Parameter Estimation for Semilinear Stochastic Partial Differential Equations

vorgelegt von
M. Sc.
Gregor Pasemann
ORCID: 0000-0002-8107-6685

an der Fakultät II – Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Tobias Breiten
Gutachter: Prof. Dr. Wilhelm Stannat
Gutachterin: Prof. Dr. Annika Lang

Tag der wissenschaftlichen Aussprache: 27.08.2021

Berlin 2021

Abstract

The problem of parametric drift estimation for semilinear stochastic partial differential equations (SPDE) is considered based on a maximum–likelihood approach. The diffusivity of such models is estimated in finite time based on a single trajectory with high resolution in space. This is implemented by observing either a large number of Fourier modes (spectral approach), a large number of spatial point evaluations of the process (discretized spectral approach) or a convolution with a kernel of small diameter (local approach). Asymptotic properties of different estimators within these observation schemes are discussed, based on a spatial regularity analysis of the solution to the underlying SPDE. Examples of the general theory include reaction-diffusion equations, Burgers equation and equations of Cahn–Hilliard type. Special emphasis is put on the issue of model misspecification, with respect to either the drift or the driving noise.

The theoretical results are supported by a numerical simulation.

As an extension, the case of simultaneous diffusivity and reaction parameter estimation from spectral observations is treated in the context of stochastic activator-inhibitor models. This is applied to experimental observations of the actin marker concentration within *Dictyostelium discoideum* giant cells, whose spatiotemporal dynamics is described as a stochastic FitzHugh–Nagumo system. The performance of different estimators is compared on synthetic data from numerical simulation as well as real data.

Zusammenfassung

Diese Arbeit befasst sich mit parametrischer Driftschätzung für semilineare stochastische partielle Differentialgleichungen (SPDE) auf der Grundlage eines Maximum-Likelihood-Ansatzes. Die Diffusivität solcher Modelle wird in endlicher Zeit unter Beobachtung eines einzelnen Pfades mit hoher räumlicher Auflösung geschätzt. Diese hohe räumliche Auflösung wird formalisiert durch eine große Anzahl an Eigenfrequenzen (Spektralansatz), eine große Anzahl beobachteter Punktauswertungen (diskretisierter Spektralansatz) oder eine Faltung mit einem Kern mit kleinem Durchmesser (lokaler Ansatz). Für verschiedene Schätzer werden die asymptotischen Eigenschaften innerhalb dieser Beobachtungsmodelle analysiert. Grundlage hierfür ist eine genaue Bestimmung der räumlichen Regularität der Lösung der zugrundeliegenden SPDE. Beispiele für die allgemeine Theorie sind Reaktions-Diffusions-Gleichungen, die Burgers-Gleichung sowie Gleichungen vom Cahn–Hilliard-Typ. Weiterhin werden Fehlspezifikationen des zugrundeliegenden Modells behandelt, bezogen sowohl auf den Driftterm als auch auf den stochastischen Term.

Die Theorie wird durch numerische Simulationen unterstützt.

Als eine Erweiterung der bisherigen Theorie wird die simultane Diffusions- und Reaktionsparameterschätzung im Spektralansatz im Kontext stochastischer Aktivator-Inhibitor-Modelle betrachtet. Dies wird angewendet auf experimentelle Beobachtungsdaten der Aktinmarkerkonzentration in *Dictyostelium discoideum*-Zellen, wobei hier eine Beschreibung als stochastisches FitzHugh–Nagumo-System angenommen wird. Die Ergebnisse verschiedener Schätzer werden für Simulationen und experimentelle Daten verglichen.

Acknowledgments

I am grateful to numerous people, with whom I discussed and from whom I learned during the last years. Especially, I would like to thank

- my PhD supervisor Wilhelm Stannat for the support and guidance through my doctorate,
- Annika Lang for examining this dissertation,
- Tobias Breiten for chairing the scientific defense,
- all my coauthors, from whom I had the chance to learn a lot: Sergio Alonso, Randolf Altmeyer, Carsten Beta, Igor Cialenco, Sven Flemming, Hyun-Jung Kim, Wilhelm Stannat (in alphabetical order),[1]
- the anonymous referees for their helpful comments on the preprints and publications,
- the SFB 1294 "Data Assimilation" as well as the Technical University Berlin (TU Berlin) for providing all the resources needed in order to conduct this work,[2]
- Igor Cialenco and the Illinois Institute of Technology (IIT) for the hospitality during a research stay from January to March 2020,
- Sven Flemming for preparing and providing the experimental giant cell data used in Section 6.2,
- Randolf Altmeyer for giving helpful feedback on the final draft of this manuscript,
- Markus Reiß for introducing me to the field of statistics for SPDEs,
- my family for supporting me and believing in my project.

[1]During my time as a PhD student, I worked on four projects [PS20], [ACP20], [PFA+21], [CKP21]. The first three works are the basis for this dissertation.

[2]This research has been funded by the Deutsche Forschungsgemeinschaft (DFG)-Project-ID 318763901 - SFB1294.

Contents

Chapter 1

Introduction

Statistical inference for stochastic partial differential equations (SPDEs) is an emerging field within the topic of statistics for stochastic processes. Formally, an SPDE can be described as an evolution equation

$$\mathrm{d}X_t = \mathcal{A}(X_t)\mathrm{d}t + B(X_t)\mathrm{d}W_t \tag{1.1}$$

on an infinite-dimensional state space, with suitable initial condition. Here, $\mathcal{A}$ determines the drift, B acts as an dispersion operator and W is a cylindrical Wiener process. Further details are given below and in Section 2.1.

It is natural to use an SPDE in order to describe the spatiotemporal dynamics of phenomena such as pattern formation or traveling waves, see [SS16] and references therein for the propagation of action potentials in neuroscience, or [FFAB20] for actin dynamics within *D. discoideum* giant cells. Such models may arise in different ways, for example by deriving them from first principles, or as a phenomenological description, by adding noise to a deterministic partial differential equation. A widely used class of models is given by stochastic reaction–diffusion systems, whose drift component combines localized reaction dynamics with diffusive coupling in space, i.e. $\mathcal{A} = \nabla \cdot (\theta\nabla) + F$, where θ describes the (possibly inhomogeneous and anisotropic) diffusivity. The reaction term F can reflect detailed knowledge of the underlying (e.g. biophysical) processes, or it may be a minimal model capable of reproducing certain features found in the observations. Different approaches may be used to describe observed patterns, and it is desirable to apply statistical methods in order to understand the advantages and limits of different models.

Despite their great variety, a common feature of many SPDE models is the presence of diffusive forcing. Consequently, a precise understanding of the diffusivity θ is crucial. As a natural first approximation, the diffusivity can be assumed to be homogeneous and isotropic, i.e. $\theta > 0$ is just a positive number. In this case, $\nabla \cdot (\theta \nabla) = \theta \Delta$, where Δ is the Laplacian. Depending on the amount and quality of the data at hand, this may be refined. For example, in [AR21], an estimation theory for a stochastic heat equation with inhomogeneous diffusivity is developed.

Context and Literature

In order to improve the understanding of a random field with spatial and temporal extension based on observed data, different techniques from statistical inference can be applied to different classes of models. We mention two related approaches, which complement the modeling ansatz outlined above: Apart from starting directly with an SPDE model, a Gaussian field with specified covariance function may be imposed, see e.g. [Yin93, Moh97]. Here, the focus lies on features of the covariance rather than the dynamics of the temporal evolution. Another related concept is a partial differential equation with boundary noise, as studied in [BST80, AB88, MP07].

Literature surveys concerning statistical inference for SPDEs are given in [Lot09, Cia18].[1] A classical problem concerns the estimation of unknown parameters of the underlying model based on observed data [IH81]. Most of the literature on statistics for SPDEs is related to parameter estimation, which we are primarily interested in. Further central tasks that have been studied include hypothesis testing [CX14, CX15], nonparametric estimation [HL00a, HL00b, PR02, HT21a] and Bayesian inference [Bis99, PR00, Bis02, CCG20]. Another different but related topic is stochastic filtering (see [LS77, LS01, BC09] for the general theory and e.g. [Ouv78, RR20] for infinite-dimensional processes). In [Lot04], the solution to a parabolic equation is interpreted as the observation process of a hidden parameter, which is assumed to satisfy itself a stochastic differential equation. Estimation of unknown quantities of the signal SPDE with maximum–likelihood methods in the context of stochastic filtering is treated in [BB84, AS88, Aih92, Aih98b].

[1] A continuously updated list of references can be found at the webpage: *https://sites.google.com/prod/view/stats4spdes/*

Inference on stochastic (ordinary) differential equations, SDEs for short, is well-established, with a huge body of literature. We refer to [Kut04] for a detailed analysis of various statistical questions for ergodic diffusion processes, which can be analyzed through their large time behavior. It is natural to consider the large time regime also for the case of SPDEs, with infinite dimensional state space, exploiting ergodicity properties of the process. This has been done by [Log84, KL85, KL86, Moh94] for parabolic equations (see also [Aih98a]), and [Jan20, Jan21] for damped wave equations. The case of fractional noise is treated in [MP07, MT13, KM19].

However, it has been noted in [HKR93, HR95] that it is possible to recover certain drift parameters (for example, the diffusivity of a stochastic heat equation) even in finite time. In fact, the presence of unbounded operators implies that the measures on path space generated by the process in finite time for different drift parameters can be singular. This is in strong contrast to the case of SDEs with finite dimensional state space, where Girsanov's theorem assures that the measures on path space are absolutely continuous. While this fragmentation of the path space into the domains of singular measures can lead to analytical difficulties (e.g. due to the lack of a dominating measure), it is helpful from a statistical point of view as it implies the identifiability of the parameters: Given an observation X, one only has to determine the measure whose support contains X. Of course, in practice this is achieved by substituting the state space by some finite-dimensional approximation and studying the asymptotics of classical estimation techniques based on that discretization.

The methods and works on parameter estimation for SPDEs can be categorized according to the observation scheme they are based on. Different discretization schemes can be applied in space and time.

We focus on the idealized assumption that the process is observed continuously in time up to a fixed $T > 0$. However, there are various works on the temporally discrete setting, see e.g. [Mar03, PR96, PR97, CDVK20] for maximum likelihood-type drift estimation, or [BT19, BT20, Cho20, Cho19, CD20, KU21b, KU21a, TTV14] for different approaches based on temporal power variations.

Under full spatial observations, it is possible to study the limit of small noise intensity, as done by [Hue99, IK99, IK00, IK01]. On the other hand, partial knowledge on the process in space can be modeled in different ways:

The *spectral approach* is based on observing an increasing number of eigenfrequencies of the process, which are associated to the highest order linear differential operator appearing in the drift. For the (finite) set of observed modes, maximum–likelihood techniques can be applied. This approach has been introduced by [HKR93, HR95, Hue93] and extended by a large number of subsequent works. If the remaining drift and dispersion terms are linear and commute with the highest order drift term, the SPDE decouples to a system of independent one-dimensional processes. Even in the non-diagonalizable case, similar techniques can be applied, see [Lot96, LR99, LR00, Lot03]. Further, in [HLR97], estimators derived from Galerkin approximations instead of spectral projections are discussed. In [LL10b, LL10a], hyperbolic equations are considered. Trajectory fitting estimators have been analyzed in [CGH18]. Different noise models are treated in [CL09, CCG20] (multiplicative noise), [CLP09, Cia10, Hui14, Kří20] (additive and multiplicative fractional noise) and [CKL20] (space-only noise).

For *spatially discrete observations*, modeled as a set of spatial point evaluations of the solution process, a natural approach is to consider power variations in space and analyze the asymptotic behavior as the mesh size of the underlying spatial grid tends to zero. This has been done in [PT07, CH20] (for a stochastic heat equation), [MKT19a, MKT19b] (for a stochastic fractional heat equation), [CKL20, CK22] (in the context of space-only noise) as well as [SST20] (for a wave equation driven by fractional noise). A joint spatiotemporal variation is considered in [HT21b, HT21a].

The new *local approach*, pioneered in [AR21], considers spatially discrete observations as local averages rather than point evaluations of the process. The process is weighted by some localized kernel, which is determined by the measurement device and its resolution. The high frequency regime from the spectral approach is substituted by a "shrinking kernel" regime that corresponds to high precision measurements. Interestingly, the diffusivity of a stochastic heat equation is identifiable from one single localized observation in finite time.

Most literature on SPDE parameter estimation is concerned with linear equations. An early treatment of nonlinear systems appears in [Hue93, Chapter 4], where estimators derived from Galerkin approximations are studied in the scope of general maximum likelihood theory [IH81], based on a similar analysis for ergodic diffusion processes [Kut04]. However, in this setting, the observations are not a functional of the full process X, but rather a finite-dimensional Markovian approximation to X. In [GM02], consistency of the maximum likelihood estimator for a class of controlled stochastic reaction–diffusion equations is shown in the large time regime (see also [DMPD00], where the case that the nonlinear term depends linearly on its parameters is treated separately). For finite observation horizon T, the first study of parameter estimation from spectral observations of a nonlinear process has been given in [CGH11] for the stochastic Navier–Stokes equations. The first result on nonparametric estimation of a reaction term based on discrete observations is given in the recent work [HT21a], where the authors consider one-dimensional stochastic reaction–diffusion equations.

So far, there are few works on application of SPDE parameter estimation techniques to experimental data. We point out [Unn89, KUP91], where parameter estimation for SPDE models arising in groundwater hydrology is studied with a formal maximum likelihood method, and diffusion and advection coefficients are calibrated from data.[2] Very recently, [ABJR21] considers stochastic cell repolarization models in the context of the local approach.

Main Results and Outline

In this work, we consider drift parameter estimation for semilinear SPDEs. We adapt the spectral and local approach to the general semilinear setting. Based on a maximum-likelihood ansatz, we construct consistent estimators for the unknown diffusivity and study their asymptotic properties in finite time in detail. Depending on the specific setting, we are able to obtain optimal convergence rates as well as asymptotic normality, which allows for the construction of confidence intervals. General results are given in Theorem 2.11 and 2.12 for the spectral approach, and Theorem 5.8 for the local approach. Examples are studied in Section 2.4. Our theory depends on a precise understanding of the spatial regularity of the solution X and related processes. Special emphasis is put on robustness of the estimators to model misspecification, either in the reaction term (Theorem 2.27) or in the specification of the dispersion operator (Theorem 3.22).

The general theory is supported by numerical evidence for the case of a stochastic Allen–Cahn equation (Section 2.5).

[2] Within a different setting, [Moh00] considers a random advection-diffusion equation in order to describe soil plutonium data.

We study how the results can be transferred to SPDEs with different noise models that are relevant for applications. For Ornstein–Uhlenbeck driven models, a detailed characterization is given in Theorem 3.7. As a new feature, even the rate of temporal correlation decay of such models can be identified in finite time (Theorem 3.10). On the other hand, the estimators are very sensitive to deviations from the semimartingale setting, as shown for the case of processes driven by integrated Wiener noise (Theorem 3.15).

In addition, we study to what extent the asymptotic results of the spectral approach can be recovered if spatially discrete observations rather than Fourier modes are available. Under mild conditions on the domain and scheme of point observations, we obtain bounds on the convergence rate of a discretized spectral estimator as the mesh size tends to zero (Theorem 4.3). Under additional assumptions on the underlying geometry, these bounds are optimal (Theorem 4.7), in the sense that they are consistent with results from related literature.

Finally, we extend the spectral estimation theory to the case of joint diffusivity and reaction parameter estimation (Theorem 6.1 and 6.4), with special emphasis on stochastic activator–inhibitor models. The results are applied to experimental *D. discoideum* giant cell observations in Section 6.2, where we discuss the effective diffusivity of intracellular actin concentration.

This work is based on the papers [PS20], [ACP20], [PFA$^+$21] and additional new material. It is structured as follows:

- Chapter 2 is based on [PS20] and develops the spectral approach for general semilinear models. In fact, the results from [PS20] are extended by means of a different approach to higher regularity (as in [ACP20]).

- Chapters 3 and 4 are new and not based on previous publications. In Chapter 3, different noise models for the spectral approach are worked out, based on SPDE models arising in biophysics literature. Special emphasis is put on the case of Ornstein–Uhlenbeck noise. Chapter 4 concerns the adaptation of the asymptotic results from the spectral approach to the case that the solution process X is observed not via its Fourier modes, but discretized in space.

- Chapter 5 is based on [ACP20]. Diffusivity estimation for semilinear SPDE models is treated from the perspective of the recently introduced local approach. A crucial tool is higher L^p-regularity of the solution process.

- Chapter 6 is based on [PFA+21]. Diffusivity and reaction parameters are jointly estimated in the scope of the spectral approach, and the results are applied to simulated and experimental cell data.

A First Example

We outline the general proceeding with a simple example. For a Gelfand triple $V \subset H \subset V^*$, consider the equation

$$\mathrm{d}X_t = \theta A(X_t)\mathrm{d}t + B(X_t)\mathrm{d}W_t, \qquad X_0 \in H, \tag{1.2}$$

with unknown $\theta > 0$, where $A : V \to V^*$ is a possibly nonlinear operator, W is a cylindrical Wiener process, and B maps V into the space of Hilbert–Schmidt operators on H. This corresponds to $\mathcal{A} = \theta A$ in (1.1). In this example, θ should be seen as the overall drift intensity rather than diffusivity. Assume that (1.2) is well-posed, e.g. under monotonicity and coercivity assumptions on A and B [LR15]. Now, given a sequence of linear projection operators $(P_N)_{N\in\mathbb{N}}$ with finite-dimensional range on H, the dynamics for $X^N := P_N X$ is given by

$$\mathrm{d}X^N = \theta A_N(X_t)\mathrm{d}t + B_N(X_t)\mathrm{d}W_t, \tag{1.3}$$

where $A_N(X) := P_N A(X)$ and $B_N(X) := P_N B(X)$. Note that in general X^N ceases to be Markovian. Assume that $X^N, A_N(X)$ and $B_N(X)$ are observed. Let $B_N(X_t)B_N(X_t)^T$ be invertible for $0 \le t \le T$ (interpreted as an operator acting on the range of P_N), and set $B_N(X_t)^+ := B_N(X_t)^T(B_N(X_t)B_N(X_t)^T)^{-1}$. This is the Moore–Penrose pseudoinverse[3] of the operator $B_N(X_t)$. Then a natural estimator for θ is given by

$$\hat{\theta}_N = \frac{\int_0^T \left\langle (B_N(X_t)B_N(X_t)^T)^{-1}A_N(X_t), \mathrm{d}X_t^N \right\rangle}{\int_0^T \|B_N(X_t)^+ A_N(X_t)\|^2 \,\mathrm{d}t}. \tag{1.4}$$

This estimator can be either derived from a maximum likelihood approach or directly justified by the decomposition

$$\hat{\theta}_N - \theta = \frac{\int_0^T \langle B_N(X_t)^+ A_N(X_t), \mathrm{d}W_t\rangle}{\int_0^T \|B_N(X_t)^+ A_N(X_t)\|^2 \,\mathrm{d}t}. \tag{1.5}$$

[3] For operators between Hilbert spaces, the Moore–Penrose pseudoinverse is defined analogously to the finite-dimensional case, cf. [WD01]. See e.g. [LS01, Chapter 13] for properties of the pseudoinverse in the finite-dimensional case.

Set $I_N := \int_0^T \|B_N(X_t)^+ A_N(X_t)\|^2 \, \mathrm{d}t$. According to Theorem A.1, $\hat{\theta}_N$ is a consistent estimator as $N \to \infty$ which is asymptotically normal with rate $(\mathbb{E}I_N)^{-1/2}$, i.e.

$$(\mathbb{E}I_N)^{1/2} \left(\hat{\theta}_N - \theta\right) \to \mathcal{N}(0,1), \tag{1.6}$$

whenever $I_N^{-1} \xrightarrow{\mathbb{P}} 0$ and $I_N/\mathbb{E}I_N \xrightarrow{\mathbb{P}} 1$ as $N \to \infty$. An explicit expression in N of the convergence rate $(\mathbb{E}I_N)^{-1/2}$ will depend on the particular projection operators P_N.

This discussion outlines the argument for maximum-likelihood based estimation theory for SPDEs with parametric drift terms. There are some comments to this approach:

(i) *Closability of the observation scheme:* In practice, it is unlikely that all three quantities $X^N, A_N(X)$ and $B_N(X)$ are observed. Usually there is just access to X^N. This problem can be addressed in different ways: Certain model assumptions can be imposed, e.g. that $B(X) \equiv B$ is constant and known. Also, the generating model and the observation scheme can be aligned in the sense that e.g. $A(X^N) = A_N(X)$, at least up to negligible terms. This is the basic idea behind the spectral approach, where A and P_N commute. When considering spatially discrete observations in Chapter 4, such commutativity relation does not hold, and we have to deal with an additional bias.

(ii) *Model refinement:* In this scenario, θ represents the overall drift intensity of the dynamics. However, it is desirable to refine this model, either by investigating a parameter linked to a specific part of the drift (e.g. spatial diffusion, as outlined above), or by considering multiparametric drift terms based on specific model knowledge. We address both questions in the sequel, with the main focus on diffusivity estimation.

(iii) *Robustness:* In the case that (1.2) is misspecified but close to the true generating dynamics, it is desirable that the asymptotic results transfer. We will look at robustness of the estimation procedure to misspecification in the drift and noise terms.

Chapter 2

The Spectral Approach

This chapter is an adaptation of the statements and results from [PS20].

The aim of this chapter is to develop an estimation theory for the diffusivity of a semilinear SPDE driven by additive noise within the spectral approach to statistical inference for SPDEs.

In the spectral approach, a finite number N of Fourier modes of the solution X to an SPDE is observed, usually continuously in time, and the asymptotic properties of estimators derived from these observations is determined as N tends to infinity. This ansatz has been pioneered by [HKR93, Hue93, HR95], where it has been noted that the coefficient of an unbounded drift operator of an SPDE can be identified in finite time, in strong contrast to the finite-dimensional case of an stochastic (ordinary) differential equation. In [CGH11], statistical inference for the stochastic Navier–Stokes equations with additive noise in two dimensions has been considered, which is the first treatment of parameter estimation in finite time for a nonlinear system within the spectral approach. The results and methods therein have been extended to general semilinear equations in [PS20], on which the present chapter is based.

In Section 2.1, we discuss the semilinear SPDE model which will be used throughout this chapter, together with some auxiliary results. Section 2.2 is concerned with optimal spatial regularity of the solution process X. In Section 2.3, we discuss the maximum-likelihood approach to diffusivity estimation and prove asymptotic results for the estimators derived from that approach. Examples including reaction-diffusion equations, the Burgers equation and equations of Cahn-Hilliard type are treated in Section 2.4. The impact of a misspecified drift term is analyzed in the same section. A numerical validation of the theory is given in Section 2.5 for the stochastic Allen-Cahn equation. Systems consisting of an observed component and an unobserved component are handled in Section 2.6.

2.1 The Setting

Let H be a separable Hilbert space with scalar product $\langle\cdot,\cdot\rangle$ and $A : D(A) \to H$ a densely defined, closed operator that is self-adjoint and negative definite with compact resolvent. For $s \geq 0$, let $H_s := D((-A)^{s/2}) \subseteq H$ be the domain of the fractional Laplacian, equipped with the norm $\|\cdot\|_s = \|(-A)^{s/2}\cdot\|_H$. For $s < 0$, let H_s be the completion of H w.r.t. the norm $\|\cdot\|_s$ given by the same term. For each $s \geq 0$, $H_s \subseteq H \subseteq H_{-s}$ forms a Gelfand triple. The dual pairing between H_s and H_{-s} is again denoted by $\langle\cdot,\cdot\rangle$. Set $V := H_1$, then V^* can be identified with H_{-1}. For $\theta > 0$, denote by $t \mapsto e^{t\theta A}$, $t > 0$, the C_0-semigroup generated by θA. Let $(\Phi_k)_{k\in\mathbb{N}}$ be an orthonormal basis of H consisting of eigenfunctions of $-A$, such that the corresponding sequence of (positive) eigenvalues $(\lambda_k)_{k\in\mathbb{N}}$ is ordered increasingly, taking into account multiplicities. Let $P_N : H \to H$ be the orthogonal projection onto the span of the first N eigenfunctions $\Phi_1, \ldots, \Phi_N$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_t)_{t\geq 0}$ a right-continuous complete filtration, then $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a stochastic basis.

In this chapter, we consider a semilinear SPDE of the form

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + B\mathrm{d}W_t \tag{2.1}$$

together with initial condition $X_0 \in H$, where W is a cylindrical Wiener process on H, $B = \sigma(-A)^{-\gamma}$ for some $\sigma, \gamma > 0$ is assumed to be of Hilbert–Schmidt type, $F : C(0, T; H) \supseteq D(F) \to L^1(0, T; H)$ is a nonlinear operator and $\theta > 0$ is an unknown parameter.

A pair of adapted processes (X, W), defined on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with $X \in D(F) \subseteq C(0, T; H)$ a.s. and W a cylindrical Wiener process, is a solution to (2.1) in the analytically and probabilistically weak sense if for all $v \in D(A) = H_2$ a.s.

$$\langle v, X_t\rangle = \langle v, X_0\rangle + \theta \int_0^t \langle Av, X_r\rangle \,\mathrm{d}r + \int_0^t \langle v, F(X)(r)\rangle \,\mathrm{d}r + \langle v, BW_t\rangle \,.$$

We always assume the following:

(W) There is a solution in the analytically and probabilistically weak sense to (2.1) in $C(0,T;H)$, which is unique in the sense of probability law.

By means of the Yamada-Watanabe theorem (see e.g. [LR15, Theorem E.0.8]), pathwise uniqueness implies uniqueness in the sense of probability law. In most examples, F will be of the form $F(X)(t) = F(X_t)$, i.e. (by abuse of notation) $F : H \supseteq D(F) \to H$, such that F extends to an operator $V \to V^*$. In this case, well-posedness of (2.1) can be handled in the context of the variational approach [LR15], cf. Section 2.4.[1]

Condition (W) alone imposes very little spatial regularity and should be considered as a minimal requirement that serves as a baseline for stronger regularity properties of X. In fact, a detailed analysis of higher regularity for X will be crucial for our statistical analysis, cf. Section 2.2. There, we need the representation of X as a mild solution:

$$X_t = e^{t\theta A}X_0 + \int_0^t e^{(t-r)\theta A}F(X)(r)\mathrm{d}r + \int_0^t e^{(t-r)\theta A}B\mathrm{d}W_r,$$

where the first integral is understood in the Bochner sense, and the second integral is a stochastic convolution.

Remark 2.1. *In general, analytically weak and mild solutions are not equivalent. However, if a.s. $X, F(X) \in L^1(0,T;H)$, it can be shown that analytically weak solutions are also mild solutions, see Proposition G.0.5 (i) and Remark G.0.6 in [LR15]. These conditions are satisfied in our setting.*

For two sequences of positive numbers $(a_N)_{N\in\mathbb{N}}$ and $(b_N)_{N\in\mathbb{N}}$, we write $a_N \asymp b_N$ if $a_N/b_N \to 1$ and $a_N \sim b_N$ if $a_N/b_N \to C$ for some $C > 0$. Further we write $a_N \lesssim b_N$ if $a_N \le Cb_N$ for some $C > 0$, $a_N \ll b_N$ if $a_N/b_N \to 0$ and $a_N \ll_p b_N$ if there is $\epsilon > 0$ such that $N^\epsilon a_N/b_N \to 0$. If the sequences are random, then these limits are meant in the almost sure sense, unless stated otherwise.

We always assume that there are $\Lambda, \beta > 0$ such that the sequence of eigenvalues of $-A$ has polynomial growth:

$$\lambda_k \asymp \Lambda k^\beta \tag{2.2}$$

[1]There is a vast literature on well-posedness and regularity for SPDEs, see [DPZ14, LR15] and references therein. In [Pes95], existence and uniqueness of semilinear equations on Banach spaces is considered. Stochastic reaction–diffusion equations are studied in detail in [Cer01]. See [Kry96, vNVW12] for a treatment of maximal L^p-regularity.

for $k \to \infty$. Note that with this notation, the condition that B is of Hilbert–Schmidt type is equivalent to

$$\gamma > \frac{1}{2\beta}. \tag{2.3}$$

We close this section by stating some auxiliary results.

Lemma 2.2. *For $s_1 \le s_2$ and $X \in H$:*

(i) $\|P_N X\|_{s_2}^2 \le \lambda_N^{s_2-s_1} \|P_N X\|_{s_1}^2$.

(ii) $\|(I-P_N)X\|_{s_1}^2 \le \lambda_{N+1}^{s_1-s_2} \|(I-P_N)X\|_{s_2}^2$.

(iii) If $X \in H_{s_1}$, then $\left\|e^{r\theta A}X\right\|_{s_2}^2 \le C_{s_2-s_1} r^{-(s_2-s_1)} \|X\|_{s_1}^2$ for some $C_{s_2-s_1} > 0$ and all $r > 0$.

Proof. All properties are clear from the spectral decomposition $\|Z\|_s^2 = \sum_{k=0}^{\infty} \lambda_k^s \langle Z, \Phi_k\rangle^2$, $s \in \mathbb{R}$, $Z \in H$. In (iii), we can choose the constant $C_{s_2-s_1} = \sup_{y>0} e^{-2y} y^{s_2-s_1}/\theta^{s_2-s_1}$. □

The statements (i) and (ii) are bounds of Bernstein and Jackson type, respectively (cf. [Sha71, BL76]). Statement (iii) is a smoothing property of the semigroup, see [Paz83, Lun95].

2.2 Spatial Regularity

In this section, we describe the precise regularity of X, and in particular, the excess regularity of its nonlinear part. In order to do so, we apply a classical splitting argument and write $X = \bar{X} + \widetilde{X}$, where $\bar{X}$ solves

$$\mathrm{d}\bar{X}_t = \theta A \bar{X}_t \mathrm{d}t + B \mathrm{d}W_t \tag{2.4}$$

with initial condition $X_0 = 0$, and $\widetilde{X}$ solves the random PDE

$$\mathrm{d}\widetilde{X}_t = (\theta A X_t + F(\bar{X} + \widetilde{X})(t))\mathrm{d}t, \quad \widetilde{X}_0 = X_0, \tag{2.5}$$

which reads as

$$\widetilde{X}_t = e^{t\theta A} X_0 + \int_0^t e^{(t-r)\theta A} F(\bar{X} + \widetilde{X})(r)\mathrm{d}r \tag{2.6}$$

in the mild formulation.

We infer higher regularity for $\widetilde{X}$ by means of the conditions $(F_{s,\eta})$ and $(F^{\mathrm{v}}_{s,\eta})$ below, which rely on the representation of $\widetilde{X}$ as a mild or weak solution, respectively. The weak solution approach has been used in [CGH11, PS20] for the spectral observation scheme, whereas the mild representation has been applied in [ACP20] in the context of the local observation scheme, cf. Chapter 5. As the mild approach yields larger excess regularity in our context, we focus mainly on $(F_{s,\eta})$. However, the weak approach using $(F^{\mathrm{v}}_{s,\eta})$ will be crucial in the examples in order to reach the level of regularity where the mild approach can be applied.

Here and in the sequel, we write $X^N := P_N X$ as well as $\bar{X}^N := P_N \bar{X}$ and $\widetilde{X} := P_N \widetilde{X}$. These projected processes satisfy

$$\begin{aligned}
\mathrm{d}\bar{X}^N_t &= \theta A \bar{X}^N_t \mathrm{d}t + B \mathrm{d}W^N_t, & \bar{X}_0 &= 0,\\
\mathrm{d}\widetilde{X}^N_t &= \theta A \widetilde{X}^N_t \mathrm{d}t + P_N F(X)(t)\mathrm{d}t, & \widetilde{X}_0 &= X_0,
\end{aligned}$$

where $W^N := P_N W$. In this section, we do not need the precise form (2.4) for the dynamics of $\bar{X}$, but only its spatial regularity. In this sense, the conclusions remain valid for Chapters 3 – 6, where the assumptions on the noise term are changed.

In order to quantify the regularity of X, we need the following spaces:

$$R(s) := L^\infty(0,T;H_s), \tag{2.7}$$

$$R^{\mathbb{E}}(s) := \bigcap_{p\geq 1} L^p(\Omega, L^\infty(0,T;H_s)), \tag{2.8}$$

i.e. $R(s)$ is a normed space with norm $\|X\|_{R(s)} = \sup_{0\leq t\leq T} \|X\|_s$, and $R^{\mathbb{E}}(s)$ is a locally convex space with $X \in R^{\mathbb{E}}(s)$ if and only if for all $p \geq 1$:

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \|X_t\|_s^p\right] < \infty. \tag{2.9}$$

While $X \in R(s)$ a.s. suffices for the purpose of diffusivity estimation, in many examples, the stronger statement $X \in R^{\mathbb{E}}(s)$ can be shown.

In order to conduct our regularity analysis, we need that for some $\eta > 0$ and $s \in \mathbb{R}$, the regularity of $F(X)$ differs from that of X by $2-\eta$ derivatives, in the following sense:

$(F_{s,\eta})$ There is $\epsilon > 0$ and a monotonous, locally bounded function $g : [0,\infty) \to [0,\infty)$ such that for all $X \in R(s)$:

$$\|F(X)\|_{R(s+\eta+\epsilon-2)} \leq g(\|X\|_{R(s)}). \tag{2.10}$$

In general, the function g has polynomial growth, cf. Section 2.4. If $F(X)(t) = F(X_t)$, i.e. F acts on every point in time separately, then it is sufficient for (2.10) that for each $0 \leq t \leq T$ and $X \in H_s$:

$$\|F(X_t)\|_{s+\eta+\epsilon-2} \leq g(\|X_t\|_s). \tag{2.11}$$

The proof of the next proposition is inspired by similar estimates in [DPDT94].

Proposition 2.3. *Let $s \in \mathbb{R}, \eta > 0$. Assume that $(F_{s,\eta})$ is true.*

(i) If $\bar{X}, \widetilde{X} \in R(s)$ and $X_0 \in H_{s+\eta}$, then $\widetilde{X} \in R(s+\eta)$ a.s.

(ii) If $\bar{X}, \widetilde{X} \in R^{\mathbb{E}}(s)$ and $X_0 \in L^p(\Omega, H_{s+\eta})$ for any $p \geq 1$ and if the function g from $(F_{s,\eta})$ is of the form $g(x) = C(1+x)^b$ for some $C, b \geq 0$, then $\widetilde{X} \in R^{\mathbb{E}}(s+\eta)$.

Proof.

(i) We have by Lemma 2.2 (iii):

$$\begin{aligned}
\left\|\widetilde{X}_t^N\right\|_{s+\eta} &\leq \left\|e^{t\theta A} X_0^N\right\|_{s+\eta} + \int_0^t \left\|e^{(t-r)\theta A} P_N F(\bar{X} + \widetilde{X})(r)\right\|_{s+\eta} \mathrm{d}r \\
&\lesssim \left\|X_0^N\right\|_{s+\eta} + \int_0^t (t-r)^{-1+\epsilon/2} \left\|P_N F(\bar{X} + \widetilde{X})(r)\right\|_{s+\eta+\epsilon-2} \mathrm{d}r \\
&\lesssim \|X_0\|_{s+\eta} + \left\|F(\bar{X} + \widetilde{X})\right\|_{R(s+\eta+\epsilon-2)} \int_0^t (t-r)^{-1+\epsilon/2} \mathrm{d}r \\
&\lesssim \|X_0\|_{s+\eta} + g\left(\left\|\bar{X} + \widetilde{X}\right\|_{R(s)}\right) \frac{2}{\epsilon} t^{\epsilon/2},
\end{aligned}$$

thus, uniformly in $N \in \mathbb{N}$,

$$\sup_{0 \leq t \leq T} \left\|\widetilde{X}_t^N\right\|_{s+\eta} \lesssim \|X_0\|_{s+\eta} + \frac{2}{\epsilon} T^{\epsilon/2} g\left(\left\|\bar{X}_t\right\|_{R(s)} + \left\|\widetilde{X}\right\|_{R(s)}\right), \tag{2.12}$$

and the right-hand side is finite by assumption.

(ii) By (2.12), for any $p \geq 1$,

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left\|\widetilde{X}_t^N\right\|_{s+\eta}^p\right] \lesssim \mathbb{E}\left\|X_0\right\|_{s+\eta}^p + \mathbb{E}\left[\left(1+\left\|\bar{X}\right\|_{R(s)}+\left\|\widetilde{X}\right\|_{R(s)}\right)^{pb}\right]$$

and further,

$$\mathbb{E}\left[\left(1+\left\|\bar{X}\right\|_{R(s)}+\left\|\widetilde{X}\right\|_{R(s)}\right)^{pb}\right] \lesssim 1+\mathbb{E}\left[\left\|\bar{X}\right\|_{R(s)}^{pb}+\left\|\widetilde{X}\right\|_{R(s)}^{pb}\right],$$

which is finite by assumption. This proves the claim.

□

We say that $s^* \in \mathbb{R}$ is the *optimal regularity* for $\bar{X}$ if a.s. $\bar{X} \in R(s)$ for all $s < s^*$ and $\bar{X} \notin R(s)$ for all $s > s^*$.

Proposition 2.4. *Let $\eta > 0$, let s^* be the optimal regularity of $\bar{X}$, let $s_0 < s^*$ such that $(F_{s,\eta})$ is true for each $s_0 \leq s < s^*$.*

(i) If a.s. $X \in R(s_0)$ and $X_0 \in H_{s^*+\eta}$, then a.s. $X \in R(s)$ for $s < s^*$ as well as $X \notin R(s)$ for $s > s^*$, and further a.s. $\widetilde{X} \in R(s+\eta)$ for $s < s^*$.

(ii) If $X \in R^{\mathbb{E}}(s_0)$, $X_0 \in L^p(\Omega, H_{s^*+\eta})$ for $p \geq 1$ and $\bar{X} \in R^{\mathbb{E}}(s)$ for $s < s^*$, and if the function g from $(F_{s,\eta})$ is of the form $g(x) = C(1+x)^b$ for some $C, b > 0$, then $X \in R^{\mathbb{E}}(s)$ for $s < s^*$, $X \notin R^{\mathbb{E}}(s)$ for $s > s^*$, and $\widetilde{X} \in R^{\mathbb{E}}(s+\eta)$ for $s < s^*$.

Proof. For (i), note that the statements $X \in R(s)$, $\widetilde{X} \in R(s+\eta)$ for $s < s^*$ follow inductively from Proposition 2.3. Further, if $X \in R(s)$ for some $s > s^*$ with positive probability, then $\bar{X} = X - \widetilde{X} \in R(s \wedge (s^* + \eta/2))$ with positive probability, in contradiction to the optimality of s^*. The reasoning for (ii) is similar. □

If it is possible to set $s_0 = 0$ in Proposition 2.4, standard existence results for SPDEs can be used as a starting point for inferring higher regularity. In contrast, if $(F_{s,\eta})$ does not hold for $s = 0$, we have to prove first that $X \in R(s_0)$ for some $s_0 > 0$. This can be achieved by modifying the regularity induction. Typically, the variational approach for SPDEs yields well-posedness of the paths of X in spaces of the form

$$R^{\mathrm{v}}(s) := L^\infty(0,T;H_{s-1}) \cap L^2(0,T;H_s). \tag{2.13}$$

It is still possible to find a condition on F which allows for an analogue of Proposition 2.3. Here, we restrict to the case $F(X)(t) = F(X_t)$.

($F^{\mathrm{v}}_{s,\eta}$) There is a locally bounded $g : [0,\infty) \to [0,\infty)$ such that for $X \in H_s$:

$$\|F(X)\|^2_{s+\eta-2} \leq (1 + \|X\|^2_s) g(\|X\|_{s-1}). \tag{2.14}$$

The next result extends a similar argument for the stochastic Navier–Stokes equations from [CGH11].

Proposition 2.5. *Let $s \in \mathbb{R}$, $\eta > 0$ such that ($F^{\mathrm{v}}_{s,\eta}$) holds true and a.s. $X_0 \in H_{s+\eta-1}$. If a.s. $\bar{X}, \widetilde{X} \in R^{\mathrm{v}}(s)$, then $\widetilde{X} \in R^{\mathrm{v}}(s+\eta)$ a.s.*

Proof. With $P_N H \simeq \mathbb{R}^N$, $\widetilde{X}^N = P_N \widetilde{X}$ is a process in $C^1(0,T;\mathbb{R}^N)$. The chain rule gives for $0 \leq t \leq T$:

$$\left\| \widetilde{X}^N_t \right\|^2_{s+\eta-1} = \|P_N X_0\|^2_{s+\eta-1} + 2 \int_0^t \left\langle \widetilde{X}^N_r, \theta A \widetilde{X}^N_r + P_N F(X_r) \right\rangle_{s+\eta-1} \mathrm{d}r,$$

and consequently,

$$\begin{aligned} \sup_{0\leq t\leq T} \left\| \widetilde{X}^N_t \right\|^2_{s+\eta-1} &+ 2\theta \int_0^T \left\| \widetilde{X}^N_t \right\|^2_{s+\eta} \mathrm{d}t \\ &\leq \|X_0\|^2_{s+\eta-1} + 2 \int_0^T \left\langle \widetilde{X}^N_t, P_N F(X_t) \right\rangle_{s+\eta-1} \mathrm{d}t. \end{aligned}$$

The last term can be estimated as

$$\begin{aligned} 2 \int_0^T \left\langle \widetilde{X}^N_t, P_N F(X_t) \right\rangle_{s+\eta-1} \mathrm{d}t &\leq 2 \int_0^T \left\| \widetilde{X}^N_t \right\|_{s+\eta} \|P_N F(X_t)\|_{s+\eta-2} \mathrm{d}t \\ &\leq \theta \int_0^T \left\| \widetilde{X}^N_t \right\|^2_{s+\eta} \mathrm{d}t + \frac{1}{2\theta} \int_0^T \|F(X_t)\|^2_{s+\eta-2} \mathrm{d}t. \end{aligned}$$

Finally, using ($F^{\mathrm{v}}_{s,\eta}$) and $X = \bar{X} + \widetilde{X} \in R^{\mathrm{v}}(s)$,

$$\begin{aligned} \sup_{0\leq t\leq T} \left\| \widetilde{X}^N_t \right\|^2_{s+\eta-1} &+ \theta \int_0^T \left\| \widetilde{X}^N_t \right\|^2_{s+\eta} \mathrm{d}t \\ &\leq \|X_0\|^2_{s+\eta-1} + \frac{1}{2\theta} \int_0^T \|F(X_t)\|^2_{s+\eta-2} \mathrm{d}t \\ &\leq \|X_0\|^2_{s+\eta-1} + \frac{1}{2\theta} \sup_{0\leq t\leq T} g\left(\|X_t\|_{s-1}\right) \int_0^T (1 + \|X_t\|^2_s) \mathrm{d}t < \infty. \end{aligned}$$

Thus $(\widetilde{X}^N)_{N\in\mathbb{N}}$ is uniformly bounded in $R^{\mathrm{v}}(s+\eta)$, and the claim follows. □

Analogously to Proposition 2.4, given $\eta > 0$ and $s_0 < s^*$ such that s^* denotes the optimal regularity of $\bar{X}$ and a.s. $X \in R^{\mathrm{v}}(s_0)$ and $X_0 \in H_{s^*+\eta-1}$, if $(F^{\mathrm{v}}_{s,\eta})$ holds for all $s_0 \leq s < s^*$, then $X \in R^{\mathrm{v}}(s)$ and $\widetilde{X} \in R^{\mathrm{v}}(s+\eta)$ for all $s < s^*$. In particular, $X \in R(s-1)$ for all $s < s^*$. The latter statement can be used as a starting point for Proposition 2.4.

Finally, we will need a pathwise regularity statement for stochastic integrals. If a.s. $U \in R(s)$, then $t \mapsto \left\langle (-A)^{s/2} U_t, \cdot \right\rangle$ has values in the space of Hilbert–Schmidt operators from H to $\mathbb{R}$, and $\int_0^T \left\langle (-A)^{s/2} U_t, \mathrm{d}W_t \right\rangle$ is well-defined [DPZ14]. Provided that $U \in L^2(\Omega \times [0,T]; H_s)$, Itô's isometry implies that this integral is approximated in $L^2(\Omega; \mathbb{R})$ by $\int_0^T \left\langle (-A)^{s/2} U^N_t, \mathrm{d}W_t \right\rangle$ as $N \to \infty$, where $U^N = P_N U$. By a classical stopping argument, we have even almost sure convergence, together with a quantification of the divergence rate of the rescaled approximants (cf. Lemma 2.2 (i)), in the following sense:

Lemma 2.6. *Let $s, s' \in \mathbb{R}$ with $s < s'$. For every process U with a.s. $U \in R(s')$, it holds that a.s.*

$$\lim_{N\to\infty} \int_0^T \left\langle (-A)^{s/2} U^N_t, \mathrm{d}W_t \right\rangle = \int_0^T \left\langle (-A)^{s/2} U_t, \mathrm{d}W_t \right\rangle, \tag{2.15}$$

and for every $a > 0$, a.s.

$$\lim_{N\to\infty} \lambda_N^{-a/2} \int_0^T \left\langle (-A)^{(s+a)/2} U^N_t, \mathrm{d}W_t \right\rangle = 0. \tag{2.16}$$

Proof. For $K \in \mathbb{N}$, let $\tau_K := \inf\{0 \leq t \leq T; \sup_{0\leq r\leq t} \|U_r\|^2_{s'} \geq K\} \wedge T$. We abbreviate $Z^N_s(t) := \int_0^t \left\langle (-A)^{s/2} U^N_r, \mathrm{d}W_r \right\rangle$ and $Z_s(t) := \int_0^t \left\langle (-A)^{s/2} U_r, \mathrm{d}W_r \right\rangle$.

Let $\epsilon > 0$ and $p \geq 4/(\beta(s'-s))$. The Burkholder–Davis–Gundy inequality and Lemma 2.2 (ii) give

$$\begin{aligned}
\mathbb{P}\left(\left| Z_s(\tau_K) - Z^N_s(\tau_K) \right| > \epsilon \right) &\leq \epsilon^{-p} \mathbb{E}\left[\left(\int_0^{\tau_K} \|(I - P_N) U_t\|^2_s \, \mathrm{d}t \right)^{p/2} \right] \\
&\leq \epsilon^{-p} \lambda_{N+1}^{p(s-s')/2} \mathbb{E}\left[\left(\int_0^{\tau_K} \|U_t\|^2_{s'} \, \mathrm{d}t \right)^{p/2} \right] \\
&\leq \epsilon^{-p} (KT)^{p/2} \lambda_{N+1}^{p(s-s')/2} \ll N^{-2}.
\end{aligned}$$

The Borel–Cantelli lemma implies that $Z_s^N(\tau_K) \to Z_s(\tau_K)$ a.s. Similarly, using Lemma 2.2 (i),

$$\begin{aligned}
\mathbb{P}\left(\left|\lambda_N^{-a/2} Z_{s+a}^N(\tau_K)\right| > \epsilon\right) &\leq \epsilon^{-p}\lambda_N^{-ap/2}\mathbb{E}\left[\left(\int_0^{\tau_K} \left\|U_t^N\right\|_{s+a}^2 \mathrm{d}t\right)^{p/2}\right] \\
&\leq \epsilon^{-p}\lambda_N^{p(s-s')/2}\mathbb{E}\left[\left(\int_0^{\tau_K} \left\|U_t^N\right\|_{s'}^2 \mathrm{d}t\right)^{p/2}\right] \\
&\leq \epsilon^{-p}(KT)^{p/2}\lambda_N^{p(s-s')/2} \ll N^{-2},
\end{aligned}$$

where we w.l.o.g. assume that $s + a > s'$ (otherwise take s' to be smaller). Again by the Borel–Cantelli lemma, $\lambda_N^{-a/2} Z_{s+a}^N(\tau_K) \to 0$ a.s.

Consequently, (2.15), (2.16) are true on the set $A_K := \{\tau_K = T\}$. The claim follows as $\bigcup_{K\in\mathbb{N}} A_K$ has probability one. $\square$

2.3 Diffusivity Estimation

For $k \in \mathbb{N}$, we set $\bar{x}^{(k)} := \langle \bar{X}, \Phi_k \rangle$. Then $(\bar{x}^{(k)})_{k\in\mathbb{N}}$ are independent one-dimensional Ornstein–Uhlenbeck processes that solve

$$\mathrm{d}\bar{x}_t^{(k)} = -\theta\lambda_k \bar{x}_t^{(k)}\mathrm{d}t + \sigma\lambda_k^{-\gamma}\mathrm{d}W_t^{(k)}, \tag{2.17}$$

$\bar{x}_0^{(k)} = 0$, where $(W^{(k)})_{k\in\mathbb{N}}$ are independent Brownian motions. $\bar{x}^{(k)}$ has the explicit representation

$$\bar{x}_t^{(k)} = \sigma\lambda_k^{-\gamma} \int_0^t e^{-\theta\lambda_k(t-r)}\mathrm{d}W_r^{(k)}, \tag{2.18}$$

and consequently,

$$\mathbb{E}[(\bar{x}_t^{(k)})^2] = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\lambda_k t})\lambda_k^{-2\gamma-1}. \tag{2.19}$$

Lemma 2.7. *For any $s < s^* := 1 + 2\gamma + 1/\beta$, it holds a.s. $\bar{X} \in R(s)$.*

Proof. For $0 < \alpha < 1/2 \wedge (s^* - s)/2$,

$$\begin{aligned}\int_0^T t^{-2\alpha} \left\|(-A)^{s/2} e^{t\theta A} B\right\|_{\mathrm{HS}}^2 \mathrm{d}t &= \sigma \sum_{k=1}^{\infty} \lambda_k^{s-2\gamma} \int_0^T t^{-2\alpha} e^{-2\theta\lambda_k t} \mathrm{d}t \\ &\lesssim \sigma \sum_{k=1}^{\infty} \lambda_k^{s-2\gamma} \int_0^{\infty} \left(\frac{r}{2\theta\lambda_k}\right)^{-2\alpha} \frac{e^{-r}}{2\theta\lambda_k} \mathrm{d}r \\ &\lesssim \Gamma(1-2\alpha) \sum_{k=1}^{\infty} \lambda_k^{s-2\gamma-1+2\alpha} \lesssim \sum_{k=1}^{\infty} k^{\beta(s-2\gamma-1+2\alpha)},\end{aligned}$$

where Γ denotes the Gamma function. The last sum is finite since $\beta(s - 2\gamma - 1 + 2\alpha) < -1$ for $\alpha < (s^* - s)/2$. Now, by [DPZ14, Theorem 5.11], $(-A)^{s/2}\bar{X} \in R(0)$, i.e. $\bar{X} \in R(s)$ a.s. $\square$

In fact, the proof of [DPZ14, Theorem 5.11] shows that even $(-A)^{s/2}\bar{X} \in R^{\mathbb{E}}(0)$ in the situation of the proof of Lemma 2.7, thus $\bar{X} \in R^{\mathbb{E}}(s)$ for $s < s^*$.

Proposition 2.8. *With $s^* = 1 + 2\gamma - 1/\beta$, let $\eta > 0$, $s_0 < s^*$ such that $(F_{s,\eta})$ holds for any $s_0 \le s < s^*$. Assume that a.s. $X \in R(s_0)$, $X_0 \in H_{s^*+\eta}$. Then a.s. for any $s > s^*$:*

$$\int_0^T \left\|(-A)^{s/2} X_t^N\right\|^2 \mathrm{d}t \asymp C_s N^{1+\beta(s-2\gamma-1)} \tag{2.20}$$

with

$$C_s = \frac{\sigma^2 T \Lambda^{s-2\gamma-1}}{2\theta(1+\beta(s-2\gamma-1))} \tag{2.21}$$

Proof. Integrating (2.19), we see that

$$\mathbb{E} \int_0^T (\bar{x}_t^{(k)})^2 \mathrm{d}t \asymp \frac{\sigma^2 T}{2\theta} \lambda_k^{-2\gamma-1}, \tag{2.22}$$

thus, using $\lambda_k \asymp \Lambda k^\beta$,

$$\begin{aligned}\mathbb{E} \int_0^T \left\|(-A)^{s/2} \bar{X}_t^N\right\|^2 \mathrm{d}t &= \sum_{k=1}^{N} \lambda_k^s \mathbb{E} \int_0^T (\bar{x}_t^{(k)})^2 \mathrm{d}t \asymp \frac{\sigma^2 T}{2\theta} \sum_{k=1}^{N} \lambda_k^{s-2\gamma-1} \\ &\asymp \frac{\sigma^2 T \Lambda^{s-2\gamma-1}}{2\theta(1+\beta(s-2\gamma-1))} N^{1+\beta(s-2\gamma-1)}.\end{aligned}$$

Lemma A.2 (ii), with $X_k^*(t) = \lambda_k^{s/2}\bar{x}_t^{(k)}$ in the notation therein, immediately gives that (2.20) is true for X^N replaced by $\bar{X}^N$. Now, for any $0 < \epsilon < \eta$, by Proposition 2.4:

$$\int_0^T \left\|(-A)^{s/2}\widetilde{X}_t^N\right\|^2 \mathrm{d}t \lesssim \lambda_N^{s-s^*-\eta+\epsilon} \int_0^T \left\|\widetilde{X}_t^N\right\|_{s^*+\eta-\epsilon}^2 \mathrm{d}t \lesssim N^{1+\beta(s-2\gamma-1-\eta+\epsilon)}.$$

This is negligible compared to the right-hand side of (2.20), and the claim follows from expanding the square on the left-hand side of (2.20) together with the Cauchy–Schwarz inequality for the mixed term. □

Remark 2.9. *In the setting of the previous proposition, it is immediate that for the limit case $s = s^*$, we have a.s.*

$$\int_0^T \left\|(-A)^{s^*/2}X_t^N\right\|^2 \mathrm{d}t \asymp \frac{\sigma^2 T}{2\theta\Lambda^{1/\beta}} \ln(N), \tag{2.23}$$

with obvious changes in the proof. In particular, as the right-hand side diverges, $X \notin R(s^*)$, and the regularity from Lemma 2.7 is optimal.

Next, we derive three maximum–likelihood type estimators for θ (cf. [CGH11]). The projected process $X^N = P_N X$ induces a measure $\mathbb{P}_\theta^{N,T}$ on the path space $C(0,T;P_N H) \simeq C(0,T;\mathbb{R}^N)$ for each value of the diffusivity $\theta > 0$. If we fix an arbitrary reference parameter $\theta_0 > 0$ and assume that each of the measures $(\mathbb{P}_\theta^{N,T})_{\theta>0}$ is absolutely continuous with respect to $\mathbb{P}_{\theta_0}^{N,T}$, we obtain a likelihood that we can use for statistical inference. According to [LS77, Section 7.6.4], the log–likelihood is formally given by

$$\ln \frac{\mathrm{d}\mathbb{P}_\theta^{N,T}}{\mathrm{d}\mathbb{P}_{\theta_0}^{N,T}}(X^N) = \frac{1}{\sigma^2}\int_0^T \left\langle (\theta-\theta_0)AX_t^N, (-A)^{2\gamma}\mathrm{d}X_t^N \right\rangle$$
$$- \frac{1}{2\sigma^2}\int_0^T \left\langle (\theta-\theta_0)AX_t^N, (-A)^{2\gamma}\left((\theta+\theta_0)AX_t^N + 2P_N F(X)(t)\right)\right\rangle \mathrm{d}t.$$

This is rigorous if $P_N F = F P_N$, otherwise it should be considered as a natural (but heuristic) approach. Maximizing for θ yields the following maximum likelihood–type estimator:

$$\hat{\theta}_N^{\mathrm{full}} := -\frac{\int_0^T \left\langle (-A)^{1+2\alpha}X_t^N, \mathrm{d}X_t^N\right\rangle}{\int_0^T \left\|(-A)^{1+\alpha}X_t^N\right\|^2 \mathrm{d}t} + \frac{\int_0^T \left\langle (-A)^{1+2\alpha}X_t^N, P_N F(X)(t)\right\rangle \mathrm{d}t}{\int_0^T \left\|(-A)^{1+\alpha}X_t^N\right\|^2 \mathrm{d}t}, \tag{2.24}$$

where we substituted γ by an additional parameter α. This estimator depends on $P_N F(X)$ and is therefore not closed in X^N. It can be modified as follows:

$$\hat{\theta}_N^{\text{part}} := -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, \mathrm{d}X_t^N \right\rangle}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|^2 \,\mathrm{d}t} + \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F(X^N)(t) \right\rangle \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|^2 \,\mathrm{d}t}. \tag{2.25}$$

Finally, the nonlinear term can be left out completely:

$$\hat{\theta}_N^{\text{lin}} := -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, \mathrm{d}X_t^N \right\rangle}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|^2 \,\mathrm{d}t}. \tag{2.26}$$

Note that the stochastic integral appearing in the numerator of each of the estimators has a robust representation

$$\begin{aligned}\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, \mathrm{d}X_t^N \right\rangle &= \sum_{k=1}^N \lambda_k^{1+2\alpha} \int_0^T x_t^{(k)} \mathrm{d}x_t^{(k)} \\ &= \frac{1}{2} \sum_{k=1}^N \lambda_k^{1+2\alpha} \left((x_T^{(k)})^2 - (x_0^{(k)})^2 - \sigma^2 \lambda_k^{-2\gamma} T \right),\end{aligned}$$

so it is a function of a single trajectory of X^N alone.

The aim of this section is to study the asymptotic properties of these estimators as $N \to \infty$. Note that one cannot directly apply the general theory for maximum likelihood estimation, as exposed e.g. in [IH81], because for $P_N F \neq F P_N$, none of these estimators is the true MLE.

Lemma 2.10. *Let $s \in \mathbb{R}$, $\epsilon > 0$. For any process $U \in R(s)$, we have a.s.:*

$$\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N U_t \right\rangle \mathrm{d}t \lesssim N^{\frac{1}{2}+\beta\left(2\alpha-\gamma+\frac{1}{2}-\frac{s}{2}+\frac{\epsilon}{2}\right)}. \tag{2.27}$$

In particular, if $s = s^* - 2 + \eta - \epsilon$ for some $\eta > 0$, then

$$\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N U_t \right\rangle \mathrm{d}t \lesssim N^{1+\beta\left(2\alpha-2\gamma+1-\frac{\eta}{2}+\epsilon\right)}. \tag{2.28}$$

Proof. This follows from

$$\begin{aligned}\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N U_t \right\rangle \mathrm{d}t &\leq \left(\int_0^T \left\| X_t^N \right\|_{2+4\alpha-s}^2 \mathrm{d}t \int_0^T \left\| P_N U_t \right\|_s^2 \mathrm{d}t \right)^{1/2} \\ &\lesssim \left(\lambda_N^{2+4\alpha-s-s^*+\epsilon} \int_0^T \left\| X_t^N \right\|_{s^*-\epsilon}^2 \mathrm{d}t \right)^{1/2} \\ &\lesssim \lambda_N^{2\alpha-\gamma-\frac{s}{2}+\frac{1}{2}+\frac{\beta^{-1}}{2}+\frac{\epsilon}{2}} \lesssim N^{\frac{1}{2}+\beta\left(2\alpha-\gamma+\frac{1}{2}-\frac{s}{2}+\frac{\epsilon}{2}\right)}.\end{aligned}$$

□

An estimator $\hat{\theta}_N$ for θ is called strongly consistent if a.s. $\hat{\theta}_N \to \theta$.

Theorem 2.11. *Let $\eta > 0$, $s_0 \in \mathbb{R}$ such that $(F_{s,\eta})$ is true for $s_0 \leq s < s^*$. Assume that $X_0 \in H_{s^*+\eta}$ and $X \in R(s_0)$. Let $\alpha > \gamma - (1 + 1/\beta)/4$.*

(i) *$\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ are strongly consistent estimators of θ.*

(ii) *$\hat{\theta}_N^{\text{full}}$ is asymptotically normal:*

$$N^{\frac{1+\beta}{2}}(\hat{\theta}_N^{\text{full}} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{2.29}$$

where

$$\Sigma = \frac{2\theta(1 + \beta(2\alpha - 2\gamma + 1))^2}{T\Lambda(1 + \beta(4\alpha - 4\gamma + 1))}. \tag{2.30}$$

(iii) *If $\eta > 1 + \beta^{-1}$, then $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ are asymptotically normal as in (2.29). Otherwise, for any $a < \beta\eta/2$,*

$$\hat{\theta}_N^{\text{part}} = \theta + o(N^{-a}) \tag{2.31}$$

almost surely, and the same is true for $\hat{\theta}_N^{\text{lin}}$.

Proof.

(i) This is a consequence of (ii) and (iii).

(ii) Plugging in the dynamics of X^N into $\hat{\theta}_N^{\text{full}}$, we obtain

$$\begin{aligned}\hat{\theta}_N^{\text{full}} - \theta &= -\sigma \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t} \\ &= -\sigma \frac{C_{2+4\alpha-2\gamma}^{1/2} N^{1/2+\beta(2\alpha-2\gamma+1/2)}}{\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t} M_T^N,\end{aligned}$$

where $M_t^N = C_{2+4\alpha-2\gamma}^{-1/2} N^{-1/2-\beta(4\alpha-4\gamma+1/2)} \int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle$ is a local martingale. By Proposition 2.8 with $s = 2+4\alpha-2\gamma$ it holds $\langle M^N\rangle_T \to 1$ in probability, thus $M_T^N / \sqrt{\langle M^N \rangle_T} \to \mathcal{N}(0,1)$ in distribution as $N \to \infty$ by Theorem A.1. An application of Slutsky's lemma together with Proposition 2.8 with $s = 2+2\alpha$ gives

$$N^{\frac{1+\beta}{2}} (\hat{\theta}_N^{\text{full}} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 C_{2+4\alpha-2\gamma} / C_{2+2\alpha}^2). \tag{2.32}$$

(iii) We write

$$\begin{aligned}\hat{\theta}_N^{\text{part}} - \theta &= -\sigma \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t} - \mathrm{bias}_N(X) + \mathrm{bias}_N(X^N), \\ \hat{\theta}_N^{\text{lin}} - \theta &= -\sigma \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t} - \mathrm{bias}_N(X),\end{aligned}$$

where, with $Y = X$ or $Y = X^N$,

$$\mathrm{bias}_N(Y) = \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F(Y)(t) \right\rangle \mathrm{d}t}{\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t}. \tag{2.33}$$

Let $\epsilon > 0$ and $s = s^* + \eta - 2 - \epsilon = 2\gamma - 1 - \beta^{-1} + \eta - \epsilon$. Using condition $(F_{s^*-2\epsilon,\eta})$, we have that $F(X) \in R(s)$. By Lemma 2.10,

$$\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F(Y)(t) \right\rangle \mathrm{d}t \lesssim N^{1+\beta\left(2\alpha-2\gamma+1-\frac{\eta}{2}+\epsilon\right)},$$

and using Proposition 2.8, a.s.

$$\begin{aligned}\mathrm{bias}_N(Y) &\asymp C_{2+2\alpha}^{-1} N^{-1-\beta(2\alpha-2\gamma+1)} \int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F(Y)(t) \right\rangle \mathrm{d}t \\ &\lesssim N^{-\frac{\beta}{2}(\eta-2\epsilon)}.\end{aligned}$$

Consequently, $N^a \text{bias}_N(Y) \to 0$ a.s. for any $a < \beta\eta/2$.

Now, if $\eta > 1 + 1/\beta$, let in addition $\epsilon < (\eta - 1 - 1/\beta)/2$, then we see that a.s. $N^{(1+\beta)/2}\text{bias}_N(Y) \to 0$, and asymptotic normality follows from an application of the Slutsky lemma.

Otherwise, in case $\eta \leq 1 + 1/\beta$, we have by Lemma 2.6 (setting $a = 2 + 4\alpha - 2\gamma - s^* + \epsilon'$ for any $\epsilon' > 0$ in the notation therein) that

$$N^b \frac{\int_0^T \langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \rangle}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|^2 \,\mathrm{d}t} \to 0 \tag{2.34}$$

almost surely for any $b < (1+\beta)/2$, and (2.31) is immediate.

□

Any asymptotically normal estimator in Theorem 2.11 allows to construct asymptotic confidence intervals for θ by using quantiles of the approximating normal distribution $\mathcal{N}(\theta, N^{-(1+\beta)}\Sigma)$ for fixed $N \in \mathbb{N}$. However, the asymptotic variance Σ depends linearly on the unknown parameter θ. Thus, in order to construct asymptotic confidence intervals for the diffusivity that do not depend on unknown quantities, the variance itself has to be estimated consistently using any of the three estimators. This is justified by Slutsky's lemma. Alternatively, a variance-stabilizing transform can be used [vdV98, Section 3.2].

Consistency of any of the three estimators implies that the measures on $C(0,T;H)$ generated by X for different values of $\theta > 0$ are mutually singular.

In the setting of this section, it is possible to determine the precise rate of almost sure convergence of the estimators by a law of iterated logarithm:

Theorem 2.12. *Let $\eta > 0$, $s_0 \in \mathbb{R}$ such that $(F_{s,\eta})$ is true for $s_0 \leq s < s^*$. Assume $X_0 \in H_{s^*+\eta}$ and $X \in R(s_0)$. Let $\alpha > \gamma - (1 + 1/\beta)/4$. Then a.s.*

$$\limsup_{N\to\infty} \frac{N^{\frac{1+\beta}{2}}}{\sqrt{\ln(\ln(N))}} \left(\hat{\theta}_N^{\text{full}} - \theta\right) = \sqrt{2\Sigma} \tag{2.35}$$

with Σ as in (2.30). If $\eta > 1 + 1/\beta$, then (2.35) is true for $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ as well.

Proof. We write

$$\bar{M}_t^N := \int_0^t \left\langle (-A)^{1+2\alpha-\gamma} \bar{X}_r^N, \mathrm{d}W_r^N \right\rangle,$$
$$\widetilde{M}_t^N := \int_0^t \left\langle (-A)^{1+2\alpha-\gamma} \widetilde{X}_r^N, \mathrm{d}W_r^N \right\rangle.$$

Then

$$\bar{M}_T^N = \sum_{k=1}^N \lambda_k^{1+2\alpha-\gamma} \int_0^T \bar{x}_t^{(k)} \mathrm{d}W_t^{(k)} =: \sum_{k=1}^N Z_k.$$

We show that we can apply the law of iterated logarithm for independent, not necessarily identically distributed random variables from [Wit85][2] to $(Z_k)_{k\in\mathbb{N}}$. To this end, we write

$$s_N := \left(\mathbb{E}\left[\sum_{k=1}^N Z_k^2\right]\right)^{\frac{1}{2}} = \left(\mathbb{E}\int_0^T \left\|(-A)^{1+2\alpha-\gamma}\bar{X}_t^N\right\|^2 \mathrm{d}t\right)^{\frac{1}{2}}.$$

Then clearly $s_N \to \infty$ and $s_N \asymp s_{N+1}$. Using the Burkholder–Davis–Gundy inequality, Jensen's inequality, Gaussianity of $\bar{x}^{(k)}$ and (2.19), we have

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^T \bar{x}_t^{(k)} \mathrm{d}W_t^{(k)}\right|^3\right] &\lesssim \mathbb{E}\left[\left(\int_0^T (\bar{x}_t^{(k)})^2 \mathrm{d}t\right)^{\frac{3}{2}}\right] \lesssim \mathbb{E}\int_0^T \left|\bar{x}_t^{(k)}\right|^3 \mathrm{d}t \\ &\lesssim \int_0^T \left(\mathbb{E}[(\bar{x}_t^{(k)})^2]\right)^{\frac{3}{2}} \mathrm{d}t \lesssim \left(\lambda_k^{-2\gamma-1}\right)^{\frac{3}{2}}.\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{k=1}^N \frac{\mathbb{E}[|Z_k|^3]}{s_k^3} &\lesssim \sum_{k=1}^N \left(\frac{\lambda_k^{2\alpha-2\gamma+1/2}}{s_k}\right)^3 \\ &\lesssim \sum_{k=1}^N k^{3\beta(2\alpha-2\gamma+1/2)-3/2-3\beta(2\alpha-2\gamma+1/2)} = \sum_{k=1}^N \frac{1}{k^{3/2}},\end{aligned}$$

[2] Although it is sufficient for our purposes, this law of iterated logarithm can be further weakened, see e.g. [Wit87] and [Che93] for a discussion.

which converges for $N \to \infty$. Therefore all the conditions from [Wit85] are satisfied, and using $\ln\ln s_N^2 \asymp \ln\ln N$, we conclude that a.s.

$$\limsup_{N\to\infty} \frac{1}{\sqrt{2\ln\ln N}} \frac{\bar{M}_T^N}{\sqrt{\langle \bar{M} \rangle_T}} = \limsup_{N\to\infty} \frac{\bar{M}_T^N}{\sqrt{2 s_N^2 \ln\ln s_N^2}} = 1.$$

In particular, $\limsup_{N\to\infty} N^{(1+\beta)/2} (\ln\ln N)^{-1/2} \bar{M}_T^N / \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t = (2\Sigma)^{1/2}/\sigma$. Further, by Lemma 2.6 (with $a = 2 + 4\alpha - 2\gamma - s^*$ in the notation therein), $N^{(1+\beta)/2} \widetilde{M}_T^N / \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t \to 0$ almost surely, where we have used that $\widetilde{X} \in R(s^* + \eta - \epsilon)$ for every $\epsilon > 0$. With $\hat{\theta}_N^{\mathrm{full}} - \theta = -\sigma(\bar{M}_T^N + \widetilde{M}_T^N) / \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t$, (2.35) is proven. The statement concerning $\hat{\theta}_N^{\mathrm{part}}$ and $\hat{\theta}_N^{\mathrm{lin}}$ follows from the proof of Theorem 2.11 (iii). □

Remark 2.13.

(i) A direct calculation shows that the asymptotic variance is minimal for $\alpha = \gamma$. In this case, $\Sigma = 2\theta(1+\beta)/T\Lambda$. However, the estimators are robust to the case $\alpha \neq \gamma$, when the spatial regularity of X is wrongly specified.

(ii) In fact, continuous observation on $[0, T]$ of any of the modes $x^{(k)}$ allows to reconstruct γ precisely via the quadratic variation $\langle x^{(k)} \rangle_T = \sigma^2 \lambda_k^{-2\gamma} T$, if σ and λ_k are known.

(iii) Σ depends linearly on T^{-1}, i.e. observation on a large time interval improves the estimate. This corresponds to the large time asymptotics with rate $T^{-1/2}$, which is well-known from statistics for stochastic differential equations under ergodicity assumptions, see e.g. [Kut04].

(iv) Following the formalism of the (heuristic) maximum–likelihood approach, the term $\sigma^{-2} \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t$ (for $\alpha = \gamma$) can be considered the observed Fisher information.

(v) The convergence rate of $\hat{\theta}_N^{\mathrm{full}}$ is upper bounded by $N^{-1/2}$ independently of the dimension d.

(vi) The convergence rate (2.31) cannot be improved for $\hat{\theta}_N^{\mathrm{lin}}$, see Section 2.4.1.

(vii) *In [PS20], a condition of the type* $(F^{\mathrm{v}}_{s,\eta})$ *has been used to infer higher regularity of* $\widetilde{X}$. *Compared to* $(F_{s,\eta})$, *this condition is more restrictive and yields lower excess regularity* η *in examples (cf. Lemma 2.22 below). Consequently, the lower bounds on the convergence rate of* $\hat{\theta}_N^{\mathrm{lin}}$ *were too pessimistic. An additional Lipschitz condition on* F *has been used in order to reduce the asymptotic behavior of* $\hat{\theta}_N^{\mathrm{part}}$ *directly to that of* $\hat{\theta}_N^{\mathrm{full}}$, *leading to better convergence rates of* $\hat{\theta}_N^{\mathrm{part}}$ *compared to* $\hat{\theta}_N^{\mathrm{lin}}$, *namely, the same rates as stated in Theorem 2.11. In the present formalism used in this chapter, this additional Lipschitz condition is no longer necessary for statistical purposes, as both* $\hat{\theta}_N^{\mathrm{part}}$ *and* $\hat{\theta}_N^{\mathrm{lin}}$ *obtain the mentioned rates using* $(F_{s,\eta})$ *alone.*

(viii) *If* F *satisfies* $(F_{s,\eta})$ *and* $P_NF - P_NFP_N$ *satisfies* $(F_{s,\eta'})$ *for some* $\eta' > \eta$, *then the convergence rate of* $\hat{\theta}_N^{\mathrm{part}}$ *can be further improved. This is trivially the case if* $[P_N, F] := P_NF - FP_N = 0$, *in this case* $\hat{\theta}_N^{\mathrm{part}}$ *coincides with* $\hat{\theta}_N^{\mathrm{full}}$.

(ix) *Consider the situation that* A *and* F *are (pseudo-) differential operators on a domain* $\mathcal{D} \subset \mathbb{R}^d$. *Theorem 2.11 implies that in order to identify* θ *in finite time, it is sufficient that* F *is of lower order compared to the leading order drift term* θA *(this is discussed in detail in Section 2.4). However, parameters describing the intensity of lower order terms may be identified in finite time as well. Consequently, it is possible that* θ *remains identifiable if* F *is of higher order than* θA. *In [HR95], a characterization for linear* F *that commute with* A *is given:* θ *is consistently (and asymptotically normal) estimated by an estimator of the type* $\hat{\theta}_N^{\mathrm{full}}$ *if and only if* $\mathrm{order}(A) \geq (\mathrm{order}(\theta A + F) - d)/2$, *i.e.* $\mathrm{order}(F) \leq 2\,\mathrm{order}(A) + d$. *This has been extended by subsequent works on the spectral approach, cf. [LR99, LR00] for the case of non-commuting operators.*

(x) *As only pathwise properties of the nonlinear process* $\widetilde{X}$ *are needed,* F *may, in fact, depend on the realization* $\omega \in \Omega$.

2.4 Discussion of Examples

Next, we discuss the validity of condition $(F_{s,\eta})$ and the resulting statements concerning diffusivity estimation for models with different nonlinear term F.

These examples are by no means exhaustive. Note that more complicated nonlinearities can be decomposed into their elementary building blocks in the following sense:

Lemma 2.14. *Let $s \in \mathbb{R}$, $\eta > 0$.*

(i) If F_1, F_2 satisfy $(F_{s,\eta})$, then the same is true for $F_1 + F_2$.

(ii) Let $\eta' > 0$ and $s' := s + \eta - 2$. If F satisfies $(F_{s,\eta})$ and G satisfies $(F_{s',\eta'})$, then $G \circ F$ satisfies $(F_{s,\eta+\eta'-2})$.

Proof. All statements are clear from (2.10). □

Consequently, a broad class of models where the present theory is applicable can be constructed from elementary components, e.g. polynomial nonlinearities (or related reaction terms), differential operators acting in spatial direction (advection or fractional diffusion), integration in time (delay terms as in [DPZ14, Example 5.6]). In the next sections, we consider different models in detail.

2.4.1 Linear Perturbations

For $r < 2$ and $c \in \mathbb{R}$, consider

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + c(-A)^{r/2} X_t \mathrm{d}t + B \mathrm{d}W_t$$

with initial condition $X_0 \in H_{s^*+2}$. Here, $F(X) = c(-A)^{r/2}X$. If $c \neq 0$, then $F : H_{s+r} \to H_s$ is an isomorphism for any $s \in \mathbb{R}$. In particular, $(F_{s,\eta})$ is true for all $s \in \mathbb{R}$ and $\eta < 2 - r$. In this setting, $\hat{\theta}_N^{\text{full}}$ coincides with $\hat{\theta}_N^{\text{part}}$. We have:

Theorem 2.15. *Let $\alpha > \gamma - (1 + 1/\beta)/4$. Then $\hat{\theta}_N^{\text{full}}$ is asymptotically normal as in (2.29). Furthermore:*

(i) If $r < 1 - 1/\beta$, then $\hat{\theta}_N^{\text{lin}}$ is asymptotically normal as in (2.29).

(ii) If $r = 1 - 1/\beta$, then

$$N^{\frac{1+\beta}{2}} \left(\hat{\theta}_N^{\text{lin}} - \theta \right) \xrightarrow{d} \mathcal{N}(\kappa, \Sigma), \tag{2.36}$$

where

$$\kappa = -c\Lambda^{r/2-1}\frac{1+\beta(2\alpha-2\gamma+1)}{1+\beta(2\alpha-2\gamma+r/2)} \tag{2.37}$$

and Σ *is given by* (2.30)*.*

(iii) If $r > 1-1/\beta$*, then a.s.*

$$N^{\beta(1-r/2)}\left(\hat{\theta}_N^{\mathrm{lin}} - \theta\right) \to \kappa. \tag{2.38}$$

Proof. (i) is a direct consequence of Theorem 2.11. For (ii), (iii), it suffices to understand the exact asymptotics of the bias term involving F in the setting of Theorem 2.11. Due to $\alpha > \gamma - (1+1/\beta)/4$ together with $r \geq 1-1/\beta$, it holds $1+\beta(2\alpha-2\gamma+r/2) > 0$. Using Proposition 2.8 and the notation therein, it holds a.s.

$$\begin{aligned}
\frac{\int_0^T \left\langle (-A)^{1+2\alpha}X_t^N, P_N F(X)(t)\right\rangle \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha}X_t^N\|^2 \mathrm{d}t} &= c\frac{\int_0^T \left\|(-A)^{1/2+r/4+\alpha}X_t^N\right\|^2 \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha}X_t^N\|^2 \mathrm{d}t} \\
&\asymp c\frac{C_{1+r/2+2\alpha}N^{1+\beta(2\alpha-2\gamma+r/2)}}{C_{2+2\alpha}N^{1+\beta(2\alpha-2\gamma+1)}} \\
&= c\Lambda^{r/2-1}\frac{1+\beta(2\alpha-2\gamma+1)}{1+\beta(2\alpha-2\gamma+r/2)}N^{-\beta(1-r/2)}.
\end{aligned}$$

Now (ii) is immediate, and (2.38) holds w.r.t. convergence in probability. Finally, Lemma 2.6 yields almost sure convergence in (2.38) by the same argument used in the proof of Theorem 2.11 (iii). □

Remark 2.16. *In particular, the convergence rate* (2.31) *for* $\hat{\theta}_N^{\mathrm{lin}}$ *as stated in Theorem 2.11 cannot be improved.*

In case $\beta = 2/d$ for $d \in \mathbb{N}$, the critical condition $r < 1-1/\beta$ is equivalent to $r < 1-d/2$, i.e. the critical order of F decreases with the dimension. Note that r is allowed to be negative here. We highlight two cases, which will be refined in the next sections:

- *Perturbation of order zero (*$r = 0$*):* In $d = 1$, $\hat{\theta}_N^{\mathrm{full}}$ and $\hat{\theta}_N^{\mathrm{lin}}$ are asymptotically normal. In $d = 2$, $\hat{\theta}_N^{\mathrm{lin}}$ still converges to θ with optimal rate. In $d \geq 3$, the convergence rate of $\hat{\theta}_N^{\mathrm{lin}}$ declines.

- *Perturbation of order one (*$r = 1$*):* In any dimension $d \geq 1$ the convergence rate of $\hat{\theta}_N^{\mathrm{lin}}$ declines compared to that of $\hat{\theta}_N^{\mathrm{full}}$, but all estimators stay consistent.

2.4.2 Reaction-Diffusion Equations

Let $d \geq 1$, and let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain with smooth boundary. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function. Consider

$$\mathrm{d}X_t = \theta \Delta X_t \mathrm{d}t + f(X_t)\mathrm{d}t + B\mathrm{d}W_t, \tag{2.39}$$

together with initial condition X_0 such that $\mathbb{E}\left[\|X_0\|_{s^*+2}^p\right] < \infty$ is true for any $p \geq 1$. W.l.o.g. we assume Dirichlet boundary conditions, i.e. $X_t = 0$ on the boundary $\partial\mathcal{D}$ for $0 \leq t \leq T$. Set $H = L^2(\mathcal{D})$. The leading order linear operator A is given by $\Delta : D(\Delta) \to H$, where $D(\Delta) = W^{2,2}(\mathcal{D}) \cap W_0^{1,2}(\mathcal{D})$. The regularity scale is given by $H_s = D((-\Delta)^{s/2})$, such that H_s consists of functions of L^2-Sobolev regularity s. It is always true that $W_0^{s,2}(\mathcal{D}) \subseteq H_s \subseteq W^{s,2}(\mathcal{D})$, $s \geq 0$. For $s \in \mathbb{N}$, a precise characterization of H_s in terms of boundary trace operators can be given [Tho06, Lemma 3.1]. Further, it is well-known that $\lambda_k \asymp \Lambda k^\beta$ with $\beta = 2/d$, see [Wey11] or [Shu01, Section 13.4].

We consider either of the following two structural assumptions:

(i) f is a polynomial of odd degree and negative leading coefficient, i.e. for $m \in 2\mathbb{N} - 1$, there are $a_0, \ldots, a_m \in \mathbb{R}$ with $a_m < 0$ such that

$$f(x) = a_m x^m + \cdots + a_1 x + a_0. \tag{2.40}$$

(ii) f is a bounded smooth functions with bounded derivatives of any order:

$$f \in C_b^\infty(\mathbb{R}). \tag{2.41}$$

Reaction–Diffusion models exhibit a broad variety of different dynamical features. Nonetheless, in terms of diffusivity estimation, they can be treated in a unified way, as explained in Theorem 2.20 below.

Proposition 2.17. *Let $d \leq 3$ and $\gamma > d/4 + 1/2$. Consider either of the following two situations:*

- *Let f be a polynomial as in* (2.40)*. If $d = 3$, assume additionally that f is at most of third order (i.e. $m \leq 3$).*
- *Let f be a smooth function with bounded derivatives as in* (2.41)*.*

Then there is a unique solution X to (2.39) *with $X \in R^{\mathbb{E}}(s)$ for some $s > d/2$.*

The proof relies on [LR15] and is given in Appendix B.1. The condition $\gamma > d/4 + 1/2$ means that B is a Hilbert–Schmidt operator from H into V. This is needed because in the proof of Proposition 2.17, coercivity is verified directly in V instead of H. However, there are situations where this requirement can be relaxed to $\gamma > d/4$, i.e. B is of Hilbert–Schmidt type from H into H:

Proposition 2.18. *Let $d = 1$, $\gamma > 1/4$ and $f(x) = x - x^3$.*

(i) There is a unique solution $X \in R^{\mathrm{v}}(1)$ to (2.39).

(ii) $(F^{\mathrm{v}}_{s,\eta})$ holds for $s = 1$ and $\eta = 1$. In particular, even $X \in R(1)$.

Note that condition $(F^{\mathrm{v}}_{s,\eta})$ instead of $(F_{s,\eta})$ is used in order to prove the basic regularity result $X \in R(1)$. From there, $(F_{s,\eta})$ can be used to infer higher regularity.

Proof.

(i) This is a special case of [LR15, Example 5.1.8].

(ii) $(F^{\mathrm{v}}_{s,\eta})$ is true due to

$$\begin{aligned}\|f(X)\|^2_{s+\eta-2} = \|f(X)\|^2_{L^2(\mathcal{D})} \lesssim \|X\|^2_{L^2(\mathcal{D})} + \|X\|^6_{L^6(\mathcal{D})} \lesssim \|X\|^2_1 + \|X\|^6_{1/3} \\ \lesssim \|X\|^2_1 + \|X\|^2_1 \|X\|^4_0 = \|X\|^2_1 \left(1 + \|X\|^4_0\right),\end{aligned}$$

where we used $\|X^3\|^2_{L^2(\mathcal{D})} = \|X\|^6_{L^6(\mathcal{D})}$ together with the Sobolev embedding $H_{1/3} \subset L^6(\mathcal{D})$ in $d = 1$ [AF03]. Proposition 2.5 implies $\widetilde{X} \in R^{\mathrm{v}}(2) \subset R(1)$. Together with $\bar{X} \in R(1)$ due to $\gamma > 1/4$, this implies the claim.

□

Since $d = 1$ in the previous proposition, we can rephrase the existence result: In particular, we have $X \in R(s)$ for some $s > d/2$, exactly as in Proposition 2.17. This condition $s > d/2$ means that X has values in a Sobolev space that is embedded into the space of continuous functions on $\mathcal{D}$. This is a natural starting point for inductively applying $(F_{s,\eta})$, cf. Proposition 2.4, as the next result illustrates:

Proposition 2.19.

(i) Let f be a polynomial as in (2.40). Then $(F_{s,\eta})$ holds for any $s > d/2$ and $0 < \eta < 2$.

(ii) Let $f \in C_b^\infty(\mathbb{R})$. Then $(F_{s,\eta})$ holds for any $s \in [0,1] \cup [d/2, \infty)$ and $0 < \eta < 2$.

Proof.

(i) For $s > d/2$ the Sobolev spaces $W^{s,2}(\mathcal{D})$ are closed under multiplication (see e.g. [AF03, Theorem 4.39], [Tri10a, p. 146]). Therefore,

$$\|f(X)\|_s \lesssim \sum_{k=1}^{m} |a_k| \, \|X\|_s^k \lesssim (1 + \|X\|_s)^m$$

for $X \in H_s$.

(ii) The case $s = 0$ is trivial since $\|f(X)\|^2_{L^2(\mathcal{D})} \leq |\mathcal{D}| \sup_{y\in\mathbb{R}} |f(y)|^2 < \infty$, so let $s > 0$. Set $\tilde{f} := f - f(0)$. By Theorem A from [AF92] and the discussion thereafter, there is $C > 0$ such that $\|\tilde{f}(X)\|_s \leq C(1 + \|X\|_s)^{1\vee s}$ for $s \in (0, 1] \cup [d/2, \infty)$, and the claim is immediate.

□

In particular, for each of the examples considered in this section, it is true that a.s. $X \in R(s)$ and $\widetilde{X} \in R(s+2)$ for any $s < s^*$. Therefore, by Theorem 2.11, we obtain the following result concerning diffusivity estimation:

Theorem 2.20. *Let $\alpha > \gamma - (d+2)/8$. Then $\hat{\theta}_N^{\text{full}}$ satisfies*

$$N^{\frac{1}{2}+\frac{1}{d}} \left(\hat{\theta}_N^{\text{full}} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\theta(d+4\alpha-4\gamma+2)^2}{T\Lambda d(d+8\alpha-8\gamma+2)}\right) \tag{2.42}$$

If $d = 1$, the same is true for $\hat{\theta}_N^{\text{part}}$, $\hat{\theta}_N^{\text{lin}}$. In $d = 2$, $\hat{\theta}_N^{\text{part}}$ is consistent with optimal rate, i.e. $\hat{\theta}_N^{\text{part}} = \theta + o(N^{-a})$ for any $a < 1$, and the same is true for $\hat{\theta}_N^{\text{lin}}$. □

It is clear that the coefficients $(a_k)_{0\leq k\leq m}$ in (2.40) may depend on $x \in \mathcal{D}$, in such a way that $a_k \in H_{s^*}$ for $0 \leq k \leq m$. This does not change the proof of $(F_{s,\eta})$ for $s < s^*$ in Proposition 2.19, thus Theorem 2.20 remains valid in that case.

Remark 2.21. *It is straightforward to include an advection term of the form $F_{\mathrm{ad}}(X) = \nabla \cdot (Xv) = \mathrm{div}(Xv)$ to the reaction–diffusion equation, where $v : \mathcal{D} \to \mathbb{R}^d$ is a vector field with components $v^{(i)} \in H_s$ for some $s > d/2$. More precisely, assume that the nonlinearity of the equation $F = F_{\mathrm{re}} + F_{\mathrm{ad}}$ splits into a reaction term $F_{\mathrm{re}}(X) = f(X)$ as before, and an advection term F_{ad} as described above. It is clear that* div *maps $(H_s)^d$ into H_{s-1} for any $s \in \mathbb{R}$.*[3] *Furthermore, if $X \in H_s$ and $v \in (H_s)^d$ for $s > d/2$, then $\left\|Xv^{(i)}\right\|_s \lesssim \|X\|_s \left\|v^{(i)}\right\|_s$ for $1 \leq i \leq d$, and consequently, $\|F_{\mathrm{ad}}(X)\|_{s-1} \lesssim \|X\|_s \sum_{i=1}^d \left\|v^{(i)}\right\|_s$, i.e. F_{ad} (and consequently $F = F_{\mathrm{re}} + F_{\mathrm{ad}}$) satisfies $(F_{s,\eta})$ with any $\eta < 1$. In terms of diffusivity estimation, this means that $\hat{\theta}_N^{\mathrm{part}} = \theta + o(N^{-a})$ for any $a < 1/d$, and similarly for $\hat{\theta}_N^{\mathrm{lin}}$, whereas $\hat{\theta}_N^{\mathrm{full}}$ is asymptotically normal with rate $N^{-1/2-1/d}$. It cannot be expected that this loss in convergence rate (compared to $\hat{\theta}_N^{\mathrm{full}}$) can be improved for $\hat{\theta}_N^{\mathrm{lin}}$, cf. Section 2.4.1.*

2.4.3 Burgers Equation

Let $d = 1$ and $\mathcal{D} = [0, L] \subset \mathbb{R}$ for some $L > 0$. Consider

$$\mathrm{d}X_t = \theta \Delta X_t \mathrm{d}t - X_t \partial_x X_t \mathrm{d}t + B \mathrm{d}W_t \tag{2.43}$$

with initial condition $X_0 \in L^p(\Omega, H_{s^*+1})$ for any $p \geq 1$, and Dirichlet boundary conditions. The nonlinearity is given by

$$F(X) = -X\partial_x X = \partial_x \left(-\frac{1}{2} X^2 \right).$$

The spaces $H = L^2(\mathcal{D})$ and $(H_s)_{s \in \mathbb{R}}$ are as in Section 2.4.2. As a special case of [LR15, Example 5.1.8], (2.43) has a unique solution in $R^{\mathrm{v}}(1)$. Higher regularity can be inferred in a stepwise manner as follows:

Lemma 2.22.

(i) F satisfies $(F^{\mathrm{v}}_{s,\eta})$ with $\eta = 3/8$ for $s = 1$.

(ii) F satisfies $(F^{\mathrm{v}}_{s,\eta})$ with $\eta = 1/2$ for any $s > 1$.

(iii) F satisfies $(F_{s,\eta})$ for all $\eta < 1$ and $s > 1/2$.

---

[3] For $s \in \mathbb{Z}$ this is obvious, for general s use the exact interpolation Theorem [AF03, Theorem 7.23].

In particular, $X \in R^{\mathrm{v}}(s)$ and $\widetilde{X} \in R^{\mathrm{v}}(s+1/2)$ for all $s < s^*$. If additionally $s^* > 3/2$ (i.e. $\gamma > 1/2$), then $X \in R(s)$ and $\widetilde{X} \in R(s+1)$ for any $s < s^*$.

Proof. In the following calculations we use repeatedly the interpolation inequality $\|X\|_{rs_1+(1-r)s_2} \lesssim \|X\|_{s_1}^r \|X\|_{s_2}^{1-r}$ for $s_1, s_2 \in \mathbb{R}$ and $0 < r < 1$, further the algebra property $\|XY\|_s \lesssim \|X\|_s \|Y\|_s$ for $s > 1/2$ and the Sobolev embedding $H_{1/4} \subset L^4(\mathcal{D})$ in $d = 1$. These estimates are standard and can be found, e.g., in [AF03].

(i) We have

$$\begin{aligned}
\|F(X)\|_{s+\eta-2}^2 &= \frac{1}{4}\left\|X^2\right\|_{3/8}^2 \lesssim \left\|X^2\right\|_{3/4}\left\|X^2\right\|_{L^2(\mathcal{D})} \lesssim \|X\|_{3/4}^2 \|X\|_{L^4(\mathcal{D})}^2 \\
&\lesssim \|X\|_{3/4}^2 \|X\|_{1/4}^2 \lesssim \|X\|_1^{3/2} \|X\|_0^{1/2} \|X\|_1^{1/2} \|X\|_0^{3/2} \\
&= \|X\|_1^2 \|X\|_0^2 ,
\end{aligned}$$

so $(F_{s,\eta}^{\mathrm{v}})$ is satisfied as stated.

(ii) For $s > 1$ and $\eta = 1/2$,

$$\|F(X)\|_{s+\eta-2}^2 = \frac{1}{4}\left\|X^2\right\|_{s-1/2}^2 \lesssim \|X\|_{s-1/2}^4 \lesssim \|X\|_s^2 \|X\|_{s-1}^2 ,$$

so condition $(F_{s,\eta}^{\mathrm{v}})$ holds.

(iii) For $s > 1/2$ and $\eta < 1$, with $\epsilon = 1 - \eta$,

$$\|F(X)\|_{s+\eta-2+\epsilon} = \frac{1}{2}\left\|X^2\right\|_s \lesssim \|X\|_s^2 ,$$

so $(F_{s,\eta})$ holds.

Concerning the regularity of X, we know already that $X \in R^{\mathrm{v}}(1)$. By (i) and Proposition 2.5, $\widetilde{X} \in R^{\mathrm{v}}(1 + 3/8)$. As $\bar{X} \in R(s)$ for all $s < s^*$, we get $X \in R^{\mathrm{v}}(s)$ for $s < s^* \wedge (1 + 3/8)$, and in particular, there is $s > 1$ with $X \in R^{\mathrm{v}}(s)$. Now (ii) and repeated use of Proposition 2.5 yields $X \in R^{\mathrm{v}}(s)$ and $\widetilde{X} \in R^{\mathrm{v}}(s + 1/2)$ for any $s < s^*$. Further, $X \in R(s-1)$ for all $s < s^*$ because $R^{\mathrm{v}}(s) \subset R(s-1)$. In case $s^* > 3/2$, we have $X \in R(s)$ for some $s > 1/2$, and Proposition 2.4 gives $X \in R(s)$, $\widetilde{X} \in R(s+1)$ for all $s < s^*$. □

This Lemma, together with Theorem 2.11, yields the asymptotic properties of $\hat{\theta}_N^{\mathrm{full}}$, $\hat{\theta}_N^{\mathrm{part}}$ and $\hat{\theta}_N^{\mathrm{lin}}$ for the Burgers equation:

Theorem 2.23. *Let $\gamma > 1/2$ and $\alpha > \gamma - 3/8$. Then $\hat{\theta}_N^{\text{full}}$ satisfies*

$$N^{\frac{3}{2}}\left(\hat{\theta}_N^{\text{full}} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\theta(3+4\alpha-4\gamma)^2}{T\Lambda(3+8\alpha-8\gamma)}\right). \tag{2.44}$$

Further, $\hat{\theta}_N^{\text{part}} = \theta + o(N^{-a})$ for $a < 1$, and the same is true for $\hat{\theta}_N^{\text{lin}}$. □

Remark 2.24. *It is possible to apply Theorem 2.11 to the stochastic Navier–Stokes equations driven by additive noise in dimension $d = 2$ with unknown viscosity. In this case, we reobtain the results from [CGH11]. It has been conjectured in [Cia18] that these results apply also to the stochastic Burgers equation.*

2.4.4 Equations of Cahn-Hilliard Type

For $d \geq 1$, fix a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ with smooth boundary. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function. Consider

$$\mathrm{d}X_t = -\theta\Delta^2 X_t \mathrm{d}t - \Delta f(X_t)\mathrm{d}t + B\mathrm{d}W_t \tag{2.45}$$

with initial condition X_0 and boundary conditions $\nabla X \cdot \nu = 0$, $\nabla(\Delta X)\cdot\nu = 0$, where $\nu : \partial\mathcal{D} \to \mathbb{R}^d$ is the unit vector pointing outwards the domain $\mathcal{D}$. We formalize this setting as in [LR15, p. 172 ff.]: Set $H = L^2(\mathcal{D})$, and let V be the closure in $W^{2,2}(\mathcal{D})$ of $\{u \in C^4(\overline{\mathcal{D}}) \mid \nabla u \cdot \nu = 0,\ \nabla(\Delta u)\cdot\nu = 0 \text{ on } \mathcal{D}\}$. Considering the Gelfand triple $V \subset H \simeq H^* \subset V^*$, we have that $A = -\Delta^2$ is a bounded operator $V \to V^*$. As before, we set $H_s := D((-A)^{s/2})$. Our standing assumption is $X_0 \in H_{s^*+1}$.

Note that the regularity counting in this section differs from the convention from the previous examples, because the leading drift term in (2.45) is of order four: This means that H_s is a closed subspace of $W^{2s,2}(\mathcal{D})$. Furthermore, in this case we have $\beta = 4/d$, i.e. $\lambda_k \asymp \Lambda k^{4/d}$, see [Shu01, Section 13.4]. We additionally introduce the "classical" regularity spaces $H_s^{\Delta} := D((-\Delta)^{s/2})$ that have been used in the previous sections. It is necessary to specify which regularity scale we are using when we speak about condition $(F_{s,\eta})$.

Proposition 2.25. *Let $s \in \mathbb{R}$, $\eta > 0$, and set $s' := 2s$, $\eta' := 2\eta$. If f satisfies $(F_{s',\eta'})$ with respect to the scale of Hilbert spaces $(H_r^{\Delta})_{r\in\mathbb{R}}$, then F, given by $F(X) = -\Delta f(X)$, satisfies $(F_{s,\eta})$ with respect to the scale of Hilbert spaces $(H_r)_{r\in\mathbb{R}}$.*

Proof. Choose $\epsilon > 0$ and $g : [0,\infty) \to [0,\infty)$ as in $(F_{s',\eta'})$. Then

$$\|-\Delta f(X)\|_{H_{s+\eta+\epsilon/2-2}} = \|f(X)\|_{H^{\Delta}_{s'+\eta'+\epsilon-2}} \leq g\left(\|X\|_{H^{\Delta}_{s'}}\right) = g\left(\|X\|_{H_s}\right).$$

□

For example, let $d \leq 2$ and assume that f is of the form

$$f(x) = cx + \phi(x) \tag{2.46}$$

for $c \in \mathbb{R}$ and $\phi \in C_b^\infty(\mathbb{R})$. In particular, f is globally Lipschitz continuous. By [LR15, Example 5.2.27], there is a unique solution X to (2.45) with a.s. $X \in R^{\mathrm{v}}(1) \subset R(0)$. As a consequence of Proposition 2.25 and Proposition 2.19 (ii), $F = -\Delta f$ satisfies $(F_{s,\eta})$ for any $s \geq 0$ with $\eta < 1$ in the regularity scale $(H_r)_{r\in\mathbb{R}}$. By Proposition 2.4 we conclude $X \in R(s)$ and $\widetilde{X} \in R(s+1)$ for any $s < s^*$. Therefore, we have:

Theorem 2.26. *Let $\gamma > d/8$ and $\alpha > \gamma - (d+4)/16$. Then $\hat{\theta}_N^{\mathrm{full}}$ satisfies*

$$N^{\frac{1}{2}+\frac{2}{d}}\left(\hat{\theta}_N^{\mathrm{full}} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\theta(d+8\alpha-8\gamma+4)^2}{T\Lambda d(d+16\alpha-16\gamma+4)}\right). \tag{2.47}$$

Further, $\hat{\theta}_N^{\mathrm{part}} = \theta + o(N^{-a})$ for all $a < 2/d$, and the same is true for $\hat{\theta}_N^{\mathrm{lin}}$. □

2.4.5 Robustness under Model Misspecification

Assume that the true dynamics of a process X is given by

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + G(X)(t)\mathrm{d}t + B\mathrm{d}W_t \tag{2.48}$$

with smooth initial condition X_0 and $F, G : C(0,T;H) \supseteq D(F) \cap D(G) \to L^1(0,T;H)$, where $D(F) \cap D(G)$ is the common domain for F and G. We assume that (2.48) is well-posed in $R(s_0)$ for some $0 \leq s_0 < s^*$, and that F satisfies (F_{s,η_F}) for some $\eta_F > 0$ and all $s_0 \leq s < s^*$. Assume further that $X_0 \in H_{s^*+\eta_F}$. We are interested in the robustness of $\hat{\theta}_N^{\mathrm{full}}$, $\hat{\theta}_N^{\mathrm{part}}$ and $\hat{\theta}_N^{\mathrm{lin}}$ with respect to the misspecification G. In this section, all three estimators are given by the same terms as in Section 2.3. In particular, $\hat{\theta}_N^{\mathrm{full}}$ and $\hat{\theta}_N^{\mathrm{part}}$ include knowledge on F but not on G.

Theorem 2.27. *Let $\alpha > \gamma - (1 + 1/\beta)/4$.*

(i) If G satisfies (F_{s,η_G}) for some $\eta_G > 0$ and $s_0 \leq s < s^*$, then $\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ are consistent.

(ii) If G satisfies (F_{s,η_G}) for some $\eta_G > 1 + 1/\beta$ and $s_0 \leq s < s^*$, then all statements on the asymptotic properties of $\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ from Theorem 2.11 transfer to the present case.

(iii) If $F + G$ satisfies $(F_{s,\eta_{F+G}})$ for some $\eta_{F+G} > 1 + 1/\beta$ and $s_0 \leq s < s^*$, then $\hat{\theta}_N^{\text{lin}}$ is asymptotically normal as in (2.29).

Proof. Note that in any of these cases, $F + G$ satisfies $(F_{s,\eta})$ for $s_0 \leq s < s^*$ and $\eta = \eta_F \wedge \eta_G$, so (2.20) remains true. Thus, all claims follow from a straightforward modification of the proof of Theorem 2.11, taking into account the additional bias of the form (2.33), with F replaced by G therein, coming from the nonlinear term $G(X)$. $\square$

The excess regularity η_{F+G} of $F+G$ clearly satisfies $\eta_{F+G} \geq \eta_F \wedge \eta_G$, but due to cancellation effects, η_{F+G} may be larger than $\eta_F \wedge \eta_G$.

Remark 2.28.

(i) The preceding examples show that a large class of nonlinearities G satisfies (F_{s,η_G}) for some $\eta_G > 0$.

(ii) As G is assumed to be unknown (or intractable), it does not make sense to construct a modified estimator that takes into account the shift coming from G in order to improve the convergence rate. Rather, G and its impact on diffusivity estimation should be understood qualitatively.

(iii) The typical situation can be described as follows: Let F^{true} be the true nonlinearity of the underlying process, which is either unknown or too complex to be handled directly. Instead, an approximate nonlinear term F^{approx} is used to model the dynamics of X. In this case $F = F^{\text{approx}}$, and $G = F^{\text{true}} - F^{\text{approx}}$ is the remainder that describes the deviation from the true model. The excess regularity η_G associated with G measures the quality of the approximate model F^{approx} for diffusivity estimation.

(iv) For example, F^{approx} may be the linearization of a nonlinear model F^{true}. In this case, η_G is related to the linearization procedure.

(v) If X is the solution to a reaction–diffusion equation (with possible advection term) as in Section 2.4.2, the excess regularity of G encodes the order of the model misspecification as a differential operator. For example, if only reaction terms (of order zero) are misspecified, but the advection mechanism (of order one) is known very precisely, then we have $\eta_G < 2$. If the description of the advection term is wrong, then $\eta_G < 1$.

(vi) In particular, for diffusivity estimation, precise knowledge on the transport term is more important than precise knowledge on the reaction term.

2.5 Numerical Illustration

We simulate the Allen-Cahn equation [CA77]

$$\mathrm{d}X_t = \theta\Delta X_t\mathrm{d}t + (X_t - X_t^3)\mathrm{d}t + (-\Delta)^{-\gamma}\mathrm{d}W_t \tag{2.49}$$

on $\mathcal{D} = [0,1]$ with Dirichlet boundary and initial condition $x \mapsto \sin(\pi x)$. We discretize the equation in Fourier space and simulate $N_0 = 100$ Fourier modes by a linear-implicit Euler scheme until $T = 1$. The temporal and spatial step size is set to $\Delta t = 2.5 \times 10^{-5}$ and $\Delta x = 5 \times 10^{-4}$, respectively. The diffusivity is given by $\theta = 0.02$. We generate $M = 1000$ Monte Carlo simulations for each of the choices $\gamma = 0.4$ and $\gamma = 0.8$. In either case, we set $\alpha = \gamma$. A detailed discussion on numerical simulation for SPDEs can be found in [LPS14].

By Theorem 2.20, all three estimators $\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ are asymptotically normal. In Figure 2.1, the simulation results concerning consistency, convergence rate and asymptotic distribution are shown. Whereas $\hat{\theta}_N^{\text{full}}$ and $\hat{\theta}_N^{\text{part}}$ perform as predicted, $\hat{\theta}_N^{\text{lin}}$ seems to exhibit non-asymptotic effects. Apparently, this depends on the impact of the noise on the dynamics, which is controlled by γ. In fact, the value of γ has two effects: It determines the spatial regularity of the noise (and consequently of X), but it also has an impact on the overall noise intensity via the magnitude of $\lambda_1^{-\gamma}$, i.e. the largest eigenvalue of $(-\Delta)^{-\gamma}$. Our interpretation is that irregular noise from low values of γ tends to cover the effect of the nonlinearity. Said another way, nonlinear effects have a larger impact under smooth noise with smaller amplitude.

We further mention that for even larger values of γ (take $\gamma = 1.3$), the estimated value from $\hat{\theta}_N^{\mathrm{lin}}$ is mostly negative and therefore not related to the true diffusivity. On the other hand, $\hat{\theta}_N^{\mathrm{full}}$ and $\hat{\theta}_N^{\mathrm{part}}$ remain consistent. It is possible that this effect depends on the number of Fourier modes N_0 used in the simulation.

2.6 The Case of Systems

Consider a partially observed system of the form

$$\begin{aligned} \mathrm{d}X_t^O &= \theta A X_t^O \mathrm{d}t + F_O(X_t^O, X_t^U)\mathrm{d}t + B_O \mathrm{d}W_t^O, \\ \mathrm{d}X_t^U &= F_U(X_t^O, X_t^U)\mathrm{d}t + B_U \mathrm{d}W_t^U, \end{aligned} \tag{2.50}$$

together with initial conditions X_0^O, X_0^U. Here, X^O denotes the observed component and X^U the unobserved component of the dynamics. We want to estimate the unknown diffusivity θ of the observed component.

More precisely, let H^O, H^U be two Hilbert spaces, and let $A : D(A) \to H^O$ be a densely defined, closed, negative definite and self-adjoint operator on H^O with compact resolvent, whose eigenvalue sequence $(\lambda_k)_{k\in\mathbb{N}}$ satisfies (2.2). F^O and F^U are nonlinear operators defined on a subset $D(F)$ of $H^O \oplus H^U$, with values in H^O and H^U, respectively. W^O and W^U are independent cylindrical Wiener processes on H^O and H^U, and B_O, B_U are Hilbert–Schmidt operators on H^O and H^U. We assume $B_O = \sigma_O(-A)^{-\gamma}$ for some $\gamma > 1/(2\beta)$, where $\sigma_O > 0$ is the noise intensity in the observed component.

$P_N : H^O \to H^O$ denotes the projection onto the span of the first N eigenvalues $\Phi_1, \dots, \Phi_N$ of A. We write $\mathscr{X} = (X^O, X^U)$ and $\mathscr{H} = H^O \oplus H^U$. Let $\mathscr{A} : D(A) \oplus H^U \to \mathscr{H}$ be the operator given by $\mathscr{A}(x, y) = (Ax, 0)$, define $\mathscr{F} : D(F) \to \mathscr{H}$ by means of $\mathscr{F}(u, v) = (F_O(u, v), F_U(u, v))$ and $\mathscr{B} : \mathscr{H} \to \mathscr{H}$ via $\mathscr{B}(u, v) = (B_O u, B_U v)$. Finally, $\mathscr{W} = (W^O, W^U)$ is a cylindrical Wiener process on $\mathscr{H}$. Then $\mathscr{X}$ satisfies

$$\mathrm{d}\mathscr{X}_t = \theta\mathscr{A}\mathscr{X}_t \mathrm{d}t + \mathscr{F}(\mathscr{X}_t)\mathrm{d}t + \mathscr{B}\mathrm{d}\mathscr{W}_t. \tag{2.51}$$

In order to capture the regularity of $\mathscr{X}$, we extend the notation from Section

2.2 and set for $s \in \mathbb{R}$:

$$H_s := D((-A)^{s/2}), \tag{2.52}$$

$$\mathscr{H}_s := D((-\mathscr{A})^{s/2}) = D((-A)^{s/2}) \oplus H^U, \tag{2.53}$$

$$R(s) := L^\infty(0,T;H_s), \tag{2.54}$$

$$\mathscr{R}(s) := L^\infty(0,T;\mathscr{H}_s). \tag{2.55}$$

In analogy to condition $(F_{s,\eta})$, we need a condition on the observed part F_O of $\mathscr{F}$:

$(F^{\mathrm{sys}}_{s,\eta})$ There is $\epsilon > 0$ and a continuous increasing function $g : [0,\infty) \to [0,\infty)$ such that for all $\mathscr{X} \in \mathscr{R}(s)$:

$$\|F_O(\mathscr{X})\|_{R(s+\eta+\epsilon-2)} \leq g\left(\|\mathscr{X}\|_{\mathscr{R}(s)}\right). \tag{2.56}$$

The splitting argument concerns only the observed part: We write $X^O = \bar{X}^O + \widetilde{X}^O$, where $\bar{X}^O$, $\widetilde{X}^O$ satisfy

$$\mathrm{d}\bar{X}^O_t = \theta A \bar{X}^O_t \mathrm{d}t + B_O \mathrm{d}W^O_t, \tag{2.57}$$

$$\mathrm{d}\widetilde{X}^O_t = \theta A \widetilde{X}^O_t \mathrm{d}t + F_O(\mathscr{X})\mathrm{d}t, \tag{2.58}$$

with $\bar{X}^O_0 = 0$, $\widetilde{X}^O_0 = X^O_0$.

In analogy to Proposition 2.3 and Proposition 2.4, we have

Proposition 2.29. *Let $\eta > 0$. If $(F^{\mathrm{sys}}_{s,\eta})$ holds for $s \in \mathbb{R}$ such that a.s. $\mathscr{X} \in \mathscr{R}(s)$ and $X^O_0 \in H_{s+\eta}$, then $\widetilde{X}^O \in R(s+\eta)$. In particular, if $s_0 < s^*$ such that $(F^{\mathrm{sys}}_{s,\eta})$ holds for $s_0 \leq s < s^*$, $\bar{X}^O \in R(s)$ for $s < s^*$, $\mathscr{X} \in \mathscr{R}(s_0)$ and $X^O_0 \in H_{s^*+\eta}$ almost surely, then $\mathscr{X} \in \mathscr{R}(s)$ and $\widetilde{X}^O \in R(s+\eta)$ for $s < s^*$.* □

Adapting the estimators from Section 2.3 to the present situation, we define

$$\hat{\theta}^{\mathrm{full}}_N := -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} P_N X^O_t, \mathrm{d}P_N X^O_t \right\rangle_{H^O}}{\int_0^T \|(-A)^{1+\alpha} P_N X^O_t\|^2_{H^O}\,\mathrm{d}t} + \frac{\int_0^T \left\langle (-A)^{1+2\alpha} P_N X^O_t, P_N F_O(\mathscr{X}) \right\rangle_{H^O} \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha} P_N X^O_t\|^2_{H^O}\,\mathrm{d}t}.$$

In fact, $\hat{\theta}_N^{\text{full}}$ is not a function of the observed process $P_N X^O$ alone, as it depends on X^U via $\mathscr{X}$. Consequently, we define

$$\begin{aligned}
\hat{\theta}_N^{\text{part}} &:= -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} P_N X_t^O, \mathrm{d}P_N X_t^O \right\rangle_{H^O}}{\int_0^T \|(-A)^{1+\alpha} P_N X_t^O\|_{H^O}^2 \,\mathrm{d}t} \\
&\quad + \frac{\int_0^T \left\langle (-A)^{1+2\alpha} P_N X_t^O, P_N F_O(P_N X_t^O, 0) \right\rangle_{H^O} \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha} P_N X_t^O\|_{H^O}^2 \,\mathrm{d}t}, \\
\hat{\theta}_N^{\text{lin}} &:= -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} P_N X_t^O, \mathrm{d}P_N X_t^O \right\rangle_{H^O}}{\int_0^T \|(-A)^{1+\alpha} P_N X_t^O\|_{H^O}^2 \,\mathrm{d}t}.
\end{aligned}$$

Analogously to Theorem 2.11, the following result is proven:

Theorem 2.30. *Let $\gamma > 1/(2\beta)$ and $\eta > 0$, $s_0 < s^*$ such that ($F_{s,\eta}^{\text{sys}}$) holds for $s_0 \leq s < s^*$. Assume a.s. $\mathscr{X} \in \mathscr{R}(s_0)$ and $X_0^O \in H_{s^*+\eta}$. Let $\alpha > \gamma - (1+1/\beta)/4$. Then $\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ are strongly consistent as $N \to \infty$, and $\hat{\theta}_N^{\text{full}}$ is asymptotically normal as in* (2.29)*. If $\eta > 1 + 1/\beta$, the same is true for $\hat{\theta}_N^{\text{part}}$, $\hat{\theta}_N^{\text{lin}}$, otherwise $\hat{\theta}_N^{\text{part}} = \theta + o(N^{-a})$ for each $a < \beta\eta/2$, and the same is true for $\hat{\theta}_N^{\text{lin}}$.* □

Example 2.31. *The theory from this section is applicable to a stochastic FitzHugh–Nagumo system [Fit61, NAY62], whose activator component is observed:*

$$\begin{aligned}
\mathrm{d}U_t &= \theta \Delta U_t \mathrm{d}t + k_1 U_t (1 - U_t)(U_t - a) - k_2 V_t \mathrm{d}t + B_O \mathrm{d}W_t^O, \\
\mathrm{d}V_t &= \epsilon \left(b U_t - V_t \right) \mathrm{d}t + B_U \mathrm{d}W_t^U,
\end{aligned}$$

with initial condition U_0, V_0, on $\mathcal{D} = [0, L]$, $L > 0$, with Neumann boundary conditions. The reaction parameters are $k_1, k_2, \epsilon, b > 0$ and $a \in (0, 1)$. The state space for both components is $H^O = H^U = L^2(\mathcal{D})$. Assume[4] that $1/2 < s^* < 2$ and $\mathscr{X} = (U, V) \in \mathscr{R}(s)$ for $s < s^*$. We verify condition ($F_{s,\eta}^{\text{sys}}$) for F_O. Here, $F_O(U, V) = k_1 U (1 - U)(U - a) - k_2 V$. The first term is a polynomial that can be treated exactly as in Section 2.4.2, with obvious changes in the notation of the norm, resulting in $\eta < 2$. Concerning the second term $-k_2 V$,

[4] In fact, $\mathscr{X} \in \mathscr{R}(0)$ together with $U \in R^{\mathrm{v}}(1)$ can be shown as in [SS15], and a direct modification of Proposition 2.18 gives $\mathscr{X} \in \mathscr{R}(1)$ in this case. Note that under the Hilbert–Schmidt assumption $\gamma > d/4$, we necessarily have $s^* > 1$. Higher regularity in $\mathscr{R}(s)$ for all $s < s^*$ follows from ($F_{s,\eta}^{\text{sys}}$).

we can find $\epsilon > 0$ such that $\|V\|_{s+\eta+\epsilon-2} = \|V\|_{L^2(\mathcal{D})}$ if and only if $\eta < 2 - s$, and this equality leads to (2.56) as well. In total, $(F^{\text{sys}}_{s,\eta})$ holds for F_O with $\eta < 2 - s^*$ for $1/2 < s < s^*$. This proves that $\hat{\theta}^{\text{full}}_N$ is asymptotically normal, and $\hat{\theta}^{\text{part}}_N$ and $\hat{\theta}^{\text{lin}}_N$ are consistent with convergence rate bounded by N^{-a} for $a < 2 - s^*$. This result can be refined if the optimal regularity of the inhibitor is taken into account. In fact, under regularity assumptions on the inhibitor noise, all three estimators will be asymptotically normal. We refer to Chapter 6, where a similar FitzHugh–Nagumo system is studied in greater detail.

Finally, we note that if $\sigma_O = 0$, i.e. if only the unobserved component is driven by noise, other methods need to be employed. We come back to that case in Remark 3.16 below.

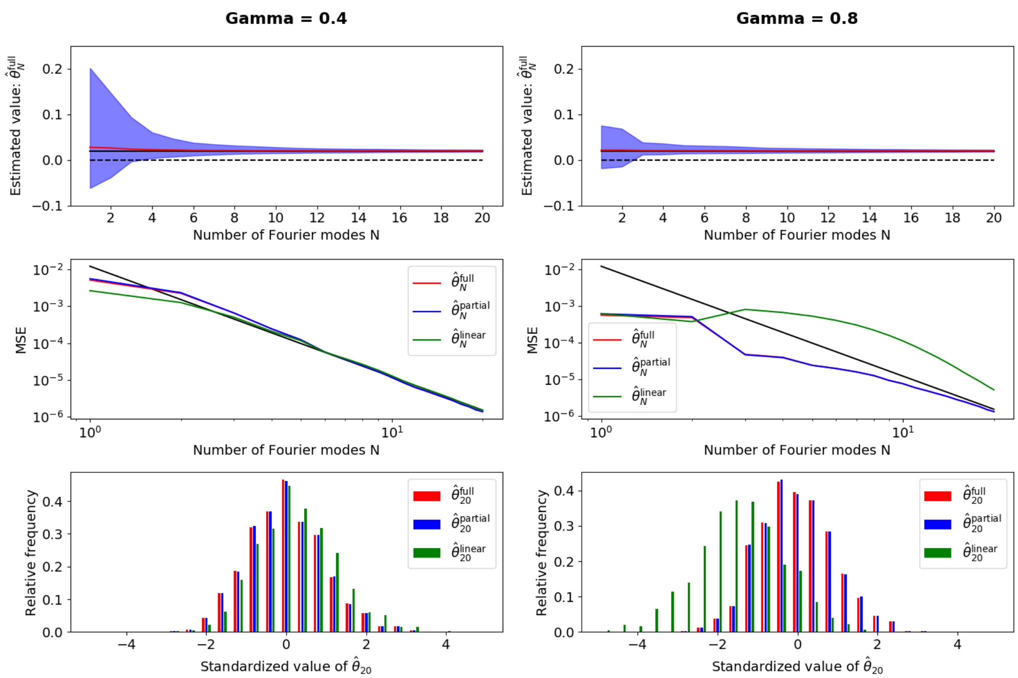

Figure 2.1: *Left column contains results for $\gamma = 0.4$, right column for $\gamma = 0.8$.* **(top row)** *Red line: Median from $M = 1000$ realizations of $\hat{\theta}_N^{\text{full}}$. The blue region is bounded above by the 97.5-percentile and below by the 2.5-percentile. Black solid line is plotted at true value $\theta = 0.02$, dashed line plotted at zero.* **(middle row)** *The mean squared error (MSE), given by $M^{-1}\sum_{k=1}^{M}(\hat{\theta}_N(k) - \theta)^2$, is plotted, where $\hat{\theta}_N(k)$ is the k-th realization of either of the estimators $\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$ or $\hat{\theta}_N^{\text{lin}}$. Black line corresponds to the squared true theoretical rate $N \mapsto (\Sigma^{1/2}N^{-3/2})^2$, with Σ from (2.30).* **(bottom row)** *Histogram for the standardized values $\Sigma^{-1/2}N^{3/2}(\hat{\theta}_N - \theta)$ at $N = 20$, where $\hat{\theta}_N$ is either of the three estimators. The width of each bin is 0.4. Outliers outside the interval $[-5, 5]$ are put into the leftmost and rightmost bin, respectively.*

Chapter 3

Extended Noise Models for the Spectral Approach

In the last chapter, we studied semilinear SPDE models driven by spatially correlated but temporally white noise. Nonetheless, in many situations it is desirable to include temporal correlation to an SPDE. There are different ways to achieve this: A standard approach is using fractional noise, as studied e.g. in [MP08, MT13, KM19] in the large time regime, [TTV14, MKT19a, MKT19b, SST20] in a spatial and/or temporal infill regime, or [CLP09, Kří20] in the spectral approach. Fractional noise impacts the temporal regularity of the solution process and can be used to model long-range dependence, see e.g. [Tud13] for a discussion of SPDEs driven by such noise.

However, in applications, there are further common approaches to include temporal correlation, which have gained little attention in literature concerning statistical inference for SPDEs. As an important example, in models appearing in biophysics literature [ASB18, FFAB20, MFF$^+$20], integrated Ornstein–Uhlenbeck noise is used.[1] While the presence of certain dynamical properties such as separation of phases or traveling waves may not be affected by substituting Brownian noise by integrated Ornstein–Uhlenbeck noise (or vice versa), the precise specification becomes important when it comes to the quantitative analysis of data. Motivated by these works, we study the cases of Ornstein–Uhlenbeck noise and integrated noise separately.

[1] As a motivation, note that the one-dimensional integrated Ornstein–Uhlenbeck process serves as an alternative model (besides the Wiener process) for describing the movement of a Brownian particle, see e.g. [HL84, Chapter 2] for a discussion.

Ornstein–Uhlenbeck noise is treated in Section 3.1. From the mathematical point of view, the statistical properties of Ornstein–Uhlenbeck driven models have not yet been investigated. We study this model both from the perspective of parameter estimation under Ornstein–Uhlenbeck assumption as well as from the point of view of model misspecification, where white noise is used in the description but the true dynamics is Ornstein–Uhlenbeck driven.

Integrated noise is studied in Section 3.2. Its statistical analysis can be reduced to the case of semimartingale-type noise. In addition, integrated noise provides a simple example of a model that cannot be handled by simply using the estimators from Chapter 2 without further modification.

Finally, in Section 3.3, we consider more general dispersion operators, which allows us to handle a certain type of multiplicative noise.

3.1 The Case of Ornstein–Uhlenbeck Noise

In this section we consider a semilinear SPDE driven by Ornstein–Uhlenbeck noise. We develop a hierarchical estimation theory for diffusivity θ and temporal correlation decay μ and compare the results to the white noise case, in particular, we consider the case of model misspecification in the noise. Our setting in this section is as follows:

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + \mathrm{d}V_t, \tag{3.1}$$

$$\mathrm{d}V_t = -\mu V_t \mathrm{d}t + B\mathrm{d}W_t, \tag{3.2}$$

with initial condition X_0 and V_0. Without loss of generality, we assume $V_0 = 0$. As before, W is a cylindrical Wiener process, $B = \sigma(-A)^{-\gamma}$ for some $\gamma > 1/(2\beta)$ and $\sigma > 0$, and $\theta > 0$ is the diffusivity. Further, $\mu \in \mathbb{R}$ is an additional parameter related to the temporal correlation length of the driving noise V. In this section we assume always $\mu \neq 0$, otherwise the equations reduce to the white noise model from Section 2.3. Additionally, we assume that w.l.o.g. for all $k \in \mathbb{N}$, $\mu \neq \pm\theta\lambda_k$. (Otherwise replace A with $A + \epsilon I$ for some $\epsilon > 0$, where $I : H \to H$ is the identity operator, and substitute F by $F - \epsilon I$. The additional perturbation is of order zero.) This will be used in Lemma 3.2 below.

Remark 3.1. *Our model is compatible with a different natural approach to Ornstein–Uhlenbeck driven SPDEs, namely:*

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + B\mathrm{d}W_t^{(\mu)} \tag{3.3}$$

with initial condition X_0, where $W^{(\mu)}$ is a cylindrical Ornstein-Uhlenbeck process in the sense that $w^{(\mu,k)} := \left\langle W^{(\mu)}, \Phi_k \right\rangle$ are independent Ornstein-Uhlenbeck processes of the form

$$\mathrm{d}w_t^{(\mu,k)} = -\mu w_t^{(\mu,k)}\mathrm{d}t + \mathrm{d}W_t^{(k)} \tag{3.4}$$

for independent Wiener processes $(W^{(k)})_{k\in\mathbb{N}}$. If $B = \sigma(-A)^{-\gamma}$, this model can be reduced to (3.1), (3.2) *by setting $V = BW^{(\mu)}$.*

The linearized model is given by

$$\begin{aligned} \mathrm{d}\bar{X}_t &= \theta A \bar{X}_t \mathrm{d}t + \mathrm{d}V_t, && (3.5)\\ \mathrm{d}V_t &= -\mu V_t \mathrm{d}t + B\mathrm{d}W_t && (3.6) \end{aligned}$$

with $\bar{X}_0 = V_0 = 0$. As in the previous chapter, we set $\widetilde{X} := X - \bar{X}$, then $\widetilde{X}$ satisfies the random PDE (2.5).

3.1.1 Covariance Structure and Asymptotic Behavior

As before, we set $\bar{x}^{(k)} = \left\langle \bar{X}, \Phi_k \right\rangle_H$ and $v^{(k)} = \langle V, \Phi_k \rangle_H$. The processes $(\bar{x}^{(k)}, v^{(k)})$, $k \in \mathbb{N}$, are independent centered Gaussian processes with

$$\begin{aligned} \mathrm{d}\bar{x}_t^{(k)} &= (-\theta\lambda_k \bar{x}_t^{(k)} - \mu v_t^{(k)})\mathrm{d}t + \sigma\lambda_k^{-\gamma}\mathrm{d}W_t^{(k)}, && (3.7)\\ \mathrm{d}v_t^{(k)} &= -\mu v_t^{(k)}\mathrm{d}t + \sigma\lambda_k^{-\gamma}\mathrm{d}W_t^{(k)}, && (3.8) \end{aligned}$$

and $\bar{x}_0^{(k)} = v_0^{(k)} = 0$, where $(W^{(k)})_{k\in\mathbb{N}}$ are independent Brownian motions.

Lemma 3.2. *With $\mu \neq 0$ and $\mu \neq \pm\theta\lambda_k$, we have the explicit representation*

$$\begin{aligned} \bar{x}_t^{(k)} &= \frac{\sigma\lambda_k^{-\gamma}}{\theta\lambda_k - \mu} \int_0^t \left(\theta\lambda_k e^{-\theta\lambda_k(t-r)} - \mu e^{-\mu(t-r)}\right) \mathrm{d}W_r^{(k)}, && (3.9)\\ v_t^{(k)} &= \sigma\lambda_k^{-\gamma} \int_0^t e^{-\mu(t-r)}\mathrm{d}W_r^{(k)}. && (3.10) \end{aligned}$$

Furthermore, for $0 \le r \le t$,

$$\begin{aligned}
\mathbb{E}[\bar{x}_r^{(k)} \bar{x}_t^{(k)}] &= \frac{\sigma^2 \lambda_k^{-2\gamma}}{(\theta\lambda_k - \mu)^2} \left(\frac{\theta\lambda_k}{2} (e^{-\theta\lambda_k(t-r)} - e^{-\theta\lambda_k(t+r)}) \right. \\
&\qquad + \frac{\mu}{2} (e^{-\mu(t-r)} - e^{-\mu(t+r)}) \\
&\qquad \left. + \frac{\mu\theta\lambda_k}{\theta\lambda_k + \mu} (e^{-\mu t - \theta\lambda_k r} + e^{-\theta\lambda_k t - \mu r} - e^{-\mu(t-r)} - e^{-\theta\lambda_k(t-r)}) \right), \\
\mathbb{E}[v_r^{(k)} v_t^{(k)}] &= \frac{\sigma^2 \lambda_k^{-2\gamma}}{2\mu} (e^{-\mu(t-r)} - e^{-\mu(t+r)}), \\
\mathbb{E}[v_r^{(k)} \bar{x}_t^{(k)}] &= \frac{\sigma^2 \lambda_k^{-2\gamma}}{\theta\lambda_k - \mu} \left(\frac{\theta\lambda_k}{\theta\lambda_k + \mu} (e^{-\theta\lambda_k(t-r)} - e^{-\theta\lambda_k t - \mu r}) \right. \\
&\qquad \left. - \frac{1}{2} (e^{-\mu(t-r)} - e^{-\mu(t+r)}) \right), \\
\mathbb{E}[\bar{x}_r^{(k)} v_t^{(k)}] &= \frac{\sigma^2 \lambda_k^{-2\gamma}}{\theta\lambda_k - \mu} \left(\frac{\theta\lambda_k}{\theta\lambda_k + \mu} (e^{-\mu(t-r)} - e^{-\mu t - \theta\lambda_k r}) \right. \\
&\qquad \left. - \frac{1}{2} (e^{-\mu(t-r)} - e^{-\mu(t+r)}) \right).
\end{aligned}$$

Proof. Fix $k \in \mathbb{N}$. With $Z = (\bar{x}^{(k)}, v^{(k)})^T$ we have $\mathrm{d}Z_t = A_Z Z_t \mathrm{d}t + B_Z \mathrm{d}W_t^{(k)}$, where $A_Z : \mathbb{R}^2 \to \mathbb{R}^2$, $B_Z : \mathbb{R} \to \mathbb{R}^2$ are linear mappings given by

$$A_Z = \begin{pmatrix} -\theta\lambda_k & -\mu \\ 0 & -\mu \end{pmatrix}, \quad B_Z = \sigma\lambda_k^{-\gamma} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

It is straightforward to verify that $A_Z = S D_Z S^{-1}$ with

$$D_Z = \begin{pmatrix} -\theta\lambda_k & 0 \\ 0 & -\mu \end{pmatrix}, \quad S = \begin{pmatrix} 1 & \mu \\ 0 & -\theta\lambda_k + \mu \end{pmatrix}.$$

It follows that for $t \in \mathbb{R}$,

$$e^{tA_Z} B_Z = S e^{tD_Z} S^{-1} B_Z = \frac{\sigma\lambda_k^{-\gamma}}{\theta\lambda_k - \mu} \begin{pmatrix} \theta\lambda_k e^{-\theta\lambda_k t} - \mu e^{-\mu t} \\ (\theta\lambda_k - \mu) e^{-\mu t} \end{pmatrix}.$$

Now, (3.9), (3.10) are clear from $Z_t = \int_0^t e^{(t-r)A_Z} B_Z \mathrm{d}W_r^{(k)}$, and the covariance terms follow from Itô's isometry. □

In particular, we have

$$\mathbb{E}[(\bar{x}_t^{(k)})^2] = \frac{\sigma^2\lambda_k^{-2\gamma}}{(\theta\lambda_k-\mu)^2}\left(\frac{\theta\lambda_k}{2}(1-e^{-2\theta\lambda_k t}) + \frac{\mu}{2}(1-e^{-2\mu t}) - \frac{2\mu\theta\lambda_k}{\theta\lambda_k+\mu}(1-e^{-(\theta\lambda_k+\mu)t})\right), \tag{3.11}$$

$$\mathbb{E}[(v_t^{(k)})^2] = \frac{\sigma^2\lambda_k^{-2\gamma}}{2\mu}(1-e^{-2\mu t}), \tag{3.12}$$

$$\mathbb{E}[v_t^{(k)}\bar{x}_t^{(k)}] = \frac{\sigma^2\lambda_k^{-2\gamma}}{\theta\lambda_k-\mu}\left(\frac{\theta\lambda_k}{\theta\lambda_k+\mu}(1-e^{-(\theta\lambda_k+\mu)t}) - \frac{1}{2}(1-e^{-2\mu t})\right). \tag{3.13}$$

From the above calculations (or the elementary observation that each $v^{(k)}$ is a classical one-dimensional Ornstein–Uhlenbeck process) it follows that $\lim_{t\to\infty}\mathbb{E}[v_t^{(k)}v_{t+d}^{(k)}]/(\mathbb{E}[(v_t^{(k)})^2]\mathbb{E}[(v_{t+d}^{(k)})^2])^{1/2} = e^{-\mu d}$ for $d \geq 0$, so μ describes the rate of exponential decay of the autocorrelation function of each noise mode in the stationary regime. Hence the name "temporal correlation decay" for μ.

Lemma 3.3. *It holds a.s. $\bar{X} \in R(s)$ for any $s < s^* := 1 + 2\gamma - 1/\beta$.*

Proof. The reasoning is similar as in Lemma 2.7. Define $C_1, C_2 : H \to H$ via $C_1\Phi_k := [\sigma\theta\lambda_k^{-\gamma+1}/(\theta\lambda_k - \mu)]\Phi_k$ and $C_2\Phi_k = -[\mu\sigma\lambda_k^{-\gamma}/(\theta\lambda_k-\mu)]\Phi_k$. Note that both operators are of Hilbert–Schmidt type due to $\gamma > 1/(2\beta)$. Using (3.9), we write $\bar{X}_t = \int_0^t e^{(t-r)\theta A}C_1 \mathrm{d}W_r + \int_0^t e^{-\mu(t-r)}C_2\mathrm{d}W_r =: \bar{X}_t^{(1)} + \bar{X}_t^{(2)}$, where, as before, $t \mapsto e^{t\theta A}$ is the C_0-semigroup generated by θA. We prove the claim for both stochastic integrals separately: For $s < s^* = 1+2\gamma-1/\beta$ and $0 < \alpha < \min\{1/2, (s^*-s)/2\}$,

$$\int_0^T t^{-2\alpha}\left\|(-A)^{s/2}e^{t\theta A}C_1\right\|_{\mathrm{HS}}^2 \mathrm{d}t = \sigma^2\theta^2\sum_{k=1}^\infty \frac{\lambda_k^{s-2\gamma+2}}{(\theta\lambda_k-\mu)^2}\int_0^T t^{-2\alpha}e^{-2\theta\lambda_k t}\mathrm{d}t$$
$$\lesssim \sum_{k=1}^\infty \lambda_k^{s-2\gamma}\int_0^\infty \lambda_k^{2\alpha-1}r^{-2\alpha}e^{-r}\mathrm{d}r \lesssim \sum_{k=1}^\infty k^{\beta(s+2\alpha-2\gamma-1)} < \infty.$$

By [DPZ14, Theorem 5.11], $(-A)^{s/2}\bar{X}^{(1)} \in C(0,T;H)$, i.e. $\bar{X}^{(1)} \in R(s)$ almost surely. With regard to $\bar{X}^{(2)}$, we have for any $s < s^*+1 = 2+2\gamma-1/\beta$

and $0 < \alpha < 1/2$,

$$\int_0^T t^{-2\alpha} \left\| (-A)^{s/2} e^{-\mu t} C_2 \right\|_{\mathrm{HS}}^2 \mathrm{d}t = \mu^2\sigma^2 \int_0^T t^{-2\alpha} e^{-2\mu t} \mathrm{d}t \sum_{k=1}^{\infty} \frac{\lambda_k^{s-2\gamma}}{(\theta\lambda_k - \mu)^2}$$
$$\lesssim \sum_{k=1}^{\infty} \lambda_k^{s-2\gamma-2} \lesssim \sum_{k=1}^{\infty} k^{\beta(s-2\gamma-2)} < \infty,$$

and consequently, again by [DPZ14, Theorem 5.11], it follows that a.s. $\bar{X}^{(2)} \in R(s)$. This proves the claim. $\square$

The next Proposition implies that a.s. $\bar{X} \notin L^2(0,T;H_s)$ for any $s > s^*$. In particular, the optimal regularity s^* is the same as in the white noise case from Section 2.3.

Proposition 3.4. *Set $C_{T,\mu}^{(\pm)} := T \pm (1 - e^{-2\mu T})/(2\mu)$.*

(i) As $k \to \infty$, we have the following asymptotic expansions:

$$\mathbb{E}\int_0^T (\bar{x}_t^{(k)})^2 \mathrm{d}t \asymp \frac{\sigma^2 T}{2\theta} \lambda_k^{-2\gamma-1}, \tag{3.14}$$

$$\mathbb{E}\int_0^T (v_t^{(k)})^2 \mathrm{d}t \asymp \frac{\sigma^2 C_{T,\mu}^{(-)}}{2\mu} \lambda_k^{-2\gamma}, \tag{3.15}$$

$$\mathbb{E}\int_0^T v_t^{(k)} \bar{x}_t^{(k)} \mathrm{d}t \asymp \frac{\sigma^2 C_{T,\mu}^{(+)}}{2\theta} \lambda_k^{-2\gamma-1}, \tag{3.16}$$

$$\mathbb{E}\int_0^T \left(\int_0^t \bar{x}_r^{(k)} \mathrm{d}r \right)^2 \mathrm{d}t \asymp \frac{\sigma^2 C_{T,\mu}^{(-)}}{2\mu\theta^2} \lambda_k^{-2\gamma-2}, \tag{3.17}$$

$$\mathbb{E}\int_0^T \bar{x}_t^{(k)} \left(\int_0^t \bar{x}_r^{(k)} \mathrm{d}r \right) \mathrm{d}t \asymp \frac{\sigma^2 (1 - e^{-2\mu T})}{4\mu\theta^2} \lambda_k^{-2\gamma-2}. \tag{3.18}$$

(ii) As $N \to \infty$, we have a.s.

$$\int_0^T \left\| (-A)^{s/2} \bar{X}_t^N \right\|^2 \mathrm{d}t \asymp \frac{\sigma^2 T \Lambda^{s-2\gamma-1}}{2\theta(1 + \beta(s - 2\gamma - 1))} N^{1+\beta(s-2\gamma-1)}, \tag{3.19}$$

$$\int_0^T \left\| (-A)^{s/2} V_t^N \right\|^2 \mathrm{d}t \asymp \frac{\sigma^2 C_{T,\mu}^{(-)} \Lambda^{s-2\gamma}}{2\mu(1 + \beta(s - 2\gamma))} N^{1+\beta(s-2\gamma)}, \tag{3.20}$$

$$\int_0^T \left\langle (-A)^s V_t^N, \bar{X}_t^N \right\rangle \mathrm{d}t \asymp \frac{\sigma^2 C_{T,\mu}^{(+)} \Lambda^{s-2\gamma-1}}{2\theta(1+\beta(s-2\gamma-1))} N^{1+\beta(s-2\gamma-1)}, \tag{3.21}$$

$$\int_0^T \left\| \int_0^t (-A)^{s/2} \bar{X}_r^N \mathrm{d}r \right\|^2 \mathrm{d}t \asymp \frac{\sigma^2 C_{T,\mu}^{(-)} \Lambda^{s-2\gamma-2}}{2\mu\theta^2(1+\beta(s-2\gamma-2))} N^{1+\beta(s-2\gamma-2)}, \tag{3.22}$$

$$\int_0^T \left\langle (-A)^s \bar{X}_t^N, \int_0^t \bar{X}_r^N \mathrm{d}r \right\rangle \mathrm{d}t \asymp \frac{\sigma^2(1-e^{-2\mu T})\Lambda^{s-2\gamma-2}}{4\mu\theta^2(1+\beta(s-2\gamma-2))} N^{1+\beta(s-2\gamma-2)}, \tag{3.23}$$

whenever s is such that the right-hand side diverges. All statements remain true if the left-hand side is replaced by its expected value.

(iii) Let $\eta > 0$ and $s_0 < s^* = 1 + 2\gamma - 1/\beta$. Assume $X_0 \in H_{s^*+\eta}$. If F satisfies $(F_{s,\eta})$ for all $s_0 \leq s < s^*$ and $X \in R(s_0)$ a.s., then (3.19) *and* (3.21) *remain true if $\bar{X}^N$ is replaced by X^N.*

Remark 3.5. *Comparing* (3.19) *and* (3.22)*, we see that $\int_0^\cdot \bar{X}_r \mathrm{d}r$ exhibits more spatial regularity than $\bar{X}$, namely one derivative in the scale of Sobolev spaces $(H_s)_{s\in\mathbb{R}}$. From the point of view of a deterministic heat equation, one may expect that one temporal derivative corresponds to two spatial derivatives. This does not apply here due to nontrivial interactions with the noise.*

Proof.

(i) First, (3.14), (3.15) and (3.16) follow from integrating the expressions (3.11), (3.12) and (3.13). Further, (3.17) and (3.18) are a direct consequence of $\int_0^t \bar{x}_r^{(k)} \mathrm{d}r = (v_t^{(k)} - \bar{x}_t^{(k)})/(\theta\lambda_k)$ and (3.14), (3.15) and (3.16).

(ii) First, if the left-hand side is replaced by its expected value, we use (i) together with $\lambda_k \asymp \Lambda k^\beta$ and the series expansion of every term, for example,

$$\mathbb{E} \int_0^T \left\langle (-A)^s V_t^N, \bar{X}_t^N \right\rangle \mathrm{d}t = \sum_{k=1}^N \lambda_k^s \left(\mathbb{E} \int_0^T v_t^{(k)} \bar{x}_t^{(k)} \mathrm{d}t \right), \tag{3.24}$$

and similar expansions for the other terms. The claim is immediate in that case. It remains to prove that the claim is still true for every realization of the left-hand side outside a set of measure zero. Now, (3.19) and (3.20) follow directly from Lemma A.2 (ii), setting $X_k^*(t) = \lambda_k^{s/2}\bar{x}_t^{(k)}$ and $X_k^*(t) = \lambda_k^{s/2} v_t^{(k)}$, respectively, in the notation therein. For (3.21), the argument is similar: From Lemma 3.2, we see that

$$\begin{aligned} A_k &:= \sup_{0\le r,t\le T} \left|\mathbb{E}[v_r^{(k)} v_t^{(k)}]\right| \lesssim \lambda_k^{-2\gamma}, \\ B_k &:= \sup_{0\le r,t\le T} \left|\mathbb{E}[v_r^{(k)} \bar{x}_t^{(k)}]\right| \lesssim \lambda_k^{-2\gamma-1}, \\ C_k &:= \sup_{0\le t\le T} \int_0^t \left|\mathbb{E}[\bar{x}_r^{(k)} \bar{x}_t^{(k)}]\right| \mathrm{d}r \lesssim \lambda_k^{-2\gamma-2}. \end{aligned}$$

Set $Y_k = \lambda_k^s \int_0^T v_t^{(k)} \bar{x}_t^{(k)} \mathrm{d}t$. Then by means of the Wick theorem [Jan97, Theorem 1.28], applied to the mixed moment $\mathbb{E}[v_t^{(k)} v_r^{(k)} \bar{x}_t^{(k)} \bar{x}_r^{(k)}]$,

$$\begin{aligned} \mathrm{Var}(Y_k) &= \lambda_k^{2s} \int_0^T \int_0^T \mathbb{E}[v_t^{(k)} v_r^{(k)} \bar{x}_t^{(k)} \bar{x}_r^{(k)}] - \mathbb{E}[v_t^{(k)} \bar{x}_t^{(k)}]\mathbb{E}[v_r^{(k)} \bar{x}_r^{(k)}] \mathrm{d}r\mathrm{d}t \\ &= 2\lambda_k^{2s} \int_0^T \int_0^t \mathbb{E}[v_t^{(k)} v_r^{(k)}]\mathbb{E}[\bar{x}_t^{(k)} \bar{x}_r^{(k)}] + \mathbb{E}[v_t^{(k)} \bar{x}_r^{(k)}]\mathbb{E}[v_r^{(k)} \bar{x}_t^{(k)}] \mathrm{d}r\mathrm{d}t \\ &\le 2\lambda_k^{2s} \left(T A_k C_k + T^2 B_k^2/2\right) \lesssim \lambda_k^{2s-4\gamma-2}. \end{aligned}$$

Now, we see that

$$\sum_{N=1}^{\infty} \frac{\mathrm{Var} Y_N}{\left(\sum_{k=1}^N \mathbb{E} Y_k\right)^2} \lesssim \sum_{N=1}^{\infty} \frac{1}{N^2} < \infty,$$

and (3.21) follows from the strong law of large numbers [Shi96, Theorem IV.3.2]. Now, (3.22) and (3.23) follow from (3.19), (3.20), (3.21) via

$$\theta \int_0^t A \bar{X}_r^N \mathrm{d}r = \bar{X}_t^N - V_t^N. \tag{3.25}$$

(iii) By Lemma 3.3, condition $(F_{s,\eta})$ for F and Proposition 2.4, $\widetilde{X} \in R(s+\eta)$ for each $s < s^*$. The analogue of (3.19) follows as in the white noise case in Proposition 2.8. For the analogue of (3.21), let $s > s^*$ and $\epsilon > 0$

with $\eta - 2(s-s^*) < \epsilon < \eta$. Then

$$\begin{aligned}\left|\int_0^T \left\langle (-A)^s V_t^N, \widetilde{X}_t^N \right\rangle \mathrm{d}t\right| \\ &\leq \sqrt{\int_0^T \|V_t^N\|^2_{2s-s^*-\eta+\epsilon}\,\mathrm{d}t}\sqrt{\int_0^T \left\|\widetilde{X}_t^N\right\|^2_{s^*+\eta-\epsilon}\mathrm{d}t} \\ &\lesssim N^{(1+\beta(2s-s^*-\eta+\epsilon-2\gamma))/2} = N^{1+\beta(s-2\gamma-1-\eta/2+\epsilon/2)}\end{aligned}$$

by (3.20). Note that the latter exponent is positive due to $\epsilon > \eta - 2(s-s^*)$. Furthermore, $1+\beta(s-2\gamma-1-\eta/2+\epsilon/2) < 1+\beta(s-2\gamma-1)$, such that

$$\begin{aligned}\int_0^T \left\langle V_t^N, X_t^N \right\rangle_s \mathrm{d}t &= \int_0^T \left\langle V_t^N, \bar{X}_t^N \right\rangle_s \mathrm{d}t + \int_0^T \left\langle V_t^N, \widetilde{X}_t^N \right\rangle_s \mathrm{d}t \\ &\asymp \frac{\sigma^2 C_{T,\mu}^{(+)} \Lambda^{s-2\gamma-1}}{2\theta(1+\beta(s-2\gamma-1))} N^{1+\beta(s-2\gamma-1)}.\end{aligned}$$

This concludes the proof.

□

Let J denote the Bochner integral operator, i.e. $JZ(t) = \int_0^t Z_r \mathrm{d}r$ for $Z \in L^1(0,T;H_s)$, $s \in \mathbb{R}$. It is desirable to transfer also (3.22) to the nonlinear case, i.e. to substitute $\bar{X}$ by X therein. In order to do so, we have to strengthen the condition on F. In addition to $(F_{s,\eta})$ for F, we need:

$(F_{s,\eta}^{\mathcal{J}})$ One of the following two conditions holds:

- (i) F satisfies $(F_{s,1+\eta})$.
- (ii) There is an operator G that satisfies $(F_{s,\eta})$ such that $JF = GJ$.

Lemma 3.6. *Let $\eta > 0$ and $s_0 < s^*$ such that F satisfies $(F_{s,\eta})$ and $(F_{s,\eta}^{\mathcal{J}})$ for all $s_0 \leq s < s^*$. Assume a.s. $X \in R(s_0)$ and $X_0 \in H_{s^*+\eta}$. Then (3.22) remains true if $\bar{X}^N$ is replaced by X^N. Furthermore, in this case, $(J \circ F)(X) \in R(s-1+\eta)$ for $s < s^*$.*

Proof. Lemma 3.3 and Proposition 2.4 yield $\widetilde{X} \in R(s+\eta)$ and therefore also $J\widetilde{X} \in R(s+\eta)$ for $s < s^*$. We distinguish the two cases from $(F_{s,\eta}^{\mathcal{J}})$ and prove that $J\widetilde{X} \in R(s+1+\eta)$, $s < s^*$, in either case:

(i) If, in fact, F satisfies even $(F_{s,1+\eta})$, another application of Proposition 2.4 proves $\widetilde{X}, J\widetilde{X} \in R(s+1+\eta)$ for $s < s^*$.

(ii) If $JF = GJ$, where G satisfies $(F_{s,\eta})$, we proceed as in the proof of Proposition 2.3: We know that $J\bar{X} \in R(s)$ for any $s < s^*+1$ due to Proposition 3.4. Let $s < s^*+1$ such that $J\widetilde{X} \in R(s)$, this is the case e.g. for $s = s^*$. Then also $JX \in R(s)$, and

$$\begin{aligned}
\left\| J\widetilde{X}_t^N \right\|_{s+\eta} &\leq \int_0^t \left\| e^{(t-r)\theta A}\left(P_N (J\circ F)(X)(r) + X_0^N\right)\right\|_{s+\eta} \mathrm{d}r \\
&\lesssim \int_0^t (t-r)^{-1+\epsilon/2} \left\|(J\circ F)(X)(r) + X_0\right\|_{s-2+\eta+\epsilon} \mathrm{d}r \\
&\lesssim \left(\sup_{0\leq r\leq T} \left\|(G\circ J)(X)(r)\right\|_{s-2+\eta+\epsilon} + \left\|X_0\right\|_{s+\eta}\right) \int_0^t (t-r)^{-1+\epsilon/2} \mathrm{d}r \\
&\lesssim \left(\sup_{0\leq r\leq T} \left\|JX_r\right\|_s + \left\|X_0\right\|_{s+\eta}\right) \frac{2}{\epsilon} T^{\epsilon/2} < \infty,
\end{aligned}$$

so $J\widetilde{X} \in R(s+\eta)$. Iterating this argument proves $J\widetilde{X} \in R(s+1+\eta)$ for all $s < s^*$.

In particular, for $s > 2 + 2\gamma - 1/\beta$ and any $\epsilon > 0$:

$$\begin{aligned}
\int_0^T \left\| \int_0^t (-A)^{s/2} \widetilde{X}_r^N \mathrm{d}r \right\|^2 \mathrm{d}t &\leq \lambda_N^{s-s^*-1-\eta+\epsilon} \int_0^T \left\| \int_0^t (-A)^{(s^*+1+\eta-\epsilon)/2} \widetilde{X}_r^N \mathrm{d}r \right\|^2 \mathrm{d}t \\
&\lesssim \lambda_N^{s-2-2\gamma+1/\beta-\eta+\epsilon} \lesssim N^{1+\beta(s-2\gamma-2-\eta+\epsilon)},
\end{aligned}$$

where we assume w.l.o.g. that the exponent is positive. As a consequence,

$$\int_0^T \left\| \int_0^t (-A)^{s/2} X_r \mathrm{d}r \right\|^2 \mathrm{d}t \asymp \int_0^T \left\| \int_0^t (-A)^{s/2} \bar{X}_r \mathrm{d}r \right\|^2 \mathrm{d}t,$$

and (3.22) holds with $\bar{X}$ replaced by X. Finally, $AJ\widetilde{X} \in R(s-1+\eta)$, thus $JF(X) = \widetilde{X} - \theta AJ\widetilde{X} - X_0 \in R(s-1+\eta)$ for $s < s^*$. □

3.1.2 The Maximum–Likelihood Approach

Heuristically, the log–likelihood is given by [LS77, Section 7.6.4]:

$$
\begin{aligned}
\ln \frac{\mathrm{d}\mathbb{P}^{N,T}_{(\theta,\mu)}}{\mathrm{d}\mathbb{P}^{N,T}_{(\theta_0,\mu_0)}}(X^N) &= \frac{1}{\sigma^2}\int_0^T \left\langle a(\theta,\mu)-a(\theta_0,\mu_0), (-A)^{2\gamma}\mathrm{d}X_t^N\right\rangle \\
&\quad - \frac{1}{2\sigma^2}\int_0^T \left\langle a(\theta,\mu)-a(\theta_0,\mu_0), (-A)^{2\gamma}(a(\theta,\mu)+a(\theta_0,\mu_0))\right\rangle \mathrm{d}t \\
&\quad - \frac{1}{\sigma^2}\int_0^T \left\langle a(\theta,\mu)-a(\theta_0,\mu_0), (-A)^{2\gamma}P_N F(X)(t)\right\rangle \mathrm{d}t,
\end{aligned}
\tag{3.26}
$$

where we abbreviate

$$
a(\theta,\mu) = \theta A X_t^N - \mu X_t^N + \mu X_0^N + \mu\theta\int_0^t A X_r^N \mathrm{d}r + \mu\int_0^t P_N F(X)(r)\mathrm{d}r.
$$

As before, this is rigorous if $P_N F = F P_N$. Maximizing for the unknown parameter θ for known μ yields the maximum likelihood–type estimator:

$$
\begin{aligned}
\hat{\theta}_N^{\mathrm{ref}} &= -\frac{\int_0^T \left\langle (-A)^{1+2\alpha}X_t^N + \mu\int_0^t (-A)^{1+2\alpha}X_r^N\mathrm{d}r, \mathrm{d}X_t^N\right\rangle}{\int_0^T \left\|(-A)^{1+\alpha}X_t^N + \mu\int_0^t(-A)^{1+\alpha}X_r^N\mathrm{d}r\right\|^2\mathrm{d}t} \\
&\quad - \mu\frac{\int_0^T \left\langle (-A)^{1+2\alpha}\left(X_t^N + \mu\int_0^t X_r^N\mathrm{d}r\right), X_t^N - X_0^N - \int_0^t P_N F(X)(r)\mathrm{d}r\right\rangle\mathrm{d}t}{\int_0^T \left\|(-A)^{1+\alpha}X_t^N + \mu\int_0^t(-A)^{1+\alpha}X_r^N\mathrm{d}r\right\|^2\mathrm{d}t} \\
&\quad + \frac{\int_0^T \left\langle (-A)^{1+2\alpha}X_t^N + \mu\int_0^t(-A)^{1+2\alpha}X_r^N\mathrm{d}r, P_N F(X)(t)\right\rangle\mathrm{d}t}{\int_0^T \left\|(-A)^{1+\alpha}X_t^N + \mu\int_0^t(-A)^{1+\alpha}X_r^N\mathrm{d}r\right\|^2\mathrm{d}t},
\end{aligned}
\tag{3.27}
$$

whereas maximizing for unknown μ and known θ yields

$$
\begin{aligned}
\hat{\mu}_N^{\mathrm{ref}} &= -\frac{\int_0^T \left\langle (-A)^{2\alpha}V_t^N, \mathrm{d}X_t^N\right\rangle}{\int_0^T \|(-A)^{\alpha}V_t^N\|^2\mathrm{d}t} - \theta\frac{\int_0^T \left\langle (-A)^{1+2\alpha}V_t^N, X_t^N\right\rangle\mathrm{d}t}{\int_0^T \|(-A)^{\alpha}V_t^N\|^2\mathrm{d}t} \\
&\quad + \frac{\int_0^T \left\langle (-A)^{2\alpha}V_t^N, P_N F(X)(t)\right\rangle\mathrm{d}t}{\int_0^T \|(-A)^{\alpha}V_t^N\|^2\mathrm{d}t},
\end{aligned}
\tag{3.28}
$$

where

$$V_t^N = X_t^N - X_0^N - \theta \int_0^t AX_r^N \mathrm{d}r - \int_0^t P_N F(X)(r) \mathrm{d}r$$

is a functional of X^N and $P_N F(X)$.

In both estimators, we substituted γ by a contrast parameter $\alpha \in \mathbb{R}$, as before. Clearly, setting $\theta = \hat{\theta}_N^{\mathrm{ref}}$ and $\mu = \hat{\mu}_N^{\mathrm{ref}}$ in the above expressions leads to a (nonlinear) system of equations for the maximum likelihood estimators in the case that both parameters are unknown. However, we are interested in a hierarchical approach of first estimating θ independently of μ and secondly estimating μ based on the previous estimator of θ, exploiting the asymptotic properties of the terms appearing in $\hat{\theta}_N^{\mathrm{ref}}$ and $\hat{\mu}_N^{\mathrm{ref}}$. This will be explained in detail in the following sections. The hierarchical approach is insightful for two reasons:

(i) From the point of view of model misspecification, the diffusivity estimators from Section 2.3 still work if the driving noise V exhibits temporal correlation which is not accounted for in the model, as long as the temporal regularity of X is not affected (as in the case of fractional or integrated noise, cf. Section 3.2).

(ii) The hierarchical approach is (at least asymptotically) as good as the direct approach in the following sense: Let $A = \Delta$ be the Laplacian. In dimension $d = 1$, Theorem 3.7 below shows that the hierarchical approach leads to an estimator for θ which is agnostic to μ, and which has the same asymptotic properties as the reference estimator $\hat{\theta}_N^{\mathrm{ref}}$ with known μ. In $d = 2$, still the optimal convergence rate is preserved. Further, by Theorem 3.10 below, a hierarchical estimator for μ behaves asymptotically as $\hat{\mu}_N^{\mathrm{ref}}$ with known θ whenever $d \leq 3$.

3.1.3 Diffusivity Estimation

As explained above, it is reasonable to consider a simplified estimator for θ which is obtained by formally setting $\mu = 0$ in $\hat{\theta}_N^{\mathrm{ref}}$:

$$\hat{\theta}_N^{\mathrm{sim}} = -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, \mathrm{d}X_t^N \right\rangle}{\int_0^T \left\| (-A)^{1+\alpha} X_t^N \right\|^2 \mathrm{d}t} + \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F(X)(t) \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{1+\alpha} X_t^N \right\|^2 \mathrm{d}t}. \tag{3.29}$$

Formally, this is the same estimator as $\hat{\theta}_N^{\text{full}}$ in the white noise setting. In particular, $\hat{\theta}_N^{\text{sim}}$ does not depend on μ. Of course, the other estimators $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ from Section 2.3 can be used, too, with conditions on the excess regularity η of F as in Theorem 2.11.

Theorem 3.7. *Let $\gamma > 1/(2\beta)$. Assume that there is $\eta > 0$ and $s_0 < s^*$ such that $(F_{s,\eta})$ is true for $s_0 \leq s < s^*$, and that $X \in R(s_0)$ and $X_0 \in H_{s^*+\eta}$ a.s. Let $\alpha > \gamma - (1 + 1/\beta)/4$.*

(i) If μ is known, $\hat{\theta}_N^{\text{ref}}$ is asymptotically normal as in the white noise case:

$$N^{\frac{1+\beta}{2}}(\hat{\theta}_N^{\text{ref}} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{3.30}$$

where Σ is given by (2.30).

(ii) If $\beta > 1$, then $\hat{\theta}_N^{\text{sim}}$ is asymptotically normal as in (3.30). *If $\beta = 1$, then*

$$N(\hat{\theta}_N^{\text{sim}} - \theta) \xrightarrow{d} \mathcal{N}(m, \Sigma), \tag{3.31}$$

where

$$m = \left(\mu + \frac{1}{2T}(1 - e^{-2\mu T})\right) \frac{1 + \beta(2\alpha - 2\gamma + 1)}{\Lambda(1 + \beta(2\alpha - 2\gamma))}. \tag{3.32}$$

If $\beta < 1$, then

$$N^{\beta}(\hat{\theta}_N^{\text{sim}} - \theta) \xrightarrow{a.s.} m. \tag{3.33}$$

If $\eta > 1 + 1/\beta$, then (3.30), (3.31), (3.33) *for $\beta > 1$, $\beta = 1$, $\beta < 1$, respectively, hold for $\hat{\theta}_N^{\text{part}}$ and $\hat{\theta}_N^{\text{lin}}$ as well. If $\eta \leq 1 + 1/\beta$, then a.s. $\hat{\theta}_N^{\text{part}} = \theta + o(N^{-a})$ for $a < \beta(\eta/2 \wedge 1)$, and the same is true for $\hat{\theta}_N^{\text{lin}}$.*

Proof.

(i) Plugging in the dynamics of X^N into (3.27), we obtain

$$\hat{\theta}_N^{\text{ref}} - \theta = -\sigma \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N + \mu \int_0^t (-A)^{1+2\alpha-\gamma} X_r^N \mathrm{d}r, \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\| (-A)^{1+\alpha} X_t^N + \mu \int_0^t (-A)^{1+\alpha} X_r^N \mathrm{d}r \right\|^2 \mathrm{d}t}.$$

Set $C_s = \sigma^2 T\Lambda^{s-2\gamma-1}/(2\theta(1+\beta(s-2\gamma-1)))$, whenever $s > 1+2\gamma-1/\beta$. By Proposition 2.4, $\widetilde{X} \in R(s+\eta)$ for $s < s^*$. As J maps $R(s+\eta)$ into itself, the same is true for $J\widetilde{X}$. From Proposition 3.4 (ii), comparing the rates of (3.19) and (3.22), we get immediately

$$\begin{aligned} I_s^N &:= \int_0^T \left\| (-A)^{s/2} X_t^N + \mu \int_0^t (-A)^{s/2} X_r^N \mathrm{d}r \right\|^2 \mathrm{d}t \\ &\asymp \int_0^T \left\| (-A)^{s/2} \bar{X}_t^N + \mu \int_0^t (-A)^{s/2} \bar{X}_r^N \mathrm{d}r \right\|^2 \mathrm{d}t \asymp C_s N^{1+\beta(s-2\gamma-1)}. \end{aligned} \tag{3.34}$$

We write

$$\hat{\theta}_N^{\mathrm{ref}} - \theta =: -\sigma \frac{C_{2+4\alpha-2\gamma}^{1/2} N^{1/2+\beta(2\alpha-2\gamma+1/2)}}{I_{2+2\alpha}^N} M_T^N,$$

such that $(M^N)_{N\in\mathbb{N}}$ is a sequence of local martingales with $\langle M^N \rangle_T \xrightarrow{\mathbb{P}} 1$. According to Theorem A.1, it follows that $M_T^N \xrightarrow{d} \mathcal{N}(0,1)$, and making use of Slutsky's lemma, we see that

$$N^{\frac{1+\beta}{2}} (\hat{\theta}_N^{\mathrm{ref}} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2 C_{2+4\alpha-2\gamma}}{C_{2+2\alpha}^2}\right). \tag{3.35}$$

(ii) We decompose $\hat{\theta}_N^{\mathrm{sim}}$ as follows:

$$\hat{\theta}_N^{\mathrm{sim}} - \theta = \mu \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, V_t^N \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{1+\alpha} X_t^N \right\|^2 \mathrm{d}t} - \sigma \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\| (-A)^{1+\alpha} X_t^N \right\|^2 \mathrm{d}t}.$$

As before, the second term converges in distribution to $\mathcal{N}(0,\Sigma)$ with rate $N^{-(1+\beta)/2}$. Using Proposition 3.4 (iii), we have

$$\mu \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, V_t^N \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{1+\alpha} X_t^N \right\|^2 \mathrm{d}t} N^\beta \xrightarrow{a.s.} \mu \frac{C_{T,\mu}^{(+)}(1+\beta(2\alpha-2\gamma+1))}{T\Lambda(1+\beta(2\alpha-2\gamma))},$$

and the right-hand side equals m. This yields the claim in the case $\beta > 1$ and $\beta = 1$, and for $\beta < 1$ note that by Lemma 2.6, we have almost surely $\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle \lesssim N^{1/2+\beta(1/2+2\alpha-2\gamma)}$, and the claim follows in this case as well. The statements regarding $\hat{\theta}_N^{\mathrm{part}}$ and $\hat{\theta}_N^{\mathrm{lin}}$ are straightforward, taking into account an additional bias term of order N^{-b} for $b < \beta\eta/2$ as in Theorem 2.11, coming from the nonlinearity.

□

Remark 3.8. *$\hat{\theta}_N^{\mathrm{sim}}$ is identical to $\hat{\theta}_N^{\mathrm{full}}$ from Section 2.3. Thus, Theorem 3.7 is revelatory for the case that the true noise model is of Ornstein–Uhlenbeck type, but the diffusivity estimator is derived under a white noise assumption. In the reference case $\beta = 2/d$, this translates as follows: In $d = 1$, $\hat{\theta}_N^{\mathrm{sim}}$ is asymptotically normal with optimal convergence rate. In particular, $\hat{\theta}_N^{\mathrm{full}}$ from Section 2.3 is asymptotically robust to noise misspecification of Ornstein–Uhlenbeck type. In $d = 2$, $\hat{\theta}_N^{\mathrm{sim}}$ converges to a non-centered normal distribution, still with optimal rate. In $d \geq 3$, $\hat{\theta}_N^{\mathrm{sim}}$ is still consistent for θ, but its convergence rate is no longer optimal.*

3.1.4 Correlation Decay Estimation

In contrast to the case of diffusivity estimation, we cannot just set the nuisance parameter θ in $\hat{\mu}_N^{\mathrm{ref}}$ to zero: According to Proposition 3.4, the term $\theta \int_0^t (-A)^{1+2\alpha} X_r^N \mathrm{d}r$ dominates the denominator of (3.28). As a consequence, estimation of μ depends on knowledge (or precise estimation) of θ. In the sequel, we set

$$V_t^{\mathrm{full}}(\vartheta) := X_t - X_0 + \vartheta \int_0^t (-A) X_r \mathrm{d}r - \int_0^t F(X)(r) \mathrm{d}r, \tag{3.36}$$

$$V_t^{\mathrm{lin}}(\vartheta) := X_t - X_0 + \vartheta \int_0^t (-A) X_r \mathrm{d}r, \tag{3.37}$$

then $V^{\mathrm{full},N}(\vartheta) = P_N V^{\mathrm{full}}(\vartheta)$ and $V^{\mathrm{lin},N}(\vartheta) = P_N V^{\mathrm{lin}}(\vartheta)$ are given by the same terms, with X and $F(X)$ replaced by X^N and $P_N F(X)$. Further, we set for $\vartheta > 0$:

$$\begin{aligned}
\hat{\mu}_N^{\mathrm{full}}(\vartheta) := &- \frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\mathrm{full},N}(\vartheta), \mathrm{d}X_t^N \right\rangle}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\vartheta) \right\|^2 \mathrm{d}t} \\
&- \vartheta \frac{\int_0^T \left\langle (-A)^{1+2\alpha} V_t^{\mathrm{full},N}(\vartheta), X_t^N \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\vartheta) \right\|^2 \mathrm{d}t} \\
&+ \frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\mathrm{full},N}(\vartheta), P_N F(X)(t) \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\vartheta) \right\|^2 \mathrm{d}t}
\end{aligned} \tag{3.38}$$

and

$$\hat{\mu}_N^{\mathrm{lin}}(\vartheta) := -\frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\mathrm{lin},N}(\vartheta), \mathrm{d}X_t^N \right\rangle}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{lin},N}(\theta) \right\|^2 \mathrm{d}t}$$
$$- \theta \frac{\int_0^T \left\langle (-A)^{1+2\alpha} V_t^{\mathrm{lin},N}(\vartheta), X_t^N \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{lin},N}(\vartheta) \right\|^2 \mathrm{d}t}. \tag{3.39}$$

Note that if $\vartheta = \theta$ is the true diffusivity, then $\hat{\mu}_N^{\mathrm{full}}(\theta) = \hat{\mu}_N^{\mathrm{ref}}$ as given in (3.28). If ϑ is close to θ, then $V^{\mathrm{full}}(\vartheta)$ and $V^{\mathrm{lin}}(\vartheta)$ should be seen as an approximation of V. This is formalized as follows:

Lemma 3.9. *Let $\eta > 0$, $s_0 \in \mathbb{R}$ such that F satisfies $(F_{s,\eta})$ and $(F_{s,\eta}^{\mathcal{J}})$ for $s_0 \leq s < s^*$. Assume $X \in R(s_0)$ and $X_0 \in H_{s^*+\eta}$. Let $(\theta^N)_{N\in\mathbb{N}}$ a sequence of estimators for θ which is a.s. consistent. Then, for $s > 2\gamma - 1/\beta$,*

$$\int_0^T \left\| (-A)^{\frac{s}{2}} \left(V_t^{\mathrm{full},N}(\theta^N) - V_t \right) \right\|^2 \mathrm{d}t = o(N^{1+\beta(s-2\gamma)}). \tag{3.40}$$

In particular,

$$\int_0^T \left\| (-A)^{\frac{s}{2}} V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t \asymp \int_0^T \left\| (-A)^{\frac{s}{2}} V_t \right\|^2 \mathrm{d}t \asymp C_s^{\mathrm{OU}} N^{1+\beta(s-2\gamma)}, \tag{3.41}$$

where $C_s^{\mathrm{OU}} = \sigma^2 C_{T,\mu}^{(-)} \Lambda^{s-2\gamma}/(2\mu(1+\beta(s-2\gamma)))$. The same statements are true for $V^{\mathrm{lin},N}(\theta^N)$.

Proof. Let $\epsilon > 0$. If $s > s^* - 1 + \eta - \epsilon$, using $JF(X) \in R(s^* - 1 + \eta - \epsilon)$ by Lemma 3.6, we have

$$\int_0^T \left\| (-A)^{\frac{s}{2}} \int_0^t P_N F(X)(r) \mathrm{d}r \right\|^2 \mathrm{d}t$$
$$\leq \lambda_N^{s-s^*+1-\eta+\epsilon} \int_0^T \left\| (-A)^{\frac{s^*-1+\eta-\epsilon}{2}} \int_0^t P_N F(X)(r) \mathrm{d}r \right\|^2 \mathrm{d}t$$
$$\lesssim N^{\beta(s-s^*+1-\eta+\epsilon)} = N^{1+\beta(s-2\gamma-\eta+\epsilon)},$$

thus for ϵ sufficiently small, this grows slower than $N^{1+\beta(s-2\gamma)}$. If $s < s^* - 1 + \eta - \epsilon$, the left-hand side is even bounded uniformly in N. The case

$s = s^* - 1 + \eta - \epsilon$ can be avoided by substituting $\epsilon \mapsto \epsilon/2$. Using that $X = X_0 + \theta JAX + JF(X) + V$, we see that

$$V_t^{\mathrm{full},N}(\theta) = V_t, \qquad V_t^{\mathrm{lin},N}(\theta) - V_t = \int_0^t F(X)(r)\mathrm{d}r,$$

and consequently,

$$\begin{aligned}\int_0^T \left\|(-A)^{\frac{s}{2}}\left(V_t^{\mathrm{lin},N}(\theta) - V_t^N\right)\right\|^2 \mathrm{d}t &\lesssim \int_0^T \left\|(-A)^{\frac{s}{2}}\int_0^t P_N F(X)(r)\mathrm{d}r\right\|^2 \mathrm{d}t \\ &\ll_p N^{1+\beta(s-2\gamma)},\end{aligned}$$

and the same estimate is trivially satisfied for $V^{\mathrm{full},N}$ instead of $V^{\mathrm{lin},N}$. Next, again by Lemma 3.6, we have

$$\begin{aligned}\int_0^T &\left\|(-A)^{\frac{s}{2}}\left(V_t^{\mathrm{full},N}(\theta^N) - V_t^{\mathrm{full},N}(\theta)\right)\right\|^2 \mathrm{d}t \\ &= (\theta^N - \theta)^2 \int_0^T \left\|(-A)^{\frac{s+2}{2}}\int_0^t X_r^N \mathrm{d}r\right\|^2 \mathrm{d}t \\ &\asymp (\theta^N - \theta)^2 \theta^{-2} C_s^{\mathrm{OU}} N^{1+\beta(s-2\gamma)},\end{aligned}$$

and the same is true for $V^{\mathrm{lin},N}$ instead of $V^{\mathrm{full},N}$. As θ^N is a consistent estimator for θ, the right-hand side is negligible compared to $N^{1+\beta(s-2\gamma)}$. Now (3.40) follows by simple norm estimates, and (3.41) is a direct consequence of (3.40). □

Theorem 3.10. *Let $\eta > 0$, $s_0 \in \mathbb{R}$ such that F satisfies $(F_{s,\eta})$ and $(F_{s,\eta}^{\mathcal{J}})$ for $s_0 \leq s < s^*$. Let $X \in R(s_0)$ and $X_0 \in H_{s^*+\eta}$ almost surely. Let $\alpha > \gamma - 1/(4\beta)$.*

(i) If $\beta > 1/2$, then

$$N^{\frac{1}{2}}\left(\hat{\mu}_N^{\mathrm{full}}\left(\hat{\theta}_N^{\mathrm{sim}}\right) - \mu\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_\mu) \tag{3.42}$$

with

$$\Sigma_\mu = \frac{4\mu^2(1+\beta(2\alpha-2\gamma))^2}{(e^{-2\mu T} - 1 + 2\mu T)(1+\beta(4\alpha-4\gamma))}. \tag{3.43}$$

(ii) If $\beta = 1/2$, then

$$N^{\frac{1}{2}}\left(\hat{\mu}_N^{\mathrm{full}}\left(\hat{\theta}_N^{\mathrm{sim}}\right) - \mu\right) \xrightarrow{d} \mathcal{N}(-m\mu/\theta, \Sigma_\mu), \tag{3.44}$$

where m is given by (3.32).

(iii) If $\beta < 1/2$, then

$$N^{\beta}\left(\hat{\mu}_N^{\mathrm{full}}\left(\hat{\theta}_N^{\mathrm{sim}}\right) - \mu\right) \xrightarrow{a.s.} -\frac{m\mu}{\theta}. \tag{3.45}$$

If F satisfies $(F_{s,\bar{\eta}})$ for some $\bar{\eta} > 1$ and all $s_0 \leq s < s^*$, $\hat{\mu}_N^{\mathrm{lin}}\left(\hat{\theta}_N^{\mathrm{lin}}\right)$ is a consistent estimator for μ. If $\bar{\eta} > 1 + 1/\beta$, (i), (ii), (iii) remain true for $\hat{\mu}_N^{\mathrm{lin}}\left(\hat{\theta}_N^{\mathrm{lin}}\right)$. Otherwise, $\hat{\mu}_N^{\mathrm{lin}}\left(\hat{\theta}_N^{\mathrm{lin}}\right) = \mu + o(N^{-b})$ for every $b < (\beta(\bar{\eta}-1)/2) \wedge \beta$.

Proof. We write $\theta^N = \hat{\theta}_N^{\mathrm{sim}}$ for short. Expanding $\hat{\mu}_N^{\mathrm{full}}\left(\theta^N\right)$ by plugging in the dynamics of X^N, we see that

$$\begin{aligned}
\hat{\mu}_N^{\mathrm{full}}\left(\theta^N\right) &= \mu \frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\mathrm{full},N}(\theta^N), V_t^N \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&\quad - \sigma \frac{\int_0^T \left\langle (-A)^{2\alpha-\gamma} V_t^{\mathrm{full},N}(\theta^N), \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&=: [I]_N - [II]_N.
\end{aligned} \tag{3.46}$$

For the second term, one would like to apply Theorem A.1 as in the previous cases. Note, however, that θ^N depends on the whole trajectory of $(X_t)_{0\leq t\leq T}$, so the integrand in the stochastic integral is not adapted. Nonetheless, as $V_t^{\mathrm{full},N}(\theta^N)$ is an affine function of θ^N, this issue is easy to avoid by decomposing the integrand as $V_t^{\mathrm{full},N}(\theta^N) = V_t^{\mathrm{full},N}(\theta) + (\theta^N - \theta)\int_0^t (-A) X_r^N \mathrm{d}r$:

$$\begin{aligned}
\sigma \frac{\int_0^T \left\langle (-A)^{2\alpha-\gamma} V_t^{\mathrm{full},N}(\theta^N), \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} &= \sigma \frac{\int_0^T \left\langle (-A)^{2\alpha-\gamma} V_t^{\mathrm{full},N}(\theta), \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&\quad + (\theta^N - \theta)\sigma \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} \int_0^t X_r^N \mathrm{d}r, \mathrm{d}W_t^N \right\rangle}{\int_0^T \left\| (-A)^{\alpha} V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&=: [IIa]_N + (\theta^N - \theta)[IIb]_N.
\end{aligned}$$

Now we rescale both terms separately with the square root of the quadratic variation processes of their stochastic integrals and apply Theorem A.1, using $\alpha > \gamma - 1/(4\beta)$ together with Lemma 3.9 and Proposition 3.4 (iii). This yields

$$N^{1/2}[IIa]_N \xrightarrow{d} \mathcal{N}(0, \sigma^2 C^{\mathrm{OU}}_{4\alpha-2\gamma}/(C^{\mathrm{OU}}_{2\alpha})^2), \tag{3.47}$$

$$N^{1/2}[IIb]_N \xrightarrow{d} \mathcal{N}(0, \sigma^2 C^{\mathrm{OU}}_{4\alpha-2\gamma}/(\theta C^{\mathrm{OU}}_{2\alpha})^2), \tag{3.48}$$

As $\theta^N \to \theta$ almost surely, (3.47) holds for $[II]_N$ instead of $[IIa]_N$ as well. The term $[I]_N$ can be treated as follows:

$$\begin{aligned}
[I]_N - \mu &= \mu \frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\mathrm{full},N}(\theta^N), V_t^N - V_t^{\mathrm{full},N}(\theta^N) \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^\alpha V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&= -(\theta^N - \theta)\mu\theta^N \frac{\int_0^T \left\| (-A)^{1+\alpha} \int_0^t X_r^N \mathrm{d}r \right\|^2 \mathrm{d}t}{\int_0^T \left\| (-A)^\alpha V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&\quad - (\theta^N - \theta)\mu \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, \int_0^t X_r^N \mathrm{d}r \right\rangle \mathrm{d}t}{\int_0^T \left\| (-A)^\alpha V_t^{\mathrm{full},N}(\theta^N) \right\|^2 \mathrm{d}t} \\
&\quad + (\theta^N - \theta)\mu \frac{\int_0^T \left\langle (-A)^{1+2\alpha} \left(X_0^N + \int_0^t P_N F(X)(r) \mathrm{d}r \right), \int_0^t X_r^N \mathrm{d}r \right\rangle \mathrm{d}t}{\int_0^T \| (-A)^\alpha V^{\mathrm{full},N}(\theta^N) \|^2 \mathrm{d}t} \\
&=: [Ia]_N + [Ib]_N + [Ic]_N.
\end{aligned}$$

For the first term,

$$[Ia]_N \asymp -(\theta^N - \theta)\mu\theta \frac{\int_0^T \left\| \int_0^t X_r^N \mathrm{d}r \right\|_{2+2\alpha}^2 \mathrm{d}t}{\int_0^T \left\| V_t^{\mathrm{full},N}(\theta^N) \right\|_{2\alpha}^2 \mathrm{d}t} \asymp -\frac{\mu}{\theta}(\theta^N - \theta).$$

For $[Ib]_N$, the Cauchy-Schwarz inequality is used: If $\alpha > \gamma + 1/4 - 1/(2\beta)$,

$$
\begin{aligned}
&\left|\int_0^T \left\langle X_t^N, \int_0^t X_r^N \mathrm{d}r \right\rangle_{1+2\alpha} \mathrm{d}t\right| \\
&\qquad \leq \sqrt{\int_0^T \|X_t^N\|_{1/2+2\alpha}^2 \,\mathrm{d}t \int_0^T \left\|\int_0^t X_r^N \mathrm{d}r\right\|_{3/2+2\alpha}^2 \mathrm{d}t} \\
&\qquad \lesssim N^{1+\beta(2\alpha-2\gamma-1/2)} \\
&\qquad \ll_p N^{1+\beta(2\alpha-2\gamma)} \lesssim \int_0^T \left\|\int_0^t X_r^N \mathrm{d}r\right\|_{2+2\alpha}^2 \mathrm{d}t.
\end{aligned}
$$

If $\alpha < \gamma + 1/4 - 1/(2\beta)$, the left-hand side is bounded uniformly in N, and for $\alpha = \gamma + 1/4 - 1/(2\beta)$ replace α by $\alpha + 1/8$ under the square root term in an additional norm estimate, and continue as before. In any case, $|[Ib]_N| \ll_p [Ia]_N$. Finally, consider $[Ic]_N$. Let $0 < \epsilon < \eta$. We can neglect $X_0 \in H_{s^*+\eta}$, which has larger spatial regularity than $JF(X) \in R(s^* - 1 + \eta - \epsilon)$. Then, if $\alpha > \gamma - 1/(2\beta) + \eta/4 - \epsilon/4$, we have with the abbreviation $\bar{r} := 2 + 4\alpha - s^* + 1 - \eta + \epsilon$:

$$
\begin{aligned}
|[Ic]_N| &\lesssim \left|\theta^N - \theta\right| \frac{\sqrt{\int_0^T \left\|\int_0^t P_N F(X)(r)\mathrm{d}r\right\|_{s^*-1+\eta-\epsilon}^2 \mathrm{d}t \int_0^T \left\|\int_0^t X_r^N \mathrm{d}r\right\|_{\bar{r}}^2 \mathrm{d}t}}{\int_0^T \left\|(-A)^\alpha V_t^{\mathrm{full},N}(\theta^N)\right\|^2 \mathrm{d}t} \\
&\lesssim \left|\theta^N - \theta\right| N^{-\frac{\beta}{2}(\eta-\epsilon)},
\end{aligned}
$$

and $|[Ic]_N| \ll_p [Ia]_N$. The case $\alpha \leq \gamma - 1/(2\beta) + \eta/4 - \epsilon/4$ is treated as before. Putting things together, we have shown $[I]_N - \mu \asymp -(\theta^N - \theta)\mu/\theta$. Now (i), (ii), (iii) for $\hat{\mu}_N^{\mathrm{full}}\left(\theta^N\right)$ follow from the asymptotic behavior of $\theta^N = \hat{\theta}_N^{\mathrm{sim}}$ from Theorem 3.7.

Now, tracing the proof for $\hat{\mu}_N^{\mathrm{lin}}\left(\theta^N\right)$ (with V^{lin} instead of V^{full} and $\theta^N = \hat{\theta}_N^{\mathrm{lin}}$ instead of $\theta^N = \hat{\theta}_N^{\mathrm{sim}}$), there are two additional bias terms that have to be controlled. First, (3.46) is replaced by

$$
\begin{aligned}
\hat{\mu}_N^{\mathrm{lin}}\left(\theta^N\right) &= [I]_N - [II]_N - \frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\mathrm{lin},N}(\theta^N), P_N F(X)(t) \right\rangle \mathrm{d}t}{\int_0^T \left\|(-A)^\alpha V_t^{\mathrm{lin},N}(\theta^N)\right\|^2 \mathrm{d}t} \\
&=: [I]_N - [II]_N - [III]_N
\end{aligned}
$$

with V^{lin} instead of V^{full} in the definition of $[I]_N$ and $[II]_N$. $[II]_N$ is treated exactly as before. Due to $(F_{s,\bar\eta})$ we have $F(X) \in R(s-2+\bar\eta)$ for all $s < s^*$. Let w.l.o.g. $2\alpha - s^* + 2 - \bar\eta > 0$, this can be achieved by choosing $\bar\eta > 1$ smaller if necessary. Then, for any $\epsilon > 0$,

$$
\begin{aligned}
|[III]_N| &\leq \left(\frac{\int_0^T \|(-A)^\alpha P_N F(X)(t)\|^2 \,\mathrm{d}t}{\int_0^T \left\|(-A)^\alpha V_t^{\text{lin},N}(\theta^N)\right\|^2 \mathrm{d}t} \right)^{\frac{1}{2}} \\
&\leq \lambda_N^{\frac{2\alpha - s^* + 2 - \bar\eta + \epsilon}{2}} \left(\frac{\int_0^T \left\|(-A)^{\frac{s^* - 2 + \bar\eta - \epsilon}{2}} P_N F(X)(t)\right\|^2 \mathrm{d}t}{\int_0^T \left\|(-A)^\alpha V_t^{\text{lin},N}(\theta^N)\right\|^2 \mathrm{d}t} \right)^{\frac{1}{2}} \\
&\lesssim N^{-\frac{1}{2} - \frac{\beta}{2}(2\alpha - 2\gamma) + \frac{1}{2} + \frac{\beta}{2}(2\alpha - 2\gamma + 1 - \bar\eta + \epsilon)} = N^{-\frac{\beta}{2}(\bar\eta - 1 - \epsilon)},
\end{aligned}
$$

and for $\bar\eta > 1$ and sufficiently small $\epsilon > 0$, this term converges to zero. If $\bar\eta > 1 + 1/\beta$, then even $N^{1/2}\,|[III]_N| \lesssim N^{-\beta(\bar\eta - 1 - 1/\beta - \epsilon)/2} \to 0$ for $N \to \infty$. Furthermore, the decomposition of $[I]_N$ changes: If we set w.l.o.g. $X_0 = 0$, then we obtain

$$
\begin{aligned}
[I]_N - \mu &= [Ia]_N + [Ib]_N - \mu \frac{\int_0^T \left\langle (-A)^{2\alpha} V_t^{\text{lin},N}(\theta^N), \int_0^t P_N F(X)(t) \mathrm{d}r \right\rangle \mathrm{d}t}{\int_0^T \left\|(-A)^\alpha V_t^{\text{lin},N}(\theta^N)\right\|^2 \mathrm{d}t} \\
&=: [Ia]_N + [Ib]_N + [Id]_N,
\end{aligned}
$$

where again V^{full} has been substituted by V^{lin} in every term. The term $[Id]_N$ is treated exactly as $[III]_N$ (note that, in fact, $JF(X)$ exhibits even larger spatial regularity than $F(X)$). The claims for the case $\bar\eta > 1 + 1/\beta$ are now immediate. For the remaining case $\bar\eta \leq 1 + 1/\beta$, note that by Theorem 3.7, $\hat\theta_N^{\text{lin}} = \theta + o(N^{-a})$ for $a < \beta\bar\eta/2 \wedge \beta$, whereas $|[III]_N| = o(N^{-b})$ for $b < \beta(\bar\eta - 1)/2$. This concludes the proof. □

Remark 3.11.

(i) *Σ_μ is minimal for $\alpha = \gamma$. In this case, $\Sigma_\mu = 4\mu^2/(e^{-2\mu T} - 1 + 2\mu T)$. In particular, $\Sigma_\mu \asymp 2\mu/T$ for $\mu \to \infty$, i.e. for large μ, the asymptotic variance grows linearly in μ. Further, $\lim_{\mu \to 0} \Sigma_\mu = 2/T^2$, i.e. for small μ, the asymptotic variance does not depend on μ, but a large observation time T is even more beneficial.*

(ii) *Similar to the case of diffusivity θ, for finite–dimensional systems, the temporal correlation decay μ is not identifiable in finite time. For example, if $F = 0$, (3.26) describes the true likelihood for the N-dimensional system, and the measures on path space are absolutely continuous for different μ.*

(iii) *If X^N and $P_N F(X)$ are observed, both $\hat{\mu}_N^{\mathrm{full}}\left(\hat{\theta}_N^{\mathrm{sim}}\right)$ and $\hat{\mu}_N^{\mathrm{lin}}\left(\hat{\theta}_N^{\mathrm{lin}}\right)$ are valid estimators. If $P_N F(X)$ is not observed, only $\hat{\mu}_N^{\mathrm{lin}}\left(\hat{\theta}_N^{\mathrm{lin}}\right)$ is feasible.*

(iv) *From the proof of Theorem 3.10, it is clear that $\hat{\mu}_N^{\mathrm{ref}} = \hat{\mu}_N^{\mathrm{full}}(\theta)$ with the true diffusivity θ is asymptotically normal as in (3.42).*

(v) *In particular, the hierarchical approach using $\hat{\mu}_N^{\mathrm{full}}\left(\hat{\theta}_N^{\mathrm{sim}}\right)$ is asymptotically as good as the direct maximum–likelihood approach with known θ whenever $\beta > 1/2$. If $\beta = 2/d$, this means $d \leq 3$.*

Example 3.12. *We close this section with a short discussion on the validity of the additional condition $(F_{s,\eta}^{\mathcal{J}})$.*

(i) *If F satisfies $(F_{s,\eta})$ with excess regularity $\eta > 1$, then $(F_{s,\eta'}^{\mathcal{J}})$ holds with $\eta' = \eta - 1$. In particular, the theory is applicable to reaction–diffusion equations as in Section 2.4.2*

(ii) *If F satisfies $(F_{s,\eta})$ and $JF = FJ$, then $(F_{s,\eta}^{\mathcal{J}})$ holds. For example, if $F(Z) = (-A)^r Z$ for some $r < 2$.*

(iii) *Let $\mathcal{D}$ be a bounded domain with smooth boundary and $A = \Delta$ the Laplacian. Let $s > d/2$. We extend Remark 2.21 as follows: Given a (possibly time-dependent) vector field $v : \mathcal{D} \times [0,T] \to \mathbb{R}^d$ with components $v^{(i)} = Jw^{(i)}$ for some $w^{(i)} \in R(s)$, consider the advection term $F(Z)(t) = \nabla\cdot(Z_t v_t)$. This term belongs to neither of the previous examples (i), (ii). We show that it satisfies $(F_{s,\eta}^{\mathcal{J}})$ for any $\eta < 1$. To this end, we use the integration by parts formula $J(Jf\cdot g) = Jf\cdot Jg - J(f\cdot Jg)$ for $f, g \in R(s)$, where multiplication is understood pointwisely for $x \in \mathcal{D}$.*[2] *Define $\bar{G}^{(i)}(Z) := v^{(i)}\cdot Z - J(w^{(i)}\cdot Z)$ and $G(Z) := \nabla\cdot\bar{G}(Z) = \sum_{i=1}^d \partial_{x_i}\bar{G}^{(i)}(Z)$. Then clearly $JF = GJ$. Further, $\bar{G}^{(i)}$ satisfies $(F_{s,\eta})$ for $\eta < 2$, thus G satisfies $(F_{s,\eta})$ with $\eta < 1$.*

[2]Note that for $f, g \in R(s)$ all terms appearing are well-defined, and for any $x \in \mathcal{D}$, the (multiplicative, bounded) point evaluation operator δ_x reduces the formula to the one-dimensional integration by parts.

3.2 The Case of Integrated Noise

In this section we consider the case that the solution process is driven by an integrated semimartingale. Such a process has pathwise Hölder regularity $3/2 - \epsilon$ in time for every $\epsilon > 0$. Apart from providing a non-standard noise model for semilinar SPDEs that is used in applications, this type of noise can arise from partially observed systems driven by Brownian noise, as explained in Remark 3.16.

A better understanding of the impact of different noise types on statistical questions may help to decide for (or against) them. This noise model provides a simple example of a model misspecification in which the natural estimator $\hat{\theta}_N^{\text{lin}}$, derived under the assumption of martingale noise, is no longer consistent. This is proven in Theorem 3.15.

For a model driven by an integrated noise term $\int_0^t W_r \mathrm{d}r$ rather than W_t, with dispersion operator $B = \sigma(-A)^{-\gamma}$, the resulting equation reads as

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + BW_t\mathrm{d}t \tag{3.49}$$

together with initial condition X_0. W.l.o.g. we assume that the Wiener process starts in zero; different (e.g. random) initial conditions can be absorbed into F. Note that X is the solution to a random PDE of the form

$$\partial_t X_t = \theta A X_t + F(X)(t) + BW_t. \tag{3.50}$$

In particular, $Y_t := \partial_t X$ satisfies

$$\mathrm{d}Y_t = \theta A Y_t \mathrm{d}t + \mathcal{S}F(Y)(t)\mathrm{d}t + B\mathrm{d}W_t \tag{3.51}$$

with initial condition $Y_0 = AX_0 + F(X)(0)$, where

$$\mathcal{S}F := \partial_t \circ F \circ (J + X_0), \tag{3.52}$$

and J is the Bochner integral operator $(JX)(t) = \int_0^t X_r \mathrm{d}r$. We make this precise as follows: Let $s \in \mathbb{R}$. For $f \in C(0,T;H_s)$ such that $\langle f, \Phi_k \rangle \in C^1(0,T;\mathbb{R})$ for $k \geq 1$, let $\partial_t f$ be given by $\langle \partial_t f, \Phi_k \rangle = \partial_t \langle f, \Phi_k \rangle$ whenever it exists. Define $C^1_\Phi(0,T;H_s) \subset C(0,T;H_s)$ to be the subspace of functions f such that $\partial_t f$ exists in the previously explained sense and belongs to $L^\infty(0,T;H_s)$. It is clear that J maps $L^\infty(0,T;H_s)$ into $C^1_\Phi(0,T;H_s)$, while ∂_t maps $C^1_\Phi(0,T;H_s)$ into $L^\infty(0,T;H_s)$. For operators of the form $F : C^1_\Phi(0,T;H_0) \supset D(F) \to C^1_\Phi(0,T;H_0)$, (3.52) is meaningful.

It is clear that the solution X to (3.49) and Y to (3.51) contain the same amount of statistical information, as both processes can be transferred into each other by means of the operators J and ∂_t.

If the nonlinear operator $\mathcal{S}F$ satisfies $(F_{s,\eta})$, we can connect to the theory from Section 2.3. Note that if X_0 is random, $\mathcal{S}F$ will depend on the realization $\omega \in \Omega$, but this does not alter any of the (pathwise) arguments. Condition $(F_{s,\eta})$ for $\mathcal{S}F$ can be deduced naturally from similar conditions on F itself, for example, in the case of reaction terms:

Lemma 3.13. *Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain with smooth boundary, $A = \Delta$ the Laplacian operator and $F(X) = f(X)$ for a differentiable function $f : \mathbb{R} \to \mathbb{R}$. W.l.o.g. let X satisfy Dirichlet boundary conditions. If $X_0 \in H_s$ and f' satisfies $(F_{s,\eta})$ for some $s > d/2$ and any $0 < \eta < 2$, then the same is true for $\mathcal{S}F$.*

Proof. In that case, $\mathcal{S}F(Z) = \partial_t f(JZ + X_0) = f'(JZ + X_0)Z$ for any $Z \in R(s)$. Choose $\epsilon > 0$ and a monotonous function g as in condition $(F_{s,\eta})$ for f'. W.l.o.g. let $s+\eta-2 > d/2$ (otherwise substitute $\eta < 2$ by a larger value) and $\epsilon \leq 2 - \eta$. Then:

$$\begin{aligned}\|\mathcal{S}F(Z)\|_{R(s+\eta+\epsilon-2)} &\lesssim \sup_{0\leq t\leq T} \|f'(JZ_t + X_0)\|_{s+\eta+\epsilon-2} \sup_{0\leq t\leq T} \|Z\|_s \\ &\leq \sup_{0\leq t\leq T} g(\|JZ_t\|_s + \|X_0\|_s) \|Z\|_{R(s)} \\ &\leq g(T\|Z\|_{R(s)} + \|X_0\|_s) \|Z\|_{R(s)},\end{aligned}$$

which proves the claim. □

Furthermore, if ∂_t commutes with F, e.g. if F itself is a linear differential operator acting in spatial direction, then $(F_{s,\eta})$ for $\mathcal{S}F$ immediately reduces to $(F_{s,\eta})$ for F.

In total, the whole theory as developed for noise of semimartingale type transfers if $Y = \partial_t X$ is considered instead of X. For example, a maximum likelihood–type estimator for the case of integrated white noise is given by

$$\begin{aligned}\hat{\theta}_N^{\text{rescaled}} &= -\frac{\int_0^T \langle (-A)^{1+2\alpha}\partial_t X_t^N, \mathrm{d}(\partial_t X^N)_t \rangle}{\int_0^T \|(-A)^{1+\alpha}\partial_t X_t^N\|^2 \,\mathrm{d}t} \\ &\quad + \frac{\int_0^T \langle (-A)^{1+2\alpha}\partial_t X_t^N, P_N(\mathcal{S}F)(\partial_t X)(t) \rangle \,\mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha}\partial_t X_t^N\|^2 \,\mathrm{d}t}.\end{aligned} \tag{3.53}$$

It is obvious that the same reduction technique applies if X is driven by integrated Ornstein–Uhlenbeck noise instead of integrated white noise.

Remark 3.14. *Taking the time derivative of X amounts to a rescaling of the Hölder regularity of X to be $1/2-\epsilon$ in time (for all $\epsilon>0$), such that semimartingale theory can be applied. A similar approach is possible for fractional noise with Hurst index $0<H<1$. In this case, the temporal regularity rescaling can be done by applying a kernel instead of taking the derivative. Based on that observation, a Girsanov transform for SODEs driven by fractional Brownian motion B^H can be derived by considering a surrogate semimartingale [NVV99, KLBR00, TV07, Mis08]. This allows for likelihood-based inference. In [CLP09, Cia10] this approach is used for parameter estimation for SPDEs driven by additive and multiplicative fractional noise.*

In the case of integrated noise, it is interesting to see how model misspecification changes the behavior of the estimator. Namely, assume that $\hat{\theta}_N^{\mathrm{full}}$ is given as in Section 2.3, but the dynamics of X is generated by integrated noise. Then even in the simplest possible case, i.e. if X satisfies a linear equation with $X_0=0$, $\hat{\theta}_N^{\mathrm{full}}$ is not consistent:

Theorem 3.15. *Let $X_0=0$, $F=0$. It holds that $\hat{\theta}_N^{\mathrm{full}}\to 0$ almost surely.*

Proof. First note that

$$\hat{\theta}_N^{\mathrm{full}}-\theta=-\sigma\frac{\int_0^T\left\langle(-A)^{1+2\alpha-\gamma}X_t^N,W_t^N\right\rangle\mathrm{d}t}{\int_0^T\left\|(-A)^{1+\alpha}X_t^N\right\|^2\mathrm{d}t}. \tag{3.54}$$

With $x^{(k)}=\langle X,\Phi_k\rangle$, it holds that

$$\mathrm{d}x_t^{(k)}=-\theta\lambda_k x_t^{(k)}\mathrm{d}t+\sigma\lambda_k^{-\gamma}W_t^{(k)}\mathrm{d}t$$

with independent Wiener processes $W^{(k)}$, and consequently,

$$x_t^{(k)}=\sigma\lambda_k^{-\gamma}\int_0^t e^{-\theta\lambda_k(t-r)}W_r^{(k)}\mathrm{d}r.$$

A straightforward calculation yields

$$\begin{aligned}\mathbb{E}[(x_t^{(k)})^2]&=2\sigma^2\lambda_k^{-2\gamma}\int_0^t\int_0^r e^{-\theta\lambda_k(t-r)}e^{-\theta\lambda_k(t-r')}r'\mathrm{d}r'\mathrm{d}r\\&=\frac{2\sigma^2}{\theta^2\lambda_k^{2\gamma+2}}\left(\frac{t}{2}-\frac{3}{4\theta\lambda_k}-\frac{e^{-2\theta\lambda_k t}}{4\theta\lambda_k}+\frac{e^{-\theta\lambda_k t}}{\theta\lambda_k}\right),\end{aligned}$$

thus

$$\mathbb{E}\int_0^T (x_t^{(k)})^2 \mathrm{d}t \asymp \frac{T^2\sigma^2}{2\theta^2}\lambda_k^{-2\gamma-2}.$$

Summing over the first N modes and using Lemma A.2 (ii), we get a.s.

$$\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|^2 \mathrm{d}t \asymp \frac{T^2\sigma^2\Lambda^{2\alpha-2\gamma}}{2\theta^2(1+\beta(2\alpha-2\gamma))} N^{1+\beta(2\alpha-2\gamma)}. \tag{3.55}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}[x_t^{(k)} W_t^{(k)}] &= \sigma\lambda_k^{-\gamma}\int_0^t e^{-\theta\lambda_k(t-r)} r \mathrm{d}r \\
&= \frac{\sigma}{\theta\lambda_k^{\gamma+1}}\left(t - \frac{1}{\theta\lambda_k}(1-e^{-\theta\lambda_k t})\right),
\end{aligned}$$

and consequently,

$$\mathbb{E}\int_0^T x_t^{(k)} W_t^{(k)} \mathrm{d}t \asymp \frac{T^2\sigma}{2\theta}\lambda_k^{-\gamma-1}.$$

By summing up, we obtain

$$\mathbb{E}\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, W_t^N \right\rangle \mathrm{d}t \asymp \frac{T^2\sigma\Lambda^{2\alpha-2\gamma}}{2\theta(1+\beta(2\alpha-2\gamma))} N^{1+\beta(2\alpha-2\gamma)}. \tag{3.56}$$

Finally, using the Wick theorem as in [Jan97, Theorem 1.28],

$$\begin{aligned}
&\mathrm{Var}\left[\int_0^T x_t^{(k)} W_t^{(k)} \mathrm{d}t\right] \\
&\qquad = \int_0^T\int_0^T \mathbb{E}[x_t^{(k)} x_r^{(k)} W_t^{(k)} W_r^{(k)}] \mathrm{d}r\mathrm{d}t - \left(\int_0^T \mathbb{E}[x_t^{(k)} W_t^{(k)}]\mathrm{d}t\right)^2 \\
&\qquad = \int_0^T\int_0^T \mathbb{E}[x_t^{(k)} x_r^{(k)}]\mathbb{E}[W_t^{(k)} W_r^{(k)}] + \mathbb{E}[x_t^{(k)} W_r^{(k)}]\mathbb{E}[x_r^{(k)} W_t^{(k)}] \mathrm{d}r\mathrm{d}t \\
&\qquad \leq 2\int_0^T\int_0^T \sqrt{rt\mathbb{E}[(x_r^{(k)})^2]\mathbb{E}[(x_t^{(k)})^2]} \mathrm{d}r\mathrm{d}t \\
&\qquad \leq 2T^2\mathbb{E}\int_0^T (x_t^{(k)})^2 \mathrm{d}t \lesssim \lambda_k^{-2\gamma-2},
\end{aligned}$$

and in particular,

$$\sum_{N=1}^{\infty} \frac{\mathrm{Var}\left[\lambda_k^{1+2\alpha-\gamma} \int_0^T x_t^{(N)} W_t^{(N)} \mathrm{d}t\right]}{\left(\mathbb{E} \int_0^T \langle (-A)^{1+2\alpha-\gamma} X_t^N, W_t^N \rangle \, \mathrm{d}t\right)^2} \lesssim \sum_{N=1}^{\infty} \frac{N^{\beta(4\alpha-4\gamma)}}{N^{2+\beta(4\alpha-4\gamma)}} < \infty,$$

such that by the strong law of large numbers [Shi96, Theorem IV.3.2], (3.56) holds a.s. for $\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, W_t^N \right\rangle \mathrm{d}t$. Now, from (3.54), we see that

$$\hat{\theta}_N^{\text{full}} - \theta \xrightarrow{a.s.} -\theta,$$

which implies the claim. □

Remark 3.16. *Integrated noise appears naturally if one considers systems such that the first component is observed, but only the second component is driven by noise. More precisely, the linear system*

$$\mathrm{d}X_t^O = \theta A_{11} X_t^O \mathrm{d}t + A_{12} X_t^U \mathrm{d}t, \tag{3.57}$$
$$\mathrm{d}X_t^U = A_{21} X_t^O \mathrm{d}t + A_{22} X_t^U \mathrm{d}t + B_2 \mathrm{d}W_t, \tag{3.58}$$

with $X_0^O = 0$, $X_0^U = 0$ *and unknown* θ, *can be formally rewritten as*

$$\mathrm{d}X_t^O = \theta A X_t^O \mathrm{d}t + F(X^O)(t)\mathrm{d}t + B W_t \mathrm{d}t, \tag{3.59}$$

where $A = A_{11}$, $B = A_{12}B_2$ *and* $F(X) = A_{12}A_{22}A_{12}^{-1}X + A_{12}A_{21}JX - \theta A_{12}A_{22}A_{12}^{-1}A_{11}JX$. *Depending on the form of* $A_{11}, A_{12}, A_{21}, A_{22}$ *and* B_2, *this reasoning can be made rigorous. In order to neglect* F, *the regularity of all terms appearing in the (linear) system* (3.57), (3.58) *can be evaluated directly, or it can be shown that* $\mathcal{S}F = F$ *satisfies* $(F_{s,\eta})$. *If either of these approaches is feasible, the reduction to the theory from Section 2.3 as described above is applicable.*

The extension of this setting to semilinear systems is possible by a regularity argument as in the previous sections, decomposing both components into their linearization and nonlinear remainder.

3.3 Structure of the Dispersion Operator

Set $\bar{B} = \sigma(-A)^{-\gamma}$. We call $\bar{B}$ the *reference dispersion operator*. In all models we have considered so far, we used $\bar{B}$ as dispersion operator. Now we study to which extent this assumption can be relaxed. W.l.o.g. we study only the white noise case. More precisely, consider

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X_t)\mathrm{d}t + B(X_t)\mathrm{d}W_t \tag{3.60}$$

with initial condition X_0. Let $L_2(H)$ denote the space of Hilbert–Schmidt operators on H, with norm $\|\cdot\|_{\mathrm{HS}}$. We demand that B maps its domain $D(B) \subseteq H$ into $L_2(H)$. In direct analogy to (W), our standing assumption is well-posedness of (3.60) in the sense of a unique probabilistically and analytically weak solution in $C(0,T;H)$. Within the variational approach, as exposed in [LR15], this can be shown under Lipschitz and growth conditions on B (and additional mild conditions on F). Here and in the sequel, we write $\widetilde{B}(Z) = B(Z) - \bar{B}$ for the deviation from the reference dispersion operator, i.e. we consider lower order (possibly multiplicative) noise of the form

$$B(Z) = \sigma(-A)^{-\gamma} + \widetilde{B}(Z). \tag{3.61}$$

In order to transfer the results from the reference case, B must be asymptotically similar to $\bar{B}$ in the following sense:

(N_η^γ) There is a locally bounded $b : [0,\infty) \to [0,\infty)$ such that for $Z \in H$ and $k \in \mathbb{N}$:

$$\left\|\widetilde{B}(Z)^T \Phi_k\right\|_H^2 \ll_p b(\|Z\|_H)\lambda_k^{-1-2\gamma-\eta}. \tag{3.62}$$

This is a natural condition, as shown in the next lemma:

Lemma 3.17. *Let $\eta > 0$. If for $Z \in H$, $\widetilde{B}(Z)$ is a linear bounded operator mapping H into H_r for some $r > 1 + 2\gamma + \eta$ such that the operator norm satisfies*

$$\left\|\widetilde{B}(Z)\right\|_{H\to H_r}^2 \leq b(\|Z\|_H), \tag{3.63}$$

then condition (N_η^γ) is satisfied.

Proof. In that case,

$$\begin{aligned}\left\|\widetilde{B}(Z)^T \Phi_k\right\|_H^2 &\lesssim \left\|\widetilde{B}(Z)^T (-A)^{r/2}\right\|_{H\to H}^2 \left\|(-A)^{-r/2}\Phi_k\right\|_H^2 \\ &\lesssim \left\|(-A)^{r/2}\widetilde{B}(Z)\right\|_{H\to H}^2 \lambda_k^{-r}.\end{aligned}$$

Now $(-A)^{r/2} : H_r \to H$ is an isometry, and therefore

$$\left\|\widetilde{B}(Z)^T\Phi_k\right\|_H^2 \lesssim \left\|\widetilde{B}(Z)\right\|_{H\to H_r}^2 \lambda_k^{-r} \lesssim b(\|Z\|_H)\lambda_k^{-r},$$

which implies the claim. $\square$

Example 3.18. *Diagonal noise of the form $B(X)\Phi_k = b_k(X)\Phi_k$ for functions $b_k : H \to \mathbb{R}$, $k \in \mathbb{N}$. Such diagonal dispersion terms have been considered e.g. in [CCG20], or in [CKL20] in the context of space-only noise. Here, condition (N_η^γ) simplifies to*

$$\left|b_k(Z)/\lambda_k^{-\gamma} - \sigma\right|^2 \ll_p b(\|Z\|_H)\lambda_k^{-1-\eta}, \tag{3.64}$$

which amounts to fast asymptotic equivalence of the modes b_k and $\lambda_k^{-\gamma}$.

As before, let $\bar{X}$ be the solution to

$$\mathrm{d}\bar{X}_t = \theta A\bar{X}_t\mathrm{d}t + \bar{B}\mathrm{d}W_t \tag{3.65}$$

with $\bar{X}_0 = 0$, and $\widetilde{X} := X - \bar{X}$. In order to control the regularity of $\widetilde{X}$, we extend the splitting argument as follows: Define $\bar{X}^F$ to be the solution of

$$\mathrm{d}\bar{X}_t^F = \theta A\bar{X}_t^F\mathrm{d}t + B(X_t)\mathrm{d}W_t \tag{3.66}$$

with $\bar{X}_0^F = 0$, such that $\widetilde{X}^F := X - \bar{X}^F$ satisfies

$$\mathrm{d}\widetilde{X}_t^F = \theta A\widetilde{X}_t^F\mathrm{d}t + F(X_t)\mathrm{d}t \tag{3.67}$$

with $\widetilde{X}_0^F = 0$. It follows from Proposition 3.19 below that $\bar{X}^F$ is well-posed. Next, with $\bar{X}^B := \bar{X}$, the process $\widetilde{X}^B := \bar{X}^F - \bar{X}^B$ satisfies

$$\mathrm{d}\widetilde{X}_t^B = \theta A\widetilde{X}_t^B\mathrm{d}t + \widetilde{B}(X_t)\mathrm{d}W_t, \tag{3.68}$$

$\widetilde{X}_0^B = 0$. This means that the nonlinear process $\widetilde{X} = \widetilde{X}^F + \widetilde{X}^B$ consists of two components, which contain the nonlinear behavior in the drift and the dispersion, respectively.

As before, we write $s^* = 1 + 2\gamma - 1/\beta$.

Proposition 3.19. *Let $\gamma > 1/(2\beta)$ and $\eta > 0$. Under condition (N_η^γ), we have $\widetilde{X}^B \in R(s+\eta)$ for any $s < s^*$.*

Proof. First, note that for any $Z \in H$, the operator $(-A)^{(s+\eta)/2}\widetilde{B}(Z)$ is a Hilbert–Schmidt operator on H:

$$\begin{aligned}\left\|(-A)^{(s+\eta)/2}\widetilde{B}(Z)\right\|_{\mathrm{HS}}^2 &= \sum_{k=1}^{\infty}\left\|\widetilde{B}(Z)^T(-A)^{(s+\eta)/2}\Phi_k\right\|_H^2 \\ &= \sum_{k=1}^{\infty}\lambda_k^{s+\eta}\left\|\widetilde{B}(Z)^T\Phi_k\right\|_H^2 \\ &\lesssim b(\|Z\|_H)\sum_{k=1}^{\infty}\lambda_k^{s+\eta-1-2\gamma-\eta-\epsilon} \lesssim \sum_{k=1}^{\infty}\lambda_k^{-1/\beta-\epsilon} < \infty\end{aligned}$$

for some $\epsilon > 0$ due to condition (N_η^γ). In particular, using $X \in C(0,T;H)$ a.s., we see that $\widetilde{B}_t^* := (-A)^{(s+\eta)/2}\widetilde{B}(X_t)$ is uniformly bounded in $\|\cdot\|_{\mathrm{HS}}$ for $t \in [0,T]$. Now, Theorem 4.2.4 of [LR15] implies that

$$\mathrm{d}Y_t = \theta AY_t\mathrm{d}t + \widetilde{B}_t^*\mathrm{d}W_t, \tag{3.69}$$

$Y_0 = 0$, has a unique solution in $C(0,T;H)$. As a consequence, $Y = (-A)^{(s+\eta)/2}\widetilde{X}^B$, and the claim follows. □

Together with Proposition 2.3 (with $\bar{X}$ replaced by $\bar{X}^F = \bar{X} + \widetilde{X}^B$ therein), we immediately obtain:

Theorem 3.20. *Let $\gamma > 1/(2\beta)$, $s \in \mathbb{R}$ and $\eta > 0$. Assume $X_0 \in H_{s+\eta}$. If $(F_{s,\eta})$ and (N_η^γ) are true and $X \in R(s)$, then $\widetilde{X} = \widetilde{X}^F + \widetilde{X}^B \in R(s+\eta)$ almost surely.* □

Remark 3.21. *If in the proof of Proposition 3.19 it can be shown that $Y = (-A)^{(s+\eta)/2}\widetilde{X}^B$ has continuous trajectories not only in H but also in V, then we can conclude even $\widetilde{X}^B \in R(s+\eta+1)$. In view of Lemma 3.17, this seems natural: There, $\widetilde{B}(Z)$ must map H into H_r for $r > 1+2\gamma+\eta$, whereas $\bar{B}$ maps H into $H_{2\gamma}$. In this sense, $1+\eta$ should be expected to be the "true" excess regularity instead of η. However, in the general setting it is not clear if Y has continuous trajectories in V, although there are sufficient criteria known in literature. For example, in the case of additive noise $B(Z) \equiv B$, according to Theorem 5.11 from [DPZ14] we have that $Y \in C(0,T;V)$ if the integral $\int_0^T t^{-2\alpha}\left\|e^{t\theta A}(-A)^{(s+\eta+1)/2}(B-\bar{B})\right\|_{\mathrm{HS}}^2\mathrm{d}t$ is finite for some $0 < \alpha < 1/2$.*

In particular, if $\hat\theta_N^{\text{full}}$, $\hat\theta_N^{\text{part}}$ and $\hat\theta_N^{\text{lin}}$ are given by (2.24), (2.25) and (2.26), the results on diffusivity estimation from Theorem 2.11 transfer directly to the model studied in this section:

Theorem 3.22. *Let $\gamma > 1/(2\beta)$, $\eta > 0 \vee (1/\beta - 1)$ and $s_0 \geq 0$, assume $X_0 \in H_{s^*+\eta}$ and $X \in R(s_0)$. Let (N_η^γ) and $(F_{s,\eta})$ be true for $s_0 \leq s < s^*$. Let $\alpha > \gamma - 1/4$. Then the following asymptotic statements are true:*

(i) $\hat\theta_N^{\text{full}}$ is asymptotically normal as in (2.29), and if $\eta > 1 + 1/\beta$, the same is true for $\hat\theta_N^{\text{part}}$, $\hat\theta_N^{\text{lin}}$.

(ii) In the case $\eta \leq 1 + 1/\beta$, $\hat\theta_N^{\text{part}}$ and $\hat\theta_N^{\text{lin}}$ are consistent in probability with convergence rate N^{-a}, $a < \beta\eta/2$, i.e.

$$N^a\left(\hat\theta_N^{\text{part}} - \theta\right) \xrightarrow{\mathbb{P}} 0, \qquad N^a\left(\hat\theta_N^{\text{lin}} - \theta\right) \xrightarrow{\mathbb{P}} 0. \tag{3.70}$$

Proof. By Theorem 3.20, $\int_0^T \left\|(-A)^{s/2}X_t^N\right\|_H^2 \mathrm{d}t$ satisfies the asymptotics from (2.20) whenever $s > s^*$. Now, by means of (N_η^γ) and $X \in C(0,T;H)$,

$$\begin{aligned}
\int_0^T \left\|\widetilde{B}(X_t)^T(-A)^{1+2\alpha}X_t^N\right\|_H^2 \mathrm{d}t &\leq \int_0^T \left(\sum_{k=1}^N \lambda_k^{1+2\alpha}\left|x_t^{(k)}\right|\left\|\widetilde{B}(X_t)^T\Phi_k\right\|_H\right)^2 \mathrm{d}t \\
&\ll_p \sup_{0\leq t\leq T} b\left(\|X_t\|_H\right)\int_0^T\left(\sum_{k=1}^N \lambda_k^{1/2+2\alpha-\gamma-\eta/2}\left|x_t^{(k)}\right|\right)^2 \mathrm{d}t \\
&\lesssim N\sum_{k=1}^N \lambda_k^{1+4\alpha-2\gamma-\eta}\int_0^T (x_t^{(k)})^2 \mathrm{d}t \\
&= N\int_0^T \left\|(-A)^{\frac12+2\alpha-\gamma-\frac{\eta}{2}}X_t^N\right\|_H^2 \mathrm{d}t.
\end{aligned} \tag{3.71}$$

In case $\alpha > \gamma + (\eta - 1/\beta)/4$ we have $1 + 4\alpha - 2\gamma - \eta > s^*$, and the latter term is dominated by $N^{2+\beta(4\alpha-4\gamma-\eta)}$. On the other hand, if $\alpha < \gamma + (\eta - 1/\beta)/4$, the last integral converges, and the latter term is dominated by N. The case $\alpha = \gamma + (\eta - 1/\beta)/4$ can be ignored by substituting $\eta \mapsto \eta - \epsilon$ for some small $\epsilon > 0$. In any of these cases, the right-hand side is negligible compared to $N^{1+\beta(1+4\alpha-4\gamma)}$, where we take into account $\eta > 1/\beta - 1$ and $\alpha > \gamma - 1/4$.

In particular, using $B(X_t)=\sigma(-A)^{-\gamma}+\widetilde{B}(X_t)$ and expanding the squared norm, we have a.s.

$$\begin{aligned}\int_0^T \left\|B(X_t)^T(-A)^{1+2\alpha}X_t^N\right\|_H^2 \mathrm{d}t &\asymp \sigma^2\int_0^T \left\|(-A)^{1+2\alpha-\gamma}X_t^N\right\|_H^2 \mathrm{d}t \\ &\asymp \sigma^2 C_{2+4\alpha-2\gamma}N^{1+\beta(1+4\alpha-4\gamma)}\end{aligned}$$

because the condition $\eta > 1/\beta-1$ ensures that $2+\beta(4\alpha-4\gamma-\eta) < 1+\beta(1+4\alpha-4\gamma)$, i.e. the remaining terms are of lower order. Consequently, the local martingale

$$M_T^N := C_{2+4\alpha-2\gamma}^{-1/2}\sigma^{-1}N^{-1/2-\beta(1+4\alpha-4\gamma)/2}\int_0^T \left\langle B(X_t)^T(-A)^{1+2\alpha}X_t^N, \mathrm{d}W_t\right\rangle$$

is such that $\langle M^N\rangle_T \to 1$ a.s., and according to Theorem A.1 and the Slutsky lemma,

$$\begin{aligned}N^{\frac{1+\beta}{2}}(\hat{\theta}_N^{\mathrm{full}}-\theta) &= -N^{\frac{1+\beta}{2}}\frac{\int_0^T \left\langle B(X_t)^T(-A)^{1+2\alpha}X_t^N, \mathrm{d}W_t^N\right\rangle}{\int_0^T \|(-A)^{1+\alpha}X_t^N\|_H^2 \mathrm{d}t} \\ &= -\sigma\frac{C_{2+4\alpha-2\gamma}^{1/2}N^{1+\beta(1+2\alpha-2\gamma)}}{\int_0^T \|(-A)^{1+\alpha}X_t^N\|_H^2 \mathrm{d}t}M_T^N\end{aligned}$$

converges to a normal distribution with mean zero and variance as given by (2.30) The remaining claims for $\hat{\theta}_N^{\mathrm{part}}$ and $\hat{\theta}_N^{\mathrm{lin}}$ follow verbatim as in Theorem 2.11 (note that the condition on α in this theorem are even more restrictive than in Theorem 2.11). □

Remark 3.23.

(i) *In comparison with Theorem 2.11, there are two additional restrictions: The excess regularity η must exceed $1/\beta-1$, and further $\alpha>\gamma-1/4$, which is always stronger than the condition on α from Theorem 2.11. Both inequalities are related to the control of the non–diagonal elements in the noise term. In the setting of Example 3.18, both restrictions can be avoided: Namely, in that case,* (3.71) *in the proof of Theorem 3.22* *is substituted by*

$$\int_0^T \left\| \widetilde{B}(X_t)^T(-A)^{1+2\alpha} X_t^N \right\|_H^2 \mathrm{d}t$$
$$= \sum_{k=1}^N \lambda_k^{2+4\alpha} \int_0^T \left(b_k(X_t) - \lambda_k^{-\gamma}\right)^2 (x_t^{(k)})^2 \mathrm{d}t$$
$$\ll_p \int_0^T \left\|(-A)^{1/2+2\alpha-\gamma-\eta/2} X_t^N\right\|_H^2 \mathrm{d}t,$$

which is always dominated by $N^{1+\beta(1+4\alpha-4\gamma)}$ *if* $\eta > 0$ *and* $\alpha > \gamma - (1 + 1/\beta)/4$*, and the rest of the proof is identical. Note that if* $A = \Delta$ *is the Laplacian on a bounded domain, then* $\beta = 2/d$*, and the additional condition* $\eta > 1/\beta - 1$ *is void in dimension* $d \leq 2$*.*

(ii) If $\widetilde{B}(X)^T$ *maps* H_s *into* $H_{s+2\gamma}$ *for all* $s \in \mathbb{R}$ *with* $\|\widetilde{B}(X)^T\|_{H_s \to H_{s+2\gamma}} \leq b_s(\|X\|_H)$ *for some locally bounded* b_s*, then it is straightforward to see that for* $s \in \mathbb{R}$*,* $(-A)^{1+2\alpha-\gamma}X \in R(s)$ *implies* $B(X)^T(-A)^{1+2\alpha}X \in R(s)$*. In this case,* (3.70) *can be strengthened to almost sure convergence using Lemma 2.6, as in the proof of Theorem 2.11.*

Chapter 4

Discretization of the Spectral Approach

In this section we adapt the spectral approach to the case that the observations consist of a set of point evaluations of the process X in space instead of Fourier modes. It is determined how much spatial information is needed in order to reconstruct the spectral asymptotics from Theorem 2.11, depending on the regularity of the process.

By now, there is plenty of literature on statistical inference for SPDEs based on spatially and/or temporally discretized observations. Various works are based on the asymptotic analysis of power variations, either in time [BT19, BT20, Cho20, Cho19, CD20, KU21a, KU21b], in space [CKL20, CK22, SST20, CKP21], in time and space [PT07, CH20, MKT19a], or extending the approach to a combined spatiotemporal variation [HT21b, HT21a]. Within the spectral approach, however, there seems to be almost no rigorous attempt to quantify the amount of spatial information needed to recover its asymptotics for diffusivity estimation. We are aware only of [Hue93, p. 44ff.], where this topic is sketched shortly (but without rigorous proof) in $d = 1$, using first-order integral approximations. On the other hand, [CDVK20] considers the discretization in time of the maximum likelihood estimator from the spectral approach.

As in the previous sections, we consider a semilinear SPDE of the form

$$\mathrm{d}X_t = \theta A X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + B\mathrm{d}W_t \tag{4.1}$$

with initial condition X_0, where A is a closed, densely defined, negative definite and self-adjoint operator with compact resolvent, $B = \sigma(-A)^{-\gamma}$ is of

Hilbert–Schmidt type, and F satisfies $(F_{s,\eta})$ for some $\eta > 0$ and $s_0 \leq s < s^*$. W.l.o.g. we set $\sigma = 1$. We always assume that (4.1) is well-posed with $X \in R(s)$ for $s < s^*$. As we are interested in spatially discrete point evaluations, we have to make the abstract setting from Chapter 2 more specific: Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain with smooth boundary, such that the state space for X is given by $H = L^2(\mathcal{D})$. For simplicity, we assume that X satisfies Dirichlet boundary conditions. The eigenvalues $(\lambda_k)_{k\in\mathbb{N}}$ of A are assumed to satisfy

$$\lambda_k \asymp \Lambda k^{\frac{D}{d}} \tag{4.2}$$

for some $D > 0$, called the order of the operator A. It is well-known that (4.2) is true if A is a (pseudo-) differential operator of order D [Shu01]. For example, for $A = \Delta$ we have $D = 2$ and for $-\Delta^2$ we have $D = 4$, cf. Section 2.4. Note that B is a Hilbert–Schmidt operator if and only if $\gamma > d/(2D)$.

For any $h > 0$, let $M_h \in \mathbb{N}$ and $(x_i^{(h)})_{i=1,\dots,M_h} \subset \mathcal{D}$. Define the evaluation operator $E_h : C(\mathcal{D}) \to \mathbb{R}^{M_h}$ via $(E_h f)_i := f(x_i^{(h)})$. Then each component of E_h is a bounded multiplicative linear form on $C(\mathcal{D})$. We write $\langle\cdot,\cdot\rangle_{(h)}$ for the Euclidean scalar product on $\mathbb{R}^{M_h}$.

In order to apply E_h to (4.1), we need AX to have values in $C(\mathcal{D})$. Therefore, we make the standing assumption

$$s^* > 2, \quad AX \in L^\infty(0,T;C(\mathcal{D})). \tag{4.3}$$

For example, if $A = \Delta$, the latter condition holds if $s^* > d/2 + 2$, i.e. $\gamma > 1/2 + d/2$, by means of the Sobolev embedding theorem. However, in many situations it is not necessary to use the Sobolev embedding theorem in order to prove continuity in space. For example, if the eigenfunctions Φ_k are uniformly bounded in $x \in \mathcal{D}$, $k \in \mathbb{N}$, then $s^* > 2$ already implies $AX \in L^\infty(0,T;C(\mathcal{D}))$, see Lemma 5.5 and Proposition 5.6 below. We note that for general A, $s^* > 2$ if and only if $\gamma > d/(2D) + 1/2$.

Similarly, we always assume that the terms $E_hF(X)$ and E_hBW are well-defined. If $A = \Delta$, the former is true e.g. if F is of the form $F(X) = f(X)$ for a function $f : \mathbb{R} \to \mathbb{R}$ by (4.3), and the latter can be enforced e.g. by imposing the additional bound $\gamma > d/2$, such that $BW \in L^2(0,T;W^{d/2,2}(\mathcal{D}))$, and spatial point evaluations are well-defined again by means of the Sobolev embedding theorem. In this situation, E_hX satisfies

$$\mathrm{d}E_hX_t = \theta E_hAX_t\mathrm{d}t + E_hF(X)(t)\mathrm{d}t + E_hB\mathrm{d}W_t. \tag{4.4}$$

Let $r^* > r_* \geq 0$ such that $r^* \geq s^* D/2$ and $r_* < (s^* - 1)D/2$. Fix a Banach space $B_r \subset H$ for each $r_* < r \leq r^*$, such that $B_{r_2} \subset B_{r_1}$ for $r_1 < r_2$. Further, let $h^* > 0$. We need the following conditions:

(D_0) For $r_* < r \leq r^*$, B_r is a Banach algebra, and $B_r \subseteq C(\mathcal{D})$.

(D_1) For any $r_* < r \leq r^*$:

$$\|\Phi_k\|_{B_r} \lesssim k^{r/d}. \tag{4.5}$$

(D_2) For any $0 < h < h^*$, there are real numbers $w_1^{(h)}, \ldots, w_{M_h}^{(h)}$ such that for any $r_* < r \leq r^*$ and $Z \in B_r$,

$$\left| \int_{\mathcal{D}} Z \mathrm{d}x - \sum_{i=1}^{M_h} w_i^{(h)} (E_h Z)_i \right| \lesssim h^r \|Z\|_{B_r}. \tag{4.6}$$

Condition (D_2) relies on higher order quadrature formulas, reflecting the regularity of X. In the examples, the scale $(B_r)_{r_* < r \leq r^*}$ will consist of Hölder spaces or L^2-based Sobolev spaces. Since these spaces are supposed to measure the regularity of X, we need

$$X \in L^\infty(0, T; B_r) \text{ for } r < s^* D/2. \tag{4.7}$$

This is immediate if B_r coincides with $H_{2r/D}$, otherwise it has to be proven separately. In all our examples, this will be valid.

Remark 4.1. *We emphasize that the index of the regularity space H_s from Chapter 2 does not count spatial derivatives, but fractional powers of $(-A)$. If $D \neq 2$, this is not the same. This is why the factor $D/2$ arises e.g. in (D_0) and related relations below.*

Denote by $\mathcal{W}_h \in \mathbb{R}^{M_h \times M_h}$ the diagonal matrix with entries $(w_i^{(h)})_{1 \leq i \leq M_h}$ and set

$$\mathcal{E}_h = \mathcal{W}_h E_h. \tag{4.8}$$

Note that for $Z_1, Z_2 \in C(\mathcal{D})$, it holds $E_h(Z_1 Z_2) = E_h Z_1 E_h Z_2$ in the sense of componentwise multiplication, and therefore:

$$\langle \mathcal{W}_h E_h Z_1, E_h Z_2 \rangle_{(h)} = \sum_{i=1}^{M_h} w_i^{(h)} E_h(Z_1 Z_2). \tag{4.9}$$

With that notation, a direct consequence of (D_0), (D_1) and (D_2) is

$$\left|\langle \Phi_k, Z\rangle - \langle \mathcal{W}_h E_h \Phi_k, E_h Z\rangle_{(h)}\right| \lesssim h^r k^{r/d} \left\|Z\right\|_{B_r}. \tag{4.10}$$

We use the following discretized version of $-A$:

$$A_{h,N}^{(s/2)} := \sum_{k=1}^{N} \lambda_k^{s/2} (E_h \Phi_k)(E_h \Phi_k)^T. \tag{4.11}$$

Based on these considerations, we want to adapt the maximum-likelihood based estimator $\hat{\theta}_N^{\mathrm{lin}}$ to the case of spatially discrete observations. Its natural discrete analogue $\hat{\theta}_{h,N}^{\mathrm{discr}}$ is given in the present setting as follows:

$$\hat{\theta}_{h,N}^{\mathrm{discr}} = -\frac{\int_0^T \left\langle A_{h,N}^{(1+2\alpha)}(\mathcal{E}_h X_t), \mathrm{d}(\mathcal{E}_h X_t)\right\rangle_{(h)}}{\int_0^T \left\langle A_{h,N}^{(2+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h X_t\right\rangle_{(h)} \mathrm{d}t}. \tag{4.12}$$

The error decomposition of $\hat{\theta}_{h,N}^{\mathrm{discr}}$ reads as:

$$\begin{aligned}
\hat{\theta}_{h,N}^{\mathrm{discr}} &= -\frac{\theta\int_0^T \left\langle A_{h,N}^{(1+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h A X_t\right\rangle_{(h)} \mathrm{d}t}{\int_0^T \left\langle A_{h,N}^{(2+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h X_t\right\rangle_{(h)} \mathrm{d}t} \\
&\quad - \frac{\int_0^T \left\langle A_{h,N}^{(1+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h F(X)(t)\right\rangle_{(h)} \mathrm{d}t}{\int_0^T \left\langle A_{h,N}^{(2+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h X_t\right\rangle_{(h)} \mathrm{d}t} - \frac{\int_0^T \left\langle A_{h,N}^{(1+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h B \mathrm{d}W_t\right\rangle}{\int_0^T \left\langle A_{h,N}^{(2+2\alpha)}\mathcal{E}_h X_t, \mathcal{E}_h X_t\right\rangle_{(h)} \mathrm{d}t} \\
&= \theta + \theta\frac{\left(I_{h,N}^{(2)}(2+2\alpha) - I_N(2+2\alpha)\right) + \left(I_N(2+2\alpha) - I_{h,N}^{(1)}(2+2\alpha)\right)}{I_{h,N}^{(1)}(2+2\alpha)} \\
&\quad - \frac{\left(F_{h,N}^{(1+2\alpha)} - F_N^{(1+2\alpha)}\right) + F_N^{(1+2\alpha)}}{I_{h,N}^{(1)}(2+2\alpha)} - \frac{\sqrt{\langle M^{(h,N)}\rangle_T}}{I_{h,N}^{(1)}(2+2\alpha)} \frac{M_T^{(h,N)}}{\sqrt{\langle M^{(h,N)}\rangle_T}},
\end{aligned} \tag{4.13}$$

where we abbreviate

$$I_N(s) = \int_0^T \left\|(-A)^{s/2} X_t^N\right\|^2 \mathrm{d}t, \tag{4.14}$$

$$I_{h,N}^{(1)}(s) = \int_0^T \left\langle A_{h,N}^{(s)} \mathcal{E}_h X_t, \mathcal{E}_h X_t \right\rangle_{(h)} \mathrm{d}t, \tag{4.15}$$

$$I_{h,N}^{(2)}(s) = \int_0^T \left\langle A_{h,N}^{(s-1)} \mathcal{E}_h X_t, \mathcal{E}_h (-A) X_t \right\rangle_{(h)} \mathrm{d}t, \tag{4.16}$$

$$F_N^{(1+2\alpha)} = \int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F(X)(t) \right\rangle \mathrm{d}t, \tag{4.17}$$

$$F_{h,N}^{(1+2\alpha)} = \int_0^T \left\langle A_{h,N}^{(1+2\alpha)} \mathcal{E}_h X_t, \mathcal{E}_h F(X)(t) \right\rangle_{(h)} \mathrm{d}t, \tag{4.18}$$

$$M_T^{(h,N)} = \int_0^T \left\langle A_{h,N}^{(1+2\alpha)} \mathcal{E}_h X_t, \mathcal{E}_h (-A)^{-\gamma} \mathrm{d}W_t \right\rangle, \tag{4.19}$$

and $(M_t^{(h,N)})_{t\geq 0}$ is a local martingale with quadratic variation

$$\langle M^{(h,N)} \rangle_T = \int_0^T \left\| (-A)^{-\gamma} (\mathcal{E}_h)^* A_{h,N}^{(1+2\alpha)} \mathcal{E}_h X_t \right\|_H^2 \mathrm{d}t. \tag{4.20}$$

Proposition 4.2. *Assume that* (D_0), (D_1), (D_2) *hold. Let* $R \geq 0$ *and* $s > s^*$.

(i) If we have

$$h \ll_p N^{-\frac{2}{d} K_{d,D,R}^{(1)}(\gamma)}, \quad K_{d,D,R}^{(1)}(\gamma) := \frac{4D\gamma + 2D - d + 2dR}{4D\gamma + 2D - 2d}, \tag{4.21}$$

then a.s.

$$\left| I_{h,N}^{(1)}(s) - I_N(s) \right| \ll N^{-R} I_N(s), \tag{4.22}$$

and in particular, $I_{h,N}^{(1)}(s) \asymp I_N(s)$.

(ii) If we have

$$h \ll_p N^{-\frac{2}{d} K_{d,D,R}^{(2)}(\gamma)}, \quad K_{d,D,R}^{(2)}(\gamma) := \frac{4D\gamma - 2D - d + 2dR}{4D\gamma - 2D - 2d}, \tag{4.23}$$

then a.s.

$$\left| I_{h,N}^{(2)}(s) - I_N(s) \right| \ll N^{-R} I_N(s), \tag{4.24}$$

and in particular, $I_{h,N}^{(2)}(s) \asymp I_N(s)$.

(iii) *Let $\alpha > (\gamma - (1 + d/D)/4) \vee (\gamma - 1/2)$. If*

$$h \ll_p N^{-\frac{2}{d} K_{d,D}^{(M)}(\gamma)}, \quad K_{d,D}^{(M)}(\gamma) := \frac{2D\gamma}{2D\gamma - d}, \tag{4.25}$$

then a.s. $\langle M^{(h,N)} \rangle_T \asymp I_N(2 + 4\alpha - 2\gamma)$ as $N \to \infty$.

(iv) *Let $\alpha > (\gamma - (1 - d/D)/4) \vee (\eta/4 - 1) \vee (-\eta/4)$, where η is as in $(F_{s,\eta})$. If $h \ll_p N^{-2K_{d,D,R}^{(2)}(\gamma)/d}$, then a.s.*

$$\left| F_{h,N}^{(1+2\alpha)} - F_N^{(1+2\alpha)} \right| \ll N^{-R} I_N(s). \tag{4.26}$$

It holds $K_{d,D,R}^{(1)}(\gamma) < K_{d,D,R}^{(2)}(\gamma)$. If $R = 0$, then $K_{d,D,R}^{(1)}(\gamma) < K_{d,D}^{(M)}(\gamma)$, and if $R > 1/2$, then $K_{d,D}^{(M)}(\gamma) < K_{d,D,R}^{(2)}(\gamma)$.

Note that the denominators in (4.21), (4.23), (4.25) are positive if and only if $s^* > 0$, $s^* > 2$, $s^* > 1$, resp., which is satisfied by (4.3).

Proof. We write for $r', s', h > 0$, $N \in \mathbb{N}$ and $Z \in L^2(0, T; H)$:

$$L_{s',r'}^{(h,N)}(Z) := h^{r'} \sum_{k=1}^{N} \lambda_k^{s'+r'/D} \int_0^T |\langle \Phi_k, Z_t \rangle| \, \mathrm{d}t. \tag{4.27}$$

Now let $Z^{(1)} \in L^\infty(0, T; B_{r_1})$ and $Z^{(2)} \in L^\infty(0, T; B_{r_2})$. Then

$$\begin{aligned} &\int_0^T \left\langle A_{h,N}^{(s')} \mathcal{E}_h Z_t^{(1)}, \mathcal{E}_h Z_t^{(2)} \right\rangle_{(h)} \mathrm{d}t \\ &\qquad = \sum_{k=1}^{N} \lambda_k^{s'} \int_0^T \left\langle \mathcal{W}_h E_h \Phi_k, E_h Z_t^{(1)} \right\rangle_{(h)} \left\langle \mathcal{W}_h E_h \Phi_k, E_h Z_t^{(2)} \right\rangle_{(h)} \mathrm{d}t \end{aligned}$$

as well as

$$\int_0^T \left\langle (-A)^{s'} P_N Z_t^{(1)}, P_N Z_t^{(2)} \right\rangle \mathrm{d}t = \sum_{k=1}^{N} \lambda_k^{s'} \int_0^T \left\langle \Phi_k, Z_t^{(1)} \right\rangle \left\langle \Phi_k, Z_t^{(2)} \right\rangle \mathrm{d}t.$$

Consequently, using $|ab - AB| \leq |a - A|\,|b| + |a|\,|b - B| + |a - A|\,|b - B|$ for $a, b, A, B \in \mathbb{R}$, together with (4.10), we obtain

$$
\begin{aligned}
&\left| \int_0^T \left\langle (-A)^{s'} P_N Z_t^{(1)}, P_N Z_t^{(2)} \right\rangle \mathrm{d}t - \int_0^T \left\langle A_{h,N}^{(s')} \mathcal{E}_h Z_t^{(1)}, \mathcal{E}_h Z_t^{(2)} \right\rangle_{(h)} \mathrm{d}t \right| \\
&\qquad \lesssim \sum_{k=1}^{N} \lambda_k^{s'} \int_0^T \Bigg(h^{r_1} k^{r_1/d} \left\| Z_t^{(1)} \right\|_{B_{r_1}} \left| \left\langle \Phi_k, Z_t^{(2)} \right\rangle \right| \\
&\qquad\qquad + h^{r_2} k^{r_2/d} \left\| Z_t^{(2)} \right\|_{B_{r_2}} \left(\left| \left\langle \Phi_k, Z_t^{(1)} \right\rangle \right| + h^{r_1} k^{r_1/d} \left\| Z_t^{(1)} \right\|_{B_{r_1}} \right) \Bigg) \mathrm{d}t \\
&\qquad \lesssim \sup_{0 \leq t \leq T} \left\| Z_t^{(1)} \right\|_{B_{r_1}} L_{s',r_1}^{(h,N)}(Z^{(2)}) + \sup_{0 \leq t \leq T} \left\| Z_t^{(2)} \right\|_{B_{r_2}} L_{s',r_2}^{(h,N)}(Z^{(1)}) \\
&\qquad\qquad + T \sup_{0 \leq t \leq T} \left\| Z_t^{(1)} \right\|_{B_{r_1}} \sup_{0 \leq t \leq T} \left\| Z_t^{(2)} \right\|_{B_{r_2}} h^{r_1 + r_2} \sum_{k=1}^{N} \lambda_k^{s'} k^{(r_1 + r_2)/d} \\
&\qquad \lesssim L_{s',r_1}^{(h,N)}(Z^{(2)}) + L_{s',r_2}^{(h,N)}(Z^{(1)}) + h^{r_1 + r_2} \sum_{k=1}^{N} k^{(s'D + r_1 + r_2)/d} \\
&\qquad \lesssim L_{s',r_1}^{(h,N)}(Z^{(2)}) + L_{s',r_2}^{(h,N)}(Z^{(1)}) + h^{r_1 + r_2} N^{1 + (s'D + r_1 + r_2)/d}.
\end{aligned}
$$

Thus, in order to bound the approximation error stemming from spatial discretization, we have to control the terms $L_{s',r_1}^{(h,N)}(Z^{(2)})$, $L_{s',r_2}^{(h,N)}(Z^{(1)})$ and

$$
L_{s',r_1,r_2}^{\mathrm{rest}} := h^{r_1 + r_2} N^{1 + (s'D + r_1 + r_2)/d}. \tag{4.28}
$$

Based on this consideration, we prove the different cases separately. We will repeatedly use Jensen's inequality in the form $\sum_{k=1}^{N} a_k^{1/2} \leq (N \sum_{k=1}^{N} a_k)^{1/2}$ (in fact, if $a_k \asymp k^r$ for some $r > -1$, then both sides grow as $N^{1+r/2}$). Further, note that for $0 < a \leq A$, the function $(A + x + y)/(a + x)$ is decreasing in $x > -a$ and increasing in $y \in \mathbb{R}$. This implies the relation between $K_{d,D,R}^{(1)}(\gamma)$, $K_{d,D,R}^{(2)}(\gamma)$ and $K_{d,D}^{(M)}(\gamma)$ from the statement, as well as similar relations used in the estimates below.

(i) In order to control $I_{h,N}^{(1)}(s)$, we set $Z^{(1)} = Z^{(2)} = X$. For $\epsilon > 0$, let

$r = r_1 = r_2 := s^* D/2 - \epsilon$. W.l.o.g. assume that $\epsilon < Ds^*$. Then

$$\begin{aligned} L_{s,r}^{(h,N)}(X) &\lesssim \sqrt{T} h^r \sum_{k=1}^{N} \left(\lambda_k^{2s+2r/D} \int_0^T (x_t^{(k)})^2 \mathrm{d}t \right)^{\frac{1}{2}} \\ &\lesssim h^r \left(N \sum_{k=1}^{N} \lambda_k^{2s+2r/D} \int_0^T (x_t^{(k)})^2 \mathrm{d}t \right)^{\frac{1}{2}} \\ &\lesssim h^r \left(N \int_0^T \left\| (-A)^{s+r/D} X_t^N \right\|^2 \mathrm{d}t \right)^{\frac{1}{2}} \\ &\lesssim h^r \left(N^{2+\frac{D}{d}\left(2s+\frac{2r}{D}-2\gamma-1\right)} \right)^{\frac{1}{2}} = h^r N^{1+\frac{D}{d}\left(s+\frac{r}{D}-\gamma-\frac{1}{2}\right)}, \end{aligned}$$

where we made use of Proposition 2.8. This is possible due to $0 < \epsilon < Ds^*$, i.e. $s + r/D > s^*/2$. Now by assumption (4.21), we can choose $\epsilon > 0$ small enough such that

$$h \ll N^{-\frac{2}{d}\frac{4D\gamma+2D-d+2Rd-2\epsilon}{4D\gamma+2D-2d-4\epsilon}}. \tag{4.29}$$

In particular, with $I_N(s) = \int_0^T \left\| (-A)^{s/2} X_t^N \right\|^2 \mathrm{d}t \sim N^{1+D(s-2\gamma-1)/d}$, and $r = s^* D/2 - \epsilon = D(1+2\gamma)/2 - d/2 - \epsilon$, it follows that $L_{s,r}^{(h,N)}(X) \ll N^{-R} I_N(s)$. It remains to bound $L_{s,r,r}^{\mathrm{rest}}$: We have $L_{s,r,r}^{\mathrm{rest}} \ll N^{-R} I_N(s)$ whenever

$$h \ll N^{-\frac{2}{d}\frac{4D\gamma+2D-d+Rd-2\epsilon}{4D\gamma+2D-2d-4\epsilon}},$$

and this follows from (4.29) for any $R \geq 0$. In total, we have shown that $\left| I_{h,N}^{(1)}(s) - I_N(s) \right| \lesssim N^{-R} I_N(s)$, and in particular, $I_{h,N}^{(1)}(s) = I_N(s) + (I_{h,N}^{(1)}(s) - I_N(s)) \asymp I_N(s)$.

(ii) Here, $Z^{(1)} = X$, $Z^{(2)} = (-A)X$, $r_1 = s^* D/2 - \epsilon$ and $r_2 = (s^* - 2)D/2 - \epsilon$ for some $\epsilon > 0$, where w.l.o.g. $\epsilon < D(s^* - 2)$. The terms can be controlled as follows:

$$\begin{aligned} L_{s-1,r_1}^{(h,N)}((-A)X) &\lesssim \sqrt{T} h^{r_1} \sum_{k=1}^{N} \left(\lambda_k^{2s-2+2r_1/D} \int_0^T (\lambda_k x_t^{(k)})^2 \mathrm{d}t \right)^{\frac{1}{2}} \\ &\lesssim h^{r_1} \left(N \sum_{k=1}^{N} \lambda_k^{2s+2r_1/D} \int_0^T (x_t^{(k)})^2 \mathrm{d}t \right)^{\frac{1}{2}}, \end{aligned}$$

which is the same term appearing in (i). As $K^{(1)}_{d,D,R}(\gamma) < K^{(2)}_{d,D,R}(\gamma)$, (4.23) implies that $L^{(h,N)}_{s-1,r_1}((-A)X) \ll N^{-R} I_N(s)$. Further,

$$\begin{aligned}
L^{(h,N)}_{s-1,r_2}(X) &\lesssim \sqrt{T} h^{r_2} \sum_{k=1}^{N} \left(\lambda_k^{2s-2+2r_2/D} \int_0^T (x_t^{(k)})^2 \mathrm{d}t \right)^{\frac{1}{2}} \\
&\lesssim h^{r_2} \left(N \sum_{k=1}^{N} \lambda_k^{2s-2+2r_2/D} \int_0^T (x_t^{(k)})^2 \mathrm{d}t \right)^{\frac{1}{2}} \\
&\lesssim h^{r_2} \left(N \int_0^T \left\| (-A)^{s-1+r_2/D} X_t^N \right\|^2 \mathrm{d}t \right)^{\frac{1}{2}} \\
&\lesssim h^{r_2} \left(N^{2+\frac{D}{d}\left(2s-2+\frac{2r_2}{D}-2\gamma-1\right)} \right)^{\frac{1}{2}} = h^{r_2} N^{1+\frac{D}{d}\left(s-\gamma-\frac{3}{2}+\frac{r_2}{D}\right)}.
\end{aligned}$$

In the last line we have used $s > s^*$ and $\epsilon < D(s^*-2)$, i.e. $s-1+r_2/D > s^*/2$, such that Proposition 2.8 is applicable. As before, we can choose $\epsilon > 0$ small enough such that

$$h \ll N^{-\frac{2}{d}\frac{4D\gamma-2D-d+2dR-2\epsilon}{4D\gamma-2D-2d-4\epsilon}}. \tag{4.30}$$

Using $r_2 = (s^*-2)D/2 - \epsilon = (-1+2\gamma)D/2 - d/2 - \epsilon$ and again the asymptotics of $I_N(s)$, we conclude $L^{(h,N)}_{s-1,r_2}(X) \ll N^{-R} I_N(s)$. Finally, with (4.28), it is clear that $L^{\mathrm{rest}}_{s-1,r_1,r_2} \ll N^{-R} I_N(s)$ whenever

$$h \ll N^{-\frac{2}{d}\frac{4D\gamma-d+dR-2\epsilon}{4D\gamma-2d-4\epsilon}}, \tag{4.31}$$

which is a consequence of (4.30) for $R \geq 0$. Now the claim follows as in (i).

(iii) First note that for $i, j \in \mathbb{N}$ and $R_{ij} := \delta_{ij} - \langle \mathcal{W}_h E_h \Phi_i, E_h \Phi_j \rangle_{(h)}$, we have by (4.10) and (D_1):

$$|R_{ij}| \lesssim (h i^{1/d} j^{1/d})^q \tag{4.32}$$

for each $r_* < q \leq r*$. We expand the quadratic variation of $M^{(h,N)}$: For $j \in \mathbb{N}$, we have by definition of $A^{(1+2\alpha)}_{h,N}$:

$$\begin{aligned}
&\left\langle \Phi_j, (-A)^{-\gamma} (\mathcal{E}_h)^* A^{(1+2\alpha)}_{h,N} \mathcal{E}_h X_t \right\rangle \\
&\qquad = \lambda_j^{-\gamma} \sum_{k=1}^{N} \lambda_k^{1+2\alpha} \left\langle \mathcal{E}_h \Phi_j, E_h \Phi_k \right\rangle_{(h)} \left\langle E_h \Phi_k, \mathcal{E}_h X_t \right\rangle_{(h)},
\end{aligned}$$

and consequently,

$$
\begin{aligned}
&\langle M^{(h,N)}\rangle_T = \int_0^T \left\|(-A)^{-\gamma}(\mathcal{E}_h)^* A_{h,N}^{(1+2\alpha)}\mathcal{E}_h X_t\right\|^2 \mathrm{d}t \\
&= \sum_{j=1}^{\infty} \lambda_j^{-2\gamma} \\
&\quad \times \int_0^T \left(\sum_{k=1}^{N} \lambda_k^{1+2\alpha} \left\langle \mathcal{W}_h E_h \Phi_j, E_h \Phi_k \right\rangle_{(h)} \left\langle \mathcal{W}_h E_h \Phi_k, E_h X_t \right\rangle_{(h)} \right)^2 \mathrm{d}t \\
&= \sum_{j=1}^{\infty} \lambda_j^{-2\gamma} \sum_{k,l=1}^{N} \lambda_k^{1+2\alpha}\lambda_l^{1+2\alpha} \\
&\quad \times \int_0^T \left\langle \mathcal{W}_h E_h \Phi_k, E_h X_t \right\rangle_{(h)} \left\langle \mathcal{W}_h E_h \Phi_l, E_h X_t \right\rangle_{(h)} \mathrm{d}t \\
&\quad \times \left\langle \mathcal{W}_h E_h \Phi_j, E_h \Phi_k \right\rangle_{(h)} \left\langle \mathcal{W}_h E_h \Phi_j, E_h \Phi_l \right\rangle_{(h)} \\
&= \sum_{k=1}^{N} \lambda_k^{2+4\alpha-2\gamma} \int_0^T \left\langle \mathcal{W}_h E_h \Phi_k, E_h X_t \right\rangle_{(h)}^2 \mathrm{d}t \\
&\quad - 2 \sum_{k,l=1}^{N} \lambda_k^{1+2\alpha-2\gamma}\lambda_l^{1+2\alpha} R_{kl} \\
&\qquad \times \int_0^T \left\langle \mathcal{W}_h E_h \Phi_k, E_h X_t \right\rangle_{(h)} \left\langle \mathcal{W}_h E_h \Phi_l, E_h X_t \right\rangle_{(h)} \mathrm{d}t \\
&\quad + \sum_{k,l=1}^{N} \lambda_k^{1+2\alpha}\lambda_l^{1+2\alpha} \int_0^T \left\langle \mathcal{W}_h E_h \Phi_k, E_h X_t \right\rangle_{(h)} \left\langle \mathcal{W}_h E_h \Phi_l, E_h X_t \right\rangle_{(h)} \mathrm{d}t \\
&\qquad \times \sum_{j=1}^{\infty} \lambda_j^{-2\gamma} R_{jk} R_{jl} \\
&=: A_1 - 2A_2 + A_3.
\end{aligned}
$$

We have $2 + 4\alpha - 2\gamma > s^*$ due to $\alpha > \gamma - (1 + d/D)/4$, and further, $K^{(1)}_{d,D,R}(\gamma) < K^{(M)}_{d,D}(\gamma)$ (with $R = 0$ in the first term), so part (i) yields

$$A_1 = I^{(1)}_{h,N}(2 + 4\alpha - 2\gamma) \asymp I_N(2 + 4\alpha - 2\gamma).$$

It remains to find bounds for A_2 and A_3. We start with the latter term. For $r_* < q < D\gamma - d/2 = (s^* - 1)D/2$, (4.32) and the Cauchy–Schwarz inequality give

$$|A_3| \lesssim h^{2q} \left(\sum_{k=1}^{N} \left(\lambda_k^{2+4\alpha} k^{2q/d} \int_0^T \langle \mathcal{W}_h E_h \Phi_k, E_h X_t \rangle_{(h)}^2 \, \mathrm{d}t \right)^{\frac{1}{2}} \right)^2$$
$$\times \sum_{j=1}^{\infty} \lambda_j^{-2\gamma} j^{2q/d},$$

where the last sum is finite. Again using (i), it follows that

$$\begin{aligned}
|A_3| &\lesssim h^{2q} N \sum_{k=1}^{N} \lambda_k^{2+4\alpha+2q/D} \int_0^T \langle \mathcal{W}_h E_h \Phi_k, E_h X_t \rangle_{(h)}^2 \, \mathrm{d}t \\
&\lesssim h^{2q} N I_{h,N}^{(1)}(2 + 4\alpha + 2q/D) \\
&\lesssim h^{2q} N I_N(2 + 4\alpha + 2q/D) \\
&\lesssim h^{2q} N^{2+\frac{D}{d}\left(2+4\alpha+\frac{2q}{D}-2\gamma-1\right)},
\end{aligned}$$

where we have used $2 + 4\alpha + 2q/D > 2 + 4\alpha - 2\gamma > s^*$. Now choose $q = (s^* - 1)D/2 - \epsilon = D\gamma - d/2 - \epsilon$ for some $\epsilon > 0$, and in addition let ϵ be small enough such that by (4.25),

$$h \ll N^{-\frac{2}{d}\frac{2D\gamma-\epsilon}{2D\gamma-d-2\epsilon}}. \tag{4.33}$$

Then we immediately obtain $|A_3| \ll I_N(2 + 4\alpha - 2\gamma)$. The bound on A_2 is similar. With (4.32) and $r_* < q \leq r^*$,

$$\begin{aligned}
|A_2| &\lesssim h^q \sum_{k=1}^{N} \left(\lambda_k^{2+4\alpha-4\gamma} k^{2q/d} \int_0^T \langle \mathcal{W}_h E_h \Phi_k, E_h X_t \rangle_{(h)}^2 \, \mathrm{d}t \right)^{\frac{1}{2}} \\
&\quad \times \sum_{l=1}^{N} \left(\lambda_l^{2+4\alpha} l^{2q/d} \int_0^T \langle \mathcal{W}_h E_h \Phi_l, E_h X_t \rangle_{(h)}^2 \, \mathrm{d}t \right)^{\frac{1}{2}} \\
&=: h^q B_1 B_2.
\end{aligned}$$

The sums B_1 and B_2 are treated as before, using part (i). For B_1,

$$
\begin{aligned}
B_1 &\lesssim \left(N \sum_{k=1}^{N} \lambda_k^{2+4\alpha-4\gamma+\frac{2q}{D}} \int_0^T \langle \mathcal{W}_h E_h \Phi_k, E_h X_t \rangle_{(h)}^2 \,\mathrm{d}t \right)^{\frac{1}{2}} \\
&\lesssim \left(N I_{h,N}^{(1)}(2+4\alpha-4\gamma+2q/D) \right)^{\frac{1}{2}} \\
&\lesssim \left(N^{2+\frac{D}{d}\left(1+4\alpha-6\gamma+\frac{2q}{D}\right)} \right)^{\frac{1}{2}} = N^{1+\frac{D}{d}\left(\frac{1}{2}+2\alpha-3\gamma+\frac{q}{D}\right)},
\end{aligned}
$$

if $2+4\alpha-4\gamma+2q/D > s^*$, which is the case for $q = s^*D/2 \le r^*$ due to $\alpha > \gamma - 1/2$. Similarly $B_2 \lesssim N^{1+D(1/2+2\alpha-\gamma+q/D)/d}$. Therefore,

$$
|A_2| \lesssim h^q N^{2+\frac{D}{d}\left(1+4\alpha-4\gamma+\frac{2q}{D}\right)}.
$$

Now (4.33) implies

$$
h \ll N^{-\frac{2}{d}\frac{2D\gamma+D}{2D\gamma-d+D}}. \tag{4.34}
$$

With $q = s^*D/2 \le r^*$, we get $|A_2| \ll I_N(2+4\alpha-2\gamma)$. Putting things together,

$$
\begin{aligned}
\langle M^{(h,N)} \rangle_T &= A_1 - 2A_2 + A_3 \\
&\asymp I_N(2+4\alpha-2\gamma) - 2A_2 + A_3 \\
&\asymp I_N(2+4\alpha-2\gamma).
\end{aligned}
$$

(iv) Set $Z^{(1)} = X$, $Z^{(2)} = F(X)$, $r_1 = s^*D/2-\epsilon$ and $r_2 = (s^*-2+\eta)D/2-\epsilon$. Let ϵ be small enough such that

$$
h \ll N^{-\frac{2}{d}\frac{4D\gamma+2D-d+2dR-D\eta}{4D\gamma+2D-2d-4\epsilon}}, \tag{4.35}
$$

$$
h \ll N^{-\frac{2}{d}\frac{4D\gamma-2D-d+D\eta+2dR-2\epsilon}{4D\gamma-2D-2d+2D\eta-4\epsilon}}, \tag{4.36}
$$

$$
h \ll N^{-\frac{2}{d}\frac{8D\gamma-2d+D\eta+2dR-4\epsilon}{8D\gamma-4d+2D\eta-8\epsilon}}, \tag{4.37}
$$

which is possible due to $h \ll_p N^{-2K_{d,D,R}^{(2)}(\gamma)/d}$. Since $\alpha > \eta/4 - 1$, it holds that $1+2\alpha+(r_1-r_2)/D = 2+2\alpha-\eta/2 > 0$, and consequently,

$$\begin{aligned}
L^{(h,N)}_{1+2\alpha,r_1}(F(X)) &\lesssim h^{r_1}\left(N\sum_{k=1}^{N}\lambda_k^{2+4\alpha+\frac{2r_1}{D}}\int_0^T\langle\Phi_k, F(X)(t)\rangle^2\,\mathrm{d}t\right)^{1/2}\\
&= h^{r_1}\left(N\int_0^T\left\|(-A)^{1+2\alpha+\frac{r_1}{D}}F(X)(t)\right\|_H^2\mathrm{d}t\right)^{1/2}\\
&\le h^{r_1}\left(N\lambda_N^{2+4\alpha+\frac{2r_1-2r_2}{D}}\int_0^T\left\|(-A)^{\frac{r_2}{D}}F(X)(t)\right\|_H^2\mathrm{d}t\right)^{1/2}\\
&\lesssim h^{r_1}N^{\frac{1}{2}+\frac{D}{d}\left(1+2\alpha+\frac{r_1-r_2}{D}\right)} = h^{r_1}N^{\frac{1}{2}+\frac{D}{d}\left(1+2\alpha+\frac{2-\eta}{2}\right)},
\end{aligned}$$

and a direct calculation using (4.35) and $r_1 = (1+2\gamma)D/2 - d/2 - \epsilon$ yields $L^{(h,N)}_{1+2\alpha,r_1}(F(X)) \ll N^{-R}I_N(2+2\alpha)$. Further, we have $\alpha > -\eta/4$, and if $\epsilon < (4\alpha+\eta)D/2$, it follows that $2+4\alpha+2r_2/D > s^*$, thus

$$\begin{aligned}
L^{(h,N)}_{1+2\alpha,r_2}(X) &\lesssim h^{r_2}\left(N\sum_{k=1}^{N}\lambda_k^{2+4\alpha+\frac{2r_2}{D}}\int_0^T x_t^{(k)}\mathrm{d}t\right)^{1/2}\\
&= h^{r_2}\left(N\int_0^T\left\|(-A)^{1+2\alpha+\frac{r_2}{D}}X_t^N\right\|_H^2\mathrm{d}t\right)^{1/2}\\
&\lesssim h^{r_2}N^{1+\frac{D}{d}\left(\frac{1}{2}+2\alpha-\gamma+\frac{r_2}{D}\right)},
\end{aligned}$$

and (4.36) together with $r_2 = (-1+2\gamma+\eta)D/2 - d/2 - \epsilon$ gives $L^{(h,N)}_{1+2\alpha,r_2}(X) \ll N^{-R}I_N(2+2\alpha)$. Finally, (4.37) can be reformulated as $L^{\mathrm{rest}}_{1+2\alpha,r_1,r_2} \ll N^{-R}I_N(2+2\alpha)$. This finishes the proof.

□

Motivated by the previous proposition, we define

$$K_{d,D}(\gamma) := K^{(2)}_{d,D,\frac{1}{2}+\frac{D}{2d}}(\gamma) = \frac{4D\gamma - D}{4D\gamma - 2D - 2d}.$$

Theorem 4.3. *In the setting of this section, assume (D_0), (D_1) and (D_2), and further*

$$h \ll_p N^{-\frac{2}{d}K_{d,D}(\gamma)}. \tag{4.38}$$

With η from $(F_{s,\eta})$, let $\alpha > (\gamma-(1+d/D)/4)\vee(\gamma-1/2)\vee(\eta/4-1)\vee(-\eta/4)$.

(i) If $\eta > 1 + d/D$, then

$$N^{\frac{1}{2}+\frac{D}{2d}}\left(\hat{\theta}^{\text{discr}}_{h,N} - \theta\right) \xrightarrow{d} \mathcal{N}(0, \Sigma) \tag{4.39}$$

as $N \to \infty$, $h \to 0$, where Σ is given by (2.30).

(ii) If $\eta \leq 1 + d/D$, then

$$N^a \left(\hat{\theta}^{\text{discr}}_{h,N} - \theta\right) \xrightarrow{\mathbb{P}} 0 \tag{4.40}$$

for any $a < D\eta/(2d)$ as $N \to \infty$, $h \to 0$.

Proof. This follows directly from the decomposition of $\hat{\theta}^{\text{discr}}_{h,N}$ as in (4.13): By Theorem A.1 and Proposition 4.2 (iii), $M_T^{(h,N)}/(\langle M^{(h,N)}\rangle_T)^{1/2} \to \mathcal{N}(0,1)$. Further, $N^{1/2+D/(2d)}(\langle M^{(h,N)}\rangle_T)^{1/2}/I^{(1)}_{h,N}(2+2\alpha) \to \Sigma^{1/2}$. Next, the term $F_N^{(1+2\alpha)}/I^{(2)}_{h,N}(2+2\alpha) \asymp F_N^{(1+2\alpha)}/I_N(2+2\alpha)$ is treated exactly as in Theorem 2.11, which leads to the case distinction $\eta > 1+d/D$ compared to $\eta \leq 1+d/D$. Finally, all other terms from (4.13) converge to zero with rate $N^{-1/2-D/(2d)}$ by Proposition 4.2, and the claim follows from the Slutsky lemma. $\square$

Remark 4.4.

(i) For fixed $d \geq 1$ and $D > 0$, we have $K_{d,D}(\gamma) \to 1$ for $\gamma \to \infty$ (or equivalently $s^* \to \infty$). This means that for large spatial regularity of X, a spatial precision of order $h \ll_p N^{-2/d}$ is sufficient in order to transfer the asymptotic results from the classical spectral approach.

(ii) On a bounded domain in dimension d, one typically has $M \sim h^{-d}$ point observations. (For example, let $\mathcal{D}$ be a hypercube in $\mathbb{R}^d$, where the point evaluation grid consists of points which are aligned along the coordinate lines in an equidistant way.) In the large regularity setting from the previous comment, this number of point observations leads to the relation

$$N^2 \ll_p M. \tag{4.41}$$

This relation does not depend on the dimension d. In this sense, a given resolution level N can be recovered from a dimension-independent number of point observations (if the spatial regularity of X is large) within the spectral approach.

(iii) On the level of the estimator for diffusivity, this means the following: In the setting of the previous comment, let additionally $F = 0$ for simplicity, such that $\hat{\theta}_N^{\text{lin}}$ converges to θ with optimal rate $N^{-1/2-D/(2d)}$. In terms of the number M of point observations, this corresponds to the convergence rate $M^{-1/4-D/(4d)}$ for $\hat{\theta}_{h,N}^{\text{discr}}$ (neglecting terms of arbitrary small polynomial order in N in the relation $N^2 \ll_p M$). This rate is upper bounded by $M^{-1/4}$. While this bound decays rather slowly in M, it holds uniformly in d.[1] *Therefore, sparse observations in an high-dimensional setting can still yield reasonable results.*

(iv) Below, we explain how to further improve this bound on the convergence rate of $\hat{\theta}_{h,N}^{\text{discr}}$ in M by tightening (4.10).

(v) We point out that $I_{h,N}^{(1)}(s)$ need not be a good approximation for $I_N(s)$. In fact, according to Proposition 4.2, the absolute error $\left|I_{h,N}^{(1)}(s) - I_N(s)\right|$ may even diverge, but slowly compared to the energy $I_N(s)$ itself. The same is true for the other approximation terms.

(vi) We highlight that there is no assumption on the shape of $\mathcal{D}$ or the distribution of the point evaluations within $\mathcal{D}$ other than the integral approximation property from (D_2). To the best of our knowledge, Theorem 4.3 is the first rigorous asymptotic result for diffusivity estimation based on such general discrete point evaluation schemes.

Next, we consider different cases in which higher order approximation estimates allow to connect to the assumptions from Theorem 4.3.

Example 4.5 (Quadrature formulas in $d = 1$)**.** *Let $L > 0$ and $\mathcal{D} = [0, L]$. Further, let $A = -(-\Delta)^{D/2}$. With Dirichlet boundary conditions, we have $\Phi_k(x) = \sqrt{2}\sin(\pi k x/L)$ and $\lambda_k = (\pi k/L)^D$. For $k \in \mathbb{N}_0$, equip the space of k times differentiable functions $C^k(\overline{\mathcal{D}})$ with the norm $\|f\|_{C^k} = \sum_{i=1}^k \|\partial_x^i f\|_\infty$. Further, for $r > 0$ let $\mathcal{C}^r(\overline{\mathcal{D}})$ be the Hölder–Zygmund space with the norm $\|\cdot\|_{\mathcal{C}^r} = \|\cdot\|_\infty + |\cdot|_{\mathcal{C}^r}$. Here, $|f|_{\mathcal{C}^r} = \sup_{x\in\mathcal{D}, 0<h<1 \text{ with } x+Kh\in\mathcal{D}} h^{-r}\left|\Delta_h^K f(x)\right|$ for any $K > r$, where $\Delta_h f = f(\cdot + h) - f$ is the difference operator. Different choices of $K > r$ lead to equivalent norms. For integer $r \in \mathbb{N}$, the spaces*

[1] Note, however, that for fixed γ, the deviation from the large spatial regularity setting still depends on d in the term $K_{d,D}(\gamma)$, i.e. the regularity of X needed to approach the large regularity regime grows with d.

$C^r(\overline{\mathcal{D}})$ and $\mathcal{C}^r(\overline{\mathcal{D}})$ do not coincide, but the former is a subspace of the latter. See e.g. [Tri10a, Tri10b] or [GN15, Chapter 4] for further details on these spaces. We will use that the Hölder–Zygmund spaces can be identified as interpolation spaces between $C(\overline{\mathcal{D}})$ and $C^k(\overline{\mathcal{D}})$ for any $k \in \mathbb{N}$, see e.g. [Lun95, Chapter 1] for details. The scale of Banach spaces $(B_r)_{r>r_*}$ is given by $(\mathcal{C}^r(\overline{\mathcal{D}}))_{r>0}$ with $r_* = 0$. The upper bound r^* is arbitrary. By Lemma 5.5 below, (4.7) is true. It is clear that the $\mathcal{C}^r(\overline{\mathcal{D}})$ are Banach algebras [Tri10a, Section 2.8.3], so (D_0) is trivially satisfied. Further, for $r > 0$,

$$\begin{aligned}\|\Phi_k\|_{\mathcal{C}^r} &\lesssim 1 + |\Phi_k|_{\mathcal{C}^r} \lesssim 1 + \sup_{x\in\mathbb{R},h>0} h^{-r}\left|(\Delta_h^K\Phi_k)(x)\right| \\ &= 1 + \sup_{x'\in\mathbb{R},h'>0} k^r h^{-r}\left|(\Delta_{h'}^K\Phi_1)(x')\right| \lesssim k^r,\end{aligned}$$

where we substituted $x' = kx$ and $h' = kh$. Thus, (D_1) is satisfied.

Fix $h^* > 0$. For $0 < h < h^*$, let $M_h \in \mathbb{N}$ such that $M_h \sim h^{-1}$ for $h \to 0$. Let $\pi(h) = \{x_0^{(h)}, x_1^{(h)}, \ldots, x_{M_h-1}^{(h)}\}$ be a partition of M_h points in $[0, L]$. Let E_h be the point evaluation operator associated to $\pi(h)$. We consider quadrature formulas of the form $Q^{(h)}(f) = \sum_{i=1}^{M_h} w_i^{(h)} f(x_i^{(h)})$ for some weights $w_i^{(h)} \in \mathbb{R}$. Let $k^* \in \mathbb{N}$, $k^* > r^*$. Typically, $Q^{(h)}$ satisfies an error estimate of the form

$$\left|\int_{\mathcal{D}} f\mathrm{d}x - Q^{(h)}(f)\right| \lesssim M_h^{-k^*}\|f\|_{C^{k^*}} \tag{4.42}$$

for $f \in C^{k^*}(\mathcal{D})$. Examples include the composite Newton-Cotes formulas of order k^* on equidistant partitions, or Gaussian quadrature formulas, where $\|\cdot\|_{C^{k^*}}$ can be even replaced by an L^2-based Sobolev norm. This is well-known, see for example [QSS00]. The right-hand side of (4.42) can be bounded by $h^{k^*}\|f\|_{C^{k^*}}$ (up to a constant), and the exact interpolation theorem [AF03, Theorem 7.23], applied to the operator $\int_{\mathcal{D}} \cdot \,\mathrm{d}x - Q^{(h)}$, extends the resulting estimate to all $0 \le r \le r^*$, where the norm on the right-hand side of (4.42) is replaced by $\|\cdot\|_{\mathcal{C}^r}$. Thus (D_2) holds with the weight matrix $\mathcal{W}_h$ determined by the quadrature weights $w_i^{(h)}$. Consequently, Theorem 4.3 is applicable in this setting.

Example 4.6 (Finite element method in $d \ge 2$)**.** *Let $A = -(-\Delta)^{D/2}$, set $B_r = H_{2r/D} = W^{r,2}(\mathcal{D})$, let $r_* = d/2$ and $r^* \in \mathbb{N}$ arbitrary. Condition (D_0) is immediate, and for (D_1), note that $\|\Phi_k\|_{B_r} = \|\Phi_k\|_{2r/D} = \lambda_k^{r/D} \lesssim k^{r/d}$. In order to describe the discretization operator E_h and the approximation property (D_2), we make use of results from the theory of finite elements. The* *finite element method is a standard approach from numerical analysis with a huge body of literature, we follow the exposition from [Cia02]. As we are interested in point evaluations, we only consider Lagrange finite elements. Let K_0 be a compact reference domain (typically a simplex or cube in d dimensions) with non-empty interior. Fix r^* points $y_0^{(1)}, \dots, y_0^{(r^*)} \in K_0$, and let P_0 be a r^*-dimensional space of polynomial functions defined on K_0, such that for $p_0 \in P_0$, $p_0(y_0^{(i)}) = 0$ for $1 \leq i \leq r^*$ implies $p_0 = 0$. Then there are r^* polynomials $p_0^{(1)}, \dots, p_0^{(r^*)} \in P_0$ such that $p_0^{(i)}(y_0^{(j)}) = \delta_{ij}$, and the interpolation operator $\Pi_0 : C(K_0) \to P_0$, given by $(P_0 f)(x) = \sum_{i=1}^{r^*} f(y_0^{(i)}) p_0^{(i)}(x)$, is well-defined and acts as the identity on P_0. Now we partition the domain $\mathcal{D}$ into a family $(K_j)_{j=1,\dots,L}$ of compact domains with open interior (which overlap only on their boundaries), such that there is a diffeomorphism $F_j : K_0 \to K_j$ for $1 \leq j \leq L$. Let P_j consist of the pullback of functions $p_0 \in P_0$ via F_j^{-1}, i.e. $P_j = \{p_0 \circ F_j^{-1} \mid p_0 \in P_0\}$, and set $y_j^{(i)} := F_j(y_0^{(i)})$. The interpolation operator on K_j is given by $\Pi_j f = (\Pi_0(f \circ F_j)) \circ F_j^{-1}$. Typically, the F_j are affine functions, which leads to a partition of polygonal domains $\mathcal{D}$, but also curved elements K_j are possible (see e.g. [Zlá73], [Cia02, Chapter 4.3]), which allow to handle a smooth boundary of $\mathcal{D}$. We assume mild compatibility criteria on the partition of $\mathcal{D}$: The images of the faces of K_0 under the F_j have disjoint $(d-1)$-dimensional interior or coincide. Further, for $f \in C(\overline{\mathcal{D}})$, the interpolation polynomials $\Pi_j f|_{K_j}$ coincide on the boundaries of the K_j, such that there is a well-defined interpolation operator Π acting on $C(\overline{\mathcal{D}})$. For each $1 \leq j \leq L$, let h_j denote the diameter of K_j and ρ_j the diameter of the largest ball contained in K_j. We assume that there is a constant $C > 0$ such that $h_j/\rho_j \leq C$ for all j. Finally, let $h := \max_{1 \leq j \leq L} h_j$ be the mesh size of the partition of $\mathcal{D}$.*

Let $\{x_i^{(h)}\}_{1 \leq i \leq M_h}$ be the set of all M_h points $y_j^{(k)}$ in the interior of $\mathcal{D}$, where $1 \leq j \leq L$, $1 \leq k \leq r^*$, and the labeling is arbitrary. Using Dirichlet boundary conditions, we can neglect evaluations at the boundary of the domain by assuming $f = 0$ on $\partial\mathcal{D}$. Now, E_h is the operator mapping $f \in C(\mathcal{D})$ to the vector of point evaluations $f(x_i^{(h)})$ for $1 \leq i \leq M_h$. The weights $w_i^{(h)}$ are given by $w_i^{(h)} := \int_{\mathcal{D}} \Pi f_i \mathrm{d}x$ for any function $f_i \in C(\overline{\mathcal{D}})$ that satisfies $f_i(x_j^{(h)}) = \delta_{ij}$, $1 \leq i, j \leq M_h$, and vanishes on $\partial\mathcal{D}$.

We denote by $|f|_{k,2}^2 = \sum_{|\alpha|=k} |\partial_\alpha f|_{L^2(\mathcal{D})}^2$ the L^2-Sobolev seminorm of order k, where α is a multi-index. It is well-known that for $0 \leq k \leq r^*$,

$$\|f - \Pi f\|_{L^2(\mathcal{D})} \lesssim h^k \, |f|_{k,2} \tag{4.43}$$

see e.g. [Cia02, Theorem 3.2.1] for the affine case. Together with the obvious estimate $|f|_{k,2} \leq \|f\|_{B_k}$ *and*

$$\left| \int_{\mathcal{D}} f \mathrm{d}x - \sum_{i=1}^{M_h} w_i^{(h)} f(x_i^{(h)}) \right| \leq \|f - \Pi f\|_{L^1(\mathcal{D})} \lesssim \|f - \Pi f\|_{L^2(\mathcal{D})}, \qquad (4.44)$$

we obtain that (4.6) *is true for integer* r*, and the exact interpolation theorem [AF03, Theorem 7.23], applied to the operator* $I - \Pi$*, extends this estimate to general* $0 \leq r \leq r^*$*. In particular,* (D_2) *is true.*

In total, Theorem 4.3 is applicable in this setting. Note that the finite element method provides a very flexible approach to discretizing $\mathcal{D}$*, which allows to handle point evaluations schemes way beyond a rectangular grid.*

Finally, we outline a possibility to further improve the results from Theorem 4.3 and Remark 4.4. An explicit understanding of the discretization error of the Fourier modes, which is not based on the universal approximation error from (D_2), can improve the bounds on the number of spatial points needed in order to recover the spectral approach from discrete observations. Our standing assumption for the rest of this section is that for each $h > 0$, the vectors $E_h\Phi_1, \dots, E_h\Phi_{M_h} \in \mathbb{R}^{M_h}$ are linearly independent. Consider the operator

$$T_h Z := \sum_{k=1}^{M_h} \langle E_h Z, E_h \Phi_k \rangle_{(h)} \Phi_k \qquad (4.45)$$

for $Z \in C(\mathcal{D})$. It clearly satisfies

$$\langle T_h Z, \Phi_k \rangle = \langle E_h Z, E_h \Phi_k \rangle_{(h)} \qquad (4.46)$$

for $1 \leq k \leq M_h$. In particular, if $\langle E_h\Phi_k, E_h\Phi_\ell \rangle_{(h)} = \langle \Phi_k, \Phi_\ell \rangle$ for $1 \leq k, \ell \leq M_h$, then the left-hand side of (4.46) can be replaced with $\langle E_h T_h Z, E_h \Phi_k \rangle_{(h)}$, the linear independence of $(E_h\Phi_k)_{1 \leq k \leq M_h}$ implies that $T_h Z$ coincides with Z at the observation points. In this case, T_h is the interpolation operator at the given observation points associated to the basis functions $(\Phi_k)_{1 \leq k \leq M_h}$. Instead of (D_2), we assume the following:

(D_2^*) For any $0 < h < h^*$, $r_* < r \leq r^*$ and $Z \in B_r$,

$$\|Z - T_h Z\|_{L^2(\mathcal{D})} \lesssim h^r \|Z\|_{B_r}. \qquad (4.47)$$

This immediately implies

$$\left|\langle Z, \Phi_k\rangle - \langle E_h Z, E_h \Phi_k\rangle_{(h)}\right| = |\langle Z - T_h Z, \Phi_k\rangle|$$
$$\leq \|Z - T_h Z\|_{L^2(\mathcal{D})} \lesssim h^r \|Z\|_{B_r} \tag{4.48}$$

for $1 \leq k \leq M_h$. If $h \ll_p N^{-1/d}$ for $N \to \infty$ and $h \to 0$, then (4.48) is true for all $1 \leq k \leq N$ at least asymptotically. Note that (4.48) is an improvement over (4.10) by a factor $k^{r/d}$.

This leads to an additional gain in rate in the proof of Proposition 4.2. As a consequence, in Theorem 4.3, condition (4.38) can be relaxed:

Theorem 4.7. *In the present setting, let* (D_0)*,* (D_1) *and* (D_2^*) *hold. Then there is a function* $K^*_{d,D}$*, with* $K^*_{d,D}(\gamma) \to 1$ *for* $\gamma \to \infty$*, such that* (4.38) *can be substituted by*

$$h \ll_p N^{-(1/d)K^*_{d,D}(\gamma)} \tag{4.49}$$

without changing the conclusions of Theorem 4.3. □

It follows that (4.41) can be improved to

$$N \ll_p M, \tag{4.50}$$

i.e. N point evaluations suffice in order to recover the spectral resolution level N for diffusivity estimation in the large regularity regime (neglecting terms of arbitrarily small polynomial order). Consequently, the convergence rate of $\hat{\theta}^{\text{discr}}_{h,N}$ is described by $M^{-1/2-D/(2d)}$ in Remark 4.4.

We shortly compare our result to related literature. Note that the construction of $\hat{\theta}^{\text{discr}}_{h,N}$ is independent of the dispersion intensity σ, which may be treated as unknown. In fact, in [HT21b] it is shown that under spatially and temporally discrete observations of a stochastic heat equation driven by space-time white noise in $d = 1$, the parameters θ and σ can be jointly estimated at rate $M^{-3/2} = M^{-1/2-D/(2d)}$ if the observation scheme is balanced, or if the resolution in time exceeds that of a balanced observation scheme. Of course, as we work with time-continuous observations, we may always assume arbitrarily high resolution in time. In this sense, Theorem 4.7 is compatible with [HT21b]. On the other hand, if σ is treated as known, the observation of the process continuously in time at a single point in space suffices to recover θ, see e.g. [PT07, CH20].

In contrast to the integral approximation estimate from (D_2), bounds on the approximation error of T_h as in (D_2^*) seem to be harder to obtain. An important example is given by a uniform observation grid on a periodic domain:

Example 4.8. *Let $d = 1$, let w.l.o.g. $\mathcal{D}$ be an interval of length 2π, and consider periodic (instead of Dirichlet) boundary conditions. Then we can identify $\mathcal{D} \simeq \mathbb{R}/(2\pi\mathbb{Z})$. Let $B_r = W^{r,2}(\mathcal{D})$. The observation grid is assumed to be spatially uniform. In this case, $\langle E_h\Phi_k, E_h\Phi_\ell\rangle_{(h)} = \langle \Phi_k, \Phi_\ell\rangle$ for $1 \leq k, \ell \leq M_h$, so T_h is a trigonometric interpolation operator, and (D_2^*) holds [KO79]. This can be extended to rectangular domains with a uniform point evaluation grid in larger dimension $d \geq 2$ [Pas80]. See also [CHQZ88, Chapter 9], [QSS00, Chapter 10.9], [SV02, Chapter 8] for discussions of trigonometric interpolation.*

Nonetheless, even for non-uniform observation point grids in $d = 1$, the situation is less clear. Recent works [Aus16, AT17] indicate that the trigonometric interpolation operator on a non-uniform grid has diminished approximation power, at least in $\|\cdot\|_\infty$, with a convergence rate depending on the deviation from the uniform point grid. While in this case we cannot expect that T_h and the trigonometric interpolation operator coincide, this gives a hint that the validity of (D_2^*) can be more involved than (D_2).

It is an interesting question for further research if (4.50) and the resulting (optimal) convergence rate in M for diffusivity estimators can be achieved in non-rectangular domains in dimension $d \geq 2$ that do not arise as the tensor product of one-dimensional intervals, or if it is possible to find a domain $\mathcal{D}$ and an observation point distribution within $\mathcal{D}$ such that (4.41) cannot be improved.

Chapter 5

The Local Approach

This chapter is an adaptation of material from [ACP20].

The local approach to parameter estimation for SPDEs is a recent development different from (and in some sense complementary to) the spectral approach. It has been introduced in [AR21] for the stochastic heat equation. [ACP20] generalizes the theory to semilinear models and [ABJR21] applies the local approach to the stochastic Meinhardt model. The novelty from the local approach is its observation scheme. It is assumed that a spatially localized average of the solution process X is observed on $[0, T]$, which is formally realized as the convolution with a compactly supported kernel. This is a physically realistic assumption in many cases. As the support of the kernel shrinks (corresponding to observing X with high resolution at a point in space) the true diffusivity can be recovered.

Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded open domain with smooth boundary. For $s \in \mathbb{N}$, $p \geq 1$ denote by $W^{s,p}(\mathcal{D})$, $W_0^{s,p}(\mathcal{D})$ the Sobolev spaces as in [AF03], and for non–integer $s \geq 0$, their complex interpolation spaces. Let Δ be the Laplacian, with domain $W^{2,p}(\mathcal{D}) \cap W_0^{1,p}(\mathcal{D})$ in $L^p(\mathcal{D})$ (see e.g. [GT01, Chapter 9.6]). For $s \geq 0$, $p \geq 1$, let $H^{s,p}(\mathcal{D}) := D_p((-\Delta)^{s/2}) \subset L^p(\mathcal{D})$ be the maximal domain of definition of the fractional Laplacian $(-\Delta)^{s/2}$ acting as a closed, densely defined operator on $L^p(\mathcal{D})$, cf. [Yag10, Chapter 16]. $H^{s,p}(\mathcal{D})$ is equipped with the norm $\|\cdot\|_{s,p} := \left\|(-\Delta)^{s/2} \cdot\right\|_{L^p(\mathcal{D})}$ and is a closed subspace of $W^{s,p}(\mathcal{D})$ for any $s \geq 0$ and $p \geq 1$. For $s < 1$, $H^{s,p}(\mathcal{D})$ is the completion of $L^p(\mathcal{D})$ w.r.t. the norm $\|\cdot\|_{s,p}$. It is clear that for $s, s' \in \mathbb{R}$ and $p \geq 1$, $(-\Delta)^{s/2}$ maps $H^{s'+s,p}(\mathcal{D})$ into $H^{s',p}(\mathcal{D})$. We write $R_p(s) := L^\infty(0, T; H^{s,p}(\mathcal{D}))$. These spaces are equipped with their natural norm $\|Z\|_{R_p(s)} = \sup_{0\leq t\leq T} \|Z_t\|_{s,p}$.

Finally, let Δ_0 be the Laplacian as a closed, densely defined operator on $L^2(\mathbb{R}^d)$. See e.g. [Tri10a, Tri10b] for more details on function spaces.

In this chapter, we consider a semilinear SPDE

$$\mathrm{d}X_t = \theta\Delta X_t \mathrm{d}t + F(X)(t)\mathrm{d}t + B\mathrm{d}W_t \tag{5.1}$$

together with Dirichlet boundary conditions $X_t = 0$ on $\partial\mathcal{D}$ for all $0 \le t \le T$, and initial condition X_0. As in the previous chapters, $F : C(0,T; L^2(\mathcal{D})) \supseteq D(F) \to L^1(0,T; L^2(\mathcal{D}))$ is a nonlinear operator. W is a cylindrical Wiener process, and B is a dispersion operator of Hilbert–Schmidt type, such that $B : L^2(\mathcal{D}) \to H^{2\gamma,2}(\mathcal{D})$ is a linear isomorphism for some $\gamma > d/4$.[1] Further assumptions on B will be given below in condition (L_B).

In order to formalize the local asymptotics, we define for $\delta > 0$, $x_0 \in \mathcal{D}$ and $Z \in L^2(\mathbb{R}^d)$:

$$\mathcal{D}_{\delta,x_0} := \delta^{-1}(\mathcal{D} - x_0) = \{\delta^{-1}(x - x_0) \mid x \in \mathcal{D}\}, \tag{5.2}$$

$$Z_{\delta,x_0}(x) := \delta^{-d/2} Z(\delta^{-1}(x - x_0)). \tag{5.3}$$

Then $Z_{\delta,x_0} \in L^2(\mathbb{R}^d)$, and $(\cdot)_{\delta,x_0}$ maps $L^2(\mathcal{D}_{\delta,x_0})$ onto $L^2(\mathcal{D})$ with

$$\|Z_{\delta,x_0}\|_{L^2(\mathcal{D})} = \|Z\|_{L^2(\mathcal{D}_{\delta,x_0})} \tag{5.4}$$

for all $\delta > 0$, $x_0 \in \mathcal{D}$. More generally, for any $1 < q < \infty$ and $Z \in L^q(\mathcal{D}_{\delta,x_0})$, (5.3) implies

$$\begin{aligned}
\|Z_{\delta,x_0}\|_{L^q(\mathcal{D})} &= \left(\int_{\mathcal{D}} |Z_{\delta,x_0}(x)|^q \,\mathrm{d}x\right)^{\frac{1}{q}} = \left(\int_{\mathcal{D}} \left(\delta^{-\frac{dq}{4}+\frac{d}{2}} \left(|Z|^{q/2}\right)_{\delta,x_0}(x)\right)^2 \mathrm{d}x\right)^{\frac{1}{q}} \\
&= \delta^{-\frac{d}{2}+\frac{d}{q}} \left\| \left(|Z|^{q/2}\right)_{\delta,x_0} \right\|_{L^2(\mathcal{D})}^{\frac{2}{q}} = \delta^{-\frac{d}{2}+\frac{d}{q}} \left\| |Z|^{q/2} \right\|_{L^2(\mathcal{D}_{\delta,x_0})}^{\frac{2}{q}} \\
&= \delta^{-\frac{d}{2}+\frac{d}{q}} \|Z\|_{L^q(\mathcal{D}_{\delta,x_0})}
\end{aligned} \tag{5.5}$$

for $\delta > 0$, $x_0 \in \mathcal{D}$. Note that the spaces $\mathcal{D}_{\delta,x_0}$ can be seen as non-asymptotic tangential spaces of $\mathcal{D}$ at x_0: Formally, for $\delta \to 0$, the tangential space $T_{x_0}\mathcal{D} \simeq \mathbb{R}^d$ is recovered. The behavior of the fractional Laplacian under localization is given by the following result:

[1] In fact, it suffices to have (5.18) for all $s > 1 + 2\gamma - d/2$.

Lemma 5.1 ([AR21], [ACP20]). *For $s \in \mathbb{R}$, $p \geq 2$, $\delta > 0$, $x_0 \in \mathcal{D}$ and $Z \in H^{s,p}(\mathcal{D}_{\delta,x_0})$, we have*

$$(-\Delta)^{s/2} Z_{\delta,x_0} = \delta^{-s}((-\Delta)^{s/2} Z)_{\delta,x_0}. \tag{5.6}$$

For $s \in 2\mathbb{N}$, this is a consequence of the chain rule.

Here and in the sequel, we fix a kernel $K \in W^{2,2}(\mathbb{R}^d)$ with compact support. We identify K_{δ,x_0}, defined via (5.3), with its restriction to $\mathcal{D}$. Then $K_{\delta,x_0} \in W^{2,2}(\mathcal{D})$ [LM72, Remark 8.1]. For small enough δ, the support of K_{δ,x_0} is compactly contained in $\mathcal{D}$. In particular, the boundary trace operators of K_{δ,x_0} of any differential order are zero on $\mathcal{D}$. Then, clearly, $K_{\delta,x_0} \in H^{2,2}(\mathcal{D})$. W.l.o.g. we restrict to that case in the sequel.

Now, by assumption, we observe a local average of X, namely X tested against K_{δ,x_0}. In addition, we also need X tested against $\Delta K_{\delta,x_0}$:

$$X^K_{\delta,x_0} := \langle X, K_{\delta,x_0}\rangle_{L^2(\mathcal{D})} = \int_{\mathcal{D}} X K_{\delta,x_0} \mathrm{d}x, \tag{5.7}$$

$$X^{\Delta K}_{\delta,x_0} := \langle X, \Delta K_{\delta,x_0}\rangle_{L^2(\mathcal{D})} = \int_{\mathcal{D}} X \Delta K_{\delta,x_0} \mathrm{d}x. \tag{5.8}$$

It is immediate from (5.1) that the dynamics of X^K_{δ,x_0} is determined by the one-dimensional stochastic differential equation

$$\mathrm{d}X^K_{\delta,x_0}(t) = \theta X^{\Delta K}_{\delta,x_0}(t)\mathrm{d}t + \langle F(X)(t), K_{\delta,x_0}\rangle \,\mathrm{d}t + \|B^* K_{\delta,x_0}\|_{L^2(\mathcal{D})} \,\mathrm{d}W^{\delta,x_0}(t) \tag{5.9}$$

with initial condition $X^K_{\delta,x_0}(0) = \langle X_0, K_{\delta,x_0}\rangle_{L^2(\mathcal{D})}$, where the process $W^{\delta,x_0} := \langle B^* K_{\delta,x_0}, W\rangle \,/\, \|B^* K_{\delta,x_0}\|_{L^2(\mathcal{D})}$ is a one-dimensional Brownian motion.

If $X^{\Delta K}_{\delta,x_0}$ and X^K_{δ,x_0} are observed on $[0, T]$, the natural MLE-type estimator, called *augmented MLE* [AR21], is given by

$$\hat{\theta}_{\delta,x_0} = \frac{\int_0^T X^{\Delta K}_{\delta,x_0}(t)\mathrm{d}X^K_{\delta,x_0}(t)}{\int_0^T X^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t}. \tag{5.10}$$

By means of (5.9), $\hat{\theta}_{\delta,x_0}$ can be decomposed as follows:

$$\begin{aligned}\hat{\theta}_{\delta,x_0} - \theta = {} & \frac{\int_0^T X^{\Delta K}_{\delta,x_0}(t) \,\langle F(X)(t), K_{\delta,x_0}\rangle_{L^2(\mathcal{D})} \,\mathrm{d}t}{\int_0^T X^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t} \\ & + \frac{\|B^* K_{\delta,x_0}\|_{L^2(\mathcal{D})} \int_0^T X^{\Delta K}_{\delta,x_0}(t)\mathrm{d}W^{\delta,x_0}(t)}{\int_0^T X^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t}.\end{aligned} \tag{5.11}$$

Note that

$$I_{\delta,x_0} := \|B^* K_{\delta,x_0}\|_{L^2(\mathcal{D})}^{-2} \int_0^T X_{\delta,x_0}^{\Delta K}(t)^2 \mathrm{d}t$$

can be interpreted as the observed Fisher information. We need the following conditions on the dispersion B and the kernel K:[2]

(L_B) There is a family of bounded linear operators $(B_{\delta,x_0})_{\delta\geq 0}$ mapping $L^2(\mathbb{R}^d)$ into itself, such that

$$B^*(-\Delta)^\gamma Z_{\delta,x_0} = (B^*_{\delta,x_0} Z)_{\delta,x_0} \tag{5.12}$$

for $Z \in C^\infty(\mathbb{R}^d)$ with support in $\mathcal{D}_{\delta,x_0}$ and $\delta > 0$, as well as

$$\left\|B^*_{\delta,x_0} Z - B^*_{0,x_0} Z\right\|_{L^2(\mathbb{R}^d)} \to 0 \tag{5.13}$$

for $Z \in L^2(\mathbb{R}^d)$ as $\delta \to 0$.

(L_K) There is $\widetilde{K} \in W^{2\lceil\gamma\rceil+2,2}(\mathbb{R}^d)$ with compact support such that $K = (-\Delta)^{\lceil\gamma\rceil}\widetilde{K}$.

(L_Ψ) We have

$$\left\|B^*_{0,x_0}(-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K}\right\|_{L^2(\mathbb{R}^d)} > 0, \tag{5.14}$$

$$\Psi_\Delta\left((-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K}\right) > 0, \tag{5.15}$$

where $\Psi_\Delta(Z) = \int_0^\infty \left\|B^*_{0,x_0} e^{r\Delta_0}\Delta_0 Z\right\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}r$.

It is straightforward to see that $\Psi_\Delta(Z) < \infty$ for $Z \in W^{2,2}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$: As B^*_{0,x_0} is bounded, we can w.l.o.g. assume that $B^*_{0,x_0} = I$. Let G_t for $t > 0$ be the heat kernel given by $e^{t\Delta_0}Z = G_t * Z$. It is elementary to verify that $\|G_t\|_{L^2(\mathbb{R}^d)} \lesssim t^{-d/4}$, given that G_t is normed in L^1 as a function. Using standard semigroup properties as stated e.g. in [EN00], we see that $\int_0^t \left\|e^{r\Delta_0}\Delta_0 Z\right\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}r = \frac{1}{2}\left\langle e^{2t\Delta_0}Z, \Delta_0 Z\right\rangle_{L^2(\mathbb{R}^d)} + \frac{1}{2}\|\nabla Z\|^2_{L^2(\mathbb{R}^d)}$. Further, using Young's inequality for convolution products, we can estimate $\left\langle e^{2t\Delta_0}Z, \Delta_0 Z\right\rangle_{L^2(\mathbb{R}^d)} \leq \|G_t\|_{L^2(\mathbb{R}^d)} \|Z\|_{L^1(\mathbb{R}^d)} \|\Delta_0 Z\|_{L^2(\mathbb{R}^d)} \lesssim t^{-d/4}$, which converges to zero for $t \to \infty$.

In (5.15), our standing assumption is $(-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K} \in W^{2,2}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. This is certainly the case if $\gamma \in \mathbb{N}$.

[2] These are the assumptions B, K and ND from [ACP20].

Example 5.2. *It is clear that $B = \sigma(-\Delta)^{-\gamma}$ for $\sigma > 0$ satisfies (L_B). In this case, we trivially have $B_{\delta,x_0}Z = \sigma Z$ for all $\delta \geq 0$ and $Z \in L^2(\mathbb{R}^d)$. In addition, the above calculation shows $\Psi_\Delta(Z) = \frac{\sigma^2}{2}\|\nabla Z\|^2_{L^2(\mathbb{R}^d)}$. This can be generalized, for example, to smooth space-dependent $\sigma : \overline{\mathcal{D}} \to [0,\infty)$, see [ACP20, Example 1].*

As before, we reduce the asymptotic analysis of $\hat{\theta}_{\delta,x_0}$ to the linear case $F = 0$ by means of the splitting argument $X = \bar{X} + \widetilde{X}$, where $\bar{X}$ satisfies (5.1) with $F = 0$ and $X_0 = 0$, and $\widetilde{X}$ solves the random PDE (2.5) with initial condition $\widetilde{X}_0 = X_0$. The terms $\bar{X}^{\Delta K}_{\delta,x_0}$ and $\widetilde{X}^{\Delta K}_{\delta,x_0}$ are defined analogously to (5.8).

Lemma 5.3. *Under (L_B), (L_K) and (L_Ψ), we have the following asymptotics as $\delta \to 0$:*

(i) $\int_0^T \bar{X}^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t \asymp C_{\mathrm{loc}}\delta^{-2+4\gamma}$*, where* $C_{\mathrm{loc}} = T\theta^{-1}\Psi_\Delta\left((-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K}\right)$.

(ii) $\|B^* K_{\delta,x_0}\|_{L^2(\mathcal{D})} \asymp C^B_{\mathrm{loc}}\delta^{2\gamma}$*, where* $C^B_{\mathrm{loc}} = \left\|B^*_{0,x_0}(-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K}\right\|_{L^2(\mathbb{R}^d)}$.

In particular, if $F = 0$ and $X_0 = 0$, $I_{\delta,x_0} \asymp (C^B_{\mathrm{loc}})^{-2}C_{\mathrm{loc}}\delta^{-2}$ as $\delta \to 0$.

Proof. (i) is a direct consequence of [ACP20, Proposition 22], and (ii) is shown in the proof of [ACP20, Proposition 2]. □

This lemma is sufficient to obtain the asymptotic properties of $\hat{\theta}_{\delta,x_0}$ in case $F = 0$ and $X_0 = 0$. In order handle the full semilinear model, we need higher regularity of $\widetilde{X}$. However, in contrast to the spectral approach, we can make use of higher L^p–type regularity in the spaces $H^{s,p}(\mathcal{D})$ instead of mere L^2–type regularity. This is further explained in Remark 5.10 below. In order to exploit L^p regularity, we have to modify Condition $(F_{s,\eta})$ as follows:

$(F^p_{s,\eta})$ There is $\epsilon > 0$ and a monotonous, locally bounded function $g : [0,\infty) \to [0,\infty)$, such that for all $Z \in R_p(s)$:

$$\|F(Z)\|_{R_p(s+\eta+\epsilon-2)} \leq g\left(\|Z\|_{R_p(s)}\right). \tag{5.16}$$

In the Markovian case $F(X)(t) = F(X_t)$, (5.16) holds if

$$\|F(Z)\|_{s+\eta+\epsilon-2,p} \lesssim g(\|Z\|_{s,p}) \tag{5.17}$$

for all $Z \in H^{s,p}(\mathcal{D})$.[3]

[3]Condition $(F^p_{s,\eta})$ in the form (5.17) is Assumption A from [ACP20].

Proposition 5.4. *Let $s \in \mathbb{R}$, $p \geq 2$ and $\eta > 0$. Assume that $X_0 \in H^{s+\eta,p}(\mathcal{D})$ and $\bar{X}, \widetilde{X} \in R_p(s)$. If $(F^p_{s,\eta})$ is true, then $\widetilde{X} \in R_p(s+\eta)$.*

Proof. Verbatim as in Proposition 2.3, using the norm $\|\cdot\|_{s,p}$ instead of $\|\cdot\|_s$. □

It remains to understand the regularity of the linear process $\bar{X}$. For $p \geq 2$, set $s^*_p := \sup\{s \in \mathbb{R} \mid \bar{X} \in R_p(s) \text{ a.s.}\}$, and further $s^*_\infty := \inf_{p\geq 2} s^*_p$, as well as $s^* := s^*_2$. As in the previous chapters, we have $s^* = 1 + 2\gamma - d/2$. This can be seen as follows: As $B^*(-\Delta)^\gamma : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ is a linear isomorphism by assumption, we have that

$$\int_0^T \left\|(-\Delta)^{s/2} e^{t\theta\Delta} B\right\|^2_{\mathrm{HS}} \,\mathrm{d}t = \infty \tag{5.18}$$

is true if and only if it is true for B replaced by $(-\Delta)^{-\gamma}$, and this holds for $s \geq 1 + 2\gamma - d/2$. This proves $s^* \leq 1 + 2\gamma - d/2$, and the opposite inequality is shown as in Lemma 2.7. It is clear that $s^*_p \leq s^*$ for all $p \geq 2$ due to $H^{s,p}(\mathcal{D}) \subset H^{s,2}(\mathcal{D})$ for $s \in \mathbb{R}$, and therefore $s^*_\infty \leq s^*$. On the other hand, we have the Sobolev embedding $H^{s,2}(\mathcal{D}) \subset H^{s-d/2,p}(\mathcal{D})$ for all $s \in \mathbb{R}$, $p \geq 2$, and thus $s^*_\infty \geq s^* - d/2$. So $0 \leq s^* - s^*_\infty \leq d/2$ is the possible regularity gap for the linear process $\bar{X}$.

Lemma 5.5. *If $\sup_{k\in\mathbb{N}} \|\Phi_k\|_{L^\infty(\mathcal{D})} < \infty$, then $s^*_\infty = s^*$.*

The proof is given in Appendix B.2. The condition $\sup_{k\in\mathbb{N}} \|\Phi_k\|_{L^\infty(\mathcal{D})} < \infty$ is true e.g. in $d = 1$, but in general, only a bound of the form $\|\Phi_k\|_{L^\infty(\mathcal{D})} \lesssim \lambda_k^{(d-1)/4}$ can be given, and this bound cannot be improved without further restrictions on $\mathcal{D}$, see [Gri02]. Proposition 5.4 together with Lemma 5.5 yields the precise L^p–excess regularity of $\widetilde{X}$:

Proposition 5.6. *Let $\eta > 0$, $s_0 < s^*_\infty$ such that $(F^p_{s,\eta})$ is true for any $s_0 \leq s < s^*$ and $p \geq 2$. Let $X \in R_2(s_0)$ and $X_0 \in H^{s^*+\eta,p}(\mathcal{D})$ for any $p \geq 2$. Then we have $X \in R_p(s)$ and $\widetilde{X} \in R_p(s+\eta)$ for all $s < s^*_\infty$ and $p \geq 2$. In particular, with*

$$\eta_\infty := \eta - (s^* - s^*_\infty), \tag{5.19}$$

we have $\widetilde{X} \in R_p(s+\eta_\infty)$ for all $s < s^*$, $p \geq 2$.

Proof. We distinguish the cases $(s^*_\infty - s_0) \geq d/2$ and $(s^*_\infty - s_0) < d/2$. In the former case, by using Proposition 5.4 iteratively, we have $X \in R_2(s)$ for all $s < s^*$, and by the Sobolev embedding theorem, $X \in R_p(s_0)$ for any $p \geq 2$. Applying Proposition 5.4 a second time repeatedly yields $X \in R_p(s)$ and $\widetilde{X} \in R_p(s+\eta)$ for all $s < s^*_\infty$ and $p \geq 2$, which implies the claim. In the case $(s^*_\infty - s_0) < d/2$, we use a similar inductive argument: If for some $p \geq 2$ it holds that $X \in R_p(s_0)$, then a repeated application of Proposition 5.4 gives $X \in R_p(s)$ for any $s < s^*_\infty \leq s^*_p$, and by the Sobolev embedding theorem, it follows that $X \in R_{p'}(s_0)$ for any $p < p' < dp/(d-2(s^*_\infty - s_0))$. In particular, there is a constant $c > 1$ such that for all $p \geq 2$ we can choose $p' = p'(p)$ in such a way that $p'/p \geq c$. Repeating this step we obtain $X \in R_p(s)$ for all $s < s^*_\infty$ and $p \geq 2$. A final application of Proposition 5.4 yields $\widetilde{X} \in R_p(s+\eta)$ for all $s < s^*_\infty$ and $p \geq 2$, which implies the claim in this case, too. ☐

The excess regularity of $\widetilde{X}$ can be used to show that the terms related to F and $\widetilde{X}$ appearing in (5.11) are of lower order:

Lemma 5.7. *Under the conditions from Proposition 5.6 and (L_K), if $\eta_\infty > 0$, the following is true with $\eta' := \eta_\infty \wedge (1 + d/4)$ and any $\epsilon > 0$:*

(i) It holds

$$\int_0^T \widetilde{X}^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t \lesssim \delta^{-2+4\gamma+2(\eta'-\epsilon)},$$

and in particular

$$\int_0^T X^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t \asymp \int_0^T \bar{X}^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t \asymp C_{\mathrm{loc}} \delta^{-2+4\gamma}. \tag{5.20}$$

(ii) It holds

$$\int_0^T X^{\Delta K}_{\delta,x_0}(t) \langle F(X)(t), K_{\delta,x_0} \rangle_{L^2(\mathcal{D})} \mathrm{d}t \lesssim \delta^{-2+4\gamma+(\eta'-\epsilon)}. \tag{5.21}$$

Proof. With $\widetilde{K}$ as in (L_K), i.e. $K = (-\Delta)^{\lceil \gamma \rceil} \widetilde{K}$, we first show that for any $q > 1$ and $r \geq 0$,

$$\sup_{0<\delta\leq 1} \left\| (-\Delta)^{r/2} \widetilde{K} \right\|_{L^q(\mathcal{D}_{\delta,x_0})} < \infty. \tag{5.22}$$

This is obvious for $r \in 2\mathbb{N}_0$, as $(-\Delta)^{r/2}\widetilde{K}$ has compact support in this case, and in particular, $\sup_{0<\delta\leq 1} \left\|(-\Delta)^{r/2}\widetilde{K}\right\|_{L^q(\mathcal{D}_{\delta,x_0})} = \left\|(-\Delta)^{r/2}\widetilde{K}\right\|_{L^q(\mathbb{R}^d)} < \infty$. For general $r \geq 0$, we use the Gagliardo–Nirenberg inequality [BM18] on the reference domain $\mathcal{D}$: For $0 < \delta \leq 1$, with $R := 2\lfloor r/2 \rfloor \in 2\mathbb{N}_0$,

$$
\begin{aligned}
\left\|(-\Delta)^{r/2}\widetilde{K}\right\|_{L^q(\mathcal{D}_{\delta,x_0})} &= \delta^{\frac{d}{2}-\frac{d}{q}+r} \left\|(-\Delta)^{r/2}\widetilde{K}_{\delta,x_0}\right\|_{L^q(\mathcal{D})} \\
&\lesssim \delta^{\frac{d}{2}-\frac{d}{q}+r} \left\|(-\Delta)^{R/2}\widetilde{K}_{\delta,x_0}\right\|_{L^q(\mathcal{D})}^{\frac{2-r+R}{2}} \left\|(-\Delta)^{R/2+1}\widetilde{K}_{\delta,x_0}\right\|_{L^q(\mathcal{D})}^{\frac{r-R}{2}} \\
&\lesssim \left\|(-\Delta)^{R/2}\widetilde{K}\right\|_{L^q(\mathcal{D}_{\delta,x_0})}^{\frac{2-r+R}{2}} \left\|(-\Delta)^{R/2+1}\widetilde{K}\right\|_{L^q(\mathcal{D}_{\delta,x_0})}^{\frac{r-R}{2}}
\end{aligned}
$$

where we used (5.5) and Lemma 5.1 repeatedly. Consequently, using that (5.22) is true for the exponents R and $R+2$, we obtain that (5.22) is true for all $r \geq 0$.

Next, by choice of η', we have $\eta' < 1 + d/2$, thus $2\gamma + 2 - s^* - \eta' > 0$, and therefore $\lceil\gamma\rceil + 1 - (s+\eta')/2 > 0$ for all $s < s^*$. Using (5.5) and (5.22),

$$
\begin{aligned}
\left\|\left((-\Delta)^{1-(s+\eta')/2}K\right)_{\delta,x_0}\right\|_{L^q(\mathcal{D})} &= \delta^{-\frac{d}{2}+\frac{d}{q}} \left\|(-\Delta)^{\lceil\gamma\rceil+1-(s+\eta')/2}\widetilde{K}\right\|_{L^q(\mathcal{D}_{\delta,x_0})} \\
&\lesssim \delta^{-\frac{d}{2}+\frac{d}{q}}. \qquad (5.23)
\end{aligned}
$$

After these preparations, we can prove the statements of the lemma.

(i) Now let $s < s^*$ and $p \geq 2$, with $q = p/(p-1) \leq 2$. By Proposition 5.6, Lemma 5.1 and (5.23), we have for all $s < s^*$

$$
\begin{aligned}
\int_0^T \widetilde{X}^{\Delta K}_{\delta,x_0}(t)^2 \mathrm{d}t &= \int_0^T \left\langle \widetilde{X}_t, \Delta K_{\delta,x_0}\right\rangle_{L^2(\mathcal{D})}^2 \mathrm{d}t \\
&= \int_0^T \left\langle (-\Delta)^{(s+\eta')/2}\widetilde{X}, (-\Delta)^{1-(s+\eta')/2}K_{\delta,x_0}\right\rangle_{L^2(\mathcal{D})}^2 \mathrm{d}t \\
&\leq T \sup_{0\leq t\leq T} \left\|\widetilde{X}\right\|_{s+\eta',p}^2 \left\|(-\Delta)^{1-(s+\eta')/2}K_{\delta,x_0}\right\|_{L^q(\mathcal{D})}^2 \\
&\lesssim \delta^{-4+2(s+\eta')} \left\|\left((-\Delta)^{1-(s+\eta')/2}K\right)_{\delta,x_0}\right\|_{L^q(\mathcal{D})}^2 \\
&\lesssim \delta^{-4+2(s+\eta')-d+2d/q}.
\end{aligned}
$$

Now, for any $\epsilon' > 0$, with $s := s^* - \epsilon' = 1 + 2\gamma - d/2 - \epsilon'$ and $p := 2d/\epsilon'$, such that $1 - 1/q = 1/p = \epsilon'/(2d)$, we have

$$\int_0^T \widetilde{X}_{\delta,x_0}^{\Delta K}(t)^2 \mathrm{d}t \lesssim \delta^{-2+4\gamma+2\eta'-2d(1-1/q)-2\epsilon'} = \delta^{-2+4\gamma+2\eta'-3\epsilon'},$$

which implies the claim with $\epsilon = 3\epsilon'/2$, using Lemma 5.3.

(ii) Condition $(F^p_{s,\eta})$ together with Proposition 5.6 gives $F(X) \in R_p(s+\eta-2)$ for all $s < s^*_\infty$ and $p \geq 2$, which is the same as $F(X) \in R_p(s+\eta_\infty-2)$ for all $s < s^*$ and $p \geq 2$. Let $\epsilon' > 0$, $s = s^* - \epsilon'$, $p = 2d/\epsilon'$ and $q = p/(p-1)$. Similar as in (i), we estimate

$$\begin{aligned}
&\int_0^T \langle F(X)(t), K_{\delta,x_0}\rangle^2_{L^2(\mathcal{D})} \,\mathrm{d}t \\
&\qquad = \int_0^T \left\langle (-\Delta)^{\frac{s+\eta'-2}{2}} F(X)(t), (-\Delta)^{-\frac{s+\eta'-2}{2}} K_{\delta,x_0} \right\rangle^2_{L^2(\mathcal{D})} \mathrm{d}t \\
&\qquad \lesssim T \, \|F(X)\|^2_{R_p(s+\eta'-2)} \left\| (-\Delta)^{1-\frac{s+\eta'}{2}} K_{\delta,x_0} \right\|^2_{L^q(\mathcal{D})} \\
&\qquad \lesssim \delta^{-2+4\gamma+2\eta'-3\epsilon'}.
\end{aligned}$$

Finally, an application of the Hölder inequality in time gives (5.21).

□

We can now state and prove the main result of this chapter:

Theorem 5.8. *Let $x_0 \in \mathcal{D}$. Assume that (L_B), (L_K) and (L_Ψ) hold. Let $\eta > 0$, $s_0 < s^*$ such that $(F^p_{s,\eta})$ is satisfied for all $s_0 \leq s < s^*$ and $p \geq 2$. Let a.s. $X \in R_2(s_0)$ and $X_0 \in H^{s^*+\eta,p}(\mathcal{D})$ for all $p \geq 2$. If $\eta_\infty = \eta - (s^* - s^*_\infty) > 0$, then the following is true:*

(i) $\hat{\theta}_{\delta,x_0}$ is a consistent estimator for θ, i.e. $\hat{\theta}_{\delta,x_0} \xrightarrow{\mathbb{P}} \theta$ as $\delta \to 0$.

(ii) If $\eta_\infty > 1$, then

$$\delta^{-1}\left(\hat{\theta}_{\delta,x_0} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\mathrm{loc}}\right), \tag{5.24}$$

where

$$\Sigma_{\mathrm{loc}} = \frac{\theta \left\| B^*_{0,x_0}(-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K}\right\|^2_{L^2(\mathbb{R}^d)}}{T\Psi_\Delta\left((-\Delta_0)^{\lceil\gamma\rceil-\gamma}\widetilde{K}\right)}. \tag{5.25}$$

In the case $\eta_\infty \le 1$, it holds

$$\delta^{-a}\left(\hat{\theta}_{\delta,x_0} - \theta\right) \xrightarrow{\mathbb{P}} 0 \tag{5.26}$$

for all $a < \eta_\infty$.

Proof. We use the representation of $\hat{\theta}_{\delta,x_0} - \theta$ as given in (5.11). First, we set $M_T^{(\delta)} := C_{\mathrm{loc}}^{-1/2}\delta^{1-2\gamma}\int_0^T X^{\Delta K}_{\delta,x_0}(t)\mathrm{d}W^{\delta,x_0}(t)$. Then $\langle M^{(\delta)}\rangle_T \to 1$ in probability as $\delta \to 0$, and Theorem A.1 implies that $M_T^{(\delta)} \to \mathcal{N}(0,1)$ in distribution. An application of Slutsky's lemma together with Lemma 5.3 (ii) and Lemma 5.7 (i) gives

$$\frac{\|B^*K_{\delta,x_0}\|_{L^2(\mathcal{D})}\int_0^T X^{\Delta K}_{\delta,x_0}(t)\mathrm{d}W^{\delta,x_0}(t)}{\int_0^T X^{\Delta K}_{\delta,x_0}(t)^2\mathrm{d}t} \xrightarrow{d} \mathcal{N}\left(0,\Sigma_{\mathrm{loc}}\right).$$

Next, again by Lemma 5.7, for any $\epsilon > 0$,

$$\frac{\int_0^T X^{\Delta K}_{\delta,x_0}(t)\left\langle F(X)(t), K_{\delta,x_0}\right\rangle_{L^2(\mathcal{D})}\mathrm{d}t}{\int_0^T X^{\Delta K}_{\delta,x_0}(t)^2\mathrm{d}t} \lesssim \delta^{\eta_\infty\wedge(1+d/4)-\epsilon}.$$

Using (5.11), this implies (ii). Finally, (i) is a consequence of (ii). □

As an important example, if $B = \sigma(-\Delta)^{-\gamma}$ for some $\sigma > 0$ and $\gamma \in \mathbb{N}$, then we immediately obtain due to $B_{0,x_0}Z = \sigma Z$ for $Z \in L^2(\mathbb{R}^d)$:

$$\Sigma_{\mathrm{loc}} = \frac{2\theta\left\|\widetilde{K}\right\|^2_{L^2(\mathbb{R}^d)}}{T\left\|\nabla\widetilde{K}\right\|^2_{L^2(\mathbb{R}^d)}}. \tag{5.27}$$

Example 5.9. *The results from Theorem 5.8 can be applied to the models from Section 2.4. For simplicity, we assume that $s^*_\infty = s^*$ (which is true, for example, in d = 1). In this case, the effective excess regularity η_∞ coincides with the optimal excess regularity η.*

(i) **Linear Perturbations.** *If $F(X) = c(-\Delta)^{r/2}$ for some $c \in \mathbb{R}$, $r < 2$, then $(F^p_{s,\eta})$ is true for all $s \in \mathbb{R}$, $p \geq 2$ and $\eta < 2 - r$. Thus, if $r < 1$, then $\hat{\theta}_{\delta,x_0}$ is asymptotically normal as in* (5.24)*. Otherwise, $\hat{\theta}_{\delta,x_0}$ is consistent as in* (5.26)*. In particular, perturbations up to order* 1 *are negligible, with first order perturbations being the critical case.*

(ii) **Reaction–Diffusion Equations.** *If $F(X) = f(x)$, where f is either a polynomial as in* (2.40) *or a bounded smooth function with bounded derivatives of any order as in* (2.41)*, then $(F^p_{s,\eta})$ holds for any $p \geq 2$, $s > d/p$ and $\eta < 2$. This can be seen verbatim as in Proposition 2.19, using the fact that $H^{s,p}(\mathcal{D})$ is closed under multiplication if $sp > d$ (cf. [AF03, Theorem 4.39]) for the case of polynomial f, as well as bounds on composition operators [AF92] for the case $f \in C^\infty_b(\mathbb{R})$. Consequently, $\hat{\theta}_{\delta,x_0}$ is asymptotically normal as in* (5.24)*.*

(iii) **Burgers Equations.** *In $d = 1$, if $F(X) = -X\partial_x X = -\partial_x(X^2/2)$, then exactly as in Lemma 2.22 (iii) it can be shown that $(F^p_{s,\eta})$ is true for $p \geq 2$, $s > 1/p$ and $\eta < 1$. In particular, $\hat{\theta}_{\delta,x_0}$ is consistent as in* (5.26)*, i.e. $\delta^{-a}(\hat{\theta}_{\delta,x_0} - \theta) \to 0$ in probability for all $a < 1$. In fact, for this particular model, it is possible to prove that the first term in* (5.11)*, representing the bias from neglecting the effect of F, converges to zero in probability even with rate δ instead of δ^a for $a < 1$, which means that asymptotic normality as in* (5.24) *transfers to $\hat{\theta}_{\delta,x_0}$ for the one–dimensional Burgers equation (see [ACP20, Theorem 11] for details).*

Remark 5.10 (L^p–regularity in the spectral approach)**.** *L^p–regularity has been a crucial tool in determining the excess regularity η of $\widetilde{X}$ in the local approach. It is a natural question to ask if L^p–regularity can improve the spectral approach as well. In order to make the two approaches comparable, we assume that a single Fourier mode of X (instead of the first N Fourier modes) is observed in the spectral approach. For simplicity, let $F(X) = (-\Delta)^r$, $r < 2$ and $B = \sigma(-\Delta)^{-\gamma}$. The natural MLE–type estimator, neglecting information on F, is given by*

$$\hat{\theta}^{\mathrm{mode}}_N = -\frac{\int_0^T \lambda_N^{1+2\gamma} x_t^{(N)} \mathrm{d}x_t^{(N)}}{\int_0^T \lambda_N^{2+2\gamma} (x_t^{(N)})^2 \mathrm{d}t} = -\frac{\int_0^T x_t^{(N)} \mathrm{d}x_t^{(N)}}{\lambda_N \int_0^T (x_t^{(N)})^2 \mathrm{d}t} \tag{5.28}$$

with $x^{(N)} = \langle X, \Phi_N \rangle_{L^2(\mathcal{D})}$ as before. This is the canonical analogon to $\hat{\theta}^{\mathrm{lin}}_N$ (with $\alpha = \gamma$) for single mode observations, and it corresponds to $\hat{\theta}_{\delta,x_0}$ if

K_{δ,x_0} is replaced by Φ_N. By (2.22) *with respect to the linear operator $A = \theta\Delta + (-\Delta)^r$, together with Lemma A.2 (i) (setting $X_k^* = \lambda_k^{1+\gamma} x^{(k)}$ therein and taking into account* $\mathrm{Var} \int_0^T (x_t^{(k)})^2 \mathrm{d}t \lesssim \lambda_k^{-4\gamma-3}$, *see e.g. [Lot09, Theorem 2.1] or [PS20, Lemma 4.1]), we have* $\lambda_N^{2+2\gamma} \int_0^T (x_t^{(N)})^2 \mathrm{d}t \asymp (\sigma^2 T\Lambda/2\theta) N^{2/d}$ *in probability. In particular, if there is $\eta' > 0$ such that*

$$\lambda_N^{2\gamma} \int_0^T \langle F(X)(t), \Phi_N \rangle_{L^2(\mathcal{D})}^2 \, \mathrm{d}t \lesssim N^{\frac{2}{d} - 2\eta'}, \tag{5.29}$$

a decomposition as in (5.11) *yields*

$$\text{If } \eta' > 1/d: \quad N^{\frac{1}{d}} \left(\hat{\theta}_N^{\mathrm{mode}} - \theta \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\theta}{T\Lambda}\right), \tag{5.30}$$

$$\text{if } \eta' \leq 1/d: \quad N^{a} \left(\hat{\theta}_N^{\mathrm{mode}} - \theta \right) \xrightarrow{\mathbb{P}} 0 \quad \text{for } a < \eta'. \tag{5.31}$$

We have that $(F_{s,\eta}^p)$ is true for any $s \in \mathbb{R}$, $p \geq 2$ and $\eta < 2 - r$, thus $F(X) \in R_p(s-r)$ for all $s < s^*$ and $p \geq 2$ (if $s_\infty^* = s^*$). Now, analogously to the proof of Lemma 5.7, the Hölder inequality gives

$$\lambda_N^{2\gamma} \int_0^T \langle F(X)(t), \Phi_N \rangle_{L^2(\mathcal{D})}^2 \, \mathrm{d}t \lesssim N^{\frac{2}{d} + 1 - \frac{2\eta}{d} + \epsilon} \left\| \Phi_N \right\|_{L^q(\mathcal{D})}^2$$

for all $q > 1$ and $\epsilon > 0$, so with $q = 2$ we can choose $\eta' = \eta/d - 1/2 - \epsilon/2$ in (5.29)*. In particular, if $\eta > 1 + 1/\beta = 1 + d/2$, then $\eta' > 1/d$, and $\hat{\theta}_N^{\mathrm{mode}}$ is asymptotically normal, in accordance with Theorem 2.11.*

Now, in order to exploit higher L^p–regularity of X in the spectral approach, the term $\left\| \Phi_N \right\|_{L^q(\mathcal{D})}$ has to converge to zero if $N \to \infty$, with convergence rate possibly depending on $1 \leq q \leq 2$. By interpolation, it suffices to understand the border case $q = 1$ in order to obtain a bound on the rate. But in general, such convergence to zero does not hold: For example, if $d = 1$ *and $\mathcal{D} = [0,1]$, we have $\Phi_N(x) = \sqrt{2}\sin(N\pi x)$, and $\left\| \Phi_N \right\|_{L^1(\mathcal{D})} = 2\sqrt{2}/\pi$ is independent of N. For general domains, results from literature on L^1–bounds for the eigenvalues of the Laplacian [vdBHV15, Vog15] do not improve the trivial estimate $\sup_{N \in \mathbb{N}} \left\| \Phi_N \right\|_{L^1(\mathcal{D})} < \infty$ in terms of the convergence rate in N. Even more, [Sog15, BS17] indicate that improved L^1–bounds (along eigenvalue subsequences) on compact manifolds without boundary are related to concentration of mass of the eigenfunctions along geodesics.* [4]

[4]Note the apparent asymmetry between the (optimal) bound $\left\| \Phi_N \right\|_{L^\infty(\mathcal{D})} \lesssim \lambda_N^{(d-1)/4}$

In contrast, $\|K_{\delta,x_0}\|_{L^q(\mathcal{D})} \lesssim \delta^{-\frac{d}{2}+\frac{d}{q}}$ *by* (5.5) *if* K *has compact support, which improves the convergence rate in* δ *as* $q \to 1$.

Remark 5.11.

(i) If the linear process $\bar{X}$ *has optimal* L^p*–regularity, i.e. if* $s_\infty^* = s^*$*, then* $\eta_\infty = \eta$*, and the convergence rate of* $\hat{\theta}_{\delta,x_0}$ *in Theorem 5.8 is determined directly by the excess regularity* η *coming from* $(F_{s,\eta}^p)$.

(ii) In view of condition (L_K)*, it is natural to consider also the case* $\gamma = 0$*. Indeed, it is possible to include that case without further modification as long as* (5.1) *is well-posed. This has been done in [ABJR21] in the context of the stochastic Meinhardt model. For instance, the results from [DPZ14, Chapter 7] show that in* $d = 1$*, reaction-diffusion equations driven by space-time white noise can be given a meaning in the mild sense. If* $\gamma = 0$*, then condition* (L_K) *is void, whereas for positive* γ*,* (L_K) *suggests that* K *approximates higher order derivatives of* X *instead of point evaluations.*

(iii) The convergence rate δ *of* $\hat{\theta}_{\delta,x_0}$ *can be recovered from the spectral approach: We can easily see how the asymptotic variance* Σ *from Theorem 2.11 for the estimator* $\hat{\theta}_N^{\text{full}}$ *behaves if the domain* $\mathcal{D}$ *is replaced by a shrunk domain* $\mathcal{D}_{1/\delta,x_0}$*.* Σ *depends linearly on* Λ^{-1}*, and this constant is characterized by* $\lambda_k \asymp \Lambda k^{2/d}$*. An explicit term for* Λ *can be found e.g. in [Shu01, Section 13.4], and using the notation therein, it holds* $\Lambda = V_1^{-2/d}$*, where* V_1 *depends linearly on* $|\mathcal{D}|$*. It is clear that* $\left|\mathcal{D}_{1/\delta,x_0}\right| \sim \delta^d |\mathcal{D}|$*. Consequently,* $\Lambda \sim \delta^{-2}$*, and finally,* $\Sigma \sim \delta^2$*. This is in accordance with Theorem 5.8.*

[Gri02] and the trivial bound $\|\Phi_N\|_{L^1(\mathcal{D})} \lesssim 1$. If the latter cannot be improved, this means that it is not possible to recover $\|\Phi_N\|_{L^2(\mathcal{D})} \equiv 1$ from interpolating the L^1 and L^∞ cases. Indeed, as pointed out in [Gri92, SS07] (see also references therein), for $p \geq 2$, optimal L^p–bounds for (linear combinations of) Φ_N do not simply follow from interpolation between the L^2 and L^∞ cases.

Chapter 6

Diffusivity Estimation for Activator-Inhibitor Models

This chapter is an adaptation of material from [PFA+21].

We apply the theory of parameter estimation for semilinear SPDEs to a particular test case from cell biology, concerning the dynamical behavior of actin concentration in *D. discoideum* giant cells. The actin cytoskeleton plays a crucial role in different processes such as motility of amoeboid cells [BBPSP14]. In spite of its complex filamental structure, at the length scale of the cell itself it may be reasonably approximated by a scalar field, representing concentration. Intracellular actin is capable of generating traveling waves. In [FFAB20], this has been described by an SPDE of FitzHugh–Nagumo type, which is coupled to a phase field representing the boundary of the cell. In order to increase the spatial resolution of experimental data, it is possible to artificially merge various cells to form a so-called giant cell, see [GEW+14]. In particular, this allows to observe the spatiotemporal actin dynamics within a cell away from the cell boundary. In this case, the describing model can be simplified by neglecting the phase field.

The reaction model employed in [FFAB20] in order to describe the actin dynamics is a minimal model capable of generating traveling waves rather than a detailed representation of the biochemical reaction pathway. Consequently, it is natural to ask to what extent the true dynamics is described by that model. In order to provide a first step towards answering that question, we extend the theory of diffusivity estimation for semilinear SPDEs from Chapter 2, including simultaneous estimation of reaction parameters. To this end, we assume that the nonlinearity is given as a parametrized term.

This can be interpreted as qualitative a priori knowledge on the behavior of the reaction terms, without knowing the magnitude of the involved parameters quantitatively. Based on these considerations, we compare the effective diffusivity, given as the value of either of different related estimators, on simulated and experimental data, in order to understand the effects of the reaction model.

In Section 6.1, we discuss joint diffusivity and reaction parameter estimation for semilinear SPDEs, extending the results from Chapter 2. We put special emphasis on the statistically linear case, i.e. when the nonlinearity depends linearly on its parameters. Finally, we state and discuss the regularity properties of an activator-inhibitor model, which is closely related to [FFAB20]. In Section 6.2, we apply the estimation theory from Section 6.1 to simulated and real data described by that activator-inhibitor model.

6.1 Joint Diffusivity and Reaction Parameter Estimation

6.1.1 The General Case

We extend the model from Chapter 2 by allowing the nonlinear term F to depend on additional parameters $\theta_1, \dots, \theta_K$, $K > 0$, which we call reaction parameters in the sequel:

$$\mathrm{d}X_t = \theta_0 A X_t \mathrm{d}t + F_{\theta_{1:K}}(X)(t)\mathrm{d}t + B(X_t)\mathrm{d}W_t \tag{6.1}$$

with initial condition X_0, where we write $\theta_{1:K} = (\theta_1, \dots, \theta_K)$ for short. Further, we write $\theta = (\theta_0, \dots, \theta_K)$ for the complete parameter vector. We fix the parameter space $\Theta \subset \mathbb{R}^{K+1}$, which encodes our usual standing assumption $\theta_0 > 0$, as well as possible restrictions on the reaction parameters coming from the bifurcation structure of (6.1) together with a priori knowledge on the dynamical regime. It is possible that an estimator for θ takes values outside Θ, in that case it should be considered void. The dispersion operator B is assumed to satisfy (N_η^γ) for some $\gamma > d/4$ and $\eta > 0$. In order to take the reaction parameters into account when controlling the nonlinearity in the drift term, we have to extend condition $(F_{s,\eta})$:

($F^{\mathrm{par}}_{s,\eta}$) There are continuous functions $g : [0,\infty) \to [0,\infty)$ and $c : \mathbb{R}^K \to [0,\infty)$ and there is $\epsilon > 0$ such that for $Z \in R(s)$:

$$\|F_{\theta_{1:K}}(Z)\|_{R(s+\eta+\epsilon-2)} \leq c(\theta_1,\ldots,\theta_K) g\left(\|Z\|_{R(s)}\right). \tag{6.2}$$

W.l.o.g. we always assume that g is increasing. The non-Markovian nature of F is crucial in our main example, as explained in Section 6.1.3.

Setting formally $B = \sigma(-A)^{-\gamma}$, by [LS77, Section 7.6.4], the log-likelihood for X^N is heuristically given by

$$\begin{aligned}
\ln \frac{\mathrm{d}\mathbb{P}^{N,T}_{\theta}}{\mathrm{d}\mathbb{P}^{N,T}_{\bar\theta}} &= \frac{1}{\sigma^2}\int_0^T \left\langle (\theta_0 - \bar\theta_0) A X^N_t, (-A)^{2\gamma} \mathrm{d}X^N_t \right\rangle \\
&\quad + \frac{1}{\sigma^2}\int_0^T \left\langle P_N F_{\theta_{1:K}}(X)(t) - P_N F_{\bar\theta_{1:K}}(X)(t), (-A)^{2\gamma} \mathrm{d}X^N_t \right\rangle \\
&\quad - \frac{1}{2\sigma^2}\int_0^T \left\langle (\theta_0 - \bar\theta_0) A X^N_t + P_N F_{\theta_{1:K}}(X)(t) - P_N F_{\bar\theta_{1:K}}(X)(t), \right. \\
&\qquad \left. (-A)^{2\gamma}\left((\theta_0 + \bar\theta_0) A X^N_t + P_N F_{\theta_{1:K}}(X)(t) + P_N F_{\bar\theta_{1:K}}(X)(t)\right)\right\rangle \mathrm{d}t,
\end{aligned}$$

where $\bar\theta = (\bar\theta_0,\ldots,\bar\theta_K) \in \Theta$ is an arbitrary reference parameter.

Maximizing this term for $\theta_0,\ldots,\theta_K$ simultaneously leads to the corresponding likelihood equations

$$\begin{aligned}
&-\int_0^T \left\langle (-A)^{1+2\alpha} X^N_t, \mathrm{d}X^N_t \right\rangle \\
&\qquad = \int_0^T \left\langle (-A)^{1+2\alpha} X^N_t, \theta_0(-A)X^N_t - P_N F_{\theta_{1:K}}(X)(t) \right\rangle \mathrm{d}t, \\
&\int_0^T \left\langle \partial_{\theta_i} P_N F_{\theta_{1:K}}(X)(t), (-A)^{2\alpha} \mathrm{d}X^N_t \right\rangle \\
&\qquad = -\int_0^T \left\langle (-A)^{2\alpha} \partial_{\theta_i} P_N F_{\theta_{1:K}}(X)(t), \theta_0(-A)X^N_t - P_N F_{\theta_{1:K}}(X)(t) \right\rangle \mathrm{d}t
\end{aligned}$$

for $i = 1,\ldots,K$, where as before we substituted γ by a free parameter α. Without further mentioning it, we assume that there is a solution to these equations. We fix any solution and call it a maximum likelihood–type estimator $\hat\theta^N$ for our problem, with components $\hat\theta^N_0,\ldots,\hat\theta^N_K$. In general, the likelihood equations cannot be solved explicitly.

Now, depending on the specific form of F as well as the eigenvalue asymptotics of A, it may happen that not all reaction parameters (if any) are identifiable in finite time. This means that $\hat{\theta}_i^N$ does not necessarily converge to θ_i if $N \to \infty$ if $i \geq 1$. Therefore, we put our main focus on diffusivity estimation, i.e. identifying θ_0, and analyze the impact of the reaction parameters on that problem. In order to control $\hat{\theta}_{1:K}^N$ when studying the asymptotics of $\hat{\theta}_0^N$, it suffices that the reaction parameter estimators are *bounded in probability* (or *tight*), i.e. $\sup_{N\in\mathbb{N}} \mathbb{P}(|\hat{\theta}_i^N| > M) \to 0$ for $M \to \infty$ and all $1 \leq i \leq K$. From the likelihood equations it is clear that $\hat{\theta}_0^N$ can be written as

$$\hat{\theta}_0^N = -\frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, \mathrm{d}X_t^N \right\rangle}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|_H^2 \, \mathrm{d}t} + \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F_{\hat{\theta}_{1:K}^N}(X)(t) \right\rangle \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|_H^2 \, \mathrm{d}t}, \tag{6.3}$$

even if this is not explicit due to the presence of $\hat{\theta}_{1:K}^N$.

Theorem 6.1. *Let $\gamma > 1/(2\beta)$ and further $s_0 \geq 0$, $\eta > 0 \vee (1/\beta - 1)$ such that (N_η^γ) and $(F_{s,\eta}^{\mathrm{par}})$ for $s_0 \leq s < s^*$ are true. Assume $X_0 \in H_{s^*+\eta}$ and $X \in R(s_0)$. Let $\alpha > \gamma - 1/4$, and assume that $(\hat{\theta}_i^N)_{N\in\mathbb{N}}$ are bounded in probability for $1 \leq i \leq K$. Then:*

(i) If $\eta > 1 + 1/\beta$, then $\hat{\theta}_0^N$ is asymptotically normal as in (2.29).

(ii) If $\eta \leq 1 + 1/\beta$, then $\hat{\theta}_0^N$ is consistent in probability with rate N^{-a}, $a < \beta\eta/2$, as in (3.70).

Proof. By Theorem 3.22, the claim is true with $\hat{\theta}_0^N$ replaced by $\hat{\theta}_N^{\mathrm{lin}}$ given by (2.26). Thus, by (6.3) it suffices to control the bias term involving $F_{\hat{\theta}_{1:K}^N}(X)$. As in the proof of Theorem 2.11, we see that by means of $(F_{s,\eta}^{\mathrm{par}})$,

$$\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F_{\hat{\theta}_{1:K}^N}(X)(t) \right\rangle \mathrm{d}t \ll_p c(\hat{\theta}_{1:K}^N) N^{1+\beta(2\alpha-2\gamma+1-\eta/2)},$$

and consequently, for (i),

$$N^{\frac{1+\beta}{2}} \frac{\int_0^T \left\langle (-A)^{1+2\alpha} X_t^N, P_N F_{\hat{\theta}_{1:K}^N}(X)(t) \right\rangle \mathrm{d}t}{\int_0^T \|(-A)^{1+\alpha} X_t^N\|_H^2 \, \mathrm{d}t} \ll_p c(\hat{\theta}_{1:K}^N) N^{\beta(1+1/\beta-\eta)/2}$$

almost surely. The right-hand side converges to zero in probability because $\eta > 1 + 1/\beta$ and $\hat{\theta}_{1:K}^N$ are bounded in probability. This implies (i). The case (ii) is similar. □

6.1.2 The Statistically Linear Case

In this section, let F depend linearly on its parameters:[1]

$$F_{\theta_{1:K}}(X) = F^*(X) + \sum_{i=1}^{K} \theta_i F_i(X) \tag{6.4}$$

for functions $F_1, \dots, F_K, F^* : C(0,T;H) \supset D(F) \to L^1(0,T;H)$. We set $D_\alpha(F) := \{Z \in D(F) \mid F_1(Z), \dots, F_K(Z) \in L^2(0,T;H_{2\alpha})\}$. Identifiability of $\theta_1, \dots, \theta_K$ is ensured by a non-degeneracy condition:

(I_α) The terms $F_1(Z), \dots, F_K(Z)$ are linearly independent in $L^2(0,T;H_{2\alpha})$ for every $Z \in D_\alpha(F)$ which is not constant in $t \in [0,T]$.

As a consequence of (I_α),

$$\int_0^T \|(-A)^\alpha F_i(X)(t)\|_H^2 \, \mathrm{d}t > 0 \tag{6.5}$$

for $i = 1, \dots, K$. In order to unify notation in the sequel, we define

$$F_0(X) := AX. \tag{6.6}$$

The maximum likelihood equations simplify to

$$A_N^{(\alpha)}(X)\hat{\theta}^N = b_N^{(\alpha)}(X), \tag{6.7}$$

where

$$\begin{aligned}
A_N^{(\alpha)}(X)_{i,j} &= \int_0^T \langle (-A)^\alpha P_N F_i(X)(t), (-A)^\alpha P_N F_j(X)(t) \rangle \, \mathrm{d}t, \\
b_N^{(\alpha)}(X)_i &= -\int_0^T \left\langle (-A)^{2\alpha} P_N F_i(X)(t), P_N F^*(X)(t) \right\rangle \mathrm{d}t \\
&\quad + \int_0^T \left\langle (-A)^{2\alpha} P_N F_i(X)(t), \mathrm{d}X_t^N \right\rangle .
\end{aligned}$$

[1] A comparable setup has been considered in [Hue93, Chapter 3] for linear SPDEs in the spectral approach with similar arguments as given below, and in [DMPD00, Section 3] for semilinear SPDEs in the large time regime.

In particular,

$$A_N^{(\alpha)}(X)_{0,0} = \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|_H^2 \,\mathrm{d}t. \tag{6.8}$$

Further, it is immediate that for $i, j = 0, \dots, K$:

$$\left|A_N^{(\alpha)}(X)_{i,j}\right| \leq \sqrt{A_N^{(\alpha)}(X)_{i,i} A_N^{(\alpha)}(X)_{j,j}}. \tag{6.9}$$

In order to connect to Theorem 6.1 and prove that the reaction parameter estimators $\hat{\theta}_{1:K}$ are bounded in probability, we have to control the rate of the determinant of $A_N^{(\alpha)}(X)$, whose square root is the volume of the $(K+1)$-dimensional parallelepiped spanned by $P_N F_0(X), \dots P_N F_K(X)$ in $L^2(0, T; H_{2\alpha})$. In order to do so, we choose α in such a way that $P_N F_0(X) = AX^N$ diverges in $L^2(0, T; H_{2\alpha})$, while $P_N F_1(X), \dots, P_N F_K(X)$ converge. This way, AX^N gets asymptotically orthogonal to the latter terms and determines the rate of volume growth. This is formalized in the next lemma.

Lemma 6.2. *Let $s_0 \geq 0$, $\eta > 0$ such that $X \in R(s_0)$, $X_0 \in H_{s^*+\eta}$ a.s. and (N_η^γ) as well as $(F_{s,\eta})$ for $s_0 \leq s < s^*$ are true. Let $\alpha \in \mathbb{R}$ such that $\gamma - (1 + 1/\beta)/2 < \alpha < \gamma + (\eta - 1 - 1/\beta)/2$. Under condition (I_α), there are $N_0 \in \mathbb{N}$ and $c, C > 0$ such that*

$$c \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|_H^2 \,\mathrm{d}t \leq \left|\det\left(A_N^{(\alpha)}(X)\right)\right| \leq C \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|_H^2 \,\mathrm{d}t$$

uniformly in $N \geq N_0$, almost surely. In particular,

$$\left|\det\left(A_N^{(\alpha)}(X)\right)\right| \sim N^{1+\beta(2\alpha - 2\gamma + 1)}. \tag{6.10}$$

Proof. First, $\alpha < \gamma + (\eta - 1 - 1/\beta)/2$ implies $2\alpha < s^* + \eta - 2$, thus $F_i(X) \in R(2\alpha)$ for $i = 1, \dots, K$. Thus, for these i, we have $\lim_{N\to\infty} A_N^{(\alpha)}(X)_{i,i} = \int_0^T \|(-A)^\alpha F_i(X)(t)\|_H^2 \,\mathrm{d}t < \infty$. In particular, $A_N^{(\alpha)}(X)_{i,i}$ are positive and finite for $i = 1, \dots, K$ and large enough N. Using (6.9), we have

$$\begin{aligned}
\left|\det\left(A_N^{(\alpha)}(X)\right)\right| &\leq (K+1)! \prod_{i=0}^K A_N^{(\alpha)}(X)_{i,i} \\
&\lesssim A_N^{(\alpha)}(X)_{0,0} = \int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|_H^2 \,\mathrm{d}t.
\end{aligned}$$

For brevity, we use the notation $\langle\cdot,\cdot\rangle_\alpha$ for the scalar product on $L^2(0,T;H_{2\alpha})$ and $\|\cdot\|_\alpha$ for the corresponding norm. We will prove that

$$\liminf_{N\to\infty}\left|\det\left(\left(\left\langle \frac{P_NF_i(X)}{\|P_NF_i(X)\|_\alpha}, \frac{P_NF_j(X)}{\|P_NF_j(X)\|_\alpha}\right\rangle_\alpha\right)_{i,j=0,\dots,K}\right)\right| > 0. \tag{6.11}$$

First note that by condition (I_α), this is true if the matrix in (6.11) is substituted by its $(0,0)$-minor, i.e. such that $1\le i,j\le K$.

Let $\epsilon>0$, let $M\in\mathbb{N}$ such that $\|(I-P_M)F_i(X)\|_\alpha < \epsilon\|F_i(X)\|$ for $i=1,\dots,K$. This is possible because $P_NF_i(X)\to F_i(X)$ in $L^2(0,T;H_{2\alpha})$. Then for $i=1,\dots,K$ and $N>M$:

$$\begin{aligned}\left\langle \frac{P_NF_0(X)}{\|P_NF_0(X)\|_\alpha}, \frac{P_NF_i(X)}{\|P_NF_i(X)\|_\alpha}\right\rangle_\alpha &= \left\langle \frac{P_MF_0(X)}{\|P_NF_0(X)\|_\alpha}, \frac{P_MF_i(X)}{\|P_NF_i(X)\|_\alpha}\right\rangle_\alpha \\ &\quad + \left\langle \frac{(P_N-P_M)F_0(X)}{\|P_NF_0(X)\|_\alpha}, \frac{(P_N-P_M)F_i(X)}{\|P_NF_i(X)\|_\alpha}\right\rangle_\alpha.\end{aligned}$$

Taking into account $\|P_NF_i(X)\|_\alpha\to\|F_i(X)\|_\alpha$ and $\|P_NF_0(X)\|_\alpha\to\infty$ in the first term as well as the Cauchy-Schwarz inequality for the second term,

$$\begin{aligned}\limsup_{N\to\infty}\left|\left\langle \frac{P_NF_0(X)}{\|P_NF_0(X)\|_\alpha}, \frac{P_NF_i(X)}{\|P_NF_i(X)\|_\alpha}\right\rangle_\alpha\right| &\le \limsup_{N\to\infty}\frac{\|(P_N-P_M)F_i(X)\|_\alpha}{\|P_NF_i(X)\|_\alpha} \\ &\le \frac{\|(I-P_M)F_i(X)\|_\alpha}{\|F_i(X)\|_\alpha} < \epsilon.\end{aligned}$$

As $\epsilon>0$ is arbitrary, we see that the $(0,i)$-entry of the matrix in (6.11) converges to zero for $i\ge 1$. Expanding the determinant in (6.11) in the first column and using the non-degeneracy of the $(0,0)$-minor, we conclude that (6.11) is true.

If D_N is the diagonal matrix with i-th diagonal entry $A_N^{(\alpha)}(X)_{i,i}$, (6.11) is equivalent to $\liminf_{N\to\infty}\left|\det(D_N^{-1/2}A_N^{(\alpha)}(X)D_N^{-1/2})\right|>0$. In particular,

$$\int_0^T \left\|(-A)^{1+\alpha}X_t^N\right\|_H^2 \mathrm{d}t \lesssim |\det(D_N)| \lesssim \left|\det\left(A_N^{(\alpha)}(X)\right)\right|. \tag{6.12}$$

Finally, (2.20) implies (6.10), and all statement in the lemma are proven. $\square$

Proposition 6.3. *Let $s_0\ge 0$, $\eta>0$ such that $X\in R(s_0)$, $X_0\in H_{s^*+\eta}$ a.s. and (N_η^γ) as well as $(F_{s,\eta})$ for every $s_0\le s<s^*$ hold. Let $\alpha\in\mathbb{R}$*

with $\gamma - (1+1/\beta)/4 < \alpha < \gamma + (\eta - 1 - 1/\beta)/4 \vee (\eta - 1 - 1/\beta)/2$ *and* $\gamma - 1/4 < \alpha \leq \gamma$. *Under condition* (I_α), *the sequences* $(\hat{\theta}_i^N)_{N\in\mathbb{N}}$, $i = 1, \ldots, K$ *are bounded in probability.*

Proof. First note that every admissible $\alpha \in \mathbb{R}$ is also admissible in Lemma 6.2. W.l.o.g. we can restrict to the case $B(X) \equiv \sigma(-A)^{-\gamma}$ due to $\alpha > \gamma - 1/4$, as explained in Theorem 3.22. Define

$$\bar{b}_N^{(\alpha)}(X)_i := \sigma \int_0^T \left\langle (-A)^{2\alpha-\gamma} P_N F_i(X)(t), \mathrm{d}W_t^N \right\rangle.$$

By Lemma 6.2, $A_N^{(\alpha)}(X)$ is invertible for all $N \geq N_0$. Plugging in the dynamics of X into the stochastic integral appearing in each component of $b_N^{(\alpha)}(X)$, it is immediate that

$$\hat{\theta}^N - \theta = A_N^{(\alpha)}(X)^{-1} \bar{b}_N^{(\alpha)}(X). \tag{6.13}$$

For simplicity of notation, denote the entries of $A_N^{(\alpha)}(X)$ by $a_{i,j}$, the entries of $A_N^{(\alpha)}(X)^{-1}$ by $a^{i,j}$ and the entries of $\bar{b}_N^{(\alpha)}(X)$ by $\bar{b}_i$, $i, j = 0, \ldots, K$. All terms implicitly depend on N. W.l.o.g. assume that $a_{i,i} > 0$ for $i = 0, \ldots, K$, which is guaranteed by (I_α) for large enough N. Then the i-th component of $\hat{\theta}^N - \theta$ reads as

$$\hat{\theta}_i^N - \theta_i = \sum_{j=0}^K a^{i,j} \bar{b}_j = \frac{1}{\det\left(A_N^{(\alpha)}(X)\right)} \sum_{j=0}^K \bar{b}_j (-1)^{i+j} \det\left(A_{j,i}\right),$$

where $A_{j,i}$ is the matrix obtained from erasing the j-th row and the i-th column in $A_N^{(\alpha)}(X)$. By means of $a_{i,j} \leq \sqrt{a_{i,i} a_{j,j}}$,

$$\begin{aligned}
\left|\hat{\theta}_i^N - \theta_i\right| &\leq \frac{1}{\left|\det\left(A_N^{(\alpha)}(X)\right)\right|} \sum_{j=0}^K \left|\bar{b}_j\right| \frac{K!}{\sqrt{a_{i,i} a_{j,j}}} \prod_{\ell=0}^K |a_{\ell,\ell}| \\
&\lesssim \sum_{j=0}^K \frac{\left|\bar{b}_j\right|}{\sqrt{a_{i,i} a_{j,j}}},
\end{aligned}$$

where we have used Lemma 6.2. Next, $\alpha < \gamma + (\eta - 1 - 1/\beta)/4$ implies $F_j(X) \in R(4\alpha - 2\gamma + \epsilon)$ for some $\epsilon > 0$. By Lemma 2.6, $\lim_{N\to\infty} \left|\bar{b}_j\right| < \infty$

a.s. for $j = 1, \dots, K$. Thus, for these j, $\bar{b}_j/\sqrt{a_{j,j}}$ is bounded almost surely and thus in probability. Finally, taking into account $\alpha > \gamma - (1+1/\beta)/4$,

$$N^{\beta(\gamma-\alpha)} \frac{\bar{b}_0}{\sqrt{a_{0,0}}} = \sigma N^{\beta(\gamma-\alpha)} \sqrt{\frac{\int_0^T \left\|(-A)^{1+2\alpha-\gamma} X_t^N\right\|_H^2 \mathrm{d}t}{\int_0^T \left\|(-A)^{1+\alpha} X_t^N\right\|_H^2 \mathrm{d}t}} \\ \times \frac{\int_0^T \left\langle (-A)^{1+2\alpha-\gamma} X_t^N, \mathrm{d}W_t^N \right\rangle}{\sqrt{\int_0^T \left\|(-A)^{1+2\alpha-\gamma} X_t^N\right\|_H^2 \mathrm{d}t}},$$

which converges to a normal distribution by (2.20) together with the choice of $\alpha > \gamma - (1+1/\beta)/4$, Theorem A.1, and the Slutsky lemma. Consequently, as $\alpha \leq \gamma$, we see that $\bar{b}_0/\sqrt{a_{0,0}}$ is bounded in probability, too. In total, for $i = 1, \dots, K$, $|\hat{\theta}_i^N - \theta_i|$ is bounded a.s. by the sum of random variables that are bounded in probability, so $(\hat{\theta}_i^N)_{N\in\mathbb{N}}$ itself is bounded in probability. $\square$

In particular, consider the case $A = \Delta$ in dimension $d \leq 2$, where F is a reaction term that satisfies $(F_{s,\eta}^{\mathrm{par}})$ for all $\eta < 2$. Combining Theorem 6.1 with Proposition 6.3, we obtain:

Theorem 6.4. *Let $d \leq 2$, $\gamma > d/4$, $s_0 \geq 0$ such that (N_η^γ) and $(F_{s,\eta}^{\mathrm{par}})$ hold for $s_0 \leq s < s^*$ and $0 < \eta < 2$. Let a.s. $X_0 \in H_{s^*+2}$ and $X \in R(s_0)$.*

- *(i) In $d = 1$, let $\gamma - 1/4 < \alpha \leq \gamma$ such that (I_α) holds. Then $\hat{\theta}_0^N$ is asymptotically normal as in (2.29).*
- *(ii) In $d = 2$, let $\gamma - 1/4 < \alpha < \gamma$ such that (I_α) holds. Then $\hat{\theta}_0^N$ converges in probability to θ_0 with rate N^{-a} for every $a < 1$.* $\square$

Remark 6.5. *The condition on α can be relaxed. For example, if $B = \sigma(-A)^{-\gamma}$, only $\alpha > \gamma - (1+d/2)/4$ is needed, and a similar result for dimension $d = 3$ can be stated, cf. Remark 3.23.*

Theorem 6.4 is applicable to the activator-inhibitor model explained in the next section.

6.1.3 A Model of FitzHugh–Nagumo Type

For $L_1, \dots, L_d > 0$, let $\mathcal{D} = [0, L_1] \times \dots \times [0, L_d] \subset \mathbb{R}^d$ a bounded domain. Motivated by [FFAB20], we consider an activator–inhibitor model of the form

$$\mathrm{d}U_t = \Big(D_U \Delta U_t + k_1 U_t(u_0 - U_t)(U_t - u_0 a(\|U_t\|_{L^2(\mathcal{D})})) - k_2 V_t\Big)\,\mathrm{d}t + B\mathrm{d}W_t, \tag{6.14}$$

$$\mathrm{d}V_t = (D_V \Delta V_t + \epsilon(bU_t - V_t))\,\mathrm{d}t, \tag{6.15}$$

together with initial conditions U_0, V_0 and periodic boundary conditions. Consequently, in this setting, the state space is given by $H := L^2(\mathcal{D})$, and $H_1 = \bar{W}^{1,2}(\mathcal{D}) = \{u \in W^{1,2}(\mathcal{D}) \mid \int_{\mathcal{D}} u \mathrm{d}x = 0\}$. Here and in the sequel we consider only the case $B = \sigma(-\Delta)^{-\gamma}$ for $\sigma > 0$. The parameters $D_U, D_V > 0$ are the diffusivity constants for the activator U and inhibitor V, resp. The parameters $k_1, k_2, u_0, \epsilon, b$ are supposed to be positive. Finally, $a : \mathbb{R} \to \mathbb{R}$ is a bounded continuously differentiable function with bounded derivative. The boundedness of a is not essential and can be modeled in practice with a cutoff. For constant a, this is the spatially extended FitzHugh–Nagumo model. We mention that a careful choice of the function a can have a stabilizing effect on the dynamics. We also introduce an additional parameter $\bar{a} \in (0, 1)$, which will describe the effective long–time average of $a(\|U\|)$. The initial conditions are assumed to be sufficiently regular, i.e. $\mathbb{E}[\|U_0\|_{s^*}^p] < \infty$, $\mathbb{E}[\|V_0\|_{s^*+2}^p] < \infty$ for all $p \geq 2$ and $s^* = 1 + 2\gamma - d/2$. This model is well-posed in dimension $d \leq 3$:

Proposition 6.6. *Let $d \leq 3$ and $\gamma > d/4 + 1/2$. Then there exists a unique solution (U, V) to (6.14), (6.15) with $U \in R^{\mathbb{E}}(s)$ and $V \in R^{\mathbb{E}}(s + 2)$ for any $s < s^*$.*

The proof is given in Appendix B.3.

This model is used to describe cell data in Section 6.2, where we assume that the observation X is given by the activator concentration $X = U$. The activator dynamics in this model can be reduced to (6.1) as follows: First, with the variation of constants formula, V is determined by U via

$$V_t = e^{t(D_V\Delta - \epsilon I)} V_0 + \epsilon b \int_0^t e^{(t-r)(D_V\Delta - \epsilon I)} U_r \mathrm{d}r, \tag{6.16}$$

where I is the identity operator acting on $L^2(\mathcal{D})$ and $t \mapsto e^{t(D_V\Delta - \epsilon I)}$ is the semigroup generated by $D_V\Delta - \epsilon I$. For simplicity, we assume $V_0 = 0$ here and in the sequel. Now, the nonlinearity F is given by

$$F_{\theta_1,\theta_2,\theta_3}(X)(t) = \theta_1 F_1(X_t) + \theta_2 F_2(X_t) + \theta_3 F_3(X)(t), \tag{6.17}$$

where

$$\theta_1 = k_1 u_0 \bar{a}, \quad \theta_2 = k_1, \quad \theta_3 = k_2 \epsilon b, \tag{6.18}$$

and

$$F_1(X_t) = -\frac{a(\|X_t\|_{L^2(\mathcal{D})})}{\bar{a}} X_t(u_0 - X_t), \tag{6.19}$$

$$F_2(X_t) = X_t^2(u_0 - X_t), \tag{6.20}$$

$$F_3(X)(t) = -\int_0^t e^{(t-r)(D_V\Delta - \epsilon I)} X_r \mathrm{d}r. \tag{6.21}$$

This matches the statistically linear case (6.4) with $F^* = 0$. Note that F_3 acts on the trajectory of X, such that the dynamics of X is not Markovian, even if the joint dynamics of (U, V) is Markovian. Finally, $\theta_0 = D_U$ is the diffusivity of the activator.

6.2 Application to Cell Data

Next, we apply the theory of joint diffusivity and reaction parameter estimation to simulated and real giant cell data. Our main assumption is that the data is generated by the FitzHugh–Nagumo model from Section 6.1.3. Why this is certainly the case for the numerical simulation (up to a discretization error), it is less clear for data coming from microscopy observation.

As a first approximation, when estimating parameters of the process, we always assume that the function a is constant, $a \equiv \bar{a}$, where the latter value is known or unknown. In this case, F_1 is replaced by

$$\widetilde{F}_1(X_t) = -X_t(u_0 - X_t). \tag{6.22}$$

For simulated data, this models a misspecification of the true generating dynamics. However, this is not severe, as $a(\|X_t\|_{L^2(\mathcal{D})})$ tends to oscillate around its effective value. In this sense, if the real data is modeled accurately by the FitzHugh–Nagumo model from Section 6.1.3, this additional assumption will have little impact.

In order to compare the effect of different model assumptions on diffusivity estimation, we construct a hierarchy of estimators, starting from the purely linear case and taking into account an increasing number of features from the FitzHugh–Nagumo model. The assumptions displayed here refer to the description of the data used to perform parameter estimation, not the generating process itself.

- $\hat{\theta}_0^{\mathrm{lin},N}$ is the estimator for θ_0 coming from the assumption of a purely linear model, i.e. a stochastic heat equation. In this case, $F = 0$, and there are no other drift parameters to be estimated.

- $\hat{\theta}_0^{\mathrm{pol},N}$ is the diffusivity estimator based on a stochastic Schlögl (or Nagumo) model [Sch72], i.e.

$$\begin{aligned} F(X) &= k_1 X(u_0 - X)(X - \bar{a}u_0) \\ &= \theta_1 \widetilde{F}_1(X) + \theta_2 F_2(X), \end{aligned}$$

  where both reaction parameters θ_1 and θ_2 are assumed to be known. This model is capable of generating spatially extending phase transitions for the concentration of X, and it arises formally from taking $\epsilon \to 0$ in the stochastic FitzHugh–Nagumo model.

- $\hat{\theta}_0^{\mathrm{full},N}$ is the diffusivity estimator under the assumption of a full FitzHugh–Nagumo model, i.e.

$$\begin{aligned} F(X) &= k_1 X(u_0 - X)(X - \bar{a}u_0) - k_2 \epsilon b \int_0^{\cdot} e^{(\cdot - r)(D_V \Delta - \epsilon I)} X_r \mathrm{d}r \\ &= \theta_1 \widetilde{F}_1(X) + \theta_2 F_2(X) + \theta_3 F_3(X), \end{aligned}$$

  This model can generate traveling waves as observed in the cell data. Again, all reaction parameters are assumed to be known.

While $\hat{\theta}_0^{\mathrm{full},N}$ incorporates the full model, the assumption of known reaction parameters will be too strong. We further relax this assumption by estimating an increasing number of reaction parameters simultaneously:

- $\hat{\theta}_0^{2,N}$ is the diffusivity estimator based on the full FitzHugh–Nagumo model as $\hat{\theta}_0^{\mathrm{full},N}$, but with unknown θ_1.

- $\hat{\theta}_0^{3,N}$ additionally treats θ_2 as unknown.

- $\hat{\theta}_0^{4,N}$ treats all reaction parameters $\theta_1, \theta_2, \theta_3$ as unknown.

The superscript denotes the number of unknown parameters, including the diffusivity θ_0. Thus, $\hat{\theta}_0^{4,N}$ is an estimator for θ_0 which uses qualitative, but little quantitative knowledge on the generating process. While all estimators are consistent with optimal rate as $N \to \infty$ by Theorem 6.4, their performance in the non-asymptotic setting may vary strongly.

For all estimators we have described here, we set $\alpha = 0$. This is reasonable if the driving noise of the activator component is close to being white noise.

The linear estimator $\hat{\theta}_0^{\mathrm{lin},N}$ is the same as $\hat{\theta}_N^{\mathrm{lin}}$ from Chapter 2, and it is given by (2.26). In contrast to the other estimators considered here, it is scale invariant in the sense that for any $c > 0$, the substitution $X \mapsto cX$ leaves the resulting estimator $\hat{\theta}_0^{\mathrm{lin},N}$ invariant. It is clear that this invariance does not hold if nonlinearities are taken into account. While the remaining estimators use detailed information on the nonlinear model, their performance depends on a careful calibration of the intensity of the input data in order to match the fixed points of the third order polynomial in the reaction term. In fact, this is a source of additional uncertainty. While the advantages of a good reaction model clearly outweighs the benefit from scale invariance for simulated data, as we will see in Section 6.2.1, this is less clear for real data, where the reaction model may not fully capture the underlying dynamics, see Section 6.2.2.

In this section, we work on two-dimensional rectangular domains of the form $\mathcal{D} = [0, L_1] \times [0, L_2]$ with $L_1, L_2 > 0$. In particular, the eigenfunctions in H of $-\Delta$ with periodic boundary conditions are given by $\Phi_{k,\ell}(x_1, x_2) = \varphi_k(x_1/L_1)\varphi_\ell(x_2/L_2)$ for $(k, \ell) \in \mathbb{Z}^2$, where $\varphi_k(x) = \cos(2\pi kx)$ for $k \leq 0$ and $\varphi_k(x) = \sin(2\pi kx)$ for $k > 0$. The eigenvalues are given by $\lambda_{k,\ell} = 4\pi^2(k^2/L_1^2 + \ell^2/L_2^2)$. As before, we choose a reordering $r : \mathbb{N} \to \mathbb{Z}^2 \backslash \{0\}$ such that $\lambda_N = \lambda_{r(N)}$ is increasing, with corresponding eigenfunction $\Phi_N = \Phi_{r(N)}$, where we exclude the case $\lambda_{0,0} = 0$.

6.2.1 Evaluation of Simulated Data

We simulate the system (6.14), (6.15) on a two-dimensional square with side length $L = 75$ and periodic boundary conditions, starting from zero

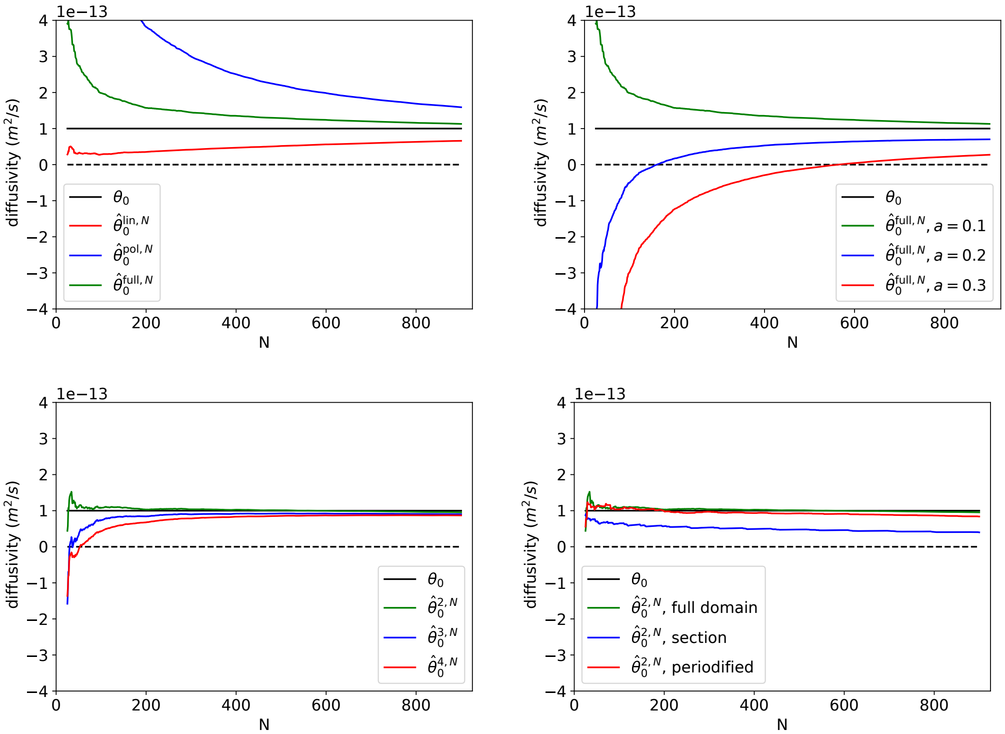

Figure 6.1: *Performance of different diffusivity estimators on a numerical simulation as N gets large. Solid black line is plotted at the true value of* $\theta_0 = 1 \times 10^{-13}$ m, *dashed black line is plotted at zero. We restrict to $N \geq 25$ in order to avoid artifacts.*

initial conditions and with $\theta_0 = D_U = 0.1$. We use an explicit finite difference scheme with spatial and temporal increment $\Delta x = 0.375$ and $\Delta t = 0.01$, respectively. In order to mitigate the impact of the initial conditions, we observe every 100th frame of the simulation in the shifted interval $[T_0, T_1)$, where $T_0 = 500$ and $T_1 = 700$. The remaining drift parameters are $D_V = 0.02, k_1 = k_2 = 1, u_0 = 1, \epsilon = 0.02, b = 0.2$. In the noise, we set $\gamma = 0$ and $\sigma = 0.1$. The unstable zero of the reaction potential is determined by $a(z) = 0.5 - b + 0.5(z/(0.33u_0L^2) - 1)$. In order to compare the simulation to real data, the unit length and unit time in these specifications are interpreted as $1\,\mu$m and $1\,s$, respectively.

In Figure 6.1 (top left), the performance of $\hat{\theta}_0^{\mathrm{lin},N}$, $\hat{\theta}_0^{\mathrm{pol},N}$ and $\hat{\theta}_0^{\mathrm{full},N}$ with

$\bar{a} = 0.1$ is compared. The linear estimator severely underestimates the true diffusivity. This can be explained as follows: The data exhibits steep concentration gradients at the wave fronts, which are interpreted by $\hat{\theta}_0^{\mathrm{lin},N}$ as coming from low diffusive forcing. In contrast, the estimator $\hat{\theta}_0^{\mathrm{pol},N}$, which takes into account the bistable polynomial from the reaction term, heavily overestimates the true value of θ_0. As an explanation, while this estimator is able to account for the phase transition at the wave front, the concentration decay due to the inhibitor in the data is interpreted as fast diffusion. Finally, $\hat{\theta}_0^{\mathrm{full},N}$ performs best of these three estimators, incorporating knowledge on the full reaction model. Note, however, that in the simulation, a is not constant, such that even $\hat{\theta}_0^{\mathrm{full},N}$ does not have perfect information on the dynamics of X. Rather, $a(\|X_t\|_{L^2(\mathcal{D})})$ oscillates around a value slightly larger than 0.15 in the simulation. Figure 6.1 (top right) shows the sensitivity of $\hat{\theta}_0^{\mathrm{full},N}$ to $\bar{a}$. Different a priori assumptions on $\bar{a}$ have a large impact on the value of the estimator, even for large N. In contrast, Figure 6.1 (bottom left) shows the performance of the estimators $\hat{\theta}_0^{2,N}$, $\hat{\theta}_0^{3,N}$ and $\hat{\theta}_0^{4,N}$, which treat the reaction parameters as unknown. All of them determine the true value of θ_0 rapidly, even if a is misspecified in their description of the dynamics.

Apart from the form of the nonlinearity F, the behavior of X at the boundary impacts the performance of the estimators. In Figure 6.1 (bottom right), we evaluate $\hat{\theta}_0^{2,N}$ on the original domain $\mathcal{D}$ of 200×200 pixels, on the restriction of X to a subdomain of 75×75 pixels and its periodification, as described below, on a square of 150×150 pixels. When evaluated on a subdomain instead of $\mathcal{D}$, the estimate deteriorates. A possible explanation is given as follows: The assumption of periodic boundary conditions on the subdomain leads to discontinuities of X at its boundary. As before, these discontinuities can be interpreted as steep gradients, which the diffusivity estimators translates into low diffusivity present in the data.

As a remedy, we use a hands-on approach and periodify the data in the sense that we take four copies of the data, mirror them on both coordinate axes and glue them together in such a way that the resulting field is continuous and fills a square with double side length compared to $\mathcal{D}$, with periodic boundary conditions. This way, we avoid the discontinuities, but the resulting field still does not satisfy the dynamics of X at the boundaries of the original domain $\mathcal{D}$. Further, due to the introduced redundancies and change in spatial extension, the estimators based on the original and periodified data should be compared for different values of N.

While periodification seems to be a natural ad-hoc approach to deal with the difficulties arising at the boundary, its performance will depend on the specific situation, and it should be used with care. In order to understand its performance better, a systematic study is needed.

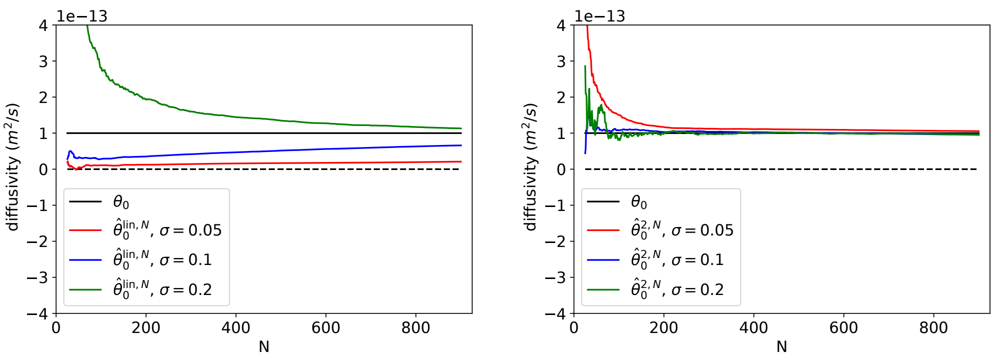

Figure 6.2: *Comparison of* **(left)** $\hat{\theta}_0^{\mathrm{lin},N}$ *and* **(right)** $\hat{\theta}_0^{2,N}$ *for different noise intensity levels. We restrict to* $N \geq 25$ *in both panels. As before, the solid black line is plotted at the true value* $\theta_0 = 1 \times 10^{-13}$ m*, and the dashed black line is plotted at zero.*

In Figure 6.2, the effect of changing the noise intensity σ on diffusivity estimation is shown. We simulate additional trajectories with $\sigma = 0.05$ and $\sigma = 0.2$, on which $\hat{\theta}_0^{\mathrm{lin},N}$ and $\hat{\theta}_0^{2,N}$ are evaluated. The former estimator is agnostic to the reaction model, whereas the latter includes the full FitzHugh–Nagumo model. While $\hat{\theta}_0^{2,N}$ performs well regardless of the noise intensity, the behavior of $\hat{\theta}_0^{\mathrm{lin},N}$ is heavily influenced by the noise level, with large noise intensity leading to better results. In this sense, a large noise intensity hides the effect of the nonlinear term. This is in accordance with the Monte–Carlo simulation for the stochastic Allen–Cahn equation in Section 2.5. We note that for $\sigma = 0.2$, the simulation is no longer capable of generating traveling waves due to large fluctuations in the driving noise.

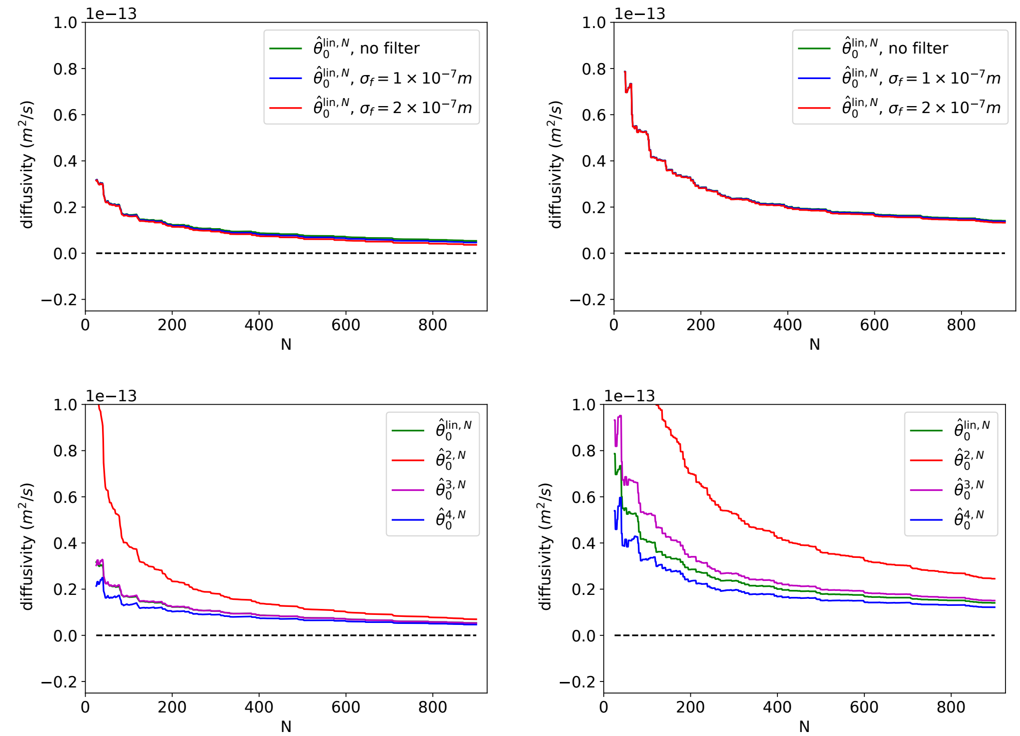

Figure 6.3: **(top row)** *Performance of* $\hat{\theta}_0^{\mathrm{lin},N}$ *on spatially smoothed observations.* **(bottom row)** *Performance of different diffusivity estimators on the data, without spatial smoothing. The data is considered* **(left column)** *without and* **(right column)** *with periodification. Dashed line is plotted at zero. As before, we restrict to* $N \geq 25$ *in all panels.*

6.2.2 Evaluation of Real Data

We first describe the data we are working with. See [PFA+21, Appendix B] for a description of the experimental setup.[2] Each observation consists of a sequence of rectangular frames of varying length, resolution and aspect ratio, describing the observed intensity of an actin marker within a *D. discoideum* giant cell. The regions considered lie completely inside the cell, i.e. no cell boundaries appear in the data. The concentration of the actin marker is given by grey values ranging from 0 to 255 at each pixel. When evaluating the data, this intensity is standardized to the interval $[0, 1]$, such that absence

[2] The giant cell data used in this chapter has been provided by Sven Flemming.

of the activator and the maximal activator intensity match the stable fixed points 0 and $u_0 = 1$, respectively.[3] As the data exhibits traveling waves, it is assumed that the actin concentration (or actin marker concentration) can be described by (6.14), (6.15).

In this section, we estimate the diffusivity on a single giant cell observation. The analysis of a population of cells is postponed to Section 6.2.3.

As a first consistency check, we consider the behavior of the data set under convolution. For $k \in L^1(\mathcal{D})$, let $T_k : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ be the convolution operator given by $(T_k f)(y) = \int_{\mathcal{D}} k(y-x)f(x)\mathrm{d}x$, where k and f are identified with their periodic continuation. It is well-known that $\Delta \circ T_k = T_k \circ \Delta$. In particular, if X is generated by a (stochastic, perturbed) heat equation with (homogeneous and isotropic) diffusivity θ_0, the same is true for $T_k X$, although the structure of the nonlinear term and the noise changes. Thus, if the describing model is reasonable, it is to expect that the estimated diffusivity of $T_k X$ is close to that of X. We use a family of kernels $k = k(\sigma_f)$ for $\sigma_f > 0$, which are constructed as a Gaussian density with standard deviation σ_f, truncated to the rectangle $[-L_1/2, L_1/2] \times [-L_2/2, L_2/2]$ and normed in $L^1(\mathcal{D})$. We use the standard deviation $\sigma_f = 1 \times 10^{-7}m$ and $\sigma_f = 2 \times 10^{-7}m$. The performance of $\hat{\theta}_0^{\mathrm{lin},N}$ on data smoothed with $k(\sigma_f)$ is shown in Figure 6.3 (top left) without periodification, and in Figure 6.3 (top right) for the periodified data. While not in perfect alignment, the estimator graphs are very close. For the data without periodification, the decrease of $\hat{\theta}_0^{\mathrm{lin},N}$ in N is slightly more highlighted. For the periodified data, which cannot be expected to satisfy (6.14), (6.15) on the boundaries of the four sub-patches of its enlarged domain, the graphs of the estimators are nonetheless almost indistinguishable. In this sense, periodification seems to retain convolution invariance. In total, these results support the hypothesis that the data is generated by a stochastic partial differential equation with diffusive forcing stemming from a second order differential operator.

Now we proceed to the nonlinear reaction model. Based on the performance of the estimators from Section 6.2.1, we compare $\hat{\theta}_0^{\mathrm{lin},N}$ with $\hat{\theta}_0^{2,N}$, $\hat{\theta}_0^{3,N}$ and $\hat{\theta}_0^{4,N}$, which incorporate knowledge on the full FitzHugh–Nagumo model. The performance on cell data is shown in Figure 6.3 (bottom left), and the performance on periodified data is shown in Figure 6.3 (bottom right). In-

[3]Such calibration is necessary for all estimators except $\hat{\theta}_0^{\mathrm{lin},N}$.

terestingly, the model-free estimator $\hat{\theta}_0^{\text{lin},N}$ behaves similar to $\hat{\theta}_0^{3,N}$ and $\hat{\theta}_0^{4,N}$, which are the most flexible estimators we consider and which do not fix the reaction rate corresponding to the bistable potential in the drift. This pattern does also appear in different sample cells. In terms of diffusivity estimation, the detailed reaction model doesn't seem to yield additional benefit, in contrast to the case of simulated data from Section 6.2.1. This can be seen as a hint that the FitzHugh–Nagumo model, while being capable of generating traveling waves, misses additional features of the true intracellular dynamics. For example, it may be helpful to consider models which mimic the biophysical reaction pathway more closely. On the other hand, $\hat{\theta}_0^{2,N}$, which fixes the parameters describing the reaction intensities in advance, deviates from the other estimators, but finally approaches them. Further evaluations suggest that changing u_0 does not alter the general picture.

The performance of the estimators on the periodified sample is similar to the the case of the original sample. In accordance with the discussion from the previous section, the estimated diffusivity increases in that case.

6.2.3 Evaluation of a Cell Population

We consider a population of 36 giant cell observations, as described in the previous section. The spatial extension of each data set is clipped in such a way that only the interior dynamics is captured, i.e. no cell boundaries appear in the data. As a consequence, the spatial resolution varies within the cell population. It is natural to assume that the range of possible N that yields meaningful results grows with the resolution of the sample. In general, while the estimate will be more precise for large N, discretization effects depending on the spatial resolution will render arbitrarily large N useless. In order to find a reasonable tradeoff, we apply the following heuristics: If each frame within a data set consists of $r_x \times r_y$ pixels, we set $N_{\text{stop}} = \lfloor 4r_x r_y / R^2 \rfloor$, where R is a parameter representing the number of pixels needed in order to extract meaningful information on $[0, 2\pi]$ by testing with a sine function. For example, if $r_x = r_y = R$, then $N_{\text{stop}} = 4$, and only the first four eigenfunctions $\Phi_{\pm 1, \pm 1}$ are taken into account, whose period is R pixels in both dimensions. We choose $R = 12$ for the cell population and $R = 24$ for the periodified population. Further, we set $N_{\text{const}} = 899$ and evaluate the estimator $\hat{\theta}_0^{3,N}$ at $N = N_{\text{const}}$ and $N = N_{\text{stop}}$. Results are shown in Figure 6.4. Note that

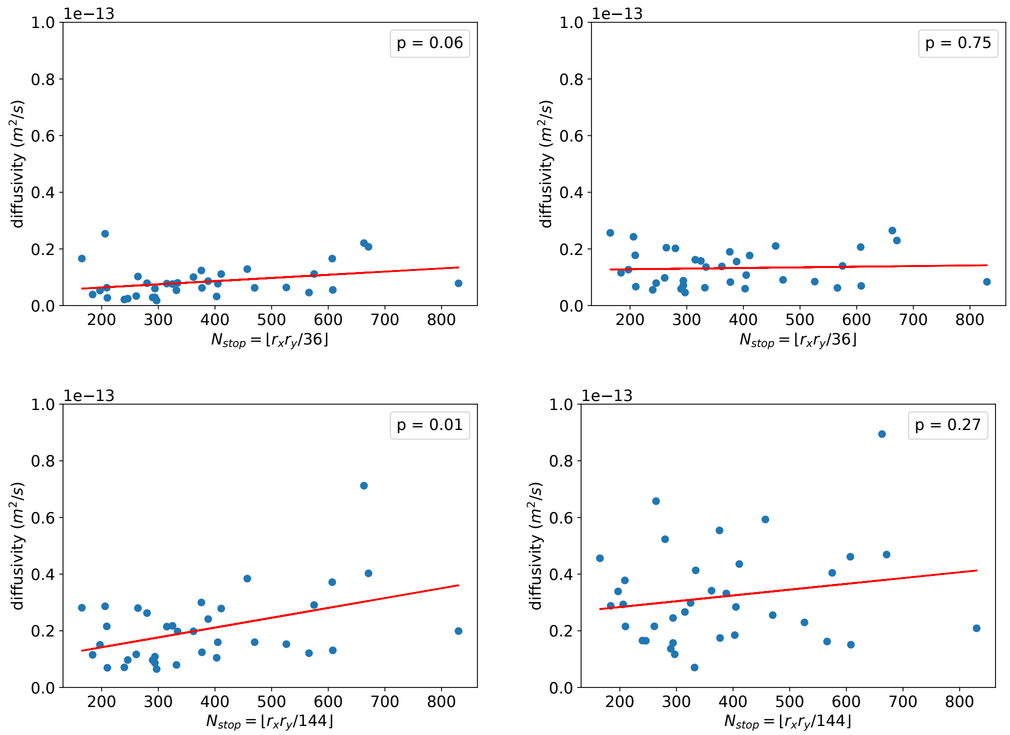

Figure 6.4: *The estimator* $\hat{\theta}_0^{3,N}$ *evaluated at* $N = N_{\mathrm{const}}$ **(left column)** *and* $N = N_{\mathrm{stop}}$ **(right column)** *is plotted against* N_{stop} *for a sample of* 36 *observations* **(top row)** *and their periodification* **(bottom row)**. *The least square regression lines are plotted in red. The p-value in each display comes from a two-sided t-test with null hypothesis that the slope is zero.*

N_{stop} encodes the resolution of the frames within a sample.[4] We see that choosing N based on the spatial resolution decorrelates the estimate for θ_0 from the resolution of the frames. Further, the estimated diffusivities for all samples considered have a similar magnitude. This indicates that the concept of effective diffusivity can be useful for statistical analysis on cell samples of the same or possibly different populations.

In addition to the inhomogeneous spatial resolution, the number of frames

[4] In fact, N_{stop} grows like the number of pixels in each frame. If each pixel is interpreted as a point evaluation of the underlying process (rather than a local average), this is in accordance with Example 4.8 and (4.50).

(i.e. the temporal resolution) varies within the population. However, the estimate tends to stabilize in time, such that this does not pose a problem.

We further note that the population is not homogeneous with respect to the side length Δx of a pixel and the temporal distance Δt between two frames. Further tests indicate that the estimated diffusivity correlates with the characteristic diffusivity $\Delta x^2/\Delta t$. However, a detailed analysis of the resulting effects, including the impact of discretization and possible scale dependence of the diffusivity, is beyond the scope of the present work.

6.2.4 The Effective Unstable Zero

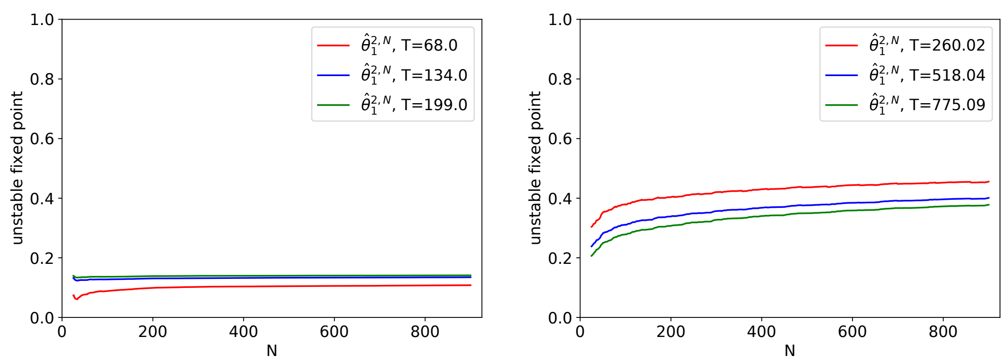

Figure 6.5: *Estimated unstable fixed point for simulated data* **(left)** *and real data* **(right)**. *Frames up to time T (in seconds) are used to calculate $\hat{\theta}_0^{2,N}$, starting from the first frame in the sample. As before, we restrict to $N \geq 25$.*

When using $\hat{\theta}_0^{2,N}$ in order to estimate the diffusivity θ_0, we simultaneously obtain an estimate $\hat{\theta}_1^{2,N}$ for θ_1 by solving (6.7). As $\theta_1 = k_1 u_0 \bar{a}$ and $k_1 = u_0 = 1$ by assumption, we can identify θ_1 with the effective unstable zero $\bar{a}$ from the reaction term. In Figure 6.5, the performance of this estimate is displayed for simulated data and an experimentally observed sample.

The term $a(\|X_t\|_{L^2(\mathcal{D})})$ oscillates around an effective value slightly larger than 0.15 in the numerically simulated trajectory. Even if this value is approximated, we see that the quality of the estimate does not improve with increasing N. Indeed, this cannot be expected, as the reaction term is of order zero: It is known [HR95] that the maximum likelihood estimate of the coefficient of a linear order zero perturbation to a stochastic heat equation converges only with logarithmic rate in dimension $d = 2$. On the other hand, also the long-time behavior can be considered, including an increasing number of frames into the evaluation. The left-hand panel in Figure 6.5 shows that the effective value is approached with larger T.[5]

In the case of real data, the results fall into the interval $(0, 1)$ and are rather stable. This indicates that the "effective unstable fixed point under the reaction model F", defined as the value at which $\hat{\theta}_1^{2,N}$ stabilizes, can be used in a meaningful way for statistically evaluating spatiotemporal cell data exhibiting traveling waves.

6.2.5 The Effective Diffusivity Outside the Cell

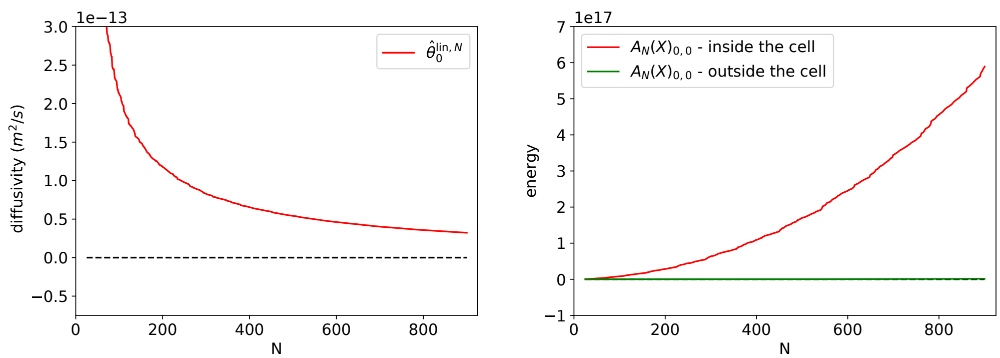

Figure 6.6: **(left)** *Performance of* $\hat{\theta}_0^{\mathrm{lin},N}$ *on a data set consisting of pure noise outside the cell.* **(right)** *Comparison of the energy of a measurement inside and outside the cell, with the same spatiotemporal extension. In both panels, the dashed line is plotted at zero. As before, we restrict to* $N \geq 25$.

When formally applying the estimation procedure to a data set consisting of pure noise, i.e. a region of a microscopy data set where no part of the cell is present, we obtain a result that can be named "effective diffusivity outside the cell" or "effective diffusivity of the noise". Here, we restrict ourselves to

[5]Typically, under ergodicity assumptions, consistency with convergence rate $T^{-1/2}$ can be expected for estimators of maximum likelihood type if T increases, see e.g. the monograph [Kut04] for SDEs, [Log84, KL85] for linear SPDEs. In [DMPD00, GM02], large time consistency is proven in the context of semilinear SPDEs.

the use of $\hat{\theta}_0^{\text{lin},N}$, i.e. $F = 0$. The result is shown in the left panel of Figure 6.6. Comparing with Figure 6.3, the effective diffusivity of a pure noise observation can even exceed the value obtained from a region inside a cell.[6] It is important to understand the order of magnitude of this effective value, as well as its impact on diffusivity estimation within the cell.

We derive the magnitude of the effective diffusivity outside the cell heuristically: A pixel can be described by a weighted indicator function of a square within $\mathcal{D}$, where the weight describes the intensity. In the case of pure noise, the instantaneous disappearance of such a pixel in the next frame can be interpreted as fast diffusion within the time between two frames. In order to understand the magnitude of θ_0 needed for that effect, we approximate the indicator function of the pixel by a Gaussian density. For $t > 0$, let $\phi_t : \mathbb{R}^2 \to \mathbb{R}$ be the centered Gaussian density in two dimensions with covariance matrix tI. This density attains its maximum at $x = 0$, with $\phi_t(0) = 1/(2\pi t)$. Let Δt be the time between two frames, let Δx be the side length of a pixel within each frame. We set $\sigma_0 = \Delta x/2$. In this case, the distance between the inflection points of the one-dimensional marginals of $\phi_{\sigma_0{}^2}$ matches the side length of a pixel, and we take $\phi_{\sigma_0{}^2}$ as an approximation for the pixel. After time Δt, the heat semigroup on $\mathbb{R}^2$ with diffusivity θ_0 maps $\phi_{\sigma_0{}^2}$ to $\phi_{\sigma_0{}^2+2\theta_0\Delta t}$. Now let the decay of the maximal value of the density, $\phi_{\sigma_0{}^2}(0)/\phi_{\sigma_0{}^2+2\theta_0\Delta t}(0)$, be at least as large as some threshold $b > 0$, i.e. $(\sigma_0{}^2 + 2\theta_0\Delta t)/\sigma_0{}^2 \geq b$. This is equivalent to

$$\theta_0 \geq \frac{b-1}{8}\frac{\Delta x^2}{\Delta t}. \tag{6.23}$$

In the data sample from Figure 6.6 (left), we have $\Delta x = 2.08 \times 10^{-7}\,\mathrm{m}$ and $\Delta t = 0.97\,s$. The decay factor b depends on the particular noise pixel and its intensity within the data set. Reasonable values are given for $b \leq 30$. For example, if $b = 30$, then $\theta_0 \geq 1.6 \times 10^{-13}\,\mathrm{m}^2/s$, if $b = 20$, then $\theta_0 \geq 1 \times 10^{-13}\,\mathrm{m}^2/s$, and if $b = 15$, then $\theta_0 \geq 7.8 \times 10^{-14}\,\mathrm{m}^2/s$. This matches the order of the observed diffusivity from Figure 6.6 (left): For example, if we apply the stopping rule from Section 6.2.3 to this case, i.e. $N_{\text{stop}} = \lfloor 4r_x r_y/R^2 \rfloor$ with $R = 12$, then we obtain $N_{\text{stop}} = 165$ and $\hat{\theta}_0^{\text{lin},N} = 1.36 \times 10^{-13}\,\mathrm{m}^2/s$ for $N = N_{\text{stop}}$, in accordance with the heuristic derivation in this section.

[6] This observation also applies to the cell sample used in the right panel of Figure 6.6.

We have seen that the observed diffusivity of the noise can be larger than the estimated diffusivity of the signal within the cell. Nonetheless, the noise described here does not interfere with the diffusivity estimate of the signal. This can be explained as follows: Assume that the signal process X^{sig} is perturbed by measurement noise W^{meas}. This means that inside the cell, we observe $X = X^{\mathrm{sig}} + W^{\mathrm{meas}}$ instead of $X = X^{\mathrm{sig}}$ as supposed previously, while outside the cell, only $X = W^{\mathrm{meas}}$ is observed. It is revelatory to consider the energy $A_N(X)_{0,0}$ for both cases separately. This is done in the right panel of Figure 6.6. The value of $A_N(W^{\mathrm{meas}})_{0,0}$ outside the cell is orders of magnitude below the energy within the cell, at least at the resolution level we consider. Consequently, $A_N(X^{\mathrm{sig}} + W^{\mathrm{meas}})_{0,0}$ is indistinguishable from $A_N(X^{\mathrm{sig}})_{0,0}$. Thus, it doesn't make a difference if $\hat{\theta}_0^{\mathrm{lin},N}$ is evaluated on $X^{\mathrm{sig}} + W^{\mathrm{meas}}$ or on X^{sig} itself, and the measurement noise has very little impact on the estimated diffusivity of the signal.

However, from a mathematical perspective, adding noise to X^{sig} has an impact on its regularity, such that the theoretical properties of $\hat{\theta}_0^{\mathrm{lin},N}$ for $N \to \infty$ will change, depending on the precise model assumptions.

Chapter 7

Further Research

As exposed in the introduction, statistical inference for SPDEs is a source for diverse mathematical research. The field is continuously expanding, and it keeps incorporating new models and methods. In this work, we considered parameter estimation for semilinear equations in different asymptotic regimes, together with possible model misspecification. To conclude, we give a list of further interesting mathematical questions related to the topic of this work. This list is by no means exhaustive, and it should be considered a suggestion for possible further research.

- Beyond semilinear models, one can consider quasilinear equations, e.g. with state-dependent diffusivity.
- Further types of model misspecification can be studied: For example, this includes the effect of an inhomogeneous or anisotropic diffusivity on the estimators from Chapter 2.
- The impact of measurement noise can be analyzed systematically.
- Apart from the spatially discretized Laplacian used in Chapter 4, which is based on a Fourier decomposition, it is interesting to consider the classical discretization on a three-point stencil or five-point stencil (in dimension one or two), and to study the asymptotics as $h \to 0$.
- In the context of Chapter 4, it remains open if the rates from Theorem 4.7 can be achieved for domains with arbitrary geometry.

Appendix A

Limit Theorems

The following martingale central limit theorem is a special case of [LS89, Theorem 5.5.4 (I)], [JS03, Theorem VIII.2.4].

Theorem A.1. *Let $(M^N)_{N\in\mathbb{N}}$ be a sequence of real-valued continuous local martingales with $M_0^N = 0$, let $T > 0$ such that $\langle M^N\rangle_T \xrightarrow{\mathbb{P}} 1$ for $N \to \infty$. Then $M_T^N \xrightarrow{d} \mathcal{N}(0,1)$ as $N \to \infty$.* □

We will repeatedly use the following version of the law of large numbers, which exploits Gaussianity:

Lemma A.2. *Let $(X_k^*)_{k\in\mathbb{N}}$ be independent centered Gaussian processes on $[0,T]$, set $Y_k^* := \int_0^T X_k^*(t)^2\mathrm{d}t$ and $Z_N^* = \sum_{k=1}^N Y_k^*$.*

(i) If $\mathrm{Var}(Y_k^*) \ll (\mathbb{E}Y_k^*)^2$ as $k \to \infty$, then $Y_k^*/\mathbb{E}Y_k^* \xrightarrow{\mathbb{P}} 1$.

(ii) If $\mathbb{E}Y_k^* \asymp Ck^\alpha$ as $k \to \infty$ for some constants $C > 0$, $\alpha \in \mathbb{R}$, then $Z_N^*/\mathbb{E}Z_N^* \xrightarrow{a.s.} 1$ as $N \to \infty$.

Proof. The first statement is a direct consequence of the Markov inequality:

$$\mathbb{P}\left(\left|\frac{Y_k^*}{\mathbb{E}Y_k^*} - 1\right| > \epsilon\right) \le \frac{\mathrm{Var}(Y_k^*)}{\epsilon^2(\mathbb{E}Y_k^*)^2} \to 0$$

for every $\epsilon > 0$. Now we prove (ii). As the (X_k^*) are Gaussian with mean zero, the Wick theorem [Jan97, Theorem 1.28] gives

$$
\begin{aligned}
\operatorname{Var}(Y_k^*) &= \int_0^T \int_0^T \mathbb{E}[X_k^*(t) X_k^*(t) X_k^*(s) X_k^*(s)] \\
&\qquad - \mathbb{E}[X_k^*(t) X_k^*(t)]\mathbb{E}[X_k^*(s) X_k^*(s)] \mathrm{d}s\mathrm{d}t \\
&= 2\int_0^T \int_0^T \mathbb{E}[X_k^*(t) X_k^*(s)]^2 \mathrm{d}s\mathrm{d}t \le 2\left(\mathbb{E}\int_0^T X_k^*(t)^2 \mathrm{d}t\right)^2 = 2(\mathbb{E}Y_k^*)^2.
\end{aligned} \tag{A.1}
$$

W.l.o.g. assume $\mathbb{E}Y_1^* > 0$, such that the denominator in the following estimates is positive. (Otherwise, if $\mathbb{E}Y_1^* = 0$, then (A.1) implies $Y_1^* = 0$ almost surely, and Y_1^* does not contribute to Z_N.) We have for $\alpha > -1$:

$$
\frac{\operatorname{Var}(Y_k^*)}{(\mathbb{E}Z_k^*)^2} \le \frac{2(\mathbb{E}Y_k^*)^2}{\left(\sum_{i=1}^k \mathbb{E}Y_k^*\right)^2} \asymp \frac{2C^2 k^{2\alpha}}{(C(\alpha+1)^{-1} k^{\alpha+1})^2} \lesssim \frac{1}{k^2}.
$$

Similarly, for $\alpha = -1$,

$$
\frac{\operatorname{Var}(Y_k^*)}{(\mathbb{E}Z_k^*)^2} \lesssim \frac{1}{k^2 \ln(k)^2}.
$$

Finally, if $\alpha < -1$, the denominator $(\mathbb{E}Z_k^*)^2$ converges for $k \to \infty$, and

$$
\frac{\operatorname{Var}(Y_k^*)}{(\mathbb{E}Z_k^*)^2} \lesssim (\mathbb{E}Y_k^*)^2 \lesssim k^{2\alpha} \ll \frac{1}{k^2}.
$$

In any case, we have $\sum_{k=1}^\infty \operatorname{Var}(Y_k^*)/(\mathbb{E}Z_k^*)^2 < \infty$, and the strong law of large numbers [Shi96, Theorem IV.3.2] implies the claim. $\square$

Appendix B

Additional Proofs

B.1 Proof of Proposition 2.17

We prove the statement in two parts.

Lemma B.1. *In the situation of Proposition 2.17, there is a unique solution X to* (2.39) *that satisfies $X \in R^{\mathbb{E}}(s)$ for $s = 1$.*

Proof. This is an application of the arguments in [LR15, Theorem 5.1.3]. More precisely, we show that the assumptions $(H1)$ (continuity) and $(H2')$ (monotonicity) therein are satisfied in the Gelfand triple $W_0^{1,2}(\mathcal{D}) \subset L^2(\mathcal{D}) \simeq L^2(\mathcal{D})^* \subset W_0^{1,2}(\mathcal{D})^*$ (i.e. $H_1 \subset H_0 \subset H_{-1}$), whereas $(H3)$ (coercivity) and $(H4')$ (boundedness) are satisfied in the shifted triple $W^{2,2}(\mathcal{D}) \cap W_0^{1,2}(\mathcal{D}) \subset W_0^{1,2}(\mathcal{D}) \subset L^2(\mathcal{D})$ (i.e. $H_2 \subset H_1 \subset H_0$). First note that in all cases considered, $\partial_x f : \mathbb{R} \to \mathbb{R}$ is bounded from above, and thus there is $c > 0$ such that for any $X \in L^2(\mathcal{D})$ and $Y : \mathcal{D} \to \mathbb{R}$:

$$\langle \partial_x f(Y) X, X \rangle_{L^2(\mathcal{D})} \leq c \left\| X \right\|^2_{L^2(\mathcal{D})}. \tag{B.1}$$

Since $\gamma > d/4 + 1/2$, $(-\Delta)^{1/2} B$ is a Hilbert–Schmidt operator on H, i.e. the dispersion operator B is a Hilbert–Schmidt operator from H to H_1. As B is constant, it suffices to test $(H1)$, $(H2')$, $(H3)$, $(H4')$ only for the drift of the SPDE (2.39).

$(H1)$ If f is a polynomial, this is a trivial consequence of the binomial theorem. On the other hand, if $f \in C_b^\infty(\mathbb{R})$ and $u, v, w \in H_1$, then $t \mapsto {}_{H_{-1}} \langle \theta \Delta(u + tv) + f(u + tv), w \rangle_{H_1}$ is continuous as a consequence of the linearity of Δ and the dominated convergence theorem.

($H2'$) This is a consequence of (B.1): For $u, v \in H_1$,

$$
\begin{aligned}
{}_{H_{-1}}\langle \theta\Delta u + f(u) - \theta\Delta v - f(v), u - v\rangle_{H_1} &\leq \langle f(u) - f(v), u - v\rangle_{H_0} \\
&= \langle \partial_x f(w)(u - v), u - v\rangle_{H_0} \\
&\leq c \left\| u - v \right\|_{H_0}^2
\end{aligned}
$$

for some $w : \mathcal{D} \to \mathbb{R}$.

($H3$) For $u \in H_2$,

$$
\begin{aligned}
{}_{L^2(\mathcal{D})}\langle \theta\Delta u + f(u), u\rangle_{H_2} &= -\theta \left\| u \right\|_{H_2}^2 + \langle f(u), u\rangle_{H_1} \\
&= -\theta \left\| u \right\|_{H_2}^2 + \langle \partial_x f(u)\nabla u, \nabla u\rangle_{L^2(\mathcal{D})} \\
&\leq -\theta \left\| u \right\|_{H_2}^2 + c \left\| u \right\|_{H_1}^2,
\end{aligned}
$$

where we used (B.1) componentwise in the last inequality.

($H4'$) First consider the case that f is a polynomial. W.l.o.g., assume $f(x) = x^k$ for $k \in \mathbb{N}$. Then for $u \in H_2$:

$$
\left\| f(u) \right\|_{L^2(\mathcal{D})}^2 = \left\| u \right\|_{L^{2k}(\mathcal{D})}^{2k},
$$

and for $d \leq 2$ this term is bounded by $\left\| u \right\|_{W^{1,2}(\mathcal{D})}^{2k} = \left\| u \right\|_{H_1}^{2k}$ up to a constant. In $d = 3$ this is still true if $k \leq 3$. This proves ($H4'$). Finally, if $f \in C_b^\infty(\mathbb{R})$, then $\left\| f(u) \right\|_{L^2(\mathcal{D})}^2 \leq |\mathcal{D}| \sup_{x\in\mathbb{R}} f(x)^2 < \infty$, and ($H4'$) is trivially satisfied.

Now as in [LR15, Lemma 5.1.4 and 5.1.5], ($H3$) and ($H4'$) imply that there is a sequence of finite dimensional approximations $X^{(n)}$ to the solution X which is bounded uniformly in $L^2(\Omega \times [0,T]; H_{s+1})$ and $L^p(\Omega; L^\infty(0,T; H_s))$ for $s = 1$, and such that $\theta\Delta X^{(n)} + f(X^{(n)})$ is bounded uniformly in $L^2(\Omega \times [0,T]; H_{s-1})$ for $s = 1$. In particular, these statements remain true for the (weaker) case $s = 0$. Based on these bounds for $s = 0$, the proof of [LR15, Theorem 5.1.3] transfers verbatim and yields a unique solution X to (2.39) with $X \in R^{\mathbb{E}}(0)$. The stronger a priori bounds ($s = 1$) imply that in fact $X \in R^{\mathbb{E}}(1)$, which concludes the proof. □

Lemma B.2. *In the situation of Proposition 2.17, there is $s > d/2$ such that $X \in R^{\mathbb{E}}(s)$.*

Proof. In $d = 1$, this has been proven in Lemma B.1, so let $d \in \{2, 3\}$. We apply the usual splitting argument and write $X = \bar{X} + \widetilde{X}$, where $\bar{X}$ is the solution to (2.39) with $f = 0$. Then $\bar{X} \in R^{\mathbb{E}}(s)$ for any $s < s^* = 1+2\gamma-d/2$, see [DPZ14, Section 5.3]. As $\gamma > d/4+1/2$, we have in particular $\bar{X} \in R^{\mathbb{E}}(2)$. As a consequence, the claim is proven once we know that for all $0 < \eta < 2$,

$$\widetilde{X} \in R^{\mathbb{E}}(\eta). \tag{B.2}$$

(i) For polynomial f, we assume w.l.o.g. $f(x) = x^k$, where k is arbitrary in $d = 2$ and $k \leq 3$ in $d = 3$. In this case, $\|f(X)\|_{L^2(\mathcal{D})} = \|X\|^k_{L^{2k}(\mathcal{D})} \lesssim \|X\|^k_{H_1}$. Consequently, $f(X) \in R^{\mathbb{E}}(0)$ because $X \in R^{\mathbb{E}}(1)$. Similar to the proof of Proposition 2.3, we estimate for $0 \leq t \leq T$:

$$\begin{aligned}\left\|\widetilde{X}_t\right\|_\eta &\leq \left\|e^{r\theta\Delta}X_0\right\|_\eta + \int_0^t \left\|e^{(t-r)\theta\Delta}f(X_r)\right\|_\eta \mathrm{d}r \\ &\lesssim \|X_0\|_\eta + \frac{2}{2-\eta}T^{1-\eta/2}\sup_{0\leq r\leq t}\|f(X_r)\|_{L^2(\mathcal{D})}.\end{aligned}$$

As $f(X) \in R^{\mathbb{E}}(0)$ and $\mathbb{E}\left[\|X_0\|^p_{s^*+2}\right] < \infty$ for any $p \geq 1$, we conclude that (B.2) holds true.

(ii) Let $f \in C_b^\infty(\mathbb{R})$. By Proposition 2.19 (ii), condition $(F_{s,\eta})$ holds for $s = 1$ and $0 < \eta < 2$. Thus, Proposition 2.3 (ii) implies (B.2).

□

This finishes the proof of Proposition 2.17.

B.2 Proof of Lemma 5.5

This section is an adaptation of the proof of [ACP20, Proposition 30], which, in turn, is a modification of [DPZ14, Theorem 5.25].

For $s < s^*$ and $\alpha > 0$, let $Y_t^{(s,\alpha)} := \int_0^t (t-r)^{-\alpha}(-\Delta)^{s/2}e^{(t-r)\theta\Delta}B\mathrm{d}W_r$. First, we prove:

Lemma B.3. *For all $s < s^*$, $0 < \alpha < (s^* - s)/2$ and $p \geq 2$, we have a.s. $Y^{(s,\alpha)} \in L^p(0, T; L^p(\mathcal{D}))$.*

Proof. For $x \in \mathcal{D}$, let δ_x be the point evaluation operator. We have for $x \in \mathcal{D}$, $0 \le t \le T$, using that $B^*(-\Delta)^\gamma$ is a bounded operator on $L^2(\mathcal{D})$:

$$\begin{aligned}
\mathbb{E}\left[Y_t^{(s,\alpha)}(x)^2\right] &= \int_0^t r^{-2\alpha}\left\|\delta_x(-\Delta)^{s/2}e^{r\theta\Delta}B\right\|_{\mathrm{HS}}^2 \mathrm{d}r \\
&= \int_0^t r^{-2\alpha}\left\|B^*(-\Delta)^\gamma e^{r\theta\Delta}(-\Delta)^{s/2-\gamma}\delta_x^*\right\|_{L^2(\mathcal{D})}^2 \mathrm{d}r \qquad \text{(B.3)} \\
&\lesssim \int_0^t r^{-2\alpha}\left\|\delta_x(-\Delta)^{s/2}e^{r\theta\Delta}(-\Delta)^{-\gamma}\right\|_{\mathrm{HS}}^2 \mathrm{d}r,
\end{aligned}$$

so w.l.o.g. we restrict to the case $B = (-\Delta)^{-\gamma}$. In that case, together with $\sup_{k\in\mathbb{N}}\|\Phi_k\|_{L^\infty(\mathcal{D})} < \infty$, a calculation as in Lemma 2.7 gives

$$\mathbb{E}\left[Y_t^{(s,\alpha)}(x)^2\right] = \sum_{k=1}^\infty \lambda_k^{s-2\gamma}\left(\int_0^t r^{-2\alpha}e^{-2\theta\lambda_k r}\mathrm{d}r\right)\Phi_k(x)^2 \lesssim \sum_{k=1}^\infty k^{\frac{2}{d}(s-2\gamma-1+2\alpha)},$$

which is finite[1] for $\alpha < (s^* - s)/2$. As $Y^{(s,\alpha)}$ is Gaussian,

$$\sup_{0\le t\le T, x\in\mathcal{D}} \mathbb{E}\left[\left|Y_t^{(s,\alpha)}(x)\right|^p\right] \lesssim \left(\sup_{0\le t\le T, x\in\mathcal{D}} \mathbb{E}\left[Y_t^{(s,\alpha)}(x)^2\right]\right)^{p/2} < \infty.$$

This leads to

$$\mathbb{E}\int_0^T\int_{\mathcal{D}}\left|Y_t^{(s,\alpha)}(x)\right|^p \mathrm{d}x\mathrm{d}t \le T\,|\mathcal{D}| \sup_{0\le t\le T, x\in\mathcal{D}} \mathbb{E}\left[\left|Y_t^{(s,\alpha)}(x)\right|^p\right] < \infty,$$

proving the claim. □

Proof of Lemma 5.5. Using the factorization formula [DPZ14, Theorem 5.10], we obtain from Lemma B.3 together with [DPZ14, Proposition 5.9] that $(-\Delta)^{s/2}\bar{X} \in C(0,T;L^p(\mathcal{D}))$ for all $s < s^*$ and $p \ge 2$ such that $1/p < (s^* - s)/2$. This means that $\bar{X} \in R_p(s)$ for all $p \ge 2$ and $s < s^* - 2/p$. As $p \ge 2$ is arbitrary, this finishes the proof. □

B.3 Proof of Proposition 6.6

The arguments are similar as in Appendix B.1, the main difference being the new inhibitor component and the concentration dependent unstable zero in the reaction polynomial. For $d \le 2$, the proof can be found in [PFA+21].

[1] In particular, the terms involving the point evaluation δ_x in (B.3) are finite.

We write $\mathscr{H}_s := H_s \oplus H_s$ for the regularity spaces describing both components[2]. Similarly to Lemma B.1, we work in the Hilbert space triples $\mathscr{H}_1 \subset \mathscr{H}_0 \subset \mathscr{H}_{-1}$ and $\mathscr{H}_2 \subset \mathscr{H}_1 \subset \mathscr{H}_0$. Further, with $f(u,z) = k_1 u(u_0 - u)(u - u_0 a(z))$, we write $A_1(U,V) = D_U \Delta U + f(U, \|U\|_{L^2(\mathcal{D})}) - k_2 V$ and $A_2(U,V) = D_V \Delta V + \epsilon(bU - V)$ as well as $\mathscr{A}(U,V) = (A_1(U,V), A_2(U,V))$. Similarly to (B.1), we have for $U \in L^2(\mathcal{D})$, $Y : \mathcal{D} \to \mathbb{R}$ and $z \in \mathbb{R}$:

$$\langle \partial_u f(Y,z) U, U \rangle_{L^2(\mathcal{D})} \le c \|U\|^2_{L^2(\mathcal{D})}, \tag{B.4}$$

because a is a bounded function. As $B = \sigma(-\Delta)^{-\gamma}$ and A_2 is linear, it suffices to consider A_1 in order to show $(H1)$, $(H2')$, $(H3)$, $(H4')$ from [LR15]. For $(H2')$, we have to take into account the dependence of f on the overall concentration: Using (B.4),

$$\begin{aligned}
&{}_{H_{-1}}\langle A_1(U_1,V_1) - A_1(U_2,V_2), U_1 - U_2\rangle_{H_1} \\
&\qquad \lesssim \Big\langle f(U_1, \|U_1\|_{L^2(\mathcal{D})}) - f(U_2, \|U_2\|_{L^2(\mathcal{D})}), U_1 - U_2 \Big\rangle_{L^2(\mathcal{D})} \\
&\qquad\quad + k_2 \|V_1 - V_2\|_{L^2(\mathcal{D})} \|U_1 - U_2\|_{L^2(\mathcal{D})} \\
&\qquad \lesssim \Big\langle \partial_u f(Y, \|U_1\|_{L^2(\mathcal{D})})(U_1 - U_2), U_1 - U_2 \Big\rangle_{L^2(\mathcal{D})} \\
&\qquad\quad + \Big\langle \partial_z f(U_2, \tilde{z}) \Big(\|U_1\|_{L^2(\mathcal{D})} - \|U_2\|_{L^2(\mathcal{D})} \Big), U_1 - U_2 \Big\rangle_{L^2(\mathcal{D})} \\
&\qquad\quad + k_2 \|V_1 - V_2\|_{L^2(\mathcal{D})} \|U_1 - U_2\|_{L^2(\mathcal{D})} \\
&\qquad \lesssim \|U_1 - U_2\|^2_{L^2(\mathcal{D})} + \|V_1 - V_2\|^2_{L^2(\mathcal{D})} \\
&\qquad\quad + \|\partial_z f(U_2, \tilde{z})\|_{L^2(\mathcal{D})} \Big| \|U_1\|_{L^2(\mathcal{D})} - \|U_2\|_{L^2(\mathcal{D})} \Big| \|U_1 - U_2\|_{L^2(\mathcal{D})} \\
&\qquad \lesssim \Big(1 + \|\partial_z f(U_2, \tilde{z})\|_{L^2(\mathcal{D})} \Big) \|(U_1,V_1) - (U_2,V_2)\|^2_{L^2(\mathcal{D}) \oplus L^2(\mathcal{D})}
\end{aligned}$$

for some $Y : \mathcal{D} \to \mathbb{R}$ and $\tilde{z} \in \mathbb{R}$. Further, using that $\partial_z a$ is a bounded function as well as the Sobolev embedding $W^{1,2}(\mathcal{D}) \subset L^4(\mathcal{D})$,

$$\begin{aligned}
\|\partial_z f(U_2, \tilde{z})\|_{L^2(\mathcal{D})} &\lesssim \|U_2(u_0 - U_2)\|_{L^2(\mathcal{D})} \lesssim \|U_2\|_{L^2(\mathcal{D})} + \|U_2\|^2_{L^4(\mathcal{D})} \\
&\lesssim \left(1 + \|U_2\|_{H_1}\right)^2.
\end{aligned}$$

Therefore we can take $\rho(U,V) = c(1 + \|U\|_{H_1})^2$ for some $c > 0$ in the notation of [LR15], and $(H2')$ is satisfied.

[2]Note that this is different from that notation in Section 2.6 as the regularity of the inhibitor component is taken into account.

Now, $(H1)$, $(H3)$ and $(H4')$ work exactly as in Lemma B.1, with obvious modifications in notation due to the presence of the inhibitor component, taking into account that a is continuous and bounded. As a consequence, we have the following analogon to Lemma B.1:

Lemma B.4. *In the situation of Proposition 6.6, there is a unique solution U, V to* (6.14), (6.15) *with $U, V \in R^{\mathbb{E}}(1)$.* $\square$

We can represent the inhibitor as $V_t = e^{t(D_V\Delta - \epsilon I)}V_0 - \epsilon b F_3(U)(t) = e^{t(D_V\Delta-\epsilon I)}V_0 + \epsilon b \int_0^t e^{(t-r)(D_V\Delta-\epsilon I)}U_r \mathrm{d}r$. Note that F_3 satisfies $(F_{s,\eta})$ for every $s \in \mathbb{R}$ and $\eta < 4$: With $\varepsilon = 2 - \eta/2$,

$$\begin{aligned}\sup_{0\le t\le T} \|F_3(U)(t)\|_{s+\eta+\varepsilon-2} &\lesssim \sup_{0\le t\le T}\int_0^t \left\|e^{(t-r)(D_V\Delta-\epsilon I)}U_r\right\|_{s+2-\varepsilon} \mathrm{d}r \\ &\lesssim \sup_{0\le t\le T}\int_0^t (t-r)^{-1+\varepsilon/2}\|U_r\|_s \,\mathrm{d}r \lesssim \frac{2}{\varepsilon}T^{\varepsilon/2}\|U\|_{R(s)}.\end{aligned}$$

Further, $\mathbb{E}[\|V_0\|^p_{s^*+2}] < \infty$ for all $p \ge 2$. Consequently, for all $s < s^*$ and $\varepsilon > 0$, we have $V \in R^{\mathbb{E}}(s+2-\varepsilon)$ whenever $U \in R^{\mathbb{E}}(s)$. In particular, from Lemma B.4 it follows that $V \in R^{\mathbb{E}}(3-\varepsilon)$.

Now, exactly as in Lemma B.2 we see that there is some $s > d/2$ such that $U \in R^{\mathbb{E}}(s)$, taking into account that a is bounded and $V \in R^{\mathbb{E}}(3-\varepsilon)$ for $\varepsilon > 0$. Finally, it is clear that $U \mapsto f(U, \|U\|_{L^2(\mathcal{D})})$ satisfies $(F_{s,\eta})$ for $d/2 < s < s^*$ and $\eta < 2$, so the same is true for $F(U) = f(U, \|U\|_{L^2(\mathcal{D})}) - k_2(e^{(\cdot)(D_V\Delta-\epsilon I)}V_0 - \epsilon b F_3(U))$. Thus, an application of Proposition 2.4 finishes the proof of Proposition 6.6.

List of Figures

All of these plots appear in one of the works on which this dissertation is based, namely [PS20] (Figure 2.1) and [PFA$^+$21] (Figure 6.1 to 6.6).

Notation

Assumptions

(W)	well-posedness of the SPDE (p. 18)
$(F_{s,\eta})$	regularity bound on the nonlinearity F (p. 21)
$(F^{\mathrm{v}}_{s,\eta})$	bound on F in variational spaces $R^{\mathrm{v}}(s)$ (p. 23)
$(F^{p}_{s,\eta})$	L^p-regularity bound on the nonlinearity F (p. 108)
$(F^{\mathrm{par}}_{s,\eta})$	regularity bound on parametrized nonlinearity F (p. 119)
$(F^{\mathrm{sys}}_{s,\eta})$	analogon of $(F_{s,\eta})$ for partially observed systems (p. 47)
$(F^{\mathcal{J}}_{s,\eta})$	bound for the integrated nonlinearity JF (p. 59)
(N^{γ}_{η})	dispersion B is asymptotically close to $\bar{B} = (-A)^{-\gamma}$ (p. 78)
(D_0)	B_r is an algebra of continuous functions (p. 86)
(D_1)	growth bound on the norm of the eigenfunctions (p. 86)
(D_2)	error bound for the integral discretization error (p. 86)
(D^*_2)	trigonometric interpolation error (p. 101)
(L_B)	local control on the dispersion (p. 107)
(L_K)	shape of the kernel (p. 107)
(L_Ψ)	non-degeneracy within the local approach (p. 107)
(I_α)	linear independence of the nonlinear components (p. 121)

Asymptotics

$a_N \sim b_N$	There is $C > 0$ such that $a_N/b_N \to C$ for $N \to \infty$.
$a_N \asymp b_N$	$a_N/b_N \to 1$ for $N \to \infty$.
$a_N \lesssim b_N$	There is $C > 0$ such that $a_N \leq C b_N$.
$a_N \ll b_N$	$a_N = o(b_N)$, i.e. $a_N/b_N \to 0$ for $N \to \infty$.
$a_N \ll_p b_N$	There is $\epsilon > 0$ such that $a_N \ll b_N N^{-\epsilon}$. (polynomial negligibility)

Similar notation is used for different asymptotic parameters (i.e. h, δ).

Vector Spaces, Norms, Scalar Products

A (fixed) norm on a Banach space B is denoted by $\|\cdot\|_B$.
If B is even a Hilbert space, the corresponding scalar product is $\langle \cdot, \cdot \rangle_B$.
Frequently used norms are abbreviated:

H	Hilbert space, typically $L^2(\mathcal{D})$ (state space)
$\|\cdot\|, \langle \cdot, \cdot \rangle$	norm and scalar product on H
H_s	$D((-A)^{s/2})$ (scale of regularity spaces)
$\|\cdot\|_s, \langle \cdot, \cdot \rangle_s$	norm and scalar product on H_s
V	$H_1 = D((-A)^{1/2})$ (energy space)
$R(s)$	$L^\infty(0, T; H_s)$
$R^{\mathbb{E}}(s)$	$\bigcap_{p \geq 1} L^p(\Omega, R(s)) = \bigcap_{p \geq 1} L^p(\Omega, L^\infty(0, T; H_s))$ (locally convex space)
$R^{\mathrm{v}}(s)$	$L^\infty(0, T; H_{s-1}) \cap L^2(0, T; H_s)$
$\|\cdot\|_{\mathrm{HS}}$	Hilbert–Schmidt norm of an operator acting on H
$H^{s,p}(\mathcal{D})$	domain of $(-\Delta)^{s/2}$ in $L^p(\mathcal{D})$
$\|\cdot\|_{s,p}$	canonical norm on $H^{s,p}(\mathcal{D})$
$R_p(s)$	$L^\infty(0, T; H^{s,p}(\mathcal{D}))$
$\|\cdot\|_{(h)}, \langle \cdot, \cdot \rangle_{(h)}$	Euclidean norm and scalar product on $\mathbb{R}^{M_h}$

Estimators

Temporally white noise (Chapter 2):	$\hat{\theta}_N^{\text{full}}$, $\hat{\theta}_N^{\text{part}}$, $\hat{\theta}_N^{\text{lin}}$ (p. 27 f.)
Ornstein–Uhlenbeck noise (Section 3.1):	$\hat{\theta}_N^{\text{ref}}$, $\hat{\mu}_N^{\text{ref}}$ (p. 61)
	$\hat{\theta}_N^{\text{sim}}$ (p. 62)
	$\hat{\mu}_N^{\text{full}}(\vartheta)$, $\hat{\mu}_N^{\text{lin}}(\vartheta)$ (p. 65 f.)
Integrated noise (Section 3.2):	$\hat{\theta}_N^{\text{rescaled}}$ (p. 74)
Discrete observations (Chapter 4):	$\hat{\theta}_{h,N}^{\text{discr}}$ (p. 87)
Local observations (Chapter 5):	$\hat{\theta}_{\delta,x_0}$ (p. 106)
	$\hat{\theta}_N^{\text{mode}}$ (p. 114)
Joint parameter estimation (Section 6.1):	$\hat{\theta}^N = (\hat{\theta}_0^N, \cdots, \hat{\theta}_K^N)$ (p. 119)
Activator–inhibitor model (Section 6.2):	$\hat{\theta}_0^{\text{lin},N}$, $\hat{\theta}_0^{\text{pol},N}$, $\hat{\theta}_0^{\text{full},N}$ (p. 128)
	$\hat{\theta}_0^{2,N}$, $\hat{\theta}_0^{3,N}$, $\hat{\theta}_0^{4,N}$ (p. 128 f.)
	$\hat{\theta}_1^{2,N}$ (p. 137)

Constants and Further Notation

Fourier decomposition of A:

λ_N	eigenvalue of $-A$
Φ_N	eigenfunction of $-A$
P_N	projection onto the span of $\Phi_1, \dots, \Phi_N$, defined on any H_s

Frequently used constants:

β	determined by $\lambda_N \sim N^\beta$, usually $\beta = 2/d$
Λ	proportionality constant given by $\lambda_N \asymp \Lambda N^\beta$, depends on $\mathcal{D}$
γ	degree of spatial correlation in the noise
s^*	optimal regularity of the solution process, i.e. $X \in R(s)$ if and only if $s < s^*$ (usually $s^* = 1 + 2\gamma - 1/\beta$)

Further notation:

J	Bochner integral operator $f \mapsto Jf = \int_0^\cdot f(r)\mathrm{d}r$

Bibliography

[AB88] S. I. Aihara and A. Bagchi, *Parameter Identification for Stochastic Diffusion Equations with Unknown Boundary Conditions*, Appl Math Optim **17** (1988), 15–36.

[ABJR21] Randolf Altmeyer, Till Bretschneider, Josef Janák, and Markus Reiß, *Parameter Estimation in an SPDE Model for Cell Repolarisation*, arXiv:2010.06340v2 [math.ST] (2021), preprint.

[ACP20] Randolf Altmeyer, Igor Cialenco, and Gregor Pasemann, *Parameter estimation for semilinear SPDEs from local measurements*, arXiv:2004.14728v2 [math.ST] (2020), preprint.

[AF92] David R. Adams and Michael Frazier, *Composition operators on potential spaces*, Proc. Amer. Math. Soc. **114** (1992), no. 1, 155–165.

[AF03] Robert A. Adams and John J. F. Fournier, *Sobolev spaces*, second ed., Pure and Applied Mathematics, vol. 140, Elsevier/Academic Press, 2003.

[Aih92] S. I. Aihara, *Regularized Maximum Likelihood Estimate for an Infinite-dimensional Parameter in Stochastic Parabolic Systems*, SIAM J. Control and Optimization **30** (1992), no. 4, 745–764.

[Aih98a] ———, *Consistency property of extended least-squares parameter estimation for stochastic diffusion equation*, Systems & Control Letters **34** (1998), 249–256.

[Aih98b] ———, *Identification of a Discontinuous Parameter in Stochastic Parabolic Systems*, Appl Math Optim **37** (1998), 43–69.

[AR21] Randolf Altmeyer and Markus Reiß, *Nonparametric Estimation for Linear SPDEs from Local Measurements*, Ann. Appl. Probab. **31** (2021), no. 1, 1 – 38.

[AS88] S. I. Aihara and Y. Sunahara, *Identification of an Infinite-dimensional Parameter for Stochastic Diffusion Equations*, SIAM J. Control and Optimization **26** (1988), no. 5, 1062–1075.

[ASB18] Sergio Alonso, Maike Stange, and Carsten Beta, *Modeling random crawling, membrane deformation and intracellular polarity of motile amoeboid cells*, PLOS ONE **13** (2018), no. 8, 1–22.

[AT17] Anthony P. Austin and Lloyd N. Trefethen, *Trigonometric Interpolation and Quadrature in Perturbed Points*, SIAM J. Numer. Anal. **55** (2017), no. 5, 2113–2122.

[Aus16] Anthony P. Austin, *Some New Results on and Applications of Interpolation in Numerical Computation*, University of Oxford, 2016, DPhil thesis.

[BB84] A. Bagchi and V. Borkar, *Parameter identification in infinite dimensional linear systems*, Stochastics **12** (1984), 201–213.

[BBPSP14] L. Blanchoin, R. Boujemaa-Paterski, C. Sykes, and J. Plastino, *Actin Dynamics, Architecture, and Mechanics in Cell Motility*, Physiol Rev **94** (2014), no. 1, 235–263.

[BC09] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability, vol. 60, Springer Science+Business Media, LLC, 2009.

[Bis99] J. P. N. Bishwal, *Bayes and Sequential Estimation in Hilbert Space Valued Stochastic Differential Equations*, Journal of the Korean Statistical Society **28** (1999), no. 1, 93–106.

[Bis02] ———, *The Bernstein-von Mises Theorem and Spectral Asymptotics of Bayes Estimators for Parabolic SPDEs*, J. Austral. Math. Soc. **72** (2002), 287–298.

[BL76] J. Bergh and J. Löfström, *Interpolation Spaces (An Introduction)*, Grundlehren der mathematischen Wissenschaften, vol. 223, Springer-Verlag Berlin Heidelberg, 1976.

[BM18] Haïm Brezis and Petru Mironescu, *Gagliardo–Nirenberg inequalities and non-inequalities: The full story*, Ann. Inst. H. Poincaré Anal. Non Linéaire **35** (2018), no. 5, 1355–1376.

[BS17] Matthew D. Blair and Christopher D. Sogge, *Refined and Microlocal Kakeya-Nikodym Bounds of Eigenfunctions in Higher Dimensions*, Comm. Math. Phys. **356** (2017), no. 2, 501–533.

[BST80] A. Bagchi, R. C. W. Strijbos, and G. Thé, *Identification of a distributed-parameter system with boundary noise*, Int. J. Systems Sci. **11** (1980), no. 1, 49–56.

[BT19] Markus Bibinger and Mathias Trabs, *On Central Limit Theorems for Power Variations of the Solution to the Stochastic Heat Equation*, Stochastic Models, Statistics and Their Applications (A. Steland, E. Rafajłowicz, and O. Okhrin, eds.), Springer Proceedings in Mathematics & Statistics, vol. 294, Springer, Cham, 2019, pp. 69–84.

[BT20] ———, *Volatility estimation for stochastic PDEs using high-frequency observations*, Stochastic Process. Appl. **130** (2020), no. 5, 3005–3052.

[CA77] J. W. Cahn and S. M. Allen, *A Microscopic Theory for Domain Wall Motion and its Experimental Verification in Fe-Al Alloy Domain Growth Kinetics*, J. Phys. Colloques **38** (1977), no. C7, C7–51 – C7–54.

[CCG20] Ziteng Cheng, Igor Cialenco, and Ruoting Gong, *Bayesian estimations for diagonalizable bilinear SPDEs*, Stochastic Process. Appl. **130** (2020), no. 2, 845–877.

[CD20] Carsten Chong and Robert C. Dalang, *Power Variations in Fractional Sobolev Spaces for a Class of Parabolic Stochastic PDEs*, arXiv:2006.15817v1 [math.PR] (2020), preprint.

[CDVK20] Igor Cialenco, Francisco Delgado-Vences, and Hyun-Jung Kim, *Drift estimation for discretely sampled SPDEs*, Stoch PDE: Anal Comp **8** (2020), no. 4, 895–920.

[Cer01] Sandra Cerrai, *Second order PDE's in finite and infinite dimension*, Lecture Notes in Mathematics, vol. 1762, Springer-Verlag, Berlin, 2001.

[CGH11] Igor Cialenco and Nathan Glatt-Holtz, *Parameter estimation for the stochastically perturbed Navier-Stokes equations*, Stochastic Process. Appl. **121** (2011), no. 4, 701–724.

[CGH18] Igor Cialenco, Ruoting Gong, and Yicong Huang, *Trajectory fitting estimators for SPDEs driven by additive noise*, Stat Inference Stoch Process **21** (2018), no. 1, 1–19.

[CH20] Igor Cialenco and Yicong Huang, *A note on parameter estimation for discretely sampled SPDEs*, Stochastics and Dynamics **20** (2020), no. 3, 2050016.

[Che93] Xia Chen, *On the law of the iterated logarithm for independent Banach space valued random variables*, The Annals of Probability **21** (1993), no. 4, 1991–2011.

[Cho19] Carsten Chong, *High-frequency analysis of parabolic stochastic PDEs with multiplicative noise: Part I*, arXiv:1908.04145v1 [math.PR] (2019), preprint.

[Cho20] ———, *High-frequency analysis of parabolic stochastic PDEs*, The Annals of Statistics **48** (2020), no. 2, 1143 – 1167.

[CHQZ88] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics*, Springer Series in Computational Physics, Springer-Verlag Berlin Heidelberg, 1988.

[Cia02] Philippe G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics in Applied Mathematics, vol. 40, Society for Industrial and Applied Mathematics, 2002, reprint of the 1978 edition.

[Cia10] Igor Cialenco, *Parameter estimation for SPDEs with multiplicative fractional noise*, Stoch. Dyn. **10** (2010), no. 4, 561–576.

[Cia18] ———, *Statistical inference for SPDEs: an overview*, Stat Inference Stoch Process **21** (2018), no. 2, 309–329.

[CK22] Igor Cialenco and Hyun-Jung Kim, *Parameter estimation for discretely sampled stochastic heat equation driven by space-only noise*, Stochastic Process. Appl. **143** (2022), 1–30.

[CKL20] Igor Cialenco, Hyun-Jung Kim, and Sergey V. Lototsky, *Statistical analysis of some evolution equations driven by space-only noise*, Stat Inference Stoch Process **23** (2020), no. 1, 83–103.

[CKP21] Igor Cialenco, Hyun-Jung Kim, and Gregor Pasemann, *Statistical analysis of discretely sampled semilinear SPDEs: a power variation approach*, arXiv:2103.04211v1 [math.PR] (2021), preprint.

[CL09] Igor Cialenco and Sergey V. Lototsky, *Parameter estimation in diagonalizable bilinear stochastic parabolic equations*, Stat Inference Stoch Process **12** (2009), no. 3, 203–219.

[CLP09] Igor Cialenco, Sergey V. Lototsky, and Jan Pospíšil, *Asymptotic properties of the maximum likelihood estimator for stochastic parabolic equations with additive fractional Brownian motion*, Stoch. Dyn. **9** (2009), no. 2, 169–185.

[CX14] Igor Cialenco and Liaosha Xu, *A note on error estimation for hypothesis testing problems for some linear SPDEs*, Stoch PDE: Anal Comp **2** (2014), no. 3, 408–431.

[CX15] ______, *Hypothesis testing for stochastic PDEs driven by additive noise*, Stochastic Process. Appl. **125** (2015), no. 3, 819–866.

[DMPD00] T. E. Duncan, B. Maslowski, and B. Pasik-Duncan, *Adaptive Control for Semilinear Stochastic Systems*, SIAM J. Control Optim. **38** (2000), no. 6, 1683–1706.

[DPDT94] Giuseppe Da Prato, Arnaud Debussche, and Roger Temam, *Stochastic Burgers' equation*, NoDEA **1** (1994), no. 4, 389–402.

[DPZ14] Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic Equations in Infinite Dimensions*, second ed., Encyclopedia of Mathematics and its Applications, vol. 152, Cambridge University Press, 2014.

[EN00] Klaus-Jochen Engel and Rainer Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Graduate Texts in Mathematics, vol. 194, Springer-Verlag New York, 2000.

[FFAB20] Sven Flemming, Francesc Font, Sergio Alonso, and Carsten Beta, *How cortical waves drive fission of motile cells*, Proceedings of the National Academy of Sciences **117** (2020), no. 12, 6330–6338.

[Fit61] R. FitzHugh, *Impulses and Physiological States in Theoretical Models of Nerve Membrane*, Biophys. J. **1** (1961), 445–466.

[GEW+14] M. Gerhardt, M. Ecke, M. Walz, A. Stengl, C. Beta, and G. Gerisch, *Actin and PIP3 waves in giant cells reveal the inherent length scale of an excited state*, Journal of Cell Science **127** (2014), no. 20, 4507–4517.

[GM02] B. Goldys and B. Maslowski, *Parameter Estimation for Controlled Semilinear Stochastic Systems: Identifiability and Consistency*, Journal of Multivariate Analysis **80** (2002), 322–343.

[GN15] E. Giné and R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2015.

[Gri92] D. Grieser, *L^p bounds for eigenfunctions and spectral projections of the Laplacian near concave boundaries*, ProQuest LLC, Ann Arbor, MI, 1992, Ph.D. thesis, University of California, Los Angeles.

[Gri02] ______, *Uniform bounds for eigenfunctions of the Laplacian on manifolds with boundary*, Commun. in Partial Differential Equations **27** (2002), no. 7-8, 1283–1299.

[GT01] David Gilbarg and Neil S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics in Mathematics, Springer-Verlag Berlin Heidelberg, 2001, reprint of the 1998 edition.

[HKR93] M. Huebner, R. Khasminskii, and B. L. Rozovskii, *Two Examples of Parameter Estimation for Stochastic Partial Differential Equations*, Stochastic Processes (S. Cambanis, J. K. Ghosh, R. L. Karandikar, and P. K. Sen, eds.), Springer-Verlag New York, 1993, pp. 149–160.

[HL84] W. Horsthemke and R. Lefever, *Noise-Induced Transitions (Theory and Applications in Physics, Chemistry, and Biology)*, Springer Series in Synergetics, Springer-Verlag Berlin Heidelberg, 1984.

[HL00a] M. Huebner and S. Lototsky, *Asymptotic analysis of a kernel estimator for parabolic SPDE's with time-dependent coefficients*, Ann. Appl. Probab. **10** (2000), no. 4, 1246–1258.

[HL00b] ———, *Asymptotic Analysis of the Sieve Estimator for a Class of Parabolic SPDEs*, Scand J Statist **27** (2000), no. 2, 353–370.

[HLR97] M. Huebner, S. Lototsky, and B. L. Rozovskii, *Asymptotic Properties of an Approximate Maximum Likelihood Estimator for Stochastic PDEs*, Statistics and Control of Stochastic Processes (Yu. M. Kabanov, B. L. Rozovskii, and A. N. Shiryaev, eds.), World Scientific, 1997, pp. 139–155.

[HR95] M. Huebner and B. L. Rozovskii, *On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's*, Probab. Theory Relat. Fields **103** (1995), no. 2, 143–163.

[HT21a] Florian Hildebrandt and Mathias Trabs, *Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations*, arXiv:2102.13415v1 [math.ST] (2021), preprint.

[HT21b] ———, *Parameter estimation for SPDEs based on discrete observations in time and space*, Electronic Journal of Statistics **15** (2021), no. 1, 2716–2776.

[Hue93] M. Huebner, *Parameter estimation for stochastic differential equations*, ProQuest LLC, Ann Arbor, MI, 1993, Ph.D. thesis, University of Southern California.

[Hue99] ———, *Asymptotic Properties of the Maximum Likelihood Estimator for Stochastic PDEs Disturbed by Small Noise*, Statistical Inference for Stochastic Processes **2** (1999), no. 1, 57–68.

[Hui14] Jiang Hui, *Moderate deviation for parameter estimator in the stochastic parabolic equations with additive fractional Brownian motion*, Stochastics and Dynamics **14** (2014), no. 3, 1450002.

[IH81] I. A. Ibragimov and R. Z. Has'minskiĭ, *Statistical Estimation (Asymptotic Theory)*, Applications of Mathematics, vol. 16, Springer-Verlag, New York-Berlin, 1981.

[IK99] I. A. Ibragimov and R. Z. Khas'minskiĭ, *Estimation Problems for Coefficients of Stochastic Partial Differential Equations. Part I*, Theory Probab. Appl. **43** (1999), no. 3, 370–387.

[IK00] ———, *Estimation Problems for Coefficients of Stochastic Partial Differential Equations. Part II*, Theory Probab. Appl. **44** (2000), no. 3, 469–494.

[IK01] ———, *Estimation Problems for Coefficients of Stochastic Partial Differential Equations. Part III*, Theory Probab. Appl. **45** (2001), no. 2, 210–232.

[Jan97] Svante Janson, *Gaussian Hilbert Spaces*, Cambridge Tracts in Mathematics, Cambridge University Press, 1997.

[Jan20] Josef Janák, *Parameter estimation for stochastic partial differential equations of second order*, Appl. Math. Optim. **82** (2020), 353–397.

[Jan21] ———, *Parameter Estimation for Stochastic Wave Equation Based on Observation Window*, Acta Appl. Math. **172** (2021), no. 2, 37 pages.

[JS03] Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften, vol. 288, Springer-Verlag, Berlin, 2003.

[KL85] T. Koski and W. Loges, *Asymptotic Statistical Inference for a Stochastic Heat Flow Problem*, Statistics & Probability Letters **3** (1985), 185–189.

[KL86] ———, *On minimum-contrast estimation for Hilbert space-valued stochastic differential equations*, Stochastics **16** (1986), no. 3-4, 217–225.

[KLBR00] M. L. Kleptsyna, A. Le Breton, and M.-C. Roubaud, *Parameter Estimation and Optimal Filtering for Fractional Type Stochastic Systems*, Stat. Inference Stoch. Process. **3** (2000), no. 1-2, 173–182.

[KM19] P. Kříž and B. Maslowski, *Central limit theorems and minimum-contrast estimators for linear stochastic evolution equations*, Stochastics **91** (2019), 1109–1140.

[KO79] Heinz-Otto Kreiss and Joseph Oliger, *Stability of the Fourier method*, SIAM J. Numer. Anal. **16** (1979), no. 3, 421–433.

[Kří20] P. Kříž, *A space-consistent version of the minimum-contrast estimator for linear stochastic evolution equations*, Stochastics and Dynamics **20** (2020), no. 3, 2050019.

[Kry96] N. V. Krylov, *On L_p-Theory of Stochastic Partial Differential Equations in the Whole Space*, SIAM Journal on Mathematical Analysis **27** (1996), no. 2, 313–340.

[KU21a] Yusuke Kaino and Masayuki Uchida, *Adaptive estimator for a parabolic linear SPDE with a small noise*, Jpn J Stat Data Sci (2021), 29 pages.

[KU21b] Yusuke Kaino and Masayuki Uchida, *Parametric estimation for a parabolic linear SPDE model based on discrete observations*, J. Statist. Plann. Inference **211** (2021), 190–220.

[KUP91] P. Kumar, T. E. Unny, and K. Ponnambalam, *Stochastic partial differential equations in groundwater hydrology (Part 2: Application to Borden aquifer)*, Stochastic Hydrol. Hydraul. **5** (1991), 239–251.

[Kut04] Yury A. Kutoyants, *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London Ltd., 2004.

[LL10a] W. Liu and S. V. Lototsky, *Estimating speed and damping in the stochastic wave equation*, Stochastic partial differential equations and applications, Quad. Mat., vol. 25, Dept. Math., Seconda Univ. Napoli, Caserta, 2010, pp. 191–206.

[LL10b] ______, *Parameter estimation in hyperbolic multichannel models*, Asymptot. Anal. **68** (2010), no. 4, 223–248.

[LM72] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Die Grundlehren der mathematischen Wissenschaften, vol. 181, Springer-Verlag, New York-Heidelberg, 1972.

[Log84] Wilfried Loges, *Girsanov's theorem in Hilbert space and an application to the statistics of Hilbert space-valued stochastic differential equations*, Stochastic Process. Appl. **17** (1984), no. 2, 243–263.

[Lot96] S. V. Lototsky, *Problems in statistics of stochastic differential equations*, ProQuest LLC, Ann Arbor, MI, 1996, Ph.D. thesis, University of Southern California.

[Lot03] ______, *Parameter estimation for stochastic parabolic equations: asymptotic properties of a two-dimensional projection-based estimator*, Stat. Inference Stoch. Process. **6** (2003), no. 1, 65–87.

[Lot04] ______, *Optimal filtering of stochastic parabolic equations*, Recent developments in stochastic analysis and related topics (S. Albeverio, Z.-M. Ma, and M. Röckner, eds.), World Scientific, 2004, pp. 330–353.

[Lot09] ______, *Statistical inference for stochastic parabolic equations: a spectral approach*, Publ. Mat. **53** (2009), no. 1, 3–45.

[LPS14] G. J. Lord, C. E. Powell, and T. Shardlow, *An Introduction to Computational Stochastic PDEs*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2014.

[LR99] S. V. Lototsky and B. L. Rosovskii, *Spectral asymptotics of some functionals arising in statistical inference for SPDEs*, Stochastic Process. Appl. **79** (1999), no. 1, 69–94.

[LR00] ———, *Parameter Estimation for Stochastic Evolution Equations with Non-commuting Operators*, Skorokhod's Ideas in Probability Theory (V. Korolyuk, N. Portenko, and H. Syta, eds.), Institute of Mathematics of the National Academy of Sciences of Ukraine, Kiev, 2000, pp. 271–280.

[LR15] Wei Liu and Michael Röckner, *Stochastic Partial Differential Equations: An Introduction*, Universitext, Springer, Cham, 2015.

[LS77] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes. I (General Theory)*, Applications of Mathematics, vol. 5, Springer-Verlag New York, 1977.

[LS89] ———, *Theory of Martingales*, Mathematics and its Applications (Soviet Series), vol. 49, Kluwer Academic Publishers, Dordrecht, 1989.

[LS01] ———, *Statistics of Random Processes. II (Applications)*, 2nd ed., Applications of Mathematics, vol. 6, Springer-Verlag Berlin Heidelberg, 2001.

[Lun95] Alessandra Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Modern Birkhäuser Classics, Birkhäuser/Springer Basel, 1995.

[Mar03] Bo Markussen, *Likelihood inference for a discretely observed stochastic partial differential equation*, Bernoulli **9** (2003), no. 5, 745–762.

[MFF$^+$20] Eduardo Moreno, Sven Flemming, Francesc Font, Matthias Holschneider, Carsten Beta, and Sergio Alonso, *Modeling cell crawling strategies with a bistable model: From amoeboid to fan-shaped cell motion*, Physica D: Nonlinear Phenomena **412** (2020), 132591.

[Mis08] Yuliya S. Mishura, *Stochastic calculus for fractional Brownian motion and related processes*, Lecture Notes in Mathematics, vol. 1929, Springer-Verlag Berlin Heidelberg, 2008.

[MKT19a] Z. Mahdi Khalil and C. A. Tudor, *Estimation of the drift parameter for the fractional stochastic heat equation via power variation*, Mod. Stoch. Theory Appl. **6** (2019), no. 4, 397–417.

[MKT19b] ______, *On the distribution and q-variation of the solution to the heat equation with fractional Laplacian*, Probab. Math. Statist. **39** (2019), no. 2, 315–335.

[Moh94] J. Mohapl, *Maximum Likelihood Estimation in Linear Infinite Dimensional Models*, Stochastic Models **10** (1994), 781–794.

[Moh97] ______, *On Estimation in the Planar Ornstein-Uhlenbeck Process*, Stochastic Models **13** (1997), 435–455.

[Moh00] ______, *A Stochastic Advection-Diffusion Model for the Rocky Flats Soil Plutonium Data*, Ann. Inst. Statist. Math. **52** (2000), no. 1, 84–107.

[MP07] B. Maslowski and J. Pospíšil, *Parameter Estimates for Linear Partial Differential Equations with Fractional Boundary Noise*, Communications in Information and Systems **7** (2007), no. 1, 1–20.

[MP08] ______, *Ergodicity and Parameter Estimates for Infinite-Dimensional Fractional Ornstein-Uhlenbeck Process*, Appl Math Optim **57** (2008), 401–429.

[MT13] B. Maslowski and C. A. Tudor, *Drift parameter estimation for infinite-dimensional fractional Ornstein–Uhlenbeck process*, Bull. Sci. math. **137** (2013), 880–901.

[NAY62] A. Nagumo, S. Arimoto, and S Yoshizawa, *An Active Pulse Transmission Line Simulating Nerve Axon*, Proc. IRE **50** (1962), no. 10, 2061–2070.

[NVV99] Ilkka Norros, Esko Valkeila, and Jorma Virtamo, *An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions*, Bernoulli **5** (1999), no. 4, 571–587.

[Ouv78] Jean-Yves Ouvrard, *Martingale Projection and Linear Filtering in Hilbert Spaces. I: The Theory*, SIAM J. Control and Optimization **16** (1978), no. 6, 912–937.

[Pas80] Joseph E. Pasciak, *Spectral and pseudospectral methods for advection equations*, Math. Comp. **35** (1980), no. 152, 1081–1092.

[Paz83] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences, vol. 44, Springer-Verlag New York, 1983.

[Pes95] Szymon Peszat, *Existence and uniqueness of the solution for stochastic equations on Banach spaces*, Stochastics and Stochastics Reports **55** (1995), no. 3-4, 167–193.

[PFA+21] Gregor Pasemann, Sven Flemming, Sergio Alonso, Carsten Beta, and Wilhelm Stannat, *Diffusivity Estimation for Activator–Inhibitor Models: Theory and Application to Intracellular Dynamics of the Actin Cytoskeleton*, Journal of Nonlinear Science **31** (2021), no. 59, 1–34.

[PR96] L. Piterbarg and B. Rozovskii, *Maximum likelihood estimators in the equations of physical oceanography*, Stochastic Modelling in Physical Oceanography (R. J. Adler, P. Müller, and B. L. Rozovskii, eds.), Progress in Probability, vol. 39, Birkhäuser Boston, 1996, pp. 397–421.

[PR97] ———, *On asymptotic problems of parameter estimation in stochastic PDE's: discrete time sampling*, Math. Methods Statist. **6** (1997), no. 2, 200–223.

[PR00] B. L. S. Prakasa Rao, *Bayes estimation for some stochastic partial differential equations*, Journal of Statistical Planning and Inference **91** (2000), no. 2, 511–524.

[PR02] ———, *Nonparametric Inference for a Class of Stochastic Partial Differential Equations Based on Discrete Observations*, Sankhyā: The Indian Journal of Statistics, Series A **64** (2002), no. 1, 1–15.

[PS20] Gregor Pasemann and Wilhelm Stannat, *Drift estimation for stochastic reaction-diffusion systems*, Electronic Journal of Statistics **14** (2020), no. 1, 547 – 579.

[PT07] Jan Pospíšil and Roger Tribe, *Parameter estimates and exact variations for stochastic heat equations driven by space-time white noise*, Stoch. Anal. Appl. **25** (2007), no. 3, 593–611.

[QSS00] Alfio Quarteroni, Riccardo Sacco, and Fausto Saleri, *Numerical Mathematics*, Texts in Applied Mathematics, vol. 37, Springer-Verlag New York, 2000.

[RR20] S. Reich and P. J. Rozdeba, *Posterior contraction rates for non-parametric state and drift estimation*, Foundations of Data Science **2** (2020), no. 3, 333–349.

[Sch72] F. Schlögl, *Chemical Reaction Models for Non-Equilibrium Phase Transitions*, Z. Physik **253** (1972), no. 2, 147–161.

[Sha71] H. S. Shapiro, *Topics in Approximation Theory*, Lecture Notes in Mathematics, vol. 187, Springer-Verlag Berlin Heidelberg, 1971.

[Shi96] A. N. Shiryaev, *Probability*, second ed., Graduate Texts in Mathematics, vol. 95, Springer-Verlag, 1996.

[Shu01] M. A. Shubin, *Pseudodifferential Operators and Spectral Theory*, second ed., Springer-Verlag Berlin Heidelberg, 2001.

[Sog15] Christopher D. Sogge, *Problems related to the concentration of eigenfunctions*, Journées EDP (2015), no. IX, 11 pages.

[SS07] Hart F. Smith and Christopher D. Sogge, *On the L^p norm of spectral clusters for compact manifolds with boundary*, Acta Math. **198** (2007), no. 1, 107–153.

[SS15] Martin Sauer and Wilhelm Stannat, *Lattice approximation for stochastic reaction diffusion equations with one-sided Lipschitz condition*, Math. Comp. **84** (2015), no. 292, 743–766.

[SS16] ———, *Analysis and approximation of stochastic nerve axon equations*, Math. Comp. **85** (2016), no. 301, 2457–2481.

[SST20] Radomyra Shevchenko, Meryem Slaoui, and C. A. Tudor, *Generalized k-variations and Hurst parameter estimation for the fractional wave equation via Malliavin calculus*, Journal of Statistical Planning and Inference **207** (2020), 155–180.

[SV02] Jukka Saranen and Gennadi Vainikko, *Periodic Integral and Pseudodifferential Equations with Numerical Approximation*, Springer Monographs in Mathematics, Springer-Verlag Berlin Heidelberg, 2002.

[Tho06] Vidar Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, second ed., Springer Series in Computational Mathematics, vol. 25, Springer-Verlag Berlin Heidelberg, 2006.

[Tri10a] Hans Triebel, *Theory of Function Spaces*, Modern Birkhäuser Classics, Birkhäuser Verlag/Springer Basel AG, 2010, reprint of the 1983 edition.

[Tri10b] ______, *Theory of Function Spaces. II*, Modern Birkhäuser Classics, Birkhäuser Verlag/Springer Basel AG, 2010, reprint of the 1992 edition.

[TTV14] S. Torres, C. A. Tudor, and F. G. Viens, *Quadratic variations for the fractional-colored stochastic heat equation*, Electron. J. Probab. **19** (2014), no. 76, 1–51.

[Tud13] C. A. Tudor, *Analysis of Variations for Self-similar Processes (A Stochastic Calculus Approach)*, Probability and Its Applications, Springer International Publishing Switzerland, 2013.

[TV07] C. A. Tudor and F. G. Viens, *Statistical aspects of the fractional stochastic calculus*, Ann. Statist. **35** (2007), no. 3, 1183–1212.

[Unn89] T. E. Unny, *Stochastic partial differential equations in groundwater hydrology (Part I: Theory)*, Stochastic Hydrol. Hydraul. **3** (1989), 135–153.

[vdBHV15] Michiel van den Berg, Rainer Hempel, and Jürgen Voigt, L_1-*estimates for eigenfunctions of the Dirichlet Laplacian*, J. Spectr. Theory **5** (2015), no. 4, 829–857.

[vdV98] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, 1998.

[vNVW12] J. van Neerven, M. Veraar, and L. Weis, *Maximal L^p-Regularity for Stochastic Evolution Equations*, SIAM J. Math. Anal. **44** (2012), no. 3, 1372–1414.

[Vog15] Hendrik Vogt, *L_1-estimates for eigenfunctions and heat kernel estimates for semigroups dominated by the free heat semigroup*, J. Evol. Equ. **15** (2015), no. 4, 879–893.

[WD01] Y. Wei and J. Ding, *Representations for Moore-Penrose Inverses in Hilbert Spaces*, Applied Mathematics Letters **14** (2001), 599–604.

[Wey11] H. Weyl, *Über die asymptotische Verteilung der Eigenwerte*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1911), 110–117.

[Wit85] Rainer Wittmann, *A General Law of Iterated Logarithm*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **68** (1985), no. 4, 521–543.

[Wit87] ______, *Sufficient Moment and Truncated Moment Conditions for the Law of the Iterated Logarithm*, Probab. Th. Rel. Fields **75** (1987), no. 4, 509–530.

[Yag10] Atsushi Yagi, *Abstract Parabolic Evolution Equations and their Applications*, Springer Monographs in Mathematics, Springer-Verlag Berlin Heidelberg, 2010.

[Yin93] Z. Ying, *Maximum Likelihood Estimation of Parameters under a Spatial Sampling Scheme*, The Annals of Statistics **21** (1993), no. 3, 1567–1590.

[Zlá73] Miloš Zlámal, *Curved elements in the finite element method. I*, SIAM J. Numer. Anal. **10** (1973), no. 1, 229–240.