Parameter estimation for SPDEs based on discrete observations in time and space

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Mathematik
Universität Hamburg

vorgelegt von

Florian Hildebrandt
Hamburg, 2021

Folgende Gutachter empfehlen die Annahme der Dissertation:

Prof. Dr. Mathias Trabs
Prof. Dr. Markus Bibinger
Prof. Dr. Mark Podolskij

Vorsitzender der Prüdungskommission:

Prof. Dr. Thomas Schmidt

Tag der Disputation: 23. März 2021

Danksagungen

Zunächst gilt mein Dank Mathias Trabs für die exzellente Betreuung meiner Doktorarbeit. Von ihm stammt die thematische Idee für diese Dissertation und er hat durch seine Anregungen und unsere Diskussionen entscheidend zu den in dieser Arbeit präsentierten Erkenntnissen beigetragen. Für mich hatte er stets ein offenes Ohr in allen Belangen und dazu eine Tasse Kaffee. Seine Zuversicht bei der Problemlösung („muss ja irgendwie gehen…“) war immer ein großer Motivator.

Neben Mathias Trabs haben Markus Bibinger und Mark Podolskij die Begutachtung dieser Arbeit übernommen. Dafür sei ihnen herzlich gedankt.

Meine Mitdoktoranden Sebastian Neblung und Maximilian Steffen hatten sich freundlicherweise bereiterklärt, diese Arbeit Korrektur zu lesen. Ihre Kommentare waren für mich bei der Fertigstellung der Dissertation sehr hilfreich. Dafür vielen Dank!

An dieser Stelle möchte ich weiterhin Sylvie Rœlly nennen, die mich während meines Studiums an der Universität Potsdam, insbesondere im Rahmen meiner Masterarbeit, betreute. Dadurch habe ich viele Dinge gelernt, die mir bei der Arbeit an dieser Dissertation auf mehreren Ebenen sehr hilfreich waren. Dafür, sowie den Zuspruch, eine Doktorarbeit in Angriff zu nehmen, bedanke ich mich herzlich.

Meinen Mitdoktoranden erster Stunde, insbesondere Daniel Ebel, Nick Kloodt, Miran Knežević und Maria Mohr, danke ich für die herzliche Aufnahme in ihren Kreis. Ohne sie wären die letzten vier Jahre nicht annähernd so angenehm und unterhaltsam gewesen.

Meinen Eltern danke ich dafür, mich stets in jeglicher Form während meines Studiums und während meiner Promotion unterstützt zu haben. Durch ihr großes Interesse, ihre Anteilnahme und ihre Besuche in Hamburg waren sie immer ein wichtiger Begleiter durch meine Promotion.

Maria Eberlein hat allein schon dadurch, dass sie mit mir den Weg nach Hamburg auf sich genommen hat, essenziell zu dieser Arbeit beigetragen. Für ihre Unterstützung, ihre Aufmunterungen, wenn es mal nicht so gut lief, sowie überragende gemeinsame Momente im Klappstuhl bin ich unendlich dankbar.

Contents

Introduction

Stochastic partial differential equations (SPDEs) combine the ability of deterministic PDE models to describe complex mechanisms with the key feature of diffusion models, namely a stochastic signal which evolves within the system. While SPDEs have been intensively studied in stochastic analysis, their statistical theory is only at its beginnings. The prototype example for the class of parabolic SPDEs is given by the stochastic heat equation on an interval, namely

$$dX_t(x) = \vartheta_2 \frac{\partial^2}{\partial x^2} X_t(x)\, dt + \sigma\, dW_t(x), \qquad (t,x) \in \mathbb{R}_+ \times [0,1], \tag{1}$$

with Dirichlet boundary conditions and driven by a white noise dW in space and time. The natural associated statistical problem is how to infer on the model parameters $\sigma^2 > 0$ and $\vartheta_2 > 0$ based on some kind of observation derived from a realization of X. Inspired by the various applications of stochastic processes in the natural sciences as well as mathematical finance, enormous efforts have been spent on the development of statistical methodology for finite dimensional processes like diffusions or semimartingales during the last decades. We refer to the textbooks Jacod and Protter [45] and Liptser and Shiryaev [52] for an extensive treatment of the topic from different perspectives. Due to fundamental structural differences, the existing theory for finite dimensional processes is not directly applicable to the infinite dimensional framework of SPDEs. Let us mention two exemplary differences: Firstly, for X defined via (1), the marginal process $t \mapsto X_t(x_0)$ at a fixed spatial location $x_0 \in (0,1)$ has a finite quartic variation (cf. Swanson [73]) and, hence, it shows a much rougher behavior in time than finite dimensional diffusion processes. Secondly, directly related to the above mentioned statistical problem, it has been shown for multiple types of observation schemes (see below) that the parameter ϑ_2 can be estimated consistently on a finite time interval, despite its resemblance in the underlying equation of a drift parameter in a finite dimensional SDE. As a consequence, it is necessary to develop new statistical tools, paying attention to the infinite dimensional nature of SPDE models. Considering multiple types of observation schemes and employing various different mathematical techniques for their analysis, this topic has become increasingly popular in the statistics literature during the last few years.

Despite the structural differences, previous works have shown that methods from the high-frequency literature on semimartingales, e.g., power variations, can be fruitfully adapted for solving statistical problems associated with SPDEs, see, e.g., [9, 15, 19]. Moving forward in this direction, in this thesis we consider the practically most relevant situation where the solution X of (a generalization of) equation (1) is observed at a discrete grid

$$(t_i, y_k)_{i=0,\dots,N,k=0,\dots,M} \subset [0,T] \times [0,1]$$

in time and space with $T > 0$ either fixed or $T \to \infty$. Our focus lies on the scenario of high frequency observations in time and space such that, in particular, both the number M of spatial observations and the number N of temporal observations tend to infinity.

Since we first need to have a thorough statistical understanding for basic SPDEs before more complex models can be studied, a large part of this thesis is concerned with the analysis of the estimation problem for the diffusivity parameter ϑ_2 and the volatility parameter σ^2 in the stochastic heat equation (1). In fact, somewhat more general than $\vartheta_2 \frac{\partial^2}{\partial x^2}$, we will consider the linear second order differential operator

$$A_\vartheta := \vartheta_2 \frac{\partial^2}{\partial x^2} + \vartheta_1 \frac{\partial}{\partial x} + \vartheta_0 \,\mathrm{id} \tag{2}$$

with additional parameters ϑ_1, $\vartheta_0 \in \mathbb{R}$. Despite its relative simplicity, developing statistical methodology for this model is not only interesting from a mathematical point of view. Apart from the stochastic heat equation, our theory covers models, e.g., from neurobiology [76] or for the description of interest rates [24]. Besides the statistical theory, this thesis also discusses the problem how discrete samples of the linear SPDE model with differential operator A_ϑ can be generated efficiently for the purpose of simulations.

Once we have achieved a good understanding for the estimation problem for the linear SPDE model, we expand the realm of our theory to the semilinear framework of reaction-diffusion systems, namely

$$dX_t(x) = \big(\vartheta_2 \frac{\partial^2}{\partial x^2} X_t(x) + f(X_t(x))\big)\, dt + \sigma \, dW_t(x), \qquad (t,x) \in \mathbb{R}_+ \times [0,1], \tag{3}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a possibly nonlinear function on which we impose no parametric assumptions. Here, our aim is to infer on the parameters ϑ_2 and σ^2 as well as on the function f. Reaction-diffusion equations describe a scenario where local production of some quantity X with the nonlinear rate function f competes with a linear diffusion effect while undergoing internal fluctuations, see [57] as well as, e.g., [35] for the physical background. Of particular interest is the case where f is a polynomial of odd degree with a negative leading coefficient. Of course, setting $f \equiv 0$, this model includes the linear stochastic heat equation (1) and, more generally, with $f(x) := \vartheta_0 x$ we arrive at the linear SPDE model with differential operator A_ϑ and $\vartheta = (\vartheta_2, 0, \vartheta_0)$. The case $\vartheta_1 \neq 0$ could be incorporated into the model by adding a first order derivative $\vartheta_1 \frac{\partial}{\partial x} X_t(x)\, dt$ on the right hand side of (3) which is excluded from our analysis for the sake of simplicity.

In order to set the scene, we will first review several approaches from the literature on statistics for SPDEs before exhibiting the contributions of this thesis. The structure of the thesis will be described at the end of this introduction.

State of the art in statistics for SPDEs

As already indicated, there still is a lot of groundwork to be done in statistics for SPDEs, even for simple linear equations. Just like the models (1) and (3), most systems treated in the literature fall into the category

$$dX_t = \big(\vartheta_2 A X_t + F(X_t)\big)\, dt + \sigma B \, dW_t$$

where the solution $(X_t)_{t\geq 0}$ is a process taking values in some separable Hilbert space H, e.g., $H = L^2(\mathcal{O})$ for some open bounded set $\mathcal{O} \subset \mathbb{R}^d$. The coefficients A and B are linear operators and $(W_t)_{t\geq 0}$ is a cylindrical Brownian motion on H. Mostly, the case where F is another linear operator, or even $F \equiv 0$, is considered in the literature. A setting where F can be a nonlinear operator acting on some subset of H has only been addressed during the last few years, for further details see also the section about estimation in a semilinear framework of this introduction.

The larger part of the existing literature on statistics for SPDEs deals with estimation of the diffusivity parameter ϑ_2 when the operator A admits a complete orthonormal system of eigenfunctions $(e_k)_{k\geq 1}$. In the so called *spectral approach*, it is assumed that observations of the first n Fourier modes $\langle X_t, e_k \rangle_H$, $t \in [0,T]$, $k \leq n$, are available and the asymptotic properties of estimators are studied when $n \to \infty$. In particular, if $F \equiv 0$ and B can be diagonalized with respect to the same basis as A, these Fourier modes are one-dimensional independent Ornstein-Uhlenbeck processes. Thus, it is possible to build statistical procedures based on the existing methodology for finite dimensional SDEs. Starting with the seminal papers Huebner et al. [40] and Huebner and Rozovskii [41], the spectral approach has been applied to various linear models. We refer to the review papers [53] and [17] for further references, an extension to a semilinear framework is treated in [18, 68].

Instead of using the eigenfunctions of A as test functions, Altmeyer and Reiß [3] considered observations $\langle X_t, K_h \rangle$, $t \leq T$, where K_h is a kernel function localizing in space as $h \to 0$. With A being a second order differential operator and $F \equiv 0$, the authors derive nonparametric estimators for a spatially varying diffusivity parameter $\vartheta_2 = \vartheta_2(y)$ based on a likelihood approach. The parametric problem for constant diffusivity has also been treated in a semilinear framework, see [2].

Starting with Koski and Loges [50], multiple authors have also considered estimation problems for observations of the full trajectory $(X_t)_{t \leq T}$ or a functional thereof when $T \to \infty$, see, e.g., [46, 51, 60].

All the approaches mentioned so far have the major disadvantage that they essentially rely on a full spatial resolution which is an unrealistic scenario in practice. The spectral approach has the further disadvantage that it requires the eigenfunctions of the operator to be known. This assumption is already violated if we consider the operator (2) with unknown $\vartheta_1 \in \mathbb{R}$. Inspired by the practically most realistic scenario, the canonical problem of parameter estimation based on fully discrete observations of the solution field of the SPDE recently attracted an increased research activity. In order to facilitate the analysis of discrete observations in a conceptual L^2-setup, authors have considered very concrete and, usually, one-dimensional systems where rather explicit representations of the solution process are available. A first step in that direction was made by Markussen [59] who considered an approximate maximum likelihood estimator when the process is observed at finitely many spatial locations at a fixed temporal frequency with $T \to \infty$. Since then, central limit theorems for method of moment type estimators based on realized variations in space or time have been studied for various linear SPDEs by, e.g., Torres et al. [74], Cialenco and Huang [19], Bibinger and Trabs [8, 9], Shevchenko et al. [71], Mahdi Khalil and Tudor [56] as well as Kaino and Uchida [48, 49]. Working in a more general framework, Chong [14, 15] studied temporal variations at finitely many spatial locations for the linear stochastic heat equation on $\mathbb{R}^d$ when the volatility σ is a random field of time and space or a function of the solution process (multiplicative noise), respectively. Estimating the integrated volatility process over time at finitely many space points, these works fall into the realm of semiparametric statistics, see also [9] for the case of a deterministic time dependent volatility $\sigma = \sigma(t)$. Note that for discrete observations to be well defined, it is necessary that the solution process admits continuous trajectories. For the stochastic heat equation driven by space-time white noise this is only the case in spatial dimension one. In the multi-dimensional case one could consider noise processes which are more regular in space, as studied by Chong [15].

Own contributions and related literature

In the following, we describe the main contributions of this thesis in the context of the existing literature on the topic. The exposition is divided into three sections, each representing one of the main chapters of this thesis.

Parametric estimation for the linear equation

For the stochastic heat equation (1), method of moment type estimators based on realized variations have been studied by [9, 15, 19, 49]. However, all these works only give partial answers to the estimation problem: even for this basic model there is neither a sharp analysis for joint estimation of ϑ_2 and σ^2 nor the case where the number of spatial observations M dominates the number of temporal observations N has been explored in general.

We provide a complete statistical analysis of parametric estimation for linear parabolic SPDEs in dimension one based on discrete observations. Assuming, first, a fixed time horizon, our main contribution on the estimation problem for the parameters (σ^2, ϑ_2) in equation (1) reveal that:

(i) ϑ_2 and σ^2 cannot be jointly estimated if N or M is fixed.

(ii) The optimal convergence rate for estimating (ϑ_2, σ^2) is $1/\sqrt{M^3 \wedge N^{3/2}}$ which is generally slower than the parametric rate $1/\sqrt{MN}$.

(iii) Realized space-time quadratic variations can be used to construct estimators which are robust with respect to the sampling frequencies N and M in time and space, respectively.

In view of (i), we consider the double asymptotic regime $M, N \to \infty$ in our analysis which results in infill asymptotics in time and space. Since the vector of observations $(X_{t_i}(y_k))_{i=0,\dots,N,k=0,\dots,M}$ is normally distributed with only two unknown parameters, it might surprise that there is no estimator with parametric rate for (ϑ_2, σ^2). The lower bound which verifies this statement is at the heart of our analysis. It shows that the parametric rate can only be achieved if N and M^2 are of the same order of magnitude. In view of the scaling invariance of the stochastic heat equation, the particular asymptotic regime $\delta \approx \sqrt{\Delta}$ with $\delta := y_{k+1} - y_k \approx 1/M$ and $\Delta := t_{i+1} - t_i \approx T/N$ implies that we add the same amount of information in time and space as N and M increase. In this sense, we have a balanced design. An unbalanced regime $\Delta = o(\delta^2)$ or $\delta = o(\sqrt{\Delta})$ causes a deterioration of the convergence rate.

More generally, our findings apply to the situation where the Laplacian is replaced by the second order differential operator A_ϑ from (2) and, possibly, to a growing time horizon $T \to \infty$. While the parameter ϑ_1 can be estimated jointly with (σ^2, ϑ_2) and without affecting the rate of convergence, ϑ_0 cannot be identified in finite time and, thus, behaves like a classical drift parameter. Also on a growing time horizon $T \to \infty$, the parameters $(\sigma^2, \vartheta_2, \vartheta_1)$ can only be estimated with parametric rate of convergence in the balanced sampling design $\delta \approx \sqrt{\Delta}$. In general,

$$\Big(\frac{T}{\delta^3 \vee \Delta^{3/2}}\Big)^{-1/2}$$

takes the role of the optimal rate of convergence for $(\sigma^2, \vartheta_2, \vartheta_1)$. The parameter ϑ_0 can be estimated with the slower but optimal rate $1/\sqrt{T}$. Our main lower and upper bounds for fixed T or $T \to \infty$ are formulated in Theorems 2.1.2 and 2.1.7 as well as Corollaries 2.2.14 and 2.2.19, respectively

Our statistical analysis also gives insights into the relation between the spectral and the discrete observation scheme. While both are heuristically comparable in view of the discrete Fourier transform, it turns out that there are important differences. In particular, the fully discrete observation scheme is not statistically equivalent (in the sense of Le Cam) to time-discrete observations of the first M Fourier modes in general.

Our estimators rely on realized quadratic variations, taking into account time and space increments,

$$(\Delta_i^N X)(y_k) := X_{t_{i+1}}(y_k) - X_{t_i}(y_k), \qquad (\delta_k^M X)(t_i) := X_{t_i}(y_{k+1}) - X_{t_i}(y_k), \tag{4}$$

respectively, as well as space-time increments or double increments,

$$D_{ik} := (\delta_k^M \circ \Delta_i^N)X = (\Delta_i^N \circ \delta_k^M)X = X_{t_{i+1}}(y_{k+1}) - X_{t_{i+1}}(y_k) - X_{t_i}(y_{k+1}) + X_{t_i}(y_k). \tag{5}$$

In contrast to the maximum likelihood approach which requires inversion of the large $MN \times MN$ covariance matrix, method of moments type estimators based on (4) and (5) are easy to implement. For one-parameter processes, power variations are a standard tool in the statistical literature, see, e.g., Barndorff-Nielsen et al. [6] or the previously mentioned textbook [45]. Also, from a probabilistic point on view, there is a certain amount of literature devoted to variations based on double increments for some random field models, see, e.g., [66, 69].

For the stochastic heat equation on a finite time interval, it is observed in Bibinger and Trabs [9] that a central limit theorem for realized temporal quadratic variations requires that the observation frequency in time dominates the observation frequency in space, more precisely, $M = o(\sqrt{N})$ is necessary. Complementarily, we show in Theorem 2.2.3 that the realized spatial quadratic variation satisfies a central limit theorem if $N = o(M)$. The remaining gap can be filled by double increments and the corresponding realized space-time quadratic variation turns out to be robust with respect to the sampling frequencies M and N, see Theorem 2.2.7. Based on these statistics, we construct method of moments estimators for σ^2, ϑ_2 and ϑ_1. Our rate optimal method for joint estimation of $(\sigma^2, \vartheta_2, \vartheta_1)$ is an M-estimator relying on double increments. Additionally, if $T \to \infty$, the parameter ϑ_0 can be estimated employing a method of moments approach based on the sum of squares of the discrete observations. Concerning the rate of convergence, our estimator for the parameter vector $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ is a considerable improvement compared to Kaino and Uchida [49] who considered method of moment estimators from [9] together with a likelihood approach for an approximation of the first Fourier mode. Our proofs rely directly on the Gaussian distribution of X which allows for an explicit covariance condition for asymptotic normality of quadratic forms of Gaussian triangular schemes, see Proposition 2.2.1. Also, our estimators could be directly generalized to a nonparametric model with time dependent coefficients, as indicated in [9, 15].

Generating fully discrete samples for the linear equation

As usual, we investigate the finite sample performance of our estimators using simulations. For our SPDE model, computing realized variations from simulated data turns out to be a delicate task, even for the linear equation (1). In the previous works [9, 19, 48, 49] on discrete observations of the stochastic heat equation the primary foundation for the statistical theory and its simulation was the fact that the solution process admits a representation $X_t(y) = \sum_{\ell\geq 1} u_\ell(t)e_\ell(y)$ where $(u_\ell)_{\ell\geq 1}$ are independent one-dimensional Ornstein-Uhlenbeck processes and $(e_\ell)_{\ell\geq 1}$ are the eigenfunctions of the differential operator in the underlying equation. In particular, in order to simulate X on a space-time grid, the approximation $X_{t_i}^{\text{trunc}}(y_k) := \sum_{\ell=1}^{K} u_\ell(t_i)e_\ell(y_k)$ for some large integer K appears natural in view of the increasing drift towards 0 of the processes u_ℓ for $\ell \to \infty$. The individual processes u_ℓ can be simulated exactly based on their AR(1)-structure. As empirically observed, e.g., by Kaino and Uchida [49], the value of K has to be chosen carefully depending on the numbers of temporal and spatial observations N and M. In fact, even for moderate sample sizes, large values of K turn out to be crucial in order to prevent a severe bias in the simulated data. This makes simulations very costly.

Again, we work with our linear SPDE model where the associated differential operator is given by (2). Generalizing an idea stated in Davie and Gaines [29], we analyze an alternative approach, leading to almost exact (in distribution) discrete samples of X at a considerably lower computational cost. The two key observations leading to the method are: Firstly, the first M eigenfunctions e_ℓ are orthorgonal with respect to the empirical inner product, which yields a representation of the spatially discrete data in terms of a finite number of eigenfunctions. Secondly, for large values of ℓ the process $(u_\ell(t_i),\, 0 \leq i \leq N)$ can be approximated well by a set of independent random variables. Here, the coefficient processes corresponding to high Fourier modes are replaced by a set of independent random variables rather than truncated, hence, we shall call this approach the *replacement method*, as opposed to the *truncation method*. In order to provide a theoretical justification for the replacement method, we measure the quality of the approximation in terms of the total variation distance of the random vector $(X_{t_i}(y_k))_{i=0,\dots,N,k=0,\dots,M}$ from its approximation. By exploiting the Gaussian property of the involved processes, we derive an explicit bound on the corresponding approximation error. Again denoting $\Delta = t_{i+1} - t_i$, our precise analysis reveals that it is sufficient to generate discrete samples of $J \geq M$ Ornstein-Uhlenbeck processes accompanied by a sample of the same size of independent normal random variables, as long as (roughly) $J\sqrt{\Delta} \to \infty$, see Theorem 3.1.3.

The literature on approximation of SPDEs usually focuses on controlling errors of the type $\mathbf{E}(\|X(T) - X^{\text{a}}(T)\|_{L^2})$ (strong sense) or $|\mathbf{E}(\phi(X(T))) - \mathbf{E}(\phi(X^{\text{a}}(T))|$ (weak sense) for an approximation X^a of X, a fixed time instance T and a continuous functional ϕ, see e.g. [47]. Our primary goal, on the other hand, is to mimic the distribution of the discrete observations $(X_{t_i}(y_k))_{i=0,\dots,N,k=0,\dots,M}$ as well as possible, particularly when at least one of the numbers M and N tends to infinity, as is the case when computing realized power variations. The corresponding functionals, mapping sample paths to the asymptotic value of their power variations, are not continuous: a function close to zero can have arbitrarily rough paths. Hence, the known bounds on the strong or weak approximation error do not provide conditions under which the approximate power variation is close to the true one, in general. Here, controlling the total variation distance between the discrete sample and its approximation is an appropriate tool. Given that the total variation distance becomes negligible, functionals computed from the approximation converge to the correct weak limit (if existent), see also the discussion following Theorem 3.1.3. We remark that Chong and Walsh [16] examined the related question how finite difference approximations affect the asymptotic value of power variations of the stochastic heat equation.

Estimation in a semilinear framework

By estimating the parameters $(\sigma^2, \vartheta_2, f)$ of the system (3), we advance the theory on statistical estimation for SPDEs based on fully discrete observations into two new directions: First of all, by considering reaction-diffusion equations, we work in a nonlinear SPDE framework. Secondly, by estimating the function f on a whole continuum, we treat a fully nonparametric problem. Furthermore, regardless of the observation scheme, our estimator for f is the first nonparametric estimator of the nonlinearity in SPDEs.

The literature on statistics for semilinear SPDEs is limited and most works have only appeared within the last two years. Within the spectral approach, Cialenco and Glatt-Holtz [18] considered drift parameter estimation for the stochastic two-dimensional Navier-Stokes equation. Their findings were later generalized for more general equations, see [67, 68]. Also, the local measurements approach was generalized to the semilinear framework, see [1, 2]. For both observation schemes the solution of the semilinear equation is regarded as a perturbation of the linear case so that the corresponding estimation methods retain their validity under certain assumptions on the nonlinearity F. More precisely, the argument is that, within the decomposition $X_t = X_t^0 + N_t$ with X_t^0 being the solution to the corresponding linear system, the regularity of the nonlinear component N_t exceeds the regularity of X_t^0 in the Sobolev spaces $\mathcal{D}((-A)^\gamma)$, $\gamma > 0$.

Statistics for semilinear SPDEs based on discrete observations is a completely new field and has not yet been addressed in the literature.[1] Assuming $f \in C^1(\mathbb{R})$ in the model (3), we show that the asymptotic properties of our parametric estimators for (σ^2, ϑ_2) based on space, time and double increments largely remain valid in the semilinear framework, see Theorems 4.3.1 − 4.3.3. As for the other observation schemes, our findings rely on the higher order regularity of the nonlinear component of the solution process. Note that, being based on power variations, our estimators exploit the exact roughness of the sample paths of X. Thus, we rely on the regularity properties of X and its nonlinear component in the spaces of Hölder continuous functions instead of the Sobolev spaces. Besides the concrete application in statistics, our detailed account of the Hölder regularity in time and space also provides structural insights from a theoretical probabilistic point of view, see Propositions 4.2.1 and 4.2.3.

The theory covered in [2, 68] also applies to more general nonlinearities, e.g., Burgers' equation with $F(u) = -u\frac{\partial}{\partial x}u$, which is possible by considering a regularizing diffusion coefficient B. In fact, in order to obtain the higher order regularity of the nonlinear component, it is necessary for the solution process to take values in the domain of the nonlinearity F. This is not the case for Burgers' equation with the diffusion coefficient $B = I$ and we restrict our analysis to nonlinearities of Nemytskii-type where $F(u) = f \circ u$.

To the author's knowledge, the only work on estimation of the nonlinearity F in semilinear SPDEs was conducted by Goldys and Maslowski [33]. In this article, consistency of the maximum likelihood

[1] Prior to the final publication of this thesis, the independent study Cialenco et al. [20] on volatility or diffusivity estimation for semilinear SPDEs based on spatial variations at a single time instance was published as a preprint.

estimator based on a full observation $(X_t)_{t\le T}$, $T\to\infty$, of a controlled SPDE is proved, assuming that the nonlinearity involves a finite dimensional unknown parameter. Somewhat related, Pasemann et al. [67] studied estimation of the diffusivity parameter when the nonlinearity is only known up to a finite dimensional nuisance parameter by using a joint maximum likelihood approach within the spectral framework.

In this thesis, we consider nonparametric estimation of the function f from the reaction-diffusion equation (3) based on fully discrete observations. In contrast to estimation of the diffusivity ϑ_2, our results turn out to be comparable to drift estimation for finite dimensional SDEs. We adapt an approach by Comte et al. [21] who derived a simple least squares estimator for high frequency observations of ergodic one-dimensional diffusion processes. Their results are a generalization of earlier works on regression models, see [22, 23]. For more involved nonparametric drift estimation methods for high frequency observations of diffusion processes see also [32, 39]. Due to the properties of the linear equation, it is clear that even the simple linear function $f(x)=\vartheta_0 x$ cannot be identified in finite time and, in order to obtain general results, we work in the regime $T\to\infty$. Our key insight is that there is a regression type decomposition

$$\frac{X_{t+\Delta}-S(\Delta)X_t}{\Delta}=f(X_t)+\text{“stochastic noise term”}+\text{“negligible remainder terms”}$$

where $(S(t))_{t\ge0}$ is the strongly continuous semigroup on $L^2((0,1))$ generated by the operator $\vartheta_2\frac{\partial^2}{\partial x^2}$. Clearly, computing $S(\Delta)X_t$ is not feasible in the discrete observation scheme, as it depends on the whole process $X_t(x)$, $x\in(0,1)$, and we replace it by an empirical counterpart $S_t^{\Delta}:=\hat S(\Delta)X_t$. Additionally, the semigroup depends on the possibly unknown parameter ϑ_2 which we address by employing a plug-in approach with the help of an appropriate estimator. A nonparametric estimator $\hat f$ is then defined as the minimizer of

$$\gamma_{N,M}(g):=\frac{1}{MN}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}\Big(g(X_{t_i}(y_k))-\frac{X_{t_{i+1}}(y_k)-S_{t_i}^{\Delta}(y_k)}{\Delta}\Big)^2$$

over the functions g from a suitable finite dimensional approximation space. Working in an ergodic regime for the process $(X_t)_{t\ge0}$, we derive oracle inequalities for the risk of the estimator when the risk is either the empirical 2-norm with evaluations at the data points or the usual L^2-norm on a compact set. Our main oracle inequalities are formulated in Theorem 4.4.9 and Corollary 4.4.11, the case of an unknown diffusivity parameter is treated in Theorem 4.4.12. If one chooses the dimension of the approximation space owing to the usual variance-bias trade-off, we provide conditions under which the estimator achieves the usual nonparametric rate $T^{-\alpha/(2\alpha+1)}$ where α is some regularity parameter associated with f. In that sense, the result from [21] for finite dimensional systems carries over, though we need to require some stricter assumption in order for remainder terms to be negligible. Recall that our parametric estimators for (σ^2,ϑ_2) are constructed as method of moments estimators directly based on the covariance structure of the discrete observations. On the other hand, our nonparametric estimator for f is based on an approximation of the spatially continuous model. As a consequence, our estimation method requires a large amount of spatial observations, namely $M\Delta^2\to\infty$ is necessary. For a discussion on the problem of estimating f directly based on the discrete model we refer to the end of Section 4.4.

Structure of the thesis

This thesis consists of the three main Chapters 2–4, accompanied by an introduction of the statistical model (Chapter 1) as well as a conclusion and outlook (Chapter 5). The main chapters are divided into sections containing our main results and their discussion and a section in which we collect proofs and auxiliary results.

In Chapter 1 we provide a short introduction to the abstract framework for infinite dimensional stochastic equations and recall fundamental existence and uniqueness results for a class of SPDEs.

Afterwards, we introduce the models (1) and (3) thoroughly by embedding it into the abstract framework. For the linear system we discuss important properties of the law and the sample paths of the solution process. The fully discrete observation scheme considered throughout this thesis is introduced at the end of the chapter.

Chapter 2 is devoted to the problem of estimating the parameters $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ of the linear system. In the first part of the chapter we discuss identifiability of the parameters and derive lower bounds on the rate of convergence of estimators. The second part focuses on the construction of estimators. To that aim, we prove central limit theorems for the realized quadratic variation based on both space and double increments. These results are used to construct asymptotically normal method of moments estimators for the parameters. In particular, the rate of convergence of our estimator based on double increments matches the lower bound derived in the first part of the chapter. Finally, we hint at the possibility of constructing approximate confidence sets by exploiting the asymptotic normality of the estimators. The results of this chapter referring to a fixed finite time horizon can be found in Hildebrandt and Trabs [38].

With the primary goal of illustrating our results on the estimators from Chapter 2, in Chapter 3 we introduce the *replacement method* for generating fully discrete samples of the linear system. We measure the accuracy of the method by bounding the total variation distance between simulated and actual discrete observations. In a numerical example we compare the replacement method with naive truncation in Fourier space. Owing to its original purpose, the chapter is concluded with a simulation study on the estimators from Chapter 2. The introduction and analysis of the replacement method can be found in Hildebrandt [36], parts of the simulation study for the estimators are taken from Hildebrandt and Trabs [38].

Chapter 4 is devoted to estimating the parameters $(\sigma^2, \vartheta_2, f)$ of the reaction-diffusion equation (3). First, we discuss further regularity assumptions on the solution process, i.e., on f, required for our further analysis. Next, we derive precise results on the Hölder regularity in time and space of the solution process and show the higher order regularity of its nonlinear component. As a consequence, we are able to conclude that the asymptotic properties of our estimators for (σ^2, ϑ_2) mainly carry over from the linear setting. Then, we turn to nonparametric estimation of f. As a first step, we introduce the approximation spaces which serve as the candidate functions for estimating f. Then, assuming that the diffusivity is known, we define our estimator and derive a corresponding oracle inequality. This is done, first, based on spatially continuous observations of the solution process and, then, based on fully discrete observations via approximation arguments. If the regularity of f is known, the oracle inequalities can be used to determine an optimal dimension for the approximation space and a corresponding rate of convergence. The chapter is concluded by showing that our estimation procedure for f can be carried out without prior knowledge on the parameters (σ^2, ϑ_2) by using a plug-in approach. The results of this chapter are part of the preprint Hildebrandt and Trabs [37].

Notations

As usual, real numbers and integers are denoted by $\mathbb{R}$ and $\mathbb{Z}$, respectively, and we use the notation $\mathbb{R}_+ := [0, \infty)$. We follow the convention that the natural numbers do not include 0 and write $\mathbb{N} := \{1, 2, \ldots\}$ as well as $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

For $a, b \in \mathbb{R}$ we use the shorthand $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$. For two sequences $(a_n), (b_n)$, we write $a_n \lesssim b_n$ to indicate that there exists some $c > 0$ such that $|a_n| \leq c \cdot |b_n|$ for all $n \in \mathbb{N}$ and we write $a_n \approx b_n$ if $a_n \lesssim b_n \lesssim a_n$. Throughout $a_n, b_n \to \infty$ is meant in the sense of $a_n \wedge b_n \to \infty$ for $n \to \infty$. If $a_n = a$ for some $a \in \mathbb{R}$ and all $n \in \mathbb{N}$, we write $(a_n) \equiv a$. When we write statements like $M, N \to \infty$, we implicitly assume that M and N depend on a common index $n \in \mathbb{N}$ such that $N_n, M_n \to \infty$ for $n \to \infty$.

When there are no ambiguities, the norm on a normed space $\mathcal{X}$ is referred to as $\|\cdot\|_{\mathcal{X}}$. The Euclidian norm on $\mathbb{R}^d$ is denoted by $\|\cdot\|$. ∂S is the boundary of a set $S \subset \mathbb{R}^d$ with respect to $\|\cdot\|$ and we write $\bar{S} = S \cup \partial S$ for its completion. For a matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, $\|A\|_2$ denotes its spectral norm and $\|A\|_F$ denotes its Frobenius norm, i.e.,

$$\|A\|_2 := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}, \qquad \|A\|_F^2 := \sum_{i,j=1}^{n} a_{ij}^2. \tag{6}$$

For sets $B \subset C$, $\mathbf{1}_B : C \to \{0, 1\}$ denotes the indicator function of B, i.e., $\mathbf{1}_B(x) = 1$ for $x \in B$ and $\mathbf{1}_B(x) = 0$ for $x \in B^c := C \setminus B$. Further, $\operatorname{sgn} : \mathbb{R} \to \{-1, 1\}$ is defined via $\operatorname{sgn} := \mathbf{1}_{[0,\infty)} - \mathbf{1}_{(-\infty,0)}$.

For Banach spaces E and F, $L(E, F)$ is the set of continuous linear mappings $E \to F$ and, as usual, we write $L(E) := L(E, E)$. Further, $C(E, F)$ is the set of all continuous mappings $E \to F$ and we denote $\|f\|_\infty := \sup_{x \in E} \|f(x)\|_F$ for $f \in C(E, F)$. We follow the convention $C(E) := C(E, \mathbb{R})$. For $I \subset \mathbb{R}$ and $\alpha > 0$, $C^\alpha(I, F)$ is the set of Hölder continuous functions $I \to F$ of order α. In particular, for $\alpha \in \mathbb{N} \cup \{\infty\}$, $C^\alpha(I, F)$ is the set of α-times continuously differentiable functions $I \to F$.

For probability measures P and Q on a common measure space $(\Omega, \mathcal{F})$, $\mathrm{TV}(P, Q)$ and $\mathcal{H}(P, Q)$ denote total variation distance and Hellinger distance, respectively, i.e.,

$$\mathrm{TV}(P, Q) := \sup_{A \in \mathcal{F}} |P(A) - Q(A)| = \frac{1}{2} \int_\Omega \left| \frac{dP}{d\mu} - \frac{dQ}{d\mu} \right| d\mu, \tag{7}$$

$$\mathcal{H}^2(P, Q) := \int_\Omega \left| \sqrt{\frac{dP}{d\mu}} - \sqrt{\frac{dQ}{d\mu}} \right|^2 d\mu, \tag{8}$$

where μ is any dominating measure for P and Q. When X and Y are random variables taking values in a common measure space we also write $\mathrm{TV}(X, Y)$ and $\mathcal{H}(X, Y)$ for the distances of their distributions.

Convergence in probability and convergence in distribution are denoted by $\xrightarrow{\mathbf{P}}$ and $\xrightarrow{\mathcal{D}}$, respectively. For a sequence (X_n) of real random variables on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ we use the usual stochastic Landau symbols: for a sequence (a_n) in $(0, \infty)$ we write $X_n = o_p(a_n)$ to indicate that $X_n/a_n \to 0$ holds in probability and $X_n = \mathcal{O}_p(a_n)$ means that for any $\varepsilon > 0$ there exists $C > 0$ such that $\mathbf{P}(|X_n/a_n| \geq C) \leq \varepsilon$ for all sufficiently large $n \in \mathbb{N}$.

Chapter 1

Essentials

The aim of this chapter is to introduce the statistical model considered throughout this thesis. As a first step, in Section 1.1 we recall general concepts from the theory of stochastic equations in infinite dimensions. Then, as an application of the abstract framework in a concrete situation, we define our SPDE model and discuss important probabilistic properties of the solution process in Section 1.2. In the same section, the introduction of the statistical model is completed by specifying the discrete observation scheme considered in this thesis.

1.1 Prerequisites on SPDEs

The following section contains a short recap on existence and uniqueness results for (semi)linear stochastic differential equations with additive noise and locally Lipschitz continuous nonlinearity. In doing so, the concept of stochastic integration in Hilbert spaces is taken for granted, the corresponding theory can be found, e.g., in Chapter 4 of Da Prato and Zabczyk [26]. The results presented here are taken from Chapters 5 and 7 of the same book.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. Further, let H be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and let $(e_l)_{l\in\mathbb{N}}$ be a corresponding complete orthonormal system. We assume that the probability space carries a cylindrical Brownian motion $W = (W_t)_{t\geq 0}$ in H, i.e., the processes $(\beta_l)_{l\in\mathbb{N}}$ defined by

$$\beta_l(t) := \langle W_t, e_l \rangle, \qquad t \geq 0,\, l \in \mathbb{N},$$

constitute a sequence of independent one-dimensional standard Brownian motions with respect to $(\mathcal{F}_t)_{t\geq 0}$. Clearly, the series $\sum_{l\geq 1} \beta_l(t) e_l$ is not convergent and, hence, W_t does not take values in H. Nevertheless, it can be regarded as a process with values in a larger Hilbert space $\tilde{H} \supset H$ and, in particular, it is possible to define a stochastic integral with respect to W. The relevant setting for this thesis is the case where $H = L^2(\mathcal{O})$ for some open bounded set $\mathcal{O} \subset \mathbb{R}^d$, in which W can be understood as the anti-derivative in time of space-time white noise.

The larger part of this thesis is concerned with linear stochastic partial differential equations with additive noise, i.e.,

$$dX_t = AX_t \, dt + BdW_t, \quad X_0 = \xi. \tag{1.1}$$

The initial value ξ is a $\mathcal{F}_0$-measurable H-valued random variable and the fundamental assumptions on the coefficients are the following:

(L1) $A : \mathcal{D}(A) \to H$ is a linear operator with domain $\mathcal{D}(A) \subset H$ generating a C_0-semigroup $S = (S(t))_{t\geq 0}$ on H.

(L2) $B : H \to H$ is a bounded linear operator such that $\int_0^t \|S(r)B\|_{\text{HS}}^2 \, dr < \infty$, $t \geq 0$.

Assumption (L1) means that S is a semigroup of bounded linear operators on H, satisfying $S(t)x \to x$ for all $x \in H$ and $\frac{S(t)x-x}{t} \to Ax$ for any $x \in \mathcal{D}(A)$ and $t \to 0$. Further, $\|R\|_{\text{HS}}^2 = \sum_{l \geq 1} \|Re_l\|_H^2$ denotes the Hilbert-Schmidt norm for a bounded linear operator R. The mild solution $(X_t)_{t \geq 0}$ of equation (1.1) is defined as the H-valued process

$$X_t = S(t)\xi + \int_0^t S(t-s)B \, dW_s, \qquad t \geq 0, \tag{1.2}$$

where the integral is a stochastic integral in the Hilbert space H, which is well-defined thanks to Assumption (L2). A priori, X is an H-valued process, though there are relevant situations in which, using the explicit representation (1.2), it can be shown that $X \in C(\mathbb{R}_+, E)$ holds almost surely for a smaller Banach space $E \subset H$ with norm $\|\cdot\|_E$, as long as $\xi \in E$. In our specific model, this will be the case with

$$H = L^2(\mathcal{O}), \qquad E = C_0(\bar{\mathcal{O}}) := \{u \in C(\bar{\mathcal{O}}) : \, u(x) = 0 \text{ for } x \in \partial\mathcal{O}\},$$

such that point evaluation $X_t(x)$ are well defined for $t \geq 0$, $x \in \mathcal{O}$. As usual, the space $C_0(\bar{\mathcal{O}})$ is equipped with the norm $\|\cdot\|_\infty$.

Later on, we will generalize our setting and consider semilinear stochastic differential equations with additive noise, i.e.,

$$dX_t = \big(AX_t + F(X_t)\big) \, dt + BdW_t, \quad X_0 = \xi, \tag{1.3}$$

where F is some, possibly, nonlinear mapping. Assuming, that the solution to the corresponding linear equation (1.1) takes values in a smaller Banach space $E \subset H$, it is desirable to have the same property for the nonlinear equation. Thus, we assume that F maps E into E and consider the part of A in E, namely

$$A_E x := Ax \quad \text{for} \quad x \in \mathcal{D}(A_E) = \{x \in \mathcal{D}(A) \cap E : \, Ax \in E\}. \tag{1.4}$$

Additionaly to (L1)-(L2), we assume the following:

(N1) The process $t \mapsto W_A(t) := \int_0^t S(t-s)B \, dW_s$ satisfies $W_A \in C(\mathbb{R}_+, E)$ almost surely.

(N2) A_E generates a C_0-semigroup $(S_E(t))_{t \geq 0}$ on E.

(N3) $F : E \to E$ is locally Lipschitz continuous and bounded on bounded subsets of E.

An adapted process $(X_t)_{t \geq 0}$ in $C(\mathbb{R}_+, E)$ is said to be a mild solution of equation (1.3) in E if it satisfies the variation of constants formula

$$X_t = S(t)\xi + \int_0^t S(t-s)F(X_s) \, ds + \int_0^t S(t-s)B \, dW_s, \qquad t \geq 0, \tag{1.5}$$

almost surely, where the first integral is a Bochner integral in the Banach space E. The above assumptions are sufficient to prove the existence of an E-valued solution locally in time. To show global existence, the concept of a subdifferential of the norm $\|\cdot\|_E$ at a point $x \in E$, denoted by $\partial\|x\|_E$, is useful: define

$$\partial\|x\|_E = \{x^* \in E^* : \, \langle x, x^* \rangle = \|x\|_E, \, \|x^*\|_{E^*} = 1\}, \tag{1.6}$$

where E^* is the topological dual space of E and $\langle x, x^* \rangle$ denotes the value of $x^* \in E^*$ applied to $x \in E$. Consider a function $u : \mathbb{R} \to E$, which is differentiable in $t_0 \in \mathbb{R}$. Then, it can be shown that the function $\gamma : \mathbb{R} \to \mathbb{R}$, $\gamma(t) := \|u(t)\|_E$, is differentiable from the left and from the right in t_0. Denoting the corresponding one-sided derivatives by $\frac{d^-\gamma}{dt}(t_0)$ and $\frac{d^+\gamma}{dt}(t_0)$, respectively, we have

$$\frac{d^+\gamma}{dt}(t_0) = \max\{\langle u'(t_0), x^*\rangle : x^* \in \partial\|u(t_0)\|_E\}, \qquad \frac{d^-\gamma}{dt}(t_0) = \min\{\langle u'(t_0), x^*\rangle : x^* \in \partial\|u(t_0)\|_E\}$$

and, in particular,

$$\frac{d^-\gamma}{dt}(t_0) \leq \langle u'(t_0), x^*\rangle \quad \text{for any} \quad x^* \in \partial\|u(t_0)\|_E. \tag{1.7}$$

Now, global existence of a solution can be deduced from the following assumptions:

(N4) $\|S_E(t)\|_{L(E)} \leq 1$ for all $t \geq 0$.

(N5) There is an increasing function $a : \mathbb{R}_+ \to \mathbb{R}_+$ such that for each $x \in E$ there exists $x^* \in \partial\|x\|_E$ with $\langle F(x+y), x^*\rangle \leq a(\|y\|_E)\big(1 + \|x\|_E\big)$ for all $y \in E$.

Theorem 1.1.1 (cf. [26, Theorem 7.7])**.** *Under Assumptions (L1)-(L2) and (N1)-(N5) equation* (1.3) *has a unique mild solution X in $C(\mathbb{R}_+, E)$ for each $\xi \in E$. Furthermore, X is an E-valued Markov process.*

Sketch of proof. For details, see Theorem 7.7 and Example 7.8 in [26]. Local existence of a solution follows from (N1)-(N3) and Banach's fixed point theorem. Next, split the solution process into its linear and its nonlinear component,

$$X_t = S(t)\xi + W_A(t) + N_t, \quad \text{with} \quad N_t = \int_0^t S(t-s)F(X_s)\, ds,$$

such that a global solution can be achieved by bounding $\|N_t\|_E$. The process $(N_t)_{t\geq 0}$ solves the integral equation $N_t = \int_0^t S(t-s)F\big(S(s)\xi + W_A(s) + N_s\big)\, ds$, $N_0 = 0$, and, using $(N4)-(N5)$ in connection with (1.7), it is possible to obtain a bound on $\frac{d^-}{dt}\|N_t\|_E$. Then, by applying Gronwall's inequality, a bound for $\|N_t\|_E$ follows. □

Let us discuss Assumption (N5) in the context of $H = L^2(\mathcal{O})$ and $E = C_0(\bar{\mathcal{O}})$ for an open bounded set $\mathcal{O} \subset \mathbb{R}^d$ with smooth boundary and a Nemytskii-type nonlinearity, i.e., $F(u) = f \circ u$ for a function $f \in C^1(\mathbb{R})$. In this situation, (N5) is satisfied if

$$f(\lambda + \eta)\operatorname{sgn}(\lambda) \leq a(|\eta|)(1 + |\lambda|), \qquad \lambda, \eta \in \mathbb{R}, \tag{1.8}$$

cf. [26, Example 7.8]: In fact, it follows from the definition (1.6) of the subdifferential of the norm that if $\|u\|_E = \sigma_u u(z_u)$ with $\sigma_u = \operatorname{sgn}(u(z_u))$ for some $z_u \in \mathcal{O}$ and $u \in E$, then the functional

$$h_u : E \ni v \mapsto \sigma_u v(z_u) \tag{1.9}$$

is an element of $\partial\|u\|_E$. Thus,

$$\langle f \circ (u+v), h_u\rangle = \operatorname{sgn}(u(z_u)) f(u(z_u) + v(z_u)) \leq a(|v(z_u)|)(1 + \|u\|_E) \leq a(\|v\|_E)(1 + \|u\|_E)$$

and, hence, (N5) holds. We remark that, formally, Theorem 1.1.1 is only applicable if $f(0) = 0$, due to the requirement that F maps E into E. This issue can be circumvented by replacing $W_A(t)$ by $W_A(t) + f(0)\int_0^t S(t-s)\mathbf{1}_{\mathcal{O}}\, ds$ and f by $f_0 := f - f(0)$.

It is shown in [26] that condition (1.8) is satisfied in the important example where f is a polynomial of odd degree with a negative leading coefficient. In fact, using the same argument, we can conclude that it is sufficient that $f \in C^1(\mathbb{R})$ is such that

$$\sup_{x\geq 0} f(x) < +\infty \quad \text{and} \quad \inf_{x\leq 0} f(x) > -\infty : \tag{1.10}$$

For $\lambda \geq 0$ we have $f(\lambda+\eta) \leq \sup_{h\geq -|\eta|} f(h)$ and for $\lambda < 0$ we have $-f(\lambda+\eta) \leq -\inf_{h\leq|\eta|} f(h)$. Thus, we can choose $a(x) := \max\Big(\sup_{h\geq -x} f(h), -\inf_{h\leq x} f(h), 0\Big)$, $x \geq 0$.

We close this section by remarking that, under the conditions of Theorem 1.1.1, the mild solution X is also a weak solution in the sense that

$$\langle X_t, z\rangle = \langle \xi, z\rangle + \int_0^t \Big(\langle X_s, A^*z\rangle + \langle F(X_s), z\rangle\Big)\, ds + \langle BW(t), z\rangle, \qquad t \geq 0,$$

holds almost surely for any $z \in \mathcal{D}(A^*)$, where A^* denotes the adjoint operator of A, cf. [26, Theorem 5.4].

1.2 Introduction of the model and basic properties

We introduce the statistical model considered throughout this thesis. The definition of our SPDE model is divided into the linear case, Section 1.2.1, and the semilinear case, Section 1.2.2. For the linear equation, we discuss important probabilistic properties of the solution process, including an Itô decomposition for the spatial process. The discrete observation scheme investigated in this thesis is introduced in Section 1.2.3. Finally, the Itô decomposition for the spatial process is proved in Section 1.2.4.

1.2.1 The linear equation

In Chapters 2 and 3 of this thesis we will study parameter estimation and simulation for the model defined by the following linear parabolic SPDE: for parameters $\sigma > 0$ and $\vartheta = (\vartheta_2, \vartheta_1, \vartheta_0) \in \mathbb{R}_+ \times \mathbb{R}^2$, we consider

$$\begin{cases} dX_t(x) = \Big(\vartheta_2 \frac{\partial^2}{\partial x^2} X_t(x) + \vartheta_1 \frac{\partial}{\partial x} X_t(x) + \vartheta_0 X_t(x)\Big)\, dt + \sigma\, dW_t(x), & x \in [0,1], t \geq 0, \\ X_t(0) = X_t(1) = 0, \\ X_0 = \xi, \end{cases} \tag{1.11}$$

driven by a cylindrical Brownian motion W and with some independent initial condition $\xi : [0,1] \to \mathbb{R}$.

To embed this SPDE into the theory presented in the previous section, we let $H = L^2((0,1))$ and, for reasons to become clear shortly, we replace the usual inner product on $L^2((0,1))$ by the weighted version

$$\langle u, v\rangle := \langle u, v\rangle_\vartheta := \int_0^1 u(x)v(x)\mathrm{e}^{\vartheta_1 x/\vartheta_2}\, dx, \qquad u, v \in L^2((0,1)).$$

We consider the mild solution of equation (1.1) with the differential operator $A_\vartheta = \vartheta_2 \frac{\partial^2}{\partial x^2} + \vartheta_1 \frac{\partial}{\partial x} + \vartheta_0$, the diffusion operator $Bx = \sigma x$, $x \in H$, and a cylindrical Brownian motion W in H. As usual, the Dirichlet boundary conditions in (1.11) are implemented by taking $\mathcal{D}(A_\vartheta) = H^2((0,1)) \cap H_0^1((0,1))$, where $H^k((0,1))$ denotes the L^2-Sobolev spaces of order $k \in \mathbb{N}$ and with $H_0^1((0,1))$ being the closure of $\{u \in C^\infty((0,1)) : u \text{ has compact support in } (0,1)\}$ in the space $H^1((0,1))$.

The reason for considering the weighted version of the inner product on $L^2((0,1))$ is that the differential operator A_ϑ now has a complete orthonormal system of eigenvectors: indeed, the corresponding eigenpairs $(-\lambda_\ell, e_\ell)_{\ell\geq 1}$ are given by

$$e_\ell(y) = \sqrt{2}\sin(\pi\ell y)\mathrm{e}^{-\kappa y/2}, \quad \lambda_\ell = \vartheta_2(\pi^2\ell^2 + \Gamma), \qquad y \in [0,1],\ \ell \in \mathbb{N},$$

$$\text{with} \qquad \kappa := \frac{\vartheta_1}{\vartheta_2} \qquad \text{and} \qquad \Gamma := \frac{\vartheta_1^2}{4\vartheta_2^2} - \frac{\vartheta_0}{\vartheta_2}.$$

Note that in absence of the first derivative in A_ϑ, i.e $\vartheta_1 = 0$, the system $(e_\ell)_{\ell\geq 1}$ reduces to the usual sine-base and $\langle\cdot,\cdot\rangle$ to the standard inner product on $L^2((0,1))$. In general, both the eigenpairs and the inner product depend on the model parameters. Hence, they are not accessible from a statistical point of view. Note also that W is a cylindrical Brownian motion with respect to $\langle\cdot,\cdot\rangle_\vartheta$ and, hence, its distribution implicitly hinges on ϑ, unless $\vartheta_1 = 0$.

Throughout, we restrict the parameter space to

$$\Theta = \Big\{(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0) \in \mathbb{R}^4 : \sigma^2, \vartheta_2, \frac{\vartheta_1^2}{4\vartheta_2^2} - \frac{\vartheta_0}{\vartheta_2} + \pi^2 > 0\Big\},$$

such that all the eigenvalues $-\lambda_\ell$ are negative and A_ϑ is a negative self-adjoint operator. In particular, A_ϑ generates a C_0-semigroup, which is explicitly given by

$$S(t)x = \sum_{\ell\geq 1} \mathrm{e}^{-\lambda_\ell t}\langle x, e_\ell\rangle e_\ell, \qquad x \in L^2((0,1)),\, t \geq 0.$$

Furthermore, we have $\|S(r)B\|_{\mathrm{HS}}^2 = \sigma^2 \sum_{\ell\geq 1} \mathrm{e}^{-2\lambda_\ell r}$ and

$$\int_0^t \|S(r)B\|_{\mathrm{HS}}^2\, dr = \sigma^2 \sum_{\ell\geq 1} \frac{1-\mathrm{e}^{-2\lambda_\ell t}}{2\lambda_\ell} < \infty.$$

Consequently, (L1) and (L2) are satisfied and for $\xi \in L^2((0,1))$, $X_t = S(t)\xi + \sigma\int_0^t S(t-s)\, dW_s$, $t \geq 0$, defines the mild solution to the SPDE (1.11). Note that, in general, the Dirichlet eigenvalues of the Laplacian on a bounded domain $\mathcal{O} \subset \mathbb{R}^d$ have growth $\lambda_\ell \approx \ell^{2/d}$, so that a function valued solution of the stochastic heat equation driven by a cylindrical Brownian motion only exists in dimension $d = 1$. Further, using the sequence of independent standard Brownian motions $(\beta_\ell)_{\ell\geq 1}$ with $\beta_\ell(t) = \langle W_t, e_\ell\rangle$, the cylindrical Brownian motion W can be realized via $W_t = \sum_{\ell\geq 1} \beta_\ell(t)e_\ell$ in the sense that $\langle W_t, \cdot\rangle = \sum_{\ell\geq 1} \beta_\ell(t)\langle\cdot, e_k\rangle$. Thus, in terms of the projections, or Fourier modes, $u_\ell(t) := \langle X_t, e_\ell\rangle$, $t \geq 0, \ell \in \mathbb{N}$, we obtain the representation

$$X_t(x) = \sum_{\ell\geq 1} u_\ell(t)e_\ell(x), \quad t \geq 0,\, x \in [0,1], \tag{1.12}$$

where $(u_\ell)_{\ell\geq 1}$ are one-dimensional processes satisfying the Ornstein-Uhlenbeck dynamics $du_\ell(t) = -\lambda_\ell u_\ell(t)\, dt + \sigma\, d\beta_\ell(t)$ or, equivalently,

$$u_\ell(t) = u_\ell(0)\mathrm{e}^{-\lambda_\ell t} + \sigma \int_0^t \mathrm{e}^{-\lambda_\ell(t-s)}\, d\beta_\ell(s), \qquad u_\ell(0) = \langle \xi, e_\ell\rangle$$

in the sense of the usual finite-dimensional stochastic integral.

Using independence of the Brownian motions driving the coefficient processes, one can explicitly compute the space-time covariance structure of X, namely

$$\mathrm{Cov}(X_s(x), X_t(y)) = \mathrm{Cov}(\xi_s(x), \xi_t(y)) + \sigma^2 \sum_{\ell\geq 1} \frac{\mathrm{e}^{-\lambda_\ell|t-s|} - \mathrm{e}^{-\lambda_\ell(t+s)}}{2\lambda_\ell} e_\ell(x)e_\ell(y),$$

for $s, t \geq 0$, $x, y \in [0,1]$ with $\xi_t := S(t)\xi$. Throughout and without further notice, we work under the standing assumption that either $X_0 = \xi = 0$, or that X is started in equilibrium. It would be possible to extend our results to more general initial conditions as long as they are sufficiently regular, as it is done, e.g., in Bibinger and Trabs [9]. This is omitted for the sake of simplicity. In order to mark results which are proved exclusively for the stationary case, we use the abbreviation (ST), i.e., we introduce the assumption

(ST) X_0 follows the stationary distribution.

Assumption (ST) can be realized by letting $(u_\ell(0))_{\ell\geq 1}$ be independent with $u_\ell(0) \sim \mathcal{N}(0, \sigma^2/(2\lambda_\ell))$, which corresponds to starting each of the coefficient processes in equilibrium. In this case, $(X_t)_{t\geq 0}$ is a strictly stationary process and the space-time covariance structure simplifies to

$$\mathrm{Cov}\,(X_s(x), X_t(y)) = \sigma^2 \sum_{\ell\geq 1} \frac{\mathrm{e}^{-\lambda_\ell |t-s|}}{2\lambda_\ell} e_\ell(x) e_\ell(y), \quad s,t \geq 0,\ x,y \in [0,1]. \tag{1.13}$$

For both initial conditions, the coefficient processes $(u_\ell)_{\ell\geq 1}$ are independent. Further, it is evident from representation (1.12) that X is a two-parameter centered Gaussian field. Therefore, the model is completely specified by its covariance structure. While σ^2 is only a multiplicative factor, the covariance structure depends on ϑ through λ_ℓ and e_ℓ.

Thanks to the explicit covariance structure, one can use Kolmogorov's criterion for random fields to show that there is a continuous version of the process $(X_t(x),\ t \geq 0, x \in [0,1])$, see [26, Chapter 5.5]. In particular, point evaluations $X_t(x)$ for fixed values of t and x are well defined and we have $X \in C(\mathbb{R}_+, E)$ almost surely with $E = C_0([0,1])$. Additionally, it can be inferred from Kolmogorov's criterion that the process $t \mapsto X_t(x)$ is locally α-Hölder continuous of any order $\alpha < 1/4$ and $x \mapsto X_t(x)$ is α-Hölder continuous of any order $\alpha < 1/2$, see also Section 4.2 of this thesis.

For a fixed spatial location x, the process $t \mapsto X_t(x)$ is not a semimartingale: since $t \mapsto X_t(x)$ is only Hölder continuous of order almost 1/4, it has infinite quadratic variation over any time interval. On the other hand, regarding X as a function of space at a fixed point in time substantially simplifies the probabilistic structure of the process:

Proposition 1.2.1. *Assume (ST) and define* $\Gamma_0 := \sqrt{|\Gamma|}$.

(i) For $x \leq y$, *we have*

$$\mathrm{Cov}\,(X_t(x), X_t(y)) = \frac{\sigma^2}{2\vartheta_2} \mathrm{e}^{-\frac{\kappa}{2}(x+y)} \cdot \begin{cases} \frac{\sin(\Gamma_0(1-y))\sin(\Gamma_0 x)}{\Gamma_0 \sin(\Gamma_0)}, & \Gamma < 0, \\ x(1-y), & \Gamma = 0, \\ \frac{\sinh(\Gamma_0(1-y))\sinh(\Gamma_0 x)}{\Gamma_0 \sinh(\Gamma_0)}, & \Gamma > 0. \end{cases}$$

(ii) The process $[0,1] \ni x \mapsto Z(x) := X_t(x)$ *is an Itô diffusion. In particular,*

$$dZ(x) = \sqrt{\frac{\sigma^2}{2\vartheta_2}} \mathrm{e}^{-\frac{\kappa}{2}x}\, dB(x) - \begin{cases} \left(\frac{\Gamma_0 \cos(\Gamma_0(1-x))}{\sin(\Gamma_0(1-x))} + \frac{\kappa}{2}\right) Z(x)\, dx, & \Gamma < 0, \\ \left(\frac{1}{1-x} + \frac{\kappa}{2}\right) Z(x)\, dx, & \Gamma = 0, \\ \left(\frac{\Gamma_0 \cosh(\Gamma_0(1-x))}{\sinh(\Gamma_0(1-x))} + \frac{\kappa}{2}\right) Z(x)\, dx, & \Gamma > 0, \end{cases}$$

where $B(\cdot) = B_t(\cdot)$ *is a standard Brownian motion.*

The above proposition is proved in Section 1.2.4. Note the similarity between the covariance structures of $X_t(\cdot)$ and the Brownian bridge, especially in the case $\Gamma = 0$. This resemblance is in line with the Dirichlet boundary conditions $X_t(0) = X_t(1) = 0$ in our model.

Remark 1.2.2. For $N \geq 2$ and fixed $0 \leq t_1 < t_2 < \ldots < t_N$ the multi-dimensional process $x \mapsto (X_{t_1}(x), \ldots, X_{t_N}(x))$ is not an Itô diffusion under (ST). Indeed, it is not even a Markov process: take $N = 2$ and let $s < t$. It is a well known fact that for Markov processes past and future are independent, given the present state. For $x < y < z$, on the other hand, using the Gaussian property of X, the (Gaussian) conditional distribution of $(X_s(x), X_t(z))$ given $(X_s(y), X_t(y))$ can be computed explicitly. From here, independence is easily disproved by checking the non-diagonal entries of the conditional covariance matrix.

We finish this section by describing the influence of the parameters $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ on the sample paths heuristically. Figure 1.1 shows exemplary realizations of temporal and spatial processes under (ST) for different parameter values. By construction, the spatial process is pinned in 0 at the spatial positions $y \in \{0, 1\}$. Furthermore, the simulations confirm that the temporal process is much rougher than the spatial process. The influence of ϑ_0 on the visual properties of the sample paths is small and, qualitatively, $-\vartheta_0$ has a comparable effect to ϑ_2. Thus, we only exhibit the case $\vartheta_0 = 0$. First, let us discuss the situation where also $\vartheta_1 = 0$. It is evident from the covariance structure (1.13) that σ^2 describes the overall noise level of the process. Furthermore, increasing ϑ_2 reduces the noise level of the solution process and, additionally, the temporal process speeds up. The first of these impacts of ϑ_2 is clearly noticeable when comparing the first two rows of plots in Figure 1.1. The second one is hard to capture visually due to the roughness of the processes. When $\vartheta_1 \neq 0$, the solution process $X_t(y)$ approximately looks like $\mathrm{e}^{-\kappa y/2}\tilde{X}_t(y)$ where $\tilde{X}$ solves $d\tilde{X}_t = \vartheta_2 \frac{\partial^2}{\partial x^2}\tilde{X}_t \, dt + \sigma dW_t$. Indeed, the covariance structure (1.13) reveals that the distributions of the two processes agree when setting $\vartheta_0 = \vartheta_1^2/(4\vartheta_2)$. As already mentioned, the latter has no strong visual effect. Thus, a parameter $\vartheta_1 \neq 0$ affects the noise level of the temporal process and produces spatial processes with different amounts of fluctuation in the two halves of the space domain $[0, 1]$. This effect is illustrated by the last two rows of plots in Figure 1.1.

1.2.2 The semilinear equation

In Chapter 4 we will complement our setting by considering the semilinear SPDE

$$\begin{cases} dX_t(x) = \left(\vartheta_2 \frac{\partial^2}{\partial x^2} X_t(x) + F(X_t)\right) dt + \sigma \, dW_t(x), \quad x \in [0,1], t \geq 0, \\ X_t(0) = X_t(1) = 0, \\ X_0 = \xi \end{cases} \tag{1.14}$$

in the Hilbert space $H = L^2((0, 1))$ equipped with its standard inner product. We will assume that the nonlinearity F is of Nemytskii-type, i.e., we have $F(u) = f \circ u$ for some function $f : \mathbb{R} \to \mathbb{R}$. Since we are working with point evaluations $X_t(x)$, it will always be assumed that $X \in C(\mathbb{R}_+, E)$ with $E = C_0([0, 1])$ holds almost surely. Sufficient conditions for the latter are provided by Theorem 1.1.1. Indeed, by setting $\vartheta_1 = \vartheta_0 = 0$, all properties of the linear component of the solution presented in the previous section remain valid. In particular, hypotheses (L1)-(L2) and (N1) are fulfilled. Also, (N2) is satisfied (see, e.g., [55]) and, as pointed out in [26, Remark A.29], (N4) holds due to the maximum principle for parabolic equations. Thus, the almost sure continuity in time and space of the solution process can be achieved by requiring that $f \in C^1(\mathbb{R})$ satisfies condition (1.10) or, more generally, condition (1.8). For the sake of coherence, the discussion of more specific regularity assumptions required for our analysis is postponed to the beginning of Chapter 4.

1.2.3 Observation scheme

If not stated otherwise, all statistical considerations in this thesis will be based on the following set of space-time-discrete observations derived from a single sample path of the process X, which is either given by the SPDE (1.11) or (1.14): we suppose to have $(M + 1)(N + 1)$ time- and space-discrete observations

$$\{X_{t_i}(y_k), \, i = 0, \ldots, N, \, k = 0, \ldots, M\}$$

on a regular grid $\{(t_i, y_k)\}_{i,k} \subset [0, T] \times [0, 1]$ with a time horizon $T > 0$ and $M, N \in \mathbb{N}_0$. More precisely, we assume that

$$y_k = b + k\delta \quad \text{and} \quad t_i = i\Delta \qquad \text{where} \qquad \delta = \frac{1 - 2b}{M}, \quad \Delta = \frac{T}{N}$$

for some fixed $b \in [0, 1/2)$. Concerning the time horizon, it will always be assumed that either $T > 0$ is fixed or that $T \to \infty$.

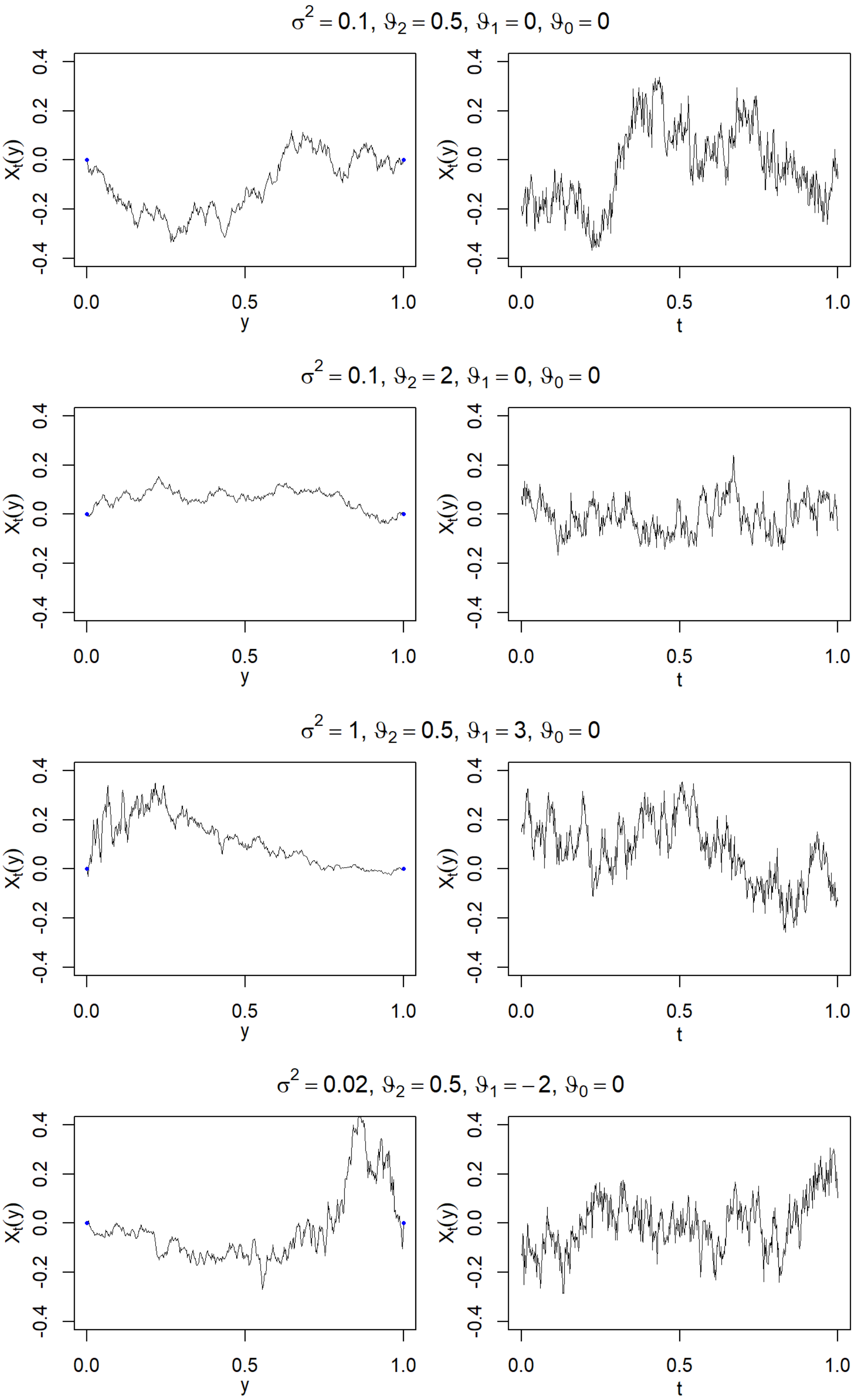


Figure 1.1: Exemplary sample paths of spatial (left) and temporal (right) processes for different parameter values. The blue points indicate the Dirichlet boundary conditions. The observation time for the spatial processes is $t = 0$, which is arbitrary, due to stationarity. The temporal processes are recorded at the spatial position $y = 0.4$. The sample paths are simulated on the space-time grid points $(i/500, k/500)$ with $i, k \in \{0, \dots, 500\}$ using the replacement method to be introduced in Chapter 3.

Due to the bounded space domain, we have high frequency observations in space whenever $M \to \infty$. In order to obtain high frequency observations in time, we will usually require that $T/N \to 0$. This is trivially satisfied if T is fixed and $N \to \infty$.

Note that the spatial locations y_k are equidistant inside a (possibly proper) sub-interval $[b, 1-b] \subset [0,1]$. For certain statistical procedures, we will exclude observations close to the boundary by requiring $b > 0$. This is done to prevent undesired boundary effects, which lead to biased estimates.

1.2.4 Proof of the Itô decomposition for the spatial process

Proof of Proposition 1.2.1. Due to (1.13) and the trigonometric identity

$$\sin(\alpha)\sin(\beta) = \frac{1}{2}\left(\cos(\alpha-\beta) - \cos(\alpha+\beta)\right), \tag{1.15}$$

we have

$$\begin{aligned}\operatorname{Cov}(X_t(x), X_t(y)) &= \sigma^2 \mathrm{e}^{-\frac{\kappa}{2}(x+y)} \sum_{\ell\geq 1} \frac{1}{2\lambda_\ell}(\cos(\pi\ell(y-x)) - \cos(\pi\ell(x+y))) \\ &= \frac{\sigma^2}{2\pi^2\vartheta_2}\mathrm{e}^{-\frac{\kappa}{2}(x+y)} \sum_{\ell\geq 1} \frac{1}{\ell^2+\Gamma/\pi^2}(\cos(\pi\ell(y-x)) - \cos(\pi\ell(x+y))).\end{aligned}$$

The claimed formulas now follow by inserting the closed expressions

$$\sum_{\ell\geq 1}\frac{1}{\ell^2+\beta}\cos(\pi\ell x) = \begin{cases} -\frac{\pi\cos(\pi\sqrt{|\beta|}(x-1))}{2\sqrt{|\beta|}\sin(\pi\sqrt{|\beta|})} + \frac{1}{2|\beta|}, & -1<\beta<0 \\ \frac{\pi^2(x-1)^2}{4} - \frac{\pi^2}{12}, & \beta = 0 \\ \frac{\pi\cosh(\pi\sqrt{\beta}(x-1))}{2\sqrt{\beta}\sinh(\pi\sqrt{\beta})} - \frac{1}{2\beta}, & \beta>0 \end{cases} \tag{1.16}$$

for $x \in [0,1]$ and again applying (1.15) as well as

$$\sinh(\alpha)\sinh(\beta) = \frac{1}{2}(\cosh(\alpha+\beta) - \cosh(\alpha-\beta)),$$

respectively. To prove the second statement, we use the ansatz $Z(x) = u(x)B(v(x))$, u, v positive and v non-decreasing, which is the general form of a Gaussian Markov process, cf. Neveu [64]. Comparison of covariance functions easily yields

$$Z(x) \overset{\mathcal{D}}{=} \sqrt{\frac{\sigma^2}{2\vartheta_2}}\mathrm{e}^{-\kappa x/2}\cdot\begin{cases}\sqrt{\frac{1}{\Gamma_0\sin\Gamma_0}}\sin\left(\Gamma_0(1-x)\right)B\left(\frac{\sin(\Gamma_0 x)}{\sin(\Gamma_0(1-x))}\right), & \Gamma<0, \\ (1-x)B\left(\frac{x}{(1-x)}\right), & \Gamma=0, \\ \sqrt{\frac{1}{\Gamma_0\sinh\Gamma_0}}\sinh\left(\Gamma_0(1-x)\right)B\left(\frac{\sinh(\Gamma_0 x)}{\sinh(\Gamma_0(1-x))}\right), & \Gamma>0.\end{cases}$$

A direct calculation shows $v' > 0$, so that v is indeed non-decreasing. Further, since $v(0) = 0$, we have $B(v(x)) \overset{\mathcal{D}}{=} \int_0^x \sqrt{v'(z)}\,dB(z)$ and, therefore, one passes to

$$Z(x) \overset{\mathcal{D}}{=} \sqrt{\frac{\sigma^2}{2\vartheta_2}}\mathrm{e}^{-\kappa x/2}\cdot\begin{cases}\sin\left(\Gamma_0(1-x)\right)\int_0^x \frac{dB(z)}{\sin(\Gamma_0(1-z))}, & \Gamma<0, \\ (1-x)\int_0^x\frac{dB(z)}{1-z}, & \Gamma=0, \\ \sinh\left(\Gamma_0(1-x)\right)\int_0^x\frac{dB(z)}{\sinh(\Gamma_0(1-z))}, & \Gamma>0.\end{cases}$$

The claimed representation now follows by applying the product rule for Itô processes. □

Chapter 2

Parametric estimation for the linear equation

This chapter discusses estimation of the parameters $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ of the linear SPDE model (1.11) based on the fully discrete observation scheme defined in Section 1.2.3. The results concerning a finite time horizon are part of Hildebrandt and Trabs [38].

Before deriving concrete parameter estimators, in Section 2.1 we answer the structural question if and how fast the parameters can be (jointly) estimated depending on both the sampling frequencies in time and space and on the time horizon T. In Section 2.2 we prove central limit theorems for realized quadratic variations based on the space and double increments from (4) and (5), respectively. These results are then used to construct method of moments estimators for the parameters. In particular, the convergence rate of our double increments based estimator for all identifiable parameters (almost) matches the lower bound derived in Section 2.1. Finally, Section 2.3 briefly discusses the possibility of constructing confidence sets for the parameters based on the asymptotic normality of the estimators. All proofs are collected in Section 2.4.

Throughout, $X = (X_t(x), t \in \mathbb{R}_+, x \in [0,1])$ denotes the solution field given by (1.12). For the results in the lower bounds section, we assume that X_0 follows the stationary distribution, the remaining results also allow for the case $X_0 = 0$.

2.1 Identifiability of parameters and lower bounds

Before discussing lower bounds on the rate of convergence of estimators for the model parameters, we derive a qualitative result concerning their identifiability on a finite time horizon. Here, we call a parameter identifiable if it can be estimated consistently from the data. Whether or not a parameter is identifiable can be assessed by studying absolute continuity properties of the solution process for different values of the parameters: let G be some parameter space and consider the situation where a sequence of statistical experiments $(P_\gamma^n)_{\gamma \in G}$ is induced by a sequence of random variables T_n on a probability space $(\Omega, \mathcal{F}, (P_\gamma)_{\gamma \in G})$, i.e., $P_\gamma^n = P_\gamma \circ T_n^{-1}$. Then, if P_{γ_1} is absolutely continuous with respect to P_{γ_2} for two different values $\gamma_1, \gamma_2 \in G$, the parameter γ is not identifiable. Indeed, if there was an estimator $\hat{\gamma}_n$ such that $\hat{\gamma}_n \to \gamma_2$ holds in P_{γ_2}-probability, then we have $\hat{\gamma}_{n_k} \to \gamma_2$ P_{γ_2}-almost surely along a subsequence (n_k). Due to the absolute continuity property, this implies $\hat{\gamma}_{n_k} \to \gamma_2$ P_{γ_1}-almost surely, which contradicts the assumption that $\hat{\gamma}_n$ is a consistent estimator.

In order to study absolute continuity properties of the process X for different parameter values

(σ^2, ϑ), we introduce the notations

$$\begin{aligned}(X_t(\cdot), t \in [0,T]) &\sim P_{(\sigma^2,\vartheta)} \text{ on } C([0,T], L^2((0,1))),\\ (X_{t_0}(x),\, x \in [0,1]) &\sim P^{(t_0,\cdot)}_{(\sigma^2,\vartheta)} \text{ on } L^2((0,1)),\\ (X_t(x_0),\, t \in [0,T]) &\sim P^{(\cdot,x_0)}_{(\sigma^2,\vartheta)} \text{ on } L^2([0,T])\end{aligned}$$

for fixed values $t_0 \geq 0$, $x_0 \in (0,1)$ and a finite time horizon $T > 0$. Further, for probability measures Q and P we write $Q \sim P$ if they are equivalent, i.e., mutually absolutely continuous.

Proposition 2.1.1. *Assume (ST) and consider a finite time horizon $T > 0$. Further, let $t_0 \geq 0$, $x_0 \in (0,1)$ be fixed. For any two sets of parameters (σ^2, ϑ), $(\tilde{\sigma}^2, \tilde{\vartheta}) \in \Theta$ we have*

(i) $P_{(\sigma^2,\vartheta)} \sim P_{(\tilde{\sigma}^2,\tilde{\vartheta})}$ *if and only if* $(\sigma^2, \vartheta_2, \vartheta_1) = (\tilde{\sigma}^2, \tilde{\vartheta}_2, \tilde{\vartheta}_1)$,

(ii) $P^{(t_0,\cdot)}_{(\sigma^2,\vartheta)} \sim P^{(t_0,\cdot)}_{(\tilde{\sigma}^2,\tilde{\vartheta})}$ *if and only if* $\left(\frac{\sigma^2}{\vartheta_2}, \kappa\right) = \left(\frac{\tilde{\sigma}^2}{\tilde{\vartheta}_2}, \tilde{\kappa}\right)$,

(iii) $P^{(\cdot,x_0)}_{(\sigma^2,\vartheta)} \sim P^{(\cdot,x_0)}_{(\tilde{\sigma}^2,\tilde{\vartheta})}$ *if and only if* $\frac{\sigma^2}{\sqrt{\vartheta_2}} \mathrm{e}^{-\kappa x_0} = \frac{\tilde{\sigma}^2}{\sqrt{\tilde{\vartheta}_2}} \mathrm{e}^{-\tilde{\kappa} x_0}$,

where $\kappa = \vartheta_1/\vartheta_2$, $\tilde{\kappa} = \tilde{\vartheta}_1/\tilde{\vartheta}_2$.

Assuming that none of the four parameters is known, the implications of Proposition 2.1.1 for discrete or even continuous observations are the following: (i) shows that it is impossible to estimate ϑ_0 consistently on a finite time horizon. (ii) shows that, based on a single temporal observation, it is impossible to estimate other parameters than $\left(\sigma^2/\vartheta_2, \kappa\right)$. (iii) reveals that, based on a single spatial observation on a finite time horizon, it is impossible to estimate any other parameter than $\frac{\sigma^2}{\sqrt{\vartheta_2}}\mathrm{e}^{-\kappa x_0}$. In fact, these restrictions on the identifiability are sharp: $\frac{\sigma^2}{\sqrt{\vartheta_2}}\mathrm{e}^{-\kappa x_0}$ can be estimated using squared time increments of the process $[0,T] \ni t \mapsto X_t(x_0)$, as in Bibinger and Trabs [9, Theorem 4.2]. $\left(\sigma^2/\vartheta_2, \kappa\right)$ can be estimated by computing the quadratic variation of the Itô process $Z = X_{t_0}(\cdot)$ from Proposition 1.2.1 on two different sub-intervals of $[0,1]$. By combining the two methods, all three parameters $(\sigma^2, \vartheta_2, \vartheta_1)$ can be obtained based on $(X_t(x),\, x \in [0,1], t \in [0,T])$.

2.1.1 Lower bound for the case of a fixed time horizon

Without loss of generality, we consider the finite time horizon $T = 1$. Due to Proposition 2.1.1, it is impossible to estimate ϑ_0 consistently on a finite time horizon and, consequently, deriving a lower bound for the whole parameter vector (σ^2, ϑ) is a trivial task. In the following, we are going to derive a lower bound for the remaining three parameters. For this purpose, it suffices to consider the sub-problem where $\vartheta_1 = \vartheta_0 = 0$ and only (σ^2, ϑ_2) has to be estimated.

Theorem 2.1.2. *Let $\vartheta_1 = \vartheta_0 = 0$ and assume (ST). Further, consider the observation scheme defined in Section 1.2.3 with $T = 1$ and $b \in [0, 1/2) \cap \mathbb{Q}$. Then:*

(i) If $\min(M,N)$ remains finite, then there is no consistent estimator for (σ^2, ϑ_2).

(ii) For any open set $H \subset (0,\infty)^2$ there is a constant $c > 0$, such that

$$\liminf_{M,N\to\infty} \inf_{\mathscr{T}} \sup_{(\sigma^2,\vartheta_2)\in H} \mathbf{P}_{(\sigma^2,\vartheta_2)}\left(\left\| \mathscr{T} - \begin{pmatrix}\sigma^2\\ \vartheta_2\end{pmatrix} \right\| > \frac{c}{\sqrt{r_{M,N}}}\right) > 0,$$

$$\text{where } r_{M,N} := \begin{cases} N^{3/2}, & \frac{M}{\sqrt{N}} \gtrsim 1,\\ M^3 \log \frac{N}{M^2}, & \frac{M}{\sqrt{N}} \to 0 \end{cases}$$

and $\inf_{\mathscr{T}}$ *is taken over all estimators* $\mathscr{T}$ *of* (σ^2, ϑ_2) *based on the observations* $\{X_{t_{i+1}}(y_k) - X_{t_i}(y_k),\ i < N,\ k \le M\}$.

Remark 2.1.3. Assertion (i) and the lower bound for the case $M/\sqrt{N} \gtrsim 1$ are also valid for estimators based on $\{X_{t_i}(y_j),\ i \le N, k \le M\}$ instead of the increments. We conjecture that this is also true for the case $M/\sqrt{N} \to 0$, see also the discussion following Proposition 2.1.6. Furthermore, we believe that the logarithmic factor appearing in the lower bound results from technical issues in our proves and could, possibly, be removed by a tighter analysis.

The above theorem shows that, in general, (σ^2, ϑ_2) cannot be estimated with the parametric rate $1/\sqrt{MN}$, in contrast to a conjecture in Cialenco and Huang [19]. Instead, we observe a phase transition with respect to the rate, depending on the sampling frequencies. The parametric rate can only be attained when $N \approx M^2$. In fact, when dealing with a large time horizon, it will be clear that this condition is more generally specified in terms of the observation intervals Δ and δ, namely $\delta \approx \sqrt{\Delta}$. In this regime, our estimators will show that it is indeed possible to identify (σ^2, ϑ_2) at the parametric rate. In that sense, when $\delta \approx \sqrt{\Delta}$, both spatial and temporal observations contain the optimal amount of information on the parameters and we call it the *balanced* sampling design.

In order to prove the lower bounds in Theorem 2.1.2, we proceed in the following way: For each sampling regime we choose a reparametrization (γ_1, γ_2) of (σ^2, ϑ_2) in such a way that γ_1 can be estimated with parametric rate, even without knowledge of γ_2. We then derive a lower bound for the simpler problem of estimating the one dimensional parameter γ_2, allowing that γ_1 is known. Clearly, the resulting lower bound for γ_2 carries over to (γ_1, γ_2) and consequently to (σ^2, ϑ_2).

The lower bound for γ_2, in turn, is obtained by the standard technique, see e.g. Tsybakov [75]. Indeed, denoting the law of the discrete observations by $\mathbf{P}^{N,M}_{\gamma_2}$, it follows from Theorem 2.2 in the same reference that if $\gamma_2 \in H_2$ for some open set H_2, then

$$\inf_{\hat{\gamma}_2} \sup_{\gamma_2 \in H_2} \mathbf{P}_{\gamma_2}\big(|\hat{\gamma}_2 - \gamma_2| \ge s\big) \ge \frac{1 - \mathcal{H}(\mathbf{P}^{N,M}_{\gamma_2'}, \mathbf{P}^{N,M}_{\gamma_2''})}{2} \tag{2.1}$$

holds for all $\gamma_2', \gamma_2'' \in H_2$ with $|\gamma_2' - \gamma_2''| \ge 2s$ and where $\mathcal{H}(\cdot,\cdot)$ is the Hellinger distance, as defined in (8). Further, using an inequality by Ibragimov and Has'minskii [43, Theorem I.7.6], the Hellinger distance can be bounded in terms of the corresponding Fisher information $J(\gamma_2)$: Let $p(\cdot, \gamma_2)$ be the Lebesgue density of $\mathbf{P}^{N,M}_{\gamma_2}$ and $g = \sqrt{p}$. Then, Jensen's inequality yields

$$\begin{aligned}
\mathcal{H}^2(\mathbf{P}^{N,M}_{\gamma_2'}, \mathbf{P}^{N,M}_{\gamma_2''}) &= \int (g(x,\gamma_2') - g(x,\gamma_2''))^2\,dx \le (\gamma_2' - \gamma_2'')^2 \int\int_0^1 \Big(\frac{\partial g}{\partial \gamma_2}(x, \bar{\gamma}_2(s))\Big)^2\,ds\,dx \\
&= \frac{(\gamma_2' - \gamma_2'')^2}{4} \int_0^1 \int \Big(\frac{\partial}{\partial \gamma_2} \log p(x, \bar{\gamma}_2(s))\Big)^2 \mathbf{P}_{\bar{\gamma}_2(s)}(dx)\,ds \\
&= \frac{(\gamma_2' - \gamma_2'')^2}{4} \int_0^1 J(\bar{\gamma}_2(s))\,ds
\end{aligned}$$

with $\bar{\gamma}_2(s) = \gamma_2' + s(\gamma_2'' - \gamma_2')$. In combination with (2.1) it follows that, in order to prove Theorem 2.1.2, it suffices to show $J(\gamma_2) \lesssim r_{N,M}$ locally uniformly. The main effort, noting that the observations are significantly correlated, is now to derive sharp upper bounds for the Fisher information in the different sampling regimes.

In the case $M/\sqrt{N} \gtrsim 1$, we apply the following bound on the Fisher information for discrete observations of the first M coefficient processes. Thanks to the Markov property, the probability density function for discrete observations of an Ornstein-Uhlenbeck process is provided by the transition density and allows for explicit computations.

Proposition 2.1.4. *Let $\vartheta_1 = \vartheta_0 = 0$ and consider a sample $(u_\ell(i/N),\, \ell \le M,\, i \le N)$, where $(u_\ell)_{\ell\in\mathbb{N}}$ are independent Ornstein-Uhlenbeck processes given by*

$$du_\ell(t) = -\lambda_\ell u_\ell(t)\,dt + \sigma\,d\beta_\ell(t), \quad u_\ell(0) \sim \mathcal{N}\left(0,\, \frac{\sigma^2}{2\lambda_\ell}\right).$$

Consider the reparametrization (σ^2, ρ^2) where $\rho^2 = \sigma^2/\vartheta_2$ and the corresponding Fisher information $J_{N,M} \in \mathbb{R}^{2\times 2}$. For $\max(M,N) \to \infty$, the diagonal entries of $J_{N,M}$ satisfy

$$J_{N,M}(\sigma^2) = \mathcal{O}(N^{3/2} \wedge (MN)) \qquad \text{and} \qquad J_{N,M}(\rho^2) = \mathcal{O}(M^3 \wedge (MN)). \tag{2.2}$$

In particular, $\min\big(J_{N,M}(\sigma^2), J_{N,M}(\rho^2)\big) \lesssim N^{3/2} \wedge M^3$ *for* $\max(N,M) \to \infty$.

Remark 2.1.5.

1. If $M \lesssim \sqrt{N}$ and σ^2 is known, Proposition 2.1.4 suggests a lower bound of $M^{-3/2}$ for estimation of ϑ_2 in the spectral approach. Indeed, this rate is achieved by the maximum likelihood estimator for time-continuous observations of the coefficient processes, cf. Lototsky [53].

2. The reparametrization was chosen since σ^2 can be computed from the quadratic variation of any coefficient process u_ℓ when $N \to \infty$, while ρ^2 can be computed from the empirical variance of $(\sqrt{2}\pi\ell)u_\ell(t_i)$, $\ell \le M$, for a fixed t_i as $M \to \infty$, even without knowledge of the other parameter, respectively.

Letting $M \to \infty$, Proposition 2.1.4 suggests that, based on observations of the coefficient processes, it is not possible to estimate σ^2 (and in particular (σ^2, ϑ_2)) at a rate faster than $N^{-3/4}$. Further, assuming $\vartheta_1 = 0$, the eigenfunctions $e_\ell(\cdot)$ do not depend on unknown parameters and hence, the space-time-discrete observations of the SPDE may be reconstructed from $\{u_\ell(t_i),\, i \le N, \ell \in \mathbb{N}\}$. Consequently, the lower bound $N^{-3/4}$ carries over to discrete observations of the SPDE.

Although the lower bounds resulting from Proposition 2.1.4 and Theorem 2.1.2 are almost the same, their proofs require a very different reasoning if $M/\sqrt{N} \to 0$: in this case, if σ^2 is known, it follows from the results in Bibinger and Trabs [9] that ϑ_2 can be estimated with parametric rate of convergence based on discrete observations of the SPDE, see also Proposition 2.2.11 in this thesis. On the other hand, Proposition 2.1.4 suggests that $\vartheta_2 = \sigma^2/\rho^2$ cannot be estimated at a faster rate than $M^{-3/2}$ based on the coefficient processes. In particular, both observation schemes are not asymptotically equivalent in the sense of Le Cam.

To derive the lower bound in the case $M/\sqrt{N} \to 0$, we consider the situation where observations are recorded at rational positions $y_k = \frac{k}{M}$, $k = 1, \ldots, M-1$, where we work with $M-1$ instead of M spatial observations to ease the notation. Thus, we potentially add spatial observations on the margin $[0,b) \cup (1-b,1]$, which can only increase the amount of information contained in the data. With this type of observation scheme, a fact that will prove to be useful in several places of this thesis is that there is also a discrete version of the orthogonality property for the eigenfunctions: it follows from basic trigonometric identities that

$$\langle e_\eta, e_\nu\rangle_M = \delta_{\eta\nu}, \qquad 1 \le \eta, \nu \le M-1,$$

with

$$\langle u, v\rangle_M := \frac{1}{M}\sum_{k=1}^{M-1} u\Big(\frac{k}{M}\Big)v\Big(\frac{k}{M}\Big)\mathrm{e}^{\kappa\frac{k}{M}},$$

see, e.g., the proof of Theorem 2.1 in Rohde [70]. By periodicity of the sine function, we further have $\bar{e}_M = 0$, $\bar{e}_{\eta+2\ell M} = \bar{e}_\eta$ and $\bar{e}_{2M-\eta+2\ell M} = -\bar{e}_\eta$ for $\eta, \nu \le M-1$ and $\bar{e}_\ell := (e_\ell(\frac{1}{M}), \ldots, e_\ell(\frac{M-1}{M}))$. Thus, for any $\eta, \nu \in \mathbb{N}$ we have

$$\langle e_\eta, e_\nu\rangle_M = \begin{cases} 1, & \text{if } \eta,\nu \in \mathcal{I}_k^+ \text{ or } \eta,\nu \in \mathcal{I}_k^- \text{ for some } k \le M-1, \\ -1, & \text{if } \eta \in \mathcal{I}_k^+,\, \nu \in \mathcal{I}_k^- \text{ or } \eta \in \mathcal{I}_k^-,\, \nu \in \mathcal{I}_k^+ \text{ for some } k \le M-1, \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

with

$$\mathcal{I}_k^+ := \{k + 2\ell M,\, \ell \in \mathbb{N}_0\}, \quad \mathcal{I}_k^- := \{2M - k + 2\ell M,\, \ell \in \mathbb{N}_0\}, \qquad k \leq M-1.$$

In particular, it follows that observing $\{X_{t_i}(y_k),\, i \leq N, k \leq M-1\}$ is equivalent to observing

$$\{U_k(t_i),\, k \leq M-1, i \leq N\}, \qquad U_k(t) := \langle X_t(\cdot), e_k \rangle_M = \sum_{\ell \in \mathcal{I}_k^+} u_\ell(t) - \sum_{\ell \in \mathcal{I}_k^-} u_\ell(t). \tag{2.4}$$

Since the sets $\mathcal{I}_k := \mathcal{I}_k^+ \cup \mathcal{I}_k^-$ are disjoint for different values of k, the processes $\{U_1, \ldots, U_{M-1}\}$ are independent which simplifies the calculation of the Fisher information considerably. Based on their spectral densities and Whittle's formula (2.25) for the asymptotic Fisher information of a stationary Gaussian time series, we obtain the following result for the increment processes $\bar{U}_k$, $k \leq M-1$, defined by

$$\bar{U}_k(j) := U_k(t_{j+1}) - U_k(t_j), \qquad j = 0, \ldots, N-1. \tag{2.5}$$

Proposition 2.1.6. *Let $\vartheta_1 = \vartheta_0 = 0$ and assume (ST). Consider the parametrization $(\sigma_0^2, \vartheta_2)$ where $\sigma_0^2 := \sigma^2/\sqrt{\vartheta_2}$. If $M/\sqrt{N} \to 0$, the Fisher information $J_{M,N}$ with respect to ϑ_2 of a sample $\{\bar{U}_k(j),\, j \leq N-1, k \leq M-1\}$ satisfies*

$$J_{M,N}(\vartheta_2) = \mathcal{O}\Big(M^3 \log \frac{N}{M^2}\Big).$$

The parametrization was chosen as it allows for estimation of $\sigma_0^2 = \sigma^2/\sqrt{\vartheta_2}$ with parameteric rate based on time increments in the regime $M/\sqrt{N} \to 0$, even when ϑ_2 is unknown. Again, we refer to [9] or Proposition 2.2.11 of this thesis. We have considered $\bar{U}_k$ instead of U_k due to the technical reason that the N-th order Fourier approximation of the spectral density of the increment process is positive and, hence, a spectral density as well. We conjecture that the same bound holds for the Fisher information of U_k.

2.1.2 Lower bound for the case $T \to \infty$

Assuming $T \to \infty$, the derivation of a lower bound for estimating (σ^2, ϑ_2) in the relevant situation $\vartheta_1 = \vartheta_0 = 0$ can be done in exactly the same way as for a fixed time horizon. Here, one obtains the lower bound $r_{\delta,\Delta,T}^{-1/2}$ with

$$r_{\delta,\Delta,T} := \begin{cases} \dfrac{T}{\Delta^{3/2}}, & \dfrac{\sqrt{\Delta}}{\delta} \gtrsim 1, \\ \dfrac{T}{\delta^3} \cdot \log \dfrac{\delta^2}{\sqrt{\Delta}}, & \dfrac{\sqrt{\Delta}}{\delta} \to 0. \end{cases} \tag{2.6}$$

In the case $T \equiv 1$, we have $\delta \approx \frac{1}{M}$ and $\Delta \approx \frac{1}{N}$ and, thus, we recover the lower bound from Theorem 2.1.2. Further, in case of a balanced sampling design $\sqrt{\Delta}/\delta \approx 1$, we have $r_{\delta,\Delta,T} = \frac{N}{\sqrt{\Delta}} \approx \frac{N}{\delta} \approx NM$, whereas for an unbalanced sampling design no parametric rate of convergence can be reached by any estimator. These findings are in line with the results for a fixed time horizon.

When considering the problem of estimating the whole parameter vector (σ^2, ϑ), this lower bound is no longer tight. In fact, even for the sub-problem of estimating ϑ_0 when the other parameters are known, we can deduce the lower bound $T^{-1/2}$, which is certainly larger than $r_{\delta,\Delta,T}^{-1/2}$. Similarly to Proposition 2.1.4, this can be shown by bounding the Fisher information for ϑ_0 of a sample of the coefficient processes $(u_\ell(t_i),\, i \leq N, \ell \leq L)$ uniformly in $L \in \mathbb{N}$. The following theorem summarizes our findings.

Theorem 2.1.7. *Assume (ST) and consider the observation scheme defined in Section 1.2.3 with $b \in [0, 1/2) \cap \mathbb{Q}$ and $T \to \infty$.*

(i) *Let $\vartheta_1 = \vartheta_0 = 0$. For any open set $H \subset (0,\infty)^2$ there is a constant $c > 0$, such that*

$$\liminf_{M,N,T\to\infty} \inf_{\mathscr{T}} \sup_{(\sigma^2,\vartheta_2)\in H} \mathbf{P}_{(\sigma^2,\vartheta_2)}\Big(\Big\|\mathscr{T} - \binom{\sigma^2}{\vartheta_2}\Big\| > \frac{c}{\sqrt{r_{\delta,\Delta,T}}}\Big) > 0,$$

with $r_{\delta,\Delta,T}$ defined in (2.6) *and where $\inf_{\mathscr{T}}$ is taken over all estimators $\mathscr{T}$ of (σ^2, ϑ_2) based on the observations $\{X_{t_{i+1}}(y_k) - X_{t_i}(y_k),\ i < N,\ k \leq M\}$.*

(ii) *For any open set $H \subset \mathbb{R}$ there is a constant $c > 0$, such that*

$$\liminf_{M,N,T\to\infty} \inf_{\mathscr{T}} \sup_{\vartheta_0\in H} \mathbf{P}_{\vartheta_0}\Big(\big|\mathscr{T} - \vartheta_0\big| > \frac{c}{\sqrt{T}}\Big) > 0,$$

where $\inf_{\mathscr{T}}$ is taken over all estimators of ϑ_0.

2.2 Method of moments estimators for the parameters

In this section we construct method of moments estimators for the parameters and prove corresponding central limit theorems. To that aim, we first study central limit theorems for realized quadratic variations based on the space and double increments from (4) and (5), respectively.

2.2.1 Central limit theorems for realized quadratic variations

The realized quadratic variations of X can be regarded as sums of squares of certain Gaussian random vectors. Hence, our central limit theorems embed into the literature on quadratic forms in random variables and their asymptotic properties, see e.g. [62]. Our key tool for proving asymptotic normality is the following proposition which is tailor made for the situation present in this thesis and which gives an explicit covariance condition that ensures convergence to the normal distribution. Recall the matrix norms from (6).

Proposition 2.2.1. *Let $(Z_{i,n},\ 1 \leq i \leq d_n,\ n \in \mathbb{N})$ be a triangular array such that $(Z_{1,n} \dots, Z_{d_n,n}) \sim \mathcal{N}(0, \Sigma_n)$ for a covariance matrix $\Sigma_n \in \mathbb{R}^{d_n \times d_n}$, $n \in \mathbb{N}$, and let $(\alpha_{i,n},\ 1 \leq i \leq d_n,\ n \in \mathbb{N})$ be a deterministic triangular array with values in $\{-1, 1\}$. Define $S_n := \sum_{i=1}^{d_n} \alpha_{i,n} Z_{i,n}^2$ for $n \geq 1$. If $\|\Sigma_n\|_2^2/\mathrm{Var}(S_n) \to 0$ as $n \to \infty$, then we have*

$$\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathrm{Var}\, S_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \textit{for} \quad n \to \infty.$$

The proof relies on the fact that S_n can be represented as a linear combination of independent $\chi^2(1)$-distributed random variables. $\|\Sigma_n\|_2^2/\mathrm{Var}(S_n) \to 0$ then implies that the corresponding Lyapunov condition is fulfilled. In this section we only require $\alpha_{i,n} = 1$ for all i and n, i.e., $S_n = \|Z_{\bullet,n}\|_2^2$. The general case will be necessary to verify asymptotic normality of the M-estimators in Section 2.2.2. It is worth noting that Proposition 2.2.1 reveals a quite elementary proof strategy to verify several central limit theorems in [9, 19, 71, 74] instead of advanced techniques from Malliavin calculus or mixing theory.

Remark 2.2.2.

1. As an application of Isserlis' theorem [44], for a 2-dimensional centered Gaussian vector (Y_1, Y_2), one obtains the formula

$$\mathrm{Cov}(Y_1^2, Y_2^2) = 2\,\mathrm{Cov}(Y_1, Y_2)^2. \tag{2.7}$$

If $(\alpha_{i,n}) \equiv 1$, it follows that $\operatorname{Var}(S_n) = 2\|\Sigma_n\|_F^2$ and, hence, the condition for asymptotic normality may be written as $\|\Sigma_n\|_2/\|\Sigma_n\|_F \to 0$. This condition is essentially optimal: in case of independent observations it is, in fact, equivalent to asymptotic negligibility of the individual normalized and centered summands and, hence, equivalent to Lindeberg's condition.

2. The spectral norm is bounded by the maximum absolute row sum. Writing $\Sigma_n = \big(\sigma_{ij}^{(n)}\big)_{i,j}$, asymptotic normality, thus, holds under the sufficient condition

$$\frac{\Big(\max_{i\le d_n}\sum_{j=1}^{d_n}\Big|\sigma_{ij}^{(n)}\Big|\Big)^2}{\operatorname{Var} S_n} \longrightarrow 0, \quad n\to\infty. \tag{2.8}$$

So far, the double asymptotic regime $M, N \to \infty$ has only been studied for time increments $(\Delta_i^N X)(y_k) = X_{t_{i+1}}(y_k) - X_{t_i}(y_k)$: if $b > 0$, T is fixed and if there exists $\rho \in (0, 1/2)$ such that $M = \mathcal{O}(\Delta^{-\rho})$, then the *rescaled realized temporal quadratic variation*

$$V_{\mathrm{t}} := \frac{1}{MN\sqrt{\Delta}}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1} \mathrm{e}^{\kappa y_k}(\Delta_i^N X)^2(y_k) \tag{2.9}$$

satisfies

$$\sqrt{MN}\left(V_{\mathrm{t}} - \frac{\sigma^2}{\sqrt{\pi\vartheta_2}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{B\sigma^4}{\pi\vartheta_2}\right), \quad N, M\to\infty, \tag{2.10}$$

where

$$B = 2 + \sum_{J=1}^{\infty}\Big(2\sqrt{J} - \sqrt{J+1} - \sqrt{J-1}\Big)^2, \tag{2.11}$$

cf. Bibinger and Trabs [9, Theorem. 3.4]. The central limit theorem (2.10) was later shown to remain valid in case of a growing time horizon, $T \to \infty$, provided that $T\Delta \to 0$, see Kaino and Uchida [49, Theorem 3]. Note that the central limit theorem is only valid under the condition (roughly) $M = o(\Delta^{-1/2})$, i.e., if the observation frequency in time is much higher than in space. This constraint is due to a non-negligible correlation of realized temporal quadratic variations at two neighboring points in space if the distance δ of these points is small compared to Δ. For fixed T, the condition translates to the requirement that there may only be few spatial compared to temporal observations.

In the situation where the spatial observation frequency dominates the temporal observations frequency the above result is not applicable. In this case, spatial increments $(\delta_k^M X)(t_i) = X_{t_i}(y_{k+1}) - X_{t_i}(y_k)$ and the corresponding rescaled realized spatial quadratic variations

$$V_{\mathrm{sp}}(t_i) := \frac{1}{M\delta}\sum_{k=0}^{M-1} \mathrm{e}^{\kappa y_k}(\delta_k^M X)^2(t_i)$$

at time t_i turn out to be useful. In contrast to squared time increments, which have to be renormalized by $\sqrt{\Delta}$ due to the roughness of $t \mapsto X_t(y)$, squared space increments have to be renormalized by δ, which is an obvious consequence of the fact that the process $y \mapsto X_t(y)$ is a semimartingale under (ST), cf. Proposition 1.2.1.

In the extreme case where observations are only available at one point $t > 0$ in time (and assuming $\vartheta_1 = \vartheta_0 = 0$ as well as $X_0 = 0$) Cialenco and Huang [19] showed that $V_{\mathrm{sp}}(t)$ is asymptotically normal with $1/\sqrt{M}$-rate of convergence. An analogous result has been proved by Shevchenko et al. [71] for the wave equation. If X_0 follows the stationary distribution, Proposition 1.2.1 reveals that $V_{\mathrm{sp}}(t)$ is in fact a rescaled realized quadratic variation of the Itô diffusion $y \mapsto X_t(y)$. Hence,

$$\sqrt{M}\left(V_{\mathrm{sp}}(t) - \frac{\sigma^2}{2\vartheta_2}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^4}{2\vartheta_2^2}\right), \qquad M\to\infty,$$

follows from standard theory on quadratic variation for semimartingales, see, e.g., [6, 45]. In order to generalize this central limit theorem to the double asymptotic regime $M, N \to \infty$, we define the time average of the *rescaled realized spatial quadratic variations*:

$$V_{\mathrm{sp}} := \frac{1}{N}\sum_{i=0}^{N-1} V_{\mathrm{sp}}(t_i) = \frac{1}{NM\delta}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1} \mathrm{e}^{\kappa y_k}(\delta_k^M X)^2(t_i). \tag{2.12}$$

In fact, in the case $X_0 = 0$, we have $V_{\mathrm{sp}}(t_0) = 0$ and, thus, we redefine V_{sp} by summing over $i \in \{1, \ldots, N\}$ instead of $i \in \{0, \ldots, N-1\}$.

Theorem 2.2.3. *Let $b \in [0, 1/2)$. If $N/M \to 0$, then*

$$\sqrt{MN}\left(V_{\mathrm{sp}} - \frac{\sigma^2}{2\vartheta_2}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^4}{2\vartheta_2^2}\right), \quad M, N \to \infty.$$

Remark 2.2.4. The condition $N/M \to 0$ is necessary in order to neglect the bias: the proof of the theorem reveals that $\delta^{-1}\mathbf{E}\left(\mathrm{e}^{-\kappa y_k}(\delta_k^M X)^2(t_i)\right) - \frac{\sigma^2}{2\vartheta_2} \approx \delta$ and, consequently, the overall bias is of the order

$$\mathbf{E}\left(\sqrt{MN}\left(V_{\mathrm{sp}} - \frac{\sigma^2}{2\vartheta_2}\right)\right) \approx \sqrt{MN}\cdot\delta \approx \sqrt{\frac{N}{M}}.$$

For fixed T, we conclude that the central limit theorem for realized temporal quadratic variations V_{t} holds when (roughly) $M = o(\sqrt{N})$, whereas the central limit theorem for realized spatial quadratic variations V_{sp} is fulfilled if $N = o(M)$. To close the remaining gap, we finally study the space-time increments D_{ik} from (5). The corresponding rescaled realized quadratic variations are robust with respect to the sampling regime, as indicated by the representation

$$D_{ik} = \sum_{\ell \geq 1}\big(u_\ell(t_{i+1}) - u_\ell(t_i)\big)\big(e_\ell(y_{k+1}) - e_\ell(y_k)\big)$$

in terms of the series expansion (1.12).

In contrast to the case of space increments (and in line with the result for time increments), we impose $b > 0$ for the remainder of this section. Inspection of the proofs suggests that this condition may be relaxed to $b \to 0$ as long as the decay is sufficiently slow. As a first step, we calculate the asymptotic expectation of the double increments. In doing so, we restrict ourselves to the stationary case, the case $X_0 = 0$ will later be dealt with by means of an approximation argument.

Proposition 2.2.5. *Assume (ST) and let $b \in (0, 1/2)$. Then:*

(i) It holds uniformly in $0 \leq k \leq M-1$ and $1 \leq i \leq N-1$ that

$$\mathbf{E}\left(D_{ik}^2\right) = \sigma^2\mathrm{e}^{-\kappa y_k}\,\Phi_\vartheta(\delta, \Delta) + \mathcal{O}\left(\delta\sqrt{\Delta}\left(\delta \wedge \sqrt{\Delta}\right)\right), \quad \max(\delta, \Delta) \to 0,$$

where

$$\Phi_\vartheta(\delta, \Delta) := F_{\vartheta_2}(0, \Delta)\left(1 + \mathrm{e}^{-\kappa\delta}\right) - 2F_{\vartheta_2}(\delta, \Delta)\mathrm{e}^{-\kappa\delta/2}$$

and

$$F_{\vartheta_2}(\delta, \Delta) := \sum_{\ell \geq 1}\frac{1 - \mathrm{e}^{-\pi^2\vartheta_2\ell^2\Delta}}{\pi^2\vartheta_2\ell^2}\cos(\pi\ell\delta).$$

(ii) Assuming that $r = \lim \delta/\sqrt{\Delta} \in [0, \infty]$ exists, Φ_ϑ admits three different asymptotic regimes:

$$\Phi_\vartheta(\delta, \Delta) = \begin{cases} \frac{1}{\vartheta_2}\cdot\delta + o(\delta), & r = 0, \\ \psi_{\vartheta_2}(r)\cdot\sqrt{\Delta} + o(\sqrt{\Delta}), & r \in (0, \infty), \\ \frac{2}{\sqrt{\vartheta_2\pi}}\cdot\sqrt{\Delta} + o(\sqrt{\Delta}), & r = \infty, \end{cases}$$

where

$$\psi_{\vartheta_2}(r) := \frac{2}{\sqrt{\pi\vartheta_2}}\Big(1 - \mathrm{e}^{-\frac{r^2}{4\vartheta_2}} + \frac{r}{\sqrt{\vartheta_2}}\int_{\frac{r}{2\sqrt{\vartheta_2}}}^{\infty} \mathrm{e}^{-z^2}\,dz\Big). \tag{2.13}$$

If, moreover, $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$, *we have*

$$\Phi_\vartheta(\delta,\Delta) = \mathrm{e}^{-\kappa\delta/2}\psi_{\vartheta_2}(r)\cdot\sqrt{\Delta} + \mathcal{O}(\Delta^{3/2}). \tag{2.14}$$

Remark 2.2.6. The first order constants appearing in the asymptotic expressions in *(ii)* stem from a first derivative of $F_{\vartheta_2}(\cdot,\Delta)$ in 0 in case $r=0$ and a Riemann sum approximation of $F_{\vartheta_2}(\delta,\Delta)$ in case $r\neq 0$, respectively. For simplicity, assuming that $\kappa=0$, the proof of Proposition 2.2.5 shows a more precise expression for the remainder terms in case $r\in\{0,\infty\}$:

$$\mathbf{E}\left(D_{ik}^2\right) = \begin{cases} \frac{1}{\vartheta_2}\cdot\delta + \mathcal{O}(\delta^2/\sqrt{\Delta}), & r=0,\\ \frac{2}{\sqrt{\pi\vartheta_2}}\cdot\sqrt{\Delta} + \mathcal{O}(\Delta^{3/2}/\delta^2), & r=\infty. \end{cases}$$

Thus, if our analysis of the remainder terms is sharp (which we believe is the case), the first order approximations have a poor quality if $\delta/\sqrt{\Delta}$ converges slowly.

Proposition 2.2.5 suggests to renormalize double increments with δ if $\delta/\sqrt{\Delta}\to 0$ and with $\sqrt{\Delta}$ otherwise, which is in line with the renormalizations of V_{sp} and V_{t}, respectively. However, this approach might not be feasible: Firstly, it requires the knowledge which asymptotic regime is present, i.e., whether or not $\delta/\sqrt{\Delta}\to 0$. Especially for one given set of observations this information may be inaccessible. In this case renormalizing with $\Phi_\vartheta(\delta,\Delta)$ automatically captures the correct asymptotic regime. Secondly, if $r\in\{0,\infty\}$, the previous remark shows that the asymptotic expressions for $\Phi_\vartheta(\delta,\Delta)$ may lead to an undesirably large bias. In fact, in order to obtain a central limit theorem with $1/\sqrt{MN}$-rate of convergence, e.g., for the case $T\equiv 1$, we would have to impose the assumptions $N^2/M\to 0$ and $M^5/N\to 0$, respectively. These constraints are even more restrictive than the ones required for time or space increments.

Therefore, we renormalize with $\Phi_\vartheta(\delta,\Delta)$ and introduce the *rescaled realized quadratic space-time variation*

$$\mathbb{V} := \frac{1}{MN\Phi_\vartheta(\delta,\Delta)}\sum_{k=0}^{M-1}\sum_{i=0}^{N-1}\mathrm{e}^{\kappa y_k}D_{ik}^2.$$

As for V_{sp}, we redefine $\mathbb{V}$ by summing over $i\in\{1,\dots,N\}$ instead of $i\in\{0,\dots,N-1\}$ in the case $X_0=0$.

Theorem 2.2.7. *Let* $b>0$ *and assume* $\Delta\to 0$ *as well as* $T=o(M)$. *If either* $\delta/\sqrt{\Delta}\to r\in\{0,\infty\}$ *or* $\delta/\sqrt{\Delta}\equiv r\in(0,\infty)$, *then*

$$\sqrt{MN}(\mathbb{V}-\sigma^2)\xrightarrow{\mathcal{D}}\mathcal{N}\big(0, C\big(r/\sqrt{\vartheta_2}\big)\sigma^4\big),\quad N,M\to\infty,$$

where $C(\cdot)$ *is a bounded continuous function on* $[0,\infty]$, *given by* (2.34), *satisfying*

$$C(0)=3 \qquad \text{and} \qquad C(\infty) = 3 + \frac{3}{2}\sum_{J=1}^{\infty}\left(\sqrt{J-1}-\sqrt{J+1}-2\sqrt{J}\right)^2.$$

Remark 2.2.8. The redefinition of $\mathbb{V}$ for the case $X_0=0$ is necessary for the conclusion of the above theorem: E.g., if $\kappa=0$, repeating the calculations from the proof of Proposition 2.2.5 with $X_0=0$ yields that $\mathbf{E}(D_{0k}^2)\approx\frac{1}{2}\Phi_\vartheta(\delta,2\Delta)$ instead of $\mathbf{E}(D_{0k}^2)\approx\Phi_\vartheta(\delta,\Delta)$. Thus, summing over $i\in\{0,\dots,N-1\}$ in the definition of $\mathbb{V}$ introduces a bias of the order $\mathbf{E}(\mathbb{V}-\sigma^2)\eqsim\frac{1}{N}$, which is not negligible for the central limit theorem, unless $M=o(N)$.

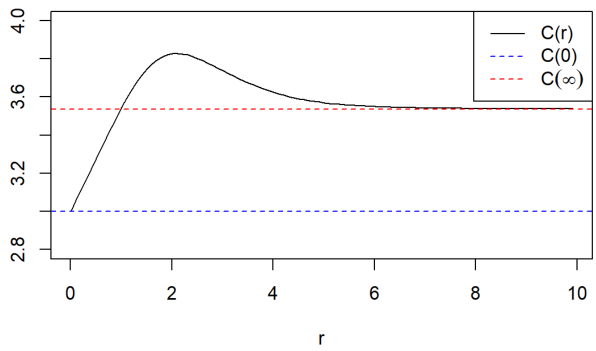


Figure 2.1: Plot of the function $C(\cdot)$ from Theorem 2.2.7.

Note that the central limit theorem for $\mathbb{V}$ holds with (almost) no assumptions on the relation of the temporal and spatial sampling frequencies. In particular, assuming a fixed time horizon, we can close the gap $\sqrt{N} \lesssim M \lesssim N$, where the central limit theorems hold neither for space nor for time increments. Also, the condition $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$ could be relaxed to $\delta/\sqrt{\Delta} \to r \in (0,\infty)$ as long as the convergence is sufficiently fast, which we omit for the sake of simplicity.

Figure 2.1 shows a plot of the function C appearing in the asymptotic variance in Theorem 2.2.7. Evidently, the asymptotic variance is minimal for $r = 0$, where it takes the value $3\sigma^4$, and maximal for $r \approx 2\sqrt{\vartheta_2}$, where it is given by roughly $3.83\sigma^4$. For larger values of r, the asymptotic variance approaches the value $C(\infty)\sigma^4 \approx 3.54\sigma^4$.

In the balanced sampling design $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$, (2.14) shows that $\Phi_\vartheta(\delta,\Delta)$ and its first order approximation are sufficiently close to be exchanged in the previous theorem. Thus, we can use a simpler renormalization which particularly does not depend on the model parameters. Noting that the condition $T = o(M)$ can be rewritten as $N\Delta^{3/2} \to 0$ if $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$, we conclude the following central limit theorem for

$$\mathbb{V}_r := \frac{1}{MN\sqrt{\Delta}} \sum_{k=0}^{M-1} \sum_{i=0}^{N-1} \mathrm{e}^{\frac{\kappa}{2}(y_k+y_{k+1})} D_{ik}^2 \tag{2.15}$$

and its obvious modification for the case $X_0 = 0$.

Corollary 2.2.9. *Let $b > 0$ and assume $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$ as well as $N\Delta^{3/2} \to 0$. Then, we have*

$$\sqrt{MN}\Big(\mathbb{V}_r - \psi_{\vartheta_2}(r)\sigma^2\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, C\big(r/\sqrt{\vartheta_2}\big)\psi_{\vartheta_2}^2(r)\sigma^4\Big), \quad N, M \to \infty,$$

with $\psi_{\vartheta_2}(r)$ from (2.13) *and $C(\cdot)$ from* (2.34).

To end this section, we compare the realized quadratic variations $V_{\mathrm{t}}, V_{\mathrm{sp}}$ and $\mathbb{V}$ and their asymptotic variances. For this purpose, we scale the statistics in such a way that they are asymptotically centered around the same mean, say σ^2:

$$V_{\mathrm{t}}' = \sqrt{\pi\vartheta_2} V_{\mathrm{t}}, \quad V_{\mathrm{sp}}' = 2\vartheta_2 V_{\mathrm{sp}}, \quad V' = \mathbb{V}. \tag{2.16}$$

For simplicity, let $\kappa = 0$. Plugging in the asymptotic expressions for $\Phi_\vartheta(\delta,\Delta)$ from Proposition 2.2.5

shows that

$$V' \approx \frac{1}{2}\sum_{k=0}^{M-1}\sum_{i=0}^{N-1} D_{ik}^2 \cdot \begin{cases} \dfrac{2\vartheta_2}{NM\delta}, & \delta/\sqrt{\Delta}\to 0,\\ \dfrac{\sqrt{\vartheta_2\pi}}{NM\sqrt{\Delta}}, & \delta/\sqrt{\Delta}\to\infty. \end{cases}$$

Therefore, V' approximately coincides with V'_{sp} and V'_{t} for $r\in\{0,\infty\}$, respectively, except for the factor $1/2$ and using double increments instead of time or space increments, respectively.

Further, denoting the asymptotic variances of $V'_{\mathrm{t}}, V'_{\mathrm{sp}}$ and V' by $\mathfrak{S}_{\mathrm{t}}$, $\mathfrak{S}_{\mathrm{sp}}$ and $\mathfrak{S}(r)$, respectively, we observe the relations

$$\mathfrak{S}(\infty)=\frac{3}{2}\mathfrak{S}_{\mathrm{t}} \quad\text{and}\quad \mathfrak{S}(0)=\frac{3}{2}\mathfrak{S}_{\mathrm{sp}}.$$

The presence of the factor $3/2$ may be explained as follows, e.g., for space increments: Since each double increment consists of two space increments and neighboring (in time) double increments have one space increment in common, the covariances that contribute to the asymptotic variance are given by

$$\begin{aligned}
\mathrm{Var}(D_{ik}^2) &= 2\mathrm{Var}(D_{ik})^2 \approx 2\big(2\mathrm{Var}((\delta_k^M X)(t_i))\big)^2 = 4\mathrm{Var}((\delta_k^M X)^2(t_i)),\\
\mathrm{Cov}(D_{ik}^2, D_{i(k+1)}^2) &= 2\mathrm{Cov}(D_{ik}, D_{i(k+1)})^2 \approx 2(\mathrm{Var}((\delta_k^M X)(t_i)))^2 = \mathrm{Var}((\delta_k^M X)^2(t_i)),\\
\mathrm{Cov}(D_{ik}^2, D_{i(k-1)}^2) &\approx \mathrm{Var}((\delta_k^M X)^2(t_i)),
\end{aligned}$$

where we have used (2.7). Hence, we get a factor of $6/4=3/2$ in the asymptotic variance of V'.

2.2.2 Construction of estimators

We exploit the central limit theorems for realized quadratic variations for the construction of estimators. First, we discuss estimation of σ^2 or ϑ_2, given that the other parameter is known, respectively. Naturally, the estimation problem becomes much harder when none of the parameters is known. Nevertheless, using double increments, we can estimate $(\sigma^2,\vartheta_2,\vartheta_1)$ in an (almost) rate optimal way. Assuming $T\to\infty$, the same holds for $(\sigma^2,\vartheta_2,\vartheta_1,\vartheta_0)$ with the additional help of the sum of squares statistic S from (2.20) below.

Estimation of σ^2 or ϑ_2

It is straightforward to use the results from the previous section to construct method of moments estimators for the volatility parameter σ^2 or the diffusivity parameter ϑ_2, provided that the other two parameters in the parametrization $(\sigma^2,\vartheta_2,\kappa)$ are known, respectively. Doing so, we generalize the spatial increments based estimator from Cialenco and Huang [19] to the double asymptotic regime and we complement the time increments based methods in Bibinger and Trabs [9] and Chong [15]. Note that assuming (ϑ_2,κ) to be known is the same as assuming $(\vartheta_2,\vartheta_1)$ to be known. Further, κ is particularly known ($\kappa=0$) in the relevant sub-model where $\vartheta_1=0$. Our estimators do not hinge on ϑ_0 such that the knowledge of its true value is not required.

Assuming, firstly, that ϑ_2 and κ are known, we obtain the following volatility estimators:

$$\hat\sigma^2_{\mathrm{sp}} := V'_{\mathrm{sp}}, \qquad \hat\sigma^2_{\mathrm{t}} := V'_{\mathrm{t}} \qquad\text{and}\quad \hat\sigma^2 := \mathbb{V}$$

where V'_{sp} and V'_{t} have been introduced in (2.16). In view of the central limit theorems from the previous section, the delta method reveals their asymptotic distributions:

Proposition 2.2.10.

(i) If $N=o(M)$, then we have

$$\sqrt{MN}\,\big(\hat\sigma^2_{\mathrm{sp}}-\sigma^2\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,2\sigma^4), \quad N,M\to\infty.$$

(ii) If $M = o(\Delta^{-\rho})$ for some $\rho \in (0, 1/2)$ and $T\Delta \to 0$, then we have with B defined in (2.11)*:*

$$\sqrt{MN}\left(\hat{\sigma}_{\mathrm{t}}^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, B\sigma^4), \quad N, M \to \infty.$$

(iii) If either $\delta/\sqrt{\Delta} \to r \in \{0, \infty\}$ or $\delta/\sqrt{\Delta} \equiv r \in (0, \infty)$ and $T = o(M)$, then we have with $C(\cdot)$ from (2.34)*:*

$$\sqrt{MN}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C(r/\sqrt{\vartheta_2})\sigma^4), \quad N, M \to \infty.$$

As discussed at the end of Section 2.2.1, the double increments estimator has a larger variance than the single increments estimators. Hence, if one of the regimes $N = o(M)$ or $M = o(\Delta^{-1/2})$ certainly applies, the single increments estimators are preferable. If none of the regimes is present or the situation is unclear, one can profit from the robustness of the double increments estimator with respect to the sampling regime.

If $N = o(M)$, the situation is close to that of N independent semimartingales (cf. Proposition 1.2.1) and the asymptotic variance $2\sigma^4$ of the spatial increments estimator equals the Cramér-Rao lower bound for estimating σ^2, as can be seen by a simple calculation. Consequently, $\hat{\sigma}_{\mathrm{sp}}^2$ is an asymptotically efficient estimator. The efficiency loss of the other estimators is due to the fact that for increasingly more temporal observations the infinite dimensional nature of the process X becomes apparent, leading to non-negligible covariances between increments.

If σ^2 and κ are known, the diffusivity ϑ_2 can be estimated by

$$\hat{\vartheta}_{2,\mathrm{sp}} := \frac{\sigma^2}{2V_{\mathrm{sp}}} \quad \text{and} \quad \hat{\vartheta}_{2,\mathrm{t}} := \frac{\sigma^4}{\pi V_{\mathrm{t}}^2},$$

using V_{sp} and V_{t} from (2.12) and (2.9), respectively. Due to the non-trivial dependence of the renormalization $\Phi_\vartheta(\delta, \Delta)$ on ϑ, it is not apparent how to construct a method of moments estimator for ϑ_2 based on Theorem 2.2.7, in general. However, in the balanced design $\delta/\sqrt{\Delta} \equiv r > 0$, the renormalization can be decoupled from the unknown parameter, as exploited in Corollary 2.2.9. Since the function $\vartheta_2 \mapsto \psi_{\vartheta_2}(r)$ has range $(0, \infty)$ and is monotonic, there is an inverse $H_r(\cdot)$ and we can define the method of moments estimator

$$\hat{\vartheta}_{2,r} = H_r(\mathbb{V}_r/\sigma^2)$$

with $\mathbb{V}_r$ from (2.15). As a direct consequence of the delta method and the relation

$$H_r'(\psi_{\vartheta_2}(r)) = \Big(\frac{\partial}{\partial \vartheta_2}\psi_{\vartheta_2}(r)\Big)^{-1} = -\vartheta_2^{3/2}\sqrt{\pi}\Big(1 - \mathrm{e}^{-\frac{r^2}{4\vartheta_2}} + \frac{2r}{\sqrt{\vartheta_2}}\int_{\frac{r}{2\sqrt{\vartheta_2}}} \mathrm{e}^{-z^2}\,dz\Big)^{-1},$$

we obtain the following proposition.

Proposition 2.2.11.

(i) If $N = o(M)$, then we have

$$\sqrt{MN}\left(\hat{\vartheta}_{2,\mathrm{sp}} - \vartheta_2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\vartheta_2^2\right), \quad N, M \to \infty.$$

(ii) If $M = o(\Delta^{-\rho})$ for some $\rho \in (0, 1/2)$ and $T\Delta \to 0$, then we have with B from (2.11)*:*

$$\sqrt{MN}\left(\hat{\vartheta}_{2,\mathrm{t}} - \vartheta_2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\vartheta_2^2 B), \quad N, M \to \infty.$$

(iii) If $\delta/\sqrt{\Delta} \equiv r > 0$ and $T\sqrt{\Delta} \to 0$, then we have with $C(\cdot)$ from (2.34)*:*

$$\sqrt{MN}(\hat{\vartheta}_{2,r} - \vartheta_2) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, C(r/\sqrt{\vartheta_2})\Big(\psi_{\vartheta_2}(r)\Big/\frac{\partial}{\partial\vartheta_2}\psi_{\vartheta_2}(r)\Big)^2\Big), \quad N, M \to \infty.$$

Estimation of $(\sigma^2, \vartheta_2, \vartheta_1)$ on a fixed time horizon

We now consider the estimation problem on a bounded time horizon when all the parameters in (σ^2, ϑ) are unknown. A first result in that direction was obtained by Bibinger and Trabs [9], who considered a least squares estimator for the parameters $(\sigma_0^2, \kappa) := (\frac{\sigma^2}{\sqrt{\vartheta_2}}, \frac{\vartheta_1}{\vartheta_2})$ based on time increments, namely

$$(\hat{\sigma}_0^2, \hat{\kappa}) = \underset{(\tilde{\sigma}_0^2, \tilde{\kappa})}{\arg\min} \frac{1}{M} \sum_{k=0}^{M-1} \Big(\frac{1}{N\sqrt{\Delta}} \sum_{i=0}^{N-1} (\Delta_i^N X)^2(y_k) - \frac{\tilde{\sigma}_0^2}{\sqrt{\pi}} \mathrm{e}^{-\tilde{\kappa} y_k} \Big)^2. \tag{2.17}$$

Using their central limit theorem (2.10) and classical M-estimation theory, the estimator is shown to be aymptotically normal with $1/\sqrt{MN}$ rate of convergence in the regime $M = o(\Delta^{-1/2})$. Note that (2.17) is an M-estimator exploiting the probabilistic structure of the processes $[0,T] \ni t \mapsto X_t(y_k)$. Indeed, our Proposition 2.1.1 (*iii*) reveals that the parameters σ_0^2 and κ are exactly the ones one can expect to identify with such an estimator. Building on (2.17), Kaino and Uchida [49] derived an estimators for $(\sigma^2, \vartheta_2, \vartheta_1)$ (as well as for $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ in case $T \to \infty$) but their thinning approach results in a convergence rate which is no faster than $1/\sqrt{N}$, see also the discussion at the end of the current section. Analogously to (2.17), it is possible to estimate the parameters appearing in Proposition 2.1.1 (*ii*), i.e. (ρ^2, κ) with $\rho^2 = \sigma^2/\vartheta_2$, using spatial increments and Theorem 2.2.3: provided that $N = o(M)$,

$$(\hat{\rho}^2, \hat{\kappa}) := \underset{(\tilde{\rho}^2, \tilde{\kappa})}{\arg\min} \frac{1}{M} \sum_{k=0}^{M-1} \left(\frac{2}{N\delta} \sum_{i=0}^{N-1} (\delta_k^M X)^2(t_i) - \tilde{\rho}^2 \mathrm{e}^{-\tilde{\kappa} y_k} \right)^2$$

satisfies a central limit theorem with rate $1/\sqrt{MN}$. We omit a detailed analysis of this estimator.

Recall from Proposition 2.1.1 and the subsequent discussion that ϑ_0 cannot be estimated consistently on a finite time horizon. Also, in order to estimate more than two parameters, we cannot rely on only the temporal or only the spatial probabilistic structure of X. To estimate all three identifiable parameters

$$\eta := (\sigma^2, \vartheta_2, \vartheta_1),$$

we employ a least squares approach based on double increments. Due to the highly nontrivial dependence of the normalization $\Phi_\vartheta(\delta, \Delta)$ on ϑ, a direct application of Theorem 2.2.7 is impossible. Assuming, however, a balanced design $\delta/\sqrt{\Delta} \equiv r \in (0, \infty)$, we can use Corollary 2.2.9 where the normalization is decoupled from the unknown parameter ϑ:

Let $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$ and define $\bar{D}_{ik} := D_{ik} + D_{(i+1)k}$ as well as $z_k = (y_k + y_{k+1})/2$. Corollary 2.2.9 suggests that for the stationary solution we have

$$\frac{1}{N\sqrt{\Delta}} \sum_{i=0}^{N-1} D_{ik}^2 \approx \mathrm{e}^{-\kappa z_k} \sigma^2 \psi_{\vartheta_2}(r) \quad \text{and} \quad \frac{1}{N\sqrt{2\Delta}} \sum_{i=0}^{N-2} \bar{D}_{ik}^2 \approx \mathrm{e}^{-\kappa z_k} \sigma^2 \psi_{\vartheta_2}(r/\sqrt{2}).$$

Now, as suggested by Figure 2.2, the function $r \mapsto \psi_{\vartheta_2}(r)$ is strictly increasing. In fact, by considering the two different sampling frequency ratios r and $r/\sqrt{2}$, it will be shown that we can distinguish σ^2 and ϑ_2 instead of recovering only the product $\sigma^2 \psi_{\vartheta_2}(r)$. To estimate η, we introduce the contrast process

$$K_{M,N}(\tilde{\eta}) := K_{M,N}^1(\tilde{\eta}) + K_{M,N}^2(\tilde{\eta}),$$

where

$$K_{M,N}^1(\tilde{\eta}) := \frac{1}{M} \sum_{k=0}^{M-1} \Big(\frac{1}{N\sqrt{\Delta}} \sum_{i=0}^{N-1} D_{ik}^2 - f_{\tilde{\eta}}^1(z_k) \Big)^2,$$

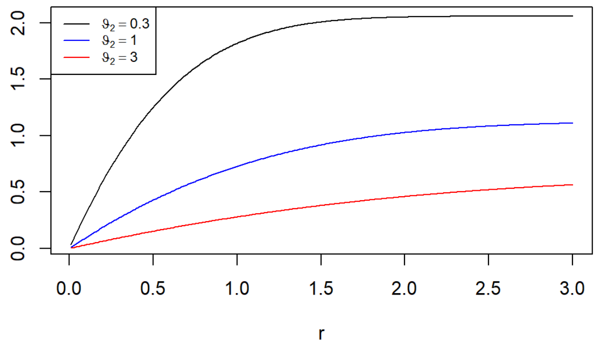


Figure 2.2: Plot of the function $r \mapsto \psi_{\vartheta_2}(r)$ for different values of ϑ_2.

$$K^2_{M,N}(\tilde{\eta}) := \frac{1}{M}\sum_{k=0}^{M-1}\Big(\frac{1}{N\sqrt{2\Delta}}\sum_{i=0}^{N-2}\bar{D}^2_{ik} - f^2_{\tilde{\eta}}(z_k)\Big)^2$$

and $f^{\nu}_{\eta}(z) := \sigma^2 \mathrm{e}^{-\kappa z}\psi_{\vartheta_2}(r/\sqrt{\nu})$, $\nu = 1, 2$. As before, we sum over $i \in \{1, \dots, N\}$ instead of $i \in \{0, \dots, N-1\}$ in the case $X_0 = 0$. The corresponding M-estimator is given by

$$\hat{\eta} := \arg\min_{\tilde{\eta}\in H} K_{M,N}(\tilde{\eta}), \tag{2.18}$$

where H is some subset of $(0,\infty)^2 \times \mathbb{R}$. Again, this estimator does not require any prior knowledge on the parameters, as it does not depend on ϑ_0.

Theorem 2.2.12. *Assume that $T > 0$ is fixed, $b > 0$ and that $\delta/\sqrt{\Delta} \equiv r > 0$. If $H \subset (0,\infty)^2 \times \mathbb{R}$ is a compact set and $\eta = (\sigma^2, \vartheta_2, \vartheta_1)$ lies in its interior, then the least squares estimator $\hat{\eta}$ from* (2.18) *satisfies*

$$\sqrt{MN}(\hat{\eta} - \eta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega^r_{\eta}), \qquad M, N \to \infty,$$

where $\Omega^r_{\eta} \in \mathbb{R}^{3\times 3}$ is a strictly positive definite covariance matrix, explicitly given by (2.38).

Remark 2.2.13. If ϑ_1 is known and the sample size is sufficiently large, the estimator for (σ^2, ϑ_2) can be computed without solving a minimization problem: For simplicity, assume $\vartheta_1 = 0$ and let

$$V^1 := \frac{1}{MN\sqrt{\Delta}}\sum_{k=0}^{M-1}\sum_{i=0}^{N-1} D^2_{ik}, \qquad V^2 := \frac{1}{MN\sqrt{2\Delta}}\sum_{k=0}^{M-1}\sum_{i=0}^{N-1}\bar{D}^2_{ik}.$$

Further, denote by G_r the inverse function of $\vartheta_2 \mapsto \psi_{\vartheta_2}(r)/\psi_{\vartheta_2}(r/\sqrt{2})$, whose existence is part of the proof of the above theorem. Then, we have

$$\hat{\vartheta}_2 = G_r(V^1/V^2), \qquad \hat{\sigma}^2 = V^1/\psi_{\hat{\vartheta}_2}(r),$$

provided that V_1/V_2 lies in the range of $\vartheta_2 \mapsto \psi_{\vartheta_2}(r)/\psi_{\vartheta_2}(r/\sqrt{2})$. Due to consistency of (V^1, V^2), the latter is true with probability tending to one.

Even when $\delta/\sqrt{\Delta} \equiv r > 0$ does not hold, there are always subsets of the data having the balanced sampling design. Hence, the estimation procedure treated in Theorem 2.2.12 can be generalized to an arbitrary set $\{X_{t_i}(y_k), i \le N,\, k \le M\}$ of discrete observations by considering an averaged version

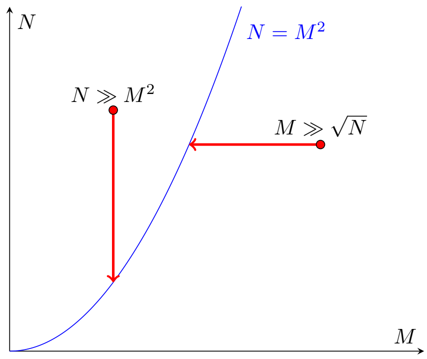


Figure 2.3: Determining the sample size of subsets admitting a balanced sampling design.

of the above contrast process. To that aim, choose $v, w \in \mathbb{N}$ such that $v \approx \max(1, \delta^2/\Delta)$ and $w \approx \max(1, \sqrt{\Delta}/\delta)$. Then, $\tilde{\Delta} := v\Delta$ and $\tilde{\delta} := w\delta$ satisfy

$$r := \tilde{\delta}/\sqrt{\tilde{\Delta}} \approx 1.$$

Using double increments on the coarser grid, namely

$$D_{v,w}(i,k) := X_{t_{i+v}}(y_{k+w}) - X_{t_i}(y_{k+w}) - X_{t_{i+v}}(y_k) + X_{t_i}(y_k),$$

we set

$$\mathcal{K}^{\nu}_{N,M}(\tilde{\eta}) := \frac{1}{M-w+1}\sum_{k=0}^{M-w}\left(\frac{1}{(N-\nu v+1)\sqrt{\nu v\Delta}}\sum_{i=0}^{N-\nu v} D^2_{\nu v,w}(i,k) - f^{\nu}_{\tilde{\eta}}\Big(\frac{y_k+y_{k+w}}{2}\Big)\right)^2,$$

with $f^{\nu}_{\eta}(z) = 2\sigma^2\psi_{\vartheta_2}(r/\sqrt{\nu})\mathrm{e}^{-\kappa z}$ and $\nu = 1, 2$. In the case $X_0 = 0$, we employ the obvious redefinition of $\mathcal{K}^{\nu}_{N,M}$. The final estimator for η is then defined as

$$\hat{\eta}_{v,w} := \operatorname*{arg\,min}_{\tilde{\eta}\in H}\big(\mathcal{K}^1_{N,M}(\tilde{\eta}) + \mathcal{K}^2_{N,M}(\tilde{\eta})\big). \tag{2.19}$$

The rate of convergence of this estimation procedure is inherited from the observations on the coarser grids $\{(t_{i+jv}, y_{k+lw}) : 0 \le j \le N/v - 1, 0 \le l \le M/w - 1\}$, $i = 0, \dots, v-1, k = 0, \dots, w-1$, on which we calculate the double increments. It follows from $\Delta \approx N^{-1}$ and $\delta \approx M^{-1}$ that each such subset consists of

$$\frac{M}{w}\cdot\frac{N}{v} \approx (M \wedge \sqrt{N})(N \wedge M^2) = M^3 \wedge N^{3/2}$$

observations and has a balanced design by construction. Figure 2.3 illustrates how the sample size of the subsets admitting a balanced design results from the total sample size: if N is much larger than M^2, then the sub-sample size is $M^2 \cdot M = M^3$ and if M is much larger than $\sqrt{N}$, then the sub-sample size is $N \cdot \sqrt{N} = N^{3/2}$. Therefore, Theorem 2.2.12 implies the convergence rate $\big(M^3 \wedge N^{3/2}\big)^{-1/2}$.

Corollary 2.2.14. *Assume that $T > 0$ is fixed, $b > 0$, and that $H \subset (0,\infty)^2 \times \mathbb{R}$ is a compact set such that $\eta = (\sigma^2, \vartheta_2, \vartheta_1)$ lies in its interior. If there exist values $v \approx \max(1, \delta^2/\Delta) \in \mathbb{N}$ and $w \approx \max(1, \sqrt{\Delta}/\delta) \in \mathbb{N}$ such that $w\delta/\sqrt{v\Delta}$ is constant, then the estimator $\hat{\eta}_{v,w}$ given by* (2.19) *satisfies*

$$\|\hat{\eta}_{v,w} - \eta\| = \mathcal{O}_p\Big(\frac{1}{\sqrt{M^3 \wedge N^{3/2}}}\Big), \quad M, N \to \infty.$$

Remark 2.2.15.

1. The rate of convergence of the estimator $\hat{\eta}_{v,w}$ matches the lower bound provided by Theorem 2.1.2 up to a logarithmic factor. Hence, our estimator is essentially rate optimal and, in particular, $(M^3 \wedge N^{3/2})^{-1/2}$ is the optimal rate of convergence for estimators of $(\sigma^2, \vartheta_2, \vartheta_1)$ on a bounded time horizon.

2. The same rate of convergence is achieved if, instead of averaging, one computes the contrast process from only *one* balanced sub-sample and discards the remaining data. Thus, if $M^2/N \to \{0, \infty\}$, the optimal rate of convergence can be reached by using only a small portion of the available data. On the other hand, our simulation study in Section 3.3 suggests that using the whole data set is beneficial for the asymptotic variance of the estimator.

3. Integer values v and w such that $w\delta/\sqrt{v\Delta}$ is constant exist, for instance, if the observations are recorded on a dyadic grid, namely when $M = 2^m$ and $N = 4^n$ with $m, n \to \infty$.

Estimation of $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ in the case $T \to \infty$

Next, we consider the regime $T \to \infty$, where all four parameters $(\sigma^2, \vartheta_2, \vartheta_1, \vartheta_0)$ can be estimated consistently. Again, let us first consider the situation of a balanced sampling design $\delta/\sqrt{\Delta} \equiv r > 0$. In fact, the central limit theorem for double increments from Corollary 2.2.9 is not limited to a bounded time horizon but only requires $T\sqrt{\Delta} \to 0$. Hence, under the latter assumption, we can use (2.18) in order to estimate $\eta = (\sigma^2, \vartheta_2, \vartheta_1)$ and, in particular, Theorem 2.2.12 carries over to an unbounded time horizon. To obtain an estimator for the remaining parameter ϑ_0, we will now define a method of moments estimator for ϑ_0 and replace the unknown parameters $(\sigma^2, \vartheta_2, \vartheta_1)$ appearing in its definition by the corresponding estimates from $\hat{\eta}$. Since the parameter ϑ_0 appears in the spatial covariance function of the stationary solution, it is natural to consider the statistic

$$S := \frac{1}{MN} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} X_{t_i}^2(y_k) \mathrm{e}^{\kappa y_k} \tag{2.20}$$

in order to derive a method of moments estimator. Based on the space-time covariance function of the process $(t, y) \mapsto X_t(y)\mathrm{e}^{\kappa y/2}$ under (ST), i.e.,

$$\begin{aligned}\rho_{xy}(t) &:= \mathrm{Cov}(X_0(x)\mathrm{e}^{\kappa x/2}, X_t(y)\mathrm{e}^{\kappa y/2}) = \mathrm{Cov}(X_t(x)\mathrm{e}^{\kappa x/2}, X_0(y)\mathrm{e}^{\kappa y/2}) \\ &= \sigma^2 \sum_{\ell \geq 1} \frac{\mathrm{e}^{-\lambda_\ell t}}{\lambda_\ell} \sin(\pi \ell x) \sin(\pi \ell y),\end{aligned}$$

the mean of S can be expressed via

$$\mathbf{E}(S) = \frac{1}{M} \sum_{k=0}^{M-1} \rho_{y_k y_k}(0) = \frac{\sigma^2}{M} \sum_{k=0}^{M-1} \sum_{\ell \geq 1} \frac{\sin^2(\pi \ell y_k)}{\lambda_\ell}.$$

Note that there is also a closed form expression for $\rho_{yy}(0)$, see Proposition 1.2.1. In the case $X_0 = 0$, we redefine S by summing over $i \in \{1, \ldots, N\}$ instead of $i \in \{0, \ldots, N-1\}$.

Proposition 2.2.16. *If $M, N, T \to \infty$ and $\Delta \to 0$, then*

$$\sqrt{T}\Big(S - \frac{1}{M} \sum_{k=0}^{M-1} \rho_{y_k y_k}(0)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, 2D^2\Big),$$

where

$$D^2 := \frac{1}{(1-2b)^2} \int_{-\infty}^{\infty} \int_b^{1-b} \int_b^{1-b} \rho_{xy}^2(t)\, dx\, dy\, dt \tag{2.21}$$

and $\rho_{xy}(t) := \rho_{xy}(-t)$ *for* $t < 0$. *If, additionally,* $\sqrt{T}/M \to 0$, *then*

$$\sqrt{T}\Big(S - \frac{1}{1-2b}\int_b^{1-b} \rho_{yy}(0)\, dy\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, 2D^2\Big).$$

Remark 2.2.17. A balanced sampling design is not assumed in Proposition 2.2.16. On the other hand, if a balanced sampling design is present, then $\sqrt{T}/M \to 0$ is implied by the condition $T\sqrt{\Delta} \to 0$ from Corollary 2.2.9.

In order to indicate the dependence of S on the unknown parameter κ, we will write $S(\kappa)$ in the sequel. It follows from Proposition 2.2.16 that

$$S(\kappa) \xrightarrow{\mathbf{P}} \frac{\sigma^2}{\vartheta_2} I_b(\Gamma)$$

where $I_b(\Gamma) := \frac{1}{1-2b}\int_b^{1-b} \sum_{\ell\geq 1} \frac{\sin^2(\pi \ell y)}{\pi^2\ell^2+\Gamma}$ and, as before, $\Gamma = \frac{\vartheta_1^2}{4\vartheta_2^2} - \frac{\vartheta_0}{\vartheta_2}$. Clearly, the function I_b is decreasing and, hence, injective. By inverting for ϑ_0 and plugging in the estimators (2.18) for η, we obtain

$$\hat{\vartheta}_0 := \hat{\vartheta}_2 \left(\frac{\hat{\kappa}^2}{4} - I_b^{-1}\Big(\frac{\hat{\vartheta}_2}{\hat{\sigma}^2} S(\hat{\kappa})\Big)\right) \quad \text{with} \quad \hat{\kappa} := \hat{\vartheta}_1/\hat{\vartheta}_2 \tag{2.22}$$

as an estimator for ϑ_0. Based on Theorem 2.2.12 and Proposition 2.2.16, we obtain the following central limit theorem for the estimator $(\hat{\eta}, \hat{\vartheta}_0)$ in case of a balanced sampling design.

Theorem 2.2.18. *Assume* $\delta/\sqrt{\Delta} \equiv r \in (0,\infty)$ *and* $T\sqrt{\Delta} \to 0$. *Further, let* H *be a compact subset of* $(0,\infty)^2 \times \mathbb{R}$ *such that* η *lies in its interior. Then, for* $T, N, M \to \infty$ *and* $\Delta \to 0$, *we have*

$$\begin{pmatrix} \sqrt{MN}(\hat{\eta} - \eta) \\ \sqrt{T}(\hat{\vartheta}_0 - \vartheta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} \Omega_\eta^r & 0 \\ 0 & \alpha^2_{\sigma^2,\vartheta} \end{pmatrix}\right),$$

where $\Omega_\eta^r \in \mathbb{R}^{3\times 3}$ *is the strictly positive definite covariance matrix from Theorem 2.2.12 and*

$$\alpha^2_{\sigma^2,\vartheta} = \frac{2\vartheta_2^4 D^2}{\sigma^4 I_b'(\Gamma)^2}.$$

As in the case of a bounded time horizon, this estimation procedure can be adapted for an unbalanced sampling design. Again, we can estimate η using $\hat{\eta}_{v,w}$ from (2.19) with $v \approx \max(1, \delta^2/\Delta)$ and $w \approx \max(1, \sqrt{\Delta}, \delta)$. ϑ_0 can then be estimated via the plug in approach (2.22) where $\hat{\eta}$ is replaced by $\hat{\eta}_{v,w}$. We refer to this estimator by $\hat{\vartheta}_0^{vw}$.

Recall that the rate of convergence of $\hat{\eta}_{v,w}$ is determined by the sample size of the subsets admitting a balanced sampling design. In the case of a growing time horizon, this sample size depends on T and is given by

$$\frac{M}{w}\cdot\frac{N}{v} \approx (\delta^{-1} \wedge \Delta^{-1/2})(N \wedge T\delta^{-2}) \approx (\delta^{-1} \wedge \Delta^{-1/2})^3\, T = \frac{T}{\max(\Delta^{3/2}, \delta^3)}.$$

Corollary 2.2.19. *Assume* $T\max(\sqrt{\Delta}, \delta) \to 0$ *and let* H *be a compact subset of* $(0,\infty)^2 \times \mathbb{R}$ *such that* η *lies in its interior. If there exist values* $v \approx \max(1, \delta^2/\Delta)$ *and* $w \approx \max(1, \sqrt{\Delta}/\delta)$ *such that* $w\delta/\sqrt{v\Delta}$ *is constant, then we have*

$$\|\hat{\eta}_{v,w} - \eta\| = \mathcal{O}_p\left(\sqrt{\frac{\max(\delta^3, \Delta^{3/2})}{T}}\right), \qquad |\hat{\vartheta}_0^{vw} - \vartheta_0| = \mathcal{O}_p\left(\frac{1}{\sqrt{T}}\right)$$

for $T, N, M \to \infty$ *and* $\Delta \to 0$. *In particular,* $\sqrt{T}(\hat{\vartheta}_0^{vw} - \vartheta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha^2_{\sigma^2,\vartheta})$.

Remark 2.2.20.

1. Comparison with Theorem 2.1.7 reveals that our estimator identifies both $(\sigma^2, \vartheta_2, \vartheta_1)$ and ϑ_0 at (almost) their optimal rates of convergence, respectively.
2. The assumption $T \max(\delta, \sqrt{\Delta}) \to 0$ is the growth condition on T from Theorem 2.2.7, relative to the coarser grid on which the double increments are calculated.

Let us compare our estimator for (σ^2, ϑ) with the thinning approach considered by Kaino and Uchida [49]. The authors work with the same observation scheme as we do, but they strictly require that the spatial margin b of the observation window is zero, i.e., their observations are $(X_{t_i}(y_k))_{i,k}$ with $t_i = i\Delta$, $0 \le i \le N$, and $y_k = \frac{k}{M}$, $1 \le k \le M-1$. Now, by using only a subset of $m \le M$ spatial observations with $m \le \Delta^{-\rho}$ for some $\rho < 1/2$, they estimate $(\sigma^2/\sqrt{\vartheta_2}, \vartheta_1/\vartheta_2)$ using $(\hat{\sigma}_0^2, \hat{\kappa})$ from (2.17) due to Bibinger and Trabs [9]. Then, exploiting the estimator $\hat{\kappa}$, they approximate the first Fourier mode u_1 from (1.12), using the empirical inner product, which is facilitated by assuming $b = 0$. Considering this approximation at $n \le N$ equidistant time points, they estimate σ^2 in terms of the corresponding quadratic variation on a finite time horizon or (σ^2, λ_1) based on a pseudo maximum likelihood estimator when $T \to \infty$. In combination with $(\hat{\sigma}_0^2, \hat{\kappa})$, they obtain asymptotically normal estimators for η and, if $T \to \infty$, for (η, ϑ_0). Since the resulting estimator only relies on the approximation of *one* Fourier mode at n time points, the rate of convergence cannot be faster than $1/\sqrt{n}$. Indeed, on a finite time horizon this rate is achieved under the assumptions

$$\frac{n^{3/2}}{M^{1-\rho_1}} \to 0, \qquad \frac{n^{3/2}}{Nm} \to 0$$

for some $\rho_1 \in (0,1)$, which are necessary to control the errors induced by the different estimation and approximation steps. As a result, $1/\sqrt{N \wedge M^{2/3}}$ is a lower bound for their rate of convergence, which is certainly larger than our rate $1/\sqrt{N^{3/2} \wedge M^3}$. On a large time horizon, the rates of convergence of the estimators for η and ϑ_0 are given by $1/\sqrt{n}$ and $1/\sqrt{T}$, respectively. Here, the assumptions are

$$T\Delta \to 0, \qquad \frac{n^{5/2}}{T^{3/2}Nm} \to 0, \qquad \frac{n^3}{T^2 M^{1-\rho_1}} \to 0$$

for some $\rho_1 \in (0,1)$. Since $m \lesssim \Delta^{-1/2}$ and, thus, $T^{3/2}Nm \lesssim TN^{3/2} = T^{5/2}/\Delta^{3/2}$, the expression $1/\sqrt{(T^{2/3}/\delta^{1/3}) \wedge (T/\Delta^{3/5})}$ is a lower bound for their rate of convergence for estimating η. Again, this is certainly larger than our rate $1/\sqrt{(T/\delta^3) \wedge (T/\Delta^{3/2})}$. The rates of convergence for estimating ϑ_0 agree in both approaches. Only requiring $T\Delta \to 0$, the result from [49] has more flexibility with respect to the length of the observation time compared to the conclusion of Corollary 2.2.19, which requires $T \max(\sqrt{\Delta}, \delta) \to 0$.

2.3 Confidence sets

All estimators considered in the previous section are consistent and the asymptotic variances (covariance matrices) in the central limit theorems are strictly positive (positive definite) as well as continuous in the parameters. Hence, it is possible to construct asymptotic confidence sets for the parameters by employing the standard procedure based on Slutsky's Lemma: Let $\gamma \in \mathbb{R}^d$ be any parameter considered in the previous section, i.e. $\gamma \in \{\sigma^2, \vartheta_2\}$ with $d = 1$, $\gamma = \eta$ with $d = 3$, or $\gamma = (\sigma^2, \vartheta)$ with $d = 4$. In the cases $d \in \{1, 3\}$, we have considered estimators $\hat{\gamma}$ such that $\sqrt{MN}(\hat{\gamma} - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(\gamma))$ for a strictly positive definite covariance matrix $\Sigma(\gamma)$ which is continuous in γ. Thus, Slutsky's Lemma implies $\sqrt{MN}\Sigma(\hat{\gamma})^{-1/2}(\hat{\gamma} - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I)$ with the identity matrix $I \in \mathbb{R}^{d \times d}$. Now, if $z_d(1-\alpha)$ is the $(1-\alpha)$-quantile of the $\chi^2(d)$-distribution, then

$$\left\{ \tilde{\gamma} \in \mathbb{R}^d : \|\Sigma(\hat{\gamma})^{-1/2}(\hat{\gamma} - \tilde{\gamma})\|^2 \le \frac{z_d(1-\alpha)}{MN} \right\}$$

is an approximate $(1-\alpha)$-confidence ellipsoid for γ. In particular, if $d=1$, we obtain the approximate confidence interval

$$\left[\hat{\gamma}-\sqrt{\frac{\Sigma(\hat{\gamma})z_1(1-\alpha)}{MN}},\hat{\gamma}+\sqrt{\frac{\Sigma(\hat{\gamma})z_1(1-\alpha)}{MN}}\right].$$

Analogously, in the case $\gamma=(\sigma^2,\vartheta)=(\eta,\vartheta_0)\in\mathbb{R}^4$, we obtain in the situation of Theorem 2.2.18 that

$$\left\{(\tilde{\eta},\tilde{\vartheta}_0)\in\Theta:\ MN\|((\hat{\Omega}^r)^{-1/2}(\hat{\eta}-\tilde{\eta})\|^2+T\frac{(\hat{\vartheta}_0-\tilde{\vartheta}_0)^2}{\hat{\alpha}^2}\le z_4(1-\alpha)\right\}$$

with $\hat{\Omega}^r=\Omega^r_{\hat{\eta}}$ and $\hat{\alpha}^2=\alpha^2_{\hat{\eta},\hat{\vartheta}_0}$ is an approximate $(1-\alpha)$-confidence set.

2.4 Proofs

The following Sections 2.4.1 - 2.4.3 contain the main proofs of the results from Sections 2.1 and 2.2.1-2.2.2, respectively. Auxiliary results and technical Lemmas are deferred to Section 2.4.4.

2.4.1 Proofs for the lower bounds

First, we proof our result on the absolute continuity properties of the solution process for different parameter values.

Proof of Proposition 2.1.1. The necessity of the conditions on the parameters follows from the discussion subsequent to Proposition 2.1.1: the parameter $\sigma^2/\sqrt{\vartheta_2}\mathrm{e}^{-\kappa x_0}$ can be estimated consistently based on a single spatial observation on a bounded time interval and the parameters $(\frac{\sigma^2}{\vartheta_2},\kappa)$ can be estimated consistently based on a single temporal observation. It remains to prove sufficiency of the conditions on the parameters:

Assertion (i) is a simple consequence of Koski and Loges [50, Proposition 1]: Set $\lambda_\ell=\vartheta_2(\pi^2\ell^2+\Gamma)$ and $\tilde{\lambda}_\ell=\vartheta_2(\pi^2\ell^2+\tilde{\Gamma})$ where $\Gamma=\frac{\vartheta_1^2}{4\vartheta_2^2}-\frac{\vartheta_0}{\vartheta_2}$ and $\tilde{\Gamma}=\frac{\vartheta_1^2}{4\vartheta_2^2}-\frac{\tilde{\vartheta}_0}{\vartheta_2}$. Then, absolute continuity follows from

$$\sum_{\ell\ge1}\frac{(\lambda_\ell-\tilde{\lambda}_\ell)^2}{\lambda_\ell}<\infty.$$

We remark that in the reference [50] the authors work with an H-valued Wiener process as opposed to a cylindrical Brownian motion. Nevertheless, inspection of their proof shows that the result remains valid as long as the condition (L2) from Section 1.1 is fulfilled. Thanks to (i), we may assume $\vartheta_0=\vartheta_1^2/(4\vartheta_2)$ and, hence, $\Gamma=\tilde{\Gamma}=0$ for the remainder of the proof.

Statement (ii) follows from the fact that $\mathrm{Cov}(X_{t_0}(x),X_{t_0}(y))$ only depends on $\left(\frac{\sigma^2}{\vartheta_2},\kappa\right)$ in view of the Gaussianity of X.

For (iii), note that $t\mapsto X_t(x_0)$ is a stationary Gaussian process with covariance function

$$\rho(t):=\sigma^2\sum_{k\ge1}\frac{\mathrm{e}^{-\lambda_k t}}{2\lambda_k}e_k^2(x_0).$$

Let

$$f_{(\sigma^2,\vartheta_2)}(u):=\frac{1}{2\pi}\int\mathrm{e}^{-iut}\rho(|t|)\,dt=\frac{1}{\pi}\int_0^\infty\cos(ut)\rho(t)\,dt=\frac{\sigma^2}{2\pi}\sum_{\ell\ge1}\frac{e_\ell^2(x_0)}{\lambda_\ell^2+u^2}$$

be the spectral density of $t\mapsto X_t(x_0)$. By Theorem 17 and its preceding discussion in Ibragimov and Rozanov [42, Chapter III], it suffices to show

$$\exists r>1:\qquad\lim_{u\to\infty}u^r f_{(\sigma^2,\vartheta_2)}(u)\in(0,\infty)\tag{2.23}$$

and

$$\frac{f_{(\sigma^2,\vartheta_2)} - f_{(\tilde\sigma^2,\tilde\vartheta_2)}}{f_{(\sigma^2,\vartheta_2)}} \in L^2(\mathbb{R}). \tag{2.24}$$

To prove these statements, we may assume $\kappa = 0$ without loss of generality. Set

$$h_{(\sigma^2,\vartheta_2)}(z) := \frac{\sigma^2}{\pi(\pi^4\vartheta_2^2 z^4 + 1)}, \quad z \in \mathbb{R}.$$

By Lemma 2.4.10 (ii), we have

$$\begin{aligned} f_{(\sigma^2,\vartheta_2)}(u) &= \frac{1}{u^2}\sum_{\ell\geq 1} h_{(\sigma^2,\vartheta_2)}\left(\frac{\ell}{\sqrt{u}}\right)\sin^2(\pi\ell x_0) \\ &= \frac{1}{u^2}\left(\frac{\sqrt{u}}{2}\int_0^\infty h_{(\sigma^2,\vartheta_2)}(z)\,dz + \mathcal{O}\left(\frac{1}{\sqrt{u}}\right)\right), \qquad u\to\infty, \end{aligned}$$

which proves (2.23). Now, if $\sigma^2/\sqrt{\vartheta_2} = \tilde\sigma^2/\sqrt{\tilde\vartheta_2}$, a substitution shows that

$$\int_0^\infty h_{(\sigma^2,\vartheta_2)}(z)\,dz = \int_0^\infty h_{(\tilde\sigma^2,\tilde\vartheta_2)}(z)\,dz.$$

Therefore, we obtain

$$\frac{f_{(\sigma^2,\vartheta_2)}(u) - f_{(\tilde\sigma^2,\tilde\vartheta_2)}(u)}{f_{(\sigma^2,\vartheta_2)}(u)} = \mathcal{O}\left(\frac{1}{u}\right), \quad u\to\infty,$$

which implies (2.24). □

Before we prove our theorems on lower bounds (Theorems 2.1.2 and 2.1.7), we verify their ingredients, Proposition 2.1.4 and Proposition 2.1.6. Further auxiliary results can be found in Section 2.4.4.

Proof of Proposition 2.1.4. By setting $a = k^2$, $\mu = \pi^2\vartheta_2$ and $\nu^2 = \frac{\sigma^2}{\pi^2\vartheta_2}$ in Lemma 2.4.2 and using independence of $(u_\ell,\ \ell\in\mathbb{N})$, we obtain the Fisher information matrix I for the parameters (μ,ν^2), namely

$$\begin{aligned} I_{11} &= N\sum_{\ell=1}^M \frac{\ell^4\Delta^2(\mathrm{e}^{-4\mu\ell^2\Delta} + \mathrm{e}^{-2\mu\ell^2\Delta})}{(1-\mathrm{e}^{-2\mu\ell^2\Delta})^2} = N\sum_{\ell=1}^M g_{11}(\ell\sqrt{\Delta}), \qquad g_{11}(x) := \frac{x^4(\mathrm{e}^{-4\mu x^2} + \mathrm{e}^{-2\mu x^2})}{(1-\mathrm{e}^{-2\mu x^2})^2}, \\ I_{12} &= N\sum_{\ell=1}^M \frac{\ell^2\Delta\mathrm{e}^{-2\mu\ell^2\Delta}}{\nu^2(1-\mathrm{e}^{-2\mu\ell^2\Delta})} = N\sum_{\ell=1}^M g_{12}(\ell\sqrt{\Delta}), \qquad g_{12}(x) := \frac{x^2\mathrm{e}^{-2\mu x^2}}{\nu^2(1-\mathrm{e}^{-2\mu x^2})}, \\ I_{22} &= \frac{(N+1)M}{2\nu^4}. \end{aligned}$$

The Fisher information matrix $J = J_{M,N}$ for the parameters (σ^2,ρ^2) can be computed via the change of variables formula $J = A^\top I A$ where

$$A = \begin{pmatrix} \pi^2/\rho^2 & -\pi^2\sigma^2/\rho^4 \\ 0 & 1/\pi^2 \end{pmatrix}$$

is the Jacobian of the function transforming (σ^2,ρ^2) to (μ,ν^2). Hence, the diagonal entries of J are given by

$$J_{11} = \frac{\pi^4}{\rho^4}I_{11}, \qquad J_{22} = \frac{\pi^4\sigma^4}{\rho^8}I_{11} - \frac{2\sigma^2}{\rho^4}I_{12} + \frac{1}{\pi^4}I_{22}.$$

If $M\sqrt{\Delta}$ is bounded away from 0, then I_{11} can be interpreted as a Riemann sum. We obtain

$$J_{11} \approx I_{11} \approx N^{3/2} \int_0^{M\sqrt{\Delta}} g_{11}(x)\, dx \approx N^{3/2}.$$

On the other hand, if $M\sqrt{\Delta} \to 0$, it follows from Lemma 2.4.11 and $g_{11}(0) = \frac{1}{2\mu^2} = \frac{\rho^4}{2\pi^4\sigma^4}$, $g_{12}(0) = \frac{1}{2\mu\nu^2} = \frac{1}{2\sigma^2}$ as well as $g_{11}'(0) = g_{12}'(0) = 0$ that

$$\begin{aligned}
I_{11} &= N^{3/2}(M\sqrt{\Delta} g_{11}(0) + \frac{M^2\Delta}{2} g_{11}'(0) + \mathcal{O}(M^3\Delta^{3/2})) = \frac{\rho^4}{2\pi^4\sigma^4} NM + \mathcal{O}(M^3),\\
I_{12} &= N^{3/2}(M\sqrt{\Delta} g_{12}(0) + \frac{M^2\Delta}{2} g_{12}'(0) + \mathcal{O}(M^3\Delta^{3/2})) = \frac{NM}{2\sigma^2} + \mathcal{O}(M^3),\\
I_{22} &= \frac{\pi^4}{2\rho^4} MN + \mathcal{O}(M).
\end{aligned}$$

Therefore, the leading terms in J_{22} cancel and, consequently, $J_{22} = \mathcal{O}(M^3)$. □

Proof of Proposition 2.1.6. For a discrete time, centered, stationary Gaussian process $(Z_j)_{j\in\mathbb{Z}}$ whose covariance function depends on an unknown parameter $\theta \in \mathbb{R}$, we denote the Fisher information of a sample $(Z_0, \ldots, Z_{n-1})$ with respect to θ by $I_n(Z)$. A particularly useful result to calculate $I_n(Z)$ for the above class of Gaussian processes is given by Whittle [78]:

$$\lim_{n\to\infty} \frac{1}{n} I_n(Z) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\frac{\partial}{\partial\theta}\phi_\theta(\omega)}{\phi_\theta(\omega)} \right)^2 d\omega, \qquad n \to \infty, \tag{2.25}$$

where

$$\phi(\omega) = \sum_{j\in\mathbb{Z}} \mathbf{E}[Z_0 Z_j] \mathrm{e}^{-ij\omega}, \qquad \omega \in [-\pi, \pi],$$

is the spectral density of Z.

Setting $\theta = \pi^2\vartheta_2$, (2.25) cannot be directly applied to the process $Z = \bar{U}_k$, for $1 \le k \le M-1$: $\bar{U}_k$ arises from high-frequency increments of the continuous time process U_k and, thus, cannot be regarded as time series. Indeed, the spectral density Φ_k^Δ of $\bar{U}_k$ hinges on $\Delta = 1/N$ and, therefore, even for large N, $I_N(\bar{U}_k)/N$ is not necessarily close to the asymptotic Fisher information defined in (2.25).

To circumvent this difficulty, consider the N-th order Fourier approximation to Φ_k^Δ:

$$\Phi_k^{N,\Delta}(\omega) = \sum_{j=1-N}^{N-1} \mathbf{E}[\bar{U}_k(0)\bar{U}_k(j)] \mathrm{e}^{-ij\omega} \ge 0, \qquad \omega \in [-\pi, \pi]. \tag{2.26}$$

Lemma 2.4.4(i) verifies that $\Phi_k^{N,\Delta}$ is positive. Therefore, there exists a stationary Gaussian process $Y_k = (Y_k(j))_{j\in\mathbb{Z}}$ with spectral density $\Phi_k^{N,\Delta}$, see, e.g., [12]. Clearly,

$$(Y_k(j), \ldots, Y_k(j+N-1))) \stackrel{\mathcal{D}}{=} \left(\bar{U}_k(0), \ldots, \bar{U}_k(N-1)\right), \qquad j \in \mathbb{N}_0,$$

and $(Y_k(j), \ldots, Y_k(j+N-1)))$ is independent of $(Y_k(h), \ldots, Y_k(h+N-1)))$ whenever $|j-h| > 2N$. Consequently, it is possible to extract L independent copies of $\left(\bar{U}_k(0), \ldots, \bar{U}_k(N-1)\right)$ from a sample $(Y_k(0), \ldots, Y_k(2NL-1))$ for any $L \in \mathbb{N}$. Now, using the fact that a statistic never has larger information than the data from which it is constructed (cf. [43, Theorem I.7.2]) yields

$$L \cdot I_N(\bar{U}_k) \le I_{2NL}(Y_k). \tag{2.27}$$

For fixed $\Delta = 1/N$, we can now apply Whittle's formula (2.25) for $L \to \infty$: For each $\varepsilon > 0$ we can choose $L \in \mathbb{N}$ such that

$$I_{2NL}(Y_k) \leq 2NL(1+\varepsilon)\mathscr{I}_k, \tag{2.28}$$

where

$$\mathscr{I}_k^{N,\Delta} := \frac{1}{4\pi}\int_{-\pi}^{\pi} S^2(\omega)\,d\omega, \qquad S := \frac{\partial}{\partial\vartheta_2}\log\Phi_k^{N,\Delta}.$$

By combining (2.27) and (2.28), we get $I_N(\bar{U}_k) \leq 2N\mathscr{I}_k$. Proving below that

$$\mathscr{I}_k^{N,\Delta} \lesssim M^2\Delta\log\frac{1}{M^2\Delta} \tag{2.29}$$

holds uniformly in $k = 0,\ldots,M-1$, we obtain $I_N(\bar{U}_k) \lesssim M^2\log\frac{1}{M^2\Delta}$ and the results follows by independence of the processes $\bar{U}_1,\ldots,\bar{U}_{M-1}$.

In order to verify (2.29), we only have to consider the integral over $[0,\pi]$, by symmetry. From Lemma 2.4.4, we can deduce for $\omega \geq k^2\Delta$ that

$$S(\omega) \lesssim \begin{cases} \frac{M\sqrt{\Delta}}{\sqrt{\omega}}, & \omega \geq M^2\Delta, \\ 1, & \omega \in [k^2\Delta, M^2\Delta] \end{cases}$$

and, hence,

$$\int_{k^2\Delta}^{\pi} S^2(\omega)\,d\omega \lesssim M^2\Delta\log\frac{1}{M^2\Delta}.$$

For $\omega \leq k^2\Delta$, Lemma 2.4.4 gives

$$S(\omega) \lesssim \frac{\frac{\omega^2}{k^4\Delta^2} + k^2\mathrm{e}^{-\theta k^2}}{\frac{\omega^2}{k^4\Delta^2} + \mathrm{e}^{-\theta k^2}}.$$

Since

$$\int_0^1 \frac{d\omega}{(\omega^2+\mathrm{e}^{-\theta k^2})^2} \leq \int_0^{\mathrm{e}^{-\theta k^2/2}} \frac{1}{\mathrm{e}^{-2\theta k^2}}\,d\omega + \int_{\mathrm{e}^{-\theta k^2/2}}^1 \frac{1}{\omega^4}\,d\omega \lesssim \exp\left(\frac{3}{2}\theta k^2\right),$$

a substitution yields

$$\begin{aligned} \int_0^{k^2\Delta} S^2(\omega)\,d\omega &\lesssim k^2\Delta\int_0^1 \left(\frac{\omega^2+k^2\mathrm{e}^{-\theta k^2}}{\omega^2+\mathrm{e}^{-\theta k^2}}\right)^2 d\omega \\ &\lesssim M^2\Delta\left(1 + k^4\mathrm{e}^{-2\theta k^2}\int_0^1 \frac{d\omega}{(\omega^2+\mathrm{e}^{-\theta k^2})^2}\right) \lesssim M^2\Delta \end{aligned}$$

and the proof is finished. □

We can now conclude the main lower bounds.

Proof of Theorem 2.1.2. As follows from the discussion subsequent to Theorem 2.1.2, it suffices to show that for each sampling regime there is a reparametrization (γ_1,γ_2) of (σ^2,ϑ_2) such that the corresponding Fisher information satisfies $J_{M,N}(\gamma_2) \lesssim r_{M,N}$ locally uniformly. Inspection of the proofs of Propositions 2.1.4 and 2.1.6 shows that the bounds on the Fisher information are indeed locally uniform.

(ii) Case $M/\sqrt{N} \gtrsim 1$. For $L \in \mathbb{N}$ define the process X^L via $X_t^L(y) = \sum_{\ell=1}^L u_\ell(t)e_\ell(y)$, $t \geq 0$, $y \in [0,1]$, and let $\mathcal{X}_{N,M}^L = \{X_{t_i}^L(y_k),\, i = 0,\ldots,N-1,\, k = 0,\ldots,M\}$ as well as $\mathcal{X}_{N,M} = \mathcal{X}_{N,M}^\infty$. Denoting the corresponding covariance matrices by $\Sigma_{N,M}^L$ and $\Sigma_{N,M}$ and using the bound on the total variation distance of Gaussian distributions due to Devroye et al. [30] (see also (3.4)), we obtain $\mathrm{TV}(\mathcal{N}(0,\Sigma_{N,M}),\mathcal{N}(0,\Sigma_{N,M}^L)) \leq \frac{3}{2}\|\Sigma_{N,M}^{-1/2}(\Sigma_{N,M}^L - \Sigma_{N,M})\Sigma_{N,M}^{-1/2}\|_F \leq \frac{3}{2}\|\Sigma_{N,M}^{-1/2}\|_F^2\|\Sigma_{N,M}^L - \Sigma_{N,M}\|_F$.

Consequently, we can pick a sequence $L_{N,M} \to \infty$ such that $\mathcal{X}_{N,M}^{L_{N,M}}$ and $\mathcal{X}_{N,M}$ are statistically equivalent in the sense of Le Cam and it is sufficient to derive a lower bound for $\mathcal{X}_{N,M}^{L_{N,M}}$, or even $\{u_\ell(t_i),\, i \leq N, \ell \leq L_{N,M}\}$. Assuming $L_{N,M} \geq M$ without loss of generality, for this observation scheme Proposition 2.1.4 yields under the parametrization $(\sigma^2/\vartheta_2, \sigma^2)$ that

$$J_{M,N}(\sigma^2) \lesssim N^{3/2} \wedge L_{N,M}^3 = N^{3/2} = r_{N,M}.$$

Case $M/\sqrt{N} \to 0$. For $b \in \mathbb{Q} \cap (0, 1/2)$, write $b = p/q$ where $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ such that $y_k = \frac{pM+k(q-2p)}{qM}$, $k \leq M$, and, consequently, $\{y_k,\, k = 0, \ldots, M\}$ is a subset of $\{z_k,\, k = 1, \ldots, qM - 1\}$ where $z_k = \frac{k}{qM}$. Now, $qM\sqrt{\Delta} \to 0$ and since $q^3M^3 \log\left(\frac{1}{q^2M^2\Delta}\right) \lesssim M^3 \log\left(\frac{1}{M^2\Delta}\right)$ Proposition 2.1.6 implies under the parametrization $(\sigma^2/\sqrt{\vartheta_2}, \vartheta_2)$ that

$$J_{M,N}(\vartheta_2) \lesssim M^3 \log(\frac{1}{M^2\Delta}) = r_{N,M}.$$

(i) If $\min(M, N)$ remains finite and $M/\sqrt{N} \gtrsim 1$, then N necessarily remains finite and the result follows from (ii). On the other hand, if $M/\sqrt{N} \to 0$, then M must remain finite. Like in the proof of (ii), extend the set of spatial locations to $\{z_k,\, k < qM\}$ and consider the corresponding processes U_k, $k = 1, \ldots, qM-1$ from (2.4). A similar calculation as in the proof of Proposition 2.1.1 shows that for any $k < qM$, the laws of the independent continuous processes $\{U_k(t),\, t \leq 1\}$ are absolutely continuous for different parameter values (σ^2, ϑ_2) and $(\tilde{\sigma}^2, \tilde{\vartheta}_2)$ as long as $\sigma^2/\sqrt{\vartheta_2} = \tilde{\sigma}^2/\sqrt{\tilde{\vartheta}_2}$ and, hence, consistent estimation of (σ^2, ϑ_2) based on continuous or discrete observations is impossible: Note that the spectral density of the time-continuous process U_k is $f_k(u) = \frac{1}{2u^2}\sum_{\ell \in \mathcal{I}_k} h_{(\sigma^2,\vartheta_2)}\left(\frac{\ell}{\sqrt{|u|}}\right)$, $u \in \mathbb{R}$, where $h_{(\sigma^2,\vartheta_2)}$ is defined in the proof of Proposition 2.1.1. Now, a Riemann sum midpoint approximation, see Lemma 2.4.10, shows that

$$f_k^+(u) := \frac{1}{2u^2}\sum_{\ell \geq 0} h_{(\sigma^2,\vartheta_2)}\left(\frac{k + 2M\ell}{\sqrt{u}}\right) = \frac{1}{2u^2}\left(\frac{\sqrt{u}}{2M}\int_{(k-M)/\sqrt{u}}^{\infty} h_{(\sigma^2,\vartheta_2)}(z)\,dz + \mathcal{O}\left(\frac{1}{\sqrt{u}}\right)\right),$$

$$f_k^-(u) := \frac{1}{2u^2}\sum_{\ell \geq 0} h_{(\sigma^2,\vartheta_2)}\left(\frac{2M - k + 2M\ell}{\sqrt{u}}\right) = \frac{1}{2u^2}\left(\frac{\sqrt{u}}{2M}\int_{(M-k)/\sqrt{u}}^{\infty} h_{(\sigma^2,\vartheta_2)}(z)\,dz + \mathcal{O}\left(\frac{1}{\sqrt{u}}\right)\right),$$

as $u \to \infty$. Since $h_{(\sigma^2,\vartheta_2)}$ is symmetric around 0 we obtain

$$f_k(u) = f_k^+(u) + f_k^-(u) = \frac{1}{u^2}\left(\frac{\sqrt{u}}{2M}\int_0^{\infty} h_{(\sigma^2,\vartheta_2)}(z)\,dz + \mathcal{O}\left(\frac{1}{\sqrt{u}}\right)\right),$$

from which equivalence follows as in Proposition 2.1.1. □

Proof of Theorem 2.1.7. The proof of (i) can be done in exactly the same way as for a finite time horizon. To prove (ii), set $a = 1$, $\sigma^2 = \nu^2\mu$ and $\mu = \lambda_\ell = \pi^2\vartheta_2\ell^2 + \vartheta_1^2/(4\vartheta_2) - \vartheta_0$ in Lemma 2.4.2. Applying the transformation rule for the Fisher information like in the proof of Proposition 2.1.4, yields that the Fisher information for ϑ_0 of the sample $(u_\ell(t_i),\, i \leq N, \ell \leq L)$ is given by

$$J_{N,\Delta,L}(\vartheta_0) = T\sum_{\ell=1}^{L} \frac{1}{\lambda_\ell} g(\lambda_\ell\Delta) + \sum_{\ell=1}^{L} \frac{1}{2\lambda_\ell^2}$$

with

$$g(x) := \frac{2x^2(\mathrm{e}^{-4x} + \mathrm{e}^{-2x}) - 4x(\mathrm{e}^{-2x} - \mathrm{e}^{-4x}) + (1 - \mathrm{e}^{2x})^2}{2x(1 - \mathrm{e}^{-2x})^2}, \qquad x > 0.$$

Insertion of the Taylor approximation of second order for the exponential function shows that g is bounded on $\mathbb{R}_+$ and, hence, $J_{N,\Delta,L}(\vartheta_0) \lesssim T$ holds uniformly in $L \in \mathbb{N}$. From here, the result follows as in the proof of Theorem 2.1.2. □

2.4.2 Proofs for the central limit theorems for realized quadratic variations

We first prove the generic central limit result in Proposition 2.2.1. Afterwards, we can verify the central limit theorems for realized quadratic variations based on spatial increments (Theorem 2.2.3) and double increments (Theorem 2.2.7).

Proof of Proposition 2.2.1. Since $\Sigma_n = Q_n^\top \Lambda_n Q_n$ for an orthogonal matrix $Q_n \in \mathbb{R}^{d_n \times d_n}$ and a diagonal matrix Λ_n, the vector $Z_{\bullet,n}$ has the same distribution as $B_n X^n$ for $B_n := Q_n^T \Lambda^{1/2}$ and $X^n := (X_1, \dots, X_{d_n})$ with independent standard normal random variables $(X_k)_{k\in\mathbb{N}}$. Denoting $A_n = \operatorname{diag}(\alpha_{1,n}, \dots, \alpha_{d_n,n})$, we obtain $S_n = Z_{\bullet,n}^\top A_n Z_{\bullet,n} \overset{\mathcal{D}}{=} X^{n\top} B_n^\top A_n B_n X^n$. Furthermore, $B_n^\top A_n B_n$ is symmetric such that $B_n^\top A_n B_n = P_n^\top \Gamma_n P_n$ where P_n is an orthogonal matrix and Γ_n is a diagonal matrix. Since $P_n X^n \sim \mathcal{N}(0, E_{d_n})$, we conclude as in Mathai and Provost [62, p. 36] that

$$S_n \overset{\mathcal{D}}{=} X^{n\top} B_n^\top A_n B_n X^n = (P_n X^n)^T \Gamma (P_n X^n) \overset{\mathcal{D}}{=} X^{n\top} \Gamma_n X^n = \sum_{i=1}^{d_n} \gamma_{i,n} X_i^2$$

where $\gamma_{i,n}$, $i \le d_n$ are the eigenvalues of $B_n^\top A_n B_n$. The statement now follows by Lyapunov's condition and $\|B_n\|_2^2 = \|\Sigma_n\|_2$:

$$\begin{aligned}\frac{\sum_{i=1}^{d_n} \gamma_{i,n}^4 \mathbf{E}\left(\left(X_k^2 - \mathbf{E}X_k^2\right)^4\right)}{(\operatorname{Var} S_n)^2} &\approx \frac{\sum_{i=1}^{d_n} \gamma_{i,n}^4}{\left(\sum_{i=1}^{d_n} \gamma_{i,n}^2\right)^2} \lesssim \frac{\max_{i\le d_n} \gamma_{i,n}^2}{\sum_{i=1}^{d_n} \gamma_{i,n}^2} = \frac{\|B_n^T A_n B_n\|_2^2}{\operatorname{Var} S_n} \\ &\le \frac{(\|B_n\|_2^2 \|A_n\|_2)^2}{\operatorname{Var} S_n} = \frac{\|\Sigma\|_2^2}{\operatorname{Var} S_n}.\end{aligned}$$

□

Throughout, we use the notation

$$\begin{aligned} D_\delta f(x) &:= f(x+\delta) - f(x), \\ D_\delta^2 f(x) &:= f(x+2\delta) - 2f(x+\delta) + f(x) \end{aligned}$$

for a function $f : \mathbb{R} \to \mathbb{R}$. We now prove the central limit theorem for space increments.

Proof of Theorem 2.2.3. In Steps 1-3 of this proof we show the central limit theorem under the assumption that X_0 follows the stationary distribution. The case $X_0 = 0$ is treated in Step 4. We abbreviate the (rescaled) space increments by

$$S_{ik} := (\delta_k^M X)(t_i) \quad \text{and} \quad \tilde{S}_{ik} := \mathrm{e}^{\kappa y_k/2} (\delta_k^M X)(t_i).$$

Step 1. We calculate the asymptotic mean of V_{sp}: Application of the trigonometric identity 1.15 yields

$$\begin{aligned} &\mathrm{e}^{\kappa x/2}(e_\ell(x+\delta) - e_\ell(x))\mathrm{e}^{\kappa y/2}(e_\ell(y+\delta) - e_\ell(y)) \\ &= 2\left(\mathrm{e}^{-\frac{\kappa}{2}\delta}\sin(\pi\ell(x+\delta)) - \sin(\pi\ell x)\right)\left(\mathrm{e}^{-\frac{\kappa}{2}\delta}\sin(\pi\ell(y+\delta)) - \sin(\pi\ell y)\right) \\ &= g(\delta)\left(2\cos(\pi\ell(y-x)) - \cos(\pi\ell(y-x-\delta)) - \cos(\pi\ell(y-x+\delta))\right) \\ &\quad + (g(2\delta) + g(0) - 2g(\delta))(\cos(\pi\ell(y-x))) \\ &\quad + 2g(\delta)\cos(\pi\ell(y+x+\delta)) - g(0)\cos(\pi\ell(y+x)) - g(2\delta)\cos(\pi\ell(x+y+2\delta)) \end{aligned} \tag{2.30}$$

where $g(x) = \exp(-\kappa x/2)$. Plugging in $x = y$ gives

$$\begin{aligned} &\mathrm{e}^{\kappa y}(e_\ell(y+\delta) - e_\ell(y))^2 \\ &= 2g(\delta)(1 - \cos(\pi\ell\delta)) + (g(2\delta) + g(0) - 2g(\delta)) \\ &\quad + 2g(\delta)\cos(\pi\ell(2y+\delta)) - g(0)\cos(2\pi\ell y) - g(2\delta)\cos(2\pi\ell(y+\delta)) \\ &= 2(1 - \cos(\pi\ell\delta)) + 2(1 - g(\delta))(\cos(\pi\ell\delta) - 1) + (g(2\delta) + g(0) - 2g(\delta)) \\ &\quad + 2g(\delta)\cos(\pi\ell(2y+\delta)) - g(2\delta)\cos(2\pi\ell(y+\delta)) - g(0)\cos(2\pi\ell y). \end{aligned} \tag{2.31}$$

Writing

$$f(y) := \sum_{\ell\geq 1} \frac{1}{2\lambda_\ell}\cos(\pi\ell y),$$

we, thus, have

$$\begin{aligned}\mathbf{E}\left(\mathrm{e}^{\kappa y}\left(X_t(y+\delta)-X_t(y)\right)^2\right) &= \sigma^2 \sum_{\ell\geq 0}\frac{1}{2\lambda_\ell}\mathrm{e}^{\kappa y}\left(e_\ell(y+\delta)-e_\ell(y)\right)^2 \\ &= \sigma^2\left(-2D_\delta f(0) - 2D_\delta g(0)D_\delta f(0) + f(0)D_\delta^2 g(0) - D_\delta^2(g(\cdot)f(2y+\cdot))(0)\right).\end{aligned}$$

Owing to its closed form expression in (1.16), we see that $f\in C_b^\infty([0,2])$ and $f'(0) = -\frac{1}{4\vartheta_2}$. Hence,

$$\mathbf{E}\left(\mathrm{e}^{\kappa y}(X_t(y+\delta)-X_t(y))^2\right) = -2\sigma^2 f'(0)\cdot\delta + \mathcal{O}(\delta^2) = \frac{\sigma^2}{2\vartheta_2}\cdot\delta + \mathcal{O}(\delta^2).$$

For $y=y_k$ we obtain the asymptotic mean

$$\mathbf{E}(V_{\mathrm{sp}}) = \frac{\sigma^2}{2\vartheta_2} + \mathcal{O}(\delta)$$

and, in particular, under the condition $N/M\to 0$,

$$\sqrt{MN}\left(V_{\mathrm{sp}} - \frac{\sigma^2}{2\vartheta_2}\right) = \sqrt{MN}(V_{\mathrm{sp}} - \mathbf{E}(V_{\mathrm{sp}})) + o(1).$$

Step 2. We calculate the asymptotic variance: Recall the relation

$$\mathrm{Cov}((\tilde{S}_{ik})^2,(\tilde{S}_{jl})^2) = 2\,\mathrm{Cov}(\tilde{S}_{ik},\tilde{S}_{jl})^2$$

from (2.7). Together with the symmetry $\mathrm{Cov}(\tilde{S}_{ik},\tilde{S}_{jl}) = \mathrm{Cov}(\tilde{S}_{jk},\tilde{S}_{il})$, this implies

$$\mathrm{Var}(V_{\mathrm{sp}}) = \frac{2}{N^2M^2\delta^2}(v_1+v_2+v_3+v_4)$$

where

$$v_1 := \sum_{i=0}^{N-1}\sum_{k=0}^{M-1}\mathrm{Var}(\tilde{S}_{ik})^2, \qquad v_2 := 2\sum_{i=0}^{N-2}\sum_{j=i+1}^{N-1}\sum_{k=0}^{M-1}\mathrm{Cov}(\tilde{S}_{ik},\tilde{S}_{jk})^2$$

$$v_3 := 2\sum_{i=0}^{N-1}\sum_{k=0}^{M-2}\sum_{l=k+1}^{M-1}\mathrm{Cov}(\tilde{S}_{ik},\tilde{S}_{il})^2, \qquad v_4 := 4\sum_{i=0}^{N-2}\sum_{j=i+1}^{N-1}\sum_{k=0}^{M-2}\sum_{l=k+1}^{M-1}\mathrm{Cov}(\tilde{S}_{ik},\tilde{S}_{jl})^2.$$

We have already shown that $\mathrm{Var}(\tilde{S}_{ik}) = \mathbf{E}((\tilde{S}_{ik})^2) = \frac{\sigma^2}{2\vartheta_2}\cdot\delta + \mathcal{O}(\delta^2)$. Therefore,

$$v_1 = NM\delta^2\cdot\frac{\sigma^4}{4\vartheta_2^2} + \mathcal{O}\left(\frac{N}{M^2}\right) = NM\delta^2\cdot\frac{\sigma^4}{4\vartheta_2^2} + o\left(\frac{N}{M}\right).$$

In the sequel, we show that the remaining covariances do not contribute to the asymptotic variance. For v_2, we define $\omega := \vartheta_2(\pi^2\wedge(\pi^2+\Gamma))>0$ such that $\lambda_\ell\geq\omega\ell^2$ for all $\ell\in\mathbb{N}$. Since $(e_\ell(y_{k+1})-e_\ell(y_k))^2 \lesssim \ell^2\delta^2$, we get for $J:=|i-j|\geq 1$ that

$$\begin{aligned}\mathrm{Cov}(\tilde{S}_{ik},\tilde{S}_{jk}) &= \sigma^2\sum_{\ell\geq 1}\frac{\mathrm{e}^{-\lambda_\ell J\Delta}}{2\lambda_\ell}\mathrm{e}^{\kappa y_k}\left(e_\ell(y_{k+1})-e_\ell(y_k)\right)^2 \lesssim \sum_{\ell\geq 1}\frac{\mathrm{e}^{-\lambda_\ell J\Delta}}{2\lambda_\ell}\ell^2\delta^2 \\ &\lesssim \delta^2\sum_{\ell\geq 1}\mathrm{e}^{-\omega\ell^2 J\Delta} \lesssim \frac{\delta^2}{\sqrt{J\Delta}}\end{aligned}$$

where the last step follows by Riemann summation with mesh size $\sqrt{J\Delta}$. Since $\frac{\log N}{M^2\Delta} \leq \frac{N}{M^2\Delta} = \frac{N^2}{M^2}\frac{1}{T} \to 0$,

$$v_2 \lesssim \frac{M\delta^4}{\Delta}\sum_{i=0}^{N-1}\sum_{j=i+1}^{N}\frac{1}{(j-i)} \leq \frac{NM\delta^4}{\Delta}\sum_{i=1}^{N}\frac{1}{i} = \mathcal{O}\left(\frac{N\log N}{M^3\Delta}\right) = o\left(\frac{N}{M}\right).$$

To bound v_3, we follow the same strategy as for the mean: since (2.30) consists exclusively of second order differences we have $\mathrm{Cov}(\tilde{S}_{ik}, \tilde{S}_{il}) = \mathcal{O}(\delta^2)$ for $k \neq l$. Therefore, $v_3 = \mathcal{O}(NM^2\delta^4) = o(N/M)$. To estimate v_4, we deduce from (2.30) for $k < l$ and $J = |i-j| \geq 1$ that

$$\begin{aligned}\mathrm{Cov}(\tilde{S}_{ik}, \tilde{S}_{jl}) = &- g(\delta)D_\delta^2 f_{J\Delta}\left(y_l - y_{k+1}\right)\\ &+ f_{J\Delta}\left(y_l - y_k\right) D_\delta^2 g(0) - D_\delta^2\left(g(\cdot)f_{J\Delta}\left(y_l + y_k + \cdot\right)\right)(0)\end{aligned}$$

where

$$f_t(y) := \sigma^2\sum_{\ell\geq 1}\frac{\mathrm{e}^{-\lambda_\ell t}}{2\lambda_\ell}\cos(\pi\ell y).$$

By Riemann summation we have $f_t''(y) \lesssim \sum_{\ell\geq 1}\mathrm{e}^{-\lambda_\ell t} \lesssim \frac{1}{\sqrt{t}}$. On the other hand, by Lemma 2.4.8,

$$f_t''(y) \lesssim \frac{1}{y\wedge(2-y)}\sup_k\left|\frac{k^2}{\lambda_k}\mathrm{e}^{-\lambda_k t}\right| \lesssim \frac{1}{y\wedge(2-y)}.$$

Therefore,

$$f_t''(y) \lesssim B(t,y) := \frac{1}{y\wedge(2-y)}\wedge\frac{1}{\sqrt{t}}.$$

Similarly, $f_t(y)$, $f_t'(y) \lesssim B(t,y)$ can be shown. We conclude

$$\begin{aligned}v_4 &\lesssim NM\sum_{i=0}^{N-1}\sum_{k=0}^{2M-2}\delta^4 B\left(i\Delta, \frac{k}{M}\right)^2 \lesssim \frac{N}{M^3}\sum_{i=1}^{N}\sum_{k=0}^{M}\frac{M^2}{k^2}\wedge\frac{1}{i\Delta}\\ &= \frac{N}{M^3}\sum_{i=1}^{N}\left(\sum_{k<M\sqrt{i\Delta}}\frac{1}{i\Delta} + \sum_{M\geq k\geq M\sqrt{i\Delta}}\frac{M^2}{k^2}\right) \lesssim \frac{N}{M^3}\sum_{i=1}^{N}\frac{M}{\sqrt{i\Delta}} \lesssim \frac{N^{3/2}}{M^2\sqrt{\Delta}} = o\left(\frac{N}{M}\right)\end{aligned}$$

where the last step follows from $\frac{\sqrt{N}}{M\sqrt{\Delta}} = \frac{N}{M}\frac{1}{\sqrt{T}} \to 0$. Summing up, we have proved that

$$\mathrm{Var}(V_{\mathrm{sp}}) = \frac{\sigma^4}{2\vartheta_2^2}\cdot\frac{1}{MN} + o\left(\frac{1}{NM}\right).$$

Step 3. To prove asymptotic normality, we interpret the number of temporal and spatial observations as sequences $M = M_n$, $N = N_n$ indexed by $n \in \mathbb{N}$ and consider the triangular array $(Z_{ik}^n,\ k < M_n, i < N_n, n \in \mathbb{N})$ where $Z_{ik}^n := \tilde{S}_{ik}/\sqrt{NM}\delta$. Since $\mathrm{Var}(\sum_{i,k}(Z_{ik}^n)^2) \approx (MN)^{-1}$, Proposition 2.2.1 applies if

$$\frac{1}{MN\delta^2}\Big(\sum_{i,k}|\mathrm{Cov}(\tilde{S}_{ik}, \tilde{S}_{jl})|\Big)^2 \to 0$$

holds uniformly in $j < N, l < M$, in view of criterion (2.8). The covariance bounds in Step 2 yield uniformly in $j < N, k < M$ that

$$\begin{aligned}\sum_{k<M}|\mathrm{Cov}(\tilde{S}_{jk}, \tilde{S}_{jl})| &= \mathcal{O}(\delta),\\ \sum_{i<N}|\mathrm{Cov}(\tilde{S}_{il}, \tilde{S}_{jl})| &= \mathcal{O}(\delta^2\sqrt{N}/\sqrt{\Delta}),\end{aligned}$$

$$\Big(\sum_{i\neq j,k\neq l}|\mathrm{Cov}(\tilde S_{ik},\tilde S_{jl})|\Big)^2\lesssim MN\sum_{i\neq j,k\neq l}|\mathrm{Cov}(\tilde S_{ik},\tilde S_{jl})|^2=o(N/M)$$

where we have used the Cauchy-Schwarz inequality to obtain the last bound. It remains to note $N/M\to 0$ and $N\Delta\gtrsim 1$.

Step 4. To show that the central limit theorem also holds for the vanishing initial condition, let X_t^0 be the process (1.12) with $\xi=0$ and $\xi_t:=S(t)\xi$, where ξ follows the stationary initial condition and is independent of X^0. For these two processes, denote the rescaled space increments by

$$\tilde S_{ik}^0:=\mathrm{e}^{\kappa y_k/2}(\delta_k^M X^0)(t_i),\qquad \tilde I_{ik}:=\mathrm{e}^{\kappa y_k/2}(\xi_{t_i}(y_{k+1})-\xi_{t_i}(y_k)).$$

We show that $\frac{1}{NM\delta}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\big((\tilde S_{ik}^0)^2-\tilde S_{ik}^2\big)=o_p(1/\sqrt{MN})$, then the result for zero initial condition follows from the result for stationary initial condition in view of Slutsky's lemma. To that aim, we expand

$$\frac{1}{NM\delta}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\big((\tilde S_{ik}^0)^2-\tilde S_{ik}^2\big)=\frac{1}{NM\delta}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\tilde I_{ik}^2-\frac{2}{NM\delta}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\tilde S_{ik}^0\tilde I_{ik}=:A_1+A_2$$

and show that $\mathbf{E}(|A_1|)=o(1/\sqrt{MN})$ and $\mathbf{E}(A_2^2)=o(1/(MN))$. First of all, using the mean value theorem and the Riemann sum argument,

$$\mathbf{E}(\tilde I_{ik}^2)=\sum_{\ell\geq 1}\frac{\sigma^2}{2\lambda_\ell}\mathrm{e}^{-2\lambda_\ell t_i}(e_\ell(y_{k+1})-e_\ell(y_k))^2\mathrm{e}^{\kappa y_k}\lesssim\delta^2\sum_{\ell\geq 1}\mathrm{e}^{-2\lambda_\ell t_i}\lesssim\frac{\delta^2}{\sqrt{t_i}}$$

for $i\geq 1$. Hence, under the condition $N/M\to 0$,

$$\mathbf{E}(|A_1|)\lesssim\frac{1}{MN\sqrt{\Delta}}\sum_{i=1}^{N-1}\frac{1}{\sqrt{i}}\lesssim\frac{1}{M\sqrt{T}}\lesssim\frac{1}{M}=o\Big(\frac{1}{\sqrt{MN}}\Big).$$

Next, we treat

$$\mathbf{E}\left(A_2^2\right)=\frac{4}{(NM\delta)^2}\sum_{i,j=0}^{N-1}\sum_{k,l=0}^{M-1}\mathbf{E}(\tilde S_{ik}^0\tilde S_{jl}^0)\mathbf{E}(\tilde I_{ik}\tilde I_{jl}).$$

Comparison of terms yields that

$$\frac{1}{(NM\delta)^2}\sum_{i\neq j}\sum_{k\neq l}\mathbf{E}(\tilde S_{ik}^0\tilde S_{jl}^0)\mathbf{E}(\tilde I_{ik}\tilde I_{jl})=o\Big(\frac{1}{MN}\Big)$$

can be shown in exactly the same way as $\frac{1}{(NM\delta)^2}\sum_{i\neq j}\sum_{k\neq l}\mathbf{E}(\tilde S_{ik}\tilde S_{jl})^2=o\Big(\frac{1}{MN}\Big)$ from Step 2 of this proof. For the other terms, it follows from

$$\mathbf{E}(\tilde I_{ik}\tilde I_{jl})\lesssim\frac{\delta^2}{\sqrt{t_i+t_j}}\ \text{ for }\ t_i+t_j>0,\quad \mathbf{E}(\tilde S_{ik}^0\tilde S_{jl}^0)\lesssim\delta^2,$$

that

$$\frac{1}{(NM\delta)^2}\sum_{i=1}^{N-1}\sum_{k,l=0}^{M-1}\mathbf{E}(\tilde S_{ik}^0\tilde S_{il}^0)\mathbf{E}(\tilde I_{ik}\tilde I_{il})\lesssim\frac{1}{(NM\delta)^2}\sum_{i=1}^{N-1}\frac{\delta^2}{\sqrt{i\Delta}}\lesssim\frac{1}{M^2N\sqrt{T}}=o\Big(\frac{1}{MN}\Big),$$

$$\frac{1}{(NM\delta)^2}\sum_{i,j=1}^{N-1}\sum_{k=0}^{M-1}\mathbf{E}(\tilde S_{ik}^0\tilde S_{jk}^0)\mathbf{E}(\tilde I_{ik}\tilde I_{jk})\lesssim\frac{1}{(NM\delta)^2}\sum_{i,j=1}^{N-1}\frac{\delta^3}{\sqrt{(i+j)\Delta}}=\frac{1}{M^3\sqrt{T}}=o\Big(\frac{1}{MN}\Big)$$

for $N/M\to 0$. □

The proof of the central limit theorem for double increments (Theorem 2.2.7) is similar to the previous one, but the more complex covariance structure of the double increments has to be taken into account carefully, see Section 2.4.4. The (asymptotic) mean of the rescaled space-time increments, as stated in Proposition 2.2.5, is the first step of our proof. In the following, we write

$$\tilde{D}_{ik} := \mathrm{e}^{\kappa y_k/2} D_{ik}. \tag{2.32}$$

Proof of Proposition 2.2.5. Step 1. We show asymptotic independence of Γ, i.e.,

$$\mathbf{E}\left(D_{ik}^2\right) = \sigma^2 \sum_{\ell\geq 1} \frac{1-\mathrm{e}^{-\pi^2\vartheta_2\ell^2\Delta}}{\pi^2\vartheta_2\ell^2}(e_\ell(y_{k+1})-e_\ell(y_k))^2 + \mathcal{O}\left(\delta\sqrt{\Delta}\left(\delta\wedge\sqrt{\Delta}\right)\right):$$

Define $f(x) := \frac{1-\mathrm{e}^{-x}}{x}$. A first order Taylor approximation of f yields

$$\mathbf{E}\left(D_{ik}^2\right) = \sigma^2\Delta \sum_{\ell\geq 1} f\left(\pi^2\vartheta_2\ell^2\Delta\right)(e_\ell(y_{k+1})-e_\ell(y_k))^2 + R$$

where

$$R \lesssim \Delta^2 \sum_{\ell\geq 1} f'(\vartheta_2(\pi^2\ell^2+\xi_\ell)\Delta)(e_\ell(x+\delta)-e_\ell(x))^2$$

for some $|\xi_\ell| \leq |\Gamma|$. Since

$$(e_\ell(y+\delta)-e_\ell(y))^2 \lesssim \left(\mathrm{e}^{-\kappa\delta/2}(\sin(\pi\ell(y+\delta))-\sin(\pi\ell y))+\sin(\pi\ell y)(\mathrm{e}^{-\kappa\delta/2}-1)\right)^2 \lesssim 1\wedge(\ell\delta)^2$$

and noting that $f'(x^2)$ and $x^2 f'(x^2)$ are integrable, we deduce

$$R \lesssim \Delta^2 \sum_{\ell\geq 1}(1\wedge(\ell\delta)^2) f'(\vartheta_2(\pi^2\ell^2+\xi_\ell)\Delta) = \mathcal{O}\big(\Delta^{3/2}\wedge(\delta^2\sqrt{\Delta})\big) = \mathcal{O}\big((\delta\Delta)\wedge(\delta^2\sqrt{\Delta})\big).$$

Step 2. We verify (i): Thanks to Step 1, we may assume $\lambda_\ell = \pi^2\vartheta_2\ell^2$. It follows from (2.31) that

$$\begin{aligned}\mathbf{E}(\tilde{D}_{ik}^2) =& \sigma^2\mathrm{e}^{-\kappa y}\left(F_{\vartheta_2}(0,\Delta)\left(1+\mathrm{e}^{-\kappa\delta}\right) - 2F_{\vartheta_2}(\delta,\Delta)\mathrm{e}^{-\kappa\delta/2}\right)\\ &- \sigma^2\mathrm{e}^{-\kappa y} D_\delta^2\Big(g(\cdot)F_{\vartheta_2}(2y_k+\cdot,\Delta)\Big)(0).\end{aligned}$$

Consequently, it remains to show

$$D_\delta^2\Big(g(\cdot)F_{\vartheta_2}(2y+\cdot,\Delta)\Big)(0) = \mathcal{O}\left(\delta\sqrt{\Delta}\left(\delta\wedge\sqrt{\Delta}\right)\right)$$

uniformly in $y\in[b,1-b]$. As before this is done by showing

$$F_{\vartheta_2}(x,\Delta), \quad \frac{\partial F_{\vartheta_2}(x,\Delta)}{\partial x} \lesssim \Delta \quad \text{and} \quad \frac{\partial^2 F_{\vartheta_2}(x,\Delta)}{\partial x^2} \lesssim \sqrt{\Delta}$$

uniformly in $x\in[2b,2(1-b)]$. By Lemma 2.4.9, we have

$$F_{\vartheta_2}(x,\Delta) = \Delta\sum_{\ell\geq 1} f(\lambda_\ell\Delta)\cos(\pi\ell x) = \mathcal{O}(\Delta).$$

In order to access the first two derivatives of $F_{\vartheta_2}(\cdot,\Delta)$, we split it into two summands, namely

$$F_{\vartheta_2}(x,\Delta) = \underbrace{\Delta\sum_{\ell\geq 1}\frac{1}{1+\lambda_\ell\Delta}\cos(\pi\ell x)}_{=:H_\Delta(x)} + \underbrace{\Delta\sum_{\ell\geq 1}\left(\frac{1-\mathrm{e}^{-\lambda_\ell\Delta}}{\lambda_\ell\Delta}-\frac{1}{1+\lambda_\ell\Delta}\right)\cos(\pi\ell x)}_{=:G_\Delta(x)}.$$

Using the cosine series formula (1.16), we can compute

$$H_\Delta(x) = \frac{1}{\vartheta_2\pi^2}\sum_{\ell\geq1}\frac{1}{\ell^2+\frac{1}{\pi^2\vartheta_2\Delta}}\cos(\pi\ell x) = \frac{\sqrt{\Delta}}{2\sqrt{\vartheta_2}}\frac{\cosh\left(\frac{1}{\sqrt{\vartheta_2\Delta}}(x-1)\right)}{\sinh\left(\frac{1}{\sqrt{\vartheta_2\Delta}}\right)} - \frac{\Delta}{2}.$$

The corresponding derivatives are bounded by

$$\begin{aligned}
H'_\Delta(x) &\eqsim \frac{\sinh\left(\frac{\pi}{\sqrt{\vartheta_2\Delta}}(x-1)\right)}{\sinh\left(\frac{\pi}{\sqrt{\vartheta_2\Delta}}\right)} \lesssim \frac{\exp\left(\frac{\pi}{\sqrt{\vartheta_2\Delta}}|x-1|\right)}{\exp\left(\frac{\pi}{\sqrt{\vartheta_2\Delta}}\right)} \\
&\lesssim \exp\left(-\frac{\pi}{\sqrt{\vartheta_2\Delta}}(x\wedge(2-x))\right) \lesssim \Delta, \\
H''_\Delta(x) &\eqsim \frac{\cosh\left(\frac{\pi}{\sqrt{\vartheta_2\Delta}}(x-1)\right)}{\sqrt{\Delta}\sinh\left(\frac{\pi}{\sqrt{\vartheta_2\Delta}}\right)} \lesssim \frac{\exp\left(-\frac{\pi}{\sqrt{\vartheta_2\Delta}}(x\wedge(2-x))\right)}{\sqrt{\Delta}} \lesssim \sqrt{\Delta}.
\end{aligned}$$

The derivatives of

$$G_\Delta(x) = \Delta\sum_{\ell\geq1}h(\ell\sqrt{\Delta})\cos(\pi\ell x) \quad \text{with} \quad h(z) := \frac{1-\mathrm{e}^{-z}(1+z)}{z(1+z)}$$

can be computed summand-wisely:

$$\begin{aligned}
G'_\Delta(x) &\eqsim \sqrt{\Delta}\sum_{\ell\geq1}(\ell\sqrt{\Delta})h(\ell\sqrt{\Delta})\sin(\pi\ell x) \lesssim \Delta, \\
G''_\Delta(x) &\eqsim \sum_{\ell\geq1}(\ell^2\Delta)h(\ell\sqrt{\Delta})\cos(\pi\ell x) \lesssim \sqrt{\Delta}
\end{aligned}$$

where the bounds follow from the Riemann sum approximations in Lemma 2.4.9, owing to $xh(x)|_{x=0} = x^2h(x)|_{x=0} = 0$.

Step 3. We show the asymptotic expressions from (*ii*): Due to a Riemann sum argument, we have $\|F_{\vartheta_2}(\cdot,\Delta)\|_\infty \lesssim \sqrt{\Delta}$ and, consequently,

$$\begin{aligned}
\Phi_\vartheta(\delta,\Delta) &= 2\left(F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta)\right) + F_{\vartheta_2}(0,\Delta)\left[1+\mathrm{e}^{-\kappa\delta} - 2\mathrm{e}^{-\kappa\delta/2}\right] \\
&\quad - 2\left(F_{\vartheta_2}(\delta,\Delta) - F_{\vartheta_2}(0,\Delta)\right)\left(\mathrm{e}^{-\kappa\delta/2}-1\right) \\
&= 2\left(F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta)\right) + \mathcal{O}(\delta\sqrt{\Delta}).
\end{aligned}$$

In the case $\delta/\sqrt{\Delta}\to0$, Taylor's formula yields

$$F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta) = -\delta\frac{\partial F_{\vartheta_2}(0,\Delta)}{\partial x} - \frac{\delta^2}{2}\frac{\partial^2 F_{\vartheta_2}(\eta,\Delta)}{\partial x^2}$$

for some $\eta\in[0,\delta]$. We now employ the representation $F_{\vartheta_2}(\cdot,\Delta) = H_\Delta + G_\Delta$ from Step 2. Since $\sin(0)=0$, we have $\frac{\partial F_{\vartheta_2}(0,\Delta)}{\partial x} = H'_\Delta(0) = -\frac{1}{2\vartheta_2}$. The Riemann sum argument yields

$$\begin{aligned}
\frac{\partial^2 F_{\vartheta_2}(\eta,\Delta)}{\partial x^2} &\lesssim \frac{1}{\sqrt{\Delta}}\frac{\cosh\left((\eta-1)/\sqrt{\vartheta_2\Delta}\right)}{\sinh\left(1/\sqrt{\vartheta_2\Delta}\right)} + \sum_{\ell\geq1}(\ell^2\Delta)h(\ell\sqrt{\Delta})\cos(\pi\ell\eta) \\
&\lesssim \frac{1}{\sqrt{\Delta}}\exp\left(-\eta/\sqrt{\vartheta_2\Delta}\right) + \sum_{\ell\geq1}(\ell^2\Delta)h(\ell\sqrt{\Delta}) \lesssim \frac{1}{\sqrt{\Delta}}.
\end{aligned}$$

Therefore,

$$F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta) = \frac{1}{2\vartheta_2}\cdot\delta + \mathcal{O}\left(\frac{\delta^2}{\sqrt{\Delta}}\right).$$

If $\delta/\sqrt{\Delta}\to\infty$, Lemma 2.4.9 implies

$$F_{\vartheta_2}(\delta,\Delta) = -\frac{\Delta}{2} + \mathcal{O}\left(\frac{\Delta^{3/2}}{\delta^2}\right)$$

and Lemma 2.4.10 yields

$$F_{\vartheta_2}(0,\Delta) = \sqrt{\Delta}\int_0^\infty \frac{1-\mathrm{e}^{-\pi^2\vartheta_2 z^2}}{\pi^2\vartheta_2 z^2}dz - \frac{\Delta}{2} + \mathcal{O}(\Delta^{3/2}). \tag{2.33}$$

Since

$$\int_0^\infty \frac{1-\mathrm{e}^{-\pi^2\vartheta_2 z^2}}{\pi^2\vartheta_2 z^2}dz = \frac{1}{\sqrt{\vartheta_2\pi}},$$

we obtain

$$F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta) = \frac{\sqrt{\Delta}}{\sqrt{\vartheta_2\pi}} + \mathcal{O}\left(\frac{\Delta^{3/2}}{\delta^2}\right).$$

Finally, we derive the asymptotic expression for the case $\delta/\sqrt{\Delta}\equiv r$, the case $\delta/\sqrt{\Delta}\to r$ can be handled similarly. We have

$$\begin{aligned}\Phi_\vartheta(\delta,\Delta) &= 2(F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta))\mathrm{e}^{-\kappa\delta/2} + F_{\vartheta_2}(0,\Delta)(1+\mathrm{e}^{-\kappa\delta} - 2\mathrm{e}^{-\kappa\delta/2})\\ &= 2(F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(\delta,\Delta))\mathrm{e}^{-\kappa\delta/2} + \mathcal{O}(\Delta^{3/2})\end{aligned}$$

and, since $1-\cos(0)=0$, Lemma 2.4.10 yields

$$\begin{aligned}F_{\vartheta_2}(0,\Delta) - F_{\vartheta_2}(r\sqrt{\Delta},\Delta) &= \sum_{\ell\geq 1}\frac{1-\mathrm{e}^{-\pi^2\vartheta_2\ell^2\Delta}}{\pi^2\vartheta_2\ell^2}\Big(1-\cos\Big(\pi\ell r\sqrt{\Delta}\Big)\Big)\\ &= \sqrt{\Delta}\int_0^\infty \frac{1-\mathrm{e}^{-\pi^2\vartheta_2 z^2}}{\pi^2\vartheta_2 z^2}\left(1-\cos\left(\pi r z\right)\right)dz + \mathcal{O}(\Delta^{3/2}).\end{aligned}$$

It remains to compute the integral. By substituting $\tilde{r} = r/\sqrt{\vartheta_2}$, we can pass to

$$\int_0^\infty \frac{1-\mathrm{e}^{-\pi^2\vartheta_2 z^2}}{\pi^2\vartheta_2 z^2}\left(1-\cos\left(\pi r z\right)\right)dz = \frac{1}{\pi\sqrt{\vartheta_2}}\Big(h_1(\tilde{r}) - h_2(\tilde{r})\Big)$$

where

$$h_1(\tilde{r}) := \int_0^\infty \frac{1-\cos(\tilde{r}z)}{z^2}dz, \qquad h_2(\tilde{r}) := \int_0^\infty \mathrm{e}^{-z^2}\frac{1-\cos(\tilde{r}z)}{z^2}dz.$$

To compute h_1, note that $S(z) + \frac{\cos(z)-1}{z}$ is an antiderivative of $\frac{1-\cos(z)}{z}$ where $S(z) := \int_0^z \frac{\sin(h)}{h}\,dh$ is the sine integral. Consequently, a substitution and $\lim_{z\to\infty} S(z) = \pi/2$ yields

$$h_1(\tilde{r}) = \tilde{r}\int_0^\infty \frac{1-\cos(z)}{z^2}\,dz = \frac{\pi\tilde{r}}{2}.$$

To treat h_2, note that $h_2(0) = h_2'(0) = 0$ and, hence, $h_2(\tilde{r}) = \int_0^{\tilde{r}}\int_0^s h_2''(u)\,du\,ds$. Now, plugging in

$$h_2''(\tilde{r}) = \int_0^\infty \mathrm{e}^{-z^2}\cos(\tilde{r}z)\,dz = \frac{\sqrt{\pi}}{2}\mathrm{e}^{-\tilde{r}^2/4}$$

and integrating by parts yields

$$h_2(\tilde r)=\frac{\sqrt{\pi}}{2}\int_0^{\tilde r}\int_0^s \mathrm{e}^{-u^2/4}\,du=\sqrt{\pi}\tilde r\int_0^{\tilde r/2}\mathrm{e}^{-u^2}\,du+\sqrt{\pi}\left(\mathrm{e}^{-\tilde r^2/4}-1\right).$$

Thus, the claim follows from

$$\begin{aligned}h_1(\tilde r)-h_2(\tilde r)&=\frac{\pi\tilde r}{2}\left(1-\frac{2}{\sqrt{\pi}}\int_0^{\tilde r/2}\mathrm{e}^{-u^2}\,du\right)+\sqrt{\pi}\left(1-\mathrm{e}^{-\tilde r^2/4}\right)\\&=\tilde r\sqrt{\pi}\int_{\tilde r/2}^{\infty}\mathrm{e}^{-u^2}\,du+\sqrt{\pi}\left(1-\mathrm{e}^{-\tilde r^2/4}\right).\end{aligned}$$

□

We now conclude the central limit theorem for the realized rescaled space-time variation.

Proof of Theorem 2.2.7. Step 1. We prove the central limit theorem for the case of a stationary initial condition: Asymptotic normality follows just like in the proof of Theorem 2.2.3 and it remains to calculate the asymptotic variance. Using the notation from the proof of the latter theorem (with space increments replaced by double increments), we have

$$\operatorname{Var}(\mathbb{V})=\frac{2}{M^2N^2\Phi_\vartheta^2(\delta,\Delta)}(v_1+v_2+v_3+v_4).$$

To determine the asymptotic variances, we have to distinguish the three different sampling regimes.
Case $\delta/\sqrt{\Delta}\to 0$: By Lemmas 2.4.5 and 2.4.6, we have

$$\begin{aligned}\operatorname{Var}(\tilde D_{ki})^2&=\frac{\sigma^4}{\vartheta_2^2}\mathrm{e}^{-\kappa\delta}\cdot\delta^2+o(\delta^2),\\ \operatorname{Cov}(\tilde D_{ki},\tilde D_{k(i+1)})^2&=\frac{\sigma^4}{4\vartheta_2^2}\mathrm{e}^{-\kappa\delta}\cdot\delta^2+o(\delta^2)\end{aligned}$$

as well as

$$\begin{aligned}\operatorname{Cov}(\tilde D_{ki},\tilde D_{kj})^2&=o\left(\frac{\delta^2}{|i-j|^5}\right),\quad |i-j|\ge 2,\\ \operatorname{Cov}(\tilde D_{ki},\tilde D_{lj})^2&=\mathcal{O}\left(\frac{\delta^4}{(|i-j|+1)^4}\left(\frac{M^2}{(k-l)^2}\wedge\frac{1}{\Delta}\right)\right),\quad k\neq l.\end{aligned}$$

The latter covariances are negligible for the asymptotic variance since

$$\sum_{k\le M}\left(\frac{M^2}{k^2}\wedge\frac{1}{\Delta}\right)\lesssim\frac{M}{\sqrt{\Delta}},$$

cf. the proof of Theorem 2.2.3. Inserting $\Phi_\vartheta^2(\delta,\Delta)=\frac{\mathrm{e}^{-\kappa\delta}}{\vartheta_2^2}\delta^2+o(\delta^2)$ from Proposition 2.2.5 yields the claim.
Case $\delta/\sqrt{\Delta}\to\infty$: By Lemmas 2.4.5 and 2.4.7, we have

$$\begin{aligned}\operatorname{Var}(\tilde D_{ki})^2&=\frac{4\sigma^4}{\pi\vartheta_2}\mathrm{e}^{-\kappa\delta}\cdot\Delta+o(\Delta),\\ \operatorname{Cov}(\tilde D_{ki},\tilde D_{(k+1)i})^2&=\frac{\sigma^4}{\pi\vartheta_2}\mathrm{e}^{-\kappa\delta}\cdot\Delta+o(\Delta).\end{aligned}$$

From $\sqrt{J-1}+\sqrt{J+1}-2\sqrt{J}=\mathcal{O}(J^{-3/2})$ and $\sqrt{\Delta}/\delta\to 0$, it follows for $J:=|i-j|\ge 1$ that

$$\operatorname{Cov}(\tilde D_{ki},\tilde D_{kj})^2=\frac{\sigma^4}{\pi\vartheta_2}\left(\sqrt{J-1}+\sqrt{J+1}-2\sqrt{J}\right)^2\mathrm{e}^{-\kappa\delta}\cdot\Delta+o\left(\frac{\Delta}{J^{3/2}}\right)+\mathcal{O}(\Delta^3),$$

$$\mathrm{Cov}(\tilde{D}_{ki}, \tilde{D}_{(k+1)j})^2 = \frac{\sigma^4}{4\pi\vartheta_2}\left(\sqrt{J-1}+\sqrt{J+1}-2\sqrt{J}\right)^2 \mathrm{e}^{-\kappa\delta}\cdot\Delta + o\left(\frac{\Delta}{J^{3/2}}\right) + \mathcal{O}(\Delta^3).$$

Note that the $\mathcal{O}(\Delta^3)$-term is negligible for the asymptotic variance since

$$N^2M\Delta^3 = MN\Delta\cdot N\Delta^2 = MN\Delta\cdot\frac{T}{M}\cdot M\sqrt{\Delta}\cdot\sqrt{\Delta} = o(NM\Delta).$$

The remaining covariances do not contribute to the asymptotic variance since for $|k-l|\geq 2$, we have

$$\mathrm{Cov}(\tilde{D}_{ki}, \tilde{D}_{lj})^2 = \mathcal{O}\left(\frac{\Delta\delta^4}{(J+1)^3}\right) + \mathcal{O}\left(\frac{\Delta^2}{(J+1)^2}\frac{M^2}{(k-l)^2}\right).$$

The claim is now proved by inserting $\Phi_\vartheta^2(\delta,\Delta) = \frac{4}{\pi\vartheta_2}\mathrm{e}^{-\kappa\delta}\Delta + o(\Delta)$ and noting that for the function $g(j) = (\sqrt{j-1}+\sqrt{j+1}-2\sqrt{j})^2$, we obtain

$$\frac{1}{N}\sum_{\substack{i,j=0\\ i\neq j}}^{N-1} g(|i-j|) = \frac{2}{N}\sum_{i=1}^{N-1}\sum_{j=1}^{i} g(j) \to 2\sum_{j\geq 1} g(j), \quad N\to\infty,$$

as the Cesàro limit.

Case $\delta/\sqrt{\Delta}\equiv r\in(0,\infty)$*:* For $f:\mathbb{R}^2\to\mathbb{R}$, we define

$$D_x^2 f(x,y) := f(x+2,y)+f(x,y)-2f(x+1,y),$$
$$D_y^2 f(x,y) := f(x,y+2)+f(x,y)-2f(x,y+1).$$

We show that the asymptotic variance is given by $C(r/\sqrt{\vartheta_2})\sigma^4$ where

$$C(h) := \frac{2}{\Lambda_{0,0}^2(h)}\sum_{j,l\in\mathbb{Z}}\Lambda_{j,l}^2(h), \qquad \Lambda_{j,l}(h) := \left(D_x^2 D_y^2 G_h\right)(|j|-1,|l|-1) \tag{2.34}$$

and

$$G_h(j,l) := \sqrt{|j|}H\left(\frac{h|l|}{\sqrt{|j|}}\right)\mathbf{1}_{\{j\neq 0\}}, \qquad H(x) := \frac{1}{2\sqrt{\pi}}\left(\exp\left(-\frac{x^2}{4}\right) - x\int_{x/2}^{\infty}\mathrm{e}^{-z^2}\,dz\right). \tag{2.35}$$

With $F_{J,\Delta}$ from Lemma 2.4.5, define

$$\xi_{i-j,k-l}^{\Delta} := \begin{cases} 2D_\delta F_{|i-j|,\Delta}(0), & l=k,\\ D_\delta^2 F_{|i-j|,\Delta}((|k-l|-1)\delta), & l\neq k\end{cases}$$

with $\delta = r\sqrt{\Delta}$. Then, Lemma 2.4.5 reads as

$$\mathrm{Cov}(\tilde{D}_{ik}, \tilde{D}_{ik}) = -\sigma^2\mathrm{e}^{-\kappa\delta/2}\xi_{i-j,k-l}^{\Delta} + \mathcal{O}\left(\frac{\Delta^{3/2}}{(J+1^{3/2})}\right). \tag{2.36}$$

Since each term $\xi_{J,L}^{\Delta}$ is a Riemann sum multiplied by $\sqrt{\Delta}$, we have for $J,L\geq 0$ that

$$\lim_{\Delta\to 0}\Delta^{-1/2}\xi_{J,L}^{\Delta} = -\begin{cases} 2(\Psi_r(J,1)) - \Psi_r(J,0)), & L=0,\\ \Psi_r(J,L-1)+\Psi_r(J,L+1)-2\Psi_r(J,L), & L\geq 1\end{cases}$$

where

$$\Psi_r(J,L) := \begin{cases} \displaystyle\int_0^\infty \frac{1-\mathrm{e}^{-\pi^2\vartheta_2 z^2}}{\pi^2\vartheta_2 z^2}\cos(\pi r L z)\,dz, & J=0,\\ \displaystyle\int_0^\infty \frac{2\mathrm{e}^{-J\pi^2\vartheta_2 z^2} - \mathrm{e}^{-(J+1)\pi^2\vartheta_2 z^2} - \mathrm{e}^{-(J-1)\pi^2\vartheta_2 z^2}}{2\pi^2\vartheta_2 z^2}\cos(\pi r L z)\,dz, & J\geq 1.\end{cases}$$

By symmetry of the cosine function,

$$\lim_{M,N\to\infty} \Delta^{-1/2}\xi_{J,L} = -\Big(\Psi_r(J,|L|-1)+\Psi_r(J,|L|+1)-2\Psi_r(J,|L|)\Big)$$

also holds for negative L. Hence, we can write for all $L\in\mathbb{Z}$, $J\geq 0$ and with G_h from (2.35):

$$\begin{aligned}\Psi_r(J,L) &= \int_0^\infty \frac{2\mathrm{e}^{-J\pi^2\vartheta_2 z^2}-\mathrm{e}^{-(J+1)\pi^2\vartheta_2 z^2}-\mathrm{e}^{-|J-1|\pi^2\vartheta_2 z^2}}{2\pi^2\vartheta_2 z^2}\cos(\pi r L z)\,dz\\ &= \Big(G_{r/\sqrt{\vartheta_2}}(J+1,L)+G_{r/\sqrt{\vartheta_2}}(J-1,L)-2G_{r/\sqrt{\vartheta_2}}(J,L)\Big)/\sqrt{\vartheta_2}\end{aligned}$$

where the last equality follows from

$$\frac{G_{r/\sqrt{\vartheta_2}}(j,l)}{\sqrt{\vartheta_2}} = \int_0^\infty \frac{1-\mathrm{e}^{-|j|\pi^2\vartheta_2 z^2}}{2\pi^2\vartheta_2 z^2}\cos(\pi r l z)\,dz, \qquad j,l\in\mathbb{Z},$$

which may be shown analogously to the calculation of $\psi_{\vartheta_2}(r)$. Consequently, for all $J\in\{1-N,\dots,N-1\}$ and $L\in\{1-M,\dots,M-1\}$, we have

$$\lim_{M,N\to\infty}\Delta^{-1/2}\xi_{J,L} = -\Lambda_{J,L}(r/\sqrt{\vartheta_2})/\sqrt{\vartheta_2}.$$

The usual Riemann sum argument shows $F_{J,\Delta}(0)\lesssim\frac{\sqrt{\Delta}}{(J+1)^{3/2}}\lesssim\frac{\sqrt{\Delta}}{(J+1)}$ for $J\geq 0$ and (2.59) from the proof of Lemma 2.4.7 yields $F_{J,\Delta}(L\delta)\lesssim\frac{\Delta}{(J+1)L\delta}\lesssim\frac{\sqrt{\Delta}}{(J+1)(L+1)}$ for $J\in\mathbb{N}_0$ and $L\geq 1$. We obtain

$$\Delta^{-1/2}\xi^{\Delta}_{J,L} = \mathcal{O}\left(\frac{1}{(|J|+1)(|L|+1)}\right), \qquad J,L\in\mathbb{Z}. \tag{2.37}$$

Therefore,

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{\sqrt{NM\Delta}}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}\tilde{D}_{ik}^2\right) &= \frac{2\sigma^4}{NM\Delta}\sum_{i,j=0}^{N-1}\sum_{k,l=0}^{M-1}\operatorname{Cov}(\tilde{D}_{ik},\tilde{D}_{jl})^2\\ &= \frac{2\sigma^4}{NM\Delta}\sum_{i,j=0}^{N-1}\sum_{k,l=0}^{M-1}(\xi^{\Delta}_{i-j,k-l})^2+o(1).\end{aligned}$$

By dominated convergence and taking Cesàro limits twice, we conclude

$$\begin{aligned}\lim_{M,N\to\infty}\operatorname{Var}\left(\frac{1}{\sqrt{NM\Delta}}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}\tilde{D}_{ik}^2\right) &= \lim_{M,N\to\infty}\frac{2\sigma^4}{\vartheta_2 NM}\sum_{i,j=0}^{N-1}\sum_{k,l=0}^{M-1}\Lambda^2_{i-j,k-l}(r/\sqrt{\vartheta_2})\\ &= \frac{2\sigma^4}{\vartheta_2}\sum_{i,k\in\mathbb{Z}}\Lambda^2_{i,k}(r/\sqrt{\vartheta_2}).\end{aligned}$$

Since $\psi_{\vartheta_2}(r) = -\Lambda_{0,0}(r/\sqrt{\vartheta_2})/\sqrt{\vartheta_2}$, we have $\Phi^2_\vartheta(\delta,\Delta) = \mathrm{e}^{-\kappa\delta}\Lambda^2_{0,0}(r/\sqrt{\vartheta_2})/\vartheta_2\cdot\Delta+o(\Delta)$. Finally, dividing by $\lim_{M,N\to\infty}\Delta^{-1}\Phi^2_\vartheta(\delta,\Delta)=\Lambda^2_{0,0}(r/\sqrt{\vartheta_2})/\vartheta_2$ yields the claimed asymptotic variance.

Step 2. To show that the central limit theorem also holds for the zero initial condition, we proceed as in Step 4 of the proof of Theorem 2.2.3: Again, let X^0_t be the process (1.12) with $\xi=0$ and $\xi_t := S(t)\xi$ where ξ follows the stationary initial condition and is independent of X^0. For these two processes, denote the rescaled double increments by $\tilde{D}^0_{ik}$ and $\tilde{I}_{ik}$, respectively. We show that $\mathbf{E}(|A_1|)=o(1/\sqrt{MN})$ and $\mathbf{E}(A_2^2)=o(1/(MN))$ with

$$A_1 := \frac{1}{NM\Phi_\vartheta(\delta,\Delta)}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\tilde{I}^2_{ik}, \qquad A_2 := -\frac{2}{NM\Phi_\vartheta(\delta,\Delta)}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\tilde{D}^0_{ik}\tilde{I}_{ik}.$$

First of all, using the Riemann sum argument,

$$\begin{aligned}\mathbf{E}(\tilde{I}_{ik}^2) &= \sum_{\ell\geq 1}\frac{\sigma^2}{2\lambda_\ell}(1-\mathrm{e}^{-\lambda_\ell\Delta})^2\mathrm{e}^{-2\lambda_\ell t_i}(e_\ell(y_{k+1})-e_\ell(y_k))^2\\ &\lesssim \Delta(\delta^2\wedge\Delta)\sum_{\ell\geq 1}\lambda_\ell \mathrm{e}^{-2\lambda_\ell t_i} \lesssim \Delta(\delta^2\wedge\Delta)t_i^{-3/2} = \frac{\delta^2}{\sqrt{\Delta}i^{3/2}}\wedge\frac{\sqrt{\Delta}}{i^{3/2}}.\end{aligned}$$

In the regime $M\sqrt{\Delta}\gtrsim 1$, we have $\Phi_\vartheta(\delta,\Delta)\approx\delta$ and, thus,

$$\mathbf{E}(|A_1|)\lesssim\frac{1}{NM}\sum_{i=1}^{N-1}\sum_{k=0}^{M-1}\frac{\delta}{\sqrt{\Delta}i^{3/2}}\lesssim\frac{M\delta}{MN\sqrt{\Delta}}\lesssim\frac{1}{M\sqrt{N}\sqrt{T}}=o\Big(\frac{1}{\sqrt{MN}}\Big).$$

In the regime $M\sqrt{\Delta}\to 0$, we have $\Phi_\vartheta(\delta,\Delta)\approx\sqrt{\Delta}$ and, thus,

$$\mathbf{E}(|A_1|)\lesssim\frac{1}{NM}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}\frac{1}{i^{3/2}}\lesssim\frac{1}{N}=\frac{1}{\sqrt{MN}}\frac{(M\sqrt{\Delta})^{1/2}}{(NT)^{1/4}}=o\Big(\frac{1}{\sqrt{MN}}\Big).$$

To treat $\mathbf{E}(A_2^2)$, we use

$$\begin{aligned}\mathbf{E}(\tilde{I}_{ik}\tilde{I}_{jl}) &= \sum_{\ell\geq 1}\frac{\sigma^2}{2\lambda_\ell}(1-\mathrm{e}^{-\lambda_\ell\Delta})^2\mathrm{e}^{-\lambda_\ell(t_i+t_j)}(e_\ell(y_{k+1})-e_\ell(y_k))(e_\ell(y_{l+1})-e_\ell(y_l))\\ &\lesssim \Delta^2\delta^2\sum_{\ell\geq 1}\lambda_\ell^2\mathrm{e}^{-2\lambda_\ell(t_i+t_j)}\lesssim\Delta^2\delta^2(t_i+t_j)^{-5/2}=\frac{\delta^2}{\sqrt{\Delta}(i+j)^{5/2}}.\end{aligned}$$

Due to

$$\sum_{i,j=1}^{N-1}\frac{1}{(i+j)^{5/2}}\leq\sum_{i=1}^{N-1}\sum_{j\geq i}\frac{1}{j^{5/2}}\lesssim\sum_{i=1}^{N-1}\frac{1}{i^{3/2}}\leq\sum_{i=1}^{\infty}\frac{1}{i^{3/2}}<\infty$$

and $\mathbf{E}(\tilde{D}_{ik}^0\tilde{D}_{jl}^0)\approx\Phi_\vartheta(\delta,\Delta)$, we obtain

$$\begin{aligned}\mathbf{E}(A_2^2) &= \frac{4}{(MN\Phi_\vartheta(\delta,\Delta))^2}\sum_{i,j=1}^{N-1}\sum_{k=0}^{M-1}\mathbf{E}(\tilde{I}_{ik}\tilde{I}_{jl})\mathbf{E}(\tilde{D}_{ik}^0\tilde{D}_{jl}^0)\\ &\lesssim\frac{1}{(MN(\delta\wedge\sqrt{\Delta}))^2}\frac{M^2\delta^2(\delta\wedge\sqrt{\Delta})}{\sqrt{\Delta}}\lesssim\frac{1}{(MN)^2\sqrt{\Delta}(\delta\wedge\sqrt{\Delta})}.\end{aligned}$$

In the regime $M\sqrt{\Delta}\gtrsim 1$, the above term is of the order $\frac{1}{MN^{3/2}\sqrt{T}}=o(\frac{1}{MN})$ and, for $M\sqrt{\Delta}\to 0$, it is of the order $\frac{1}{M^2NT}=o(\frac{1}{MN})$. This finishes the proof. □

2.4.3 Proofs for the estimators

Propositions 2.2.10 and 2.2.11 follow immediately from the central limit theorems for the realized quadratic variations and the delta method.

Before proving Theorem 2.2.12, we introduce some notation that will be used throughout the proof and we state the asymptotic covariance matrix explicitly. To simplify calculations, we will assume that the target function $K_{N,M}$ is parameterized in terms of

$$\eta_0 := (\sigma^2,\vartheta_2,\kappa)$$

such that minimizing $K_{N,M}$ with respect to $\eta \in H$ is the same as minimizing with respect to $\eta_0 \in H_0$ and then multiplying the second and third coordinates of the minimizer to obtain the estimator for ϑ_1. The corresponding parameter space is given by the compact set $H_0 := \{z \in \mathbb{R}^3 : z = (\sigma^2, \vartheta_2, \vartheta_1/\vartheta_2)$ for some $(\sigma^2, \vartheta_2, \vartheta_1) \in H\}$. Recall the definition of $\Lambda_{i,k}(\cdot)$ from (2.34) and for any $i, k \in \mathbb{Z}$ let

$$
\begin{aligned}
A_{ik}^r &:= -\Lambda_{ik}(r/\sqrt{\vartheta_2})/\sqrt{\vartheta_2}, & A_r &:= \sum_{i,k\in\mathbb{Z}} (A_{ik}^r)^2,\\
B_{ik}^r &:= 2A_{ik}^r + A_{(i-1)k}^r + A_{(i+1)k}^r, & B_r &:= \sum_{i,k\in\mathbb{Z}} (B_{ik}^r)^2,\\
C_{ik}^r &:= A_{ik}^r + A_{(i-1)k}^r, & C_r &:= \sum_{i,k\in\mathbb{Z}} (C_{ik}^r)^2.
\end{aligned}
$$

In terms of

$$
H(x) := \frac{4x}{\sqrt{\pi}}\left(1 - \mathrm{e}^{-x^2} + 2x\int_x^\infty \mathrm{e}^{-z^2}\,dz\right), \qquad H'(x) = \frac{4}{\sqrt{\pi}}\left(1 - \mathrm{e}^{-x^2} + 4x\int_x^\infty \mathrm{e}^{-z^2}\,dz\right),
$$

$x \geq 0$, we have $\psi_{\vartheta_2}(r) = \frac{1}{r}H\big(\frac{r}{2\sqrt{\vartheta_2}}\big)$ and $\frac{\partial}{\partial\vartheta_2}\psi_{\vartheta_2}(r) = -H'\big(\frac{r}{2\sqrt{\vartheta_2}}\big)\frac{1}{4\vartheta_2^{3/2}}$. Denoting $r_i := r/\sqrt{i}$, the gradient of $\eta_0 \mapsto f_{\eta_0}^i(z)$ is, thus, given by

$$
g_{\eta_0}^i(z) := \mathrm{e}^{-\kappa z}\left(\frac{1}{r_i}H\Big(\frac{r_i}{2\sqrt{\vartheta_2}}\Big), -\frac{\sigma^2}{4\vartheta_2^{3/2}}H'\Big(\frac{r_i}{2\sqrt{\vartheta_2}}\Big), -z\frac{\sigma^2}{r_i}H\Big(\frac{r_i}{2\sqrt{\vartheta_2}}\Big)\right)^\top
$$

and we define

$$
h_{\eta_0}^i(z) := \mathrm{e}^{-\kappa z} g_{\eta_0}^i(z)
$$

for $i = 1, 2$ and $z \in [b, 1-b]$. Moreover, we write $\langle f, g\rangle_b := \frac{1}{1-2b}\int_b^{1-b} f(x)g(x)dx$ for $f, g \in L^2([b, 1-b])$. We will prove that the asymptotic covariance matrix is given by

$$
\Omega_\eta^r = JV^{-1}UV^{-1}J^\top \tag{2.38}
$$

where $U = U(\eta_0), V = V(\eta_0)$ and $J = J(\eta_0)$ are defined via

$$
\begin{aligned}
U_{ij} &:= 4\sigma^4\Big(2A_r\langle (h_{\eta_0}^1)_i, (h_{\eta_0}^1)_j\rangle_b + B_r\langle (h_{\eta_0}^2)_i, (h_{\eta_0}^2)_j\rangle_b + \sqrt{2}C_r\big(\langle (h_{\eta_0}^1)_i, (h_{\eta_0}^2)_j\rangle_b + \langle (h_{\eta_0}^2)_i, (h_{\eta_0}^1)_j\rangle_b\big)\Big),\\
V_{ij} &:= 2\left(\langle (g_{\eta_0}^1)_i, (g_{\eta_0}^1)_j\rangle_b + \langle (g_{\eta_0}^2)_i, (g_{\eta_0}^2)_j\rangle_b\right), \qquad i, j \in \{1, 2, 3\},
\end{aligned}
$$

and

$$
J := \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & \kappa & \vartheta_2 \end{pmatrix}.
$$

Proof of Theorem 2.2.12. The proof uses the classical theory on minimum contrast estimators, see, e.g., [27]. In particular, the mean value theorem yields

$$
-\dot{K}_{N,M}(\eta_0) = \dot{K}_{N,M}(\hat{\eta}_0) - \dot{K}_{N,M}(\eta_0) = \left(\int_0^1 \ddot{K}_{N,M}(\eta_0 + \tau(\hat{\eta}_0 - \eta_0))\,d\tau\right)(\hat{\eta}_0 - \eta_0)
$$

as soon as $[\hat{\eta}_0, \eta_0] \subset H_0$, where $\dot{K}_{N,M}$ and $\ddot{K}_{N,M}$ denote gradient and Hessian with respect to η_0, respectively. In the sequel, we will verify that $K_{N,M}$ is associated with the contrast function

$$
K(\eta_0, \tilde{\eta}_0) := K^1(\eta_0, \tilde{\eta}_0) + K^2(\eta_0, \tilde{\eta}_0) \quad \text{where} \quad K^i(\eta_0, \tilde{\eta}_0) := \frac{1}{1-2b}\int_b^{1-b} (f_{\eta_0}^i(z) - f_{\tilde{\eta}_0}^i(z))^2\,dz
$$

(Steps 1-2), show consistency of $\hat{\eta}_0$ (Step 3), prove asymptotic normality of $\dot{K}_{N,M}(\eta_0)$ with covariance matrix U (Steps 4-7) and deduce stochastic convergence of $\int_0^1 \ddot{K}_{N,M}(\eta_0+\tau(\hat{\eta}_0-\eta_0))\,d\tau$ to the invertable matrix V (Steps 8-9). Then, Slutsky's Lemma yields $\sqrt{MN}(\hat{\eta}_0-\eta_0) = -\sqrt{MN}V(\eta_0)^{-1}\dot{K}_{N,M}(\eta_0) + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V^{-1}UV^{-1})$. Finally, since J is the Jacobian matrix of the mapping $\eta_0 = (\sigma^2, \vartheta_2, \kappa) \mapsto (\sigma^2, \vartheta_2, \kappa\vartheta_2) = \eta$, the result follows from the delta method. Throughout, we work under Assumption (ST). The case $X_0 = 0$ can be treated by employing similar approximation arguments as in the proof of Theorem 2.2.7 in Steps 2, 3, 7 and 8 of this proof.

Step 1. We show that K is a contrast function in the sense that, for each η_0, the function $\tilde{\eta}_0 \mapsto K(\eta_0, \tilde{\eta}_0)$ attains its unique minimum in $\tilde{\eta}_0 = \eta_0$. Since $f^i_{\eta_0}(\cdot)$ is continuous, it is sufficient to show that $(f^1_{\eta_0}, f^2_{\eta_0}) = (f^1_{\tilde{\eta}_0}, f^2_{\tilde{\eta}_0})$ holds if and only if $\eta_0 = \tilde{\eta}_0$. Clearly, $(f^1_{\eta_0}, f^2_{\eta_0}) = (f^1_{\tilde{\eta}_0}, f^2_{\tilde{\eta}_0})$ holds if and only if $\kappa = \tilde{\kappa}$ and $\sigma^2\psi_{\vartheta_2}(r_i) = \tilde{\sigma}^2\psi_{\tilde{\vartheta}_2}(r_i)$ for $i = 1, 2$. Therefore, in order to prove identifiability, it is sufficient to show that $\vartheta_2 \mapsto \psi_{\vartheta_2}(r_1)/\psi_{\vartheta_2}(r_2)$ is injective, which, in turn, is implied by strict monotonicity of $H(r_1 z)/H(r_2 z)$ in $z > 0$. We show that the corresponding derivative or, equivalently, the function $z \mapsto H'(r_1 z)H(r_2 z)r_1 - H'(r_2 z)H(r_1 z)r_2$, is strictly negative for all $z > 0$: For $x > 0$, define $p(x) = \int_x^\infty \mathrm{e}^{-z^2}\,dz$ and $q(x) = 1 - \mathrm{e}^{-x^2}$. A simple calculation shows that

$$H'(r_1 z)H(r_2 z)r_1 - H'(r_2 z)H(r_1 z)r_2 = \frac{32}{\pi} r_1 r_2 z\Big(p(r_1 z)q(r_2 z)r_1 z - p(r_2 z)q(r_1 z)r_2 z\Big)$$

which is strictly negative if we can show that $p(b)q(a)b - p(a)q(b)a < 0$ for all $0 < a < b$. Now, a substitution yields $p(x) = x\int_1^\infty \mathrm{e}^{-x^2t^2}\,dt$ and $q(x) = 2x^2\int_0^1 s\mathrm{e}^{-x^2s^2}\,ds$ and, therefore,

$$p(b)q(a)b - p(a)q(b)a = 2a^2b^2 \int_0^1 \int_1^\infty s\left(\mathrm{e}^{-b^2t^2-a^2s^2} - \mathrm{e}^{-a^2t^2-b^2s^2}\right)\,dt\,ds < 0$$

follows from negativity of the integrand.

In the sequel, we follow the series of arguments from Theorem 5.1 of Bibinger and Trabs [9].

Step 2. K is the contrast function associated with the process $K_{N,M}$ in the sense that $K_{N,M}(\tilde{\eta}_0) \xrightarrow{\mathbf{P}_{\eta_0}} K(\eta_0, \tilde{\eta}_0)$, $N, M \to \infty$, for all $\tilde{\eta}_0 \in H_0$: Recall from the proof of Theorem 2.2.7 that for $i, j, k, l \in \mathbb{N}_0$, we have

$$\operatorname{Cov}(D_{ik}, D_{jl}) = \sigma^2 \mathrm{e}^{-\kappa\frac{z_k+z_l}{2}} \xi^\Delta_{i-j,k-l} + \mathcal{O}\left(\frac{\Delta^{3/2}}{(|i-j|+1)^{3/2}}\right), \tag{2.39}$$

$$\xi^\Delta_{i,k} = \mathcal{O}\left(\frac{\sqrt{\Delta}}{(|i|+1)(|k|+1)}\right) \tag{2.40}$$

and $\lim_{N,M\to\infty} \Delta^{-1/2}\xi^\Delta_{i-j,k-l} = A^r_{ik} = -\Lambda_{ik}(r/\sqrt{\vartheta_2})/\sqrt{\vartheta_2}$. Now, in terms of

$$r_{ik}(\eta_0) := D^2_{ik}/\sqrt{\Delta} - f^1_{\eta_0}(z_k), \qquad R_k(\eta_0) := \frac{1}{N}\sum_{i=0}^{N-1} r_{ik}(\eta_0),$$

we can write

$$\begin{aligned} K^1_{N,M}(\tilde{\eta}_0) =& \frac{1}{M}\sum_{k=0}^{M-1} \left(f^1_{\eta_0}(z_k) - f^1_{\tilde{\eta}_0}(z_k)\right)^2 \\ &+ \frac{2}{M}\sum_{k=0}^{M-1} R_k(\eta_0)\left(f^1_{\eta_0}(z_k) - f^1_{\tilde{\eta}_0}(z_k)\right) + \frac{1}{M}\sum_{k=0}^{M-1} R^2_k(\eta_0). \end{aligned} \tag{2.41}$$

Clearly, the first summand converges to $K^1(\eta_0, \tilde{\eta}_0)$. To prove that the other two summands are negligible, note that

$$\mathbf{E}(r_{ik}r_{jl}) = \mathbf{E}\left((D^2_{ik}/\sqrt{\Delta} - \mathbf{E}(D^2_{ik}/\sqrt{\Delta}) + \mathcal{O}(\Delta))(D^2_{jl}/\sqrt{\Delta} - \mathbf{E}(D^2_{jl}/\sqrt{\Delta}) + \mathcal{O}(\Delta))\right)$$

$$
\begin{aligned}
&= \frac{1}{\Delta}\text{Cov}(D_{ik}^2, D_{jl}^2) + \mathcal{O}(\Delta^2) \\
&= \frac{2}{\Delta}\text{Cov}(D_{ik}, D_{jl})^2 + \mathcal{O}(\Delta^2) = \mathcal{O}\left(\frac{1}{(|i-j|+1)^2(|k-l|+1)^2}\right) + \mathcal{O}(\Delta^2).
\end{aligned}
$$

By Markov's inequality and boundedness of $\phi(\cdot) = f_{\eta_0}^1(\cdot) - f_{\tilde{\eta}_0}^1(\cdot)$, we have for any $\varepsilon > 0$ that

$$
\begin{aligned}
\mathbf{P}\left(\left|\frac{1}{M}\sum_{k=0}^{M-1} R_k\phi(z_k)\right| \geq \varepsilon\right) &\leq \frac{1}{\varepsilon^2 M^2}\sum_{k,l=0}^{M-1} |\mathbf{E}\left(R_k R_l\right)\phi(z_k)\phi(z_l)| \lesssim \frac{1}{M^2}\sum_{k,l=0}^{M-1} |\mathbf{E}\left(R_k R_l\right)| \\
&\leq \frac{1}{M^2N^2}\sum_{k,l=0}^{M-1}\sum_{i,j=0}^{N-1} |\mathbf{E}\left(r_{ik}r_{jl}\right)| = o(1),
\end{aligned}
$$

and, hence, the second summand in (2.41) converges to zero in probability. For the third summand, the same conclusion holds since

$$
\mathbf{E}\left(\frac{1}{M}\sum_{k=0}^{M-1} R_k^2\right) = \frac{1}{M}\sum_{k=0}^{M-1}\mathbf{E}\left(R_k^2\right) = \frac{1}{MN^2}\sum_{k=0}^{M-1}\sum_{i,j=0}^{N-1}\mathbf{E}\left(r_{ik}r_{jk}\right) = o(1)
$$

and L^1-convergence implies convergence in probability. $K_{N,M}^2$ can be handled similarly by considering a decomposition into two sums of non-overlapping increments:

$$
\bar{R}_k(\eta_0) = 2\left(\frac{1}{2N}\sum_{\substack{i\leq N-1 \\ i \text{ is even}}} \bar{r}_{ik}(\eta_0) + \frac{1}{2N}\sum_{\substack{i\leq N-1 \\ i \text{ is odd}}} \bar{r}_{ik}(\eta_0)\right)
$$

where $\bar{r}_{ik} = \bar{D}_{ik}^2/\sqrt{2\Delta} - f_{\eta_0}^2(z_k)$.

Step 3. Consistency of $\hat{\eta}_0$ follows from uniform convergence in probability of the contrast process. Since $K_{N,M}$ and K are continuous, this, in turn, follows from

$$
\forall\varepsilon > 0: \quad \lim_{h\to 0}\limsup_{M,N\to\infty}\mathbf{P}_{\eta_0}\left(\sup_{|\eta_0'-\eta_0''|<h} |K_{N,M}(\eta_0') - K_{N,M}(\eta_0'')| \geq \epsilon\right) = 0:
$$

By compactness of the parameter space, for each $a > 0$ there exists $h > 0$ such that $\|f_{\eta_0'}^i - f_{\eta_0''}^i\|_\infty, \|(f_{\eta_0'}^i)^2 - (f_{\eta_0''}^i)^2\|_\infty \leq a$ for all $|\eta_0' - \eta_0''| < h$. Therefore,

$$
\begin{aligned}
&|K_{N,M}^1(\eta_0') - K_{N,M}^1(\eta_0'')| \\
&\quad\leq \frac{2}{M}\sum_{k=0}^{M-1}\left(\frac{1}{N\sqrt{\Delta}}\sum_{i=0}^{N-1} D_{ik}^2\right)|f_{\eta_0''}^1(z_k) - f_{\eta_0'}^1(z_k)| + \frac{1}{M}\sum_{k=0}^{M-1}|f_{\eta_0'}^1(z_k)^2 - f_{\eta_0''}^1(z_k)^2| \\
&\quad\leq a\left(\frac{2}{M}\sum_{k=0}^{M-1}\left(\frac{1}{N\sqrt{\Delta}}\sum_{i=0}^{N-1} D_{ik}^2\right) + 1\right)
\end{aligned}
$$

and, hence,

$$
\begin{aligned}
&\limsup_{M,N\to\infty}\mathbf{P}_{\eta_0}\left(\sup_{|\eta_0'-\eta_0''|<h} |K_{N,M}^1(\eta_0') - K_{N,M}^1(\eta_0'')| \geq \varepsilon\right) \\
&\qquad\leq \limsup_{M,N\to\infty}\frac{1}{\varepsilon}\mathbf{E}\left(\sup_{|\eta_0'-\eta_0''|<h} |K_{N,M}^1(\eta_0') - K_{N,M}^1(\eta_0'')|\right)
\end{aligned}
$$

$$\leq \limsup_{M,N\to\infty} \frac{a}{\varepsilon}\mathbf{E}\left(\frac{2}{M}\sum_{k=0}^{M-1}\left(\frac{1}{N\sqrt{\Delta}}\sum_{i=0}^{N-1} D_{ik}^2\right)+1\right) \lesssim \frac{a}{\varepsilon}.$$

The same argument applies to $K^2_{N,M}$ and the result follows.

Step 4. Let $F_1, F_2 \in C^1([0,1])$ and $(a_k)_{k\in\mathbb{Z}}$ be absolutely summable. Then, we can write

$$\begin{aligned}\frac{1}{n}\sum_{k,l=0}^{n-1} a_{k-l}F_1(z_k)F_2(z_l) =& \frac{a_0}{n}\left(F_1(z_0)F_2(z_0)+\cdots+F_1(z_{n-1})F_2(z_{n-1})\right)\\ &+\frac{a_1}{n}\left(F_1(z_1)F_2(z_0)+\cdots+F_1(z_{n-1})F_2(z_{n-2})\right)\\ &+\frac{a_{-1}}{n}\left(F_1(z_0)F_2(z_1)+\cdots+F_1(z_{n-2})F_2(z_{n-1})\right)+\cdots\end{aligned}$$

and, consequently, we have $\frac{1}{n}\sum_{k,l=0}^{n-1} a_{k-l}F_1(z_k)F_2(z_l) \to \langle F_1, F_2\rangle_b \cdot \sum_{k\in\mathbb{Z}} a_k$, $n\to\infty$, by dominated convergence.

Step 5. We show that the asymptotic covariance matrix of $\sqrt{NM}\dot{K}_{N,M}(\eta_0)$ is given by U: We have $\dot{K}_{N,M}(\eta_0) = \dot{K}^1_{N,M}(\eta_0) + \dot{K}^1_{N,M}(\eta_0)$ as well as

$$\dot{K}^1_{N,M}(\eta_0) = -\frac{2}{M}\sum_{k=0}^{M-1}\left(\frac{1}{N\sqrt{\Delta}}\sum_{i=0}^{N-1} D_{ik}^2 - f^1_{\eta_0}(z_k)\right) g^1_{\eta_0}(z_k)$$

and similarly for $\dot{K}^2_{N,M}(\eta_0)$. Using Isserlis' covariance formula (2.7) in connection with (2.39) and noting $\bar{D}_{ik} = D_{ik} + D_{(i+1)k}$, it follows that

$$\begin{aligned}\mathrm{Cov}(D_{ik}^2, D_{jl}^2) &= 2\left(\sigma^2 \mathrm{e}^{-\frac{z_k+z_l}{2}}\xi^{\Delta}_{i-j,k-l} + \mathcal{O}\left(\frac{\Delta^{3/2}}{(|i-j|+1)^{3/2}}\right)\right)^2,\\ \mathrm{Cov}(\bar{D}_{ik}^2, \bar{D}_{jl}^2) &= 2\left(\sigma^2 \mathrm{e}^{-\frac{z_k+z_l}{2}}(2\xi^{\Delta}_{i-j,k-l} + \xi^{\Delta}_{i-j-1,k-l} + \xi^{\Delta}_{i-j+1,k-l}) + \mathcal{O}\left(\frac{\Delta^{3/2}}{(|i-j|+1)^{3/2}}\right)\right)^2,\\ \mathrm{Cov}(D_{ik}^2, \bar{D}_{jl}^2) &= 2\left(\sigma^2 \mathrm{e}^{-\frac{z_k+z_l}{2}}(\xi^{\Delta}_{i-j,k-l} + \xi^{\Delta}_{i-j-1,k-l}) + \mathcal{O}\left(\frac{\Delta^{3/2}}{(|i-j|+1)^{3/2}}\right)\right)^2.\end{aligned}$$

Now, for any $1 \leq e, f \leq 3$, the first summand in the expansion

$$\begin{aligned}\mathrm{Cov}((\dot{K}_{N,M})_e, (\dot{K}_{N,M})_f) =& \mathrm{Cov}((\dot{K}^1_{N,M})_e, (\dot{K}^1_{N,M})_f) + \mathrm{Cov}((\dot{K}^2_{N,M})_e, (\dot{K}^2_{N,M})_f)\\ &+ \mathrm{Cov}((\dot{K}^1_{N,M})_e, (\dot{K}^2_{N,M})_f) + \mathrm{Cov}((\dot{K}^2_{N,M})_e, (\dot{K}^1_{N,M})_f)\end{aligned} \tag{2.42}$$

is given by

$$\mathrm{Cov}((\dot{K}^1_{N,M})_e, (\dot{K}^1_{N,M})_f) = \frac{4}{M^2N^2\Delta}\sum_{i,j=0}^{N-1}\sum_{k,l=0}^{M-1}\mathrm{Cov}(D_{ik}^2, D_{jl}^2)\,(g^1_{\eta_0})_e(z_k)\,(g^1_{\eta_0})_f(z_l).$$

Like in the proof of Theorem 2.2.7, the covariances may be replaced by their asymptotic expressions, due to dominated convergence. Further, using $(h^i_{\eta_0})_e(z) = \mathrm{e}^{-\kappa z}(g^i_{\eta_0})_e(z)$ and Step 4, we have

$$MN\cdot \mathrm{Cov}((\dot{K}^1_{N,M})_e, (\dot{K}^1_{N,M})_f) \to 8\sigma^4\sum_{i,k\in\mathbb{Z}}(A^r_{i,k})^2\cdot\langle(h^1_{\eta_0})_e, (h^1_{\eta_0})_f\rangle_b, \qquad M,N\to\infty.$$

Analogously,

$$MN\cdot \mathrm{Cov}((\dot{K}^2_{N,M})_e, (\dot{K}^2_{N,M})_f) \to 4\sigma^4\sum_{i,k\in\mathbb{Z}}(B^r_{i,k})^2\cdot\langle(h^2_{\eta_0})_e, (h^2_{\eta_0})_f\rangle_b, \qquad M,N\to\infty,$$

$$MN \cdot \mathrm{Cov}((\dot{K}^1_{N,M})_e, (\dot{K}^2_{N,M})_f) \to 4\sqrt{2}\sigma^4 \sum_{i,k\in\mathbb{Z}} (C^r_{i,k})^2 \cdot \langle (h^1_{\eta_0})_e, (h^2_{\eta_0})_f\rangle_b, \qquad M, N \to \infty,$$

and insertion into (2.42) yields the claimed asymptotic covariance matrix.

Step 6. U is strictly positive definite: It is sufficient to show $C_r < \sqrt{A_r B_r}$, then it follows for any $\alpha \in \mathbb{R}^3 \setminus \{0\}$ and $H^i_\alpha = \sum_{j=1}^3 \alpha_j (h^i_{\eta_0})_j$, $i = 1, 2$, that

$$\begin{aligned}\alpha^\top U\alpha &= 4\sigma^4 \left(2A_r\|H^1_\alpha\|_b^2 + B_r\|H^2_\alpha\|_b^2 + 2\sqrt{2}C_r\langle H^1_\alpha, H^2_\alpha\rangle_b\right)\\ &> 4\sigma^4 \left(2A_r\|H^1_\alpha\|_b^2 + B_r\|H^2_\alpha\|_b^2 + 2\sqrt{2A_rB_r}\langle H^1_\alpha, H^2_\alpha\rangle_b\right)\\ &= 8\sigma^4 \left\|\sqrt{2A_r}H^1_\alpha + \sqrt{B}_r H^2_\alpha\right\|_b^2 \geq 0\end{aligned}$$

where we may assume $\langle H^1_\alpha, H^2_\alpha\rangle_b < 0$ since, otherwise, $\alpha^\top U\alpha > 0$ follows immediately from the first equality. Now, consider $(A^r_{i,k})$ and $(B^r_{i,k})$ as elements in the Hilbert space ℓ^2 of square summable sequences indexed by $\mathbb{Z}\times\mathbb{Z}$. Clearly, $A_r = \|(A^r_{i,k})\|^2_{\ell^2}$, $B_r = \|(B^r_{i,k})\|^2_{\ell^2}$ and a direct calculation shows that $C_r = \langle (A^r_{i,k}), (B^r_{i,k})\rangle_{\ell^2}$. Thus, by the Cauchy-Schwarz inequality, we have $C_r \leq \sqrt{A_rB_r}$ and equality is ruled out by the fact that $(A^r_{i,k})$ and $(B^r_{i,k})$ are not linearly dependent.

Step 7. We show $\sqrt{NM}\dot{K}^1_{N,M}(\eta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, U)$ under $\mathbf{P}_{\eta_0}$: In view of the Cramér-Wold device, we have to prove $\sqrt{NM}\alpha^\top \dot{K}_{N,M} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha^\top U\alpha)$ for any $\alpha \in \mathbb{R}^3$. Let s_{ik} and Z_{ik} be given by the relation $s_{ik}Z^2_{ik} = -\frac{2\alpha^T \dot{f}^1_{\eta_0}(z_k)}{\sqrt{NM\Delta}} D^2_{ik}$ where $s_{ik} \in \{-1, 1\}$ is deterministic. Analogously, define $\bar{s}_{ik}$ and $\bar{Z}^2_{i,k}$. Then, $\mathcal{Z}_{N,M} = (Z_{ik}, \bar{Z}_{j,l})_{i,j,k,l}$ is a Gaussian vector and from Proposition 2.2.5, it follows that

$$\sqrt{NM}\alpha^\top \dot{K}_{N,M}(\eta_0) = S_{N,M} - \mathbf{E}(S_{N,M}) + o(1)$$

where $S_{N,M} = \sum_{i=0}^{N-1}\sum_{k=0}^{M-1} s_{ik}Z^2_{ik} + \sum_{i=0}^{N-1}\sum_{k=0}^{M-2} \bar{s}_{ik}\bar{Z}^2_{ik}$. From Steps 5 and 6, we can deduce that $\mathrm{Var}\,(S_{N,M}) \to \alpha^\top U\alpha > 0$, $N, M \to \infty$, and, thus, in view of criterion (2.8), asymptotic normality follows if the absolute row sums of the covariance matrix of $\mathcal{Z}_{N,M}$ vanish uniformly. This, in turn, is a simple consequence of (2.39) and (2.40).

Step 8. In order to prove $\int_0^1 \ddot{K}_{N,M}(\eta_0 + \tau(\hat{\eta}_0 - \eta_0))\,d\tau \xrightarrow{\mathbf{P}_{\eta_0}} V(\eta_0)$, we show $\ddot{K}_{N,M}(\check{\eta}_0) \xrightarrow{\mathbf{P}_{\eta_0}} V(\eta_0)$ for any consistent estimator $\check{\eta}_0$ of η_0: We have

$$\ddot{K}^1_{N,M}(\eta_0) = \frac{2}{M}\sum_{k=0}^{M-1} g^1_{\eta_0}(z_k)g^1_{\eta_0}(z_k)^\top - \frac{2}{M}\sum_{k=0}^{M-1}\left(\frac{1}{N\sqrt{\Delta}}\sum_{i=0}^{N-1} D^2_{ik} - f^1_{\eta_0}(z_k)\right)\ddot{f}^1_{\eta_0}(z_k)$$

and analogously for $\ddot{K}^2_{N,M}$. By using $\mathbf{P}_{\eta_0}(\check{\eta}_0 \in H_0) \to 1$ and the uniform continuity of $f^i_{\eta_0}(z)$ and its derivatives in the parameter $(z, \eta_0) \in [0,1] \times H_0$, it is straightforward to show $\ddot{K}_{N,M}(\check{\eta}_0) - \ddot{K}_{N,M}(\eta_0) \xrightarrow{\mathbf{P}_{\eta_0}} 0$. Now, write $V = 2(V^1 + V^2)$ where V^i is the Gram matrix of the functions $\{(g^i_{\eta_0})_1, (g^i_{\eta_0})_2, (g^i_{\eta_0})_3\}$ with respect to the inner product $\langle\cdot,\cdot\rangle_b$, i.e., $V^i_{ef} = \langle (g^i_{\eta_0})_e, (g^i_{\eta_0})_f\rangle_b$, $1 \leq e, f \leq 3$. Clearly, the first summand of $\ddot{K}^1_{N,M}(\eta_0)$ converges to $2V^1$ while the calculations of Step 2 show that the second summand converges to 0 in probability. The same reasoning holds for $\ddot{K}^2_{N,M}(\eta_0)$ and the result follows.

Step 9. V is strictly positive definite: Being Gram matrices, V^1 and V^2 are positive semidefinite and, consequently, the same holds for V. Clearly, the only way V can be singular is if there exists $\alpha \in \mathbb{R}^3$ such that $0 = \alpha^\top V^i\alpha = \left\|\sum_{e=1}^3 \alpha_e(g^i_{\eta_0})_e\right\|_b^2$ holds for both $i \in \{1, 2\}$. From the particular form of the functions $(g^i_{\eta_0})_e$ it is apparent that this would imply that $\alpha_1\psi_{\vartheta_2}(r_i) + \alpha_2\sigma^2\frac{\partial\psi_{\vartheta_2}(r_i)}{\partial\vartheta_2} = \alpha_3 = 0$ for both $i \in \{1, 2\}$, which is impossible. □

Corollary 2.2.14 is a direct consequence of the previous proof.

Proof of Corollary 2.2.14. We have to prove

$$\forall \varepsilon > 0 \, \exists C > 0 : \limsup_{N,M\to\infty} \mathbf{P}_\eta \left(\sqrt{M^3 \wedge N^{3/2}} \|\hat{\eta}_{v,w} - \eta\| \geq C \right) \leq \varepsilon$$

or, equivalently,

$$\forall \varepsilon > 0 \, \exists C > 0 : \limsup_{N,M\to\infty} \mathbf{P}_{\eta_0} \left(\sqrt{M^3 \wedge N^{3/2}} \|\hat{\eta}_0^{v,w} - \eta_0\| \geq C \right) \leq \varepsilon$$

where $\hat{\eta}_0^{v,w}$ is the minimizer of $\mathcal{K}_{N,M}$ in terms of the parametrization $\eta_0 = (\sigma^2, \vartheta_2, \kappa)$, as in the proof of Theorem 2.2.12. In fact, similar calculations as in the latter proof show that its Steps 1-3 and 8-9 remain valid. Consequently, we have the representation $-\dot{\mathcal{K}}_{N,M}(\eta_0) = V_{N,M}(\hat{\eta}_0^{v,w}, \eta_0)(\hat{\eta}_0^{v,w} - \eta_0)$ where $V_{N,M}(\tilde{\eta}_0, \eta_0) = \int_0^1 \ddot{\mathcal{K}}_{N,M}(\eta_0 + \tau(\tilde{\eta}_0 - \eta_0))\, d\tau$ as well as $V_{N,M}(\hat{\eta}_0^{v,w}, \eta_0) \xrightarrow{\mathbf{P}_{\eta_0}} V(\eta_0)$, where $V(\eta_0)$ is an invertible deterministic matrix. In particular, the set

$$A_{N,M} := \left\{ V_{N,M}(\hat{\eta}_0^{v,w}, \eta_0) \text{ is invertible with } \|V_{N,M}(\hat{\eta}_0^{v,w}, \eta_0)^{-1}\|_2 \leq \|V(\eta_0)^{-1}\|_2 + 1 \right\}$$

satisfies $\mathbf{P}_{\eta_0}(A_{N,M}) \to 1$. Further, $\dot{\mathcal{K}}_{N,M}(\eta_0)$ can be written as an average of expressions of the type $\dot{K}_{N,M}$ from Theorem 2.2.12 so that the calculations of Step 5 show together with the Cauchy-Schwarz inequality that $\mathbf{E}_{\eta_0}\left(\|\dot{\mathcal{K}}_{N,M}(\eta_0)\|^2\right) = \mathcal{O}((M^3 \wedge N^{3/2})^{-1})$. Now,

$$\begin{aligned}\mathbf{P}_{\eta_0} \left(\sqrt{M^3 \wedge N^{3/2}} \|\hat{\eta}_0^{v,w} - \eta_0\| \geq C \right) \leq & \mathbf{P}_{\eta_0} \left(\left\{ \sqrt{M^3 \wedge N^{3/2}} \|\hat{\eta}_0^{v,w} - \eta_0\| \geq C \right\} \cap A_{N,M} \right) \\ & + \mathbf{P}_{\eta_0}(A_{N,M}^{\mathrm{c}}).\end{aligned}$$

The second summand becomes arbitrarily small as $M, N \to \infty$. For the first summand, let $\gamma(\eta_0) := \|V(\eta_0)^{-1}\|_2 + 1$, then it follows from Markov's inequality that

$$\begin{aligned}&\mathbf{P}_{\eta_0} \left(\{ \sqrt{M^3 \wedge N^{3/2}} \|\hat{\eta}_0^{v,w} - \eta_0\| \geq C \} \cap A_{N,M} \right) \\ &= \mathbf{P}_{\eta_0} \left(\{ \sqrt{M^3 \wedge N^{3/2}} \|V_{N,M}(\hat{\eta}_0^{v,w}, \eta_0)^{-1} \dot{\mathcal{K}}_{N,M}(\eta_0)\| \geq C \} \cap A_{N,M} \right) \\ &\leq \mathbf{P}_{\eta_0} \left(\{ \sqrt{M^3 \wedge N^{3/2}} \|\dot{\mathcal{K}}_{N,M}(\eta_0)\| \geq \frac{C}{\gamma(\eta_0)} \} \cap A_{N,M} \right) \\ &\leq \mathbf{P}_{\eta_0} \left(\sqrt{M^3 \wedge N^{3/2}} \|\dot{\mathcal{K}}_{N,M}(\eta_0)\| \geq \frac{C}{\gamma(\eta_0)} \right) \leq (M^3 \wedge N^{3/2}) \mathbf{E}_{\eta_0}(\|\dot{\mathcal{K}}_{N,M}(\eta_0)\|^2) \frac{\gamma(\eta_0)^2}{C^2} \lesssim \frac{1}{C^2}. \quad \square\end{aligned}$$

Before analyzing our estimator for the whole parameter vector (σ^2, ϑ), we prove the central limit theorem for the rescaled sum of squares corresponding to the discrete observations of X.

Proof of Proposition 2.2.16. First, let us consider the case of a stationary initial distribution: By regarding the mean of S as a Riemann sum and noting that $y \mapsto \rho_{yy}(0)$ is continuously differentiable, we get

$$\mathbf{E}(S) = \frac{1}{1-2b} \int_b^{1-b} \rho_{yy}(0)\, dy + \mathcal{O}\Big(\frac{1}{M}\Big).$$

Hence, the second statement is a direct consequence of the first. To compute the asymptotic variance, define $S_x := \frac{1}{N} \sum_{i=0}^{N-1} X_{t_i}^2(x) \mathrm{e}^{\kappa x}$. We have

$$\begin{aligned}\mathrm{Var}(S) &= \frac{1}{M^2} \sum_{k,l=0}^{M-1} \mathrm{Cov}(S_{y_k}, S_{y_l}) \\ &= \frac{1}{M^2} \sum_{k,l=0}^{M-1} \frac{2}{T} \int_{-\infty}^{\infty} \rho_{y_k y_l}^2(t)\, dt + \frac{1}{M^2} \sum_{k,l=0}^{M} R_{y_k y_l}\end{aligned}$$

where it follows from Isserlis' Covariance formula (2.7) that

$$\begin{aligned}R_{xy} &= \frac{2}{N^2}\sum_{i,j=0}^{N-1}\rho_{xy}^2(t_i-t_j) - \frac{2}{T}\int_{-\infty}^{\infty}\rho_{xy}^2(t)\,dt\\ &= \frac{2}{N^2}\sum_{j=0}^{N-1}\sum_{i=-j}^{j}\rho_{xy}^2(i\Delta) - \frac{2}{T}\int_{-\infty}^{\infty}\rho_{xy}^2(t)\,dt.\end{aligned}$$

For $t>0$, we have $|\rho_{xy}(t)| \lesssim \mathrm{e}^{-\lambda_1 t}\sum_{\ell\geq 1}\frac{1}{\lambda_\ell} \lesssim \mathrm{e}^{-\pi^2\vartheta_2 t}$ and, therefore, $\int_T^\infty \rho_{xy}^2(t)\,dt \to 0$ for $T\to\infty$ uniformly in $x,y\in[b,1-b]$. Consequently, by taking a Cesàro limit,

$$\begin{aligned}R_{xy} &= \frac{2}{T}\frac{1}{N}\sum_{j=0}^{N-1}\left(\Delta\sum_{i=-j}^{j}\rho_{xy}^2(i\Delta) - \int_{-\infty}^{\infty}\rho_{xy}^2(t)\,dt\right)\\ &= \frac{2}{T}\frac{1}{N}\sum_{j=0}^{N-1}\left(\Delta\sum_{i=-j}^{j}\rho_{xy}^2(i\Delta) - \int_{-(j+1/2)\Delta}^{(j+1/2)\Delta}\rho_{xy}^2(t)\,dt\right) + o\Big(\frac{1}{T}\Big).\end{aligned}$$

Since $\rho_{xy}(\cdot)$ is bounded uniformly in x and y, we have $|\frac{d}{dt}\rho_{xy}^2(t)| = \big|2\rho_{xy}(t)\frac{d}{dt}\rho_{xy}(t)\big| \lesssim \big|\frac{d}{dt}\rho_{xy}(t)\big|$. Further, there exists $C>0$ such that

$$\big|\frac{d}{dt}\rho_{xy}(t)\big| \lesssim \sum_{\ell\geq 1}\mathrm{e}^{-C\ell^2|t|} \lesssim \mathrm{e}^{-C|t|} + \int_1^\infty \mathrm{e}^{-C|t|u^2}\,du \lesssim \begin{cases}|t|^{-1/2}, & |t|\leq 1,\\ \mathrm{e}^{-C|t|}, & |t|\geq 1\end{cases}$$

and, hence, $\frac{d}{dt}\rho_{xy}^2(\cdot)$ is integrable over $\mathbb{R}$. Now, using some intermediate points $\xi_i\in[(i-\frac{1}{2})\Delta,(i+\frac{1}{2})\Delta]$, we can deduce that

$$\begin{aligned}\left|\Delta\sum_{i=-j}^{j}\rho_{xy}^2(i\Delta) - \int_{-(j+1/2)\Delta}^{(j+1/2)\Delta}\rho_{xy}^2(t)\,dt\right| &= \left|\sum_{i=-j}^{j}\int_{(i-1/2)\Delta}^{(i+1/2)\Delta}\big(\rho_{xy}^2(i\Delta)-\rho_{xy}^2(t)\big)\,dt\right|\\ &\leq \sum_{i=-\infty}^{\infty}\int_{(i-1/2)\Delta}^{(i+1/2)\Delta}\big|\rho_{xy}^2(i\Delta)-\rho_{xy}^2(t)\big|\,dt\\ &\leq \Delta^2\sum_{i=-\infty}^{\infty}\big|\frac{d}{dt}\rho_{xy}^2(\xi_i)\big| \lesssim \Delta\end{aligned}$$

uniformly in x,y and j. Therefore, $R_{xy}=o(1/T)$ holds uniformly in x and y and

$$\begin{aligned}\mathrm{Var}(S) &= \frac{1}{M^2}\sum_{k,l=0}^{M}\frac{2}{T}\int_{-\infty}^{\infty}\rho_{y_k y_l}^2(t)\,dt + o\Big(\frac{1}{T}\Big)\\ &= \frac{2}{T(1-2b)^2}\int_{-\infty}^{\infty}\int_b^{1-b}\int_b^{1-b}\rho_{xy}^2(t)\,dx\,dy\,dt + o\Big(\frac{1}{T}\Big).\end{aligned}$$

Finally, in order to prove asymptotic normality, we make use of Proposition 2.2.1, which states that is sufficient to prove that the maximum absolute row sum of the covariance matrix of the random vector $(\frac{T^{1/4}}{\sqrt{MN}}X_{t_i}(y_k)\mathrm{e}^{\kappa y_k/2})_{i,k}$ vanishes asymptotically. Now, using $|\rho_{xy}(t)|\lesssim \mathrm{e}^{-\lambda_1 t}$, the absolute row sum can be bounded by

$$\sup_{i,k}\frac{\sqrt{T}}{MN}\sum_{j=0}^{N-1}\sum_{l=0}^{M-1}|\rho_{y_k y_l}(t_i-t_j)| \leq \sup_k \frac{\sqrt{T}}{MN}\sum_{l=0}^{M-1}\sum_{j=-\infty}^{\infty}|\rho_{y_k y_l}(t_j)|$$

$$\leq \frac{\sqrt{T}}{N} \sum_{j=-\infty}^{\infty} \mathrm{e}^{-\lambda_1 t_j} \lesssim \frac{1}{\sqrt{T}} \to 0$$

where the last bound follows by regarding the sum as a Riemann sum with lag Δ. Hence, asymptotic normality follows, finishing the proof for the stationary case.

To show the statement for the case $X_0 = 0$, we proceed as in Step 4 of the proof of Theorem 2.2.3. Again, let X_t^0 be the process (1.12) with $\xi = 0$ and $\xi_t := S(t)\xi$ where ξ follows the stationary distribution and is independent of X^0. We have

$$\mathbf{E}(\xi_{t_i}(y_k)\xi_{t_j}(y_l)) = \sum_{\ell\geq 1} \frac{\sigma^2}{2\lambda_\ell} \mathrm{e}^{-\lambda_\ell(t_i+t_j)} e_\ell(y_k)e_\ell(y_l) \leq \mathrm{e}^{-\lambda_1(t_i+t_j)} \sum_{\ell\geq 1} \frac{\sigma^2}{2\lambda_\ell} |e_\ell(y_k)e_\ell(y_l)| \lesssim \mathrm{e}^{-\lambda_1(t_i+t_j)}$$

and, therefore, the Riemann sum argument yields

$$\frac{1}{MN} \sum_{i=1}^{N} \sum_{k=0}^{M-1} \mathbf{E}(\xi_{t_i}^2(y_k)) \lesssim \frac{1}{N} \sum_{i=1}^{N} \mathrm{e}^{-2\lambda_1 t_i} = \frac{\Delta}{T} \sum_{i=1}^{N} \mathrm{e}^{-2\lambda_1 t_i} \lesssim \frac{1}{T} = o\Big(\frac{1}{\sqrt{T}}\Big).$$

The cross terms are also negligible due to $\mathbf{E}\big(X_{t_i}^0(y_k)X_{t_j}^0(y_l)\big) = \mathcal{O}(1)$ and

$$\frac{1}{(MN)^2} \sum_{i,j=1}^{N1} \sum_{k,l=0}^{M-1} \mathbf{E}\big(\xi_{t_i}(y_k)\xi_{t_j}(y_l)\big)\mathbf{E}\big(X_{t_i}^0(y_k)X_{t_j}^0(y_l)\big) \lesssim \frac{1}{N^2} \sum_{i,j=1}^{N} \mathrm{e}^{-\lambda_1(t_i+t_j)} \lesssim \frac{1}{T^2} = o\Big(\frac{1}{T}\Big).$$

Thus, in view of Slutsky's Lemma, the central limit theorem carries over from the stationary case. □

Next, we will prove Theorem 2.2.18. Like in the proof of Theorem 2.2.12, we consider the estimator $\hat{\eta}_0$, which is defined as the minimizer of $K_{N,M}$ in terms of $\eta_0 = (\sigma^2, \vartheta_2, \kappa)$. Due to the delta method, it is sufficient to prove

$$\begin{pmatrix} \sqrt{MN}(\hat{\eta}_0 - \eta_0) \\ \sqrt{T}(\hat{\vartheta}_0 - \vartheta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} \Pi_{\eta_0}^r & 0 \\ 0 & \alpha_{\sigma^2,\vartheta}^2 \end{pmatrix}\right), \tag{2.43}$$

with $\Pi_{\eta_0}^r := V^{-1}UV^{-1}$ and V and U from (2.38). Again, we assume that X_0 follows the stationary distribution and refer to the approximation steps from the proofs of Proposition 2.2.16 and Theorem 2.2.7 for the case $X_0 = 0$. The first step of the proof is given by the following lemma.

Lemma 2.4.1. *Consider the situation of Theorem 2.2.18 and assume (ST). For $N, M, T \to \infty$, we have*

$$\begin{pmatrix} \sqrt{MN}(\hat{\eta}_0 - \eta_0) \\ \sqrt{T}(S(\kappa) - \sigma^2 I_b(\Gamma)/\vartheta_2) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} \Pi_{\eta_0}^r & 0 \\ 0 & 2D^2 \end{pmatrix}\right)$$

with D^2 from (2.21).

Proof. The central limit theorem for $\hat{\eta}_0$ in the case $T \to \infty$ can be shown in exactly the same way as for a bounded time horizon, noting that the underlying central limit theorem for double increments (Theorem 2.2.7) is also valid in the regime $T \to \infty$, as long as $T\sqrt{\Delta} \to 0$. In particular, like in the proof of Theorem 2.2.12, we have the representation

$$\sqrt{MN}(\hat{\eta}_0 - \eta_0) = -\sqrt{MN} V(\eta_0)^{-1} \dot{K}_{M,N}(\eta_0) + o_p(1)$$

where $V(\eta_0) \in \mathbb{R}^{3\times 3}$ is a strictly positive definite matrix such that $\ddot{K}_{M,N}(\eta_0) \xrightarrow{\mathbf{P}} V(\eta_0)$. Hence, it is sufficient to show the statement with $\sqrt{MN}(\hat{\eta}_0 - \eta_0)$ replaced by $-\sqrt{MN}V(\eta_0)^{-1}\dot{K}_{M,N}(\eta_0)$. Furthermore, as discussed in Step 8 of the proof of Theorem 2.2.12, the bias of the latter expression is negligible for the central limit theorem. The bias of $S(\kappa)$ is negligible since $\sqrt{T}/M \approx \sqrt{T\Delta} \to 0$ holds under the condition $T\sqrt{\Delta} \to 0$.

Step 1. We have

$$\begin{aligned}\operatorname{Cov}(D_{ik}, X_{t_j}(y_l)) &= \sigma^2 \sum_{\ell\geq 1} \frac{\mathrm{e}^{-\lambda_\ell |i-j+1|\Delta} - \mathrm{e}^{-\lambda_\ell |i-j|\Delta}}{2\lambda_\ell} \big(e_\ell(y_{k+1}) - e_\ell(y_k)\big) e_\ell(y_l) \\ &\lesssim \sum_{\ell\geq 1} \frac{\mathrm{e}^{-\lambda_\ell |i-j+1|\Delta} - \mathrm{e}^{-\lambda_\ell |i-j|\Delta}}{\lambda_\ell}.\end{aligned}$$

Thus, for $|i-j| \leq 1$, the usual Riemann sum argument shows that $\operatorname{Cov}(D_{ik}, X_{t_j}(y_l)) = \mathcal{O}(\sqrt{\Delta})$. In case $|i-j| > 1$ the mean value theorem yields $\operatorname{Cov}(D_{ik}, X_{t_j}(y_l)) \lesssim \Delta \sum_{\ell\geq 1} \mathrm{e}^{-\lambda_\ell |i-j|\Delta} = \mathcal{O}\Big(\sqrt{\frac{\Delta}{|i-j|}}\Big)$. Consequently, we have for all i, j, k, l that

$$\operatorname{Cov}(D_{ik}, X_{t_j}(y_l)) \lesssim \Delta \sum_{\ell\geq 1} \mathrm{e}^{-\lambda_\ell |i-j|\Delta} = \mathcal{O}\Big(\sqrt{\frac{\Delta}{|i-j|+1}}\Big). \tag{2.44}$$

Step 2. We calculate the asymptotic (co)variance of the estimator: Since the individual asymptotic (co)variances of $\hat{\eta}_0$ and $S(\kappa)$ are already given by Theorem 2.2.12 and Proposition 2.2.16, respectively, it remains to show $\sqrt{MNT}\operatorname{Cov}((V(\eta_0)^{-1}\dot{K}_{M,N}(\eta_0))_i, S(\kappa)) \to 0$ for $i \in \{1,2,3\}$. Letting $W := V^{-1}(\eta_0)$, we have

$$\operatorname{Cov}((W\dot{K}_{M,N}(\eta_0))_i, S(\kappa)) = \sum_{j=1}^{3} W_{ij} \operatorname{Cov}(\partial_j K^1_{M,N}(\eta_0), S(\kappa)) + \sum_{j=1}^{3} W_{ij} \operatorname{Cov}(\partial_j K^2_{M,N}(\eta_0), S(\kappa))$$

as well as

$$\begin{aligned}\operatorname{Cov}(\partial_j K^1_{M,N}(\eta_0), S(\kappa)) &= \frac{2}{MN\sqrt{\Delta}} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \operatorname{Cov}(D^2_{ik}, S(\kappa))\, \partial_j f^1_{\eta_0}(z_k) \\ &= \frac{2}{M^2N^2\sqrt{\Delta}} \sum_{i,j=0}^{N-1} \sum_{k,l=0}^{M-1} \operatorname{Cov}(D^2_{ik}, X^2_{t_j}(y_l))\, \mathrm{e}^{\kappa y_l} \partial_j f^1_{\eta_0}(z_k) \\ &\lesssim \frac{1}{M^2N^2\sqrt{\Delta}} \sum_{i,j=0}^{N-1} \sum_{k,l=0}^{M-1} \operatorname{Cov}(D_{ik}, X_{t_j}(y_l))^2\end{aligned}$$

and similarly for $K^2_{M,N}$. Using Step 1, we can bound

$$\begin{aligned}\operatorname{Cov}(\partial_j K_{M,N}(\eta_0), S(\kappa)) &\lesssim \frac{1}{M^2N^2\sqrt{\Delta}} \sum_{i,j=0}^{N-1} \sum_{k,l=0}^{M-1} \frac{\Delta}{|i-j|+1} \\ &\lesssim \frac{1}{M^2N^2\sqrt{\Delta}} M^2 N \log(N)\Delta = \frac{\sqrt{\Delta}\log N}{N} \lesssim \frac{\sqrt{\Delta}}{N^{1-\alpha}}\end{aligned}$$

for any $\alpha > 0$. Consequently, using $M \approx \Delta^{-1/2}$ and taking $\alpha = 1/4$,

$$\sqrt{MNT}\operatorname{Cov}((V(\eta_0)^{-1}\dot{K}_{M,N}(\eta_0))_i, S(\kappa)) \lesssim \sqrt{MNT}\frac{\sqrt{\Delta}}{N^{1-\alpha}} = N^\alpha \Delta^{3/4} = (T\Delta)^{1/4}\Delta^{1/4} \to 0.$$

Step 3. To prove asymptotic normality, we employ Proposition 2.2.1. A similar reasoning as in Step 7 of the proof of Theorem 2.2.12 shows that it is sufficient to prove that the maximal absolute row sum of the covariance matrix of the random vector

$$\Big(\underbrace{\frac{1}{(MN\Delta)^{1/4}}D_{ik}}_{=:A_{ik}}, \underbrace{\frac{1}{(MN\Delta)^{1/4}}\bar{D}_{ik}}_{=:B_{ik}}, \underbrace{\frac{T^{1/4}}{\sqrt{MN}}X_{t_i}(y_k)}_{=:C_{ik}}\Big)_{\substack{0\leq i<N\\ 0\leq k<M}}$$

converges to zero. As in Step 2, it only remains to consider the covariances between the double increments and the values of X. We show

$$\sup_{i,k}\sum_{j,l}\Big(|\mathrm{Cov}(C_{ik},A_{jl})|+|\mathrm{Cov}(C_{ik},B_{jl})|+|\mathrm{Cov}(A_{ik},C_{jl})|+|\mathrm{Cov}(B_{ik},C_{jl})|\Big)\to 0:$$

We only consider the first term, the other three can be treated similarly. Following Step 1, we have

$$\begin{aligned}\sum_{j,l}|\mathrm{Cov}(C_{ik},A_{jl})| &\lesssim \frac{T^{1/4}}{(MN)^{3/4}\Delta^{1/4}}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}\sqrt{\frac{\Delta}{|i-j|+1}} \lesssim \frac{T^{1/4}}{(MN)^{3/4}\Delta^{1/4}}M\sqrt{N}\sqrt{\Delta}\\ &\lesssim M^{1/4}\sqrt{\Delta}\approx \Delta^{3/8}\to 0.\end{aligned}$$

□

We conclude the central limit theorem for our estimator of the whole parameter vector:

Proof of Theorem 2.2.18. As already remarked, the claim follows if we show (2.43) under assumption (ST).

Step 1. $\sqrt{T}(\hat{\eta}_0-\eta_0)\xrightarrow{\mathbf{P}}0$: By Slustsky's Lemma, we have

$$\sqrt{T}(\hat{\eta}_0-\eta_0) = \underbrace{\frac{\sqrt{T}}{\sqrt{MN}}}_{=\sqrt{\Delta/M}\to 0}\sqrt{MN}(\hat{\eta}_0-\eta_0)\xrightarrow{\mathcal{D}}0.$$

The assertion follows since convergence in distribution to a constant implies convergence in probability.

Step 2. $\sqrt{T}(S(\hat{\kappa})-S(\kappa))\xrightarrow{\mathbf{P}}0$. In particular, $S(\hat{\kappa})$ is a consistent estimator for $S_\infty:=\frac{\sigma^2}{\vartheta_2}I_b(\Gamma)$: We have

$$\sqrt{T}|S(\kappa)-S(\hat{\kappa})| = \frac{1}{MN}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}X_{t_i}^2(y_k)\mathrm{e}^{\bar{\kappa}y_k}\sqrt{T}|\hat{\kappa}-\kappa|$$

for some $\bar{\kappa}$ between $\hat{\kappa}$ and κ. The expression $\mathrm{e}^{\bar{\kappa}y_k}$ is bounded due to the compactness assumption on the parameter space and $\frac{1}{MN}\sum_{i=0}^{N-1}\sum_{k=0}^{M-1}X_{t_i}^2(y_k)\mathrm{e}^{\kappa y_k}$ converges in probability to S_∞. Thus, the claim follows from Step 1.

Step 3. Define $H(\eta_0,S):=\vartheta_2\left(\frac{\kappa^2}{4}-I_b^{-1}(\frac{\vartheta_2}{\sigma^2}S)\right)$ so that $\hat{\vartheta}_0=H(\hat{\eta}_0,S(\hat{\kappa}))$. We show $\sqrt{T}(\hat{\vartheta}_0-\vartheta_0)=\sqrt{T}\partial_S H(\eta_0,S_\infty)(S(\kappa)-S_\infty)+o_p(1)$ for some $\bar{S}$ between $S(\hat{\kappa})$ and S_∞: By the mean value theorem, there exist $\bar{\eta}_0$ between $\hat{\eta}_0$ and η_0 and $\bar{S}$ between $S(\hat{\kappa})$ and S_∞ such that

$$\begin{aligned}\sqrt{T}(\hat{\vartheta}_0-\vartheta_0) &= \sqrt{T}(H(\hat{\eta}_0,S(\hat{\kappa}))-H(\eta_0,S_\infty))\\ &= \sqrt{T}(H(\hat{\eta}_0,S(\hat{\kappa}))-H(\eta_0,S(\hat{\kappa})))+\sqrt{T}(H(\eta_0,S(\hat{\kappa}))-H(\eta_0,S_\infty))\\ &= \sqrt{T}\nabla_{\eta_0}H(\bar{\eta}_0,S(\hat{\kappa}))(\hat{\eta}_0-\eta_0)+\sqrt{T}\partial_S H(\eta_0,\bar{S})(S(\hat{\kappa})-S(\kappa))\\ &\quad+\sqrt{T}\big(\partial_S H(\eta_0,\bar{S})-\partial_S H(\eta_0,S_\infty)\big)(S(\kappa)-S_\infty)\\ &\quad+\sqrt{T}\partial_S H(\eta_0,S_\infty)(S(\kappa)-S_\infty). \qquad (2.45)\end{aligned}$$

Steps 1 and 2 as well as consistency show that the first three terms converge to 0 in probability, from which the result follows.

Step 4. Step 3 in combination with Slutsky's Lemma shows that the limiting distributions of

$$\begin{pmatrix} \sqrt{MN}(\hat{\eta}_0 - \eta_0) \\ \sqrt{T}(\hat{\vartheta}_0 - \vartheta_0) \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sqrt{MN}(\hat{\eta}_0 - \eta_0) \\ \sqrt{T}\big(\partial_S H(\eta_0, S_\infty)\big)(S(\kappa) - S_\infty) \end{pmatrix}$$

agree and, following Lemma 2.4.1 and the delta method, it is given by

$$\mathcal{N}\left(0, \begin{pmatrix} \Pi^r_{\eta_0} & 0 \\ 0 & 2D^2\big(\partial_S H(\eta_0, S_\infty)\big)^2 \end{pmatrix}\right).$$

The proof is now finalized by noting

$$\partial_S H(\eta_0, S_\infty) = -\frac{\vartheta_2^2}{\sigma^2 I_b'(\Gamma)}. \qquad \square$$

The rate of convergence in case of a general, not necessarily balanced, sampling design is a simple consequence:

Proof of Corollary 2.2.19. The statement on $\hat{\eta}_{vw}$ can be proved in the same way as for a bounded time horizon. The validity of the central limit theorem for $\hat{\vartheta}_0^{vw}$ follows from the fact that the first three terms in (2.45) also tend to zero in probability when $\hat{\eta}_0$ is replaced by $\hat{\eta}_0^{vw}$. $\square$

2.4.4 Auxiliary results

Auxiliary results for the lower bounds

For Proposition 2.1.4 and Theorem 2.1.7 (*ii*) we require the following auxiliary lemma.

Lemma 2.4.2. *Consider a discrete sample $(u(i\Delta),\ i = 0, \ldots, N)$ with $\Delta = 1/N$ of the Ornstein-Uhlenbeck process given by*

$$du(t) = -a\mu u(t)\,dt + \nu\sqrt{\mu}\,dB_t, \quad u(0) \sim \mathcal{N}\left(0, \frac{\nu^2}{2a}\right).$$

Then, the Fisher information matrix I for the parameter (μ, ν^2) is given by

$$I_{11} = \frac{a^2\Delta(\mathrm{e}^{-4\mu a\Delta} + \mathrm{e}^{-2\mu a\Delta})}{(1 - \mathrm{e}^{-2\mu a\Delta})^2}, \qquad I_{12} = \frac{a\mathrm{e}^{-2\mu a\Delta}}{\nu^2(1 - \mathrm{e}^{-2\mu a\Delta})}, \qquad I_{22} = \frac{N+1}{2\nu^4}.$$

Proof. Thanks to the Markov property of u, the log-likelihood function for (μ, ν^2) is given by

$$\ell(\mu, \nu^2) = \log \pi_0(u(0)) + \sum_{i=0}^{N-1} \log p_\Delta(u(i\Delta), u((i+1)\Delta))$$

where

$$p_t(x, y) := \frac{1}{\sqrt{\pi\nu^2(1 - \mathrm{e}^{-2\mu a t})/a}} \exp\left(-\frac{(y - x\mathrm{e}^{-\mu a t})^2}{\nu^2(1 - \mathrm{e}^{-2\mu a t})/a}\right)$$

is the transition density of u and π_0 is the density of the initial distribution $\mathcal{N}\left(0, \frac{\nu^2}{2a}\right)$. For

$$\log p_\Delta(x, y) \eqsim -\frac{1}{2}\log \nu^2 - \frac{1}{2}\log(1 - \mathrm{e}^{-2\mu a\Delta}) - \frac{a(y - x\mathrm{e}^{-\mu a\Delta})^2}{\nu^2(1 - \mathrm{e}^{-2\mu a\Delta})},$$

the partial derivatives of first order are given by

$$\partial_\mu \left[\log p_\Delta(x, y)\right] = -\frac{a\Delta\mathrm{e}^{-2\mu a\Delta}}{1 - \mathrm{e}^{-2\mu a\Delta}} - \frac{2a^2\Delta\mathrm{e}^{-\mu a\Delta}x(y - x\mathrm{e}^{-\mu a\Delta})}{\nu^2(1 - \mathrm{e}^{-2\mu a\Delta})}$$

$$
\begin{aligned}
&+\frac{2a^2\Delta \mathrm{e}^{-2\mu a\Delta}(y-x\mathrm{e}^{-\mu a\Delta})^2}{\nu^2(1-\mathrm{e}^{-2\mu a\Delta})^2},\\
\partial_{\nu^2}\left[\log p_\Delta(x,y)\right] &= -\frac{1}{2\nu^2}+\frac{a(y-x\mathrm{e}^{-\mu a\Delta})^2}{\nu^4(1-\mathrm{e}^{-2\mu a\Delta})}.
\end{aligned}
$$

The second order derivatives are

$$
\begin{aligned}
\partial_\mu^2\left[\log p_\Delta(x,y)\right] &= \frac{2a^2\Delta^2\mathrm{e}^{-2\mu a\Delta}}{1-\mathrm{e}^{-2\mu a\Delta}}+\frac{2a^2\Delta^2\mathrm{e}^{-4\mu a\Delta}}{(1-\mathrm{e}^{-2\mu a\Delta})^2}\\
&-\partial_\mu\left[\frac{2a^2\Delta \mathrm{e}^{-\mu a\Delta}}{\nu^2(1-\mathrm{e}^{-2\mu a\Delta})}\right]x(y-x\mathrm{e}^{-\mu a\Delta})-\frac{2a^3\Delta^2\mathrm{e}^{-2\mu a\Delta}}{\nu^2(1-\mathrm{e}^{-2\mu a\Delta})}x^2\\
&-\left(\frac{4a^3\Delta^2\mathrm{e}^{-2\mu a\Delta}}{\nu^2(1-\mathrm{e}^{-2\mu a\Delta})^2}+\frac{8a^3\Delta^2\mathrm{e}^{-4\mu a\Delta}}{\nu^2(1-\mathrm{e}^{-2\mu a\Delta})^3}\right)(y-x\mathrm{e}^{-\mu a\Delta})^2\\
&+\frac{4a^3\Delta^2\mathrm{e}^{-3\mu a\Delta}}{\nu^2(1-\mathrm{e}^{-2\mu a\Delta})^2}x(y-x\mathrm{e}^{-\mu a\Delta}),\\
\partial_{\nu^2}^2\left[\log p_\Delta(x,y)\right] &= \frac{1}{2\nu^4}-\frac{2a(y-x\mathrm{e}^{-\mu a\Delta})^2}{\nu^6(1-\mathrm{e}^{-2\mu a\Delta})},\\
\partial_{\nu^2}\partial_\mu\left[\log p_\Delta(x,y)\right] &= \frac{2a^2\Delta \mathrm{e}^{-\mu a\Delta}x(y-x\mathrm{e}^{-\mu a\Delta})}{\nu^4(1-\mathrm{e}^{-2\mu a\Delta})}-\frac{2a^2\Delta \mathrm{e}^{-2\mu a\Delta}(y-x\mathrm{e}^{-\mu a\Delta})^2}{\nu^4(1-\mathrm{e}^{-2\mu a\Delta})^2}.
\end{aligned}
$$

Finally, for the initial distribution,

$$
\partial_{\nu^2}^2\log\pi_0(x)=\frac{1}{2\nu^4}-2a\frac{x^2}{\nu^6}.
$$

By stationarity of u, the Fisher information simplifies to

$$
I=-\mathbf{E}\left(D^2\ell(\mu,\nu^2)\right)=-\mathbf{E}\left(D^2\log\pi_0(u(0))\right)-N\mathbf{E}\left(D^2\log p_\Delta(u(0),u(\Delta))\right)
$$

where we write D^2g for the Hessian of a function g. Insertion of

$$
\begin{aligned}
\mathbf{E}\left(u(0)^2\right) &= \frac{\nu^2}{2a},\\
\mathbf{E}\left((u(\Delta)-u(0)\mathrm{e}^{-\mu a\Delta})^2\right) &= \frac{\nu^2}{2a}\left(1-\mathrm{e}^{-2\mu a\Delta}\right),\\
\mathbf{E}\left(u(0)(u(\Delta)-u(0)\mathrm{e}^{-\mu a\Delta})\right) &= 0
\end{aligned}
$$

finishes the calculation. □

To investigate the spectral density of the processes $\bar{U}_k$ from (2.5), the following auxiliary lemma is necessary.

Lemma 2.4.3. *The function $g:[0,\infty)\times[-\pi,\pi]\to\mathbb{R}$, defined by*

$$
g(x,\omega):=\frac{2x^2-\sinh(x^2)\cosh(x^2)+\cos(\omega)(\sinh(x^2)-2x^2\cosh(x^2))}{x^2(\cosh(x^2)-\cos(\omega))^2}(1-\cos(\omega)),
$$

satisfies

(i) $\int_0^\infty g(x,\omega)\,dx=0$, *for all* $\omega\in[-\pi,\pi]$,

(ii) $\sup_{|\omega|\le\pi}\|\frac{\partial}{\partial x}g(\cdot,\omega)\|_{L^1}<\infty$.

(iii) $|g(x,\omega)|\lesssim\frac{1+x^2}{x^4}\omega^2$ *uniformly in* $\omega\in[-\pi,\pi]$, $x>0$.

Proof. Assertion (i) follows from the fact that

$$G(x,\omega) := \frac{\sinh(x^2)(1-\cos(\omega))}{x(\cosh(x^2)-\cos(\omega))}, \qquad x>0,\ \omega\in[-\pi,\pi],$$

is a primitive of g in x and since $\lim_{x\to\infty} G(x,\omega) = \lim_{x\to 0} G(x,\omega) = 0$ for all $\omega \in [-\pi,\pi]$.

To show (ii), we decompose $g = \sum_{i=1}^5 g_i$ where

$$\begin{aligned}
g_1(x,\omega) &:= \frac{2}{(\cosh(x^2)-\cos(\omega))^2}(1-\cos(\omega))^2,\\
g_2(x,\omega) &:= -\frac{\sinh(x^2)}{x^2(\cosh(x^2)-\cos(\omega))}(1-\cos(\omega))^2,\\
g_3(x,\omega) &:= -\frac{\sinh(x^2)\cos(\omega)}{x^2(\cosh(x^2)-\cos(\omega))^2}(1-\cos(\omega))^2,\\
g_4(x,\omega) &:= \frac{2x^2+\sinh(x^2)}{x^2(\cosh(x^2)-\cos(\omega))^2}(1-\cos(\omega))^2\cos(\omega),\\
g_5(x,\omega) &:= -\frac{2x^2+\sinh(x^2)}{x^2(\cosh(x^2)-\cos(\omega))}(1-\cos(\omega))\cos(\omega).
\end{aligned}$$

By taking derivatives, one can show that for any $a\in[-1,1]$, the function

$$z \mapsto \frac{\sinh(z)}{z(\cosh(z)-a)}$$

is positive and decreasing in $z>0$. From this observation it can be easily deduced that, for fixed ω, the first derivatives of the functions g_i with respect to x are either negative or positive on all of $\mathbb{R}_+$. Consequently,

$$\left\|\frac{\partial}{\partial x} g(\cdot,\omega)\right\|_{L^1} \le \sum_{i=1}^5 \left|\int_0^\infty \frac{\partial}{\partial x} g_i(x,\omega)\,dx\right| = \sum_{i=1}^5 |\lim_{x\to\infty} g_i(x,\omega) - \lim_{x\to 0} g_i(x,\omega)| \le 11.$$

For (iii), we use the decomposition

$$g(x,\omega) = \frac{1-\cos(\omega)}{x^2}(h_1(x,\omega)+h_2(x,\omega))$$

where

$$\begin{aligned}
h_1(x,\omega) &:= \frac{(2x^2+\sinh(x^2))(1-\cosh(x^2))}{(\cosh(x^2)-\cos(\omega))^2},\\
h_2(x,\omega) &:= \frac{(\cos(\omega)-1)(\sinh(x^2)-2x^2\cosh(x^2))}{(\cosh(x^2)-\cos(\omega))^2}.
\end{aligned}$$

Then, the result follows from

$$|h_1(x,\omega)| \le \frac{2x^2+\sinh(x^2)}{\cosh(x^2)-1} \lesssim \frac{1}{x^2}\vee 1$$

and

$$|h_2(x,\omega)| \lesssim \frac{2x^2\cosh(x^2)-\sinh(x^2)}{\cosh(x^2)-1} \wedge \frac{2(2x^2\cosh(x^2)-\sinh(x^2))}{(\cosh(x^2)-1)^2} \lesssim \frac{1}{x^2}\vee 1. \qquad \square$$

The following lemma analyzes the N-th order Fourier approximation to the spectral density of the processes $\{\bar U_k(j),\ j\in\mathbb{N}_0\}$ for $k=1,\dots,M-1$.

Lemma 2.4.4. *Consider the parametrization from Proposition 2.1.6 and the function* $\Phi_k^{N,\Delta}$ *from* (2.26)*. If* $M\sqrt{\Delta} \to 0$*, then*

(i) $\Phi_k^{N,\Delta}(\omega) > 0$ *for all* $\omega \in [-\pi, \pi]$,

(ii)

$$\Phi_k^{N,\Delta}(\omega) \gtrsim \begin{cases} \frac{\sqrt{\Delta}}{M}\sqrt{|\omega|}, & |\omega| \geq M^2\Delta, \quad (2.46a) \\ \Delta, & k^2\Delta \leq |\omega| \leq M^2\Delta, \quad (2.46b) \\ \frac{\omega^2}{k^4\Delta} + \Delta e^{-\vartheta_2 k^2}, & |\omega| \leq k^2\Delta, \quad (2.46c) \end{cases}$$

(iii)

$$\frac{\partial}{\partial\vartheta_2}\Phi_k^{N,\Delta}(\omega) \lesssim \begin{cases} \Delta, & \omega \in [-\pi, \pi], \quad (2.47a) \\ \frac{\omega^2}{k^4\Delta} + \Delta k^2 e^{-\vartheta_2 k^2}, & |\omega| \leq k^2\Delta. \quad (2.47b) \end{cases}$$

Proof. Without loss of generality, we consider the parameters $\theta = \pi^2\vartheta_2$ and $\sigma_0^2 = \pi^2$. We denote the covariance function of $\bar{U}_k$ by $\rho_k : \mathbb{Z} \to \mathbb{R}$ and write Φ_k^N instead of $\Phi_k^{N,\Delta}$, i.e. , $\Phi_k^N(\omega) = \sum_{j=1-N}^{N-1} \rho_k(j)e^{-ij\omega}, \omega \in [-\pi, \pi]$.

(i) Let r_k be the covariance function of the process $(U_k(t_0), U_k(t_1), \ldots)$, i.e.,

$$r_k(j) = \sum_{\ell \in \mathcal{I}_k} \frac{e^{-\theta\ell^2|j|\Delta}}{2\sqrt{\theta}\ell^2}, \qquad j \in \mathbb{Z},$$

where $\mathcal{I}_k = \mathcal{I}_k^+ \cup \mathcal{I}_k^-$. Note that r_k and ρ_k are related by

$$\rho_k(j) = 2r_k(j) - r_k(j-1) - r_k(j+1), \qquad j \in \mathbb{Z},$$

which is a second order difference if $j \neq 0$. Since $x \mapsto e^{-x}$ has a positive second derivative, it follows that $\rho_k(j) < 0$. On the other hand, for $j = 0$, we have $\rho_k(0) = \mathrm{Var}(\bar{U}_k(t_0)) > 0$ and, therefore,

$$\begin{aligned} \Phi_k^N(\omega) &= \rho_k(0) + 2\sum_{j=1}^{N-1} \rho_k(j)\cos(j\omega) \\ &\geq \rho_k(0) + 2\sum_{j=1}^{N-1} \rho_k(j) = 2(r_k(N-1) - r_k(N)) > 0. \end{aligned}$$

To treat *(ii)* and *(iii)*, we calculate

$$\begin{aligned} \Phi_k^N(\omega) &= \sum_{j=1-N}^{N-1} \rho_k(j)e^{-ij\omega} \\ &= 2(1-\cos(\omega))\sum_{j=2-N}^{N-2} r_k(j)e^{-ij\omega} + 4r_k(N-1)\cos((N-1)\omega) \\ &\quad - 2r_k(N)\cos((N-1)\omega) - 2r_k(N-1)\cos((N-2)\omega). \end{aligned}$$

By using

$$\sum_{\mathrm{j}=1-\mathrm{J}}^{\mathrm{J}-1} e^{-\theta\ell^2|j|\Delta}e^{-ij\omega} = \sum_{\mathrm{j}=0}^{\mathrm{J}-1} e^{-\theta\ell^2 j\Delta}e^{-ij\omega} + \sum_{\mathrm{j}=0}^{\mathrm{J}-1} e^{-\theta\ell^2 j\Delta}e^{ij\omega} - 1$$

$$
\begin{aligned}
&= \frac{1-\mathrm{e}^{-J(\theta\ell^2\Delta - i\omega)}}{1-\mathrm{e}^{-(\theta\ell^2\Delta - i\omega)}} + \frac{1-\mathrm{e}^{-J(\theta\ell^2\Delta + i\omega)}}{1-\mathrm{e}^{-(\theta\ell^2\Delta + i\omega)}} - 1 \\
&= \frac{1-\mathrm{e}^{-2\theta\ell^2\Delta} + 2\mathrm{e}^{-(J+1)\theta\ell^2\Delta}\cos((J-1)\omega) - 2\mathrm{e}^{-J\theta\ell^2\Delta}\cos(J\omega)}{1+\mathrm{e}^{-2\theta\ell^2\Delta} - 2\mathrm{e}^{-\theta\ell^2\Delta}\cos(\omega)} \\
&= \frac{\sinh(\theta\ell^2\Delta) + \mathrm{e}^{-J\theta\ell^2\Delta}\cos((J-1)\omega) - \mathrm{e}^{-(J-1)\theta\ell^2\Delta}\cos(J\omega)}{\cosh(\theta\ell^2\Delta) - \cos(\omega)}
\end{aligned}
$$

for $J \geq 1$ and elementary manipulations, we can pass to the representation

$$
\Phi_k^N = \Phi + R_N
$$

where

$$
\Phi(\omega) := (1-\cos(\omega)) \sum_{\ell \in \mathcal{I}_k} \frac{1}{\sqrt{\theta}\ell^2} \frac{\sinh(\theta\ell^2\Delta)}{\cosh(\theta\ell^2\Delta) - \cos(\omega)}
$$

and

$$
R_N(\omega) := \sum_{\ell \in \mathcal{I}_k} (1-\cosh(\theta\ell^2\Delta)) \frac{\mathrm{e}^{-\theta\ell^2(N-1)\Delta}}{\sqrt{\theta}\ell^2} \frac{\mathrm{e}^{-\theta\ell^2\Delta}\cos((N-1)\omega) - \cos(N\omega)}{\cosh(\theta\ell^2\Delta) - \cos(\omega)}.
$$

Note that we have suppressed the dependence on k to ease the notation and that

$$
\Phi(\omega) = \sum_{j \in \mathbb{Z}} \rho_k(j)\mathrm{e}^{-ij\omega}, \qquad \omega \in [-\pi, \pi],
$$

is the spectral density of the process $(\bar{U}_k(j))_{j \geq 0}$.

(*ii*) To prove (2.46a), we note that for $\omega \geq M^2\Delta$ we have

$$
\begin{aligned}
&\left|\mathrm{e}^{-\theta\ell^2\Delta}\cos((N-1)\omega) - \cos(N\omega)\right| \\
&= \left|(\mathrm{e}^{-\theta\ell^2\Delta} - 1)\cos((N-1)\omega) + \cos((N-1)\omega) - \cos(N\omega)\right| \\
&\lesssim \ell^2\Delta + \omega \\
&\lesssim \ell^2\omega.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
R_N(\omega) &\lesssim \sum_{\ell \in \mathcal{I}_k} \ell^2\Delta \sinh(\theta\ell^2\Delta) \frac{\mathrm{e}^{-\theta\ell^2(N-1)\Delta}}{\sqrt{\theta}\ell^2} \frac{\ell^2\omega}{\cosh(\theta\ell^2\Delta) - \cos(\omega)} \\
&\lesssim \frac{\Delta}{\omega} \sum_{\ell \in \mathcal{I}_k} \frac{\sinh(\theta\ell^2\Delta)}{\ell^2(\cosh(\theta\ell^2\Delta) - \cos(\omega))}(1-\cos(\omega)) \\
&\lesssim \frac{1}{M^2}\Phi(\omega)
\end{aligned}
$$

and, hence, R_N is negligible compared to Φ. In order to compute an asymptotic expression for Φ, set

$$
h(x,\omega) := \frac{\sinh(\theta x^2)(1-\cos(\omega))}{x^2(\cosh(\theta x^2) - \cos(\omega))}, \quad x > 0,\, \omega \in [-\pi, \pi].
$$

As already remarked in the proof of Lemma 2.4.3, we have $\frac{\partial h}{\partial x} \leq 0$ and, therefore,

$$
\left\| \frac{\partial}{\partial x} h(\cdot, \omega) \right\|_{L^1} = h(0,\omega) - \lim_{x \to \infty} h(x,\omega) = \theta
$$

is uniformly bounded in ω. Thus, using the mean value theorem and a Riemann sum approximation with mesh size $M\sqrt{\Delta}$ for $\frac{\partial}{\partial x}h(\cdot,\omega)$, we obtain

$$\Phi(\omega) \eqsim \Delta \sum_{\ell\in\mathcal{I}_k} h(\ell\sqrt{\Delta},\omega) = \Delta \sum_{\ell=1}^{\infty} h(2\ell M\sqrt{\Delta},\omega) + \mathcal{O}(\Delta).$$

Further, since

$$\left| \varepsilon \sum_{\ell\geq 1} f(\ell\varepsilon) - \int_0^\infty f(x)\,dx \right| \leq \varepsilon \|f'\|_{L^1} \tag{2.48}$$

for any function $f \in C^1[0,\infty)$, we get

$$\Phi(\omega) \eqsim \frac{\sqrt{\Delta}}{M} \int_0^\infty h(x,\omega)\,dx + \mathcal{O}(\Delta).$$

Finally, due to

$$a + b \eqsim \max(a,b), \qquad a,b > 0, \tag{2.49}$$

we have

$$(\cosh(\theta\omega x^2) - \cos(\omega)) \eqsim \max\left(\cosh(\theta\omega x^2) - 1, 1 - \cos(\omega)\right)$$

and, consequently, for $x \leq \theta^{-1/2}$:

$$h(\sqrt{\omega}x,\omega) = \frac{\sinh(\theta\omega x^2)(1-\cos(\omega))}{\omega x^2(\cosh(\theta\omega x^2) - \cos(\omega))} \gtrsim \frac{\sinh(\theta\omega x^2)}{\omega x^2} \gtrsim 1.$$

Therefore,

$$\int_0^\infty h(x,\omega)\,dx = \sqrt{\omega}\int_0^\infty h(\sqrt{\omega}x,\omega)\,dx \gtrsim \sqrt{\omega},$$

finishing the proof of (2.46a).
To prove (2.46b) and (2.46c), let us write

$$\Phi = \sum_{\ell\in\mathcal{I}_k} \varphi_\ell, \qquad R_N = \sum_{\ell\in\mathcal{I}_k} \varrho_\ell^N.$$

Since the argument in the proof of (i) was on a summand-wise level, also each of the functions $\varphi_\ell + \varrho_\ell^N$ is positive, $\ell \in \mathbb{N}$. Therefore, we can bound Φ_k^N from below using the first summand, namely

$$\Phi_k^N \geq \varphi_k + \varrho_k^N = \varrho_k^N(0) + \varphi_k + \left(\varrho_k^N - \varrho_k^N(0)\right).$$

We show that there exists an environment U around zero and some $\delta \in (0,1)$ such that

$$|\varrho_k^N(\omega) - \varrho_k^N(0)| \leq (1-\delta)\varphi_k(\omega), \qquad \omega \in U : \tag{2.50}$$

A simple calculation yields

$$\begin{aligned}
\varrho_k^N(\omega) - \varrho_k^N(0) &= \mathrm{e}^{-(N-1)\theta k^2\Delta} \frac{(\cos((N-1)\omega)) - \cos(N\omega))(1-\cosh(\theta k^2\Delta))}{\sqrt{\theta}k^2(\cosh(\theta k^2\Delta) - \cos(\omega))} \\
&\quad + \mathrm{e}^{-(N-1)\theta k^2\Delta} \frac{(1-\mathrm{e}^{-\theta k^2\Delta})(1-\cos((N-1)\omega)))(1-\cosh(\theta k^2\Delta))}{\sqrt{\theta}k^2(\cosh(\theta k^2\Delta) - \cos(\omega))} \\
&\quad + \mathrm{e}^{-(N-1)\theta k^2\Delta} \frac{\left(\mathrm{e}^{-\theta k^2\Delta} - 1\right)(1-\cos(\omega))}{\sqrt{\theta}k^2(\cosh(\theta k^2\Delta) - \cos(\omega))}.
\end{aligned}$$

Since $\cos(x) - \cos(y) = -2\sin\frac{x+y}{2}\sin\frac{x-y}{2}$, $x, y \in \mathbb{R}$, we have

$$\Big|\cos((N-1)\omega) - \cos(N\omega)\Big| = \Big|2\sin\left(\frac{(2N-1)\omega}{2}\right)\sin\left(\frac{\omega}{2}\right)\Big| \leq N\omega^2. \tag{2.51}$$

Therefore, for any $\alpha > 0$ there exists an environment U of 0 such that

$$|\cos((N-1)\omega) - \cos(N\omega)| \leq N\omega^2 \leq N(1-\cos(\omega))(2+\alpha),$$
$$1 - \cos((N-1)\omega) \leq \frac{N^2\omega^2}{2} \leq \frac{N^2}{2}(1-\cos(\omega))(2+\alpha)$$

holds for all $\omega \in U$. Further, for all $x \geq 0$ we have

$$\cosh(x) - 1 \leq \frac{\sinh(x)x}{2}, \qquad 1 - \mathrm{e}^{-x} \leq \sinh(x),$$

and, consequently,

$$\begin{aligned}\frac{|\varrho_k^N(\omega) - \varrho_k^N(0)|}{\varphi_k(\omega)} &\leq \mathrm{e}^{-(N-1)\theta k^2\Delta}(1 + \frac{2+\alpha}{2}\theta k^2 + \frac{2+\alpha}{4}\theta^2 k^4) \\ &\leq \frac{2+\alpha}{2}\mathrm{e}^{\Delta\theta k^2}\underbrace{\mathrm{e}^{-\theta k^2}(1 + \theta k^2 + \frac{\theta^2 k^4}{2})}_{<1}.\end{aligned}$$

Clearly, for Δ sufficiently small, one can choose α in such a way that this bound is strictly less than 1 for all $k \leq M-1$, yielding (2.50). Consequently, it is sufficient to prove (2.46b) and (2.46c) with Φ_k^N replaced by $\varphi_k + \varrho_k^N(0)$. Now,

$$\varphi_k(0) + \varrho_k^N(0) = \varrho_k^N(0) = \mathrm{e}^{-\theta k^2(N-1)\Delta}\frac{1 - \mathrm{e}^{-\theta k^2\Delta}}{k^2} \eqsim \Delta\mathrm{e}^{-\theta k^2}$$

and, again by using (2.49), we get

$$\varphi_k(\omega) \gtrsim \frac{\sinh(\theta k^2\Delta)}{k^2} \gtrsim \Delta, \qquad \omega \geq k^2\Delta,$$

and

$$\varphi_k(\omega) \gtrsim (1-\cos(\omega))\frac{1}{\sqrt{\theta}k^2}\frac{\sinh(\theta k^2\Delta)}{\cosh(\theta k^2\Delta) - 1} \gtrsim \frac{\omega^2}{k^4\Delta}, \qquad \omega \leq k^2\Delta.$$

(*iii*) We show (2.47a): We have

$$\frac{\partial}{\partial\theta}\Phi(\omega) = \frac{\Delta}{2\sqrt{\theta}}\sum_{\ell\in\mathcal{I}_k} g(\ell\sqrt{\theta\Delta}, \omega)$$

with g defined in Lemma 2.4.3. Using the properties of g derived in Lemma 2.4.3 and the Riemann sum approximation (2.48) with mesh size $M\sqrt{\Delta}$, we obtain

$$\begin{aligned}\frac{\partial}{\partial\theta}\Phi(\omega) &\eqsim \Delta\sum_{\ell\geq 1} g(\ell M\sqrt{\Delta}, \omega) + \mathcal{O}(\Delta) \\ &= \frac{\sqrt{\Delta}}{M}\int_0^\infty g(x,\omega)\,dx + \mathcal{O}(\Delta) = \mathcal{O}(\Delta).\end{aligned}$$

To show that $\frac{\partial}{\partial\theta} R_N$ is of the claimed order, we write

$$\varrho_\ell^N = \alpha_\ell \beta_\ell \quad \text{where} \quad \alpha_\ell(\omega) := \frac{1-\cosh(\theta\ell^2\Delta)}{\sqrt{\theta}\ell^2\left(\cosh(\theta\ell^2\Delta)-\cos(\omega)\right)},$$
$$\beta_\ell(\omega) := \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\left(\mathrm{e}^{-\theta\ell^2\Delta}\cos((N-1)\omega)-\cos(N\omega)\right).$$

The corresponding derivatives are given by

$$\frac{\partial}{\partial\theta}\alpha_\ell(\omega) = \underbrace{\frac{\cosh(\theta\ell^2\Delta)-1}{2\theta^{3/2}\ell^2\left(\cosh(\theta\ell^2\Delta)-\cos(\omega)\right)}}_{=:a_\ell^1(\omega)} \underbrace{-\frac{\Delta\sinh(\theta\ell^2\Delta)\left(1-\cos(\omega)\right)}{\sqrt{\theta}\left(\cosh(\theta\ell^2\Delta)-\cos(\omega)\right)^2}}_{=:a_\ell^2(\omega)}$$

and

$$\frac{\partial}{\partial\theta}\beta_\ell(\omega) = \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\left(-\ell^2 N\Delta\mathrm{e}^{-\theta\ell^2\Delta}\cos((N-1)\omega)+\ell^2(N-1)\Delta\cos(N\omega)\right)$$
$$=: b_\ell(\omega).$$

By the product rule, we have

$$\frac{\partial}{\partial\theta}R_N = \sum_{\ell\in\mathcal{I}_k} a_\ell^1\beta_\ell + a_\ell^2\beta_\ell + \alpha_\ell b_\ell. \tag{2.52}$$

To bound each of the terms, we estimate

$$\frac{\cosh(x)-1}{\cosh(x)-\cos(\omega)} \le 1, \quad x>0,$$
$$\frac{\cosh(x)-1}{\cosh(x)-\cos(\omega)} \le \frac{\cosh(x)-1}{1-\cos(\omega)} \lesssim \frac{x^2}{\omega^2}, \qquad x \le |\omega| \le \pi,$$

resulting in

$$\frac{\cosh(x)-1}{\cosh(x)-\cos(\omega)} \lesssim \frac{x^2}{x^2 \vee \omega^2} \tag{2.53}$$

and

$$\frac{x\sinh(x)(1-\cos(\omega))}{(\cosh(x)-\cos(\omega))^2} \lesssim \frac{x\sinh(x)}{(\cosh(x)-1)^2} \wedge \frac{x\sinh(x)}{\cosh(x)-1} \lesssim 1, \qquad x>0,$$
$$\frac{x\sinh(x)(1-\cos(\omega))}{(\cosh(x)-\cos(\omega))^2} \lesssim \frac{x\sinh(x)}{1-\cos(\omega)} \lesssim \frac{x^2}{\omega^2}, \qquad x \le |\omega| \le \pi,$$

resulting in

$$\frac{x\sinh(x)(1-\cos(\omega))}{(\cosh(x)-\cos(\omega))^2} \lesssim \frac{x^2}{x^2\vee\omega^2}. \tag{2.54}$$

In combination with

$$\beta_\ell(\omega) \lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\left((\ell^2\Delta)\vee\omega\right),$$
$$b_\ell(\omega) \lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^2\left((\ell^2\Delta)\vee\omega\right),$$

the bounds (2.53) and (2.54) show that any of the three products in (2.52) can be bounded by

$$\sum_{\ell\in\mathcal{I}_k}\mathrm{e}^{-\theta\ell^2(N-1)\Delta}\frac{\ell^4\Delta^2}{(\ell^4\Delta^2)\vee\omega^2}\left((\ell^2\Delta)\vee\omega\right) \le \Delta\sum_{\ell\in\mathcal{I}_k}\mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^2 \lesssim \Delta.$$

Consequently, we have $\frac{\partial}{\partial\theta}R_N = \mathcal{O}(\Delta)$ which finishes the proof of (2.47a).
To prove (2.47b), we use property *(iii)* of Lemma 2.4.3 to deduce

$$\frac{\partial}{\partial\theta}\Phi(\omega) \lesssim \omega^2\Delta \sum_{\ell\in\mathcal{I}_k} \frac{1+\theta\ell^2\Delta}{\theta^2\ell^4\Delta^2}$$
$$\lesssim \frac{\omega^2}{\Delta}\left(\frac{1+\theta k^2\Delta}{\theta^2 k^4} + \sum_{\ell\geq 1}\frac{1+\theta(2\ell M)^2\Delta}{\theta^2(2\ell M)^4}\right) \lesssim \frac{\omega^2}{k^4\Delta}$$

where the last step follows from $k^2\Delta \leq M^2\Delta \to 0$. Further, using decomposition (2.52),

$$\begin{aligned}\frac{\partial}{\partial\theta}(R_N(\omega) - R_N(0)) = &\sum_{\ell\in\mathcal{I}_k} a_\ell^1(\omega)(\beta_\ell(\omega)-\beta_\ell(0)) + \sum_{\ell\in\mathcal{I}_k}(a_\ell^1(\omega)-a_\ell^1(0))\beta_\ell(0)\\ &+ \sum_{\ell\in\mathcal{I}_k} a_\ell^2(\omega)\beta_\ell(\omega)\\ &+ \sum_{\ell\in\mathcal{I}_k}\alpha_\ell(\omega)(b_\ell(\omega)-b_\ell(0)) + \sum_{\ell\in\mathcal{I}_k}(\alpha_\ell(\omega)-\alpha_\ell(0))b_\ell(0).\end{aligned} \tag{2.55}$$

Now, by (2.51), we obtain

$$\begin{aligned}&\beta_\ell(\omega) - \beta_\ell(0)\\ &= \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\left((\mathrm{e}^{-\theta\ell^2\Delta}-1)(\cos((N-1)\omega)-1) + \cos((N-1)\omega) - \cos(N\omega)\right)\\ &\lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^2 N\omega^2.\end{aligned}$$

In a similar way, we can bound

$$\begin{aligned}\beta_\ell(0) &\lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^2\Delta,\\ \beta_\ell(\omega) &\lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\left((\ell^2\Delta)\vee\omega\right) \lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^2\Delta,\\ b_\ell(\omega) - b_\ell(0) &\lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^4 N\omega^2,\\ b_\ell(0) &\lesssim \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^4\Delta\end{aligned}$$

where the second inequality uses $\omega \leq k^2\Delta \leq \ell^2\Delta$ for $\ell \in \mathcal{I}_k$. Also,

$$a_\ell^1(\omega) - a_\ell^1(0) \lesssim \frac{1-\cos(\omega)}{\cosh(\theta\ell^2\Delta)-\cos(\omega)} \lesssim \frac{1-\cos(\omega)}{(\cosh(\theta\ell^2\Delta)-1)} \lesssim \frac{\omega^2}{k^4\Delta^2}$$

and, similarly,

$$\alpha_\ell(\omega) - \alpha_\ell(0) \lesssim \frac{\omega^2}{k^4\Delta^2}, \qquad a_\ell^2(\omega) \lesssim \frac{\omega^2}{k^4\Delta^2}$$

as well as

$$a_\ell^1(\omega) \lesssim 1, \qquad \alpha_\ell(\omega) \lesssim 1.$$

Using the bounds just developed in combination with $\mathrm{e}^{-\theta\ell^2(N-1)\Delta} \lesssim \frac{1}{k^4\ell^m}$, $m \in \mathbb{N}$, shows that any of the five terms in (2.55) is of the order $\mathcal{O}(\frac{\omega^2}{k^4\Delta})$ and, hence,

$$\frac{\partial}{\partial\theta}(R_N(\omega) - R_N(0)) \lesssim \frac{\omega^2}{k^4\Delta}.$$

Now, the proof of (2.47b) is finalized by

$$\begin{aligned}\frac{\partial}{\partial\theta}R_N(0) &= \sum_{\ell\in\mathcal{I}_k} \mathrm{e}^{-\theta\ell^2(N-1)\Delta}\frac{2\theta\ell^2(N-1)\Delta(\mathrm{e}^{-\theta\ell^2\Delta}-1)+2\theta\ell^2\Delta\mathrm{e}^{-\theta\ell^2\Delta}+\mathrm{e}^{-\theta\ell^2\Delta}-1}{2\theta^{3/2}\ell^2}\\ &\lesssim \Delta\sum_{\ell\in\mathcal{I}_k}\mathrm{e}^{-\theta\ell^2(N-1)\Delta}\ell^2 \lesssim \Delta k^2\mathrm{e}^{-\theta k^2}.\end{aligned}$$

□

Covariances of double increments

The following three lemmas are used to calculate the asymptotic variance of $\mathbb{V}$. As can be seen from the main proof of Theorem 2.2.7, it is sufficient to consider the case where X_0 follows the stationary distribution. The first lemma identifies the relevant terms in the covariance structure and shows independence of Γ. Recall the definition of $\tilde{D}_{ik}$ from (2.32).

Lemma 2.4.5. *Assume (ST) and let $b\in(0,1/2)$. For $J\geq 1$ define*

$$F_{J,\Delta}(z) := \sum_{\ell\geq1}\frac{2\mathrm{e}^{-\pi^2\vartheta_2 J\ell^2\Delta}-\mathrm{e}^{-\pi^2\vartheta_2(J+1)\ell^2\Delta}-\mathrm{e}^{-\pi^2\vartheta_2(J-1)\ell^2\Delta}}{2\pi^2\vartheta_2\ell^2}\cos(\pi\ell z)$$

and $F_{0,\Delta}=F_{\vartheta_2}(\cdot\,,\Delta)$. Then,

$$\begin{aligned}\mathrm{Cov}(\tilde{D}_{ik},\tilde{D}_{jl}) = &-\sigma^2\mathrm{e}^{-\kappa\delta/2}\cdot\begin{cases}2D_\delta F_{J,\Delta}(0), & l=k,\\ D_\delta^2 F_{J,\Delta}(y_l-y_{k+1}), & l>k\end{cases}\\ &+\mathcal{O}\left(\frac{\sqrt{\Delta}\delta^2}{(J+1)^{3/2}}\right)\end{aligned}$$

where $J=|i-j|$.

Proof. First of all, it immediately follows from the covariance structure of the coefficient processes, $\mathrm{Cov}(u_\ell(s),u_\ell(t))=\frac{\sigma^2}{2\lambda_\ell}\mathrm{e}^{-\lambda_\ell|t-s|}$, $s,t\geq0$, that

$$\begin{aligned}\mathrm{Cov}(D_{ik},\,D_{jl}) = \sigma^2\sum_{\ell\geq1}&(e_\ell(y_{k+1})-e_\ell(y_k))(e_\ell(y_{l+1}+\delta)-e_\ell(y_l))\\ &\cdot\begin{cases}\dfrac{1-\mathrm{e}^{-\lambda_\ell\Delta}}{\lambda_\ell}, & J=0,\\ \dfrac{2\mathrm{e}^{-\lambda_\ell J\Delta}-\mathrm{e}^{-\lambda_\ell(J+1)\Delta}-\mathrm{e}^{-\lambda_\ell(J-1)\Delta}}{2\lambda_\ell}, & J\geq1.\end{cases}\end{aligned}$$

Step 1. We show negligibilty of Γ: We already know from the first step of the proof of Proposition 2.2.5 that

$$\mathrm{Cov}(D_{ik},D_{il}) = \sigma^2\sum_{\ell\geq1}\frac{1-\mathrm{e}^{-\pi^2\vartheta_2\ell^2\Delta}}{\pi^2\vartheta_2\ell^2}(e_\ell(y_{k+1})-e_\ell(y_k))(e_\ell(y_{l+1})-e_\ell(y_l))+\mathcal{O}\left(\sqrt{\Delta}\delta^2\right).$$

For $J\geq1$, we will show now that

$$\begin{aligned}\mathrm{Cov}(D_{ik},D_{jl}) = \sigma^2\sum_{\ell\geq1}&\frac{2\mathrm{e}^{-\pi^2\vartheta_2\ell^2 J\Delta}-\mathrm{e}^{-\pi^2\vartheta_2\ell^2(J+1)\Delta}-\mathrm{e}^{-\pi^2\vartheta_2\ell^2(J-1)\Delta}}{2\pi^2\vartheta_2\ell^2}\\ &\cdot(e_\ell(y_{k+1})-e_\ell(y_k))(e_\ell(y_{l+1})-e_\ell(y_l))+\mathcal{O}\left(\frac{\sqrt{\Delta}\delta^2}{(J+1)^{3/2}}\right):\end{aligned}$$

If $J = 1$, this directly follows from the case $J = 0$ in view of

$$\frac{2e^{-\lambda_\ell \Delta} - e^{-2\lambda_\ell \Delta} - 1}{2\lambda_\ell} = \frac{1 - e^{-2\lambda_\ell \Delta}}{2\lambda_\ell} - \frac{1 - e^{-\lambda_\ell \Delta}}{\lambda_\ell}. \tag{2.56}$$

For $J \geq 2$, define

$$g_J(x) := \frac{2e^{-Jx} - e^{-(J+1)x} - e^{-(J-1)x}}{2x}.$$

A first order Taylor approximation of g_J gives

$$\begin{aligned}\mathrm{Cov}(D_{ik}, D_{jl}) &= \Delta \sum_{\ell \geq 1} g_J(\lambda_\ell \Delta)(e_\ell(y_{k+1}) - e_\ell(y_k))(e_\ell(y_{l+1}) - e_\ell(y_l)) \\ &= \Delta \sum_{\ell \geq 1} g_J(\pi^2 \vartheta_2 \ell^2 \Delta)(e_\ell(y_{k+1}) - e_\ell(y_k))(e_\ell(y_{l+1}) - e_\ell(y_l)) + R\end{aligned}$$

where

$$R \lesssim \Delta^2 \sum_{\ell \geq 1} g_J'(\vartheta_2(\pi^2 \ell^2 + \xi_\ell)\Delta)\ell^2 \delta^2$$

for some $|\xi_\ell| \leq |\Gamma|$. Writing

$$g_J(x) = e^{-(J-1)x} h(x), \quad h(x) := \frac{2e^{-x} - e^{-2x} - 1}{2x},$$

and noting $h(x) \lesssim x$ as well as $h'(x) \lesssim 1$ shows that

$$g_J'(x) = -(J-1)e^{-(J-1)x} h(x) + e^{-(J-1)x} h'(x) \lesssim e^{-(J-1)x/2}.$$

Therefore, for some $\omega > 0$ and by regarding R as a Riemann sum with lag $\sqrt{(J-1)\Delta}$,

$$\begin{aligned}R \lesssim &\Delta^2 \sum_{\ell \geq 1} e^{-\omega(J-1)\ell^2 \Delta} \ell^2 \delta^2 \\ = &\frac{\Delta^2 \delta^2}{((J-1)\Delta)^{3/2}} \cdot \sqrt{(J-1)\Delta} \sum_{\ell \geq 1} e^{-\omega(J-1)\ell^2 \Delta}(J-1)\ell^2 \Delta \\ \lesssim &\frac{\sqrt{\Delta}\delta^2}{(J-1)^{3/2}} \lesssim \frac{\sqrt{\Delta}\delta^2}{(J+1)^{3/2}}.\end{aligned}$$

Step 2. By Step 1, we may assume $\lambda_\ell = \pi^2 \vartheta_2 \ell^2$. By (2.31), we have

$$\begin{aligned}\mathrm{Cov}(\tilde{D}_{ik}, \tilde{D}_{jk}) = &- 2\sigma^2 g(\delta) D_\delta F_{J,\Delta}(0) + \sigma^2 F_{J,\Delta}(0) D_\delta^2 g(0) \\ &- \sigma^2 D_\delta^2 \left(g(\cdot) F_{J,\Delta}(2y_k + \cdot)\right)(0)\end{aligned}$$

and, by (2.30), for $l > k$,

$$\begin{aligned}\mathrm{Cov}(\tilde{D}_{ik}, \tilde{D}_{jl}) = &- \sigma^2 g(\delta) D_\delta^2 F_{J,\Delta}(y_l - y_{k+1}) + \sigma^2 F_{J,\Delta}(y_l - y_k) D_\delta^2 g(0) \\ &- \sigma^2 D_\delta^2 \left(g(\cdot) F_{J,\Delta}(y_l + y_k + \cdot)\right)(0).\end{aligned}$$

Hence, as in previous results, it is sufficient to establish

$$F_{J,\Delta}(0),\, F_{J,\Delta}(z),\, F_{J,\Delta}'(z)\, F_{J,\Delta}''(z) \lesssim \frac{\sqrt{\Delta}}{J^{3/2}}, \quad z \in [2b, 2(1-b)].$$

For $J = 0$, this was already proven in Proposition 2.2.5. The case $J = 1$ follows from the case $J = 0$ since (2.56) shows

$$F_{1,\Delta}(z) = \frac{1}{2} F_{2\Delta}(z) - F_\Delta(z). \tag{2.57}$$

For $J \geq 2$, we have

$$2\mathrm{e}^{-\lambda_\ell J\Delta} - \mathrm{e}^{-\lambda_\ell (J+1)\Delta} - \mathrm{e}^{-\lambda_\ell (J-1)\Delta} = \mathrm{e}^{-\lambda_\ell (J-1)\Delta}(2\mathrm{e}^{-\lambda_\ell \Delta} - \mathrm{e}^{-\lambda_\ell 2\Delta} - 1)$$
$$\lesssim \mathrm{e}^{-\lambda_\ell (J-1)\Delta}(\lambda_\ell \Delta)^2, \tag{2.58}$$

and, therefore,

$$\begin{aligned} F_{J,\Delta}(z) \lesssim F_{J,\Delta}(0) &\lesssim \sum_{\ell\geq 1} \lambda_\ell \Delta^2 \mathrm{e}^{-\lambda_\ell (J-1)\Delta} \\ &= \frac{\sqrt{\Delta}}{(J-1)^{3/2}} \sqrt{(J-1)\Delta} \sum_{\ell \geq 1} ((J-1)\lambda_\ell \Delta)\mathrm{e}^{-\lambda_\ell (J-1)\Delta} \\ &= \mathcal{O}\left(\frac{\sqrt{\Delta}}{(J-1)^{3/2}}\right). \end{aligned}$$

The bound on the first derivative is provided by Lemma 2.4.8, namely

$$\begin{aligned} F'_{J,\Delta}(z) &\lesssim \sum_{\ell\geq 1} \frac{2\mathrm{e}^{-\lambda_\ell J\Delta} - \mathrm{e}^{-\lambda_\ell (J+1)\Delta} - \mathrm{e}^{-\lambda_\ell (J-1)\Delta}}{2\lambda_\ell} \ell \sin(\pi \ell z) \\ &\lesssim \sup_\ell \left| \frac{2\mathrm{e}^{-\lambda_\ell J\Delta} - \mathrm{e}^{-\lambda_\ell (J+1)\Delta} - \mathrm{e}^{-\lambda_\ell (J-1)\Delta}}{2\lambda_\ell} \ell \right| \frac{1}{z \wedge (2-z)} \\ &\lesssim \sup_\ell \left| \lambda_\ell \Delta^2 \mathrm{e}^{-\lambda_\ell J \Delta} \ell \right| \lesssim \frac{\sqrt{\Delta}}{J^{3/2}}. \end{aligned}$$

Finally, to bound $F''_{J,\Delta}$, we define $h_J(z) = 2\mathrm{e}^{-Jz^2} - \mathrm{e}^{-(J+1)z^2} - \mathrm{e}^{-(J-1)z^2}$. Clearly, $h_J(0) = 0$ and

$$\begin{aligned} \frac{d}{dz} h_J(z) &= \frac{d}{dz} \mathrm{e}^{-(J-1)z^2}(2\mathrm{e}^{-z^2} - \mathrm{e}^{-2z^2} - 1) \\ &= -2(J-1)z\mathrm{e}^{-(J-1)z^2} \underbrace{(2\mathrm{e}^{-z^2} - \mathrm{e}^{-2z^2} - 1)}_{\lesssim z^4} + \mathrm{e}^{-(J-1)z^2} \underbrace{(-4z\mathrm{e}^{-z^2} + 4z\mathrm{e}^{-2z^2})}_{\lesssim z^3} \\ &\lesssim (J-1)^{-3/2}, \end{aligned}$$

i.e., $\|h'_J\|_\infty \lesssim J^{-3/2}$. In view of Lemma 2.4.9, this shows

$$\begin{aligned} F''_{J,\Delta}(z) &\lesssim \sum_{\ell\geq 1} \frac{2\mathrm{e}^{-\lambda_\ell J\Delta} - \mathrm{e}^{-\lambda_\ell (J+1)\Delta} - \mathrm{e}^{-\lambda_\ell (J-1)\Delta}}{2\lambda_\ell} \ell^2 \cos(\pi \ell z) \\ &\lesssim \sum_{\ell\geq 1} h_J(\sqrt{\lambda_\ell \Delta}) \cos(\pi \ell z) \\ &= \mathcal{O}\left(\frac{1}{(z\wedge(2-z))^2} \frac{\sqrt{\Delta}}{J^{3/2}}\right). \end{aligned}$$

□

The following Lemma is useful for calculating the asymptotic variance in case $\delta/\sqrt{\Delta} \to 0$.

Lemma 2.4.6. *For $J \in \mathbb{N}_0$ and $z \in (0,2)$, it holds that*

(i) $F_{J,\Delta}(0) - F_{J,\Delta}(\delta) = \delta \frac{1}{2\vartheta_2} \mathbf{1}_{\{J=0\}} - \delta \frac{1}{4\vartheta_2} \mathbf{1}_{\{J=1\}} + \mathcal{O}\left(\frac{\delta^2}{(J+1)^{5/2}\sqrt{\Delta}}\right)$,

(ii) $2F_{J,\Delta}(z) - F_{J,\Delta}(z+\delta) - F_{J,\Delta}(z-\delta) = \mathcal{O}\left(\frac{\delta^2}{(J+1)^2}\left(\frac{1}{\sqrt{\Delta}} \wedge \frac{1}{z\wedge(2-z)}\right)\right)$.

Proof. (*i*) The validity of the case $J = 0$ follows from the proof of Proposition 2.2.5 (ii), the case $J = 1$ follows from (2.57). For $J \geq 2$, we have

$$F_{J,\Delta}(0) - F_{J,\Delta}(\delta) = -\delta F'_{J,\Delta}(0) - \frac{\delta^2}{2} F''_{J,\Delta}(\xi)$$

for some $\xi \in [0, \delta]$ by Taylor's theorem. Now, the claim is proved by inserting $F'_{J,\Delta}(0) = 0$ and noting that (2.58) implies

$$\begin{aligned} \left\| F''_{J,\Delta} \right\|_\infty &\lesssim \sum_{\ell \geq 1} \left(2\mathrm{e}^{-\lambda_\ell J \Delta} - \mathrm{e}^{-\lambda_\ell (J+1)\Delta} - \mathrm{e}^{-\lambda_\ell (J-1)\Delta} \right) \\ &\lesssim \sum_{\ell \geq 1} \lambda_\ell^2 \Delta^2 \mathrm{e}^{-\lambda_\ell (J-1)\Delta} \lesssim \frac{1}{J^{5/2}\sqrt{\Delta}}. \end{aligned}$$

(*ii*) As previously, it suffices to establish

$$F''_{J,\Delta}(z) \lesssim \frac{1}{(J+1)^2} \left(\frac{1}{\sqrt{\Delta}} \wedge \frac{1}{z \wedge (2-z)} \right).$$

For the case $J = 0$, we employ the representation $F_\Delta = H_\Delta + G_\Delta$ from the proof of Proposition 2.2.5. Then, the validity of the bound on H''_Δ follows from

$$H''_\Delta(z) \lesssim \frac{\exp\left(-\frac{\pi}{\sqrt{\vartheta_2 \Delta}} (y \wedge (2-y)) \right)}{\sqrt{\Delta}} \lesssim \frac{1}{\sqrt{\Delta}} \wedge \frac{1}{y \wedge (2-y)}.$$

The bound on $G''_\Delta(z)$ follow from $\|G''_\Delta\|_\infty \lesssim 1/\sqrt{\Delta}$ and

$$G''_\Delta(z) \lesssim \sup_\ell \left| \frac{1 - \mathrm{e}^{-\lambda_\ell \Delta}(1 + \lambda_\ell \Delta)}{1 + \lambda_\ell \Delta} \right| \frac{1}{z \wedge (2-z)} \lesssim \frac{1}{z \wedge (2-z)},$$

by Lemma 2.4.8. The case $J = 1$ follows from the case $J = 0$, see (2.57). For $J \geq 2$, we proceed in the same way: In the proof of (*i*) it was shown that $\left\| F''_{\Delta,J} \right\|_\infty \lesssim \frac{1}{J^{5/2}\sqrt{\Delta}} \lesssim \frac{1}{J^2\sqrt{\Delta}}$. Finally, by Lemma 2.4.8,

$$\begin{aligned} F''_{J,\Delta}(z) &\lesssim \sup_\ell \left| 2\mathrm{e}^{-\lambda_\ell J \Delta} - \mathrm{e}^{-\lambda_\ell (J+1)\Delta} - \mathrm{e}^{-\lambda_\ell (J-1)\Delta} \right| \frac{1}{z \wedge (2-z)} \\ &\lesssim \sup_\ell \left| (\lambda_\ell \Delta)^2 \mathrm{e}^{-\lambda_\ell (J-1)\Delta} \right| \frac{1}{z \wedge (2-z)} \\ &\lesssim \frac{1}{(J+1)^2} \frac{1}{z \wedge (2-z)}. \end{aligned}$$

□

The following Lemma is useful for calculating the asymptotic variance in case $\delta/\sqrt{\Delta} \to \infty$.

Lemma 2.4.7. *For $J \in \mathbb{N}_0$ and $z \in (0, 2)$, we have*

(i)

$$F_{J,\Delta}(0) - F_{J,\Delta}(\delta) = \begin{cases} \frac{\sqrt{\Delta}}{\sqrt{\vartheta_2 \pi}} + \mathcal{O}\left(\frac{\Delta^{3/2}}{\delta^2} \right), & J = 0, \\ \frac{\sqrt{\Delta}}{2\sqrt{\pi \vartheta_2}} \left(\sqrt{J-1} + \sqrt{J+1} - 2\sqrt{J} \right) + \mathcal{O}\left(\Delta^{3/2} + \frac{\Delta}{(J+1)\delta} \right), & J \geq 1, \end{cases}$$

(ii)

$$\begin{aligned} &2F_{J,\Delta}(\delta) - F_{J,\Delta}(0) - F_{J,\Delta}(2\delta) \\ &\quad = \begin{cases} -\frac{\sqrt{\Delta}}{\sqrt{\vartheta_2 \pi}} + \mathcal{O}\left(\frac{\Delta^{3/2}}{\delta^2} \right), & J = 0, \\ -\frac{\sqrt{\Delta}}{2\sqrt{\pi \vartheta_2}} \left(\sqrt{J-1} + \sqrt{J+1} - 2\sqrt{J} \right) + \mathcal{O}\left(\Delta^{3/2} + \frac{\Delta}{(J+1)\delta} \right), & J \geq 1, \end{cases} \end{aligned}$$

(iii)

$$2F_{J,\Delta}(z) - F_{J,\Delta}(z-\delta) - F_{J,\Delta}(z+\delta) = \mathcal{O}\left(\frac{\Delta}{J+1}\frac{1}{z\wedge(2-z)}\right).$$

Proof. (iii) It is sufficient to show

$$F_{J,\Delta}(z) = \mathcal{O}\left(\frac{\Delta}{J+1}\frac{1}{z\wedge(2-z)}\right) \tag{2.59}$$

for $J \in \mathbb{N}_0$ and $z \in (0,2)$: If $J=0$, Lemma 2.4.8 gives

$$\begin{aligned}
F_{\Delta}(z) &= \sum_{\ell\geq 1} \frac{1-\mathrm{e}^{-\lambda_\ell \Delta}}{\lambda_\ell}\cos(\pi\ell z) \\
&\lesssim \sup_{\ell\geq 1}\left|\frac{1-\mathrm{e}^{-\lambda_\ell\Delta}}{\lambda_\ell}\right| \frac{1}{z\wedge(2-z)} \lesssim \frac{\Delta}{z\wedge(2-z)}.
\end{aligned}$$

By (2.57) this bound is also valid for $F_{1,\Delta}(z)$. For $J\geq 2$, the same method gives

$$\begin{aligned}
F_{J,\Delta}(z) &= \sum_{\ell\geq 1}\frac{2\mathrm{e}^{-\lambda_\ell J\Delta} - \mathrm{e}^{-\lambda_\ell(J+1)\Delta} - \mathrm{e}^{-\lambda_\ell(J-1)\Delta}}{2\lambda_\ell}\cos(\pi\ell z) \\
&\lesssim \sup_{\ell\geq 1}\left|\frac{2\mathrm{e}^{-\lambda_\ell J\Delta} - \mathrm{e}^{-\lambda_\ell(J+1)\Delta} - \mathrm{e}^{-\lambda_\ell(J-1)\Delta}}{\lambda_\ell}\right|\frac{1}{z\wedge(2-z)} \\
&\lesssim \sup_{\ell\geq 1}\left|\lambda_\ell\Delta^2\mathrm{e}^{-\lambda_\ell J\Delta}\right|\frac{1}{z\wedge(2-z)} \\
&\lesssim \frac{\Delta}{J}\frac{1}{z\wedge(2-z)}
\end{aligned}$$

where we have used (2.58).

(i) The case $J=0$ was already shown in the proof of Proposition 2.2.5. For $J\geq 1$, we prove

$$F_{J,\Delta}(0) = \frac{\sqrt{\Delta}}{2\sqrt{\pi\vartheta_2}}\left(\sqrt{J-1}+\sqrt{J+1}-2\sqrt{J}\right) + \mathcal{O}(\Delta^{3/2}),$$

then (ii) follows in view of (2.59): If $J=1$, we use (2.33) to calculate

$$\begin{aligned}
F_{1,\Delta}(0) &= \frac{1}{2}F_{2\Delta}(0) - F_{\Delta}(0) \\
&= \frac{1}{2}\left(\frac{\sqrt{2\Delta}}{\sqrt{\pi\vartheta_2}} - \Delta\right) - \left(\frac{\sqrt{\Delta}}{\sqrt{\pi\vartheta_2}} - \frac{\Delta}{2}\right) + \mathcal{O}\left(\Delta^{3/2}\right) \\
&= \frac{\sqrt{\Delta}}{2\sqrt{\pi\vartheta_2}}\left(\sqrt{2}-2\right) + \mathcal{O}\left(\Delta^{3/2}\right).
\end{aligned}$$

For $J\geq 2$, define

$$g_J(z) := \frac{2\mathrm{e}^{-J\pi^2\vartheta_2 z^2} - \mathrm{e}^{-(J+1)\pi^2\vartheta_2 z^2} - \mathrm{e}^{-(J-1)\pi^2\vartheta_2 z^2}}{2\pi^2\vartheta_2 z^2}.$$

Then,

$$\int_0^\infty g_J(z)\,dz = \frac{1}{2\sqrt{\pi\vartheta_2}}\left(\sqrt{J-1}+\sqrt{J+1}-2\sqrt{J}\right)$$

and, since $g_J(0)=0$, we have by Lemma 2.4.10 that

$$F_{J,\Delta}(0)= \Delta\sum_{\ell\geq 1} g_J(\ell\sqrt{\Delta}) = \sqrt{\Delta}\int_0^\infty g_J(z)\,dz + \mathcal{O}\left(\Delta^{3/2}\right)$$
$$= \frac{\sqrt{\Delta}}{2\sqrt{\pi\vartheta_2}}\left(\sqrt{J-1}+\sqrt{J+1}-2\sqrt{J}\right) + \mathcal{O}(\Delta^{3/2}).$$

Finally, (ii) is a direct consequence of (i). □

Bounds on Fourier series and Riemann summation

The Lemmas in this section provide bounds for Fourier series and Taylor expansions for Riemann sums. They are our basic tools for computing and bounding covariances. Similar results are stated in Lemma 7.2 of Bibinger and Trabs [9]. Instead of the derivation in [9], which uses the decay of the Fourier transforms of L^1-functions, we show that these results also follow from the following simple lemma.

Lemma 2.4.8. *Let (a_n) be a real sequence and $\tau \in \{\sin, \cos\}$. Then,*

$$\left|\sum_{k=1}^N a_k\tau(ky)\right| \leq \frac{1+2K_N}{y\wedge(2\pi-y)}\sup_{n\leq N}|a_n|$$

holds for any $y\in(0,2\pi)$ where K_N is the number of monotone sections of $(a_n)_{1\leq n\leq N}$.

Proof. By Lagrange's trigonometric identities,

$$\sum_{k=1}^N \cos(ky) = \frac{\sin\left((N+1/2)y\right)-\sin(y/2)}{2\sin(y/2)},$$
$$\sum_{k=1}^N \sin(ky) = \frac{\cos(y/2)-\cos\left((N+1/2)y\right)}{2\sin(y/2)},$$

we have uniformly in $M\leq N$ that

$$\left|\sum_{k=M}^N \tau(ky)\right| \leq \frac{1}{\sin(y/2)} \leq \frac{1}{y\wedge(2\pi-y)}.$$

Therefore,

$$\begin{aligned}\left|\sum_{k=1}^N a_k\tau(ky)\right| &= \Bigg|a_1\sum_{k=1}^N\tau(ky) + (a_2-a_1)\sum_{k=2}^N\tau(ky)\\ &\qquad + (a_3-a_2)\sum_{k=3}^N\tau(ky)+\cdots+(a_N-a_{N-1})\tau(Ny)\Bigg|\\ &\leq |a_1|\left|\sum_{k=1}^N\tau(ky)\right| + |a_2-a_1|\left|\sum_{k=2}^N\tau(ky)\right|\\ &\qquad + |a_3-a_2|\left|\sum_{k=3}^N\tau(ky)\right|+\cdots+|a_N-a_{N-1}|\,|\tau(Ny)|\\ &\leq \frac{1}{y\wedge(2\pi-y)}\left(|a_1|+\sum_{k=1}^{N-1}|a_{k+1}-a_k|\right) \leq \frac{1+2K_N}{y\wedge(2\pi-y)}\sup_{n\leq N}|a_n|\end{aligned}$$

where the last inequality follows from the fact that, if $(a_k)_{N_0\le k\le N_1}$ is monotone for some $N_0 \le N_1 \le N$, then,

$$\sum_{k=N_0}^{N_1-1} |a_{k+1}-a_k| = |a_{N_1}-a_{N_0}| \le 2\sup_{n\le N} |a_n|. \qquad \square$$

Lemma 2.4.9. *Let $g \in C^1(\mathbb{R}_+)$ be such that g' is bounded and has a finite number K of monotone sections. Then, we have*

$$\sum_{k=1}^{\infty} g(k\varepsilon)\cos(ky) = -\frac{g(0)}{2} + \mathcal{O}\left(\frac{\varepsilon\,\|g'\|_\infty}{(y\wedge(2\pi-y))^2}\right),$$
$$\sum_{k=1}^{\infty} g(k\varepsilon)\sin(ky) = \frac{g(0)}{2}\cot\left(\frac{y}{2}\right) + \mathcal{O}\left(\frac{\varepsilon\,\|g'\|_\infty}{(y\wedge(2\pi-y))^2}\right),$$

as $\varepsilon\to 0$, for $y\in(0,2\pi)$.

Proof. We use the formula $\sin(\alpha)-\sin(\beta) = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$, $\alpha,\beta\in\mathbb{R}$, to calculate

$$\begin{aligned}
&\frac{g(0)}{2} + \sum_{k=1}^{\infty} g(k\varepsilon)\cos(ky)\\
&= \frac{g(0)}{2} + \frac{1}{2\sin\frac{y}{2}}\sum_{k=1}^{\infty} g(k\varepsilon)\big(\sin((k+1/2)y) - \sin((k-1/2)y)\big)\\
&= \frac{g(0)}{2} - \frac{g(\varepsilon)}{2} + \frac{1}{2\sin\frac{y}{2}}\sum_{k=1}^{\infty}\sin((k+1/2)y)\big(g(k\varepsilon)-g((k+1)\varepsilon)\big)\\
&= -\frac{1}{2}\left(g'(\xi_0^\varepsilon) + \frac{1}{\sin\frac{y}{2}}\sum_{k=1}^{\infty}\sin((k+1/2)y)\,g'(\xi_k^\varepsilon)\right)\varepsilon\\
&\le \frac{1+2K}{(y\wedge(2\pi-y))^2}\|g'\|_\infty\,\varepsilon
\end{aligned}$$

where $\xi_k^\varepsilon\in[k\varepsilon,(k+1)\varepsilon]$. Here, the last step follows from

$$\sin((k+1/2)y) = \sin(ky)\cos(y/2) + \cos(ky)\sin(y/2)$$

and then applying Lemma 2.4.8. Similarly, the second statement follows from $\cos(\alpha)-\cos(\beta) = -2\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)$, $\alpha,\beta\in\mathbb{R}$, and

$$\begin{aligned}
&-\frac{g(0)}{2}\cot\left(\frac{y}{2}\right) + \sum_{k=1}^{\infty} g(k\varepsilon)\sin(ky)\\
&= -\frac{g(0)}{2}\cot\left(\frac{y}{2}\right) - \frac{1}{2\sin\frac{y}{2}}\sum_{k=1}^{\infty} g(k\varepsilon)\big(\cos((k+1/2)y)-\cos((k-1/2)y)\big)\\
&= -\frac{g(0)}{2}\cot\left(\frac{y}{2}\right) + \frac{g(\varepsilon)}{2}\cot\left(\frac{y}{2}\right) - \frac{1}{2\sin\frac{y}{2}}\sum_{k=1}^{\infty}\cos((k+1/2)y)\big(g(k\varepsilon)-g((k+1)\varepsilon)\big)\\
&= \left(\frac{g'(\xi_0^\varepsilon)}{2}\cot\left(\frac{y}{2}\right) + \frac{1}{2\sin\frac{y}{2}}\sum_{k=1}^{\infty}\cos((k+1/2)y)\,g'(\xi_k^\varepsilon)\right)\varepsilon. \qquad \square
\end{aligned}$$

Lemma 2.4.10. *Let $g\in C^2(\mathbb{R}_+)\cap L^1(\mathbb{R}_+)$ be such that $g'\in L^\infty(\mathbb{R}_+)$ and $g''\in L^1(\mathbb{R}_+)$. Then, as $\varepsilon\to 0$,*

(i)
$$\varepsilon \sum_{k\geq 1} g(k\varepsilon) = \int_0^\infty g(z)\,dz - \frac{g(0)}{2}\varepsilon + \mathcal{O}(\varepsilon^2 \|g''\|_{L^1}),$$

(ii)
$$\varepsilon \sum_{k\geq 1} g(k\varepsilon)\sin^2(ky) = \frac{1}{2}\int_0^\infty g(z)\,dz + \mathcal{O}\left(\varepsilon^2 \left(\frac{\|g'\|_\infty}{(y\wedge(\pi-y))^2} \wedge \|g''\|_{L^1}\right)\right).$$

Proof. For a detailed proof of (i), we refer to [9, Lemma 7.2]. The main idea is to regard each term $\varepsilon g(k\varepsilon)$ as a midpoint integral approximation, see also the proof of the following emma. Since $\sin^2(y) = (1-\cos(2y))/2$, statement (ii) is a direct consequence of (i) and the previous lemma. □

Lemma 2.4.11. *Let $g \in C^2(\mathbb{R}_+)$. Then, as $M \to \infty$, $M\epsilon \to 0$,*
$$\epsilon \sum_{k=1}^M g(k\epsilon) = M\epsilon g(0) + \frac{(M^2+M)\epsilon^2}{2} g'(0) + \mathcal{O}((M\epsilon)^3).$$

Proof. First of all, by the midpoint rule, there exist $\eta_k \in [(k-1/2)\epsilon, (k+1/2)\epsilon]$ such that
$$\begin{aligned}\left|\epsilon \sum_{k=1}^M g(k\epsilon) - \int_{\epsilon/2}^{(M+1/2)\epsilon} g(x)\,dx\right| &= \left|\sum_{k=1}^M \int_{(k-1/2)\epsilon}^{(k+1/2)\epsilon} (g(k\epsilon)-g(x))\,dx\right| \\ &\leq \epsilon^3 \sum_{k=1}^M |g''(\eta_k)| \lesssim M\epsilon^3 \lesssim M^3\epsilon^3.\end{aligned}$$

Secondly, a Taylor approximation shows that
$$\begin{aligned}\int_{\epsilon/2}^{(M+1/2)\epsilon} g(x)\,dx &= \int_0^{(M+1/2)\epsilon} g(x)\,dx - \int_0^{\epsilon/2} g(x)\,dx \\ &= (M+1/2)\epsilon g(0) + \frac{((M+1/2)\epsilon)^2}{2} g'(0) + \mathcal{O}((M\epsilon)^3) \\ &\quad - \epsilon g(0)/2 - \frac{1}{8}\epsilon^2 g'(0) + \mathcal{O}(\epsilon^3) \\ &= M\epsilon g(0) + \frac{(M^2+M)\epsilon^2}{2} g'(0) + \mathcal{O}((M\epsilon)^3).\end{aligned}$$
□

Chapter 3

Generating fully discrete samples for the linear equation

Inspired by the application of performing simulations on the estimators derived in Section 2.2, in this chapter we develop and analyze a suitable method for generating fully discrete samples of the solution to the linear SPDE (1.11). As already pointed out in the introduction, simulation methods in the literature generally do not capture the exact Hölder regularity properties of the underlying model and, hence, are not suited for computing power variations. Our simulation method, which we call the *replacement method*, generalizes an idea stated in Davie and Gaines [29] without providing a theoretical justification. Using the Gaussian property of the true and the approximate model, we derive conditions for the corresponding total variation distance to tend to zero as the number of observations tends to infinity. In particular, this preserves the asymptotic properties of our estimators. Except for the simulation study on the estimators, the results of this chapter can be found in Hildebrandt [36]. The simulations for the estimators on a fixed time horizon are taken from Hildebrandt and Trabs [38].

In Section 3.1 we introduce the replacement method and state our convergence result. Section 3.2 is devoted to a numerical example illustrating the accuracy of the replacement method. In particular, it is compared to the *truncation method*, i.e., naive truncation in Fourier space. The results provided by the replacement method turn out to be more accurate at a considerably lower computational cost. Owing to our original purpose, Section 3.3 contains a simulation study for the estimators derived in Section 2.2. Finally, the proofs are collected in Section 3.4.

Throughout, $X = (X_t(x), t \in \mathbb{R}_+, x \in [0,1])$ denotes the solution field given by (1.12) with stationary or zero initial condition.

3.1 Simulation method and convergence result

Our aim is to generate discrete samples $(X_{t_i}(y_k), i \leq N, k \leq M)$ where the points of observation (t_i, y_k) are as defined in the observation scheme from Section 1.2.3 with $b = 0$, i.e.,

$$y_k = \frac{k}{M},\ k = 0, \dots, M, \qquad t_i = \frac{iT}{N},\ i = 0, \dots N,$$

and all of the numbers $N, M \in \mathbb{N}_0$ and $T > 0$ are allowed to tend to infinity, in general. Assuming $b = 0$ allows us to take advantage of the discrete version of the orthogonality property (2.3) of the eigenfunctions e_ℓ. In fact, with U_m defined in (2.4), we have $X_{t_i}(y_k) = \sum_{m=1}^{M-1} U_m(t_i) e_m(y_k)$ for all $i \leq N$, $k \leq M$ and, thus, sampling from X at the grid points (t_i, y_k) is equivalent to sampling from the independent processes U_m, $m \leq M-1$, at times $t_0, \dots, t_N$. Further, any coefficient process u_ℓ may be simulated exactly using its AR(1)-structure, namely

$$u_\ell(0) = \langle \xi, \mathrm{e}_\ell \rangle_\vartheta, \qquad u_\ell(t_{i+1}) = \mathrm{e}^{-\lambda_\ell \Delta} u_\ell(t_i) + \sigma \sqrt{\frac{1 - \mathrm{e}^{-2\lambda_\ell \Delta}}{2\lambda_\ell}} N_i^\ell, \quad i \in \mathbb{N}_0, \tag{3.1}$$

where (N_i^ℓ) are independent standard normal random variables.

To derive the simulation method, let us first assume that $X_0 = 0$. In this case, the coefficient processes u_ℓ are centered Gaussian with covariance function

$$\mathrm{Cov}(u_\ell(t_i), u_\ell(t_j)) = \frac{\sigma^2}{2\lambda_\ell} \mathrm{e}^{-\lambda_\ell |i-j| \Delta} \left(1 - \mathrm{e}^{-2\lambda_\ell \min(i,j)\Delta}\right), \qquad 0 \leq i,j \leq N.$$

Thus, when $\lambda_\ell \approx \ell^2$ is large compared to $1/\Delta$, the random variables $(u_\ell(t_i), 1 \leq i \leq N)$ effectively behave like iid Gaussian random variables with variance

$$\mathrm{Var}(u_\ell(t_i)) \approx \frac{\sigma^2}{2\lambda_\ell}, \quad 1 \leq i \leq N,$$

due to the exponential factor $\mathrm{e}^{-\lambda_\ell |i-j| \Delta}$ in the covariance. Now, in order to define the approximation of the processes U_m, choose $L = L_{M,N} \in \mathbb{N}$ and replace all coefficient processes $(u_\ell(t_i),\, 1 \leq i \leq N)$ with $\ell \geq LM$ by a vector of independent normal random variables with variance $\sigma^2/(2\lambda_\ell)$. Note that counting in multiples of M is convenient due to the particular form of the index sets $\mathcal{I}_m = \mathcal{I}_m^+ \cup \mathcal{I}_m^-$ from the definition of U_m. Since the normal distribution is stable with respect to summation, for each $m < M$ it is sufficient to generate one set $(R_m^L(i),\, 1 \leq i \leq N)$ of independent random variables with $R_m^L(i) \sim \mathcal{N}(0, s_m^2)$ where

$$s_m^2 := \sum_{\ell \in \mathcal{I}_m,\, \ell \geq LM} \frac{\sigma^2}{2\lambda_\ell}. \tag{3.2}$$

The resulting approximation is defined by

$$U_m^L(0) := 0, \qquad U_m^L(t_i) := \sum_{\ell \in \mathcal{I}_m, \ell < LM} u_\ell(t_i) + R_m^L(i), \quad 1 \leq i \leq N.$$

Similarly, if X is the stationary solution, the coefficient processes u_ℓ are centered Gaussian with covariance function

$$\mathrm{Cov}(u_\ell(t_i), u_\ell(t_j)) = \frac{\sigma^2}{2\lambda_\ell} \mathrm{e}^{-\lambda_\ell |i-j| \Delta}, \qquad 0 \leq i,j \leq N.$$

Consequently, for iid random variables $(R_m^L(i),\, 0 \leq i \leq N)$ with $R_m^L(i) \sim \mathcal{N}(0, s_m^2)$ we define the approximation

$$U_m^L(t_i) := \sum_{\ell \in \mathcal{I}_m, \ell < LM} u_\ell(t_i) + R_m^L(i), \quad 0 \leq i \leq N.$$

In order to generate samples based on the replacement method, it is necessary to calculate the variances s_m^2. In fact, approximating the infinite series (3.2) can be avoided thanks to the closed form expression provided by the following lemma. We use the notation $\rho : [0,1] \to \mathbb{R}$ for the covariance function of the stationary initial condition with $\kappa = 0$, i.e., for the symmetric function given by

$$\rho(x,y) := \frac{\sigma^2}{2\vartheta_2} \cdot \begin{cases} \frac{\sin(\Gamma_0(1-y))\sin(\Gamma_0 x)}{\Gamma_0 \sin(\Gamma_0)}, & \Gamma < 0, \\ x(1-y), & \Gamma = 0, \\ \frac{\sinh(\Gamma_0(1-y))\sinh(\Gamma_0 x)}{\Gamma_0 \sinh(\Gamma_0)}, & \Gamma > 0, \end{cases} \qquad \text{for } x \leq y$$

with $\Gamma = \frac{\vartheta_1^2}{4\vartheta_2^2} - \frac{\vartheta_0}{\vartheta_2}$ and $\Gamma_0 = \sqrt{|\Gamma|}$, cf. Proposition 1.2.1.

Lemma 3.1.1. *Define* $\Sigma \in \mathbb{R}^{(M+1)\times(M+1)}$ *via* $\Sigma_{kl} := \rho(y_k, y_l)$ *for* $0 \le k, l \le M$ *and let* $b_m := \sqrt{2}(\sin(\pi m y_0), \dots, \sin(\pi m y_M))^\top \in \mathbb{R}^{M+1}$. *The variance* s_m^2 *defined by* (3.2) *satisfies*

$$s_m^2 = \frac{1}{M^2} b_m^\top \Sigma b_m - \sum_{\ell \in \mathcal{I}_m,\, \ell < LM} \frac{\sigma^2}{2\lambda_\ell}. \tag{3.3}$$

Our simulation method is summarized in the following algorithm:

Algorithm 3.1.2 (Replacement method)**.** Choose $L \in \mathbb{N}$.

For $1 \le m < M$ do the following:

(1) For $\ell \in \mathcal{I}_m \cap (0, LM)$ simulate $(u_\ell(t_i),\, 0 \le i \le N)$ according to (3.1).

(2) Compute s_m^2 according to (3.3) and generate $R_m^L(0), \dots, R_m^L(N) \sim \mathcal{N}(0, s_m^2)$ independently. For the zero initial condition replace $R_m^L(0)$ by 0.

(3) Compute
$$U_m^L(t_i) = \sum_{\ell \in \mathcal{I}_m, \ell < LM} u_\ell(t_i) + R_m^L(i), \quad 0 \le i \le N.$$

Output: $X_{t_i}^L(y_k) = \sum_{m=1}^{M-1} U_m^L(t_i) e_m(y_k)$ for $0 \le k \le M$ and $0 \le i \le N$.

Assuming a finite set of observations, Davie and Gaines [29] proposed the replacement method with $L = 1$, while omitting a theoretical analysis. Based on a bound on the total variation distance of Gaussians by Devroye et al. [30], namely

$$\mathrm{TV}\Big(\mathcal{N}(0, A), \mathcal{N}(0, B)\Big) \le \frac{3}{2} \|A^{-1/2}(B - A)A^{-1/2}\|_F \tag{3.4}$$

for non-singular covariance matrices A and B, we are able to theoretically justify their approach. In particular, allowing for $M, N \to \infty$, the following theorem provides a condition on L for the validity of the approximation in total variation distance.

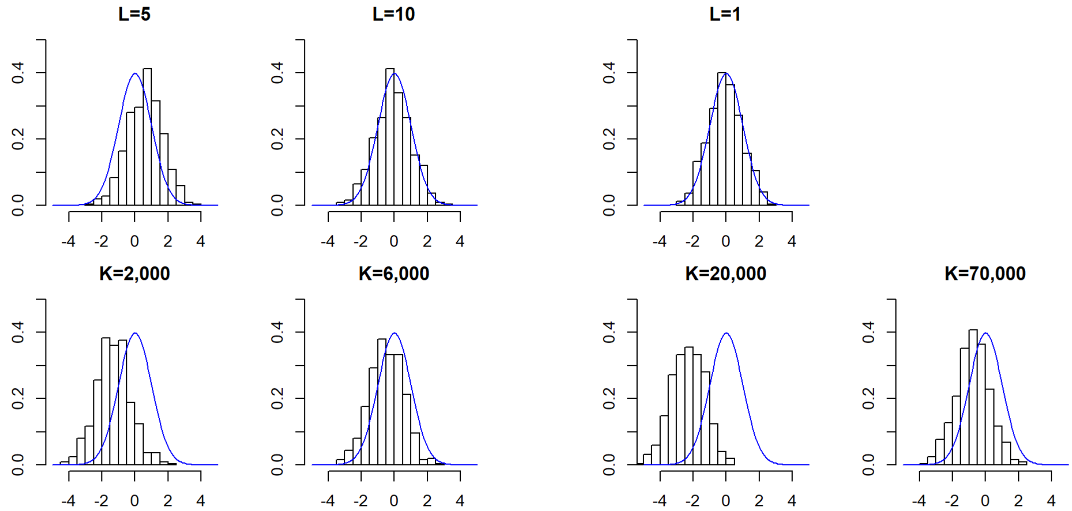
Theorem 3.1.3. *Let* $\mathcal{X}$ *be the vector of observations, i.e.,* $\mathcal{X} := (X_{t_i}(y_k),\, i \le N, k \le M)$ *either with zero or with stationary initial condition and let* $\mathcal{X}^L$ *be its approximation computed via Algorithm 3.1.2.*

(i) There exist constants $c, C > 0$ *only depending on the parameters* (σ^2, ϑ) *such that*
$$\mathrm{TV}(\mathcal{X}, \mathcal{X}^L) \le C\sqrt{MN} \mathrm{e}^{-cL^2M^2\Delta}.$$

(ii) Assume $T\Delta^q \to 0$ *for some* $q > 0$. *If there exists* $\alpha > 1/2$ *such that* $LM\Delta^\alpha \to \infty$, *then* $\mathrm{TV}(\mathcal{X}, \mathcal{X}^L) \to 0$. *In particular, if* $T = \mathrm{const.}$ *and* $M/N^\alpha \to \infty$ *for some* $\alpha > 1/2$, *then* $\mathrm{TV}(\mathcal{X}, \mathcal{X}^1) \to 0$.

A negligible total variation distance is exactly what is required for statistical simulations since functionals based on true and approximate data share the same limiting distribution: Let $(X_{n,k})$ and $(Y_{n,k})$ be triangular arrays of the same size and assume that $\phi_n(X_{n,\bullet})$ has a weak limit Z for some sequence (ϕ_n) of functionals. Then, if $\mathrm{TV}(X_{n,\bullet}, Y_{n,\bullet}) \to 0$, the sequence $\phi_n(Y_{n,\bullet})$ also converges to Z weakly. In fact, if μ_n is a dominating measure for the laws of $X_{n,\bullet}$ and $Y_{n,\bullet}$ with corresponding Radon-Nikodym derivatives $f_{X_{n,\bullet}}$ and $f_{Y_{n,\bullet}}$, then

$$\begin{aligned}\big|\mathbf{E}(\mathrm{e}^{it\phi_n(X_{n,\bullet})}) - \mathbf{E}(\mathrm{e}^{it\phi_n(Y_{n,\bullet})})\big| &= \Big|\int \mathrm{e}^{it\phi_n(z)}(f_{X_{n,\bullet}}(z) - f_{Y_{n,\bullet}}(z))\,\mu(dz)\Big| \\ &\le \|f_{X_{n,\bullet}} - f_{Y_{n,\bullet}}\|_{L^1(\mu)} = 2\mathrm{TV}(X_{n,\bullet}, Y_{n,\bullet}).\end{aligned}$$



(a) temporal quadratic variation for $N = 5,000$, $M = 10$

(b) spatial quadratic variation for $N = 100$, $M = 1,000$

Figure 3.1: Histograms based on 500 Monte Carlo iterations for normalized quadratic variations based on the replacement (top) and truncation method (bottom). The solid line corresponds to the standard normal density function.

Thus, the limiting characteristic functions coincide.

Another aspect worth noting is that there is no statistical test that can consistently distinguish between two models whose total variation distance tends to zero. Indeed, in such a case, the maximum of type one and type two error of any test for the true model is asymptotically bounded from below by 1/2, see e.g. the proof of [75, Theorem 2.2].

3.2 Simulations on the accuracy of the replacement method

In order to test the performance of the replacement method and compare it to truncation of the Fourier series, we compute rescaled realized temporal and spatial quadratic variations, namely V_{t} from (2.9) and V_{sp} from (2.12), based on both methods on the finite time horizon $T = 1$. The outcomes are then compared with the corresponding theoretical limiting distributions given by (2.10) and Theorem 2.2.3. For the simulations we have set the parameters to the values $\sigma^2 = 0.1$, $\vartheta_2 = 0.5$, $\vartheta_1 = -0.4$, $\vartheta_0 = 0.3$ and have considered the stationary initial condition. Each of the plots in Figures 3.1a and 3.1b shows a histogram of the centered and normalized (with respect to theoretical asymptotic means and variances) realized quadratic variations based on 500 Monte Carlo iterations. The solid line corresponds to the standard normal density function.

For the temporal quadratic variation (Figure 3.1a) we have considered $M = 10$ spatial and $N = 5,000$ temporal observations. As long as $N \to \infty$, the central limit theorem for time increments also holds for finite M. Hence, one can expect that the asymptotic regime is reached with these values for N and M. Indeed, Figure 3.1a shows that the values provided by the replacement method with $L = 10$ (corresponding to $LM = 100$ simulated Ornstein-Uhlenbeck processes) are already in good accordance with the theoretical limit. Note that $LM\sqrt{\Delta} \approx 3.2$ is far from infinity, so the method works better than predicted by Theorem 3.1.3. The truncation method, on the other hand, requires simulation of more than $6,000$ coefficient processes in order to produce accurate results and prevent a severe bias in the simulated values.

Examining the results for the spatial quadratic variation (Figure 3.1b), this effect becomes even more apparent. Here, we considered $M = 1,000$ spatial and $N = 100$ temporal observations. Consequently, $M\sqrt{\Delta} = 100$ and Theorem 3.1.3 suggests that $L = 1$ (i.e., $LM = 1,000$ simulated coefficient processes) is sufficient for the replacement method. Figure 3.1b confirms this prediction. On the other hand, even with $K = 70,000$ coefficient processes, the simulated values based on the truncation method still suffer from a severe bias.

In fact, the bias in the central limit theorems introduced by truncation can be explained analytically: A simple calculation shows that for the normalized temporal quadratic variation, the bias is of the order $\sqrt{MN}\frac{1}{\sqrt{\Delta}}\sum_{\ell\geq K}\frac{1}{\lambda_\ell} \approx \frac{\sqrt{MN}}{K\sqrt{\Delta}}$ and in our simulation for the temporal quadratic variation we have $\frac{\sqrt{MN}}{\sqrt{\Delta}} \approx 16,000$. Similarly, the bias for the spatial quadratic variation is of the order $\sqrt{MN}\frac{1}{\delta}\sum_{\ell\geq K}\frac{1}{\lambda_\ell} \approx \frac{\sqrt{MN}}{K\delta}$, in our simulation we have $\frac{\sqrt{MN}}{\delta} \approx 316,000$.

3.3 Simulation study for the estimators from Section 2.2

The following numerical examples illustrate the asymptotic results for the estimators derived in Section 2.2. In order to simulate X on a grid in time and space, we use the replacement method developed in Section 3.1. The estimators applied to the simulated data then have the correct limiting distribution, as follows from Theorem 3.1.3 and its subsequent discussion.

3.3.1 The case of a fixed time horizon

Letting $T = 1$, we have considered a fixed number $N = 2^{10}$ or $N = 2^{14}$ of temporal observations, while M varies in the set $\{15, 29, 57, 113, 225, 449, 897, 1793, 3585, 7169\}$. The precise values for M stem from the procedure of lying a dyadic grid on $[0,1]$ and then removing the points on the margin $[0,b)\cup(b-1,1]$ where $b = 2^{-4}$. In fact, all observations are obtained as subsets of a simulation of X on the full grid $\big((i/\bar{N}, k/\bar{M}),\, i \leq \bar{N}, k \leq \bar{M}\big)$ with $\bar{M} = 2^{13}$ and $\bar{N} = 2^{14}$. We have used the replacement method with $L = 1$ which is justified by Theorem 3.1.3 in view of $\bar{M}^2/\bar{N} = 2^{12} \gg 1$. The parameters are chosen to be $\sigma^2 = 0.1$, $\vartheta_2 = 0.5$, $\vartheta_1 = -0.4$ and $\vartheta_0 = 0.3$.

First, we consider the estimators for the volatility σ^2 and the diffusivity ϑ_2 which have been analyzed in Propositions 2.2.10 and 2.2.11, respectively. Figure 3.2 shows the normalized (with respect to $1/(MN)$ as well as the constants σ^4 and ϑ_2^2, respectively) mean squared errors based on 500 Monte Carlo iterations plotted against the logarithm of the sampling ratio $\sqrt{N}/M$. The simplified double increments estimator $\hat{\vartheta}_{2,r}$ is computed with $r = (1-2b)\frac{\sqrt{N}}{M}$. Using the same value for r, the simplified double increments estimator for σ^2 is computed by replacing the normalization $\Phi_\vartheta(\delta,\Delta)$ with $\mathrm{e}^{-\kappa\delta/2}\psi_{\vartheta_2}(r)\sqrt{\Delta}$.

As expected, the estimators based on temporal increments only achieve the parametric rate of convergence as long as M is not too large, whereas estimators based on space increments only work well when M is not too small. The estimators based on double increments perform very well throughout any regime depicted in the plot. Even the simplified versions work surprisingly well, although their applicability is only supported by our theory as long as $M \approx \sqrt{N}$. In particular, the double increments estimator for σ^2 can barely be distinguished from the simplified one. Furthermore, as suggested by the theory, the simulations show that the estimators based on space increments or time increments have a smaller mean squared error than the double increments estimators in the regimes $\sqrt{N}/M \to 0$ or $\sqrt{N}/M \to \infty$, respectively.

The above estimators require all but one of the parameters $(\sigma^2, \vartheta_2, \kappa)$ to be known. Within the more difficult statistical problem where all parameters are unknown, $\eta = (\sigma^2, \vartheta_2, \vartheta_1)$ can be estimated by $\hat{\eta}$ from (2.18) and by $\hat{\eta}_{v,w}$ from (2.19). Furthermore, we have implemented a data-

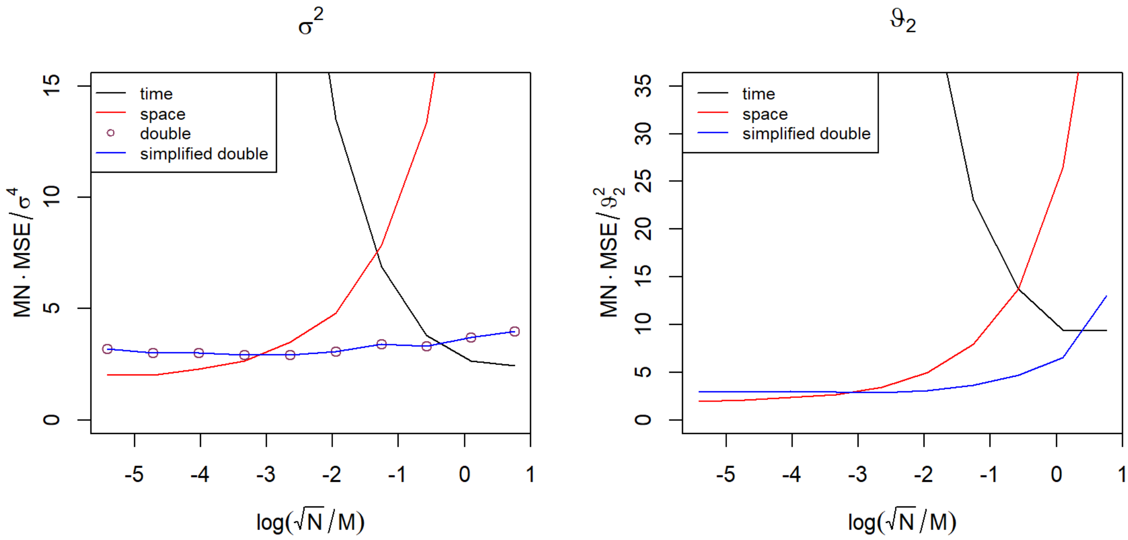


Figure 3.2: Normalized mean squared errors of estimators for σ^2 (left) and ϑ_2 (right) with $T = 1$, $N = 2^{10}$ and $M \in [15, 7169]$.

thinning version of the estimator where the contrast process uses only *one* balanced sub-sample and discards the remaining data instead of averaging. For the estimator $\hat{\eta}_{v,w}$ and its thinning version, we set $v = [\max(1, \frac{N}{M^2})]$ and $w = [\max(1, M/\sqrt{N})]$ where $[\cdot]$ indicates rounding to the next integer. The minimization problems were numerically solved using the nonlinear least squares function `nls` from `R`. Figure 3.3 shows the logarithm of the mean squared errors plotted against the logarithm of the sampling ratio $\sqrt{N}/M$, again based on 500 Monte Carlo iterations. Here, displaying the mean squared errors on the logarithmic scale helps in distinguishing the different curves and provides a close-up view at their behavior when they are very small.

For the fixed value $N = 2^{14}$, taking $M = 113$ results in a balanced regime and, in particular, we have $v = w = 1$. Thus, the definitions of all estimators agree, leading to an intersection of the three curves at $\log(\sqrt{N}/M) \approx 0.12$. In contrast to the double increments estimators for single parameters, $\hat{\eta}$ only produces good results as long as $M \approx \sqrt{N}$, which is covered by the theoretical foundation. In fact, with the smallest number of spatial observations, $M = 15$, the optimization algorithm was even unable to detect a minimum in almost 3/5 of the simulation runs and the mean squared error is computed based on the remaining data. Unsurprisingly, the other two estimators have, overall, a much better performance. On the contrary, when $M = 57$ ($\log(\sqrt{N}/M) \approx -0.56$) the estimator $\hat{\eta}$ works slightly better. This can be explained by the fact that, here, the choice of v and w is too conservative in the sense that $v \vee w > 1$ although the regime is still reasonably balanced. Furthermore, we see that it is only possible to profit from an increasing number of spatial observations up to a certain degree: For $M \leq \sqrt{N}$, the optimal rate is $M^{-3/2}$ and the empirical mean squared error of $\hat{\eta}_{v,w}$ as well as its thinning version becomes increasingly smaller. For $M \geq \sqrt{N}$, the optimal rate is $N^{-3/4}$ and, indeed, the empirical mean squared errors become stationary. Furthermore, while the latter two estimators have a similar qualitative behavior, the mean squared error of $\hat{\eta}_{v,w}$ is consistently smaller. As announced in Remark 2.2.15, this indicates that using the whole data results in an improved asymptotic variance.

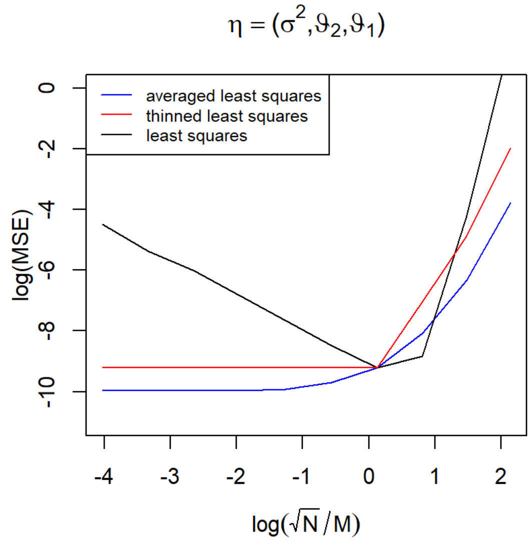


Figure 3.3: Logarithm of the mean squared errors of the least squares estimator $\hat{\eta}$, its averaging version $\hat{\eta}_{v,w}$ and a thinning version exploiting only one balanced sub-sample. The time horizon is $T = 1$ and the sample sizes are $N = 2^{14}$ and $M \in [15, 7169]$.

3.3.2 The case $T \to \infty$

In order to illustrate the conclusion of Theorem 2.2.18 on the estimator $(\hat{\eta}, \hat{\vartheta}_0)$ in case of a balanced sampling design, we generated 2,500 samples of $(X_{t_i}(\bar{y}_k))_{k,i}$ where $\bar{y}_k = k/\bar{M}$, $t_i = i\Delta$ with $\Delta = 4/\bar{M}^2$ and $T \in \{25, 50, 75, 100\}$ as well as $\bar{M} \in \{75, 90, 100\}$. To that aim, we used the replacement method with parameter $L = 4$. For computing the estimators, we only used the spatial observations satisfying $\bar{y}_k \in [b, 1-b]$ with $b = 0.05$. The parameter values are $\sigma^2 = 0.1$, $\vartheta_2 = 0.5$, $\vartheta_1 = -0.4$ and $\vartheta_0 = 4$. For smaller values of ϑ_0, the estimation problem becomes increasingly harder. This is caused by the fact that the function $I_b : (-\pi^2, \infty) \to \mathbb{R}$, whose inverse is used for the estimator (2.22) for ϑ_0, is quite flat around arguments not close to $-\pi^2$. Thus, larger sample sizes would be required to illustrate the precise features of the estimator.

The top row of Figure 3.4 shows the empirical mean squared errors of $\hat{\eta}$ and $\hat{\vartheta}_0$ as a function of T, normalized with respect to their theoretical orders of magnitude, $1/(MN)$ and $1/T$, respectively. As a baseline, the plot for ϑ_0 also shows the normalized mean squared errors when the estimators for ϑ_0 are computed using the actual values of η instead of their estimates (circular dots). Note that for different values of $\bar{M}$, the mean squared errors are renormalized in a different way, namely, we multiply by $MN = TM/\Delta \approx TM^{3/2}$, in view of the balanced sampling design. Furthermore, the dotted lines in the plots represent the estimated mean squared errors plus/minus a Monte Carlo estimate of their standard deviations. E.g., when $Z_1, \dots, Z_n$ with $n = 2,500$ are the Monte Carlo realizations of $\hat{\vartheta}_0$, define $Y_i := (Z_i - \vartheta_0)^2$. An estimate for the mean squared error of $\hat{\vartheta}_0$ is then given by the empirical mean of $Y_1, \dots, Y_n$. The standard deviation of this empirical mean squared error, in turn, can be estimated by the sample standard deviation of $Y_1, \dots, Y_n$ divided by $\sqrt{n}$. Let us refer to the regions between the dotted lines as confidence bands. The bottom row in Figure 3.4 shows the decomposition of the mean squared error of $\hat{\eta}$ into the empirical variance (i.e., the sum of the three individual variances) and the squared empirical bias (i.e., the sum of squares of the three individual biases).

The normalized mean squared errors of the estimators for ϑ_0 seem to approach a finite value as

$T \to \infty$, which confirms the $1/\sqrt{T}$-rate of convergence of $\hat{\vartheta}_0$. In view of the confidence bands, the slight deviation in the case $\bar{M} = 90$ does not contradict this general behavior. Furthermore, we see that with $\bar{M} = 100$ spatial observations, our estimator for ϑ_0 is already very close to the estimator employing the true values of η, which shows that our plug-in approach with $\hat{\eta}$ works very well.

The asymptotic properties of the estimator $\hat{\eta}$ are largely confirmed by the simulation study as well, though some precise features are obscured by the fluctuations in the renormalized system. Indeed, as suggested by the displayed empirical mean squared errors of $\hat{\eta}$, the asymptotic distribution of $\hat{\eta}$ has a fairly large variance, which seems acceptable in view of the fast rate of convergence. In particular, the asymptotic mean squared error itself is hard to assess precisely, even based on 2,500 Monte Carlo iterations. This is also reflected in the rather wide confidence bands. They indicate that based on a second Monte Carlo simulation, the displayed curves might as well be in a different vertical ordering. On the other hand, it seems plausible from the plot that for a fixed value of T, the mean squared errors remain bounded as a function of M, which is in line with the parametric rate of convergence for fixed T. For individual values of M, the mean squared errors show a linear growth in T. Inspection of the decomposition into variance and bias in the bottom row of Figure 3.4, indicates that this is purely due to a bias effect. This is covered by our theory: It is suggested by Proposition 2.2.5 that the estimator $\hat{\eta}$ has a bias of the order $\mathcal{O}(\Delta)$. Thus, when multiplying the squared bias by MN, we get an overall error of the order $\mathcal{O}(MN\Delta^2) = \mathcal{O}(T/M)$ in the balanced design. For a fixed value of M, this exactly explains the linear growth observed for the squared bias. The bias is not present in our central limit theorem for $\hat{\eta}$ since it is proved under the condition that $T/M \approx T\sqrt{\Delta} \to 0$. Illustrating the decay of the bias in M in simulations would require very large samples sizes with $M \gg T$, which is beyond the scope of this simulation study.

Figure 3.5 shows standard normal QQ-plots for the estimators where, as usual, the sample quantiles are plottet against the theoretical quantiles of the standard normal distribution, accompanied by a solid line through the first and third quartiles. Due to the scaling invariance of the normal distribution, QQ-plots following a straight line indicate that the empirical distribution is approximately normal. QQ-plots for the estimators of the four parameters with $\bar{M} = 90$ and $T = 25$ as well as $T = 100$ are displayed.

Our theory predicts that the estimators for $\eta = (\sigma^2, \vartheta_2, \vartheta_1)$ should be asymptotically normal as soon as M is sufficiently large, no matter the value of T. This prediction is reflected well in the QQ-plots. Asymptotic normality of $\hat{\vartheta}_0$, on the other hand, can only be expected when T is large. Indeed, at time $T = 25$, the QQ-plot shows a clear deviation from a straight line. At time $T = 100$, the empirical distribution already seems much closer to the normal distribution, although some deviation remains. This can be explained by the fact that $T = 100$ is still not very large, in particular, when comparing to the number $MN \approx T\bar{M}^3 \approx 7.3 \cdot 10^7$ governing the asymptotics for $\hat{\eta}$.

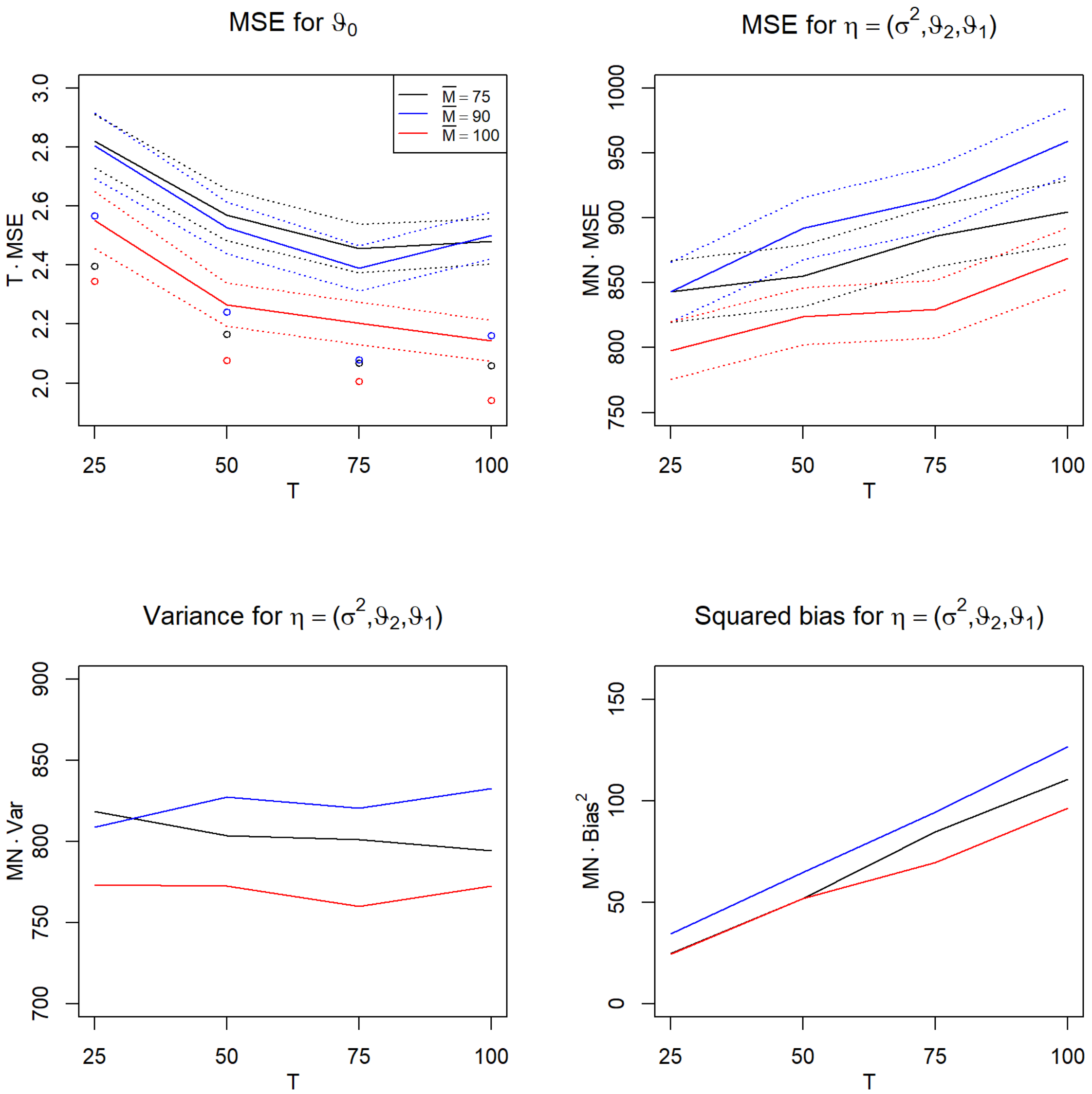


Figure 3.4: Normalized (with respect to the theoretical rates of convergence) mean squared errors of $\hat{\vartheta}_0$ and $\hat{\eta}$ (top) as well as normalized empirical variance and squared bias for $\hat{\eta}$ (bottom) in the balanced sampling design $\Delta = 4/\bar{M}^2$. Circular dots correspond to the estimators for ϑ_0 when using the actual values of η instead of $\hat{\eta}$. Dotted lines represent the estimated mean squared errors plus/minus their empirical (with respect to the Monte Carlo simulation) standard deviations.

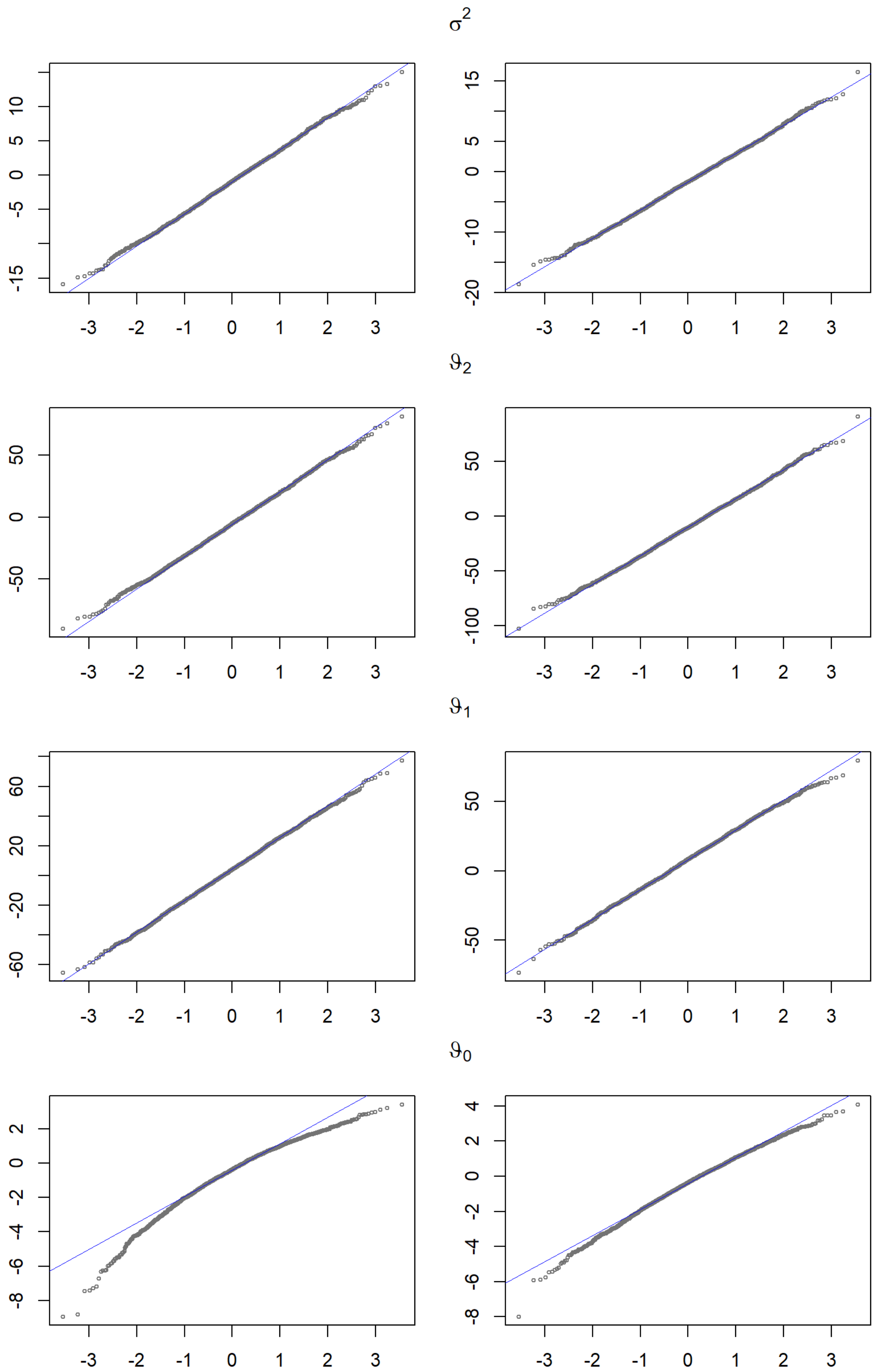


Figure 3.5: Normal QQ-plots for the centered (around the true values) and normalized (with respect to the theoretical rates of convergence) estimators from Theorem 2.2.18 with $\bar{M} = 90$, $\Delta = 4/\bar{M}^2 \approx 4.9 \cdot 10^{-4}$ and $T = 25$ (left) as well as $T = 100$ (right).

3.4 Proofs

First, we prove the closed form expression for the variance term s_m^2:

Proof of Lemma 3.1.1. It follows from Proposition 1.2.1 that Σ is the covariance matrix of the vector $Z := (\mathrm{e}^{\kappa y_1/2}X_0(y_0), \dots, \mathrm{e}^{\kappa y_M/2}X_0(y_M))^\top$ when X_0 follows the stationary distribution. Therefore, the claimed formula follows from

$$\sum_{\ell\in\mathcal{I}_m} \frac{\sigma^2}{2\lambda_\ell} = \mathrm{Var}\left(\langle X_0(\cdot), e_m\rangle_M\right) = \frac{1}{M^2}\mathrm{Var}\left(b_m^\top Z\right) = \frac{1}{M^2} b_m^\top \Sigma b_m$$

where the exponential factors cancel in the second step. □

In the following, we prove our convergence result for the replacement method:

Proof of Theorem 3.1.3. First, we treat the case of a stationary initial condition. It follows directly from the definition of the total variation distance that $\mathrm{TV}(f(X), f(Y)) \leq \mathrm{TV}(X, Y)$ holds for any random vectors X and Y and any measurable function f. Thus, the problem can be reduced to bounding the total variation distance of $(U_m(t_i),\, i \leq N, m \leq M-1)$ from its approximation. Furthermore, since both U_m and U_m^L are made up of independent summands, it is sufficient to consider the parts of the sums in which the two differ. To that aim, define $\mathcal{R}^L := (R_m^L(i),\, i \leq N, m \leq M-1)$ and $\mathcal{V}^L := (V_m^L(t_i),\, i \leq N, m \leq M-1)$ where $V_m^L(t) := \sum_{\ell\in\mathcal{I}_m,\, \ell\geq LM} u_\ell(t)$. Let Ξ_m be the covariance matrix of $(V_m^L(t_i),\, i \leq N)$ and $\Xi_m^{\perp\!\!\!\perp}$ be the covariance matrix of $(R_m^L(t_i),\, i \leq N)$ as well as $\Xi := \mathrm{diag}(\Xi_1, \dots, \Xi_{M-1})$, $\Xi^{\perp\!\!\!\perp} := \mathrm{diag}(\Xi_1^{\perp\!\!\!\perp}, \dots, \Xi_{M-1}^{\perp\!\!\!\perp})$. Since $\mathcal{V}^L$ and $\mathcal{R}^L$ are centered Gaussian random vectors with covariance matrices Ξ and $\Xi^{\perp\!\!\!\perp}$, respectively, we can use (3.4) together with the block structure of the matrices to bound

$$\mathrm{TV}(\mathcal{V}^L, \mathcal{R}^L)^2 \leq \frac{9}{4}\|(\Xi^{\perp\!\!\!\perp})^{-\frac{1}{2}}(\Xi - \Xi^{\perp\!\!\!\perp})(\Xi^{\perp\!\!\!\perp})^{-\frac{1}{2}}\|_F^2 = \frac{9}{4}\sum_{m=1}^{M-1}\|(\Xi_m^{\perp\!\!\!\perp})^{-\frac{1}{2}}(\Xi_m - \Xi_m^{\perp\!\!\!\perp})(\Xi_m^{\perp\!\!\!\perp})^{-\frac{1}{2}}\|_F^2. \tag{3.5}$$

We now treat each term in the sum separately. Note that $\Xi_m^{\perp\!\!\!\perp}$ is a diagonal matrix with the same diagonal elements as Ξ_m, namely s_m^2. Therefore, by the monotonicity of the exponential function,

$$\begin{aligned}\|(\Xi_m^{\perp\!\!\!\perp})^{-\frac{1}{2}}(\Xi_m - \Xi_m^{\perp\!\!\!\perp})(\Xi_m^{\perp\!\!\!\perp})^{-\frac{1}{2}}\|_F^2 &= \frac{1}{s_m^4}\sum_{i\neq j}\left(\sigma^2 \sum_{\ell\in\mathcal{I}_m,\, \ell\geq LM} \frac{\mathrm{e}^{-\lambda_\ell|i-j|\Delta}}{2\lambda_\ell}\right)^2 \\ &\leq \frac{1}{s_m^4}\sum_{i\neq j}\left(\sum_{\ell\in\mathcal{I}_m,\, \ell\geq LM}\frac{\sigma^2}{2\lambda_\ell}\right)^2 \mathrm{e}^{-2\lambda_{LM}|i-j|\Delta} \\ &= \sum_{i\neq j}\mathrm{e}^{-2\lambda_{LM}|i-j|\Delta}.\end{aligned}$$

Using $\sum_{i=1}^\infty q^i = \frac{q}{1-q}$ for $|q| < 1$, we can proceed to

$$\sum_{i\neq j}\mathrm{e}^{-2\lambda_{LM}|i-j|\Delta} \leq 2N\sum_{i=1}^{\infty}\mathrm{e}^{-2\lambda_{LM}i\Delta} = 2N\frac{\mathrm{e}^{-2\lambda_{LM}\Delta}}{1-\mathrm{e}^{-2\lambda_{LM}\Delta}} \lesssim N\mathrm{e}^{-2\lambda_{LM}\Delta}$$

where the last step follows from the fact that $L^2M^2\Delta \geq (LM\Delta^\alpha)^2 \to \infty$. Now, letting $c > 0$ be such that $c\ell^2 \leq \lambda_\ell$ for all $\ell \in \mathbb{N}$, we get the overall bound on the total variation distance claimed in (i), namely

$$\mathrm{TV}(\mathcal{X}, \mathcal{X}^L)^2 \leq \mathrm{TV}(\mathcal{V}^L, \mathcal{R}^L)^2 \lesssim MN\mathrm{e}^{-2\lambda_{LM}\Delta} \leq MN\mathrm{e}^{-2cL^2M^2\Delta}.$$

To prove (ii), choose $r>0$ such that $\frac{r+q+1}{2r-1}\le\alpha$. Then, using (i) and $\exp(-x)\lesssim x^{-r}, x>0$, for any $r>0$, we find

$$\mathrm{TV}(\mathcal{X},\mathcal{X}^L)^2\lesssim MN\mathrm{e}^{-2cL^2M^2\Delta}\lesssim\frac{MT}{(LM)^{2r}\Delta^{r+1}}=\frac{T\Delta^q}{L}\left(\frac{1}{LM\Delta^{\frac{r+q+1}{2r-1}}}\right)^{2r-1}\to 0,$$

finishing the proof for the stationary case.

Also for the case $X_0=0$, let Ξ_m be the covariance matrix of $(V_m^L(t_i),\, i\le N, m\le M-1)$ and $\Xi_m^{\perp\!\!\!\perp}$ be the covariance matrix of $(R_m^L(i),\, i\le N, m\le M-1)$ (without the initial deterministic value). Clearly, bound (3.5) remains valid and

$$\begin{aligned}
\|(\Xi_m^{\perp\!\!\!\perp})^{-\frac12}(\Xi_m-\Xi_m^{\perp\!\!\!\perp})(\Xi_m^{\perp\!\!\!\perp})^{-\frac12}\|_F^2 &= \frac{1}{s_m^4}\sum_{i,j=1}^N\left(\sigma^2\sum_{\ell\in\mathcal{I}_m,\,\ell\ge LM}\frac{\mathrm{e}^{-\lambda_\ell|i-j|\Delta}}{2\lambda_\ell}\left(1-\delta_{ij}-\mathrm{e}^{-2\lambda_\ell(i\wedge j)\Delta}\right)\right)^2\\
&=\frac{1}{s_m^4}\sum_{i\ne j}\left(\sigma^2\sum_{\ell\in\mathcal{I}_m,\,\ell\ge LM}\frac{\mathrm{e}^{-\lambda_\ell|i-j|\Delta}}{2\lambda_\ell}\left(1-\mathrm{e}^{-2\lambda_\ell(i\wedge j)\Delta}\right)\right)^2\\
&\quad+\frac{1}{s_m^4}\sum_{i=1}^N\left(\sigma^2\sum_{\ell\in\mathcal{I}_m,\,\ell\ge LM}\frac{1}{2\lambda_\ell}\mathrm{e}^{-2\lambda_\ell i\Delta}\right)^2\\
&\le\frac{1}{s_m^4}\sum_{i\ne j}\left(\sigma^2\sum_{\ell\in\mathcal{I}_m,\,\ell\ge LM}\frac{\mathrm{e}^{-\lambda_\ell|i-j|\Delta}}{2\lambda_\ell}\right)^2\\
&\quad+\frac{1}{s_m^4}\sum_{i=1}^N\left(\sigma^2\sum_{\ell\in\mathcal{I}_m,\,\ell\ge LM}\frac{1}{2\lambda_\ell}\mathrm{e}^{-2\lambda_\ell i\Delta}\right)^2\\
&\le\frac{2}{s_m^4}\sum_{i\ne j}\left(\sigma^2\sum_{\ell\in\mathcal{I}_m,\,\ell\ge LM}\frac{\mathrm{e}^{-\lambda_\ell|i-j|\Delta}}{2\lambda_\ell}\right)^2,
\end{aligned}$$

from which the result follows as in the stationary case. □

Chapter 4

Estimation in a semilinear framework

In this chapter we advance the theory on statistical estimation for SPDEs based on fully discrete observations to the semilinear framework of reaction-diffusion systems. Our aim is to estimate the diffusivity and volatility coefficients as well as the nonlinearity in the underlying equation. The latter is done in a fully nonparametric way. The results of this chapter are part of the preprint Hildebrandt and Trabs [37].

In Section 4.1 we state the different regularity assumptions on the solution process, required for the different problems treated in this section. These include a boundedness assumption (B), needed whenever working with large time asymptotics and a mixing assumption (M) needed for the nonparametric estimation of the nonlinearity. For the latter problem, we will also require an assumption (E) on the existence of densities for the random variables $X_t(y)$. Using analytic tools from semigroup theory, in Section 4.2 we discuss the Hölder regularity in time and space of the solution process and show higher order regularity of its nonlinear component. In Section 4.3 we revisit the parametric estimators for (σ^2, ϑ_2) from Chapter 2 in the semilinear framework. Based on the higher order Hölder regularity of the nonlinear component of the solution, we show that their asymptotic properties largely persist. Section 4.4 discusses nonparametric estimation of the nonlinearity in the underlying reaction-diffusion equation. We treat the estimation problem by adapting an approach used by Comte et al. [21] in the context of one-dimensional diffusion processes to our infinite dimensional framework. To that aim, we proceed in two steps: First, we consider observations that are discrete in time but continuous in space. By implementing an approximation step, the method is then adapted to fully discrete observations. We derive oracle inequalities for the expected risk when the risk is either the empirical 2-norm with evaluations at the data points or the usual L^2-norm on a compact set. All proofs are collected in Section 4.5. For most results the position of their proofs in the latter section is evident, otherwise it is indicated in the text.

As announced in Section 1.2.2, we consider a reaction-diffusion system on $[0, 1]$, namely the mild solution X to the equation

$$dX_t = (\vartheta_2 \frac{\partial^2}{\partial x^2} X_t + F(X_t))\, dt + \sigma dW_t, \quad X_0 = \xi, \tag{4.1}$$

with Dirichlet boundary conditions and where the nonlinearity F is given by

$$F(u) = f \circ u$$

for a function $f \in C^1(\mathbb{R})$. For simplicity, we will also refer to the functional by f, i.e., we write $f(u) = f \circ u$ for functions $u : [0, 1] \to \mathbb{R}$.

4.1 Further assumptions

Without further notice, in the sequel we will work under the standing assumption that there is a unique mild solution

$$X_t = S(t)\xi + \sigma \int_0^t S(t-s)\, dW_s + \int_0^t S(t-s) f(X_s)\, ds, \quad t \geq 0, \tag{4.2}$$

such that X is a Markov process with state space $E = C_0([0,1])$ and $X \in C(\mathbb{R}_+, E)$ holds almost surely for $\xi \in E$. Specifically, we will always assume that either $\xi = 0$ or that ξ follows the stationary distribution on E associated with the Markov process X, provided that it exists. Note that, in case of existence, the stationary distribution for the nonlinear equation is different from the stationary distribution for the linear equation considered in the previous chapters. We will denote the linear and nonlinear component of X by

$$X_t^0 := \sigma \int_0^t S(t-s)\, dW_s, \qquad N_t := \int_0^t S(t-s) f(X_s)\, ds, \quad t \geq 0.$$

Furthermore, it will be assumed throughout that f and its derivative are at most of polynomial growth, i.e., there exist constants $c > 0$ and $d \in \mathbb{N}$ such that

$$|f(x)|, |f'(x)| \leq c(1 + |x|^d), \quad x \in \mathbb{R}. \tag{4.3}$$

The set of basic assumptions just introduced is sufficient for generalizing the estimation methods from Section 2.2.2 to the semilinear framework as long as the time horizon T remains bounded. When dealing with the case $T \to \infty$, on the other hand, we need to impose a stricter assumption, that ensures that the error induced by the nonlinearity remains negligible uniformly in time, namely:

(B) The process X from (4.2) with zero or, in case of existence, stationary initial condition satisfies $\sup_{t\geq 0} \mathbf{E}(\|X_t\|_\infty^p) < \infty$ for any $p \geq 1$.

When dealing with nonparametric estimation of the nonlinearity f, our analysis will, in particular, rely on a concentration inequality derived via the mixing property of a stationary process. Hence, we will later assume:

(M) For the SPDE (4.1) there exists a stationary distribution π on E and the mild solution X from (4.2) with $\xi \sim \pi$ satisfies $\mathbf{E}(\|X_t\|_\infty^p) = \mathbf{E}(\|X_0\|_\infty^p) < \infty$ for any $p \geq 1$. Furthermore, X is exponentially β-mixing, i.e., there exist constants $L, \gamma > 0$ such that

$$\beta_X(t) := \int_E \|P_t(u,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}\, \pi(du) \leq L e^{-\gamma t} \tag{4.4}$$

where $(P_t)_{t\geq 0}$ is the transition semigroup on E associated with the Markov process X.

Sufficient conditions for Assumptions (B) and (M) to be satisfied are given in the following Proposition which is strongly based upon results derived in Goldys and Maslowski [34].

Proposition 4.1.1. *If there are constants $a, b, c, \beta \geq 0$ such that*

$$\operatorname{sgn}(x) f(x+y) \leq -a|x| + b|y|^\beta + c \tag{4.5}$$

holds for all $x, y \in \mathbb{R}$, then Assumptions (B) and (M) are satisfied.

A proof for the above proposition is given in Section 4.5.4. Condition (4.5) requires that $-f$ has at least linear growth at infinity and is, not surprisingly, stronger than the condition (1.8) from the general existence result. Still, it covers a large class of systems, including the case where f is a polynomial of odd degree with a negative leading coefficient.

Finally, for the nonparametric estimation of f on a compact set $A \subset \mathbb{R}$, we will need that the $L^2(A)$-norm is comparable to the empirical norm induced by the process X. This can be achieved by requiring the following equivalence condition.

(E) For the SPDE (4.1) there exists a stationary distribution π on E and, if $\xi \sim \pi$, the random variables $\xi(x)$ admit a Lebesgue density μ_x for each $x \in (0,1)$. Further, for any compact set $A \subset \mathbb{R}$ there are constants $c_0, c_1 > 0$ and $b \in (0, \frac{1}{2})$ such that

$$\mu_x(z) \leq c_1 \quad \text{for all} \quad z \in A,\, x \in (0,1),$$
$$\mu_x(z) \geq c_0 \quad \text{for all} \quad z \in A,\, x \in (b, 1-b).$$

The presence of the constant b in the lower bound is required due to the degeneracy induced by the Dirichlet boundary conditions. Assumption (E) will be necessary in order to conclude the existence of constants $c, C > 0$ such that

$$c\|t\|^2_{L^2(A)} \leq \mathbf{E}\Big(\int_0^1 t^2(X_0(y))\,dy\Big) \leq C\|t\|^2_{L^2(A)},$$
$$c\|t\|^2_{L^2(A)} \leq \mathbf{E}\Big(\frac{1}{M}\sum_{k=1}^{M-1} t^2(X_0(y_k))\Big) \leq C\|t\|^2_{L^2(A)}$$

holds for all functions $t \in L^2(\mathbb{R})$ with support in the compact set A. Assumption (E) is clearly satisfied in the case where f is a linear function, $f(x) = \vartheta_0 x$ for some $\vartheta_0 < 0$, see Section 1.2.1. Concerning a more general framework, there is a large amount of literature concerned with the existence and regularity of Lebesgue densities corresponding to the marginal distributions associated with various SPDE models, see, e.g., [4, 58, 63, 65]. However, to the author's best knowledge, there are so far no estimates on the densities of the random variables $X_t(x)$ that hold uniformly in $x \in \mathcal{X}$ for some infinite set $\mathcal{X} \subset (0,1)$. Deriving a sufficient condition on f to ensure (E) goes beyond the scope of this thesis and is postponed to further research.

4.2 Hölder regularity of the solution process

In this section, we discuss the Hölder regularity of the process $(X_t(y),\, t \geq 0, y \in [0,1])$ in time and space and, in particular, we show the higher order regularity of its nonlinear component $(N_t(y),\, t \geq 0, y \in [0,1])$. For $\alpha > 0$, we consider the Hölder spaces $C^\alpha := C^\alpha([0,1])$ consisting of all $u \in C^{[\alpha]}$ such that

$$\|u\|_{C^\alpha} := \sum_{k=0}^{[\alpha]} \|u^{(k)}\|_\infty + \sup_{x,y\in[0,1]} \frac{|u(x)-u(y)|}{|x-y|^{\alpha-[\alpha]}}$$

is finite where $[\alpha] := \max\{k \in \mathbb{N}_0 : k \leq \alpha\}$. Further, the Hölder continuous functions with Dirichlet boundary conditions are denoted by

$$C^\alpha_0 := \{u \in C^\alpha,\, u(0) = u(1) = 0\}.$$

Recall that the linear component $(X^0_t(x),\, x \in [0,1], t \geq 0)$ of X is a Gaussian process and

$$\begin{aligned}
\mathbf{E}((X^0_t(\xi) - X^0_t(\eta))^2) &= \sum_{\ell\geq 1} \frac{\sigma^2}{2\lambda_\ell}(1-\mathrm{e}^{-2\lambda_\ell t})(e_\ell(\xi) - e_\ell(\eta))^2 \\
&\leq \sum_{\ell\geq 1} \frac{\sigma^2}{2\lambda_\ell}(e_\ell(\xi) - e_\ell(\eta))^2 \approx |\xi - \eta|, && (4.6)\\
\mathbf{E}((X^0_t(x) - X^0_s(x))^2) &= \sum_{\ell\geq 1} \frac{\sigma^2}{\lambda_\ell}(1-\mathrm{e}^{-\lambda_\ell|t-s|})\left(1 - \frac{1-\mathrm{e}^{-\lambda_\ell|t-s|}}{2}\mathrm{e}^{-2\lambda_\ell s}\right) e^2_\ell(x) \\
&\leq \sum_{\ell\geq 1} \frac{\sigma^2}{\lambda_\ell}(1-\mathrm{e}^{-\lambda_\ell|t-s|}) \approx \sqrt{|t-s|}. && (4.7)
\end{aligned}$$

From this observation it follows that, almost surely, $X_t^0 \in E$ for any $t \geq 0$, $x \mapsto X_t^0(x)$ is 2γ-Hölder continuous and $t \mapsto X_t^0(x)$ is locally γ-Hölder continuous for any $\gamma < 1/4$. The following proposition generalizes this fact to the semilinear setting and shows that, under Assumption (B), the corresponding Hölder norms are $L^p(\mathbf{P})$-bounded as functions of time.

Proposition 4.2.1. *For any $p \in [1, \infty)$ the following hold.*

(i) For any $\gamma < 1/2$, we have $X \in C(\mathbb{R}_+, C_0^\gamma)$ a.s. and, if Assumption (B) is satisfied, then $\sup_{t\geq 0} \mathbf{E}(\|X_t\|_{C_0^\gamma}^p) < \infty$.

(ii) For any $\gamma < 1/4$ and $T > 0$, we have $(X_t)_{0\leq t\leq T} \in C^\gamma([0,T], E)$ a.s. and, if Assumption (B) is satisfied, then there exists a constant $C > 0$ such that $\mathbf{E}(\|X_t - X_s\|_\infty^p) \leq C|t-s|^{\gamma p}$ for all $s, t \geq 0$.

Furthermore, the same results hold for X replaced by $f_0(X)$ where $f_0(x) := f(x) - f(0)$.

We remark that a norm bound as in (i) with $p = 1$ is also derived in Cerrai [13, Proposition 4.2]. To prove the above proposition, we analyze the linear and the nonlinear component of X separately. The regularity of (X_t^0) can be assessed using properties (4.6) and (4.7) together with techniques from Da Prato and Zabczyk [26] for the study of continuity properties for linear equations, see Lemma 4.5.1. The regularity of (N_t) is a consequence of the regularizing property of the semigroup $(S(t))_{t\geq 0}$ in view of the fact that, due to our basic assumptions, the process $f(X)$ is continuous as a function of time and space. To deal with the situation where $f(0) \neq 0$ and, hence, $f(X_t) \notin E$, we need to consider the semigroup on the space $\tilde{E} = C([0,1])$. Note that the semigroup S is not strongly continuous on $\tilde{E}$. Indeed, we have $\lim_{t\to 0} S(t)x = x$ in $\tilde{E}$ if and only if $x \in E$. Nevertheless, $(S(t))_{t\geq 0}$ defines a so called *analytic* semigroup on $\tilde{E}$ which retains many properties of C_0-semigroups. For a precise definition of analytic semigroups we refer to, e.g., [55]. The following inequalities, which are particular cases of results derived in Sinestrari [72], are our main tool to study the regularity of (N_t). Recall the definition $A_\vartheta = \vartheta_2 \frac{\partial^2}{\partial x^2}$.

Lemma 4.2.2. *We fix an element $\lambda_0 \in (0, \lambda_1)$. For any $\alpha, \beta, \in (0,2) \setminus \{1\}$ and $n \in \mathbb{N}_0$ there exists a constant $C > 0$ such that*

(i) $\|A_\vartheta^n S(t)x\|_\infty \leq C\mathrm{e}^{-\lambda_0 t} t^{-n} \|x\|_\infty$ for all $x \in \tilde{E}$,

(ii) $\|S(t)x\|_{C_0^\alpha} \leq C\mathrm{e}^{-\lambda_0 t} t^{-\alpha/2} \|x\|_\infty$ for all $x \in \tilde{E}$,

(iii) $\|A_\vartheta S(t)x\|_\infty \leq Ct^{-(1-\alpha/2)} \|x\|_{C_0^\alpha}$ for all $x \in C_0^\alpha$,

(iv) $\|A_\vartheta^n S(t)x\|_{C_0^\beta} \leq C\mathrm{e}^{-\lambda_0 t} t^{-(n+\frac{\beta-\alpha}{2})} \|x\|_{C_0^\alpha}$ for all $x \in C_0^\alpha$ where it is required that either $n \geq 1$ or $\alpha \leq \beta$.

For a proof of (i), (ii) and (iv) we refer to [55, Proposition 2.3.1], (iii) follows from [72, Proposition 1.11]. Further, in order to transfer the spatial to the temporal regularity, of particular importance for our study are the so called intermediate spaces, defined by

$$D_{A_\vartheta}(\alpha, \infty) := \left\{ x \in \tilde{E},\ \|x\|_{D_{A_\vartheta}(\alpha,\infty)} := \|x\|_{\tilde{E}} + \sup_{t>0} \frac{\|S(t)x - x\|_{\tilde{E}}}{t^\alpha} < \infty \right\}, \qquad \alpha \in (0,1),$$

which are Banach spaces with the norm $\|\cdot\|_{D_{A_\vartheta}(\alpha,\infty)}$. These spaces can be defined for arbitrary analytic semigroups on a Banach space, see, e.g. [72]. For our concrete choice of A_ϑ and $\tilde{E}$, they are given by the Dirichlet-Hölder spaces

$$D_{A_\vartheta}(\alpha, \infty) = C_0^{2\alpha}([0,1]), \qquad \alpha \neq \frac{1}{2},$$

where the norms are equivalent, see Lunardi [54].

Having derived the Hölder regularity of the process X and, in particular, of $f_0(X)$, we can use Lemma 4.2.2 once more to show that the regularity of (N_t) exceeds the regularity of X. A related strategy has been pursued by Pasemann and Stannat [68] who studied the higher order regularity of the nonlinear component of X in the spaces $\mathcal{D}((-A_\vartheta)^\varepsilon)$, $\varepsilon > 0$. For our purpose, we can proceed similarly to Sinestrari [72] who studied the Hölder regularity of mild solutions to deterministic systems. We use the decomposition $N_t = N_t^0 + M_t$ where

$$N_t^0 := \int_0^t S(t-s) f_0(X_s)\, ds, \quad M_t := \int_0^t S(r) m\, dr \tag{4.8}$$

for $m \equiv f(0)$ and $f_0(x) = f(x) - f(0)$. Note that f_0 maps E and, in particular, $D_{A_\vartheta}(\alpha, \infty) = C_0^{2\alpha}$ into itself.

Proposition 4.2.3. *For any $T > 0$, $p \geq 1$ and $\gamma < 1/4$ the following hold.*

(i) For any $t \geq 0$, we have $N_t^0 \in C_0^{2+2\gamma}$ and $\sup_{t\leq T} \|A_\vartheta N_t^0\|_{C_0^{2\gamma}} < \infty$ almost surely. In particular, if Assumption (B) is satisfied, then $\sup_{t\geq 0} \mathbf{E}(\|A_\vartheta N_t^0\|_{C_0^{2\gamma}}^p) < \infty$.

(ii) We have $(N_t^0)_{t\leq T} \in C^{1+\gamma}([0,T], E)$ and $\frac{d}{dt} N_t^0 = f_0(X_t) + A_\vartheta N_t^0$ in E almost surely. In particular, under Assumption (B), there exists $C > 0$ such that $\mathbf{E}(\|\frac{d}{dt}(N_t^0 - N_s^0)\|_\infty^p) \leq C(t-s)^{\gamma p}$ holds for all $s, t \geq 0$.

Furthermore, the same results hold for (N_t) and f instead of (N_t^0) and f_0, provided that we replace E by $C([b, 1-b])$ and $C_0^{2\gamma}$ by $C^{2\gamma}([b, 1-b])$ for some $b \in (0, \frac{1}{2})$.

4.3 Diffusivity and volatility estimation

Using the results of the previous section, we are now able to carry the central limit theorems for space and double increments from Section 2.2.1 as well as the result (2.10) for time increments, derived in Bibinger and Trabs [9], over to the semilinear framework. As a consequence, the estimators considered in Section 2.2.2 for the volatility σ^2 and the diffusivity ϑ_2 can be used in the semilinear framework and, under quite general assumptions, their asymptotic properties remain unchanged. As before, X denotes the mild solution of equation (4.1) and we consider the observation scheme defined in Section 1.2.3. The constant b defining the minimal distance of spatial observations to the boundary of $[0, 1]$ is assumed to be strictly positive so that Proposition 4.2.3 provides the regularity of the process $(N_t(x), x \in [b, 1-b], t \geq 0)$ in space and time.

First, let us consider the realized quadratic variation based on time increments, i.e.,

$$\bar{V}_{\mathrm{t}} := \frac{1}{MN\sqrt{\Delta}} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} (X_{t_{i+1}}(y_k) - X_{t_i}(y_k))^2.$$

Recall that in the case $f \equiv 0$, $\bar{V}_{\mathrm{t}} = V_{\mathrm{t}}$ satisfies the central limit theorem (2.10) under the assumptions

$$M = o(\Delta^{-\rho}) \text{ for some } \rho < \frac{1}{2} \quad \text{and} \quad T\Delta \to 0. \tag{4.9}$$

Theorem 4.3.1. *Grant assumption* (4.9).

(i) If T is fixed and finite, the central limit theorem (2.10) *remains valid for $\bar{V}_{\mathrm{t}}$. If $T \to \infty$, it remains valid if Assumption (B) is satisfied and there exists $\rho < 1/2$ such that $TM = o(\Delta^{-\rho})$.*

(ii) Under the same assumptions as in (i), the central limit theorems for the time increments-based estimators for σ^2 or ϑ_2 from Section 2.2.2 remain valid.

While the result for a fixed time horizon carries over from the linear setting without any extra assumptions on the interplay between M, N and T, the additional assumption for the case $T \to \infty$ is much stricter. In the proof of the above theorem we show that $R_{\mathrm{t}} := \bar{V}_{\mathrm{t}} - V_{\mathrm{t}} = o_p(1/\sqrt{MN})$, which proves the result in view of Slutsky's Lemma. In fact, it follows from the temporal regularity properties of the processes (X_t^0) and (N_t), that R_{t} is of the order $\mathcal{O}_p(\Delta^\alpha)$ for any $\alpha < 3/4$. Hence, the reason for the additional assumption in the case $T \to \infty$ is that $\sqrt{MN}\Delta^\alpha = \sqrt{MT\Delta^{2\alpha-1}}$ is required to tend to 0.

Next, we consider the realized quadratic variation based on space increments, i.e.,

$$\bar{V}_{\mathrm{sp}} := \frac{1}{MN\delta} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} (X_{t_i}(y_{k+1}) - X_{t_i}(y_k))^2.$$

Like in the linear setting, we sum over $i \in \{1, \dots, N\}$ instead of $\{0, \dots, N-1\}$ if $X_0 = 0$. The central limit theorem for V_{sp} in the case $f \equiv 0$, Theorem 2.2.3, holds under the condition

$$N = o(M). \tag{4.10}$$

Theorem 4.3.2. *Grant assumption* (4.10).

(i) If T is fixed and finite, the conclusion of Theorem 2.2.3 remains valid for $\bar{V}_{\mathrm{sp}}$. If $T \to \infty$, it remains valid under Assumption (B).

(ii) Under the same assumptions as in (i), the central limit theorems for the space increments-based estimators for σ^2 or ϑ_2 from Section 2.2.2 remain valid.

Although our proof strategy for the above theorem is the same as for time increments, here, the result carries over from the linear setting with no extra conditions on M, N and T, at all. Indeed, by using a summation by parts formula to rewrite $R_{\mathrm{sp}} := \bar{V}_{\mathrm{sp}} - V_{\mathrm{sp}}$, we can profit from the fact that the second order spatial increments of (N_t), namely $N_{t_i}(y_{k+1}) - 2N_{t_i}(y_k) + N_{t_i}(y_{k-1})$, are of the order $\mathcal{O}_p(\delta^2)$, thanks to the spatial regularity of the process (N_t).

Finally, we consider the realized quadratic variation based on double increments, i.e.,

$$\bar{\mathbb{V}} := \frac{1}{MN\Phi_\vartheta(\delta, \Delta)} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \bar{D}_{ik}^2$$

with $\bar{D}_{ik} := X_{t_{i+1}}(y_{k+1}) - X_{t_i}(y_{k+1}) - X_{t_{i+1}}(y_k) + X_{t_i}(y_k)$. As in the linear case, if a balanced sampling design is present, i.e. $\delta/\sqrt{\Delta} \equiv r$ for some $r > 0$, we can also consider

$$\bar{\mathbb{V}}_r := \frac{1}{MN\sqrt{\Delta}} \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} \bar{D}_{ik}^2.$$

In the case $X_0 = 0$, $\bar{\mathbb{V}}$ and $\bar{\mathbb{V}}_r$ are redefined in the obvious way. If $f \equiv 0$, the conditions for the central limit theorem for $\bar{\mathbb{V}} = \mathbb{V}$, Theorem 2.2.7, are

$$\delta/\sqrt{\Delta} \to r \in \{0, \infty\} \quad \text{or} \quad \delta/\sqrt{\Delta} \equiv r > 0 \tag{4.11}$$

as well as

$$\Delta \to 0 \quad \text{and} \quad T = o(M). \tag{4.12}$$

Theorem 4.3.3. *Grant assumptions* (4.11) *and* (4.12).

(i) If T is fixed and finite, the conclusions of Theorem 2.2.7 and Corollary 2.2.9 remain valid for $\bar{\mathbb{V}}$ and $\bar{\mathbb{V}}_r$, respectively. If $T \to \infty$, they remain valid if Assumption (B) is satisfied and there exists $a \in (0, 1)$ such that $T = o(M^a)$.

(ii) Under the same assumptions as in (i), the central limit theorems for the double increments-based estimators for σ^2, ϑ_2 or (σ^2, ϑ_2) from Section 2.2.2 remain valid.

Remark 4.3.4. In particular, in view of Remark 2.2.13, the conclusions of Corollaries 2.2.14 and 2.2.19 remain valid. Hence, the estimator (2.19) for (σ^2, ϑ_2) defines a rate optimal estimator in our semilinear framework.

As for space increments, there are essentially no additional assumptions compared to the linear setting. The influence induced by the nonlinearity is negligible, in particular, since the double increments computed from the process (N_t) decay in both Δ and δ at the same time, as opposed to the double increments computed from (X_t^0) which are roughly of the order $(\delta \wedge \sqrt{\Delta})^{1/2}$, see also Lemma 4.5.2.

4.4 Nonparametric estimation of the nonlinearity

The following section discusses nonparametric estimation of f. We adapt an estimation procedure considered by Comte et al. [21] in the context of one-dimensional diffusions to our SPDE setting. First, we treat observations that are discrete in time but continuous in space and, then, by implementing an approximation step, an estimation procedure for fully discrete observations will be introduced. Note that the parameters (σ^2, ϑ_2) can be estimated well using the methods analyzed in the previous section. Thus, in a first step, we will assume that these parameters (in fact, only ϑ_2 is necessary) are known. Later, a plug-in approach will be considered.

In contrast to the previous section, we will strictly require that the mild solution X from (4.2) admits a stationary distribution, denoted by π, and, moreover, that the mixing assumption (M) is satisfied. Furthermore, it will be essential for the derivation of our oracle inequalities that we have $T \to \infty$. From now on, let $A \subset \mathbb{R}$ be a fixed compact set on which we want to estimate f.

Before treating the actual estimation problem, we introduce the approximation spaces serving as candidate functions for the estimation of f in the following section.

4.4.1 Spaces of approximation

In order to estimate f on the set A, we consider a sequence $(S_m)_{m\in\mathbb{N}}$ of finite dimensional sub-spaces of $L^2(A)$ such that $D_m := \dim(S_m) \to \infty$ for $m \to \infty$. The intuition is that we choose m depending on the sample size and, if accessible, depending on the regularity of f and then estimate f by taking the function $\hat{f}_m \in S_m$ that matches the data in the best possible way with respect to a suitable criterion to be defined later. Like in [21], our key assumption on the approximation spaces S_m is the following.

(N) There is a constant $C > 0$ such that for each $m \in \mathbb{N}$ there is an orthonormal basis $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ of S_m, $|\Lambda_m| = D_m$, with

$$\Big\| \sum_{\lambda\in\Lambda_m} \varphi_\lambda^2 \Big\|_\infty \leq C D_m.$$

It is shown in Birgé and Massart [10] that Assumption (N) is equivalent to requiring $\|t\|_\infty^2 \leq C D_m \|t\|_{L^2(A)}^2$ for all $t \in S_m$ and $m \in \mathbb{N}$.

Let us briefly recall some examples of approximation spaces with property (N) that are considered in [21]. We assume that A is a closed interval and take $A = [-a, a]$ for some $a > 0$ without loss of generality.

Example 4.4.1.

[T] The trigonometric spaces

$$S_m = \operatorname{span}\Big(\Big\{\frac{1}{\sqrt{2a}},\ \frac{1}{\sqrt{a}}\sin\big(\frac{k\pi}{a}\cdot\big),\ \frac{1}{\sqrt{a}}\cos\big(\frac{k\pi}{a}\cdot\big),\ 1\le k\le m\Big\}\Big)$$

have dimension $D_m = 2m+1$ and property (N) follows directly from the fact that trigonometric base functions are uniformly bounded.

[P] Piecewise polynomials on a dyadic grid: Most conveniently, these spaces are parameterized in terms of a pair $m = (p, r)$ with $p \in \mathbb{N}$, $r \in \{0, \dots, r_{\max}\}$ and $r_{\max} \in \mathbb{N}_0$ being some fixed value. Let $(p_l)_{l\in\mathbb{N}_0}$ be the complete orthonormal system in $L^2([0,1])$ such that p_l is the rescaled Legendre polynomial of degree l for $l \in \mathbb{N}_0$. Further, for $p \in \mathbb{N}$ and $j \in \{-2^p, \dots, 2^p - 1\}$ let $I_j^p := [ja2^{-p}, (j+1)a2^{-p})$. Then, for $m = (p, r)$, we define

$$S_{(p,r)} := \operatorname{span}\Big(\{\varphi_{j,l}^p,\ l \le r, -2^p \le j \le 2^p - 1\}\Big)$$

with

$$\varphi_{j,l}^p(x) := \sqrt{\frac{2^p}{a}}\, p_l\Big(\frac{2^p x}{a} - j\Big)\mathbf{1}_{I_j^p}(x), \qquad x \in A.$$

Clearly, $\dim(S_{p,r}) = (r+1)2^{p+1} \le (r_{max}+1)2^{p+1}$ and property (N) holds with a constant C depending on $r_{\max}$.

[W] The dyadic wavelet generated spaces: For arbitrary $r \in \mathbb{N}$, there are a scaling and a wavelet function $\phi, \psi \in C^\alpha(\mathbb{R})$, respectively, for some $\alpha > 0$ with support in $[0,1]$ such that ψ has r vanishing moments and

$$\Big\{\frac{1}{\sqrt{a}}\phi\Big(\frac{\cdot}{a}\Big),\ \frac{1}{\sqrt{a}}\phi\Big(\frac{\cdot}{a}+1\Big), \sqrt{\frac{2^p}{a}}\psi\Big(\frac{2^p\cdot}{a} - j\Big),\ -2^p \le j < 2^p,\ p \in \mathbb{N}\Big\}$$

is a complete orthonormal system in $L^2(A)$, see [28]. Then, the subspace

$$S_m = \operatorname{span}\Big(\Big\{\frac{1}{\sqrt{a}}\phi\Big(\frac{\cdot}{a}\Big),\ \frac{1}{\sqrt{a}}\phi\Big(\frac{\cdot}{a}+1\Big), \sqrt{\frac{2^p}{a}}\psi\Big(\frac{2^p\cdot}{a} - j\Big),\ -2^p \le j < 2^p,\ p \le m\Big\}\Big)$$

satisfies $\dim(S_m) = 2^{m+2}$ and property (N) is fulfilled.

The following definition due to Baraud et al. [5] proves to be useful in analyzing our nonparametric estimator. We fix an orthonormal basis $(\varphi_\lambda,\ \lambda \in \Lambda_m)$ of S_m according to Assumption (N) and define the matrices $V^m, B^m \in \mathbb{R}^{\Lambda_m\times\Lambda_m}$ by

$$V_{\lambda,\lambda'}^m := \|\varphi_\lambda\varphi_{\lambda'}\|_{L^2(A)}, \qquad B_{\lambda,\lambda'}^m := \|\varphi_\lambda\varphi_{\lambda'}\|_\infty.$$

These expressions are convenient in order to express certain estimates, e.g., $|\varphi_\lambda(Z)\varphi_{\lambda'}(Z)| \le B_{\lambda,\lambda'}^m$ and $\mathbf{E}(|\varphi_\lambda(Z)\varphi_{\lambda'}(Z)|^2) \lesssim (V_{\lambda,\lambda'}^m)^2$ when Z is an A-valued random variable with a bounded Lebesgue density. Further, let

$$L_m := \max(\rho^2(V^m), \rho(B^m)), \qquad \rho(H) := \sup_{a\in\mathbb{R}^{\Lambda_m},\, \sum_\lambda a_\lambda^2\le 1} \sum_{\lambda,\lambda'} |a_\lambda a_{\lambda'} H_{\lambda,\lambda'}|,\ H \in \{V^m, B^m\}. \quad (4.13)$$

For our main oracle inequalities, we will have to require that L_m is asymptotically negligible with respect to the time horizon T. For the previous examples of approximation spaces, the quantity L_m can be linked directly to the dimension D_m. In fact, it is shown in [5] that $L_m \lesssim D_m^2$ for [T] and $L_m \lesssim D_m$ for [P] and [W].

4.4.2 Estimation based on space-continuous observations

In this section, we consider observations that are discrete in time and continuous in space, i.e., the data is given by

$$\{X_{t_i}(x),\, x \in [0,1],\, i = 0, \dots, N\}.$$

From (4.2) it is evident that we can decompose

$$X_{t+\Delta} = S(\Delta)X_t + \sigma \int_t^{t+\Delta} S(t+\Delta-s)\, dW_s + \int_t^{t+\Delta} S(t+\Delta-s) f(X_s)\, ds.$$

By rearranging, we can pass to

$$\begin{aligned}\frac{X_{t+\Delta} - S(\Delta)X_t}{\Delta} &= f(X_t) + \frac{\sigma}{\Delta}\int_t^{t+\Delta} S(t+\Delta-s)\, dW_s \\ &\quad + \frac{1}{\Delta}\int_t^{t+\Delta} \Big(S(t+\Delta-s)f(X_s) - f(X_t)\Big)\, ds,\end{aligned}$$

yielding the regression model

$$Y_i = f(X_{t_i}) + R_i + \varepsilon_i, \qquad 0 \le i \le N-1,$$

with

$$\begin{aligned}Y_i &:= \frac{X_{t_{i+1}} - S(\Delta)X_{t_i}}{\Delta}, \\ \varepsilon_i &:= \frac{\sigma}{\Delta}\int_{t_i}^{t_{i+1}} S(t_{i+1}-s)\, dW_s, \\ R_i &:= \frac{1}{\Delta}\int_{t_i}^{t_{i+1}} \Big(S(t_{i+1}-s)f(X_s) - f(X_{t_i})\Big)\, ds.\end{aligned}$$

The main term in the regression model is given by $f(X_{t_i})$, ε_i is the stochastic noise term and R_i is a negligible bias. Note that the stochastic noise term is stochastically independent of the covariate X_{t_i}. The corresponding least squares estimator is defined by

$$\hat{f}_m := \arg\min_{g \in S_m} \gamma_N(g), \qquad \gamma_N(g) := \frac{1}{N}\sum_{i=0}^{N-1} \|Y_i - g(X_{t_i})\|_{L^2}^2$$

with $L^2 := L^2((0,1))$. Note that this estimator hinges on the parameter ϑ_2 through the semigroup $S(\cdot)$ appearing in the response variables Y_i and, for now, we assume that it is known. The natural empirical norm associated with the observations scheme is given by

$$\|g\|_N^2 := \frac{1}{N}\sum_{i=0}^{N-1} \|g(X_{t_i})\|_{L^2}^2$$

and, in the sequel, we derive a bound on $\mathbf{E}(\|\hat{f}_m - f_A\|_N^2)$ with $f_A := f\mathbf{1}_A$. As before, π denotes the stationary distribution for X and, for nonrandom $g \in L^2(A)$, let

$$\|g\|_\pi^2 := \mathbf{E}(\|g\|_N^2) = \mathbf{E}(\|g(X_0)\|_{L^2}^2).$$

Recall that, under Assumption (E), there are constants $c, C > 0$ such that

$$c\|g\|_{L^2(A)}^2 \le \|g\|_\pi^2 \le C\|g\|_{L^2(A)}^2 \tag{4.14}$$

holds for all $g \in L^2(A)$. The oracle choice for an estimator of f_A from the space S_m is given by

$$f_m := \arg\min_{g \in S_m} \|f - g\|_{L^2(A)}^2.$$

Our main result for space-continuous observations is the following.

Theorem 4.4.2. *Grant Assumptions (M), (E) and (N). Further assume that* $\frac{N\Delta}{\log^2 N} \to \infty$ *as well as* $L_m = o(\frac{N\Delta}{\log^2 N})$ *and* $D_m \leq N$*. Then, for any* $\gamma < 1/2$,

$$\mathbf{E}\big(\|\hat{f}_m - f_A\|_N^2\big) \lesssim \|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma.$$

Remark 4.4.3. In contrast to the term Δ^γ, $\gamma < 1/2$, on the right hand side of the above inequality, Comte et al. [21] obtain the smaller bound Δ in their corresponding result in the context of SODEs. In fact, this error term arises from bounding the bias terms R_i in the underlying regression model. The difference in the order of magnitude is due to the fact that for the SPDE model there only is temporal Hölder regularity up to exponent 1/4, as opposed to exponent 1/2 in the finite dimensional setting.

We encounter the usual variance-bias trade-off in nonparametric estimation: When we choose m too small the estimator is not sufficiently versatile, leading to a large bias term $\|f_A - f_m\|_{L^2(A)}^2$. On the other hand, when choosing m too large, the estimated function will closely follow the concrete realization of the data, leading to a large variance D_m/T. Now, assuming that $\|f_m - f_A\|_{L^2(A)} \approx D_m^{-\alpha}$, balancing the bias and the variance term shows that it is optimal to choose $D_m \approx T^{\frac{1}{1+2\alpha}}$. Under the additional assumption that $T\Delta^\gamma \to 0$ holds for some $\gamma < 1/2$, the term Δ^γ appearing on the right hand side of the oracle inequality can be regarded as a negligible remainder and we obtain the usual (squared) nonparametric rate

$$\mathbf{E}\big(\|\hat{f}_m - f_A\|_N^2\big) \lesssim T^{-\frac{2\alpha}{2\alpha+1}}.$$

In fact, there might be additional conditions due to the assumption $L_m = o\big(\frac{T}{\log^2 N}\big)$ in the theorem. For instance, for the trigonometric spaces [T], we have $L_m \approx D_m^2$ and, thus, it is only possible to take $D_m \approx T^{\frac{1}{1+2\alpha}}$, provided that $\alpha > \frac{1}{2}$, as already pointed out in [21]. For instance, when using [T], the k-th Fourier coefficients of a function $f \in C^1$ are generally of the order $1/k$. A faster decay is only expectable in the exceptional case where the function f is periodic on A. Thus, we have $\|f_m - f_A\|_{L^2(A)} \approx D_m^{-\alpha}$ with $\alpha = \frac{1}{2}$ and it is possible to achieve a convergence rate of $T^{-\frac{\tilde{\alpha}}{2\tilde{\alpha}+1}}$ for any $\tilde{\alpha} < \alpha$. We remark that, in general, the true value of the regularity parameter α is unknown, as it is a property of the unknown function f_A. Building upon the results presented in this thesis, this issue is addressed in Hildebrandt and Trabs [37] where adaptivity of the estimator on the regularity of f is achieved via model selection. See also the outlook in Chapter 5 of this thesis.

In order to prove Theorem 4.4.2, we adapt the proof strategy from [21] to our infinite dimensional setting. The main steps of the proof are explained in the following.

For an arbitrary function g, we can write

$$\begin{aligned}\gamma_N(g) - \gamma_N(f) &= \|g - f\|_N^2 + \frac{2}{N}\sum_{i=0}^{N-1}\langle Y_i - f(X_{t_i}), f(X_{t_i}) - g(X_{t_i})\rangle_{L^2} \\ &= \|g - f\|_N^2 + \frac{2}{N}\sum_{i=0}^{N-1}\langle \varepsilon_i + R_i, f(X_{t_i}) - g(X_{t_i})\rangle_{L^2}.\end{aligned}$$

By definition of $\hat{f}_m$, we have $\gamma_N(\hat{f}_m) - \gamma_N(f) \leq \gamma_N(f_m) - \gamma_N(f)$ and using the above expansion on both sides of this inequality yields

$$\|\hat{f}_m - f\|_N^2 \leq \|f_m - f\|_N^2 + \frac{2}{N}\sum_{i=0}^{N-1}\langle \varepsilon_i + R_i, \hat{f}_m(X_{t_i}) - f_m(X_{t_i})\rangle_{L^2}.$$

Since both $\hat{f}_m$ and f_m are A-supported, if we insert $f = f\mathbf{1}_A + f\mathbf{1}_{A^c}$ in the above equation, then the terms $\|f1_{A^c}\|_N^2$ on both sides of the inequality cancel. We arrive at the fundamental oracle inequality

$$\|\hat{f}_m - f_A\|_N^2 \leq \|f_m - f_A\|_N^2 + \frac{2}{N}\sum_{i=0}^{N-1}\langle \varepsilon_i, \hat{f}_m(X_{t_i}) - f_m(X_{t_i})\rangle_{L^2}$$

$$+\frac{2}{N}\sum_{i=0}^{N-1}\langle R_i, \hat{f}_m(X_{t_i}) - f_m(X_{t_i})\rangle_{L^2}. \tag{4.15}$$

By treating each of the three terms appearing on the right hand side above individually, we can derive the following proposition.

Proposition 4.4.4. *Grant Assumptions (M) and (N) and for $\underline{c} > 0$ let*

$$\Omega_{N,m} := \Omega_{N,m,\underline{c}} := \Big\{\|t\|_N^2 \geq \underline{c}\|t\|_{L^2(A)}^2 \text{ for all } t \in S_m\Big\}.$$

Then, for any $\gamma < 1/2$, we have

$$\mathbf{E}\Big(\|\hat{f}_m - f_A\|_N^2 \mathbf{1}_{\Omega_{N,m}}\Big) \lesssim \|f_A - f_m\|_\pi^2 + \frac{D_m}{T} + \Delta^\gamma.$$

The event $\Omega_{N,m}$ has been introduced since the proof requires bounding the stochastic noise term $\frac{1}{N}\sum_{i=0}^{N-1}\langle\varepsilon_i, t(X_{t_i})\rangle_{L^2}$ uniformly over all $\|\cdot\|_N$-normalized $t \in S_m$. The latter is difficult since both the object to be bounded and the norm are random objects. On the event $\Omega_{N,m}$, it is sufficient to bound it uniformly over all $\|\cdot\|_{L^2(A)}$-normalized $t \in S_m$ which is possible thanks to Assumption (N).

Under Assumption (E), we can further bound $\|f_A - f_m\|_\pi^2 \lesssim \|f_A - f_m\|_{L^2(A)}^2$, hence, Proposition 4.4.4 already provides the relevant terms appearing in the oracle inequality from Theorem 4.4.2. The second main step of the proof of the theorem is to verify that the event $\Omega_{N,m}^c$ has negligible probability. To that aim, let us consider

$$\Xi_{N,m} := \Big\{\Big|\frac{\|t\|_N^2}{\|t\|_\pi^2} - 1\Big| \leq \frac{1}{2}\,\forall t \in S_m\Big\}$$

which satisfies $\Xi_{N,m} \subset \Omega_{N,m,\frac{1}{2}}$. Since $\mathbf{E}(\|t\|_N^2) = \|t\|_\pi^2$ and

$$\Big|\frac{\|t\|_N^2}{\|t\|_\pi^2} - 1\Big| \approx \Big|\frac{\|t\|_N^2 - \|t\|_\pi^2}{\|t\|_{L^2(A)}^2}\Big|$$

under Assumption (E), bounding the probability of $\Xi_{N,m}^c$ is equivalent to deriving a concentration inequality for $\|t\|_N^2$ uniformly over all $L^2(A)$-normalized $t \in S_m$. This can be done using the standard techniques for β-mixing sequences, see, e.g., [31]: by means of the mixing assumption (4.4) in (M), we approximate $(X_{t_0}, \ldots, X_{t_N})$ by a process with independent blocks and, then, apply a variant of Bernstein's inequality. We obtain the following bound for $\mathbf{P}(\Xi_{N,m}^c)$.

Lemma 4.4.5. *Grant Assumptions (M), (E) and (N). Then, there are constants $K, K' > 0$ such that*

$$\mathbf{P}(\Xi_{N,m}^c) \leq K\Big(N\beta_X(q_N\Delta) + D_m^2 \exp\Big(-K'\frac{p_N}{L_m}\Big)\Big)$$

holds for any $p_N, q_N \in \mathbb{N}$ with $N = 2p_Nq_N$. In particular, with the constants γ and L from the β-mixing condition (4.4) as well as $\tilde{K} := K\max(L, 1)$, we have

$$\mathbf{P}(\Xi_{N,m}^c) \leq \tilde{K}\Big(N\exp\big(-\gamma q_N\Delta\big) + D_m^2\exp\Big(-K'\frac{p_N}{L_m}\Big)\Big).$$

The conclusion of the main theorem is a straightforward consequence of Proposition 4.4.4 and Lemma 4.4.5, we refer to Section 4.5.3 for further details on the proof.

Next, we assess the quality of our estimator in terms of the more intuitive distance measure $\|\hat{f}_m - f\|_{L^2(A)}$, rather then $\|\hat{f}_m - f_A\|_N$. Using the triangle inequality as well as the equivalence of the empirical and the $L^2(A)$-norm on $\Xi_{N,m}$, we can bound

$$\|\hat{f}_m - f\|_{L^2(A)}^2 \leq 2\|\hat{f}_m - f_m\|_{L^2(A)}^2 + 2\|f_m - f\|_{L^2(A)}^2$$

$$= 2\|\hat f_m - f_m\|^2_{L^2(A)}\mathbf{1}_{\Xi_{N,m}} + 2\|\hat f_m - f_m\|^2_{L^2(A)}\mathbf{1}_{\Xi^c_{N,m}} + 2\|f_m - f\|^2_{L^2(A)}$$
$$\approx 2\|\hat f_m - f_m\|^2_{N}\mathbf{1}_{\Xi_{N,m}} + 2\|\hat f_m - f_m\|^2_{L^2(A)}\mathbf{1}_{\Xi^c_{N,m}} + 2\|f_m - f\|^2_{L^2(A)}.$$

Then, thanks to Proposition 4.4.4 and Lemma 4.4.5, it is straightforward, to derive an upper bound in probability. Bounding $\mathbf{E}(\|\hat f_m - f\|^2_{L^2(A)})$, on the other hand, is more challenging since the behavior of $\|\hat f_m - f\|^2_{L^2(A)}$ on the set $\Xi^c_{N,m}$ is a priori unclear. This issue can be circumvented by considering the truncated version

$$\hat f_m^{K_N} := (-K_N) \vee (\hat f_m \wedge K_N)$$

where (K_N) is a sequence of positive numbers with $K_N \to \infty$ such that $K_N \mathbf{P}(\Xi^c_{N,m}) \to 0$ sufficiently fast.

Corollary 4.4.6. *Grant Assumptions (M), (E) and (N). Further assume that $\frac{N\Delta}{\log^2 N} \to \infty$ as well as $L_m = o(\frac{N\Delta}{\log^2 N})$ and $D_m \le N$. Then, for any $\gamma < 1/2$,*

$$\|\hat f_m - f\|^2_{L^2(A)} = \mathcal{O}_p\Big(\|f - f_m\|^2_{L^2(A)} + \frac{D_m}{T} + \Delta^\gamma\Big).$$

Furthermore, for a sequence (K_N) with $K_N \to \infty$ and $K_N/N^\beta \to 0$ for some $\beta > 0$, we have

$$\mathbf{E}(\|\hat f_m^{K_N} - f\|^2_{L^2(A)}) \lesssim \|f - f_m\|^2_{L^2(A)} + \frac{D_m}{T} + \Delta^\gamma.$$

4.4.3 Estimation based on fully discrete observations

We return to our fully discrete observation scheme described in Section 1.2.3. In order to derive a discretized version of the estimator discussed in the previous section, we assume that the temporal and spatial observations are recorded at the locations

$$t_i = i\Delta \quad \text{and} \quad y_k = \frac{k}{M}$$

for $0 \le i \le N$ and $0 \le k \le M$, i.e., the parameter b specifying the margin of the spatial observation window is set to $b = 0$. With observations distributed throughout the whole space domain $(0,1)$, we have the possibility to approximate the coefficient processes $x_k(t) := \langle X_t, e_k\rangle$ by their empirical counterpart $\langle X_t, e_k\rangle_M := \frac{1}{M}\sum_{l=1}^{M-1} X_t(y_l)e_k(y_l)$. Recall from (2.4) that for $k \le M-1$ we have the relation

$$\langle X_t, e_k\rangle_M = \sum_{\ell\in\mathcal{I}_k^+} x_\ell(t) - \sum_{\ell\in\mathcal{I}_k^-} x_\ell(t)$$

where $\mathcal{I}_k^+ = k + 2M\cdot\mathbb{N}_0$ and $\mathcal{I}_k^- = 2M - k + 2M\cdot\mathbb{N}_0$. In order to approximate the expression $S(\Delta)X_{t_i}$ appearing in the estimator based on space-continuous observations, we define $\hat S(\Delta) := \hat S_M(\Delta)$ by

$$\hat S(\Delta)X_t = \sum_{\ell=1}^{M-1} \mathrm{e}^{-\lambda_\ell\Delta}\langle X_t, e_\ell\rangle_M e_\ell$$

which only hinges on X_t through the discrete data $(X_t(y_k),\, k = 1,\dots,M-1)$. Now, from the space-continuous regression model

$$\frac{X_{t_{i+1}} - S(\Delta)X_{t_i}}{\Delta} = f(X_{t_i}) + R_i + \varepsilon_i$$

we pass to

$$\frac{\hat S(0)X_{t_{i+1}} - \hat S(\Delta)X_{t_i}}{\Delta} = \hat S(0)f(X_{t_i}) + \tilde R_i + \varepsilon_i \tag{4.16}$$

with

$$\begin{aligned}\tilde{R}_i :=& f(X_{t_i}) - \hat{S}(0)f(X_{t_i}) + \frac{S(\Delta)X_{t_i} - \hat{S}(\Delta)X_{t_i}}{\Delta} + \frac{\hat{S}(0)X_{t_{i+1}} - X_{t_{i+1}}}{\Delta} \\ &+ \frac{1}{\Delta}\int_{t_i}^{t_{i+1}} \Big(S(t_{i+1}-s)f(X_s) - f(X_{t_i})\Big)\, ds, \\ \varepsilon_i =& \frac{\sigma}{\Delta}\int_{t_i}^{t_{i+1}} S(t_{i+1}-s)\, dW_s.\end{aligned}$$

We consider the corresponding least squares estimator

$$\begin{aligned}\hat{f}_m :=& \arg\min_{g\in S_m} \frac{1}{N}\sum_{i=0}^{N-1} \Big\| \frac{\hat{S}(0)X_{t_{i+1}} - \hat{S}(\Delta)X_{t_i}}{\Delta} - \hat{S}(0)g(X_{t_i})\Big\|_{L^2}^2 \\ =& \arg\min_{g\in S_m} \frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=1}^{M-1}\Big(\frac{\langle X_{t_{i+1}}, e_k\rangle_M - \mathrm{e}^{-\lambda_k\Delta}\langle X_{t_i}, e_k\rangle_M}{\Delta} - \langle g(X_{t_i}), e_k\rangle_M\Big)^2\end{aligned}$$

which is purely based on the space-time-discrete observations.

In the following, we will require a minimal continuity property for the approximation spaces S_m and impose the following assumption.

(H) For any $g \in \bigcup_{m\in\mathbb{N}} S_m$, let $\bar{g}\colon \mathbb{R}\to\mathbb{R}$ be the extension of g by zero on the set A^c. Then, the function $\bar{g}$ is piecewise Hölder continuous, i.e., there are constants $\alpha > 0$ and $-\infty = a_0 < a_1 < \ldots < a_L = \infty$, $L \in \mathbb{N}$, such that $\bar{g}|_{(a_l,a_{l+1})} \in C^\alpha((a_l, a_{l+1}))$ for any $1 \le l \le L-1$.

Note that all approximation spaces from Example 4.4.1 also meet the additional requirement (H).

Under Assumption (H), it is possible to derive a convenient and intuitive representation for our estimator $\hat{f}_m$ based on the following lemma.

Lemma 4.4.7. *Let $H\colon [0,1]\to\mathbb{R}$ be Hölder continuous in an environment of y_k for each $1 \le k \le M-1$ and set $h_k := \langle H, e_k\rangle_{L^2}$. Then, the series $H_k := \sum_{l\in\mathcal{I}_k^+} h_l - \sum_{l\in\mathcal{I}_k^-} h_l$ converges and we have $\langle H, e_k\rangle_M = H_k$ as well as*

$$\frac{1}{M}\sum_{k=1}^{M-1} H^2(y_k) = \|H^M\|_{L^2}^2 = \sum_{l=1}^{M-1} H_l^2$$

with $H^M := \hat{S}(0)H = \sum_{l=1}^{M-1} H_l e_l$.

Remark 4.4.8. The Hölder condition in the above lemma can be relaxed to requiring convergence of the Fourier series of H at y_k to $H(y_k)$ for each $1 \le k \le M-1$.

Under Assumptions (H) and (E), the random variables $X_{t_i}(y_k)$ hit a discontinuity of the extension $\bar{g}$ of some $g \in \bigcup_{m\in\mathbb{N}} S_m$ with probability zero and, hence, the above lemma is applicable with $H := \frac{X_{t_{i+1}} - \hat{S}(\Delta)X_{t_i}}{\Delta} - g(X_{t_i})$. In particular, the estimator $\hat{f}_m$ can, almost surely, be expressed via

$$\hat{f}_m = \arg\min_{g\in S_m} \gamma_{N,M}(g), \qquad \gamma_{N,M}(g) := \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{k=1}^{M-1}\Big(\frac{X_{t_{i+1}}(y_k) - S_{t_i}^{\Delta}(y_k)}{\Delta} - g(X_{t_i}(y_k))\Big)^2 \quad (4.17)$$

where $S_{t_i}^{\Delta} := \hat{S}(\Delta)X_{t_i}$.

As in the space-continuous case, our main effort is to bound $\hat{f}_m - f_A$ in terms of the empirical norm which is now given by

$$\|g\|_{N,M}^2 := \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{k=1}^{M-1} g^2(X_{t_i}(y_k)).$$

Using analogous steps as in the space-continuous case, we can conclude the following theorem.

Theorem 4.4.9. *Grant Assumptions (M), (E), (N) and (H). Further assume that* $\frac{N\Delta}{\log^2 N} \to \infty$ *as well as* $L_m = o(\frac{N\Delta}{\log^2 N})$ *and* $D_m \leq N$*. Then, for any* $\gamma < 1/2$ *and* $\rho < 1/4$*, we have*

$$\mathbf{E}\big(\|\hat{f}_m - f_A\|_{N,M}^2\big) \lesssim \|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho}.$$

Remark 4.4.10. The conclusion of the theorem is only useful when $M\Delta^2 \to \infty$. In the latter regime, the $\mathcal{O}(M^{-\rho})$-term is negligible with respect to the $\mathcal{O}(\Delta^\gamma)$-term. Note, furthermore, that the condition $M\Delta^2 \to \infty$ rules out a balanced sampling design.

The additional terms (compared to the space-continuous case) on the right hand side of the above inequality arise from bounding the additional parts in the bias terms $\tilde{R}_i$ of the underlying regression model: In the proof we show that for $h \geq 0$ we have $\mathbf{E}(\|S(h)X_t - \hat{S}(h)X_t\|_{L^2}^2) = \mathcal{O}(1/M)$ and the approximation error gets amplified by dividing by the squared renormalization Δ^2. Furthermore, the process $f(X_t)$ has the same regularity properties as X_t^0 but we cannot profit from independence of the coefficient processes. This results in the additional error term $\mathbf{E}(\|f(X_t) - \hat{S}(0)f(X_t)\|_{L^2}^2) = \mathcal{O}(1/M^\rho)$ for $\rho < 1/4$.

As in the discussion following Theorem 4.4.2, it can be concluded from the above theorem that if $\|f_m - f\|_{L^2(A)}^2 \approx D_m^{-2\alpha}$, the estimator $\hat{f}_m$ with $D_m \approx T^{\frac{1}{1+2\alpha}}$ admits the usual nonparametric rate of convergence $T^{-\frac{\alpha}{2\alpha+1}}$ under the conditions $L_m = o(\frac{T}{\log^2 N})$ and

$$T\Big(\Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho}\Big) \to 0$$

for some $\gamma < 1/2$ and $\rho < 1/4$.

Figure 4.1 shows four exemplary realizations of the estimator $\hat{f}_m$ with the trigonometric basis [T] when f is a linear function such that discrete observations of the solution process can be generated by means of the replacement method from Chapter 3. The compact set on which f is estimated is $A = [-1, 1]$. One can see that the general shape of the function f is captured accurately inside some interval containing the origin, roughly $[-0.6, 0.6]$. It is evident from the histograms that areas further away from the origin do not contain as many data points which, naturally, affects the quality of the estimator there. Also, there is a boundary effect caused by the fact that the functions in S_m are necessarily periodic over $[-1, 1]$.

With the same reasoning as in the space-continuous case, setting $\hat{f}_m^{K_N} := (-K_N) \vee (\hat{f}_m \wedge K_N)$, we get the following bound on the $L^2(A)$-risk.

Corollary 4.4.11. *Grant Assumptions (M), (E), (N) and (H). Further assume that* $\frac{N\Delta}{\log^2 N} \to \infty$ *as well as* $L_m = o(\frac{N\Delta}{\log^2 N})$ *and* $D_m \leq N$*. Then, for any* $\gamma < 1/2$ *and* $\rho < 1/4$*, we have*

$$\|\hat{f}_m - f\|_{L^2(A)}^2 = \mathcal{O}_p\Big(\|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho}\Big).$$

Furthermore, for a sequence (K_N) *with* $K_N \to \infty$ *and* $K_N/N^\beta \to 0$ *for some* $\beta > 0$*, we have*

$$\mathbf{E}(\|\hat{f}_m^{K_N} - f\|_{L^2(A)}^2) \lesssim \|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho}.$$

Estimation of f with unknown diffusivity and volatility

In practice, the diffusivity parameter ϑ_2 appearing in our nonparametric estimator for f will generally be unknown and has to be replaced by an estimator. To that aim, we make use of the double increments based estimator $\hat{\vartheta}_2 := \hat{\vartheta}_2^{vw}$ as defined in (2.19) with $\vartheta_1 = 0$, while omitting the spatial observations

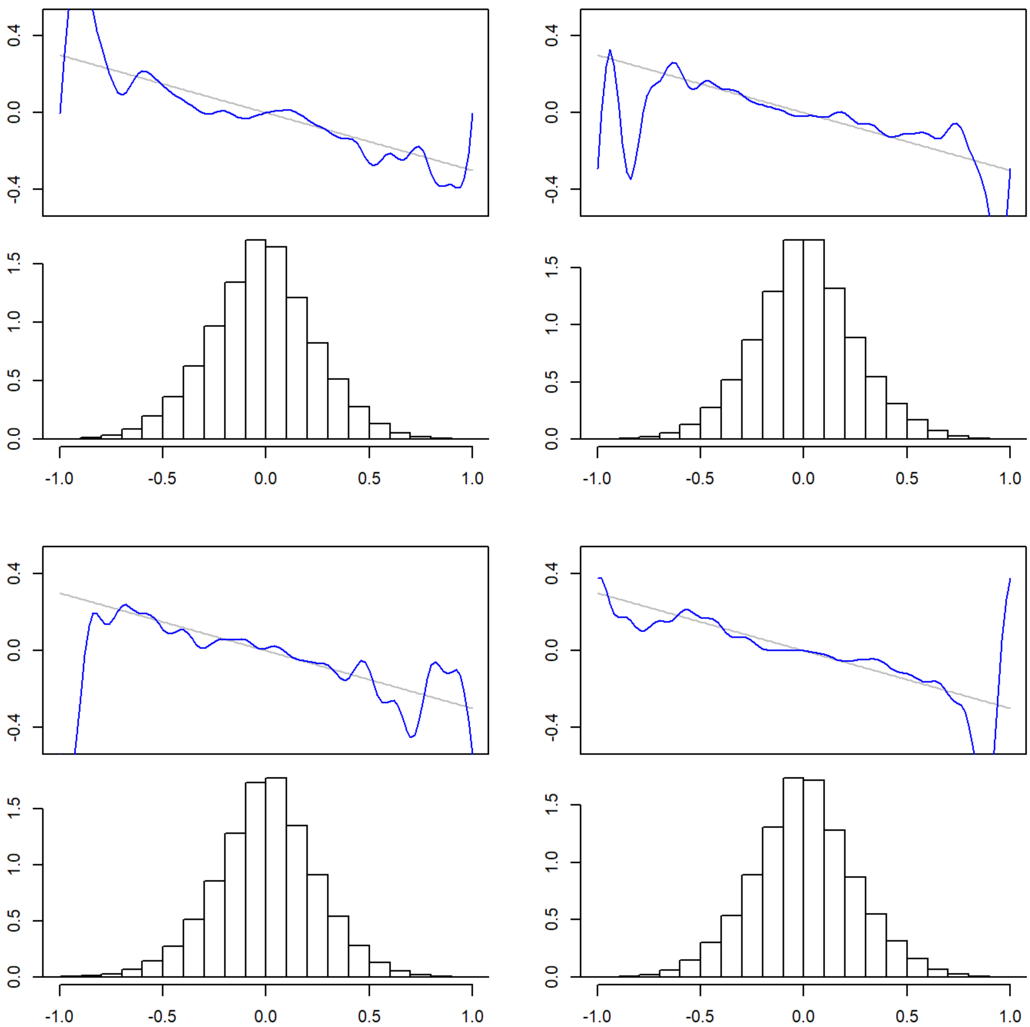


Figure 4.1: Four realizations of the estimator $\hat{f}_m$ from (4.17) with the trigonometric basis [T] on $A = [-1, 1]$ (blue) along with the true underlying function $f(x) = -0.3 \cdot x$ (gray). The barplot below each realization shows a histogram of the corresponding discrete observations $(X_{t_i}(y_k))_{i,k}$. The sample sizes are given via $M = 500$, $T = 200$, $\Delta = 0.05$. The dimension of the approximation space was chosen to be $D_m = 2m + 1 = 29$, which corresponds to $D_m \approx \sqrt{T}$. The discrete observations of X are obtained by means of the replacement method with parameter $L = 1$, see Chapter 3. The remaining parameter values are $\sigma^2 = \vartheta_2 = 0.05$.

$y_k \notin [b, 1-b]$ for an arbitrary but fixed $b > 0$. Recall that the computation of $\hat{\vartheta}_2$ does not require prior knowledge of the volatility parameter σ^2 and Corollary 2.2.19 together with Section 4.3 reveal the (squared) convergence rate $(\vartheta_2 - \hat{\vartheta}_2)^2 = \mathcal{O}_p(\max(\Delta^{3/2}, \delta^3)/T)$. Since the conclusion of Theorem 4.4.9 is only useful as long as $M\Delta^2 \to \infty$, we work in the regime $M\sqrt{\Delta} \to \infty$ where the (squared) convergence rate is given by $(\vartheta_2 - \hat{\vartheta}_2)^2 = \mathcal{O}_p(\Delta^{3/2}/T)$. Now, based on the estimator $\hat{\vartheta}_2$, we can define an approximation $\check{S}(\Delta)$ of the discretized semigroup $\hat{S}(\Delta)$, namely

$$\check{S}(\Delta)u := \sum_{\ell=1}^{M-1} \mathrm{e}^{-\hat{\lambda}_\ell \Delta} \langle u, e_\ell \rangle_M e_\ell \quad \text{with} \quad \hat{\lambda}_\ell := \pi^2 \hat{\vartheta}_2 \ell^2$$

for continuous functions $u : [0,1] \to \mathbb{R}$. The corresponding nonparametric estimator for f is then given by

$$\check{f}_m := \underset{g \in S_m}{\arg\min} \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{k=1}^{M-1} \Big(\frac{X_{t_{i+1}}(y_k) - \check{S}_{t_i}^{\Delta}(y_k)}{\Delta} - g(X_{t_i}(y_k)) \Big)^2$$

where $\check{S}_{t_i}^{\Delta} := \check{S}(\Delta) X_{t_i}$. In order to analyze the convergence rate of $\check{f}_m$, we incorporate the approximation of the semigroup into the regression model. Due to $\hat{S}(0) = \check{S}(0)$, it is now given by

$$\frac{\hat{S}(0) X_{t_{i+1}} - \check{S}(\Delta) X_{t_i}}{\Delta} = \hat{S}(0) f(X_{t_i}) + R_i' + \varepsilon_i$$

with

$$R_i' := \tilde{R}_i + \frac{\hat{S}(\Delta) X_{t_i} - \check{S}(\Delta) X_{t_i}}{\Delta}.$$

Based on this representation, we are now going to show that the approximation of the discretized semigroup does not affect the convergence rate of the nonparametric estimator. In fact, since our error bound for $\hat{\vartheta}_2$ from Corollary 2.2.19 is a priori only valid in probability sense, the same holds for $\check{f}_m$.

Theorem 4.4.12. *Grant Assumptions (M), (E), (N) and (H). Further assume that* $M\sqrt{\Delta} \to \infty$, $T\sqrt{\Delta} \to 0$ *and* $\frac{N\Delta}{\log^2 N} \to \infty$ *as well as* $L_m = o(\frac{N\Delta}{\log^2 N})$ *and* $D_m \le N$. *Then, for any* $\gamma < 1/2$ *and* $\rho < 1/4$, *we have*

$$\|\check{f}_m - f_A\|_{N,M}^2 = \mathcal{O}_p\Big(\|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho} \Big).$$

Furthermore, the same bound holds for $\|\check{f}_m - f\|_{L^2(A)}^2$.

Remark 4.4.13. The proof does not make use of any properties of $\hat{\vartheta}_2$ apart from its convergence rate. In general, if $\tilde{\vartheta}_2$ is any estimator for ϑ_2 with $|\tilde{\vartheta}_2 - \vartheta_2| = \mathcal{O}_p(a_{M,N,T})$ for some rate of convergence $a_{M,N,T}$, we get the additional term $\frac{a_{N,M,T}^2}{\Delta^{3/2}}$ in the above upper bound. For $\hat{\vartheta}_2$, this additional term does not appear in the theorem, as it is dominated by D_m/T.

Circumventing spectral approximations?

A drawback of our nonparametric estimation method based on fully discrete observations is that its validity is only supported by the theory when the observation frequency in space is much larger than in time. Indeed, Theorem 4.4.9 only serves as consistency result as long as $M\Delta^2 \to \infty$. This issue results from the bias introduced by approximating the coefficient processes of X by their empirical counterparts, in order to get an approximation of the semigroup. Due to the roughness of the paths $x \mapsto X_t(x)$, the corresponding approximation quality is rather poor and the error gets amplified by the renormalization. A similar effect can be observed in Kaino and Uchida [49] where a spectral approximation is used for parametric estimation for the linear equation, see also the end of Section 2.2. Thus, it could be beneficial to set up an estimation method that does not require the spectral approximation step. Also, this would enable considering observations that are not distributed throughout the whole space domain $[0, 1]$ which corresponds to allowing $b > 0$ in our observation scheme from Section 1.2.3. In the following, we describe an approach which we have pursued unsuccessfully. Nevertheless, it would be interesting to consider for other models, in particular, when there is more regularity in the driving noise.

For simplicity, set $X_0 = 0$ and recall that the nonlinear component (N_t) of the solution process satisfies the equation $dN_t = A_\vartheta N_t + f(X_t)$. Thus, treating the terms involving X_t^0 as stochastic noise, we have the expansion

$$\begin{aligned}\frac{X_{t+\Delta}(x) - X_t(x)}{\Delta} &= \frac{N_{t+\Delta}(x) - N_t(x)}{\Delta} + \varepsilon_t^\Delta(x)\\ &= A_\vartheta N_t(x) + f(X_t(x)) + R_t^\Delta(x) + \varepsilon_t^\Delta(x)\end{aligned}$$

with

$$\varepsilon_t^\Delta(x) := \frac{X_{t+\Delta}^0(x) - X_t^0(x)}{\Delta}, \qquad R_t^\Delta(x) := \frac{N_{t+\Delta}(x) - N_t(x)}{\Delta} - \frac{d}{dt}N_t(x).$$

On the other hand, for sufficiently smooth functions, the second order differential operator A_ϑ can be approximated by means of second order differences. Thus, we consider the expansion

$$\begin{aligned}\vartheta_2\frac{X_t(x+\delta) - 2X_t(x) + X_t(x-\delta)}{\delta^2} &= \vartheta_2\frac{N_t(x+\delta) - 2N_t(x) + N_t(x-\delta)}{\delta^2} + \tilde{\varepsilon}_t^\delta(x)\\ &= A_\vartheta N_t(x) + \tilde{R}_t^\delta(x) + \tilde{\varepsilon}_t^\delta(x)\end{aligned}$$

with

$$\begin{aligned}\tilde{\varepsilon}_t^\delta(x) &:= \vartheta_2\frac{X_t^0(x+\delta) - 2X_t^0(x) + X_t^0(x-\delta)}{\delta^2},\\ \tilde{R}_t^\delta(x) &:= \vartheta_2\frac{N_t(x+\delta) - 2N_t(x) + N_t(x-\delta)}{\delta^2} - A_\vartheta N_t(x).\end{aligned}$$

Setting $X_{ik} := X_{t_i}(y_k)$ and

$$Y_{ik} := \frac{X_{t_{i+1}}(y_k) - X_{t_i}(y_k)}{\Delta} - \vartheta_2\frac{X_{t_i}(x_{k+1}) - 2X_{t_i}(y_k) + X_{t_i}(x_{k-1})}{\delta^2},$$

we obtain the regression model

$$Y_{ik} = f(X_{ik}) + R_{ik} + \varepsilon_{ik}$$

with the bias and stochastic noise terms

$$R_{ik} := R_{t_i}^\Delta(y_k) - \tilde{R}_{t_i}^\delta(y_k), \qquad \varepsilon_{ik} := \varepsilon_{t_i}^\Delta(y_k) - \tilde{\varepsilon}_{t_i}^\delta(y_k).$$

The corresponding least squares estimator is then given by

$$\bar{f}_m = \operatorname*{arg\,min}_{g\in\mathcal{S}_m}\frac{1}{MN}\sum_{k=0}^{M-1}\sum_{i=0}^{N-1}(Y_{ik} - g(X_{ik}))^2.$$

The bias terms R_{ik} can be controlled thanks to the following lemma:

Lemma 4.4.14. *Grant Assumption (B). For any $p \geq 1$ and $\gamma < 1/2$ there exists $C > 0$ such that*

$$\mathbf{E}\Big(\Big|\Delta^{-1}\Big(N^0_{t+\Delta}(x) - N^0_t(x)\Big) - \partial_t N^0_t(x)\Big|^p\Big) \leq C\Delta^{p\frac{\gamma}{2}}, \tag{4.18}$$

$$\mathbf{E}\Big(\Big|\frac{\vartheta_2}{\delta^2}\Big(N^0_t(x+\delta) - 2N^0_t(x) + N^0_t(x-\delta)\Big) - A_\vartheta N^0_t(x)\Big|^p\Big) \leq C\delta^{p\gamma}. \tag{4.19}$$

for any $(t,x) \in \mathbb{R}_+ \times (0,1)$.

A proof of this lemma is given in Section 4.5.1. The remainder terms R_{ik} are well-behaved but we run into problems with the stochastic noise terms. In fact, these are correlated to the covariates and do not meet the fundamental requirement of a regression model that $\mathbf{E}(\varepsilon_{ik}|X_{ik}) = 0$. Indeed, since $e_\ell(y_{k+1}) - 2e_\ell(y_k) + e_\ell(y_{k-1}) = -4\sin^2(\pi\ell\delta/2)e_\ell(y_k)$, we have

$$\begin{aligned}\varepsilon_{ik} &= \frac{X^0_{t_{i+1}}(y_k) - X^0_{t_i}(y_k)}{\Delta} - \vartheta_2 \frac{X^0_{t_i}(y_{k+1}) - 2X^0_t(y_k) + X^0_t(y_{k-1})}{\delta^2} \\ &= \sum_{\ell \geq 1}\Big(\frac{u_\ell(t_{i+1}) - u_\ell(t_i)}{\Delta} + \frac{4\vartheta_2 \sin^2(\pi\ell\delta/2)u_\ell(t_i)}{\delta^2}\Big)e_\ell(y_k)\end{aligned}$$

and there is no reason why this expression should be independent of $X_{t_i}(y_k)$. In fact, assuming a balanced sampling design $\Delta = \delta^2$, we can further decompose

$$\varepsilon_{ik} = \sum_{\ell \geq 1} \frac{u_\ell(t_{i+1}) - \mathrm{e}^{-\lambda_\ell \Delta}u_\ell(t_i)}{\Delta} e_\ell(y_k) + \sum_{\ell \geq 1} \frac{\big(4\vartheta_2 \sin^2(\pi\ell\sqrt{\Delta}/2) - 1 + \mathrm{e}^{-\lambda_\ell\Delta}\big)u_\ell(t_i)}{\Delta} e_\ell(y_k)$$

such that the first term is independent of $\mathcal{F}_{t_i}$ due to the AR(1)-structure of the Ornstein-Uhlenbeck processes u_ℓ. Still, the second term is not negligible, as its variance is of the order

$$\sum_{\ell \geq 1} \frac{(4\vartheta_2 \sin^2(\pi\ell\sqrt{\Delta}/2) - 1 + \mathrm{e}^{-\lambda_\ell\Delta})^2}{\lambda_\ell \Delta} \approx \frac{1}{\sqrt{\Delta}}$$

due to the usual Riemann sum argument.

As already mentioned, the estimation procedure might be fruitful, on the other hand, if one considers a noise process with more regularity than a cylindrical Brownian motion. Indeed, if the denominator in the above sum grew faster than λ_k, one could profit from the fact that

$$4\vartheta_2 \sin^2(\frac{x}{2}) - 1 + \mathrm{e}^{-\vartheta_2 x^2} \approx x^4 \quad \text{for} \quad x \to 0.$$

Of course, a more regular noise process also results in an improved regularity of the solution process X. This, in turn, also facilitates spectral approximations so that further comparisons of the two approaches will be interesting.

4.5 Proofs

We start by proving the results on the Hölder regularity of the linear and nonlinear component of X. The two subsequent sections contain the proofs of our main results, namely on diffusivity and volatility estimation, Section 4.5.2, and on nonparametric estimation of the nonlinearity f, Section 4.5.3. Further proofs and auxiliary results are deferred to Section 4.5.4.

4.5.1 Proofs for the Hölder regularity of X

We verify the results on the Hölder regularity of the processes X and (N_t) claimed in Propositions 4.2.1 and 4.2.3 of Section 4.2, respectively. To that aim, recall that for $s \geq 0$ and $p \geq 1$ the Sobolev spaces $W^{s,p} := W^{s,p}((0,1))$ are defined as the set of all $[s]$-times weakly differentiable functions $u : (0,1) \to \mathbb{R}$ such that

$$\|u\|_{W^{s,p}} := \sum_{k=0}^{[s]} \|u^{(k)}\|_{L^p} + \left(\int_0^1 \int_0^1 \frac{|u^{([s])}(\xi) - u^{([s])}(\eta)|^p}{|\xi - \eta|^{1+(s-[s])p}} \, d\xi \, d\eta \right)^{1/p} < \infty.$$

Further, the Sobolev embedding theorem states that for any $s \geq 0$, $p \geq 1$ and $\alpha > 0$, we have

$$s - \frac{1}{p} > \alpha \quad \Rightarrow \quad W^{s,p} \subset C^\alpha \tag{4.20}$$

and the embedding is continuous. Our first step is an analysis of the Hölder regularity of the linear component X^0. We remark that the norm bounds in statements (i) and (ii) of the following lemma are also stated in Cerrai [13] as well as Da Prato and Zabczyk [26].

Lemma 4.5.1. *For any $p \in [1, \infty)$, the following hold.*

(i) $\sup_{t \geq 0} \mathbf{E}(\|X_t^0\|_\infty^p) < \infty$.

(ii) For any $\gamma < \frac{1}{2}$, we have $X^0 \in C(\mathbb{R}_+, C_0^\gamma)$ a.s. and $\sup_{t \geq 0} \mathbf{E}(\|X_t^0\|_{C_0^\gamma}^p) < \infty$.

(iii) For any $\gamma < \frac{1}{4}$ and $T > 0$, we have $(X_t^0)_{0 \leq t \leq T} \in C^\gamma([0,T], E)$ a.s. and there exists a constant $C > 0$ such that $\mathbf{E}(\|X_t^0 - X_s^0\|_\infty^p) \leq C|t-s|^{\gamma p}$ for all $s, t \geq 0$.

Proof. (iii) The property $(X_t^0)_{0 \leq t \leq T} \in C^\gamma([0,T], E)$ is a consequence of Kolmogorov's criterion and $\mathbf{E}(\|X_t^0 - X_s^0\|_\infty^p) \leq C|t-s|^{\gamma p}$. To verify the latter statement, we proceed similarly as in the proof of Kolmogorov's test from [26, Theorem 3.5], see also [26, Remark 11.35]. Without loss of generality, assume $s, t \in (a, a+1)$ for some $a \geq 0$ and define $\mathcal{U} := (a, a+1) \times (0,1)$. Using (4.6) and (4.7), we see that

$$\begin{aligned} \mathbf{E}(|X_t^0(x) - X_s^0(y)|^2) &\lesssim \sqrt{|t-s|} + |x-y| \leq \sqrt{|t-s|} + \sqrt{|x-y|} \\ &\lesssim ((t-s)^2 + (x-y)^2)^{1/4} \end{aligned}$$

holds uniformly in $x, y \in (0,1)$ and $s, t \geq 0$. The last step follows from the equivalence of norms on $\mathbb{R}^2$. Now, since $(t,x) \mapsto X_t^0(x)$ is a continuous function, we can use the following bound on its increments from Da Prato and Zabczyk [25, Theorem B.1.5]: for any $\alpha > 0, \beta > 4$, there exists a constant $c > 0$ (independent of a) such that

$$|X_s^0(x) - X_t^0(y)| \leq c((x-y)^2 + (t-s)^2)^{\frac{\beta-4}{2\alpha}} \left(\int_{\mathcal{U} \times \mathcal{U}} \frac{|X_u^0(\xi) - X_v^0(\eta)|^\alpha}{(|\xi-\eta|^2 + |u-v|^2)^{\beta/2}} \, d\xi \, d\eta \, du \, dv \right)^{\frac{1}{\alpha}} \tag{4.21}$$

for all $(x,s), (y,t) \in \mathcal{U}$. Note that for $x = y$, the right hand side of the above inequality is independent of x. Now, choose $\alpha = 2m$ for some $m \in \mathbb{N}$ in such a way that $\alpha = 2m > p$. Then, by applying Jensen's inequality to the concave function $\mathbb{R}_+ \ni h \mapsto h^{p/\alpha}$, we obtain

$$\begin{aligned} \mathbf{E}(\sup_x |X_s^0(x) - X_t^0(x)|^p) &\leq c^p (t-s)^{\frac{\beta-4}{\alpha} p} \left(\int_{\mathcal{U} \times \mathcal{U}} \frac{\mathbf{E}(|X_u^0(\xi) - X_v^0(\eta)|^{2m})}{(|\xi-\eta|^2 + |u-v|^2)^{\beta/2}} \, d\xi \, d\eta \, du \, dv \right)^{\frac{p}{\alpha}} \\ &\leq c^p (t-s)^{\frac{\beta-4}{\alpha} p} \left(\int_{\mathcal{U} \times \mathcal{U}} \frac{(|\xi-\eta|^2 + |u-v|^2)^{m/4}}{(|\xi-\eta|^2 + |u-v|^2)^{\beta/2}} \, d\xi \, d\eta \, du \, dv \right)^{\frac{p}{\alpha}}. \end{aligned}$$

The above integral is finite as long as $\beta - \frac{m}{2} < 2$. Now, the result follows since for any given $\gamma < 1/4$, we can pick $m \in \mathbb{N}$ and $\beta < 2 + \frac{m}{2}$ such that $\frac{\beta-4}{2m} \leq \gamma$.

Assertion (i) can be proved similarly by taking $s = t$ and $y = 1$ in (4.21) to obtain a bound for $\sup_x |X_t^0(x)| = \sup_x |X_t^0(x) - X_s^0(y)|$. Note that in order to be able to chose $y = 1$, we have to modify the set $\mathcal{U}$ by taking, e.g., $\mathcal{U} = (a, a+1) \times (-\varepsilon, 1+\varepsilon)$ for some $\varepsilon > 0$, and extend X^0 by defining $X_t(z) := 0$ for $z \notin [0,1]$ such that X^0 is a continuous function on $\mathcal{U}$.

(ii) Clearly, A_ϑ is a second order differential operator whose eigenvalues satisfy the condition $\sum_{\ell \geq 1} \lambda_\ell^{-\rho}$ for any $\rho > 1/2$. Thus, by [26, Theorem 5.25], $X^0 \in C(\mathbb{R}_+, W^{2\alpha,p})$ holds for any $\alpha > 0$ and $p > 1$ such that $1/p + \alpha < 1/4$. Now, by choosing α close to $1/4$ and p sufficiently large, $X^0 \in C(\mathbb{R}_+, C_0^\gamma)$ follows from the Sobolev embedding theorem (4.20). In order to bound the norm we use the following argument taken from [13]: with the bound (4.6) for the Gaussian process X^0, we get for any $h \in (0,1)$ that

$$\begin{aligned} \mathbf{E}(\|X_t^0\|_{W^{h,q}}^q) &\lesssim \mathbf{E}(\|X_t^0\|_\infty^q) + \int_0^1 \int_0^1 \frac{\mathbf{E}(|X_t^0(\xi) - X_t^0(\eta)|^q)}{|\xi - \eta|^{1+hq}} \, d\xi \, d\eta \\ &\leq \mathbf{E}(\|X_t^0\|_\infty^q) + \int_0^1 \int_0^1 \frac{|\xi - \eta|^{q/2}}{|\xi - \eta|^{1+hq}} \, d\xi \, d\eta. \end{aligned}$$

In view of (i), this shows that $\sup_{t \geq 0} \mathbf{E}(\|X_t^0\|_{W^{h,q}}^q) < +\infty$ as long as $h < 1/2$. Further, by the Sobolev embedding theorem, we have $\|X_t^0\|_{C_0^\gamma} \lesssim \|X_t^0\|_{W^{h,q}}$, provided that $h - \frac{1}{q} > \gamma$. Thus, choosing $h \in (\gamma, 1/2)$ and $q > \max((h-\gamma)^{-1}, p)$, we get

$$\mathbf{E}(\|X_t^0\|_{C_0^\gamma}^p) \lesssim \mathbf{E}(\|X_t^0\|_{W^{h,q}}^{q\frac{p}{q}}) \leq \mathbf{E}(\|X_t^0\|_{W^{h,q}}^q)^{\frac{p}{q}}$$

by Jensen's inequality. The claim now follows by taking the supremum over $t \geq 0$ in the above inequality. □

Before proving Proposition 4.2.1, we recall some facts from semigroup theory. For details, in particular, on semigroups generated by differential operators we refer to, e.g., [55]. For $\tilde{E} = C([0,1])$, consider the part $A_{\tilde{E}}$ of $A_\vartheta = \vartheta_2 \Delta$ in $\tilde{E}$, as defined in (1.4). As already mentioned, $A_{\tilde{E}}$ generates an analytic semigroup $S_{\tilde{E}}$ on $\tilde{E}$ which is not strongly continuous at 0. Indeed, this follows from $\overline{\mathcal{D}(A_{\tilde{E}})}^{\tilde{E}} = E$ and the semigroup $S_{\tilde{E}}$ satisfies $\lim_{t \to 0} S_{\tilde{E}}(t)x = x$ if and only if $x \in E$. Nevertheless, for any $x \in \tilde{E}$ it holds that $\int_0^t S_{\tilde{E}}(r)x \, dr \in \mathcal{D}(A_{\tilde{E}})$ and we have the representation

$$S_{\tilde{E}}(t)x - x = A_{\tilde{E}} \int_0^t S_{\tilde{E}}(r)x \, dr. \tag{4.22}$$

Hence, if $r \mapsto \|A_{\tilde{E}} S_{\tilde{E}}(r)x\|_{\tilde{E}}$ is integrable over $[0,t]$, then $S_{\tilde{E}}(t)x - x = \int_0^t A_{\tilde{E}} S_{\tilde{E}}(r)x \, dr$. Since the definitions of the semigroups S, S_E and $S_{\tilde{E}}$ and their generators agree on the intersection of their domains, respectively, we will refer to all three by S and A from now on.

Proof of Proposition 4.2.1. Due to Lemma 4.5.1, in order to prove the statements for X, they have to be proved for (N_t) and, if ξ follows the stationary distribution, for $(\xi_t)_{t \geq 0}$ with $\xi_t := S(t)\xi$.

(i) *Step 1.* We show $\|N_t\|_{C_0^\gamma} < \infty$ a.s. for all $t \geq 0$ and, under Assumption (B), $\sup_{t \geq 0} \mathbf{E}(\|N_t\|_{C_0^\gamma}^p) < \infty$: From Lemma 4.2.2 (ii) we have that

$$\|N_t\|_{C_0^\gamma} \leq \int_0^t \|S(t-s)f(X_s)\|_{C_0^\gamma} \, ds \lesssim \int_0^t \mathrm{e}^{-\lambda_0(t-s)} (t-s)^{-\frac{\gamma}{2}} \|f(X_s)\|_\infty \, ds$$

and, consequently,

$$\|N_t\|_{C_0^\gamma} \lesssim \sup_{s \leq t} \|f(X_s)\|_\infty \int_0^t \mathrm{e}^{-\lambda_0 r} r^{-\frac{\gamma}{2}} \, dr$$

is finite almost surely by our basic assumptions. Also, using Jensen's inequality and the fact that $r \mapsto a(r) := \mathrm{e}^{-\lambda_0 r} r^{-\frac{\gamma}{2}}$ is integrable over $\mathbb{R}_+$, we get

$$\|N_t\|_{C_0^\gamma}^p \le \int_0^t a(t-s)\|f(X_s)\|_\infty^p \, ds \cdot \left(\int_0^t a(r)\, dr\right)^{p-1} \lesssim \int_0^t a(t-s)\|f(X_s)\|_\infty^p \, ds.$$

Thus, Fubini's theorem and the polynomial growth condition on f from (4.3) yield

$$\sup_{t\ge 0} \mathbf{E}(\|N_t\|_{C_0^\gamma}^p) \lesssim \sup_{s\ge 0} \mathbf{E}(\|f(X_s)\|_\infty^p) \lesssim 1 + \sup_{s\ge 0} \mathbf{E}(\|X_s\|_\infty^{dp})$$

which is finite under Assumption (B).

Step 2: We show $(N_t) \in C(\mathbb{R}_+, C_0^\gamma)$: In order to verify $\|N_{t+h} - N_t\|_{C_0^\gamma} \to 0$ for $h \to 0$ a.s., we use the decomposition

$$N_{t+h} - N_t = (S(h) - I)N_t + \int_t^{t+h} S(t+h-r) f(X_r)\, dr.$$

To treat the first term, choose $\alpha \in (\gamma, 1/2)$. Then, using (4.22) and property (iv) of Lemma 4.2.2, we can bound

$$\|(S(h) - I)N_t\|_{C_0^\gamma} \lesssim \int_0^h \|A_\vartheta S(r) N_t\|_{C_0^\gamma}\, dr \le \|N_t\|_{C_0^\alpha} \int_0^h \mathrm{e}^{-\lambda_0 r} r^{-(1+\frac{\gamma-\alpha}{2})}\, dr$$

which tends to 0 for $h \to 0$. For the second term, it follows from bound (ii) in Lemma 4.2.2 that

$$\Big\| \int_t^{t+h} S(t+h-r) f(X_r)\, dr \Big\|_{C_0^\gamma} \lesssim \sup_{r\le T} \|f(X_r)\|_\infty \int_t^{t+h} \mathrm{e}^{-\lambda_0 r} r^{-\frac{\gamma}{2}}\, dr$$

which also tends to 0 a.s. for $h \to 0$.

Step 3: Steps 1 and 2 verify claim (i) in the case $\xi = 0$. To treat the case where ξ follows the stationary distribution, we use the fact that X has the same distribution as $\tilde{X} = (X_{1+t})_{t\ge 0}$. Again, we have the decomposition

$$\tilde{X}_t = S(1+t)\xi + X_{1+t}^0 + N_{1+t}$$

and (i) has already been proved for the second and third term. For the first term, the result follows from $\|S(1+t)\xi\|_{C_0^\gamma} \lesssim \|\xi\|_\infty = \|X_0\|_\infty$ by inequality (ii) in Lemma 4.2.2.

Step 4. We transfer the result (i) from X to $f_0(X)$: First of all, $f_0(X) \in C(\mathbb{R}_+, C_0^\gamma)$ a.s. holds due to the result for X and the assumption $f_0 \in C^1(\mathbb{R})$. Further, we have

$$\|f_0(X_t)\|_{C_0^\gamma} = \|f_0(X_t)\|_\infty + \sup_{\xi \ne \eta} \frac{|f(X_t(\xi)) - f(X_t(\eta))|}{|\xi - \eta|^\gamma} \le \|f_0(X_t)\|_\infty + \|f'(X_t)\|_\infty \|X_t\|_{C_0^\gamma}$$

and, under Assumption (B),

$$\begin{aligned} \mathbf{E}(\|f_0(X_t)\|_{C_0^\gamma}^p) &\lesssim \mathbf{E}(\|f_0(X_t)\|_\infty^p) + \mathbf{E}(\|f'(X_t)\|_\infty^{2p}) + \mathbf{E}(\|X_t\|_{C_0^\gamma}^{2p}) \\ &\lesssim 1 + \mathbf{E}(\|X_t\|_\infty^{2dp}) + \mathbf{E}(\|X_t\|_{C_0^\gamma}^{2p}) < \infty \end{aligned}$$

uniformly in $t \ge 0$.

(ii) *Step 1:* We show the claim for (N_t): Using the same decomposition for the increments of N as in the proof of (i), we get

$$\|N_t - N_s\|_\infty \le \|(S(t-s) - I)N_s\|_\infty + \int_s^t \|S(t-r) f(X_r)\|_\infty\, dr$$

for $s < t$. For the first term, by definition of the intermediate spaces, it holds that

$$\|(S(t-s)-I)N_s\|_\infty \lesssim \|N_s\|_{D_{A_\vartheta}(\gamma,\infty)}\,(t-s)^\gamma \lesssim \|N_s\|_{C_0^{2\gamma}}\,(t-s)^\gamma. \tag{4.23}$$

By Lemma 4.2.2 (i) and Hölder's inequality, we have

$$\begin{aligned}\Big\|\int_s^t S(t-r)f(X_r)\,dr\Big\|_\infty^p &\le \Big(\int_s^t \|S(t-r)f(X_r)\|_\infty\,dr\Big)^p\\ &\lesssim \Big(\int_s^t \mathrm{e}^{-\lambda_0(t-r)}\|f(X_r)\|_\infty\,dr\Big)^p\\ &\le (t-s)^{p-1}\int_s^t \mathrm{e}^{-p\lambda_0(t-r)}\|f(X_r)\|_\infty^p\,dr. \end{aligned}\tag{4.24}$$

By combining (4.23) and (4.24), we obtain $(N_t)_{0\le t\le T}\in C^\gamma([0,T],E)$ a.s. and, under Assumption (B),

$$\mathbf{E}\left(\|N_t-N_s\|_\infty^p\right) \lesssim (t-s)^{\gamma p}\mathbf{E}(\|N_s\|^p_{C_0^{2\gamma}}) + (t-s)^p(1+\sup_{h\ge 0}\mathbf{E}(\|X_h\|_\infty^{pd}),$$

from which the result for (N_t) follows due to (i).

Step 2. The case where ξ follows the stationary distribution can be treated as in (i) since

$$\|S(1+t)\xi - S(1+s)\xi\|_\infty \lesssim (t-s)^\gamma\|S(1)\xi\|_{C_0^{2\gamma}} \lesssim (t-s)^\gamma\|X_0\|_\infty.$$

Step 3. We transfer the result (ii) from X to $f_0(X)$: First of all, the pathwise property is again a consequence of the assumption $f_0\in C^1(\mathbb{R})$. Next, without loss of generality, assume that d from (4.3) is given by $d=2m$ for some $m\in\mathbb{N}$. Then, using the formula $a^n-b^n=(a-b)\sum_{k=0}^{n-1}a^kb^{n-1-k}$ for $a,b\in\mathbb{R}$ and $n\in\mathbb{N}$, yields

$$\begin{aligned}|f(X_t(x))-f(X_s(x))| &\le \int_{X_s(x)}^{X_t(x)}|f'(h)|\,dh \lesssim \int_{X_s(x)}^{X_t(x)}(1+h^{2m})\,dh\\ &= |X_t(x)-X_s(x)| + \frac{1}{2m+1}|X_t(x)^{2m+1}-X_s(x)^{2m+1}|\\ &\lesssim |X_t(x)-X_s(x)|\underbrace{\left(1+\sum_{k=0}^{2m}|X_t(x)^kX_s(x)^{2m-k}|\right)}_{=:Z_{s,t}}\end{aligned}$$

where we have assumed $X_t(x)\ge X_s(x)$ without loss of generality. Consequently, since $(s,t)\mapsto\|Z_{s,t}\|_\infty$ is bounded in $L^p(\mathbf{P})$ for any $p\ge 1$ under Assumption (B), we obtain

$$\mathbf{E}(\|f(X_t)-f(X_s))\|_\infty^p) \lesssim \mathbf{E}(\|X_t-X_s\|_\infty^{2p})^{1/2}\mathbf{E}(\|Z_{s,t}\|_\infty^{2p})^{1/2} \lesssim (t-s)^{\gamma p}. \qquad \square$$

We turn to the excess Hölder regularity of the nonlinear component (N_t) of X. Since (N_t) is the pathwise solution of the equation $dN_t = A_\vartheta N_t + f(S(t)\xi + X_t^0 + N_t)$, $N_0=0$, the almost sure properties are a consequence of the results of Sinestrari [72] on the regularity of solutions to deterministic systems. In the following, we give a direct proof for them, both for the sake of completeness and since we require its steps in order to bound the respective norms in $L^p(\mathbf{P})$.

Proof of Proposition 4.2.3. (i) Due to Proposition 4.2.1, we have $f_0(X_t)\in D_{A_\vartheta}(\gamma,\infty)=C_0^{2\gamma}$ for any $\gamma<1/4$. Further, for any $\tilde\gamma\in(\gamma,\frac14)$, Lemma 4.2.2 (iv) yields that

$$\|A_\vartheta N_t^0\|_{C_0^{2\gamma}} \lesssim \int_0^t \|A_\vartheta S(t-s)f_0(X_s)\|_{C_0^{2\gamma}}\,ds \le \int_0^t h(t-s)\|f_0(X_s)\|_{C_0^{2\tilde\gamma}}\,ds$$

with $h(r) := \mathrm{e}^{-\lambda_0 r} r^{-1+\tilde{\gamma}-\gamma}$. Since h is integrable over $\mathbb{R}_+$, the pathwise properties immediately follow from $f_0(X) \in C(\mathbb{R}_+, C_0^{2\tilde{\gamma}})$, cf. Proposition 4.2.1. Further, Jensen's inequality gives

$$\|A_\vartheta N_t^0\|_{C_0^{2\gamma}}^p \lesssim \int_0^t h(t-s)\|f_0(X_s)\|_{C_0^{2\tilde{\gamma}}}^p \, ds \left(\int_0^t h(r)\, dr\right)^{p-1}.$$

Consequently, under Assumption (B), $\sup_{t\geq 0} \mathbf{E}(\|A_\vartheta N_t^0\|_{C_0^{2\gamma}}^p) \lesssim \sup_{t\geq 0} \mathbf{E}(\|f_0(X_t)\|_{C_0^{2\tilde{\gamma}}}^p)$ is finite by Proposition 4.2.1.

(ii) In order to prove $\frac{d}{dt} N_t = A_\vartheta N_t^0 + f_0(X_t)$ in E, note that the usual decomposition for the increments of (N_t^0) and formula (4.22) yield the representation

$$\begin{aligned}\Delta^{-1}(N_{t+\Delta}^0 - N_t^0) - A_\vartheta N_t^0 - f_0(X_t) =& \frac{1}{\Delta}\int_0^\Delta (S(r)-I)A_\vartheta N_t^0 \, dr \\ &+ \frac{1}{\Delta}\int_t^{t+\Delta} \Big(S(t+\Delta-r)f_0(X_r) - f_0(X_t)\Big)\, dr.\end{aligned}$$

We have $\|(S(r)-I)A_\vartheta N_t^0\|_\infty \leq r^\gamma \|A_\vartheta N_t^0\|_{C_0^{2\gamma}}$ and

$$\begin{aligned}\|S(h)f_0(X_r) - f_0(X_t)\|_\infty \leq& \|S(h)(f_0(X_r) - f_0(X_t))\|_\infty + \|(S(h)-I)f_0(X_t)\|_\infty \\ \lesssim& \|f_0(X_r) - f_0(X_t)\|_\infty + h^\gamma \|f_0(X_t)\|_{C_0^{2\gamma}}.\end{aligned}$$

Thus, *(i)* and Proposition 4.2.1 yield $\|\Delta^{-1}(N_{t+\Delta}^0 - N_t^0) - A_\vartheta N_t^0 - f_0(X_t)\|_\infty \lesssim \Delta^\gamma \to 0$ uniformly on bounded time intervals, almost surely.

The properties claimed for $\frac{d}{dt}N_t^0$ now follow from the properties of $f_0(X_t)$ provided by Proposition 4.2.1 and

$$\begin{aligned}\|A_\vartheta N_{t+\Delta}^0 - A_\vartheta N_t^0\|_\infty \leq& \|(S(\Delta)-I)A_\vartheta N_t^0\|_\infty + \int_t^{t+\Delta} \|A_\vartheta S(t+\Delta-r)f_0(X_r)\|_\infty \, dr \\ \lesssim& \Delta^\gamma \|A_\vartheta N_t^0\|_{C_0^{2\gamma}} + \int_t^{t+\Delta} (t+\Delta-r)^{-1+\gamma}\|f_0(X_r)\|_{C_0^{2\gamma}}\, dr\end{aligned}$$

where the bound on the integrand is taken from result *(iii)* in Lemma 4.2.2.

It remains to analyze the regularity of the process M. First of all, by (4.22), we have $A_\vartheta M_t = S(t)m - m$ and $m \in C^{2\gamma}([b, 1-b])$ is trivially fulfilled. Further, setting $m_t := S(t)m$, we have $m_t(x) = \frac{2\sqrt{2}}{\pi}\sum_{\ell\geq 0} \frac{\mathrm{e}^{-\lambda_{2\ell+1}t}}{2\ell+1} e_{2\ell+1}(x)$. The mean value theorem yields

$$m_t(x) - m_t(y) = (x-y)8\sum_{\ell\geq 0} \mathrm{e}^{-\lambda_{2\ell+1}t}\cos(\pi(2\ell+1)\xi)$$

for some ξ between x and y. Thanks to Lemma 2.4.8, the sum $\sum_{\ell\geq 0}\mathrm{e}^{-\lambda_{2\ell+1}t}\cos(\pi(2\ell+1)\xi)$ is uniformly bounded in $t > 0$ and $\xi \in [b, 1-b]$ and we can conclude $\sup_{t\geq 0}\|A_\vartheta M_t\|_{C^{2\gamma}([b,1-b])} < \infty$. The same argument shows that

$$\|S(t+\Delta)m - S(t)m\|_{C([b,1-b])} \lesssim \sup_{\ell\geq 0} \frac{1-\mathrm{e}^{-\lambda_{2\ell+1}\Delta}}{2\ell+1} \lesssim \sqrt{\Delta}.$$

Hence,

$$\|\Delta^{-1}(M_{t+\Delta} - M_t) - S(t)m\|_{C([b,1-b])} \leq \frac{1}{\Delta}\int_t^{t+\Delta} \|(S(r)m - S(t)m)\|_{C([b,1-b])}\, dr \lesssim \sqrt{\Delta}$$

and, in particular, $\frac{d}{dt}M_t = S(t)m$ in $C([b, 1-b])$ as well as $\|\frac{d}{dt}(M_{t+\Delta} - M_t)\|_{C([b,1-b])} \lesssim \sqrt{\Delta} \lesssim \Delta^\gamma$. $\square$

We conclude this section by analyzing the approximation quality of the difference quotients for the derivatives of (N_t).

Proof of Lemma 4.4.14. For the first statement, note that by Jensen's inequality applied to the Lebesgue integral, we have

$$\begin{aligned}\mathbf{E}(|\Delta^{-1}(N^0_{t+\Delta}(x) - N^0_t(x)) - \partial_t N^0_t(x)|^p) &= \mathbf{E}\Big(\Big|\frac{1}{\Delta}\int_t^{t+\Delta} \partial_t(N^0_s(x) - N^0_t(x))\, ds\Big|^p\Big)\\ &\leq \frac{1}{\Delta}\int_t^{t+\Delta} \mathbf{E}\Big(\Big|\partial_t(N^0_s(x) - N^0_t(x))\Big|^p\Big)\, ds.\end{aligned}$$

Further, $|\partial_t(N^0_s(x) - N^0_t(x)|^p \lesssim |f_0(X_s(x)) - f_0(X_t(x))|^p + |A_\vartheta(N^0_s(x) - N^0_t(x))|^p$. From Proposition 4.2.1 we get for $s \in [t, t+\Delta]$ that

$$\mathbf{E}(|f_0(X_s(x)) - f_0(X_t(x))|^p) \lesssim \Delta^{p\frac{\gamma}{2}}.$$

To estimate the second term, we use the usual decomposition $N^0_s - N^0_t = (S(s-t) - I)N^0_t + \int_t^s S(s-r)f_0(X_r)\, dr$ and

$$\|A_\vartheta(S(s-t) - I)N^0_t\|_\infty \leq \Delta^{\gamma/2}\|A_\vartheta N^0_t\|_{D_{A_\vartheta}(\gamma/2,\infty)} \lesssim \Delta^{\gamma/2}\|A_\vartheta N^0_t\|_{C^\gamma_0}$$

as well as

$$\Big\|A_\vartheta \int_t^s S(s-r)f_0(X_r)\, dr\Big\|_\infty \lesssim \int_t^s (s-r)^{-(1-\gamma/2)}\|f_0(X_r)\|_{C^\gamma_0}\, dr,$$

see Lemma 4.2.2 (iii). Thus, we have by Hölder's inequality that

$$\mathbf{E}(|A_\vartheta(N^0_s(x) - N^0_t(x))|^p) \lesssim \Delta^{p\frac{\gamma}{2}}\mathbf{E}(\|A_\vartheta N^0_t\|^p_{C^\gamma}) + (s-t)^{p\frac{\gamma}{2}-1}\int_t^s \mathbf{E}(\|f_0(X_r)\|^p_{C^\gamma})\, dr$$

which, in view of Propositions 4.2.1 and 4.2.3, finishes the proof of (4.18). For the proof of (4.19), recall that by Taylor's formula, we have the expansion $h(x+\delta) = h(x) + \delta h'(x) + \int_x^{x+\delta}(x+\delta-z)h''(z)\, dz$ for any $h \in C^2(\mathbb{R})$. Hence, we can write

$$\delta^{-2}(h(x+\delta) - 2h(x) + h(x-\delta)) = \int K_\delta(z-x)h''(z)\, dz \tag{4.25}$$

with $K_\delta(z) := \delta^{-1}K(\delta^{-1}z)$ and the triangular kernel $K(z) := (1-|z|)\mathbf{1}_{\{-1\leq z\leq 1\}}$. Since K_δ integrates to 1, it follows that

$$\begin{aligned}\frac{\vartheta_2}{\delta^2}(N^0_t(x+\delta) - 2N^0_t(x) + N^0_t(x-\delta)) - A_\vartheta N^0_t(x) &= \int_{x-\delta}^{x+\delta} K_\delta(z-x)(A_\vartheta N^0_t(z) - A_\vartheta N^0_t(x))\, dz\\ &\leq (2\delta)^\gamma\|A_\vartheta N^0_t\|_{C^\gamma_0},\end{aligned}$$

from where the result follows due to Proposition 4.2.3. □

4.5.2 Proofs for the estimators of σ^2 and ϑ_2

In the following, we prove the central limit theorems for the realized quadratic variations in the semi-linear framework, as claimed in Theorems 4.3.1,4.3.2 and 4.3.3 of Section 4.3. The corresponding results for the estimators then follow directly in view of the delta method. This also applies to the joint estimator of (σ^2, ϑ_2), as we assume $\vartheta_0 = 0$ and, hence, $(\hat{\sigma}^2, \hat{\vartheta}_2)$ can be directly calculated from two realized quadratic variations based on double increments when the sample size is sufficiently large, see Remark 2.2.13.

We start with the result for time increments.

Proof of Theorem 4.3.1. It is sufficient to consider the process $X^0_{t+t_0} + N_{t+t_0}$ instead of X_t: if ξ follows the stationary distribution, then X has the same distribution as $\tilde{X}$ with $\tilde{X}_t := X_{t_0+t} = S(t)S(t_0)\xi + X^0_{t+t_0} + N_{t+t_0}$ for any $t_0 > 0$ and $(S(t)S(t_0)\xi)_{t\geq 0}$ can be chosen arbitrarily regular by choosing t_0 sufficiently large. In fact, since the properties of (N_t) used in the sequel, are the same under each of the initial conditions, we can assume $\xi = 0$ for simplicity. Then, we have

$$\begin{aligned}
\bar{V}_{\mathrm{t}} &= \frac{1}{MN\sqrt{\Delta}} \sum_{i=0}^{N-1}\sum_{k=0}^{M-1} (X^0_{t_{i+1}}(y_k) - X^0_{t_i}(y_k))^2 + \frac{1}{MN\sqrt{\Delta}} \sum_{i=0}^{N-1}\sum_{k=0}^{M-1} (N_{t_{i+1}}(y_k) - N_{t_i}(y_k))^2 \\
&\quad + \frac{2}{MN\sqrt{\Delta}} \sum_{i=0}^{N-1}\sum_{k=0}^{M-1} (X^0_{t_{i+1}}(y_k) - X^0_{t_i}(y_k))(N_{t_{i+1}}(y_k) - N_{t_i}(y_k)) \\
&=: V_{\mathrm{t}} + R_1 + R_2.
\end{aligned}$$

Since V_{t} satisfies the claimed central limit theorem (2.10), due to Slutsky's lemma, it suffices to prove that R_1 and R_2 are of the order $o_p(1/\sqrt{MN})$.

If T is finite, it follows from Lemma 4.5.1 and Proposition 4.2.3 that for all $\gamma < 1/4$ and $\mathbf{P}$-almost all realizations $\omega \in \Omega$, there exists a constant $C = C(\omega, T)$ such that $|X^0_{t_{i+1}}(y_k) - X^0_{t_i}(y_k)| \leq C\Delta^\gamma$ and $|N_{t_{i+1}}(y_k) - N_{t_i}(y_k)| \leq C\Delta$ for all $i \leq N$, $k \leq M$ and $N, M \in \mathbb{N}$. Consequently, R_1 and R_2 are of the order $o_p(\Delta^{\frac{1}{2}+\gamma})$ and the statement follows due to the condition $M = o(\Delta^{-\rho})$ for some $\rho < 1/2$.

If $T \to \infty$ and Assumption (B) is satisfied, Lemma 4.5.1 and Proposition 4.2.3 yield $\mathbf{E}(|R_1|) \lesssim \Delta^{3/2}$ and, by applying the Cauchy-Schwarz inequality to the cross terms, we get $\mathbf{E}(|R_2|) \lesssim \Delta^{\frac{1}{2}+\gamma}$ for any $\gamma < 1/4$. The claim follows since $\sqrt{MN}\Delta^{\frac{1}{2}+\gamma} = \sqrt{TM\Delta^{2\gamma}}$ converges to 0 for any $\gamma \in (\rho/2, 1/4)$ and the fact that convergence in $L^1(\mathbf{P})$ implies convergence in probability. □

Next, we prove the result for space increments.

Proof of Theorem 4.3.2. We only consider the case of a finite time horizon, the case $T \to \infty$ can be treated similarly by taking expectations. Further, it suffices to consider the case $\xi = 0$, see also the proof of Theorem 4.3.1. We have

$$\begin{aligned}
\bar{V}_{\mathrm{sp}} &= \frac{1}{MN\delta} \sum_{i=1}^{N}\sum_{k=0}^{M-1} (X^0_{t_i}(y_{k+1}) - X^0_{t_i}(y_k))^2 + \frac{1}{MN\delta} \sum_{i=1}^{N}\sum_{k=0}^{M-1} (N_{t_i}(y_{k+1}) - N_{t_i}(y_k))^2 \\
&\quad + \frac{2}{MN\delta} \sum_{i=1}^{N}\sum_{k=0}^{M-1} (X^0_{t_i}(y_{k+1}) - X^0_{t_i}(y_k))(N_{t_i}(y_{k+1}) - N_{t_i}(y_k)) \\
&=: V_{\mathrm{sp}} + R_1 + R_2
\end{aligned}$$

and the claim follows if R_1 and R_2 are of the order $o_p(1/\sqrt{MN})$.

To bound the term R_2, we use the formula

$$\sum_{k=0}^{M-1} a_k(b_{k+1} - b_k) = -\sum_{k=0}^{M-2} (a_{k+1} - a_k)b_{k+1} + a_{M-1}b_M - a_0 b_0. \tag{4.26}$$

Setting $a_k := N_{t_i}(y_{k+1}) - N_{t_i}(y_k)$ and $b_k := X^0_{t_i}(y_k)$, we get

$$\begin{aligned}
R_2 &= \frac{2}{MN\delta} \sum_{i=1}^{N}\sum_{k=0}^{M-2} X^0_{t_i}(y_{k+1})(N_{t_i}(y_{k+2}) - 2N_{t_i}(y_{k+1}) + N_{t_i}(y_k)) \\
&\quad + \frac{1}{MN\delta} \sum_{i=1}^{N} \left((N_{t_i}(y_N) - N_{t_i}(y_{N-1}))X^0_{t_i}(y_N) + (N_{t_i}(y_1) - N_{t_i}(y_0))X^0_{t_i}(y_0)\right).
\end{aligned}$$

By Lemma 4.5.1 and Proposition 4.2.3, we have $X^0 \in C(\mathbb{R}_+, C([b, 1-b]))$ and $\sup_{t \le T} \|A_\vartheta N_t\|_{C([b,1-b])} < \infty$ a.s. Thus, there exists a random variable $C = C(\omega, T)$ with $|N_{t_i}(y_{k+2}) - 2N_{t_i}(y_{k+1}) - N_{t_i}(y_k)| \le C\delta^2$, $|N_{t_i}(y_{k+1}) - N_{t_i}(y_k)| \le C\delta$ and $|X^0_{t_i}| \lesssim C$ for all $i \le N$, $k \le M-1$ and $M, N \in \mathbb{N}$ a.s. It follows that $|R_1| \le C^2\delta$ and $|R_2| \lesssim C^2\delta$ hold a.s. and, therefore, the claim follows from the fact that $\sqrt{MN}\delta \approx \sqrt{N/M}$ tends to 0, by assumption. □

Finally, we prove the result for double increments. To that aim, let $\mathbb{N}_{ik}$ denote the double increments computed from the process (N_t), i.e., $\mathbb{N}_{ik} := N_{t_{i+1}}(y_{k+1}) - N_{t_{i+1}}(y_k) - N_{t_i}(y_{k+1}) + N_{t_i}(y_k)$. The first step of the proof of Theorem 4.3.3 is given by the following lemma.

Lemma 4.5.2. *Assume that the constant b defined in Section 1.2.3 is strictly positive and let $p \ge 1$.*

(i) Let $\alpha \in (0,1)$ and $\beta \in (0,1]$ be such that $\alpha + \beta < 3/2$. If T is finite, then there exists a random variable $C = C(\omega, T) > 0$ such that

$$|\mathbb{N}_{ik}| \le C\delta^\alpha \Delta^{\frac{1+\beta}{2}}$$

holds for all $i \le N, k \le M$ and $N, M \in \mathbb{N}$ a.s. If Assumption (B) is satisfied, then there exists a constant $C > 0$ such that

$$\mathbf{E}(|\mathbb{N}_{ik}|^p) \le C\Big(\delta^\alpha \Delta^{\frac{1+\beta}{2}}\Big)^p$$

holds for all $i \le N, k \le M$, $N, M \in \mathbb{N}$ uniformly in $T > 0$.

(ii) Let $\gamma < 2$ and $\varepsilon < \frac{1}{4}$. If T is finite, then there exists a random variable $C = C(\omega, T) > 0$ such that

$$|\mathbb{N}_{i(k+1)} - \mathbb{N}_{ik}| \le C\delta^\gamma \Delta^\varepsilon$$

holds for all $i \le N, k \le M$ and $N, M \in \mathbb{N}$ a.s. If Assumption (B) is satisfied, then there exists a constant $C > 0$ such that

$$\mathbf{E}(|\mathbb{N}_{i(k+1)} - \mathbb{N}_{ik}|^p) \le C\Big(\delta^\gamma \Delta^\varepsilon\Big)^p$$

holds for all $i \le N, k \le M$, $N, M \in \mathbb{N}$ uniformly in $T > 0$.

Proof. We write $\mathbb{N}_{ik} = \mathbb{N}^0_{ik} + \mathbb{M}_{ik}$, where $\mathbb{N}^0_{ik}$ and $\mathbb{M}_{ik}$ are the double increments computed from the processes (N^0_t) and (M_t) defined by (4.8), respectively. In the following, these double increments are estimated separately.

(i) For $\alpha \in (0,1)$, we have

$$\begin{aligned} |\mathbb{N}^0_{ik}| &\le \delta^\alpha \|N^0_{t_{i+1}} - N^0_{t_i}\|_{C^\alpha_0} \\ &\le \delta^\alpha \left(\|S(\Delta) - I)N^0_{t_i}\|_{C^\alpha_0} + \Big\| \int_{t_i}^{t_{i+1}} S(t_{i+1} - s) f_0(X_s)\, ds \Big\|_{C^\alpha_0} \right). \end{aligned}$$

Further, using formula (4.22) and Lemma 4.2.2 (iv) in combination with $\alpha + \beta - 1 \le \alpha$ yields

$$\begin{aligned} \|S(\Delta) - I)N^0_{t_i}\|_{C^\alpha_0} &= \Big\| \int_0^\Delta A_\vartheta S(r) N^0_{t_i}\, dr \Big\|_{C^\alpha_0} \\ &\le \int_0^\Delta \|S(r)\|_{L(C^{\alpha+\beta-1}_0, C^\alpha_0)} \|A_\vartheta N^0_{t_i}\|_{C^{\alpha+\beta-1}_0}\, dr \\ &\lesssim \int_0^\Delta r^{-\frac{1-\beta}{2}} \|A_\vartheta N^0_{t_i}\|_{C^{\alpha+\beta-1}_0}\, dr \lesssim \Delta^{\frac{1+\beta}{2}} \|A_\vartheta N^0_t\|_{C^{\alpha+\beta-1}_0}. \end{aligned}$$

Similarly, by Lemma 4.2.2 (*iii*) and Hölder's inequality,

$$\Big\| \int_{t_i}^{t_{i+1}} S(t_{i+1}-s) f_0(X_s)\, ds \Big\|_{C_0^\alpha} \lesssim \int_{t_i}^{t_{i+1}} (t_{i+1}-s)^{-\frac{1-\beta}{2}} \|f_0(X_s)\|_{C_0^{\alpha+\beta-1}}\, ds$$
$$\lesssim \Big(\int_{t_i}^{t_{i+1}} \|f_0(X_s)\|^p_{C_0^{\alpha+\beta-1}}\, ds \Big)^{\frac{1}{p}} \Delta^{1-\frac{1}{p}-\frac{1-\beta}{2}}.$$

Thus, noting $\alpha+\beta-1<1/2$, Propositions 4.2.1 and 4.2.3 yield the claim for the case of a finite time horizon and, under Assumption (B),

$$\mathbf{E}(|\mathbb{N}^0_{ik}|^p) \lesssim \delta^{p\alpha} \Big(\Delta^{p\frac{1+\beta}{2}} \mathbf{E}(\|A_\vartheta N^0_t\|^p_{C_0^{\alpha+\beta-1}}) + \Delta^{p-1-p\frac{1-\beta}{2}} \int_{t_i}^{t_{i+1}} \mathbf{E}(\|f_0(X_s)\|^p_{C_0^{\alpha+\beta-1}})\, ds \Big)$$
$$\lesssim \delta^{p\alpha} \Delta^{p\frac{1+\beta}{2}}.$$

To verify that $\mathbb{M}_{ik}$ is of the claimed order, recall that in the proof of Proposition 4.2.3 it is shown that $\frac{d}{dt}M_t = S(t)m =: m_t$ and that $|m_t(x)-m_t(y)| \lesssim |x-y|$ holds uniformly in $t>0$ and $x,y \in [b,1-b]$. Thus, we have $\mathbb{M}_{ik} = \int_{t_i}^{t_{i+1}} (m_s(y_{k+1}) - m_s(y_k))\, ds$ and, consequently, $|\mathbb{M}_{ik}| \lesssim \Delta\delta \lesssim \delta^\alpha \Delta^{\frac{1+\beta}{2}}$.

(*ii*) For $\gamma \in (1,2)$, we have

$$|\mathbb{N}^0_{i(k+1)} - \mathbb{N}^0_{ik}| \le \delta^\gamma \|N^0_{t_{i+1}} - N^0_{t_i}\|_{C_0^\gamma}.$$

Using the decomposition

$$N^0_{t_{i+1}} - N^0_{t_i} = \int_0^{t_i} S(t_i-s)(f_0(X_{s+\Delta}) - f_0(X_s))\, ds + \int_0^\Delta S(t_{i+1}-s) f_0(X_s)\, ds,$$

we get from Lemma 4.2.2 (*i*) that

$$\Big\| \int_0^{t_i} S(t_i-s)(f_0(X_{s+\Delta}) - f_0(X_s))\, ds \Big\|_{C_0^\gamma} = \Big\| \int_0^{t_i} S(r)(f_0(X_{t_{i+1}-r}) - f_0(X_{t_i-r}))\, dr \Big\|_{C_0^\gamma}$$
$$\lesssim \int_0^{t_i} \mathrm{e}^{-\lambda_0 r} r^{-\frac{\gamma}{2}} \|f_0(X_{t_{i+1}-r}) - f_0(X_{t_i-r})\|_\infty\, dr.$$

Further, for $h<\frac{1}{2}$, Lemma 4.2.2 (*iii*) gives

$$\Big\| \int_0^\Delta S(t_{i+1}-s) f_0(X_s)\, ds \Big\|_{C_0^\gamma} \lesssim \int_0^\Delta (t_{i+1}-r)^{-\frac{\gamma-h}{2}} \|f_0(X_r)\|_{C_0^h}\, dr.$$

Now, the result in case of a fixed T follows from the path regularity of $f_0(X)$. Further, under Assumption (B), we can use Jensen's and Hölder's inequality to estimate

$$\mathbf{E}(|\mathbb{N}^0_{i(k+1)} - \mathbb{N}^0_{ik}|^p) \lesssim \delta^\gamma \sup_{t\ge 0} \mathbf{E}(\|f_0(X_{t+\Delta}) - f_0(X_t)\|^p_\infty) + \delta^\gamma \Delta^{p-1-\frac{\gamma-h}{2}p} \int_0^\Delta \mathbf{E}(\|f_0(X_{t_{i+1}-r})\|^p_{C_0^h})\, dr$$
$$\lesssim \delta^\gamma \Delta^{p\varepsilon} + \delta^\gamma \Delta^{p(1-\frac{\gamma-h}{2})}.$$

The result follows, since one can pick $h \in (0,\frac{1}{2})$ such that $1-\frac{\gamma-h}{2} \ge \varepsilon$.

To estimate $|\mathbb{M}_{ik}|$, recall that in the proof of Proposition 4.2.3 it is shown that $\frac{\partial^2}{\partial x^2}M_t = \frac{1}{\vartheta_2}A_\vartheta M_t = \frac{1}{\vartheta_2}(S(t)-I)m$ and that $\|\frac{\partial^2}{\partial x^2}M_t - \frac{\partial^2}{\partial x^2}M_s\|_{C([b,1-b])} = \frac{1}{\vartheta_2}\|S(t)m - S(s)m\|_{C([b,1-b])} \lesssim \sqrt{|t-s|}$. Application of (4.25) to the double increments yields

$$\mathbb{M}_{i(k+1)} - \mathbb{M}_{ik} = \delta^2 \int_{x-\delta}^{x+\delta} K_\delta(z-x) \frac{\partial^2}{\partial z^2}(M_{t_{i+1}}(z) - M_{t_i}(z))\, dz$$

and, consequently, $|\mathbb{M}_{i(k+1)} - \mathbb{M}_{ik}| \lesssim \delta^2\sqrt{\Delta} \lesssim \Delta^\gamma \Delta^\varepsilon$. □

The above lemma is the main ingredient for the following proof of the central limit theorem for double increments.

Proof of Theorem 4.3.3. As for time and space increments, we can assume $\xi = 0$ and the claim follows if we verify $|R_i| = o_p(1/\sqrt{MN})$, $i \in \{1,2\}$, with

$$R_1 := \frac{1}{MN\Phi_\vartheta(\delta,\Delta)} \sum_{i=1}^{N} \sum_{k=0}^{M-1} \mathbb{N}_{ik}^2, \qquad R_2 := \frac{1}{MN\Phi_\vartheta(\delta,\Delta)} \sum_{i=1}^{N} \sum_{k=0}^{M-1} \mathbb{N}_{ik} D_{ik}$$

and where D_{ik} are the double increments computed from X^0. In the following, we verify the claim under Assumption (B). The result for the case of a fixed T can be shown analogously by using the pathwise properties of (N_t) derived in Lemma 4.5.2. We treat the cases $M\sqrt{\Delta} = \mathcal{O}(1)$ and $M\sqrt{\Delta} \to \infty$ separately.

Case $M\sqrt{\Delta} = \mathcal{O}(1)$: Using Lemma 4.5.2 with $\alpha = 0$ and $\beta = 1$ yields $\mathbf{E}(\mathbb{N}_{ik}^2) \lesssim \Delta^2$ and, hence,

$$\sqrt{MN}\mathbf{E}(|R_1|) \lesssim \frac{\sqrt{MN}}{MN\sqrt{\Delta}} \sum_{i,k} \mathbf{E}(\mathbb{N}_{ik}^2) \lesssim \sqrt{MN}\Delta^{3/2} = \sqrt{\frac{T}{M}} \cdot M\sqrt{\Delta} \cdot \sqrt{\Delta} \to 0$$

since each of the three factors tends to zero. For the cross terms, we take $\beta = 1$ and $\alpha = \frac{a}{2} < \frac{1}{2}$ in Lemma 4.5.2 to bound

$$\mathbf{E}(|D_{ik}\mathbb{N}_{ik}|) \le \mathbf{E}(D_{ik}^2)^{1/2}\mathbf{E}(\mathbb{N}_{ik}^2)^{1/2} \lesssim \Delta^{1/4}\Delta\delta^{a/2}.$$

Consequently,

$$\sqrt{MN}\mathbf{E}(|R_1|) \lesssim \sqrt{MN}\frac{\Delta^{5/4}\delta^{a/2}}{\sqrt{\Delta}} \approx \sqrt{\frac{T}{M^a}} \cdot \sqrt{M}\Delta^{1/4} \to 0.$$

Case $M\sqrt{\Delta} \to \infty$: With $\beta = \frac{1}{2}$ and $\alpha = \frac{a+3}{4} < 1$ in Lemma 4.5.2, we get $\mathbf{E}(\mathbb{N}_{ik}^2) \lesssim \Delta^{3/2}\delta^{2\alpha}$ and, hence,

$$\sqrt{MN}\mathbf{E}(|R_1|) \lesssim \sqrt{MN}\frac{\Delta^{3/2}\delta^{2\alpha}}{\delta} \approx \sqrt{\frac{T}{M^a}} \cdot \Delta \to 0.$$

To treat the cross terms, we use formula (4.26) with $a_k := \mathbb{N}_{ik}$ and $b_k := H_{ik} := X^0_{t_{i+1}}(y_k) - X^0_{t_i}(y_k)$ to deduce

$$\sum_{i=1}^{N}\sum_{k=0}^{M-1} D_{ik}\mathbb{N}_{ik} = -\sum_{i=1}^{N}\sum_{k=0}^{M-2} (\mathbb{N}_{i(k+1)} - \mathbb{N}_{ik})H_{i(k+1)} + \sum_{i=1}^{N} \mathbb{N}_{i(M-1)}H_{iM} - \sum_{i=1}^{N} \mathbb{N}_{i0}H_{i0}. \tag{4.27}$$

Since $\mathbf{E}(H_{ik}^2) \lesssim \sqrt{\Delta}$, Lemma 4.5.2 gives for any $\gamma < 2$ and $\varepsilon < \frac{1}{4}$ that

$$\begin{aligned}
&\mathbf{E}\left(\Big|\frac{1}{MN\delta}\sum_{i=1}^{N}\sum_{k=0}^{M-2}(\mathbb{N}_{i(k+1)} - \mathbb{N}_{ik})H_{i(k+1)}\Big|\right) \\
&\le \frac{1}{MN\delta}\sum_{i=1}^{N}\sum_{k=0}^{M-2}\mathbf{E}((\mathbb{N}_{i(k+1)} - \mathbb{N}_{ik})^2)^{1/2}\mathbf{E}(H_{i(k+1)}^2)^{1/2} \\
&\lesssim \frac{\delta^\gamma\Delta^\varepsilon\sqrt{\Delta}}{\delta} = \delta^{\gamma-1}\Delta^{\varepsilon+1/4}.
\end{aligned}$$

Further, by picking ε and γ in such a way that $2\gamma - 4 + 4\varepsilon > a$, we get

$$\sqrt{MN}\delta^{\gamma-1}\Delta^{\varepsilon+1/4} = \sqrt{\frac{T}{M^{2\gamma-4+4\varepsilon}}} \cdot \frac{1}{(M\sqrt{\Delta})^{1/2-2\varepsilon}} \to 0.$$

For the remaining two terms in (4.27), take $\alpha = \frac{a+1}{2} < 1$ and $\beta = \frac{1}{2}$ in Lemma 4.5.2. Then,

$$\mathbf{E}\left(\Big|\frac{1}{MN\delta}\sum_{i=1}^{N}\mathbb{N}_{i(M-1)}H_{iM}\Big|\right) \lesssim \frac{\Delta^{1/4}\delta^{\alpha}\Delta^{3/4}}{M\delta} \approx \Delta\delta^{\gamma}$$

and, finally,

$$\sqrt{MN}\Delta\delta^{\alpha} = \sqrt{\frac{T}{M^{a}}} \cdot \sqrt{\Delta} \to 0.$$

Summarizing, we have shown that $\sqrt{MN}\mathbf{E}(|R_2|) \to 0$, which finishes the proof. □

4.5.3 Proofs for the nonparametric estimator of f

Space-continuous observations

We prove our results for the space-continuous observation scheme. To that aim, we follow the main proof strategy from Comte et al. [21, Proposition 1]. First, we verify our estimate for $\|\hat{f}_m - f_A\|_N^2$ on the event $\Omega_{N,m}$.

Proof of Proposition 4.4.4. By applying the Cauchy-Schwarz inequality and Young's inequality to (4.15), we can bound

$$\begin{aligned}
\|\hat{f}_m - f_A\|_N^2 \leq & \|f_m - f_A\|_N^2 + \frac{2}{N}\sum_{i=0}^{N-1}\big\|\hat{f}_m(X_{t_i}) - f_m(X_{t_i})\big\|_{L^2}\big\|R_i\big\|_{L^2} \\
& + \|\hat{f}_m - f_m\|_N \sup_{t\in S_m,\,\|t\|_N=1}\frac{2}{N}\sum_{i=0}^{N-1}\big\langle t(X_{t_i}), \varepsilon_i\big\rangle_{L^2} \\
\leq & \|f_m - f_A\|_N^2 + \frac{1}{\eta}\|\hat{f}_m - f_m\|_N^2 + \frac{\eta}{N}\sum_{i=0}^{N-1}\big\|R_i\big\|_{L^2}^2 \\
& + \frac{1}{\eta}\|\hat{f}_m - f_m\|_N^2 + \eta\left(\sup_{t\in S_m,\,\|t\|_N=1}\frac{1}{N}\sum_{i=0}^{N-1}\big\langle t(X_{t_i}), \varepsilon_i\big\rangle_{L^2}\right)^2 \\
\leq & \Big(1 + \frac{4}{\eta}\Big)\|f_m - f_A\|_N^2 + \frac{4}{\eta}\|\hat{f}_m - f_A\|_N^2 + \frac{\eta}{N}\sum_{i=0}^{N-1}\big\|R_i\big\|_{L^2}^2 \\
& + \eta\left(\sup_{t\in S_m,\,\|t\|_N=1}\frac{1}{N}\sum_{i=0}^{N-1}\big\langle t(X_{t_i}), \varepsilon_i\big\rangle_{L^2}\right)^2
\end{aligned}$$

for any $\eta > 0$. Taking $\eta = 8$ and rearranging gives

$$\|\hat{f}_m - f_A\|_N^2 \leq 3\|f_m - f_A\|_N^2 + \frac{16}{N}\sum_{i=0}^{N-1}\big\|R_i\big\|_{L^2}^2 + 16\left(\sup_{t\in S_m,\,\|t\|_N=1}\frac{1}{N}\sum_{i=0}^{N-1}\big\langle t(X_{t_i}), \varepsilon_i\big\rangle_{L^2}\right)^2. \tag{4.28}$$

The claim of the proposition follows by bounding the expectation on $\Omega_{N,m}$ of the three terms on the right hand side of the above inequality. For the first term, we have

$$\mathbf{E}(\|f_m - f_A\|_N^2\mathbf{1}_{\Omega_{N,m}}) \leq \mathbf{E}(\|f_m - f_A\|_N^2) = \|f_m - f_A\|_\pi^2.$$

For the second term, note that with $f_0 := f - f(0)$ and $\mathbf{1} := \mathbf{1}_{[0,1]}$, we have

$$\begin{aligned}\Big\|S(h)f(X_s)-f(X_t)\Big\|_{L^2}^2 \lesssim & \Big\|S(h)f_0(X_s)-f_0(X_s)\Big\|_{\infty}^2 + f(0)^2\Big\|S(h)\mathbf{1}-\mathbf{1}\Big\|_{L^2}^2 \\ & + \Big\|f(X_s)-f(X_t)\Big\|_{\infty}^2 \\ \lesssim & h^{\gamma}\|f_0(X_s)\|^2_{D_A(\gamma/2,\infty)} + f(0)^2\sum_{\ell\geq 1}(1-\mathrm{e}^{-\lambda_\ell\Delta})^2\langle \mathbf{1}, e_\ell\rangle^2 \\ & + \Big\|f(X_s)-f(X_t)\Big\|_{\infty}^2.\end{aligned}$$

Using Jensen's inequality, $D_A(\gamma/2,\infty) = C_0^{\gamma}$ and $\langle \mathbf{1}, e_\ell\rangle^2 \lesssim \ell^{-2}$, we get for any $\gamma < 1/2$

$$\begin{aligned}\mathbf{E}\left(\|R_i\|_{L^2}^2\right) \leq & \frac{1}{\Delta}\mathbf{E}\left(\int_{t_i}^{t_{i+1}}\Big\|S(t_{i+1}-s)f(X_s)-f(X_{t_i})\Big\|_{L^2}^2\,ds\right) \\ \lesssim & \frac{1}{\Delta}\int_{t_i}^{t_{i+1}}\Delta^{\gamma}\mathbf{E}\left(\Big\|f_0(X_s)\Big\|_{C_0^{\gamma}}^2\right)ds + f(0)^2\sqrt{\Delta} \\ & + \frac{1}{\Delta}\int_{t_i}^{t_{i+1}}\mathbf{E}\left(\Big\|f(X_s)-f(X_{t_i})\Big\|_{\infty}^2\right)ds \\ \lesssim & \Delta^{\gamma} \end{aligned} \tag{4.29}$$

in view of Proposition 4.2.1. To treat the third term on the right hand side of (4.28), consider an orthonormal system $\{\varphi_\lambda,\, \lambda\in\Lambda_m\}$ of S_m with the property $\|\sum_{\lambda\in\Lambda_m}\varphi_\lambda^2\|_\infty \leq CD_m$ which exists due to Assumption (N). Since on $\Omega_{N,m}$, $\|t\|_N = 1$ implies $\|t\|^2_{L^2(A)} \leq 1/\underline{c}$, we obtain

$$\begin{aligned}\sup_{t\in S_m,\,\|t\|_N=1}\left(\frac{1}{N}\sum_{i=0}^{N-1}\langle t(X_{t_i}),\varepsilon_i\rangle_{L^2}\right)^2\mathbf{1}_{\Omega_{N,m}} \leq & \frac{1}{\underline{c}}\sup_{t\in S_m,\,\|t\|_{L^2}\leq 1}\left(\frac{1}{N}\sum_{i=0}^{N-1}\langle t(X_{t_i}),\varepsilon_i\rangle_{L^2}\right)^2 \\ = & \frac{1}{\underline{c}}\sup_{\alpha\in\mathbb{R}^{\Lambda_m},\,\|\alpha\|\leq 1}\left(\sum_{\lambda\in\Lambda_m}\alpha_\lambda\frac{1}{N}\sum_{i=0}^{N-1}\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}\right)^2 \\ \lesssim & \sum_{\lambda\in\Lambda_m}\left(\frac{1}{N}\sum_{i=0}^{N-1}\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}\right)^2.\end{aligned}$$

To handle the expectation of the above bound, note that $\varepsilon_i = \frac{\sigma}{\Delta}\int_{t_i}^{t_{i+1}} S(t_{i+1}-s)\,dW_s$ is independent of $\mathcal{F}_{t_i}$, implying

$$\begin{aligned}\mathbf{E}\left(\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}|\mathcal{F}_{t_i}\right) = & \int_0^1\mathbf{E}\big(\varphi_\lambda(X_{t_i}(x))\varepsilon_i(x)|\mathcal{F}_{t_i}\big)\,dx = \int_0^1\varphi_\lambda(X_{t_i}(x))\mathbf{E}\left(\varepsilon_i(x)|\mathcal{F}_{t_i}\right)dx \\ = & \int_0^1\varphi_\lambda(X_{t_i}(x))\mathbf{E}\left(\varepsilon_i(x)\right)dx = 0.\end{aligned}$$

Hence, for $j < i$, we have

$$\mathbf{E}\left(\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}\langle\varphi_\lambda(X_{t_j}),\varepsilon_j\rangle_{L^2}\right) = \mathbf{E}\left(\mathbf{E}\left(\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}\Big|\mathcal{F}_{t_i}\right)\left\langle\varphi_\lambda(X_{t_j}),\varepsilon_j\right\rangle_{L^2}\right) = 0$$

and, consequently,

$$\mathbf{E}\left(\sup_{t\in S_m,\,\|t\|_N=1}\left(\frac{1}{N}\sum_{i=0}^{N-1}\langle t(X_{t_i}),\varepsilon_i\rangle_{L^2}\right)^2\mathbf{1}_{\Omega_{N,m}}\right) \leq \sum_{\lambda\in\Lambda_m}\frac{1}{N^2}\sum_{i=0}^{N-1}\mathbf{E}\left(\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}^2\right).$$

Further, Parseval's relation yields

$$
\begin{aligned}
\mathbf{E}\left(\langle\varphi_\lambda(X_{t_i}),\varepsilon_i\rangle_{L^2}^2\right) &= \mathbf{E}\Bigg(\Big(\sum_{k\geq 1}\langle\varphi_\lambda(X_{t_i}),e_k\rangle_{L^2}\langle\varepsilon_i,e_k\rangle_{L^2}\Big)^2\Bigg)\\
&= \frac{\sigma^2}{\Delta^2}\mathbf{E}\Bigg(\Big(\sum_{k\geq 1}\langle\varphi_\lambda(X_{t_i}),e_k\rangle_{L^2}\int_{t_i}^{t_{i+1}}\mathrm{e}^{-\lambda_k(t_{i+1}-s)}\,d\beta_k(s)\Big)^2\Bigg)\\
&= \frac{\sigma^2}{\Delta^2}\sum_{k\geq 1}\mathbf{E}(\langle\varphi_\lambda(X_{t_i}),e_k\rangle_{L^2}^2)\mathbf{E}\left(\Big(\int_{t_i}^{t_{i+1}}\mathrm{e}^{-\lambda_k(t_{i+1}-s)}\,d\beta_k(s)\Big)^2\right)\\
&= \frac{\sigma^2}{\Delta^2}\sum_{k\geq 1}\frac{1-\mathrm{e}^{-2\lambda_k\Delta}}{2\lambda_k}\mathbf{E}(\langle\varphi_\lambda(X_{t_i}),e_k\rangle_{L^2}^2)\\
&\leq \frac{\sigma^2}{\Delta}\mathbf{E}\left(\|\varphi_\lambda(X_{t_i})\|_{L^2}^2\right).
\end{aligned}
$$

Above, we have used independence of the (one-dimensional) stochastic integrals from $\mathcal{F}_{t_i}$ and pairwise independence of $\{\beta_k,\, k\geq 1\}$ in the third step. In view of Assumption (N), we have shown

$$
\begin{aligned}
\mathbf{E}\left(\sup_{t\in S_m,\,\|t\|_N=1}\left(\frac{1}{N}\sum_{i=0}^{N-1}\langle t(X_{t_i}),\varepsilon_i\rangle_{L^2}\right)^2\mathbf{1}_{\Omega_{N,m}}\right) &\lesssim \sum_{\lambda\in\Lambda_m}\frac{1}{N\Delta}\mathbf{E}(\|\varphi_\lambda(X_0)\|_{L^2}^2)\\
&= \frac{1}{T}\mathbf{E}\left(\int_0^1\sum_{\lambda\in\Lambda_m}\varphi_\lambda^2(X_0(x))\,dx\right)\lesssim\frac{D_m}{T},
\end{aligned}
$$

which finishes the proof. □

The following proof verifies our bound on the probability of the event $\Xi_{N,m}$.

Proof of Lemma 4.4.5. We follow the steps of the proof of Lemma 1 in [21] which employs the standard technique for deriving concentration inequalities for β-mixing sequences, see, e.g., Theorem 4 in Doukhan [31, Section 1.4.2]. In contrast to the result derived in [31], a different version of Bernstein's inequality is used which is convenient to work with in our situation: it directly follows from Massart [61, Proposition 2.9] that for independent real-valued random variables $Z_1,\dots,Z_n$ with $|Z_i|\leq B$ and $\mathbf{E}(Z_i^2)\leq\nu^2$ for some constants $B,\nu>0$, we have

$$
\mathbf{P}\big(|S_n-\mathbf{E}(S_n)|\geq\nu\sqrt{2x}+Bx\big)\leq 2\mathrm{e}^{-nx}\quad\text{where}\quad S_n:=\frac{1}{n}\sum_{i=1}^n Z_i \tag{4.30}
$$

for any $x>0$. In order to be able to make use of (4.30) in our context, we need to approximate the observations $X_{t_0},\dots,X_{t_N}$ by independent blocks. In fact, using Berbee's coupling Lemma [7], it can be shown (see, e.g., the discussion following Lemma 5.1 in [77]) that there exists a process $(X^*_{i\Delta},\,0\leq i\leq N-1)$ with the following properties. For every $j=0,\dots,p_N-1$, we have

$$
\begin{aligned}
U_{j,1}&:=(X_{[2jq_N+1]\Delta},\dots,X_{[(2j+1)q_N]\Delta})\overset{\mathcal{D}}{=}(X^*_{[2jq_N+1]\Delta},\dots,X^*_{[(2j+1)q_N]\Delta})=:U^*_{j,1},\\
U_{j,2}&:=(X_{[(2j+1)q_N]\Delta},\dots,X_{[2(j+1)q_N]\Delta})\overset{\mathcal{D}}{=}(X^*_{[(2j+1)q_N]\Delta},\dots,X^*_{[2(j+1)q_N]\Delta})=:U^*_{j,2}
\end{aligned}
$$

and for each $a\in\{1,2\}$, $U^*_{0,a},\dots,U^*_{p_N-1,a}$ are independent and $\mathbf{P}(U_{j,a}\neq U^*_{j,a})\leq\beta_X(q_N\Delta)$. Here, β_X is the β-mixing coefficient of X which is in our case given by (4.4). Set $\Omega^*:=\{X_{i\Delta}=X^*_{i\Delta},\,i=0,\dots,N-1\}$ and $\mathbf{P}^*:=\mathbf{P}(\cdot\cap\Omega^*)$. Clearly,

$$
\mathbf{P}(\Xi^c_{N,m})\leq\mathbf{P}(\Omega^*\cap\Xi^c_{N,m})+\mathbf{P}((\Omega^*)^c)
$$

and using the union bound, we get

$$\mathbf{P}((\Omega^*)^c) \leq 2p_N\beta_X(q_N\Delta) \leq N\beta_X(q_N\Delta).$$

It remains to show $\mathbf{P}^*(\Xi^c_{N,m}) \lesssim D_m^2 \exp(-K'\frac{p_N}{L_m})$. To that aim, set

$$v_N(t) := \frac{1}{N}\sum_{i=0}^{N-1}\Big(\int_0^1 t(X_{i\Delta}(x))\,dx - \mathbf{E}(\int_0^1 t(X_{i\Delta}(x))\,dx)\Big)$$

so that $v_N(t^2) = \|t\|_N^2 - \|t\|_\pi^2$. Recall the constants $0 < c < C < \infty$ from the implication (4.14) of Assumption (E). We have

$$\begin{aligned}\mathbf{P}^*\left(\Xi^c_{N,m}\right) &= \mathbf{P}^*\left(\sup_{t\in S_m\setminus\{0\}}\Big|\frac{\|t\|_N^2-\|t\|_\pi^2}{\|t\|_\pi^2}\Big| \geq \frac{1}{2}\right) \leq \mathbf{P}^*\left(\sup_{t\in S_m\setminus\{0\}}\Big|\frac{\|t\|_N^2-\|t\|_\pi^2}{c\|t\|_{L^2(A)}^2}\Big| \geq \frac{1}{2}\right)\\ &= \mathbf{P}^*\left(\sup_{t\in S_m,\|t\|_{L^2(A)}=1}|v_N(t^2)| \geq \frac{c}{2}\right).\end{aligned}$$

Now, each $t \in S_m$ with $\|t\|_{L^2(A)} = 1$ has a representation $t = \sum_{\lambda\in\Lambda_m}\alpha_\lambda\varphi_\lambda$ with $\sum_{\lambda\in\Lambda_m}\alpha_\lambda^2 = 1$ and

$$v_N(t^2) = \sum_{\lambda,\lambda'}\alpha_\lambda\alpha_{\lambda'}v_N(\varphi_\lambda\varphi_{\lambda'}).$$

On the set $\{|v_N(\varphi_\lambda\varphi_{\lambda'})| \leq 2V^m_{\lambda\lambda'}(2Cx)^{1/2} + 2B^m_{\lambda\lambda'}x, \forall\lambda,\lambda' \in \Lambda_m\}$ with $x := \frac{c^2}{64CL_m}$, we have

$$\sum_{\lambda,\lambda'}|\alpha_\lambda\alpha_{\lambda'}||v_N(\varphi_\lambda\varphi_{\lambda'})| \leq 2(2Cx)^{1/2}\rho(V^m) + 2x\rho(B^m) \leq \frac{c}{2\sqrt{2}} + \frac{c^2}{32C} \leq \frac{c}{2}$$

where the last bound is due to $c \leq C$ and, hence, $\sup_{t\in S_m,\|t\|_{L^2(A)}=1}|v_N(t^2)| \leq \frac{c}{2}$ is fulfilled. Consequently,

$$\begin{aligned}\mathbf{P}^*\left(\Xi^c_{N,m}\right) &\leq \mathbf{P}^*\Big(\exists\lambda,\lambda' \in \Lambda_m : |v_N(\varphi_\lambda\varphi_{\lambda'})| \geq 2V^m_{\lambda\lambda'}(2Cx)^{1/2} + 2B^m_{\lambda\lambda'}x\Big)\\ &\leq \sum_{\lambda,\lambda'\in\Lambda_m}\mathbf{P}^*\Big(|v_N(\varphi_\lambda\varphi_{\lambda'})| \geq 2V^m_{\lambda\lambda'}(2Cx)^{1/2} + 2B^m_{\lambda\lambda'}x\Big).\end{aligned}$$

We decompose $v_N(\varphi_\lambda\varphi_{\lambda'}) = v_N^1(\varphi_\lambda\varphi_{\lambda'}) + v_N^2(\varphi_\lambda\varphi_{\lambda'})$ where

$$v_N^a(t) := \frac{1}{p_N}\sum_{j=0}^{p_N-1}\Big(Z_{j,a}(t) - \mathbf{E}(Z_{j,a}(t))\Big), \qquad Z_{j,a}(t) := \frac{1}{q_N}\sum_{i=1}^{q_N}\int_0^1 t(U^i_{j,a}(x))\,dx$$

and $U^i_{j,a}$ denotes the i-th entry of $U_{j,a}$. Under $\mathbf{P}^*$, the family $(Z_{0,a}(t),\ldots,Z_{p_N-1,a}(t))$ is independent for $a \in \{1,2\}$ by construction and satisfies

$$\begin{aligned}|Z_{j,a}(\varphi_\lambda\varphi_{\lambda'})| &\leq B^m_{\lambda,\lambda'}\\ \mathbf{E}\Big(Z^2_{j,a}(\varphi_\lambda\varphi_{\lambda'})\Big) &\leq \frac{1}{q_N}\sum_{i=1}^{q_N}\mathbf{E}\left(\Big(\int_0^1\varphi_\lambda(U^i_{j,a}(x))\varphi_{\lambda'}(U^i_{j,a}(x))\,dx\Big)^2\right)\\ &\leq \mathbf{E}\left(\int_0^1\varphi_\lambda^2(X_0(x))\varphi_{\lambda'}^2(X_0(x))\,dx\right) = \|\varphi_\lambda\varphi_{\lambda'}\|_\pi^2 \leq C\|\varphi_\lambda\varphi_{\lambda'}\|^2_{L^2(A)} = C(V^m_{\lambda,\lambda'})^2\end{aligned}$$

where we have used Jensen's inequality twice in the second line. Thus, by the Bernstein inequality (4.30), we get

$$\mathbf{P}^*\Big(|v_N(\varphi_\lambda\varphi_{\lambda'})| \geq 2V^m_{\lambda\lambda'}(2Cx)^{1/2} + B^m_{\lambda\lambda'}x\Big) \leq \sum_{a=1}^{2}\mathbf{P}^*\Big(|v^a_N(\varphi_\lambda\varphi_{\lambda'})| \geq V^m_{\lambda\lambda'}(2Cx)^{1/2} + B^m_{\lambda\lambda'}x\Big)$$
$$\leq 4\mathrm{e}^{-p_N x}.$$

Summing up, we have shown

$$\mathbf{P}^*\left(\Xi^c_{N,m}\right) \leq 4D^2_m \exp\left(-p_N\frac{c^2}{64CL_m}\right),$$

which finishes the proof. □

Based on the previous results, we are now ready to verify the conclusion of our main theorem on space-continuous observations.

Proof of Theorem 4.4.2. Consider $\Omega_{N,m} = \Omega_{N,m,\frac{c}{2}}$ as defined in Proposition 4.4.4 with $c > 0$ from the implication (4.14) of Assumption (E). Then, on $\Xi_{N,m}$, we have $\|t\|^2_N \geq \frac{1}{2}\|t\|^2_\pi \geq \frac{c}{2}\|t\|^2_{L^2(A)}$ for all $t \in S_m$, implying $\Xi_{N,m} \subset \Omega_{N,m}$. Thus,

$$\begin{aligned}\mathbf{E}\big(\|\hat{f}_m - f_A\|^2_N\big) &= \mathbf{E}\big(\|\hat{f}_m - f_A\|^2_N\mathbf{1}_{\Omega_{N,m}}\big) + \mathbf{E}\big(\|\hat{f}_m - f_A\|^2_N\mathbf{1}_{\Omega^c_{N,m}}\big)\\ &\lesssim \|f_A - f_m\|^2_\pi + \frac{D_m}{T} + \Delta^\gamma + \mathbf{E}\big(\|\hat{f}_m - f_A\|^2_N\mathbf{1}_{\Omega^c_{N,m}}\big)\\ &\lesssim \|f_A - f_m\|^2_{L^2(A)} + \frac{D_m}{T} + \Delta^\gamma + \mathbf{E}\big(\|\hat{f}_m - f_A\|^2_N\mathbf{1}_{\Xi^c_{N,m}}\big)\end{aligned}$$

by Proposition 4.4.4 and Assumption (E). In the following, we conclude the theorem by showing that

$$\mathbf{E}(\|\hat{f}_m - f_A\|^2_N\mathbf{1}_{\Xi^c_{N,m}}) = o(\Delta^\gamma).$$

We consider the Hilbert space $H^N := (L^2(0,1))^N$ equipped with the inner product $\langle u, v\rangle_{H^N} := \frac{1}{N}\sum_{i=1}^N\langle u_i, v_i\rangle_{L^2}$ for $u, v \in H^N$. Note that $\|t\|^2_N = \|\bar{t}\|^2_{H^N}$ with $\bar{t} := (t(X_0), \ldots, t(X_{(N-1)\Delta}))$. Clearly, the vector $\hat{\bar{f}}_m := (\hat{f}_m(X_0), \ldots, \hat{f}_m(X_{(N-1)\Delta}))$ is the orthogonal projection in H^N of $\bar{Y} := (Y_0, \ldots, Y_{N-1})$ onto the subspace $\{(t(X_0), \ldots, t(X_{(N-1)\Delta})),\ t \in S_m\}$. Denoting the corresponding projection operator by Π_m, we have

$$\begin{aligned}\|\hat{f}_m - f_A\|^2_N \leq \|\hat{f}_m - f\|^2_N = \|\Pi_m\bar{Y} - \bar{f}\|^2_{H^N} &= \|(I - \Pi_m)\bar{f}\|^2_{H^N} + \|\Pi_m(\bar{Y} - \bar{f})\|^2_{H^N}\\ &\leq \|\bar{f}\|^2_{H^N} + \|\bar{Y} - \bar{f}\|^2_{H^N}\end{aligned}$$

since the operator norm of the projections is given by one. Now,

$$\mathbf{E}(\|\bar{f}\|^2_{H^N}\mathbf{1}_{\Xi^c_{N,m}}) = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{E}(\|f(X_{i\Delta})\|^2_{L^2}\mathbf{1}_{\Xi^c_{N,m}}) \leq \mathbf{E}(\|f(X_0)\|^4_\infty)^{1/2}\mathbf{P}(\Xi^c_{N,m})^{1/2} \lesssim \mathbf{P}(\Xi^c_{N,m})^{1/2}$$

and

$$\mathbf{E}(\|\bar{Y} - \bar{f}\|^2_{H^N}\mathbf{1}_{\Xi^c_{N,m}}) = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{E}(\|R_i + \varepsilon_i\|^2_{L^2}\mathbf{1}_{\Xi^c_{N,m}}) \lesssim (\mathbf{E}(\|R_i\|^4_{L^2})^{1/2} + \mathbf{E}(\|\varepsilon_i\|^4_{L^2})^{1/2})\mathbf{P}(\Xi^c_{N,m})^{1/2}.$$

It can be shown just like in (4.29) that for any $\gamma < 1/2$, we have

$$\mathbf{E}(\|R_i\|^4_{L^2}) \lesssim \Delta^{2\gamma} = \mathcal{O}(1)$$

and an explicit calculation yields

$$\begin{aligned}
\mathbf{E}(\|\varepsilon_i\|_{L^2}^4) &= \frac{\sigma^4}{\Delta^4}\mathbf{E}\left(\Big(\sum_{\ell\geq 1}\Big(\int_{t_i}^{t_{i+1}} \mathrm{e}^{-\lambda_\ell(t_{i+1}-s)}\, d\beta_\ell(s)\Big)^2\Big)^2\right)\\
&= \frac{\sigma^4}{\Delta^4}\sum_{\ell,\ell'\geq 1}\mathbf{E}\left(\Big(\int_{t_i}^{t_{i+1}} \mathrm{e}^{-\lambda_\ell(t_{i+1}-s)}\, d\beta_\ell(s)\Big)^2\Big(\int_{t_i}^{t_{i+1}} \mathrm{e}^{-\lambda_{\ell'}(t_{i+1}-s)}\, d\beta_{\ell'}(s)\Big)^2\right)\\
&\lesssim \frac{1}{\Delta^4}\Big(\sum_{\ell\geq 1}\frac{1-\mathrm{e}^{-2\lambda_\ell\Delta}}{2\lambda_\ell}\Big)^2 = \mathcal{O}(\Delta^{-3}).
\end{aligned}$$

Gathering bounds, we have shown

$$\mathbf{E}\big(\|\hat{f}_m - f_A\|_N^2 \mathbf{1}_{\Xi_{N,m}^c}\big) \lesssim \Delta^{-3/2}\mathbf{P}(\Xi_{N,m}^c)^{1/2}.$$

Using Lemma 4.4.5 in view of $D_m \leq N$, we get

$$\mathbf{P}(\Xi_{N,m}^c) \lesssim N\exp(-\gamma q_N\Delta) + N^2\exp\Big(-K'\frac{N}{2q_N L_m}\Big).$$

Under the condition $\frac{N\Delta}{\log^2 N} \to \infty$ it is possible to choose (q_N) such that $q_N/(\frac{\nu\log N}{\Delta}) \to 1$ for some fixed $\nu > 0$. Then, we have

$$N\exp(-\gamma q_N\Delta) \leq N\exp(-\frac{\gamma\nu}{2}\log(N)) = N^{-(\frac{\gamma\nu}{2}-1)}.$$

Further, since $L_m = o(\frac{N\Delta}{\log^2 N})$, for any $\beta > 0$ we have $L_m \leq \beta\frac{N\Delta}{\log^2 N}$ for N sufficiently large, implying

$$K'\frac{N}{2q_N L_m} \geq K'\frac{N\Delta}{4\nu\log N L_m} \geq \frac{K'}{4\nu\beta}\log(N)$$

as well as

$$N^2\exp\Big(-K'\frac{N}{2q_N L_m}\Big) \leq N^{2-\frac{K'}{4\nu\beta}}.$$

Hence, for arbitrary $\alpha > 0$, we can choose ν sufficiently large and β sufficiently small such that $\mathbf{P}(\Xi_{N,m}^c) \leq N^{-(2\alpha+3)}$ and, thus,

$$\mathbf{E}\big(\|\hat{f}_m - f_A\|_N^2 \mathbf{1}_{\Omega_{N,m}^c}\big) \lesssim \Delta^{-3/2}\mathbf{P}(\Xi_{N,m}^c)^{1/2} \lesssim \frac{1}{T^{3/2}N^\alpha} \lesssim \frac{1}{N^\alpha} = \frac{\Delta^\alpha}{T^\alpha} = o(\Delta^\alpha).$$

From here, the claim follows by choosing $\alpha = \gamma$. □

We prove our final result for the space-continuous observation scheme, namely the bound on the L^2-risk.

Proof of Corollary 4.4.6. First, we prove the bound in probability: For any $a > \|f\|_{L^2(A)}^2$, the triangle inequality yields

$$\begin{aligned}
\mathbf{P}\big(\|\hat{f}_m - f\|_{L^2(A)}^2 \geq 2a\big) &\leq \mathbf{P}\big(\|\hat{f}_m - f_m\|_{L^2(A)}^2 \geq a - \|f_m - f\|_{L^2(A)}^2\big)\\
&\leq \mathbf{P}\big(\mathbf{1}_{\Xi_{N,m}}\|\hat{f}_m - f_m\|_{L^2(A)}^2 \geq a - \|f_m - f\|_{L^2(A)}^2\big) + \mathbf{P}\big(\Xi_{N,m}^c\big).
\end{aligned}$$

Now, as in the proof of Theorem 4.4.2, we have $\|t\|_N^2 \geq \frac{1}{2}\|t\|_\pi^2 \geq \frac{c}{2}\|t\|_{L^2(A)}^2$ on $\Xi_{N,m}$ for all $t \in S_m$ where $c > 0$ is the constant from the equivalence condition (4.14). Thus, using $\hat{f}_m - f_m \in S_m$ as well as Markov's inequality and the triangle inequality,

$$\mathbf{P}\big(\mathbf{1}_{\Xi_{N,m}}\|\hat{f}_m - f_m\|_{L^2(A)}^2 \geq a - \|f_m - f\|_{L^2(A)}^2\big) \leq \mathbf{P}\big(\mathbf{1}_{\Xi_{N,m}}\|\hat{f}_m - f_m\|_N^2 \geq \frac{c}{2}(a - \|f_m - f\|_{L^2(A)}^2)\big)$$

$$\leq \frac{2\mathbf{E}\big(\mathbf{1}_{\Xi_{N,m}}\|\hat{f}_m - f_m\|_N^2\big)}{c(a - \|f_m - f\|_{L^2(A)}^2)}$$
$$\leq \frac{4\mathbf{E}\big(\mathbf{1}_{\Xi_{N,m}}\|\hat{f}_m - f_A\|_N^2\big) + 4\|f_m - f_A\|_\pi^2}{c(a - \|f_m - f\|_{L^2(A)}^2)}.$$

Now, we set $a := K\Big(\|f - f_m\|_{L^2(A)}^2 + D_m/T + \Delta^\gamma\Big)$ for some $K > 1$. By Proposition 4.4.4 and Assumption (E), the above bound can be estimated up to a constant by

$$\frac{\|f_m - f\|_{L^2(A)}^2 + D_m/T + \Delta^\gamma}{a - \|f_m - f\|_{L^2(A)}^2} \leq \frac{\|f_m - f\|_{L^2(A)}^2 + D_m/T + \Delta^\gamma}{(K-1)\Big(\|f - f_m\|_{L^2(A)}^2 + D_m/T + \Delta^\gamma\Big)} = \frac{1}{K-1}.$$

Using $\mathbf{P}(\Xi_{N,m}^c) \lesssim N^{-\alpha}$ for any power $\alpha > 0$ from the proof of Theorem 4.4.2, we can conclude that for any $\varepsilon > 0$ there exists $K > 0$ such that

$$\limsup_{N\to\infty} \mathbf{P}\big(\|\hat{f}_m - f\|_{L^2(A)}^2 \geq 2K\big(\|f_A - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma\big)\big) \leq \varepsilon$$

which verifies the claimed bound in probability.

Next, we consider the truncated estimator $\hat{f}_m^{K_N}$: We have

$$\|\hat{f}_m^{K_N} - f\|_{L^2(A)}^2 \leq \|\hat{f}_m^{K_N} - f\|_{L^2(A)}^2 \mathbf{1}_{\Xi_{N,m}} + 2(\|f\|_{L^\infty(A)}^2 + K_N^2)\mathbf{1}_{\Xi_{N,m}^c}$$

and, thus, as soon as $K_N \geq \|f\|_{L^\infty(A)}$, we can further bound

$$\|\hat{f}_m^{K_N} - f\|_{L^2(A)}^2 \leq \|\hat{f}_m - f\|_{L^2(A)}^2 \mathbf{1}_{\Xi_{N,m}} + 4K_N^2 \mathbf{1}_{\Xi_{N,m}^c}.$$

For the expectation of the first term, we get like in the above derivation of the bound in probability that $\mathbf{E}(\|\hat{f}_m - f\|_{L^2(A)}^2 \mathbf{1}_{\Xi_{N,m}}) \lesssim \|f_A - f_m\|_{L^2(A)}^2 + D_m/T + \Delta^\gamma$. The expectation of the second term, $4K_N^2\mathbf{P}(\Xi_{N,m}^c)$, decreases faster than any power of N, thanks to Lemma 4.4.5 and the growth assumption on K_N. Thus, $4K_N^2\mathbf{P}(\Xi_{N,m}^c) \lesssim N^{-\gamma} \lesssim \Delta^\gamma$, which finishes the proof. □

Fully discrete observations

In this section, we prove Theorems 4.4.9 and 4.4.12. The proof for Corollary 4.4.11 follows exactly the same arguments as for the space-continuous case and is omitted. Further technical Lemmas and a proof of Lemma 4.4.7 are postponed to Section 4.5.4.

Before proving the main theorem for fully discrete observations, we state and prove the discrete analogons of Proposition 4.4.4 and Lemma 4.4.5. To that aim, we define

$$\|g\|_{\pi,M}^2 := \frac{1}{M}\sum_{k=1}^{M-1} \mathbf{E}\Big(g^2(X_0(y_k))\Big)$$

for nonrandom $g \in L^2(A)$. Again, we have

$$c\|g\|_{L^2(A)}^2 \leq \|g\|_{\pi,M}^2 \leq C\|g\|_{L^2(A)}^2$$

for all $g \in L^2(A)$ and some constants $c, C > 0$ under Assumption (E).

Proposition 4.5.3. *Grant Assumptions (M), (N) and (H). For $\underline{c} > 0$, define*

$$\Omega_{N,M,m} := \Omega_{N,M,m,\underline{c}} := \left\{\|t\|_{N,M}^2 \geq \underline{c}\|t\|_{L^2}^2 \text{ for all } t \in S_m\right\}.$$

Then, for any $\gamma < 1/2$ and $\rho < 1/4$, we have

$$\mathbf{E}\Big(\|\hat{f}_m - f_A\|_{N,M}^2 \mathbf{1}_{\Omega_{N,M,m}}\Big) \lesssim \|f_A - f_m\|_{\pi,M}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho}.$$

Proof. Recall the underlying regression model from (4.16). By Lemma 4.4.7, we have $\|g\|_{N,M}^2 = \frac{1}{N}\sum_{i=0}^{N-1} \|\hat{S}(0)g(X_{t_i})\|_{L^2}^2$ for any $g \in S_m$ and, thus, we can derive the basic inequality

$$\|\hat{f}_m - f_A\|_{N,M}^2 \le 3\|f_m - f_A\|_{N,M}^2 + \frac{16}{N}\sum_{i=0}^{N-1} \|\tilde{R}_i\|_{L^2}^2 + 16 \sup_{t\in S_m,\, \|t\|_{N,M}=1} \Big(\frac{1}{N}\sum_{i=0}^{N-1} \Big\langle \hat{S}(0)t(X_i), \varepsilon_i\Big\rangle_{L^2}\Big)^2$$

just as in the proof for the space-continuous case. Also the variance term can be handled analogously: With an orthonormal basis $(\varphi_\lambda,\ \lambda \in \Lambda_m)$ of S_m, the same line of arguments as in the proof Proposition 4.4.4 leads to

$$\mathbf{E}\Big(\sup_{t\in S_m,\, \|t\|_{N,M}=1} \Big(\frac{1}{N}\sum_{i=0}^{N-1} \Big\langle \hat{S}(0)t(X_i), \varepsilon_i\Big\rangle_{L^2}\Big)^2 \mathbf{1}_{\Omega_{N,M,m}}\Big) \lesssim \frac{1}{N^2\Delta}\sum_{\lambda\in\Lambda_m}\sum_{i=0}^{N-1} \mathbf{E}(\|\hat{S}(0)\varphi_\lambda(X_{t_i})\|_{L^2}^2).$$

Further, since $\|\hat{S}(0)\varphi_\lambda(X_{t_i})\|_{L^2}^2 = \frac{1}{M}\sum_{k=1}^{M-1} \varphi_\lambda^2(X_{t_i}(y_k))$, we get $\sum_{\lambda\in\Lambda_m} \|\hat{S}(0)\varphi_\lambda(X_{t_i})\|_{L^2}^2 \le D_m$ by Assumption (N). Consequently,

$$\mathbf{E}\Big(\sup_{t\in S_m,\, \|t\|_{N,M}=1} \Big(\frac{1}{N}\sum_{i=0}^{N-1} \Big\langle \hat{S}(0)t(X_i), \varepsilon_i\Big\rangle_{L^2}\Big)^2 \mathbf{1}_{\Omega_{N,M,m}}\Big) \lesssim \frac{D_m}{T}.$$

We finish the proof by showing that

$$\mathbf{E}(\|\tilde{R}_i\|_{L^2}^2) \lesssim \frac{1}{M\Delta^2} + \frac{1}{M^\rho} + \Delta^\gamma \tag{4.31}$$

holds for any $\rho < 1/4$ and $\gamma < 1/2$: First of all, by Lemma 4.5.7, we have

$$\mathbf{E}(\|f(X_t) - \hat{S}(0)f(X_t)\|_{L^2}^2) \lesssim \mathbf{E}(\|f(X_t)\|_{C^{2\alpha}}^2 + \|f(X_t)\|_\infty^2 + \|f(X_t)\|_{D_\alpha}^2)\delta^{\frac{8\alpha^2}{4\alpha+1}}$$

with the space D_α defined in (4.34). The expectation on the right hand side is finite as long as $\alpha < 1/4$, due to Lemma 4.5.5 and Proposition 4.2.1. Thus, by picking α sufficiently close to 1/4, we get

$$\mathbf{E}(\|f(X_t) - \hat{S}(0)f(X_t)\|_{L^2}^2) \lesssim \delta^\rho = M^{-\rho}.$$

To bound $\Delta^{-2}\mathbf{E}(\|S(h)X_t - \hat{S}(h)X_t\|_{L^2}^2)$ for $h \in \{0, \Delta\}$, we use the usual decomposition $X_t = S(t)X_0 + X_t^0 + N_t$ where we can fix a convenient value for $t > 0$, due to stationarity. Since the decomposition is trivial for $t = 0$, we pick $t := t_1 = \Delta$. For $S(t)X_0$, we have

$$\begin{aligned}\|S(h)S(t)X_0 - \hat{S}(h)S(t)X_0\|_{L^2}^2 = &\sum_{k=1}^{M-1} \mathrm{e}^{-2\lambda_k h}\Big(\langle S(t)X_0, e_k\rangle_{L^2} - \langle S(t)X_0, e_k\rangle_M\Big)^2 \\ &+ \sum_{k\ge M} \mathrm{e}^{-2\lambda_k(h+t)}\langle X_0, e_k\rangle_{L^2}^2.\end{aligned}$$

For the first sum, the Cauchy-Schwarz inequality yields

$$\begin{aligned}\Big(\langle S(t)X_0, e_k\rangle_{L^2} - \langle S(t)X_0, e_k\rangle_M\Big)^2 =&\Big(\sum_{l\in\mathcal{I}_k^+\setminus\{k\}} \mathrm{e}^{-\lambda_l t}\langle X_0, e_l\rangle_{L^2} - \sum_{l\in\mathcal{I}_k^-} \mathrm{e}^{-\lambda_l t}\langle X_0, e_l\rangle_{L^2}\Big)^2 \\ \le& \|X_0\|_{L^2}^2 \sum_{l\in(\mathcal{I}_k^+\cup\mathcal{I}_k^-)\setminus\{k\}} \mathrm{e}^{-2\lambda_l t}\end{aligned}$$

and, thus, with $t = \Delta$,

$$\sum_{k=1}^{M-1} \mathrm{e}^{-2\lambda_k h}\Big(\langle S(t_1)X_0, e_k\rangle_{L^2} - \langle S(t_1)X_0, e_k\rangle_M\Big)^2$$

$$\leq \|X_0\|_{L^2}^2 \sum_{l\geq M} \mathrm{e}^{-2\lambda_l\Delta} \leq \|X_0\|_{L^2}^2 \frac{1}{\sqrt{\Delta}} \int_{M\sqrt{\Delta}}^{\infty} \mathrm{e}^{-2\pi^2\vartheta x^2}\,dx \lesssim \|X_0\|_{L^2}^2 \frac{1}{M^2\Delta^{3/2}}.$$

The same bound holds for the second sum since

$$\sum_{k\geq M} \mathrm{e}^{-2\lambda_k(h+\Delta)} \langle X_0, e_k\rangle_{L^2}^2 \leq \|X_0\|_{L^2}^2 \mathrm{e}^{-2\lambda_M\Delta} \lesssim \|X_0\|_{L^2}^2 \frac{1}{M^2\Delta} \lesssim \|X_0\|_{L^2}^2 \frac{1}{M^2\Delta^{3/2}}.$$

Therefore, assuming $M\Delta^2 \to \infty$, we get

$$\Delta^{-2}\mathbf{E}(\|S(h)S(t_1)X_0 - \hat{S}(h)S(t_1)X_0\|_{L^2}^2) \lesssim \mathbf{E}(\|X_0\|_{L^2}^2) \frac{1}{M^2\Delta^{7/2}} = o\Big(\frac{1}{M\Delta^2}\Big).$$

The linear component X_t^0 can easily be treated due to independence:

$$\begin{aligned}\mathbf{E}(\|S(h)X_t^0 - \hat{S}(h)X_t^0\|_{L^2}^2) &= \mathbf{E}\Big(\sum_{k=1}^{M-1} \mathrm{e}^{-2\lambda_k h}\Big(\sum_{\ell\in\mathcal{I}_k^+\setminus\{k\}} u_\ell(t) - \sum_{\ell\in\mathcal{I}_k^-} u_\ell(t)\Big)^2\Big) + \mathbf{E}\Big(\sum_{\ell\geq M} \mathrm{e}^{-2\lambda_\ell h} u_\ell^2(t)\Big)\\ &\leq 2\sum_{\ell\geq M} \mathbf{E}(u_\ell^2(t)) \lesssim \frac{1}{M}.\end{aligned}$$

For the nonlinear part, set $B_k := \sum_{\ell\in\mathcal{I}_k^+\setminus\{k\}} n_\ell(t) - \sum_{\ell\in\mathcal{I}_k^-} n_\ell(t)$ with $n_\ell(t) := \langle N_t, e_\ell\rangle_{L^2}$. Then, by the Cauchy-Schwarz inequality,

$$B_k^2 \leq \Big(\sum_{\ell\in(\mathcal{I}_k^+\cup\mathcal{I}_k^-)\setminus\{k\}} \lambda_\ell^{2\alpha} n_\ell^2\Big)\Big(\sum_{\ell\in(\mathcal{I}_k^+\cup\mathcal{I}_k^-)\setminus\{k\}} \lambda_\ell^{-2\alpha}\Big) \leq \|N_t\|_{D_\alpha}^2 \Big(\sum_{\ell\in(\mathcal{I}_k^+\cup\mathcal{I}_k^-)\setminus\{k\}} \lambda_\ell^{-2\alpha}\Big).$$

Since, furthermore, $\sum_{\ell\geq M} n_\ell^2(t) \leq \lambda_M^{-2\alpha}\|N_t\|_{D_\alpha}^2$, we have

$$\begin{aligned}\mathbf{E}(\|S(h)N_t - \hat{S}(h)N_t\|_{L^2}^2) &\leq \mathbf{E}\Big(\sum_{k=1}^{M-1} B_k^2\Big) + \mathbf{E}\Big(\sum_{\ell\geq M} n_\ell^2(t)\Big) \lesssim \mathbf{E}(\|N_t\|_{D_\alpha}^2)\Big(\sum_{k\geq M} \lambda_k^{-2\alpha} + \lambda_M^{-2\alpha}\Big)\\ &\lesssim \frac{1}{M^{4\alpha-1}}\mathbf{E}(\|N_t\|_{D_\alpha}^2).\end{aligned}$$

Now, by Remark 4.5.6, we have $\mathbf{E}(\|N_t\|_{D_\alpha}^2) < \infty$ for $\alpha = 1/2$ and, thus, $\mathbf{E}(\|S(h)N_t - \hat{S}(h)N_t\|_{L^2}^2) \lesssim \frac{1}{M}$. Finally, the bound (4.31) follows in view of

$$\mathbf{E}\left(\Big\|\frac{1}{\Delta}\int_{t_i}^{t_{i+1}} \Big(S(t_{i+1}-s)f(X_s) - f(X_{t_i})\Big)\,ds\Big\|_{L^2}^2\right) \lesssim \Delta^\gamma$$

for any $\gamma < 1/2$, which is shown in the proof of Proposition 4.4.4. □

The remaining steps of the proof closely follow the space-continuous case.

Lemma 4.5.4. *Grant Assumptions (M), (E), (N) and (H) and and let*

$$\Xi_{N,M,m} := \left\{\Big|\frac{\|t\|_{N,M}^2}{\|t\|_{\pi,M}^2} - 1\Big| \leq \frac{1}{2}\,\forall t\in S_m\right\}.$$

There are constants $K, K' > 0$ such that

$$\mathbf{P}(\Xi_{N,M,m}^c) \leq K\Big(N\beta_X(q_N\Delta) + D_m^2 \exp\Big(-K'\frac{p_N}{L_m}\Big)\Big)$$

holds for any $p_N, q_N \in \mathbb{N}$ with $N = 2p_Nq_N$. In particular, with the constants γ and L from the β-mixing condition (4.4) as well as $\tilde{K} := K\max(L,1)$, we have

$$\mathbf{P}(\Xi_{N,M,m}^c) \leq \tilde{K}\Big(N\exp\big(-\gamma q_N\Delta\big) + D_m^2\exp\Big(-K'\frac{p_N}{L_m}\Big)\Big).$$

Proof. After replacing integrals by their empirical counterpart, the proof can be carried out in exactly the same way as for space-continuous observations. In particular, $v_N(t)$ has to be replaced by

$$v_{N,M}(t) := \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{k=1}^{M-1}\big(t(X_{i\Delta}(y_k)) - \mathbf{E}\big(t(X_{i\Delta}(y_k))\big)\big)$$

and $Z_{j,a}(t)$ by

$$\tilde{Z}_{j,a}(t) := \frac{1}{q_N}\sum_{i=1}^{q_N}\frac{1}{M}\sum_{k=1}^{M-1} t(U^i_{j,a}(y_k)).$$

□

Now, we can finish the proof for the main theorem on fully discrete observations.

Proof of Theorem 4.4.9. As in the space-continuous case, we prove that

$$\mathbf{E}(\|\hat{f}_m - f_A\|^2_{N,M}\mathbf{1}_{\Xi^c_{N,M,m}}) \lesssim N^{-\alpha}$$

holds for any $\alpha > 0$. Since $\hat{f}_m = \arg\min_{g\in S_m}\frac{1}{N}\sum_{i=0}^{N-1}\|\tilde{Y}_i - \hat{S}(0)g(X_{t_i})\|^2_{L^2}$ with $\tilde{Y}_i := \Delta^{-1}\big(\hat{S}(0)X_{t_{i+1}} - \hat{S}(\Delta)X_{t_i}\big)$, we have that $(\hat{S}(0)\hat{f}_m(X_{t_0}),\dots,\hat{S}(0)\hat{f}_m(X_{t_{N-1}}))$ is the orthogonal projection in $H^N = \big(L^2(0,1)\big)^N$ of $(\tilde{Y}_0,\dots,\tilde{Y}_{N-1})$ onto the subspace $\{(\hat{S}(0)t(X_{t_0}),\dots,\hat{S}(0)t(X_{t_N})),\ t\in S_m\}$. Using the fact that projection operators have norm 1 and inserting $\tilde{Y}_i = \hat{S}(0)f(X_{t_i}) + \tilde{R}_i + \varepsilon_i$, we get

$$\begin{aligned}
\|\hat{f}_m - f_A\|^2_{N,M} &\le \frac{1}{N}\sum_{i=0}^{N-1}\|\hat{S}(0)\hat{f}_m(X_{t_i}) - \hat{S}(0)f(X_{t_i})\|^2_{L^2}\\
&\le \frac{1}{N}\sum_{i=0}^{N-1}\|\hat{S}(0)f(X_{t_i})\|^2_{L^2} + \frac{1}{N}\sum_{i=0}^{N-1}\|\tilde{R}_i + \varepsilon_i\|^2_{L^2}\\
&\lesssim \frac{1}{N}\sum_{i=0}^{N-1}\|f(X_{t_i})\|^2_{\infty} + \frac{1}{N}\sum_{i=0}^{N-1}\|\tilde{R}_i\|^2_{L^2} + \frac{1}{N}\sum_{i=0}^{N-1}\|\varepsilon_i\|^2_{L^2}.
\end{aligned}$$

From here, the conclusion follows in exactly the same way as in the proof of Theorem 4.4.2. □

The following proof verifies that the rate of convergence is not affected when the parameter ϑ_2 appearing in the estimator for f is replaced by an appropriate estimator.

Proof of Theorem 4.4.12. We verify the bound for $\|\check{f}_m - f_A\|^2_{N,M}$, the bound for $\|\check{f}_m - f_A\|^2_{L^2(A)}$ then follows as in the proof of Corollary 4.4.6. We define

$$\Psi^h_{N,M} := \Big\{(\hat{\vartheta}_2 - \vartheta_2)^2 \le h\frac{\Delta^{3/2}}{T}\Big\}.$$

Step 1: We show that

$$\mathbf{E}(\mathbf{1}_{\Psi^h_{N,M}}\Delta^{-2}\|\hat{S}(\Delta)X_{t_i} - \check{S}(\Delta)X_{t_i}\|^2_{L^2}) \lesssim \frac{h}{T}:$$

For fixed $\underline{\vartheta}_2 \in (0,\vartheta_2)$, we have $\hat{\vartheta}_2 \ge \underline{\vartheta}_2$ on the event $\Psi^h_{N,M}$ as soon as T is sufficiently large. Thus, we can estimate

$$\|\hat{S}(\Delta)X_{t_i} - \check{S}(\Delta)X_{t_i}\|^2_{L^2} = \sum_{k=1}^{M-1}(\mathrm{e}^{-\lambda_k\Delta} - \mathrm{e}^{-\hat{\lambda}_k\Delta})^2\langle X_{t_i}, e_k\rangle^2_M$$

$$\lesssim (\vartheta_2 - \hat{\vartheta}_2)^2 \Delta^2 \sum_{k=1}^{M-1} \lambda_k^2 \mathrm{e}^{-\underline{\vartheta}_2 \pi^2 k^2 \Delta} \langle X_{t_i}, e_k \rangle_M^2$$
$$\leq h \frac{\Delta^{7/2}}{T} \sum_{k=1}^{M-1} \lambda_k^2 \mathrm{e}^{-\underline{\vartheta}_2 \pi^2 k^2 \Delta} \langle X_{t_i}, e_k \rangle_M^2$$

on $\Psi_{N,M}^h$. Therefore,

$$\mathbf{E}(\mathbf{1}_{\Psi_{N,M}^h} \|\hat{S}(\Delta) X_{t_i} - \check{S}(\Delta) X_{t_i}\|_{L^2}^2) \lesssim h \frac{\Delta^{7/2}}{T} \sum_{k=1}^{M-1} \lambda_k^2 \mathrm{e}^{-\underline{\vartheta}_2 \pi^2 k^2 \Delta} \mathbf{E}(\langle X_{t_i}, e_k \rangle_M^2)$$

and the claim follows from the usual Riemann sum argument if we show that $\mathbf{E}(\langle X_{t_i}, e_k \rangle_M^2) \lesssim \lambda_k^{-1}$. To that aim, we apply the decomposition $X_t = S(t)\xi + X_t^0 + N_t$. As in previous results, $S(t)\xi$ is negligible since we can choose t arbitrarily large due to stationarity. For the linear part, we have

$$\mathbf{E}(\langle X_{t_i}^0, e_k \rangle_M^2) \lesssim \sum_{\ell \in \mathcal{I}_k^+ \cup \mathcal{I}_k^-} \frac{1}{\lambda_\ell} \lesssim \sum_{\ell \geq 0} \frac{1}{(k + 2\ell M)^2} \leq \frac{1}{k^2} \sum_{\ell \geq 0} \frac{1}{(1 + 2\ell)^2} \lesssim \frac{1}{\lambda_k}.$$

Finally, for the nonlinear part, define $n_\ell(t) := \langle N_t, e_\ell \rangle_{L^2}$. Then, using the Cauchy-Schwarz inequality and the spaces D_α from (4.34), we have

$$\langle N_t, e_k \rangle_M^2 \leq \Big(\sum_{\ell \in (\mathcal{I}_k^+ \cup \mathcal{I}_k^-)} \lambda_\ell^{2\alpha} n_\ell^2 \Big) \Big(\sum_{\ell \in (\mathcal{I}_k^+ \cup \mathcal{I}_k^-)} \lambda_\ell^{-2\alpha} \Big) \leq \|N_t\|_{D_\alpha}^2 \Big(\sum_{\ell \in (\mathcal{I}_k^+ \cup \mathcal{I}_k^-)} \lambda_\ell^{-2\alpha} \Big)$$
$$\leq \frac{1}{\lambda_k} \|N_t\|_{D_\alpha}^2 \Big(\sum_{\ell \geq 1} \lambda_\ell^{-(2\alpha - 1)} \Big) \lesssim \frac{1}{\lambda_k} \|N_t\|_{D_\alpha}^2,$$

provided that $\alpha > 3/4$. Now, by picking $\alpha \in (\frac{3}{4}, 1)$, we get $\mathbf{E}(\langle N_t, e_k \rangle_M^2) \lesssim \lambda_k^{-1}$ in view of Remark 4.5.6.

Step 2: By Markov's inequality, we can estimate

$$\mathbf{P}\Big(\|\check{f}_m - f_A\|_{N,M}^2 \geq a \Big) \leq \mathbf{P}\Big(\{ \|\check{f}_m - f_A\|_{N,M}^2 \geq a \} \cap \Psi_{N,M}^h \cap \Xi_{N,M,m} \Big) + \mathbf{P}((\Psi_{N,M}^h)^c) + \mathbf{P}(\Xi_{N,M,m}^c)$$
$$\leq a^{-1} \mathbf{E}\Big(\mathbf{1}_{\Psi_{N,M}^h \cap \Xi_{N,M,m}} \|\check{f}_m - f_A\|_{N,M}^2 \Big) + \mathbf{P}((\Psi_{N,M}^h)^c) + \mathbf{P}(\Xi_{N,M,m}^c)$$

for any $a > 0$. Now, using Step 1, we can show

$$\mathbf{E}\Big(\mathbf{1}_{\Psi_{N,M}^h \cap \Xi_{N,M,m}} \|\check{f}_m - f_A\|_{N,M}^2 \Big) \lesssim \|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho} + \frac{h}{T}$$

just like in the proof of Proposition 4.5.3. Further, $\mathbf{P}(\Xi_{N,M,m}^c)$ converges to 0 under the assumptions of this theorem and, thanks to Corollary 2.2.19, $\mathbf{P}((\Psi_{N,M}^h)^c)$ can be made arbitrarily small by choosing h sufficiently large. Since $h/T \lesssim D_m/T$ holds for any fixed h, we have shown that for arbitrary $\varepsilon > 0$, we can pick $K > 0$ such that

$$\limsup_{M,N \to \infty} \mathbf{P}\Big(\|\check{f}_m - f_A\|_{N,M}^2 \geq K\big(\|f - f_m\|_{L^2(A)}^2 + \frac{D_m}{T} + \Delta^\gamma + \frac{1}{M\Delta^2} + \frac{1}{M^\rho} \big) \Big) < \varepsilon. \qquad \square$$

4.5.4 Further proofs and auxiliary results

Before turning to the auxiliary results for nonparametric estimation of the nonlinearity f, we prove that condition (4.5) implies Assumptions (B) and (M).

Proof of Proposition 4.1.1. First, we sketch the existence proof and show that Assumption (B) is satisfied for $\xi = 0$. To that aim, we follow the line of arguments from [26, Theorem 7.7], see also [34, Propositon 6.1]. As before, write $m \equiv f(0)$ as well as $f_0(x) = f(x) - m$ and decompose $X_t = w(t) + v(t)$ with $w(t) := X_t^0 + \int_0^t S(r)m\,dr$ and $v(t) := S(t)\xi + \int_0^t S(t-s)f_0(X_s)\,dt$. It follows from Lemma 4.5.1 and $\|S(r)m\|_\infty \lesssim \mathrm{e}^{-\lambda_0 r}\|m\|_\infty$ (cf. Lemma 4.2.2) that $w \in C(\mathbb{R}_+, E)$ holds almost surely and

$$\sup_{t\geq 0} \mathbf{E}(\|w_t\|_\infty^p) < \infty. \tag{4.32}$$

As in Theorem 1.1.1, it follows from the fact that $F_0(u) := f_0 \circ u$ is a locally Lipschitz continuous function from E into itself, that there exists a solution to equation (4.1) up to a terminal time $t_{\max} = t_{\max}(\omega) > 0$. Thus, global existence follows from an a-priori estimate on $\|v(\cdot)\|_\infty$. We consider the approximation $v_n := nR(n, A_\vartheta)S(t)\xi + \int_0^t nR(n, A_\vartheta)S(t-s)f_0(v(s) + w(s))\,ds$ where $R(n, A_\vartheta) := (nI - A_\vartheta)^{-1}$ is the resolvent operator of A_ϑ. Then, v_n is differentiable in time, even when v is not. Now, for any $x \in E$ and $x^* \in \partial\|x\|$, it follows like in [26, Example 7.8] from the condition (N4) in Section 1.1 that $\langle A_\vartheta x, x^*\rangle \leq 0$ where $\partial\|x\|$ is the subdifferential of the norm defined in (1.6). Thus, setting $\delta_n(t) := v_n'(t) - A_\vartheta v_n - f(v_n(t) + w(t))$ and using the bound (1.7) with $h_{v_n(t)} \in \partial\|v_n(t)\|$ from (1.9), we can estimate

$$\begin{aligned}
\frac{d^-}{dt}\|v_n(t)\|_\infty &\leq \langle \frac{d}{dt}v_n(t), h_{v_n(t)}\rangle = \langle A_\vartheta v_n(t), h_{v_n(t)}\rangle + \langle f(v_n(t) + w(t)), h_{v_n(t)}\rangle + \langle \delta_n(t), h_{v(t)}\rangle \\
&\leq \langle f(v_n(t) + w(t)), h_{v_n(t)}\rangle + \|\delta_n(t)\|_\infty \\
&\leq -a\|v_n(t)\|_\infty + b\|w(t)\|_\infty^\beta + c + \|\delta_n(t)\|_\infty.
\end{aligned}$$

Using Gronwall's inequality and the fact that $v_n(t) \to v(t)$ and $\delta_n(t) \to 0$ uniformly on compact time intervals yields

$$\|v(t)\|_\infty \leq \mathrm{e}^{-at}\|\xi\|_\infty + \int_0^t \mathrm{e}^{-a(t-s)}(b\|w(s)\|_\infty^\beta + c)\,ds.$$

By Jensen's inequality, we pass to

$$\|v(t)\|_\infty^p \lesssim e^{-apt}\|\xi\|_\infty^p + \int_0^t \mathrm{e}^{-a(t-s)}(b\|w(t)\|_\infty^\beta + c)^p\,ds \cdot \left(\int_0^t \mathrm{e}^{-as}\,ds\right)^{p-1}$$

and Fubini's theorem as well as (4.32) show that there exists $K > 0$ such that for nonrandom initial conditions $\xi = x \in E$, we have

$$\mathbf{E}(\|X_t\|_\infty^p) \lesssim \mathrm{e}^{-apt}\|x\|_\infty^p + K. \tag{4.33}$$

In particular, Assumption (B) with $\xi = 0$ is satisfied.

Further, based on their derivation of lower bounds for the transition densities associated with the Markov semigroup (P_t), Goldys and Maslowski [34, Theorem 6.3] show the existence of an invariant measure π on E and of constants $C, \gamma > 0$ such that

$$\|P_t^*\nu - \pi\|_{\mathrm{TV}} \leq C\Big(\int_E \|u\|_\infty\,\nu(du) + 1\Big)\mathrm{e}^{-\gamma t} \quad \text{with} \quad P_t^*\nu := \int_E P_t(u, \cdot)\,\nu(du)$$

holds for all probability measures ν on E. Thus, we have $\|P_t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq C(\|x\|_\infty + 1)\mathrm{e}^{-\gamma t}$ and $P_t(x, \cdot)$ converges weakly to $\pi(\cdot)$ as $t \to \infty$ for all $x \in E$. By Skorokhod's representation theorem, there exists a probability space on which there are E-valued random variables $Z, Z_1, Z_2, \ldots$ with $Z_i \sim P_i(x, \cdot)$, $Z \sim \pi$ and $Z_i \to Z$ almost surely. Denoting the expectation on the second probability space by $\tilde{\mathbf{E}}$, Fatou's Lemma yields

$$\int_E \|u\|_\infty^p\,\pi(du) = \tilde{\mathbf{E}}(\|Z\|_\infty^p) \leq \liminf_{i\to\infty} \tilde{\mathbf{E}}(\|Z_i\|_\infty^p) = \liminf_{i\to\infty} \int_E \|u\|_\infty^p\,P_i(x, du) < \infty$$

by (4.33). Thus, if $X_0 = \xi \sim \pi$, then $\mathbf{E}(\|X_t\|_\infty^p) = \mathbf{E}(\|X_0\|_\infty^p) < \infty$ and

$$\int_E \|P_t(u,\cdot) - \pi\|_{\mathrm{TV}}\,\pi(du) \le C\Big(\int_E \|u\|_\infty\,\pi(du) + 1\Big)\mathrm{e}^{-\gamma t} \lesssim \mathrm{e}^{-\gamma t},$$

as required for (M) as well as (B) in case of a stationary initial condition. □

Technical Lemmas for the nonparametric estimator of f

The following proof verifies the connection between the L^2-norm on $[0,1]$ and its empirical counterpart.

Proof of Lemma 4.4.7. In view of Dini's test, the Hölder condition implies convergence of the Fourier series of H at the points y_k, i.e., $\bar{H}^n(y_k) := \sum_{l=1}^n h_l e_l(y_k) \to H(y_k)$ as $n \to \infty$ for any $1 \le k \le M-1$. Therefore, $|\langle H, e_k\rangle_M - \langle \bar{H}^n, e_k\rangle_M| \le \frac{1}{M}\sum_{l=1}^{M-1} |H(y_l) - \bar{H}^n(y_l)||e_k(y_l)|$ tends to 0 as $n \to \infty$. Hence, the sequence $\langle \bar{H}^n, e_k\rangle_M = \sum_{l\in\mathcal{I}_k^+\cap[1,n]} h_l - \sum_{l\in\mathcal{I}_k^-\cap[1,n]} h_l$ converges to the limit $\langle H, e_k\rangle_M$, proving the first part of the lemma. In the same way, using $e_l(y_k) = \pm e_j(y_k)$ for $l \in \mathcal{I}_j^{\pm}$, one can show that $H(y_k) = \sum_{l=1}^{M-1} H_l e_l(y_k)$. Consequently,

$$\frac{1}{M}\sum_{k=1}^{M-1} H^2(y_k) = \frac{1}{M}\sum_{k=1}^{M-1}\Big(\sum_{l=1}^{M-1} H_l e_l(y_k)\Big)^2 = \sum_{l,l'=1}^{M-1} H_l H_{l'}\langle e_l, e_{l'}\rangle_M = \sum_{l=1}^{M-1} H_l^2 = \|H^M\|_{L^2}^2. \qquad □$$

The following lemma analyzes the regularity of X_t in the spaces

$$D_\varepsilon := \mathcal{D}((-A_\vartheta)^\varepsilon) := \big\{u \in L^2((0,1)), \sum_{k\ge1} \lambda_k^{2\varepsilon}\langle u, e_k\rangle^2 < \infty\big\} \tag{4.34}$$

endowed with the norm $\|u\|_{D_\varepsilon} := \|(-A_\vartheta)^\varepsilon u\|_{L^2}$. For $\varepsilon < 1/4$, these spaces can be identified with L^2-Sobolov spaces on $(0,1)$, namely $D_\varepsilon = W^{2\varepsilon,2}$ and the norms are equivalent. For a proof of this characterization, we refer to, e.g., [11].

Lemma 4.5.5. *Under Assumption (M) we have* $\mathbf{E}(\|X_t\|_{D_\varepsilon}^p) = \mathbf{E}(\|X_0\|_{D_\varepsilon}^p) < \infty$ *and* $\mathbf{E}(\|f(X_t)\|_{D_\varepsilon}^p) = \mathbf{E}(\|f(X_0)\|_{D_\varepsilon}^p) < \infty$ *for all* $\varepsilon < 1/4$ *and* $p \ge 1$.

Proof. We use the usual decomposition $X_t = S(t)X_0 + X_t^0 + N_t$. By stationarity, we may choose $t = 1$. As before, $\mathbf{E}(\|X_1^0\|_{D_\varepsilon}^p) < \infty$ can be shown by a direct calculation. Further, $\mathbf{E}(\|S(1)X_0\|_{D_\varepsilon}^p) < \infty$ follows from

$$\|S(1)X_0\|_{D_\varepsilon}^2 = \sum_{k\ge1} \mathrm{e}^{-2\lambda_k}\lambda_k^{2\varepsilon}\langle X_0, e_k\rangle^2 \le \|X_0\|_{L^2}^2 \sum_{k\ge1} \mathrm{e}^{-2\lambda_k}\lambda_k^{2\varepsilon} \lesssim \|X_0\|_\infty^2.$$

To treat $N_1 = \int_0^1 S(1-s)f(X_s)\,ds$, note that

$$\|(-A_\vartheta)^\varepsilon S(h)u\|_{L^2}^2 = \sum_{k\ge1} \lambda_k^{2\varepsilon}\mathrm{e}^{-2\lambda_k h}\langle u, e_k\rangle^2 \le \sup_{\lambda\ge\lambda_1} \lambda^{2\varepsilon}\mathrm{e}^{-2\lambda h}\|u\|_{L^2}^2.$$

The function $\lambda \mapsto \lambda^{2\varepsilon}\mathrm{e}^{-2\lambda h}$ attains its maximum over $\mathbb{R}_+$ in $\lambda^* := \varepsilon/h$ and is monotonically decreasing on $[\lambda^*, \infty)$. Thus, we have $\sup_{\lambda\ge\lambda_1} \lambda^{2\varepsilon}\mathrm{e}^{-2\lambda h} \le g^2(h)$ with $g(h) := (\frac{\varepsilon}{\mathrm{e}h})^\varepsilon$ for $h \le \varepsilon/\lambda_1$ and $g(h) := \lambda_1^\varepsilon \mathrm{e}^{-\lambda_1 h}$ for $h > \varepsilon/\lambda_1$. Since $g \in L^1(\mathbb{R}_+)$, we can use Jensen's inequality to show

$$\begin{aligned}\|N_1\|_{D_\varepsilon}^p &\le \Big(\int_0^1 g(1-s)\|f(X_s)\|_{L^2}\,ds\Big)^p \le \Big(\int_0^1 g(s)\,ds\Big)^{p-1}\Big(\int_0^t g(1-s)\|f(X_s)\|_{L^2}^p\,ds\Big)\\ &\lesssim \int_0^1 g(1-s)\|f(X_s)\|_{L^2}^p\,ds.\end{aligned}$$

Therefore, $\mathbf{E}(\|N_1\|_{D_\varepsilon}^p) \lesssim \mathbf{E}(\|f(X_0)\|_{L^2}^p) \lesssim \mathbf{E}(\|f(X_0)\|_\infty^p) < \infty$ by Assumption (M) which shows the claim for X_t. In order to transfer the result to $f(X_t)$, we estimate

$$\begin{aligned}\|f(X_t)\|_{D_\varepsilon}^2 \lesssim \|f(X_t)\|_{W^{2\varepsilon,2}}^2 &= \|f(X_t)\|_{L^2}^2 + \int_0^1\int_0^1 \frac{(f(X_t(x)) - f(X_t((y)))^2}{|x-y|^{1+4\varepsilon}}\,dxdy \\ &\leq \|f(X_t)\|_{L^2}^2 + \|f'(X_t)\|_\infty^2 \|X_t\|_{D_\varepsilon}^2 \\ &\lesssim \|f(X_t)\|_{L^2}^2 + \|f'(X_t)\|_\infty^4 + \|X_t\|_{D_\varepsilon}^4,\end{aligned}$$

from where the claim follows by Assumptions (M) and (4.3) in view of the first part of this proof. □

Remark 4.5.6. The treatment of the nonlinear component in the above proof shows that $\mathbf{E}(\|N_t\|_{D_\varepsilon}^p) < \infty$ holds for all $\varepsilon < 1$.

The following lemma is useful for bounding the expression $\|\hat{S}(0)f(X_t) - f(X_t)\|_{L^2}^2$ appearing in the remainder term $\tilde{R}_i$ from the regression model (4.16). Of particular interest to us is the situation where α is close to 1/4 and, hence, the exponent $\frac{8\alpha^2}{4\alpha+1}$ can be chosen close to 1/4.

Lemma 4.5.7. *Let $H \in C^{2\alpha}([0,1]) \cap D_\alpha$ for some $\alpha \in (0, \frac{1}{2})$. Further, let $H^M := \sum_{k=1}^{M-1} H_k e_k$ where $H_k := \langle H, e_k\rangle_M = \frac{1}{M}\sum_{l=1}^{M-1} H(y_l)e_k(y_l)$. Then, there exists a constant $C > 0$ such that*

$$\|H - H^M\|_{L^2}^2 \leq CK^2\delta^{\frac{8\alpha^2}{4\alpha+1}}$$

where $K := \max(\|H\|_\infty, \|H\|_{C^{2\alpha}}, \|H\|_{D_\alpha})$.

Proof. First of all, by regarding H_k as a Riemann sum, we can bound

$$\begin{aligned}|H_k - h_k| = \Big|\frac{1}{M}\sum_{l=1}^{M} H(y_l)e_k(y_l) - \int_0^1 H(y)e_k(y)\,dy\Big| &\leq \sum_{l=0}^{M}\int_{y_l}^{y_{l+1}} |H(y_l)e_k(y_l) - H(y)e_l(y)|\,dy \\ \lesssim (\|e_k\|_\infty\|H\|_{C^{2\alpha}} + \|H\|_\infty\|e_k\|_{C^{2\alpha}})\delta^{2\alpha} &\lesssim (\|H\|_{C^{2\alpha}} + \|H\|_\infty k^{2\alpha})\delta^{2\alpha} \lesssim K\lambda_k^\alpha\delta^{2\alpha}. \quad (4.35)\end{aligned}$$

Similarly, since $\frac{1}{M}\sum_{k=1}^{M-1} H^2(y_k) = \|H^M\|_{L^2}^2 = \sum_{k=1}^{M-1} H_k^2$ holds by Lemma 4.4.7, we have

$$\begin{aligned}\Big|\|H^M\|_{L^2}^2 - \|H\|_{L^2}^2\Big| = \Big|\frac{1}{M}\sum_{k=1}^{M-1} H^2(y_k) - \|H\|_{L^2}^2\Big| &\leq \sum_{k=0}^{M-1}\int_{y_k}^{y_{k+1}} |H^2(y_k) - H^2(y)|\,dy \\ &\leq \|H^2\|_{C^{2\alpha}}\delta^{2\alpha} \leq 2\|H\|_\infty\|H\|_{C^{2\alpha}}\delta^{2\alpha} \lesssim K^2\delta^{2\alpha}. \quad (4.36)\end{aligned}$$

Also, note that for $h_k := \langle H, e_k\rangle_{L^2}$ and any $R \in \mathbb{N}$, we have

$$\sum_{l\geq R} h_l^2 \leq \lambda_R^{-2\alpha}\sum_{l\geq R}\lambda_l^{2\alpha}h_l^2 \leq \|H\|_{D_\alpha}^2\lambda_R^{-2\alpha} \lesssim K^2/R^{4\alpha}. \quad (4.37)$$

The three inequalities just derived are now used to bound

$$\begin{aligned}\|H - H^M\|_{L^2}^2 &= \|H^M\|_{L^2}^2 - \|H\|_{L^2}^2 + 2\langle H - H^M, H\rangle_{L^2} \\ &\leq |\|H^M\|_{L^2}^2 - \|H\|_{L^2}^2| + 2|\langle H - H^M, H\rangle_{L^2}| :\end{aligned}$$

Due to (4.36), the first term can be bounded by $K^2\delta^{2\alpha} \lesssim K^2\delta^{\frac{8\alpha^2}{4\alpha+1}}$ up to a constant. For the second term, using Parseval's identity, we get

$$|\langle H - H^M, H\rangle_{L^2}| = \Big|\sum_{l=1}^{M-1}(h_l - H_l)h_l + \sum_{l=M}^{\infty} h_l^2\Big| \leq \Big|\sum_{l=1}^{M-1}(h_l - H_l)h_l\Big| + \sum_{l=M}^{\infty} h_l^2 =: T_1 + T_2.$$

It follows directly from (4.37) that $T_2 \lesssim K^2/M^4 \lesssim K^2 \delta^{\frac{8\alpha^2}{4\alpha+1}}$. To estimate T_1, we decompose

$$T_1 \leq \Big| \sum_{l=1}^{M_0-1} (h_l - H_l) h_l \Big| + \Big| \sum_{l=M_0}^{M-1} (h_l - H_l) h_l \Big| =: T_{11} + T_{12}$$

for some intermediate value $M_0 \in \{1, \dots M-1\}$. Now, using the Cauchy-Schwarz inequality and (4.35), we get

$$T_{11}^2 \leq \Big(\sum_{l=1}^{M_0-1} \lambda_l^{-2\alpha} (h_l - H_l)^2 \Big) \Big(\sum_{l=1}^{M_0-1} \lambda_l^{2\alpha} h_l^2 \Big) \lesssim K^2 M_0 \delta^{4\alpha} \|H\|_{D_\alpha}^2 \lesssim K^4 M_0 \delta^{4\alpha}$$

and, by (4.37),

$$T_{12}^2 \leq \sum_{l=M_0}^{M-1} (h_l - H_l)^2 \sum_{l=M_0}^{M-1} h_l^2 \lesssim (\|H\|_{L^2}^2 + \|H^M\|_{L^2}^2) \sum_{l=M_0}^{\infty} h_l^2 \lesssim K^4 / M_0^{4\alpha}.$$

Balancing the bounds for T_{11} and T_{12} shows that it is optimal to take $M_0 \eqsim \delta^{-\frac{4\alpha}{4\alpha+1}}$ and with this choice we obtain the overall bound $T_1 \lesssim K^2 \delta^{\frac{8\alpha^2}{4\alpha+1}}$, which finishes the proof. □

Chapter 5

Conclusion and outlook

This thesis provides genuinely new insights to the theory of parameter estimation for SPDEs based on discrete observations in time and space. The heart of this thesis is given by the interplay of our results on estimation of (σ^2, ϑ_2). By providing matching upper and lower bounds, we have, in a sense, completely solved the problem of simultaneous estimation of the diffusivity and the volatility coefficient of the stochastic heat equation on an interval. The resulting optimal rate is remarkable since it differs from the usual parametric rate of convergence, in general. Furthermore, we have complemented the existing literature on statistics for SPDEs based on discrete observations by setting first steps into two new directions: by considering reaction-diffusion equations, we are the first to treat estimation in a semilinear framework and, by estimating the associated nonlinearity, we are the first to consider a fully nonparametric estimation problem. In fact, the latter is also the first account of nonparametric estimation of the nonlinearity in SPDEs, regardless of the observation scheme. In contrast to the parameter ϑ_2, estimation of the nonlinearity f turns out to be comparable to drift estimation for finite dimensional SDEs. The parametric and nonparametric estimators constructed in this thesis are either directly given by a closed form expression or via a simple least squares criterion, which makes them easy to implement and practically relevant. Apart from statistical contributions, this thesis introduces the replacement method for generating fully discrete samples of the linear stochastic heat equation on an interval. Based on this method, it is possible to generate almost exact (in distribution) samples at a comparably low computational cost. Employing the replacement method, simulation studies on our estimators confirm our theoretical results.

Based on our insights and developed techniques, it will be possible to extend the theory to more general models and estimation problems. This is particularly important with respect to practical applicability of the statistical methods. In the following, we state some further research directions that we consider to be interesting. Of course, the list is not exhaustive and, in particular, it would be possible to consider any combination of these extensions.

5.1 Adaptive nonparametric estimation of the nonlinearity

In Section 4.4 we have derived a nonparametric estimator for the nonlinearity f and, under certain assumptions, the estimator achieves the usual nonparametric rate $T^{-\frac{\alpha}{2\alpha+1}}$ where α determines the regularity of f. The estimator has the limitation that choosing the optimal value for the dimension D_m of the approximation space requires knowledge of the regularity α of f. This problem can usually be circumvented by considering an adaptive version of the estimator where the dimension of the approximation space is chosen in a data driven way. As in Comte et al. [21] this should be possible in our case by introducing a penalization $\mathrm{pen}(m) = \mathrm{pen}_{N,M}(m)$ for too large dimensions D_m, leading to overfitting. The resulting estimator is then given by

$$\hat{f} := \hat{f}_{\hat{m}} \quad \text{with} \quad \hat{m} := \arg\min_{m} \big(\min_{g \in S_m} \gamma_{N,M}(g) + \mathrm{pen}(m) \big) \tag{5.1}$$

where m belongs to an appropriate subset of $\mathbb{N}$ and $\gamma_{N,M}$ is the risk process from (4.17). In analogy to [21], it can be expected that choosing a penalization $\mathrm{pen}(m) \geq C\frac{D_m}{T}$ for some constant $C > 0$ will automatically realize the bias-variance compromise.[1]

5.2 Nonparametric estimation of the nonlinearity under low spatial resolution

Our nonparametric estimator of the nonlinearity f from Section 4.4 relies on an approximation of the heat semigroup by means of replacing the Fourier modes of the solution process by their empirical counterparts. This approximation step is only possible for discrete observations distributed throughout the whole spatial domain and with a much higher spatial than temporal resolution. Indeed, our oracle inequalities only serve as a consistency result when $M\Delta^2 \to \infty$. Thus, in future research it should be explored how nonparametric estimation of the nonlinearity can be accomplished while avoiding spectral approximations. For equations driven by a more regular noise process than a cylindrical Brownian motion, a possible approach could be to directly approximate the Laplacian rather than the corresponding semigroup, as discussed at the end of Section 4.4. For reaction-diffusion equations driven by space-time white noise, this remains a completely open problem.

5.3 Spatially varying diffusivity

Let us consider the stochastic heat equation with a spatially varying diffusivity parameter $\vartheta_2 : (0, 1) \to \mathbb{R}_+$, namely

$$dX_t(x) = \vartheta_2(x)\frac{\partial^2}{\partial x^2}X_t(x)\,dt + \sigma\,dW_t(x), \qquad (t, x) \in \mathbb{R}_+ \times (0, 1).$$

Without parametric assumptions on ϑ_2, using the local measurements approach due to Altmeyer and Reiß [3], it is possible to recover the value $\vartheta_2(x_0)$ based on the observation $\langle X_t, K_{h,x_0}\rangle$, $t \in [0, T]$, as $h \to 0$, where the kernel K_{h,x_0} is supported in a radius h around $x_0 \in (0, 1)$. Thus, one can say that the information on the value $\vartheta_2(x_0)$ is stored in the process locally around x_0. It will be interesting to explore how this property is reflected in space-time-discrete observations that are recorded locally around a point $x_0 \in (0, 1)$. Indeed, this is not only interesting from the practical perspective of deriving estimation methods for more general models but it will also clarify the structural connection between the discrete and the functional observation scheme.

5.4 Multi-dimensional space domains

Central limit theorems for time increments at finitely many spatial locations in $\mathbb{R}^d$ have been studied by Chong [15]. The situation where the number of spatial observations tends to infinity has not yet been explored. Neither have other types of increments been studied in the literature in a multi-dimensional setting, yet. Let us consider an open bounded space domain $\mathcal{O} \subset \mathbb{R}^d$ for some $d \geq 1$. Since the stochastic heat equation driven by a cylindrical Brownian motion only has a function valued solution in space dimension $d = 1$, we need to consider a noise process that is more regular in space. This can be implemented by considering

$$dX_t(x) = \vartheta_2\Delta X_t(x)\,dt + \sigma B\,dW_t(x), \qquad (t, x) \in \mathbb{R}_+ \times \mathcal{O},$$

[1] This approach has indeed proved to be successful, see the preprint Hildebrandt and Trabs [37] which appeared prior to the final publication of this thesis.

where the diffusion coefficient B has a regularizing effect in the sense that the condition (L2) from Section 1.1 is satisfied. E.g., by considering $B = (-\Delta)^{-\gamma}$ for an appropriate power $\gamma > 0$, we can again obtain a diagonalizable equation where the coefficient processes are independent. Furthermore, when considering a rectangular bounded space domain, e.g., $\mathcal{O} = (0,1)^d$, there are explicit expressions for the eigenfunctions of the Laplacian. In fact, they are given by products of the eigenfunctions in the one-dimensional case. This enables explicit calculations on the covariance structure of discrete observations but they will certainly be more tedious compared to dimension one.

While time increments are defined just like in the one-dimensional case, a generalization of the other types of increments hinges on the geometry of the points of observation in the space domain. If these form a rectangular grid, at least the generalization of double increments is straight forward. We conjecture that it is possible to estimate (σ^2, ϑ_2) based on a generalization of double increments in a rate optimal way. It will be interesting to see how this rate is affected by the dimension d as well as the regularity parameter γ. Of course, with the introduction of the new parameter γ also arises the question how it can be inferred from the data.

5.5 General nonlinearities

As for multi-dimensional space domains, a noise process that is more regular than a cylindrical Brownian motion also allows for more flexibility with respect to the class of nonlinearities F in systems of the type

$$dX_t(x) = \big(\vartheta_2 \frac{\partial^2}{\partial x^2} X_t(x) + F(X_t)(x)\big)\, dt + \sigma B\, dW_t(x), \qquad (t,x) \in \mathbb{R}_+ \times (0,1).$$

So far, we have only considered nonlinearities of Nemytskii-type, i.e., $F(u) = f \circ u$ for some function $f : \mathbb{R} \to \mathbb{R}$, but there are also other important examples to consider, e.g., Burgers' equation where $F(u) = -u \frac{\partial}{\partial x} u$. As in Pasemann and Stannat [68], by taking $B = (-\Delta)^{-\gamma}$ for some sufficiently large $\gamma > 0$, it is possible to force the solution process to take values in the domain of the operator F. Then, if the Hölder regularity of the solution process is preserved by F to a certain extent, one can conclude that the nonlinear component exceeds the linear component of the solution process in regularity. Thus, similar estimators for (σ^2, ϑ_2) as in Section 4.3 will keep their validity in the general semilinear framework.

Bibliography

[1] Altmeyer, R., Bretschneider, T., Janák, J., and Reiß, M. (2020a). Parameter estimation in an SPDE model for cell repolarisation. *arXiv preprint arXiv:2010.06340.*

[2] Altmeyer, R., Cialenco, I., and Pasemann, G. (2020b). Parameter estimation for semilinear SPDEs from local measurements. *arXiv preprint arXiv:2004.14728.*

[3] Altmeyer, R. and Reiß, M. (2021). Nonparametric estimation for linear SPDEs from local measurements. *Ann. Appl. Probab.*, 31(1):1–38.

[4] Bally, V. and Pardoux, E. (1998). Malliavin calculus for white noise driven parabolic SPDEs. *Potential Anal.*, 9(1):27–64.

[5] Baraud, Y., Comte, F., and Viennet, G. (2001). Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.*, 29(3):839–875.

[6] Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M., and Shephard, N. (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From stochastic calculus to mathematical finance*, pages 33–68. Springer, Berlin.

[7] Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory.* Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam.

[8] Bibinger, M. and Trabs, M. (2019). On central limit theorems for power variations of the solution to the stochastic heat equation. In *Stochastic Models, Statistics and Their Applications. Springer Proceedings in Mathematics & Statistics*, volume 294, pages 69–84.

[9] Bibinger, M. and Trabs, M. (2020). Volatility estimation for stochastic PDEs using high-frequency observations. *Stochastic Process. Appl.*, 130(5):3005–3052.

[10] Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York.

[11] Bonforte, M., Sire, Y., and Vázquez, J. L. (2015). Existence, uniqueness and asymptotic behaviour for fractional porous medium equations on bounded domains. *Discrete Contin. Dyn. Syst.*, 35(12):5725–5767.

[12] Brockwell, P. J., Davis, R. A., and Fienberg, S. E. (1991). *Time series: theory and methods.* Springer Science & Business Media.

[13] Cerrai, S. (1999). Ergodicity for stochastic reaction-diffusion systems with polynomial coefficients. *Stochastics Stochastics Rep.*, 67(1-2):17–51.

[14] Chong, C. (2019). High-frequency analysis of parabolic stochastic PDEs with multiplicative noise: Part I. *arXiv preprint arXiv:1908.04145.*

[15] Chong, C. (2020). High-frequency analysis of parabolic stochastic PDEs. *Ann. Statist.*, 48(2):1143–1167.

[16] Chong, Y. and Walsh, J. B. (2012). The roughness and smoothness of numerical solutions to the stochastic heat equation. *Potential Anal.*, 37(4):303–332.

[17] Cialenco, I. (2018). Statistical inference for SPDEs: an overview. *Stat. Inference Stoch. Process.*, 21(2):309–329.

[18] Cialenco, I. and Glatt-Holtz, N. (2011). Parameter estimation for the stochastically perturbed Navier-Stokes equations. *Stochastic Process. Appl.*, 121(4):701–724.

[19] Cialenco, I. and Huang, Y. (2020). A note on parameter estimation for discretely sampled SPDEs. *Stoch. Dyn.*, 20(3).

[20] Cialenco, I., Kim, H.-J., and Pasemann, G. (2021). Statistical analysis of discretely sampled semilinear SPDEs: a power variation approach. *arXiv preprint arXiv:2103.04211.*

[21] Comte, F., Genon-Catalot, V., Rozenholc, Y., et al. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543.

[22] Comte, F. and Rozenholc, Y. (2002). Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.*, 97(1):111–145.

[23] Comte, F. and Rozenholc, Y. (2004). A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.*, 56(3):449–473.

[24] Cont, R. (2005). Modeling term structure dynamics: an infinite dimensional approach. *Int. J. Theor. Appl. Finance*, 8(3):357–380.

[25] Da Prato, G. and Zabczyk, J. (1996). *Ergodicity for infinite-dimensional systems*, volume 229 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge.

[26] Da Prato, G. and Zabczyk, J. (2014). *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, Cambridge.

[27] Dacunha-Castelle, D. and Duflo, M. (1986). *Probability and statistics. Vol. II.* Springer-Verlag, Berlin Heidelberg New York.

[28] Daubechies, I. (1992). *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[29] Davie, A. and Gaines, J. (2001). Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations. *Math. Comp.*, 70(233):121–134.

[30] Devroye, L., Mehrabian, A., and Reddad, T. (2020). The total variation distance between high-dimensional Gaussians. *arXiv preprint arXiv:1810.08693v5.*

[31] Doukhan, P. (1994). *Mixing*, volume 85 of *Lecture Notes in Statistics.* Springer-Verlag, New York. Properties and examples.

[32] Galtchouk, L. I. and Pergamenshchikov, S. M. (2015). Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions. *Bernoulli*, 21(4):2569–2594.

[33] Goldys, B. and Maslowski, B. (2002). Parameter estimation for controlled semilinear stochastic systems: identifiability and consistency. *J. Multivariate Anal.*, 80(2):322–343.

[34] Goldys, B. and Maslowski, B. (2006). Lower estimates of transition densities and bounds on exponential ergodicity for stochastic PDE's. *Ann. Probab.*, 34(4):1451–1496.

[35] Haken, H. (2013). *Synergetics: Introduction and advanced topics.* Springer Science & Business Media.

[36] Hildebrandt, F. (2020). On generating fully discrete samples of the stochastic heat equation on an interval. *Statist. Probab. Lett.*, 162. Article 108750.

[37] Hildebrandt, F. and Trabs, M. (2021a). Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations. *arXiv preprint arXiv:2102.13415.*

[38] Hildebrandt, F. and Trabs, M. (2021b). Parameter estimation for SPDEs based on discrete observations in time and space. *Electron. J. Stat.* Forthcoming.

[39] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79(1):135–163.

[40] Huebner, M., Khasminskii, R., and Rozovskii, B. (1993). Two examples of parameter estimation for stochastic partial differential equations. In *Stochastic processes*, pages 149–160. Springer.

[41] Huebner, M. and Rozovskii, B. L. (1995). On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's. *Probab. Theory Related Fields*, 103(2):143–163.

[42] Ibragimov, I. and Rozanov, Y. (1978). *Gaussian random processes.* Springer-Verlag, Berlin Heidelberg New York.

[43] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York Berlin. Asymptotic theory, Translated from the Russian by Samuel Kotz.

[44] Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–139.

[45] Jacod, J. and Protter, P. (2012). *Discretization of processes*, volume 67 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg.

[46] Janák, J. (2020). Parameter estimation for stochastic partial differential equations of second order. *Appl. Math. Optim.*, 82(1):353–397.

[47] Jentzen, A. and Kloeden, P. E. (2009). The numerical approximation of stochastic partial differential equations. *Milan J. Math.*, 77:205–244.

[48] Kaino, Y. and Uchida, M. (2021a). Adaptive estimator for a parabolic linear SPDE with a small noise. *Jpn. J. Stat. Data Sci.* Forthcoming.

[49] Kaino, Y. and Uchida, M. (2021b). Parametric estimation for a parabolic linear SPDE model based on discrete observations. *J. Statist. Plann. Inference*, 211:190–220.

[50] Koski, T. and Loges, W. (1985). Asymptotic statistical inference for a stochastic heat flow problem. *Statist. Probab. Lett.*, 3:185–189.

[51] Kriz, P. and Maslowski, B. (2019). Central limit theorems and minimum-contrast estimators for linear stochastic evolution equations. *Stochastics*, 91(8):1109–1140.

[52] Liptser, R. S. and Shiryaev, A. N. (2013). *Statistics of Random Processes II - Applications.* Springer Science & Business Media, Berlin Heidelberg.

[53] Lototsky, S. V. (2009). Statistical inference for stochastic parabolic equations: a spectral approach. *Publ. Mat.*, 53(1):3–45.

[54] Lunardi, A. (1985). Interpolation spaces between domains of elliptic operators and spaces of continuous functions with applications to nonlinear parabolic equations. *Math. Nachr.*, 121(1):295–318.

[55] Lunardi, A. (2012). *Analytic semigroups and optimal regularity in parabolic problems.* Springer Science & Business Media.

[56] Mahdi Khalil, Z. and Tudor, C. (2019). Estimation of the drift parameter for the fractional stochastic heat equation via power variation. *Mod. Stoch. Theory Appl.*, 6(4):397–417.

[57] Manthey, R. (1986). Existence and uniqueness of a solution of a reaction-diffusion equation with polynomial nonlinearity and white noise disturbance. *Math. Nachr.*, 125:121–133.

[58] Marinelli, C., Nualart, E., and Quer-Sardanyons, L. (2013). Existence and regularity of the density for solutions to semilinear dissipative parabolic SPDEs. *Potential Anal.*, 39(3):287–311.

[59] Markussen, B. (2013). Likelihood inference for a discretely observed stochastic partial differential equation. *Bernoulli*, 9(5):745 – 762.

[60] Maslowski, B. and Pospisil, J. (2008). Ergodicity and parameter estimates for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.*, 57(3):401–429.

[61] Massart, P. (2007). *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003.

[62] Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables.* Marcel Dekker, inc., New York.

[63] Mueller, C. and Nualart, D. (2008). Regularity of the density for the stochastic heat equation. *Electron. J. Probab.*, 13(74):2248–2258.

[64] Neveu, J. (1968). Processus aléatoires gaussiens. Seminaire de mathematiques superieures. Les Presses de l'Université de Montréal.

[65] Nualart, D. and Quer-Sardanyons, L. (2009). Gaussian density estimates for solutions to quasi-linear stochastic partial differential equations. *Stochastic Process. Appl.*, 119(11):3914–3938.

[66] Pakkanen, M. S. (2014). Limit theorems for power variations of ambit fields driven by white noise. *Stochastic Process. Appl.*, 124(5):1942–1973.

[67] Pasemann, G., Flemming, S., Alonso, S., Beta, C., and Stannat, W. (2021). Diffusivity estimation for activator-inhibitor models: Theory and application to intracellular dynamics of the actin cytoskeleton. *J. Nonlinear Sci.*, 31. Paper No. 59.

[68] Pasemann, G. and Stannat, W. (2020). Drift estimation for stochastic reaction-diffusion systems. *Electron. J. Stat.*, 14(1):547–579.

[69] Réveillac, A., Stauch, M., and Tudor, C. A. (2012). Hermite variations of the fractional brownian sheet. *Stoch. Dyn.*, 12(3):1150021.

[70] Rohde, A. (2004). On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statist. Decisions*, 22(3):235–243.

[71] Shevchenko, R., Slaoui, M., and Tudor, C. A. (2020). Generalized k-variations and Hurst parameter estimation for the fractional wave equation via Malliavin calculus. *J. Statist. Plann. Inference*, 207:155–180.

[72] Sinestrari, E. (1985). On the abstract cauchy problem of parabolic type in spaces of continuous functions. *J. Math. Anal. Appl.*, 107(1):16–66.

[73] Swanson, J. (2007). Variations of the solution to a stochastic heat equation. *Ann. Probab.*, 35(6):2122–2159.

[74] Torres, S., Tudor, C., Viens, F., et al. (2014). Quadratic variations for the fractional-colored stochastic heat equation. *Electron. J. Probab.*, 19.

[75] Tsybakov, A. B. (2010). *Introduction to Nonparametric Estimation.* Springer, New York.

[76] Tuckwell, H. C. (2013). Stochastic partial differential equations in neurobiology: Linear and non-linear models for spiking neurons. In *Stochastic Biomathematical Models*, pages 149–173. Springer.

[77] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492.

[78] Whittle, P. (1953). The analysis of multiple stationary time series. *J. R. Stat. Soc. Ser. B*, 15:125 – 139.

Appendix

Abstract

Parameter estimation for parabolic stochastic partial differential equations in one space dimension is studied, observing the solution field on a discrete grid in time and space. We focus on an infill asymptotic regime in both the time and the space coordinate.

First of all, we consider the linear stochastic heat equation on an interval. While temporal power variations have already been studied in the literature, we prove central limit theorems for realized quadratic variations based on spatial increments as well as on double increments in time and space. Resulting method of moments estimators for the diffusivity and the volatility parameter inherit the asymptotic normality and can be constructed robustly with respect to the sampling frequencies in time and space. Upper and lower bounds reveal that, in general, the optimal convergence rate for joint estimation of the parameters is slower than the usual parametric rate.

Then, the semilinear framework of reaction-diffusion equations is considered, where the nonlinearity is given by a smooth function $f : \mathbb{R} \to \mathbb{R}$. Noting that the solution process is exceeded by its nonlinear component in Hölder regularity, we show that the asymptotic properties of our diffusivity and volatility estimators largely carry over from the linear setup. Furthermore, we derive a nonparametric estimator for the nonlinearity f of the underlying equation. The estimate is chosen from a finite dimensional function space based on a simple least squares criterion. We derive oracle inequalities both with respect to the empirical 2-norm with evaluations at the data points and with respect to the L^2-risk. Our results provide conditions for the estimator to achieve the usual nonparametric rate of convergence.

Reverting to the linear setup, our theoretical results are illustrated in numerical examples. In order to perform precise simulations, we develop the *replacement method* for generating fully discrete samples of the solution to the stochastic heat equation on an interval. Our approach generalizes a method proposed by Davie and Gaines (2001). In order to provide a theoretical justification of the method, we derive a condition for the validity of the approximation which is particularly applicable when the number of temporal and spatial observations tends to infinity. The quality of the approximation is measured in total variation distance. Simulation results indicate that samples provided by the replacement method are more accurate and considerably less computationally expensive than those obtained by naive truncation in Fourier space.

Zusammenfassung

Wir behandeln Parameterschätzung für parabolische stochastische partielle Differentialgleichungen in einer Raumdimension, wenn der Lösungsprozess an diskreten Gitterpunkten in Raum und Zeit beobachtet wird. Der Fokus liegt dabei auf hochfrequenten Beobachtungen, sowohl in der Zeit- als auch in der Raumkoordinate.

Zuerst betrachten wir die lineare stochastische Wärmeleitungsgleichung auf einem Intervall. Während zeitliche Variationsprozesse bereits in der Literatur untersucht wurden, beweisen wir zentrale Grenzwertsätze für empirische quadratische Variationen basierend sowohl auf Orts- als auch auf

Doppelinkrementen in Raum und Zeit. Die resultierenden Momentenschätzer für den Diffusivitäts- und den Volatilitätsparameter erben die asymptotische Normalität und können robust bezüglich der Beobachtungsfrequenz im Ort und in der Zeit konstruiert werden. Obere und untere Schranken zeigen, dass die optimale Konvergenzrate für die gemeinsame Schätzung von Volatilität und Diffusivität im Allgemeinen langsamer als die parametrische Konvergenzrate ist.

Anschließend betrachten wir ein semilineares Modell, und zwar Reaktions-Diffusions-Gleichungen, deren Nichlinearität durch eine glatte Funktion $f : \mathbb{R} \to \mathbb{R}$ gegeben ist. Wir stellen fest, dass die Hölder-Regularität der nichtlinearen Komponente der Lösung höher als die des gesamten Lösungsprozesses ist. So können wir zeigen, dass sich die asymptotischen Eigenschaften unserer Diffusivitäts- und Volatilitätsschätzer größtenteils nicht vom linearen Fall unterscheiden. Desweiteren behandeln wir nichtparametrische Schätzung der Nichtlinearität f der zugrundeliegenden Gleichung. Unser Schätzer wird aus einem endlichdimensionalen Funktionenraum basierend auf einem leicht zu implementierenden kleinste-Quadrate-Kriterium gewählt. Wir leiten Orakel-Ungleichungen bezüglich der empirischen 2-Norm, die sich durch Auswertung einer Funktion an den Datenpunkten ergibt, und bezüglich des L^2-Risikos her. Dadurch erhalten wir Bedingungen, unter denen der Schätzer die übliche nichtparametrische Konvergenzrate erreicht.

Unsere theoretischen Resultate werden anhand der linearen Gleichung mit numerischen Beispielen illustriert. Um präzise Simulationen durchführen zu können, entwickeln wir die *Ersetzungsmethode* zur Erzeugung raum- und zeitdiskreter Beobachtungen für die stochastische Wärmeleitungsgleichung auf einem Intervall. Unser Verfahren verallgemeinert eine Methode, die von Davie und Gaines (2001) vorgeschlagen wurde. Zur theoretischen Rechtfertigung der Methode leiten wir eine Bedingung für die Gültigkeit der Approximation her, die insbesondere anwendbar ist, wenn die Anzahl an zeitlichen und örtlichen Beobachtungen gegen unendlich konvergiert. Dabei wird die Approximationsgüte bezüglich des Totalvariationsabstandes gemessen. Simulationionsergebnisse legen nahe, dass anhand der Ersetzungsmethode generierte Beobachtungen sowohl präziser als auch deutlich weniger rechenaufwändig sind im Vergleich zum bloßen Abschneiden der Fourierreihe.

List of publications derived from the dissertation

The results of Chapter 2 referring to a fixed finite time horizon as well as the corresponding simulations from Chapter 3 are taken from Hildebrandt and Trabs [38]. The remaining results of Chapter 3 can be found in the publication Hildebrandt [36]. The results of Chapter 4 are part of the preprint Hildebrandt and Trabs [37].

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Florian Hildebrandt
Hamburg, den 10.05.2021