Langevin equation for $d = 1$, given by (1.2.10), falls into the setting of this example. As a matter of fact, in Section 2.4, we will see that this Langevin equation satisfies all conditions of Assumptions (QET).

Theorem 2.1.14. *Let Assumptions (QET) hold. If $T_\varepsilon \to \infty$ as $\varepsilon \to 0$ in an arbitrary way, then, for any $\varphi \in \mathscr{P}$ and $p \in [1,\infty)$, we have*

$$\int_{\mathbb{R}} \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^{\varepsilon,y})\,dt - \int_{\mathbb{R}} \varphi(z)\mu_\varepsilon(z)\,dz\right|^p \mu_\varepsilon(y)\,dy = \mathcal{O}\left(C_{q_1,\varphi}^p T_\varepsilon^{-p/2}\right), \quad \varepsilon \to 0, \tag{2.1.82}$$

where $C_{q_1,\varphi} > 0$ is the constant from the polynomial growth bound of φ and $X^{\varepsilon,y}$ denotes the solution to (2.1.15) *with initial condition $X_0^{\varepsilon,y} = y$.*

Proof. As already mentioned before, we will only give a proof sketch. The main proof ideas are contained in [Kut04, Lemma 1.17]. We first define the function

$$\Phi_\varepsilon(x) = -\int_0^x \frac{2}{\Sigma(y)^2\mu_\varepsilon(y)}\int_{-\infty}^y \overline{\varphi}_\varepsilon(z)\mu_\varepsilon(z)\,dz\,dy, \quad x \in \mathbb{R}, \tag{2.1.83}$$

and note that it solves the Poisson equation

$$-\mathcal{L}_\varepsilon\Phi_\varepsilon = \overline{\varphi}_\varepsilon := \varphi - \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy, \qquad \mathcal{L}_\varepsilon := [F + H_\varepsilon]\,\partial_x + \frac{\Sigma^2}{2}\partial_{xx}. \tag{2.1.84}$$

Thus, Itô's formula gives

$$\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt = \frac{\Phi_\varepsilon(y) - \Phi_\varepsilon(X_{T_\varepsilon}^\varepsilon)}{T_\varepsilon} + \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \Sigma(X_t^\varepsilon)\Phi_\varepsilon'(X_t^\varepsilon)\,dW_t. \tag{2.1.85}$$

For $x \in \mathbb{R}$, the RHS of the Poisson equation satisfies

$$|\overline{\varphi}_\varepsilon(x)| \le C_{q_1,\varphi}(1 + |x|^{q_1}) + \left|\int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy\right|.$$

With (QET) iv) and the uniform lower bound (2.1.79), we obtain

$$\begin{aligned}
&\left|\int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy\right| \\
&\le \frac{C_{q_1,\varphi}C_{q_2,\Sigma^{-2}}}{\inf_{\varepsilon>0} Z_\varepsilon}\left[\int_{-N}^N (1+|y|^{q_1})(1+|y|^{q_2})\exp\left(2\int_0^y \frac{F(z)}{\Sigma(z)^2}\,dz\right)\exp(2b_N(y))\,dy\right. \\
&\qquad \left. + \int_{|y|>N} (1+|y|^{q_1})(1+|y|^{q_2})\exp\left(2\int_0^y \frac{F(z)}{\Sigma(z)^2}\,dz\right)\exp\left(2\int_0^y \frac{H_\varepsilon(z)}{\Sigma(z)^2}\,dz\right)\,dy\right]
\end{aligned}$$

for any $N \in \mathbb{N}$. The first term is obviously bounded independently of ε and the second term can be dealt with as follows. By Assumption (QET) iii), there exist $r > 0$ and $s \in \mathbb{R}$ with $r > s$ such that for $\delta_{r,s} := (r-s)/4 > 0$, there exists an $A > 0$ with

$$\frac{\operatorname{sgn}(x)F(x)}{\Sigma(x)^2} \le -r + \delta_{r,s}, \qquad \sup_{\substack{|v|\ge 1,\\ \operatorname{sgn}(x)v>0}} \frac{1}{v}\int_{x\wedge x+v}^{x\vee x+v} \frac{H_\varepsilon(u)}{\Sigma(u)^2}\,du \le s + \delta_{r,s}, \qquad |x| > A.$$

The second inequality can be written here, depending on the sign of x, as

$$\sup_{v\geq 1}\frac{1}{v}\int_x^{x+v}\frac{H_\varepsilon(u)}{\Sigma(u)^2}\,du\leq s+\delta_{r,s},\quad x>A,\qquad \sup_{v\leq -1}\frac{1}{v}\int_{x+v}^{x}\frac{H_\varepsilon(u)}{\Sigma(u)^2}\,du\leq s+\delta_{r,s},\quad x<-A.$$

For $N-1\geq A$, it then holds

$$\begin{aligned}
&\int_{y>N}(1+|y|^{q_1})(1+|y|^{q_2})\exp\left(2\int_0^y\frac{F(z)+H_\varepsilon(z)}{\Sigma(z)^2}\,dz\right)dy\\
\leq&4\exp\left(2\int_0^{N-1}\frac{|F(z)|}{\Sigma(z)^2}\,dz+2b_{N-1}(N-1)\right)\int_N^\infty|y|^{q_1+q_2}\exp\left(2\int_{N-1}^y\frac{F(z)+H_\varepsilon(z)}{\Sigma(z)^2}\,dz\right)dy\\
\leq&C_N\int_0^\infty|N+w|^{q_1+q_2}\exp\left(2\int_{N-1}^{N+w}\frac{F(z)}{\Sigma(z)^2}\,dz\right)\exp\left(2\int_{N-1}^{N+w}\frac{H_\varepsilon(z)}{\Sigma(z)^2}\,dz\right)dw\\
\leq&C_N\int_0^\infty|N+w|^{q_1+q_2}\exp\left(-2(r-s-2\delta_{r,s})(w+1)\right)\,dw\\
=&C_N\int_0^\infty|N+w|^{q_1+q_2}\exp\left(-(r-s)w\right)\,dw,
\end{aligned}$$

which is again a finite constant independent of ε. A similar estimate holds for the integration region $\{y\in\mathbb{R}\,|\,y<-N\}$ due to the symmetry of our assumptions. Thus, we have just proved that there exists a constant $D_1>0$, depending only on $C_{q_2,\Sigma^{-2}},q_1,q_2,r,s,A,N,b_N$, such that

$$|\overline{\varphi}_\varepsilon(x)|\leq C_{q_1,\varphi}D_1(1+|x|^{q_1}),\quad x\in\mathbb{R}. \tag{2.1.86}$$

Using this estimate, the preceding argument, and the fact that

$$\int_{-\infty}^y\overline{\varphi}_\varepsilon(z)\mu_\varepsilon(z)\,dz=\int_{\mathbb{R}}\overline{\varphi}_\varepsilon(z)\mu_\varepsilon(z)\,dz-\int_y^\infty\overline{\varphi}_\varepsilon(z)\mu_\varepsilon(z)\,dz=-\int_y^\infty\overline{\varphi}_\varepsilon(z)\mu_\varepsilon(z)\,dz,\quad y\in\mathbb{R},$$

it is easy to obtain the estimate

$$|\Phi_\varepsilon'(x)|\leq C_{q_1,\varphi}D_2(1+|x|^{q_1+q_2}),\quad x\in\mathbb{R}, \tag{2.1.87}$$

where $D_2>0$ is a constant depending only on $C_{q_2,\Sigma^{-2}},q_1,q_2,r,s,A,N,b_N$, cf. [Kut04, Lemma 1.17]. The preceding arguments and estimate (2.1.87) enable us to bound

$$\sup_{\varepsilon>0}\int_{\mathbb{R}}|\Phi_\varepsilon(x)|^p\mu_\varepsilon(x)\,dx+\int_{\mathbb{R}}|\Sigma(x)\Phi_\varepsilon'(x)|^p\mu_\varepsilon(x)\,dx\leq C_{q_1,\varphi}^pD_3, \tag{2.1.88}$$

where $D_3>0$ only depends on $C_{q_2,\Sigma^{-2}},q_1,q_2,p,r,s,A,N,b_N$. With all this preliminary work, we can now prove the theorem. Equation (2.1.85), stationarity with respect to the invariant measure $\mu_\varepsilon\,dx$, cf. Lemma 2.1.44, Burkholder-Davis-Gundy's inequalites (A.1.8) with universal constant $B_{p/2}>0$, Jensen's inequality, and estimate (2.1.88) yield

$$\begin{aligned}
\int_{\mathbb{R}}\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^p\mu_\varepsilon(y)\,dy\leq\frac{3^p}{T_\varepsilon^p}\Bigg[&\int_{\mathbb{R}}|\Phi_\varepsilon(y)|^p\,\mu_\varepsilon(y)\,dy+\int_{\mathbb{R}}\mathbb{E}\left|\Phi_\varepsilon(X_{T_\varepsilon}^\varepsilon)\right|^p\mu_\varepsilon(y)\,dy\\
&+\int_{\mathbb{R}}\mathbb{E}\left|\int_0^{T_\varepsilon}\Sigma(X_t^\varepsilon)\Phi_\varepsilon'(X_t^\varepsilon)\,dW_t\right|^p\mu_\varepsilon(y)\,dy\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{3^p}{T_\varepsilon^p}\left[B_{p/2}\int_{\mathbb{R}}\mathbb{E}\left|\int_0^{T_\varepsilon}\Sigma(X_t^\varepsilon)^2\Phi_\varepsilon'(X_t^\varepsilon)^2\,dt\right|^{p/2}\mu_\varepsilon(y)\,dy\right.\\
&\qquad\left.+\,2\int_{\mathbb{R}}|\Phi_\varepsilon(y)|^p\,\mu_\varepsilon(y)\,dy\right]\\
&\leq \frac{3^p}{T_\varepsilon^p}\left[B_{p/2}T_\varepsilon^{p/2-1}\int_0^{T_\varepsilon}\int_{\mathbb{R}}\mathbb{E}\left|\Sigma(X_t^\varepsilon)\Phi_\varepsilon'(X_t^\varepsilon)\right|^p\mu_\varepsilon(y)\,dy\,dt\right.\\
&\qquad\left.+\,2\int_{\mathbb{R}}|\Phi_\varepsilon(y)|^p\,\mu_\varepsilon(y)\,dy\right]\\
&\leq \frac{3^pC_{q_1,\varphi}^pD_3}{T_\varepsilon^p}[B_{p/2}T_\varepsilon^{p/2}+2],
\end{aligned}
$$

which gives the claimed estimate (2.1.82). □

Remark 2.1.15. *Under Assumptions (QET) and $\varphi\in\mathscr{P}$, the above proof shows that the first derivative of the solution Φ_ε of the Poisson equation*

$$
-\mathcal{L}_\varepsilon\Phi_\varepsilon=\overline{\varphi}_\varepsilon=\varphi-\int_{\mathbb{R}}\varphi(y)\mu_\varepsilon(y)\,dy,\qquad \mathcal{L}_\varepsilon:=[F+H_\varepsilon]\,\partial_x+\frac{\Sigma^2}{2}\partial_{xx},
$$

satisfies the polynomial growth bound in (2.1.87)*, namely*

$$
|\Phi_\varepsilon'(x)|\leq C_{q_1,\varphi}D_2(1+|x|^{q_1+q_2}),\quad x\in\mathbb{R},
$$

where $D_2>0$ is a constant depending only on $C_{q_2,\Sigma^{-2}},q_1,q_2,r,s,A,N,b_N$. In particular, if $q_1=q_2=0$, then Φ_ε' is uniformly bounded in ε on $\mathbb{R}$ and Φ_ε is globally Lipschitz-continuous with a Lipschitz constant independent of ε.

Remark 2.1.16. *The same proof arguments work even if the test function explicitly depends on $\varepsilon>0$, but has a polynomial growth bound uniform in ε, i.e., $\varphi=\varphi_\varepsilon$ and*

$$
|\varphi_\varepsilon(x)|\leq C_{q,\varphi_\varepsilon}(1+|x|^q),\quad x\in\mathbb{R}, \tag{2.1.89}
$$

where $C_{q,\varphi_\varepsilon}>0$ does not depend on ε. For example, Theorem 2.1.14 applies to Φ_ε' as in Remark 2.1.15.

Evidently, the last theorem gives us $L^p(\Omega)$-estimates only in stationarity, that is, as if we initialized the process X^ε with the invariant distribution $\mu_\varepsilon\,dy$. We would like to extend this result to the non-stationary case where we start the process in an arbitrary deterministic state $x_0\in\mathbb{R}$. Unfortunately, we need more assumptions for that, and we also need to take a step back in order to prove estimates for the transition probability density function p_ε of the process X^ε. To this end, we will assume that the diffusion coefficient is constant with $\Sigma\equiv 1$. It is not necessary to assume a unit diffusion for the following results to hold, but it improves readability. However, it is thus far not clear how to obtain the results for non-constant diffusion coefficients. That being said, suitable estimates for transition probability density functions in the non-constant case might be found in [SV79; Str08]. We will present a more elementary approach here. In this particular 1-dimensional setting, p_ε is given by the formula

$$p_\varepsilon(h,x,y) = \frac{1}{\sqrt{2\pi h}} \left(\frac{\mu_\varepsilon(y)}{\mu_\varepsilon(x)}\right)^{1/2} \exp\left(-\frac{(y-x)^2}{2h}\right) \mathbb{E}\exp\left(h\int_0^1 B_\varepsilon(N(h,x,y,u))\,du\right), \tag{2.1.90}$$

with

$$B_\varepsilon(x) = -\frac{1}{2}F_\varepsilon(x)^2 - \frac{1}{2}F_\varepsilon'(x), \quad F_\varepsilon(x) = F(x) + H_\varepsilon(x).$$

Here, $N(h,x,y,u)$ is a shorthand notation for

$$N(h,x,y,u) \sim \mathcal{N}_1(x+u(y-x), hu(1-u)), \quad u \in (0,1). \tag{2.1.91}$$

The formula (2.1.90) can be obtained from the results in [GS68, §13], in particular, from equation (9) of that paragraph. The proof of the following lemma is interesting in the sense that we encounter a reoccurring phenomenon when dealing with multiscale SDE problems. More precisely, in order to leverage against the $1/\varepsilon^2$ factor in the exponential estimate appearing in the proof, one chooses a time step of $h \in (0,\varepsilon^2]$, i.e., at most at the fastest scale of the multiscale SDE. In other words, to get good estimates on the transition probability density function, the time step h between the two transitional states x and y should be of order $\mathcal{O}(\varepsilon^{-2})$.

Lemma 2.1.17. *Assume (QET), $\Sigma \equiv 1 > 0$, and that there exist constants $K_1, K_2 > 0$ such that for all $x \in \mathbb{R}$ and $\varepsilon > 0$,*

$$-F_\varepsilon'(x) = -(F+H_\varepsilon)'(x) \le K_1 x^2 + \frac{K_2}{\varepsilon^2}. \tag{2.1.92}$$

Then there exists a constant $C_p > 0$, independent of ε and $h > 0$, such that for all $x, y \in \mathbb{R}$, sufficiently small ε, and $h \in (0,\varepsilon^2]$, the following bound holds

$$p_\varepsilon(h,x,y) \le C_p \mu_\varepsilon(y)^{1/2} \phi_h(y-x), \tag{2.1.93}$$

where ϕ_h is the probability density function of the centered normal distribution $\mathcal{N}_1(0, h/(1-4K_1h^2))$.

Proof. Let $x, y \in \mathbb{R}$ be arbitrary and $h \in (0,\varepsilon^2]$. Using Jensen's inequality and condition (2.1.92), we can estimate the expectation figuring in equation (2.1.90) as follows

$$\begin{aligned}
\mathbb{E}\exp\left(h\int_0^1 B_\varepsilon(N(h,x,y,u))\,du\right) &\le \int_0^1 \mathbb{E}\exp\left(hB_\varepsilon(N(h,x,y,u))\right)du \\
&\le \int_0^1 \mathbb{E}\exp\left(-\frac{h}{2}F_\varepsilon'(N(h,x,y,u))\right)du \\
&\le \exp\left(\frac{K_2 h}{2\varepsilon^2}\right)\int_0^1 \mathbb{E}\exp\left(\frac{K_1 h}{2}N(h,x,y,u)^2\right)du.
\end{aligned}$$

Hence, because $h \in (0, \varepsilon^2]$, we obtain

$$\mathbb{E} \exp\left(h \int_0^1 B_\varepsilon(N(h,x,y,u))\, du\right) \le \exp\left(\frac{K_2}{2}\right) \int_0^1 \mathbb{E} \exp\left(\frac{K_1 h}{2} N(h,x,y,u)^2\right) du.$$

We now use

$$N(h,x,y,u) \sim x + u(y-x) + \sqrt{hu(1-u)}Z, \quad Z \sim \mathcal{N}_1(0,1),$$

in order to bound

$$\begin{aligned}
\mathbb{E} \exp\left(\frac{K_1 h}{2} N(h,x,y,u)^2\right) &\le \mathbb{E} \exp\left(K_1 h \left([x+u(y-x)]^2 + hu(1-u)Z^2\right)\right) \\
&\le \exp\left(K_1 h\, [x+u(y-x)]^2\right) \mathbb{E} \exp\left(\frac{K_1 h^2}{4} Z^2\right) \\
&\le \exp\left(2K_1 h \left[x^2 + (y-x)^2\right]\right) \mathbb{E} \exp\left(\frac{K_1 h^2}{4} Z^2\right) \\
&= \exp\left(2K_1 h \left[x^2 + (y-x)^2\right]\right) \sqrt{\frac{2}{2 - K_1 h^2}},
\end{aligned}$$

where $u \in (0,1)$ and $\varepsilon > 0$ is sufficiently small so that $K_1 h^2 < 2$. Hence, for sufficiently small $\varepsilon > 0$, we have the following estimate

$$\mathbb{E} \exp\left(h \int_0^1 B_\varepsilon(N(h,x,y,u))\, du\right) \le \exp\left(\frac{K_2}{2}\right) \exp\left(2K_1 h \left[x^2 + (y-x)^2\right]\right) \sqrt{\frac{2}{2 - K_1 h^2}}. \tag{2.1.94}$$

Moreover, a short calculation shows that

$$\frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{1}{2h}(y-x)^2\right) \exp\left(2K_1 h (y-x)^2\right) = \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{1 - 4K_1 h^2}{2h}(y-x)^2\right),$$

where $\varepsilon > 0$ is sufficiently small so that $4K_1 h^2 < 1$. In conclusion, we have proved the following bound, whenever ε is sufficiently small such that for $h \in (0, \varepsilon^2]$ and $K_1 h^2/2 \vee 4K_1 h^2 < 1/2$,

$$q_\varepsilon(h,x,y) \le \exp\left(\frac{K_2}{2}\right) \exp\left(2K_1 h x^2\right) \left(\frac{\mu_\varepsilon(y)}{\mu_\varepsilon(x)}\right)^{1/2} \frac{1}{\sqrt{\pi h}} \exp\left(-\frac{1 - 4K_1 h^2}{2h}(y-x)^2\right). \tag{2.1.95}$$

Using (QET) iv), it is easy to see that $\mu_\varepsilon(x)^{-1}$ is uniformly bounded in ε, so that a simple calculation eventually yields

$$q_\varepsilon(h,x,y) \le C_p \mu_\varepsilon(y)^{1/2}\, \phi_h(y-x), \tag{2.1.96}$$

with a constant $C_p > 0$ that is independent of ε and h, and a centered Gaussian density ϕ_h with variance $h/(1 - 4K_1 h^2)$, that is,

$$\phi_h(y) = \sqrt{\frac{1 - 4K_1 h^2}{2\pi h}} \exp\left(-\frac{1 - 4K_1 h^2}{2h} y^2\right), \quad y \in \mathbb{R}. \qquad \square$$

With these bounds, we can now handle the non-stationary case.

Theorem 2.1.18. *Assume (QET), $\Sigma \equiv 1 > 0$, and that there exist constants $K_1, K_2 > 0$ such that for all $x \in \mathbb{R}$ and $\varepsilon > 0$,*

$$-(F + H_\varepsilon)'(x) \leq K_1 x^2 + \frac{K_2}{\varepsilon^2}. \tag{2.1.97}$$

Furthermore, for all $x \in \mathbb{R}$ and $\varepsilon > 0$, assume the Lyapunov condition

$$\mathcal{L}_\varepsilon v(x) \leq -\lambda v(x) + \frac{c}{\varepsilon^{2m}}, \quad \mathcal{L}_\varepsilon = [F + H_\varepsilon]\,\partial_x + \frac{\Sigma^2}{2}\partial_{xx}, \tag{2.1.98}$$

with a nonnegative function $v \in C^2(\mathbb{R}) \cap \mathscr{P}(2m)$ for some $m \in \mathbb{N}$ and constants $\lambda, c > 0$. Then, for any continuous function φ such that $|\varphi|^2 \leq v$ and $T_\varepsilon = \varepsilon^{-\alpha}$ with $\alpha > m/3 \vee 1/2$, we have, as $\varepsilon \to 0$,

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy\right|^2 = \begin{cases} \mathcal{O}\left((1 \vee C_{m,\varphi}^2)\varepsilon^{(6\alpha-2m)/5}\right), & \text{if } \alpha \leq 2m - \frac{5}{2}, \\ \mathcal{O}\left((1 \vee C_{m,\varphi}^2)\varepsilon^{\alpha-1/2}\right), & \text{if } \alpha > 2m - \frac{5}{2}, \end{cases} \tag{2.1.99}$$

where $C_{m,\varphi} > 0$ is the constant from the polynomial growth bound of φ. If, instead of the Lyapunov condition, we simply assume that $\varphi \in \mathscr{P}(0)$, then, for $T_\varepsilon = \varepsilon^{-\alpha}$ with $\alpha > 1/2$, we have, as $\varepsilon \to 0$,

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy\right|^2 = \mathcal{O}\left(\frac{C_{0,\varphi}^2}{T_\varepsilon\sqrt{\varepsilon}}\right) = \mathcal{O}\left(C_{0,\varphi}^2 \varepsilon^{\alpha-1/2}\right). \tag{2.1.100}$$

Proof. Let $\overline{\varphi}_\varepsilon$ be defined as in (2.1.84). We first note that the Lyapunov condition implies the following standard estimate

$$\mathbb{E}\, v(X_t^\varepsilon) \leq v(x_0)\exp(-\lambda t) + \frac{c}{\varepsilon^{2m}\lambda}\left(1 - \exp(-\lambda t)\right), \quad t \in [0, \infty),$$

whose argument can be found in [KS91, Exercise 4.5.35] for example. Therefore, we also have

$$\int_0^t \mathbb{E}\, v(X_s^\varepsilon)\,ds \leq \frac{v(x_0)}{\lambda} + \frac{c}{\varepsilon^{2m}\lambda}\left(t + \frac{1}{\lambda}\right), \quad t \in [0, \infty). \tag{2.1.101}$$

We can estimate the relevant quantity for any $h \in (0, \varepsilon^2]$ as follows

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^2 &\leq \frac{2}{T_\varepsilon^2}\mathbb{E}\left|\int_0^h \overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^2 + \frac{2}{T_\varepsilon^2}\mathbb{E}\left|\int_h^{T_\varepsilon} \overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^2 \\
&\leq \frac{2h}{T_\varepsilon^2}\int_0^h \mathbb{E}\,|\overline{\varphi}_\varepsilon(X_t^\varepsilon)|^2\,dt + \frac{2}{T_\varepsilon^2}\mathbb{E}\left[\mathbb{E}\left[\left|\int_0^{T_\varepsilon-h} \overline{\varphi}_\varepsilon(X_{t+h}^{\varepsilon,x_0})\,dt\right|^2 \middle| X_h^{\varepsilon,x_0}\right]\right] \\
&= \frac{2h}{T_\varepsilon^2}\int_0^h \mathbb{E}\,|\overline{\varphi}_\varepsilon(X_t^\varepsilon)|^2\,dt + \frac{2}{T_\varepsilon^2}\mathbb{E}\left|\int_0^{T_\varepsilon-h} \overline{\varphi}_\varepsilon(X_t^{\varepsilon,X_h^{\varepsilon,x_0}})\,dt\right|^2 \\
&= \frac{2h}{T_\varepsilon^2}\int_0^h \mathbb{E}\,|\overline{\varphi}_\varepsilon(X_t^\varepsilon)|^2\,dt
\end{aligned}$$

$$+\frac{2}{T_\varepsilon^2}\int_{\mathbb{R}}\mathbb{E}\left|\int_0^{T_\varepsilon-h}\overline{\varphi}_\varepsilon(X_t^{\varepsilon,y})\,dt\right|^2 p_\varepsilon(h,x_0,y)\,dy,$$

where we used the Cauchy–Schwartz inequality once and the Markov property in the last line. Here, $X^{\varepsilon,y}$ denotes the solution to (2.1.15) with initial condition $X_0^{\varepsilon,y}=y$. The number h will soon be chosen to depend on ε such that $h\to 0$ in a suitable way as $\varepsilon\to 0$. Just as estimate (2.1.86), we can show that there exist constants $D_1,D_2>0$ such that

$$|\overline{\varphi}_\varepsilon(x)|^2\le C_{2m,v}D_1+D_2v(x),\quad x\in\mathbb{R},$$

Thus, we can apply (2.1.101) to obtain for the term above

$$\begin{aligned}\frac{h}{T_\varepsilon^2}\int_0^h\mathbb{E}\,|\overline{\varphi}_\varepsilon(X_t^\varepsilon)|^2\,dt&\le\frac{h}{T_\varepsilon^2}\int_0^h C_{2m,v}D_1+D_2\mathbb{E}\,v(X_t^\varepsilon))\,dt\\&\le\frac{h}{T_\varepsilon^2}\left[hC_{2m,v}D_1+\frac{D_2v(x_0)}{\lambda}+\frac{D_2c}{\varepsilon^{2m}\lambda}\left(h+\frac{1}{\lambda}\right)\right]=\mathcal{O}\left(\frac{h}{\varepsilon^{2m}T_\varepsilon^2}\right).\end{aligned}\tag{2.1.102}$$

For the other term, we use the estimate in Lemma 2.1.17, the Cauchy–Schwartz inequality, and Theorem 2.1.14 to obtain

$$\begin{aligned}&\frac{1}{T_\varepsilon^2}\int_{\mathbb{R}}\mathbb{E}\left|\int_0^{T_\varepsilon-h}\overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^2 p_\varepsilon(h,x_0,y)\,dy\\&\quad\le\int_{\mathbb{R}}\mathbb{E}\left[\left|\frac{1}{T_\varepsilon-h}\int_0^{T_\varepsilon-h}\overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^2\right]p_\varepsilon(h,x_0,y)\,dy\\&\quad\le C_p\int_{\mathbb{R}}\mathbb{E}\left[\left|\frac{1}{T_\varepsilon-h}\int_0^{T_\varepsilon-h}\overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^2\right]\mu_\varepsilon(y)^{1/2}\phi_h(y-x_0)\,dy\\&\quad\le C_p\left(\int_{\mathbb{R}}\mathbb{E}\left[\left|\frac{1}{T_\varepsilon-h}\int_0^{T_\varepsilon-h}\overline{\varphi}_\varepsilon(X_t^\varepsilon)\,dt\right|^4\right]\mu_\varepsilon(y)\,dy\right)^{1/2}\left(\int_{\mathbb{R}}\phi_h(y-x_0)^2\,dy\right)^{1/2}\\&\quad=\mathcal{O}(C_{m,\varphi}^2T_\varepsilon^{-1}h^{-1/4})\end{aligned}\tag{2.1.103}$$

for $h\in(0,\varepsilon^2]$ and sufficiently small $\varepsilon>0$. We now choose $h=\varepsilon^{4(2m-\alpha)/5}\wedge\varepsilon^2$ and combine the previous estimates to obtain the claim. For the other case where φ is bounded, we can use this boundedness in (2.1.102) to obtain the estimate

$$\frac{h}{T_\varepsilon^2}\int_0^h\mathbb{E}\,|\overline{\varphi}_\varepsilon(X_t^\varepsilon)|^2\,dt=\mathcal{O}\left(C_{0,\varphi}^2\frac{h^2}{T_\varepsilon^2}\right),\tag{2.1.104}$$

which, in turn, enables us to choose $h=\varepsilon^2$ in (2.1.103). This proves the final claim. □

Remark 2.1.19. *Using similar ideas, one can obtain bounds in $L^p(\Omega)$-spaces with $p\in[1,\infty)$, but since we only need the MET, we will not delve into these.*

Remark 2.1.20. *As we already noted in Remark 2.1.16, the same proof arguments work even if the test function explicitly depends on $\varepsilon > 0$ and has an upper bound uniform in ε, i.e., $\varphi = \varphi_\varepsilon$ and*

$$|\varphi_\varepsilon(x)|^2 \le v(x), \quad x \in \mathbb{R}, \tag{2.1.105}$$

in the unbounded case with the Lyapunov condition, or $\sup_{\varepsilon>0} \|\varphi_\varepsilon\|_\infty < \infty$ *in the bounded case without the Lyapunov condition.*

Remarks on multidimensional extensions II.

The proofs in this section rely heavily on the Markov property of X^ε and stationarity with respect to the invariant density μ_ε as one can see, for example, in the proof of Theorem 2.1.14 or Theorem 2.1.18. In addition to that, we utilize estimates for the transition probability density function of X^ε and for the solution to a Poisson equation based on the generator $\mathcal{L}_\varepsilon$ of X^ε. On the one hand, all of this essentially prohibits an extension of the given proofs to multidimensional multiscale SDE systems (1.2.1) where X^ε and Y^ε are coupled, because we lose the Markovian structure for X^ε alone. On the other hand, the use of Markovian structure might suggest that analyzing X^ε alone is not productive for ergodic theorems. The full multiscale system with generator $\mathcal{L}_\varepsilon$ is in any case Markovian, so that one could advocate for proving the MET with respect to $(X^\varepsilon, Y^\varepsilon)$, i.e., proving something like

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^\varepsilon, Y_t^\varepsilon)\,dt - \int_{\mathbb{R}^{d_x}\times\mathbb{R}^{d_y}} \varphi(x,y)\,d\nu(x,y)\right|^2 \to 0, \quad \varepsilon \to 0, \tag{2.1.106}$$

where, most likely, $d\nu(x,y) = d\mu(x)\,d\mu^x(y)$ with the invariant measure μ of (1.2.2) and the invariant measure μ^x of the frozen SDE (1.2.4), cf. [PPS09, Theorem 3.8]. This would obviously imply an MET for X^ε if φ only depends on the first argument, but proving (2.1.106) brings a whole set of different problems with it. Just to give an example, the generator of $(X^\varepsilon, Y^\varepsilon)$ is often expressed in the form

$$\mathcal{L}_\varepsilon = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2, \tag{2.1.107}$$

with

$$\mathcal{L}_0 = \frac{1}{2}\sum_{i,j=1}^{d_y} a_{ij}\frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{d_y} f_0^{(i)}\frac{\partial}{\partial y_i}, \quad a = \sigma\sigma^\top, \tag{2.1.108}$$

$$\mathcal{L}_1 = \sum_{i=1}^{d_x} H^{(i)}\frac{\partial}{\partial x_i} + \sum_{i=1}^{d_y} f_1^{(i)}\frac{\partial}{\partial y_i}, \tag{2.1.109}$$

$$\mathcal{L}_2 = \frac{1}{2}\sum_{i,j=1}^{d_x} A_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d_x} F^{(i)}\frac{\partial}{\partial x_i}, \quad A := \Sigma\Sigma^\top, \tag{2.1.110}$$

so that analyzing the Poisson equation with respect to the full generator, i.e.,

$$-\mathcal{L}_\varepsilon\Phi_\varepsilon = \varphi - \int_{\mathbb{R}^{d_x}\times\mathbb{R}^{d_y}} \varphi(x,y)\,d\nu_\varepsilon(x,y), \tag{2.1.111}$$

with the invariant measure ν_ε of $(X^\varepsilon, Y^\varepsilon)$ if it exists, usually goes against the grain in homogenization theory, where Poisson equations are commonly studied for generators without explicit ε-dependence. Although we essentially analyzed exactly such Poisson equations for a multiscale system with decoupled slow and fast variable, our proofs could be adapted rather easily due to the explicit formula for the solution to said Poisson equation, see (2.1.83). This is something that is most certainly not true for the multi-dimensional setting. However, initial reference points for Poisson equations in the whole Euclidean space could be the well-known series of papers [PV01; PV05; PV03] and the more recent work [RX21].

2.2 Asymptotic Vanishing of Tail Probabilities

This section is concerned with the quantity $X^{\varepsilon,x}_{T_\varepsilon}/R_\varepsilon$ and its convergence in probability to zero as $\varepsilon \to 0$, where $R_\varepsilon \to \infty$ satisfies a sufficient asymptotic condition with respect to $\varepsilon \to 0$ and $X^{\varepsilon,x}$ is the solution to (2.1.15) with initial condition $X^{\varepsilon,x}_0 = x$. The final result is contained in Corollary 2.2.5 at the end of this section. Most of the crucial proof ideas of this section originate from [KMN12, Chapter 4], but, as in the preceding sections, we must practice caution due to the ε-dependence. Because the results of this section are primarily used to argue a specific point in the next section about CLTs, which itself is a qualitative convergence statement, we will mostly consider the setting and the assumptions of Subsection 2.1.1. That being said, with the necessary techniques that are outlined here and in the preceding section, it should be possible to obtain the exact same results under Assumptions (QET), as well.

We make the following adjustment to Assumptions (C) where the only difference between (C) and (C*) is the strict nondegeneracy of Σ^2 on $\mathbb{R}$ and the explicit local linear growth condition on F_ε.

Assumptions (C*).

i) The functions $\bar{F}$ and $\bar{\Sigma}$ are locally Lipschitz-continuous on $\mathbb{R}$, $\bar{\Sigma}^2$ is strictly positive, and $\bar{F}/\bar{\Sigma}^2$ is locally integrable, that is, for all $x \in \mathbb{R}$, there exists a $\delta > 0$ such that

$$\int_{x-\delta}^{x+\delta} \frac{|\bar{F}(y)|}{\bar{\Sigma}(y)^2}\, dy < \infty. \tag{2.2.1}$$

ii) The function Σ is locally Lipschitz-continuous on $\mathbb{R}$ and satisfies $\Sigma^2 \geq m_0$ on $\mathbb{R}$ for some $m_0 > 0$. Moreover, for any $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists a constant $L_N > 0$ such that for all $x, y \in [-N, N]$,

$$|F_\varepsilon(x) - F_\varepsilon(y)| \leq \frac{L_N}{\varepsilon^2}|x - y|, \qquad |F_\varepsilon(x)| \leq \frac{L_N}{\varepsilon^2}(1 + |x|).$$

The first major step towards the main result of this section is establishing a priori interior estimates of the Schauder type for the parabolic partial differential equation (PDE)

$$\partial_t u_\varepsilon(t, x) - \mathcal{L}_\varepsilon u_\varepsilon(t, x) = 0, \quad t \in (0, T),\ x \in U, \tag{2.2.2}$$

where $\mathcal{L}_\varepsilon$ is the generator defined in (2.1.18), $T > 0$, and U is an open, bounded interval in $\mathbb{R}$. Let us introduce additional terminology that is commonly used in the analysis of parabolic PDEs, see for example [Fri64; Fri58; Lor17; Lie96]. For $\alpha \in (0,1)$, we say that a function $f\colon (0,T) \times U \to \mathbb{R}$ is locally Hölder-continuous on $(0,T) \times U$ with exponent $\alpha \in (0,1)$ if, for any closed subset $D \subset (0,T) \times U$, there exists a constant $H > 0$ such that for all $(t,x),(r,y) \in D$,

$$|f(t,x) - f(r,y)| \leq H \left[|x-y|^2 + |t-r|\right]^{\alpha/2} \tag{2.2.3}$$

holds. Furthermore, we define the function $d\colon [0,\infty) \times \overline{U} \to \mathbb{R}$ by

$$d(t,x) := \operatorname{dist}(x, \partial U) \wedge \sqrt{t}, \quad t \in [0,\infty),\ x \in \overline{U}. \tag{2.2.4}$$

At last, for $m \in \mathbb{N}_0$ and $\alpha \in (0,1)$, we define the norms

$$\begin{aligned} |d^m, f|_\alpha := & \sup_{(t,x)\in(0,T)\times U} |d(t,x)^m f(t,x)| \\ & + \sup_{\substack{(t,x),(r,y)\in(0,T)\times U \\ (t,x)\neq(r,y)}} |d(t,x) \wedge d(r,y)|^{m+\alpha} \frac{|f(t,x) - f(r,y)|}{[|x-y|^2 + |t-r|]^{\alpha/2}}. \end{aligned} \tag{2.2.5}$$

Observe that whenever the norm $|d^m, f|_\alpha$ is finite for some $m \in \mathbb{N}_0$ and $\alpha \in (0,1)$, then f is locally Hölder-continuous on $(0,T) \times U$ with exponent α.

The following result, whose proof is standard in the literature for PDEs, see [Fri64, Chapter 4], requires, although very technical and tedious, merely an explicit tracking of the constant appearing in the upper bound with respect to ε. A tabular derivation of the final constant is given in Appendix C. The acquired estimates on the constant are far from being optimal in any sense, but, given our assumptions, this will serve our purposes, as we will eventually see.

Proposition 2.2.1. *Let Σ be bounded away from zero on U and let the coefficients Σ, F_ε of the operator $\mathcal{L}_\varepsilon$ satisfy*

$$|d^0, \Sigma|_\alpha \leq K_\Sigma, \quad |d^1, F_\varepsilon|_\alpha \leq K_F\, \varepsilon^{-2}, \tag{2.2.6}$$

for an exponent $\alpha \in (0,1)$ and some constants $K_\Sigma, K_F > 0$. Assume that $u_\varepsilon \in C^{1,2}((0,T)\times U)$ is a solution of (2.2.2) such that $\sup_{(0,T)\times U} |u_\varepsilon| < \infty$ and u_ε, $\partial_x u_\varepsilon$, $\partial_x^2 u_\varepsilon$, $\partial_t u_\varepsilon$ are locally Hölder-continuous on $(0,T) \times U$ with exponent $\alpha \in (0,1)$. Then there exists a constant $K > 0$, only depending on K_Σ, K_F and α, such that for sufficiently small $\varepsilon > 0$,

$$|d^0, u_\varepsilon|_\alpha + |d^1, \partial_x u_\varepsilon|_\alpha + |d^2, \partial_x^2 u_\varepsilon|_\alpha + |d^2, \partial_t u_\varepsilon|_\alpha \leq K \varepsilon^{-(10+4/\alpha)} \sup_{(0,T)\times U} |u_\varepsilon|. \tag{2.2.7}$$

Proof. Carefully tracking the constants in [Fri64, Chapter 4, Section 4] with respect to ε yields the claim; refer to Appendix C for more details. □

Our aim is to apply Proposition 2.2.1 to the time-homogeneous Feller transition probability function P_ε of X^{ε,x_0}, which exists by Theorem A.2.3 and is given by the relation

$$P_\varepsilon(x,t,A) = \mathbb{P}(X_t^{\varepsilon,x} \in A), \quad x \in \mathbb{R},\ t \in (0,\infty),\ A \in \mathscr{B}(\mathbb{R}). \tag{2.2.8}$$

Evidently, it is stochastically continuous, which is to say, by definition, that for all $x \in \mathbb{R}$ and $\delta > 0$

$$P_\varepsilon(x,t,B_\delta(x)) = \mathbb{P}(|X_t^{\varepsilon,x} - x| < \delta) \to 1, \quad t \to 0^+.$$

Hence, [KMN12, Lemma 3.1] implies that $(t,x) \mapsto P_\varepsilon(x,t,A)$ is a $\mathscr{B}([0,\infty)\times\mathbb{R})$-measurable function for any $A \in \mathscr{B}(\mathbb{R})$, which makes it possible to write down certain integrals.

Before we proceed, we introduce the following parabolic PDE with initial and Dirichlet boundary conditions and coefficients in the differential operator that do not depend on time

$$\begin{cases} \partial_t u_\varepsilon(t,x) - \mathcal{L}_\varepsilon u_\varepsilon(t,x) = 0, & t \in (0,T],\ x \in U, \\ u_\varepsilon(0,x) = g(x), & x \in \overline{U}, \\ u_\varepsilon(t,x) = h(t,x), & t \in (0,T],\ x \in \partial U, \end{cases} \tag{2.2.9}$$

where $g \in C(\overline{U})$ and $h \in C([0,T]\times\partial U)$ are given functions satisfying $g(x) = h(0,x)$ for $x \in \partial U$. If there exists a solution to this initial-boundary value problem, then it satisfies the Feynman-Kac formula, cf. [KMN12, Lemma 3.3, Remark 3.13], that is,

$$u_\varepsilon(t,x) = \mathbb{E}\left[g(X_t^{\varepsilon,x})\mathbb{1}_{\{t\le\tau_U\}} + h(t-\tau_U, X_{\tau_U}^{\varepsilon,x})\mathbb{1}_{\{t>\tau_U\}}\right], \quad t\in(0,T],\ x\in U, \tag{2.2.10}$$

where $\tau_U := \inf\{t \ge 0 \,|\, X_t^{\varepsilon,x} \notin U\}$ is the first exit time of the process $X^{\varepsilon,x}$ from U. This representation can be directly proved with the help of Itô's formula; the needed techniques to carry out this proof are standard and can be found in [KS91; PR14]. With these facts at our disposal, we may now state and prove the following lemma.

Lemma 2.2.2. *Assume (C*) and (MET). Then, for any closed set $A \in \mathscr{B}(\mathbb{R})$, the function $(t,x) \mapsto P_\varepsilon(x,t,A)$ is a solution of*

$$\partial_t u_\varepsilon(t,x) - \mathcal{L}_\varepsilon u_\varepsilon(t,x) = 0, \quad (t,x) \in D, \tag{2.2.11}$$

where $D \subset (0,T)\times U$ is an arbitrary subdomain with closure in $(0,T]\times U$. Moreover, for $(t_0,T)\times C \subset (0,T)\times U$ with $t_0 \in (0,T)$ and closed C, and for any $\alpha \in (0,1)$ and sufficiently small $\varepsilon > 0$, we have

$$\sup_{(t,x)\in(t_0,T)\times C} |\partial_x P_\varepsilon(x,t,A)| \le \frac{K\varepsilon^{-(10+4/\alpha)}}{\operatorname{dist}(\partial C,\partial U) \wedge \sqrt{t_0}}, \tag{2.2.12}$$

where $K > 0$ is the same constant as in (2.2.7).

Proof. First, note that (C*) ii) implies that the coefficients Σ, F_ε of the operator $\mathcal{L}_\varepsilon$ satisfy (2.2.6) for any exponent $\alpha \in (0,1)$. Indeed, since the coefficients do not depend on time, it is easy to prove

$$|d^0,\Sigma|_\alpha \le \sup_{x\in\overline{U}} |\Sigma(x)| + \operatorname{diam}(U)^\alpha \sup_{\substack{(x,y)\in\times\overline{U}\\ x\ne y}} \frac{|\Sigma(x)-\Sigma(y)|}{|x-y|^\alpha},$$

$$|d^1,F_\varepsilon|_\alpha \le \operatorname{diam}(U)\sup_{x\in\overline{U}} |F_\varepsilon(x)| + \operatorname{diam}(U)^{1+\alpha} \sup_{\substack{(x,y)\in\times\overline{U}\\ x\ne y}} \frac{|F_\varepsilon(x)-F_\varepsilon(y)|}{|x-y|^\alpha},$$

which implies the inequalities in (2.2.6) after applying (C*) ii). Observe also that these constants do not depend on T. With these estimates and the strict nondegeneracy of Σ, the existence of a solution u_ε to (2.2.9) with locally Hölder-continuous derivatives of all relevant orders for any exponent $\alpha \in (0,1)$ is secured whenever the initial and boundary data are continuous, see [Fri64, Chapter 3, Section 4, Corollary 2]. For $t \in (0,T]$ and $x \in U$, we can write

$$P_\varepsilon(x,t,A) = \mathbb{E}\left[\mathbb{1}_A(X_t^{\varepsilon,x})\mathbb{1}_{\{t\le\tau_U\}}\right] + \mathbb{P}\left(X_t^{\varepsilon,x} \in A, t > \tau_U\right). \tag{2.2.13}$$

Using the strong Markov property, see Appendix A, the second term can be expressed as follows

$$\begin{aligned} \mathbb{P}\left(X_t^{\varepsilon,x} \in A, t > \tau_U\right) &= \int_{\{\tau_U<t\}} \mathbb{P}\left(X_t^{\varepsilon,x} \in A \,\middle|\, \mathcal{F}_{\tau_U}\right)(\omega)\, d\mathbb{P}(\omega) \\ &= \int_\Omega \mathbb{1}_{\{\tau_U(\omega)<t\}} P_\varepsilon\left(X_{\tau_U}^{\varepsilon,x}, t-\tau_U, A\right)(\omega)\, d\mathbb{P}(\omega) \\ &= \int_{\partial U}\int_0^t P_\varepsilon\left(y, t-s, A\right)\, d\mathbb{P}^{\left(\tau_U, X_{\tau_U}^{\varepsilon,x}\right)}(s,y). \end{aligned} \tag{2.2.14}$$

In the last step, we used the $\mathscr{B}([0,\infty)\times\mathbb{R})$-measurability of $(t,x)\mapsto P_\varepsilon(x,t,A)$. Hence, it follows that

$$\mathbb{P}\left(X_t^{\varepsilon,x} \in A, t > \tau_U\right) = \mathbb{E}\left[P_\varepsilon(X_{\tau_U}^{\varepsilon,x}, t-\tau_U, A)\mathbb{1}_{\{t>\tau_U\}}\right]. \tag{2.2.15}$$

For $t \in (0,T]$ and $x \in U$, inserting (2.2.15) into (2.2.13) gives

$$P_\varepsilon(x,t,A) = \mathbb{E}\left[\mathbb{1}_A(X_t^{\varepsilon,x})\mathbb{1}_{\{t\le\tau_U\}} + P_\varepsilon(X_{\tau_U}^{\varepsilon,x}, t-\tau_U, A)\mathbb{1}_{\{t>\tau_U\}}\right], \tag{2.2.16}$$

which is (2.2.10), but with initial and boundary data that are only Borel-measurable. We can take care of this with the following approximation argument. To this end, consider the sequence of functions

$$\kappa_A^n(x) := \max\{0, 1-n\operatorname{dist}(x,A)\}, \quad x\in\mathbb{R}, \quad n\in\mathbb{N}. \tag{2.2.17}$$

Since A is closed, it follows that $\kappa_A^n \downarrow \mathbb{1}_A$ pointwise on $\mathbb{R}$ as $n\to\infty$. Moreover, κ_A^n is continuous and $0 \le \kappa_A^n \le 1$ for each n. Now define for $n \in \mathbb{N}$,

$$u_\varepsilon^n(t,x) := \mathbb{E}\,\kappa_A^n(X_t^{\varepsilon,x}), \quad (t,x)\in[0,T]\times\partial U. \tag{2.2.18}$$

Since U is a bounded interval, ∂U consists of only two points. Hence, in order to check continuity of u_ε^n on $[0,T]\times\partial U$, it suffices to check the continuity in t at each boundary point $x\in\partial U$. But this follows directly from the sample path continuity of $X^{\varepsilon,x}$, continuity of κ_A^n, and the dominated convergence theorem. Thus $u_\varepsilon^n \in C([0,T]\times\partial U)$. By the discussion at the beginning of this proof, for each $n\in\mathbb{N}$, there exists a unique solution v_ε^n to (2.2.9) with initial data $g = \kappa_A^n \in C(\overline{U})$ and boundary data $h = u_\varepsilon^n \in C([0,T]\times\partial U)$ and a representation given by the Feynman-Kac formula, namely

$$v_\varepsilon^n(t,x) = \mathbb{E}\left[\kappa_A^n(X_t^{\varepsilon,x})\mathbb{1}_{\{t\le\tau_U\}} + u_\varepsilon^n(t-\tau_U, X_{\tau_U}^{\varepsilon,x})\mathbb{1}_{\{t>\tau_U\}}\right], \quad t\in(0,T],\ x\in U. \tag{2.2.19}$$

A similar application of the strong Markov property as above shows that

$$v_\varepsilon^n(t,x) = \mathbb{E}\,\kappa_A^n(X_t^{\varepsilon,x}) = u_\varepsilon^n(t,x), \quad t \in (0,T],\ x \in U. \tag{2.2.20}$$

Hence, u_ε^n solves $\partial_t u_\varepsilon^n - \mathcal{L}_\varepsilon u_\varepsilon^n = 0$ on $(0,T] \times U$ and has locally Hölder-continuous derivatives of all relevant orders for any exponent $\alpha \in (0,1)$. A well-known result from the PDE literature, e.g., [Fri64, Chapter 3, Section 6, Theorem 15], provides us with a subsequence $u_\varepsilon^{n_j}$ of u_ε^n that converges pointwise on D to a function u_ε^∞ as $j \to \infty$, where $D \subset (0,T) \times U$ is an arbitrary subdomain with closure in $(0,T] \times U$. Furthermore, this limit function satisfies $\partial_t u_\varepsilon^\infty - \mathcal{L}_\varepsilon u_\varepsilon^\infty = 0$ on D and also has locally Hölder-continuous derivatives of all relevant orders for any exponent $\alpha \in (0,1)$. We will now show that $u_\varepsilon^\infty(t,x) = P_\varepsilon(x,t,A)$ for $(t,x) \in D$. On the one hand, we have the pointwise convergence

$$\lim_{j\to\infty} u_\varepsilon^{n_j}(t,x) = u_\varepsilon^\infty(t,x), \quad (t,x) \in D,$$

and, on the other hand, the dominated convergence theorem gives

$$\lim_{j\to\infty} u_\varepsilon^{n_j}(t,x) = \lim_{j\to\infty} \mathbb{E}\,\kappa_A^{n_j}(X_t^{\varepsilon,x}) = \mathbb{E}\,\mathbb{1}_A(X_t^{\varepsilon,x}), \quad (t,x) \in (0,T] \times U.$$

Eventually, this implies

$$u_\varepsilon^\infty(t,x) = P_\varepsilon(x,t,A), \quad (t,x) \in D. \tag{2.2.21}$$

Hence, by way of this approximation argument, we have shown that the function $(t,x) \mapsto P_\varepsilon(x,t,A)$ fulfills (2.2.11) and, in particular, has locally Hölder-continuous derivatives of all relevant orders for any exponent $\alpha \in (0,1)$. At last, we can derive (2.2.12), which is a consequence of (2.2.7) since

$$\sup_{(t,x)\in(t_0,T)\times C} |\partial_x P_\varepsilon(x,t,A)| \le \frac{|d^1, \partial_x P_\varepsilon(\cdot,\cdot,A)|_\alpha}{\operatorname{dist}(\partial C, \partial U) \wedge \sqrt{t_0}} \le \frac{K\varepsilon^{-(10+4/\alpha)}}{\operatorname{dist}(\partial C, \partial U) \wedge \sqrt{t_0}}. \qquad \square$$

For the remainder of this subsection and the next subsection, we will need the following additional assumption.

Assumptions (R). *There exist numbers $S, \gamma > 0$ such that for all $|y| > S$,*

$$\operatorname{sgn}(y) \frac{\bar{F}(y)}{\bar{\Sigma}(y)^2} \le -\gamma. \tag{2.2.22}$$

Informally speaking, the inequality above indicates that the drift coefficient of the limit SDE (2.1.16), when sufficiently far away from the origin, stays below the threshold $-\gamma$ and pulls the dynamics of X towards the origin, resulting in a solution with infinite explosion time and strong recurrence properties. This condition already implies the conditions (MET) i) and ii), so that, by Remark 2.1.2, the invariant densities μ and μ_ε of X and X^ε, respectively, exist. Furthermore, if (MET) iii) holds, then it is easy to prove that, for all $\varepsilon > 0$,

$$c_\mu \mu \le \mu_\varepsilon \le C_\mu \mu \quad \text{on } \mathbb{R}, \tag{2.2.23}$$

with some constants $c_\mu, C_\mu > 0$ that are independent of ε.

In the proof of the next Proposition 2.2.3, we will make use of recurrence properties of the process X^{ε,x_0}; refer to [KMN12; Dur96] for different aspects of recurrence properties of SDE solutions. First, note that, under Assumptions (C*) and (MET), the process X^{ε,x_0} is recurrent for each $\varepsilon > 0$, i.e., for any $y \in \mathbb{R}$ and $\varepsilon > 0$, it is true that

$$\mathbb{P}(\tau^y\left(X^{\varepsilon,x_0}\right) < \infty) = 1, \tag{2.2.24}$$

where we recall $\tau^y\left(X^{\varepsilon,x_0}\right) = \inf\{t > 0 \,|\, X_t^{\varepsilon,x_0} = y\}$. This is a consequence of [KMN12, Lemma 3.9, Remark 15], cf. with [KMN12, Example 3.10], as well. We even have positive recurrence of X^{ε,x_0} uniformly in ε, which implies recurrence as described above. That is to say, for any $y \in \mathbb{R}$, we have

$$\sup_{\varepsilon>0} \mathbb{E}\,\tau^y\left(X^{\varepsilon,x_0}\right) < \infty. \tag{2.2.25}$$

Indeed, for all $\varepsilon > 0$ and $y \in \mathbb{R}$, Lemma 2.1.4 implies

$$\mathbb{E}\,\tau^y\left(X^{\varepsilon,x_0}\right) = \mathbb{E}\,\tau^{f_\varepsilon(y)}(\xi^{\varepsilon,f_\varepsilon(x_0)}) \le 2D_\rho|f_\varepsilon(x_0) - f_\varepsilon(y)| \le 2D_\rho(|f(x_0)| + |f(y)|).$$

Proposition 2.2.3. *If Assumptions (C*), (MET), and (R) hold, then*

$$\mathbb{P}\left(\left|X_{T_\varepsilon}^{\varepsilon,x_0}\right| > R(\varepsilon)\right) \to 0, \quad \varepsilon \to 0, \tag{2.2.26}$$

for any initial condition $x_0 \in \mathbb{R}$, where $R(\varepsilon) = \Omega(\varepsilon^{-\eta})$ with $\eta > 0$.

Proof. Fix $I := (-y, y)$ for some $y > 1$. Consider the initial datum $x_0 \in I$ first, and let $T > 0$ be arbitrary. For all $x \in B_{r(\varepsilon)}(x_0)$, where $r(\varepsilon) \downarrow 0$ as $\varepsilon \to 0$ will be determined soon, $t_0 \in (0, T)$, and all $\alpha \in (0, 1)$, the mean value theorem and Lemma 2.2.2 yield

$$\left|P_\varepsilon(x, T, B_{R(\varepsilon)}(0)^c) - P_\varepsilon(x_0, T, B_{R(\varepsilon)}(0)^c)\right| \le \frac{K\varepsilon^{-(10+4/\alpha)}}{\operatorname{dist}(\partial B_{r(\varepsilon)}(x_0), \partial I) \wedge \sqrt{t_0}}|x - x_0|. \tag{2.2.27}$$

Evidently, we have $\operatorname{dist}(\partial B_{r(\varepsilon)}(x_0), \partial I) \to \operatorname{dist}(x_0, \partial I) = |x_0 - y| > 0$ as $\varepsilon \to 0$, so that choosing $r(\varepsilon) := \varepsilon^{11+4/\alpha}$ gives

$$\left|P_\varepsilon(x, T, B_{R(\varepsilon)}(0)^c) - P_\varepsilon(x_0, T, B_{R(\varepsilon)}(0)^c)\right| \le K\varepsilon, \tag{2.2.28}$$

with a new constant $K > 0$ independent of ε. As the next step towards verifying the claim, we may estimate, using (2.2.23) and stationarity as in (2.1.44),

$$\int_{\mathbb{R}} P_\varepsilon(x, T, B_{R(\varepsilon)}(0)^c)\mu_\varepsilon(x)\,dx = \int_{B_{R(\varepsilon)}(0)^c} \mu_\varepsilon(x)\,dx \le C_\mu \int_{B_{R(\varepsilon)}(0)^c} \mu(x)\,dx. \tag{2.2.29}$$

Also, using (2.2.28), it follows

$$\begin{aligned}\int_{\mathbb{R}} P_\varepsilon(x, T, B_{R(\varepsilon)}(0)^c)\mu_\varepsilon(x)\,dx &\ge c_\mu \int_{B_{r(\varepsilon)}(x_0)} P_\varepsilon(x, T, B_{R(\varepsilon)}(0)^c)\mu(x)\,dx \\ &\ge c_\mu \int_{B_{r(\varepsilon)}(x_0)} \mu(x)\,dx\left[P_\varepsilon(x_0, T, B_{R(\varepsilon)}(0)^c) - K\varepsilon\right].\end{aligned}$$

Rearranging terms yields

$$\mathbb{P}\left(|X_T^{\varepsilon,x_0}| > R(\varepsilon)\right) = P_\varepsilon(x_0, T, B_{R(\varepsilon)}(0)^c) \le K\varepsilon + \frac{C_\mu \int_{B_{R(\varepsilon)}(0)^c} \mu(x)\, dx}{c_\mu \int_{B_{r(\varepsilon)}(x_0)} \mu(x)\, dx} \tag{2.2.30}$$

for all $T > 0$ and $x_0 \in I$. Observe that, by continuity, we have

$$\lim_{\varepsilon \to 0} \frac{1}{2r(\varepsilon)} \int_{B_{r(\varepsilon)}(x_0)} \mu(x)\, dx = \mu(x_0) > 0, \quad \varepsilon \to 0, \tag{2.2.31}$$

on the one hand. On the other hand, for $x > S$, Assumption (R) implies

$$\begin{aligned}\mu(x) = \frac{1}{Z} \exp\left(2\int_0^x \frac{\bar{F}(y)}{\bar{\Sigma}(y)^2}\, dy\right) &\le C(\bar{F}, \bar{\Sigma}, S) \exp\left(2\int_S^x \frac{\bar{F}(y)}{\bar{\Sigma}(y)^2}\, dy\right)\\ &\le C(\bar{F}, \bar{\Sigma}, S, \gamma) \exp(-2\gamma x),\end{aligned}$$

and, similarly, for $x < -S$ but with different signs. For sufficiently small $\varepsilon > 0$ such that $R(\varepsilon) \ge S$, the last estimate gives

$$\int_{B_{R(\varepsilon)}(0)^c} \mu(x)\, dx \le C(\bar{F}, \bar{\Sigma}, S, \gamma) \exp(-2\gamma R(\varepsilon)). \tag{2.2.32}$$

Combining (2.2.30), (2.2.31), and (2.2.32) eventually yields

$$\mathbb{P}\left(|X_T^{\varepsilon,x_0}| > R(\varepsilon)\right) \le K\varepsilon + C(\bar{F}, \bar{\Sigma}, S, \gamma, \mu) \frac{\exp(-2\gamma R(\varepsilon))}{r(\varepsilon)\mu(x_0)} \to 0, \quad \varepsilon \to 0, \tag{2.2.33}$$

while using our condition on $R(\varepsilon)$. Since the appearing estimates did not depend on T, this preceding result also holds when $T = T_\varepsilon$.
Now consider $x_0 \in I^c$, assume $x_0 > y$ without loss of generality, and let $\delta > 0$. By (2.2.25) and Markov's inequality, we can choose $T_0 := T_0(\delta, x_0, y) > 0$, independent of ε, such that for all $\varepsilon > 0$,

$$\mathbb{P}(\tau^{y-1}(X^{\varepsilon,x_0}) > T_0) \le \frac{\sup_{\varepsilon > 0} \mathbb{E}\, \tau^{y-1}(X^{\varepsilon,x_0})}{T_0} < \delta. \tag{2.2.34}$$

Notice that we use $y - 1 \in I$ here, because we soon want to utilize (2.2.33), which was only proved for interior points $x_0 \in I$. Next, for sufficiently small $\varepsilon > 0$ such that $T_\varepsilon \ge T_0$, we have

$$\mathbb{P}\left(\left|X_{T_\varepsilon}^{\varepsilon,x_0}\right| > R(\varepsilon)\right) \le \delta + \mathbb{P}\left(\left|X_{T_\varepsilon}^{\varepsilon,x_0}\right| > R(\varepsilon), \tau^{y-1}(X^{\varepsilon,x_0}) \le T_\varepsilon\right).$$

Using (2.2.33) and the strong Markov property again, the last term can be estimated as follows

$$\begin{aligned}\mathbb{P}\left(\left|X_{T_\varepsilon}^{\varepsilon,x_0}\right| > R(\varepsilon), \tau^{y-1}(X^{\varepsilon,x_0}) \le T_\varepsilon\right) &= \int_0^{T_\varepsilon} P_\varepsilon\left(y-1, T_\varepsilon - s, B_{R(\varepsilon)}(0)^c\right) d\mathbb{P}^{\tau^{y-1}(X^{\varepsilon,x_0})}(s)\\ &\le K\varepsilon + C(\bar{F}, \bar{\Sigma}, S, \gamma, \mu) \frac{\exp(-2\gamma R(\varepsilon))}{r(\varepsilon)\mu(y-1)}.\end{aligned}$$

Hence, we obtain

$$\lim_{\varepsilon \to 0} \mathbb{P}\left(\left|X_{T_\varepsilon}^{\varepsilon,x_0}\right| > R(\varepsilon)\right) \le \delta,$$

which yields the claim for the case $x_0 \in I^c$, as well. □

Remark 2.2.4. *In the proof of Proposition 2.2.3, instead of $R(\varepsilon) = \Omega(\varepsilon^{-\eta})$, one can also choose $R(\varepsilon) = -(2\gamma)^{-1}\Omega(\log(r(\varepsilon)\varepsilon^{\eta}))$ for some $\eta > 0$ as $\varepsilon \to 0$ and obtain the same convergence result.*

Corollary 2.2.5. *Let Assumptions (C*), (MET), and (R) hold, and consider $R_\varepsilon = \Omega(\varepsilon^{-\eta})$ for some $\eta > 0$ as $\varepsilon \to 0$. Then, for all $\delta > 0$ and any $x_0 \in \mathbb{R}$, we have*

$$\lim_{\varepsilon\to 0} \mathbb{P}\left(\frac{\left|X_{T_\varepsilon}^{\varepsilon,x_0}\right|}{R_\varepsilon} > \delta\right) = 0. \tag{2.2.35}$$

In other words, $X_{T_\varepsilon}^{\varepsilon,x_0}/R_\varepsilon$ converges in probability to zero as $\varepsilon \to 0$.

Remarks on multidimensional extensions III.
As in the preceding remarks on multidimensional extensions, we observe that Markov structure plays an important role in this section, as well. Fortunately, some other results in this section, although formulated for the 1-dimensional case only, are based on standard PDE arguments which are easily extendable to the multidimensional setting of a fully coupled multiscale SDE system (1.2.1). For example, the interior Schauder estimates can be copied almost word-for-word, see Appendix C. But also results like Proposition 2.2.3 may be proved for the multidimensional case since they are based on arguments in [KMN12, Chapter 4], which were originally proved in multiple dimensions. However, as always with multiscale systems, one needs good, ideally uniform, estimates of relevant quantities with respect to ε. In the context of this section, these relevant quantities are the invariant measure of the full multiscale system $(X^\varepsilon, Y^\varepsilon)$ and mean hitting times for bounded domains of these coupled processes.

2.3 An ε-Dependent Central Limit Theorem

As we already motivated at the beginning of Section 2.1, our ultimate goal in this chapter is a rigorous CLT for time integrals. Up until now, most of the efforts were concentrated on ergodic theorems for time integrals, so that it should not come as a surprise that the ergodic theorem is needed for the CLT. In fact, it is often formulated as a sufficient condition for the CLT of time integrals, cf. [KLO12; Kut94; Kut76; Kut04], and we will follow the same route. Interestingly though, this stands in contrast to the classical CLT by Lindeberg-Feller and the law of large numbers for i.i.d. random variables, see, for example, [Chu01].

For a given function $h_\varepsilon \colon \mathbb{R} \to \mathbb{R}$ and the solution X^ε of (2.1.15) with fixed initial condition $X_0^\varepsilon = x_0$, we want to prove

$$\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\, dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0, \tau^2), \quad \varepsilon \to 0, . \tag{2.3.1}$$

We will achieve this by employing the technique of the Poisson equation, which is classical in the literature, cf. [KLO12; MST10], and was already used in Subsection 2.1.2. The idea works as follows.

Let $L^1(\mu_\varepsilon) := L^1(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu_\varepsilon)$ be the weighted L^1-space with respect to the invariant density μ_ε of Remark 2.1.2, and define

$$L^1_0(\mu_\varepsilon) := \left\{ f \in L^1(\mu_\varepsilon) \,\middle|\, \int_{\mathbb{R}} f(x)\mu_\varepsilon(x)\,dx = 0 \right\}. \tag{2.3.2}$$

For $\varepsilon > 0$, we consider a continuous function $h_\varepsilon \in L^1_0(\mu_\varepsilon)$. If there exists a solution $\Phi_\varepsilon \in C^2(\mathbb{R})$ to the Poisson equation

$$-\mathcal{L}_\varepsilon \Phi_\varepsilon = h_\varepsilon, \tag{2.3.3}$$

where $\mathcal{L}_\varepsilon$ is the generator of X^ε introduced in (2.1.18), then applying the Itô formula to Φ_ε yields

$$\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} h_\varepsilon(X^\varepsilon_t)\,dt = \frac{\Phi_\varepsilon(x_0) - \Phi_\varepsilon(X^\varepsilon_{T_\varepsilon})}{\sqrt{T_\varepsilon}} + \frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \Sigma(X^\varepsilon_t)\Phi'_\varepsilon(X^\varepsilon_t)\,dW_t. \tag{2.3.4}$$

Inspecting equation (2.3.4), we notice that, in order to obtain the CLT (2.3.1), we need to control the first term on the RHS in such a way that it vanishes in probability as $\varepsilon \to 0$ and use a CLT for the remaining stochastic integral. The latter claim is contained in [RY98, Exercise (IV.3.33)], with proofs given in [Kut76] and [Pra99, Appendix B], so that we merely state the result in the way that we require.

Proposition 2.3.1. *For every $\varepsilon > 0$, let B^ε be an adapted process with respect to the filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$ such that $\int_0^{T_\varepsilon} (B^\varepsilon_t)^2\,dt < \infty$, $\mathbb{P}$-a.s. If there exists $\tau > 0$ such that*

$$\frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} (B^\varepsilon_t)^2\,dt \xrightarrow{\mathbb{P}} \tau^2, \quad \varepsilon \to 0, \tag{2.3.5}$$

then

$$\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} B^\varepsilon_t\,dW_t \xrightarrow{\mathcal{D}} \mathcal{N}_1(0, \tau^2), \quad \varepsilon \to 0. \tag{2.3.6}$$

To apply this proposition to (2.3.4), we first provide a lemma concerning the properties of Φ_ε and its derivative. The main ideas for this lemma are borrowed from the proof of [Kut04, Lemma 1.17] and are modified to our setting.

Lemma 2.3.2. *Let Assumptions (C*), (MET), and (R) hold. Assume that the functions $h_\varepsilon \in L^1_0(\mu_\varepsilon)$, $\varepsilon > 0$, are continuous and satisfy $\sup_{\varepsilon>0} \|h_\varepsilon\|_\infty < \infty$. Then the sequence of functions Φ'_ε, where Φ_ε solves (2.3.3), is uniformly bounded, i.e., $\sup_{\varepsilon>0} \|\Phi'_\varepsilon\|_\infty < \infty$.*

Proof. In this particular 1-dimensional setting, the function Φ_ε is given by

$$\Phi_\varepsilon(x) = -\int_0^x \frac{2}{\Sigma(y)^2\mu_\varepsilon(y)} \int_{-\infty}^y h_\varepsilon(z)\mu_\varepsilon(z)\,dz\,dy, \quad x \in \mathbb{R}. \tag{2.3.7}$$

Its first derivative equals

$$\Phi'_\varepsilon(x) = -\frac{2}{\Sigma(x)^2\mu_\varepsilon(x)} \int_{-\infty}^x h_\varepsilon(z)\mu_\varepsilon(z)\,dz, \quad x \in \mathbb{R}. \tag{2.3.8}$$

Recall the inequality (2.2.23), namely

$$c_\mu \mu \leq \mu_\varepsilon \leq C_\mu \mu \quad \text{on } \mathbb{R},$$

and observe that $h_\varepsilon \in L_0^1(\mu_\varepsilon)$ implies

$$\Phi_\varepsilon'(x) = -\frac{2}{\Sigma(x)^2 \mu_\varepsilon(x)} \int_{-\infty}^{x} h_\varepsilon(z)\mu_\varepsilon(z)\, dz = \frac{2}{\Sigma(x)^2 \mu_\varepsilon(x)} \int_{x}^{\infty} h_\varepsilon(z)\mu_\varepsilon(z) dz, \quad x \in \mathbb{R}.$$

Now, let S, γ be as in Asumptions (R). By using (C*) ii), the inequality in (R), and the uniform boundedness of h_ε, we obtain

$$\begin{aligned} |\Phi_\varepsilon'(x)| &\leq \frac{2C_\mu}{c_\mu} \int_x^\infty \frac{|h_\varepsilon(z)|}{\Sigma(z)^2} \exp\left(\int_x^z \frac{2\bar{F}(y)}{\bar{\Sigma}(y)^2}\, dy \right) dz \\ &\leq \frac{2C_\mu \sup_{\varepsilon>0} \|h_\varepsilon\|_\infty}{c_\mu m_0} \int_x^\infty \exp\left(-2\gamma(z-x)\right) dz = \frac{C_\mu \sup_{\varepsilon>0} \|h_\varepsilon\|_\infty}{\gamma c_\mu m_0} \end{aligned}$$

for $x > S$. A similar estimate is also true for $x < -S$. For $x \in [-S, S]$, it obviously holds

$$|\Phi_\varepsilon'(x)| \leq \frac{\sup_{\varepsilon>0} \|h_\varepsilon\|_\infty}{c_\mu m_0 \mu(x)},$$

which implies the claim. □

For the final result, we define $L^1(\mu)$ and $L_0^1(\mu)$ with respect to the invariant density μ of Remark 2.1.2 in the same way as with μ_ε before. For a continuous function $h \in L_0^1(\mu)$, we define

$$\Phi(x) := -\int_0^x \frac{2}{\bar{\Sigma}(y)^2 \mu(y)} \int_{-\infty}^{y} h(z)\mu(z)\, dz\, dy, \quad x \in \mathbb{R}. \tag{2.3.9}$$

This function solves the Poisson equation $-\mathcal{L}\Phi = h$ with the generator $\mathcal{L}$ of X given in (2.1.18).

Theorem 2.3.3. *Let Assumptions (C*), (MET), and (R) hold, and consider* $T_\varepsilon = \Omega(\varepsilon^{-\eta})$ *for some* $\eta > 0$ *as* $\varepsilon \to 0$. *Assume that* Σ *is bounded on* $\mathbb{R}$ *and that the functions* $h_\varepsilon \in L^1(\mu_\varepsilon)$, $\varepsilon > 0$, *and* $h \in L_0^1(\mu)$ *are continuous and satisfy* $\sup_{\varepsilon>0} \|h_\varepsilon\|_\infty < \infty$ *and* $\|h\|_\infty < \infty$. *Additionally, assume that*

$$\tau_\varepsilon^2 := \int_\mathbb{R} \Sigma(x)^2 \Phi_\varepsilon'(x)^2 \mu_\varepsilon(x)\, dx \to \int_\mathbb{R} \bar{\Sigma}(x)^2 \Phi'(x)^2 \mu(x)\, dx =: \tau^2, \quad \varepsilon \to 0, \tag{2.3.10}$$

where Φ_ε *and* Φ *solve the Poisson equations*

$$-\mathcal{L}_\varepsilon \Phi_\varepsilon = h_\varepsilon, \qquad -\mathcal{L}\Phi = h, \tag{2.3.11}$$

respectively. Then

$$\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\, dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0, \tau^2), \quad \varepsilon \to 0\,. \tag{2.3.12}$$

Proof. Recall the equation (2.3.4)

$$\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} h_\varepsilon(X_t^\varepsilon)\,dt = \frac{\Phi_\varepsilon(x_0)-\Phi_\varepsilon(X_{T_\varepsilon}^\varepsilon)}{\sqrt{T_\varepsilon}} + \frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} \Sigma(X_t^\varepsilon)\Phi_\varepsilon'(X_t^\varepsilon)\,dW_t.$$

By virtue of Corollary 2.2.5, the first term on the right-hand side vanishes in probability as $\varepsilon \to 0$ because Φ_ε' is uniformly bounded by Lemma 2.3.2. For the second term, we first bound

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\Sigma(X_t^\varepsilon)^2\Phi_\varepsilon'(X_t^\varepsilon)^2\,dt-\tau^2\right| \le \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\Sigma(X_t^\varepsilon)^2\Phi_\varepsilon'(X_t^\varepsilon)^2\,dt-\tau_\varepsilon^2\right| + |\tau_\varepsilon^2-\tau^2|,$$

and then we apply Theorem 2.1.10 to get

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\Sigma(X_t^\varepsilon)^2\Phi_\varepsilon'(X_t^\varepsilon)^2\,dt-\tau_\varepsilon^2\right| \to 0, \quad \varepsilon\to 0.$$

By condition (2.3.10) it therefore follows

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\Sigma(X_t^\varepsilon)^2\Phi_\varepsilon'(X_t^\varepsilon)^2\,dt-\tau^2\right| \to 0, \quad \varepsilon\to 0.$$

Hence, in view of Proposition 2.3.1 and Slutzky's lemma B.1.5, the claim is established. □

Remarks on multidimensional extensions IV.
It should be clear by now that the prevailing approach of this chapter is the Poisson equation. And this is especially true for the proof of the CLT for time integrals. In multiple dimensions, the idea carries out in the same way. If we have a multiscale system (1.2.1) with generator $\mathcal{L}_\varepsilon$ as given in (2.1.107) and a function $h\colon \mathbb{R}^{d_x}\to\mathbb{R}$, then we consider the Poisson equation with respect to the full multiscale generator, that is,

$$-\mathcal{L}_\varepsilon\Phi_\varepsilon(x,y) = \bar{h}_\varepsilon(x) := h(x) - \int_{\mathbb{R}^{d_x}\times\mathbb{R}^{d_y}} h(x)\,d\nu_\varepsilon(x,y),$$

where ν_ε is the invariant measure of $(X^\varepsilon, Y^\varepsilon)$. Applying the Itô formula to Φ_ε, inserting the Poisson equation, and rearranging terms yields

$$\begin{aligned}\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{h}_\varepsilon(X_t^\varepsilon)\,dt &= \frac{\Phi_\varepsilon(X_0^\varepsilon,Y_0^\varepsilon)-\Phi_\varepsilon(X_{T_\varepsilon}^\varepsilon,Y_{T_\varepsilon}^\varepsilon)}{\sqrt{T_\varepsilon}}\\ &\quad+\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\nabla_x\Phi_\varepsilon(X_t^\varepsilon,Y_t^\varepsilon)\Sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(x)}\\ &\quad+\frac{1}{\varepsilon\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\nabla_y\Phi_\varepsilon(X_t^\varepsilon,Y_t^\varepsilon)\sigma(X_t^\varepsilon,Y_t^\varepsilon)\,dW_t^{(y)}.\end{aligned}$$

As we can see from this representation, it is unavoidable to have quantities of the full multiscale system $(X^\varepsilon, Y^\varepsilon)$ on the RHS, even if the LHS only depends on the slow variable X^ε. The first term might be controllable with the results of Section 2.2 extended to multiple dimensions, but even then, it is unclear if one can get good estimates for the solution Φ_ε with respect to ε. Considering Proposition 2.3.1, which can be generalized to multiple dimensions, see [Kut76], the representation also suggests that one is anyway forced to obtain an ergodic theorem for the full multiscale system; at least if one aims for a CLT. A possible approach into this direction could be the work in [MSH02] in which the authors establish geometric ergodicity for various SDE systems. From geometric ergodicity as desribed in that paper, one can then derive a mean ergodic theorem, see the main arguments in the proof by Gikhman and Skorokhod [GS68, §18]. Apart from these problems, the convergence of asymptotic variances as assumed in (2.3.10) poses considerable difficulties, too. Except for specific examples, which we will present in the next section, it is not clear how to obtain the convergence of the unhomogenized asymptotic variance

$$\int_{\mathbb{R}^{d_x}\times\mathbb{R}^{d_y}} \left[\left|\nabla_x\Phi_\varepsilon(x,y)\Sigma(x,y)\right|_2^2 + \varepsilon^{-2}\left|\nabla_y\Phi_\varepsilon(x,y)\sigma(x,y)\right|_2^2\right] d\nu_\varepsilon(x,y)$$

to the homogenized asymptotic variance

$$\int_{\mathbb{R}^{d_x}} \left|\nabla_x\Phi(x)\bar{\Sigma}(x)\right|_2^2 \, d\mu(x)$$

as $\varepsilon \to 0$, where μ is the invariant measure of the homogenized SDE solution X and Φ solves the Poisson equation

$$-\mathcal{L}\Phi = h - \int_{\mathbb{R}^{d_x}} h(x)\, d\mu(x).$$

All in all, unless one comes up with an alternative approach to the limit theorems in multiple dimensions with fully coupled multiscale systems, one is probably bound to deal with some challenging theoretical problems.

2.4 Multiscale Overdamped Langevin Equation

In this section, we derive several results surrounding our accompanying example introduced in Chapter 1. We will see that, under a small set of assumptions, the multiscale overdamped Langevin equation (1.2.10) and its homogenized limit equation (1.2.12) fulfill the conditions of the preceding sections, so that the limit theorems are directly applicable in this case.

2.4.1 Multidimensional Case

We will use the following set of assumptions for the coefficient functions of the multiscale overdamped Langevin equation (1.2.10) and the homogenized Langevin equation (1.2.12). In the context of homogenization, it is customary to call V the large-scale potential and p the fast-scale periodic oscillation.

Assumptions (L).

i) $V, p \in C^\infty(\mathbb{R}^d)$, p is $\mathbb{L}$-periodic in all directions with period $\mathbb{L} > 0$, and $\alpha, \sigma > 0$.

ii) For any multi-index $\beta \in \mathbb{N}_0^d$, the derivative $\partial^\beta V$ is at most polynomially bounded, i.e., there exists an $N \in \mathbb{N}_0$ and a constant $C_{N,\beta} > 0$ such that for all $x \in \mathbb{R}^d$,

$$|\partial^\beta V(x)| \leq C_{N,\beta}(1 + |x|_2^N). \tag{2.4.1}$$

iii) There exist $a, b > 0$ and $R > 0$ such that for all $x \in \mathbb{R}^d$ with $|x| \geq R$,

$$-\nabla V(x)x \leq a - b|x|_2^2. \tag{2.4.2}$$

It is not necessary to impose so much regularity on the coefficients, but it makes the presentation of the results here and in later chapters slightly less technical. That being said, it should become clear from the statement and the proofs of the results how much regularity is actually required.

Remark 2.4.1. *Most of our results are established for* 1*-dimensional SDEs, so that we will rarely consider the Langevin equations* (1.2.10) *and* (1.2.12) *in arbitrary dimensions, but let us remark that, under Assumptions (L), there exists a unique strong solution to* (1.2.10) *and* (1.2.12)*, respectively. This is a consequence of the local Lipschitz-continuity of the coefficients and the fact that (L) iii) gives the explicit Lyapunov function $x \to |x|_2^2$ for each of the generators $\mathcal{L}_\varepsilon$ and $\mathcal{L}$, see [KMN12, Theorem 3.5] or Theorem A.2.3. In this case, the generators of X^ε and X are given by*

$$\mathcal{L}_\varepsilon = \sigma \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} - \sum_{i=1}^d \left[\alpha \partial_{x_i} V + \frac{1}{\varepsilon} \partial_{x_i} p\left(\frac{\cdot}{\varepsilon}\right)\right] \frac{\partial}{\partial x_i}, \tag{2.4.3}$$

$$\mathcal{L} = \sigma \mathcal{K} \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} - \sum_{i=1}^d \alpha \mathcal{K} \partial_{x_i} V \frac{\partial}{\partial x_i}, \tag{2.4.4}$$

respectively, with $\mathcal{K}$ defined in (1.2.13)*. Furthermore, although the homogenization results for the Langevin equations are based on globally Lipschitz-continuous coefficients, cf. [DDP23], one can suspect that it also holds in the local case, as has been indicated in [PSV77].*

We will first prove a multitude of analytical results in arbitrary dimensions and later specialize to the case $d = 1$ as the limit theorems of the previous sections were proved for 1-dimensional SDEs only. We start with the following simple fact. Here, $\mathscr{S}(\mathbb{R}^d)$ stands for the space of Schwartz functions defined on $\mathbb{R}^d$.

Lemma 2.4.2. *Let Assumptions (L) hold. Then $x \mapsto \exp(-V(x)) \in \mathscr{S}(\mathbb{R}^d)$.*

Proof. We first derive a lower bound for V. For this, fix a vector $\theta \in \mathbb{S}^{d-1}$, where $\mathbb{S}^{d-1}$ is the d-dimensional unit sphere in $\mathbb{R}^d$, and define $f(r) := V(r\theta)$ for $r > 0$. Then

$$f'(r) = \nabla V(r\theta)\theta = \frac{1}{r} \nabla V(r\theta) r\theta.$$

For $r \geq R$, condition (L) iii) gives

$$f'(r) \geq \frac{1}{r}(br^2 - a) = br - \frac{a}{r}.$$

Integrating from R to $r \geq R$ yields

$$f(r) - f(R) \geq \int_R^r \left(bs - \frac{a}{s}\right) ds = \frac{b}{2}(r^2 - R^2) - a \log \frac{r}{R}.$$

Hence

$$V(r\theta) \geq V(R\theta) + \frac{b}{2}(r^2 - R^2) - a \log \frac{r}{R}.$$

Since $\theta \mapsto V(R\theta)$ is continuous on the unit sphere $\mathbb{S}^{d-1}$, which is compact, it attains a minimum m. Therefore, for all $x \in \mathbb{R}^d$ with $|x|_2 = r \geq R$, we have

$$V(x) \geq m + \frac{b}{2}(r^2 - R^2) - a \log \frac{r}{R}. \tag{2.4.5}$$

From this estimate, it follows that there exists a constant $C > 0$ such that for all $|x| \geq R$,

$$\exp(-V(x)) \leq C|x|_2^a \exp\left(-\frac{b}{2}|x|_2^2\right). \tag{2.4.6}$$

Let $\alpha \in \mathbb{N}_0^d$ be a multi-index. By induction on $|\alpha| = \sum_{i=1}^d \alpha_i$, we have

$$\partial^\alpha(\exp(-V))(x) = \exp(-V(x))P_\alpha(x), \tag{2.4.7}$$

where $P_\alpha(x)$ is a finite sum of products of derivatives of V of order between 1 and $|\alpha|$. Indeed, for $|\alpha| = 0$, this is trivial with $P_0 \equiv 1$. So, assume the statement holds for $|\alpha|$. Let $i \in \{1, \dots, d\}$ and consider $\alpha + e_i$ where e_i is the i-th standard unit vector. Then

$$\partial^{\alpha+e_i}(\exp(-V)) = \partial_{x_i}\Big(\exp(-V)P_\alpha\Big) = (\partial_{x_i}\exp(-V))P_\alpha + \exp(-V)\partial_{x_i}P_\alpha.$$

Since $\partial_{x_i}\exp(-V) = -(\partial_{x_i}V)\exp(-V)$, we obtain

$$\partial^{\alpha+e_i}(\exp(-V)) = \exp(-V)\left(-(\partial_{x_i}V)P_\alpha + \partial_{x_i}P_\alpha\right),$$

which has the required form. This completes the induction. With formula (2.4.7), estimate (2.4.6), and the polynomial growth condition (2.4.1), we therefore obtain the claim. $\square$

In the next result, we derive an estimate for an integral that is related to certain oscillatory integrals and which produces quantitative weak convergence rates for the invariant measures of our considered Langevin equations. In the proof, we will make heavy use of multi-index notation and Fourier analysis on the d-dimensional torus $\mathbb{T}^d$, that is, we fix $\mathbb{L} = 1$. A standard text for the latter topics is [Gra14, Chapter 3]. Recall here that if $\lambda \in \mathbb{R}^d \backslash \{0\}$, then successive integration by parts gives the formula

$$\int_{\mathbb{R}^d} (\partial^\beta f)(x) \exp\left(2\pi i \lambda^\top x\right) dx = (-1)^{|\beta|}(2\pi i \lambda)^\beta \int_{\mathbb{R}^d} f(x) \exp\left(2\pi i \lambda^\top x\right) dx, \tag{2.4.8}$$

for any multi-index $\beta \in \mathbb{N}_0^d$ and $f \in \mathscr{S}(\mathbb{R}^d)$. This is obviously just the differentiation rule for the Fourier transform on $\mathbb{R}^d$ and will be useful for what follows. We denote by $W^{k,1}(\mathbb{R}^d)$ the Sobolev space of functions defined on $\mathbb{R}^d$ whose weak derivatives are integrable up to order k. More concretely, let

$$W^{k,1}(\mathbb{R}^d) := \left\{ f \in L^1(\mathbb{R}^d) \,\middle|\, \partial^\beta f \in L^1(\mathbb{R}^d)\ \forall \beta \in \mathbb{N}_0^d \text{ with } |\beta| \leq k \right\} \tag{2.4.9}$$

be equipped with the standard Sobolev norm $\|\cdot\|_{W^{k,1}(\mathbb{R}^d)}$ given by

$$\|f\|_{W^{k,1}(\mathbb{R}^d)} := \sum_{\substack{\beta \in \mathbb{N}_0^d \\ |\beta| \leq k}} \|\partial^\beta f\|_{L^1(\mathbb{R}^d)}, \quad f \in W^{k,1}(\mathbb{R}^d). \tag{2.4.10}$$

We now come to the previously announced integral estimate.

Proposition 2.4.3. *Recall* $\mathbb{T}^d_{\mathbb{L}} = \mathbb{R}^d/(\mathbb{L}\mathbb{Z})^d$ *and set*

$$\hat{p}(l) := \int_{\mathbb{T}^d_{\mathbb{L}}} p(y) \exp\left(2\pi i l^\top y\right) dy, \quad l \in \mathbb{Z}^d. \tag{2.4.11}$$

If Assumptions (L) hold, then, for all $\varepsilon > 0$*, any* $k \in \mathbb{N}$*, and any* $f \in \mathscr{S}(\mathbb{R}^d)$*, we have*

$$\left| \int_{\mathbb{R}^d} f(x) \left[\exp\left(p\left(x/\varepsilon\right)\right) - \frac{1}{\mathbb{L}^d} \int_{\mathbb{T}^d_{\mathbb{L}}} \exp\left(p(y)\right) dy \right] dx \right| \leq \frac{\varepsilon^k}{(2\pi)^k} \exp\left(\sum_{l \in \mathbb{Z}^d} |\hat{p}(l)| \right) \|f\|_{W^{k,1}(\mathbb{R}^d)}. \tag{2.4.12}$$

Proof. By rescaling we may assume without loss of generality that $\mathbb{L} = 1$, so that $\mathbb{T}^d_{\mathbb{L}} = \mathbb{T}^d$. We will do an approximation argument. To this end, we first consider the case of a general trigonometric polynomial on the torus given by

$$T_N(x) := \sum_{\substack{l \in \mathbb{Z}^d \\ |l| \leq N}} a(l) \exp(2\pi i l^\top x), \quad x \in \mathbb{T}^d, \quad N \in \mathbb{N},$$

with coefficients $a(l) \in \mathbb{C}$. We will occasionally suppress writing $l \in \mathbb{Z}^d$ and likewise notation in the appearing sums whenever it does not create confusion. Given the set of coefficient indices $A := \{l \in \mathbb{Z}^d \,|\, |l| \leq N\}$, there exists a bijective map

$$I \colon A \to \{1, \dots, M\}; \quad l \mapsto I(l),$$

where $M := |A|$ is the cardinality of A. Using a series expansion and the multinomial theorem, we can write

$$\begin{aligned}
\int_{\mathbb{T}^d} \exp\left(T_N(y)\right) dy &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{T}^d} \left(T_N(y)\right)^n dy \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{T}^d} \left(\sum_{m=1}^{M} a(I^{-1}(m)) \exp(2\pi i I^{-1}(m)^\top y) \right)^n dy \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\substack{\alpha \in \mathbb{N}_0^M \\ |\alpha| = n}} \binom{n}{\alpha} \int_{\mathbb{T}^d} \prod_{m=1}^{M} \left[a(I^{-1}(m)) \exp(2\pi i I^{-1}(m)^\top y) \right]^{\alpha_m} dy \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{|\alpha| = n} \binom{n}{\alpha} \prod_{m=1}^{M} a(I^{-1}(m))^{\alpha_m} \int_{\mathbb{T}^d} \exp\left(2\pi i y^\top \sum_{m=1}^{M} I^{-1}(m) \alpha_m \right) dy.
\end{aligned}$$

We set $S^M_\alpha := \sum_{m=1}^{M} I^{-1}(m)\alpha_m$ and note that $S^M_\alpha \in \mathbb{Z}^d$. By virtue of the periodicity of the complex exponential on $\mathbb{T}^d$,

$$\int_{\mathbb{T}^d} \exp\left(2\pi i y^\top S^M_\alpha\right) dy = \mathbb{1}_{\{S^M_\alpha = 0\}}$$

holds, so that the above calculation simplifies to

$$\int_{\mathbb{T}^d} \exp\left(T_N(y)\right) dy = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\substack{|\alpha| = n \\ S^M_\alpha = 0}} \binom{n}{\alpha} \prod_{m=1}^{M} a(I^{-1}(m))^{\alpha_m}. \tag{2.4.13}$$

For any $\varepsilon > 0$ and $x \in \mathbb{R}^d$, we obtain with similar steps as before

$$\exp\left(T_N(x/\varepsilon)\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\substack{\alpha \in \mathbb{N}_0^M \\ |\alpha|=n}} \binom{n}{\alpha} \prod_{m=1}^{M} a(I^{-1}(m))^{\alpha_m} \exp\left(2\pi i x^\top (S_\alpha^M/\varepsilon)\right).$$

Observe that by (2.4.13)

$$\begin{aligned}
\exp\left(T_N(x/\varepsilon)\right) &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\substack{|\alpha|=n \\ S_\alpha^M=0}} \binom{n}{\alpha} \prod_{m=1}^{M} a(I^{-1}(m))^{\alpha_m} \\
&\quad + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\substack{|\alpha|=n \\ S_\alpha^M\neq 0}} \binom{n}{\alpha} \prod_{m=1}^{M} a(I^{-1}(m))^{\alpha_m} \exp\left(2\pi i x^\top (S_\alpha^M/\varepsilon)\right) \\
&= \int_{\mathbb{T}^d} \exp\left(T_N(y)\right) dy \\
&\quad + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\substack{|\alpha|=n \\ S_\alpha^M\neq 0}} \binom{n}{\alpha} \prod_{m=1}^{M} a(I^{-1}(m))^{\alpha_m} \exp\left(2\pi i x^\top (S_\alpha^M/\varepsilon)\right).
\end{aligned}$$

For the next step, we note that for any multi-index $\alpha \in \mathbb{N}_0^M$ with $|\alpha| = n$ and $S_\alpha^M \neq 0$, there exist indices $j_1, \ldots, j_m \in \{1, \ldots, d\}$, $m \leq d$, such that $(S_\alpha^M)_{j_1}, \ldots, (S_\alpha^M)_{j_m} \neq 0$. Hence, using formula (2.4.8), we obtain the following estimate, for any multi-index $\beta \in \mathbb{N}_0^d$ with $|\beta| = \sum_{j=1}^d \beta_j = \sum_{i=1}^m \beta_{j_i} = k$ and $\beta_{j_i} > 0$,

$$\left|\int_{\mathbb{R}^d} f(x) \exp\left(2\pi i x^\top (S_\alpha^M/\varepsilon)\right) dx\right| \leq \frac{\varepsilon^k \|f\|_{W^{k,1}(\mathbb{R}^d)}}{(2\pi)^k |(S_\alpha^M)^\beta|} \leq \frac{\varepsilon^k \|f\|_{W^{k,1}(\mathbb{R}^d)}}{(2\pi)^k}, \tag{2.4.14}$$

where we also used the fact that $|(S_\alpha^M)^\beta| \geq 1$. Using this last estimate (2.4.14) and the preceding calculations, we get

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} f(x) \left[\exp\left(T_N\left(x/\varepsilon\right)\right) - \int_{\mathbb{T}^d} \exp\left(T_N(y)\right) dy\right] dx\right| \\
&\leq \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\substack{|\alpha|=n \\ S_\alpha^M\neq 0}} \binom{n}{\alpha} \prod_{m=1}^{M} \left|a(I^{-1}(m))\right|^{\alpha_m} \left|\int_{\mathbb{R}^d} f(x) \exp\left(2\pi i x^\top (S_\alpha^M/\varepsilon)\right) dx\right| \\
&\leq \frac{\varepsilon^k \|f\|_{W^{k,1}(\mathbb{R}^d)}}{(2\pi)^k} \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\substack{|\alpha|=n \\ S_\alpha^M\neq 0}} \binom{n}{\alpha} \prod_{m=1}^{M} \left|a(I^{-1}(m))\right|^{\alpha_m}.
\end{aligned}$$

By another application of the multinomial theorem, we have

$$\sum_{\substack{|\alpha|=n \\ S_\alpha^M\neq 0}} \binom{n}{\alpha} \prod_{m=1}^{M} \left|a(I^{-1}(m))\right|^{\alpha_m} \leq \sum_{|\alpha|=n} \binom{n}{\alpha} \prod_{m=1}^{M} \left|a(I^{-1}(m))\right|^{\alpha_m}$$

$$= \left(\sum_{m=1}^{M} \left| a(I^{-1}(m)) \right| \right)^n = \left(\sum_{|l| \le N} |a(l)| \right)^n .$$

Hence, we obtain the following estimate for a general trigonometric polynomial

$$\left| \int_{\mathbb{R}^d} f(x) \left[\exp\left(T_N\left(x/\varepsilon\right)\right) - \int_{\mathbb{T}^d} \exp\left(T_N(y)\right) dy \right] dx \right| \le \frac{\varepsilon^k}{(2\pi)^k} \exp\left(\sum_{|l| \le N} |a(l)| \right) \|f\|_{W^{k,1}(\mathbb{R}^d)}. \tag{2.4.15}$$

Note that, by $p \in C^\infty(\mathbb{T}^d)$[1], the Fourier coefficients of p defined in (2.4.11) satisfy $(\hat{p}(l))_{l \in \mathbb{Z}^d} \in \ell^1(\mathbb{Z}^d)$, see [Gra14, Theorem 3.3.9]. Next, recall that the Cesàro means defined by

$$\sigma_N(x) = \sum_{\substack{l \in \mathbb{Z}^d \\ |l_j| \le N}} \left(1 - \frac{|l_1|}{N+1}\right) \cdots \left(1 - \frac{|l_d|}{N+1}\right) \hat{p}(l) \exp(2\pi i l^\top x), \quad x \in \mathbb{T}^d, \quad N \in \mathbb{N},$$

are trigonometric polynomials, whose coefficients are evidently bounded by the absolute value of the Fourier coefficients of p. So, using (2.4.15), we have the following uniform bound in N

$$\left| \int_{\mathbb{R}^d} f(x) \left[\exp\left(\sigma_N\left(x/\varepsilon\right)\right) - \int_{\mathbb{T}^d} \exp\left(\sigma_N(y)\right) dy \right] dx \right| \le \frac{\varepsilon^k}{(2\pi)^k} \exp\left(\sum_{l \in \mathbb{Z}^d} |\hat{p}(l)| \right) \|f\|_{W^{k,1}(\mathbb{R}^d)}.$$

Using this uniform bound, the mean value theorem, and the triangle inequality, we eventually arrive at

$$\begin{aligned} \left| \int_{\mathbb{R}^d} f(x) \left[\exp\left(p\left(x/\varepsilon\right)\right) - \int_{\mathbb{T}^d} \exp\left(p(y)\right) dy \right] dx \right| &\le \frac{\varepsilon^k}{(2\pi)^k} \exp\left(\sum_{l \in \mathbb{Z}^d} |\hat{p}(l)| \right) \|f\|_{W^{k,1}(\mathbb{R}^d)} \\ &+ 2 \exp\left(\|p\|_{C(\mathbb{T}^d)} + \sum_{l \in \mathbb{Z}^d} |\hat{p}(l)| \right) \|f\|_{L^1(\mathbb{R}^d)} \|p - \sigma_N\|_{C(\mathbb{T}^d)}. \end{aligned}$$

It is a well-known fact, see [Gra14, Chapter 3], that the Cesàro means converge to p in $C(\mathbb{T}^d)$ as $N \to \infty$, so that we finally arrive at the concluding estimate (2.4.12). □

The next corollary gives weak convergence rates for the invariant densities of our considered Langevin equations. The fact that they are indeed the invariant densities has been proved in [PS07, Proposition 5.1 and 5.2].

Corollary 2.4.4. *Define the d-dimensional Lebesgue-densities*

$$\mu_\varepsilon(x) := \frac{1}{Z_\varepsilon} \exp\left(-\frac{\alpha}{\sigma} V(x) - \frac{p\left(x/\varepsilon\right)}{\sigma} \right), \qquad x \in \mathbb{R}^d, \quad \varepsilon > 0, \tag{2.4.16}$$

$$\mu(x) := \frac{1}{Z} \exp\left(-\frac{\alpha}{\sigma} V(x) \right), \qquad x \in \mathbb{R}^d, \tag{2.4.17}$$

[1] the Hölder space $C^{d,\delta}(\mathbb{T}^d)$ with $\delta > 0$ would be enough

where $Z, Z_\varepsilon > 0$ are normalization constants. Set two random variables $\zeta_\varepsilon \sim \mu_\varepsilon \, dx$, $\zeta \sim \mu \, dx$. If Assumptions (L) hold, then, for any function $g \in C^\infty(\mathbb{R}^d)$ with all of its derivatives growing at most polynomially, for sufficiently small $\varepsilon > 0$, and any $k \in \mathbb{N}$, we have

$$|\mathbb{E}\, g(\zeta_\varepsilon) - \mathbb{E}\, g(\zeta)| = \mathcal{O}\left(\varepsilon^k\right). \tag{2.4.18}$$

Proof. Let $\mathbb{L} = \alpha = \sigma = 1$ without loss of generality. Recall the following normalization constant from (1.2.14)

$$Z_p = \int_{\mathbb{T}^d} \exp(-p(y))\, dy. \tag{2.4.19}$$

Notice that

$$\begin{aligned}
\mathbb{E}\, g(\zeta_\varepsilon) - \mathbb{E}\, g(\zeta) &= \int_{\mathbb{R}^d} g(x) \left[\mu_\varepsilon(x) - \mu(x)\right] dx \\
&= \int_{\mathbb{R}^d} g(x) \left[\mu_\varepsilon(x) - \frac{Z_p}{Z_\varepsilon} \exp(-V(x))\right] dx \\
&\quad + \int_{\mathbb{R}^d} g(x) \left[\frac{Z_p}{Z_\varepsilon} \exp(-V(x)) - \mu(x)\right] dx \\
&= \frac{1}{Z_\varepsilon} \int_{\mathbb{R}^d} g(x) \exp(-V(x)) \left[\exp(-p(x/\varepsilon)) - Z_p\right] dx \\
&\quad + \frac{Z Z_p - Z_\varepsilon}{Z Z_\varepsilon} \int_{\mathbb{R}^d} g(x) \exp(-V(x))\, dx.
\end{aligned}$$

Observe that

$$Z_\varepsilon - Z Z_p = \int_{\mathbb{R}^d} \exp(-V(x)) \left[\exp(-p(x/\varepsilon)) - \int_{\mathbb{T}^d} \exp(-p(y))\, dy\right] dx. \tag{2.4.20}$$

Lemma 2.4.2 and Proposition 2.4.3 yield

$$|Z_\varepsilon - Z Z_p| \le \frac{\varepsilon^k}{(2\pi)^k} \exp\left(\sum_{l \in \mathbb{Z}^d} |\hat{p}(l)|\right) \|\exp(-V)\|_{W^{k,1}(\mathbb{R}^d)}. \tag{2.4.21}$$

In particular, $Z_\varepsilon \ge (Z Z_p)/2 > 0$ for sufficiently small $\varepsilon > 0$. Furthermore, using the estimate (2.4.6) and the assumption that all derivatives of g grow at most polynomially, it follows that $x \mapsto g(x) \exp(-V(x)) \in \mathscr{S}(\mathbb{R}^d)$. Therefore, Proposition 2.4.3 eventually implies

$$\begin{aligned}
|\mathbb{E}\, g(\zeta_\varepsilon) - \mathbb{E}\, g(\zeta)| &\le \frac{2\varepsilon^k}{Z Z_p (2\pi)^k} \exp\left(\sum_{l \in \mathbb{Z}^d} |\hat{p}(l)|\right) \|g \exp(-V)\|_{W^{k,1}(\mathbb{R}^d)} \\
&\quad + \frac{2\varepsilon^k}{Z^2 Z_p (2\pi)^k} \exp\left(\sum_{l \in \mathbb{Z}^d} |\hat{p}(l)|\right) \|\exp(-V)\|_{W^{k,1}(\mathbb{R}^d)} \|g \exp(-V)\|_{L^1(\mathbb{R}^d)} \\
&= \mathcal{O}\left(\varepsilon^k\right),
\end{aligned} \tag{2.4.22}$$

for sufficiently small $\varepsilon > 0$. □

2.4.2 One-Dimensional Case

From now on, we assume $d = 1$. In this case, we can derive stronger results and the first one is a special case of Proposition 2.4.3 and Corollary 2.4.4 with explicit large-scale potential given by a quadratic potential and a fast-scale periodic oscillation given by a cosine function. The proof is essentially the same, but it might be more illuminating than the general case as we do not need multi-index notation. We will need this special case later in Chapter 3.

Lemma 2.4.5. *Assume the setting of Corollary 2.4.4 with $d = 1$, $\alpha, \sigma > 0$, $\mathbb{L} = 2\pi$, $V(x) = x^2/2$, and $p(x) = -\cos(x)$. Set $\phi_u(x) := \exp(iux)$ for $x, u \in \mathbb{R}$. Then*

$$|\mathbb{E}\,\phi_u(\zeta_\varepsilon) - \mathbb{E}\,\phi_u(\zeta)| = \mathcal{O}\left(\exp\left(-\frac{2\sigma}{\alpha\varepsilon^2}\right)\right). \tag{2.4.23}$$

Proof. In this proof, the appearing constant $C(\alpha, \sigma)$ may change from line to line but will, in any case, remain independent of ε. As before, we can write

$$\begin{aligned}\mathbb{E}\,\phi_u(\zeta_\varepsilon) - \mathbb{E}\,\phi_u(\zeta) = \frac{1}{Z_\varepsilon}\int_{\mathbb{R}} \exp(iux)\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\left[\exp\left(\frac{\cos(x/\varepsilon)}{\sigma}\right) - \Pi^-\right]dx \\ + \frac{Z\Pi^- - Z_\varepsilon}{ZZ_\varepsilon}\int_{\mathbb{R}^d}\exp(ix)\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\,dx,\end{aligned}$$

where Π^- was defined in (1.2.16). Observe that

$$Z_\varepsilon - Z\Pi^- = \int_{\mathbb{R}^d}\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\left[\exp\left(\frac{\cos(x/\varepsilon)}{\sigma}\right) - \frac{1}{2\pi}\int_0^{2\pi}\exp\left(\frac{\cos(y)}{\sigma}\right)\,dy\right]dx.$$

We will prove that

$$|Z_\varepsilon - Z\Pi^-| = \mathcal{O}\left(\exp\left(-\frac{2\sigma}{\alpha\varepsilon^2}\right)\right). \tag{2.4.24}$$

For this, we can use a series expansion and the binomial theorem to calculate the integral

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}\exp\left(\frac{\cos(y)}{\sigma}\right)dy &= \frac{1}{2\pi}\sum_{n=0}^{\infty}\frac{1}{\sigma^n n!}\int_0^{2\pi}\cos(y)^n dy \\ &= \frac{1}{2\pi}\sum_{n=0}^{\infty}\frac{1}{(2\sigma)^n n!}\int_0^{2\pi}\left(\exp(iy) + \exp(-iy)\right)^n dy \\ &= \frac{1}{2\pi}\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{(2\sigma)^n n!}\binom{n}{k}\int_0^{2\pi}\exp\left(i(n-2k)y\right)dy \\ &= \frac{1}{2\pi}\sum_{m=0}^{\infty}\sum_{k=0}^{2m}\frac{1}{(2\sigma)^{2m}}\frac{1}{k!(2m-k)!}\int_0^{2\pi}\exp\left(2i(m-k)y\right)dy \\ &= \sum_{m=0}^{\infty}\frac{1}{(2\sigma)^{2m}(m!)^2}.\end{aligned}$$

Similarly, for $\varepsilon > 0$ and $x \in \mathbb{R}$, we obtain

$$\begin{aligned}\exp\left(\frac{\cos(x/\varepsilon)}{\sigma}\right) &= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{(2\sigma)^n n!}\binom{n}{k}\exp\left(i(n-2k)x/\varepsilon\right)\\ &= \sum_{m=0}^{\infty}\frac{1}{(2\sigma)^{2m}(m!)^2} + \sum_{m=0}^{\infty}\sum_{\substack{k=0\\k\neq m}}^{2m}\frac{\exp\left(2i(m-k)x/\varepsilon\right)}{(2\sigma)^{2m}k!(2m-k)!}\\ &\quad+\sum_{m=0}^{\infty}\sum_{k=0}^{2m+1}\frac{\exp\left(i(2(m-k)+1)x/\varepsilon\right)}{(2\sigma)^{2m+1}k!(2m-k+1)!}.\end{aligned}$$

We will estimate the last two terms. For the middle term in the last appearing equation, we obtain, by Fubini's theorem, the fact that $\mathbb{E}\exp(itZ) = \exp(-t^2/2)$ when $Z \sim \mathcal{N}_1(0,1)$, $|m-k| \geq 1$, and the binomial theorem,

$$\begin{aligned}&\left|\int_{\mathbb{R}}\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\sum_{m=0}^{\infty}\sum_{\substack{k=0\\k\neq m}}^{2m}\frac{\exp\left(2i(m-k)x/\varepsilon\right)}{(2\sigma)^{2m}k!(2m-k)!}\,dx\right|\\ &\leq \sum_{m=0}^{\infty}\sum_{\substack{k=0\\k\neq m}}^{2m}\frac{1}{(2\sigma)^{2m}k!(2m-k)!}\left|\int_{\mathbb{R}}\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\exp\left(2i(m-k)x/\varepsilon\right)\,dx\right|\\ &\leq C(\alpha,\sigma)\sum_{m=0}^{\infty}\sum_{\substack{k=0\\k\neq m}}^{2m}\frac{1}{(2\sigma)^{2m}k!(2m-k)!}\exp\left(-\frac{2(m-k)^2\sigma}{\alpha\varepsilon^2}\right)\\ &\leq C(\alpha,\sigma)\exp\left(-\frac{2\sigma}{\alpha\varepsilon^2}\right)\sum_{m=0}^{\infty}\frac{1}{\sigma^{2m}(2m)!}.\end{aligned} \tag{2.4.25}$$

An analog estimate holds for the third term, namely

$$\begin{aligned}&\left|\int_{\mathbb{R}}\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\sum_{m=0}^{\infty}\sum_{k=0}^{2m+1}\frac{\exp\left(i(2(m-k)+1)x/\varepsilon\right)}{(2\sigma)^{2m+1}k!(2m-k+1)!}\,dx\right|\\ &\leq C(\alpha,\sigma)\exp\left(-\frac{2\sigma}{\alpha\varepsilon^2}\right)\sum_{m=0}^{\infty}\frac{1}{\sigma^{2m+1}(2m+1)!}.\end{aligned}$$

With these equations and estimates, we can therefore arrive at (2.4.24). Note also that we can get a similar estimate for the term

$$\begin{aligned}&\int_{\mathbb{R}}\exp(iux)\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\left[\exp\left(\frac{\cos(x/\varepsilon)}{\sigma}\right)-\Pi^-\right]dx\\ &=\int_{\mathbb{R}}\exp(iux)\exp\left(-\frac{\alpha}{2\sigma}x^2\right)\left[\exp\left(\frac{\cos(x/\varepsilon)}{\sigma}\right)-\frac{1}{2\pi}\int_0^{2\pi}\exp\left(\frac{\cos(y)}{\sigma}\right)\,dy\right]dx.\end{aligned}$$

This is because we only have to replace $\exp(-\alpha x^2/2\sigma)$ with $\exp(iux)\exp(-\alpha x^2/2\sigma)$ in (2.4.25), so that we end up with the characteristic function of an unnormalized Gaussian again. □

Our last objective in this section is to prove that the Langevin equations (1.2.10) and (1.2.12) satisfy the assumptions of the preceding sections on limit theorems so that every main result is directly applicable.

Lemma 2.4.6. *Under Assumptions (L), the set of Assumptions (C*), (MET), (R), and (QET) are all satisfied for the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12)*.*

Proof. The conditions on the diffusion coefficients $\Sigma^2 = 2\sigma$ and $\bar{\Sigma}^2 = 2\sigma\mathcal{K}$ in all stated assumptions are trivial as they are merely constants. The local Lipschitz-continuity in Assumptions (C*) i) and ii) is immediate from the mean value theorem. In particular, for $x, y \in [-N, N]$ for some $N \in \mathbb{N}$, the periodicity of p, hence boundedness, implies

$$\left|\alpha V'(x) + \frac{1}{\varepsilon}p'(x) - \alpha V'(y) - \frac{1}{\varepsilon}p'(y)\right| \leq \alpha \max_{z\in[-N,N]} |V''(z)||x-y| + \frac{1}{\varepsilon^2}\|p''\|_\infty |x-y|.$$

Moreover, the local integrability condition in (2.2.1) holds by continuity and the fact that the diffusion coefficient $\bar{\Sigma}^2$ is just a constant. Hence, Assumptions (C*) holds. Next, we will prove Assumptions (R). For this, note that, by (L) iii), there exist $a, b, R > 0$ such that for all $x \in \mathbb{R}$ with $|x| \geq R$,

$$-\operatorname{sgn}(x)V'(x) \leq \frac{a}{|x|} - b|x| \leq \frac{a}{R} - b|x|.$$

Thus, we can find $S, \beta > 0$ such that for $x \in \mathbb{R}$ with $|x| > S$,

$$-\operatorname{sgn}(x)\frac{\alpha\mathcal{K}V'(x)}{\bar{\Sigma}^2} \leq -\beta, \tag{2.4.26}$$

which is exactly Assumptions (R). We will now prove Assumptions (MET). In this case, the scale functions are given by

$$\begin{aligned} f_\varepsilon(x) &= \int_0^x \exp\left(\frac{\alpha}{\sigma}V(y) + \frac{1}{\sigma}p\left(\frac{y}{\varepsilon}\right)\right)\, dy, \\ f(x) &= \int_0^x \exp\left(\frac{\alpha}{\sigma}V(y)\right)\, dy. \end{aligned} \tag{2.4.27}$$

By periodicity of p, hence boundedness, we therefore obtain the pointwise lower and upper bounds in (MET) iii). After noting

$$C_\rho = \int_{\mathbb{R}} \rho(x)^2\, dx = \frac{1}{\bar{\Sigma}^2}\int_{\mathbb{R}} \frac{1}{f'(x)}\, dx,$$

Condition ii) is a consequence either of Lemma 2.4.2 or (2.4.26). Condition i) follows from 2.4.26. For $x > S$, we have

$$f(x) = \int_0^x \exp\left(\frac{2\alpha\mathcal{K}}{\bar{\Sigma}^2}\int_0^y V'(z)\, dz\right)\, dy \geq f(S) + f'(S)\int_S^x \exp\left(2\beta\right)\, dy,$$

which implies $\lim_{x\to\infty} f(x) = \infty$. Similarly for $\lim_{x\to-\infty} f(x) = -\infty$ but with different signs. Hence, Assumptions (MET) holds, too. At last, we come to Assumptions (QET).

Considering what we have proved so far, there is not much left to verify. Concerning the recurrence conditions in (QET), we mainly need to observe that (L) iii) yields

$$\limsup_{|x|\to\infty} -\operatorname{sgn}(x)V'(x) = -\infty, \tag{2.4.28}$$

yielding $(-V', 1) \in \mathscr{R}(r)$ for all $r > 0$. This readily implies that $(-\alpha V' - p'(\cdot/\varepsilon)/\varepsilon, \Sigma^2) \in \mathscr{R}$ and $(-\alpha\mathcal{K}V', \bar{\Sigma}^2) \in \mathscr{R}$. Assumption (QET) iii) follows by Example 2.1.13. Since p is bounded on $\mathbb{R}$ by $\mathbb{L}$-periodicity, we have $(-p'(\cdot/\varepsilon)/\varepsilon, \Sigma^2) \in \mathscr{I}(2\Sigma^{-2}\|p\|_\infty)$. Choosing $r > 2\Sigma^{-2}\|p\|_\infty$, which is possible due to $(-\alpha V', \Sigma^2) \in \cap_{r>0} R(r)$, indeed gives (QET) iii). Finally, (QET) iv) is quite straightforward because

$$\left|\int_0^x \frac{p'(y/\varepsilon)}{\Sigma^2\varepsilon}\, dy\right| = \frac{1}{\Sigma^2}\,|p(x/\varepsilon) - p(0)| \le \frac{2\|p\|_\infty}{\Sigma^2}, \quad x \in \mathbb{R}.$$

This concludes the proof. □

To properly apply Theorem 2.3.3 to the Langevin equations, one also needs to verify the convergence condition in (2.3.10). Hence, in the rest of this section, we will be primarily concerned with explicit convergence rates for the difference of the asymptotic variances

$$\left|\Sigma^2 \int_{\mathbb{R}} \Phi_\varepsilon'(x)^2 \mu_\varepsilon(x)\, dx - \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi'(x)^2 \mu(x)\, dx\right| \tag{2.4.29}$$

as $\varepsilon \to 0$, where $\Sigma = \sqrt{2\sigma}$, $\bar{\Sigma} = \sqrt{2\sigma\mathcal{K}}$, and Φ_ε, Φ solve the Poisson equations

$$-\mathcal{L}_\varepsilon\Phi_\varepsilon = \bar{h}_\varepsilon := h - \int_{\mathbb{R}} h(x)\mu_\varepsilon(x)\, dx, \tag{2.4.30}$$

$$-\mathcal{L}\Phi = \bar{h} := h - \int_{\mathbb{R}} h(x)\mu(x)\, dx, \tag{2.4.31}$$

respectively, with the generators $\mathcal{L}_\varepsilon$ and $\mathcal{L}$ given in (2.4.3) and (2.4.4). One could prove qualitative convergence in the sense of a limit result, but we will need explicit convergence rates in Chapter 4.

We start with two lemmas that will be the main key for obtaining estimates for (2.4.29). The estimates we are about to obtain are nowhere near optimal in any sense, but they are at least sufficient for our purposes. We will formulate and prove the results without loss of generality for $\mathbb{L} = \alpha = \sigma = 1$ as the handling of these constants is only a matter of rescaling or absorption through the potential V and the oscillation p. The general case works exactly the same and is merely an additional notational burden. The first lemma is a slight generalization of Proposition 2.4.3 to finite intervals.

Lemma 2.4.7. *Let Assumptions (L) hold and consider $a, b \in [-\infty, \infty]$ with $a < b$. If $h \in C^\infty(\mathbb{R})$ such that $h\exp(-V) \in \mathscr{S}(\mathbb{R})$, then*

$$\begin{aligned}&\left|\int_a^b h(x)(\mu_\varepsilon(x) - \mu(x))\, dx\right| \\ &= \mathcal{O}\left(\varepsilon\left[|h(a)\exp(-V(a))| + |h(b)\exp(-V(b))| + \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\right]\right),\end{aligned} \tag{2.4.32}$$

with the understanding that $|h(-\infty)\exp(-V(-\infty))| = |h(\infty)\exp(-V(\infty))| = 0$. *Moreover, for any* $k \in \mathbb{N}$, *we have*

$$\left|\int_{\mathbb{R}} h(x)(\mu_\varepsilon(x) - \mu(x))\, dx\right| = \mathcal{O}\left(\varepsilon^k \|h \exp(-V)\|_{W^{k,1}(\mathbb{R})}\right). \tag{2.4.33}$$

Proof. Assume without loss of generality that $\mathbb{L} = \alpha = \sigma = 1$. The proof is very similar to the proofs of Proposition 2.4.3 and Corollary 2.4.4, which is why we will only give a sketch here. First, note that

$$\begin{aligned}\int_a^b h(x)(\mu_\varepsilon(x) - \mu(x))\, dx = \frac{1}{Z_\varepsilon}\int_a^b h(x)\exp(-V(x))\left[\exp(-p(x/\varepsilon)) - \Pi^-\right] dx \\ + \frac{Z\Pi^- - Z_\varepsilon}{ZZ_\varepsilon}\int_a^b h(x)\exp(-V(x))\, dx,\end{aligned} \tag{2.4.34}$$

and

$$Z_\varepsilon - Z\Pi^- = \int_{\mathbb{R}} \exp(-V(x))\left[\exp(-p(x/\varepsilon) - \Pi^-\right] dx, \quad \Pi^- = \int_0^1 \exp(-p(y))\, dy. \tag{2.4.35}$$

As we can see from these two expressions, this puts us into the setting of Proposition 2.4.3, with the only difference being that we might have a finite interval (a, b) in some integrals. Thoroughly inspecting the proof of that proposition reveals that the relevant estimate appears in (2.4.14), which, in the case of $a = -\infty$, $b = \infty$, turns into

$$\left|\int_{\mathbb{R}} h(x)\exp(-V(x))\exp\left(2\pi i x(S_\alpha^M/\varepsilon)\right) dx\right| \leq \varepsilon^k \|h\exp(-V)\|_{W^{k,1}(\mathbb{R})},$$

with a number $S_\alpha^M \in \mathbb{Z}\backslash\{0\}$ determined in the proof of that proposition. In the case of $a, b \in [-\infty, \infty]$ with at least one of these two numbers being finite, we can use integration by parts once to obtain

$$\begin{aligned}\left|\int_a^b h(x)\exp(-V(x))\exp\left(2\pi i x(S_\alpha^M/\varepsilon)\right) dx\right| \leq \varepsilon\Big[|h(a)\exp(-V(a))| + |h(b)\exp(-V(b))| \\ + \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\Big],\end{aligned}$$

Continuing verbatim with the proof of that proposition, we obtain the estimates

$$\begin{aligned}&\left|\int_{\mathbb{R}} h(x)\exp(-V(x))\left[\exp(-p(x/\varepsilon)) - \Pi^-\right] dx\right| = \mathcal{O}\left(\varepsilon^k \|h\exp(-V)\|_{W^{k,1}(\mathbb{R})}\right), \\ &\left|\int_a^b h(x)\exp(-V(x))\left[\exp(-p(x/\varepsilon)) - \Pi^-\right] dx\right| \\ &= \mathcal{O}\Big(\varepsilon\Big[|h(a)\exp(-V(a))| + |h(b)\exp(-V(b))| \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\Big]\Big), \quad a, b \in [-\infty, \infty].\end{aligned}$$

By using these in (2.4.34) and (2.4.35), we arrive at the conclusion. □

The second lemma shows, in particular, that Φ_ε converges uniformly on compact sets of $\mathbb{R}$ to Φ as $\varepsilon \to 0$, proving a type of homogenization result for the solution to the Poisson equation.

Lemma 2.4.8. *Let Assumptions (L) hold and consider $x \in \mathbb{R}$. Then, for any bounded function $h \in C^\infty(\mathbb{R})$ with $h\exp(-V) \in \mathscr{S}(\mathbb{R})$, we have*

$$|\Phi_\varepsilon(x) - \Phi(x)| = \mathcal{O}\left(\varepsilon \operatorname{sgn}(x)\left[x + \int_0^x |V'(y)|\,dy + \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\int_0^x \exp(V(y))\,dy\right]\right), \tag{2.4.36}$$

as $\varepsilon \to 0$, where Φ_ε and Φ solve the Poisson equations

$$-\mathcal{L}_\varepsilon\Phi_\varepsilon = \bar{h}_\varepsilon = h - \int_{\mathbb{R}} h(x)\mu_\varepsilon(x)\,dx, \tag{2.4.37}$$

$$-\mathcal{L}\Phi = \bar{h} = h - \int_{\mathbb{R}} h(x)\mu(x)\,dx. \tag{2.4.38}$$

Proof. Let $x > 0$. The case $x < 0$ works similarly due to symmetry in our assumptions, especially due to (L) iii). Assume without loss of generality that $\mathbb{L} = \alpha = \sigma = 1$. First, recall the explicit solution formulas for Φ_ε and Φ given in (2.3.7) and (2.3.9) by

$$\Phi_\varepsilon(x) = -\int_0^x \frac{1}{\mu_\varepsilon(y)}\int_{-\infty}^y \bar{h}_\varepsilon(z)\mu_\varepsilon(z)\,dz\,dy, \quad \Phi(x) = -\int_0^x \frac{1}{\mathcal{K}\mu(y)}\int_{-\infty}^y \bar{h}(z)\mu(z)\,dz\,dy.$$

Since p is bounded, we have the simple inequality

$$c_\mu\mu \le \mu_\varepsilon \le C_\mu\mu \quad \text{on } \mathbb{R}, \tag{2.4.39}$$

with constants $c_\mu, C_\mu > 0$ that are independent of ε. We define

$$H_\varepsilon(y) := \int_{-\infty}^y \bar{h}_\varepsilon(z)\mu_\varepsilon(z)\,dz, \quad H(y) := \int_{-\infty}^y \bar{h}(z)\mu(z)\,dz, \quad y \in \mathbb{R}. \tag{2.4.40}$$

Then, for $y \in \mathbb{R}$, Lemma 2.4.7 yields

$$\begin{aligned}
|H_\varepsilon(y) - H(y)| &= \left|\int_{-\infty}^y (\bar{h}_\varepsilon(z) - \bar{h}(z))\mu_\varepsilon(z)\,dz + \int_{-\infty}^y \bar{h}(z)(\mu_\varepsilon(z) - \mu(z))\,dz\right| \\
&\le \left\|\bar{h}_\varepsilon - \bar{h}\right\|_\infty + \left|\int_{-\infty}^y \bar{h}(z)(\mu_\varepsilon(z) - \mu(z))\,dz\right| \\
&= \left|\int_{\mathbb{R}} h(z)(\mu_\varepsilon(z) - \mu(z))\,dz\right| + \left|\int_{-\infty}^y \bar{h}(z)(\mu_\varepsilon(z) - \mu(z))\,dz\right| \\
&= \mathcal{O}\left(\varepsilon^k\|h\exp(-V)\|_{W^{k,1}(\mathbb{R})}\right) \\
&\quad + \mathcal{O}\left(\varepsilon\left[|\bar{h}(y)\exp(-V(y))| + \|\bar{h}\exp(-V)\|_{W^{1,1}(\mathbb{R}}\right]\right) \\
&= \mathcal{O}\left(\varepsilon\left[|\bar{h}(y)|\exp(-V(y)) + \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\right]\right).
\end{aligned}$$

It follows

$$\begin{aligned}
|\Phi_\varepsilon(x) - \Phi(x)| &= \left|\int_0^x \frac{H_\varepsilon(y)}{\mu_\varepsilon(y)} - \frac{H(y)}{\mathcal{K}\mu(y)}\,dy\right| \\
&= \left|\int_0^x \frac{H_\varepsilon(y) - H(y)}{\mu_\varepsilon(y)}\,dy + \int_0^x H(y)\left(\frac{1}{\mu_\varepsilon(y)} - \frac{1}{\mathcal{K}\mu(y)}\right)dy\right|
\end{aligned}$$

$$\leq \frac{1}{c_\mu}\int_0^x \frac{|H_\varepsilon(y) - H(y)|}{\mu(y)}\,dy + \left|\int_0^x H(y)\left(\frac{1}{\mu_\varepsilon(y)} - \frac{1}{\mathcal{K}\mu(y)}\right)\,dy\right|.$$

For the first term, using the estimates above and the boundedness of h, we get

$$\begin{aligned}\frac{1}{c_\mu}\int_0^x \frac{|H_\varepsilon(y) - H(y)|}{\mu(y)}\,dy &= \mathcal{O}\left(\varepsilon\left[\int_0^x |\bar{h}(y)| + \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\exp(V(y))\,dy\right]\right)\\ &= \mathcal{O}\left(\varepsilon\left[x + \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\int_0^x \exp(V(y))\,dy\right]\right),\end{aligned} \tag{2.4.41}$$

and, for the second term, we notice that

$$\begin{aligned}\int_0^x H(y)\left(\frac{1}{\mu_\varepsilon(y)} - \frac{1}{\mathcal{K}\mu(y)}\right)\,dy &= Z_\varepsilon\int_0^x H(y)\exp\left(V(y)\right)\left(\exp(p(y/\varepsilon)) - \Pi^+\right)\,dy\\ &\quad + (Z_\varepsilon - Z\Pi^-)\Pi^+\int_0^x H(y)\exp\left(V(y)\right)\,dy,\end{aligned} \tag{2.4.42}$$

with Π^+ as in (1.2.16). With the same ideas as in the proof of Lemma 2.4.7 for the finite interval case, we can show that

$$\begin{aligned}&\left|\int_0^x H(y)\exp\left(V(y)\right)\left(\exp(p(y/\varepsilon)) - \Pi^+\right)\,dy\right|\\ &= \mathcal{O}\left(\varepsilon\left[|H(x)\exp(V(x))| + \int_0^x |(H\exp(V))'|(y)\,dy\right]\right).\end{aligned}$$

Obviously, $H\exp(V)$ is bounded on $[0,S]$ where $S>0$ comes from (2.4.26). For $x > S$, we can use (2.4.26) and the fact that $\lim_{y\to\infty} H(y) = 0$ to estimate

$$\begin{aligned}|H(x)\exp(V(x))| &= \frac{1}{Z}\left|\int_x^\infty \bar{h}(y)\exp(-V(y)+V(x))\,dy\right|\\ &= \frac{1}{Z}\left|\int_x^\infty \bar{h}(y)\exp\left(-\int_x^y V'(z)\,dz\right)\,dy\right|\\ &\leq \frac{\|\bar{h}\|_\infty}{Z}\int_x^\infty \exp(-2\beta(y-x))\,dy\\ &= \frac{2\|h\|_\infty}{Z}\int_0^\infty \exp(-2\beta y)\,dy = \frac{\|h\|_\infty}{\beta Z},\end{aligned}$$

that is, $H\exp(V)$ is bounded on $\mathbb{R}$. By using this fact, $H'\exp(V) = \bar{h}/Z$, and the boundedness of h on $\mathbb{R}$, we obtain

$$\left|\int_0^x H(y)\exp\left(V(y)\right)\left(\exp(p(y/\varepsilon)) - \Pi^+\right)\,dy\right| = \mathcal{O}\left(\varepsilon\left[x + \int_0^x |V'(y)|\,dy\right]\right),$$

providing us, in conjunction with Lemma 2.4.7 (recall (2.4.35)), with the following result for the second term (2.4.42)

$$\left|\int_0^x H(y)\left(\frac{1}{\mu_\varepsilon(y)} - \frac{1}{\mathcal{K}\mu(y)}\right)\,dy\right| = \mathcal{O}\left(\varepsilon\left[x + \int_0^x |V'(y)|\,dy\right]\right). \tag{2.4.43}$$

Combining (2.4.41) and (2.4.43) yields the claim. □

We come now to convergence rates for the difference of the asymptotic variances.

Proposition 2.4.9. *Let Assumptions (L) hold. Then, for any bounded function* $h \in C^\infty(\mathbb{R})$ *with* $h\exp(-V) \in \mathscr{S}(\mathbb{R})$ *and any* $\gamma \in (0,1)$*, we have*

$$\left|\Sigma^2 \int_{\mathbb{R}} \Phi_\varepsilon'(x)^2 \mu_\varepsilon(x)\,dx - \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi'(x)^2 \mu(x)\,dx\right| = \mathcal{O}\left(\varepsilon^\gamma \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\right), \tag{2.4.44}$$

as $\varepsilon \to 0$*, where* Φ_ε *and* Φ *solve the Poisson equations*

$$-\mathcal{L}_\varepsilon \Phi_\varepsilon = \bar{h}_\varepsilon = h - \int_{\mathbb{R}} h(x)\mu_\varepsilon(x)\,dx, \tag{2.4.45}$$

$$-\mathcal{L}\Phi = \bar{h} = h - \int_{\mathbb{R}} h(x)\mu(x)\,dx. \tag{2.4.46}$$

Proof. Recall that we assume $\mathbb{L} = \alpha = \sigma = 1$ without loss of generality. Integration by parts and the Poisson equation give

$$\begin{aligned}\mathcal{K}\int_{\mathbb{R}} \Phi'(x)^2\mu(x)\,dx &= \mathcal{K}\left[\Phi(x)\Phi'(x)\mu(x)\right]_{-\infty}^{\infty} - \mathcal{K}\int_{\mathbb{R}} \Phi(x)\left[-V'(x)\Phi'(x) + \Phi''(x)\right]\mu(x)\,dx \\ &= \int_{\mathbb{R}} \Phi(x)(-\mathcal{L}\Phi)(x)\mu(x)\,dx = \int_{\mathbb{R}} \Phi(x)\bar{h}(x)\mu(x)\,dx,\end{aligned}$$

where $\Phi\Phi'\mu$ vanishes at infinity due to Remark 2.1.15 and Lemma 2.4.2. A completely analog calculation shows

$$\int_{\mathbb{R}} \Phi_\varepsilon'(x)^2\mu_\varepsilon(x)\,dx = \int_{\mathbb{R}} \Phi_\varepsilon(x)\bar{h}_\varepsilon(x)\mu_\varepsilon(x)\,dx.$$

We can do the following splitting

$$\begin{aligned}\int_{\mathbb{R}} \Phi_\varepsilon(x)\bar{h}_\varepsilon(x)\mu_\varepsilon(x)\,dx - \int_{\mathbb{R}} \Phi(x)\bar{h}(x)\mu(x)\,dx &= \int_{\mathbb{R}} \Phi_\varepsilon(x)(\bar{h}_\varepsilon(x) - \bar{h}(x))\mu_\varepsilon(x)\,dx \\ &\quad + \int_{\mathbb{R}} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx \\ &\quad + \int_{\mathbb{R}} \Phi(x)\bar{h}(x)(\mu_\varepsilon(x) - \mu(x))\,dx \\ &=: I_\varepsilon + J_\varepsilon + K_\varepsilon.\end{aligned}$$

By Remark 2.1.15 and $\Phi_\varepsilon(0) = \Phi(0) = 0$, we know that $|\Phi_\varepsilon(x)|, |\Phi(x)| \le C|x|$ for all $x \in \mathbb{R}$, with a constant $C > 0$ that does not depend on ε. With this fact, Lemma 2.4.7, and (2.4.39), the first and the last term can be estimated, for any $k \in \mathbb{N}$, as follows

$$I_\varepsilon \le \left\|\bar{h}_\varepsilon - \bar{h}\right\|_\infty \int_{\mathbb{R}} |x|\mu(x)\,dx = \mathcal{O}\left(\varepsilon^k \|h\exp(-V)\|_{W^{k,1}(\mathbb{R})}\right), \tag{2.4.47}$$

$$K_\varepsilon = \mathcal{O}\left(\varepsilon^k \|\Phi\bar{h}\exp(-V)\|_{W^{k,1}(\mathbb{R})}\right), \tag{2.4.48}$$

where in the first line we also used the fact that $\int_{\mathbb{R}} |x|\mu(x)\,dx < \infty$ by Lemma 2.4.2. For the middle term, we first introduce a sufficiently large $R_\varepsilon > S$, which we will choose depending on ε soon, such that (2.4.26) holds and

$$J_\varepsilon = \int_{|x|\le R_\varepsilon} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx + \int_{|x|>R_\varepsilon} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx.$$

Then, using the linear bounds on Φ_ε and Φ as mentioned above, (2.4.39), and the boundedness of h, yields

$$\int_{|x|>R_\varepsilon} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx = \mathcal{O}\left(\int_{|x|>R_\varepsilon} |x| \exp(-V(x))\,dx\right)$$

Note that, by (2.4.26), we have

$$\begin{aligned}
\int_{R_\varepsilon}^{\infty} x\exp(-V(x))\,dx &= \exp(-V(R_\varepsilon)) \int_{R_\varepsilon}^{\infty} x\exp\left(-\int_{R_\varepsilon}^{x} V'(y)\,dy\right)\,dx \\
&\leq \exp(-V(R_\varepsilon)) \int_0^{\infty} (x+R_\varepsilon)\exp\left(-2\beta x\right)\,dx \\
&\leq \exp(-2\beta R_\varepsilon)\exp(2\beta S - V(S)) \int_0^{\infty} (x+R_\varepsilon)\exp\left(-\beta x\right)\,dx \\
&= \mathcal{O}\left(R_\varepsilon \exp(-2\beta R_\varepsilon)\right).
\end{aligned}$$

We can get similar estimates for the integral on the negative axis due to the symmetric nature of (2.4.26). Therefore, we obtain

$$\int_{|x|>R_\varepsilon} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx = \mathcal{O}\left(R_\varepsilon \exp(-2\beta R_\varepsilon)\right). \tag{2.4.49}$$

For the first term in J_ε, we utilize Lemma 2.4.8, (2.4.39), and the boundedness of h to obtain

$$\begin{aligned}
&\int_{|x|\leq R_\varepsilon} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx \\
&= \mathcal{O}\left(\varepsilon \int_{|x|\leq R_\varepsilon} \operatorname{sgn}(x)\left[x + \int_0^x |V'(y)|\,dy\right]\exp(-V(x))\,dx\right) \\
&\quad + \mathcal{O}\left(\varepsilon \int_{|x|\leq R_\varepsilon} \operatorname{sgn}(x)\left[\|h\exp(-V)\|_{W^{1,1}(\mathbb{R})} \int_0^x \exp(V(y))\,dy\right]\exp(-V(x))\,dx\right).
\end{aligned}$$

By (L) ii), it follows that $\int_0^x |V'(y)|\,dy$ is at most polynomially bounded, so that, by Lemma 2.4.2, we may upper bound

$$\int_{|x|\leq R_\varepsilon} \operatorname{sgn}(x) \int_0^x |V'(y)|\,dy \exp(-V(x))\,dx \leq \int_{\mathbb{R}} \int_0^{|x|} |V'(y)|\,dy\exp(-V(x))\,dx < \infty.$$

Moreover, by L'Hôpital's rule and (2.4.26), we obtain

$$\limsup_{|x|\to\infty} \left(\operatorname{sgn}(x)\int_0^x \exp(V(y))\,dy\exp(-V(x))\right) = \limsup_{|x|\to\infty} \frac{1}{\operatorname{sgn}(x)V'(x)} \leq \frac{1}{2\beta}, \tag{2.4.50}$$

giving us at least the boundedness of that quotient on $\mathbb{R}$, so that we eventually arrive at

$$\int_{|x|\leq R_\varepsilon} (\Phi_\varepsilon(x) - \Phi(x))\bar{h}(x)\mu_\varepsilon(x)\,dx = \mathcal{O}\left(\varepsilon R_\varepsilon \|h\exp(-V)\|_{W^{1,1}(\mathbb{R})}\right). \tag{2.4.51}$$

Choosing $R_\varepsilon = \varepsilon^{\eta}$ for any $\eta \in (0,1)$, we get the final result by combining (2.4.47), (2.4.48), (2.4.49), (2.4.51), and taking the worst among these estimates, that is, (2.4.51). $\square$

Remark 2.4.10. *Closely inspecting the last proof reveals that an improvement of the convergence rate primarily relies on an improvement of the bound in* (2.4.51)*. To be more precise, if one were to find a function* $a\colon \mathbb{R} \to \mathbb{R}$ *giving the exact asymptotic growth of the quotient in* (2.4.50)*, namely*

$$\lim_{|x|\to\infty} \frac{a(x)\,\mathrm{sgn}(x) f(x)}{f'(x)} = \lim_{|x|\to\infty} \frac{a(x)\,\mathrm{sgn}(x) \int_0^x \exp(V(y))\,dy}{\exp(V(x))} = 1, \tag{2.4.52}$$

where f *is the scale function in* (2.4.27)*, then such an improvement is feasible, but it depends on the exact condition in (L) iii) or* (2.4.26)*. We settled for this explicit condition, but more stringent conditions can certainly yield better control.*

We end this chapter with a summary and restatement of the main limit theorems in the case of the Langevin equations so that we can reference it more easily in the following chapters.

Corollary 2.4.11. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L). Let* h *be a bounded function such that* $h \in C^\infty(\mathbb{R})$ *and* $h \exp(-V) \in \mathscr{S}(\mathbb{R})$*. Then, for* $T_\varepsilon \to \infty$ *as* $\varepsilon \to 0$*, we have*

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} h(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} h(x)\mu(x)\,dx\right|^2 \to 0, \quad \varepsilon \to 0. \tag{2.4.53}$$

Proof. This is an immediate consequence of Lemma 2.4.6, Theorem 2.1.10, and Lemma 2.4.7. □

Corollary 2.4.12. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L). Assume that* $V'' \in \mathscr{P}(2)$*, i.e., there exists a constant* $K > 0$ *such that for all* $x \in \mathbb{R}$*,*

$$|V''(x)| \le K(1 + |x|^2). \tag{2.4.54}$$

Let $m \in \mathbb{N}$ *be such that* $b > (2m-1)/(2m)^2$*, where* $b > 0$ *comes from (L) iii). Then, for any continuous function* $\varphi \in \mathscr{P}(m)$ *and* $T_\varepsilon = \varepsilon^{-\alpha}$ *with* $\alpha > m/3 \vee 1/2$*, we have*

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy\right|^2 = \begin{cases} \mathcal{O}\left((1 \vee C_{m,\varphi}^2)\varepsilon^{(6\alpha-2m)/5}\right), & \text{if } \alpha \le 2m - \frac{5}{2}, \\ \mathcal{O}\left((1 \vee C_{m,\varphi}^2)\varepsilon^{\alpha-1/2}\right), & \text{if } \alpha > 2m - \frac{5}{2}, \end{cases} \tag{2.4.55}$$

as $\varepsilon \to 0$*, where* $C_{m,\varphi} > 0$ *is the constant from the polynomial growth bound of* φ*. If* $\varphi \in \mathscr{P}(0)$*, then, for* $T_\varepsilon = \varepsilon^{-\alpha}$ *with* $\alpha > 1/2$*, we have*

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \varphi(y)\mu_\varepsilon(y)\,dy\right|^2 = \mathcal{O}\left(\frac{C_{0,\varphi}^2}{T_\varepsilon\sqrt{\varepsilon}}\right) = \mathcal{O}\left(C_{0,\varphi}^2\varepsilon^{\alpha-1/2}\right), \tag{2.4.56}$$

as $\varepsilon \to 0$*.*

Proof. This follows from Lemma 2.4.6 and Theorem 2.1.18 as soon as we verify the two remaining conditions in Theorem 2.1.18. Inequality (2.1.97) holds by boundedness of p'' and the assumption (2.4.54), since, for $\varepsilon > 0$ small enough,

$$V''(x) + \frac{p''(x/\varepsilon)}{\varepsilon^2} \le K|x|^2 + \frac{K + \|p''\|_\infty}{\varepsilon^2}.$$

The other assumption one has to verify is the Lyapunov condition (2.1.98). To this end, define $v(x) := x^{2m}$ for $x \in \mathbb{R}$. Using (L) iii), we can calculate and estimate

$$\mathcal{L}_\varepsilon v \leq -2mbx^{2m} + (2ma + m(2m-1)\Sigma^2)x^{2m-2} + \frac{2m\|p'\|_\infty}{\varepsilon} x^{2m-1}.$$

Young's inequality for products with $p = 2m/2m-1$ and $q = 2m$ yields

$$\frac{2m\|p'\|_\infty}{\varepsilon} x^{2m-1} \leq \frac{2m-1}{2m} x^{2m} + \frac{2m^{2m-1}\|p'\|_\infty^{2m}}{\varepsilon^{2m}}.$$

Moreover, for any $\eta > 0$, we have the basic fact that $x^{2m-2} \leq \eta x^{2m} + C(\eta, m)$ with $C(\eta, m) := \sup_{x\in\mathbb{R}}(x^{2m-2} - \eta x^{2m})$. With these two estimates, we can conclude

$$\mathcal{L}_\varepsilon v \leq -\lambda v(x) + \frac{2m^{2m-1}\|p'\|_\infty^{2m} + (2ma + m(2m-1)\Sigma^2)}{\varepsilon^{2m}},$$

where

$$\lambda := 2mb - \frac{2m-1}{2m} - \eta(2ma + m(2m-1)\Sigma^2), \quad c := 2m^{2m-1}\|p'\|_\infty^{2m} + (2ma + m(2m-1)\Sigma^2),$$

which are both positive by the assumption $b > (2m-1)/(2m)^2$ and after choosing η sufficiently small. □

Corollary 2.4.13. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L). Let $R_\varepsilon = \Omega(\varepsilon^{-\eta})$ for some $\eta > 0$ as $\varepsilon \to 0$. Then, for all $\delta > 0$, we have*

$$\lim_{\varepsilon\to 0} \mathbb{P}\left(\frac{\left|X_{T_\varepsilon}^\varepsilon\right|}{R_\varepsilon} > \delta\right) = 0. \tag{2.4.57}$$

Proof. This follows immediately from Lemma 2.4.6 and Corollary 2.2.5. □

Corollary 2.4.14. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L) and the Poisson equations*

$$-\mathcal{L}_\varepsilon \Phi_\varepsilon = \bar{h}_\varepsilon = h - \int_{\mathbb{R}} h(x)\mu_\varepsilon(x)\,dx, \tag{2.4.58}$$

$$-\mathcal{L}\Phi = \bar{h} = h - \int_{\mathbb{R}} h(x)\mu(x)\,dx, \tag{2.4.59}$$

with a bounded function $h \in C^\infty(\mathbb{R})$ such that $h\exp(-V) \in \mathscr{S}(\mathbb{R})$. Then, for $T_\varepsilon = \Omega(\varepsilon^{-\eta})$ with $\eta > 0$, we have

$$\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon)\,dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0, \tau^2), \quad \varepsilon \to 0, \tag{2.4.60}$$

where

$$\tau^2 = \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi'(x)^2 \mu(x)\,dx. \tag{2.4.61}$$

Proof. This readily follows from Lemma 2.4.6, Theorem 2.3.3, and Proposition 2.4.9. □

Remark 2.4.15. *We want to repeat that it is not necessary to assume so much regularity on the coefficient functions of the generators and test functions in the limit theorems. For example, in Corollary 2.4.11 and Corollary 2.4.14, it would, in fact, suffice to assume* $V, p \in C^2(\mathbb{R})$ *and a bounded function* $h \in C^1(\mathbb{R})$ *such that* $h \exp(-V)$ *and its derivative* $(h \exp(-V))'$ *have sufficient decrease at infinity.*

Overall, this section shows us that, although we proved general limit theorems in the previous sections, there still remains some work even when considering a specific class of multiscale equations with their homogenized limit equation. This added work is mostly due to the condition on the asymptotic variances (2.3.10) in Theorem 2.3.3. Hence, it would be highly valuable to find a proof for the limit theorems that avoids such a condition, but, for now, this remains an open problem.

Chapter 3

Parametric Estimation

With the necessary asymptotic techniques developed in the previous chapter, we can now present and analyze several estimators for parameters in homogenized SDE models under the aspect of robustness and asymptotic normality. The estimators all share some common features. One feature is that we are aiming to retrieve parameters from the invariant densities of the homogenized limit SDE through observations of the unhomogenized slow variable of the multiscale system. The idea is that, as shown in the previous chapter, invariant densities and averaged functionals, that is, time averages, of the weakly perturbed observations are asymptotically well-behaved. The other feature is that we will identify these parameters through minimization of a suitable distance quantity. This idea lends itself to the so-called minimum distance approach, which is an ubiquitous method for statistical inference in parametric estimation problems and often also goes under the name of M-estimation, cf. [Wol57; Kut94; Vaa98; Pra99]. In certain contexts, it encapsulates many commonly known inference methods as special cases, e.g., maximum likelihood estimation, least squares methods, or method of moments.

In the first section, we introduce a general estimation framework for models under weak misspecification, define the notion of robustness, and provide an example showcasing the failure of the maximum likelihood estimator when it is built based on the limit model, but confronted with data from the perturbed model. Afterwards, we present a first simple example of a parametric estimator that is robust and asymptotically normal. Although the considered test case is quite simple, the example already highlights what is necessary in order to prove these two asymptotic properties and how the results in Chapter 2 come into play. In the second section, we present a more sophisticated estimation method, prove several interesting asymptotic results, and substantiate them with a numerical simulation study. Since this whole chapter deals with several different notions regarding the theory of tightness and weak convergence of measures, we recommend Appendix B for continuous reference to these topics.

3.1 A Primer on Parametric Inference under Weak Misspecification

As we are in a setting of parameter estimation under weak misspecification, we first extend some basic terminology that is commonly used in asymptotic estimation theory, cf. [IK81; Hen24; Kut94]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying complete probability space and $(\mathscr{X}_T, \mathscr{B}(\mathscr{X}_T))$ a measurable space with a non-empty set $\mathscr{X}_T$ of functions from $[0, T]$

to $\mathbb{R}^d$, equipped with a σ-algebra $\mathscr{B}(\mathscr{X}_T)$. Moreover, let $\Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$, be an open set and $\psi\colon \Theta \to \mathbb{R}^p$ a map. We consider the stochastic processes $X, X^\varepsilon \in \mathscr{X}_T$, $\varepsilon > 0$, and assume that their laws with respect to $\mathbb{P}$ can be parameterized through a finite-dimensional parameter $\vartheta \in \Theta$, i.e., there exist two sets of probability measures on $(\mathscr{X}_T, \mathscr{B}(\mathscr{X}_T))$, say

$$\mathcal{P} := \left\{\mathbb{P}^T_\vartheta \,|\, \vartheta \in \Theta\right\} \quad \text{and} \quad \mathcal{P}_\varepsilon := \left\{\mathbb{P}^T_{\psi(\vartheta)} \,|\, \vartheta \in \Theta\right\}, \tag{3.1.1}$$

serving as distributional models for X and X^ε, respectively, such that $\mathbb{P}^X \in \mathcal{P}$ and $\mathbb{P}^{X^\varepsilon} \in \mathcal{P}_\varepsilon$. In our specific context, we will refer to $\mathcal{P}$ as limit model or effective model and to $\mathcal{P}_\varepsilon$ as weakly perturbed model. The map ψ, which is independent of ε, draws a connection between the parameters of the two distributional models, see the beginning of Subsection 3.1.1 for an example. For identifiability purposes, we assume that the maps $\vartheta \mapsto \mathbb{P}^T_\vartheta$ and $\vartheta \mapsto \mathbb{P}^T_{\psi(\vartheta)}$ are injective. In particular, this implies that ψ is injective. If we, for a second, interpret X^ε as a realization of a random experiment on $(\mathscr{X}_T, \mathscr{B}(\mathscr{X}_T))$, then our statistical task is to find the parameter $\vartheta_0 \in \Theta$ which gives the distributional model $\mathbb{P}^T_{\psi(\vartheta_0)}$ and, thus, also $\mathbb{P}^T_{\vartheta_0}$. We will call such a ϑ_0 the true value or true parameter. Finding the true parameter can be accomplished with a parametric estimator, which is just a measurable map $\widehat{\theta}_T\colon \mathscr{X}_T \to \Theta$.

We will either consider the following family of T-dependent parametric models

$$\mathscr{M}_T := \left(\mathscr{X}_T, \mathscr{B}(\mathscr{X}_T), \left\{\mathbb{P}^T_{\psi(\vartheta)} \,|\, \vartheta \in \Theta\right\}\right), \quad T > 0, \tag{3.1.2}$$

or the family of ε-dependent parametric models

$$\mathscr{M}_{T_\varepsilon} := \left(\mathscr{X}_{T_\varepsilon}, \mathscr{B}(\mathscr{X}_{T_\varepsilon}), \left\{\mathbb{P}^{T_\varepsilon}_{\psi(\vartheta)} \,|\, \vartheta \in \Theta\right\}\right), \quad \varepsilon > 0, \tag{3.1.3}$$

where $T_\varepsilon > 0$ depends on the parameter ε. Note also that in the classical parameter estimation setting for well-specified models, we would rather work with the family of parametric models

$$\left(\mathscr{X}_T, \mathscr{B}(\mathscr{X}_T), \left\{\mathbb{P}^T_\vartheta \,|\, \vartheta \in \Theta\right\}\right), \quad T > 0.$$

We will say that the estimator $(\widehat{\theta}_T) := (\widehat{\theta}_T)_{T>0}$ is ε-T-robust for the true value $\theta_0 \in \Theta$ if for all $\delta > 0$,

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \mathbb{P}(|\widehat{\theta}_T(X^\varepsilon) - \theta_0|_2 > \delta) = 0. \tag{3.1.4}$$

Similarly, we will say that the estimator $(\widehat{\theta}_T)$ is T-ε-robust for the true value $\theta_0 \in \Theta$ if for all $\delta > 0$,

$$\lim_{T\to\infty}\limsup_{\varepsilon\to 0} \mathbb{P}(|\widehat{\theta}_T(X^\varepsilon) - \theta_0|_2 > \delta) = 0. \tag{3.1.5}$$

The reason to use lim sup instead of lim here is purely technical and essentially motivated by Alexandroff's theorem B.1.4. Lastly, we will say that the estimator $(\widehat{\theta}_{T_\varepsilon}) := (\widehat{\theta}_{T_\varepsilon})_{\varepsilon>0}$ is ε-robust for the true value $\theta_0 \in \Theta$ if for all $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mathbb{P}(|\widehat{\theta}_{T_\varepsilon}(X^\varepsilon) - \theta_0|_2 > \delta) = 0. \tag{3.1.6}$$

In these cases, we also write $\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{\theta}_T(X^\varepsilon) \stackrel{\mathbb{P}}{=} \theta_0$ and likewise for the other limits.

Observe that these definitions bear resemblance to the concept of consistency of estimators, see [Kut04; Kut94; IK81]; however, since we do not assume that the given observations are generated from a distribution of the limit model, the nomenclature is adjusted to our setting.

3.1.1 Example – Maximum Likelihood Estimation

Consider the multiscale overdamped Langevin equation (1.2.10) with $d = 1$, $\alpha, \sigma > 0$, $\mathbb{L} = 2\pi$, $V(x) = x^2/2$, and $p(x) = -\cos(x)$, that is,

$$dX_t^\varepsilon = \left[-\alpha X_t^\varepsilon - \frac{1}{\varepsilon}\sin\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\right] dt + \sqrt{2\sigma}\, dW_t, \quad X_0^\varepsilon = x_0, \tag{3.1.7}$$

and its homogenized limit equation (1.2.12), restated here for convenience,

$$dX_t = -\alpha\mathcal{K} X_t\, dt + \sqrt{2\sigma\mathcal{K}}\, d\bar{W}_t, \quad X_0 = x_0, \tag{3.1.8}$$

with the constant homogenization factor $\mathcal{K}$ defined in (1.2.16). If the statistical task is the estimation of the drift parameter $\vartheta := \alpha\mathcal{K} \in (0, \infty) =: \Theta$ in (3.1.8) through observations coming from (3.1.7), then we may consider the ε-dependent model

$$\mathscr{M}_{T_\varepsilon} := \left(C([0, T_\varepsilon]; \mathbb{R}), \mathscr{B}(C([0, T_\varepsilon]; \mathbb{R})), \left\{\mathbb{P}_{\psi(\vartheta)}^{T_\varepsilon} \,|\, \vartheta \in \Theta_0\right\}\right), \quad \varepsilon > 0, \tag{3.1.9}$$

where $\mathscr{B}(C([0, T_\varepsilon]; \mathbb{R}))$ is the Borel σ-algebra on $C([0, T_\varepsilon]; \mathbb{R})$, $T_\varepsilon \to \infty$ as $\varepsilon \to 0$, and Θ_0 is a compact interval in $\Theta = (0, \infty)$. In this case, the map ψ is given by $\psi(\vartheta) = \vartheta/\mathcal{K}$. We want to prove that, even when T_ε depends on the scale parameter $\varepsilon > 0$, the MLE $(\widehat{A}_{T_\varepsilon})$ defined in (2.1.3) is asymptotically biased, i.e., it does not recover the true parameter ϑ_0 when confronted with observations from (3.1.7). The precise statement is formulated in the next proposition.

Proposition 3.1.1. *Consider the multiscale overdamped Langevin equation* (3.1.7) *and its homogenized limit equation* (3.1.8)*. Let $\vartheta_0 = \alpha_0\mathcal{K} \in \Theta_0 \subset (0, \infty)$ be the true parameter for some $\alpha_0 > 0$. Then, for $T_\varepsilon = \Omega(\varepsilon^{-1})$ as $\varepsilon \to 0$ and under the true parameter ϑ_0, that is, whenever $\mathbb{P}^{X^\varepsilon} = \mathbb{P}_{\psi(\vartheta_0)}^{T_\varepsilon}$, we have*

$$\lim_{\varepsilon \to 0} \widehat{A}_{T_\varepsilon}(X^\varepsilon) \stackrel{\mathbb{P}}{=} \alpha_0. \tag{3.1.10}$$

Proof. We first note that the MLE is given by the formula

$$\widehat{A}_{T_\varepsilon}(X^\varepsilon) = -\frac{\int_0^{T_\varepsilon} X_t^\varepsilon\, dX_t^\varepsilon}{\int_0^{T_\varepsilon} |X_t^\varepsilon|^2\, dt}, \tag{3.1.11}$$

recall (2.1.3). We can apply the Itô formula to obtain

$$|X_{T_\varepsilon}^\varepsilon|^2 - |X_0^\varepsilon|^2 = 2\int_0^{T_\varepsilon} X_t^\varepsilon \left[-\alpha_0 X_t^\varepsilon - \frac{1}{\varepsilon}\sin\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\right] dt + 2\sqrt{2\sigma}\int_0^{T_\varepsilon} X_t^\varepsilon\, dW_t + 2\sigma T_\varepsilon,$$

and, after rearranging, we get the expression

$$\begin{aligned} \frac{2}{\varepsilon T_\varepsilon}\int_0^{T_\varepsilon} X_t^\varepsilon \sin\left(\frac{X_t^\varepsilon}{\varepsilon}\right) dt = \frac{|X_0^\varepsilon|^2 - |X_{T_\varepsilon}^\varepsilon|^2}{T_\varepsilon} + \frac{2\sqrt{2\sigma}}{T_\varepsilon}\int_0^{T_\varepsilon} X_t^\varepsilon\, dW_t \\ - \frac{2\alpha_0}{T_\varepsilon}\int_0^{T_\varepsilon} |X_t^\varepsilon|^2\, dt + 2\sigma. \end{aligned} \tag{3.1.12}$$

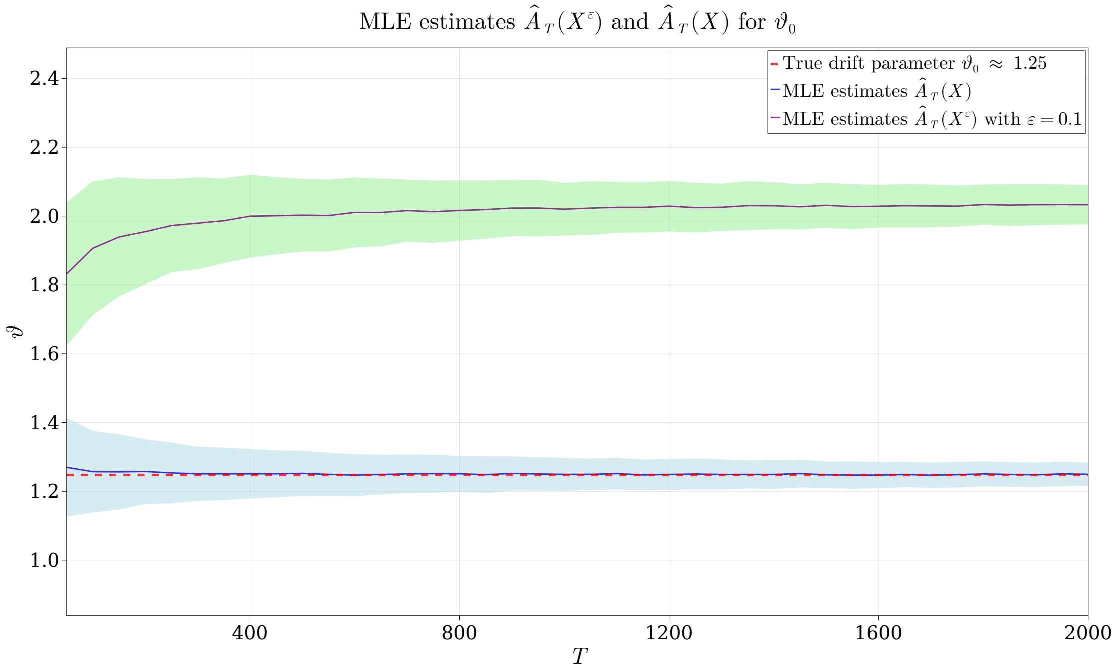


Figure 3.1: MLE estimates $\widehat{A}_T$ for ϑ_0 when the observations come from the Euler–Maruyama approximation of the multiscale overdamped Langevin equation (3.1.7) and the homogenized limit equation (3.1.8).

The first term on the RHS converges in probability to zero as $\varepsilon \to 0$ by Corollary 2.4.13 since $T_\varepsilon = \Omega(\varepsilon^{-1})$ as $\varepsilon \to 0$. For the second term on the RHS, we use a standard result for stochastic integrals, cf. [KS91, Problem 1.5.25], which states that if the quadratic variation process of a sequence of local martingales converges in probability to zero, so does the sequence of local martingales. By Corollary 2.4.12 with $m = 2$ (note that we have $b = 1$) and Lemma 2.4.7, it follows that

$$\lim_{\varepsilon \to 0} \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} |X_t^\varepsilon|^2 \, dt \overset{L^2}{=} \int_{\mathbb{R}} x^2 \mu(x) \, dx = \frac{\sigma}{\alpha_0}. \tag{3.1.13}$$

If we multiply the time integral on the LHS of the last limit statement (3.1.13) by $8\sigma/T_\varepsilon$, then we get exactly the quadratic variation process of the stochastic integral in (3.1.12). By the preceding arguments, this quadratic variation process converges in probability to zero as $\varepsilon \to 0$. So the stochastic integral, too, converges in probability to zero. Hence, we obtain

$$\lim_{\varepsilon \to 0} \frac{2}{\varepsilon T_\varepsilon} \int_0^{T_\varepsilon} X_t^\varepsilon \sin\left(\frac{X_t^\varepsilon}{\varepsilon}\right) dt \overset{\mathbb{P}}{=} -2\alpha_0 \frac{\sigma}{\alpha_0} + 2\sigma = 0.$$

Now we only need to observe that

$$\begin{aligned} \widehat{A}_{T_\varepsilon}(X^\varepsilon) &= \left(\frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} |X_t^\varepsilon|^2 \, dt\right)^{-1} \left[\frac{\alpha_0}{T_\varepsilon} \int_0^{T_\varepsilon} |X_t^\varepsilon|^2 \, dt + \frac{1}{\varepsilon T_\varepsilon} \int_0^{T_\varepsilon} X_t^\varepsilon \sin\left(\frac{X_t^\varepsilon}{\varepsilon}\right) dt\right] \\ &\quad + \left(\frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} |X_t^\varepsilon|^2 \, dt\right)^{-1} \frac{\sqrt{2\sigma}}{T_\varepsilon} \int_0^{T_\varepsilon} X_t^\varepsilon \, dW_t \end{aligned}$$

$$= \alpha_0 + \left(\frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} |X_t^\varepsilon|^2 \, dt \right)^{-1} \left[\frac{1}{\varepsilon T_\varepsilon} \int_0^{T_\varepsilon} X_t^\varepsilon \sin\left(\frac{X_t^\varepsilon}{\varepsilon} \right) dt + \frac{\sqrt{2\sigma}}{T_\varepsilon} \int_0^{T_\varepsilon} X_t^\varepsilon \, dW_t \right].$$

Applying the previous arguments yields $\lim_{\varepsilon \to 0} \widehat{A}_{T_\varepsilon}(X^\varepsilon) \stackrel{\mathbb{P}}{=} \alpha_0$. □

Figure 3.1 shows simulation results for the MLE illustrating the consequence of Proposition 3.1.1. We plugged in discrete observations from the Euler–Maruyama approximation, for which we have provided more details in Section 3.2.4, of both SDEs, the multiscale overdamped Langevin equation (3.1.7) and the homogenized limit equation (3.1.8), to compare the results. We set $\alpha_0 = 2$, $\sigma = 1$, $\varepsilon = 0.1$, and a time discretization step of $\Delta t = 10^{-3}$ for the Euler–Maruyama approximation. With these values, the true value equals approximately $\vartheta_0 \approx 1.248$. Each estimated value is an average of the same simulation experiment repeated for 1000 times. The shaded regions show empirical standard deviation bands of 1 standard deviation computed from the acquired estimates. As we can clearly and unsurprisingly see, the MLE estimates the true parameter ϑ_0 very well when the observations come from the homogenized limit equation. However, in concordance with Proposition 3.1.1, it is biased under observations of the multiscale overdamped Langevin equation and instead recovers the unhomogenized parameter α_0.

3.1.2 Example – Minimum Distance Estimation

We stay in the same statistical setting of the previous example introduced in Subsection 3.1.1 but propose a different estimator. For identifiability reasons of the homogenized limit equation (3.1.8), we assume that we know the homogenized diffusion parameter $\bar{\Sigma}^2 = 2\sigma\mathcal{K}$, e.g., from a prior estimation procedure. Such diffusion parameter estimation problems under multiscale data were analyzed, for example, in [PS07; KPK13; MP19; Abd+23]. We consider the following simple minimum distance estimator

$$\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) \in \underset{\vartheta \in \Theta_0}{\arg\inf} \left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon) \, dt - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta} u^2 \right) \right|^2, \tag{3.1.14}$$

where we recall that Θ_0 is a compact interval in $\Theta = (0, \infty)$ and $\vartheta = \alpha\mathcal{K}$. Evidently, the proposed estimator is based on the characteristic function of the invariant density μ at the point $u \in \mathbb{R}\backslash\{0\}$, which equals an unnormalized Gaussian density in this setting, since μ is the density of a centered normal distribution with variance σ/α. Since this estimator is a special case of the general minimum distance estimator which we will study more thoroughly in the next section, we will not discuss estimation-related issues, such as existence or measurability, here. These will be addressed in the more general setting of the next Section 3.2. Using our established limit theorems from Chapter 2, we can now prove ε-robustness and asymptotic normality of $(\widehat{\vartheta}_{T_\varepsilon})$ with ease.

Proposition 3.1.2. *Consider the multiscale overdamped Langevin equation* (3.1.7) *and its homogenized limit equation* (3.1.8)*. Let $u \in \mathbb{R}\backslash\{0\}$ and $\vartheta_0 = \alpha_0\mathcal{K} \in \Theta_0^\circ \subset (0, \infty)$ be the true parameter for some $\alpha_0 > 0$. Assume that $\bar{\Sigma}^2 = 2\sigma\mathcal{K}$ is known. Consider the time horizon T_ε such that $T_\varepsilon = \Omega(\varepsilon^{-\eta})$ for some $\eta > 0$ and $T_\varepsilon = o(\exp(4\sigma/(\alpha_0\varepsilon^2))$ as $\varepsilon \to 0$. Then, under the true parameter $\vartheta_0 = \alpha_0\mathcal{K}$, that is, whenever $\mathbb{P}^{X^\varepsilon} = \mathbb{P}^{T_\varepsilon}_{\psi(\vartheta_0)}$, we have*

$$\lim_{\varepsilon \to 0} \widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) \stackrel{\mathbb{P}}{=} \vartheta_0, \tag{3.1.15}$$

$$\sqrt{T_\varepsilon}(\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) - \vartheta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_1\left(0, \left(\frac{\bar{\Sigma}^2 u^2 \tau_u}{4\vartheta_0 \log(\vartheta_0)^2}\right)^2\right), \quad \varepsilon \to 0, \tag{3.1.16}$$

where

$$\tau_u^2 := \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi_u'(x)^2 \mu(x)\, dx = \sqrt{\frac{\vartheta_0}{\pi}} \bar{\Sigma} \int_{\mathbb{R}} \Phi_u'(x)^2 \exp\left(-\frac{\vartheta_0}{\bar{\Sigma}^2} x^2\right) dx, \tag{3.1.17}$$

with

$$\Phi_u(x) := -\int_0^x \frac{2}{\bar{\Sigma}^2 \mu(y)} \int_{-\infty}^y \left(\operatorname{Re}(\exp(iuz)) - \int_{\mathbb{R}} \exp(iuv)\mu(v)\, dv\right) \mu(z)\, dz\, dy, \quad x \in \mathbb{R}. \tag{3.1.18}$$

Proof. We first note that we can rewrite the minimization problem in (3.1.14) as follows

$$\begin{aligned} &\underset{\vartheta\in\Theta_0}{\arg\inf} \left| \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon)\, dt - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta} u^2\right)\right|^2 \\ =& \underset{\vartheta\in\Theta_0}{\arg\inf} \left(\operatorname{Re} \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon)\, dt - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta} u^2\right)\right)^2 \end{aligned}$$

because the imaginary part does not depend on the parameter ϑ. Observe that

$$\begin{aligned} &\operatorname{Re} \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon)\, dt - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta_0} u^2\right) \\ =& \operatorname{Re} \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon)\, dt - \int_{\mathbb{R}} \exp(iux)\mu(x)\, dx, \end{aligned} \tag{3.1.19}$$

which converges in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ to zero as $\varepsilon \to 0$ by Corollary 2.4.11. Moreover, using differentiation on Θ_0°, it is straightforward to obtain an explicit expression for the estimator, namely

$$\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) = -\frac{\bar{\Sigma}^2 u^2}{4}\left(\log\left(\operatorname{Re} \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon)\, dt\right)\right)^{-1}. \tag{3.1.20}$$

Hence, convergence of (3.1.19) to zero implies

$$\lim_{\varepsilon\to0} \widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) \overset{\mathbb{P}}{=} \lim_{\varepsilon\to0} -\frac{\bar{\Sigma}^2 u^2}{4}\left(\log\left(\operatorname{Re} \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon)\, dt\right)\right)^{-1} \overset{\mathbb{P}}{=} \vartheta_0,$$

proving the first claim (3.1.15). Moving on to asymptotic normality, we can write

$$\operatorname{Re} \frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon) - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta_0} u^2\right) dt = \frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon)\, dt + \sqrt{T_\varepsilon}(\bar{h}_\varepsilon(x) - \bar{h}(x)), \tag{3.1.21}$$

where for $x \in \mathbb{R}$,

$$\bar{h}_\varepsilon(x) := \operatorname{Re}(\exp(iux)) - \int_{\mathbb{R}} \exp(iuy)\mu_\varepsilon(y)\, dy,$$

$$\bar{h}(x) := \operatorname{Re}(\exp(iux)) - \int_{\mathbb{R}} \exp(iuy)\mu(y)\,dy,$$

since μ and μ_ε are symmetric densities. By Lemma 2.4.5, we already know that

$$|\bar{h}_\varepsilon(x) - \bar{h}(x)| \le \|\bar{h}_\varepsilon - \bar{h}\|_\infty = \mathcal{O}\left(\exp\left(-\frac{2\sigma}{\alpha_0\varepsilon^2}\right)\right),$$

so that by the assumption that $T_\varepsilon = o(\exp(4\sigma/(\alpha_0\varepsilon^2))$ as $\varepsilon \to 0$, we realize that the second term on the RHS in (3.1.21) goes to zero as $\varepsilon \to 0$. The first term on the RHS in (3.1.21) converges weakly to $\mathcal{N}_1(0,\tau_u^2)$ by Corollary 2.4.14, so that, with Slutzky's lemma B.1.5, this implies

$$\operatorname{Re}\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} \exp(iuX_t^\varepsilon) - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta_0}u^2\right)dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0,\tau_u^2), \quad \varepsilon \to 0.$$

The final claim (3.1.16) is just a simple application of the delta method, cf. [Hen24], to the function $g_u(\vartheta) := -\bar{\Sigma}^2u^2/(4\log(\vartheta))$, $\vartheta > 0$, with $g_u'(\vartheta) = \bar{\Sigma}^2u^2/(4\vartheta\log^2(\vartheta))$, and yields

$$\begin{aligned}\sqrt{T_\varepsilon}(\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) - \vartheta_0) &= \sqrt{T_\varepsilon}\left[g_u\left(\operatorname{Re}\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\exp(iuX_t^\varepsilon)\,dt\right) - g_u\left(\exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta_0}u^2\right)\right)\right]\\ &\xrightarrow{\mathcal{D}} \mathcal{N}_1\left(0,\left(\frac{\bar{\Sigma}^2u^2\tau_u}{4\vartheta_0\log(\vartheta_0)^2}\right)^2\right), \quad \varepsilon \to 0.\end{aligned}$$

□

Remark 3.1.3. *This example motivates clearly why the time horizon should depend on ε. It is less about the robustness result and more about the asymptotic normality result. More precisely, the decomposition on the RHS of* (3.1.21) *shows that T_ε cannot grow to infinity in an arbitrary way, because, otherwise, a bias emerges that can be made arbitrarily large. Interestingly though, this is something that is special to model misspecification. To illustrate this, suppose we plug in observations X coming from* (3.1.8) *into the estimator, that is, we consider the estimator*

$$\widehat{\vartheta}_T(X) \in \operatorname*{arg\,inf}_{\vartheta\in\Theta_0}\left|\frac{1}{T}\int_0^T \exp(iuX_t)\,dt - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta}u^2\right)\right|^2,$$

where T is now independent of ε, and we carry out the steps of the previous proof. In this case, (3.1.21) *turns into*

$$\operatorname{Re}\frac{1}{\sqrt{T}}\int_0^T \exp(iuX_t) - \exp\left(-\frac{\bar{\Sigma}^2}{4\vartheta_0}u^2\right)dt = \frac{1}{\sqrt{T}}\int_0^T \bar{h}(X_t)\,dt,$$

which actually has no bias term and, under standard assumptions, converges weakly to $\mathcal{N}_1(0,\tau_u^2)$ as $T \to \infty$. Thus, the time horizon, which represents the amount of data and, therefore, usually controls sampling error, acquires a new interpretation of bias amplification in a setting of model misspecification. As a result, when discussing distributional statements about estimators, it is important to find a value for the time horizon that minimizes both sampling error and bias amplification. Unfortunately, precise statements depend on the unknown scale separation parameter ε, whose estimation problem we will touch upon briefly in the next Chapter 4.

3.2 A General Minimum Distance Estimator Approach

The approach in this section is a direct generalization of the minimum distance estimator in the preceding section. They are both inspired by the classical BHEP goodness-of-fit test for multivariate normality, which was thoroughly studied in [HW97] and later generalized to the setting of random elements in a separable Hilbert space in [HJ21]. Minimum distance estimation was originally developed and introduced by Wolfowitz in [Wol57]. Since then, the method has been extensively studied and applied; see, for example, [Kut94] and the references therein for a comprehensive analysis of minimum distance estimators in the context of statistical inference for SDE processes. Although minimum distance estimators enjoy several favorable theoretical properties, such as consistency or asymptotic normality, a prevailing shortcoming is their computational cost; cf. [CM11] for a brief discussion of this aspect. We will partially address this issue in the following by providing a computationally tractable formula for our minimum distance estimator in a multivariate Gaussian setting. In non-Gaussian settings, we will utilize a fast Fourier transform algorithm for the efficient and accurate approximation of appearing convolution integrals.

3.2.1 Problem Setting and Approach

One can formulate the minimum distance estimator and prove results in very general settings, where the state space is, for example, a Polish space, but, since the stochastic processes that we consider are continuous, we will refrain from doing so and just present everything for the state space $C([0,T];\mathbb{R}^d)$. We adopt the notation from the previous Section 3.1 and consider the following set of assumptions.

Assumptions (MDE).

i) There exists a compact set $\Theta_0 \subset \Theta$ where the open set $\Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$, is the ambient parameter space.

ii) ν is a probability measure on $\mathbb{R}^d$.

iii) The stochastic processes $X, X^\varepsilon \in C([0,T];\mathbb{R}^d)$ are such that X^ε converges weakly to X in $C([0,T];\mathbb{R}^d)$ as $\varepsilon \to 0$.

iv) The two families of probability measures $\mathcal{P}$ and $\mathcal{P}_\varepsilon$ in (3.1.1) act as parametric distributional models for X and X^ε, respectively.

v) If $\mathbb{P}^X = \mathbb{P}^T_\vartheta$ for some $\vartheta \in \Theta$, then there exists a probability measure $\mu(\vartheta,\cdot)$ on $\mathbb{R}^d$ such that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(X_t)\,dt \overset{L^2}{=} \int_{\mathbb{R}^d} \varphi(x)\,d\mu(\vartheta,x) \tag{3.2.1}$$

for all $\varphi \in C_b(\mathbb{R}^d)$.

vi) The function $\Theta \ni \vartheta \mapsto \mu(\vartheta,\cdot)$ is weakly continuous, i.e., if $\vartheta_n \to \vartheta$, then $\mu(\vartheta_n,\cdot)$ converges weakly to $\mu(\vartheta,\cdot)$ on $\mathbb{R}^d$.

As before, we call a realization of the process X^ε a weakly perturbed observation of

X. Based on Assumptions (MDE), we define the following bounded, jointly continuous maps:

$$\hat{\mathscr{C}}_T : C([0,T];\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{C}; \quad (f,u) \mapsto \frac{1}{T}\int_0^T \exp\left(iu^\top f_t\right)\, dt, \tag{3.2.2}$$

$$\mathscr{C} : \Theta \times \mathbb{R}^d \to \mathbb{C}; \quad (\vartheta,u) \mapsto \int_{\mathbb{R}^d} \exp\left(iu^\top x\right)\, d\mu(\vartheta,x). \tag{3.2.3}$$

For notational brevity, we will occasionally write $\mu(\vartheta) := \mu(\vartheta,\cdot)$, $\mathscr{C}_\vartheta := \mathscr{C}(\vartheta,\cdot)$, and $\hat{\mathscr{C}}_T(X) := \hat{\mathscr{C}}_T(X,\cdot)$ whenever $X \in C([0,T];\mathbb{R}^d)$ is a fixed stochastic process. Observe that $\mathscr{C}_\vartheta$ is simply the characteristic function of the measure $\mu(\vartheta,\cdot)$ and $\hat{\mathscr{C}}_T$ is the empirical counterpart. Let $L^2(\nu) := L^2(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \nu)$ be the weighted L^2-space with respect to the weight measure ν. For $\vartheta \in \Theta$ and $X \in C([0,T];\mathbb{R}^d)$, we define the (random) distance quantity

$$\Delta_T(\vartheta, X) := \| \hat{\mathscr{C}}_T(X) - \mathscr{C}_\vartheta \|^2_{L^2(\nu)} = \int_{\mathbb{R}^d} \left| \hat{\mathscr{C}}_T(X,u) - \mathscr{C}_\vartheta(u) \right|^2 d\nu(u). \tag{3.2.4}$$

For fixed $X \in C([0,T];\mathbb{R}^d)$, the map $\Theta \ni \vartheta \mapsto \Delta_T(\vartheta,X)$ is continuous by the dominated convergence theorem. Hence, we can define the minimum distance estimator $\widehat{\vartheta}_T$ (MDE) by means of the minimization task

$$\widehat{\vartheta}_T(X) \in \operatorname*{arg\,inf}_{\vartheta \in \Theta_0} \Delta_T(\vartheta, X), \tag{3.2.5}$$

where the set on the RHS is non-empty by compactness of Θ_0 and continuity of the distance quantity. In other words, the existence of a minimizer is secured under Assumptions (MDE). To avoid dwelling too much on measurability questions, we assume the existence of a measurable minimizer $\widehat{\vartheta}_T$ and only consider those from now on; under the present assumptions, this can be justified by measurable selection arguments. Asymptotic uniqueness will be obtained separately in the results because we consider different limit orders in our robustness notion. Although we used X to exemplify the definitions, we will almost always analyze the MDE under the process X^ε.

Choosing the weight measure ν as a Dirac measure δ_u at a point $u \in \mathbb{R}\backslash\{0\}$ recovers the type of minimum distance estimator that is defined in (3.1.14). Minimum distance estimators are well known in the statistics literature. For example, in [Kut94] several minimum distance estimation approaches were rigorously discussed and studied for diffusion processes with small noise. Furthermore, questions of (asymptotic) existence and uniqueness of minimum distance estimators have been addressed and answered under mild regularity and identifiability conditions, see [Mil84; Kut94]. We will not dwell on these more general assumptions for the MDE at this stage. Instead, we want to provide a specific example that will become important in the numerical experiments presented later in the section.

Example 3.2.1. Let us first introduce useful notation for this example. We denote by φ_Q the density of a centered multivariate normal distribution with positive-definite covariance matrix $Q \in \mathbb{R}^{d\times d}$, i.e.,

$$\varphi_Q(x) := \frac{1}{(2\pi)^{d/2}\det(Q)^{-1/2}} \exp\left(-\frac{1}{2}x^\top Q^{-1} x\right), \quad x \in \mathbb{R}^d. \tag{3.2.6}$$

We further abbreviate $\varphi_\beta := \varphi_{\beta^2 I_d}$ for $\beta > 0$ and the identity matrix $I_d \in \mathbb{R}^{d\times d}$. Now let $\vartheta \in \Theta_0$. Choose the weight measure ν in (3.2.4) through the density φ_β, that is,

$$d\nu(u) = \varphi_\beta(u)\,du = \frac{1}{(2\pi\beta^2)^{d/2}} \exp\left(-\frac{|u|_2^2}{2\beta^2}\right) du, \quad u \in \mathbb{R}^d, \tag{3.2.7}$$

where $\det(\cdot)$ denotes the determinant. Note that the characteristic function of ν equals $k_\beta(x) := \exp(-\beta^2|x|_2^2/2)$, $x \in \mathbb{R}^d$. We can expand the complex inner product inside the integral in (3.2.4) and apply Fubini's theorem to obtain

$$\begin{aligned}
\Delta_T(\vartheta, X) &= \| \hat{\mathscr{C}}_T(X) \|_{L^2(\nu)}^2 - 2\,\mathrm{Re}\langle \hat{\mathscr{C}}_T(X), \mathscr{C}_\vartheta\rangle_{L^2(\nu)} + \| \mathscr{C}_\vartheta \|_{L^2(\nu)}^2 \\
&= \frac{1}{T^2}\int_0^T\int_0^T\int_{\mathbb{R}^d} \exp(iu^\top(X_t - X_s))\varphi_\beta(u)\,du\,dt\,ds \\
&\quad - \frac{2}{T}\int_0^T\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp(iu^\top(X_t - x))\varphi_\beta(u)\,du\,d\mu(\vartheta,x)\,dt \\
&\quad + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp(iu^\top(y - x))\varphi_\beta(u)\,du\,d\mu(\vartheta,x)\,d\mu(\vartheta,y) \\
&= \frac{1}{T^2}\int_0^T\int_0^T k_\beta(X_t - X_s)\,dt\,ds \\
&\quad - \frac{2}{T}\int_0^T\int_{\mathbb{R}^d} k_\beta(X_t - x)\,d\mu(\vartheta,x)\,dt + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k_\beta(y-x)\,d\mu(\vartheta,x)\,d\mu(\vartheta,y).
\end{aligned}$$

Hence, we get the formula

$$\begin{aligned}
\Delta_T(\vartheta, X) =& \frac{1}{T^2}\int_0^T\int_0^T k_\beta(X_t - X_s)\,dt\,ds \\
&- \frac{2}{T}\int_0^T (k_\beta * \mu(\vartheta))(X_t)\,dt + \int_{\mathbb{R}^d}(k_\beta * \mu(\vartheta))(x)\,d\mu(\vartheta,x),
\end{aligned} \tag{3.2.8}$$

where $*$ denotes the convolution operator on $\mathbb{R}^d$. Going a step further, if $\mu(\vartheta)$ is given by the density $\varphi_{Q(\vartheta)}$ of a centered multivariate normal distribution with a positive-definite covariance matrix $Q(\vartheta) \in \mathbb{R}^{d\times d}$, then formula (3.2.8) simplifies to

$$\begin{aligned}
\Delta_T(\vartheta, X) =& \frac{1}{T^2}\int_0^T\int_0^T k_\beta(X_t - X_s)\,dt\,ds \\
&- \frac{2}{T\sqrt{\det(I_d + \beta^2 Q(\vartheta))}}\int_0^T \exp\left(-\frac{\beta^2}{2}X_t^\top\left(I_d + \beta^2 Q(\vartheta)\right)^{-1} X_t\right) dt \\
&+ \frac{1}{\sqrt{\det(I_d + 2\beta^2 Q(\vartheta))}},
\end{aligned} \tag{3.2.9}$$

The last formula is a consequence of the convolution formula for the densities of two normal distributions, which states $\varphi_{Q_1} * \varphi_{Q_2} = \varphi_{Q_1+Q_2}$, and the following integral formula for the product of two such densities

$$\int_{\mathbb{R}^d} \varphi_{Q_1}(x)\varphi_{Q_2}(x)\,dx = \frac{1}{(2\pi)^{d/2}\sqrt{\det(Q_1+Q_2)}}. \tag{3.2.10}$$

In our case, we have $k_\beta = (2\pi/\beta^2)^{d/2}\, \varphi_{1/\beta}$, so that

$$k_\beta * \mu(\vartheta) = \frac{(2\pi)^{d/2}}{\beta^d}\varphi_{1/\beta} * \mu(\vartheta) = \frac{(2\pi)^{d/2}}{\beta^d}\varphi_{Q+\beta^{-2}I_d}.$$

From this formula, we immediately obtain the middle term on the RHS of (3.2.9), whereas for the third term on the RHS of (3.2.8), we simply calculate

$$\begin{aligned}\int_{\mathbb{R}^d}(k_\beta * \mu(\vartheta))(x)\, d\mu(\vartheta, x) &= \frac{(2\pi)^{d/2}}{\beta^d}\int_{\mathbb{R}^d}\varphi_{Q+\beta^{-2}I_d}(x)\varphi_Q(x)\, dx \\ &= \frac{1}{\beta^d\sqrt{\det(\beta^{-2}I_d + 2Q(\vartheta))}} = \frac{1}{\sqrt{\det(I_d + 2\beta^2 Q(\vartheta))}}.\end{aligned}$$

In this particular Gaussian case, there is a remarkable connection to nonparametric kernel density estimators, compare with the results in [HW97]. Namely, we can define the following kernel density estimator

$$K_{T,\beta}(X, x) := \frac{1}{T}\int_0^T \varphi_h(X_t - x)\, dt = \frac{1}{Th^d}\int_0^T \frac{1}{(2\pi)^{d/2}}\exp\left(-\frac{|X_t - x|_2^2}{2h^2}\right) dt, \qquad (3.2.11)$$

with $h := 1/(\sqrt{2}\beta)$. For $D := 2^{-1}\beta^{-2}I_d + Q(\vartheta)$ and $K_{T,\beta}(X) := K_{T,\beta}(X, \cdot)$, we can calculate

$$\begin{aligned}\frac{(2\pi)^{d/2}}{\beta^d}\| K_{T,\beta}(X) - \varphi_D \|_{L^2(\mathbb{R}^d)}^2 = \frac{(2\pi)^{d/2}}{\beta^d}\Big(&\| K_{T,\beta}(X) \|_{L^2(\mathbb{R}^d)}^2 \\ &- 2\langle K_{T,\beta}(X), \varphi_D\rangle_{L^2(\mathbb{R}^d)} + \| \varphi_D \|_{L^2(\mathbb{R}^d)}^2\Big).\end{aligned}$$

After completing the square, the first term equals

$$\begin{aligned}\| K_{T,\beta}(X) \|_{L^2(\mathbb{R}^d)}^2 &= \frac{1}{T^2h^{2d}}\int_0^T\int_0^T\int_{\mathbb{R}^d}\varphi_h(X_t - x)\varphi_h(X_s - x)\, dx\, dt\, ds \\ &= \frac{1}{(2\pi)^d T^2 h^{2d}}\int_0^T\int_0^T \exp\left(-\frac{|X_t - X_s|_2^2}{4h^2}\right) \\ &\quad \int_{\mathbb{R}^d}\exp\left(-\frac{|(X_t + X_s)/2 - x|_2^2}{h^2}\right) dx\, dt\, ds \\ &= \frac{\pi^{d/2}h}{(2\pi)^d T^2 h^{2d}}\int_0^T\int_0^T \exp\left(-\frac{\beta^2}{2}\,|X_t - X_s|_2^2\right) dt\, ds \\ &= \frac{\beta^d}{(2\pi)^{d/2}}\frac{1}{T^2}\int_0^T\int_0^T k_\beta(X_t - X_s)\, dt\, ds.\end{aligned}$$

With the previously used convolution formula, the middle inner product term equals

$$\langle K_{T,\beta}(X), \varphi_D\rangle_{L^2(\mathbb{R}^d)} = \frac{1}{T}\int_0^T \varphi_h * \varphi_D(X_t)\, dt = \frac{1}{T}\int_0^T \varphi_{D+h^2I_d}(X_t)\, dt$$

$$= \frac{1}{T}\int_0^T \varphi_{Q(\vartheta)+\beta^{-2}I_d}(X_t)\,dt$$

$$= \frac{\beta^d}{T(2\pi)^{d/2}\sqrt{\det(I_d+\beta^2 Q(\vartheta))}}\int_0^T \exp\left(-\frac{\beta^2}{2}X_t^\top\left(I_d+\beta^2 Q(\vartheta)\right)^{-1}X_t\right)dt.$$

Finally, by the formula (3.2.10), the third term equals

$$\int_{\mathbb{R}^d}\varphi_D(x)^2\,dx = \frac{1}{(2\pi)^{d/2}\sqrt{\det(2D)}} = \frac{\beta^d}{(2\pi)^{d/2}\sqrt{\det(I_d+2\beta^2 Q(\vartheta))}}\,.$$

Hence, multiplying these three terms by $(2\pi)^{d/2}/\beta^d$ proves the representation

$$\Delta_T(\vartheta, X) = \frac{(2\pi)^{d/2}}{\beta^d}\|\,K_{T,\beta}(X)-\varphi_D\,\|^2_{L^2(\mathbb{R}^d)}. \tag{3.2.12}$$

This formula allows for the interpretation of β as a smoothing parameter through the defined bandwidth $h = 1/(\sqrt{2}\beta)$ of the kernel density estimator. Although the formula suggests that one could analyze asymptotic properties of the MDE through the kernel density estimator – using material such as [Bos98; GN16; Tsy08] – we will not take this path as we aim for more general asymptotic statements.

The example demonstrates that even in the simple case of a multivariate normal distribution, it is impossible to derive a closed-form expression for the MDE, which, from a numerical point of view, necessitates the use of optimization algorithms. Fortunately, in a multivariate Gaussian setting, the formula (3.2.9) makes the computation quite tractable, which need not be the case for minimum distance estimators in general, cf. [CM11].

3.2.2 Asymptotic Properties

We come to questions of robustness and asymptotic normality of the proposed MDE. As we will see, the proofs for the various robustness properties that we defined in Section 3.1 are quite straightforward and certainly rely on Assumptions (MDE), but hinge primarily upon limit theorems for time averages of the observations, which, in contrast, can be difficult to obtain as we have realized in Chapter 2. Considering the fact that we try to retrieve the parameter of interest from the invariant measure of the limit model, it is rather unsurprising that limit theorems for time averages play a significant role here.

Proposition 3.2.2. *Let Assumptions (MDE) hold and fix $\vartheta_0 \in \Theta_0$ as the true parameter. Assume that, whenever $\mathbb{P}^{X^\varepsilon} = \mathbb{P}^T_{\psi(\vartheta_0)}$, there exists another probability measure $\mu_\varepsilon(\vartheta_0,\cdot) := \mu_\varepsilon(\psi(\vartheta_0),\cdot)$ on $\mathbb{R}^d$ such that $\mu_\varepsilon(\vartheta_0,\cdot)$ converges weakly to $\mu(\vartheta_0,\cdot)$ on $\mathbb{R}^d$ as $\varepsilon \to 0$, and such that for all $\varphi \in C_b(\mathbb{R}^d)$ and $\varepsilon > 0$,*

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \varphi(X_t^\varepsilon)\,dt \overset{L^2}{=} \int_{\mathbb{R}^d}\varphi(x)\,d\mu_\varepsilon(\vartheta_0, x). \tag{3.2.13}$$

Suppose further that the separation condition

$$d_\delta(\vartheta_0) := \inf_{|\vartheta-\vartheta_0|_2>\delta}\|\,\mathscr{C}_\vartheta - \mathscr{C}_{\vartheta_0}\,\|_{L^2(\nu)} > 0, \tag{3.2.14}$$

and the ν-identifiability condition

$$\| \mathscr{C}_{\vartheta_1} - \mathscr{C}_{\vartheta_2} \|_{L^2(\nu)} = 0 \quad \Rightarrow \quad \vartheta_1 = \vartheta_2 \quad \forall \vartheta_1, \vartheta_2 \in \Theta. \tag{3.2.15}$$

hold. Then, under the true parameter ϑ_0, the MDE $(\widehat{\vartheta}_T)$ is ε-T-robust for ϑ_0.

Proof. The ν-identifiability condition is needed for global model identifiability as required in the beginning of Section 3.1. The separation condition is used to prove the robustness of the MDE, but it also gives model identifiability in the neighborhood of the true parameter. By definition of the MDE, we have

$$\left\{ |\widehat{\vartheta}_T(X^\varepsilon) - \vartheta_0|_2 > \delta \right\} \subset \left\{ \inf_{|\vartheta - \vartheta_0|_2 \le \delta} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_\vartheta \|_{L^2(\nu)} \ge \inf_{|\vartheta - \vartheta_0|_2 > \delta} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_\vartheta \|_{L^2(\nu)} \right\}$$

for all $\delta > 0$. It follows that

$$\begin{aligned}
&\mathbb{P}(|\widehat{\vartheta}_T(X^\varepsilon) - \vartheta_0|_2 > \delta) \\
&\le \mathbb{P}\left(\inf_{|\vartheta - \vartheta_0|_2 \le \delta} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_\vartheta \|_{L^2(\nu)} \ge \inf_{|\vartheta - \vartheta_0|_2 > \delta} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_\vartheta \|_{L^2(\nu)} \right) \\
&\le \mathbb{P}\Bigg(\inf_{|\vartheta - \vartheta_0|_2 \le \delta} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \\
&\qquad + \inf_{|\vartheta - \vartheta_0|_2 \le \delta} \| \mathscr{C}_\vartheta - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \ge \inf_{|\vartheta - \vartheta_0|_2 \ge \delta} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_\vartheta \|_{L^2(\nu)} \Bigg).
\end{aligned}$$

We use the fact that $\inf_{|\vartheta - \vartheta_0|_2 \le \delta} \| \mathscr{C}_\vartheta - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} = 0$, the triangle inequality, and Markov's inequality to deduce

$$\begin{aligned}
\mathbb{P}(|\widehat{\vartheta}_T(X^\varepsilon) - \vartheta_0|_2 > \delta) &\le \mathbb{P}\Bigg(\| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \ge \inf_{|\vartheta - \vartheta_0|_2 > \delta} \| \mathscr{C}_\vartheta - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \\
&\qquad\qquad - \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \Bigg) \\
&= \mathbb{P}\Bigg(2 \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \ge d_\delta(\vartheta_0) \Bigg) \le \frac{4}{d_\delta(\vartheta_0)^2} \mathbb{E}\, \Delta_T(\vartheta_0, X^\varepsilon),
\end{aligned}$$

Next, we define

$$\mathscr{C}_{\varepsilon,\vartheta_0}(u) := \int_{\mathbb{R}^d} \exp\left(i u^\top x \right) \, d\mu_\varepsilon(\vartheta_0, x), \quad u \in \mathbb{R}^d. \tag{3.2.16}$$

and bound the last expectation by

$$\mathbb{E}\, \Delta_T(\vartheta_0, X^\varepsilon) \le 2 \left[\mathbb{E} \| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\varepsilon,\vartheta_0} \|^2_{L^2(\nu)} + \| \mathscr{C}_{\varepsilon,\vartheta_0} - \mathscr{C}_{\vartheta_0} \|^2_{L^2(\nu)} \right].$$

By Tonelli's theorem, assumption (3.2.13), and the dominated convergence theorem, we obtain for the first term on the RHS of the last inequality

$$\mathbb{E}\,\| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\varepsilon,\vartheta_0} \|^2_{L^2(\varphi)} = \int_{\mathbb{R}^d} \mathbb{E}\left| \frac{1}{T}\int_0^T \exp\left(iu^\top X_t^\varepsilon\right)\,dt - \mathscr{C}_{\varepsilon,\vartheta_0}(u)\right|^2 d\nu(u) \to 0, \quad T\to\infty,$$

and, for the second term, we have that $\mu_\varepsilon(\vartheta_0,\cdot)$ converges weakly to $\mu(\vartheta_0,\cdot)$ on $\mathbb{R}^d$ as $\varepsilon\to 0$, so that $\mathscr{C}_{\varepsilon,\vartheta_0}\to\mathscr{C}_{\vartheta_0}$ pointwise on $\mathbb{R}^d$ as $\varepsilon\to 0$. Another application of the dominated convergence theorem gives

$$\| \mathscr{C}_{\varepsilon,\vartheta_0} - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \to 0, \quad \varepsilon\to 0.$$

Putting everything together, we obtain

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \mathbb{P}(|\widehat{\vartheta}_T(X^\varepsilon) - \vartheta_0|_2 > \delta) = 0. \qquad \square$$

Remark 3.2.3. *If we exchange the ergodicity assumption* (3.2.13) *by the condition that, for any $\varphi\in C_b(\mathbb{R}^d)$,*

$$\mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\varphi(X_t^\varepsilon)\,dt - \int_{\mathbb{R}^d}\varphi(x)\mu_\varepsilon(\vartheta_0,x)\,dx\right|^2 \to 0, \quad \varepsilon\to 0, \tag{3.2.17}$$

is satisfied, then, by the same proof, we obtain the ε-robustness of $(\widehat{\vartheta}_{T_\varepsilon})$ for ϑ_0. We have already proved results of this kind in Chapter 2.

We briefly want to recall that the space $C(\mathbb{R}^d;\mathbb{C})$ is equipped with the metric

$$\rho(f,g) := \sum_{n=1}^\infty \frac{1}{2^n}\frac{\rho_n(f,g)}{1+\rho_n(f,g)}, \quad \rho_n(f,g) := \sup_{x\in[-n,n]^d}|f(x)-g(x)|_2, \quad f,g\in C(\mathbb{R}^d;\mathbb{C}), \tag{3.2.18}$$

which makes $C(\mathbb{R}^d;\mathbb{C})$ a complete, separable metric space. For details on tightness and weak convergence of random elements in $C(\mathbb{R}^d;\mathbb{C})$, we refer the reader to Appendix B.

In the next proposition, we exchange the order of the limits and prove the T-ε-robustness under different assumptions. Heuristically speaking, the assumptions in the following proposition make sense as we no longer need the ergodicity on the "perturbation level" but only on the "limit level". Nonetheless, we require other additional assumptions which secure tightness of a naturally emerging random element in $C(\mathbb{R}^d;\mathbb{C})$.

Proposition 3.2.4. *Let Assumptions (MDE), the separation condition* (3.2.14), *and the ν-identifiability condition* (3.2.15) *hold. Fix $\vartheta_0\in\Theta_0$ as the true parameter. Furthermore, assume that for some $p>d$,*

$$M(\vartheta_0) := \int_{\mathbb{R}^d}|x|_2\,d\mu(\vartheta_0,x) < \infty, \qquad \sup_{\varepsilon>0}\int_0^T \mathbb{E}\,|X_t^\varepsilon|_2^p\,dt \le C_T, \tag{3.2.19}$$

with a constant $C_T>0$ that can depend on T. Then, under the true parameter ϑ_0, the MDE $(\widehat{\vartheta}_T)$ is T-ε-robust for ϑ_0.

Proof. In the preceding proof, we already established that

$$\mathbb{P}(|\widehat{\vartheta}_T(X^\varepsilon) - \vartheta_0|_2 > \delta) \leq \mathbb{P}\left(2\| \hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \geq d_\delta(\vartheta_0) \right) \tag{3.2.20}$$

for any $\varepsilon > 0$ and $T > 0$. The last expression motivates the definition of the map

$$Z_T(X^\varepsilon, u) := \hat{\mathscr{C}}_T(X^\varepsilon, u) - \mathscr{C}_{\vartheta_0}(u) = \frac{1}{T}\int_0^T \exp\left(iu^\top X_t^\varepsilon\right) - \mathscr{C}_{\vartheta_0}(u)\, dt, \quad u \in \mathbb{R}^d, \tag{3.2.21}$$

with $T > 0$ fixed. We claim that the map $Z_T(X^\varepsilon, \cdot)\colon \mathbb{R}^d \to \mathbb{C}$ is a random element in $C(\mathbb{R}^d; \mathbb{C})$, which means that $\Omega \ni \omega \mapsto Z_T(\omega) := Z_T(X^\varepsilon(\omega), \cdot)$ takes values in $C(\mathbb{R}^d; \mathbb{C})$ and is $\mathcal{F}$-$\mathscr{B}(C(\mathbb{R}^d; \mathbb{C}))$-measurable, where $\mathscr{B}(C(\mathbb{R}^d; \mathbb{C}))$ is the Borel σ-algebra of the metric space $C(\mathbb{R}^d; \mathbb{C})$ and $\mathcal{F}$ is the σ-algebra of the underlying probability space. For fixed $\omega \in \Omega$, the map $\mathbb{R}^d \ni u \mapsto Z_T(X^\varepsilon(\omega), u) \in \mathbb{C}$ is continuous by the dominated convergence theorem, hence, an element of $C(\mathbb{R}^d; \mathbb{C})$. Regarding the measurability, we only need to prove that $\Omega \ni \omega \mapsto \rho(Z_T(\omega), f)$ is $\mathcal{F}$-$\mathscr{B}(\mathbb{R})$-measurable for any $f \in C(\mathbb{R}^d; \mathbb{C})$. This is because

$$Z_T^{-1}(B_r(f)) = \{\omega \in \Omega \mid \rho(Z_T(\omega, f)) < r\},$$

where $B_r(f)$ is the open ball of f in $C(\mathbb{R}^d; \mathbb{C})$ with radius $r > 0$, and these open balls generate the Borel σ-algebra $\mathscr{B}(C(\mathbb{R}^d; \mathbb{C}))$. To see that the map $\omega \mapsto \rho(Z_T(\omega), f)$ is indeed $\mathcal{F}$-$\mathscr{B}(\mathbb{R})$-measurable, one only needs to realize that $\omega \mapsto \rho_n(Z_T(\omega, f)$ is $\mathcal{F}$-$\mathscr{B}(\mathbb{R})$-measurable, which is immediate from the following identities

$$\rho_n(Z_T(\omega, f) = \sup_{u \in [-n,n]^d} |Z_T(\omega, u) - f(u)| = \sup_{u \in \mathbb{Q} \cap [-n,n]^d} |Z_T(\omega, u) - f(u)|.$$

The last equality in the line above is a standard fact for continuous functions. The map $\omega \mapsto |Z_T(\omega, u) - f(u)|$ for fixed $u \in \mathbb{R}^d$ is measurable by Fubini's theorem. Hence, $\omega \mapsto \rho_n(Z_T(\omega, f)$ is measurable as the supremum of measurable functions over a countable set and, in turn, $\omega \mapsto \rho(Z_T(\omega), f)$ is measurable as the countable series of measurable functions. Completely analog considerations hold for $Z_T(X, \cdot)$. On the other hand, $Z_T(\cdot, u)\colon (C([0,T]; \mathbb{R}^d), \|\cdot\|_\infty) \to \mathbb{C}$ with fixed $u \in \mathbb{R}^d$ is bounded and Lipschitz-continuous. Indeed, for fixed $u \in \mathbb{R}^d$ and any $f, g \in C([0,T]; \mathbb{R}^d)$, we have $|Z_T(f, u)| \leq 2$ and

$$\begin{aligned} |Z_T(f,u) - Z_T(g,u)| &= \left| \frac{1}{T}\int_0^T \exp\left(iu^\top f_t\right) - \exp\left(iu^\top g_t\right)\, dt \right| \\ &\leq \frac{1}{T}\int_0^T \left| \int_{u^\top g_t}^{u^\top f_t} i \exp(is)\, ds \right| dt \\ &\leq \frac{1}{T}\int_0^T \left| u^\top (f_t - g_t) \right| dt \leq |u|_2 \|f - g\|_\infty . \end{aligned}$$

In the following, we prove, for $T > 0$,

$$Z_T(X^\varepsilon, \cdot) \xrightarrow{\mathcal{D}} Z_T(X, \cdot) \quad \text{in } C(\mathbb{R}^d; \mathbb{C}) \tag{3.2.22}$$

as $\varepsilon \to 0$. If not stated otherwise, the subsequent limit behavior is always understood as $\varepsilon \to 0$. We fix arbitrary vectors $u_1, \dots, u_k \in \mathbb{R}^d$, $k \in \mathbb{N}$, and define the map

$$F\colon (C([0,T];\mathbb{R}^d), \|\cdot\|_\infty) \to (\mathbb{C}^k, \|\cdot\|_{\mathbb{C}^k}); \quad f \mapsto (Z_T(f,u_1), \dots, Z_T(f,u_k)).$$

It immediately follows by the calculation before that F is Lipschitz-continuous:

$$\|F(f) - F(g)\|^2_{\mathbb{C}^k} = \sum_{i=1}^k |Z_T(f,u_i) - Z_T(g,u_i)|_2 \le \sum_{i=1}^k |u_i|_2^2 \|f-g\|^2_\infty, \quad f,g \in C([0,T];\mathbb{R}^d).$$

Weak convergence of X^ε to X in $C([0,T];\mathbb{R}^d)$ and the continuous mapping theorem B.1.6 yield the convergence of the finite-dimensional distributions of $Z_T(X^\varepsilon, \cdot)$:

$$(Z_T(X^\varepsilon,u_1), \dots Z_T(X^\varepsilon,u_k)) \xrightarrow{\mathcal{D}} (Z_T(X,u_1), \dots Z_T(X,u_k)) \quad \text{in } \mathbb{C}^k.$$

To finish the proof of (3.2.22), it suffices to establish tightness of $(Z_T(X^\varepsilon, \cdot))_{\varepsilon>0}$ in the space $C(\mathbb{R}^d;\mathbb{C})$ as weak convergence then follows from Theorem B.2.1. For this purpose, we aim at establishing the two conditions of Lemma B.2.2, which require that $(Z_T(X^\varepsilon, \cdot))_{\varepsilon>0}$ satisfies, for some positive constants $\alpha, \beta, \gamma > 0$,

$$\sup_{\varepsilon>0} \mathbb{E}\, |Z_T(X^\varepsilon, 0)|^\alpha < \infty, \tag{3.2.23}$$

and, for all $k \in \mathbb{N}$ and $u, v \in \{x \in \mathbb{R}^d \mid |x|_2 \le k\}$,

$$\sup_{\varepsilon>0} \mathbb{E}\, |Z_T(X^\varepsilon, u) - Z_T(X^\varepsilon, v)|^\beta \le C_k\, |u-v|_2^{d+\gamma}, \tag{3.2.24}$$

where $C_k > 0$ can depend on k. We will establish these conditions for $\alpha = 1$, $\beta = p$, and $\gamma = p - d$. The first condition (3.2.23) is obvious since $Z_T(X^\varepsilon, 0) = 0$, $\mathbb{P}$-a.s., for all $\varepsilon > 0$. For the second condition (3.2.24), we estimate

$$\begin{aligned}
|Z_T(X^\varepsilon, u) - Z_T(X^\varepsilon, v)| &\le \Bigg[\left| \frac{1}{T} \int_0^T \exp\left(iu^\top X_t^\varepsilon\right) - \exp\left(iv^\top X_t^\varepsilon\right)\, dt \right| \\
&\qquad + |\mathscr{C}_{\vartheta_0}(u) - \mathscr{C}_{\vartheta_0}(v)| \Bigg] \\
&\le \left[\left(\frac{1}{T} \int_0^T |X_t^\varepsilon|_2\, dt \right) |u-v|_2 + M(\vartheta_0)\, |u-v|_2 \right] \\
&\le \left[\frac{1}{T} \int_0^T |X_t^\varepsilon|_2\, dt + M(\vartheta_0) \right] |u-v|_2\,,
\end{aligned}$$

which turns into

$$\mathbb{E}\, |Z_T(X^\varepsilon, u) - Z_T(X^\varepsilon, v)|^p \le 2^{p-1} \left[\frac{1}{T} \int_0^T \mathbb{E}\, |X_t^\varepsilon|_2^p\, dt + M(\vartheta_0)^p \right] |u-v|_2^p$$

after exponentiation by p, taking expectations, and applying Jensen's inequality. Thus, by the assumptions (3.2.19), condition (3.2.24) is satisfied, as well. Hence, (3.2.22) follows by Lemma B.2.2 and Theorem B.2.1. The next natural step would be to apply the continuous mapping theorem to $\|\cdot\|_{L^2(\nu)}$, but this function is not even well-defined on $C(\mathbb{R}^d;\mathbb{C})$. Luckily, we can remedy this as follows. First, note that $Z_T(X^\varepsilon,\cdot), Z_T(X,\cdot) \in S_0 := \{f \in C(\mathbb{R}^d;\mathbb{C}) \mid \|f\|_\infty \le 2\}$, $\mathbb{P}$-a.s., where $\|f\|_\infty = \sup_{u\in\mathbb{R}^d}|f(u)|$. It is easy to see that S_0 is closed in $C(\mathbb{R}^d;\mathbb{C})$. We simply have to notice that

$$S_0 = \bigcap_{n=1}^{\infty}\{f \in C(\mathbb{R}^d;\mathbb{C}) \mid \sup_{u\in[-n,n]^d}|f(u)| \le 2\},$$

and this is just the countable intersection of preimages of the closed set $[0,2]$ under the continuous map $C(\mathbb{R}^d;\mathbb{C}) \ni f \mapsto \sup_{u\in[-n,n]^d}|f(u)|$. We now define $\phi : S_0 \to [0,\infty)$ by $\phi(f) := \|f\|_{L^2(\nu)}$. The map ϕ is bounded with $|\phi(f)| \le 2$ for all $f \in S_0$. It is also continuous, since, for any $\delta > 0$, we can first find $n \in \mathbb{N}$ large enough such that $\nu(\mathbb{R}^d\backslash[-n,n]^d) < \delta$ and, thus, for any $f_k \to f$ in $C(\mathbb{R}^d;\mathbb{C})$ with $f_k, f \in S_0$, it follows that

$$\begin{aligned}|\phi(f_k)-\phi(f)| &= \int_{\mathbb{R}^d}|f_k(u)|^2 - |f(u)|^2\,d\nu(u)\\ &= \int_{\mathbb{R}^d}(|f_k(u)|-|f(u)|)(|f_k(u)|+|f(u)|)\,d\nu(u)\\ &= \int_{\mathbb{R}^d}|f_k(u)-f(u)|(|f_k(u)|+|f(u)|)\,d\nu(u)\\ &\le 4\int_{[-n,n]^d}|f_k(u)-f(u)|\,d\nu(u) + 16\nu(\mathbb{R}^d\backslash[-n,n]^d)\\ &\le 4\rho_n(f_k,f)+16\delta.\end{aligned}$$

By the Tietze–Urysohn theorem B.1.7, ϕ extends to a bounded continuous function $\tilde\phi$ on the whole of $C(\mathbb{R}^d;\mathbb{C})$ such that $\tilde\phi|_{S_0} = \phi$. We can now apply the continuous mapping theorem B.1.6 to obtain

$$\phi(Z_T(X^\varepsilon,\cdot)) = \tilde\phi(Z_T(X^\varepsilon,\cdot)) \xrightarrow{\mathcal{D}} \tilde\phi(Z_T(X,\cdot)) = \phi(Z_T(X,\cdot)),$$

which is equivalent to

$$\|\hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0}\|_{L^2(\nu)} \xrightarrow{\mathcal{D}} \|\hat{\mathscr{C}}_T(X) - \mathscr{C}_{\vartheta_0}\|_{L^2(\nu)} \quad \text{in } \mathbb{R}.$$

Returning back to (3.2.20), Alexandroff's theorem B.1.4 implies

$$\limsup_{\varepsilon\to 0}\mathbb{P}\left(2\|\hat{\mathscr{C}}_T(X^\varepsilon) - \mathscr{C}_{\vartheta_0}\|_{L^2(\nu)} \ge d_\delta(\vartheta_0)\right) \le \mathbb{P}\left(2\|\hat{\mathscr{C}}_T(X) - \mathscr{C}_{\vartheta_0}\|_{L^2(\nu)} \ge d_\delta(\vartheta_0)\right),$$

so that the final claim now follows from Markov's inequality and (MDE) v), the mean ergodic theorem for X. □

Our next result concerns an asymptotic representation of the centered and scaled quantity $\sqrt{T_\varepsilon}(\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) - \vartheta_0)$ in a way that eventually allows for a proof of the asymptotic normality of the MDE as $\varepsilon \to 0$.

Theorem 3.2.5. *Assume the setting of Proposition 3.2.2 with condition* (3.2.17) *instead of* (3.2.13) *and assume additionally that* Θ_0 *is convex. Fix* $\vartheta_0 \in \Theta_0^\circ$ *as the true parameter. Suppose that the function* $\Theta_0^\circ \ni \vartheta \mapsto \mathscr{C}_\vartheta(u) \in \mathbb{C}^p$, $u \in \mathbb{R}^d$, *is twice continuously differentiable with respect to* ϑ *and*

$$\sup_{\vartheta\in\Theta_0^\circ} |\partial_\vartheta \mathscr{C}_\vartheta(u)|_2 \le B^{(1)}(u), \quad \sup_{\vartheta\in\Theta_0^\circ} \left\|\partial_\vartheta^2 \mathscr{C}_\vartheta(u)\right\|_F \le B^{(2)}(u), \tag{3.2.25}$$

with $B^{(1)}, B^{(2)} \in L^2(\nu)$. *Here,* $\|A\|_F^2 := \sum_{i,j=1}^p |a_{ij}|^2$ *denotes the Frobenius norm of a matrix* $A \in \mathbb{C}^{p\times p}$. *We interpret the first derivative* $\partial_\vartheta \mathscr{C}_\vartheta$ *as a column vector in* $\mathbb{C}^p$ *whereas* $\partial_\vartheta^2 \mathscr{C}_\vartheta$ *is the* $p\times p$*-Jacobian matrix of* $\mathscr{C}_\vartheta$. *Assume further that for any* $a \in \mathbb{R}^p$,

$$a^\top \partial_\vartheta \mathscr{C}_{\vartheta_0} = 0 \quad \nu\text{-a.e.} \quad \Rightarrow \quad a = 0. \tag{3.2.26}$$

Then, under the true parameter ϑ_0 *and for any* $x \in \mathbb{R}$, *we have*

$$\begin{aligned}&\sqrt{T_\varepsilon}(\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) - \vartheta_0)\\ &= \left(J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)\right)\left[\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon)\,dt + \sqrt{T_\varepsilon}(\bar{h}_\varepsilon(x) - \bar{h}(x))\right] + o_{\mathbb{P}}(1),\end{aligned} \tag{3.2.27}$$

as $\varepsilon \to 0$. *Here,* J *is the positive-definite* $p\times p$*-Jacobian matrix of the estimating equation defined by*

$$J(\vartheta_0) := \operatorname{Re}\int_{\mathbb{R}^d} \partial_\vartheta \mathscr{C}_{\vartheta_0}(u)\,\overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}^\top d\nu(u), \tag{3.2.28}$$

while $\bar{h}_\varepsilon$ *and* $\bar{h}$ *are given by*

$$\bar{h}_\varepsilon(x) := \operatorname{Re}\int_{\mathbb{R}^d}\left(\exp\left(iu^\top x\right) - \mathscr{C}_{\varepsilon,\vartheta_0}(u)\right)\overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}\,d\nu(u), \quad x \in \mathbb{R}, \tag{3.2.29}$$

$$\bar{h}(x) := \operatorname{Re}\int_{\mathbb{R}^d}\left(\exp\left(iu^\top x\right) - \mathscr{C}_{\vartheta_0}(u)\right)\overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}\,d\nu(u), \quad x \in \mathbb{R}, \tag{3.2.30}$$

where

$$\mathscr{C}_{\varepsilon,\vartheta_0}(u) = \int_{\mathbb{R}^d} \exp\left(iu^\top x\right)\, d\mu_\varepsilon(\vartheta_0, x), \quad u \in \mathbb{R}^d. \tag{3.2.31}$$

Proof. In the following, we write $\widehat{\vartheta} := \widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon)$ and $\widehat{\mathscr{C}}(\cdot) := \widehat{\mathscr{C}}_{T_\varepsilon}(X^\varepsilon, \cdot)$ for the sake of readability. By our integrability assumptions (3.2.25), we can differentiate under the integral sign in (3.2.4) with respect to ϑ and obtain

$$\begin{aligned}\partial_{\vartheta_i}\Delta_{T_\varepsilon}(\vartheta, X^\varepsilon) &= \partial_{\vartheta_i}\left[\int_{\mathbb{R}^d}(\widehat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{(\widehat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))}\,d\nu(u)\right]\\ &= \int_{\mathbb{R}^d}(-\partial_{\vartheta_i}\mathscr{C}_\vartheta(u))\overline{(\widehat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))} + (\widehat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{(-\partial_{\vartheta_i}\mathscr{C}_\vartheta(u))}\,d\nu(u)\\ &= -2\operatorname{Re}\int_{\mathbb{R}^d}(\widehat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{\partial_{\vartheta_i}\mathscr{C}_\vartheta(u)}\,d\nu(u),\end{aligned}$$

for $i = 1, \ldots, p$, where we used the fact that $w\overline{z} + z\overline{w} = 2\operatorname{Re} w\overline{z}$ for any $w, z \in \mathbb{C}$. Hence, in vector notation, we have

$$\partial_\vartheta \Delta_{T_\varepsilon}(\vartheta, X^\varepsilon) = -2\operatorname{Re}\int_{\mathbb{R}^d}(\widehat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{\partial_\vartheta \mathscr{C}_\vartheta(u)}\,d\nu(u), \quad \vartheta \in \Theta_0^\circ.$$

We define

$$\Psi_\varepsilon(\vartheta) := \operatorname{Re}\int_{\mathbb{R}^d} (\hat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{\partial_\vartheta \mathscr{C}_\vartheta(u)}\, d\nu(u), \quad \vartheta \in \Theta_0^\circ. \tag{3.2.32}$$

Let A_ε be the event on which $\widehat{\vartheta} \in \Theta_0^\circ$ and the line segment $S[\vartheta_0, \widehat{\vartheta}]$ that connects ϑ_0 and $\widehat{\vartheta}$ lies in Θ_0°. Recall that, under the setting of Proposition 3.2.2 with condition (3.2.17), the MDE $\widehat{\vartheta}$ is ε-robust for ϑ_0. Thus, by convexity of Θ_0, we have $\mathbb{P}(A_\varepsilon) \to 1$ as $\varepsilon \to 0$. We also have $\Psi_\varepsilon(\widehat{\vartheta}) = 0$ on A_ε because $\widehat{\vartheta}$ is the minimizer of $\Delta_T(\vartheta, X^\varepsilon)$. By the integrability conditions (3.2.25), the function $\Psi_\varepsilon\colon \Theta_0^\circ \to \mathbb{R}^p$ is differentiable, so that we can apply the mean value theorem around ϑ_0 as follows

$$0 = \Psi_\varepsilon(\widehat{\vartheta}) = \Psi_\varepsilon(\vartheta_0) + \partial_\vartheta \Psi_\varepsilon(\bar{\vartheta}_\varepsilon)(\widehat{\vartheta} - \vartheta_0), \tag{3.2.33}$$

with an intermediate point $\bar{\vartheta}_\varepsilon \in S[\vartheta_0, \widehat{\vartheta}]$. For the partial derivatives of Ψ_ε, we have

$$\partial_{\vartheta_i}\Psi_\varepsilon^{(j)}(\vartheta) = \operatorname{Re}\int_{\mathbb{R}^d} (-\partial_{\vartheta_i}\mathscr{C}_\vartheta(u))\overline{\partial_{\vartheta_j}\mathscr{C}_\vartheta(u)} + (\hat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{\partial_{\vartheta_i}\partial_{\vartheta_j}\mathscr{C}_\vartheta(u)}\, d\nu(u).$$

for $i, j = 1, \ldots, p$. In compact matrix notation, we therefore have

$$\partial_\vartheta \Psi_\varepsilon(\vartheta) = -J(\vartheta) + R_\varepsilon(\vartheta), \tag{3.2.34}$$

with J defined as in (3.2.28) and

$$R_\varepsilon(\vartheta) := \operatorname{Re}\int_{\mathbb{R}^d} (\hat{\mathscr{C}}(u) - \mathscr{C}_\vartheta(u))\overline{\partial_\vartheta^2 \mathscr{C}_\vartheta(u)}\, d\nu(u). \tag{3.2.35}$$

This last remainder term satisfies, by the Cauchy–Schwartz inequality and (3.2.25),

$$\begin{aligned}\|R_\varepsilon(\bar{\vartheta}_\varepsilon)\|_F &\le \| \hat{\mathscr{C}} - \mathscr{C}_{\bar{\vartheta}_\varepsilon} \|_{L^2(\nu)} \| B^{(2)} \|_{L^2(\nu)} \\ &\le \| B^{(2)} \|_{L^2(\nu)} \left(\| \hat{\mathscr{C}} - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} + \| \mathscr{C}_{\bar{\vartheta}_\varepsilon} - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} \right).\end{aligned}$$

Recall here that, in Proposition 3.2.2, we have proved that $\| \hat{\mathscr{C}} - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)}$ converges in probability to zero as $\varepsilon \to 0$. For the other term in the above estimate of the remainder, we have, by the mean value theorem and (3.2.25) again,

$$\begin{aligned}\mathbb{P}\left(\| \mathscr{C}_{\bar{\vartheta}_\varepsilon} - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} > \delta \right) &\le \mathbb{P}\left(\| B^{(1)} \|_{L^2(\nu)} |\bar{\vartheta}_\varepsilon - \vartheta_0|_2 > \delta \right) \\ &\le \mathbb{P}\left(|\widehat{\vartheta} - \vartheta_0|_2 > \delta / \| B^{(1)} \|_{L^2(\nu)} \right) \to 0, \quad \varepsilon \to 0,\end{aligned}$$

for any $\delta > 0$. Here, we also used the ε-robustness of $\widehat{\vartheta}$. Hence,

$$R_\varepsilon(\bar{\vartheta}_\varepsilon) = o_{\mathbb{P}}(1), \quad \varepsilon \to 0. \tag{3.2.36}$$

Observe that $J(\bar{\vartheta}_\varepsilon)$ converges in probability to $J(\vartheta_0)$ by continuity of the map $\Theta_0 \ni \vartheta \mapsto J(\vartheta)$ and $\lim_{\varepsilon\to 0} \bar{\vartheta}_\varepsilon \overset{\mathbb{P}}{=} \vartheta_0$. The condition (3.2.26) implies that $J(\vartheta_0)$ is positive-definite. Indeed, this can be seen from

$$a^\top J(\vartheta_0) a = \operatorname{Re}\int_{\mathbb{R}^d} a^\top \partial_\vartheta \mathscr{C}_{\vartheta_0}(u)\, \overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}^\top a\, d\nu(u)$$

$$= \mathrm{Re} \int_{\mathbb{R}^d} a^\top \partial_\vartheta \mathscr{C}_{\vartheta_0}(u) \overline{a^\top \partial_\vartheta \mathscr{C}_{\vartheta_0}(u)} \, d\nu(u)$$

$$= \int_{\mathbb{R}^d} |a^\top \partial_\vartheta \mathscr{C}_{\vartheta_0}(u)|_2^2 \, d\nu(u), \quad a \in \mathbb{R}^p,$$

which is zero if and only if

$$a^\top \partial_\vartheta \mathscr{C}_{\vartheta_0} = 0 \quad \nu\text{-a.e.}$$

Thus, by assumption, a must be zero. We now use the standard fact that if A_n is a sequence of matrices converging to an invertible matrix A as $n \to \infty$, then A_n must be invertible for sufficiently large n, and $A_n^{-1} \to A^{-1}$ as $n \to \infty$. Since we have justified that

$$\partial_\vartheta \Psi_\varepsilon(\bar{\vartheta}_\varepsilon) = -J(\bar{\vartheta}_\varepsilon) + R_\varepsilon(\bar{\vartheta}_\varepsilon) = -J(\vartheta_0) + o_{\mathbb{P}}(1), \quad \varepsilon \to 0, \tag{3.2.37}$$

this continuity of the matrix inversion yields, on another event B_ε on which $\partial_\vartheta \Psi_\varepsilon(\bar{\vartheta}_\varepsilon)$ is invertible, that $\mathbb{P}(B_\varepsilon) \to 1$ and

$$\partial_\vartheta \Psi_\varepsilon(\bar{\vartheta}_\varepsilon)^{-1} = -J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1). \tag{3.2.38}$$

All that is left to do now is rearrange the terms in (3.2.33) and multiply both sides by $\sqrt{T_\varepsilon}$ to obtain, on $A_\varepsilon \cap B_\varepsilon$,

$$\begin{aligned}
\sqrt{T_\varepsilon}(\widehat{\vartheta} - \vartheta_0) &= -\partial_\vartheta \Psi_\varepsilon(\bar{\vartheta}_\varepsilon)^{-1} \sqrt{T_\varepsilon} \Psi_\varepsilon(\vartheta_0) \\
&= (J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)) \sqrt{T_\varepsilon} \Psi_\varepsilon(\vartheta_0) \\
&= \frac{J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \mathrm{Re} \int_{\mathbb{R}^d} \left(\exp\left(iu^\top X_t^\varepsilon\right) - \mathscr{C}_{\vartheta_0}\right) \overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)} \, d\nu(u) \, dt \\
&= \frac{J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \bar{h}(X_t^\varepsilon) \, dt \\
&= \frac{J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)}{\sqrt{T_\varepsilon}} \left[\int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon) \, dt + \int_0^{T_\varepsilon} \bar{h}(X_t^\varepsilon) - \bar{h}_\varepsilon(X_t^\varepsilon) \, dt\right] \\
&= \left(J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)\right) \left[\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon) \, dt + \sqrt{T_\varepsilon}(\bar{h}_\varepsilon(x) - \bar{h}(x))\right].
\end{aligned}$$

Since $\mathbb{P}(A_\varepsilon \cap B_\varepsilon) \to 1$ as $\varepsilon \to 0$ and the last asymptotic expansion is exact on the event $A_\varepsilon \cap B_\varepsilon$, we obtain the final claim (3.2.27). □

Inspecting the asymptotic representation (3.2.27), it is tempting to use a CLT for the time integral

$$\frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon) \, dt,$$

and conclude asymptotic normality for the MDE. This would be justified by Slutzky's lemma B.1.5 as long as one can guarantee $\sqrt{T_\varepsilon}(\bar{h}_\varepsilon(x) - \bar{h}(x)) = o_{\mathbb{P}}(1)$ at the same time. However, while such a CLT for the time integral would provide us with the asymptotic normality of the MDE, it cannot simply follow from our general setting without imposing assumptions that might be difficult to verify. Instead, we will return to our accompanying example in the following subsection and carefully investigate the limit behavior of the stochastic fluctuation and deterministic error in (3.2.27) with the help of our established limit theorems of Chapter 2 and, especially, Section 2.4.

3.2.3 Multiscale Overdamped Langevin Equation

In this section, we want to prove that the MDE is ε-robust and asymptotically normal when the weakly perturbed model is given by the multiscale overdamped Langevin equation (1.2.10) and the limit model by the homogenized limit equation (1.2.12). As we have seen in more generality in the preceding section, this problem reduces to verifying the assumptions of Theorem 3.2.5 and, ultimately, to proving weak convergence of the stochastic fluctuation and deterministic error figuring in (3.2.27) to a normal distribution. Since we will apply the limit theorems of Section 2.4, which were, for the most parts, restricted to the 1-dimensional case, we will also set $d = 1$ for the Langevin equations.

For the purpose of precision, let us reiterate the statistical framework. Our aim is to estimate the drift parameter $\vartheta := \alpha\mathcal{K} > 0$, $\alpha > 0$, in the homogenized limit equation (1.2.12). Therefore, it is natural to consider the ambient parameter space $\Theta := (0, \infty)$ and a compact interval $\Theta_0 \subset \Theta$. We do not assume that we know the fast-scale periodic oscillation p, but, at least for identifiability reasons of the homogenized limit equation, we must assume that the homogenized diffusion coefficient $\bar{\Sigma}^2 = 2\sigma\mathcal{K}$ is known or has been attained through a prior estimation method, cf. [PS07; KPK13; MP19; Abd+23]. For simplicity, let us assume that $\bar{\Sigma}^2$ is a fixed positive constant. We take the ε-dependent parametric model

$$\mathscr{M}_{T_\varepsilon} := \left(C([0,T_\varepsilon];\mathbb{R}), \mathscr{B}(C([0,T_\varepsilon];\mathbb{R})), \left\{\mathbb{P}^{T_\varepsilon}_{\psi(\vartheta)} \,|\, \vartheta \in \Theta_0\right\}\right), \quad \varepsilon > 0, \tag{3.2.39}$$

where $\mathscr{B}(C([0,T_\varepsilon];\mathbb{R}))$ is the Borel σ-algebra of $C([0,T_\varepsilon];\mathbb{R})$, $T_\varepsilon \to \infty$ as $\varepsilon \to 0$, and $\psi(\vartheta) = \vartheta/\mathcal{K}$. The set of probability measures in this parametric model serves as a distributional model for X^ε. Let us also recall the invariant densities of X^ε and X from Corollary 2.4.4

$$\mu_\varepsilon(\vartheta, x) = \frac{1}{Z_{\varepsilon,\vartheta}} \exp\left(-\frac{\psi(\vartheta)}{\sigma}V(x) - \frac{p\,(x/\varepsilon)}{\sigma}\right), \qquad x \in \mathbb{R}, \quad \varepsilon > 0,$$

$$\mu(\vartheta, x) = \frac{1}{Z_\vartheta} \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2}V(x)\right), \qquad x \in \mathbb{R},$$

where we obviously abuse the notation regarding μ and μ_ε from the preceding Subsection 3.2.1, but we will stick to the notation at least for this subsection.

Before we analyze the MDE under this statistical framework, we need to establish the following short lemma which gives weak convergence rates of the invariant measures of X^ε and X in terms of their characteristic functions. Although we already know that they converge weakly by Corollary 2.4.4 and the fact that $C_c^\infty(\mathbb{R}^d)$, the space of compactly supported, infinitely differentiable real-valued functions on $\mathbb{R}^d$, is a convergence determining class for probability measures on $\mathbb{R}^d$, see [Bog18, Theorem 1.5.3], we still need such a refined estimate for the proof of asymptotic normality later.

Lemma 3.2.6. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L). Then, for any $k \in \mathbb{N}$, $u \in \mathbb{R}$, and $\vartheta \in \Theta$, we have*

$$|\mathscr{C}_{\varepsilon,\vartheta}(u) - \mathscr{C}_\vartheta(u)| = \mathcal{O}\left(\sum_{n=0}^{k}(1+|u|)^n\varepsilon^k\right), \quad \varepsilon \to 0. \tag{3.2.40}$$

Proof. The estimate (2.4.22) at the end of the proof of Corollary 2.4.4 applied to the function $g(x) := \exp(iux)$ with $u, x \in \mathbb{R}$ gives

$$\begin{aligned}
&|\mathscr{C}_{\varepsilon,\vartheta}(u) - \mathscr{C}_\vartheta(u)| \\
&\le \frac{2\varepsilon^k}{Z_\vartheta \Pi^-(2\pi)^k} \exp\left(\frac{1}{\sigma}\sum_{l\in\mathbb{Z}} |\widehat{p}(l)|\right) \left\| g \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{W^{k,1}(\mathbb{R})} \\
&\quad + \frac{2\varepsilon^k}{Z_\vartheta^2 \Pi^-(2\pi)^k} \exp\left(\frac{1}{\sigma}\sum_{l\in\mathbb{Z}} |\widehat{p}(l)|\right) \left\| \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{W^{k,1}(\mathbb{R})} \left\| g \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{L^1(\mathbb{R})},
\end{aligned}$$

for sufficiently small $\varepsilon > 0$, where Π^- was defined in (1.2.16). Using Leibniz's differentiation rule, Hölder's inequality, and the fact that $|\partial_x^j g| \le |u|^j$ on $\mathbb{R}$ for $j = 0, \ldots, k$, yields

$$\begin{aligned}
\left\| g \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{W^{k,1}(\mathbb{R})} &\le \sum_{n=0}^{k}\sum_{j=0}^{n} \binom{n}{j} |u|^{n-j} \left\| \partial_x^j \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{L^1(\mathbb{R})} \\
&\le \sum_{n=0}^{k} (1+|u|)^n \left\| \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{W^{k,1}(\mathbb{R})},
\end{aligned}$$

and therefore

$$\begin{aligned}
|\mathscr{C}_{\varepsilon,\vartheta}(u) - \mathscr{C}_\vartheta(u)| \le{}& \frac{2}{(Z_\vartheta \wedge Z_\vartheta^2)\Pi^-(2\pi)^k} \exp\left(\frac{1}{\sigma}\sum_{l\in\mathbb{Z}} |\widehat{p}(l)|\right) \left\| \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{W^{k,1}(\mathbb{R})} \\
&\times \left(1 + \left\| \exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2} V\right)\right\|_{L^1(\mathbb{R})}\right) \sum_{n=0}^{k} (1+|u|)^n \varepsilon^k,
\end{aligned}$$

producing the claimed estimate. □

Proposition 3.2.7. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L). Let Θ_0 be a compact interval in $\Theta = (0, \infty)$ and ν a probability measure on $\mathbb{R}$ with* $\operatorname{supp}(\nu) = \mathbb{R}$. *Then, under the true parameter $\vartheta_0 \in \Theta_0$, the MDE $(\widehat{\vartheta}_{T_\varepsilon})$ is ε-robust for ϑ_0.*

Proof. We will prove the validity of the assumptions in Proposition 3.2.2 where we exchange (3.2.13) with (3.2.17), recall Remark 3.2.3. Concerning Assumptions (MDE), the ergodic theorem for X as stated in v) is essentially a special case of Theorem 2.1.10 but is anyway classical and can be found in [GS68; Sun89; DZ96; MSH02]. Condition vi) follows from Scheffé's theorem, see for example [Bog07, Theorem 2.8.9], because $\Theta \ni \vartheta \mapsto \mu(\vartheta, x)$, $x \in \mathbb{R}$, is continuous. The remaining assumptions in (MDE) are obvious from our setting.

Convergence of the invariant measures follows from Lemma 3.2.6 and Lévy's continuity theorem, see, for example, [Bog18, Theorem 1.6.6]. Condition (3.2.17) is a consequence of Corollary 2.4.11. We have to remark here that, in the proof of Proposition 3.2.2, condition (3.2.17) was only applied to the function $x \mapsto \exp(iux)$, $u \in \mathbb{R}$, which satisfies the requirements of Corollary 2.4.11. In other words, a mean ergodic theorem as in Corollary

2.4.11 is already sufficient for the proof of Proposition 3.2.2.

It remains to prove the separation condition (3.2.14) and the ν-identifiability condition (3.2.15). To accomplish this, we assume that there exists a $\delta^* > 0$ such that (3.2.14) is not satisfied. This means that there exists a $\vartheta^* \in \Theta_0$ such that $|\vartheta^* - \vartheta_0| > \delta^*$ and $\| \mathscr{C}_{\vartheta^*} - \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)} = 0$. This is possible by using a minimizing sequence and the fact that Θ_0 is compact. Thus, the characteristic functions $\mathscr{C}_{\vartheta^*}$ and $\mathscr{C}_{\vartheta_0}$ coincide ν-a.e. on $\mathbb{R}$. In particular, they must coincide on $\operatorname{supp}(\nu) = \mathbb{R}$ by continuity. This is because if $u \in \operatorname{supp}(\nu)$ such that $\mathscr{C}_{\vartheta^*}(u) \neq \mathscr{C}_{\vartheta_0}(u)$, then, by continuity of the characteristic functions and the support definition, there must be an open interval $U \ni u$ for which $\mathscr{C}_{\vartheta^*} \neq \mathscr{C}_{\vartheta_0}$ on U and $\nu(U) > 0$, contradicting $\mathscr{C}_{\vartheta^*} = \mathscr{C}_{\vartheta_0}$ ν-a.e. Hence, they must indeed coincide everywhere, which, in turn, means that the densities $\mu(\vartheta^*)$ and $\mu(\vartheta_0)$ coincide Lebesgue-a.e. and, thus, everywhere by continuity. This implies

$$\exp\left(-\frac{2(\vartheta^* - \vartheta_0)}{\bar{\Sigma}^2} V(x)\right) = \frac{Z_{\vartheta^*}}{Z_{\vartheta_0}}, \quad x \in \mathbb{R}.$$

Now, choose any two distinct points $x_1, x_2 \in \mathbb{R}$ for which $V(x_1) \neq V(x_2)$ so that

$$\exp\left(-\frac{2(\vartheta^* - \vartheta_0)}{\bar{\Sigma}^2} (V(x_1) - V(x_2))\right) = 1,$$

from which it easily follows that $\vartheta^* = \vartheta_0$, a contradiction. Hence, (3.2.14) holds. The ν-identifiability condition (3.2.15) follows by the same arguments. Therefore, Proposition 3.2.2 with condition (3.2.17) implies that the MDE $(\widehat{\vartheta}_{T_\varepsilon})$ is ε-robust for ϑ_0. □

Remark 3.2.8. *We assumed that the weight measure ν has a support that equals the whole real line. This was only used to obtain the separation condition* (3.2.14) *and the ν-identifiability condition* (3.2.15)*, but it is possible to somewhat weaken this assumption in exchange for more technical arguments. To give an example. Consider a probability measure ν on $\mathbb{R}^d$ such that* $\operatorname{supp}(\nu) = \{u \in \mathbb{R} \,|\, \forall \delta > 0 : \nu(B_\delta(u)) > 0\}$ *has an accumulation point a_0 in $\mathbb{R}$. If $\mathscr{C}_{\vartheta_1} = \mathscr{C}_{\vartheta_2}$ ν-a.e., then $D := \{u \in \mathbb{R} \,|\, \mathscr{C}_{\vartheta_1}(u) = \mathscr{C}_{\vartheta_2}(u)\}$ has full ν-measure. Since a_0 is an accumulation point of* $\operatorname{supp}(\nu)$*, there exists a sequence $(a_n) \subset$* $\operatorname{supp}(\nu)$ *such that $a_n \neq a_0$ and $a_n \to a_0$. Because $\nu(D) = 1$ and $a_n \in$* $\operatorname{supp}(\nu)$*, we know that $D \cap B_{\frac{1}{2}|a_n - a_0|}(a_n) \neq \emptyset$. Take a sequence $\tilde{a}_n \in D \cap B_{\frac{1}{2}|a_n - a_0|}(a_n)$. This is another sequence approximating the accumulation point a_0 and, more importantly, along which $\mathscr{C}_{\vartheta_1}$ and $\mathscr{C}_{\vartheta_2}$ coincide. Now, if $\mathscr{C}_\vartheta$, $\vartheta \in \Theta$, is a real-analytic function, then the identity theorem for analytic functions, see, for example, [Sas13, Theorem C.1.8], implies that $\mathscr{C}_{\vartheta_1}$ and $\mathscr{C}_{\vartheta_2}$ must coincide on all of $\mathbb{R}$. The rest of the proof then works as before.*

Theorem 3.2.9. *Consider the multiscale overdamped Langevin equation* (1.2.10) *and its homogenized limit equation* (1.2.12) *under Assumptions (L). Let Θ_0 be a compact interval in $\Theta = (0, \infty)$ and ν a probability measure on $\mathbb{R}$ with* $\operatorname{supp}(\nu) = \mathbb{R}$*. For $k \in \mathbb{N}$, let T_ε be such that $T_\varepsilon = \mathcal{O}(\varepsilon^{-\eta})$ for some $\eta > 0$ and $T_\varepsilon = o(\varepsilon^{-2k})$ as $\varepsilon \to 0$. Then, under the true parameter $\vartheta_0 \in \Theta_0^\circ$, the MDE $(\widehat{\vartheta}_{T_\varepsilon})$ is asymptotically normal, i.e.,*

$$\sqrt{T_\varepsilon}\left(\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) - \vartheta_0\right) \xrightarrow{\mathcal{D}} J(\vartheta_0)^{-1} \mathcal{N}_1(0, \tau^2(\vartheta_0)), \quad \text{as } \varepsilon \to 0, \tag{3.2.41}$$

with

$$J(\vartheta_0) = \| \partial_\vartheta \mathscr{C}_{\vartheta_0} \|_{L^2(\nu)}^2, \quad \tau^2(\vartheta_0) := \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi'_{\vartheta_0}(x)^2 \mu(\vartheta_0, x)\, dx.$$

The function Φ_{ϑ_0} is given by

$$\Phi_{\vartheta_0}(x) := -\int_0^x \frac{2}{\bar{\Sigma}^2\mu(\vartheta_0, y)} \int_{-\infty}^y \bar{h}(z)\mu(\vartheta_0, z)\, dz\, dy, \quad x \in \mathbb{R}, \tag{3.2.42}$$

where $\bar{h}$ was defined in (3.2.30).

Proof. Since we already proved that $(\widehat{\vartheta}_{T_\varepsilon})$ is ε-robust for ϑ_0 in Proposition 3.2.7, we want to verify the remaining assumptions of Theorem 3.2.5 and then prove weak convergence of the RHS in (3.2.27). First, recall the following estimate (2.4.5) from the proof of Lemma 2.4.2

$$V(x) \geq m + \frac{b}{2}(|x|^2 - R^2) - a\log\left(\frac{|x|}{R}\right), \quad |x| \geq R.$$

For $\vartheta \in \Theta_0$, this implies the following estimate

$$\exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2}V(x)\right) \leq \frac{2\vartheta}{\bar{\Sigma}^2 R}\exp\left(\frac{\vartheta}{\bar{\Sigma}^2}(2|m| + bR^2)\right)|x|^{2\vartheta/\bar{\Sigma}^2}\exp\left(-\frac{\vartheta b}{\bar{\Sigma}^2}|x|^2\right). \tag{3.2.43}$$

From this upper bound and the compactness of Θ_0, we can obtain a ϑ-independent dominating function for $V\mu(\vartheta,\cdot)$, $\partial_\vartheta\mu(\vartheta,\cdot)$, and $\partial_\vartheta^2\mu(\vartheta,\cdot)$, enabling us to differentiate under integral signs, so that $\vartheta \mapsto \mathscr{C}_\vartheta$ and $\vartheta \mapsto Z_\vartheta$ are twice continuously differentiable with respect to ϑ. A calculation reveals that

$$\partial_\vartheta\, \mu(\vartheta, x) = -\mu(\vartheta, x)\left(\frac{2}{\bar{\Sigma}^2}V(x) + \frac{\partial_\vartheta Z_\vartheta}{Z_\vartheta}\right), \tag{3.2.44}$$

$$\partial_\vartheta^2\, \mu(\vartheta, x) = \mu(\vartheta, x)\left[\left(\frac{2}{\bar{\Sigma}^2}V(x) + \frac{\partial_\vartheta Z_\vartheta}{Z_\vartheta}\right)^2 - \frac{\partial_\vartheta^2 Z_\vartheta}{Z_\vartheta} + \left(\frac{\partial_\vartheta Z_\vartheta}{Z_\vartheta}\right)^2\right], \tag{3.2.45}$$

and

$$\partial_\vartheta Z_\vartheta = -\frac{2}{\bar{\Sigma}^2}\int_{\mathbb{R}} V(y)\exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2}V(y)\right) dy, \tag{3.2.46}$$

$$\partial_\vartheta^2 Z_\vartheta = \left(\frac{2}{\bar{\Sigma}^2}\right)^2\int_{\mathbb{R}} V(y)^2\exp\left(-\frac{2\vartheta}{\bar{\Sigma}^2}V(y)\right) dy. \tag{3.2.47}$$

By the above arguments, we obtain the integrability conditions (3.2.25) as follows

$$\sup_{\vartheta\in\Theta_0^\circ}|\partial_\vartheta\mathscr{C}_\vartheta(u)| \leq \sup_{\vartheta\in\Theta_0^\circ}\int_{\mathbb{R}}|\partial_\vartheta\mu(\vartheta, x)|\, dx =: B^{(1)}(u) < \infty,$$

$$\sup_{\vartheta\in\Theta_0^\circ}|\partial_\vartheta^2\mathscr{C}_\vartheta(u)| \leq \sup_{\vartheta\in\Theta_0^\circ}\int_{\mathbb{R}}|\partial_\vartheta^2\mu(\vartheta, x)|\, dx =: B^{(2)}(u) < \infty.$$

Since $B^{(1)}$ and $B^{(2)}$ do not depend on u, they must belong to $L^2(\nu)$. It remains to prove condition (3.2.26). For this, note that $u \mapsto \partial_\vartheta\mathscr{C}_{\vartheta_0}(u) \in \mathscr{S}(\mathbb{R})$ because $\partial_\vartheta\mathscr{C}_{\vartheta_0} = 2\pi\mathscr{F}(\partial_\vartheta\mu(\vartheta_0, (-2\pi)\cdot))$, where $\mathscr{F}$ is the Fourier transform operator, and $\partial_\vartheta\mu(\vartheta_0,\cdot) \in$

$\mathscr{S}(\mathbb{R})$ by Lemma 2.4.2. By continuity and $\operatorname{supp}(\nu) = \mathbb{R}$, we obtain $J(\vartheta_0) > 0$, which implies (3.2.26). Hence, Theorem 3.2.5 applies and we get the asymptotic representation

$$\begin{aligned}&\sqrt{T_\varepsilon}(\widehat{\vartheta}_{T_\varepsilon}(X^\varepsilon) - \vartheta_0)\\&= \left(J(\vartheta_0)^{-1} + o_{\mathbb{P}}(1)\right)\left[\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon)\,dt + \sqrt{T_\varepsilon}(\bar{h}_\varepsilon(x) - \bar{h}(x))\right] + o_{\mathbb{P}}(1),\end{aligned} \tag{3.2.48}$$

for $\varepsilon \to 0$, where $x \in \mathbb{R}$ is arbitrary. We use Lemma 3.2.6 to estimate

$$\begin{aligned}|\bar{h}_\varepsilon(x) - \bar{h}(x)| &= \left|\operatorname{Re}\int_{\mathbb{R}^d} (\mathscr{C}_{\varepsilon,\vartheta_0}(u) - \mathscr{C}_{\vartheta_0}(u))\,\overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}\,d\nu(u)\right|\\&= \mathcal{O}\left(\int_{\mathbb{R}}\sum_{n=0}^{k}(1+|u|)^n|\mathscr{C}_{\vartheta_0}(u)|\,d\nu(u)\,\varepsilon^k\right) = \mathcal{O}\left(\varepsilon^k\right),\end{aligned}$$

which holds because $\sup_{u\in\mathbb{R}}\sum_{n=0}^{k}(1+|u|)^n|\mathscr{C}_{\vartheta_0}(u)| < \infty$ by virtue of $u \mapsto \mathscr{C}_{\vartheta_0}(u) \in \mathscr{S}(\mathbb{R})$. Thus, by our assumption $T_\varepsilon = o(\varepsilon^{-2k})$, we obtain

$$\sqrt{T_\varepsilon}(\bar{h}_\varepsilon(x) - \bar{h}(x)) = o(1), \quad \varepsilon \to 0. \tag{3.2.49}$$

It remains to prove the weak convergence

$$\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon} \bar{h}_\varepsilon(X_t^\varepsilon)\,dt \xrightarrow{\mathcal{D}} \mathcal{N}_1(0, \tau^2(\vartheta_0)), \quad \varepsilon \to 0. \tag{3.2.50}$$

If we define the function

$$h(x) := \operatorname{Re}\int_{\mathbb{R}^d} \exp(iux)\overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}\,d\nu(u), \quad x \in \mathbb{R},$$

then it holds that

$$\begin{aligned}\bar{h}_\varepsilon(x) &= h(x) - \int_{\mathbb{R}} h(y)\mu_\varepsilon(\vartheta_0, y)\,dy,\\ \bar{h}(x) &= h(x) - \int_{\mathbb{R}} h(y)\mu(\vartheta_0, y)\,dy,\end{aligned}$$

after applying Fubini's theorem. Moreover, $h \in C^\infty$ is together with all of its derivatives bounded on $\mathbb{R}$ by the dominated convergence theorem and $u \mapsto \mathscr{C}_{\vartheta_0}(u) \in \mathscr{S}(\mathbb{R})$. In particular, the boundedness of all derivatives implies $h\exp(-V) \in \mathscr{S}(\mathbb{R})$ by Lemma 2.4.2. As a result, Corollary 2.4.14 applies directly to this case, so that (3.2.50) indeed holds. The asymptotic normality of the MDE now follows from combining (3.2.49) and (3.2.50) with Slutzky's lemma B.1.5 and tightness of the weakly convergent sequences in the asymptotic representation (3.2.48). □

Remark 3.2.10. *If we consider the multiscale overdamped Langevin equation and its homogenized limit equation in multiple dimensions, then the proof of robustness of the MDE with sequential limit orders, that is, Proposition 3.2.2 and 3.2.4, may be obtained with existing results. For example, the required mean ergodic theorems in (MDE) v) and* (3.2.13) *are well established in the literature [DZ96; KMN12; MT59]. The more difficult* *conditions would be the separation condition* (3.2.14) *and the moment bounds* (3.2.19) *that are uniform in ε. At least the latter one could potentially be obtained from corrected Lyapunov functions, see [BP78] for this idea. However, for an asymptotic normality result such as Theorem 3.2.9, we would need ε-robustness of the MDE, which remains unclear due to the lack of an ε-dependent mean ergodic and central limit theorem in higher dimensions. As we have outlined in various remarks in Chapter 2, the aforementioned limit theorems would be highly valuable in multiple dimensions but fairly difficult to obtain rigorously.*

3.2.4 Numerical Experiments

In this subsection, we will, on the one hand, substantiate the theoretical findings with numerical simulations for different example pairs of a weakly perturbed model and a limit model and, on the other hand, investigate how the MDE performs empirically in a setting which is not yet fully covered by the provided theory. The simulation results have been realized in the programming language Julia, see [Bez+17], and the code with full documentation can be accessed publicly through the GitHub repository [Bor25].

Implementation details. In order to validate robustness and asymptotic normality through a simulation, we have to generate realizations of the path observations from the weakly perturbed model. In almost all cases, we will be simulating realizations from SDE solutions for which we employ the standard Euler–Maruyama scheme, see [KP92, Chapter 9] or [PB10, Chapter 5]. For this, we fix an initial value $z_0 \in \mathbb{R}^d$, $d \in \mathbb{N}$, and a time discretization step $\Delta t > 0$ such that we have an equidistant time interval partition $0 =: t_0 < t_1 < \cdots < t_N := T$ with $t_{i+1} - t_i = \Delta t$ for all $i = 0, \ldots, N-1$ and $N = \lceil T/\Delta t \rceil$. If Z is the solution to the generic d-dimensional SDE

$$dZ_t = b(Z_t)\,dt + c(Z_t)\,dW_t, \quad t \in [0,T], \quad Z_0 = z_0,$$

then the Euler–Maruyama approximation is given by

$$Z_{i+1} = Z_i + b(Z_i)\,\Delta t + c(Z_i)(W_{i+1} - W_i), \quad i = 0, \ldots, N-1,$$

where $Z_i := Z_{i\Delta t}$, $W_i := W_{i\Delta t}$, and $W_{i+1} - W_i \sim \sqrt{\Delta t}\,\mathcal{N}_1(0,1)$ for all $i = 0, \ldots, N-1$. This approximation produces a set of discrete observation points $(Z_i)_{i=0}^N$, which we will plug in into the MDE. Under certain regularity conditions on the coefficients of the SDE, the Euler–Maruyama approximation achieves a strong order of convergence of 1/2 and a weak order of convergence of 1. Since our theory is built upon continuous stochastic processes and ergodicity, we will be considering T large and Δt small, which corresponds to high-frequency observations over a long time horizon.

Regarding the MDE implementation, we will suppose a setting as in Example 3.2.1 where we choose the weight measure ν according to (3.2.7), but we also assume that the invariant measure $\mu(\vartheta)$ is absolutely continuous with respect to Lebesgue measure and, by abuse of notation, denote the corresponding invariant density by $\mu(\vartheta)$ again. With these assumptions we get the simplified distance formula

$$\Delta_T(\vartheta, X^\varepsilon) = \frac{1}{T^2} \int_0^T \int_0^T k_\beta(X_t^\varepsilon - X_s^\varepsilon)\,dt\,ds$$

$$-\frac{2}{T}\int_0^T (k_\beta * \mu(\vartheta))(X_t^\varepsilon)\,dt + \int_{\mathbb{R}^d}(k_\beta * \mu(\vartheta))(x)\mu(\vartheta,x)\,dx,$$

where we recall $k_\beta(x) = \exp(-\beta^2|x|_2^2/2)$, $x \in \mathbb{R}^d$, with $\beta > 0$ and X^ε comes from a weakly perturbed model. Since we are optimizing with respect to ϑ and the double time integral above does not depend on ϑ, we can safely discard that term in the numerical simulations. The computational implementation of the distance quantity is straightforward and realized through the discretized distance formula

$$\Delta_N(\vartheta,(X_i^\varepsilon)) := -\frac{2}{N+1}\sum_{i=0}^N (k_\beta * \mu(\vartheta))(X_i^\varepsilon) + \int_{\mathbb{R}^d}(k_\beta * \mu(\vartheta))(x)\mu(\vartheta,x)\,dx, \tag{3.2.51}$$

with $(X_i^\varepsilon) := (X_i^\varepsilon)_{i=0}^N$ produced from the Euler–Maruyama approximation of a weakly perturbed model equation. If no analytical expression is available for the space integral in (3.2.51), then we use an adaptive integration rule to approximate this integral for the given parameters. Specifically, we used the Gauss-Kronrod quadrature rule, see [DR84], that is available in the Julia package Cubature.jl [Joh05]. When it comes to the convolution terms in the sum, then a naive direct convolution would be very expensive to evaluate, especially, considering that N is quite large in our cases. Hence, we use the fast Fourier transform algorithm, see, for example, [TAL97], to evaluate the discrete-time convolution as the inverse discrete-time Fourier transform of the product of the individual discrete-time Fourier transforms. In practice, we need to add a sufficiently large cutoff of the functions k_β and $\mu(\vartheta)$ for given parameters β and ϑ in order to avoid erroneous artifacts caused by non-negligible tails. We also need to include zero-padding so that we obtain the true linear convolution instead of the circular convolution. We implemented this step through the Julia package DSP.jl [Kor+25]. Furthermore, since the convolution outputs are given on a deterministic spatial grid, but we require the convolution evaluated at the observation points $(X_i^\varepsilon)_{i=1}^N$, we additionally applied a linear interpolation at these points.

Concerning the actual optimization task (3.2.5), we used different gradient descent algorithms from the Julia package Optim.jl [MR18], such as the limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS) and the interior-point Newton method (IPNewton), see [NW06, Chapter 7, Chapter 19]. For both of these methods, we need gradients of the cost functional with respect to ϑ, i.e., $\partial_\vartheta \Delta_N(\vartheta,(X_i^\varepsilon))$, which we will either compute with the forward mode of automatic differentiation, cf. [NW06, Chapter 8], during optimization or implement directly through the gradient formula

$$\Delta_N(\vartheta,(X_i^\varepsilon)) := -\frac{2}{N+1}\sum_{i=0}^N (k_\beta * \partial_\vartheta\mu(\vartheta))(X_i^\varepsilon) + \int_{\mathbb{R}^d}(k_\beta * \partial_\vartheta\mu(\vartheta))(x)\mu(\vartheta,x)\,dx, \tag{3.2.52}$$

for whose computation we can use the numerical approaches described in the preceding paragraph. In any case, we will specify the exact parameter constellations separately for each considered example and simulation.

As we have realized in Example 3.2.1, the distance formula simplifies drastically in the Gaussian setting, i.e., when the invariant density $\mu(\vartheta)$ is given by the density $\varphi_{Q(\vartheta)}$ of a centered multivariate normal distribution with a positive-definite covariance matrix

$Q(\vartheta) \in \mathbb{R}^{d\times d}$. Likewise, the discretized formula (3.2.51) simplifies to

$$\Delta_N(\vartheta, (X_i^\varepsilon)) = -\frac{2}{(N+1)\sqrt{\det(I_d + \beta^2 Q(\vartheta))}} \sum_{i=0}^{N} \exp\left(-\frac{\beta^2}{2}(X_i^\varepsilon)^\top \left(I_d + \beta^2 Q(\vartheta)\right)^{-1} X_i^\varepsilon\right) + \frac{1}{\sqrt{\det(I_d + 2\beta^2 Q(\vartheta))}}, \tag{3.2.53}$$

which now only involves a simple sum, allowing for very fast evaluation and, hence, parameter estimation.

Multiscale overdamped Langevin equation in one dimension. We are in the setting of Subsection 3.2.3 and consider the 1-dimensional multiscale overdamped Langevin equation (1.2.10) with either a quadratic potential (QdP1) or a quartic potential (QrP)

$$V(x) = \frac{1}{2}x^2, \tag{QdP1}$$

$$V(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2, \tag{QrP}$$

and a fast scale 2π-periodic oscillation $p = \sin$. When choosing the quadratic potential (QdP1), the homogenized limit SDE process X, defined through (1.2.12), is a simple Ornstein–Uhlenbeck process with invariant density given by

$$\mu(\vartheta, x) = \frac{\sqrt{\vartheta}}{\sqrt{\pi}\bar{\Sigma}} \exp\left(-\frac{\vartheta}{\bar{\Sigma}^2}x^2\right),$$

and when choosing the quartic potential the invariant density equals

$$\mu(\vartheta, x) = \frac{1}{Z_\vartheta} \exp\left(-\frac{\vartheta}{2\bar{\Sigma}^2}x^4 + \frac{\vartheta}{\bar{\Sigma}^2}x^2\right), \quad Z_\vartheta = \int_{\mathbb{R}} \exp\left(-\frac{\vartheta}{2\bar{\Sigma}^2}y^4 + \frac{\vartheta}{\bar{\Sigma}^2}y^2\right) dy.$$

Recall here that we want to estimate the true homogenized drift coefficient $\vartheta_0 = \alpha_0\mathcal{K}$ from observations (X_i^ε), while assuming that we know the homogenized diffusion coefficient $\bar{\Sigma}^2 = 2\sigma\mathcal{K}$. Hence, the optimization is performed with respect to $\vartheta \in (0, \infty)$. The specific parameter constellation we have used for this simulation is as follows.

- Multiscale and homogenized SDE parameters:

$$\alpha_0 = 2, \quad \sigma = 1, \quad \vartheta_0 \approx 1.248, \quad \bar{\Sigma}^2 \approx 1.248, \quad X_0^\varepsilon = 10.$$

- Small scale parameter and time parameters:

$$(\text{QdP1}) : \varepsilon \in \{0.25, 0.1\}, \quad T \in \{50m \,|\, m = 1, \dots, 40\}, \quad \Delta t = \varepsilon^3;$$

$$(\text{QrP}) : \varepsilon = 0.1, \quad T \in \{100m \,|\, m = 1, \dots, 20\}, \quad \Delta t = \varepsilon^3.$$

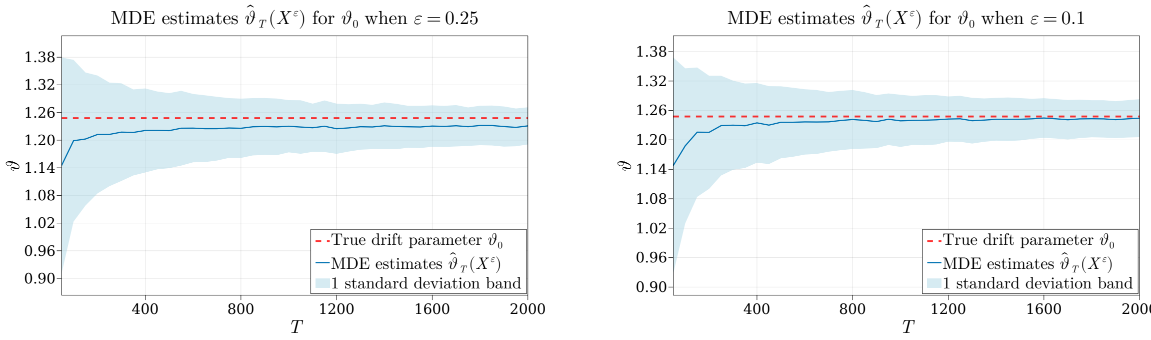


Figure 3.2: MDE estimates $\widehat{\vartheta}_T(X^\varepsilon)$ for ϑ_0 when the observations (X_i^ε) come from the Euler–Maruyama approximation of the multiscale overdamped Langevin equation (1.2.10) with $d = 1$, a fast scale 2π-periodic oscillation $p = \sin$, and a linear drift term corresponding to (QdP1). The left figure shows the estimation results for $\varepsilon = 0.25$ and the right figure for $\varepsilon = 0.1$.

- Gradient descent algorithm parameters:

$$\beta = 1, \quad \text{initial point: } \vartheta_{\text{initial}} = 10, \quad \text{box interval: } (0, \infty), \quad \text{optimizer: L-BFGS.}$$

Note that we make a couple of distinctions depending on the chosen drift term. This is because the optimization is numerically more involved for the quartic potential (QrP). To be more precise, for the quartic potential, we optimize the discretized distance formula (3.2.51), while for the quadratic potential we use the simplified formula (3.2.53). For the optimization itself, we use the L-BFGS algorithm with box-constraints from Optim.jl [MR18], which require an initial starting point $\vartheta_{\text{initial}} \in (0, \infty)$ and the gradient of the cost functional with respect to ϑ in (3.2.52).

Figure 3.2 and 3.3 show estimation results on the robustness of the MDE in the case (QdP1) and (QrP). Here, every single estimated value is, in fact, an average of the same experiment repeated for 1000 times. We can observe that the MDE succeeds in estimating the true parameter ϑ_0 for decreasing ε and increasing T. This is in agreement with the established theoretical results of Subsection 3.2.3.

Before we move on to the next example, we briefly want to draw attention to numerical results regarding the asymptotic normality of the MDE. We consider the case (QdP1) with the same parameter configuration as before but this time only for $\varepsilon = 0.1$ and $T \in \{100, 1000\}$. The two plots in Figure 3.4 show normalized histograms of 1000 estimates, together with the (approximated) density of the asymptotic normal distribution of Proposition 3.2.9. The histogram bin width has been chosen according to the Freedman–Diaconis rule, see [FD81], and normalized means that the total area of the bars equals 1, which allows for a proper comparison with a probability density function. To compare the MDE estimates with the density of the asymptotic normal distribution, we need the asymptotic variance of that normal distribution, for which we can calculate the following formulas. Since $\mu(\vartheta_0)$ is the density of a centered normal distribution with variance $\bar{\Sigma}^2/(2\vartheta_0)$, the characteristic function equals $\mathscr{C}_{\vartheta_0}(u) = \exp(-\bar{\Sigma}^2 u^2/(4\vartheta))$, $u \in \mathbb{R}$,

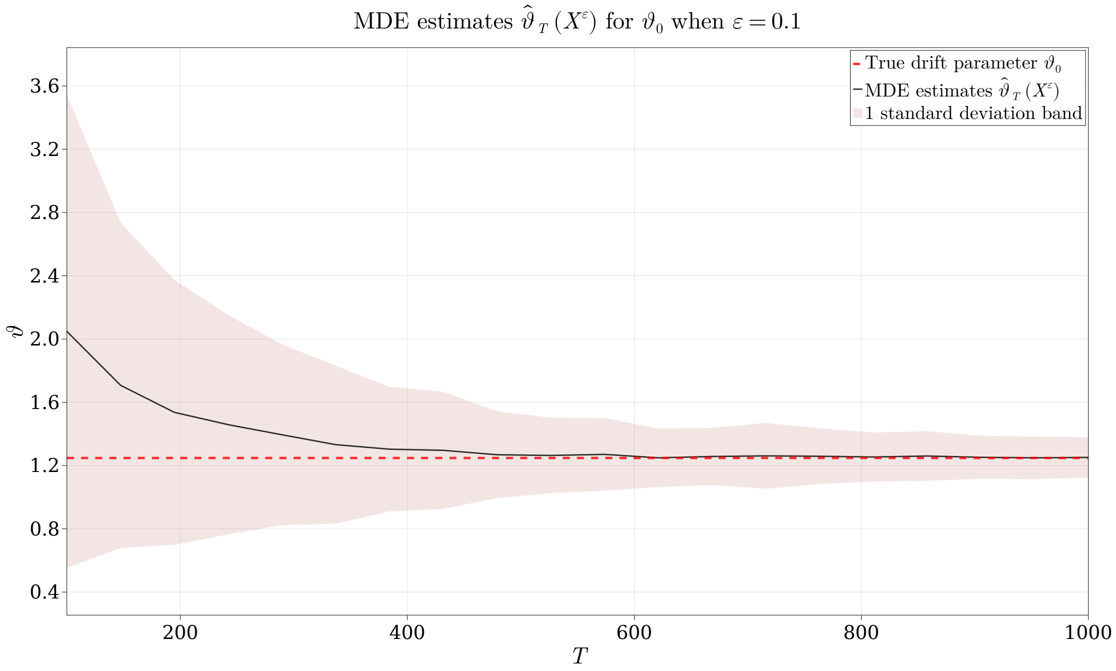


Figure 3.3: MDE estimates $\widehat{\vartheta}_T(X^\varepsilon)$ for ϑ_0 when the observations (X_i^ε) come from the Euler–Maruyama approximation of the multiscale overdamped Langevin equation (1.2.10) with $d = 1$, a fast scale 2π-periodic oscillation $p = \sin$, and a nonlinear drift term corresponding to (QrP).

from which we obtain

$$J(\vartheta_0) = \int_{\mathbb{R}} |\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)|^2 \varphi_\beta(u)\,du = \frac{3}{4\beta}\left(\frac{\bar{\Sigma}^2}{2\vartheta_0^2}\right)^2 \sigma_1^5, \quad \sigma_1^2 := \frac{\beta^2\vartheta_0}{\vartheta_0 + \bar{\Sigma}^2\beta^2},$$

after a short calculation, and

$$\begin{aligned}\bar{h}(y) &= \operatorname{Re}\int_{\mathbb{R}^d}\left(\exp\left(iu^\top y\right) - \mathscr{C}_{\vartheta_0}(u)\right)\overline{\partial_\vartheta \mathscr{C}_{\vartheta_0}(u)}\,\varphi_\beta(u)\,du \\ &= \frac{\bar{\Sigma}^2}{4\vartheta_0^2\beta}\left[\sigma_2^3(1-\sigma_2^2 z^2)\exp\left(-\frac{\sigma_2^2 z^2}{2}\right) - \sigma_1^3\right], \quad \sigma_2^2 := \frac{\beta^2\vartheta_0}{\vartheta_0 + \bar{\Sigma}^2\beta^2/2},\end{aligned}$$

for $y \in \mathbb{R}$. We have approximated $\tau^2(\vartheta_0)$ with the help of formula (3.2.42) and the Gauss-Kronrod quadrature rule. In this particular case, we computed an approximate value of $\tau^2(\vartheta_0)/J(\vartheta_0)^2 \approx 2.670$ for the asymptotic variance. As we can see from Figure 3.4, the results are in relatively good agreement with the theoretical results.

Multiscale overdamped Langevin equation in two dimensions. Consider the equations (1.2.10) and (1.2.12) for $d = 2$ with

$$V(x) = \frac{1}{2}x^\top Mx, \quad p(x) := p_1(x_1) + p_2(x_2) := \sin(x_1) + \frac{1}{2}\sin(x_2), \tag{QdP2}$$

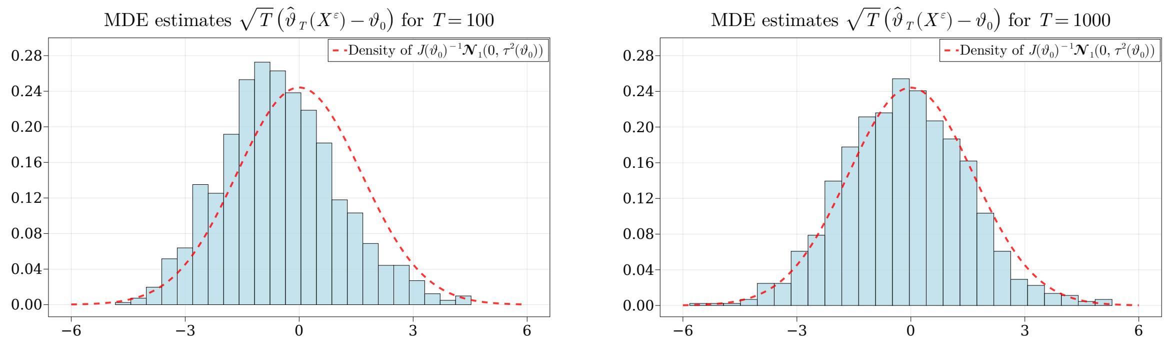


Figure 3.4: Comparison between the density of the asymptotic normal distribution $J(\vartheta_0)^{-1}\mathcal{N}_1(0,\tau^2(\vartheta_0))$ and normalized histogram plots of centered and scaled MDE estimates $\sqrt{T}(\widehat{\vartheta}_T(X^\varepsilon)-\vartheta_0)$ when the observations (X_i^ε) come from the Euler–Maruyama approximation of the multiscale over-damped Langevin equation (1.2.10) with $d=1$, a fast scale 2π-periodic oscillation $p=\sin$, and a linear drift term corresponding to (QdP1).

where $x:=(x_1,x_2)^\top\in\mathbb{R}^2$ and $M\in\mathbb{R}^{2\times 2}$ is positive-definite. In this case, the Langevin equations read as follows, for $X^\varepsilon:=(X^{1,\varepsilon},X^{2,\varepsilon})^\top$ and $X:=(X^1,X^2)^\top$,

$$\begin{pmatrix} dX_t^{1,\varepsilon} \\ dX_t^{2,\varepsilon}\end{pmatrix} = \left[-\alpha\begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2}\end{pmatrix}\begin{pmatrix} X_t^{1,\varepsilon} \\ X_t^{2,\varepsilon}\end{pmatrix} - \begin{pmatrix} \cos\left(X_t^{1,\varepsilon}/\varepsilon\right) \\ \frac{1}{2}\cos\left(X_t^{2,\varepsilon}/\varepsilon\right)\end{pmatrix}\right] dt + \sqrt{2\sigma}\begin{pmatrix} dW_t^1 \\ dW_t^2\end{pmatrix}, \tag{3.2.54}$$

$$\begin{pmatrix} dX_t^{1} \\ dX_t^{2}\end{pmatrix} = -\alpha\begin{pmatrix} \mathcal{K}_{1,1} & 0 \\ 0 & \mathcal{K}_{2,2}\end{pmatrix}\begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2}\end{pmatrix}\begin{pmatrix} X_t^{1} \\ X_t^{2}\end{pmatrix} dt + \begin{pmatrix} \sqrt{2\sigma\mathcal{K}_{1,1}} & 0 \\ 0 & \sqrt{2\sigma\mathcal{K}_{2,2}}\end{pmatrix}\begin{pmatrix} d\bar{W}_t^1 \\ d\bar{W}_t^2\end{pmatrix}, \tag{3.2.55}$$

and the homogenization matrix $\mathcal{K}\in\mathbb{R}^{2\times 2}$ equals

$$\mathcal{K} := \begin{pmatrix} \mathcal{K}_{1,1} & 0 \\ 0 & \mathcal{K}_{2,2}\end{pmatrix} := \begin{pmatrix} (\Pi_1^+\Pi_1^-)^{-1} & 0 \\ 0 & (\Pi_2^+\Pi_2^-)^{-1}\end{pmatrix}, \tag{3.2.56}$$

$$\Pi_i^\pm := \frac{1}{2\pi}\int_0^{2\pi}\exp\left(\pm\frac{p_i(y)}{\sigma}\right)dy, \quad i=1,2. \tag{3.2.57}$$

We set $\alpha=1$ and formulate our goal as the estimation of the positive-definite matrix $\vartheta_0:=\mathcal{K}M_0$ for some positive-definite matrix $M_0\in\mathbb{R}^{2\times 2}$ while assuming that we know the homogenized diffusion matrix $\bar{\Sigma}^2:=2\sigma\mathcal{K}$. Since X is a 2-dimensional Ornstein–Uhlenbeck process, by the results in [Lor17, Chapter 10], the invariant density $\mu(\vartheta_0)$ is given by the density of a centered 2-dimensional normal distribution with positive-definite covariance matrix $Q_\infty\in\mathbb{R}^{2\times 2}$ satisfying

$$Q_\infty = \int_0^\infty \exp\left(-t\mathcal{K}M_0\right)(2\sigma\mathcal{K})\exp\left(-t(KM_0)^\top\right)dt. \tag{3.2.58}$$

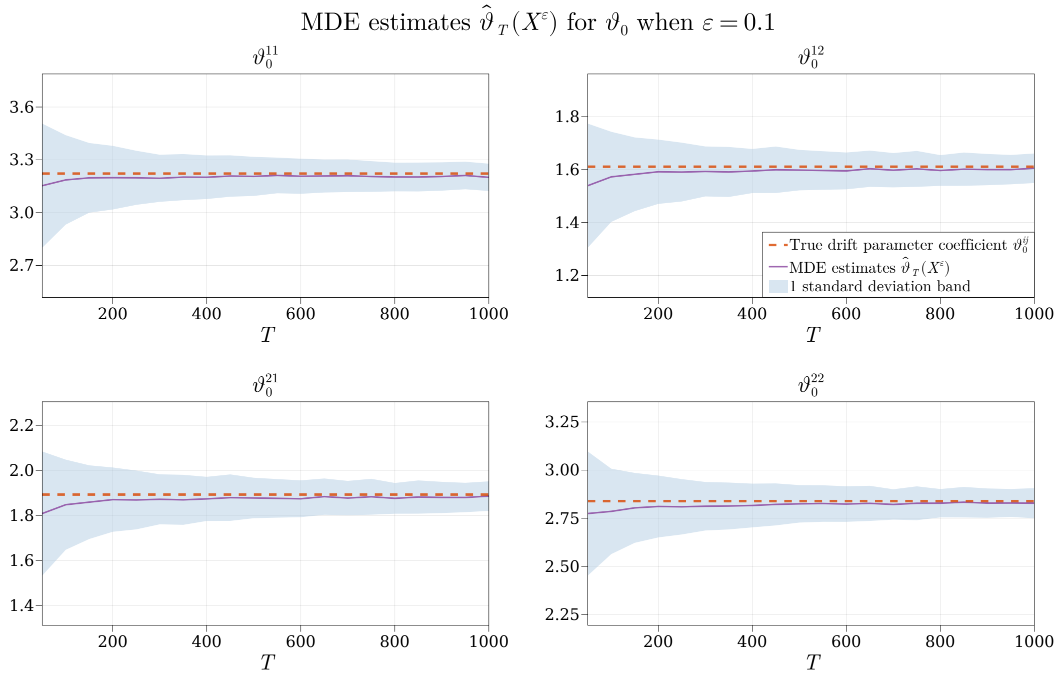


Figure 3.5: MDE estimates $\widehat{\vartheta}_T(X^\varepsilon)$ for $\vartheta_0 \in \mathbb{R}^{2\times 2}$ when the observations X^ε come from the Euler–Maruyama approximation of the multiscale overdamped Langevin equation (1.2.10) with $d = 2$ and a linear drift term with fast scale 2π-periodic oscillation corresponding to (QdP2).

Using the properties of the matrix exponential and a diagonalization of the positive-definite matrix $\sqrt{\mathcal{K}}M\sqrt{\mathcal{K}} = SDS^\top$ with orthogonal matrix S and diagonal matrix $D = \operatorname{diag}(d_1, d_2)$, we can simplify the previous formula

$$
\begin{aligned}
Q_\infty &= 2\sigma\sqrt{\mathcal{K}} \int_0^\infty \sqrt{\mathcal{K}}^{-1} \exp\left(-t\mathcal{K}M_0\right)\mathcal{K}\exp\left(-tM_0\mathcal{K}\right)\sqrt{\mathcal{K}}^{-1} dt\,\sqrt{\mathcal{K}} \\
&= 2\sigma\sqrt{\mathcal{K}} \int_0^\infty \exp\left(-t\sqrt{\mathcal{K}}M_0\sqrt{\mathcal{K}}\right)\exp\left(-t\sqrt{\mathcal{K}}M_0\sqrt{\mathcal{K}}\right) dt\,\sqrt{\mathcal{K}} \\
&= 2\sigma\sqrt{\mathcal{K}} \int_0^\infty S\exp\left(-tD\right)S^\top S\exp\left(-tD\right)S^\top dt\,\sqrt{\mathcal{K}} \\
&= 2\sigma\sqrt{\mathcal{K}}S \int_0^\infty \exp(-2tD)\,dt\,S^\top\sqrt{\mathcal{K}} \\
&= 2\sigma\sqrt{\mathcal{K}}S\operatorname{diag}\left(\int_0^\infty \exp(-2t/d_1)dt, \int_0^\infty \exp(-2t/d_2)dt\right)S^\top\sqrt{\mathcal{K}} \\
&= \sigma\sqrt{\mathcal{K}}SD^{-1}S^\top\sqrt{\mathcal{K}} \\
&= \sigma M_0^{-1} = (\mathcal{K}M_0)^{-1}\sigma\mathcal{K} = (2\vartheta_0)^{-1}\bar{\Sigma}^2.
\end{aligned}
$$

We can utilize the simplified formula (3.2.53) for the optimization task. The optimization was performed using an interior-point Newton algorithm with suitable nonlinear constraints which ensure that the cost functional of the optimization stays well-defined.

More precisely, we optimized with respect to the four entries of a matrix $A \in \mathbb{R}^{2\times 2}$ that satisfy the following nonlinear constraints

$$\begin{aligned} c_1(A) &:= A_{11} > 0, \\ c_2(A) &:= A_{11}A_{22} - A_{21}A_{12} > 0, \\ c_3(A) &:= \frac{A_{12}}{(\bar{\Sigma}^2)_{11}} - \frac{A_{21}}{(\bar{\Sigma}^2)_{22}} = 0. \end{aligned} \tag{3.2.59}$$

These constraints are derived from the requirements that A must be positive-definite, which corresponds to $c_1(A) > 0$ and $c_2(A) > 0$, and $(2A)^{-1}\bar{\Sigma}^2$ must be symmetric, which corresponds to $c_3(A) = 0$. The exact parameter configuration for this example is as follows.

- Multiscale and homogenized SDE parameters:

$$M_0 = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}, \quad \sigma = 1.5, \quad \vartheta_0 \approx \begin{pmatrix} 3.222 & 1.611 \\ 1.893 & 2.839 \end{pmatrix},$$

$$\bar{\Sigma}^2 \approx \begin{pmatrix} 2.416 & 0 \\ 0 & 2.838 \end{pmatrix}, \quad X_0^\varepsilon = \begin{pmatrix} 10 \\ 10 \end{pmatrix}.$$

- Small scale parameter and time parameters:

$$\varepsilon = 0.1, \quad T \in \{50k \mid k = 1, \dots, 20\}, \quad \Delta t = 10^{-3}.$$

- Interior-point Newton algorithm parameters:

$$\beta = 1, \quad \text{initial point: } \vartheta_{\text{initial}} = \begin{pmatrix} 3 & (\bar{\Sigma}^2)_{11}/4 \\ (\bar{\Sigma}^2)_{22}/4 & 6 \end{pmatrix}, \quad \text{optimizer: IPNewton.}$$

The interior-point Newton algorithm requires an initial starting point, in this case, a matrix $\vartheta_{\text{initial}} \in \mathbb{R}^{2\times 2}$ satisfying the nonlinear constraints above, and gradients of the cost functional Δ_N, which we have computed with the forward mode of automatic differentiation during optimization. Figure 3.5 shows the estimation results of the same experiment repeated for 500 times. We can once again see that the MDE is capable of estimating the true parameter matrix, which validates the robustness results for a multidimensional example that is not yet covered by the presented theory.

Fast Chaotic Noise. Consider the following system of ordinary differential equations (ODEs) with a slow component X^ε and a fast component $Y^\varepsilon := (Y_1^\varepsilon, Y_2^\varepsilon, Y_3^\varepsilon)$

$$\begin{aligned} \frac{dX^\varepsilon}{dt} &= AX - BX^3 + \frac{\lambda}{\varepsilon} Y_2^\varepsilon, \\ \frac{dY_1^\varepsilon}{dt} &= \frac{10}{\varepsilon^2} \left(Y_2^\varepsilon - Y_1^\varepsilon \right), \\ \frac{dY_2^\varepsilon}{dt} &= \frac{1}{\varepsilon^2} \left(28 Y_1^\varepsilon - Y_2^\varepsilon - Y_1^\varepsilon Y_3^\varepsilon \right), \\ \frac{dY_3^\varepsilon}{dt} &= \frac{1}{\varepsilon^2} \left(Y_1^\varepsilon Y_2^\varepsilon - \frac{8}{3} Y_3^\varepsilon \right), \end{aligned} \tag{3.2.60}$$

where $A, B \in \mathbb{R}$, and $\lambda > 0$. According to [PS08, Chapter 11], see also [GKS04], by eliminating the fast component Y^ε, which by itself solves the Lorenz equation at parameter values where the solution is chaotic, see [Wig03], the slow dynamics X^ε can be approximated by the solution of an SDE given by

$$dX_t = (AX_t - BX_t^3)dt + \bar{\Sigma}\, d\bar{W}_t. \tag{3.2.61}$$

Here, the diffusion coefficient $\bar{\Sigma}^2$ can be expressed in terms of the Green–Kubo formula, see [PS08, Chapter 11],

$$\bar{\Sigma}^2 = 2\lambda^2 \int_0^\infty \lim_{T \to \infty} \frac{1}{T} \int_0^T Y_2^{\varepsilon=1}(s) Y_2^{\varepsilon=1}(s+t)\, ds\, dt, \tag{3.2.62}$$

where $Y_2^{\varepsilon=1}(t)$ denotes the solution of the fast process' second component at time t when $\varepsilon = 1$. This whole approximation can be rigorously justified with the proofs in [MS11]. Looking at the Green–Kubo formula for $\bar{\Sigma}^2$, it would be a computationally expensive task to approximate it numerically. Instead, we will use the MDE for the estimation of that diffusion coefficient by plugging in discrete observations coming from the slow component X^ε of the deterministic system (3.2.60). This is possible because, when keeping the formulas (3.2.51) and (3.2.53) in mind, it is immediate that there is no need to change anything about the methodology. We merely have to optimize these distance formulas with respect to the parameter $\bar{\Sigma}^2$ now. However, as before, we have to assume that the true drift coefficients A and B are given or obtained from a prior estimation procedure.

Due to the stiffness of the ODE system (3.2.60), a simple fourth-order Runge–Kutta scheme, see [DB02, Chapter 4], is sufficient to generate a discrete set of observation points of the slow component. The considered parameter constellation in this simulation example is as follows.

- ODE parameters:

$$A = -1, \quad B = 0, \quad \lambda = \frac{2}{45}, \quad (X^\varepsilon(0), Y^\varepsilon(0)) = (1, 1, 1, 1); \tag{3.2.63}$$

$$A = B = 1, \quad \lambda = \frac{2}{45}, \quad (X^\varepsilon(0), Y^\varepsilon(0)) = (1, 1, 1, 1). \tag{3.2.64}$$

	$\varepsilon \setminus T$	100	500	1000		$\varepsilon \setminus T$	100	500	1000
$A=-1$,	10^{-1}	0.131	0.115	0.118	$A=1$,	10^{-1}	0.069	0.104	0.105
$B=0$	$10^{-3/2}$	0.123	0.118	0.109	$B=1$	$10^{-3/2}$	0.100	0.127	0.118

Table 3.1: MDE estimates $\widehat{\vartheta}_T(X^\varepsilon)$ for $\bar{\Sigma}_0^2$ when the observations (X_i^ε) is a discrete approximation of the slow component X^ε of the system (3.2.60) under different parameter configurations.

- Small scale parameter and time parameters:

$$\varepsilon \in \{10^{-1}, 10^{-3/2}\}, \quad T \in \{100, 500, 1000\}, \quad h = 10^{-3}.$$

- Gradient descent algorithm parameters:

$$\beta = 1, \quad \text{initial point: } \vartheta_{\text{initial}} = 0.8, \quad \text{box interval: } (0,\infty), \quad \text{optimizer: L-BFGS.}$$

Table 3.1 shows estimated values for the true diffusion coefficient $\vartheta_0 := \bar{\Sigma}_0^2$ under the specified parameter configurations. Since there is no analytic expression for $\bar{\Sigma}_0^2$, we have to compare the obtained estimates directly with existing numerical results given in [GKS04]. There, they adopted a setting with $A = B = 1$ and $\varepsilon = 10^{-3/2}$, giving an estimate of 0.126 ± 0.003 via Gaussian moment approximations based on a modified Euler–Maruyama discretization of the limit equation (3.2.61). Another estimate is 0.13 ± 0.01 which is based on the heterogeneous multiscale method (HMM) with a discretization of the Green–Kubo formula for $\bar{\Sigma}_0^2$; see [WLV05; FV04] and the references therein for more details on the HMM method and further applications to the Lorenz 96 model. When $T \in \{500, 1000\}$, we observe similar values for the diffusion coefficient $\bar{\Sigma}_0^2$, indicating good agreement of our estimation with previously reported values. All in all, this example solidifies the robustness and usefulness of the MDE for a chaotic deterministic ODE that is not at all covered by the theoretical results.

Bilinearly coupled fast-slow SDE system with Ornstein–Uhlenbeck limit. We return to the general fully coupled multiscale SDE system (1.2.1) and its homogenized limit SDE (1.2.2). More specifically, let us consider the system

$$dX_t^\varepsilon = \left[\frac{1}{\varepsilon}\gamma^\top Y_t^\varepsilon - \alpha X_t^\varepsilon\right] dt + \Sigma\, dW_t^{(x)}, \quad X_0^\varepsilon = x_0 \in \mathbb{R}, \tag{3.2.65a}$$

$$dY_t^\varepsilon = \left[-\frac{1}{\varepsilon^2} D Y_t^\varepsilon + \frac{c}{\varepsilon} X_t^\varepsilon Y_t^\varepsilon\right] dt + \frac{1}{\varepsilon}\sigma\, dW_t^{(y)}, \quad Y_0^\varepsilon = y_0 \in \mathbb{R}^m, \tag{3.2.65b}$$

with $m \geq 1$, $\gamma \in \mathbb{R}^m$, $\alpha > 0$, $\Sigma, c \in \mathbb{R}$, $D \in \mathbb{R}^{m\times m}$ such that all eigenvalues are contained in $\{\lambda \in \mathbb{C} \mid \operatorname{Re}\lambda > 0\}$, and $\sigma \in \mathbb{R}^{m\times m}$ such that $a = \sigma\sigma^\top$ is positive-definite. In this case, the frozen SDE (1.2.4) takes the form

$$dY_t^x = -DY_t^x\, dt + \sigma dW_t^{(y)}, \quad Y_0^x = y_0, \quad x \in \mathbb{R}, \tag{3.2.66}$$

which is an m-dimensional Ornstein–Uhlenbeck process. The assumption on the eigenvalues of D secures the existence and uniqueness of the invariant measure μ^x of Y^x, cf. [Lor17, Section 10.3], which is given by the density of a centered normal distribution with covariance matrix

$$Q_\infty = \int_0^\infty \exp(-tD)\sigma \exp(-tD^\top)\, dt. \tag{3.2.67}$$

We will only need the covariance matrix in the following, but let us remark that, after differentiation, Q_∞ solves the matrix equation

$$DQ_\infty + Q_\infty D^\top = \sigma.$$

Since μ^x has mean zero, we immediately obtain

$$\int_{\mathbb{R}^m} \gamma^\top y\, d\mu^x(y) = 0, \quad x \in \mathbb{R},$$

showing that the centering condition (1.2.3) is satisfied. Hence, we can solve for the solution of the cell problem (1.2.6)

$$-\mathcal{L}_0(x,y)\chi(x,y) = \gamma^\top y, \quad x \in \mathbb{R}, \tag{3.2.68}$$

where

$$\mathcal{L}_0(x,y) = \frac{1}{2}\sum_{i,j=1}^m a_{ij}\frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^m (Dy)_i \frac{\partial}{\partial y_i}. \tag{3.2.69}$$

One can check that the unique solution to the cell problem (3.2.68) satisfying the centering condition (1.2.3) is given by $\chi(x,y) = \gamma^\top D^{-1} y$, where D^{-1} exists by the eigenvalue assumption. With the solution of the cell problem, we can write down the coefficients of the homogenized limit SDE. The homogenized drift is given by

$$\begin{aligned}\bar{F}(x) &= \int_{\mathbb{R}^m} \Big[-\alpha x + \partial_x \chi(x,y)\gamma^\top y + \nabla_y \chi(x,y) xy\Big]\, d\mu^x(y) \\ &= -\alpha x + \int_{\mathbb{R}^m} \gamma^\top D^{-1} xy\, d\mu^x(y) = -\alpha x, \quad x \in \mathbb{R},\end{aligned}$$

and the homogenized diffusion coefficient by

$$\begin{aligned}\bar{\Sigma}^2 &= \int_{\mathbb{R}^m} \Big[\Sigma^2 + \nabla_y \chi(x,y)\, a\, \nabla_y \chi(x,y)^\top\Big]\, d\mu^x(y) \\ &= \Sigma^2 + \gamma^\top D^{-1} a D^{-T}\gamma.\end{aligned}$$

Hence, the homogenized limit SDE is a 1-dimensional Ornstein–Uhlenbeck process

$$dX_t = -\alpha X_t\, dt + \bar{\Sigma}\, d\bar{W}_t, \quad X_0 = x_0. \tag{3.2.70}$$

Assuming we know α, we aim to estimate the true diffusion coefficient $\vartheta_0 := \bar{\Sigma}_0^2$ with the MDE. Once again, we can do the optimization efficiently with the simplified distance formula (3.2.53) since the invariant density of X equals

$$\mu(\vartheta, x) = \sqrt{\frac{\alpha}{\pi\vartheta}} \exp\left(-\frac{\alpha}{\vartheta}x^2\right). \tag{3.2.71}$$

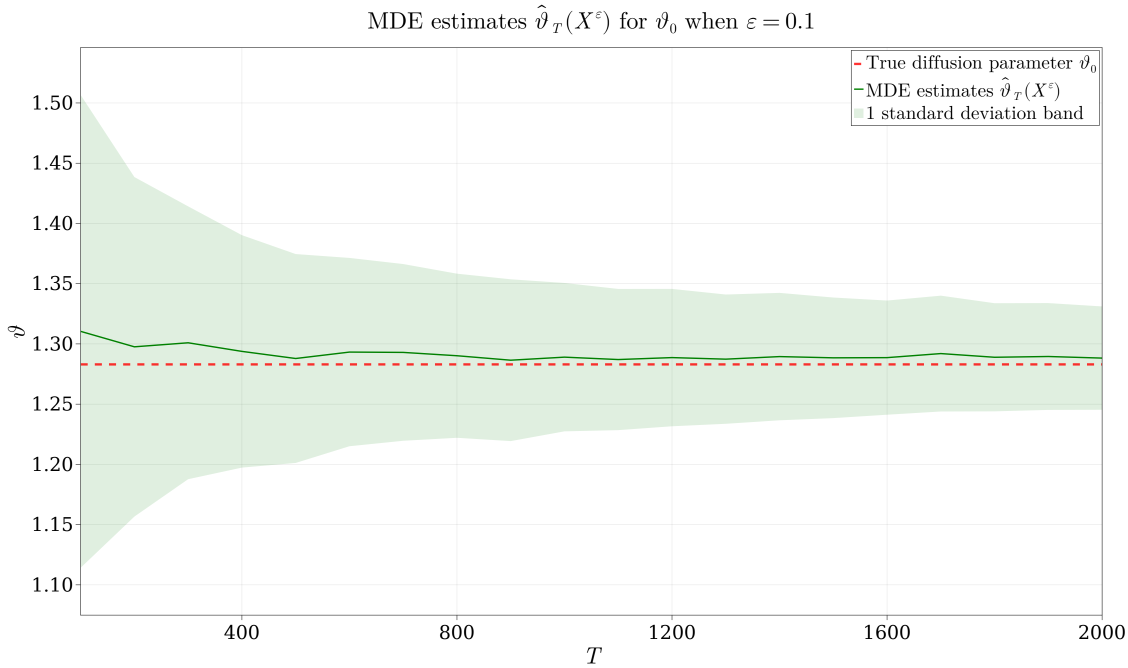


Figure 3.6: MDE estimates $\widehat{\vartheta}_T(X^\varepsilon)$ for ϑ_0 when the observations (X_i^ε) come from the Euler–Maruyama approximation of the slow variable X^ε of the coupled fast-slow SDE system (3.2.65).

We chose the following set of parameters for the simulation. The results are shown in Figure 3.6. They are comparable to the simulation results of the previous examples and showcase that the MDE is capable of estimating homogenized parameters even from observations of fully coupled multiscale SDE systems.

- Multiscale and homogenized SDE parameters:

$$m = 3, \quad \gamma = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad D = \begin{pmatrix} 2.5 & 1 & 0 \\ 0 & 2 & 0.8 \\ 0 & 0 & 1.5 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 1 & 0.3 & 0 \\ 0.2 & 0.9 & 0.4 \\ 0 & 0.1 & 0.8 \end{pmatrix},$$

$$\alpha = 1, \quad \Sigma = 0.5, \quad c = 1, \quad X_0^\varepsilon = 1, \quad Y_0^\varepsilon = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}, \quad \vartheta_0 \approx 1.283.$$

- Small scale parameter and time parameters:

$$\varepsilon = 0.1, \quad T \in \{100k \mid k = 1, \ldots, 20\}, \quad \Delta t = 10^{-3}.$$

- Gradient descent algorithm parameters:

$$\beta = 1, \quad \text{initial point: } \vartheta_{\text{initial}} = 10, \quad \text{box interval: } (0, \infty), \quad \text{optimizer: L-BFGS.}$$

Chapter 4

Nonparametric Estimation

This chapter contributes to the vast field of nonparametric estimation [Tsy08; IK81; Pra99; Was05; GN16] and considers the prototypical problem of estimating densities [Vaa98; Kut04; Bos98] but under our predominating perspective of weak misspecification. We will present a projection-based density estimator for the estimation of the homogenized invariant density of an overdamped Langevin SDE process. The approach is driven substantially by a truncation of the Fourier expansion of the homogenized invariant density with respect to the orthonormal basis of $L^2(\mathbb{R})$ consisting of the Hermite functions and an ergodic approximation of the Fourier coefficients through time averages of continuous observations.

The first two sections are meant to introduce the setting and the estimation approach. In the third section, we prove that the proposed density estimator is, upon fine-tuning, robust and satisfies a CLT in the Hilbert space $L^2(\mathbb{R})$; the main results are contained in Theorem 4.3.2 and Theorem 4.3.6. We also explore robustness and questions of weak convergence in different spaces beyond $L^2(\mathbb{R})$. As in the preceding chapter on parametric estimation, any rigorous asymptotic statement is only possible thanks to the limit theorems of Chapter 2. Finally, in the fourth section, we present a multitude of numerical results validating and supplementing the approach and the corresponding theoretical results.

4.1 The Density of Interest

Although the estimation approach is not particularly constrained by the underlying SDEs and can be practically implemented for many different SDEs, the theoretical results and many technical considerations do require a certain set of assumptions. As we have seen in the preceding chapters, the multiscale overdamped Langevin equation (1.2.10) and its homogenized limit equation (1.2.12) put us into a fairly convenient setting where most of the established limit theorems of Chapter 2 are directly applicable. For that reason, we will consider the following 1-dimensional multiscale overdamped Langevin equation as the running example, that is,

$$dX_t^\varepsilon = \left[-V'(X_t^\varepsilon) - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\right]dt + \Sigma\, dW_t, \quad X_0^\varepsilon = x_0, \tag{4.1.1}$$

with $\Sigma = \sqrt{2\sigma}$, and the homogenized limit equation

$$dX_t = -\mathcal{K}V'(X_t)\,dt + \bar{\Sigma}\,d\bar{W}_t, \quad X_0 = x_0, \tag{4.1.2}$$

with $\bar{\Sigma} = \sqrt{2\sigma\mathcal{K}}$, where $\mathcal{K}$ was defined in (1.2.16). In this chapter, we will work under the following set of assumptions.

Assumptions (L*).

i) $V, p \in C^\infty(\mathbb{R})$*, p is $\mathbb{L}$-periodic with period $\mathbb{L} > 0$, and $\sigma > 0$.*

ii) Any derivative of V is at most polynomially bounded and there exists a constant $K > 0$ such that for all $x \in \mathbb{R}$,

$$|V''(x)| \leq K(1 + |x|^2). \tag{4.1.3}$$

iii) There exist $a, b > 0$ and $R > 0$ such that for all $x \in \mathbb{R}$ with $|x| \geq R$

$$-V'(x)x \leq a - b|x|^2. \tag{4.1.4}$$

They are slightly stronger than the original Assumptions (L) in the sense that we allow the second derivative of V to grow at most quadratically. This is simply motivated by condition (2.4.54) in Corollary 2.4.12, so that all the results of Section 2.4 are directly applicable. Let us also recall the invariant densities of X^ε and X, which were initially defined in Corollary 2.4.4,

$$\mu_\varepsilon(x) = \frac{1}{Z_\varepsilon} \exp\left(-\frac{1}{\sigma}V(x) - \frac{p\,(x/\varepsilon)}{\sigma}\right), \qquad x \in \mathbb{R}, \quad \varepsilon > 0, \tag{4.1.5}$$

$$\mu(x) = \frac{1}{Z} \exp\left(-\frac{1}{\sigma}V(x)\right), \qquad x \in \mathbb{R}. \tag{4.1.6}$$

Our goal is the nonparametric estimation of the homogenized density μ from continuous observations of the process X^ε defined by (4.1.1).

4.2 Density Estimation by Hermite Expansion

Since the density $\mu \in L^2(\mathbb{R})$, one can expand μ in an orthonormal Fourier basis and, thanks to the ergodic theorem, approximate the Fourier coefficients with available observations. This is what the authors of [CGL25] proposed and studied in their Section 4 and we will base our approach on a similar idea. However, we have to handle the truncation of the estimator's Fourier expansion more carefully as we are, as always, confronted with weak misspecification.

We focus on the orthonormal basis (ONB) of $L^2(\mathbb{R})$, which is equipped with the usual inner product $\langle \cdot, \cdot \rangle_{L^2(\mathbb{R})}$, generated by the Hermite functions

$$\psi_n(x) := (2^n n! \sqrt{\pi})^{-1/2} \exp\left(-\frac{x^2}{2}\right) H_n(x), \quad x \in \mathbb{R}, \quad n \in \mathbb{N}_0, \tag{4.2.1}$$

where H_n denotes the n-th physicist's Hermite polynomial, see, for example, [Sze75, Chapter V] or [CGO21], given by

$$H_n(x) := (-1)^n \exp(x^2) \frac{d^n}{dx^n} \exp(-x^2), \quad x \in \mathbb{R}, \quad n \in \mathbb{N}_0. \tag{4.2.2}$$

The Fourier expansion of μ with respect to this ONB is

$$\mu = \sum_{n=0}^{\infty} \alpha_n \psi_n, \qquad \alpha_n := \int_{\mathbb{R}} \psi_n(x)\mu(x)\,dx. \tag{4.2.3}$$

Recall here our underlying assumption that we consider continuous observations of the SDE paths over a long time horizon. Hence, by ergodicity and homogenization of the multiscale overdamped Langevin equation, it is natural to approximate each Fourier coefficient α_n by time averages $\widehat{\alpha}^{\varepsilon}_{n,T}$ given by

$$\widehat{\alpha}^{\varepsilon}_{n,T} := \frac{1}{T}\int_0^T \psi_n(X_t^{\varepsilon})\,dt, \quad T > 0, \tag{4.2.4}$$

since

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \frac{1}{T}\int_0^T \psi_n(X_t^{\varepsilon})\,dt = \lim_{\varepsilon\to 0}\int_{\mathbb{R}} \psi_n(x)\mu_{\varepsilon}(x)\,dx = \int_{\mathbb{R}} \psi_n(x)\mu(x)\,dx, \tag{4.2.5}$$

which is the same as

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{\alpha}^{\varepsilon}_{n,T} = \alpha_n, \tag{4.2.6}$$

where the convergence is at least in $\mathbb{P}$-probability. Based on this initial observation, we introduce the estimator $\widehat{\mu}^{\varepsilon}_{N,T}$ for the invariant density μ by truncating the Fourier series and retaining only the first N approximated Fourier coefficients

$$\widehat{\mu}^{\varepsilon}_{N,T} := \sum_{n=0}^{N-1} \widehat{\alpha}^{\varepsilon}_{n,T}\,\psi_n. \tag{4.2.7}$$

Before we investigate the asymptotic behavior of this estimator, let us take a step back and see what has been established about this type of estimator in [CGL25] when there is no weak misspecification, that is, when one uses observations coming from X defined by (4.1.2). In [CGL25, Section 4, Proposition 7], the authors proved the following bound

$$\mathbb{E}\,\|\widehat{\mu}_{N,T} - \mu\|^2_{L^2(\mathbb{R})} \le \|\mu_N - \mu\|^2_{L^2(\mathbb{R})} + \frac{C}{T}, \tag{4.2.8}$$

for a constant $C > 0$ independent of N and T. Here, $\widehat{\mu}_{N,T}$ is

$$\widehat{\mu}_{N,T} := \sum_{n=0}^{N-1} \widehat{\alpha}_{n,T}\,\psi_n, \qquad \widehat{\alpha}_{n,T} := \frac{1}{T}\int_0^T \psi_n(X_t)\,dt, \tag{4.2.9}$$

and μ_N is the truncated Fourier expansion of μ given by $\mu_N := \sum_{n=0}^{N-1} \alpha_n\,\psi_n$. As we can see, the bound on the mean squared error (MSE) of $\widehat{\mu}_{N,T}$ in (4.2.8) consists of a truncation error of the Fourier expansion of μ and a sampling error controlled by the time horizon T. We will observe a similar error decomposition of the MSE in our asymptotic analysis of $\widehat{\mu}^{\varepsilon}_{N,T}$, see Theorem 4.3.2 and Remark 4.3.4, but we will have one additional error term that accounts for the weak misspecification.

4.3 Estimator Robustness and Asymptotic Normality

The estimator $\widehat{\mu}^{\varepsilon}_{N,T}$ depends on three key parameters: the scale separation parameter ε, the time horizon or observation time T, and the number of Fourier coefficients N. Among these, ε is typically unknown, T reflects the amount of available observations, and N is a tunable hyperparameter of the method. Intuitively speaking, one might expect the estimator to converge to the true invariant density μ in the limit of infinite data ($T \to \infty$), infinitely many Fourier coefficients ($N \to \infty$), and vanishing scale separation parameter ($\varepsilon \to 0$). However, this convergence does not always hold. In fact, the estimator is robust, i.e., asymptotically unbiased, only under specific conditions reflected by certain couplings of T, N, and ε. In particular, the number of Fourier coefficients N should not grow too quickly relative to the scale separation parameter ε. In other words, N should be carefully chosen as a function of ε. Conceptually, this can be seen from the truncated series of the Fourier expansion of μ_ε

$$\mu_{N,\varepsilon} := \sum_{n=0}^{N-1} \alpha_{n,\varepsilon}\, \psi_n, \qquad \alpha_{n,\varepsilon} := \int_{\mathbb{R}} \psi_n(x)\mu_\varepsilon(x)\, dx, \tag{4.3.1}$$

and its convergence

$$\lim_{N\to\infty} \lim_{\varepsilon\to 0} \|\mu_{N,\varepsilon} - \mu\|^2_{L^2(\mathbb{R})} = \lim_{N\to\infty} \lim_{\varepsilon\to 0} \sum_{n=0}^{N-1} |\alpha_{n,\varepsilon} - \alpha_n|^2 + \sum_{n=N}^{\infty} |\alpha_n|^2 = 0,$$

which is a consequence of weak convergence of the invariant measures and the fact that Hermite functions are bounded; recall Corollary 2.4.4 or Lemma 2.4.7, and refer to (4.3.11). However, if we exchange the limit order, then we obtain

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty} \|\mu_{N,\varepsilon} - \mu\|^2_{L^2(\mathbb{R})} = \lim_{\varepsilon\to 0} \|\mu_\varepsilon - \mu\|^2_{L^2(\mathbb{R})},$$

but the last limit does not even exist as we only have weak convergence of the invariant measures, which is strictly weaker than convergence in $L^2(\mathbb{R})$.
On the other hand, the final time T must diverge sufficiently fast; a condition we have required multiple times not only in the limit theorems of Chapter 2 but also in the parametric estimation problems of Chapter 3. Hence, in order to get robust estimates and a proper convergence result, it is important to impose meaningful scaling restrictions between these key parameters. These restrictions should encapsulate the interplay of the parameters in the sense of "T must grow sufficiently fast while ε must vanish before N diverges" as mentioned earlier. The robustness result under these requirements, while at the same time being the main theoretical contribution of the joint work [Bor+25], is stated in the following theorem.

Theorem 4.3.1. *Consider the SDEs* (4.1.1) *and* (4.1.2). *Assume the following conditions:*

i) $V, p \in C^2(\mathbb{R})$, *p is $\mathbb{L}$-periodic, and $V(0) = 0 = p(0)$ holds without loss of generality.*

ii) *V' is globally Lipschitz on $\mathbb{R}$ with Lipschitz constant $L_V > 0$, and there exists a constant $K > 0$ such that for all $x \in \mathbb{R}$,*

$$-V''(x) \le K(1 + |x|^2). \tag{4.3.2}$$

iii) There exist $b > 0$ *and* $R \geq 1$ *such that for all* $x \in \mathbb{R}$ *with* $|x| \geq R$,

$$-\operatorname{sgn}(x)V'(x) \leq -b\,|x|\,. \tag{4.3.3}$$

Suppose further that the number of Fourier coefficients and the final time of observation scale with ε *as*

$$N = N_\varepsilon = \left\lfloor \frac{\pi^2}{\gamma \mathbb{L}^2 \varepsilon^2} \right\rfloor, \qquad T = T_\varepsilon = \kappa \varepsilon^{-\zeta}, \tag{4.3.4}$$

respectively, for some $\kappa > 0$*, where the parameters* γ *and* ζ *satisfy*

$$\gamma > \begin{cases} 3 + \log(8), & \text{if } \sigma = 1, \\ c^2 + \max\left\{16e^{\frac{3}{2}}, \frac{c}{4}\left(\log\left|\frac{2}{c} - 1\right| + 2\log(4)\right)\right\}, & \text{if } \sigma \neq 1, \end{cases} \qquad \zeta > \begin{cases} 5, & \text{if } r \geq l, \\ 5\frac{l}{r}, & \text{if } r < l, \end{cases} \tag{4.3.5}$$

with

$$c = \frac{\sigma + 1}{\sigma}, \qquad l = \frac{L_V + |V'(0)|}{\sigma}, \qquad r = \frac{b}{\sigma}. \tag{4.3.6}$$

Then

$$\lim_{\varepsilon \to 0} \mathbb{E} \left\| \widehat{\mu}^\varepsilon_{N_\varepsilon, T_\varepsilon} - \mu \right\|^2_{L^2(\mathbb{R})} = 0. \tag{4.3.7}$$

The key insight of this theorem is that the MSE of $\widehat{\mu}^\varepsilon_{N,T}$ indeed converges to zero as $\varepsilon \to 0$ as long as $N = N_\varepsilon$ and $T = T_\varepsilon$ satisfy the prescribed scaling conditions. We will not prove this specific result here, because, unfortunately, it is not sufficient for our second goal, proving the asymptotic normality of the density estimator. That being said, the rather technical proof of Theorem 4.3.1 is contained in [Bor+25]. Instead of the qualitative convergence statement of Theorem 4.3.1, we aim for a quantitative statement with explicit convergence rates, which will serve our purpose of establishing asymptotic normality.

More precisely, building from a quantitative robustness result, we can tackle the asymptotic normality of the estimator $\widehat{\mu}^\varepsilon := \widehat{\mu}^\varepsilon_{N_\varepsilon, T_\varepsilon}$ by proving weak convergence of $\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon - \mu)$ to a centered Gaussian random element taking values in $L^2(\mathbb{R})$. To properly motivate what we are trying to accomplish, note that one can decompose the estimator's error using a stochastic part and a deterministic part

$$\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon - \mu) = \sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}) + \sqrt{T_\varepsilon}\,(\mu_{N_\varepsilon,\varepsilon} - \mu) =: \mathbb{S}_\varepsilon + D_\varepsilon, \tag{4.3.8}$$

with $\mu_{N_\varepsilon,\varepsilon}$ defined in (4.3.1). This is a decomposition that we have encountered many times before in other estimation settings, cf. (3.1.21), (3.2.27), or (3.2.48), making it a reoccurring theme of statistical estimation under weak misspecification. For the given setting at hand, the idea is to prove weak convergence of $\mathbb{S}_\varepsilon$ in $L^2(\mathbb{R})$ and convergence of D_ε to zero as $\varepsilon \to 0$ so that Slutzky's lemma B.1.5 can yield weak convergence of $\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon - \mu)$. However, this is challenging for at least two reasons. First, we do not have explicit $L^2(\mathbb{R})$ estimates for the term $(\mu_{N_\varepsilon,\varepsilon} - \mu)$ yet such that it would allow us to leverage them against the exploding $\sqrt{T_\varepsilon}$ factor in D_ε. Second, it is not clear yet why, under the standing assumptions, an asymptotic Gaussian random element should always exist in $L^2(\mathbb{R})$. We will address and answer these questions in the following.

We need to introduce some preliminary facts about Hermite functions and scale-dependent Sobolev-type spaces defined through these. We start with the latter. For $k \in \mathbb{N}_0$, we define the spaces

$$\mathbb{H}^k := \left\{ f \in L^2(\mathbb{R}) \mid \sum_{n=0}^{\infty} (2n+1)^k |\langle f, \psi_n \rangle_{L^2(\mathbb{R})}|^2 < \infty \right\}, \tag{4.3.9}$$

which emerge from the completion of the pre-Hilbert space $(\mathscr{S}(\mathbb{R}), \|\cdot\|_k)$ endowed with the inner product

$$\langle f, g \rangle_k := \sum_{n=0}^{\infty} (2n+1)^k \langle f, \psi_n \rangle_{L^2(\mathbb{R})} \langle g, \psi_n \rangle_{L^2(\mathbb{R})}, \quad f, g \in \mathscr{S}(\mathbb{R}), \tag{4.3.10}$$

and the corresponding induced norm $\|\cdot\|_k := \sqrt{\langle \cdot, \cdot \rangle_k}$. The construction of these spaces has been carried out in [Itô84, Chapter 1] and has, in fact, been defined more generally for arbitrary exponents $s \in \mathbb{R}$. Note that $\mathbb{H}^0 = L^2(\mathbb{R})$. It can be shown that $\mathscr{S}(\mathbb{R}) \subset \mathbb{H}^k$ for all $k \in \mathbb{N}_0$, see [Str23, Section 4.3].

We also need the following two standard facts about Hermite functions, see, for example, [Ind61] and [Tha93, Chapter 1]. For $n \in \mathbb{N}_0$, it is true that

$$\|\psi_n\|_\infty \leq \pi^{-1/4}, \tag{4.3.11}$$

and

$$\psi_n'(x) = \sqrt{\frac{n}{2}} \psi_{n-1}(x) - \sqrt{\frac{n+1}{2}} \psi_{n+1}(x), \quad x \in \mathbb{R}. \tag{4.3.12}$$

As announced earlier, before we can prove asymptotic normality, we need to establish the following quantitative restatement of Theorem 4.3.1 and prove an explicit convergence rate of the estimator to the homogenized invariant density μ in terms of the MSE of $\widehat{\mu}^\varepsilon_{N_\varepsilon, T_\varepsilon}$.

Theorem 4.3.2. *Consider the SDEs* (4.1.1) *and* (4.1.2) *under Assumptions (L*). Suppose that the number of Fourier coefficients and the final time of observation scale with ε as*

$$N_\varepsilon = \left\lfloor \varepsilon^{-\delta} \right\rfloor, \qquad T_\varepsilon = \varepsilon^{-\alpha}, \tag{4.3.13}$$

where $\alpha > 1/2$ and $\delta \in (0, 2 \wedge (\alpha - 1/2))$. Then, for all $k \in \mathbb{N}$ and as $\varepsilon \to 0$, we have

$$\mathbb{E} \left\| \widehat{\mu}^\varepsilon_{N_\varepsilon, T_\varepsilon} - \mu \right\|^2_{L^2(\mathbb{R})} = \mathcal{O}\left(\frac{N_\varepsilon}{T_\varepsilon \sqrt{\varepsilon}} + \varepsilon^{2k} N_\varepsilon^{k+1} + N_\varepsilon^{-k} \right). \tag{4.3.14}$$

In particular, upon choosing $k \geq \max\{(\alpha - 1/2 - \delta)/\delta, (\alpha - 1/2)/(2 - \delta)\}$, there exists a constant $C_{\alpha,k} > 0$ such that

$$\mathbb{E} \left\| \widehat{\mu}^\varepsilon_{N_\varepsilon, T_\varepsilon} - \mu \right\|^2_{L^2(\mathbb{R})} \leq C_{\alpha,k} \varepsilon^{\alpha - 1/2 - \delta} \tag{4.3.15}$$

for sufficiently small $\varepsilon > 0$.

Proof. Just like in (4.3.8), we can write

$$\widehat{\mu}^{\varepsilon} - \mu = \widehat{\mu}^{\varepsilon} - \mu_{N_\varepsilon,\varepsilon} + \mu_{N_\varepsilon,\varepsilon} - \mu,$$

where $\widehat{\mu}^{\varepsilon} = \mu^{\varepsilon}_{N_\varepsilon,T_\varepsilon}$ and $\mu_{N_\varepsilon,\varepsilon}$ is defined in (4.3.1). We will estimate these two differences separately. For the first difference, orthonormality and estimate (2.4.56) of Corollary 2.4.12, which is applicable to the Hermite functions thanks to inequality (4.3.11), imply

$$\begin{aligned}
\mathbb{E}\, \|\widehat{\mu}^{\varepsilon} - \mu_{N_\varepsilon,\varepsilon}\|^2_{L^2(\mathbb{R})} &= \sum_{n=0}^{N_\varepsilon-1} \mathbb{E}\, \left|\widehat{\alpha}^{\varepsilon}_{n,T_\varepsilon} - \alpha_{n,\varepsilon}\right|^2 \\
&= \sum_{n=0}^{N_\varepsilon-1} \mathbb{E}\, \left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \psi_n(X_t^\varepsilon)\, dt - \int_{\mathbb{R}} \psi_n(x)\mu_\varepsilon(x)\, dx\right|^2 \qquad (4.3.16) \\
&= \mathcal{O}\left(N_\varepsilon \|\psi_n\|^2_\infty T_\varepsilon^{-1}\varepsilon^{-1/2}\right) = \mathcal{O}\left(\frac{N_\varepsilon}{T_\varepsilon\sqrt{\varepsilon}}\right).
\end{aligned}$$

Now let $k \in \mathbb{N}$ be arbitrary. For the second difference, observe that

$$\|\mu_{N_\varepsilon,\varepsilon} - \mu\|^2_{L^2(\mathbb{R})} \le 2\left[\sum_{n=0}^{N_\varepsilon-1} |\alpha_{n,\varepsilon} - \alpha_n|^2 + \sum_{n=N_\varepsilon}^{\infty} |\alpha_n|^2\right].$$

With Lemma 2.4.7, which is applicable since Hermite functions are Schwartz functions, we can estimate the coefficients in the first term by

$$|\alpha_{n,\varepsilon} - \alpha_n| = \left|\int_{\mathbb{R}} \psi_n(x)(\mu_\varepsilon(x) - \mu(x))\, dx\right| = \mathcal{O}\left(\varepsilon^k \|\psi_n \exp(-V)\|_{W^{k,1}(\mathbb{R})}\right).$$

From (4.3.11) and (4.3.12) one can deduce, inductively for $n, l \in \mathbb{N}_0$,

$$\|\psi_n^{(l)}\|_\infty = \mathcal{O}\left((n+1)^{l/2}\right), \qquad (4.3.17)$$

so that Leibniz's differentiation rule and Lemma 2.4.2 yield

$$\|\psi_n \exp(-V)\|_{W^{k,1}(\mathbb{R})} \le \sum_{j=0}^{k}\sum_{l=0}^{j} \binom{j}{l} \|\psi_n^{(l)} \left[\exp(-V)\right]^{(j-l)}\|_{L^1(\mathbb{R})} = \mathcal{O}\left((n+1)^{k/2}\right).$$

Hence,

$$\sum_{n=0}^{N_\varepsilon-1} |\alpha_{n,\varepsilon} - \alpha_n|^2 = \mathcal{O}\left(\varepsilon^{2k}\sum_{n=0}^{N_\varepsilon-1}(n+1)^k\right) = \mathcal{O}\left(\varepsilon^{2k}N_\varepsilon^{k+1}\right). \qquad (4.3.18)$$

Since $\mu \in \mathscr{S}(\mathbb{R}) \subset \mathbb{H}^k$ by Lemma 2.4.2, the tail of the Fourier series expansion of μ can be estimated by

$$\sum_{n=N_\varepsilon}^{\infty} |\alpha_n|^2 \le \frac{1}{(2N_\varepsilon+1)^k}\sum_{n=N_\varepsilon}^{\infty}(2n+1)^k \left|\langle \mu, \psi_n\rangle_{L^2(\mathbb{R})}\right|^2 \le \frac{\|\mu\|_k^2}{N_\varepsilon^k} = \mathcal{O}\left(N_\varepsilon^{-k}\right). \qquad (4.3.19)$$

Combining (4.3.16), (4.3.18), and (4.3.19) yields (4.3.14). The last bound (4.3.15) is obtained from (4.3.14) by choosing $k \ge \max\{(\alpha - 1/2 - \delta)/\delta, (\alpha - 1/2)/(2-\delta)\}$, so that $2k - \delta(k+1) \ge \alpha - 1/2 - \delta$ and $\delta k \ge \alpha - 1/2 - \delta$ simultaneously. Hence,

$$\mathcal{O}\left(\frac{N_\varepsilon}{T_\varepsilon\sqrt{\varepsilon}}\right) = \mathcal{O}\left(\varepsilon^{\alpha-1/2}N_\varepsilon\right) = \mathcal{O}\left(\varepsilon^{\alpha-1/2-\delta}\right),$$

$$\mathcal{O}\left(\varepsilon^{2k}N_\varepsilon^{k+1}\right) = \mathcal{O}\left(\varepsilon^{2k-\delta(k+1)}\right) = \mathcal{O}\left(\varepsilon^{\alpha-1/2-\delta}\right),$$
$$\mathcal{O}\left(N_\varepsilon^{-k}\right) = \mathcal{O}\left(\varepsilon^{\delta k}\right) = \mathcal{O}\left(\varepsilon^{\alpha-1/2-\delta}\right),$$

which yields the final claim. □

Remark 4.3.3. *The number $k \in \mathbb{N}$ may be understood as a smoothness parameter that controls the regularity of μ through the spaces $\mathbb{H}^k$. It may be chosen arbitrarily large here because $\mu \in \mathscr{S}(\mathbb{R})$ by our assumptions, but it certainly has to be restricted if we impose less regularity on μ.*

Remark 4.3.4. *The bound in* (4.3.14) *consists of three terms that correspond to three different error sources. The first term comes from*

$$\mathbb{E}\,\|\widehat{\mu}^{\varepsilon} - \mu_{N_\varepsilon,\varepsilon}\|_{L^2(\mathbb{R})}^2 = \sum_{n=0}^{N_\varepsilon-1} \mathbb{E}\left|\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon} \psi_n(X_t^\varepsilon)\,dt - \int_{\mathbb{R}} \psi_n(x)\mu_\varepsilon(x)\,dx\right|^2, \tag{4.3.20}$$

which is the stochastic sampling error that decreases with growing observation time $T_\varepsilon = \varepsilon^{-\alpha}$, that is, more observation points. Note here that the constant that is obscured by the Landau notation in (4.3.16) *does not actually depend on α; one can verify this claim from the proof of Corollary 2.4.11 or Corollary 2.4.12. The second term corresponds to the weak model error, namely*

$$\left\|\sum_{n=0}^{N_\varepsilon-1} \alpha_{n,\varepsilon}\psi_n - \sum_{n=0}^{N_\varepsilon-1}\alpha_n\psi_n\right\|_{L^2(\mathbb{R})}^2 = \sum_{n=0}^{N_\varepsilon-1}\left|\int_{\mathbb{R}} \psi_n(x)(\mu_\varepsilon(x)-\mu(x))\,dx\right|^2, \tag{4.3.21}$$

which goes to zero in the limit for $\varepsilon \to 0$ as long as N_ε obeys the appropriate scaling regime. The last term corresponds to the truncation error of the Fourier series of μ, that is,

$$\left\|\mu - \sum_{n=0}^{N_\varepsilon-1}\alpha_n\psi_n\right\|_{L^2(\mathbb{R})}^2 = \sum_{n=N_\varepsilon}^{\infty} |\alpha_n|^2, \tag{4.3.22}$$

which converges to zero as long as N_ε goes to infinity. Also, keep in mind that (4.3.15) *does not mean that we can increase α indefinitely and hereby get an arbitrarily small error, because the constant $C_{\alpha,k}$ will blow up as $\alpha \to \infty$, which can be verified from the estimates leading up to* (4.3.18).

Remark 4.3.5. *Since $\mu \in \mathscr{S}(\mathbb{R})$ and $\widehat{\mu}_{N,T}^{\varepsilon} \in \mathscr{S}(\mathbb{R})$, $\mathbb{P}$-a.s., it is actually possible to prove robustness in the Schwartz space $\mathscr{S}(\mathbb{R})$. By [Str23, Section 4.3], the Schwartz space can be equipped with the metric*

$$d(f,g) := \sum_{l=0}^{\infty} 2^{-l}\frac{\|f-g\|_l}{1+\|f-g\|_l}, \quad f,g \in \mathscr{S}(\mathbb{R}), \tag{4.3.23}$$

where $\|\cdot\|_l$ is defined through (4.3.10)*, turning the Schwartz space into a complete, separable metric space. For $\eta > 0$, we have the basic estimate*

$$\mathbb{P}\left(d(\widehat{\mu}_{N_\varepsilon,T_\varepsilon}^{\varepsilon},\mu) > \eta\right) \le \frac{\mathbb{E}\,d(\widehat{\mu}_{N_\varepsilon,T_\varepsilon}^{\varepsilon},\mu)}{\eta},$$

which, according to the definition of the metric d, converges to zero as $\varepsilon \to 0$ if and only if

$$\mathbb{E}\,\|\widehat{\mu}^{\varepsilon}_{N_\varepsilon,T_\varepsilon} - \mu\|_l \to 0, \quad \varepsilon \to 0,$$

for all $l \in \mathbb{N}_0$. By similar arguments as in the proof of Theorem 4.3.2, we can prove, for $l \in \mathbb{N}_0$,

$$\mathbb{E}\,\|\widehat{\mu}^{\varepsilon} - \mu_{N_\varepsilon,\varepsilon}\|_l^2 = \sum_{n=0}^{N_\varepsilon-1} (2n+1)^l \left|\widehat{\alpha}^{\varepsilon}_{n,T_\varepsilon} - \alpha_{n,\varepsilon}\right|^2 = \mathcal{O}\left(\varepsilon^{\alpha-\frac{1}{2}} N_\varepsilon^{l+1}\right),$$

$$\left\|\sum_{n=0}^{N_\varepsilon-1} \alpha_{n,\varepsilon}\psi_n - \sum_{n=0}^{N_\varepsilon-1} \alpha_n\psi_n\right\|_l^2 = \sum_{n=0}^{N_\varepsilon-1} (2n+1)^l |\alpha_{n,\varepsilon} - \alpha_n|^2 = \mathcal{O}\left(\varepsilon^{2(l+1)} N_\varepsilon^{2(l+1)}\right),$$

$$\left\|\mu - \sum_{n=0}^{N_\varepsilon-1} \alpha_n\psi_n\right\|_{L^2(\mathbb{R})}^2 = \sum_{n=N_\varepsilon}^{\infty} (2n+1)^l |\alpha_n|^2 = \mathcal{O}\left(N_\varepsilon^{-1}\right),$$

where one can take $k = l+1$ in the proof of Theorem 4.3.2 to achieve the latter two estimates. Depending on l, one can therefore choose the scaling regime for N_ε such that all three terms vanish for $\varepsilon \to 0$. Since l was arbitrary, this shows that $\widehat{\mu}^{\varepsilon}_{N_\varepsilon,T_\varepsilon}$ converges in $\mathbb{P}$-probability to μ in the metric space $(\mathscr{S}(\mathbb{R}), d)$.

Our next goal is to prove a CLT for $\sqrt{T_\varepsilon}\,(\widehat{\mu}^{\varepsilon} - \mu)$ to a centered Gaussian random element in $L^2(\mathbb{R})$ and thereby show that, under Assumptions (L*), the asymptotic random element always exists in $L^2(\mathbb{R})$. To prove this CLT, we verify the two conditions given in Theorem B.3.1, which we recall here in the form that we require. Suppose we identified the asymptotic random element $\mathbb{G} \in L^2(\mathbb{R})$, then, in order to conclude $\sqrt{T_\varepsilon}\,(\widehat{\mu}^{\varepsilon} - \mu) \xrightarrow{\mathcal{D}} \mathbb{G}$ in $L^2(\mathbb{R})$, we need to prove

$$\left\langle \sqrt{T_\varepsilon}\,(\widehat{\mu}^{\varepsilon} - \mu), \sum_{n=0}^{l} \beta_n\psi_n \right\rangle_{L^2(\mathbb{R})} \xrightarrow{\mathcal{D}} \left\langle \mathbb{G}, \sum_{n=0}^{l} \beta_n\psi_n \right\rangle_{L^2(\mathbb{R})}, \quad \varepsilon \to 0, \quad \forall l \in \mathbb{N}_0, \tag{4.3.24}$$

$$\lim_{l\to\infty} \limsup_{\varepsilon\to 0} \mathbb{P}\left(\left\| \sqrt{T_\varepsilon}\,(\widehat{\mu}^{\varepsilon} - \mu) - \Pi_l\left(\sqrt{T_\varepsilon}\,(\widehat{\mu}^{\varepsilon} - \mu)\right) \right\|_{L^2(\mathbb{R})} \geq \eta \right) = 0, \quad \forall \eta > 0, \tag{4.3.25}$$

where $\beta_0, \ldots, \beta_l \in \mathbb{R}$ are arbitrary and

$$\Pi_l \colon L^2(\mathbb{R}) \to L^2(\mathbb{R}); \quad f \mapsto \sum_{n=0}^{l} \langle f, \psi_n \rangle_{L^2(\mathbb{R})}\, \psi_n,$$

is the orthogonal projection onto the l-th subspace of $L^2(\mathbb{R})$ spanned by the orthonormal Hermite basis vectors $\psi_0, \ldots, \psi_l$.

Since $L^2(\mathbb{R})$ is a separable Banach space, the notions of strong measurability of a function $Z \colon \Omega \to L^2(\mathbb{R})$, i.e., the pointwise approximation of Z by simple functions $Z_n \colon \Omega \to L^2(\mathbb{R})$, and weak measurability, i.e., the measurability of $\Omega \ni \omega \mapsto \langle Z(\omega), f \rangle_{L^2(\mathbb{R})}$ for all $f \in L^2(\mathbb{R})$, are equivalent. This is a consequence of the Pettis measurability theorem. Moreover, in the following we denote by $L^2(\Omega; L^2(\mathbb{R}))$ the Bochner space of strongly measurable functions $Z\colon \Omega \to L^2(\mathbb{R})$, that is, the space of random elements Z taking values in $L^2(\mathbb{R})$ such that

$$\mathbb{E}\,\|Z\|^2_{L^2(\mathbb{R})} = \int_\Omega \|Z(\omega)\|^2_{L^2(\mathbb{R})}\,d\mathbb{P}(\omega) < \infty.$$

A reference for the different notions of measurability, the Pettis measurability theorem, and Bochner spaces is [Hyt+16]. Recall also that a random element $Z\colon \Omega \to L^2(\mathbb{R})$ is called a centered Gaussian random element if for every $f \in L^2(\mathbb{R})$ the random variable $\langle Z, f\rangle_{L^2(\mathbb{R})}$ has a 1-dimensional normal distribution with mean zero. We present the main result now.

Theorem 4.3.6. *Consider the SDEs* (4.1.1) *and* (4.1.2) *under Assumptions (L*). Suppose that the number of Fourier coefficients and the final time of observation scale with ε as*

$$N_\varepsilon = \left\lfloor \varepsilon^{-\delta} \right\rfloor, \qquad T_\varepsilon = \varepsilon^{-\alpha}, \tag{4.3.26}$$

with $\alpha > 1/2$ and $\delta \in (0, (2/3)\wedge(\alpha-1/2))$. Then there exists a centered Gaussian random element $\mathbb{G}\colon \Omega \to L^2(\mathbb{R})$ such that

$$\sqrt{T_\varepsilon}\left(\widehat{\mu}^{\varepsilon}_{N_\varepsilon,T_\varepsilon} - \mu\right) \xrightarrow{\mathcal{D}} \mathbb{G}, \quad \text{in } L^2(\mathbb{R}), \quad \varepsilon \to 0. \tag{4.3.27}$$

The Gaussian random element $\mathbb{G}$ is given by the Fourier series

$$\mathbb{G} = \sum_{n=0}^{\infty} \xi_n \psi_n, \tag{4.3.28}$$

where $(\xi_0, \dots, \xi_n)^\top$, $n \in \mathbb{N}_0$, is an $(n+1)$-dimensional random vector with a multivariate normal distribution with mean zero and covariance matrix entries given by

$$\mathbb{E}\,\xi_n \xi_m = \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi_n'(x)\Phi_m'(x)\mu(x)\,dx =: \tau_{nm}, \quad m, n \in \mathbb{N}_0. \tag{4.3.29}$$

Here, $\Phi_n \in C^2(\mathbb{R})$ is the solution to the Poisson equation

$$-\mathcal{L}\Phi_n = \bar{\psi}_n := \psi_n - \int_{\mathbb{R}} \psi_n(x)\mu(x)\,dx, \quad \mathcal{L} = -\mathcal{K}V'\partial_x + \sigma\mathcal{K}\partial^2_{xx}, \tag{4.3.30}$$

which is explicitly given by

$$\Phi_n(x) = -\int_0^x \frac{2}{\bar{\Sigma}^2\mu(y)} \int_{-\infty}^y \bar{\psi}_n(z)\mu(z)\,dz\,dy, \quad x \in \mathbb{R}. \tag{4.3.31}$$

Before we prove this result, we want to motivate, through an example, why $L^2(\mathbb{R})$ is indeed a natural space to consider for the asymptotics and not a larger space, e.g., a Sobolev-type space as in (4.3.9) with negative scale exponent. Although we leverage specific eigenfunction structure of the Poisson equation (4.3.30) in the following example of an Ornstein-Uhlenbeck process, it reveals certain techniques that can be utilized in the more general setting of Theorem 4.3.6.

Example 4.3.7. Consider the Ornstein-Uhlenbeck process for the homogenized limit equation (4.1.2) with $\sigma = 1$, that is,

$$dX_t = -\mathcal{K} X_t \, dt + \sqrt{\mathcal{K}} \, d\bar{W}_t. \tag{4.3.32}$$

The invariant density is given by

$$\mu(x) = \frac{1}{\sqrt{\pi}} \exp\left(-x^2\right). \tag{4.3.33}$$

We want to show that

$$\sum_{n=0}^{\infty} \mathbb{E} \left|\xi_n\right|^2 < \infty, \tag{4.3.34}$$

because $\mathbb{G}$ would then exist as a random element in $L^2(\mathbb{R})$ by a simple Cauchy sequence argument, see the proof later. We may consider the generator $\mathcal{L} = -\mathcal{K}x\partial_x + (\mathcal{K}/2)\,\partial_{xx}^2$ as an unbounded operator on $L^2(\mu)$ with domain $\operatorname{dom}(\mathcal{L}) \subset L^2(\mu)$, cf. [Lor17, Chapter 10], where $L^2(\mu) := L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu)$ is the weighted L^2-space with respect to the invariant density μ. The normalized Hermite polynomial basis $(\widehat{H}_m)_{m\in\mathbb{N}_0}$ defined by

$$\widehat{H}_m(x) := \frac{1}{\sqrt{2^m m!}} H_m(x), \quad x \in \mathbb{R},$$

is an ONB of $L^2(\mu)$. By the following well-known recurrence relation for Hermite polynomials, see [Sze75],

$$H_{m+1}(x) = 2xH_m(x) - H_m'(x), \quad x \in \mathbb{R}, \quad m \in \mathbb{N}_0,$$

one can easily verify that $\widehat{H}_m$ is an eigenfunction of $-\mathcal{L}$ with eigenvalue $\mathcal{K}m$ for $m \in \mathbb{N}_0$. We can expand $\bar{\psi}_n$ in this ONB of eigenfunctions of $-\mathcal{L}$ via

$$\bar{\psi}_n = \sum_{m=1}^{\infty} \left\langle \psi_n, \widehat{H}_m \right\rangle_{L^2(\mu)} \widehat{H}_m,$$

since $\langle \bar{\psi}_n, \widehat{H}_0 \rangle_{L^2(\mu)} = \langle \bar{\psi}_n, 1 \rangle_{L^2(\mu)} = 0$ and $\langle \bar{\psi}_n, \widehat{H}_0 \rangle_{L^2(\mu)} = \langle \psi_n, \widehat{H}_m \rangle_{L^2(\mu)}$ for $m > 0$ due to centeredness of $\bar{\psi}_n$ with respect to μ and orthogonality of $(\widehat{H}_m)_{m\in\mathbb{N}_0}$. If we define

$$\Phi_n := \sum_{m=1}^{\infty} \frac{1}{\mathcal{K}m} \left\langle \psi_n, \widehat{H}_m \right\rangle_{L^2(\mu)} \widehat{H}_m, \tag{4.3.35}$$

then one can verify that $\Phi_n \in D(\mathcal{L})$ by the spectral theorem for unbounded operators, see [BS18], but, more importantly, the equation

$$-\mathcal{L}\Phi_n = \sum_{m=1}^{\infty} \frac{1}{\mathcal{K}m} \left\langle \psi_n, \widehat{H}_m \right\rangle_{L^2(\mu)} (-\mathcal{L}\widehat{H}_m) = \sum_{m=1}^{\infty} \left\langle \psi_n, \widehat{H}_m \right\rangle_{L^2(\mu)} \widehat{H}_m = \bar{\psi}_n$$

holds in $L^2(\mu)$. Thus, integration by parts, the preceding Poisson equation in $L^2(\mu)$, and the ONB expansions yield

$$\mathbb{E}\left|\xi_n\right|^2 = \mathcal{K} \int_{\mathbb{R}} \Phi_n'(x)^2 \mu(x)\, dx = \int_{\mathbb{R}} \Phi_n(x)(-\mathcal{L}\Phi_n)(x)\mu(x)\, dx$$

$$= \langle \Phi_n, -\mathcal{L}\Phi_n \rangle_{L^2(\mu)} = \left\langle \Phi_n, \bar{\psi}_n \right\rangle_{L^2(\mu)} = \frac{1}{\mathcal{K}} \sum_{m=1}^{\infty} \frac{1}{m} \left| \langle \psi_n, \widehat{H}_m \rangle_{L^2(\mu)} \right|^2 .$$

Furthermore, Tonelli's theorem and Parseval–Plancherel's identity imply

$$\sum_{n=0}^{\infty} \mathbb{E}\,|\xi_n|^2 = \frac{1}{\mathcal{K}} \sum_{m=1}^{\infty} \frac{1}{m} \sum_{n=0}^{\infty} \left| \langle \psi_n, \widehat{H}_m \rangle_{L^2(\mu)} \right|^2 = \frac{1}{\mathcal{K}} \sum_{m=1}^{\infty} \frac{1}{m} \|\widehat{H}_m \mu\|^2_{L^2(\mathbb{R})}.$$

So the verification of (4.3.34) rests upon the finiteness of the last series. The remaining estimates follow from the following integral formula for Hermite polynomials, see [GR07, 7.374(2)],

$$\int_{\mathbb{R}} H_m(x)^2 \exp(-2x^2)\, dx = 2^{m-\frac{1}{2}} \Gamma\left(m + \frac{1}{2}\right),$$

and the asymptotic approximation of the Gamma function

$$\Gamma\left(m + \frac{1}{2}\right) \sim_a \Gamma(m) m^{1/2} = (m-1)! \sqrt{m}$$

for $m \to \infty$, where $\sim_a$ denotes asymptotic equivalence. The last asymptotic formula can be proved directly with Stirling's formula $\Gamma(m+1) \sim_a \sqrt{2\pi m}(m/e)^m$ for $m \to \infty$. With these facts at our disposal, we obtain

$$\begin{aligned}
\|\widehat{H}_m \mu\|^2_{L^2(\mathbb{R})} &= \frac{1}{\pi 2^m m!} \int_{\mathbb{R}} H_m(x)^2 \exp(-2x^2)\, dx \\
&= \frac{\Gamma\left(m + \frac{1}{2}\right)}{\sqrt{2}\pi m!} \sim_a \frac{(m-1)!\sqrt{m}}{\sqrt{2}\pi m!} = \frac{1}{\sqrt{2}\pi \sqrt{m}},
\end{aligned}$$

and therefore

$$\sum_{n=0}^{\infty} \mathbb{E}\,|\xi_n|^2 \sim_a \frac{1}{\sqrt{2}\pi \mathcal{K}} \sum_{m=1}^{\infty} \frac{1}{m^{3/2}} < \infty.$$

Proof of Theorem 4.3.6. To prove the assertion of the theorem, we proceed in four steps.
Step 1. We first show that $\sum_{n=0}^{\infty} \mathbb{E}\,|\xi_n|^2 < \infty$. For $n \in \mathbb{N}_0$, we define

$$\bar{\Psi}_n(x) := \int_{-\infty}^{x} \bar{\psi}_n(y)\mu(y)\, dy, \qquad F_\mu(x) := \int_{-\infty}^{x} \mu(y)\, dy, \quad x \in \mathbb{R}.$$

Then

$$\begin{aligned}
\int_{\mathbb{R}} \psi_n(y)\mu(y)(\mathbb{1}_{(-\infty,x]}(y) - F_\mu(x))\, dy &= \int_{-\infty}^{x} \psi_n(y)\mu(y)\, dy - \int_{-\infty}^{x} \int_{\mathbb{R}} \psi_n(z)\,\mu(z)\, dz\, \mu(y)\, dy \\
&= \int_{-\infty}^{x} \left(\psi_n(y) - \int_{\mathbb{R}} \psi_n(z)\mu(z)\, dz\right) \mu(y)\, dy = \bar{\Psi}_n(x),
\end{aligned}$$

which is simply $\bar{\Psi}_n(x) = \langle \psi_n, k_x \rangle_{L^2(\mathbb{R})}$ with

$$k_x(y) := \mu(y)(\mathbb{1}_{(-\infty,x]}(y) - F_\mu(x)), \quad x, y \in \mathbb{R}. \tag{4.3.36}$$

Since

$$\Phi_n'(x) = -\frac{2}{\bar{\Sigma}^2} \frac{\bar{\Psi}_n(x)}{\mu(x)},$$

we obtain

$$\mathbb{E}\,|\xi_n|^2 = \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi_n'(x)^2 \mu(x)\,dx = \left(\frac{2}{\bar{\Sigma}^2}\right)^2 \int_{\mathbb{R}} \frac{\bar{\Psi}_n(x)^2}{\mu(x)}\,dx = \left(\frac{2}{\bar{\Sigma}^2}\right)^2 \int_{\mathbb{R}} \frac{|\langle \psi_n, k_x\rangle_{L^2(\mathbb{R})}|^2}{\mu(x)}\,dx.$$

Applying Tonelli's theorem and Parseval–Plancherel's identity yields

$$\sum_{n=0}^{\infty} \mathbb{E}\,|\xi_n|^2 = \left(\frac{2}{\bar{\Sigma}^2}\right)^2 \int_{\mathbb{R}} \frac{1}{\mu(x)} \sum_{n=0}^{\infty} |\langle \psi_n, k_x\rangle_{L^2(\mathbb{R})}|^2\,dx = \left(\frac{2}{\bar{\Sigma}^2}\right)^2 \int_{\mathbb{R}} \frac{1}{\mu(x)} \|k_x\|^2_{L^2(\mathbb{R})}\,dx. \tag{4.3.37}$$

Observe that

$$\begin{aligned}
&(1-F_\mu(x))^2 \int_{-\infty}^{x} \mu(y)^2\,dy + F_\mu(x)^2 \int_x^{\infty} \mu(y)^2\,dy \\
&= \Big(1 - 2F_\mu(x) + F_\mu(x)^2\Big) \int_{-\infty}^{x} \mu(y)^2\,dy + F_\mu(x)^2 \int_x^{\infty} \mu(y)^2\,dy \\
&= \int_{\mathbb{R}} \mu(y)^2 \Big[(1-2F_\mu(x))\mathbb{1}_{(-\infty,x]}(y) + F_\mu(x)^2\Big]\,dy \\
&= \int_{\mathbb{R}} \mu(y)^2 \Big(\mathbb{1}_{(-\infty,x]}(y) - F_\mu(x)\Big)^2\,dy = \|k_x\|^2_{L^2(\mathbb{R})}.
\end{aligned} \tag{4.3.38}$$

By Assumptions (L*) iii), there exist $a, b, R > 0$ such that for all $|x| \geq R$

$$\operatorname{sgn}(x)\frac{\mu'(x)}{\mu(x)} = -\operatorname{sgn}(x)\frac{1}{\sigma}V'(x) \leq \frac{1}{\sigma}\left(\frac{a}{|x|} - b|x|\right).$$

Suppose $x \geq R$. Then, integrating from x to $y \geq x$ yields

$$\mu(y) \leq \mu(x)\left(\frac{y}{x}\right)^{a/\sigma} \exp\left(-\frac{b}{2\sigma}(y^2 - x^2)\right).$$

From this estimate, one obtains, for some constant $C_1 > 0$ and sufficiently large $x \geq R$,

$$\begin{aligned}
\int_x^{\infty} \mu(y)^2\,dy &\leq \mu(x)^2 x^{-2a/\sigma} \exp\left(\frac{b}{\sigma}x^2\right) \int_x^{\infty} y^{2a/\sigma} \exp\left(-\frac{b}{\sigma}y^2\right) dy \\
&\leq C_1 \mu(x)^2 x^{-2a/\sigma} \exp\left(\frac{b}{\sigma}x^2\right) x^{2a/\sigma - 1} \exp\left(-\frac{b}{\sigma}x^2\right) \\
&= C_1 \frac{\mu(x)^2}{x} \leq \frac{C_1}{R}\mu(x)^2,
\end{aligned}$$

since, by L'Hôpital's rule,

$$\lim_{x\to\infty} \frac{2a_2 \int_x^{\infty} y^{a_1} \exp\left(-a_2 y^2\right)\,dy}{x^{a_1 - 1}\exp(-a_2 x^2)} = 1$$

for any $a_1, a_2 > 0$. Moreover, we have

$$1 - F_\mu(x) = \int_x^{\infty} \mu(y)\,dy \leq \frac{C_2}{R}\mu(x)$$

for some constant $C_2 > 0$ and sufficiently large $x \geq R$. Similar bounds can be obtained for $x \leq -R$, as well, but, in this case, one should establish the bounds for $\int_{-\infty}^x \mu(y)^2\,dy$ and $F_\mu(x)$. With these estimates and formula (4.3.38), we can conclude, for sufficiently large $\tilde{R} \geq R$,

$$
\begin{aligned}
\int_{\mathbb{R}} \frac{1}{\mu(x)} \|k_x\|^2_{L^2(\mathbb{R})}\,dx &\leq \int_{\mathbb{R}} \frac{(1-F_\mu(x))^2}{\mu(x)} \int_{-\infty}^x \mu(y)^2\,dy + \frac{F_\mu(x)^2}{\mu(x)} \int_x^\infty \mu(y)^2\,dy\,dx \\
&\leq C_3 + \int_{x\geq \tilde{R}} \frac{(1-F_\mu(x))^2}{\mu(x)} \|\mu\|_{L^2(\mathbb{R})} + \frac{1}{\mu(x)} \int_x^\infty \mu(y)^2\,dy\,dx \\
&\quad + \int_{x\leq -\tilde{R}} \frac{1}{\mu(x)} \int_{-\infty}^x \mu(y)^2\,dy + \frac{F_\mu(x)^2}{\mu(x)} \|\mu\|_{L^2(\mathbb{R})}\,dx \\
&\leq C_3 + C_4 \int_{|x|\geq \tilde{R}} \mu(x)\,dx,
\end{aligned}
$$

with constants $C_3, C_4 > 0$ that may have changed from line to line. Returning back to (4.3.37), we therefore have $\sum_{n=0}^\infty \mathbb{E}\,|\xi_n|^2 < \infty$.

Step 2. Building on the previous step, we can now prove that there exists this well-defined centered Gaussian random element $\mathbb{G}$ taking values in $L^2(\mathbb{R})$. To this end, consider the partial sums $S_L := \sum_{n=0}^L \xi_n \psi_n$ for $L \in \mathbb{N}_0$, where we recall that $(\xi_0, \dots, \xi_L)^\top$ have been defined in the statement of the theorem. For $L \in \mathbb{N}_0$, S_L takes values in $L^2(\mathbb{R})$ and is strongly measurable, because, for any $f \in L^2(\mathbb{R})$, the inner product $\langle S_L, f\rangle_{L^2(\mathbb{R})} = \sum_{n=0}^L \xi_n \langle \psi_n, f\rangle_{L^2(\mathbb{R})}$ is measurable as a finite linear combination of the random variables ξ_n. Moreover, we have

$$
\mathbb{E}\,\|S_L\|^2_{L^2(\mathbb{R})} = \sum_{n=0}^L \mathbb{E}\,|\xi_n|^2 < \infty.
$$

Hence, $S_L \in L^2(\Omega; L^2(\mathbb{R}))$. For $M > L$, it follows

$$
\begin{aligned}
\mathbb{E}\,\|\, S_M - S_L \,\|^2_{L^2(\mathbb{R})} &= \mathbb{E}\,\langle \sum_{m=L+1}^M \xi_m \psi_m, \sum_{l=L+1}^M \xi_l \psi_l \rangle_{L^2(\mathbb{R})} = \mathbb{E} \sum_{m,\,l=L+1}^M \xi_m \xi_l \delta_{ml} \\
&= \sum_{n=L+1}^M \mathbb{E}\,\xi_n^2 \leq \sum_{n=L+1}^\infty \mathbb{E}\,\xi_n^2 \to 0, \quad L \to \infty,
\end{aligned}
$$

where the last convergence is justified by step 1. This shows that $(S_L)_{L\in\mathbb{N}_0}$ is a Cauchy sequence in $L^2(\Omega; L^2(\mathbb{R}))$, so that $\mathbb{G} := \lim_{L\to\infty} S_L$ exists as an element in $L^2(\Omega; L^2(\mathbb{R}))$ and, therefore, as a random element in $L^2(\mathbb{R})$. It remains to show that $\mathbb{G}$ is centered Gaussian. We take an arbitrary $f \in L^2(\mathbb{R})$ and observe that

$$
\langle S_L, f\rangle_{L^2(\mathbb{R})} = \sum_{n=0}^L \xi_n \langle \psi_n, f\rangle_{L^2(\mathbb{R})}
$$

has a 1-dimensional normal distribution with mean zero because $(\xi_0, \dots, \xi_n)^\top$ is normally distributed with mean zero. The continuity of the inner product implies $\langle S_L, f\rangle_{L^2(\mathbb{R})} \to \langle \mathbb{G}, f\rangle_{L^2(\mathbb{R})}$ as $L \to \infty$. Hence, we can conclude that the limit $\langle \mathbb{G}, f\rangle_{L^2(\mathbb{R})}$ also has a 1-dimensional normal distribution with mean zero because normal distributions are closed

even under weak convergence, which is a consequence of the convergence of their characteristic functions. This proves that $\mathbb{G}$ is a centered Gaussian random element in $L^2(\mathbb{R})$.
Step 3. We now move on to prove condition (4.3.24). First, recall the splitting in (4.3.8)

$$\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon-\mu)=\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon-\mu_{N_\varepsilon,\varepsilon})+\sqrt{T_\varepsilon}\,(\mu_{N_\varepsilon,\varepsilon}-\mu)=\mathbb{S}_\varepsilon+D_\varepsilon.$$

By the estimates (4.3.18) and (4.3.19), we know that

$$\|\mu_{N_\varepsilon,\varepsilon}-\mu\|^2_{L^2(\mathbb{R})}=\mathcal{O}\left(\varepsilon^{2k}N_\varepsilon^{k+1}+N_\varepsilon^{-k}\right),\quad \varepsilon\to 0,$$

for any $k\in\mathbb{N}$. With the assumption on T_ε and N_ε, we therefore obtain

$$T_\varepsilon\|\mu_{N_\varepsilon,\varepsilon}-\mu\|^2_{L^2(\mathbb{R})}=\mathcal{O}\left(\varepsilon^{2k-\alpha-\delta(k+1)}+\varepsilon^{\delta k-\alpha}\right),$$

so that we get $\|D_\varepsilon\|_{L^2(\mathbb{R})}\to 0$ as $\varepsilon\to 0$ after choosing $k>\max\{(\alpha+\delta)/(2-\delta),\alpha/\delta\}$. Concerning the stochastic part, we can write it in the form

$$\begin{aligned}\mathbb{S}_\varepsilon&=\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon-\mu_{N_\varepsilon,\varepsilon})=\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\sum_{n=0}^{N_\varepsilon-1}\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt\,\psi_n,\\ \bar{\psi}_{n,\varepsilon}(x)&:=\psi_n(x)-\int_{\mathbb{R}}\psi_n(y)\mu_\varepsilon(y)\,dy,\quad x\in\mathbb{R}.\end{aligned}\tag{4.3.39}$$

For sufficiently small ε such that $N_\varepsilon-1>l$, the last representation (4.3.39) gives

$$\left\langle\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon-\mu_{N_\varepsilon,\varepsilon}),\sum_{n=0}^{l}\beta_n\psi_n\right\rangle_{L^2(\mathbb{R})}=\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\sum_{n=0}^{l}\beta_n\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt,$$

where $\beta_0,\dots,\beta_l\in\mathbb{R}$, $l\in\mathbb{N}_0$, are arbitrary numbers. This motivates the definition

$$\bar{h}_{l,\varepsilon}(x):=\sum_{n=0}^{l}\beta_n\bar{\psi}_{n,\varepsilon}(x)=\sum_{n=0}^{l}\beta_n\psi_n(x)-\int_{\mathbb{R}}\sum_{n=0}^{l}\beta_n\psi_n(y)\mu_\varepsilon(y)\,dy,\quad x\in\mathbb{R}.$$

Note that $\sum_{n=0}^{l}\beta_n\psi_n\in C^\infty(\mathbb{R})$ with all of its derivatives bounded uniformly on $\mathbb{R}$ by inequality (4.3.17), so that $\sum_{n=0}^{l}\beta_n\psi_n\exp(-V)\in\mathscr{S}(\mathbb{R})$. Hence, we can use Corollary 2.4.14 to conclude

$$\left\langle\sqrt{T_\varepsilon}\,(\widehat{\mu}^\varepsilon-\mu_{N_\varepsilon,\varepsilon}),\sum_{n=0}^{l}\beta_n\psi_n\right\rangle_{L^2(\mathbb{R})}=\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{h}_{l,\varepsilon}(X_t^\varepsilon)\,dt\xrightarrow{\mathcal{D}}\mathcal{N}_1(0,\tau_l^2),\quad\varepsilon\to 0,\tag{4.3.40}$$

where

$$\tau_l^2:=\bar{\Sigma}^2\int_{\mathbb{R}}\Phi_l'(x)^2\mu(x)\,dx>0,\tag{4.3.41}$$

and Φ_l solves the Poisson equation $-\mathcal{L}\Phi_l=\bar{h}_l$ with

$$\bar{h}_l(x):=\sum_{n=0}^{l}\beta_n\bar{\psi}_n(x)=\sum_{n=0}^{l}\beta_n\psi_n(x)-\int_{\mathbb{R}}\sum_{n=0}^{l}\beta_n\psi_n(y)\mu(y)\,dy,\quad x\in\mathbb{R}.$$

The solution to that Poisson equation is explicitly given by

$$\Phi_l(x)=-\int_0^x\frac{2}{\bar{\Sigma}^2\mu(y)}\int_{-\infty}^{y}\bar{h}_l(z)\mu(z)\,dz\,dy,\quad x\in\mathbb{R}.$$

Observe that the asymptotic variance in (4.3.41) satisfies

$$\tau_l^2 = \sum_{n=0}^{l}\sum_{m=0}^{l}\beta_n\beta_m\tau_{nm} = \sum_{n=0}^{l}\sum_{m=0}^{l}\beta_n\beta_m\mathbb{E}\,\xi_n\xi_m = \mathbb{E}\left[\left(\sum_{n=0}^{l}\beta_n\xi_n\right)^2\right] = \mathbb{V}\left(\sum_{n=0}^{l}\beta_n\xi_n\right).$$

Now we only need to notice that

$$\left\langle \mathbb{G}, \sum_{n=0}^{l}\beta_n\psi_n\right\rangle_{L^2(\mathbb{R})} = \left\langle \sum_{m=0}^{\infty}\xi_m\psi_m, \sum_{n=0}^{l}\beta_n\psi_n\right\rangle_{L^2(\mathbb{R})} = \sum_{n=0}^{l}\beta_n\xi_n,$$

which is a normally distributed random variable with mean zero and variance τ_l^2, i.e., it has the same distribution as the limit distribution in (4.3.40). This, together with Slutzky's lemma B.1.5, proves (4.3.24).

Step 4. It remains to prove condition (4.3.25). Because of the linearity of Π_l and $\|D_\varepsilon\|_{L^2(\mathbb{R})} \to 0$ as $\varepsilon \to 0$, we only have to prove

$$\lim_{l\to\infty}\limsup_{\varepsilon\to 0}\mathbb{P}\left(\left\|\sqrt{T_\varepsilon}\left(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}\right) - \Pi_l\left(\sqrt{T_\varepsilon}\left(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}\right)\right)\right\|_{L^2(\mathbb{R})} \geq \eta\right) = 0, \quad \forall \eta > 0. \tag{4.3.42}$$

By (4.3.39) and sufficiently small $\varepsilon > 0$ such that $N_\varepsilon - 1 > l$, we have

$$\begin{aligned}\Pi_l\left(\sqrt{T_\varepsilon}\left(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}\right)\right) &= \sum_{n=0}^{l}\left\langle \sqrt{T_\varepsilon}\left(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}\right), \psi_n\right\rangle_{L^2(\mathbb{R})}\psi_n \\ &= \sum_{n=0}^{l}\sum_{m=0}^{N_\varepsilon-1}\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt\,\delta_{nm}\psi_n \\ &= \sum_{n=0}^{l}\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt\,\psi_n,\end{aligned}$$

which implies

$$\begin{aligned}\left\|\sqrt{T_\varepsilon}\left(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}\right) - \Pi_l\left(\sqrt{T_\varepsilon}\left(\widehat{\mu}^\varepsilon - \mu_{N_\varepsilon,\varepsilon}\right)\right)\right\|_{L^2(\mathbb{R})}^2 &= \left\|\sum_{n=l+1}^{N_\varepsilon-1}\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt\,\psi_n\right\|_{L^2(\mathbb{R})}^2 \\ &= \sum_{n=l+1}^{N_\varepsilon-1}\left|\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt\right|^2.\end{aligned}$$

We can, once again, consider the Poisson equation $-\mathcal{L}_\varepsilon\Phi_{n,\varepsilon} = \bar{\psi}_{n,\varepsilon}$. Since ψ_n is uniformly bounded in n on $\mathbb{R}$ by (4.3.11), we obtain uniform bounds of the kind

$$M := \sup_{\varepsilon>0}\sup_{n\in\mathbb{N}_0}\|(\Phi_{n,\varepsilon})'\|_\infty < \infty \tag{4.3.43}$$

by Remark 2.1.15, which is applicable due to Lemma 2.4.6. Itô's formula entails the $\mathbb{P}$-almost sure representation

$$\frac{1}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}\bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt = \frac{\Phi_{n,\varepsilon}(x_0) - \Phi_{n,\varepsilon}(X_{T_\varepsilon}^\varepsilon)}{\sqrt{T_\varepsilon}} + \frac{\Sigma}{\sqrt{T_\varepsilon}}\int_0^{T_\varepsilon}(\Phi_{n,\varepsilon})'(X_t^\varepsilon)\,dW_t.$$

On the one hand, we have

$$\sum_{n=l+1}^{N_\varepsilon-1} \frac{\left|\Phi_{n,\varepsilon}(x_0) - \Phi_{n,\varepsilon}(X^\varepsilon_{T_\varepsilon})\right|^2}{T_\varepsilon} \le 2M^2 \left|X^\varepsilon_{T_\varepsilon} - x_0\right|^2 \frac{N_\varepsilon}{T_\varepsilon}$$

by (4.3.43). On the other hand, Itô's isometry and the triangle inequality give

$$\begin{aligned}
&\mathbb{E}\left|\frac{\Sigma}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} (\Phi_{n,\varepsilon})'(X^\varepsilon_t)\, dW_t\right|^2 \\
&= \Sigma^2 \mathbb{E}\, \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} (\Phi_{n,\varepsilon})'(X^\varepsilon_t)^2\, dt \\
&\le \Sigma^2 \left|\mathbb{E}\, \frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} (\Phi_{n,\varepsilon})'(X^\varepsilon_t)^2\, dt - \int_{\mathbb{R}} (\Phi_{n,\varepsilon})'(x)^2 \mu_\varepsilon(x)\, dx\right| \\
&+ \left|\Sigma^2 \int_{\mathbb{R}} (\Phi_{n,\varepsilon})'(x)^2 \mu_\varepsilon(x)\, dx - \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi_n'(x)^2 \mu(x)\, dx\right| + \bar{\Sigma}^2 \int_{\mathbb{R}} \Phi_n'(x)^2 \mu(x)\, dx \\
&=: A_{\varepsilon,n} + B_{\varepsilon,n} + \tau_{n,n}.
\end{aligned}$$

By Corollary 2.4.12, Remark 2.1.20, and estimate (4.3.43), we have

$$A_{\varepsilon,n} \le \mathbb{E}\left|\frac{1}{T_\varepsilon} \int_0^{T_\varepsilon} (\Phi_{n,\varepsilon})'(X^\varepsilon_t)^2\, dt - \int_{\mathbb{R}} (\Phi_{n,\varepsilon})'(x)^2 \mu_\varepsilon(x)\, dx\right|^2 = \mathcal{O}\left(\varepsilon^{\alpha-1/2}\right).$$

By Proposition 2.4.9 and (4.3.17), we also have

$$B_{\varepsilon,n} = \mathcal{O}\left(\varepsilon^\gamma \|\psi_n \exp(-V)\|_{W^{1,1}(\mathbb{R})}\right) = \mathcal{O}\left(\varepsilon^\gamma \sqrt{n+1}\right)$$

for any $\gamma \in (0,1)$. We choose $\gamma \in (3\delta/2, 1)$ to obtain

$$\sum_{n=l+1}^{N_\varepsilon-1} A_{\varepsilon,n} = \mathcal{O}\left(\varepsilon^{\alpha-1/2} N_\varepsilon\right) = \mathcal{O}\left(\varepsilon^{\alpha-1/2-\delta}\right), \qquad \sum_{n=l+1}^{N_\varepsilon-1} B_{\varepsilon,n} = \mathcal{O}\left(\varepsilon^\gamma N_\varepsilon^{3/2}\right) = \mathcal{O}\left(\varepsilon^{\gamma-3\delta/2}\right).$$

With Markov's inequality, Itô's isometry, and the triangle inequality, we therefore obtain, for any $\eta > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sum_{n=l+1}^{N_\varepsilon-1} \left|\frac{\Sigma}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} (\Phi_{n,\varepsilon})'(X^\varepsilon_t)\, dW_t\right|^2 \ge \frac{\eta^2}{4}\right) &\le \frac{4}{\eta^2} \sum_{n=l+1}^{N_\varepsilon-1} \left(A_{n,\varepsilon} + B_{n,\varepsilon} + \tau_{n,n}\right) \\
&\le \frac{4}{\eta^2}\left(\mathcal{O}\left(\varepsilon^{\alpha-1/2-\delta} + \varepsilon^{\gamma-3\delta/2}\right) + \sum_{n=l+1}^{\infty} \tau_{n,n}\right).
\end{aligned}$$

Moreover, we have

$$\mathbb{P}\left(\sum_{n=l+1}^{N_\varepsilon-1} \frac{\left|\Phi_{n,\varepsilon}(x_0) - \Phi_{n,\varepsilon}(X^\varepsilon_{T_\varepsilon})\right|^2}{T_\varepsilon} \ge \frac{\eta^2}{4}\right) \le \mathbb{P}\left(2M^2 \left|X^\varepsilon_{T_\varepsilon} - x_0\right|^2 \frac{N_\varepsilon}{T_\varepsilon} \ge \frac{\eta^2}{4}\right),$$

which converges to zero as $\varepsilon \to 0$ by Corollary 2.4.13 since $T_\varepsilon/N_\varepsilon = \varepsilon^{-(\alpha-\delta)}$ and $0 < \delta < \alpha$. By our assumptions on α and δ, and by the choice of γ, we finally obtain

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \mathbb{P}\left(\left\| \sqrt{T_\varepsilon}\,(\widehat{\mu}^{\,\varepsilon} - \mu_{N_\varepsilon,\varepsilon}) - \Pi_l\left(\sqrt{T_\varepsilon}\,(\widehat{\mu}^{\,\varepsilon} - \mu_{N_\varepsilon,\varepsilon})\right) \right\|_{L^2(\mathbb{R})} \geq \eta \right) \\
&= \limsup_{\varepsilon\to 0} \mathbb{P}\left(\sum_{n=l+1}^{N_\varepsilon-1} \left| \frac{1}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} \bar{\psi}_{n,\varepsilon}(X_t^\varepsilon)\,dt \right|^2 \geq \eta^2 \right) \\
&\leq \limsup_{\varepsilon\to 0} \mathbb{P}\left(\sum_{n=l+1}^{N_\varepsilon-1} \frac{\left|\Phi_{n,\varepsilon}(x_0) - \Phi_{n,\varepsilon}(X_{T_\varepsilon}^\varepsilon)\right|^2}{T_\varepsilon} \geq \frac{\eta^2}{4} \right) \\
&\quad + \limsup_{\varepsilon\to 0} \mathbb{P}\left(\sum_{n=l+1}^{N_\varepsilon-1} \left| \frac{\Sigma}{\sqrt{T_\varepsilon}} \int_0^{T_\varepsilon} (\Phi_{n,\varepsilon})'(X_t^\varepsilon)\,dW_t \right|^2 \geq \frac{\eta^2}{4} \right) \\
&\leq \frac{4}{\eta^2} \sum_{n=l+1}^{\infty} \tau_{n,n} = \frac{4}{\eta^2} \sum_{n=l+1}^{\infty} \mathbb{E}\,\xi_n^2.
\end{aligned}$$

By step 1, the last series tail converges to zero as $l \to \infty$, finishing the whole proof. $\square$

Remark 4.3.8. *One might be able to prove weak convergence not only in $L^2(\mathbb{R})$ but also in a scale-dependent Hilbert space up to a certain scale threshold. To be more precise, if one considers the harmonic oscillator on $\mathbb{R}$ given formally by*

$$\mathcal{H} := -\partial_{xx} + |x|^2, \tag{4.3.44}$$

then it is well-known that this unbounded operator is essentially self-adjoint on the complex Hilbert space $L^2_{\mathbb{C}}(\mathbb{R})$ with compact resolvent and eigenfunctions given by the Hermite basis functions $(\psi_n)_{n\in\mathbb{N}_0}$ with eigenvalues $\lambda_n := 2n+1$ for $n \in \mathbb{N}_0$, see [CR21]. The spectral theorem and the functional calculus for unbounded operators, see [BS18], enable us to characterize the domain of s-powers of $\mathcal{H}$, $s \geq 0$, via

$$\operatorname{dom}(\mathcal{H}^{s/2}) = \left\{ f = \sum_{n=0}^{\infty} \langle f, \psi_n\rangle_{L^2_{\mathbb{C}}(\mathbb{R})}\, \psi_n \;\middle|\; \sum_{n=0}^{\infty} \lambda_n^s |\, \langle f, \psi_n\rangle_{L^2_{\mathbb{C}}(\mathbb{R})} \,|^2 < \infty \right\}. \tag{4.3.45}$$

With this domain characterization and for any $s \geq 0$, we can define the scale-dependent Sobolev-type Hilbert space

$$\mathbb{H}^s := \left\{ f \in L^2(\mathbb{R}) \;\middle|\; \sum_{n=0}^{\infty} \lambda_n^s |\, \langle f, \psi_n\rangle_{L^2(\mathbb{R})} \,|^2 < \infty \right\}, \tag{4.3.46}$$

which is a closed subspace of $\operatorname{dom}(\mathcal{H}^{s/2})$ *and a direct generalization of the spaces introduced in* (4.3.9)*. The inner product on $\mathbb{H}^s$ is given by*

$$\langle f, g\rangle_{\mathbb{H}^s} := \sum_{n=0}^{\infty} \lambda_n^s \, \langle f, \psi_n\rangle_{L^2(\mathbb{R})} \, \langle g, \psi_n\rangle_{L^2(\mathbb{R})}\,, \quad f, g \in \mathbb{H}^s.$$

Proving weak convergence in $\mathbb{H}^s$ amounts to justifying the two conditions (4.3.24) *and* (4.3.25) *with the Hilbert space structure provided by $\mathbb{H}^s$. As we have seen in the previous* *proof, this only works if we can also prove that* $\mathbb{G}$ *is a random element in* $\mathbb{H}^s$*, which is a question of verifying* $\sum_{n=0}^{\infty} \lambda_n^s \mathbb{E}\,|\xi_n|^2 < \infty$*. We are led to believe that this is the case if and only if* $s < 1/2$*, since, by step 1 of the proof of Theorem 4.3.6 while using similar arguments, we would have,*

$$\sum_{n=0}^{\infty} \lambda_n^s \mathbb{E}\,|\xi_n|^2 = \left(\frac{2}{\bar{\Sigma}^2}\right)^2 \int_{\mathbb{R}} \frac{1}{\mu(x)} \|k_x\|_{\mathbb{H}^s}^2 \, dx, \tag{4.3.47}$$

with k_x *exactly as in* (4.3.36) *and* $\|\cdot\|_{\mathbb{H}^s} := \sqrt{\langle \cdot, \cdot \rangle_{\mathbb{H}^s}}$*. Note that* k_x *has a jump discontinuity of height* $\mu(x)$ *at* x*, but, outside a neighborhood of this jump discontinuity, it is actually a Schwartz function. Since* $\mathbb{H}^s$ *is locally equivalent to the classical Sobolev space* H^s*, see [BT06, Theorem 3], and, according to [Sic21, Theorem 1, Lemma 3],* $k_x \in H^s$ *if and only if* $s < 1/2$*, it does not seem too far-fetched to obtain weak convergence in* $\mathbb{H}^s$ *for* $s < 1/2$*, with a sharp scale threshold of* $1/2$*. Although these hints appear promising, one would still need more specific control of the norm* $\|k_x\|_{\mathbb{H}^s}$ *appearing in* (4.3.47)*.*

In the same vein as Remark 4.3.5, one might think that proving weak convergence in the Schwartz space $\mathscr{S}(\mathbb{R})$ *is more appropriate, and although the Schwartz space equipped with* d *as in* (4.3.23) *is a complete, separable metric space for which weak convergence is well-studied, see [Bog18; Bil99; Hen24], it cannot follow from our setting. There is no such Gaussian limit element in* $\mathscr{S}(\mathbb{R})$ *as we have seen from the above discussion and Example 4.3.7.*

4.4 Numerical Illustration

In this section, we present a series of numerical experiments to assess the empirical performance of $\widehat{\mu}_{N,T}^{\varepsilon}$ in approximating the invariant density μ of the homogenized limit equation (4.1.2). Since our theory is established for continuous observations and is inherently asymptotic in nature, but numerical results would be pre-asymptotic, that is, ε is small but fixed, the presentation of those numerical findings serves more as an illustration rather than a rigorous numerical study. We begin by demonstrating the influence of the observation time T and the truncation level N on the accuracy of the estimation when the scale separation parameter ε is small. Although not covered by the theoretical results, we also show through an example simulation that the methodology extends naturally to multidimensional SDE processes. Then, following Theorem 4.3.6, we present various distributional plots validating the CLT in $L^2(\mathbb{R})$. Finally, through a heuristic approach, we illustrate how the scale separation parameter ε can be inferred from our density estimator by considering a large number of Fourier coefficients. The numerical results have been implemented in the programming language Python and the code is available through the public GitHub repository [BZH26].

4.4.1 General Implementation Details

Throughout all experiments, path observations are generated by numerically solving the multiscale overdamped Langevin equation (4.1.1) with deterministic initial condition $X_0^{\varepsilon} = 0$. The SDEs are discretized using the Euler–Maruyama scheme, as already described in Section 3.2.4, with a fine time discretization step of $\Delta t = \varepsilon^3$. The invariant densities μ and μ_ε are computed from the equations (4.1.5) and (4.1.6), where the normalization constants Z and Z_ε are evaluated via a Gauss-Kronrod quadrature rule, see [DR84], as provided by the Python function `scipy.integrate.quad` from the SciPy library [Vir+20]. The implementation of $\widehat{\mu}^\varepsilon_{N,T}$ is straightforward and realized through the following natural discretization of (4.2.7)

$$\widehat{\mu}^L_N((X^\varepsilon_i)) := \sum_{n=0}^{N-1} \left(\frac{1}{L+1} \sum_{i=0}^{L} \psi_n(X^\varepsilon_i) \right) \psi_n, \tag{4.4.1}$$

where $(X^\varepsilon_i) := (X^\varepsilon_{i\Delta t})_{i=0}^L$ comes from the Euler–Maruyama discretization of (4.1.1) with $L = \lceil T/\Delta t \rceil$. Furthermore, the physicist's Hermite polynomials (4.2.2), thus ψ_n, are computed using the Python function `scipy.special.eval_hermite` from the aforementioned SciPy library [Vir+20].

4.4.2 Robustness

We consider the multiscale overdamped Langevin equation (4.1.1) with a large-scale potential and a fast-scale 2π-periodic oscillation given by

$$V(x) = \frac{x^4}{4} - \frac{x^2}{2}, \quad p(y) = \cos(y), \quad \sigma = 1, \quad \varepsilon = 0.1, \tag{4.4.2}$$

for $x \in \mathbb{R}$ and $y \in \mathbb{T}_{2\pi} = \mathbb{R}/(2\pi\mathbb{Z})$. The discretized version of the density estimator $\widehat{\mu}^\varepsilon_{N,T}$ is then computed for varying values of the observation time $T \in \{50, 500, 5000\}$ and number of Fourier coefficients $N \in \{4, 16, 64\}$.

In Figure 4.1, we compare the resulting estimates[1] of $\widehat{\mu}^\varepsilon_{N,T}$ with both the target homogenized invariant density μ and the unhomogenized invariant density μ_ε through density plots. Strictly speaking, these plots are not corresponding to expected $L^2(\mathbb{R})$-convergence as presented in Theorem 4.3.1 and Theorem 4.3.2, but they make for a better illustrative tool than plain $L^2(\mathbb{R})$-norm plots. However, recall Remark 4.3.5 where we justified the convergence of our estimator in the Schwartz space, which may be a more convincing reason for considering density plots. From the plots, we observe that the estimator fails to accurately approximate the invariant density when the number of Fourier coefficients N is too small. This is due to the truncation error in the Fourier expansion as characterized by the remainder term, cf. proof of Theorem 4.3.2 and Remark 4.3.4. On the contrary, when N is chosen appropriately and the trajectory is sufficiently long, the estimator $\widehat{\mu}^\varepsilon_{N,T}$ accurately captures the homogenized density μ. However, an excessive number of Fourier coefficients causes the estimator to break down, consistent with the theoretical limitations established in Theorem 4.3.1 and Theorem 4.3.2. In particular, $\widehat{\mu}^\varepsilon_{N,T}$ begins to capture the fine-scale oscillations of the multiscale dynamics instead of the coarse-grained behavior of the homogenized equation. This leads the estimator to approximate μ_ε rather than μ. Additionally, the estimation quality improves overall as the observation time T increases, which is in concordance with Remark 4.3.4.

Next, we examine a two-dimensional test case to demonstrate that the methodology extends beyond the 1-dimensional setting. Recall the multiscale overdamped Langevin

[1] actually estimates of the discretized version as given by (4.4.1)

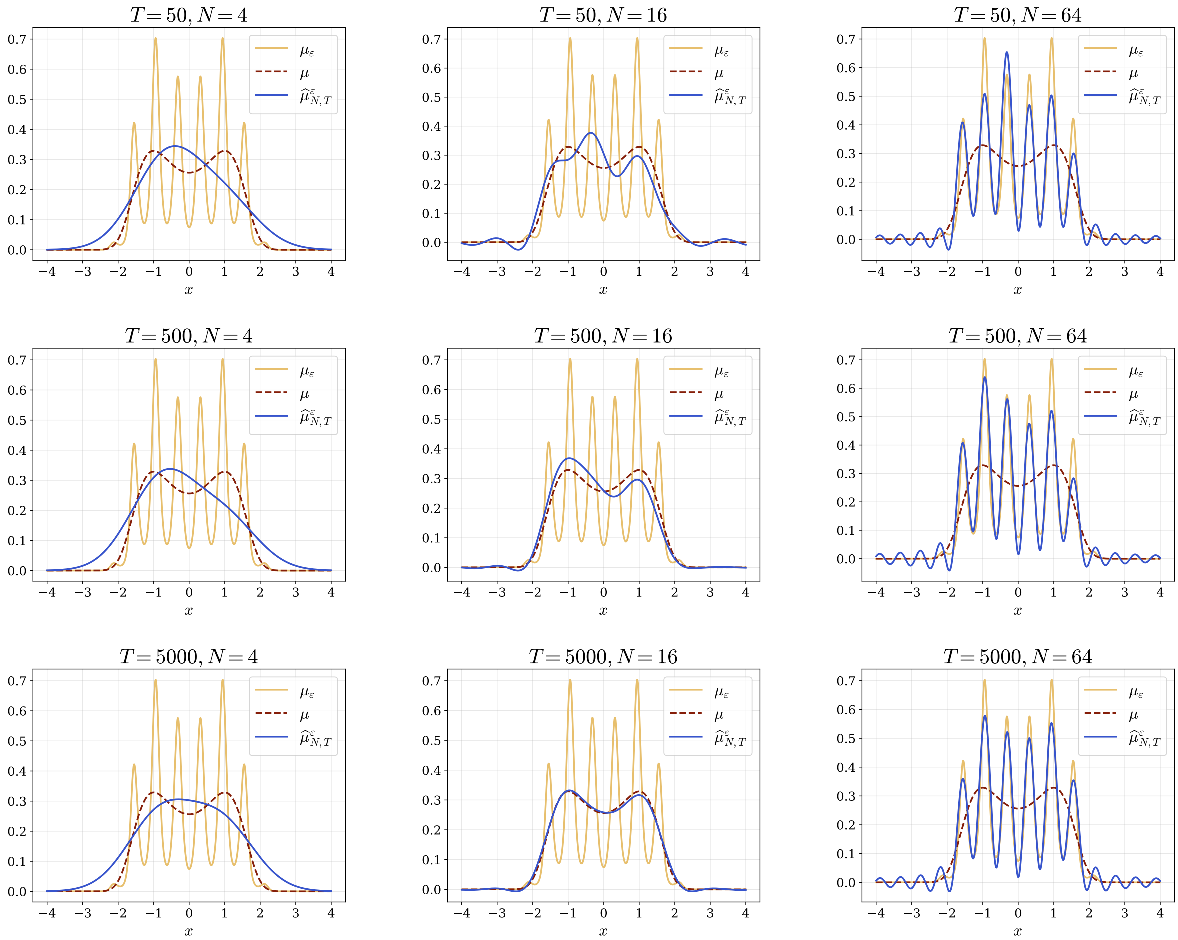


Figure 4.1: Estimates of $\widehat{\mu}^{\varepsilon}_{N,T}$ across varying values of the observation time $T \in \{50, 500, 5000\}$ and number of Fourier coefficients $N \in \{4, 16, 64\}$ for observations (X^{ε}_i) coming from the Euler–Maruyama discretization of (4.1.1) under the setting (4.4.2).

equation (1.2.10), its homogenized limit equation (1.2.12) for $d = 2$, and the corresponding invariant densities μ_ε and μ given in (2.4.16) and (2.4.17), respectively. We consider a two-dimensional analog of the previous 1-dimensional test case given by

$$V(x) = \frac{x_1^4 + x_2^4}{4} - \frac{x_1^2 + x_2^2}{2}, \quad p(y) = \sin(y_1) + \sin^2(y_2), \quad \sigma = 1.5, \quad \varepsilon = 0.1, \tag{4.4.3}$$

where $x = (x_1, x_2)^\top \in \mathbb{R}^2$ and $y = (y_1, y_2)^\top \in \mathbb{T}^2_{2\pi} = \mathbb{R}^2/(2\pi\mathbb{Z})^2$. The basis functions for the Fourier expansion are constructed through tensor products of 1-dimensional Hermite functions. Specifically, the estimator $\widehat{\mu}^{\varepsilon}_{N_1 N_2,T}$ with respect to continuous observations is given by

$$\widehat{\mu}^{\varepsilon}_{N_1,N_2,T} := \sum_{m=0}^{N_1-1} \sum_{n=0}^{N_2-1} \widehat{\alpha}^{\varepsilon}_{mn,T} \Psi_{mn}, \tag{4.4.4}$$

where $N_1, N_2 \in \mathbb{N}$ are the number of Fourier coefficients in dimension one and two, respectively, and the basis functions $\Psi_{mn} \colon \mathbb{R}^2 \to \mathbb{R}$ are defined by

$$\Psi_{mn}(x) := \psi_m(x_1)\psi_n(x_2), \quad x \in \mathbb{R}^2, \tag{4.4.5}$$

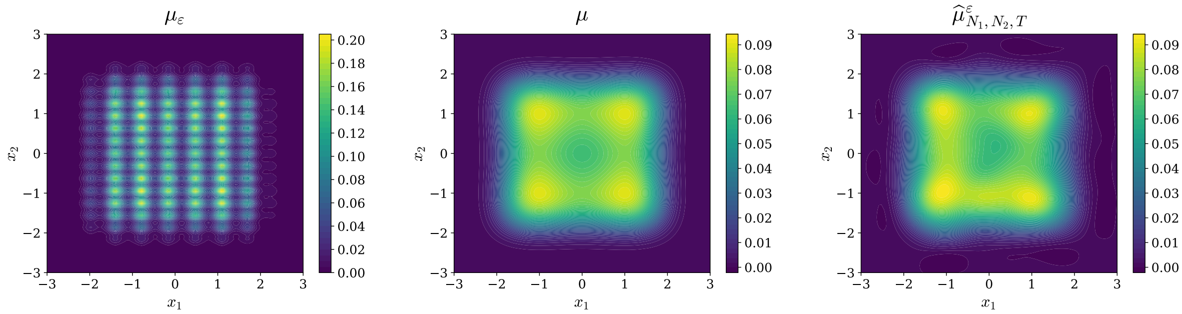


Figure 4.2: Estimates of $\widehat{\mu}^\varepsilon_{N_1,N_2,T}$ with $T = 2000$ and $N_1 = N_2 = 16$ for observations (X_i^ε) coming from the Euler–Maruyama discretization of (1.2.10) under the setting (4.4.3).

so that the coefficients $\alpha^\varepsilon_{mn,T}$ are given by

$$\widehat{\alpha}^\varepsilon_{mn,T} := \frac{1}{T}\int_0^T \Psi_{mn}(X_t^\varepsilon)\,dt. \tag{4.4.6}$$

Likewise, the discretized version of this estimator is defined by

$$\widehat{\mu}^L_{N_1,N_2}((X_i^\varepsilon)) := \sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}\left(\frac{1}{L+1}\sum_{i=0}^{L}\Psi_{mn}(X_i^\varepsilon)\right)\Psi_{mn}, \tag{4.4.7}$$

where $(X_i^\varepsilon) = (X^\varepsilon_{i\Delta t})_{i=0}^L$ are two-dimensional observation points coming from the Euler–Maruyama discretization of (1.2.10) with $L = \lceil T/\Delta t\rceil$. A trajectory of the multiscale SDE is generated over a time interval of length $T = 2000$ with a time discretization step of $\Delta t = \varepsilon^3$, and the estimator is computed using $N_1 = N_2 = 16$ Fourier coefficients in each dimension. In Figure 4.2, we compare a single estimate of $\widehat{\mu}^\varepsilon_{N_1,N_2,T}$ and the invariant densities μ_ε and μ with the help of contour plots. We can observe that $\widehat{\mu}^\varepsilon_{N_1N_2,T}$ provides a reasonably accurate approximation of μ by capturing its main features, demonstrating that the methodology produces numerically sound robustness results even in multidimensional settings.

4.4.3 Asymptotic Normality

To investigate aspects of the asymptotic normality of $\widehat{\mu}^\varepsilon_{N,T}$ numerically, we must draw samples from the random vectors $(\xi_0,\dots,\xi_n)^\top$, $n \in \mathbb{N}_0$, which, as we recall, have a multivariate normal distribution with mean zero and covariance structure prescribed by (4.3.29). For this, we need to solve the Poisson equation (4.3.30) for a finite number of centered Hermite functions. While an explicit formula is given by (4.3.31), it can become numerically unstable as we have to divide a very small number, the inner integral, by a very large number, the inverse of μ, which is why we choose to solve the Poisson equation more efficiently with the finite element method (FEM), see, for example, [Hac92]. More specifically, for $n \in \mathbb{N}_0$, we can first write the Poisson equation

$$-\mathcal{L}\Phi_n = \bar{\psi}_n = \psi_n - \int_{\mathbb{R}}\psi_n(x)\mu(x)\,dx, \quad \mathcal{L} = -\mathcal{K}V'\partial_x + \sigma\mathcal{K}\partial^2_{xx}, \tag{4.4.8}$$

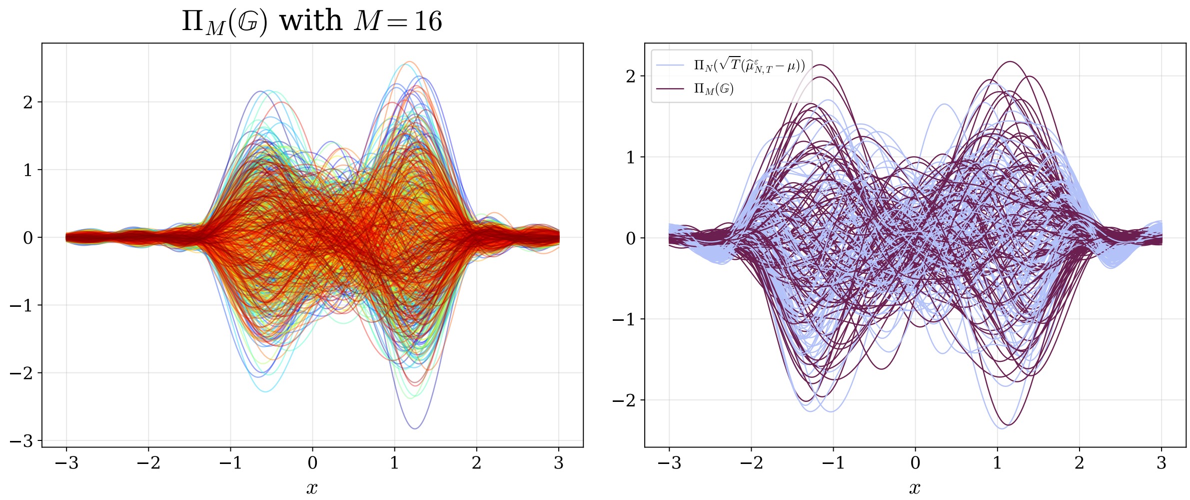


Figure 4.3: Left: 500 sample paths of the projection $\Pi_M(\mathbb{G})$ with $M = 50$ under setting (4.4.2) but with large-scale potential as in (4.4.16). Right: 100 sample paths of both $\Pi_N(\sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu))$ and $\Pi_M(\mathbb{G})$ with $N = M = 16$ and $T = 5000$ under setting (4.4.2).

in the equivalent form

$$-\frac{\bar{\Sigma}^2}{2}(\mu\Phi_n')' = \bar{\psi}_n\mu, \tag{4.4.9}$$

since $\mu'/\mu = -V/\sigma$. For sufficiently large $R > 0$ such that $\mu(R)\Phi_n'(R)$ and $\mu(-R)\Phi_n'(-R)$ become negligible, we can turn (4.4.9) into a weak problem with Neumann (zero-flux) boundary conditions corresponding to $\mu(R)\Phi_n'(R) = \mu(-R)\Phi_n'(-R) = 0$, see [Hac92; EG04], as follows:

Seek $\Phi_n \in H^1((-R, R))$ such that

$$a(\Phi_n, v) = l_n(v), \quad \forall v \in H^1((-R, R)), \tag{4.4.10}$$

where

$$H^1((-R, R)) := \left\{ v \in L^2((-R, R)) \,|\, v' \in L^2((-R, R)) \right\}, \tag{4.4.11}$$

and

$$a(u, v) := \frac{\bar{\Sigma}^2}{2} \int_{-R}^{R} u'(x)v'(x)\mu(x)\, dx, \tag{4.4.12}$$

$$l_n(v) := \int_{-R}^{R} \bar{\psi}_n(x)v(x)\mu(x)\, dx. \tag{4.4.13}$$

for $u, v \in H^1((-R, R))$. The solution to (4.4.10) is only unique up to an additive constant. To fix this constant, we impose the following normalization

$$\int_{-R}^{R} \Phi_n(x)\mu(x)\, dx = 0, \tag{4.4.14}$$

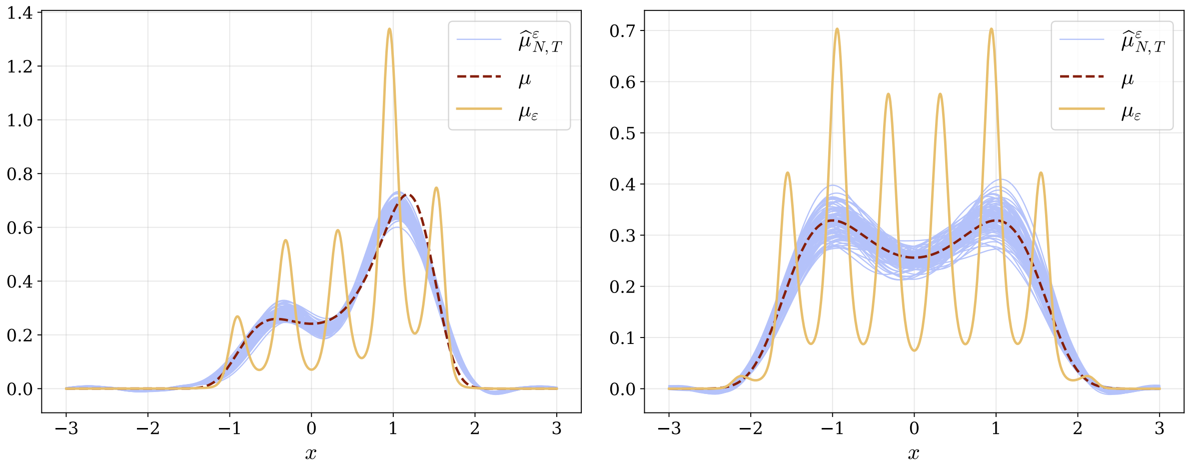


Figure 4.4: Left: 100 estimates of $\widehat{\mu}^{\varepsilon}_{N,T}$ for $N = 16$ and $T = 1000$ under setting (4.4.2) but with a large-scale potential as in (4.4.16). Right: 100 estimates of $\widehat{\mu}^{\varepsilon}_{N,T}$ for $N = 16$ and $T = 1000$ under setting (4.4.2).

i.e., $\Phi_n \in H^1_\mu((-R,R))$, where

$$H^1_\mu((-R,R)) := \left\{ v \in H^1((-R,R)) \;\middle|\; \int_{-R}^{R} v(x)\mu(x)\,dx = 0 \right\}. \tag{4.4.15}$$

Using this mathematical setup, one can build the FEM implementation to compute the relevant covariances, which correspond to $\tau_{nm} = \mathbb{E}\,\xi_n\xi_m = 2a(\Phi_n', \Phi_m')$ for $n, m \in \mathbb{N}_0$. In our setup and for the sole purpose of illustration, we chose an ad-hoc value for R such that the approximation of τ_{nm} stabilizes. We solved for the FEM approximation in a finite-dimensional subspace of $H^1((-R,R))$ consisting of 1-dimensional Lagrange linear elements. The normalization (4.4.14) was enforced a posteriori by subtracting the μ-weighted mean from the acquired FEM solution. Although this is more aligned with the centering condition of the original Poisson problem (4.4.8), the normalization does not affect the computed quantities τ_{nm}, which only depend on derivatives of the solutions. The actual implementation itself has been realized with the open-source computing platform FEniCS [Bar+25; Aln+14] and their provided Python interface.

In order to draw sample paths from the centered Gaussian random element $\mathbb{G}$ of Theorem 4.3.6, we have to compute the covariances τ_{nm} for finitely many $n, m \in \mathbb{N}_0$ up to a cutoff of $M \in \mathbb{N}_0$, assemble the covariance matrix $\tau_M := (\tau_{nm})_{n,m=0}^{M}$, and then draw samples from a multivariate normal distribution $\mathcal{N}_{M+1}(0, \tau_M)$. Strictly speaking, this only gives us the orthogonal projection $\Pi_M(\mathbb{G})$ onto the M-th subspace of $L^2(\mathbb{R})$ spanned by the orthonormal Hermite basis vectors $\psi_0, \dots, \psi_M$, but it is a meaningful approximation to $\mathbb{G}$ and is sufficient for our illustrative purposes. On the left hand side of Figure 4.3, we can see an example of 500 sample paths of this projection with $M = 50$. On the right hand side of Figure 4.3, a rough comparison of 100 sample paths of both the projection of the centered and scaled estimator $\widehat{\mu}^{\varepsilon}_{N,T}$ and the projection of $\mathbb{G}$ with $N = M = 16$ and $T = 5000$ indicates a visual similarity between their respective distributions. The right plot has been generated under the setting (4.4.2) and the left plot, too, but with a

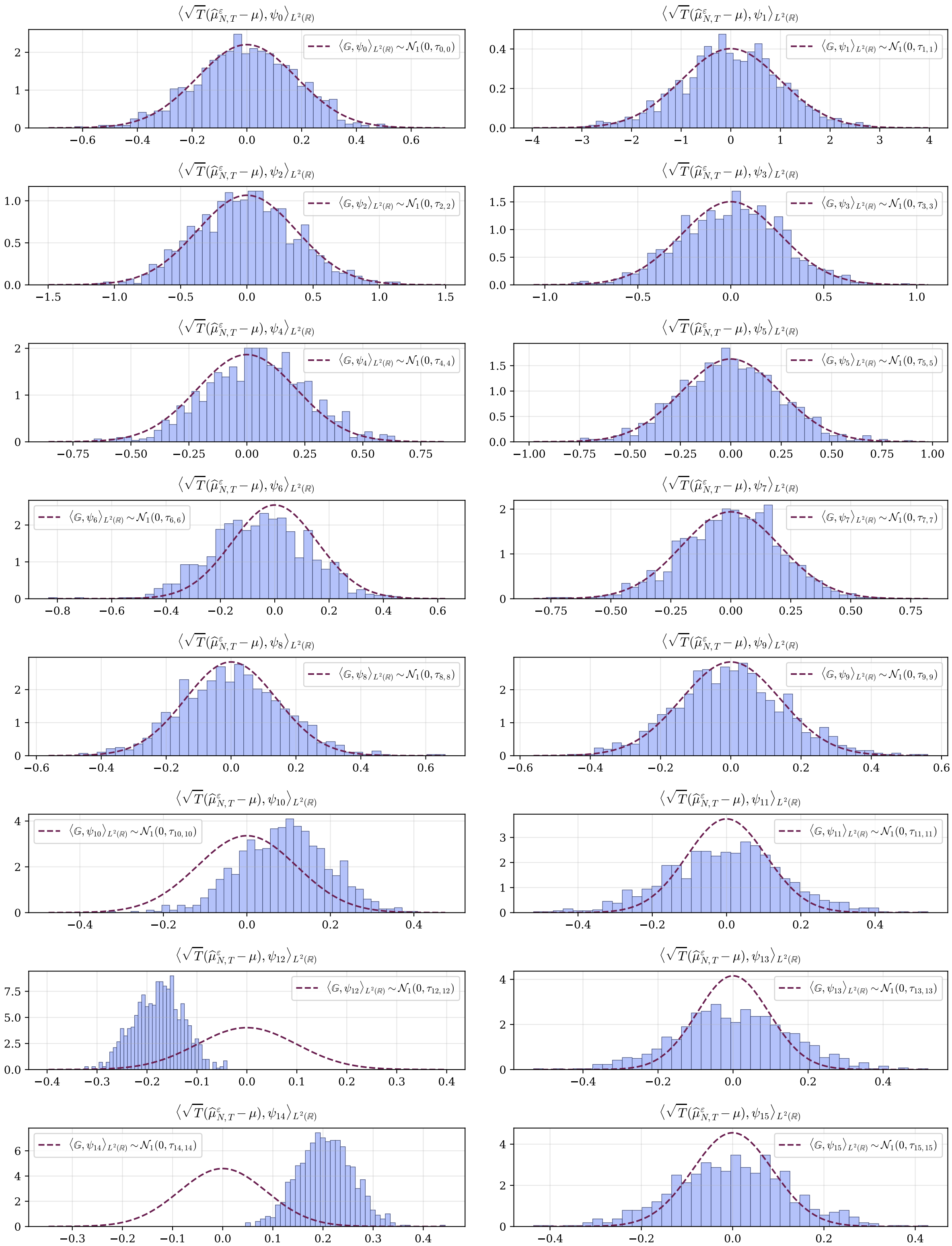


Figure 4.5: Comparison between the density of the asymptotic random variable $\langle \mathbb{G}, \psi_n \rangle_{L^2(\mathbb{R})} \sim \mathcal{N}_1(0, \tau_{n,n})$ for $n \in \{0, \ldots, 15\}$ and normalized histogram plots of the Fourier coefficients $\langle \sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu), \psi_n \rangle_{L^2(\mathbb{R})}$ for $N = 16$ and $T = 1000$ when the observations (X^{ε}_i) come from the Euler–Maruyama discretization of (4.1.1) under the setting (4.4.2) with $\varepsilon = 0.1$.

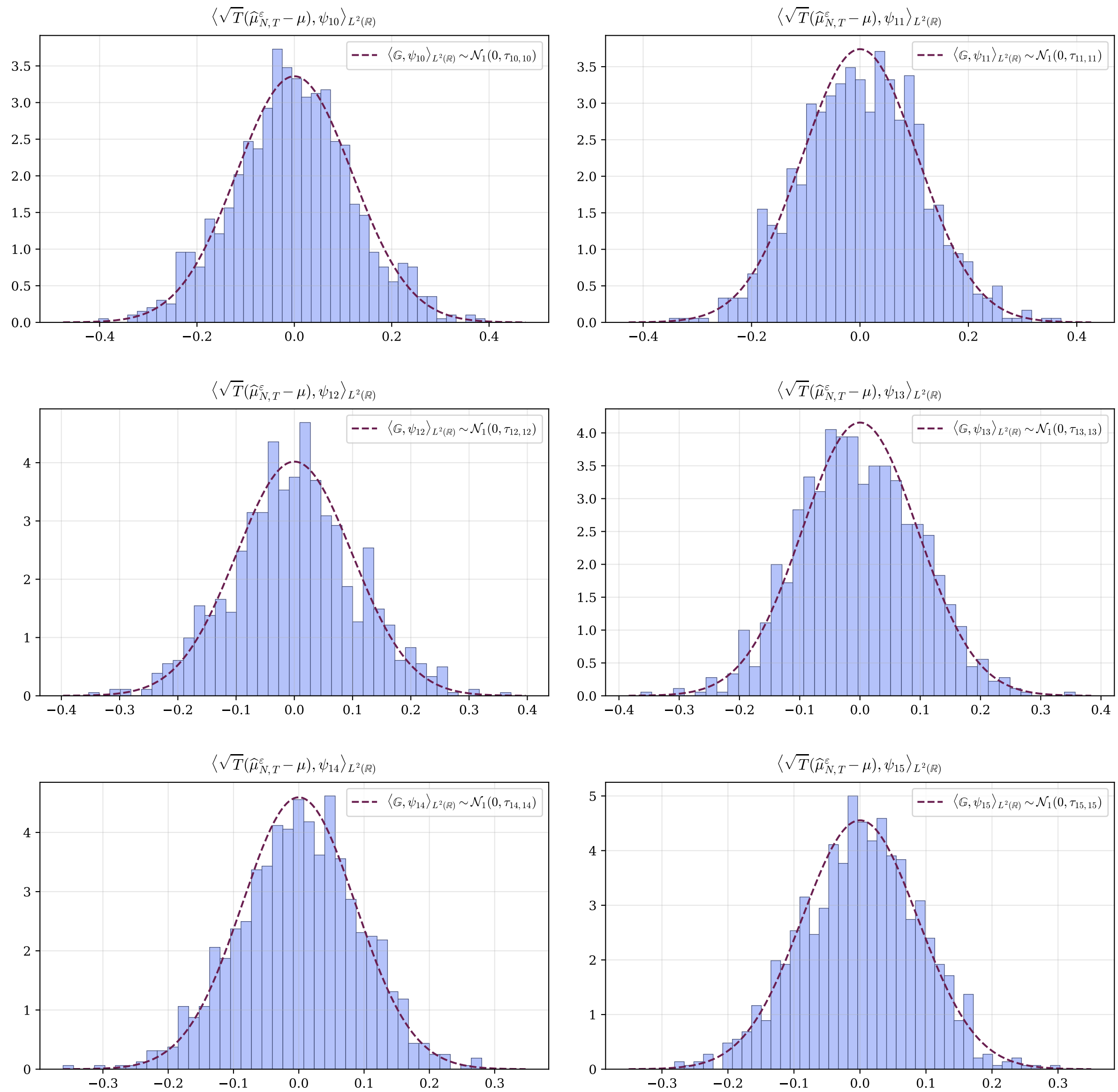


Figure 4.6: Comparison between the density of the asymptotic random variable $\langle \mathbb{G}, \psi_n\rangle_{L^2(\mathbb{R})} \sim \mathcal{N}_1(0, \tau_{n,n})$ for $n \in \{10, \dots, 15\}$ and normalized histogram plots of the Fourier coefficients $\langle \sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu), \psi_n\rangle_{L^2(\mathbb{R})}$ for $N = 16$ and $T = 1000$ when the observations (X_i^{ε}) come from the Euler–Maruyama discretization of (4.1.1) under the setting (4.4.2) with $\varepsilon = 0.05$.

different large-scale potential given by

$$V(x) = x^4 - x^3 - x^2, \quad x \in \mathbb{R}. \tag{4.4.16}$$

The humps of the samples paths in Figure 4.3 can be attributed to the specific forms of the homogenized invariant density μ. For a concrete visual comparison, Figure 4.4 shows density plots as in the previous Subsection 4.4.2 but with 100 estimates of $\widehat{\mu}^{\varepsilon}_{N,T}$ using $N = 16$ and $T = 1000$. Here, the left plot corresponds to setting (4.4.2) with large-scale potential (4.4.16) and the right plot to the already familiar case (4.4.2) of the previous

Subsection 4.4.2.

Let us now take a closer look at individual Fourier coefficients of the centered and scaled estimator $\sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu)$, that is, $\langle \sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu), \psi_n \rangle_{L^2(\mathbb{R})} = \sqrt{T}(\widehat{\alpha}^{\varepsilon}_{n,T} - \alpha_n)$ for $n \in \{0, \dots, N\}$, and compare their distribution with the asymptotic normal distribution given by the random variables $\langle \mathbb{G}, \psi_n \rangle_{L^2(\mathbb{R})} = \xi_n \sim \mathcal{N}_1(0, \tau_{n,n})$. We assume setting (4.4.2) with $\varepsilon \in \{0.1, 0.05\}$. Figure 4.5 shows normalized histograms of 1000 samples of the random variable $\langle \sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu), \psi_n \rangle_{L^2(\mathbb{R})}$ for $n \in \{0, \dots, 15\}$ and compares them to the (approximated) density of the asymptotic normal distribution $\mathcal{N}_1(0, \tau_{n,n})$. Most of the histograms resemble the correct asymptotic normal distribution, but the histograms of $\langle \sqrt{T}(\widehat{\mu}^{\varepsilon}_{N,T} - \mu), \psi_n \rangle_{L^2(\mathbb{R})}$ for $n \in \{10, \dots, 15\}$ do not attain the correct asymptotic shape yet. Because ε is not small enough yet, we can still see a mismatch in estimated mean and variance. This is perfectly in line with the established asymptotic theory since the final asymptotic normal distribution is only reached as soon as the deterministic bias term, i.e., D_ε in (4.3.8), vanishes and the unhomogenized asymptotic variance converges to the homogenized asymptotic variance, recall condition (2.3.10) in Theorem 2.3.3. If we reduce the scale separation parameter to $\varepsilon = 0.05$, then we can indeed observe that these aforementioned mismatches in mean and variance start to disappear, as illustrated in Figure 4.6.

4.4.4 Inference of the Scale Separation Parameter

Theorem 4.3.1 shows that the optimal number of Fourier coefficients N depends on the wavelength $\mathbb{L}\varepsilon$, which is typically unknown. A heuristic approach to estimate this wavelength involves analyzing the Fourier transform of the estimator $\widehat{\mu}^{\varepsilon}_{N,T}$ in the regime where $N \gg 1$ and the scaling condition in Theorem 4.3.1 is not fulfilled. In this case, fast-scale oscillations emerge in the estimator, and the wavelength $\mathbb{L}\varepsilon$ can be inferred by identifying the most significant positive frequency $\bar{\xi} > 0$ in the frequency domain, see [Bra00], that is, $\bar{\xi} := \arg\max_{\xi \in \mathcal{M}} |\mathscr{F}(\widehat{\mu}^{\varepsilon}_{N,T})(\xi)|$ where $\mathscr{F}$ is the Fourier transform operator and

$$\mathcal{M} := \left\{ \xi > 0 \,|\, \xi \text{ is a local maximum of } |\mathscr{F}(\widehat{\mu}^{\varepsilon}_{N,T})(\xi)| \right\}. \tag{4.4.17}$$

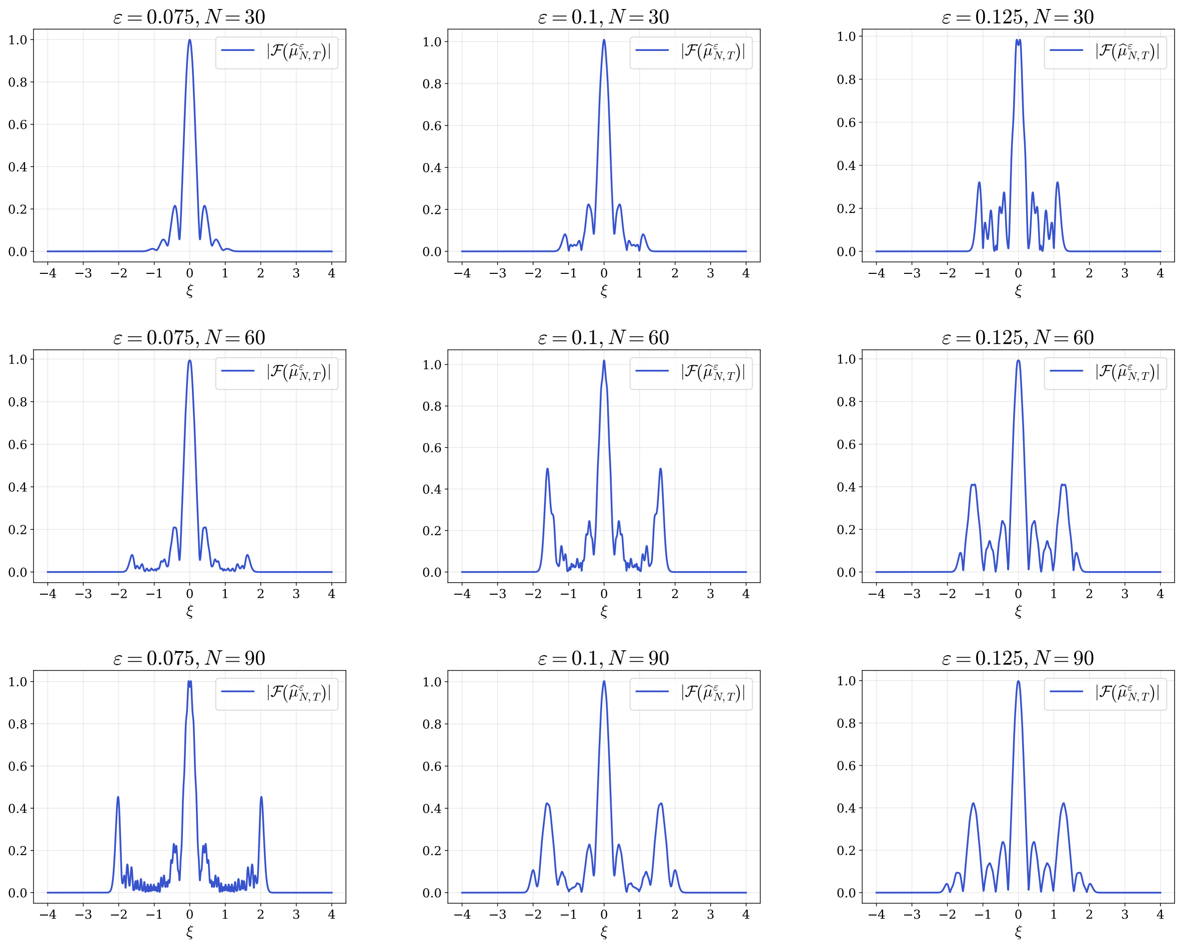
For the sake of simplicity, we assume that $\bar{\xi}$ is well-defined. The Fourier transform of $\widehat{\mu}^{\varepsilon}_{N,T}$ reads

$$\mathscr{F}(\widehat{\mu}^{\varepsilon}_{N,T})(\xi) = \int_{\mathbb{R}} \widehat{\mu}^{\varepsilon}_{N,T}(x) e^{-2\pi i \xi x}\, dx = \sum_{n=0}^{N-1} \widehat{\alpha}^{\varepsilon}_{n,T} \int_{\mathbb{R}} \psi_n(x) e^{-2\pi i \xi x}\, dx, \quad \xi \in \mathbb{R}, \tag{4.4.18}$$

and, using the fact that the Hermite functions ψ_n are eigenfunctions of the Fourier transform [Tha93], we obtain

$$\mathscr{F}(\widehat{\mu}^{\varepsilon}_{N,T})(\xi) = \sqrt{2\pi} \sum_{n=0}^{N-1} (-i)^n \widehat{\alpha}^{\varepsilon}_{n,T} \psi_n(2\pi\xi). \tag{4.4.19}$$

Then, by plotting the magnitude of the Fourier transform $|\mathscr{F}(\widehat{\mu}^{\varepsilon}_{N,T})(\xi)|$, we can identify the dominant frequency $\bar{\xi}$. From this, the dominant wavelength can be estimated as $\mathbb{L}\varepsilon \simeq 1/\bar{\xi}$. Note also that knowing the wavelength $\mathbb{L}\varepsilon$ enables selecting a suitable observation time T, as prescribed in Theorem 4.3.1 and Theorem 4.3.2.



ε	0.075	0.1	0.125
$\widehat{\varepsilon} = 1/(\mathbb{L}\bar{\xi})$	~ 0.079	~ 0.099	~ 0.125

Figure 4.7: Top: Magnitude of the Fourier transform of the density estimator, namely $|\mathcal{F}(\widehat{\mu}^{\varepsilon}_{N,T})|$, across varying values of $\varepsilon \in \{0.075, 0.1, 0.125\}$ and $N \in \{30, 60, 90\}$ with fixed $T = 1000$, where the observations (X_i^{ε}) come from the Euler–Maruyama discretization of (4.1.1) under the setting (4.4.2). Bottom: Inference of the scale separation parameter ε from the dominant frequency $\bar{\xi} = \arg\max_{\xi \in \mathcal{M}} |\mathscr{F}(\widehat{\mu}^{\varepsilon}_{N,T})(\xi)|$ with known period $\mathbb{L} = 2\pi$.

Following this procedure, we compute the magnitude of the Fourier transform of the estimator, namely $|\mathcal{F}(\widehat{\mu}^{\varepsilon}_{N,T})|$, for various values of $\varepsilon \in \{0.075, 0.1, 0.125\}$ and $N \in \{30, 60, 90\}$ through a discretization of (4.4.19). In Figure 4.7, we observe that the dominant frequency $\bar{\xi}$ emerges only when N is sufficiently large. In particular, smaller values of ε require larger values of N for this frequency to become visible. From the final set of plots with $N = 90$ Fourier coefficients, we extract the dominant frequency $\bar{\xi}$ and, assuming the fast-scale period $\mathbb{L} = 2\pi$ is known, we estimate the scale separation parameter via $\widehat{\varepsilon} = 1/(\mathbb{L}\bar{\xi})$. Comparing $\widehat{\varepsilon}$ with the true value of ε, we find strong agreement, demonstrating the effectiveness of this heuristic approach for inferring the wavelength $\mathbb{L}\varepsilon$. This, in turn, provides a practical method for guiding the selection of N and T in the density estimator.

Chapter 5

Concluding Remarks and Outlook

In this thesis, we have considered the problem of statistical estimation for misspecified SDE models when the observations come from a stochastic process, depending on a scale separation parameter, that is assumed to converge weakly to the solution of the SDE model of interest as that scale separation parameter goes to zero. In this sense, we have referred to this setting as estimation under weak misspecification. The main mathematical framework we have been using is the one of homogenization for SDEs and a running example for our theoretical and numerical analysis was the multiscale overdamped Langevin equation. Based on ergodic assumptions of the underlying SDE processes, we have proved a variety of novel limit theorems, with which we were then able to study asymptotic properties of parametric and nonparametric estimators for quantities of homogenized SDE models but under the aspect of weak misspecification.

In Chapter 2, we derived the limit theorems which then became indispensable tools for the later chapters. Starting from the development of a qualitative and a quantitative mean ergodic theorem for SDE solutions, we were able to eventually prove a central limit theorem for time averages of weakly perturbed observations. Since we aimed for results where we do not need to start the stochastic process in stationarity, a larger intermediate step in this endeavor was proving that certain tail probabilities of the weakly perturbed observations vanish as the small scale parameter goes to zero. We then continued to establish the validity of the various limit theorems for a specific class of SDEs, the multiscale overdamped Langevin equation and its homogenized limit Langevin equation.

Although the established limit theorems turned out be useful for proving asymptotic properties of statistical estimators, their applicability is nonetheless still limited. On the one hand, we have, for the sake of simplicity, only considered continuous observations, which is, of course, not properly aligned with numerical implementations of estimators. Thus, it would be desirable to extend the gist of the limit theorems to the more realistic case of discrete observations and perhaps even to low-frequency observations. On the other hand, most of the results have only been proved for 1-dimensional SDE processes, which, as we recall, has bought us some additional structural properties, such as the Markov property and the existence of a generator for the slow variable, and thus enabled our ideas to flourish more easily. However, multidimensional versions of the limit theorems remain unanswered for now. Regarding this matter, a first starting point for future research may be the several remarks on multidimensional extensions in Chapter 2.

In Chapter 3, we first introduced a statistical estimation framework for parametric models under weak misspecification. Under this framework, we reconfirmed the known fact from the existing literature that maximum likelihood estimation is asymptotically biased when confronted with multiscale observations. We then proposed a general minimum distance estimator for parameters of homogenized SDE models under ergodic assumptions. The estimator is based on a minimization problem for the L^2-distance between the characteristic function of the invariant measure of the homogenized SDE and an empirical characteristic function depending on the observations. We proved that the minimum distance estimator exhibits many favorable asymptotic properties, like robustness under weakly perturbed observations for different limit orders and even asymptotic normality. These findings were also numerically reinforced in a comprehensive empirical Monte Carlo parameter study.

Apart from theoretical questions, which primarily rely on an extension of the limit theorems in the sense we have already indicated in the last paragraph, there are certain methodological and computational challenges concerning the MDE that ought to be remarked. A striking methodological constraint of the MDE is the requirement of a prior estimation of a homogenized quantity, say the diffusion coefficient of the homogenized SDE model, if we are trying to estimate the drift coefficient. This is per se an identification problem of the underlying parametric models because we are trying to estimate quantities from the invariant measure, for which identifiability issues arise even in settings without misspecification. This is usually deemed unproblematic in specified model settings because one can, for example, estimate the diffusion coefficient consistently with the classical quadratic variation estimator purely from observations and then use this estimate in a subsequent drift estimation method. However, as we recall from the review on related works in Section 1.3, the quadratic variation estimator is asymptotically biased under weak misspecification, so that the identifiability issue persists for the MDE. In a way, this shifts the actual estimation problem under weak misspecification from the MDE to the prior estimation method. That being said, if one has an asymptotically unbiased prior estimator, then the MDE performs robustly, so that the approach of combining two robust estimators might be a direction worth investigating.

Concerning computational aspects, the MDE is overall numerically tractable, even in specific multidimensional settings, but it does suffer from the curse of dimensionality in non-Gaussian settings. One potential way of mitigating this problem lies in the following idea. Let us return to the distance formula (3.2.8) of Example 3.2.1, while ignoring the first double integral as it is irrelevant for the minimization and assuming that the invariant measure $\mu(\vartheta)$ has a Lebesgue-density μ_ϑ, and consider

$$\Delta_T^\circ(\vartheta, X) := -\frac{2}{T}\int_0^T (k_\beta * \mu_\vartheta)(X_t)\, dt + \int_{\mathbb{R}^d} (k_\beta * \mu_\vartheta)(x)\mu_\vartheta(x)\, dx.$$

Using Fubini's theorem, we can rewrite this to

$$\Delta_T^\circ(\vartheta, X) = \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} \mu_\vartheta(x-y)\mu_\vartheta(x)\, dx - \frac{2}{T}\int_0^T \mu_\vartheta(X_t - y)\, dt\right] k_\beta(y)\, dy.$$

Recalling that k_β is given by $k_\beta = (2\pi/\beta^2)^{d/2}\varphi_{1/\beta}$ and thus corresponding to the Lebesgue-density of a finite measure, we obtain

$$\Delta_T^\circ(\vartheta, X) = (2\pi^2)^{d/2}\mathbb{E}_Z\left[\int_{\mathbb{R}^d} \mu_\vartheta(x-Z)\mu_\vartheta(x)\, dx - \frac{2}{T}\int_0^T \mu_\vartheta(X_t - Z)\, dt\right],$$

conditioned on a fixed sample path X, where $Z \sim \mathcal{N}_d(0, \beta^{-2}I_d)$ is independent of X. Since we may draw samples of Z independently of X, the last expression for $\Delta_T^\circ(\vartheta, X)$ gives passage to optimization schemes through Monte Carlo estimators and stochastic gradient descent, or, more precisely, through stochastic approximation as originally introduced in [RM51; KW52]. This would be computationally more tractable as we no longer have to deal with the integral involving the convolution $k_\beta * \mu_\vartheta$ and essentially reduced it to pointwise evaluations of the convolution $\mu_\vartheta * \mu_\vartheta$. Though this approach seems promising, it would still need to come under proper mathematical scrutiny.

In Chapter 4, we proposed a projection-based density estimator for the estimation of the invariant density of a homogenized Langevin equation with observations coming from the multiscale overdamped Langevin equation. The estimator is based on a finite Fourier expansion of the homogenized invariant density in $L^2(\mathbb{R})$ with respect to the Hermite orthonormal basis and an approximation of the Fourier coefficients via time averages of the observations. As before, the driving underlying principle here is ergodicity. We proved that the density estimator is robust and converges to a Gaussian random element in $L^2(\mathbb{R})$ as long as the number of Fourier coefficients and the final observation time scale appropriately with the scale separation parameter. A numerical simulation study further corroborated and supplemented the established theoretical results.

As with the MDE, most generalizations of the theoretical results regarding the density estimator depend on an extension of the limit theorems of Chapter 2. However, Remark 4.3.8 already shows a possible extension of the acquired central limit theorem to a space beyond plain $L^2(\mathbb{R})$, which should be a result attainable without new limit theorems. Another possible direction for future research concerns nonparametric drift estimation. More specifically, for the homogenized drift $\mathcal{V} := -\mathcal{K}V$, the following identities hold

$$\mathcal{V}' = \frac{\bar{\Sigma}^2}{2}\frac{\mu'}{\mu}, \qquad \mathcal{V} = \frac{\bar{\Sigma}^2}{2}\left(\log(\mu) + \log(Z)\right), \qquad \mu = \frac{1}{Z}\exp\left(-\frac{2\mathcal{K}}{\bar{\Sigma}^2}V\right),$$

which express both $\mathcal{V}'$ and $\mathcal{V}$ in terms of the homogenized invariant density μ. Estimating $\mathcal{V}'$ requires the additional estimation of μ' while estimating $\mathcal{V}$ necessitates an extra condition, e.g., a zero mean condition with respect to μ, to fix the additive constant. Moreover, knowledge of the homogenized diffusion coefficient $\bar{\Sigma}^2 = 2\sigma\mathcal{K}$ is required in both cases and must therefore be estimated in advance, which brings us to the same challenge that we have discussed beforehand with the MDE. Nonparametric estimators for $\mathcal{V}'$ and $\mathcal{V}$ would then be obtained by substituting μ and μ' with their respective estimators in the previous identities.

Finally, the nonparametric estimation has revealed that the procedure depends sensitively on the scale separation parameter as reflected by the choice of the number of Fourier coefficients and the final observation time. As we recall from Section 1.3, this is also true for numerous other statistical methodologies where hyperparameters are incorporated into estimators that depend on the scale separation parameter. Although we provided a heuristic inference method for identifying the small scale parameter from observations via Fourier transforms and setting a high number of Fourier coefficients, the estimation of this parameter remains highly non-trivial for more complex multiscale systems with higher dimensionality and explicit coupling between slow and fast variable. One reason for this difficulty can probably be attributed to the fact that we, in general, do not observe the fast variable of the multiscale system, or, at least, this is what is often assumed. Hence, an extremely valuable future contribution could be the investigation and analysis of data-driven inference methods for this scale separation parameter.

Chapter A

Stochastic Calculus

In this appendix, we collect several different results and concepts from stochastic calculus and the theory of Markov processes which were used in the main material. Even though these topics are standard in the existing literature, they are somewhat more advanced in the sense that they go beyond Itô's formula and are therefore worth mentioning for the purpose of easier reference to the main text. We will primarily present and take material from [KS91; KMN12] in the way that we require, but most of it can also be found in other standard references, such as [RY98; Pro03; PR14; SV79; Øks03; KS14; Lor17; DZ96; Dyn65; IM96; Kal21]. The first section is based on [KS91, Section 1.5, Section 3.2, Section 3.3, Section 3.7] and the second section on [KMN12, Section 3.1, Section 3.3, Section 3.4].

A.1 Martingales, Quadratic Variation, and Local Time

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space equipped with a filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$ that satisfies the usual conditions, which, as we recall, means that $(\mathcal{F}_t)_{t\in[0,\infty)}$ is right-continuous and $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets of $\mathcal{F}$. The appearing stochastic processes in this section are all 1-dimensional real-valued stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition A.1.1. *A continuous process $X = (X_t)_{t\in[0,\infty)}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, adapted to the filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$, with $\mathbb{E}\,|X_t| < \infty$ for every $t \in [0,\infty)$ is a martingale if, for every $0 \leq s < t < \infty$, we have*

$$\mathbb{E}\left[X_t \,|\, \mathcal{F}_s\right] = X_s, \quad \mathbb{P}\text{-}a.s. \tag{A.1.1}$$

If $\mathbb{E}\,|X_t|^2 < \infty$ for every $t \in [0,\infty)$, then we say that X is a continuous square-integrable martingale and if, in addition, $X_0 = 0$, $\mathbb{P}$-a.s., then we write $X \in \mathscr{M}_2^{\mathrm{c}}$.

For any $X \in \mathscr{M}_2^{\mathrm{c}}$, it is true, by the Doob-Meyer decomposition, that $X^2 = M + A$ with a continuous martingale M and a natural increasing process A such that $M_0 = A_0 = 0$, $\mathbb{P}$-a.s. As we define next, this latter process A constitutes the quadratic variation process of X. See [KS91, Section 1.4] for the concepts of increasing and natural processes.

Definition A.1.2. *For $X \in \mathscr{M}_2^{\mathrm{c}}$, we define the quadratic variation process of X to be the process $\langle X\rangle := A$, where A is the natural increasing process in the Doob-Meyer decomposition of X^2, that is, $\langle X\rangle$ is the unique (up to indistinguishability) adapted, natural increasing, continuous process such that $\langle X\rangle_0 = 0$, $\mathbb{P}$-a.s., and $X^2 - \langle X\rangle$ is a martingale.*

Definition A.1.3. *For $X, Y \in \mathscr{M}_2^{\mathrm{c}}$, we define their cross-variation process $\langle X, Y\rangle$ by*

$$\langle X, Y\rangle_t := \frac{1}{4}\left[\langle X+Y\rangle_t - \langle X-Y\rangle_t\right], \quad t \in [0,\infty). \tag{A.1.2}$$

Definition A.1.4. *Let $X = (X_t)_{t\in[0,\infty)}$ be a continuous process that is adapted to the filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$. If there exists a non-decreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times of $(\mathcal{F}_t)_{t\in[0,\infty)}$ such that $(X_{t\wedge\tau_n})_{[0,\infty)} \in \mathscr{M}_2^{\mathrm{c}}$ for each $n \in \mathbb{N}$ and $\mathbb{P}(\lim_{n\to\infty}\tau_n = \infty) = 1$, then we say that X is a continuous local martingale. If, in addition, $X_0 = 0$, $\mathbb{P}$-a.s., then we write $X \in \mathscr{M}^{\mathrm{c,\,loc}}$.*

For $X \in \mathscr{M}^{\mathrm{c,\,loc}}$, one can show that there exists a unique (up to indistinguishability) adapted, increasing, continuous process, also denoted $\langle X\rangle$, such that $\langle X\rangle_0 = 0$, $\mathbb{P}$-a.s., and $X^2 - \langle X\rangle \in \mathscr{M}^{\mathrm{c,\,loc}}$. We call this process $\langle X\rangle$ the quadratic variation process of X. Likewise, the cross-variation process $\langle X, Y\rangle$ of $X, Y \in \mathscr{M}^{\mathrm{c,\,loc}}$ can be defined through the same polarization identity (A.1.2) as in Definition A.1.3.

In the definition of the stochastic integral, one can take $M \in \mathscr{M}_2^{\mathrm{c}}$ as an integrator as long as the progressively measurable integrand X satisfies

$$\mathbb{E}\int_0^T X_t^2\, d\langle M\rangle_t < \infty, \quad \forall\, T \in [0,\infty).$$

However, one can weaken these conditions and extend the definition of the stochastic integral to integrators given by local martingales $M \in \mathscr{M}^{\mathrm{c,\,loc}}$ and to integrands that are only $\mathbb{P}$-a.s. locally integrable. The last requirement can be formalized via the collection $\mathscr{P}^*(M)$ of (equivalence classes of) all progressively measurable processes $X = (X_t)_{t\in[0,\infty)}$ satisfying

$$\mathbb{P}\left(\int_0^T X_t^2\, d\langle M\rangle_t < \infty\right) = 1, \quad \forall\, T \in [0,\infty). \tag{A.1.3}$$

Through a localization procedure, one can then define the stochastic integral with respect to a local martingale integrator $M \in \mathscr{M}^{\mathrm{c,\,loc}}$ and integrands from $\mathscr{P}^*(M)$ as follows.

Definition A.1.5. *Consider $M \in \mathscr{M}^{\mathrm{c,\,loc}}$ and $X \in \mathscr{P}^*(M)$. The stochastic integral, denoted $I(X) := (\int_0^t X_s\, dM_s)_{t\in[0,\infty)}$, is the unique (up to indistinguishability) continuous local martingale $\Phi \in \mathscr{M}^{\mathrm{c,\,loc}}$ that satisfies*

$$\langle \Phi, N\rangle_t = \int_0^t X_s\, d\langle M, N\rangle_s, \quad \mathbb{P}\text{-a.s.}, \tag{A.1.4}$$

for every $N \in \mathscr{M}^{\mathrm{c,\,loc}}$ and $t \in [0,\infty)$. In particular, we have

$$\langle I(X)\rangle_t = \int_0^t X_s^2\, d\langle M\rangle_s, \quad \mathbb{P}\text{-a.s.}, \tag{A.1.5}$$

for all $t \in [0,\infty)$.

We come now to the definition of a continuous semimartingale. The most prominent example of a continuous (non-Markovian) semimartingale that we have encountered in the main text is the slow variable X^ε of the multiscale system (1.2.1).

Definition A.1.6. *An adapted process $X = (X_t)_{t\in[0,\infty)}$ is a continuous semimartingale if it has the decomposition*

$$X_t = X_0 + M_t + B_t, \quad t \in [0,\infty), \quad \mathbb{P}\text{-a.s.}, \tag{A.1.6}$$

where $M := (M_t)_{t\in[0,\infty)} \in \mathscr{M}^{\mathrm{c,\,loc}}$ and $B := (B_t)_{t\in[0,\infty)}$ is a continuous, adapted process with $B_0 = 0$, $\mathbb{P}$-a.s., and finite total variation on $[0,t]$ for any $t \in [0,\infty)$. This decomposition is unique up to indistinguishability.

A direct consequence of Itô's formula in multiple dimensions is the following stochastic integration by parts formula for continuous semimartingales.

Lemma A.1.7. *Suppose we have two continuous semimartingales with their respective decompositions*

$$X_t = X_0 + M_t + B_t, \quad Y_t = Y_0 + N_t + C_t, \quad t \in [0,\infty), \quad \mathbb{P}\text{-a.s.},$$

where $M, N \in \mathscr{M}^{\mathrm{c,\,loc}}$ and B, C are adapted, continuous processes with $B_0 = C_0 = 0$, $\mathbb{P}$-a.s., and finite total variation on $[0,t]$ for any $t \in [0,\infty)$. Then, for any $t \in [0,\infty)$, we have

$$\int_0^t X_s\, dY_s = X_tY_t - X_0Y_0 - \int_0^t Y_s\, dX_s - \langle M, N\rangle_t, \quad \mathbb{P}\text{-a.s.} \tag{A.1.7}$$

Another result that follows from Itô's formula are the so-called Burkholder-Davis-Gundy inequalities, which are very useful martingale moment inequalities.

Theorem A.1.8. *Let $M \in \mathscr{M}^{\mathrm{c,\,loc}}$. For every $m \in (0,\infty)$, there exist universal constants $b_m, B_m > 0$, depending only on m, such that*

$$b_m \mathbb{E}\left[\langle M\rangle_\tau^m\right] \le \mathbb{E}\left[\left(\sup_{0\le s\le\tau} |M_s|\right)^{2m}\right] \le B_m \mathbb{E}\left[\langle M\rangle_\tau^m\right] \tag{A.1.8}$$

holds for any stopping time τ of $(\mathcal{F}_t)_{t\in[0,\infty)}$.

Finally, we present several results and formulas surrounding the local time of a continuous semimartingale.

Theorem A.1.9. *Let X be a continuous semimartingale of the form* (A.1.6). *Then there exists local time for X, i.e., a nonnegative random field $L(X) := (L_t^z(X))_{t\in[0,\infty),z\in\mathbb{R}}$ such that the following hold:*

i) The map $(t,z,\omega) \mapsto L_t^z(X)(\omega)$ is measurable and, for each fixed (t,z), the random variable $L_t^z(X)$ is $\mathcal{F}_t$-measurable.

ii) For every fixed $z \in \mathbb{R}$ and $\omega \in \Omega$, the map $t \mapsto L_t^z(X)(\omega)$ is continuous and non-decreasing with $L_0^z(X)(\omega) = 0$.

iii) For every Borel-measurable function $k\colon \mathbb{R} \to [0,\infty)$, the occupation density formula

$$\int_0^t k(X_s)\, d\langle M\rangle_s = \int_{\mathbb{R}} k(z) L_t^z(X)\, dz, \quad \mathbb{P}\text{-a.s.}, \tag{A.1.9}$$

holds for every $t \in [0,\infty)$.

iv) For every $z \in \mathbb{R}$, the Meyer-Tanaka formula

$$|X_t - z| = |X_0 - z| + \int_0^t \operatorname{sgn}(X_s - z)\, dX_s + +L_t^z(Y), \quad \mathbb{P}\text{-}a.s., \tag{A.1.10}$$

holds for every $t \in [0, \infty)$, where the sign function is defined to be left-continuous, that is, $\operatorname{sgn} := \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0]}$.

A.2 Markov Processes and Strong SDE Solutions

Let the underlying probability space and the filtration be as in the previous section. An adapted process $(X_t)_{t\in[0,\infty)}$ with values in $\mathbb{R}^d$, $d \in \mathbb{N}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with $(\mathcal{F}_t)_{t\in[0,\infty)}$, is called a Markov process if, for all $B \in \mathscr{B}(\mathbb{R}^d)$ and $0 \leq s < t$,

$$\mathbb{P}(X_t \in B \,|\, \mathcal{F}_s) = \mathbb{P}(X_t \in B \,|\, X_s), \quad \mathbb{P}\text{-a.s.} \tag{A.2.1}$$

Associated to a Markov process, it can be proved that there exists a so-called transition probability function, which we define as follows.

Definition A.2.1. *A function $(s, x, t, B) \mapsto P(s, x, t, B)$, defined for $0 \leq s < t < \infty$, $x \in \mathbb{R}^d$, and $B \in \mathscr{B}(\mathbb{R}^d)$, is called a transition probability function if it satisfies the following set of conditions:*

i) $P(s, x, s, \mathbb{R}^d \backslash \{x\}) = 0$ for all $s \geq 0$ and $x \in \mathbb{R}^d$,

ii) $\mathscr{B}(\mathbb{R}^d) \ni B \mapsto P(s, x, t, B)$ is a probability measure on $\mathbb{R}^d$ for all $0 \leq s < t$ and $x \in \mathbb{R}^d$,

iii) $x \ni \mathbb{R}^d \mapsto P(s, x, t, B)$ is Borel-measurable for all $0 \leq s < t$ and $B \in \mathscr{B}(\mathbb{R}^d)$,

iv) for all $B \in \mathscr{B}(\mathbb{R}^d)$ and $0 \leq s < t$,

$$\mathbb{P}(X_t \in B \,|\, \mathcal{F}_s) = P(s, X_s, t, B), \quad \mathbb{P}\text{-}a.s., \tag{A.2.2}$$

v) for all $0 \leq r < s < t < \infty$ and $B \in \mathscr{B}(\mathbb{R}^d)$,

$$P(r, x, t, B) = \int_{\mathbb{R}^d} P(s, y, t, B) P(r, x, s, dy). \tag{A.2.3}$$

It is said to be time-homogeneous if $P(s, x, t+s, B) = P(0, x, t, B)$ for all $0 \leq s < t < \infty$, $x \in \mathbb{R}^d$, and $B \in \mathscr{B}(\mathbb{R}^d)$. In this case, one can use the notation $P(s, x, t, B) = P(x, t - s, B)$ without harm. A time-homogeneous transition probability function P is stochastically continuous if $\lim_{t\to 0^+} P(x, t, B_\delta(x)) = 1$ *for any $\delta > 0$ and $x \in \mathbb{R}^d$.*

Equation (A.2.3) is known as the Chapman-Kolmogorov equation. Based on a time-homogeneous transition probability function P, one can associate the following family of operators defined on $C_0(\mathbb{R}^d)$, the space of continuous functions on $\mathbb{R}^d$ vanishing at infinity equipped with the supremum-norm, and $t \in [0, \infty)$

$$T_t f(x) := \int_{\mathbb{R}^d} f(y) P(x, t, dy), \quad f \in C_0(\mathbb{R}^d). \tag{A.2.4}$$

A time-homogeneous transition probability function P for which T_t maps $C_0(\mathbb{R}^d)$ into itself, i.e., $T_t(C_0(\mathbb{R}^d)) \subset C_0(\mathbb{R}^d)$, is called a Feller transition probability function. The Chapman-Kolmogorov equation (A.2.3) yields the semigroup property $T_{t+s} = T_t T_s$. Furthermore, if P is also stochastically continuous, then $\lim_{t\to 0^+} T_t f = f$ for any $f \in C_0(\mathbb{R}^d)$, so that $(T_t)_{t\in[0,\infty)}$ forms a strongly continuous semigroup in $C_0(\mathbb{R}^d)$. As a result of semigroup theory, one can thus define the infinitesimal generator $\mathcal{A}$ of the strongly continuous semigroup $(T_t)_{t\in[0,\infty)}$ by the standard formula

$$\mathcal{A}f := \lim_{t\to 0^+} \frac{T_t f - f}{t}, \quad f \in \operatorname{dom}(\mathcal{A}), \tag{A.2.5}$$

with the domain given by

$$\operatorname{dom}(\mathcal{A}) := \left\{ f \in C_0(\mathbb{R}^d) \,\middle|\, \exists g \in C_0(\mathbb{R}^d) : g = \lim_{t\to 0^+} \frac{T_t f - f}{t} \right\}. \tag{A.2.6}$$

A standard example of a Markov process with Feller transition probability function is given by the solution to an SDE. In order to present this and related results, we choose a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$ such that the filtration satisfies the usual conditions and the whole space can accommodate an r-dimensional standard Brownian motion $W := (W_t)_{t\in[0,\infty)}$, adapted to $(\mathcal{F}_t)_{t\in[0,\infty)}$, where $r \in \mathbb{N}$; refer to [KS91, Section 5.2] for this construction. We consider the time-homogeneous d-dimensional SDE

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \quad t \in [0,\infty), \tag{A.2.7}$$

with Borel-measurable functions $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times r}$ and initial condition $X_0 = x_0 \in \mathbb{R}^d$. Let us recall the concept of a strong solution to SDE (A.2.7).

Definition A.2.2. *A strong solution of the stochastic differential equation* (A.2.7)*, on the given probability space with filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$ and with respect to the fixed Brownian motion W and initial condition $x_0 \in \mathbb{R}^d$, is a continuous process $X = (X_t)_{t\in[0,\infty)}$ with the following properties:*

i) X is adapted to the filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$,

ii) $X_0 = x_0$, $\mathbb{P}$-a.s.,

iii) for every $1 \le i \le d$, $1 \le j \le r$, and $t \in [0,\infty)$,

$$\mathbb{P}\left(\int_0^t |b_i(X_s)| + |\sigma_{ij}(X_s)|^2\, ds < \infty \right) = 1 \tag{A.2.8}$$

holds, where b_i and σ_{ij} are the component functions of b and σ in (A.2.7)*.*

iv) the integral version of (A.2.7)*, namely*

$$X_t = X_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s, \quad \mathbb{P}\text{-a.s.}, \tag{A.2.9}$$

holds for all $t \in [0,\infty)$.

Whenever W is an r-dimensional standard Brownian motion on the given probability space with filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$, the initial condition $x_0 \in \mathbb{R}^d$ is fixed, and $X, \tilde{X}$ are two strong solutions of (A.2.7) *such that they are unique up to indistinguishability, i.e.,* $\mathbb{P}(X = \tilde{X}$ *on* $[0,\infty)) = 1$, *then we say that X is the unique strong solution of* (A.2.7) *with $X_0 = x_0$.*

The following theorem gives a general existence and uniqueness result for SDEs with only local regularity assumptions of the coefficients. At least in the context of this theorem, the function v that satisfies the conditions (A.2.12) and (A.2.13) is commonly called a Lyapunov function.

Theorem A.2.3. *Let $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times r}$, $d, r \in \mathbb{N}$, be continuous functions such that for every $N > 0$ there exists some constant $K_N > 0$ with*

$$|b_i(x) - b_i(y)| + |\sigma_{ij}(x) - \sigma_{ij}(y)| \le K_N|x-y|, \qquad x, y \in B_N(0), \tag{A.2.10}$$

$$|b_i(x)| + |\sigma_{ij}(x)| \le K_N(1+|x|), \qquad x \in B_N(0), \tag{A.2.11}$$

for all $1 \le i \le d$ and $1 \le j \le r$. Moreover, suppose that there exists a nonnegative function $v \in C^2(\overline{B_R(0)}^c)$ for some $R > 0$ such that for a constant $c > 0$

$$\frac{1}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2 v}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial v}{\partial x_i} \le cv, \quad \text{on } \overline{B_R(0)}^c, \tag{A.2.12}$$

$$\lim_{R\to\infty} \inf_{|x|>R} v(x) \to \infty, \tag{A.2.13}$$

where $a := \sigma\sigma^\top \in \mathbb{R}^{d\times d}$ is the positive semi-definite diffusion matrix. Then there exists a unique strong solution X to (A.2.7) *with $X_0 = x_0$ satisfying the inequality*

$$\mathbb{E}\, v(X_t) \le v(x_0)\exp(ct), \quad t \in [0,\infty). \tag{A.2.14}$$

This solution is a Markov process with a time-homogeneous, stochastically continuous Feller transition probability function that is given, for $t > 0$, $x \in \mathbb{R}^d$, and $B \in \mathscr{B}(\mathbb{R}^d)$, by the relation $P(x,t,B) = \mathbb{P}(X_t^x \in B)$, where X^x is the unique strong solution to the SDE

$$dX_t^x = b(X_t^x)\,dt + \sigma(X_t^x)\,dW_t, \quad t \in [0,\infty), \quad X_0^x = x. \tag{A.2.15}$$

Based on the last theorem, it is customary to associate a second-order differential operator with drift coefficient b and diffusion coefficient a given by

$$\mathcal{L}f := \frac{1}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial f}{\partial x_i}. \tag{A.2.16}$$

It turns out, see, for example, [Øks03], that the infinitesimal generator $\mathcal{A}$ of the Markov process given by the unique strong solution X to (A.2.7) coincides with the second-order differential operator $\mathcal{L}$ on $C_c^2(\mathbb{R}^d)$, the space of twice continuously differentiable functions having compact support, that is,

$$\mathcal{A}f = \mathcal{L}f, \quad f \in C_c^2(\mathbb{R}^d). \tag{A.2.17}$$

By a slight abuse of terminology, it is therefore common to call $\mathcal{L}$ the generator of X.

Lastly, when X^x is like in Theorem A.2.3, then it is not only a Markov process but even a strong Markov process, which is to say that, for any $x \in \mathbb{R}^d$, $B \in \mathscr{B}(\mathbb{R}^d)$, any stopping time τ of $(\mathcal{F}_t)_{t\in[0,\infty)}$, and any $\mathcal{F}_\tau$-measurable random variable η satisfying $\eta \geq \tau$, $\mathbb{P}$-a.s., we have

$$\mathbb{P}(X^x_\eta \in B \,|\, \mathcal{F}_\tau) = P(X^x_\tau, \eta - \tau, B), \quad \mathbb{P}\text{-a.s.} \tag{A.2.18}$$

Chapter B

Weak Convergence of Measures

Similar to the previous appendix, we present here general concepts and especially useful sufficient conditions for tightness and weak convergence of probability measures on different spaces that have been considered in the main text. This is a very well-studied branch of mathematics and standard references include, among others, [Bog18; Bil99; VW23; Kal21]. Particularly interesting cases of weak convergence of diffusion processes and related questions, such as tightness in the context of martingales, are thoroughly discussed in [SV79]. We will mostly present material, adjusted to our purposes, from [Bog18, Chapter 2] and [KS91, Section 2.4] but also state results from [Hen24, Chapter 17] for the case of a separable Hilbert space.

B.1 General Metric Spaces

Let (E, d) be a metric space equipped with the metric d and the Borel σ-algebra $\mathscr{B}(E)$.

Definition B.1.1. *Let $(\mathbb{P}_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on E and $\mathbb{P}$ another probability measure on this metric space. We say that $\mathbb{P}_n$ converges weakly to $\mathbb{P}$ on E if*

$$\lim_{n\to\infty} \int_E f(x)\, d\mathbb{P}_n(x) = \int_E f(x)\, d\mathbb{P}(x) \tag{B.1.1}$$

holds for every $f \in C_b(E)$, the space of bounded, continuous functions on E.

Definition B.1.2. *If X_n, $n \in \mathbb{N}$, and X are E-valued random elements that are defined on $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$, respectively, then we say that X_n converges weakly to X in E as $n \to \infty$ if the induced probability measures $\mathbb{P}_n^{X_n}$ converge weakly to $\mathbb{P}^X$ on E as $n \to \infty$. In this case, we write $X_n \xrightarrow{\mathcal{D}} X$ in E as $n \to \infty$. Equivalently, $X_n \xrightarrow{\mathcal{D}} X$ in E as $n \to \infty$ if and only if*

$$\lim_{n\to\infty} \mathbb{E}_n f(X_n) = \mathbb{E} f(X), \tag{B.1.2}$$

for every $f \in C_b(E)$, where $\mathbb{E}_n$ and $\mathbb{E}$ denote expectations with respect to $\mathbb{P}_n$ and $\mathbb{P}$, respectively.

We stress the fact again that X_n and X need not be defined on the same probability space. If they are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then we may consider the following convergence mode.

Definition B.1.3. *Let X_n, $n \in \mathbb{N}$, and X be E-valued random elements defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Then X_n converges in probability to X if, for every $\delta > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(d(X_n, X) > \delta) = 0. \tag{B.1.3}$$

In this case, we write $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$. In particular, if $\mathbb{P}(X = a) = 1$ for some $a \in E$, then we also write $X_n \xrightarrow{\mathbb{P}} a$ as $n \to \infty$.

The following characterization theorem is due to Alexandroff but often goes under the name of Portmanteau theorem.

Theorem B.1.4. *Let $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $\mathbb{P}$ be probability measures on E. Then the following conditions are equivalent:*

i) $\mathbb{P}_n$ *converges weakly to* $\mathbb{P}$ *on* E.

ii) For every bounded, uniformly continuous function f on E,

$$\lim_{n\to\infty} \int_E f(x)\, d\mathbb{P}_n(x) = \int_E f(x)\, d\mathbb{P}(x). \tag{B.1.4}$$

iii) For every closed subset F of E, $\limsup_{n\to\infty} \mathbb{P}(F) \leq \mathbb{P}(F)$.

iv) For every open subset G of E, $\liminf_{n\to\infty} \mathbb{P}(G) \geq \mathbb{P}(G)$.

v) For every subset A of E with $\mathbb{P}(\partial A) = 0$, $\lim_{n\to\infty} \mathbb{P}(A) = \mathbb{P}(A)$.

A result we heavily make use of is Slutzky's lemma.

Theorem B.1.5. *Let (X_n, Y_n), $n \in \mathbb{N}$, and X be E-valued random elements defined on the same probability space. If $X_n \xrightarrow{\mathcal{D}} X$ in E and $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$, then $Y_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.*

We also need the following most basic form of the continuous mapping theorem.

Theorem B.1.6. *Let (E, d), (E', d') be metric spaces and $h\colon E \to E'$ a continuous map. Suppose X_n, $n \in \mathbb{N}$, and X are E-valued random elements such that $X_n \xrightarrow{\mathcal{D}} X$ in E as $n \to \infty$. Then $h(X_n) \xrightarrow{\mathcal{D}} h(X)$ in E' as $n \to \infty$.*

The following result is the well-known Tietze–Urysohn extension theorem from topology, which is quite useful together with the continuous mapping theorem. For example, we have seen a combined application of these two theorems in the proof of Proposition 3.2.4.

Theorem B.1.7. *Every bounded, continuous function f defined on a closed set F of a metric space E extends to a bounded, continuous function $\tilde{f}$ on the whole space with the same supremum-norm, that is, $\tilde{f}|_F = f$ and $\sup_{x\in F} |f(x)| = \sup_{x\in E} |\tilde{f}(x)|$.*

We conclude this section with the definition of tightness and relative compactness of a family of probability measures.

Definition B.1.8. *Let $\mathscr{P}$ be a non-empty family of probability measures on the metric space E. We say that $\mathscr{P}$ is tight if, for every $\delta > 0$, there exists a compact set $K_\delta \subset E$ such that $\mathbb{P}(K_\delta) \geq 1-\delta$ for every $\mathbb{P} \in \mathscr{P}$. We say that $\mathscr{P}$ is relatively compact if every sequence of elements in $\mathscr{P}$ contains a weakly convergent subsequence. If $(X_n)_{n\in\mathbb{N}}$ is a family of E-valued random elements, then we say that $(X_n)_{n\in\mathbb{N}}$ is tight (relatively compact) if the family of induced probability measures $(\mathbb{P}^{X_n})_{n\in\mathbb{N}}$ is tight (relatively compact).*

B.2 Continuous Functions on $\mathbb{R}^d$

Let us consider $C(\mathbb{R}^d)$, the space of continuous functions on $\mathbb{R}^d$, equipped with the topology of compact convergence induced by the metric

$$\rho(f,g) := \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\rho_n(f,g)}{1+\rho_n(f,g)}, \quad \rho_n(f,g) := \sup_{x\in[-n,n]^d} |f(x)-g(x)|, \quad f,g \in C(\mathbb{R}^d), \tag{B.2.1}$$

turning $C(\mathbb{R}^d)$ into a complete, separable metric space. Recall here that convergence of a sequence $(f_k)_{k\in\mathbb{N}} \subset C(\mathbb{R}^d)$ to an element $f \in C(\mathbb{R}^d)$ with respect to the given metric ρ can be determined by the criterion

$$\rho(f_k, f) \to 0 \quad \text{as } k \to \infty \quad \Leftrightarrow \quad \rho_n(f_k, f) \to 0 \quad \forall n \in \mathbb{N}, \quad \text{as } k \to \infty. \tag{B.2.2}$$

If X_n, $n \in \mathbb{N}$, and X are $C(\mathbb{R}^d)$-valued random elements, then it is a well-known fact that convergence of the finite-dimensional distributions of X_n to those of X, i.e.,

$$(X_n(u_1), \ldots, X_n(u_l)) \xrightarrow{\mathcal{D}} (X(u_1), \ldots, X(u_l))\,, \quad \text{as } n \to \infty, \tag{B.2.3}$$

for any finite points $u_1, \ldots, u_l \in \mathbb{R}^d$, $l \in \mathbb{N}$, is not enough for weak convergence of X_n to X in $C(\mathbb{R}^d)$. The missing link here is tightness, which essentially ensures that no probability mass escapes to infinity. More concretely, we have the following result, which is a consequence of the famous theorem by Prokhorov about the equivalence of tightness and relative compactness of measures on complete, separable metric spaces.

Theorem B.2.1. *Let $(X_n)_{n\in\mathbb{N}}$ be a tight sequence of $C(\mathbb{R}^d)$-valued random elements and X a $C(\mathbb{R}^d)$-valued random element such that the finite-dimensional distributions of X_n converge to those of X in the sense of* (B.2.3)*. Then $X_n \xrightarrow{\mathcal{D}} X$ in $C(\mathbb{R}^d)$ as $n \to \infty$.*

The convergence of the finite-dimensional distributions can often be verified by studying characteristic functions, so that one only needs to verify tightness. A set of practical sufficient conditions is given in the following lemma, which is a corollary of a variant of the Arzelà–Ascoli theorem and the Kolmogorov–Chentsov continuity theorem.

Lemma B.2.2. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $C(\mathbb{R}^d)$-valued random elements on $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following conditions*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\,|X_n(0)|^\alpha < \infty, \tag{B.2.4}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}\,|X_n(u) - X_n(v)|^\beta \leq C_k |u-v|^{d+\gamma}, \quad \forall u,v \in \{x \in \mathbb{R}^d \,|\, |x| \leq k\}, \quad \forall k \in \mathbb{N}, \tag{B.2.5}$$

for some constants $\alpha, \beta, \gamma > 0$ and $C_k > 0$, which may depend on k. Then $(X_n)_{n\in\mathbb{N}}$ is tight.

B.3 Separable Hilbert Spaces

Let $\mathbb{H}$ be a separable Hilbert space over $\mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and induced norm $\| \cdot \|_{\mathbb{H}}$. Consider an ONB $(e_k)_{k \in \mathbb{N}_0}$ of $\mathbb{H}$ and, for $l \in \mathbb{N}_0$, let

$$\Pi_l \colon \mathbb{H} \to \mathbb{H}; \quad x \mapsto \sum_{k=0}^{l} \langle x, e_k \rangle_{\mathbb{H}} e_k, \tag{B.3.1}$$

be the orthogonal projection onto the l-th subspace

$$\left\{ \sum_{k=0}^{l} \beta_k e_k \,\middle|\, \beta_1, \ldots, \beta_l \in \mathbb{R} \right\}$$

of $\mathbb{H}$ spanned by $\{e_1, \ldots, e_l\}$. We have the following simple result concerning the weak convergence of random elements in $\mathbb{H}$.

Theorem B.3.1. *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of $\mathbb{H}$-valued random elements and X an $\mathbb{H}$-valued random element such that*

$$\left\langle X_n, \sum_{k=0}^{l} \beta_k e_k \right\rangle_{\mathbb{H}} \xrightarrow{\mathcal{D}} \left\langle X, \sum_{k=0}^{l} \beta_k e_k \right\rangle_{\mathbb{H}}, \quad n \to \infty, \quad \forall \beta_1, \ldots, \beta_l \in \mathbb{R}, \quad \forall l \in \mathbb{N}_0, \tag{B.3.2}$$

$$\lim_{l \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\| X_n - \Pi_l (X_n) \|_{\mathbb{H}} \geq \delta \right) = 0, \quad \forall \delta > 0. \tag{B.3.3}$$

Then $X_n \xrightarrow{\mathcal{D}} X$ in $\mathbb{H}$ as $n \to \infty$.

Taking a closer look at the preceding two conditions, then (B.3.2) is reminiscent of a condition corresponding to convergence of the finite-dimensional distributions while (B.3.3) resembles a tightness condition.

Chapter C

Interior Estimates of the Schauder Type

This appendix is entirely devoted to the derivation of the final estimate appearing in Proposition 2.2.1 of Chapter 2, which is a classical result in the theory of parabolic PDEs, see [Fri64; Fri58; Lor17; Lie96], and goes under the name of a priori interior estimates of the Schauder type. We will follow along the proof by Friedman given in [Fri64, Chapter 4, Section 4]. Since we are literally only concerned with the tracking of constants with respect to ε, and, other than that, change nothing about the proof, we will only give a detailed table with the relevant constants and estimates. Another reason for not laying out the whole proof is that it would add a substantial amount of material that is not pertinent to the main subject of the main text. That being said, the table, which we will thoroughly explain shortly, should serve as a reference aid if one were to reconstruct said proof line by line. Also, because this material is, except for the aforementioned Proposition 2.2.1, not used elsewhere in the main text, we will completely adhere to the notation by Friedman [Fri64] without explaining it here in detail. This also means that, unless otherwise stated, any equation number, theorem, lemma, etc., appearing here corresponds one-to-one to the one in [Fri64]. As such, the present appendix is considered to be contextually independent from the main text and it is very important to keep in mind that it only makes sense alongside [Fri64].

Let us first start with some useful comments. In [Fri64, Chapter 1, Section 1] and [Fri64, Chapter 3, Section 2], the main definitions on Hölder regularity, Hölder norms, and parabolic PDEs are introduced, and it is explained what interior estimates are and what they entail. The result that we are interested in is [Fri64, Chapter 4, Section 1, Theorem 1], which we state here again for clarity with additional information enclosed by the parentheses [ ].

Theorem 1. *[Let $D \subset \mathbb{R}^{n+1}$, $n \in \mathbb{N}$, be any bounded, open set.] Assume that*

$$|a_{ij}|_{0,\alpha} \leq K_1, \quad |b_i|_{1,\alpha} \leq K_1, \quad |c|_{2,\alpha} \leq K_1, \tag{1.8}$$

that for any real vector ξ and $(x,t) \in D$,

$$\sum_{i,j=1}^{n} a_{ij}(x,t)\xi_i\xi_j \geq K_2|\xi|^2, \qquad (K_2 > 0), \tag{1.9}$$

and that $|f|_{2,\alpha} < \infty$. There exists a constant K, depending only on K_1, K_2, n, α, such that if u is any solution of (1.1) in D such that $D_x u, D_x^2 u, D_t u$ are [locally] Hölder continuous (exponent α) in D and such that $|u|_0 < \infty$, then $u \in C_{2+\alpha}$ and

$$|u|_{0,2+\alpha} \leq K(|u|_0 + |f|_{2,\alpha}). \tag{1.10}$$

Here, u is the solution to the following parabolic PDE, which was defined in equation [Fri64, Chapter 4, (1.1)],

$$Lu \equiv \sum_{i,j=1}^{n} a_{ij}(x,t)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x,t)\frac{\partial u}{\partial x_i} + c(x,t)u - \frac{\partial u}{\partial t} = f(x,t). \tag{1.1}$$

In our case, instead of (1.8), we want to assume that, for $\varepsilon > 0$,

$$|a_{ij}|_{0,\alpha} \leq K_1, \quad |b_i|_{1,\alpha} \leq K_1 \varepsilon^{-2}, \quad |c|_{2,\alpha} \leq K_1, \tag{1.8*}$$

that is, we assume explicit ε-dependence for the first-order terms b_i. Therefore, the main task is to simply track how this ε-dependence propagates to the final estimate (1.10). For convenience, we write out the Hölder norms of the coefficient functions and the solution to the parabolic PDE in the notation by Friedman. By the equations [Fri64, Chapter 4, (1.4), (1.5), (1.6), (1.7)], we have

$$\begin{aligned}
|a_{ij}|_{0,\alpha} = |a_{ij}|_{0,0} + M_{0,\alpha}[a_{ij}] &= |a_{ij}|_0 + H_\alpha[a_{ij}] \\
&= \operatorname*{l.u.b.}_{P\in D} |a_{ij}(P)| + \operatorname*{l.u.b.}_{P,Q\in D} d_{PQ}^{\alpha} \frac{|a_{ij}(P) - a_{ij}(Q)|}{d(P,Q)^\alpha}, \\
|b_i|_{1,\alpha} = |b_i|_{1,0} + M_{1,\alpha}[b_i] &= |db_i|_0 + H_\alpha[db_i] \\
&= \operatorname*{l.u.b.}_{P\in D} d_P|b_i(P)| + \operatorname*{l.u.b.}_{P,Q\in D} d_{PQ}^{1+\alpha} \frac{|b_i(P) - b_i(Q)|}{d(P,Q)^\alpha}, \\
|c|_{2,\alpha} = |c|_{2,0} + M_{2,\alpha}[c] &= |d^2 c|_0 + H_\alpha[d^2 c] \\
&= \operatorname*{l.u.b.}_{P\in D} d_P^2|c(P)| + \operatorname*{l.u.b.}_{P,Q\in D} d_{PQ}^{2+\alpha} \frac{|c(P) - c(Q)|}{d(P,Q)^\alpha}, \\
|f|_{2,\alpha} = |f|_{2,0} + M_{2,\alpha}[f] &= |d^2 f|_0 + H_\alpha[d^2 f] \\
&= \operatorname*{l.u.b.}_{P\in D} d_P^2|f(P)| + \operatorname*{l.u.b.}_{P,Q\in D} d_{PQ}^{2+\alpha} \frac{|f(P) - f(Q)|}{d(P,Q)^\alpha},
\end{aligned}$$

and

$$\begin{aligned}
|u|_{0,2+\alpha} &= |u|_{0,2} + M_{0,\alpha}[u] + M_{1,\alpha}[u] + M_{2,\alpha}[u] \\
&= M_{0,0}[u] + M_{0,1}[u] + M_{0,2}[u] + M_{0,\alpha}[u] + M_{1,\alpha}[u] + M_{2,\alpha}[u] \\
&= |u|_0 + \sum |dD_x u|_0 + \sum |d^2 D_x^2 u|_0 + H_\alpha[u] + \sum H_\alpha[dD_x u] + \sum H_\alpha[d^2 D_x^2 u] \\
&= |u|_\alpha + \sum |dD_x u|_\alpha + \sum |d^2 D_x^2 u|_\alpha,
\end{aligned}$$

where the summation is taken over all partial derivatives and d_P, d_{PQ}, and $d(P,Q)$ have been defined in [Fri64, Chapter 3, Section 2]. As already mentioned, we will not write the whole proof of [Fri64, Chapter 4, Section 1, Theorem 1] again, but let us at least illustrate by an exemplary estimate, which is quite prototypical for the proof, how the ε-dependence enters the interior estimates. Consider estimate [Fri64, Chapter 4, (4.8)] for example. There appears the term $|b_i||\partial u/\partial x_i|$. For $Q \in N$, where N is a semicube with top P and edge $d = \lambda d_P$, it can be estimated as follows

$$
\begin{aligned}
|b_i(Q)| \left| \frac{\partial u}{\partial_{x_i}}(Q) \right| &= d_P^{-2} d_P |b_i(Q)| d_P \left| \frac{\partial u}{\partial_{x_i}}(Q) \right| \leq 4 d_P^{-2} d_Q |b_i(Q)| d_Q \left| \frac{\partial u}{\partial_{x_i}}(Q) \right| \\
&\leq 4 d_P^{-2} |b_i|_{1,0} M_{0,1}[u] \leq 4 d_P^{-2} |b_i|_{1,\alpha} |u|_{0,1} \leq 4 K_1 d_P^{-2} \varepsilon^{-2} |u|_{0,1}.
\end{aligned}
$$

Here, we additionally used the fact that $d_Q \geq d_P/2$ if $Q \in N$, as stated in the proof of [Fri64, Chapter 4, Section 1, Theorem 1]. Although tedious, other estimates can be obtained in a similar fashion. Another noteworthy point in our adjustment is that the parameter λ that is introduced and determined in the proof will be chosen to depend on ε at some point, but this will be explicitly indicated in the table.

A couple of last remarks are in order. Although [Fri64, Chapter 4, Section 2] and [Fri64, Chapter 4, Section 3] are crucial for establishing the actual proof in [Fri64, Chapter 4, Section 4], the estimates do not depend on ε as there are no first-order terms involved in the proofs, so that one can safely skip those two sections. Hence, the focus is solely on the proof that is contained in [Fri64, Chapter 4, Section 4]. Since we consider the more general setting of Friedman here, the derivation of the estimates applies to multidimensional and time-inhomogeneous settings, so that Proposition 2.2.1 from the main text follows as a simple corollary. The obsolete notation l. u. b. in [Fri64] stands for the supremum-norm by context. The result [Fri64, Chapter 4, Section 1, Theorem 1] has been stated without proof prior to that chapter as [Fri64, Chapter 3, Section 2, Theorem 5]. There seems to be an editorial slip between that Theorem 5 in Chapter 3 and Theorem 1 in Chapter 4 because Friedman writes locally Hölder continuous in D for the former one but Hölder continuous in D for the latter one. The former is the precise statement, considering the weighted norms that are used in the proof. This can also be confirmed through his article [Fri58] with a more general statement of said theorem, where he actually defines Hölder continuity in D as local Hölder continuity in D.

Table C.1 shows a comprehensive tracking of the constants in [Fri64, Chapter 4, Section 4] with respect to ε. The table is ordered according to the equation numbers as they appear in that section starting from equation (4.8). To reduce redundancy, any equation or estimate staying the same as in the original material has been skipped. Sometimes, there are intermediate steps indicated by (+) which lead up to a subsequent equation. These were added for more transparent tracking because some details are not fully written out in the original proof by Friedman. Compared to the original material, we also added two additional numbered estimates, namely (4.26) and (4.27), around the end of the table which correspond to the last two unnumbered estimates at the end of the proof after equation (4.25). The field ”LHS” and ”RHS” mean the left-hand side and the right-hand side corresponding to that equation or estimate. Inequalities should always be understood as $\leq$. A constant such as $K^{(8)}(\varepsilon^{-2})$ indicates that it contains at least one term of order $\mathcal{O}(\varepsilon^{-2})$. The occasionally appearing constant $K > 0$ in the table is not always the same. It is loosely used for any positive constant that is independent of ε. Lastly, the field ”Notes” provides further remarks if necessary and ”case (a)” and ”case (b)” have been defined in the actual proof.

Equation/ Estimate	LHS	RHS	Notes
(4.8)	$\mathrm{l.u.b.}_N \lvert F\rvert$	$K^{(8)}(\varepsilon^{-2})d_P^{-2}\left(\lvert f\rvert_{2,\alpha}+\lvert u\rvert_{0,1}+\lambda^\alpha M\right)$	-
(4.9)	$d^\alpha H_{P,N}[F]$	$K\lambda^\alpha d_P^{-2}\lvert f\rvert_{2,\alpha}+K^{(9)}(\varepsilon^{-2})\lambda^\alpha d_P^{-2}\lvert u\rvert_{0,2}+K^{(9)}(\varepsilon^{-2})\lambda^\alpha d_P^\alpha\sum_{i=0}^1 d_P^{i-2}H_{P,N}[D_x^i u]$	-
(4.10)	$d^\alpha H_{P,N}[F]$	$K^{(10)}(\varepsilon^{-2})\lambda^\alpha d_P^{-2}\left(\lvert f\rvert_{2,\alpha}+\lvert u\rvert_{0,2}+\sum_{i=0}^1 M_{0,i+\alpha}\right)$	-
(4.11)	$d_P^2 I$	$K\lambda^{-2}\lvert u\rvert_{0,2}+K^{(11)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}+K^{(11)}(\varepsilon^{-2})\lvert u\rvert_{0,1}+K^{(11)}(\varepsilon^{-2})\lambda^\alpha\left(M+M_0'+M_1'\right)$	-
(4.13)	$d_P^2\lvert D_x^2u(P)\rvert$	$K^{(13)}(\varepsilon^{-2})\lambda^{-2}\lvert u\rvert_0+K^{(13)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}+K^{(13)}(\varepsilon^{-2})\lambda^\alpha\left(M+M_0'+M_1'\right)$	-
(4.14)	M	$K^{(14)}(\varepsilon^{-2})\lambda^{-2}\lvert u\rvert_0+K^{(14)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}+K^{(14)}(\varepsilon^{-2})\lambda^\alpha\left(M+M_0'+M_1'\right)$	-
(4.18)	$d^\alpha H_{Q,N}[F]$	$K^{(18)}(\varepsilon^{-2})\lambda^\alpha d_P^{-2}\left(\lvert f\rvert_{2,\alpha}+\lvert u\rvert_{0,2}+\sum_{i=0}^1 M_{0,i+\alpha}\right)+Kd^\alpha\sum\lvert a_{ij}(P)-a_{ij}(Q)\rvert H_{Q,N}[D_x^2u]$	-
(+)	$d^\alpha H_{Q,N}[F]$	$K^{(18)}(\varepsilon^{-2})\lambda^\alpha d_P^{-2}\left(\lvert f\rvert_{2,\alpha}+\lvert u\rvert_{0,2}+\sum_{i=0}^1 M_{0,i+\alpha}\right)+K\lambda^{2\alpha}d_P^{-2}M_2'$	-
(4.19)	M_2'	$K^{(19)}(\varepsilon^{-2})\lambda^{-2-\alpha}\lvert u\rvert_0+K^{(19)}(\varepsilon^{-2})\lambda^{-\alpha}\lvert f\rvert_{2,\alpha}+K^{(19)}(\varepsilon^{-2})\left(M+M_0'+M_1'\right)+K\lambda^\alpha M_2'$	(4.12) is used to handle $\lvert u\rvert_{0,2}$
(4.21)	M_i'	$K^{(21)}(\varepsilon^{-2})\lambda^{-i-\alpha}\lvert u\rvert_0+K^{(21)}(\varepsilon^{-2})\lambda^{2-i-\alpha}\lvert f\rvert_{2,\alpha}+K^{(21)}(\varepsilon^{-2})\lambda^{2-i}\left(M+M_0'+M_1'\right)+K\lambda^{2-i+\alpha}M_2'$	-
(+)	M_i'	$K_+^{(21)}(\varepsilon^{-2})\lambda^{-1-\alpha}\lvert u\rvert_0+K_+^{(21)}(\varepsilon^{-2})\lambda^{1-\alpha}\lvert f\rvert_{2,\alpha}+K_+^{(21)}(\varepsilon^{-2})\lambda\left(M+M_0'+M_1'\right)+K\lambda^{1+\alpha}M_2'+K\lambda^{-\alpha}M_{0,1}$	sum of RHS of (4.20) and (4.21)
(4.22)	$M_0'+M_1'$	$K^{(22)}(\varepsilon^{-2})\lambda^{-1-\alpha}\lvert u\rvert_0+K^{(22)}(\varepsilon^{-2})\lambda^{1-\alpha}\lvert f\rvert_{2,\alpha}+K^{(22)}(\varepsilon^{-2})\lambda M+K\lambda^{1+\alpha}M_2'+K\lambda^{-\alpha}M_{0,1}$	choosing λ suff. small with $\lambda=\lambda_\varepsilon=\mathcal{O}\left(\varepsilon^2\right)$
(4.23)	M_2'	$K^{(23)}(\varepsilon^{-4})\lambda^{-2-\alpha}\lvert u\rvert_0+K^{(23)}(\varepsilon^{-4})\lambda^{-\alpha}\lvert f\rvert_{2,\alpha}+K^{(23)}(\varepsilon^{-2})M+K^{(23)}(\varepsilon^{-2})\lambda^{-\alpha}M_{0,1}$	multiplying constants; choosing λ even smaller with $\lambda=\lambda_\varepsilon=\mathcal{O}\left(\varepsilon^{2/\alpha}\right)$
(4.24)	$M+M_0'+M_1'$	$K^{(24)}(\varepsilon^{-2})\lambda^{-2}\lvert u\rvert_0+K^{(24)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}+K^{(24)}(\varepsilon^{-2})\lambda^\alpha(M+M_0'+M_1')+K\lambda^{-\alpha}M_{0,1}$	-
(+)	$K\lambda^{-\alpha}M_{0,1}$	$\frac14 M+4C_0K\lambda^{-\alpha}\lvert u\rvert_0$	by (2.3)
(4.25)	$M+M_0'+M_1'$	$K^{(25)}(\varepsilon^{-2})\lambda^{-2}\lvert u\rvert_0+K^{(25)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}$	**final estimate case (a)**
(+)	M_2'	$K_+^{(25)}(\varepsilon^{-2})\lambda^{-2-\alpha}\lvert u\rvert_0+K_+^{(25)}(\varepsilon^{-2})\lambda^{-\alpha}\lvert f\rvert_{2,\alpha}$	**final estimate case (a)**
(4.26)	$M_0'+M_1'$	$K^{(26)}(\varepsilon^{-4})\lambda^{-1-\alpha}\lvert u\rvert_0+K^{(26)}(\varepsilon^{-2})\lambda^{1-\alpha}\lvert f\rvert_{2,\alpha}+K^{(26)}(\varepsilon^{-2})\lambda M+K^{(26)}(\varepsilon^{-2})\lambda^{-\alpha}M_{0,1}$	-
(4.27)	$M+M_0'+M_1'$	$K^{(27)}(\varepsilon^{-4})\lambda^{-2}\lvert u\rvert_0+K^{(27)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}+K^{(27)}(\varepsilon^{-2})\lambda^\alpha(M+M_0'+M_1')+K^{(27)}(\varepsilon^{-2})\lambda^{-\alpha}M_{0,1}$	-
(+)	$K^{(27)}(\varepsilon^{-2})\lambda^{-\alpha}M_{0,1}$	$\frac14 M+4C_0K^{(27)}(\varepsilon^{-2})\lambda^{-\alpha}\lvert u\rvert_0$	by (2.3)
(+)	$M+M_0'+M_1'$	$K_+^{(27)}(\varepsilon^{-4})\lambda^{-2}\lvert u\rvert_0+K_+^{(27)}(\varepsilon^{-2})\lvert f\rvert_{2,\alpha}$	**final estimate case (b)**
(+)	M_2'	$K_{++}^{(27)}(\varepsilon^{-6})\lambda^{-2-\alpha}\lvert u\rvert_0+K_{++}^{(27)}(\varepsilon^{-4})\lambda^{-\alpha}\lvert f\rvert_{2,\alpha}$	**final estimate case (b)**

Table C.1: Tracking of the final constant in estimate (1.10) of [Fri64, Chapter 4, Section 1, Theorem 1] with respect to ε when (1.8) is replaced by (1.8*).

Let us now derive the full interior estimate from the table. Summing over all final estimates in case (a) and (b) from Table C.1 implies

$$M + M_0' + M_1' + M_2' \leq K^{(28)}(\varepsilon^{-6})\lambda_\varepsilon^{-2-\alpha}(|u|_0 + |f|_{2,\alpha}),$$

and using [Fri64, Chapter 4, Section 1, Lemma 1] in the form

$$|u|_{0,1} \leq \frac{1}{2}M + K|u|_0$$

yields

$$|u|_{0,1} + M + M_0' + M_1' + M_2' \leq K^{(29)}(\varepsilon^{-6})\lambda_\varepsilon^{-2-\alpha}(|u|_0 + |f|_{2,\alpha})$$

after a rearrangement of terms. By definition of the quantities on the LHS, the last inequality is equivalent to

$$|u|_{0,2+\alpha} = |u|_\alpha + \sum |dD_x u|_\alpha + \sum |d^2 D_x^2 u|_\alpha \leq K^{(29)}(\varepsilon^{-6})\lambda_\varepsilon^{-2-\alpha}(|u|_0 + |f|_{2,\alpha}).$$

For the sake of completeness, to get a similar bound on the weighted norm $|d^2 D_t u|_\alpha$, one uses the PDE (1.1) and the bounds [Fri64, Chapter 3, (5.8)] and [Fri64, Chapter 4, (1.3)], namely

$$|d^2(gh)|_\alpha \leq |d^2 g|_\alpha |h|_\alpha, \qquad |d^2(gh)|_\alpha \leq |dg|_\alpha |dh|_\alpha,$$

to obtain

$$\begin{aligned} |d^2 D_t u|_\alpha &\leq |d^2(cu)|_\alpha + |d^2 f|_\alpha + \sum |d^2(b_i D_x u)|_\alpha + \sum |d^2(a_{ij} D_x^2 u)|_\alpha \\ &\leq |d^2 c|_\alpha |u|_\alpha + |d^2 f|_\alpha + \sum |db_i|_\alpha |dD_x u|_\alpha + \sum |a_{ij}|_\alpha |d^2 D_x^2 u|_\alpha \\ &\leq K^{(30)}(\varepsilon^{-8})\lambda_\varepsilon^{-2-\alpha}(|u|_0 + |f|_{2,\alpha}), \end{aligned}$$

where the last inequality followed from the previously derived interior estimates and (1.8*). After recalling that $\lambda_\varepsilon = \mathcal{O}\left(\varepsilon^{2/\alpha}\right)$, we eventually arrive at the final estimate

$$|u|_\alpha + \sum |dD_x u|_\alpha + \sum |d^2 D_x^2 u|_\alpha + |d^2 D_t u|_\alpha \leq K\varepsilon^{-10-4/\alpha}(|u|_0 + |f|_{2,\alpha}) \qquad (1.10^*)$$

for sufficiently small $\varepsilon > 0$, which corresponds to (2.2.7) of Proposition 2.2.1 from the main material with $f \neq 0$.

Bibliography

[Abd+23] A. Abdulle, G. Garegnani, G. A. Pavliotis, A. M. Stuart, and A. Zanoni. "Drift estimation of multiscale diffusions based on filtered data." In: *Found. Comput. Math.* 23 (2023), pp. 33–84. DOI: 10.1007/s10208-021-09541-9.

[APZ22] A. Abdulle, G. A. Pavliotis, and A. Zanoni. "Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions." In: *Stat. Comput.* 32 (2022), p. 34. DOI: 10.1007/s11222-022-10081-7.

[Abr+02] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. "Multiscale nature of network traffic." In: *IEEE Signal Process. Mag.* 19.3 (2002), pp. 28–46. DOI: 10.1109/79.998080.

[AMZ05] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. "How often to sample a continuous-time process in the presence of market microstructure noise." In: *Rev. Financ. Stud.* 18.2 (2005), pp. 351–416. DOI: 10.1093/rfs/hhi016.

[Aln+14] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells. "Unified form language: a domain-specific language for weak formulations of partial differential equations." In: *ACM Trans. Math. Softw.* 40.2 (2014), 9:1–9:37. DOI: 10.1145/2566630.

[Aze+13] R. Azencott, A. Beri, A. Jain, and I. Timofeyev. "Sub-sampling and parametric estimation for multiscale dynamics." In: *Commun. Math. Sci.* 11.4 (2013), pp. 939–970. DOI: 10.4310/CMS.2013.v11.n4.a3.

[Bal+06] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. "Asymptotic analysis of multiscale approximations to reaction networks." In: *Ann. Appl. Probab.* 16.4 (2006), pp. 1925–1961. DOI: 10.1214/105051606000000420.

[Bar+25] I. A. Baratta et al. *DOLFINx: the next generation FEniCS problem solving environment.* 2025. DOI: 10.5281/zenodo.1812201307.

[BLP78] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures.* Vol. 5. Studies in Mathematics and its Applications. North-Holland, 1978. ISBN: 978-0-444-85172-7.

[Bez+17] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. "Julia: a fresh approach to numerical computing." In: *SIAM Rev.* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671.

[Bil99] P. Billingsley. *Convergence of Probability Measures.* 2nd ed. Wiley Series in Probability and Statistics. John Wiley & Sons, 1999. ISBN: 978-0-471-19745-4. DOI: 10.1002/9780470316962.

[BP78] G. Blankenship and G. C. Papanicolaou. "Stability and control of stochastic systems with wide-band noise disturbances. I." In: *SIAM J. Appl. Math.* 34.3 (1978), pp. 437–476. DOI: 10.1137/0134036.

[Bog18] V. I. Bogačev. *Weak Convergence of Measures.* Vol. 234. Mathematical Surveys and Monographs. Providence, Rhode Island: American Mathematical Society, 2018. ISBN: 9781470447380. DOI: 10.1090/surv/234.

[Bog07] V. I. Bogachev. *Measure Theory.* Vol. 1. Springer, 2007. ISBN: 978-3-540-34513-8. DOI: 10.1007/978-3-540-34514-5.

[BT06] B. Bongioanni and J. L. Torrea. "Sobolev spaces associated to the harmonic oscillator." In: *Proc. Indian Acad. Sci. (Math. Sci.)* 116 (2006), pp. 337–360. DOI: 10.1007/BF02829750.

[Bor25] J. I. Borodavka. *MDEforM.jl*. Version v0.1. 2025. URL: https://github.com/UQatKIT/MDEforM.jl.git.

[BZH26] J. I. Borodavka, A. Zanoni, and M. Hirsch. *Hermite-Multiscale-Estimator*. Version v0.1. 2026. URL: https://github.com/UQatKIT/Hermite-Multiscale-Estimator.git.

[Bor+25] J. I. Borodavka, M. Hirsch, S. Krumscheid, and A. Zanoni. "Nonparametric estimation of homogenized invariant measures from multiscale data via Hermite expansion." arXiv:2510.25521. 2025.

[BK26] J. I. Borodavka and S. Krumscheid. "Limit theorems for one-dimensional homogenized diffusion processes." In: *J. Stat. Phys.* 193.35 (2026). DOI: 10.1007/s10955-026-03592-4.

[BKP25] J. I. Borodavka, S. Krumscheid, and G. A. Pavliotis. "A minimum distance estimator approach for misspecified ergodic processes." arXiv:2506.12432. 2025.

[Bos98] D. Bosq. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. 2nd ed. Vol. 110. Lecture Notes in Statistics. Springer, 1998. ISBN: 978-0-387-98590-9. DOI: 10.1007/978-1-4612-1718-3.

[Bra00] R. Bracewell. *The Fourier Transform and its Applications*. 3rd ed. McGraw-Hill Higher Education, 2000. ISBN: 0-07-303938-1.

[BS18] P. Bühler and D. A. Salamon. *Functional Analysis*. Vol. 191. Graduate Studies in Mathematics. American Mathematical Society, 2018. ISBN: 978-1-4704-4190-6. DOI: 10.1090/gsm/191.

[CGO21] E. Celeghini, M. Gadella, and M. A. del Olmo. "Hermite functions and Fourier series." In: *Symmetry* 13.5 (2021). DOI: 10.3390/sym13050853.

[CM11] C. Cervellera and D. Macciò. "A numerical method for minimum distance estimation problems." In: *J. Multivar. Anal.* 102.4 (2011), pp. 789–800. DOI: 10.1016/j.jmva.2010.12.006.

[CR21] C. Cheverry and N. Raymond. *A Guide to Spectral Theory: Applications and Exercises*. 1st ed. Vol. 338. Birkhäuser Advanced Texts Basler Lehrbücher. Birkhäuser Cham, 2021. ISBN: 978-3-030-67461-8. DOI: 10.1007/978-3-030-67462-5.

[Chu01] K. L. Chung. *A Course in Probability Theory*. 3rd ed. Academic Press, 2001. ISBN: 978-0-12-174151-8.

[CGL25] Comte, F., Genon-Catalot, V., and Larédo, C. "Nonparametric moment method for scalar McKean-Vlasov stochastic differential equations." In: *ESAIM: PS* 29 (2025), pp. 400–449. DOI: 10.1051/ps/2025012.

[Cri+26] D. Crisan, P. Dobson, B. Goddard, M. Ottobre, and I. Souttar. "Poisson equations with locally-Lipschitz coefficients and uniform in time averaging for stochastic differential equations via strong exponential stability." In: *Ann. Inst. H. Poincaré Probab. Statist.* 62.1 (2026), pp. 667–721. DOI: 10.1214/24-AIHP1526.

[CV11] D. Crommelin and E. Vanden-Eijnden. "Diffusion estimation from multiscale data by operator eigenpairs." In: *Multiscale Model. Simul.* 9.4 (2011), pp. 1588–1623. DOI: 10.1137/100795917.

[DZ96] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems.* Vol. 229. London Mathematical Society Lecture Note Series. Cambridge University Press, 1996. ISBN: 9780511662829. DOI: 10.1017/CBO9780511662829.

[DR84] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration.* 2nd ed. Computer Science and Applied Mathematics. Academic Press, 1984. ISBN: 978-0-12-206360-2. DOI: 10.1016/C2013-0-10566-1.

[DB02] P. Deuflhard and F. Bornemann. *Scientific Computing with Ordinary Differential Equations.* 1st ed. Vol. 42. Texts in Applied Mathematics. Springer New York, NY, 2002. ISBN: 978-0-387-95462-2. DOI: 10.1007/978-0-387-21582-2.

[DDP23] A. B. Duncan, M. H. Duong, and G. A. Pavliotis. "Brownian motion in an N-scale periodic potential." In: *J. Stat. Phys.* 190 (2023). DOI: 10.1007/s10955-023-03092-9.

[Dur96] R. Durrett. *Stochastic Calculus: A Practical Introduction.* 1st ed. Probability and Stochastics Series. CRC Press, 1996, p. 341. ISBN: 9780203738283. DOI: 10.1201/9780203738283.

[Dyn65] E. B. Dynkin. *Markov Processes I.* 1st ed. Die Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 1965. ISBN: 978-3-662-00033-5. DOI: 10.1007/978-3-662-00031-1.

[EG04] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements.* Vol. 159. Applied Mathematical Sciences. Springer New York, NY, 2004. ISBN: 978-0-387-20574-8. DOI: 10.1007/978-1-4757-4355-5.

[FV04] I. Fatkullin and E. Vanden-Eijnden. "A computational strategy for multiscale systems with applications to Lorenz 96 model." In: *J. Comput. Phys.* 200.2 (2004), pp. 605–638. DOI: 10.1016/j.jcp.2004.04.013.

[FR11] Y. Frederix and D. Roose. "A drift-filtered approach to diffusion estimation for multiscale processes." In: *Coping with Complexity: Model Reduction and Data Analysis.* Ed. by A. N. Gorban and D. Roose. Vol. 75. Lecture Notes in Computational Science and Engineering. Springer Berlin, Heidelberg, 2011, pp. 269–286. ISBN: 978-3-642-14940-5. DOI: 10.1007/978-3-642-14941-2.

[FD81] D. A. Freedman and P. Diaconis. "On the histogram as a density estimator: L_2 theory." In: *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 57 (1981), pp. 453–476. DOI: 10.1007/BF01025868.

[FW12] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems.* 3rd ed. Vol. 260. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 2012. ISBN: 978-3-642-25846-6. DOI: 10.1007/978-3-642-25847-3.

[Fri58] A. Friedman. "Interior estimates for parabolic systems of partial differential equations." In: *Indiana Univ. Math. J.* 7.3 (1958), pp. 393–417. DOI: 10.1512/iumj.1958.7.57025.

[Fri64] A. Friedman. *Partial Differential Equations of Parabolic Type.* Prentice-Hall, 1964.

[GS17] S. Gailus and K. Spiliopoulos. "Statistical inference for perturbed multiscale dynamical systems." In: *Stoch. Process. Their Appl.* 127.2 (2017), pp. 419–448. DOI: 10.1016/j.spa.2016.06.013.

[GS18] S. Gailus and K. Spiliopoulos. "Discrete-time statistical inference for multiscale diffusions." In: *Multiscale Model. Simul.* 16.4 (2018), pp. 1824–1858. DOI: 10.1137/17M1147408.

[GZ23] G. Garegnani and A. Zanoni. "Robust estimation of effective diffusions from multiscale data." In: *Commun. Math. Sci.* 21.2 (2023), pp. 405–435. DOI: 10.4310/CMS.2023.v21.n2.a5.

[GS68] I. I. Gikhman and A. Skorokhod. *Stochastic Differential Equations.* In Russian. Naukova Dumka, 1968.

[GN16] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Vol. 40. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016. ISBN: 978-1-107-04316-9. DOI: 10.1017/CBO9781107337862.

[GKS04] D. Givon, R. Kupferman, and A. Stuart. "Extracting macroscopic dynamics: model problems and algorithms." In: *Nonlinearity* 17.6 (2004), R55. DOI: 10.1088/0951-7715/17/6/R01.

[GR07] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* 7th ed. Translated from the Russian. Elsevier/Academic Press, 2007. ISBN: 978-0-12-373637-6.

[Gra14] L. Grafakos. *Classical Fourier Analysis.* 3rd ed. Vol. 249. Graduate Texts in Mathematics. Springer New York, NY, 2014. ISBN: 978-1-4939-1193-6. DOI: 10.1007/978-1-4939-1194-3.

[Hac92] W. Hackbusch. *Elliptic Differential Equations: Theory and Numerical Treatment.* 2nd ed. Vol. 18. Springer Series in Computational Mathematics. Springer Berlin, Heidelberg, 1992. ISBN: 978-3-662-54960-5. DOI: 10.1007/978-3-662-54961-2.

[Hen24] N. Henze. *Asymptotic Stochastics: An Introduction with a View towards Statistics.* 1st ed. Vol. 10. Mathematics Study Resources. Springer Berlin, Heidelberg, 2024. ISBN: 978-3-662-68922-6. DOI: 10.1007/978-3-662-68923-3.

[HJ21] N. Henze and M. D. Jiménez-Gamero. "A test for Gaussianity in Hilbert spaces via the empirical characteristic functional." In: *Scand. J. Statist.* 48.2 (2021), pp. 406–428. DOI: 10.1111/sjos.12470.

[HW97] N. Henze and T. Wagner. "A new approach to the BHEP tests for multivariate normality." In: *J. Multivar. Anal.* 62.1 (1997), pp. 1–23. DOI: 10.1006/jmva.1997.1684.

[HZ24] M. Hirsch and A. Zanoni. "Stochastic gradient descent in continuous time for drift identification in multiscale diffusions." In: *ESAIM Math. Model. Numer. Anal.* 58.6 (2024), pp. 2387–2414. DOI: 10.1051/m2an/2024079.

[Hyt+16] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces. Volume I: Martingales and Littlewood–Paley Theory.* Vol. 63. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer Cham, 2016. ISBN: 978-3-319-48519-5. DOI: 10.1007/978-3-319-48520-1.

[IK81] I. A. Ibragimov and R. Khas'minskii. *Statistical Estimation: Asymptotic Theory.* 1st ed. Vol. 16. Stochastic Modelling and Applied Probability. Springer New York, NY, 1981. ISBN: 978-1-4899-0027-2. DOI: 10.1007/978-1-4899-0027-2.

[Ind61] J. Indritz. "An inequality for Hermite polynomials." In: *Proc. Amer. Math. Soc.* 12 (1961), pp. 981–983. DOI: 10.1090/S0002-9939-1961-0132852-2.

[Itô84] K. Itô. *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*. Society for Industrial and Applied Mathematics, 1984. ISBN: 0-89871-193-2. DOI: 10.1137/1.9781611970234.

[IM96] K. Itô and H. P. McKean. *Diffusion Processes and Their Sample Paths*. 1st ed. Classics in Mathematics. Springer Berlin, Heidelberg, 1996. ISBN: 978-3-540-60629-1. DOI: 10.1007/978-3-642-62025-6.

[Joh05] S. G. Johnson. *Multi-dimensional adaptive integration in C: the Cubature package*. Version v1.0.4. 2005. URL: https://github.com/stevengj/cubature.

[Kal21] O. Kallenberg. *Foundations of Modern Probability*. 3rd ed. Vol. 99. Probability Theory and Stochastic Modelling. Springer Cham, 2021. ISBN: 978-3-030-61870-4. DOI: 10.1007/978-3-030-61871-1.

[KKP15] S. Kalliadasis, S. Krumscheid, and G. Pavliotis. "A new framework for extracting coarse-grained models from time series with multiscale structure." In: *J. Comput. Phys.* 296 (2015), pp. 314–328. DOI: 10.1016/j.jcp.2015.05.002.

[KS14] G. Kallianpur and P. Sundar. *Stochastic Analysis and Diffusion Processes*. Vol. 24. Oxford Graduate Texts in Mathematics. Oxford University Press, 2014. ISBN: 978–0–19–965706–3.

[KS91] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. 2nd ed. Vol. 113. Graduate Texts in Mathematics. Springer New York, NY, 1991. ISBN: 978-0-387-97655-6. DOI: 10.1007/978-1-4612-0949-2.

[Kha66a] R. Z. Khas'minskii. "A limit theorem for the solutions of differential equations with random right-hand sides." In: *Theory Probab. Appl.* 11.3 (1966), pp. 390–406. DOI: 10.1137/1111038.

[Kha66b] R. Z. Khas'minskii. "On stochastic processes defined by differential equations with a small parameter." In: *Theory Probab. Appl.* 11.2 (1966), pp. 211–228. DOI: 10.1137/1111018.

[KMN12] R. Z. Khas'minskii, G. N. Milshteĭn, and M. Nevelson. *Stochastic Stability of Differential Equations*. 2nd ed. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 2012. ISBN: 978-3-642-23279-4. DOI: 10.1007/978-3-642-23280-0.

[KW52] J. Kiefer and J. Wolfowitz. "Stochastic estimation of the maximum of a regression function." In: *Ann. Math. Stat.* 23.3 (1952), pp. 462–466. DOI: 10.1214/aoms/1177729392.

[KRK17] J. K. Kim, G. A. Rempala, and H.-W. Kang. "Reduction for stochastic biochemical reaction networks with multiscale conservations." In: *Multiscale Model. Simul.* 15.4 (2017), pp. 1376–1403. DOI: 10.1137/16M1099443.

[KP92] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. 1st ed. Vol. 23. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 1992. ISBN: 978-3-540-54062-5. DOI: 10.1007/978-3-662-12616-5.

[KLO12] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation*. 1st ed. Vol. 345. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 2012. ISBN: 978-3-642-29879-0. DOI: 10.1007/978-3-642-29880-6.

[Kor+25] S. Kornblith et al. *JuliaDSP/DSP.jl*. Version v0.8.4. 2025. DOI: 10.5281/zenodo.15632710.

[KPK13] S. Krumscheid, G. A. Pavliotis, and S. Kalliadasis. "Semiparametric drift and diffusion estimation for multiscale diffusions." In: *Multiscale Model. Simul.* 11.2 (2013), pp. 442–473. DOI: 10.1137/110854485.

[Kru+15] S. Krumscheid, M. Pradas, G. A. Pavliotis, and S. Kalliadasis. "Data-driven coarse graining in action: modeling and prediction of complex systems." In: *Phys. Rev. E* 92.4 (2015), p. 042139. DOI: 10.1103/PhysRevE.92.042139.

[Kru18] S. Krumscheid. "Perturbation-based inference for diffusion processes: obtaining effective models from multiscale data." In: *Math. Models Methods Appl. Sci.* 28.8 (2018), pp. 1565–1597. DOI: 10.1142/S0218202518500434.

[Kue15] C. Kuehn. *Multiple Time Scale Dynamics*. 1st ed. Vol. 191. Applied Mathematical Sciences. Springer Cham, 2015. ISBN: 978-3-319-12315-8. DOI: 10.1007/978-3-319-12316-5.

[Kut76] Y. A. Kutoyants. "On a problem of testing hypotheses and asymptotic normality of stochastic integrals." In: *Theory Probab. Appl.* 20.2 (1976), pp. 376–384. DOI: 10.1137/1120042.

[Kut94] Y. A. Kutoyants. *Identification of Dynamical Systems with Small Noise*. 1st ed. Vol. 300. Mathematics and Its Applications. Springer Dordrecht, 1994. ISBN: 978-0-7923-3053-0. DOI: 10.1007/978-94-011-1020-4.

[Kut04] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. 1st ed. Springer Series in Statistics. Springer London, 2004. ISBN: 978-1-85233-759-9. DOI: 10.1007/978-1-4471-3866-2.

[Lie96] G. M. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific, 1996. ISBN: 978-981-02-2883-5. DOI: 10.1142/3302.

[LS00] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes II: Applications*. 2nd ed. Vol. 6. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 2000. ISBN: 978-3-540-63928-2. DOI: 10.1007/978-3-662-10028-8.

[Lor17] L. Lorenzi. *Analytical Methods for Kolmogorov Equations*. 2nd ed. Monographs and Research Notes in Mathematics. Chapman & Hall/CRC, 2017. ISBN: 978-1-4822-4332-1.

[MP19] T. Manikas and A. Papavasiliou. "Diffusion parameter estimation for the homogenized equation." In: *Multiscale Model. Simul.* 17.2 (2019), pp. 675–695. DOI: 10.1137/18M120138X.

[MT59] G. Maruyama and H. Tanaka. "Ergodic property of N-dimensional recurrent Mrkov processes." In: *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics* 13.2 (1959), pp. 157–172. DOI: 10.2206/kyushumfs.13.157.

[MSH02] J. Mattingly, A. Stuart, and D. Higham. "Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise." In: *Stoch. Process. Their Appl.* 101.2 (2002), pp. 185–232. ISSN: 0304-4149. DOI: 10.1016/S0304-4149(02)00150-3.

[MST10] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov. "Convergence of numerical time-averaging and stationary measures via Poisson equations." In: *SIAM J. Numer. Anal.* 48.2 (2010), pp. 552–577. DOI: 10.1137/090770527.

[MS11] I. Melbourne and A. M. Stuart. "A note on diffusion limits of chaotic skew-product flows." In: *Nonlinearity* 24.4 (2011), p. 1361. DOI: 10.1088/0951-7715/24/4/018.

[Mil84] P. W. Millar. "A general approach to the optimality of minimum distance estimators." In: *Trans. Amer. Math. Soc.* 286 (1984), pp. 377–418. DOI: 10.1090/S0002-9947-1984-0756045-0.

[MR18] P. K. Mogensen and A. N. Riseth. "Optim: a mathematical optimization package for Julia." In: *J. Open Source Softw.* 3.24 (2018), p. 615. DOI: 10.21105/joss.00615.

[NW06] J. Nocedal and S. J. Wright. *Numerical Optimization*. 2nd ed. Springer Series in Operations Research and Financial Engineering. Springer New York, NY, 2006. ISBN: 978-0-387-30303-1. DOI: 10.1007/978-0-387-40065-5.

[Øks03] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. 6th ed. Universitext. Springer Berlin, Heidelberg, 2003. ISBN: 978-3-540-04758-2. DOI: 10.1007/978-3-642-14394-6.

[OSP10] 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." In: *Multiscale Model. Simul.* 8.2 (2010), pp. 393–427. DOI: 10.1137/090756363.

[PSV77] G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan. "Martingale approach to some limit theorems." In: *Duke Turbulence Conference*. Duke University Mathematics Series 3. Duke University, 1977, pp. ii+120.

[Pap76] G. Papanicolaou. "Some probabilistic problems and methods in singular perturbations." In: *Rocky Mountain J. Math.* 6.4 (1976), pp. 653–674. DOI: 10.1216/RMJ-1976-6-4-653.

[PPS09] A. Papavasiliou, G. Pavliotis, and A. Stuart. "Maximum likelihood drift estimation for multiscale diffusions." In: *Stoch. Process. Their Appl.* 119.10 (2009), pp. 3173–3210. DOI: 10.1016/j.spa.2009.05.003.

[Par99] É. Pardoux. "Homogenization of linear and semilinear second order parabolic PDEs with periodic coefficients: A probabilistic approach." In: *J. Funct. Anal.* 167.2 (1999), pp. 498–520. DOI: 10.1006/jfan.1999.3441.

[PR14] É. Pardoux and A. Răşcanu. *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*. Vol. 69. Stochastic Modelling and Applied Probability. Springer Cham, 2014. ISBN: 978-3-319-05713-2. DOI: 10.1007/978-3-319-05714-9.

[PV01] É. Pardoux and A. Y. Veretennikov. "On the Poisson equation and diffusion approximation. I." In: *Ann. Probab.* 29.3 (2001), pp. 1061–1085. DOI: 10.1214/aop/1015345596.

[PV03] É. Pardoux and A. Y. Veretennikov. "On the Poisson equation and diffusion approximation. II." In: *Ann. Probab.* 31.3 (2003), pp. 1166–1192. DOI: 10.1214/aop/1055425774.

[PV05] É. Pardoux and A. Y. Veretennikov. "On the Poisson equation and diffusion approximation. III." In: *Ann. Probab.* 33.3 (2005), pp. 1111–1133. DOI: 10.1214/009117905000000062.

[PS07] G. A. Pavliotis and A. M. Stuart. "Parameter estimation for multiscale diffusions." In: *J. Stat. Phys.* 127 (2007), pp. 741–781. DOI: 10.1007/s10955-007-9300-6.

[PS08] G. A. Pavliotis and A. M. Stuart. *Multiscale Methods: Averaging and Homogenization*. Vol. 53. Texts in Applied Mathematics. Springer New York, NY, 2008. ISBN: 978-0-387-73828-4. DOI: 10.1007/978-0-387-73829-1.

[PPS12] G. A. Pavliotis, Y. Pokern, and A. M. Stuart. "Parameter estimation for multiscale diffusions: An overview." In: *Statistical Methods for Stochastic Differential Equations*. Ed. by M. Kessler, A. Lindner, and M. Sørensen. Monographs on Statistics and Applied Probability 124. Chapman and Hall/CRC, 2012. Chap. 7. ISBN: 978-0-4290-6325-1. DOI: 10.1201/b12126.

[PRZ25] G. A. Pavliotis, S. Reich, and A. Zanoni. "Filtered data based estimators for stochastic processes driven by colored noise." In: *Stoch. Process. Their Appl.* 181 (2025), p. 104558. DOI: 10.1016/j.spa.2024.104558.

[PB10] E. Platen and N. Bruti-Liberati. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Vol. 64. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 2010. ISBN: 978-3-642-12057-2. DOI: 10.1007/978-3-642-13694-8.

[Pra99] B. L. S. Prakasa Rao. *Statistical Inference for Diffusion Type Processes*. Vol. 8. Kendall's Library of Statistics. Arnold, 1999. ISBN: 0-340-74149-X.

[Pro03] P. Protter. *Stochastic Integration and Differential Equations*. 2nd ed. Vol. 21. Stochastic Modelling and Applied Probability. Springer Berlin, Heidelberg, 2003. ISBN: 978-3-540-00313-7. DOI: 10.1007/978-3-662-10061-5.

[RY98] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. 3rd ed. Vol. 293. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 1998. ISBN: 978-3-540-64325-8. DOI: 10.1007/978-3-662-06400-9.

[RM51] H. Robbins and S. Monro. "A stochastic approximation method." In: *Ann. Math. Statist.* 22.3 (1951), pp. 400–407. DOI: 10.1214/aoms/1177729586.

[RX21] M. Röckner and L. Xie. "Diffusion approximation for fully coupled stochastic differential equations." In: *Ann. Probab.* 49.3 (2021), pp. 1205–1236. DOI: 10.1214/20-AOP1475.

[Sas13] Z. Sasvari. *Multivariate Characteristic and Correlation Functions*. Vol. 50. De Gruyter Studies in Mathematics. De Gruyter, 2013. DOI: 10.1515/9783110223996.

[Sch10] T. Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Vol. 21. Interdisciplinary Applied Mathematics. Springer New York, NY, 2010. ISBN: 978-1-4419-6350-5. DOI: 10.1007/978-1-4419-6351-2.

[Sic21] W. Sickel. "On the regularity of characteristic functions." In: *Anomalies in Partial Differential Equations*. Ed. by M. Cicognani, D. Del Santo, A. Parmeggiani, and M. Reissig. Vol. 43. Springer INdAM Series. Springer Cham, 2021, pp. 395–441. ISBN: 978-3-030-61346-4. DOI: 10.1007/978-3-030-61346-4_18.

[SC13] K. Spiliopoulos and A. Chronopoulou. "Maximum likelihood estimation for small noise multiscale diffusions." In: *Stat. Inference Stoch. Process.* 16 (2013), pp. 237–266. DOI: 10.1007/s11203-013-9088-8.

[Str08] D. W. Stroock. *Partial Differential Equations for Probabilists*. Vol. 112. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2008. ISBN: 9780511755255. DOI: 10.1017/CBO9780511755255.

[Str23] D. W. Stroock. *Gaussian Measures in Finite and Infinite Dimensions*. 1st ed. Universitext. Springer Cham, 2023. ISBN: 978-3-031-23121-6. DOI: 10.1007/978-3-031-23122-3.

[SV79] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Vol. 233. Classics in Mathematics. Reprint edition, 2014. Springer Berlin, Heidelberg, 1979. ISBN: 978-3-662-22201-0. DOI: 10.1007/3-540-28999-2.

[Sun89] P. Sundar. "Ergodic solutions of stochastic differential equations." In: *Stochastics Stoch. Rep.* 28.2 (1989), pp. 65–83. DOI: 10.1080/17442508908833584.

[Sze75] G. Szegő. *Orthogonal Polynomials*. 4th ed. Vol. 23. American Mathematical Society Colloquium Publications. American Mathematical Society, 1975. ISBN: 978-0-8218-1023-1.

[Tha93] S. Thangavelu. *Lectures on Hermite and Laguerre Expansions*. Vol. 42. Mathematical Notes. Princeton University Press, 1993. ISBN: 978-0-691-00048-0.

[TAL97] R. Tolimieri, M. An, and C. Lu. *Mathematics of Multidimensional Fourier Transform Algorithms*. 2nd ed. Signal Processing and Digital Filtering. Springer New York, NY, 1997. ISBN: 978-0-387-98260-1. DOI: 10.1007/978-1-4612-1948-4.

[Tsa10] R. S. Tsay. *Analysis of Financial Time Series*. 3rd ed. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2010. ISBN: 9780470414354. DOI: 10.1002/9780470644560.

[Tsy08] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. 1st ed. Springer Series in Statistics. Springer New York, NY, 2008. ISBN: 978-0-387-79051-0. DOI: 10.1007/b13794.

[Vaa98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. ISBN: 9780511802256. DOI: 10.1017/CBO9780511802256.

[VW23] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. 2nd ed. Springer Series in Statistics. Springer Cham, 2023. ISBN: 978-3-031-29038-1. DOI: 10.1007/978-3-031-29040-4.

[Ver97] A. Y. Veretennikov. "On polynomial mixing bounds for stochastic differential equations." In: *Stoch. Process. Their Appl.* 70.1 (1997), pp. 115–127. DOI: 10.1016/S0304-4149(97)00056-2.

[Ver00] A. Y. Veretennikov. "On polynomial mixing and convergence rate for stochastic difference and differential equations." In: *Theory Probab. Appl.* 44.2 (2000), pp. 361–374. DOI: 10.1137/S0040585X97977550.

[Vir+20] P. Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." In: *Nat. Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[Was05] L. Wasserman. *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer, 2005. ISBN: 978-0-387-25145-5. DOI: 10.1007/0-387-30623-4.

[WLV05] E. Weinan, D. Liu, and E. Vanden-Eijnden. "Analysis of multiscale methods for stochastic differential equations." In: *Comm. Pure Appl. Math.* 58.11 (2005), pp. 1544–1585. DOI: 10.1002/cpa.20088.

[Wig03] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. 2nd ed. Vol. 2. Texts in Applied Mathematics. Springer New York, NY, 2003. ISBN: 978-0-387-00177-7. DOI: 10.1007/b97481.

[Wol57] J. Wolfowitz. "The minimum distance method." In: *Ann. Math. Statist.* 28.1 (1957), pp. 75–88. DOI: 10.1214/aoms/1177707038.

[Zan22] A. Zanoni. "Filtered Data and Eigenfunction Estimators for Statistical Inference of Multiscale and Interacting Diffusion Processes." PhD thesis. École Polytechnique Fédérale de Lausanne, 2022. DOI: 10.5075/epfl-thesis-10458.

[ZMA05] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. "A tale of two time scales: determining integrated volatility with noisy high-frequency data." In: *J. Am. Stat. Assoc.* 100.472 (2005), pp. 1394–1411. DOI: 10.1198/016214505000000169.