Statistical inference for stochastic partial differential equations from local measurements

PhD Dissertation

Anton Tiepner

Department of Mathematics

Aarhus University

July 30, 2024

Statistical inference for stochastic partial differential equations from local measurements

PhD Dissertation by
Anton Tiepner

Department of Mathematics, Aarhus University
Ny Munkegade 118, 8000 Aarhus C, Denmark

Supervised by
Associate Professor Claudia Strauch, Aarhus University

Submitted to Graduate School of Natural Sciences, Aarhus, July 30, 2024

AARHUS
UNIVERSITY
DEPARTMENT OF MATHEMATICS

Contents

Preface

With this dissertation, I conclude my PhD studies at the Department of Mathematics, Aarhus University, lasting from August 2021 to July 2024. It consists of the following four papers:

Paper A Optimal parameter estimation for linear SPDEs from multiple measurements.
To appear in the *Annales of Statistics*.
Preprint available at arXiv (arXiv:2211.02496v2).

Paper B Nonparametric velocity estimation in stochastic convection-diffusion equations from multiple local measurements.
Preprint available at arXiv (arXiv:2402.08353v1).

Paper C Multivariate change estimation for a stochastic heat equation from local measurements.
Working paper, a preprint will be available soon.

Paper D Parameter estimation in hyperbolic linear SPDEs from multiple measurements.
Preprint available at arXiv (arXiv:2407.13461v1).

Besides minor changes in layout, numbering, typesetting and correction of typing errors, those papers correspond to their revised or current versions, respectively. All of the articles were joint projects where both research and writing stage did not follow any strict work division and I have contributed extensively to those phases in all four papers.

That being said, I could not have written these articles nor this dissertation without all the wonderful people who supported me on that path during the last three years and beyond that.

Let me start with my main supervisor Claudia Strauch. I am very grateful for everything you helped me with. Not only did you find the time to discuss many of my numerous scientific problems and assisted me in the best imaginary way in all bureaucratic matters a PhD consists of, but you also cared for my general well-being. Clearly, the dissertation would not have been possible without you.

Next, I want to thank all my colleagues here in Aarhus for little chats during the day and for teaching me the hard way that ropes and hands don't go well together. Special thanks goes to my office mates Péter Juhász and Emil Dare for sharing the struggle of a PhD's life together. I also want to point out Niklas Dexheimer and Lukas Trottner for your valuable ability to cheer me up, for refreshing coffee breaks and for 'lending' me your thesis template.

Furthermore, I want to express my gratitude to Markus Reiß for the opportunity to join his working group during my change of research environment at Humboldt University of Berlin in autumn 2023. It was a warm welcome right from the start and a really productive time period. In this context, I would also like to mention Gregor Pasemann, Sascha Gaudlitz and Eric Ziebell with whom I had many fruitful discussions on SPDEs and related topics.

Many thanks goes to the people I met and the friends I made in Aarhus in the last three years through the local volleyball club, during countless boulder-sessions or in the language school, exploring the difficulty of Danish together. Special thanks goes to Lara for your sunny disposition and wisdom to life and to Daniela for being the lovely and inspiring human being you are. Without you I probably would never have got to know Mellemfolk - a non-profit plant-based café in Aarhus operated by volunteers - where I spend many hours of my free time in the last six months. I would like to name all of the people working there but then this preface goes on forever. So let me just say that you creating such a wonderful and welcoming community, that you all are doing amazing work and that I will miss that place very much.

But speaking of friends, I would also like to acknowledge all the constant support that I received from my friends from Germany. To reply to one of them—Lukas—hopefully you can call me officially 'Doktor Anton' very soon. Thank you all for visits in Denmark, for wonderful evenings in Berlin, for dramatic DnD-sessions, and also for long-distance activities such as scientific settlement placements, magnificent monster hunts, chivalrous crusades, legendary league clubs, sensational shootouts and abyssal adventures. Special thanks goes to Thorsten, Ally and Robert 'der Hühne' as even moving to Aarhus would have been difficult without your help.

Lastly, I want to thank my family. I thank my older sister Marzelline, her husband Robert and their two little kids Aurelia and Konstantin for their support in reaching my goals, my younger sister Frenilla for always bringing joy to my life and my parents Karoline and Sven for long phone calls, advice to all kinds of situations or, more general, for raising me to be the person I am.

Anton Tiepner
Aarhus, July 2024

Abstract

Stochastic partial differential equations are a multifaceted field where both theoretical and applied problems arise. While there is a rich literature on analytical and probabilistic matters, works on statistical aspects are limited, leaving many research questions unanswered. This dissertation aims to bridge some of these gaps by exploring the statistical potential of the novel local measurement approach.

Paper A is devoted to the joint parameter estimation for coefficients in a linear stochastic convection-diffusion equation. A modified log-likelihood approach leads to an asymptotically normal estimator and the derived central limit theorem generalises previous results. Robustness and applicability of the estimator are discussed. Moreover, minimax rate-optimality, i.e., a lower bound with the same rate of convergence, is established based on innovative insights on the reproducing kernel Hilbert space of the stochastic heat equation and its relation to the Hellinger distance between Gaussian measures.

Paper B examines nonparametric estimation of a spatially varying velocity. The constructed pointwise estimator is motivated through a local log-likelihood approach, and weight functions known from nonparametric regression are introduced. The estimator is decomposed into bias and variance components which are balanced through an additional bandwidth parameter. Under Hölder smoothness conditions, classical nonparametric convergence rates are achieved and their optimality is verified through an adaptation of the lower bounds approach in Paper A. Furthermore, the estimation procedure is extended to both integrated risk and unknown diffusivity level.

Paper C addresses multivariate change estimation for the stochastic heat equation where the discontinuous diffusivity has a jump occurring at some hypersurface. An estimator for the change area is constructed by a CUSUM approach. It consists of the union of optimally chosen pixels. The quality of the estimator is evaluated in terms of the symmetric difference pseudometric. Its analysis depends on the area's underlying complexity, i.e., its boundary roughness, and on the concentration of empirical processes. The results are discussed for the special cases of both graph representation and convexity of the change area.

Paper D focuses on hyperbolic stochastic partial differential equations, and the considered second-order Cauchy problem is given by an elastic system, whose intensity and energy development is characterised by unknown parameters. Combining methods and ideas from both parabolic and hyperbolic equations, a joint central limit theorem for the unknown coefficients is established. The derived convergence rate reflects the impact of the system's damping, i.e., energy loss, as underlying coefficients are more difficult to identify when the magnitude of the damping increases.

Resumé

Stokastiske partielle differentialligninger er et komplekst felt, hvor både teoretiske og anvendte problemer opstår. Selv om der findes en omfattende litteratur om analytiske og sandsynlighedsteoretiske emner, er der begrænset arbejde inden for statistiske aspekter, hvilket efterlader mange forskningsspørgsmål ubesvarede. Denne afhandling sigter mod at udfylde nogle af disse huller ved at udforske det statistiske potentiale ved nye lokale målemetoder.

Artikel A er dedikeret til simultan parameterestimation for koefficienter i en lineær stokastisk konvektions-diffusionsligning. En modificeret log-likelihood tilgang fører til en asymptotisk normal estimator, hvor den afledte centrale grænseværdisætning generaliserer tidligere resultater. Robustheden og anvendeligheden af estimatoren diskuteres. Desuden fastlægges minimaks rate-optimalitet, dvs. en nedre grænse med samme konvergensrate, baseret på innovative indsigter om reproducing kernel Hilbert space for den stokastiske varmeligning og dens relation til Hellinger-afstanden mellem Gaussiske mål.

Artikel B undersøger ikke-parametrisk estimering af en rumvarierende hastighed. Den konstruerede punktvise estimator er motiveret gennem en lokal log-likelihood tilgang, og vægtfunktioner kendt fra ikke-parametrisk regression introduceres. Estimatoren dekomponeres i bias- og varianskomponenter, som balanceres gennem en ekstra båndbreddeparameter. Under Hölder glathedsbetingelser opnås klassiske ikke-parametriske konvergensrater, og deres optimalitet verificeres gennem en tilpasning af den ovenfor nævnte nedre grænse tilgang. Endvidere udvides estimeringsproceduren til både integreret risiko og ukendt diffusivitetsniveau.

Artikel C adresserer multivariat ændringsestimation for den stokastiske varmeligning, hvor en diskontinuerlig diffusivitet har et spring, der forekommer ved en hyperflade. En estimator for ændringsområdet konstrueres ved en CUSUM-tilgang som foreningen af optimalt valgte pixels. Kvaliteten af estimatoren evalueres i forhold til den symmetriske mængdedifference. Analysen afhænger af den underliggende kompleksitet, dvs. områdets grænseruhed, og koncentrationen af empiriske processer. Resultaterne diskuteres for de specielle tilfælde af både grafrepræsentation og konveksitet af ændringsområdet.

Artikel D fokuserer på hyperbolske stokastiske partielle differentialligninger, og det resulterende andetordens Cauchy problem gives ved et elastisk system, hvis intensitet og energiforbrug karakteriseres ved ukendte parametre. Ved at kombinere metoder og ideer fra både paraboliske og hyperbolske ligninger udledes en simultan central grænseværdisætning for de ukendte koefficienter. Den afledte konvergensrate afspejler systemets dæmpning, dvs. energitab, idet de underliggende koefficienter er sværere at identificere, når dæmpningens størrelse øges.

Introduction

1

In my PhD studies, I worked in the project 'Exploring the potential of nonparametric modelling of complex systems via stochastic partial differential equations' financed by the Carlsberg Foundation Young Researcher Fellowship grant. Two key ingredients of this thesis project are given by stochastic partial differential equations (SPDEs) and by (nonparametric) statistics which unite my written papers in this time period.

In the following chapter, I give a brief introduction into the world of SPDEs in Section 1.1, and I also present some of the current state of the art on their statistical aspects in Section 1.2. There I will discuss, among other things, the observation scheme of *local measurements* which was employed in all four papers. The subsequent sections provide an overview of the frameworks, main results and methodologies in the different articles.

1.1 Introduction to SPDEs

When explaining SPDEs to non-experts on that field, I often motivate them either as deterministic partial differential equations (PDEs) with additional dynamical noise or as the limit $N \to \infty$ of N-dimensional stochastic differential equations (SDEs) with a specific type of drift structure. I will portray these motivations in detail below.

While trying to avoid unnecessary technicalities, let me start with a brief and abstract definition of SPDEs, see, e.g., [19, Chapter 6]. Consider the equation

$$\begin{cases} \dot{X}(t,x) = AX(t,x) + f(t,x) + B\dot{W}(t,x), & 0 < t \leq T, \quad x \in H, \\ X(0) = X_0, \end{cases} \tag{1.1}$$

on a Hilbert space H where $A\colon \mathcal{D}(A) \subset H \to H$, $B\colon \mathcal{D}(B) \subset H \to H$ are linear operators, X_0 is an H-valued $\mathcal{F}_0$-measurable random variable, f is a predictable processes and W is a cylindrical Brownian motion on H inducing the white noise $\dot{W}$ in time and space. The equation (1.1) belongs to the class of additive SPDEs, meaning that the volatility operator B is independent of the H-valued predictable solution process $X = (X(t))_{0\leq t\leq T}$. Heuristically, the first temporal derivative of the process X corresponds to some spatial derivatives of X, given through the (differential) operator A, a source or nonlinear term induced by the process f and a driving force, i.e., the (colored) noise $B\dot{W}$.

SPDEs find application in many real life phenomena. Examples include microscopic particle movement [42], surface temperature fluctuation [34], fluid dynamics [52], neuronal response [56], biomass concentration [20] or wave evolution [10] on, for instance, mechanical, electromagnetic or acoustic level. As a common factor, SPDEs model systems which have both spatial and temporal changes while also accounting for random effects.

As the cylindrical Brownian motion W is not differentiable in time, $\dot{W}$ is mathematically not well-defined and (1.1) has to be understood in the sense of a stochastic integral. From a technical point of view, one distinguishes between the three solution concepts of analytically *strong*, *weak* and *mild* solutions, each characterising the solution to (1.1) in a different sense. Interestingly, the terms of analytically strong and weak solution do not correspond to probabilistically strong/weak solution concepts in classical SDEs. Rather, the weak solution in an SPDE sense is closely related to weak solutions in the PDE theory, that is (1.1) only holds for functionals $\langle X(t), \varphi \rangle$, i.e., when X is tested against certain (smooth) test functions φ. Furthermore, mild solutions can be obtained from weak solutions by the variation of constants formula and vice versa given that some regularity assumptions are satisfied. A more detailed overview on underlying solution theory to SPDEs, cylindrical Brownian motions and definitions of the stochastic integral can be found in the books [19, 41, 45].

In what follows, the Hilbert space H is given by $L^2(\Lambda)$ for some open subset $\Lambda \subset \mathbb{R}^d$, A is some differential operator parameterised by an unknown function ϑ, which is the target of estimation, the source term f vanishes and B corresponds to the identity id on $L^2(\Lambda)$. These restrictions are made to unify the framework throughout this introduction. They are aligned with the settings of the four papers.

1.1.1 *Motivation via PDEs*

As outlined, I informally explain in this section, how deterministic partial differential equations are connected to their stochastic counterpart (1.1). Prototypical PDEs are discussed in Example 1.1 and a graphical illustration is present in Figure 1.1.

PDEs arise in a vast variety of phenomena in, for instance, physics, chemistry, biology or geoscience. They are processes in both time and space. An important class is given by the continuity (or transport) equations, where the time derivative equals the divergence (spatial derivative) of some vector field. For a fruitful and detailed discussion on PDEs, I can recommend the monographs [21, 53].

Linear parabolic PDEs are usually formulated as an abstract first order Cauchy problem

$$\dot{U}(t,x) = AU(t,x), \quad 0 < t \leq T, \quad x \in \Lambda, \tag{1.2}$$

which can be solved under the specification of initial and boundary conditions. The differential operator A is the infinitesimal generator of a strongly continuous semigroup $(S(t))_{t \geq 0}$ and the mild solution to (1.2) is given by

$$U(t,x) = S(t)U(0,x), \quad 0 \leq t \leq T, \quad x \in \Lambda.$$

Some simple examples of linear PDEs include the following.

Example 1.1.

(i) Heat equations

$$\dot{U}(t,x) = \vartheta \Delta U(t,x), \quad 0 < t \leq T, \quad x \in \Lambda,$$

model the diffusion of heat or particles. The diffusion speed is determined by the *diffusivity* $\vartheta > 0$. The spatial movement is entailed by the Laplace operator $\Delta = \sum_{i=1}^d \partial^2/\partial x_i^2$. Given Dirichlet boundary conditions and the initial condition $U(0) = U_0 \in L^2(\mathbb{R}^d)$, the fundamental solution on $\Lambda = \mathbb{R}^d$ is explicitly given as a convolution with the heat kernel $q_t(y) = (4\pi t)^{-d/2} \exp(-\|y\|^2/(4t))$, i.e.,

$$U(t,x) = (q_{\vartheta t} * U_0)(x), \quad 0 \leq t \leq T, \quad x \in \mathbb{R}^d.$$

(ii) Convection-reaction-diffusion equations

$$\dot{U}(t,x) = (\nabla \cdot \vartheta_1 \nabla + \vartheta_2 \cdot \nabla + \vartheta_3)U(t,x), \quad 0 < t \leq T, \quad x \in \Lambda, \tag{1.3}$$

additionally account for spatial transportation via the *velocity* $\vartheta_2 \colon \Lambda \longrightarrow \mathbb{R}^d$ and damping (or amplifying) described by the reaction coefficient $\vartheta_3 \colon \Lambda \longrightarrow \mathbb{R}$. Stochastic versions of (1.3) are discussed in Paper A, B and C.

(iii) Linear second-order Cauchy problems result in a coupled PDE

$$\begin{cases} \ddot{u}(t,x) = Au(t,x) + Bv(t,x), & 0 < t \le T, \quad x \in \Lambda, \\ \dot{u}(t,x) = v(t,x), & 0 < t \le T, \quad x \in \Lambda, \end{cases}$$

which can be transformed into a first-order problem by the transformation $U(t,x) = (u(t,x), v(t,x))^\top$, leading to

$$\dot{U}(t,x) = \mathcal{A}U(t,x), \quad 0 < t \le T, \quad x \in \Lambda,$$

for the matrix-valued differential operator

$$\mathcal{A} = \begin{pmatrix} 0 & I \\ A & B \end{pmatrix}.$$

The elasticity operator A models the deformation and the dissipative operator B describes the energy development of the coupled system. Prototypical examples are provided by

(a) the (undamped) wave equation: $A = \vartheta\Delta$, $B = 0$ with the *wave speed* $\vartheta > 0$.

(b) the structurally damped plate equation: $A = \vartheta_1\Delta^2$, $B = \vartheta_2\Delta$, where $\vartheta_1 < 0$, $\vartheta_2 > 0$.

Stochastic second-order Cauchy problems are investigated in Paper D.

When adding noise to PDEs to account for model misspecifications or observational errors, there is a fundamental difference whether the noise enters (1.2) dynamically (SPDE) or observationally (PDE with added observation noise). A graphical illustration in the case of the convection-reaction-diffusion equation in Example 1.1 (ii) can be found in Figure 1.1. It can be seen that both the PDE with added noise and the SPDE retain the striking bright heat flow of the deterministic PDE. On the other hand, the dynamical noise in the SPDE provides more complex structures, leading to visible 'butterfly effects' and highly flexible data dynamics, for which the noisy PDE does not account for.

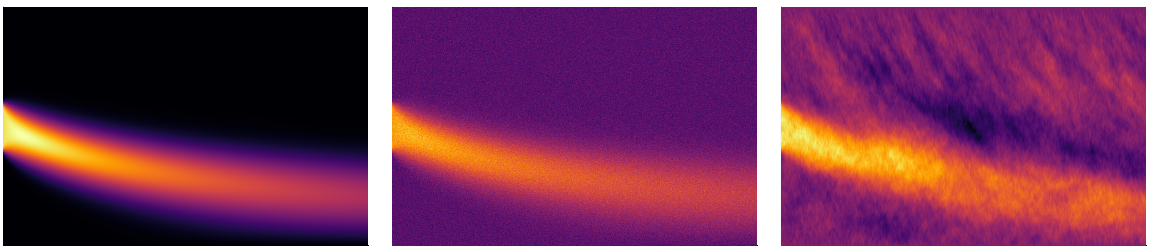


Figure 1.1: Comparison of a realisation of a convection-reaction-diffusion equation. Time is marked on the x-axis and space on the y-axis. Initial heat concentration in the one-dimensional domain $\Lambda = (0,1)$ diffuses and is transported in space over time; (left) no noise; (middle) observational noise; (right) dynamical noise.

1.1.2 *Motivation via SDEs*

Another starting point to motivate SPDEs is given by stochastic differential equations which already include a random forcing term. Conceptually, the solution to an SPDE is a function-valued process whereas the solution to an SDE is only an N-dimensional process in time. When N is large and the drift function follows a specific structure, the SDE solution can be interpreted as an approximate point evaluation of the function-valued SPDE solution at N distinct spatial grid points. I illustrate this motivation using the example of the stochastic heat equation. Furthermore, I will discuss statistical aspects of this toy model with regard to the results in Section 1.2.

Generally speaking, SDEs provide a flexible tool to describe the temporal evolution of processes and find application in, for instance, stock prices or particle movements. The solution process $(\widetilde{Y}(t))_{t\geq 0}$ to

$$\mathrm{d}\widetilde{Y}(t) = b(t, \widetilde{Y}(t))\,\mathrm{d}t + \sigma(t, \widetilde{Y}(t))\,\mathrm{d}\widetilde{W}(t), \quad 0 < t \leq T, \tag{1.4}$$

with an $\mathbb{R}^N$-valued Brownian motion $(\widetilde{W}(t))_{t\geq 0}$, a drift function $b\colon [0,T] \times \mathbb{R}^N \to \mathbb{R}^N$ and dispersion $\sigma\colon [0,T] \times \mathbb{R}^N \to \mathbb{R}^{N\times N}$ is also referred to as (Itô) diffusion, already resulting in a terminological overlap with the diffusivity ϑ in the heat equation from Example 1.1 (i). The reason for the different meaning of 'diffusion' lies in the considered perspective. From a microscopic or atomistic point of view, the movement of particles is characterised by a random walk. On the other hand, Fick's law describes diffusion as the movement of quantities from regions of higher concentration to lower concentrated ones, thus resulting in the macroscopic PDE perspective.

Under a certain drift structure, which describes specific component interactions, SDEs can be understood as approximations to SPDEs on a fine spatial grid. Consider the N-dimensional Ornstein–Uhlenbeck (OU) process starting in 0 and given by the following SDE

$$\mathrm{d}\widetilde{Y}(t) = \vartheta A \widetilde{Y}(t)\,\mathrm{d}t + \mathrm{d}\widetilde{W}(t), \quad 0 < t \leq T. \tag{1.5}$$

The drift matrix $A \in \mathbb{R}^{N\times N}$ is assumed to be of the form

$$A := (N+1)^2 \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix}, \tag{1.6}$$

i.e., A forms an approximation based on finite differences of the Laplace operator Δ with Dirichlet boundary conditions on $(0,1)$. When $N \to \infty$, the process $Y(t) := (0, \widetilde{Y}^\top, 0)^\top \in \mathbb{R}^{N+2}$ approximates the $L^2((0,1))$-valued process $X(t)$ solving the following heat equation on $(0,1)$

$$\mathrm{d}X(t) = \vartheta \Delta X(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \leq T, \tag{1.7}$$

with a cylindrical Brownian motion $(W(t))_{t\geq 0}$ on $L^2((0,1))$, cf. [44, Theorem 3.34 and Theorem 5.13].

Such spatial approximations are important, for instance, in the simulations of PDEs and SPDEs, cf. [44]. They are also useful to motivate the case of continuous space. The one-dimensional wave equation, for example, can be derived from Hooke's law. Then, in the overdamped regime, ignoring the mass of particles, reaction-diffusion equations arise as discussed in [24, Section 2.1]. More generally, SPDEs can be obtained from *lifts* of diffusion processes, cf. [19, Chapter 0].

Statistical inference for the toy example

Consider again the equation (1.4). Given a time-continuous record of observations $(\widetilde{Y}(t))_{0\le t\le T}$, the statistician 'knows' the exact value of the diffusion coefficient $\sigma(t,\widetilde{Y}(t))\sigma^\top(t,\widetilde{Y}(t))$ through differentiation of the observable quadratic variation process. This trivial identifiability criterion holds since the measures induced by different diffusion coefficients are singular. Note that inference on $\sigma(s,x)\sigma^\top(s,x)$, $s\le T$, is impossible if $x\in\mathbb{R}^N$ is not visited by $(\widetilde{Y}(t))_{t\le T}$. On the other hand, the (non)parametric estimation of the drift coefficient b is not possible in finite time $T<\infty$ and fixed dimension N, even in the simple case of linear drift. I refer to [37] for a general overview on statistical matters for diffusion processes. Now, if we were to increase the dimension N, this results in an identifiability criterion in finite time.

I will finish this section with a little teaser on statistical aspects and an enlightening observation regarding matching convergence rates for the maximum likelihood estimator (MLE) $\widehat{\vartheta}$ in the OU model (1.5) compared to the MLE obtained from local measurement observations of the stochastic heat equation.

It was shown in paper A that the unknown diffusivity $\vartheta>0$ in (1.7) is identifiable with optimal rate of convergence $N^{-3/2}$, given N local measurement observations, described in detail in Section 1.2, at points $x_1,\dots,x_N$ separated by an Euclidean distance of order N^{-1}. The next result shows that the same rate is obtained in the SDE setting (1.5), when the dimension N tends to infinity.

Theorem 1.2. *The maximum-likelihood estimator $\widehat{\vartheta}$ for ϑ in the SDE (1.5), given by*

$$\widehat{\vartheta}=\left(\int_0^T \widetilde{Y}(t)^\top A^2\widetilde{Y}(t)\,\mathrm{d}t\right)^{-1}\int_0^T \widetilde{Y}(t)^\top A\,\mathrm{d}\widetilde{Y}(t), \tag{1.8}$$

satisfies

$$N^{3/2}(\widehat{\vartheta}-\vartheta)\xrightarrow{d}\mathcal{N}\left(0,\frac{\vartheta}{T}\right).$$

Proof. The expression of the MLE (1.8) follows by maximising the Radon–Nikodym derivative, obtained through Girsanov's theorem, with respect to ϑ, cf. [37]. Plugging (1.5) into (1.8) yields the error decomposition

$$\widehat{\vartheta}=\vartheta+\left(\int_0^T \widetilde{Y}(t)^\top A^2\widetilde{Y}(t)\,\mathrm{d}t\right)^{-1}\int_0^T \widetilde{Y}(t)^\top A\,\mathrm{d}\widetilde{W}(t).$$

Since the observed Fisher-information

$$\mathcal{I}:=\int_0^T \widetilde{Y}(t)^\top A^2\widetilde{Y}(t)\,\mathrm{d}t$$

forms the quadratic variation of the (time-)martingale

$$\int_0^T \widetilde{Y}(t)^\top A\,\mathrm{d}\widetilde{W}(t),$$

it suffices to show the convergence of $N^{-3}\mathcal{I}\to T\vartheta^{-1}$ due to a general martingale CLT, e.g., Theorem A.25 in Paper A below. I prove

$$N^{-3}\mathbb{E}[\mathcal{I}]\to\frac{T}{\vartheta},\quad N^{-6}\operatorname{Var}(\mathcal{I})\to 0, \tag{1.9}$$

which implies $N^{-3}\mathcal{I}_\delta \to T\vartheta^{-1}$ in probability as $N \to \infty$ by Chebyshev's inequality. By the variation of constants formula, it holds

$$\widetilde{Y}(t) = \int_0^t e^{\vartheta(t-s)A}\,\mathrm{d}\widetilde{W}(s), \quad 0 \le t \le T.$$

Furthermore, the eigenvalues of the tridiagonal Toeplitz matrix A from (1.6) are given by

$$\lambda_k = (N+1)^2\Big(-2 + 2\cos(k\pi(N+1)^{-1})\Big) \in (-4(N+1)^2, 0), \quad k = 1, \dots, N.$$

Note that for a symmetric matrix $B \in \mathbb{R}^{N\times N}$ with eigenvalues $\mu_1, \dots, \mu_N$ the Hilbert–Schmidt-norm $\|B\|_{\mathrm{HS}}$ (or Frobenius-norm $\|B\|_{\mathrm{F}}$) satisfies

$$\|B\|_{\mathrm{HS}}^2 = \mathrm{tr}(B^2) = \sum_{k=1}^N \mu_i^2.$$

Thus, it follows

$$\begin{aligned}\mathbb{E}[\mathcal{I}] = \mathbb{E}\left[\int_0^T \widetilde{Y}(t)^\top A^2 \widetilde{Y}(t)\,\mathrm{d}t\right] &= \int_0^T\int_0^t \|e^{\vartheta(t-s)A}A\|_{\mathrm{HS}}^2\,\mathrm{d}s\,\mathrm{d}t \\ &= \sum_{k=1}^N \int_0^T\int_0^t e^{2\vartheta\lambda_k s}\lambda_k^2\,\mathrm{d}s\,\mathrm{d}t \\ &= \sum_{k=1}^N \int_0^T \lambda_k(2\vartheta)^{-1}\Big(e^{2\vartheta\lambda_k t} - 1\Big)\,\mathrm{d}t \\ &= \sum_{k=1}^N (2\vartheta)^{-2}\Big(e^{2\vartheta T\lambda_k} - 2\vartheta T\lambda_k - 1\Big).\end{aligned}$$

Observe further that

$$N^{-3}\sum_{k=1}^N \lambda_k = -2N^{-2}(N+1)^2 + 2\frac{(N+1)^2}{N^3}\sum_{k=1}^N \cos\left(\frac{k\pi}{N+1}\right) \to -2, \quad N \to \infty,$$

due to symmetry properties of the cosine function. Since $|e^{2\vartheta T\lambda_k} - 1| \le 1$ when $\lambda_k < 0$, this yields

$$N^{-3}\mathbb{E}[\mathcal{I}] \to \frac{T}{\vartheta}, \quad N \to \infty.$$

As for the variance, Wick's formula [31, Theorem 1.28] and the Cauchy–Schwarz inequality imply

$$\begin{aligned}\mathrm{Var}(\mathcal{I}) &= 4\int_0^T\int_0^t \mathrm{Cov}(\widetilde{Y}(t), \widetilde{Y}(s))^2\,\mathrm{d}s\,\mathrm{d}t \\ &= 4\int_0^T\int_0^t\left(\int_0^{t-s}\|e^{\vartheta(s/2+r)A}A\|_{\mathrm{HS}}^2\,\mathrm{d}r\right)^2\mathrm{d}s\,\mathrm{d}t \\ &= 4\int_0^T\int_0^t\left(\int_0^{t-s}\sum_{k=1}^N e^{\vartheta(s+2r)\lambda_k}\lambda_k^2\,\mathrm{d}r\right)^2\mathrm{d}s\,\mathrm{d}t\end{aligned}$$

$$= 4\int_0^T\int_0^t\left(\sum_{k=1}^N\frac{\lambda_k e^{s\vartheta\lambda_k}}{2\vartheta}(e^{2\vartheta(t-s)\lambda_k}-1)\right)^2 \mathrm{d}s\,\mathrm{d}t$$
$$\le 4\int_0^T\int_0^t N\sum_{k=1}^N\frac{\lambda_k^2 e^{2s\vartheta\lambda_k}}{4\vartheta^2}\,\mathrm{d}s\,\mathrm{d}t$$
$$\lesssim \int_0^T N\sum_{k=1}^N|\lambda_k|\,\mathrm{d}t$$
$$\lesssim N^4.$$

Hence, $N^{-6}\operatorname{Var}(\mathcal{J}) \longrightarrow 0$, which proves (1.9). ■

Theorem 1.2 demonstrates that time-continuous observations of the high-dimensional SDE model (1.5) with interacting nodes, i.e., points $x_1,\dots,x_N$, on a spatial grid of $[0,1]$, allow for the consistent estimation of ϑ where the MLE $\widehat{\vartheta}$ achieves the rate of convergence $N^{-3/2}$. In Section 1.2, I briefly present the spectral, discrete and local observation approaches, each leading to the same convergence rate. Interestingly, the asymptotics of the SDE setting describe a system of increasing interaction and intensity in the space-discontinuous process $\widetilde{Y}$ whereas in the mentioned SPDE observation schemes N functionals of the underlying space-continuous process X from (1.7) are observed which asymptotically results in an improved observational accuracy.

Remark 1.3. When the heat equation is approximated on a d-dimensional domain instead, the finite difference scheme gives an $N^d \times N^d$ dimensional drift matrix A where each entry is again of order N^2. Repeating the previous steps results in the optimal convergence rate $N^{-1/2-1/d}$.

1.2 Statistical inference for SPDEs

To be best of my knowledge, the first articles on statistics for SPDEs were published in the mid 80s ([36, 43]). Since then, many researchers have studied all kinds of problems in that field employing various estimation methods, utilising different techniques and assuming certain observable structures. While there is plenty of material available by now, many questions are still open. In the following section, I present an excerpt of some relevant literature related to my own research, thus I will limit myself on studies of (non)parametric drift estimation in SPDEs in finite time and non-vanishing noise. Beyond that, I recommend the survey paper [14] and the website [3] for a more comprehensive overview of the existing literature.

The estimation quality of an unknown quantity ϑ, e.g., the diffusivity in the stochastic heat equation (1.7), relies on the amount of observed data. But what is actually observable in an SPDE? As the solution $X(t,x)$ is a process in time and space, it seems ideal to observe it continuously in both components, i.e., $X(t,x)$ is known for any $0 \le t \le T$ and $x \in \Lambda$, which provides the largest amount of information possible while also keeping all fundamental properties of the solution. Unfortunately, a full record of observations is rarely possible and usually less data is available to the statistician. In the following, I will present three different approaches where discretisations are used.

1.2.1 *Spectral measurements*

In the spectral approach, firstly mentioned in [27, 28] and concretised in [29], an SPDE on $\Lambda \subset \mathbb{R}^d$ of the form

$$\mathrm{d}X(t,x) = (A_0 + \vartheta A_1)X(t,x)\,\mathrm{d}t + \mathrm{d}W(t,x), \quad 0 < t \leq T, \quad x \in \Lambda,$$

is considered where A_0 and A_1 are known diagonalisable differential operators having a common system of orthonormal eigenvectors $(e_k)_{k\geq 1} \subset L^2(\Lambda)$ with eigenvalues $(\mu_k)_{k\geq 1}$ and $(\lambda_k)_{k\geq 1}$, respectively. It is assumed that the projection of the solution on the first N eigenvectors are given, i.e., the first N Fourier-modes $(X_k(t))_{t\leq T} := (\langle X(t,\cdot), e_k(\cdot)\rangle)_{t\leq T}$, $1 \leq k \leq N$, are observed continuously in time. This leads to the study of N independent Ornstein–Uhlenbeck processes satisfying

$$\mathrm{d}X_k(t) = (\mu_k + \vartheta\lambda_k)X_k(t)\,\mathrm{d}t + \mathrm{d}\beta_k(t), \quad 0 < t \leq T, \quad 1 \leq k \leq N,$$

for independent Brownian motions $(\beta_k)_{k\geq 1}$. The MLE $\widehat{\vartheta}$ is constructed through Girsanov's theorem. The analysis of the estimator $\widehat{\vartheta}$ relies on the size of the eigenvalues and, consequently, consistency is only possible if a certain relation between dimension d, the differential order $\mathrm{ord}(A_0)$ and $\mathrm{ord}(A_1)$ is satisfied. As the eigenvalues of higher-order differential operators are larger, the corresponding higher-order coefficients become more visible, and they are identified at faster rate.

Extensions of the spectral approach include, e.g., joint parameter estimation for two unknown parameters [46], hyperbolic equations [39, 40], fractional noise [13], testing [18], lower-order nonliniarities [50], temporal discretisation [15] or to the Bayesian framework [11]. A detailed overview on the methodology is provided by [47].

Due to the elegant decoupling of the spectral observation scheme, the underlying estimation problem is separated into the study of independent OU processes which simplifies the estimation analysis. However, there are some drawbacks to the approach. There is, for instance, no known estimator yet for a velocity from Example 1.1 (ii) in the stochastic convection-diffusion equation. Furthermore, nonparametric estimation of space-dependent coefficients has not been studied so far as the eigenvalues and eigenfunctions will no longer be independent of ϑ. As a final note, the Fourier modes are usually not directly observable and have to be approximated through discrete spatial observations instead, cf. [16, p. 2] and [17, p. 2], but a complete analysis of this idea is still an open problem.

1.2.2 *Discrete observations*

The access to discrete measurements, on the other hand, is quite natural to assume. Suppose X solves the stochastic convection-diffusion equation

$$\mathrm{d}X(t,x) = (\vartheta_1\Delta + \vartheta_2\nabla + \vartheta_3)X(t,x)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t,x), \quad 0 < t \leq T, \quad x \in (0,1),$$

with unknown constants $\vartheta_1, \vartheta_2, \vartheta_3, \sigma$. Given that $X(t_i, x_k)$, $1 \leq i \leq M$, $1 \leq k \leq N$, is observed on a time-space grid, the following can be found.

The estimation of the volatility σ relies on the quadratic variation and can be realised via method of moments or power variation approaches [8, 9, 12]. Furthermore, it is impossible to estimate both ϑ_1 and σ if only temporal or spatial increments are available [26, Proposition 2.3]. The reaction coefficient ϑ_3, however, cannot be identified in one spatial dimension, even if the covariance structure is exploited both in time and space. In that case other asymptotic regimes such as large time [25], small diffusivity [23], spatial ergodicity [22] or small noise [32] have to be considered. Joint estimation is treated, amongst others, in [7, 26, 33] leading to similar rates obtained in the spectral approach if both M and N tend to infinity sufficiently fast. Discrete measurements are also utilised in the study of a nonparametric reaction-terms [25], in the two-dimensional framework [54] with colored noise or hyperbolic equations [48].

Despite the accessibility of observations, the methodology is not yet viable in arbitrary space dimensions since point evaluations are no longer well-defined in dimension $d \geq 2$ as the rough space-time white noise $\dot{W}(t)$ only induces a distribution-valued solution $X(t)$. Moreover, nonparametric estimation of the coefficients is again complicated due to their influence on the eigenvalues in the spectral representation.

1.2.3 *Local measurements*

Observing the exact value of a process $(X(t,x))_{0\leq t\leq T}$ at a discrete point $x \in \Lambda$ in space is, in general, not possible due to physical limitations. Besides from measurement errors induced by the used measurement device, often only a blurred image is available in optical systems, which is modeled through a convolution of X with a so-called *point-spread* function $K_{\delta,x}$ [5, 6]. Examples of such locally blurred averages can, for instance, occur when an infrared thermometer with laser diameter δ is used to evaluate the temperature at a small area around x or, alternatively, in microscopic scaling limits with resolution level δ.

Mathematically formulated, a continuous time local in space measurement is defined through

$$X_{\delta,x}(t) := \langle X(t), K_{\delta,x}\rangle, \quad 0 \leq t \leq T, \quad x \in \Lambda, \quad \delta > 0,$$

where

$$K_{\delta,x}(y) := \delta^{-d/2} K(\delta^{-1}(y-x)), \quad y \in \Lambda, \quad \delta > 0,$$

for some compactly supported function $K \in L^2(\mathbb{R}^d)$. The scaling $\delta^{-d/2}$ is just taken as a convenient normalising factor as $\|K_{\delta,x}\|_{L^2(\Lambda)} = \|K\|_{L^2(\mathbb{R}^d)}$ and does not affect the estimation procedure.

Local measurements were introduced in [4], where it was shown that a spatially varying diffusivity ϑ in the perturbed heat equation (convection-reaction-diffusion equation)

$$\mathrm{d}X(t) = (\nabla \cdot \vartheta \nabla + b \cdot \nabla + c)X(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \leq T, \tag{1.10}$$

can already be identified at $x_0 \in \Lambda$ with rate δ upon observing $(X_{\delta,x_0}(t))_{0\leq t\leq T}$.

Subsequently, extensions of (1.10) have been studied and the local measurement approach was successfully applied to semilinear (coupled) equations [1, 2], and practical relevance to cell repolarisation was investigated in [1]. Furthermore, [30] explored the heat equation driven by multiplicative noise, [51] analysed a one-dimensional change-point problem arising from a discontinuous diffusivity and [57] considered the nonparametric wave equation.

The full anisotropic case of (1.10) was studied in Paper A, where we were also interested in estimation of the lower-order transport and reaction coefficients b and c. It turned out that consistency for those parameters necessarily requires an increasing amount $N = N(\delta) \to \infty$ of local measurements centred at $x_1, \ldots, x_N \in \Lambda$. Given that the corresponding point-spread functions K_{δ,x_k} have non-overlapping support, this allows for at most $N \asymp \delta^{-d}$ observation points.

In other words, only data of distinct pixel provides adequate new information. Under such assumptions, we established a minimax-optimal convergence rate. The verification of the lower bound relies on a novel approach relating the Hellinger distance of Gaussian measures with the reproducing kernel Hilber space (RKHS) of the local measurements using the Feldman–Hájek theorem. This rate matches convergence results obtained in, for instance, the spectral approach [29], the discrete setting [26] or the described high-dimensional SDE setup in Section 1.1.

To the best of my knowledge, Paper B is the only paper in the literature devoted to the nonparametric velocity estimation. Based on the results of Paper A, we derived a weighted estimator from a local log-likelihood approach which is designed to automatically reduce the bias appearing in the nonparametric framework. This led to the introduction of the bandwidth $h = h(\delta) \to 0$. Naturally, we had to face the classical bias-variance trade off and we obtained minimax-optimality through the optimal bandwidth choice.

In Paper C, we extended the one-dimensional change point problem studied in [51] to its multivariate equivalent. Primarily interested in the estimation of the boundary of some set Λ_+, i.e., the change interface of a discontinuous diffusivity, we constructed an estimator $\widehat{\Lambda}_+$ as the union of optimal pixels (cubes). The estimation procedure was motivated by a CUSUM approach, common in change point problems, which we combined with ideas from image reconstruction, cf. [35].

Lastly, Paper D was devoted to the study of second-order stochastic Cauchy problems, extending the work [57] on the nonparametric stochastic wave equation. Our obtained convergence rate coincides with the spectral approach [39], but the proofs required different arguments in contrast to previous contributions. Due to additional smoothing properties of an associated semigroup, we arrived eventually in between semigroup-structures of heat equations and undamped wave equations, and neither of the established approaches in these setting was applicable. We dealt with the emerging statistical problems through functional calculus for operators, discovering new analytical results along the way.

1.3 Paper A

Paper A was written in collaboration with Randolf Altmeyer and Martin Wahl. It has started as an extension of my master's thesis on parametric velocity estimation from local measurements. The derived central limit theorem (CLT) and accompanying material has seen major changes throughout the beginning and its final state which itself forms an exiting result. The significant novelty, however, lied in the lower bound approach also bypassing a gap in the original paper [4]. In order to stay consistent throughout the introduction, I replace the number M of observed measurements in Paper A by N in the following section.

1.3.1 *Framework*

We studied the stochastic convection-diffusion equation

$$\mathrm{d}X(t) = A_\vartheta X(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \leq T, \tag{1.11}$$

on some open, bounded domain $\Lambda \subset \mathbb{R}^d$, starting in $X(0) = X_0 \in L^2(\Lambda)$, driven by space-time white noise and with Dirichlet boundary conditions. A prototypical example of the second-order linear elliptic operator A_ϑ is given by

$$A_\vartheta = \vartheta_1 \Delta + \vartheta_2 \cdot \nabla + \vartheta_3$$

for $\vartheta_1 > 0$, $\vartheta_2 \in \mathbb{R}^d$, $\vartheta_3 \in \mathbb{R}$. The goal was to recover the unknown parameter $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3)^\top \in \mathbb{R}^{d+2}$.

We assumed access to the continuously observed local measurement process $X^A_{\delta,k}(t) = \left(X_{\delta,k}(t), -X^\nabla_{\delta,k}(t), X^\Delta_{\delta,k}(t)\right)^\top$, where for $0 \le t \le T$, $1 \le k \le N$,

$$\begin{aligned} X_{\delta,k}(t) &= \langle X(t), K_{\delta,x_k} \rangle, \\ X^\nabla_{\delta,k}(t) &= \langle X(t), \nabla K_{\delta,x_k} \rangle, \\ X^\Delta_{\delta,k}(t) &= \langle X(t), \Delta K_{\delta,x_k} \rangle. \end{aligned}$$

Thus, each local measurement $X_{\delta,k}$ is an Itô process fulfilling

$$\mathrm{d}X_{\delta,k}(t) = \vartheta X^A_{\delta,k}(t)\,\mathrm{d}t + \|K\|_{L^2(\mathbb{R}^d)}\,\mathrm{d}W_k(t), \quad 0 < t \le T, \quad 1 \le k \le N, \tag{1.12}$$

with initial values $X_{\delta,k}(0) = \langle X_0, K_{\delta,x_k} \rangle$ and driven by the scalar Brownian motions $W_k(t) = \langle W(t), K_{\delta,x_k} \rangle / \|K\|_{L^2(\mathbb{R}^d)}$. As neither $X_{\delta,k}$ nor $X^A_{\delta,k}$ are Markov processes due to the infinite speed of propagation in space induced by the semigroup $(S_\vartheta(t))_{t\ge 0}$ generated by A_ϑ, the processes $X_{\delta,k}$ are not independent, even when the driving Brownian motions W_k are. Hence, classical statistical methods for diffusion processes, see, e.g., [37], fail. However, using a general Girsanov theorem for multivariate Itô processes and ignoring conditional expectations, we obtained the *augmented MLE*

$$\widehat{\vartheta}_\delta = \mathcal{I}_\delta^{-1} \sum_{k=1}^N \int_0^T X^A_{\delta,k}(t)\,\mathrm{d}X_{\delta,k}(t), \tag{1.13}$$

with the *observed Fisher information matrix*

$$\mathcal{I}_\delta = \sum_{k=1}^N \int_0^T X^A_{\delta,k}(t) X^A_{\delta,k}(t)^\top\,\mathrm{d}t. \tag{1.14}$$

1.3.2 *Main results*

In Paper A, we derived three major results. First of all, we were interested in the asymptotic properties of the estimator (1.13). If the observation points $x_1, \ldots, x_N$ belong to a compact subset $\mathcal{J} \subset \Lambda$ and the corresponding localising functions K_{δ,x_k} have disjoint supports, then, under suitable smoothness assumptions on the initial condition X_0 and K, for some deterministic, invertible matrix Σ_ϑ the following CLT unfolds.

$$\begin{pmatrix} N^{1/2}\delta^{-1}(\widehat{\vartheta}_1 - \vartheta_1) \\ N^{1/2}(\widehat{\vartheta}_2 - \vartheta_2) \\ N^{1/2}\delta(\widehat{\vartheta}_3 - \vartheta_3) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \|K\|^2_{L^2(\mathbb{R}^d)} \Sigma_\vartheta^{-1}), \quad , \delta \to 0. \tag{1.15}$$

As an immediate consequence, the consistent estimation of ϑ_2 and ϑ_3 necessarily requires $N^{1/2} \to \infty$ and $N^{1/2}\delta \to \infty$, respectively. Combined with the fact that the disjoint support assumption on K_{δ,x_k} restricts the number of measurements to $N \asymp \delta^{-d}$, the seemingly best convergence rates for ϑ_i are given by $\delta^{d/2+2-i}$, $i \in \{1,2,3\}$. They mirror the rates from both spectral and discrete setting. In case of the reaction coefficient ϑ_3, the rate $\delta^{d/2-1}$ explodes in $d = 1$ and a boundary case occurs in $d = 2$, where we instead proved logarithmic rates, when less restricting smoothing assumptions on the kernel K are imposed.

Since we were also interested in the verification of minimax-optimal in the sense of [55], lower bounds had to be proven. A key argument for those lower bound considerations is a precise characterisation of the RKHS of the Gaussian measure induced by the solution to the stochastic heat equation and the related RKHS-norm. Such findings were derived by a generalisation of the finite-dimensional Ornstein–Uhlenbeck case and are of independent interest.

It turned out that the rates in (1.15) are indeed optimal. Not only is it impossible to consistently estimate the velocity ϑ_2 from a single local measurement in finite time, but even a full observation scheme does not allow consistent estimation of ϑ_3 in $d \leq 2$. This, however, is no contradiction to the mentioned logarithmic rate in $d = 2$ due to different assumptions on the kernel K.

1.3.3 *Methodology*

A key observation in the analysis of (1.13) lied in the study of the observed Fisher information (1.14). Indeed, plugging (1.12) into (1.13) results in the error decomposition

$$\widehat{\vartheta}_\delta = \vartheta + \|K\|_{L^2(\mathbb{R}^d)} \mathcal{I}_\delta^{-1} \mathcal{M}_\delta$$

with the martingale term

$$\mathcal{M}_\delta = \sum_{k=1}^{N} \int_0^T X_{\delta,k}^A(t)\,\mathrm{d}W_k(t).$$

When the Brownian motions W_k, $k = 1, \ldots, N$, are mutually independent (which holds under the imposed disjoint support condition) the observed Fisher information $\mathcal{I}_\delta$ forms the quadratic covariation matrix of $\mathcal{M}_\delta$. By a martingale CLT and Slutsky's lemma, the CLT (1.15) holds if we find a deterministic matrix Σ_ϑ as limiting object of the rescaled Fisher information. In other words, given a diagonal matrix ρ_δ of rescaling coefficients, we had to show that

$$\mathbb{E}[\rho_\delta \mathcal{I}_\delta \rho_\delta] \to \Sigma_\vartheta, \quad \operatorname{Var}(\rho_\delta \mathcal{I}_\delta \rho_\delta) \to 0, \quad \delta \to 0.$$

These convergences relied upon pointwise convergence of the semigroup to the heat kernel via a Feynman–Kac approach, rescaling properties of semigroups and operators, (integrable) semigroup bounds to utilise dominated convergence and Wick's formula for the variance part.

As the information geometry of local measurements is complex and a Markovian structure of standard diffusion processes is not maintained, standard MLE optimality results for continuously observed processes were no longer available. Hence, the lower bounds are based on an innovative strategy, relating the Hellinger distance of Gaussian measures to properties of their RKHS. The RKHS computations were mainly achieved under basic operations such as linear transformations. The combination of the Feldman–Hájek theorem [19, Theorem 2.25] with basic properties of the Hellinger distance resulted in the desired lower bound, once underlying RKHS-norms, written in terms of covariance operators between local measurements, were sufficiently well bounded.

1.4 Paper B

Paper B was a joint work with my supervisor Claudia Strauch. The main goal was the construction and analysis of an estimator for a spatially varying velocity field ϑ. This estimation problem, to the best of my knowledge, has not been investigated before.

1.4.1 *Framework*

We considered the stochastic convection-diffusion equation (1.11) where the differential operator A_ϑ is defined through

$$A_\vartheta = a\Delta + \vartheta \cdot \nabla + c$$

for the (possibly unknown) constant diffusivity $a > 0$, the unknown velocity $\vartheta : \Lambda \to \mathbb{R}^d$ and the nuisance reaction function $c : \Lambda \to \mathbb{R}$. Due to the nonparametric nature of the problem, we had to adjust the estimation procedure to account for bias-reduction. Motivated through the augmented MLE in Paper A and a local constant log-likelihood approach, we constructed the *weighted augmented MLE* $\widehat{\vartheta}(x)$

$$\widehat{\vartheta}_\delta(x) = -(\mathcal{I}_\delta^x)^{-1} \sum_{k=1}^N w_k(x) \left(\int_0^T X_{\delta,k}^\nabla(t)\, \mathrm{d}X_{\delta,k}(t) - \int_0^T a X_{\delta,k}^\Delta(t) X_{\delta,k}^\nabla(t)\, \mathrm{d}t \right)$$

with the weighted observed Fischer information matrix

$$\mathcal{I}_\delta^x = \sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\nabla(t) X_{\delta,k}^\nabla(t)^\top \, \mathrm{d}t$$

and weight functions $w_k : \Lambda \to \mathbb{R}$ as an estimator for ϑ evaluated at $x \in \Lambda$. The estimator is decomposable into

$$\widehat{\vartheta}_\delta(x) = \vartheta(x) + (\mathcal{I}_\delta^x)^{-1}\mathcal{R}_\delta^x - \|K\|_{L^2(\mathbb{R}^d)} (\mathcal{I}_\delta^x)^{-1}\mathcal{M}_\delta^x, \tag{1.16}$$

with the martingale part and remainder defined through

$$\mathcal{M}_\delta^x = \sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\nabla(t)\, \mathrm{d}W_k(t),$$
$$\mathcal{R}_\delta^x = \sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\nabla(t) \langle X(t), ((\vartheta - \vartheta(x)) \cdot \nabla + \mathrm{div}(\vartheta) - c) K_{\delta,x_k} \rangle \, \mathrm{d}t.$$

An appropriate choice of weight functions was vital and we found inspiration in the classical local polynomial regression, cf. [55, Chaper 1.6]. Introducing an extra tuning parameter, the bandwidth h, which enters the weight functions to control the contribution of each local measurement, the analysis of the decomposition (1.16) was studied under Hölder-smoothness assumptions on ϑ.

1.4.2 *Main results*

Given that ϑ belongs to the class of β-Hölder continuous functions for $\beta \in (1, 2]$, i.e., ϑ is continuously differential with continuous partial derivatives having Hölder-exponent $\beta - 1$, the pointwise estimator satisfies

$$\widehat{\vartheta}_\delta(x) - \vartheta(x) = O_{\mathbb{P}}(h^\beta + (Nh^d)^{-1/2}). \tag{1.17}$$

Clearly, (1.17) is optimised for the choice $h \asymp N^{-1/(2\beta+d)}$, leading to the standard nonparametric rate $N^{-\beta/(2\beta+d)}$ for the estimator or, in terms of a full observation scheme $N \asymp \delta^{-d}$, the equivalent rate $\delta^{\beta d/(2\beta+d)}$. Adapting the lower bound method of Paper A, we were able to verify that $N^{-\beta/(2\beta+d)}$ is indeed the minimax-optimal rate of convergence. Moreover, we also studied estimation when the diffusivity a is unknown and investigated the integrated risk.

1.4.3 *Methodology*

Due to the incorporation of the weights w_k in the estimation procedure, the weighted Fisher information $\mathcal{I}_\delta^x$ no longer forms the quadratic covariation of $\mathcal{M}_\delta^x$. In fact, due to the convergence $\mathcal{I}_\delta^x \to \Sigma$ as $\delta \to 0$, which does not involve any normalisation, the estimation error is solely induced by the bias term $\mathcal{R}_\delta^x$ and the variance part $\mathcal{M}_\delta^x$.

Since the square root of the quadratic covariation bounds the order of the martingale $\mathcal{M}_\delta^x$ itself, it was enough to verify for the quadratic covariation process $[\mathcal{M}_\delta^x]$ that

$$[\mathcal{M}_\delta^x]_T = \sum_{k=1}^{N} w_k(x)^2 \int_0^T X_{\delta,k}^{\nabla}(t) X_{\delta,k}^{\nabla}(t)^\top \,\mathrm{d}t = O_{\mathbb{P}}((Nh^d)^{-1}), \quad \delta \to 0,$$

to derive the order of the variance term. The proof combines the imposed structure on the weights, namely $\sum_{k=1}^N w_k(x)^2 \lesssim (Nh^d)^{-1}$, and calculations close to those made in Paper A for the summands in the observed Fisher-information.

On the other hand, verification of $\mathcal{R}_\delta^x = O_{\mathbb{P}}(h^\beta)$ was more involved and it was particularly challenging to prove $\mathbb{E}[\mathcal{R}_\delta^x] = O(h^\beta)$. The covariance structure of local measurements depends on the semigroup $(S_\vartheta^*(t))_{t\geq 0}$ generated by A_ϑ^*. Precisely controlling the semigroup approximations arising from both the variation of constants formula and a Feynman–Kac argument, we considered the heat kernel instead of S_ϑ^* in the covariance structure. A major difficulty lied in the difference $\vartheta(\cdot) - \vartheta(x)$ appearing in the remainder $\mathcal{R}_\delta^x$. This difference was rewritten by a first-order Taylor expansion with Peano-remainder. Choosing reproducing weights of order one imply under (anti-)symmetry kernel assumptions that the first-order term of the Taylor series vanishes. To achieve that we exploited that a convolution with the heat kernel keeps the orientation of an even (respectively odd) function. Thus, the size of $\mathbb{E}[\mathcal{R}_\delta^x]$ is entirely determined by the Peano-reminder of order $O(h^\beta)$.

1.5 Paper C

Paper C was a collaboration with Lukas Trottner. We studied the multivariate version of the change point problem in [51]. As the change interface no longer consists of a single point but instead is characterised by a hypersurface in space, a different estimation construction had to be developed.

1.5.1 *Framework*

We considered the heat equation

$$\mathrm{d}X(t) = \Delta_\vartheta X(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \le T,$$

with weighted Laplace operator $\Delta_\vartheta = \nabla \cdot \vartheta \nabla$ and discontinuous diffusivity

$$\vartheta(x) = \vartheta_- \mathbf{1}_{\Lambda_-}(x) + \vartheta_+ \mathbf{1}_{\Lambda_+}(x), \quad x \in (0,1)^d,$$

where the jump occurs at the hypersurface

$$\Gamma := \partial\Lambda_- \cap \partial\Lambda_+ \subset [0,1]^d$$

which separates $\overline{\Lambda} = [0,1]^d$ into a partition of sets $\Lambda_\pm$. Local measurement observations $X_{\delta,\alpha}$, $X^\Delta_{\delta,\alpha}$ were assumed on a regular spatial grid with observation points $x_\alpha = \delta(\alpha - \frac{1}{2}\mathbb{1})$, $\alpha \in [n]^d$ for $n = \delta^{-1} \in \mathbb{N}$. Estimating Γ is intrinsically related to the estimation of Λ_+, and the proposed estimator $\widehat{\Lambda}_+$ consists of a union of closed hypercubes $\mathrm{Sq}(\alpha)$ decomposing $[0,1]^d$ into n^d cubes of edge length δ. It is constructed by a CUSUM approach in the following way.

We introduced the modified log-likelihood

$$\ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+) = \vartheta_{\delta,\alpha}(\Lambda_+) \int_0^T X^\Delta_{\delta,\alpha}(t)\,\mathrm{d}X_{\delta,\alpha}(t) - \frac{\vartheta_{\delta,\alpha}(\Lambda_+)^2}{2} \int_0^T X^\Delta_{\delta,\alpha}(t)^2\,\mathrm{d}t,$$

for the decision rule

$$\vartheta_{\delta,\alpha}(\Lambda_+) := \begin{cases} \vartheta_+, & \mathrm{Sq}(\alpha) \subset \Lambda_+, \\ \vartheta_-, & \text{else}, \end{cases} = \begin{cases} \vartheta_+, & x_\alpha \in \Lambda_+, \\ \vartheta_-, & \text{else}, \end{cases}$$

determining where to assign the value $\vartheta_\pm$. Then, the M-estimator is given by

$$(\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+) \in \operatorname*{arg\,max}_{(\vartheta_-,\vartheta_+,\Lambda_+) \in \Theta_- \times \Theta_+ \times \mathcal{A}_+} \sum_{\alpha \in [n]^d} \ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+). \tag{1.18}$$

1.5.2 *Main results*

To quantify how well the estimator $\widehat{\Lambda}_+$ from (1.18) approximates the truth Λ^0_+, we considered the expectation of the symmetric difference's Lebesgue measure $\boldsymbol{\lambda}(\widehat{\Lambda}_+ \triangle \Lambda^0_+)$. The convergence rate depends on the underlying complexity of the boundary $\partial\Lambda^0_+$, measured by the set

$$\mathcal{B} = \{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap \partial\Lambda^0_+ \neq \emptyset\},$$

which quantifies cubes $\mathrm{Sq}(\alpha)$ whose interior intersects the boundary $\partial\Lambda^0_+$. Given that for some $\beta \in (0,1]$ and some constant $c > 0$ the size of $\mathcal{B}$ is bounded by

$$|\mathcal{B}| \le c\delta^{-d+\beta}, \tag{1.19}$$

we obtained for some absolute constant C independent of δ our main result

$$\mathbb{E}\big[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \triangle \Lambda^0_+)\big] \le C\delta^\beta.$$

We subsequently showed consistency for the nuisance parameters $\vartheta^0_\pm$, establishing $|\widehat{\vartheta}_\pm - \vartheta^0_\pm| = \mathcal{O}_{\mathbb{P}}(\delta^{\beta/2})$. The main results was also discussed in two special cases where the hypersurface has either a graph representation, given by a β-Hölder continuous function τ^0, or is assumed to be the boundary of a convex set Λ^0_+. In both cases, the number of tiles covering the boundary is bounded in the sense of (1.19).

1.5.3 *Methodology*

The estimation method was based on a CUSUM approach commonly used in change point problems. The form of the M-estimator $\widehat{\Lambda}_+$, on the other hand, was motivated by related edge estimation techniques, cf. [35].

To analyse the estimator, we first introduced a minimal deterministic tiling $\Lambda_+^\updownarrow$, consisting of unions of hypercubes $\mathrm{Sq}(\alpha)$, which covered the truth Λ_+^0 as tight as possible. This enabled the study of the symmetric difference pseudometric between $\widehat{\Lambda}_+$ and $\Lambda_+^\updownarrow$ instead which we could relate to the size of $\mathcal{B}$. In a last step, we controlled the concentration of some underlying empirical processes, resulting from an estimator decomposition, similarly to [51].

When proving the consistency of the nuisance parameters $\vartheta_\pm^0$, it is crucial that the underlying partition is visible, i.e., $\lambda(\Lambda_\pm^0) > 0$, to identify a jump in the diffusivity correctly.

The smoothness of the boundary $\partial\Lambda_+^0$ determines (1.19). In case of the graph representation, the boundary fragment was assumed to be β-Hölder continuous, $\beta \in (0,1]$, providing an upper bound directly. When Λ_+^0 is a convex set, the boundary can be described locally by a Lipschitz function. In particular, a vertical ray through the interior of Λ_+^0 intersects the boundary at exactly two points and those intersections form an upper concave and a lower convex function. An upper bound of $|\mathcal{B}|$ is thus intuitively given with $\beta = 1$ in (1.19), cf. also [38].

1.6 Paper D

The work on Paper D begun during my research stay at the Humboldt University of Berlin in autumn 2023 and evolved as a fruitful collaboration with Eric Ziebell. We were interested in parameter estimation for general hyperbolic equations where both characteristics from parabolic problems (Paper A) and phenomena of wave equations [57] arise.

1.6.1 *Framework*

In contrast to the other projects Paper A, Paper B and Paper C, we considered now a hyperbolic equation, i.e., a second-order stochastic Cauchy problem of the form

$$\begin{cases} \mathrm{d}v(t) = \big(A_\vartheta u(t) + B_\eta v(t)\big)\,\mathrm{d}t + \mathrm{d}W(t), & 0 < t \le T, \\ \mathrm{d}u(t) = v(t)\,\mathrm{d}t, \end{cases} \tag{1.20}$$

with amplitude u and velocity v. The elastic and dissipative differential operators A_ϑ and B_η are linear combinations of fractional Laplace operators, i.e.,

$$A_\vartheta = \sum_{i=1}^{p} \vartheta_i(-\Delta)^{\alpha_i}, \quad B_\eta = \sum_{j=1}^{q} \eta_j(-\Delta)^{\beta_j},$$

for strictly increasing nonnegative sequences $(\alpha_i)_{i\le p}$ and $(\beta_j)_{j\le q}$. We assumed local measurement observations

$$u_{\delta,k}(t) = \langle u(t), K_{\delta,x_k}\rangle, \quad u_{\delta,k}^{\Delta_i}(t) = \langle u(t), (-\Delta)^{\alpha_i}K_{\delta,x_k}\rangle, \quad 0 \le t \le T, \quad 1 \le i \le p, \quad 1 \le k \le N$$
$$v_{\delta,k}(t) = \langle v(t), K_{\delta,x_k}\rangle, \quad v_{\delta,k}^{\Delta_j}(t) = \langle v(t), (-\Delta)^{\beta_j}K_{\delta,x_k}\rangle, \quad 0 \le t \le T, \quad 1 \le j \le q, \quad 1 \le k \le N,$$

which are collected in the vector-valued observation processes

$$Y_{\delta,k}(t) = \begin{pmatrix} u_{\delta,k}^{\Delta_1}(t) & \dots & u_{\delta,k}^{\Delta_p}(t) & v_{\delta,k}^{\Delta_1}(t) & \dots & v_{\delta,k}^{\Delta_q}(t) \end{pmatrix}^\top \in \mathbb{R}^{p+q}, \quad 0 \le t \le T, \quad k = 1, \dots, N.$$

The augmented MLE $(\widehat{\vartheta}_\delta, \widehat{\eta}_\delta)^\top \in \mathbb{R}^{p+q}$ is given by

$$\begin{pmatrix} \widehat{\vartheta}_\delta \\ \widehat{\eta}_\delta \end{pmatrix} = \mathcal{I}_\delta^{-1} \sum_{k=1}^N \int_0^T Y_{\delta,k}(t)\, \mathrm{d}v_{\delta,k}(t)$$

where the observed Fisher information matrix is defined through

$$\mathcal{I}_\delta = \sum_{k=1}^N \int_0^T Y_{\delta,k}(t) Y_{\delta,k}(t)^\top\, \mathrm{d}t.$$

An error decomposition results in

$$\begin{pmatrix} \widehat{\vartheta}_\delta \\ \widehat{\eta}_\delta \end{pmatrix} = \begin{pmatrix} \vartheta_\delta \\ \eta_\delta \end{pmatrix} + \|K\|_{L^2(\mathbb{R}^d)} \mathcal{I}_\delta^{-1} \mathcal{M}_\delta$$

with the martingale part

$$\mathcal{M}_\delta = \sum_{k=1}^N \int_0^T Y_{\delta,k}(t)\, \mathrm{d}W_k(t). \tag{1.21}$$

1.6.2 *Main results*

Under similar assumptions as imposed in Paper A, we established asymptotic normality, i.e., we derived

$$\rho_\delta^{-1} \begin{pmatrix} \widehat{\vartheta}_\delta - \vartheta \\ \widehat{\eta}_\delta - \eta \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \|K\|_{L^2(\mathbb{R}^d)}^2 \Sigma_{\vartheta,\eta}^{-1}), \quad \delta \to 0,$$

with normalising diagonal matrix $\rho_\delta \in \mathbb{R}^{(p+q)\times(p+q)}$ given by

$$(\rho_\delta)_{ii} := \begin{cases} N^{-1/2}\delta^{2\alpha_i - \alpha_1 - \beta_1}, & i \le p, \\ N^{-1/2}\delta^{2\beta_{i-p} - \beta_1}, & p < i \le p+q. \end{cases}$$

The limiting matrix $\Sigma_{\vartheta,\eta}$ is block-diagonal, implying the asymptotic independence between elasticity and damping coefficients. It depends on the time horizon T like the MLE of the drift in an Ornstein–Uhlenbeck process in the explosive, stable and ergodic case, cf. [37, Proposition 3.46]. Under a full observation scheme $N \asymp \delta^{-d}$, the convergence rates match the rates achieved in the spectral approach [39] up to some logarithmic boundary cases.

1.6.3 *Methodology*

We rewrote the coupled system of equations (1.20) as a first-order system

$$\mathrm{d}X(t) = \mathcal{A}_{\vartheta,\eta} X(t)\, \mathrm{d}t + \begin{pmatrix} 0 \\ I \end{pmatrix} \mathrm{d}W(t), \quad 0 < t \le T,$$

for $X(t) = (u(t), v(t))^\top$. The differential operator $\mathcal{A}_{\vartheta,\eta}$, defined through

$$\mathcal{A}_{\vartheta,\eta} := \begin{pmatrix} 0 & I \\ A_\vartheta & B_\eta \end{pmatrix},$$

generates a strongly continuous semigroup $(J_{\vartheta,\eta}(t))_{t\geq 0}$ given by

$$J_{\vartheta,\eta}(t) := \begin{pmatrix} M(t) & N(t) \\ A_\vartheta N(t) & M(t) + B_\eta N(t) \end{pmatrix} = \begin{pmatrix} M(t) & N(t) \\ M'(t) & N'(t) \end{pmatrix}, \quad t \geq 0.$$

The M, N-functions appearing in the semigroup, cf. [49], are generalisations of cosine and sine operator functions which are related to the solution of the deterministic undamped wave equation. They have additional smoothing properties in similarity to parabolic problems due to the damping operator B_η or rather its generated semigroup $(e^{tB_\vartheta})_{t\geq 0}$ emerging in these functions. Therefore, controlling the observed Fisher information, i.e., showing

$$\mathbb{E}[\rho_\delta \mathcal{I}_\delta \rho_\delta] \to \Sigma_{\vartheta,\eta}, \quad \operatorname{Var}(\rho_\delta \mathcal{I}_\delta \rho_\delta) \to 0, \quad \delta \to 0, \tag{1.22}$$

required a new method as we could neither rely on the asymptotic equipartition of energy [57], nor the pointwise convergence of a certain integrand as utilised in Paper A. Our solution is based on functional calculus. Combined with sufficiently good semigroup upper bounds, we verified (1.22) by the fundamental theorem of calculus.

References

[1] R. Altmeyer, T. Bretschneider, J. Janák, and M. Reiß. "Parameter Estimation in an SPDE Model for Cell Repolarisation". In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1 (2022), pp. 179–199.

[2] R. Altmeyer, I. Cialenco, and G. Pasemann. "Parameter estimation for semilinear SPDEs from local measurements". In: *Bernoulli* 29.3 (2023), pp. 2035–2061.

[3] R. Altmeyer, I. Cialenco, and M. Reiß. *Statistics for SPDEs*. https://sites.google.com/view/stats4spdes. Accessed: 2024-07-30.

[4] R. Altmeyer and M. Reiß. "Nonparametric estimation for linear SPDEs from local measurements". In: *Annals of Applied Probability* 31.1 (2021), pp. 1–38.

[5] T. Aspelmeier, A. Egner, and A. Munk. "Modern statistical challenges in high-resolution fluorescence microscopy". In: *Annual Reviews of Statistics and Its Applications* 2 (2015), pp. 163–202.

[6] A. S. Backer and W. E. Moerner. "Extending Single-Molecule Microscopy Using Optical Fourier Processing". In: *The Journal of Physical Chemistry B* 118.28 (2014), pp. 8313–8329.

[7] M. Bibinger and P. Bossert. "Efficient parameter estimation for parabolic SPDEs based on a log-linear model for realized volatilities". In: *Japanse Journal of Statistics and Data Science* 6 (2023), pp. 407–429.

[8] M. Bibinger and M. Trabs. "On central limit theorems for power variations of the solution to the stochastic heat equation". In: *Stochastic Models, Statistics and Their Applications*. Ed. by A. Steland, E. Rafajłowicz, and O. Okhrin. Springer, 2019, pp. 69–84.

[9] M. Bibinger and M. Trabs. "Volatility estimation for stochastic PDEs using high-frequency observations". In: *Stochastic Processes and their Applications* 130.5 (2020), pp. 3005–3052.

[10] A. Charous and P. Lermusiaux. "Dynamically Orthogonal Differential Equations for Stochastic and Deterministic Reduced-Order Modeling of Ocean Acoustic Wave Propagation". In: *OCEANS 2021 IEEE/MTS San Diego* (2021), pp. 1–7.

[11] Z. Cheng, I. Cialenco, and R. Gong. "Bayesian estimations for diagonalizable bilinear SPDEs". In: *Stochastic Processes and their Applications* 130.2 (2020), pp. 845–877.

[12] C. Chong. "High-frequency analysis of parabolic stochastic PDEs". In: *Annals of Statistics* 48.2 (2020), pp. 1143–1167.

[13] I. Cialenco. "Parameter estimation for SPDEs with multiplicative fractional noise". In: *Stochastics and Dynamics* 10.04 (2010), pp. 561–576.

[14] I. Cialenco. "Statistical inference for SPDEs: an overview". In: *Statistical Inference for Stochastic Processes* 21.2 (2018), pp. 309–329.

[15] I. Cialenco, F. Delgado-Vences, and H.-J. Kim. "Drift Estimation for Discretely Sampled SPDEs". In: *Stochastics and Partial Differential Equations: Analysis and Computations* 8 (2020), pp. 895–920.

[16] I. Cialenco and Y. Huang. “A note on parameter estimation for discretely sampled SPDEs”. In: *Stochastics and Dynamics* 24 (2019). Publisher: World Scientific Publishing Company, Paper No. 2050016.

[17] I. Cialenco and H.-J. Kim. “Parameter estimation for discretely sampled stochastic heat equation driven by space-only noise”. In: *Stochastic Processes and their Applications* 143 (2022), pp. 1–30.

[18] I. Cialenco and L. Xu. “A note on error estimation for hypothesis testing problems for some linear SPDEs”. In: *Stochastic Partial Differential Equations: Analysis and Computations* 2.3 (2014), pp. 408–431.

[19] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.

[20] G. Denaro, D. Valenti, A. La Cognata, B. Spagnolo, A. Bonanno, G. Basilone, S. Mazolla, S. W. Zgozi, S. Aronica, and C. Brunet. “Spatio-temporal behaviour of the deep chlorophyll maximum in Mediteran Sea: Development of a stochastic model for picophytoplankton dynamics”. In: *Ecological Complexity* 13 (2013), pp. 21–34.

[21] L. C. Evans. *Partial Differential Equations*. American Mathematical Soc., 2010.

[22] S. Gaudlitz. *Non-parametric estimation of the reaction term in semi-linear SPDEs with spatial ergodicity*. 2024. arXiv: 2307.05457 [math.PR].

[23] S. Gaudlitz and M. Reiß. “Estimation for the reaction term in semi-linear SPDEs under small diffusivity”. In: *Bernoulli* 29.4 (2023), pp. 3033 –3058.

[24] M. Hairer. *An Introduction to Stochastic PDEs*. 2023. arXiv: 0907.4178 [math.PR].

[25] F. Hildebrandt and M. Trabs. “Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations”. In: *Stochastic Processes and their Applications* 162 (2023), pp. 171–217.

[26] F. Hildebrandt and M. Trabs. “Parameter estimation for SPDEs based on discrete observations in time and space.” In: *Electronic Journal of Statistics* 15 (2021), pp. 2716–2776.

[27] M. Hübner, R. Khasminskii, and B. L. Rozovskii. “Two Examples of Parameter Estimation for Stochastic Partial Differential Equations”. In: *Stochastic Processes: A Festschrift in Honour of Gopinath Kallianpur*. Ed. by S. Cambanis, J. K. Ghosh, R. L. Karandikar, and P. K. Sen. Springer, 1993, pp. 149–160.

[28] M. Huebner. “Parameter Estimation for SPDEs”. PhD Thesis. University of Southern Calofornia, 1993.

[29] M. Huebner and B. Rozovskii. “On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE’s”. In: *Probability Theory and Related Fields* 103.2 (1995), pp. 143–163.

[30] J. Janák and M. Reiß. “Parameter estimation for the stochastic heat equation with multiplicative noise from local measurements”. In: *Stochastic Process. Appl.* 175 (2024), Paper No. 104385.

[31] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[32] Y. Kaino and M. Uchida. “Adaptive estimator for a parabolic linear SPDE with small noise”. In: *Japanese Journal of Statistics and Data Science* 4 (2021), pp. 513–541.

[33] Y. Kaino and M. Uchida. “Parametric estimation for a parabolic linear SPDE model based on discrete observations”. In: *Journal of Statistical Planning and Inference* 211 (2021), pp. 190–220.

[34] K.-Y. Kim and G. R. North. “Surface temperature fluctuations in a stochastic climate model”. In: *Journal of Geophysical Research: Atmospheres* 96.D10 (1991), pp. 18573–18580.

[35] A. P. Korostelëv and A. B. Tsybakov. *Minimax theory of image reconstruction*. Vol. 82. Lecture Notes in Statistics. Springer, 1993.

[36] T. Koski and W. Loges. “Asymptotic statistical inference for a stochastic heat flow problem”. In: *Statistics & Probability Letters* 3.4 (1985), pp. 185–189.

[37] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2013.

[38] M. Lassak. “Covering the Boundary of a Convex Set by Tiles”. In: *Proceedings of the American Mathematical Society* 104.1 (1988), pp. 269–272.

[39] W Liu and S. Lototsky. “Parameter estimation in hyperbolic multichannel models”. In: *Asymptotic Analysis* 68 (2010), pp. 223–248.

[40] W. Liu and S. V. Lototsky. *Estimating Speed and Damping in the Stochastic Wave Equation*. 2008. arXiv: 0810.0046 [math.PR].

[41] W Liu and M Röckner. *Stochastic Partial Differential Equations: An Introduction*. Springer, 2015.

[42] X. Liu, K. Yeo, Y. Hwang, J. Singh, and J. Kalagnanam. “A statistical modeling approach for air quality data based on physical dispersion processes and its application to ozone modeling”. In: *Annals of Applied Statistics* 10.2 (2016), pp. 756–785.

[43] W. Loges. “Girsanov’s theorem in Hilbert space and an application to the statistics of Hilbert space- valued stochastic differential equations”. In: *Stochastic Processes and their Applications* 17.2 (1984), pp. 243–263.

[44] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. English. Cambridge University Press, 2014.

[45] S. Lototsky and B. Rozovsky. *Stochastic partial differential equations*. Springer, 2017.

[46] S. V. Lototsky. “Parameter Estimation for Stochastic Parabolic Equations: Asymptotic Properties of a Two-Dimensional Projection-Based Estimator”. In: *Statistical Inference for Stochastic Processes* 6.1 (2003), pp. 65–87.

[47] S. V. Lototsky. “Statistical inference for stochastic parabolic equations: a spectral approach”. In: *Publicacions Matemàtiques* 53.1 (2009), pp. 3–45.

[48] B. Markussen. “Likelihood inference for a discretely observed stochastic partial differential equations”. In: *Bernoulli* 9.5 (2003), pp. 745–762.

[49] V. Melnikova Irina and A. Filinkov. *Abstract Cauchy Problems: Three Approaches*. Chapman and Hall, 2001.

[50] G. Pasemann and W. Stannat. “Drift estimation for stochastic reaction-diffusion systems”. In: *Electronic Journal of Statistics* 14.1 (2020), pp. 547–579.

[51] M. Reiß, C. Strauch, and L. Trottner. *Change point estimation for a stochastic heat equation*. 2023. arXiv: 2307.10960 [math.ST].

[52] S. E. Serrano. “Random Evolution Equations in Hydrology”. In: *Applied Mathematics and Computation* 38 (1990), pp. 201–226.

[53] W. Strauss. *Partial Differential Equations: An Introduction*. Wiley, 2007.

[54] Y. Tonaki, Y. Kaino, and M. Uchida. “Parameter estimation for linear parabolic SPDEs in two space dimensions based on high frequency data”. In: *Scandinavian Journal of Statistics* 50.4 (2023), pp. 1568–1589.

[55] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

[56] J. B. Walsh. “A stochastic model of neural response”. In: *Advances in Applied Probability* 13.2 (1981), pp. 231–281.

[57] E. Ziebell. *Non-parametric estimation for the stochastic wave equation*. 2024. arXiv: 2404.18823 [math.ST].

Optimal parameter estimation for linear SPDEs from multiple measurements

A

Randolf Altmeyer, Anton Tiepner and Martin Wahl

Abstract

The coefficients in a second order parabolic linear stochastic partial differential equation (SPDE) are estimated from multiple spatially localised measurements. Assuming that the spatial resolution tends to zero and the number of measurements is non-decreasing, the rate of convergence for each coefficient depends on its differential order and is faster for higher order coefficients. Based on an explicit analysis of the reproducing kernel Hilbert space of a general stochastic evolution equation, a Gaussian lower bound scheme is introduced. As a result, minimax optimality of the rates as well as sufficient and necessary conditions for consistent estimation are established.

A.1 Introduction

Stochastic partial differential equations (SPDEs) form a flexible class of models for space-time data. They combine phenomena such as diffusion and transport that occur naturally in many processes, but also include random forcing terms, which may arise from microscopic scaling limits or account for model uncertainty. Quantifying the size of these different effects is an important step in model validation.

Suppose that $X = (X(t))_{0\leq t\leq T}$ solves the linear parabolic SPDE

$$\mathrm{d}X(t) = A_\vartheta X(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 \leq t \leq T, \tag{A.1}$$

on an open, bounded and smooth domain $\Lambda \subset \mathbb{R}^d$ with some initial value X_0, a space-time white noise $\mathrm{d}W$ and a second order elliptic operator

$$A_\vartheta = \sum_{i=1}^{p} \vartheta_i A_i + A_0 \tag{A.2}$$

satisfying zero Dirichlet boundary conditions. The A_i are known differential operators of differential order $n_i \in \{0, 1, 2\}$ and we aim at recovering the unknown parameter $\vartheta \in \mathbb{R}^p$. A prototypical example is

$$A_\vartheta = \vartheta_1\Delta + \vartheta_2(\nabla\cdot b) + \vartheta_3, \qquad \vartheta \in (0,\infty)\times\mathbb{R}\times(-\infty,0], \tag{A.3}$$

with diffusivity, transport and reaction coefficients ϑ_1, ϑ_2, ϑ_3 in front of the Laplace operator Δ and the divergence operator $\nabla\cdot$ such that $n_1 = 2$, $n_2 = 1$, $n_3 = 0$, with a known unit velocity vector $b \in \mathbb{R}^d$. The general form in (A.2) allows for wide range of models affected by a mixture of different, possibly anisotropic, mechanisms. Equations such as (A.1) are also called stochastic advection–diffusion equations and often serve as building blocks for more complex models, with applications in different areas such as neuroscience [50, 57, 60], biology [1, 2], spatial statistics

[41, 53] and data assimilation [42]. For concrete examples of (A.2) with a mixture of known and unknown model coefficients from fluid dynamics and engineering see [11, 29, 45].

While the estimation of a scalar parameter in front of the highest order operator A_i is well studied in the literature [13, 14, 23, 27, 32], there is little known about estimating the lower order coefficients or the full multivariate parameter ϑ. Relying on discrete space-time observations $X(t_k, x_j)$ in case of (A.3) and in dimension $d = 1$, [9, 26, 54] have analysed power variations and contrast estimators. For two parameters in front of operators A_1 and A_2, [44] computed the maximum likelihood estimator from M spectral measurements $(\langle X(t), e_j\rangle)_{0\leq t\leq T}$, $j = 1, \ldots, M$, where the e_j are the eigenfunctions of A_ϑ and $\langle\cdot,\cdot\rangle$ is the inner product on $L^2(\Lambda)$. This leads as $M \to \infty$ to rates of convergence depending on the differential order of the operators A_1, A_2, but is restricted to domains and diagonalisable operators with known e_j, independent of ϑ. In particular, in the spectral approach there is no known estimator for the transport coefficient ϑ_2 in (A.3). Estimators for nonlinearities or noise specifications are studied e.g. by [8, 12, 21, 25].

In contrast, we construct an estimator $\widehat{\vartheta}_\delta$ of ϑ on general domains and with arbitrary possibly anisotropic A_ϑ from local measurement processes

$$X_{\delta,k} = (\langle X(t), K_{\delta,x_k}\rangle)_{0\leq t\leq T}, \quad X^{A_i}_{\delta,k} = (\langle X(t), A_i^* K_{\delta,x_k}\rangle)_{0\leq t\leq T}$$

for $i = 0, \ldots, p$ and locations $x_1, \ldots, x_M \in \Lambda$. The K_{δ,x_k}, also known as *point spread functions* in optical systems [6, 7], are compactly supported functions on subsets of Λ with radius $\delta > 0$ and centred at the x_k. They are part of the observation scheme and describe the physical limitation that point sources $X(t_k, x_j)$ can only be measured up to a convolution with the point spread function. Local measurements were introduced in a recent contribution by [4] to demonstrate that a nonparametric diffusivity can already be identified at x_k from the spatially localised information $X_{\delta,k}$ as $\delta \to 0$ with $T > 0$ fixed. See [3] for robustness to semilinear perturbations and different noise configurations besides space-time white noise. For more details on practical aspects of local measurements, as well as a concrete example from cell biology [2], see Section A.5.3 below.

Let us briefly describe our main contributions. Our first result extends the augmented MLE $\widehat{\vartheta}_\delta$ and the CLT of [4] to $M = M(\delta)$ measurements and joint asymptotic normality of

$$(M^{1/2}\delta^{1-n_i}(\widehat{\vartheta}_{\delta,i} - \vartheta_i))_{i=1}^p, \quad \delta \to 0.$$

This yields the convergence rates $M^{1/2}\delta^{1-n_i}$ for ϑ_i, with the fastest rate for diffusivity terms with $n_i = 2$ and the slowest rate for reaction terms with $n_i = 0$. We then turn to the problem of establishing optimality of these rates in case of (A.3). We compute the reproducing kernel Hilbert space (RKHS) of the Gaussian measures induced by the laws of X and of the local measurements. From this we derive minimax lower bounds, implying that the rates in the CLT are indeed optimal, and provide conditions under which consistent estimation is impossible. Combined with our CLT we deduce for general point spread functions K_{δ,x_k} with non-intersecting supports that consistent estimation is possible if and only if $M^{1/2}\delta^{1-n_i} \to \infty$. Since M is at most of order δ^{-d}, reaction terms cannot be estimated when $d = 1$.

Conceptually, spectral measurements can be obtained approximately from local measurements on a dense grid over the entire domain by a discrete Fourier transform and we recover the rates of convergence of [27] by taking M of maximal order δ^{-d}.

The information geometry underlying local measurements is complex due to the non-linear dependence of the solution X on ϑ (cf. (A.4)) and the non-Markovian dynamics of the processes $X_{\delta,k}, X_{\delta,k}^{A_i}$. This leads to a non-explicit likelihood function, making standard MLE-based estimation and optimality results for continuously observed diffusion processes [36] non-applicable in this context. Instead, we introduce a novel lower bound scheme for Gaussian measures, which exploits that the Hellinger distance of their laws can be related to properties of their RKHS. This is different from the lower bound approach of [4] for $M = 1$ and paves the way to rigorous lower bounds for each coefficient and an arbitrary number of measurements. One of our key results states that the RKHS of the Gaussian measure induced by X with $A_\vartheta = \Delta$ consists of all absolutely continuous $h \in L^2([0,T]; L^2(\Lambda))$ with $\Delta h, h' \in L^2([0,T]; L^2(\Lambda))$ and its squared RKHS norm equals

$$\|\Delta h\|^2_{L^2([0,T];L^2(\Lambda))} + \|h'\|^2_{L^2([0,T];L^2(\Lambda))} + \|(-\Delta)^{1/2} h(0)\|^2_{L^2(\Lambda)} + \|(-\Delta)^{1/2} h(T)\|^2_{L^2(\Lambda)}.$$

This surprisingly simple formula generalises the finite-dimensional Ornstein–Uhlenbeck case [38], and provides a route to obtain the RKHS of local measurements as linear transformations of X. To the best of our knowledge the RKHS of X has not been stated before in the literature, and may be of independent interest, e.g. in constructing Bayesian procedures with Gaussian process priors, cf. [58].

The paper is organised as follows. Section A.2 deals with the local measurement scheme, the construction of our estimator and the CLT. Section A.3 addresses the RKHS of X and of the local measurements, while Section A.4 presents the lower bounds for the rates established in the CLT. Section A.5 covers model examples, the boundary case for estimating zero order terms in $d = 2$ and some practical aspects. All proofs are deferred to Section A.6 and Section A.7.

Basic notation

Throughout the paper, we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$. We write $a \lesssim b$ if $a \le Cb$ for a universal constant C not depending on δ, but possibly depending on other quantities such as T and Λ. Unless stated otherwise, all limits are understood as $\delta \to 0$ with non-decreasing $M = M(\delta)$ possibly depending on δ.

The Euclidean inner product and distance of two vectors $a, b \in \mathbb{R}^p$ is denoted by $a \cdot b$ and $|b - a|$, $I_{p\times p}$ is the identity matrix in $\mathbb{R}^{p\times p}$. We write $\|\cdot\|_{\mathrm{op}}$ for the operator norm of a matrix. For an open set $U \subset \mathbb{R}^d$ and $p \ge 1$, $L^p(U)$ is the usual L^p-space with norm $\|\cdot\|_{L^p(U)}$ and the inner product on $L^2(U)$ is denoted $\langle\cdot,\cdot\rangle_{L^2(U)}$. We write $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{L^2(\Lambda)}$, $\|\cdot\| = \|\cdot\|_{L^2(\Lambda)}$. Let $H^k(U)$ denote the usual Sobolev spaces and let $H_0^1(U)$ be the completion of the space of smooth compactly supported functions $C_c^\infty(U)$ relative to the $H^1(U)$-norm.

We write D_i, D_{ij} for partial derivatives. The gradient and Laplace operators are ∇, $\Delta = \sum_{i=1}^d D_{ii}$. The divergence of a d-dimensional vector field v is $\nabla \cdot v = \sum_{i=1}^d D_i v_i$. The Laplace operator Δ will be considered with domain $H_0^1(\Lambda) \cap H^2(\Lambda)$, while with domain $H^2(\mathbb{R}^d)$ it will be denoted by Δ_0.

For a Hilbert space $\mathcal{H}$, the space $L^2([0,T]; \mathcal{H})$ consists of all measurable functions $h : [0,T] \to \mathcal{H}$ with $\int_0^T \|h(t)\|^2_{\mathcal{H}}\, \mathrm{d}t < \infty$. We write $\|T\|_{\mathrm{HS}(\mathcal{H}_1,\mathcal{H}_2)}$ for the Hilbert–Schmidt norm of a linear operator $T : \mathcal{H}_1 \to \mathcal{H}_2$ between two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$.

A.2 Joint parameter estimation

A.2.1 *Setup*

Let $\vartheta \in \Theta \subset \mathbb{R}^p$ be an unknown parameter. For $i = 0, \dots, p$, suppose that the operators in (A.2) are of the form $A_i = \nabla \cdot a^{(i)} \nabla + \nabla \cdot b^{(i)} + c^{(i)}$ for symmetric $a^{(i)} \in \mathbb{R}^{d\times d}$, $b^{(i)} \in \mathbb{R}^d$ and $c^{(i)} \in \mathbb{R}$, where for each $i = 1, \dots, p$ only one of the coefficients $a^{(i)}$, $b^{(i)}$, $c^{(i)}$ is non-vanishing. For each A_i, the formal adjoint is $A_i^* = \nabla \cdot a^{(i)} \nabla - \nabla \cdot b^{(i)} + c^{(i)}$, and its differential order $n_i = \mathrm{ord}(A_i) \in \{0, 1, 2\}$ is the number of non-vanishing derivatives. With $a_\vartheta = \sum_{i=1}^p \vartheta_i a^{(i)} + a^{(0)}$, $b_\vartheta = \sum_{i=1}^p \vartheta_i b^{(i)} + b^{(0)}$ and $c_\vartheta = \sum_{i=1}^p \vartheta_i c^{(i)} + c^{(0)}$, (A.2) gives

$$A_\vartheta = \nabla \cdot a_\vartheta \nabla + \nabla \cdot b_\vartheta + c_\vartheta.$$

We suppose that a_ϑ is positive definite for all $\vartheta \in \Theta$. Then A_ϑ is a strongly elliptic operator and generates with domain $H_0^1(\Lambda) \cap H^2(\Lambda)$ an analytic semigroup $(S_\vartheta(t))_{t\geq 0}$ on $L^2(\Lambda)$ [46]. Considered with the same domain, the adjoint $A_\vartheta^* = \sum_{i=1}^p \vartheta_i A_i^* + A_0^*$ generates the adjoint semigroup $(S_\vartheta^*(t))_{t\geq 0}$ [62, Section 2.5.3].

With an $\mathcal{F}_0$-measurable initial value X_0 and a cylindrical Wiener process W on $L^2(\Lambda)$ define a process $X = (X(t))_{0\leq t\leq T}$ by

$$X(t) = S_\vartheta(t) X_0 + \int_0^t S_\vartheta(t-t')\,\mathrm{d}W(t'), \quad 0 \leq t \leq T. \tag{A.4}$$

Due to the low spatial regularity of W this process is understood as a random element with values in $L^2(\Lambda) \subset \mathcal{H}_1$ almost surely for a larger Hilbert space $\mathcal{H}_1$ with an embedding $\iota : L^2(\Lambda) \to \mathcal{H}_1$ such that $\int_0^t \|\iota S_\vartheta(t')\|^2_{\mathrm{HS}(L^2(\Lambda),\mathcal{H}_1)}\,\mathrm{d}t' < \infty$ [24, Remark 6.6]. Such an embedding always exists. For example, $\mathcal{H}_1$ can be realised as a negative Sobolev space (see Section A.6.2 below). Let $\mathcal{H}_1'$ denote the dual space of $\mathcal{H}_1$ with the associated dual pairing $\langle \cdot, \cdot \rangle_{\mathcal{H}_1 \times \mathcal{H}_1'}$. Let $(e_k)_{k\geq 1}$ be an orthonormal basis of $L^2(\Lambda)$ and let β_k be independent scalar Brownian motions. Then, realising the Wiener process as $W = \sum_{k\geq 1} e_k \beta_k$, we find for all $z \in \mathcal{H}_1' \subset L^2(\Lambda)$, $0 \leq t \leq T$, that (see, e.g. [40, Lemma 2.4.1 and Proposition 2.4.5])

$$\begin{aligned}\langle X(t) - S_\vartheta(t) X_0, z\rangle_{\mathcal{H}_1\times\mathcal{H}_1'} &= \sum_{k\geq 1} \int_0^t \langle S_\vartheta(t-t') e_k, z\rangle_{\mathcal{H}_1\times\mathcal{H}_1'}\,\mathrm{d}\beta_k(t') \\ &= \int_0^t \langle S_\vartheta^*(t-t') z, \mathrm{d}W(t')\rangle.\end{aligned}$$

According to [4, Proposition 2.1] and [40, Lemma 2.4.2] this allows us to extend the dual pairings $\langle X(t), z\rangle_{\mathcal{H}_1\times\mathcal{H}_1'}$ to a real-valued Gaussian process $(\langle X(t), z\rangle)_{0\leq t\leq T, z\in L^2(\Lambda)}$ by

$$\langle X(t), z\rangle = \langle S_\vartheta(t) X_0, z\rangle + \int_0^t \langle S_\vartheta^*(t-t') z, \mathrm{d}W(t')\rangle \tag{A.5}$$

(the notation $\langle X(t), z\rangle$ is used for convenience and indicates that the process does not depend on the embedding space $\mathcal{H}_1$). This process solves the SPDE (A.1) in the sense that for all $z \in H_0^1(\Lambda) \cap H^2(\Lambda)$ and $0 \leq t \leq T$

$$\langle X(t), z\rangle = \langle X_0, z\rangle + \int_0^t \langle X(t'), A_\vartheta^* z\rangle\,\mathrm{d}t' + \langle W(t), z\rangle, \tag{A.6}$$

where $\langle W(t), z\rangle / \|z\|_{L^2(\Lambda)}$ is a scalar Brownian motion.

A.2.2 *Local measurements, construction of the estimator*

Introduce for $z \in L^2(\mathbb{R}^d)$ the scale and shift operation

$$z_{\delta,x}(y) = \delta^{-d/2} z(\delta^{-1}(y-x)), \quad x, y \in \Lambda, \quad \delta > 0. \tag{A.7}$$

Suppose that $K \in H^2(\mathbb{R}^d)$ is an (unscaled) point spread function with compact support (see Section A.5 for concrete examples). Consider locations $x_1, \dots, x_M \in \Lambda$, $M \in \mathbb{N}$, and a resolution level $\delta > 0$, which is small enough to ensure that the point spread functions K_{δ,x_k} are supported on Λ. Local measurements of X at the locations $x_1, \dots, x_M$ at resolution δ correspond to the continuously observed processes $X_\delta, X_\delta^{A_0} \in L^2([0,T];\mathbb{R}^M)$, $X_\delta^A \in L^2([0,T];\mathbb{R}^{p\times M})$, where for $i = 1, \dots, p$, $k = 1, \dots, M$

$$\begin{aligned}
(X_\delta)_k &= X_{\delta,k} = (\langle X(t), K_{\delta,x_k}\rangle)_{0\le t\le T},\\
(X_\delta^{A_0})_k &= X_{\delta,k}^{A_0} = (\langle X(t), A_0^* K_{\delta,x_k}\rangle)_{0\le t\le T},\\
(X_\delta^{A})_{ik} &= X_{\delta,k}^{A_i} = (\langle X(t), A_i^* K_{\delta,x_k}\rangle)_{0\le t\le T}.
\end{aligned}$$

According to (A.6), every local measurement is an Itô process

$$\mathrm{d}X_{\delta,k}(t) = \left(\sum_{i=1}^p \vartheta_i X_{\delta,k}^{A_i}(t) + X_{\delta,k}^{A_0}(t)\right)\mathrm{d}t + \|K\|_{L^2(\mathbb{R}^d)}\,\mathrm{d}W_k(t) \tag{A.8}$$

with initial values $X_{\delta,k}(0) = \langle X_0, K_{\delta,x_k}\rangle$ and scalar Brownian motions

$$W_k(t) = \langle W(t), K_{\delta,x_k}\rangle / \|K\|_{L^2(\mathbb{R}^d)}.$$

It should be noted that neither (A.8) nor the system of equations augmented with $X_\delta^A, X_\delta^{A_0}$ are Markov processes, because the time evolution at x_k depends on the spatial structure of the whole process X, and not only of X_δ. This is due to the infinite speed of propagation in space by $S_\vartheta(t)$. This also means the processes $X_{\delta,k}$ are *not* independent, even if the driving noise processes W_k are, e.g., due to non-overlapping supports of the K_{δ,x_k} as will be assumed below. Therefore, standard results for estimating the parameters ϑ_i from continuously observed diffusion processes by the maximum likelihood estimator (e.g., [36]) do not apply here. Instead, a general Girsanov theorem for multivariate Itô processes, cf. [39, Section 7.6], yields after ignoring conditional expectations, the initial value and possible correlations between measurements the modified log-likelihood function

$$\begin{aligned}
\ell_\delta(\vartheta) = \|K\|_{L^2(\mathbb{R}^d)}^{-2} \sum_{k=1}^M \Bigg(\int_0^T &\left(\sum_{i=1}^p \vartheta_i X_{\delta,k}^{A_i}(t) + X_{\delta,k}^{A_0}(t)\right)\mathrm{d}X_{\delta,k}(t)\\
&- \frac{1}{2}\int_0^T \left(\sum_{i=1}^p \vartheta_i X_{\delta,k}^{A_i}(t) + X_{\delta,k}^{A_0}(t)\right)^2 \Bigg)\mathrm{d}t.
\end{aligned}$$

Maximising $\ell_\delta(\vartheta)$ with respect to ϑ leads to the estimator

$$\widehat{\vartheta}_\delta = \mathcal{I}_\delta^{-1} \sum_{k=1}^M \left(\int_0^T X_{\delta,k}^A(t)\,\mathrm{d}X_{\delta,k}(t) - \int_0^T X_{\delta,k}^A(t) X_{\delta,k}^{A_0}(t)\,\mathrm{d}t \right), \tag{A.9}$$

which we call *augmented MLE* generalising [2, Section 4.1], with *observed Fisher information*

$$\mathcal{I}_\delta = \sum_{k=1}^M \int_0^T X_{\delta,k}^A(t) X_{\delta,k}^A(t)^\top\,\mathrm{d}t. \tag{A.10}$$

A.2.3 *A central limit theorem*

We show now that the augmented MLE $\widehat{\vartheta}_\delta$ satisfies a CLT as $\delta \to 0$. Replacing $\mathrm{d}X_{\delta,k}(t)$ in the definition of the augmented MLE by the right hand side in (A.8) yields the basic decomposition

$$\widehat{\vartheta}_\delta = \vartheta + \|K\|_{L^2(\mathbb{R}^d)} \mathcal{I}_\delta^{-1} \mathcal{M}_\delta \tag{A.11}$$

with the martingale term

$$\mathcal{M}_\delta = \sum_{k=1}^{M} \left(\int_0^T X^A_{\delta,k}(t)\,\mathrm{d}W_k(t) \right). \tag{A.12}$$

If the Brownian motions W_k are independent, then the matrix $\mathcal{I}_\delta$ corresponds to the quadratic covariation process of $\mathcal{M}_\delta$ and we therefore expect $\mathcal{I}_\delta^{-1/2}\mathcal{M}_\delta$ to follow approximately a multivariate normal distribution. The rate at which the estimation error in (A.11) vanishes corresponds to the speed at which the components of the observed Fisher information diverge. Exploiting scaling properties of the underlying semigroup (cf. Lemma A.13), we will see that this depends on its action on the point spread functions K_{δ,x_k}. We define a diagonal matrix of scaling coefficients $\rho_\delta \in \mathbb{R}^{p\times p}$,

$$(\rho_\delta)_{ii} = M^{-1/2}\delta^{n_i - 1}, \tag{A.13}$$

and make the following additional structural assumptions.

Assumption H.

(i) The functions $A_i K$ are linearly independent for all $i = 1, \dots, p$.

(ii) $n_i > 1 - d/2$ for all $i = 1, \dots, p$.

(iii) The locations x_k, $k = 1, \dots, M$, belong to a fixed compact set $\mathcal{J} \subset \Lambda$, which is independent of δ and M. There exists $\delta' > 0$ such that $\operatorname{supp}(K_{\delta,x_k}) \cap \operatorname{supp}(K_{\delta,x_l}) = \emptyset$ for $k \neq l$ and all $\delta \leq \delta'$.

(iv) $\sup_{x\in\mathcal{J}} \int_0^T \mathbb{E}[\langle X_0, S_\vartheta^*(t) A_i^* K_{\delta,x}\rangle^2]\,\mathrm{d}t = o(\delta^{2-2n_i})$ for all $i = 1, \dots, p$.

Assumption H(i) guarantees invertibility of the observed Fisher information, for a proof see Section A.7.1.

LEMMA A.1. *Under Assumption H(i), $\mathcal{I}_\delta$ is $\mathbb{P}$-almost surely invertible.*

The support condition in Assumption H(iii) is natural in applications, e.g., in microscopy. It guarantees $\langle K_{\delta,x_k}, K_{\delta,x_l}\rangle = 0$ and thus independence of the Brownian motions W_k in (A.8) as $\delta \to 0$. It holds for x_k, which are separated by a Euclidean distance of at least $C\delta$ for a fixed constant C, hence there are at most $M = O(\delta^{-d})$ such locations. The next lemma shows that Assumption H(iv) on the initial value is satisfied in most relevant situations. For a proof see again Section A.7.1.

LEMMA A.2. *Assumption H(ii) implies Assumption H(iv) for any $X_0 \in L^q(\Lambda)$, $q > 2$, and if $c_\vartheta \leq 0$ also for the stationary initial condition $X_0 = \int_{-\infty}^0 S_\vartheta(-t')\,\mathrm{d}W(t')$.*

We establish now the asymptotic behaviour of the observed Fisher information and a CLT for the augmented MLE as the resolution δ tends to zero. To this extent, consider the positive operator $-\nabla\cdot a_\vartheta\nabla$ with domain $H^2(\mathbb{R}^d)$. Its spectral calculus induces for each $s\in\mathbb{R}$ the fractional operator $(-\nabla\cdot a_\vartheta\nabla)^s$, which acts in the Fourier domain as the multiplication operator with multiplier $\xi\mapsto(-\xi^\top a_\vartheta\xi)^s$, cf. [37] or [18, Chapter VI.5]. By positive definiteness of a_ϑ, this means $(-\nabla\cdot a_\vartheta\nabla)^s z\in L^2(\mathbb{R}^d)$ as soon as $\xi\mapsto|\xi|^{2s}\mathcal{F}z(\xi)\in L^2(\mathbb{R}^d)$ with the Fourier transform $\mathcal{F}z$. By usual Fourier calculus [18, Lemma VI.5.4], $\mathcal{F}D_jz=i\xi_j\mathcal{F}z$. Together with Assumption H(ii), this means $(-\nabla\cdot a_\vartheta\nabla)^{-1/2}A_i^*K\in L^2(\mathbb{R}^d)$ for all $i=1,\dots,p$.

Theorem A.3. *Under Assumption H the matrix $\Sigma_\vartheta\in\mathbb{R}^{p\times p}$ with entries*

$$(\Sigma_\vartheta)_{ij}=(T/2)\langle(-\nabla\cdot a_\vartheta\nabla)^{-1/2}A_i^*K,(-\nabla\cdot a_\vartheta\nabla)^{-1/2}A_j^*K\rangle_{L^2(\mathbb{R}^d)}$$

is invertible and $\rho_\delta\mathcal{I}_\delta\rho_\delta\xrightarrow{\mathbb{P}}\Sigma_\vartheta$ as $\delta\to0$. Moreover, the augmented MLE satisfies the CLT

$$(\rho_\delta\mathcal{I}_\delta\rho_\delta)^{1/2}\rho_\delta^{-1}(\widehat{\vartheta}_\delta-\vartheta)\xrightarrow{d}\mathcal{N}(0,\|K\|^2_{L^2(\mathbb{R}^d)}I_{p\times p}),\quad\delta\to0,$$

or, equivalently,

$$(M^{1/2}\delta^{1-n_i}(\widehat{\vartheta}_{\delta,i}-\vartheta_i))_{i=1}^p\xrightarrow{d}\mathcal{N}(0,\|K\|^2_{L^2(\mathbb{R}^d)}\Sigma_\vartheta^{-1}).$$

Theorem A.3 shows that parameters ϑ_i of an operator A_i with differential order n_i can be estimated at the rate of convergence $M^{1/2}\delta^{1-n_i}$. Consistency requires $M^{1/2}\delta^{1-n_i}\to\infty$. This excludes reaction terms ϑ_i in $d=1$ with $n_i=0$ and $M=O(\delta^{-1})$, but in $d=2$ a logarithmic rate holds for a restricted class of functions K, see Proposition A.12. The asymptotic variances for two parameters ϑ_i,ϑ_j are independent if A_i^*K and A_j^*K are orthogonal in the geometry induced by $\|(-\nabla\cdot a_\vartheta\nabla)^{-1/2}\cdot\|_{L^2(\mathbb{R}^d)}$. The theorem generalises [4, Theorem 5.3] in the parametric case to the anisotropic setting with M measurement locations.

A.3 The RKHS

In Section A.4, we show optimality of the rates of convergence appearing in Theorem A.3. A crucial ingredient for these lower bound considerations is a good understanding of the reproducing kernel Hilbert space (RKHS) of the Gaussian measure induced by the law of the observations when $A_\vartheta=\Delta$.

We first derive the RKHS of the stochastic convolution (A.4) in a more general setting. Suppose that A is an (unbounded) negative self-adjoint closed operator on a Hilbert space $(\mathcal{H},\|\cdot\|_{\mathcal{H}})$ with domain $\mathcal{D}(A)\subset\mathcal{H}$ such that $Ae_j=-\lambda_je_j$ for a non-decreasing sequence $(\lambda_j)_{j\geq1}$ of positive real numbers with $0<\lambda_1\leq\lambda_2\leq\cdots$ and an orthonormal basis $(e_j)_{j\geq1}$ of $\mathcal{H}$, and such that A generates a strongly continuous semigroup $(S(t))_{t\geq0}$ on $\mathcal{H}$ [18]. With a cylindrical Wiener process W, consider the stationary stochastic convolution

$$X(t)=\int_{-\infty}^t S(t-t')\,\mathrm{d}W(t'),\quad t\geq0.\tag{A.14}$$

As discussed after (A.4) the process $X=(X(t))_{0\leq t\leq T}$ is understood as a random element with values in $\mathcal{H}\subset\mathcal{H}_1$ almost surely for some larger Hilbert space $\mathcal{H}_1$.

In what follows, we use the convention that a RKHS is denoted by the letter H. Moreover, we add a subscript to indicate the process which is under consideration. For instance, H_X denotes the RKHS of X considered as a Gaussian random variable taking values in the Hilbert space $L^2([0,T];\mathcal{H}_1)$. Since the RKHS of X depends only on its distribution, the RKHS, as well as its norm, in the next theorem are independent of the embedding space $\mathcal{H}_1$ (see, e.g., Exercise 2.6.5 in [22]). For the proof and some background on the RKHS of a Gaussian measure see Section A.6.2.

THEOREM A.4. *Let $(H_X, \|\cdot\|_X)$ be the RKHS of the process X in (A.14). Let $T \geq 1$. Then*

$$H_X = \{h \in L^2([0,T];\mathcal{H}) : h \text{ absolutely continuous, } Ah, h' \in L^2([0,T];\mathcal{H})\}$$

and for $h \in H_X$

$$\|h\|_X^2 = \|Ah\|^2_{L^2([0,T];\mathcal{H})} + \|h'\|^2_{L^2([0,T];\mathcal{H})} + \|(-A)^{1/2}h(0)\|^2_{\mathcal{H}} + \|(-A)^{1/2}h(T)\|^2_{\mathcal{H}},$$

as well as

$$\|h\|_X^2 \leq 3\|Ah\|^2_{L^2([0,T];\mathcal{H})} + \|h\|^2_{L^2([0,T];\mathcal{H})} + 2\|h'\|^2_{L^2([0,T];\mathcal{H})}.$$

Note that $h, h', Ah \in L^2([0,T];\mathcal{H})$ implies that the map $t \mapsto \langle Ah(t), h(t)\rangle$ is absolutely continuous (cf. the proof of [19, Theorem 5.9.3] and the proof of Theorem A.4), so that the norm $\|\cdot\|_X$ is indeed well-defined. Theorem A.4 generalises the result for scalar Ornstein–Uhlenbeck processes to the infinite dimensional process X, cf. Lemma A.20 below.

Next, as in (A.6), consider the Gaussian process $(\langle X(t), z\rangle_{\mathcal{H}})_{t\geq 0, z\in\mathcal{H}}$, where the 'inner product' here corresponds to

$$\langle X(t), z\rangle_{\mathcal{H}} = \int_{-\infty}^{t} \langle S(t-t')z, \mathrm{d}W(t')\rangle_{\mathcal{H}},$$

satisfying (A.6) for $z \in \mathcal{D}(A^*) = \mathcal{D}(A)$ by analogous arguments. We study the RKHS of $(\langle X(t), z\rangle_{\mathcal{H}})_{0\leq t\leq T}$ for finitely many z. A first proof considers z from the dual space of $\mathcal{H}_1$. In that case, we can realise $\langle X, z\rangle_{\mathcal{H}}$ as a linear transformation of X by a bounded linear map L from $L^2([0,T];\mathcal{H}_1)$ to $L^2([0,T])^M$, and this allows for relating the RKHS of X and $\langle X, z\rangle_{\mathcal{H}}$ using Theorem A.4. Another proof is presented in Section A.7.7, which circumvents this by an approximation argument.

THEOREM A.5. *For $K_1, \ldots, K_M \in \mathcal{D}(A)$ and with X in (A.14) consider the process X_K with $X_K(t) = (\langle X(t), K_k\rangle_{\mathcal{H}})_{k=1}^M$. Suppose that the Gram matrix $G = (\langle K_k, K_l\rangle_{\mathcal{H}})_{1\leq k,l\leq M}$ is non-singular, and let $G_A = (\langle AK_k, AK_l\rangle_{\mathcal{H}})_{1\leq k,l\leq M}$. Let $T \geq 1$. Then the RKHS $(H_{X_K}, \|\cdot\|_{X_K})$ of X_K satisfies $H_{X_K} = H^M$, where*

$$H = \{h \in L^2([0,T]) : h \text{ absolutely continuous}, h' \in L^2([0,T])\}$$

and for $h = (h_k)_{k=1}^M \in H_{X_K}$

$$\|h\|^2_{X_K} \leq (3\|G^{-1}\|^2_{\mathrm{op}}\|G_A\|_{\mathrm{op}} + \|G^{-1}\|_{\mathrm{op}}) \sum_{k=1}^M \|h_k\|^2_{L^2([0,T])} + 2\|G^{-1}\|_{\mathrm{op}} \sum_{k=1}^M \|h'_k\|^2_{L^2([0,T])}.$$

Theorem A.5 (and its slight generalization in (A.31)) can be used to compute the RKHS of quite general observation schemes. In the specific case $A = \Delta$ and local measurements with $K_k = K_{\delta,x_k}$ we obtain the following.

Corollary A.6. *Let $(H_{X_\delta}, \|\cdot\|_{X_\delta})$ be the RKHS of X_δ with respect to $A = \Delta$, $K \in H^2(\mathbb{R}^d)$ with $\|K\|_{L^2(\mathbb{R}^d)} = 1$, and points $x_1, \dots, x_M$ such that $\operatorname{supp}(K_{\delta,x_k}) \subset \Lambda$ for all $k = 1, \dots, M$ and $\operatorname{supp}(K_{\delta,x_k}) \cap \operatorname{supp}(K_{\delta,x_l}) = \emptyset$ for all $1 \le k \ne l \le M$. Suppose that $\delta^2 \le \|\Delta K\|_{L^2(\mathbb{R}^d)}$ and $T \ge 1$. Then $H_{X_\delta} = H^M$ and for $h = (h_k)_{k=1}^M \in H_{X_\delta}$*

$$\|h\|_{X_\delta}^2 \le 4\frac{\|\Delta K\|_{L^2(\mathbb{R}^d)}^2}{\delta^4}\sum_{k=1}^M \|h_k\|_{L^2([0,T])}^2 + 2\sum_{k=1}^M \|h_k'\|_{L^2([0,T])}^2.$$

Similar results hold for the RKHS of (X_δ, X_δ^A), see Corollary A.28.

A.4 Optimality

In this section, we show that the rates of convergence $M^{1/2}\delta^{1-n_i}$ achieved by the augmented MLE for parameters ϑ_i with respect to operators A_i of order $n_i = \operatorname{ord}(A_i)$ are indeed optimal and cannot be improved in our general setup. The proof strategy (presented in Section A.6.3) relies on a novel lower bound scheme for Gaussian measures by relating the Hellinger distance of their laws to properties of their RKHS. The Gaussian lower bound is then applied to one-dimensional submodels $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta_i}$ with A_ϑ from (A.3) assuming a sufficiently regular kernel function K and a stationary initial condition.

Assumption L. Suppose that $\mathbb{P}_\vartheta$ corresponds to the law of the stationary solution X to the SPDE (A.1) and assume that the following conditions hold:

(i) The kernel function satisfies $K = \Delta^2\widetilde{K}$ with $\widetilde{K} \in C_c^\infty(\mathbb{R}^d)$.

(ii) The models are $A_\vartheta = \vartheta_1\Delta + \vartheta_2(\nabla\cdot b) + \vartheta_3$ for $\vartheta \in \mathbb{R}^3$, a fixed unit vector $b \in \mathbb{R}^d$, and where ϑ lies in one of the parameter classes

$$\Theta_1 = \{\vartheta = (\vartheta_1, 0, 0) : \vartheta_1 \ge 1\},$$
$$\Theta_2 = \{\vartheta = (1, \vartheta_2, 0) : \vartheta_2 \in [0,1]\},$$
$$\Theta_3 = \{\vartheta = (1, 0, \vartheta_3) : \vartheta_3 \le 0\}.$$

(iii) Let $x_1, \dots, x_M$ be δ-separated points in Λ, that is, $|x_k - x_l| > \delta$ for all $1 \le k \ne l \le M$. Moreover, suppose that $\operatorname{supp}(K_{\delta,x_k}) \subset \Lambda$ for all $k = 1, \dots, M$ and $\operatorname{supp}(K_{\delta,x_k}) \cap \operatorname{supp}(K_{\delta,x_l}) = \emptyset$ for all $1 \le k \ne l \le M$.

The parameter classes Θ_i correspond to the cases of estimating the diffusivity ϑ_1, transport coefficient ϑ_2 and reaction coefficient ϑ_3 in front of operators A_i with differential orders $n_1 = 2$, $n_2 = 1$, $n_3 = 0$. We start with a non-asymptotic lower bound when only X_δ is observed.

Theorem A.7. *Grant Assumption L with $M \ge 1$, $T \ge 1$ and let $i \in \{1,2,3\}$. Then there exist constants $c_1, c_2 > 0$ depending only on K and an absolute constant $c_3 > 0$ such that the following assertions hold:*

(i) *If* $\delta^{n_i-1}/\sqrt{TM} < 1$ *and* $\delta \leq c_1$*, then*

$$\inf_{\widehat{\vartheta}_i} \sup_{\substack{\vartheta\in\Theta_i \\ |\vartheta-(1,0,0)^\top|\leq c_2 \frac{\delta^{n_i-1}}{\sqrt{TM}}}} \mathbb{P}_\vartheta\left(|\widehat{\vartheta}_i - \vartheta_i| \geq \frac{c_2}{2}\frac{\delta^{n_i-1}}{\sqrt{TM}}\right) > c_3.$$

(ii) *If* $\delta^{n_i-1}/\sqrt{TM} \geq 1$ *and* $\delta \leq c_1$*, then*

$$\inf_{\widehat{\vartheta}_i} \sup_{\substack{\vartheta\in\Theta_i \\ |\vartheta-(1,0,0)^\top|\leq c_2}} \mathbb{P}_\vartheta(|\widehat{\vartheta}_i - \vartheta_i| \geq c_2/2) > c_3.$$

In (i) and (ii), the infimum is taken over all real-valued estimators $\widehat{\vartheta}_i = \widehat{\vartheta}_i(X_\delta)$.

Several comments are in order for the above result. First, by Markov's inequality Theorem A.7 also implies lower bounds for the squared risk. Second, part (ii) detects settings under which consistent estimation is impossible. For instance, if $i = 2$, then consistent estimation is impossible for $T = 1$ (resp. T bounded) and $M = 1$, that is, if only a single spatial measurement is observed in a bounded time interval. A similar conclusion holds in the case $i = 3$, in which case consistent estimation is even impossible in a full observation scheme with $M = \lceil c\delta^{-d}\rceil$ locations for $d \leq 2$ and T bounded. Third, part (i) of Theorem A.7 shows that the different rates in our CLT are minimax optimal. In particular, it easily implies an asymptotic minimax lower bound when $\delta \to 0$. A first important case is $M = 1$ and $i = 1$ in which case Theorem A.7 also follows from Proposition 5.12 in [4] and gives the rate of convergence δ. For $M = \lceil c\delta^{-d}\rceil$ we get the following.

COROLLARY A.8. *Grant Assumption L with* $M = \lceil c\delta^{-d}\rceil$, $\delta \to 0$ *and* $T \geq 1$*, and let* $i \in \{1, 2, 3\}$*. If* $n_i - 1 + d/2 > 0$*, then*

$$\liminf_{\delta\to 0} \inf_{\widehat{\vartheta}_i} \sup_{|\vartheta-(1,0,0)^\top|\leq c_1} \mathbb{P}_\vartheta(\delta^{-n_i+1-d/2}|\widehat{\vartheta}_i - \vartheta_i| \geq c_2) > 0,$$

where the infimum is taken over all real-valued estimators $\widehat{\vartheta}_i = \widehat{\vartheta}_i(X_\delta)$.

Similar optimality results have been derived in [27] for the case of M spectral measurements. Provided there exists an orthonormal basis of eigenfunctions $(e_j)_{j=1}^\infty$ of A_ϑ independent of ϑ (e.g., in the case $i = 1$ or $i = 3$), it is possible to estimate ϑ_i from M spectral measurements $(\langle X(t), e_j\rangle)_{0\leq t\leq T, 1\leq j\leq M}$ with rates $M^{-\tau}$ or $\log M$ if $\tau = n_i/d - 1/d + 1/2 > 0$ or $\tau = 0$, respectively. Consistent estimation fails to hold for $\tau < 0$. While [27] obtained asymptotic efficiency by combining Girsanov's theorem with LAN techniques, these rates can also be derived from Lemma A.29 combined with a version of Lemma A.23. For $\delta = cM^{-1/d}$ the rate in Corollary A.8 and Theorem A.3 coincides with $M^{-\tau}$ if $\tau > 0$, and $\tau = 0$ is again a boundary case. Regarding the latter case, we briefly discuss in Section A.5 that a non-negative point spread function achieves the $\log M$-rate when $i = 3$ and $d = 2$.

Recall that the augmented MLE $\widehat{\vartheta}_\delta$ depends also on the measurements X_δ^A. We show next that including them into the lower bounds does not change the optimal rates of convergence.

THEOREM A.9. *Theorem A.7 remains valid when the infimum is taken over all real-valued estimators* $\widehat{\vartheta}_i = \widehat{\vartheta}_i(X_\delta, X_\delta^\Delta, X_\delta^{\nabla\cdot b})$*, provided that* K*,* ΔK *and* $(\nabla \cdot b)K$ *are linearly independent and Assumption L holds for* K*,* ΔK *and* $(\nabla \cdot b)K$*.*

A.5 Applications and extensions

A.5.1 *Examples*

Let us illustrate the main results in two examples.

Example A.10. Suppose $A_\vartheta = \vartheta_1\Delta + \vartheta_2\nabla\cdot b + c$ for $\vartheta_1 > 0$. This corresponds to (A.2) with $A_0 = c$, $A_1 = \Delta$, $A_2 = \nabla\cdot b$ for $c \in \mathbb{R}$ and a unit vector $b \in \mathbb{R}^d$, and with differential orders $n_1 = 2$, $n_2 = 1$. A typical realisation of the solution X in $d = 1$ can be seen in Figure A.1(left). For known c, the augmented MLE $\widehat{\vartheta}_\delta$ is a consistent estimator of $\vartheta \in \mathbb{R}^2$ by Theorem A.3, attaining the optimal rates of convergence $M^{1/2}\delta^{-1}$, $M^{1/2}$ for the diffusivity and the transport terms, respectively according to the lower bounds in Theorem A.7. If we suppose for simplicity $\|K\|_{L^2(\mathbb{R}^d)} = 1$, then the CLT holds with a diagonal matrix

$$\Sigma_\vartheta = \frac{T}{2\vartheta_1}\operatorname{diag}\Big(\|\nabla K\|^2_{L^2(\mathbb{R}^d)}, \|(-\Delta_0)^{-1/2}(\nabla\cdot b)K\|^2_{L^2(\mathbb{R}^d)}\Big),$$

implying that $\widehat{\vartheta}_{\delta,1}$ and $\widehat{\vartheta}_{\delta,2}$ are asymptotically independent.

Figure A.1(right) presents root mean squared errors in $d = 1$ for local measurements obtained from the data displayed in the left part of the figure with $K(x) = \exp(-5/(1-x^2))\mathbf{1}(-1 < x < 1)$ and the maximal choice of $M \asymp \delta^{-1}$. We see that the optimal rates of convergence, and even the exact asymptotic variances (blue dashed lines) are approached quickly as $\delta \to 0$. For comparison, we have included in Figure A.1(right) estimation errors for an estimator $\bar{\vartheta}_\delta$ without the correction factor depending on the lower order 'nuisance operator' A_0 in (A.9). We can see that this introduces only a small bias, which is negligible as $\delta \to 0$.

Example A.11. Consider now $A_\vartheta = \vartheta_1\Delta + \vartheta_2\nabla\cdot b + \vartheta_3$ such that A_1, A_2 are as in the last example, but now also $A_0 = 0$, $A_3 = 1$ with $n_3 = 0$. If $d \geq 3$ and $M^{1/2}\delta \to \infty$, then the CLT in Theorem A.3 applies with optimal rates of convergence as in the last example for $\vartheta_{\delta,1}$, $\vartheta_{\delta,2}$ and with rate $M^{1/2}\delta$ for the reaction term ϑ_3. Using integration by parts we find

$$\Sigma_\vartheta = \frac{T}{2\vartheta_1}\begin{pmatrix} \|\nabla K\|^2_{L^2(\mathbb{R}^d)} & 0 & -1 \\ 0 & \|(-\Delta_0)^{-1/2}(\nabla\cdot b)K\|^2_{L^2(\mathbb{R}^d)} & 0 \\ -1 & 0 & \|(-\Delta_0)^{-1/2}K\|^2_{L^2(\mathbb{R}^d)} \end{pmatrix},$$

so we have pairwise asymptotic independence of diffusion and transport estimators, as well as of transport and reaction estimators. Similar numerical results as in the first example were obtained, but details are omitted.

A.5.2 *A boundary case: estimation in $d = 2$*

Theorem A.3 is not valid for $d \leq 2$ and reaction terms ϑ_i with differential order $n_i = 0$. The singularities of the heat kernel on $\mathbb{R}^d$ in $d \leq 2$ (cf. the discussion before Theorem A.3) can be avoided for sufficiently regular K, e.g., by assuming $K = \Delta\widetilde{K}$ for some $\widetilde{K} \in H^4(\mathbb{R}^d)$. In that case, the CLT still holds with the same proof, but consistency towards ϑ_i is lost, because $M^{1/2}\delta$ does not diverge. Nevertheless, we show now that in $d = 2$ for non-negative K, a logarithmic rate holds. This is consistent with results for the MLE from spectral observations in $d = 2$, cf. [27]. For a proof see Section A.7.2. For simplicity, only a simplified model is considered.

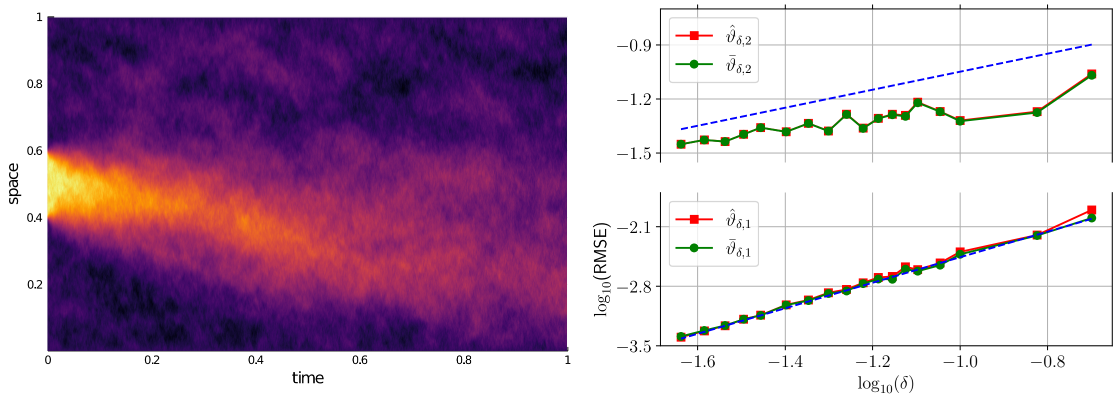


Figure A.1: (left) heat map for a typical realisation of $X(t,x)$ corresponding to (A.1) in $d=1$ with domain $\Lambda=(0,1)$ in Example A.10; (right) $\log_{10}\log_{10}$ plot of the root mean squared errors for estimating ϑ_1 and ϑ_2 in Example A.10.

PROPOSITION A.12. *Suppose that $d=2$, $A_\vartheta=\Delta+\vartheta$ for $\vartheta\in\mathbb{R}$, $X(0)=0$ and $M\delta^2\to 1$ as $\delta\to 0$. If $K\geq 0$ and $K\neq 0$, then $\widehat{\vartheta}_\delta=\vartheta+O_{\mathbb{P}}(\log(\delta^{-1})^{-1/2})$.*

A.5.3 *Practical aspects*

In this section, we outline a precise situation where local measurements arise, and how the augmented MLE can be applied, even if the additional measurements $X_{\delta,k}^{A_i}$ are not available.

Optical measurements of physical or chemical concentrations $X(t)$ at a focal point $x_k\in\Lambda$ are obtained as (normalised) counts of certain markers, e.g., photons [17]. According to classical microscopy [35], diffraction leads to a blurred image of $X(t)$, and the blur pattern can be described by convolution with a specific point spread function, which can be written as

$$X_{\delta,k}(t)=\langle X(t),K_{\delta,x_k}\rangle=(X(t)*\bar{K}_\delta)(x_k),\qquad \bar{K}_\delta(y)=\delta^{d/2}K(-\delta^{-1}y). \tag{A.15}$$

It is reasonable to assume that the additional measurement noise due to photon counting is negligible and that measuring happens on faster time scales than the dynamics of X.

The resolution δ is specific to the measurement device and determines how far focal points can be apart to distinguish them [35, Definition 2.5]. The point spread function depends inversely on δ, and is often approximated by a normal density with standard deviation δ [6]. This phenomenon is the source for the large number of statistical works on Gaussian deconvolution. In applications, both the point spread function and the resolution δ are usually known, and can even be engineered to meet desired specifications [7]. Note that multiplicative constants such as the scaling of K_{δ,x_k} cancel out in the augmented MLE and therefore play no role for parameter estimation.

If we have (time discrete) local measurements (A.15) at our disposal, then exchanging differentiation and convolution gives

$$X_{\delta,k}^{A_i}=(X(t)*A_i^*\bar{K}_\delta)(x_k)=A_i^*(X(t)*\bar{K}_\delta)(x_k).$$

This can be approximated by finite differences. For example, if $A_i=\Delta$ and $x_{k-1}=x_k-\delta e_i$, $x_{k+1}=x_k+\delta e_i$ are 'neighbours' of x_k in the i-th coordinate with the unit vector e_i, separated by a distance δ, then $X^{A_i}_{\delta,k}(t)$ can be approximated by $\delta^{-2}(X_{\delta,k+1}(t) - 2X_{\delta,k}(t) + X_{\delta,k-1}(t))$. Using suitable Riemann sum approximations for Lebesgue and stochastic integrals, we thus obtain a discretised version of the augmented MLE $\widehat{\vartheta}_\delta$.

While a full analysis of such discretisation schemes is beyond the scope of this paper, we shortly report on a recent case study for cell motility, using the augmented MLE for real and simulated data [2, Sections 5 and 6]. There, the first component of a coupled stochastic activator-inhibitor system (X_1, X_2) follows a semi-linear SPDE with diffusity ϑ, reaction function f and noise level $\sigma > 0$,

$$\mathrm{d}X_1(t) = (\vartheta\Delta X_1(t) + f(X_1(t), X_2(t)))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t).$$

The equation models the change in actine concentration along the cell cortex during cell repolarisation. In [43], on a time grid of up to 256 seconds $M = 100$ measurements for 18 different cells, expected to have about the same diffusivities, were used to fit parameters in the deterministic PDE with $\sigma = 0$. In [2], the same data were taken as local measurements from $X_1(t)$, and ϑ was estimated by the discretised augmented MLE as discussed above, providing a biologically reasonable magnitude for ϑ, which can be used to distinguish the mechanisms contributing to diffusion. The resolution δ was found as an upper bound on the spatial mesh size. The estimates are stable across the cell populations as opposed to [43], which averaged the different estimates across cells to reduce 'noise', and obtained in this way a much inflated average diffusivity.

A.6 Core Proofs

A.6.1 *Proof of the central limit theorem*

Preliminaries

We write $\Lambda_{\delta,x} = \{\delta^{-1}(y-x) : y \in \Lambda\}$, $\Lambda_{0,x} = \mathbb{R}^d$ and introduce with domains $H^1_0(\Lambda_{\delta,x}) \cap H^2(\Lambda_{\delta,x})$ the operators

$$A_{\vartheta,\delta,x} = \nabla \cdot a_\vartheta \nabla + \delta \nabla \cdot b_\vartheta + \delta^2 c_\vartheta, \quad \widetilde{A}_{\vartheta,\delta,x} = \nabla \cdot a_\vartheta \nabla. \tag{A.16}$$

They generate the analytic semigroups $(S_{\vartheta,\delta,x}(t))_{t\geq 0}$ and $(\widetilde{S}_{\vartheta,\delta,x}(t))_{t\geq 0}$ on $L^2(\Lambda_{\delta,x})$. Similarly, the adjoint operators $A^*_{\vartheta,\delta,x}$ and $\widetilde{A}^*_{\vartheta,\delta,x}$ generate with the same domains the adjoint semigroups $(S^*_{\vartheta,\delta,x}(t))_{t\geq 0}$ and $(\widetilde{S}^*_{\vartheta,\delta,x}(t))_{t\geq 0}$. When a_ϑ is the identity matrix, then we also write $\Delta_{\delta,x} = \widetilde{A}_{\vartheta,\delta,x}$ and $e^{t\Delta_{\delta,x}} = \widetilde{S}_{\vartheta,\delta,x}(t)$. Moreover, $e^{t\Delta_0}$ and $e^{t\nabla\cdot a_\vartheta \nabla}$ are semigroups on $L^2(\mathbb{R}^d)$ generated by Δ_0 and $\nabla \cdot a_\vartheta \nabla$, respectively, with domain $H^2(\mathbb{R}^d)$. We often use implicitly that $z \in L^2(\Lambda_{\delta,x})$ extends to an element of $L^2(\mathbb{R}^d)$ by setting $z(y) = 0$ outside of $\Lambda_{\delta,x}$. The A_i and their formal adjoints A^*_i are considered as differential operators on sufficiently weakly differentiable functions without boundary conditions.

The rescaled semigroup

In this section we collect some results on the semigroup operators $S_\vartheta(t)$ and their actions on localised functions $z_{\delta,x}(\cdot) = \delta^{-d/2} z(\delta^{-1}(\cdot - x))$.

By a standard-PDE result (see, e.g., [31, Example III.6.11] or [48, equation (5.1)]), the operator $A_{\vartheta,\delta,x}$ and the generated semigroup are diagonalizable [18, Example 2.1 in Section

II.2]. This yields the useful representations

$$A_{\vartheta,\delta,x} = U_{\vartheta,\delta,x}(\widetilde{A}_{\vartheta,\delta,x} + \delta^2\widetilde{c}_\vartheta)U_{\vartheta,\delta,x}^{-1}, \quad S_{\vartheta,\delta,x}(t) = e^{t\delta^2\widetilde{c}_\vartheta}U_{\vartheta,\delta,x}\widetilde{S}_{\vartheta,\delta,x}(t)U_{\vartheta,\delta,x}^{-1} \tag{A.17}$$

with the multiplication operators $U_{\vartheta,\delta,x}z(y) = \exp(-(a_\vartheta^{-1}b_\vartheta)\cdot(\delta y + x)/2)z(y)$ and with $\widetilde{c}_\vartheta = c_\vartheta - \frac{1}{4}b_\vartheta\cdot(a_\vartheta^{-1}b_\vartheta)$. Observe the following scaling properties.

LEMMA A.13. *Let $\delta' \geq \delta \geq 0$, $x \in \Lambda$, $i = 1,\ldots,p$.*

(i) If $z \in H_0^1(\Lambda_{\delta,x}) \cap H^2(\Lambda_{\delta,x})$, then $A_i^* z_{\delta,x} = \delta^{-n_i}(A_i^* z)_{\delta,x}$, $A_\vartheta^* z_{\delta,x} = \delta^{-2}(A_{\vartheta,\delta,x}^* z)_{\delta,x}$.

(ii) If $z \in L^2(\Lambda_{\delta,x})$, $t \geq 0$, then $S_\vartheta^*(t)z_{\delta,x} = (S_{\vartheta,\delta,x}^*(t\delta^{-2})z)_{\delta,x}$.

Proof. Part (i) is clear, part (ii) follows analogously to [4, Lemma 3.1]. ∎

The semigroup on the bounded domain $\Lambda_{\delta,x}$ is after zooming in as $\delta \to 0$ intuitively close to the semigroup on $\mathbb{R}^d$. The next result makes this precise, uniformly in $x \in \mathcal{J}$.

LEMMA A.14. *Under Assumption H(iii) the following holds:*

(i) There exists $C > 0$ such that if $z \in C_c(\mathbb{R}^d)$ is supported in $\bigcap_{x\in\mathcal{J}}\Lambda_{\delta,x}$ for some $\delta \geq 0$, then for all $t \geq 0$

$$\sup_{x\in\mathcal{J}}\left|(S_{\vartheta,\delta,x}^*(t)z)(y)\right| \leq Ce^{\widetilde{c}_\vartheta t\delta^2}(e^{t\nabla\cdot a_\vartheta\nabla}|z|)(y), \quad y \in \mathbb{R}^d.$$

(ii) If $z \in L^2(\mathbb{R}^d)$, then as $\delta \to 0$ for all $t > 0$

$$\sup_{x\in\mathcal{J}}\|S_{\vartheta,\delta,x}^*(t)(z|_{\Lambda_{\delta,x}}) - e^{t\nabla\cdot a_\vartheta\nabla}z\|_{L^2(\mathbb{R}^d)} \to 0.$$

Proof. (i). By (A.17) and noting that the function $y \mapsto \exp(-(a_\vartheta^{-1}b_\vartheta)\cdot(\delta y + x)/2)$ is uniformly upper and lower bounded on $\bigcap_{x\in\mathcal{J}}\Lambda_{\delta,x}$, we get

$$\sup_{x\in\mathcal{J}}\left|(S_{\vartheta,\delta,x}^*(t)z)(y)\right| \lesssim e^{t\delta^2\widetilde{c}_\vartheta}(\widetilde{S}_{\vartheta,\delta,x}(t)|z|)(y), \quad y \in \mathbb{R}^d.$$

It is therefore enough to prove the claim with respect to $\widetilde{S}_{\vartheta,\delta,x}$ and with $|z|$ instead of z. By the classical Feynman–Kac formulas (cf. [30, Chapter 4.4], the anisotropic case is an easy generalisation, which can also be obtained by a change of variables leading to a diagonal diffusivity matrix a_ϑ, which corresponds to d scalar heat equations) we have with a process $Y_t = y + a_\vartheta^{1/2}\widetilde{W}_t$ and a d-dimensional Brownian motion $\widetilde{W}$, all defined on another probability space with expectation and probability operators $\widetilde{\mathbb{E}}_y$, $\widetilde{\mathbb{P}}_y$, that $(e^{t\nabla\cdot a_\vartheta\nabla}z)(y) = \widetilde{\mathbb{E}}_y[z(Y_t)]$ and $\widetilde{S}_{\vartheta,\delta,x}(t)z(y) = \widetilde{\mathbb{E}}_y\left[z(Y_t)\mathbf{1}(t < \tau_{\delta,x})\right]$ with the stopping times $\tau_{\delta,x} := \inf\{t \geq 0 : Y_t \notin \Lambda_{\delta,x}\}$. The claim follows now from

$$\sup_{x\in\mathcal{J}}(\widetilde{S}_{\vartheta,\delta,x}(t)|z|)(y) \leq \widetilde{\mathbb{E}}_y[|z(Y_t)|] = (e^{t\nabla\cdot a_\vartheta\nabla}|z|)(y).$$

(ii). By an approximation argument it is enough to consider $z \in C_c(\bar{\Lambda})$ and $0 < \delta \leq \delta'$ such that z is supported in $\Lambda_{\delta',x}$, hence $z|_{\Lambda_{\delta,x}} = z$ for all such δ. Compactness of $\mathcal{J}$ according to

Assumption H(iii) guarantees for sufficiently small δ the existence of a ball with centre 0 and radius $\rho\delta^{-1}$ for some $\rho > 0$, contained in $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$. With this and the representation formulas in (i), combined with the Cauchy–Schwarz inequality, we have for all $y \in \mathbb{R}^d$

$$\begin{aligned}
&\sup_{x\in\mathcal{J}} |(\widetilde{S}_{\vartheta,\delta,x}(t)z)(y) - (e^{t\nabla\cdot a_\vartheta\nabla}z)(y)|^2 = \sup_{x\in\mathcal{J}} |\widetilde{\mathbb{E}}_y\big[z(Y_t)\mathbf{1}(\tau_{\delta,x} \le t)\big]|^2 \\
&\le \sup_{x\in\mathcal{J}} \widetilde{\mathbb{E}}_y[z^2(Y_t)]\widetilde{\mathbb{P}}_y(\tau_{\delta,x} \le t) \le (e^{t\nabla\cdot a_\vartheta\nabla}z^2)(y)\widetilde{\mathbb{P}}_y(\max_{0\le s\le t} |Y_s| \ge \rho\delta^{-1}) \\
&\le (e^{t\nabla\cdot a_\vartheta\nabla}z^2)(y)\widetilde{\mathbb{P}}_y(\max_{0\le s\le t} |\widetilde{W}_s| \ge \widetilde{\rho}\delta^{-1}) \lesssim (e^{t\nabla\cdot a_\vartheta\nabla}z^2)(y)(\delta t^{1/2}e^{-\delta^{-2}t^{-1}}) \to 0
\end{aligned}$$

as $\delta \to 0$ for another constant $\widetilde{\rho}$, concluding by [30, equation (2.8.4)]. Since $\|e^{t\nabla\cdot a_\vartheta\nabla}z^2\|_{L^1(\mathbb{R}^d)} \le \|z\|^2_{L^2(\mathbb{R}^d)}$, dominated convergence proves the claim when $b_\vartheta = 0$, $c_\vartheta = 0$. The general case is then an easy consequence of the last display and (A.17). ∎

We require frequently quantitative statements on the decay of the action of the semigroup operators $S^*_{\vartheta,\delta,x}(t)$ as $t \to \infty$ when applied to functions of a certain smoothness and integrability. This is well-known for an analytic semigroup, but is shown here to hold true for all δ and uniformly in $x \in \mathcal{J}$.

Lemma A.15. *Let $0 \le \delta \le 1$, $t > 0$, $x \in \mathcal{J}$ and $1 < p \le \infty$. Moreover, let $z \in L^p(\Lambda_{\delta,x})$ if $1 < p < \infty$ and $z \in C(\Lambda_{\delta,x})$ with $z = 0$ on $\partial\Lambda_{\delta,x}$ if $p = \infty$. Then it holds with implied constants not depending on x:*

$$\|A^*_{\vartheta,\delta,x}S^*_{\vartheta,\delta,x}(t)z\|_{L^p(\Lambda_{\delta,x})} \lesssim t^{-1}\|z\|_{L^p(\Lambda_{\delta,x})}.$$

Proof. Apply first the scaling in Lemma A.13 in reverse order such that with $1 < p \le \infty$

$$\|A^*_{\vartheta,\delta,x}S^*_{\vartheta,\delta,x}(t)z\|_{L^p(\Lambda_{\delta,x})} = \delta^{d(1/2-1/p)+2}\|A^*_\vartheta S^*_\vartheta(t\delta^2)z_{\delta,x}\|_{L^p(\Lambda)}.$$

If $p < \infty$, by the semigroup property for analytic semigroups in [5, Theorem V.2.1.3], the $L^p(\Lambda)$-norm is up to a constant upper bounded by $(t\delta^2)^{-1}\|z_{\delta,x}\|_{L^p(\Lambda)}$, and the claim follows. The same proof applies to $p = \infty$, noting that A^*_ϑ generates an analytic semigroup on $\{u \in C(\Lambda), u = 0$ on $\partial\Lambda\}$, cf. [47, Theorem 7.3.7]. ∎

The proof for the next result relies on the Bessel-potential spaces $H^{s,p}_0(\Lambda_{\delta,x})$, $1 < p < \infty$, $s \in \mathbb{R}$, defined for $\delta > 0$ as the domains of the fractional Dirichlet–Laplacian $(-\Delta_{\delta,x})^{s/2}$ on $\Lambda_{\delta,x}$ with norms $\|\cdot\|_{H^{s,p}(\Lambda_{\delta,x})} = \|(-\Delta_{\delta,x})^{s/2}\cdot\|_{L^p(\Lambda_{\delta,x})}$, see [16] for details and also Section A.6.2 below. Since a_ϑ is positive definite, the norms $\|\cdot\|_{H^{s,p}(\Lambda_{\delta,x})}$ are equivalently generated by the fractional powers of $-\widetilde{A}_{\vartheta,\delta,x}$ [62, Theorem 16.15].

Lemma A.16. *Let $0 < \delta \le 1$, $t > 0$, $1 < p \le 2$ and grant Assumption H(iii). Let $z \in H^s_0(\mathbb{R}^d)$, $s \ge 0$, be compactly supported in $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$, suppose that $V_{\delta,x} : L^p(\Lambda_{\delta,x}) \to H^{-s,p}_0(\Lambda_{\delta,x})$ are bounded linear operators with $\|V_{\delta,x}z\|_{H^{-s,p}(\Lambda_{\delta,x})} \le V_{\mathrm{op}}\|z\|_{L^p(\Lambda_{\delta,x})}$ for some V_{op} independent of δ, x. Then for $1 < p \le 2$ and $\gamma = (1/p - 1/2)d/2$ there exists a constant $C > 0$, depending on p and s such that*

$$\sup_{x\in\mathcal{J}}\|S^*_{\vartheta,\delta,x}(t)V_{\delta,x}z\|_{L^2(\Lambda_{\delta,x})} \le Ce^{\widetilde{c}_\vartheta t\delta^2}\sup_{x\in\mathcal{J}}\Big(\|V_{\delta,x}z\|_{L^2(\Lambda_{\delta,x})} \wedge (V_{\mathrm{op}}t^{-s/2-\gamma}\|z\|_{L^p(\Lambda_{\delta,x})})\Big).$$

If $s = 0$, then this holds also for $p = 1$.

Proof. Set $u = V_{\delta,x}z$, $v = U_{\vartheta,\delta,x}u$. The $U_{\vartheta,\delta,x}$ are bounded operators on $L^2(\Lambda_{\delta,x})$ uniformly in $\delta \geq 0$ and $x \in \mathcal{J}$ and thus by (A.17)

$$\|S^*_{\vartheta,\delta,x}(t)u\|_{L^2(\Lambda_{\delta,x})} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}\|\widetilde{S}_{\vartheta,\delta,x}(t)v\|_{L^2(\Lambda_{\delta,x})}. \tag{A.18}$$

Let first $s = 0$ such that $H_0^{-s,p}(\mathbb{R}^d) = L^p(\mathbb{R}^d)$. Ellipticity and symmetry of a_ϑ show

$$\|e^{t\nabla\cdot a_\vartheta\nabla}|v|\|_{L^2(\mathbb{R}^d)} \leq \|e^{tC'\Delta_0}|v|\|_{L^2(\mathbb{R}^d)}$$

for a constant $C' > 0$ (use either [51] or argue that the semigroup $e^{t\nabla\cdot a_\vartheta\nabla}$ on acts on $L^2(\mathbb{R}^d)$ as a multiplication operator in the Fourier domain according to [18, Section VI.5], which can be upper bounded by the identity operator). Approximating u by continuous and compactly supported functions, we thus find from Lemma A.14(i) and hypercontractivity of the heat kernel on $\mathbb{R}^d$ uniformly in $x \in \mathcal{J}$

$$\|S^*_{\vartheta,\delta,x}(t)u\|_{L^2(\Lambda_{\delta,x})} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}\|e^{C't\Delta_0}|v|\|_{L^2(\mathbb{R}^d)} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}t^{-\gamma}\|u\|_{L^p(\mathbb{R}^d)} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}t^{-\gamma}\|z\|_{L^p(\mathbb{R}^d)}.$$

This yields the result for $s = 0$. These inequalities hold also for $p = 1$, thus proving the supplement of the statement. For $s > 0$ and $p > 1$ note first that by [3, Proposition 17(i)] we have $\|(-t\widetilde{A}_{\vartheta,\delta,x})^{s/2}\widetilde{S}_{\vartheta,\delta,x}(t)z\|_{L^2(\Lambda_{\delta,x})} \lesssim \|z\|_{L^2(\Lambda_{\delta,x})}$. Inserting this and then the last display with u replaced by $(-\widetilde{A}_{\vartheta,\delta,x})^{-s/2}v$ into (A.18) we get

$$\begin{aligned}\|S^*_{\vartheta,\delta,x}(t)u\|_{L^2(\Lambda_{\delta,x})} &\lesssim e^{\widetilde{c}_\vartheta t\delta^2}\|(-\widetilde{A}_{\vartheta,\delta,x})^{s/2}\widetilde{S}_{\vartheta,\delta,x}(t)(-\widetilde{A}_{\vartheta,\delta,x})^{-s/2}v\|_{L^2(\Lambda_{\delta,x})}\\ &\lesssim e^{\widetilde{c}_\vartheta t\delta^2}t^{-s/2}\|\widetilde{S}_{\vartheta,\delta,x}(t/2)(-\widetilde{A}_{\vartheta,\delta,x})^{-s/2}v\|_{L^2(\Lambda_{\delta,x})} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}t^{-s/2-\gamma}\|v\|_{H^{-s,p}(\Lambda_{\delta,x})},\end{aligned}$$

uniformly in $x \in \mathcal{J}$. Note that the $U_{\vartheta,\delta,x}$ also induce a family of multiplication operators on $H_0^{s,p}(\mathbb{R}^d)$ for $s \geq 0$ with operator norms uniformly bounded in $x \in \mathcal{J}$, cf. [55, Theorem 2.8.2]. By duality and restriction this transfers to $H_0^{s,p}(\Lambda_{\delta,x})$ for general s according to [55, Theorem 3.3.2]. Hence,

$$\|S^*_{\vartheta,\delta,x}(t)u\|_{L^2(\Lambda_{\delta,x})} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}t^{-s/2-\gamma}\|u\|_{H^{-s,p}(\Lambda_{\delta,x})} \lesssim e^{\widetilde{c}_\vartheta t\delta^2}t^{-s/2-\gamma}V_{\mathrm{op}}\|z\|_{L^p(\Lambda_{\delta,x})}. \qquad \blacksquare$$

Covariance structure of multiple local measurements

LEMMA A.17.

(i) *If $X_0 = 0$, then the Gaussian process from* (A.5) *has mean zero and covariance function*

$$\mathrm{Cov}(\langle X(t),z\rangle,\langle X(t'),z'\rangle) = \int_0^{t\wedge t'}\langle S^*_\vartheta(t-s)z, S^*_\vartheta(t'-s)z'\rangle\,\mathrm{d}s.$$

(ii) *If X_0 is the stationary initial condition from Lemma A.2, then the Gaussian process from* (A.5) *has mean zero and covariance function*

$$\mathrm{Cov}(\langle X(t),z\rangle,\langle X(t'),z'\rangle) = \int_0^{\infty}\langle S^*_\vartheta((t-t')+s)z, S^*_\vartheta(s)z'\rangle\,\mathrm{d}s, \quad t \geq t'.$$

Proof. Part (i) follows from (A.5) and Itô's isometry [15, Proposition 4.28]. For part (ii) we conclude in the same way from noting that the stationary solution given by

$$\langle X(t), z\rangle = \int_{-\infty}^{t} \langle S_\vartheta^*(t-s)z, \mathrm{d}W(s)\rangle$$

has mean zero. ∎

Introduce for $i, j = 1, \ldots, p$

$$\Psi_\vartheta(A_i^*K, A_j^*K) = \frac{1}{2}\langle(-\nabla\cdot a_\vartheta\nabla)^{-1/2}A_i^*K, (-\nabla\cdot a_\vartheta\nabla)^{-1/2}A_j^*K\rangle_{L^2(\mathbb{R}^d)},$$

which is well-defined under Assumption H. by the discussion before Theorem A.3.

LEMMA A.18. *Grant Assumption H and let $X_0 = 0$. We have as $\delta \to 0$*

$$\delta^{-2+n_i+n_j}(MT)^{-1}\sum_{k=1}^{M}\int_0^T \mathbb{E}\Big[\langle X(t), A_i^*K_{\delta,x_k}\rangle\langle X(t), A_j^*K_{\delta,x_k}\rangle\Big]\,\mathrm{d}t \to \Psi_\vartheta(A_i^*K, A_j^*K).$$

Proof. Fix i, j with $n_i + n_j > 2 - d$. Then, applying Lemma A.17(i), the scaling from Lemma A.13 and changing variables give

$$\delta^{-2+n_i+n_j}(MT)^{-1}\sum_{k=1}^{M}\int_0^T \mathbb{E}\Big[\langle X(t), A_i^*K_{\delta,x_k}\rangle\langle X(t), A_j^*K_{\delta,x_k}\rangle\Big]\,\mathrm{d}t = \int_0^\infty f_\delta(t')\,\mathrm{d}t'$$

with

$$f_\delta(t') = (MT)^{-1}\sum_{k=1}^{M}\langle S_{\vartheta,\delta,x_k}^*(t')A_i^*K, S_{\vartheta,\delta,x_k}^*(t')A_j^*K\rangle_{L^2(\Lambda_{\delta,x_k})}\int_0^T \mathbf{1}_{\{0\le t'\le t\delta^{-2}\}}\,\mathrm{d}t.$$

Consider now the differential operators $V_{\delta,x_k} = A_i^*$. If D^m is a composition of m partial differential operators, then Theorem 1.43 of [62] yields that D^m is a bounded linear operator from $L^p(\Lambda)$ to $H_0^{-m,p}(\Lambda)$, implying $\|D^m K_{\delta,x_k}\|_{H^{-m,p}(\Lambda)} \lesssim \delta^{-m}\|K_{\delta,x_k}\|_{L^p(\Lambda)}$. Since $(D^m K)_{\delta,x_k} = \delta^m D^m K_{\delta,x_k}$, changing variables gives $\|D^m K\|_{H^{-m,p}(\Lambda_{\delta,x_k})} \lesssim \|K\|_{L^p(\Lambda_{\delta,x_k})}$. From this we find $\|V_{\delta,x_k}K\|_{H^{-n_i,p}(\Lambda_{\delta,x})} \le \|K\|_{L^p(\Lambda_{\delta,x_k})}$, $\|V_{\delta,x_k}K\|_{L^2(\Lambda_{\delta,x_k})} \lesssim \|K\|_{H^{n_i}(\mathbb{R}^d)}$, and Lemma A.16 shows for $0 \le t' \le T\delta^{-2}$, $\varepsilon > 0$ and all sufficiently small $\delta > 0$

$$\sup_{x\in\mathcal{J}}\|S_{\vartheta,\delta,x}^*(t')A_i^*K\|_{L^2(\Lambda_{\delta,x})} \lesssim 1 \wedge (t')^{-n_i/2-d/4+\varepsilon}. \tag{A.19}$$

By the Cauchy–Schwarz inequality we get $|f_\delta(t')| \lesssim 1 \wedge (t')^{-n_i/2-n_j/2-d/2+2\varepsilon}$. In particular, taking ϵ so small that $n_i + n_j > 2 - d - 4\epsilon$ yields $\sup_{\delta>0}|f_\delta| \in L^1([0,\infty))$. Lemma A.14(ii), Lemma A.13(ii) and continuity of the L^2-scalar product show now pointwise for all $t' > 0$ that $f_\delta(t') \to \langle e^{2t'\nabla\cdot a_\vartheta\nabla}A_i^*K, A_j^*K\rangle_{L^2(\mathbb{R}^d)}$. Conclude by the dominated convergence theorem and $\int_0^\infty \langle e^{2t'\nabla\cdot a_\vartheta\nabla}A_i^*K, A_j^*K\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}t' = \Psi_\vartheta(A_i^*K, A_j^*K)$. ∎

LEMMA A.19. *Grant Assumption H and let $X_0 = 0$. If $n_i + n_j > 2 - d$ for $i, j = 1, \ldots, p$, then* $\sup_{x\in\mathcal{J}}\operatorname{Var}(\int_0^T \langle X(t), A_i^*K_{\delta,x}\rangle\langle X(t), A_j^*K_{\delta,x}\rangle\,\mathrm{d}t) = o(\delta^{4-2n_i-2n_j})$.

Proof. Applying the scaling from Lemma A.13 and using Wicks theorem [28, Theorem 1.28] we have for $x \in \mathcal{J}$

$$\delta^{2n_i+2n_j} \operatorname{Var}\left(\int_0^T \langle X(t), A_i^* K_{\delta,x}\rangle \langle X(t), A_j^* K_{\delta,x}\rangle \,\mathrm{d}t\right)$$
$$= \operatorname{Var}\left(\int_0^T \langle X(t), (A_i^* K)_{\delta,x}\rangle \langle X(t), (A_j^* K)_{\delta,x}\rangle \,\mathrm{d}t\right) = V_1 + V_2$$

with $V_1 = V_{\delta,x}(A_i^*K, A_i^*K, A_j^*K, A_j^*K)$, $V_2 = V_{\delta,x}(A_i^*K, A_j^*K, A_j^*K, A_i^*K)$, and where for $v, v', z, z' \in L^2(\Lambda_{\delta,x})$

$$V_{\delta,x}(v, v', z, z') = \int_0^T \int_0^T \mathbb{E}[\langle X(t), v_{\delta,x}\rangle \langle X(t'), v'_{\delta,x}\rangle]\mathbb{E}[\langle X(t), z_{\delta,x}\rangle \langle X(t'), z'_{\delta,x}\rangle] \,\mathrm{d}t' \,\mathrm{d}t.$$

We only upper bound V_1, the arguments for V_2 are similar. Set

$$f_{i,j}(s, s') = \langle S^*_{\vartheta,\delta,x}(s + s')A_i^* K, S^*_{\vartheta,\delta,x}(s')A_j^* K\rangle_{L^2(\Lambda_{\delta,x})}.$$

Using Lemma A.17(i) and the scaling in Lemma A.13 we have

$$V_1 = 2\delta^6 \int_0^T \int_0^{t\delta^{-2}} \left(\int_0^{t\delta^{-2}-s} f_{i,i}(s, s') \,\mathrm{d}s'\right)\left(\int_0^{t\delta^{-2}-s} f_{j,j}(s, s'') \,\mathrm{d}s''\right) \mathrm{d}s \,\mathrm{d}t,$$

cf. [4, Proof of Proposition A.9]. From (A.19) and the Cauchy–Schwarz inequality we infer

$$\sup_{x\in\mathcal{J}} |f_{i,i}(s, s')f_{j,j}(s, s'')| \lesssim (1 \wedge s^{-(n_i+n_j)/2-d/2+2\varepsilon})(1 \wedge s'^{-n_i/2-d/4+\varepsilon})(1 \wedge s''^{-n_j/2-d/4+\varepsilon})$$

for $\varepsilon > 0$, which gives

$$\begin{aligned}\sup_{x\in\mathcal{J}}|V_1| &\lesssim \delta^6 \int_0^{T\delta^{-2}} (1 \wedge s^{-n_i/2-n_j/2-d/2+2\varepsilon}) \,\mathrm{d}s \int_0^{T\delta^{-2}} (1 \wedge s'^{-n_i/2-d/4+\varepsilon}) \,\mathrm{d}s' \\ &\quad \cdot \int_0^{T\delta^{-2}} (1 \wedge s''^{-n_j/2-d/4+\varepsilon}) \,\mathrm{d}s'' \\ &\lesssim \delta^6(1 \vee \delta^{n_i+n_j+d-2-4\varepsilon})(1 \vee \delta^{n_i+d/2-2-2\varepsilon})(1 \vee \delta^{n_j+d/2-2-2\varepsilon}).\end{aligned}$$

Without loss of generality let $n_i \le n_j$. For ε small enough, we can ensure $\delta^{n_i+n_j+d-2-4\varepsilon} \le 1$, as $n_i + n_j > 2 - d$. In $d \le 2$ only the pairs $(n_i, n_j) \in \{(0, 0), (0, 1)\}$ are excluded, and in every case the claimed bound holds. The same applies to $d \ge 3$ for all pairs (n_i, n_j). ■

Proof of Theorem A.3

Proof. We begin with the observed Fisher information. Suppose first $X_0 = 0$. Under Assumption H we find that $n_i + n_j > 2 - d$ for all $i, j = 1, \dots, p$ in all dimensions $d \ge 1$. It follows from Lemmas A.18 and A.19 that

$$(\rho_\delta \mathcal{I}_\delta \rho_\delta)_{ij} = \delta^{-2+n_i+n_j} M^{-1} \sum_{k=1}^{M} \int_0^T \langle X(t), A_i^* K_{\delta,x_k}\rangle \langle X(t), A_j^* K_{\delta,x_k}\rangle \,\mathrm{d}t$$

$$= T\Psi_\vartheta(A_i^*K, A_j^*K) + o_{\mathbb{P}}(1) = (\Sigma_\vartheta)_{ij} + o_{\mathbb{P}}(1).$$

This yields for $X_0 = 0$ the wanted convergence $\rho_\delta \mathcal{I}_\delta \rho_\delta \xrightarrow{\mathbb{P}} \Sigma_\vartheta$. In order to extend this to the general X_0 from Assumption H, let $\bar{X}$ be defined as X, but starting in $\bar{X}(0) = 0$ such that for $v \in L^2(\Lambda)$, $\langle X(t), v\rangle = \langle \bar{X}(t), v\rangle + \langle S_\vartheta(t)X_0, v\rangle$. If $\bar{\mathcal{I}}_\delta$ is the observed Fisher information corresponding to $\bar{X}$, then by the Cauchy–Schwarz inequality, with $v_i = \sup_k \delta^{2n_i-2}\int_0^T \langle X_0, S_\vartheta^*(t)A_i^*K_{\delta,x_k}\rangle^2\,\mathrm{d}t$,

$$|(\rho_\delta \mathcal{I}_\delta \rho_\delta)_{ij} - (\rho_\delta \bar{\mathcal{I}}_\delta \rho_\delta)_{ij}| \lesssim (\rho_\delta \bar{\mathcal{I}}_\delta \rho_\delta)_{ii}^{1/2} v_j^{1/2} + (\rho_\delta \bar{\mathcal{I}}_\delta \rho_\delta)_{jj}^{1/2} v_i^{1/2} + v_i^{1/2} v_j^{1/2}.$$

By the first part, $(\rho_\delta \bar{\mathcal{I}}_\delta \rho_\delta)_{ii}$ is bounded in probability and Assumption H(iv) gives $v_i = o_{\mathbb{P}}(1)$ for all i. From this obtain again the convergence of the observed Fisher information. Regarding the invertibility of Σ_ϑ, let $\lambda \in \mathbb{R}^p$ such that

$$0 = \sum_{i,j=1}^p \lambda_i \lambda_j (\Sigma_\vartheta)_{ij} = T\Psi_\vartheta\Big(\sum_{i=1}^p \lambda_i A_i^* K, \sum_{i=1}^p \lambda_i A_i^* K\Big).$$

By the definition of Ψ_ϑ this implies $e^{t\nabla\cdot a_\vartheta \nabla}(\sum_{i=1}^p \lambda_i A_i^* K) = 0$ for all $t \geq 0$ and thus $\sum_{i=1}^p \lambda_i A_i^* K = 0$. Since the functions $A_i^* K$ are linearly independent by Assumption H(i), conclude that Σ_ϑ is invertible.

We proceed next to the proof of the CLT. The augmented MLE and the statement of the limit theorem remain unchanged when K is multiplied by a scalar factor. We can therefore assume without loss of generality that $\|K\|_{L^2(\mathbb{R}^d)} = 1$. By the basic error decomposition (A.11) and because Σ_ϑ is invertible, this means

$$(\rho_\delta \mathcal{I}_\delta \rho_\delta)^{1/2} \rho_\delta^{-1}(\widehat{\vartheta}_\delta - \vartheta) = (\rho_\delta \mathcal{I}_\delta \rho_\delta)^{-1/2} \Sigma_\vartheta^{1/2} (\Sigma_\vartheta^{-1/2} \rho_\delta \mathcal{M}_\delta). \tag{A.20}$$

Note that $\mathcal{M}_\delta = \mathcal{M}_\delta(T)$ corresponds to a p-dimensional continuous and square integrable martingale $(\mathcal{M}_\delta(t))_{0\leq t\leq T}$ with respect to the filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ evaluated at $t = T$. In view of Assumption H(iii) let $\delta \leq \delta'$ such that for $s, t \geq 0$ and k, k' with the Kronecker delta $\delta_{k,k'}$

$$\mathbb{E}[W_k(s)W_{k'}(t)] = (s \wedge t)\langle K_{\delta,x_k}, K_{\delta,x_{k'}}\rangle = (s \wedge t)\delta_{k,k'}.$$

This means that the Brownian motions W_k and $W_{k'}$ are independent for $k \neq k'$ and thus their quadratic co-variation process at t is $[W_k, W_{k'}]_t = t\delta_{k,k'}$. From this infer that the quadratic co-variation process of the martingale $(\mathcal{M}_\delta(t))_{0\leq t\leq T}$ at $t = T$ for $\delta \leq \delta'$ is equal to

$$[\mathcal{M}_\delta]_T = \sum_{k,k'=1}^M \int_0^T X_{\delta,k}^A(t) X_{\delta,k'}^A(t)^\top d[W_k, W_{k'}]_t = \mathcal{I}_\delta.$$

Theorem A.25 now implies $\Sigma_\vartheta^{-1/2}\rho_\delta \mathcal{M}_\delta \xrightarrow{d} \mathcal{N}(0, I_{p\times p})$. Conclude in (A.20) by $\rho_\delta \mathcal{I}_\delta \rho_\delta \xrightarrow{\mathbb{P}} \Sigma_\vartheta$ and Slutsky's lemma. ∎

A.6.2 *RKHS computations*

The proofs of the RKHS results from Section A.3 are achieved by basic operations on RKHS, in particular under linear transformation (see, e.g., [38, Chapter 4] or [59, Chapter 12]).

Recall the stationary process X in (A.14) and that $Ae_j = -\lambda_j e_j$ with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ and an orthonormal basis $(e_j)_{j\geq 1}$ of $\mathcal{H}$. The cylindrical Wiener process can be realised as $W = \sum_{j\geq 1} e_j \beta_j$ for independent scalar Brownian motions β_j and we obtain

$$X(t) = \sum_{j\geq 1} \int_{-\infty}^{t} e^{-\lambda_j(t-t')} \,\mathrm{d}\beta_j(t') e_j = \sum_{j\geq 1} Y_j(t) e_j, \tag{A.21}$$

with independent stationary Ornstein–Uhlenbeck processes Y_j satisfying

$$\mathrm{d}Y_j(t) = -\lambda_j Y_j(t)\,\mathrm{d}t + \mathrm{d}\beta_j(t).$$

For a sequence (μ_j) of non-decreasing, positive real numbers, take $\mathcal{H}_1$ to be the closure of $\mathcal{H}$ under the norm

$$\|z\|_{\mathcal{H}_1}^2 = \sum_{j\geq 1} \frac{1}{\mu_j^2} \langle z, e_j \rangle_{\mathcal{H}}^2,$$

such that $\mathcal{H}$ is continuously embedded in $\mathcal{H}_1$. If for $0 \leq t \leq T$

$$\int_{-\infty}^{t} \|S(t')\|_{\mathrm{HS}(\mathcal{H},\mathcal{H}_1)}^2 \,\mathrm{d}t' = \sum_{j\geq 1} \int_{-\infty}^{t} \|S(t')e_j\|_{\mathcal{H}_1}^2 \,\mathrm{d}t' = \sum_{j\geq 1} \frac{1}{\mu_j^2} \int_{-\infty}^{t} e^{-2\lambda_j t'} \,\mathrm{d}t' < \infty, \tag{A.22}$$

then we conclude by [15, Theorem 5.2] that the law of X induces a Gaussian measure on the Hilbert space $L^2([0,T];\mathcal{H}_1)$. A first universal choice is given by $\mu_j = j$ for all $j \geq 1$. Moreover, if A is a second order elliptic differential operator, then Weyl's law [52, Lemma 2.3] says that the λ_j are positive real numbers of the order $j^{2/d}$, meaning that the choice $\mu_j = \lambda_j^{s/2}$ is possible whenever $s \geq 0$ and $s + 1 > d/2$. In this case, $\mathcal{H}_1$ corresponds to a Sobolev space of negative order $-s$ induced by the eigensequence $(\lambda_j, e_j)_{j\geq 1}$.

Let us introduce some background on the RKHS of a centred Gaussian random variable Z, defined on a separable Hilbert space $\mathcal{Z}$. Its covariance operator C_Z is necessarily positive self-adjoint and trace-class. This means, by the spectral theorem, there exist strictly positive eigenvalues $(\sigma_j^2)_{j\geq 1}$ and an associated orthonormal system of eigenvectors $(u_j)_{j\geq 1}$ such that $C_Z = \sum_{j\geq 1} \sigma_j^2 (u_j \otimes u_j)$. Associate with Z (or rather with the induced centred Gaussian measure) the so-called kernel or RKHS $(H_Z, \|\cdot\|_Z)$, where

$$H_Z = \{h \in \mathcal{Z} : \|h\|_Z < \infty\}, \qquad \|h\|_Z^2 = \sum_{j\geq 1} \frac{\langle u_j, h \rangle_{\mathcal{Z}}^2}{\sigma_j^2} \tag{A.23}$$

(see, e.g., [38, Example 4.2] and also [38, Chapters 4.1 and 4.3] and [22, Chapter 3.6] for other characterizations of the RKHS of a Gaussian measure or process). Alternatively, we have $H_Z = C_Z^{1/2}\mathcal{Z}$ and $\|h\|_Z = \|C_Z^{-1/2}h\|_{\mathcal{Z}}$ for $h \in H_Z$. A useful tool to compute the RKHS is the fact that the RKHS behaves well under linear transformation. More precisely, if $L : \mathcal{Z} \to \mathcal{Z}'$ is a bounded linear operator between Hilbert spaces, then the image $L(Z)$ is a centred Gaussian random variable with RKHS $L(H_Z)$ and norm $\|h\|_{L(Z)} = \inf\{\|f\|_Z : f \in L^{-1}h\}$ (see Proposition 4.1 in [38] and also Chapter 3.6 in [22]).

RKHS of an Ornstein–Uhlenbeck process

We start by computing the RKHS $(H_{Y_j}, \|\cdot\|_{Y_j})$ of the processes Y_j. We show that the RKHS is equal to the set H from Theorem A.5, and therefore independent of j, while the corresponding norm depends on λ_j.

LEMMA A.20. *For every $j \geq 1$ we have $H_{Y_j} = H$ and*

$$\|h\|_{Y_j}^2 = \lambda_j^2 \|h\|_{L^2([0,T])}^2 + \lambda_j (h^2(T) + h^2(0)) + \|h'\|_{L^2([0,T])}^2. \tag{A.24}$$

Proof. By Example 4.4 in [38], a scalar Brownian motion $(\beta(t))_{0\leq t\leq T}$ starting in zero has RKHS $H_\beta = \{h : h(0) = 0, h \text{ absolutely continuous}, h, h' \in L^2([0,T])\}$ with norm

$$\|h\|_\beta^2 = \int_0^T (h'(t))^2 \,\mathrm{d}t.$$

Moreover, the Brownian motion $(\bar\beta(t))_{0\leq t\leq T}$ with $\bar\beta(t) = X_0 + \beta(t)$, where X_0 is a standard Gaussian random variable independent of $(\beta(t))_{0\leq t\leq T}$ has RKHS

$$H_{\bar\beta} = \{\alpha + h : \alpha \in \mathbb{R}, h \in H_\beta\} = H, \quad \|h\|_{\bar\beta}^2 = \int_0^T (h'(t))^2 \,\mathrm{d}t + h^2(0),$$

as can be seen from Proposition 4.1 in [38] or Example 12.28 in [59]. To compute the RKHS of Y_j we now proceed similarly as in Example 4.8 in [38]. Define the bounded linear map $L : L^2([0, e^{2\lambda_j T} - 1]) \to L^2([0,T])$, $(Lf)(t) = (2\lambda_j)^{-1/2} e^{-\lambda_j t} f(e^{2\lambda_j t} - 1)$. Then we have $L\bar\beta = Y_j$ in distribution and L is bijective with inverse $L^{-1}h(s) = \sqrt{2\lambda_j(s+1)}\,h((2\lambda_j)^{-1}\log(s+1))$ for $0 \leq s \leq e^{2\lambda_j T} - 1$. By Proposition 4.1 in [38] (see also the discussion after (A.23)), we conclude that $H_{Y_j} = L(H_{\bar\beta}) = L(H) = H$ with

$$\begin{aligned}
\|h\|_{Y_j}^2 = \|L^{-1}h\|_{\bar\beta}^2 &= \int_0^{e^{2\lambda_j T}-1} \left(\frac{\mathrm{d}}{\mathrm{d}s}\sqrt{2\lambda_j(s+1)}\,h\Big(\frac{1}{2\lambda_j}\log(s+1)\Big)\right)^2 \mathrm{d}s + 2\lambda_j h^2(0) \\
&= \int_0^T (\lambda_j h(t) + h'(t))^2 \,\mathrm{d}t + 2\lambda_j h^2(0) \\
&= \lambda_j^2 \int_0^T h^2(t)\,\mathrm{d}t + \lambda_j (h^2(T) + h^2(0)) + \int_0^T (h'(t))^2 \,\mathrm{d}t.
\end{aligned}$$

■

RKHS of the SPDE

We compute next the RKHS of the process X. Let us start with the following series representation, which is independent of $\mathcal{H}_1$.

LEMMA A.21. *The RKHS $(H_X, \|\cdot\|_X)$ of the process X in (A.14) satisfies*

$$H_X = \Big\{h = \sum_{j\geq 1} h_j e_j : h_j \in H, \|h\|_X < \infty\Big\} \quad \textit{and} \quad \|h\|_X^2 = \sum_{j\geq 1} \|h_j\|_{Y_j}^2.$$

Note that $h \in L^2([0,T];\mathcal{H})$ if and only if $h = \sum_{j\geq 1} h_j e_j$ with $h_j \in L^2([0,T])$ and where $\sum_{j\geq 1} \|h_j\|_{L^2([0,T])}^2 < \infty$. In this case we have $h_j = \langle h, e_j \rangle$ for all $j \geq 1$. Moreover, since the λ_j are bounded from below by a positive constant, we conclude that H_X is indeed a subspace of $L^2([0,T];\mathcal{H})$.

Proof of Lemma A.21. Choose $\mu_j = j$ for all $j \geq 1$. Then $X = \sum_{j\geq 1} j^{-1} Y_j \widetilde{e}_j$ with orthonormal basis $\widetilde{e}_j = j e_j$ of $\mathcal{H}_1$ and the covariance operator C_X of X is given by

$$C_X : L^2([0,T];\mathcal{H}_1) \to L^2([0,T];\mathcal{H}_1), \quad \sum_{j\geq 1} f_j \widetilde{e}_j \mapsto \sum_{j\geq 1} j^{-2} (C_{Y_j} f_j) \widetilde{e}_j.$$

with $C_{Y_j} : L^2([0,T]) \to L^2([0,T])$ being the covariance operator of Y_j. Hence, using the definition of the RKHS given after (A.23), the RKHS of X consists of all elements of the form

$$h = C_X^{1/2} f = \sum_{j\geq 1} j^{-1} (C_{Y_j}^{1/2} f_j) \widetilde{e}_j = \sum_{j\geq 1} (C_{Y_j}^{1/2} f_j) e_j$$

with $f = \sum_{j\geq 1} f_j \widetilde{e}_j \in L^2([0,T];\mathcal{H}_1)$ and we have

$$\|h\|_X^2 = \|C_X^{-1/2} h\|^2_{L^2([0,T];\mathcal{H}_1)} = \|f\|^2_{L^2([0,T];\mathcal{H}_1)} = \int_0^T \|f(t)\|^2_{\mathcal{H}_1}\, dt = \sum_{j\geq 1} \|f_j\|^2_{L^2([0,T])} < \infty.$$

Using Lemma A.20, we can write $h = \sum_{j\geq 1} h_j e_j$ with $h_j = C_{Y_j}^{1/2} f_j \in H$ and

$$\|h_j\|^2_{Y_j} = \|C_{Y_j}^{-1/2} h_j\|^2_{L^2([0,T])} = \|f_j\|^2_{L^2([0,T])}.$$

Inserting this above, the claim follows. ■

Proof of Theorem A.4. We first show

$$H_X = \Big\{ h = \sum_{j\geq 1} h_j e_j : h_j \in H, \sum_{j\geq 1} (\lambda_j^2 \|h_j\|^2_{L^2([0,T])} + \|h_j'\|^2_{L^2([0,T])}) < \infty \Big\}, \tag{A.25}$$

meaning that the middle term in the squared RKHS norm $\|\cdot\|^2_{Y_j}$ can be dropped. By the calculus rules for Sobolev functions (cf. [20, Theorem 4.4]), we have

$$\forall s, u \in [0,T] : \quad h_j^2(s) - h_j^2(u) = 2 \int_u^s h_j'(t) h_j(t)\, dt.$$

Fix $j \geq 1$ for the moment and choose $t_0 \in [0,T]$ such that $h_j^2(t_0) = T^{-1} \|h_j\|^2_{L^2([0,T])}$. Then

$$\begin{aligned}
\lambda_j (h_j^2(T) + h_j^2(0)) &= 2\Big(\int_{t_0}^0 + \int_{t_0}^T \Big) h_j'(t) h_j(t)\, dt + 2\lambda_j \|h_j\|^2_{L^2([0,T])} \\
&\leq 2\lambda_j \|h_j'\|_{L^2([0,T])} \|h_j\|_{L^2([0,T])} + 2\lambda_j \|h_j\|^2_{L^2([0,T])} \\
&\leq 2\lambda_j^2 \|h_j\|^2_{L^2([0,T])} + \|h_j'\|^2_{L^2([0,T])} + \|h_j\|^2_{L^2([0,T])},
\end{aligned}$$

where we also used the Cauchy–Schwarz inequality, the fact that $T \geq 1$, and the inequality $2ab \leq a^2 + b^2$, $a, b \in \mathbb{R}$. Summing over $j \geq 1$, we get

$$\sum_{j\geq 1} \lambda_j (h_j^2(T) + h_j^2(0)) \leq \sum_{j\geq 1} (2\lambda_j^2 \|h_j\|^2_{L^2([0,T])} + \|h_j'\|^2_{L^2([0,T])} + \|h_j\|^2_{L^2([0,T])}), \tag{A.26}$$

from which (A.25) follows, taking into account the discussion after Lemma A.21.

Let us now write

$$\widetilde{H}_X = \{h \in L^2([0,T];\mathcal{H}) : Ah, h' \in L^2([0,T];\mathcal{H})\}$$

and

$$\|h\|_{\widetilde{H}_X}^2 = \|Ah\|_{L^2([0,T];\mathcal{H})}^2 + \|h'\|_{L^2([0,T];\mathcal{H})}^2 + \langle -Ah(0), h(0)\rangle_{\mathcal{H}} + \langle -Ah(T), h(T)\rangle_{\mathcal{H}}.$$

By the RKHS computations in Lemma A.21 it remains to check that $H_X = \widetilde{H}_X$ and $\|h\|_X = \|h\|_{\widetilde{H}_X}$ for all $h \in H_X$. First, let $h = \sum_{j\geq 1} h_j e_j \in H_X$. Then h is absolutely continuous with

$$h' = \sum_{j\geq 1} h_j' e_j \in L^2([0,T];\mathcal{H}), \quad Ah = -\sum_{j\geq 1} \lambda_j h_j e_j \in L^2([0,T];\mathcal{H}). \tag{A.27}$$

Hence, $h \in \widetilde{H}_X$ and therefore $H_X \subset \widetilde{H}_X$. To see the second claim in (A.27), set $h^{(m)} = \sum_{j=1}^m h_j e_j$ and $g^{(m)} = -\sum_{j=1}^m \lambda_j h_j e_j$ for $m \geq 1$. Then, $h^{(m)}(t)$ and $g^{(m)}(t)$ are in $\mathcal{H}$ for all $t \in [0,T]$ and we have $Ah^{(m)} = g^{(m)}$ because $(\lambda_j, e_j)_{j\geq 1}$ is an eigensequence of $-A$. Moreover $h^{(m)}(t) \to h(t)$ and $Ah^{(m)}(t) = g^{(m)}(t) \to g(t) = \sum_{j\geq 1} \lambda_j h_j(t) e_j$ for a.e. t. Since A is closed, we conclude that $Ah(t) = g(t)$ for a.e. t.

Next, let $h \in \widetilde{H}_X$. Then we can write $h = \sum_{j\geq 1} h_j e_j$ with $h_j = \langle h, e_j\rangle \in L^2([0,T])$. Using also [40, Proposition A.22], the h_j are absolutely continuous with $h_j' = \langle h, e_j\rangle' = \langle h', e_j\rangle \in L^2([0,T])$. Hence, $h_j \in H$ for all $j \geq 1$. Moreover, the relations in (A.27) continue to hold, as can be seen from the identities $\langle Ah(t), e_j\rangle = \lambda_j h_j(t)$ and $\langle h', e_j\rangle = h_j'$, and we have

$$\|Ah\|_{L^2([0,T];\mathcal{H})}^2 = \sum_{j\geq 1} \lambda_j^2 \|h_j\|_{L^2([0,T])}^2, \quad \|h'\|_{L^2([0,T];\mathcal{H})}^2 = \sum_{j\geq 1} \|h_j'\|_{L^2([0,T])}^2. \tag{A.28}$$

Hence, $h \in H_X$ and therefore also $\widetilde{H}_X \subset H_X$. We conclude that $\widetilde{H}_X = H_X$ and that the norms coincide, where the latter follows from (A.28) and (A.27). Moreover, inserting

$$\begin{aligned}&\langle -Ah(0), h(0)\rangle_{\mathcal{H}} + \langle -Ah(T), h(T)\rangle_{\mathcal{H}} \\ &\qquad \leq 2\|Ah\|_{L^2([0,T];\mathcal{H})}^2 + \|h'\|_{L^2([0,T];\mathcal{H})}^2 + \|h\|_{L^2([0,T];\mathcal{H})}^2,\end{aligned}$$

the upper RKHS norm bound follows, as can be seen from (A.26), (A.27) and (A.28). ∎

RKHS of multiple measurements

In this section we deduce Theorem A.5 from Theorem A.4. This requires the $K_1, \dots, K_M$ to lie in the dual space $\mathcal{H}_1'$. When $A = \Delta$ this is a Sobolev space of order $s > d/2 - 1$ (see the beginning of Section A.6.2). In Section A.7.7, we give a second more technical proof based on an approximation argument, which provides the claim under the weaker assumption $K_1, \dots, K_M \in \mathcal{D}(A)$.

First proof of Theorem A.5. For a non-decreasing sequence (μ_j) of positive real numbers, take

$$V_\mu = \{f \in \mathcal{H} : \|f\|_{V_\mu}^2 = \sum_{j\geq 1} \mu_j^2 \langle f, e_j\rangle_{\mathcal{H}}^2 < \infty\},$$

and take $\mathcal{H}_1 = V'_\mu$ to be the closure of $\mathcal{H}$ under the norm

$$\|z\|_{V'_\mu}^2 = \sum_{j\geq 1} \frac{1}{\mu_j^2} \langle z, e_j \rangle_{\mathcal{H}}^2.$$

Then, V_μ is continuously embedded in $\mathcal{H}$ and $(V_\mu, \mathcal{H}, V'_\mu)$ forms a Gelfand triple, i.e., $\mathcal{H}$ is identified with its dual and thus $\mathcal{H}$ is also continuously embedded in V'_μ. Moreover, we can extend $\langle f, g \rangle = \langle f, g \rangle_{\mathcal{H}}$ to pairs $f \in V_\mu$ and $g \in V'_\mu$ and we have the (generalised) Cauchy–Schwarz inequality

$$|\langle f, g \rangle| \leq \|f\|_{V_\mu} \|g\|_{V'_\mu}. \tag{A.29}$$

We choose the sequence (μ_j) such that (A.22) holds, meaning that X can be considered as a Gaussian random variable in $L^2([0,T]; V'_\mu)$. For $K_1, \ldots, K_M \in V_\mu$, consider the linear map

$$L : L^2([0,T]; V'_\mu) \to L^2([0,T])^M, \quad Lf(t) = (\langle K_k, f(t) \rangle)_{k=1}^M, t \in [0,T].$$

Then, $LX = X_K$ in distribution. Using (A.29), it is easy to see that L is a bounded operator with norm bounded by $(\sum_{k=1}^M \|K_k\|_{V_\mu}^2)^{1/2}$:

$$\sum_{k=1}^M \int_0^T \langle K_k, f(t) \rangle^2 \,\mathrm{d}t \leq \Big(\sum_{k=1}^M \|K_k\|_{V_\mu}^2 \Big) \|f\|_{L^2([0,T]; V'_\mu)}^2.$$

Next, we show that $L(H_X) = H^M$. First, for $(h_k)_{k=1}^M \in H^M$, the function

$$f = \sum_{k,l=1}^M G_{k,l}^{-1} K_k h_l \in H_X \quad \text{satisfies} \quad Lf = (h_k)_{k=1}^M. \tag{A.30}$$

Hence $H^M \subset L(H_X)$. To see the reverse inclusion, let $f \in H_X$. Set $(h_k)_{k=1}^M = Lf$ such that $h_k(t) = \langle K_k, f(t) \rangle$. By the definition of H_X and properties of the Bochner integral (see, e.g., [40, Proposition A.22]), the h_k are absolutely continuous with derivatives $h'_k(t) = \langle K_k, f'(t) \rangle$, and we have

$$\int_0^T (h'_k(t))^2 \,\mathrm{d}t \leq \|K_k\|_{\mathcal{H}}^2 \int_0^T \|f'(t)\|_{\mathcal{H}}^2 \,\mathrm{d}t = \|K_k\|_{\mathcal{H}}^2 \|f'\|_{L^2([0,T];\mathcal{H})}^2 < \infty.$$

We get $h_k \in H$ for all $k = 1, \ldots, M$. Hence, $L(H_X) \subset H^M$ and therefore $L(H_X) = H^M$. It remains to prove the bound for the norm. Using (A.30), the behavior of the RKHS under linear transformation (see [38, Proposition 4.1]) and Theorem A.4, we have

$$\begin{aligned}\|(h_k)_{k=1}^M\|_{X_K}^2 \leq \Big\|\sum_{k,l=1}^M G_{k,l}^{-1} K_k h_l\Big\|_X^2 &\leq 3\Big\|\sum_{k,l=1}^M G_{k,l}^{-1} A K_k h_l\Big\|_{L^2([0,T];\mathcal{H})}^2 \\ &\quad + \Big\|\sum_{k,l=1}^M G_{k,l}^{-1} K_k h_l\Big\|_{L^2([0,T];\mathcal{H})}^2 + 2\Big\|\sum_{k,l=1}^M G_{k,l}^{-1} K_k h'_l\Big\|_{L^2([0,T];\mathcal{H})}^2.\end{aligned}$$

Using the definition of G_A, the last display becomes

$$\|(h_k)_{k=1}^M\|_{X_K}^2 \leq 3\int_0^T \sum_{k,l=1}^M (G^{-1}G_AG^{-1})_{kl}h_k(t)h_l(t)\ \mathrm{d}t + \int_0^T \sum_{k,l=1}^M (G^{-1})_{kl}h_k(t)h_l(t)\ \mathrm{d}t$$
$$+ 2\int_0^T \sum_{k,l=1}^M (G^{-1})_{kl}h_k'(t)h_l'(t)\ \mathrm{d}t, \tag{A.31}$$

and the claim follows from standard results for the operator norm of symmetric matrices. ■

Proof of Corollary A.6. Since the Laplace operator Δ is negative and self-adjoint, the stochastic convolution (A.14) is just the weak solution in (A.4) and $\mathcal{H} = L^2(\Lambda)$. If $(K_k)_{k=1}^M = (K_{\delta,x_k})_{k=1}^M$ with $\|K\|_{L^2(\mathbb{R}^d)} = 1$, then $K_1, \dots, K_M$ have disjoint supports and satisfy the assumptions of Theorem A.5 with $G = I_{M\times M}$ and G_Δ being a diagonal matrix with $(G_\Delta)_{kk} = \|\Delta K_{\delta,x_k}\|_{L^2(\mathbb{R}^d)}^2$. By construction and the Cauchy–Schwarz inequality, we have $\|K_{\delta,x_k}\| = 1$ and $\|\Delta K_{\delta,x_k}\| \leq \delta^{-2}\|\Delta K\|_{L^2(\mathbb{R}^d)}$. From Theorem A.5, we obtain the RKHS $H_{X_K} = H^M$ of X_K with the claimed upper bound on its norm, where we also used that $\delta^2 \leq \|\Delta K\|_{L^2(\mathbb{R}^d)}$ by assumption. ■

A.6.3 *Proof of the lower bounds*

In this section, we give the main steps of the proof of Theorem A.7, which follows a Gaussian route. First, we combine a classical Gaussian lower bound (based on two hypotheses) with arguments from the Feldman–Hájek theorem to formulate a lower bound scheme that is expressed in terms of covariance operators and RKHS norms. Second, we invoke the RKHS computations from Section A.3 to further reduce our analysis to L^2-distances of the involved (cross-)covariance kernels and their first and second derivatives. Finally, we use semigroup perturbation arguments to compute these distances in the setting of Assumption L. The proofs of three key lemmas are deferred to the appendix.

Gaussian minimax lower bounds

Let $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ be a family of probability measures defined on the same measurable space with a parameter set $\Theta \subset \mathbb{R}^p$. For $\vartheta^0, \vartheta^1 \in \Theta$, the (squared) Hellinger distance between $\mathbb{P}_{\vartheta^0}$ and $\mathbb{P}_{\vartheta^1}$ is defined by $H^2(\mathbb{P}_{\vartheta^0}, \mathbb{P}_{\vartheta^1}) = \int(\sqrt{\mathbb{P}_{\vartheta^0}} - \sqrt{\mathbb{P}_{\vartheta^1}})^2$ (see, e.g. [56, Definition 2.3]). Moreover, if $\vartheta^0, \vartheta^1 \in \Theta$ satisfy

$$H^2(\mathbb{P}_{\vartheta^0}, \mathbb{P}_{\vartheta^1}) \leq 1, \tag{A.32}$$

then we have the lower bound

$$\inf_{\widehat{\vartheta}} \max_{\vartheta\in\{\vartheta^0,\vartheta^1\}} \mathbb{P}_\vartheta\left(|\widehat{\vartheta} - \vartheta| \geq \frac{|\vartheta^0 - \vartheta^1|}{2}\right) \geq \frac{1}{4}\frac{2-\sqrt{3}}{4} =: c_3, \tag{A.33}$$

where the infimum is taken over all $\mathbb{R}^p$-valued estimators $\widehat{\vartheta}$ and $|\cdot|$ denotes the Euclidean norm. For a proof of this lower bound, see [56, Theorem 2.2(ii)].

Next, let $\mathbb{P}_{\vartheta^0}$ and $\mathbb{P}_{\vartheta^1}$ be two Gaussian measures defined on a separable Hilbert space $\mathcal{Z}$ with expectation zero and positive self-adjoint trace-class covariance operators C_{ϑ^0} and C_{ϑ^1}, respectively. By the spectral theorem, there exist strictly positive eigenvalues $(\sigma_j^2)_{j\geq 1}$ and an associated orthonormal system of eigenvectors $(u_j)_{j\geq 1}$ such that $C_{\vartheta^0} = \sum_{j\geq 1} \sigma_j^2 (u_j \otimes u_j)$. Given the Gaussian measure $\mathbb{P}_{\vartheta^0}$, we can associate the RKHS $(H_{\vartheta^0}, \|\cdot\|_{H_{\vartheta^0}})$ of $\mathbb{P}_{\vartheta^0}$ given by $H_{\vartheta^0} = \{h \in \mathcal{Z} : \|h\|_{H_{\vartheta^0}} < \infty\}$ and $\|h\|_{H_{\vartheta^0}}^2 = \sum_{j\geq 1} \sigma_j^{-2} \langle u_j, h\rangle_{\mathcal{Z}}^2$ (cf. the beginning of Section A.6.2). Combining (A.32) with the RKHS machinery, we get the following lower bound.

LEMMA A.22. *In the above Gaussian setting, suppose that $(u_j)_{j\geq 1}$ is an orthonormal basis of $\mathcal{Z}$ and that*

$$\sum_{j\geq 1} \sigma_j^{-2} \|(C_{\vartheta^1} - C_{\vartheta^0})u_j\|_{H_{\vartheta^0}}^2 \leq 1/2. \tag{A.34}$$

Then the lower bound in (A.33) holds, that is

$$\inf_{\widehat{\vartheta}} \max_{\vartheta\in\{\vartheta^0,\vartheta^1\}} \mathbb{P}_\vartheta\Big(|\widehat{\vartheta} - \vartheta| \geq \frac{|\vartheta^0 - \vartheta^1|}{2}\Big) \geq c_3.$$

Lemma A.22 is a consequence of the proof of the Feldman–Hájek theorem [15, Theorem 2.25] in combination with basic properties of the Hellinger distance and the minimax risk. A proof is given in Section A.7.3.

Proof of Theorem A.7

Our goal is to apply Lemma A.22 and Corollary A.6 to the Gaussian process X_δ under Assumption L. We assume without loss of generality that $\|K\|_{L^2(\mathbb{R}^d)} = 1$. We choose $\vartheta^0 = (1,0,0)$ and $\vartheta^1 \in \Theta_1 \cup \Theta_2 \cup \Theta_3$, meaning that the null model is $A_{\vartheta^0} = \Delta$ and the alternatives are $A_{\vartheta^1} = \vartheta_1^1\Delta + \vartheta_2^1(\nabla\cdot b) + \vartheta_3^1$ for $\vartheta^1 \in \mathbb{R}^3$, where ϑ^1 lies in one of the parameter classes Θ_1, Θ_2 or Θ_3. For $\vartheta \in \{\vartheta^0, \vartheta^1\}$, let $\mathbb{P}_{\vartheta,\delta}$ be the law of X_δ on $\mathcal{Z} = L^2([0,T])^M$, let $C_{\vartheta,\delta}$ be its covariance operator, and let $(H_{\vartheta,\delta}, \|\cdot\|_{H_{\vartheta,\delta}})$ be the associated RKHS. For $(f_k)_{k=1}^M \in L^2([0,T])^M$, we have $C_{\vartheta,\delta}(f_k)_{k=1}^M = (\sum_{l=1}^M C_{\vartheta,\delta,k,l} f_l)_{k=1}^M$ with (cross-)covariance operators defined by

$$C_{\vartheta,\delta,k,l} : L^2([0,T]) \to L^2([0,T]),$$
$$C_{\vartheta,\delta,k,l} f_l(t) = \mathbb{E}_\vartheta[\langle X_{\delta,l}, f_l\rangle_{L^2([0,T])} X_{\delta,k}(t)], \qquad 0 \leq t \leq T$$

(see also Section A.7.4 for more details). By stationarity of X_δ under Assumption L

$$C_{\vartheta,\delta,k,l} f_l(t) = \int_0^t c_{\vartheta,\delta,k,l}(t-t') f_l(t')\, \mathrm{d}t' + \int_t^T c_{\vartheta,\delta,l,k}(t'-t) f_l(t')\, \mathrm{d}t', \quad 0 \leq t \leq T$$

with covariance kernels $c_{\vartheta,\delta,k,l}(t) = \mathbb{E}_\vartheta[X_{\delta,k}(t) X_{\delta,l}(0)]$, $0 \leq t \leq T$. Following the notation of Section A.6.3, let $(\sigma_j^2)_{j\geq 1}$ be the strictly positive eigenvalues of $C_{\vartheta^0,\delta}$ and let $(u_j)_{j\geq 1}$ with $u_j = (u_{j,k})_{k=1}^m \in L^2([0,T])^M$ be a corresponding orthonormal system of eigenvectors. By Corollary A.6, we have $H_{\vartheta^0,\delta} = H^M$ as sets. Since H^M is dense in $L^2([0,T])^M$, $(u_j)_{j\geq 1}$ forms an orthonormal basis of $L^2([0,T])^M$. This means that the first assumption of Lemma A.22 is satisfied. To verify the second assumption in (A.34), we will use the bound for the RKHS norm in Corollary A.6.

LEMMA A.23. *In the above setting, we have*

$$\sum_{j=1}^\infty \sigma_j^{-2} \|(C_{\vartheta^0,\delta} - C_{\vartheta^1,\delta})u_j\|_{H_{\vartheta^0,\delta}}^2$$

$$\leq cT\sum_{k,l=1}^{M}\Big(\frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8}\|c_{\vartheta^0,\delta,k,l}-c_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])}+\|c''_{\vartheta^0,\delta,k,l}-c''_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])}\Big)$$

for all $\delta^2\leq\|\Delta K\|_{L^2(\mathbb{R}^d)}$ and all $T\geq 1$, where $c>0$ is an absolute constant.

The proof of Lemma A.23 can be found in Section A.7.4. Moreover, combining Lemma A.17(ii) with perturbation arguments for semigroups, we prove the following upper bound in Section A.7.5.

LEMMA A.24. *In the above setting let $\vartheta^1=(\vartheta_1,\vartheta_2,\vartheta_3)\in\Theta_1\cup\Theta_2\cup\Theta_3$ with $M\geq 1$. Then there exists a constant $c>0$, depending only on K such that*

$$\sum_{k,l=1}^{M}\Big(\delta^{-8}\|c_{\vartheta^0,\delta,k,l}-c_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])}+\|c''_{\vartheta^0,\delta,k,l}-c''_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])}\Big)$$
$$\leq cM(\delta^{-2}(1-\vartheta_1)^2+\vartheta_2^2+\delta^2\vartheta_3^2).$$

Choosing consecutively

$$\vartheta^1=(\vartheta_1,0,0)\in\Theta_1,\qquad \vartheta_1=1+c_2\frac{\delta}{\sqrt{TM}},$$
$$\vartheta^1=(1,\vartheta_2,0)\in\Theta_2,\qquad \vartheta_2=c_2\frac{1}{\sqrt{TM}},$$
$$\vartheta^1=(1,0,\vartheta_3)\in\Theta_3,\qquad \vartheta_3=c_2\min\Big(1,\frac{\delta^{-1}}{\sqrt{TM}}\Big),$$

Theorem A.7 follow from Lemma A.22 in combination with Lemmas A.23 and A.24. ∎

A.7 ADDITIONAL PROOFS

A.7.1 *Additional proofs from Section A.2*

The proof of invertibility of the observed Fisher information is classical when the solution process is a multivariate Ornstein–Uhlenbeck process [33], but requires a different proof for the Itô processes $X^A_{\delta,k}$.

Proof of Lemma A.1. It is enough to show that the first summand with $k=1$ in the definition of the observed Fisher information is $\mathbb{P}$-almost surely positive definite. By a density argument we can assume without loss of generality that $K\in C_c^\infty(\mathbb{R}^d)$. Define a symmetric matrix $\beta\in\mathbb{R}^{p\times p}$, $\beta_{ij}=\langle A_i^*K_{\delta,x_1},A_j^*K_{\delta,x_1}\rangle$ and suppose for $\lambda\in\mathbb{R}^p$ that

$$0=\sum_{i,j=1}^{p}\lambda_i\lambda_j\beta_{ij}=\|\sum_{i=1}^{p}\lambda_iA_i^*K_{\delta,x_1}\|^2.$$

By linear independence this yields $\lambda=0$, and so β is invertible. It follows that

$$dX^A_{\delta,1}(t)=(\langle X(t),A^*_\vartheta A_i^*K_{\delta,x_1}\rangle)_{i=1}^p\,dt+\beta^{1/2}\,d\bar W(t)$$

with a p-dimensional Brownian motion $\bar{W}(t) = \beta^{-1/2}(\langle W(t), A_i^* K_{\delta,x_1}\rangle)_{i=1}^p$. Then $Y = \beta^{-1/2} X_{\delta,1}^A$ satisfies $\mathrm{d}Y(t) = \alpha(t)\,\mathrm{d}t + \mathrm{d}\bar{W}(t)$ for some p-dimensional Gaussian process α. Invertibility of $\int_0^T X_{\delta,1}^A(t) X_{\delta,1}^A(t)^\top \mathrm{d}t$ is equivalent to the invertibility of $\int_0^T Y(t)Y(t)^\top \mathrm{d}t$. Applying first the innovation theorem, cf. [39, Theorem 7.18], componentwise and then the Girsanov theorem for multivariate diffusions, this is further equivalent to the $\mathbb{P}$-almost sure invertibility of $\int_0^T \bar{W}(t)\bar{W}(t)^\top \mathrm{d}t$. The result is now obtained from noting that the determinant of the $p \times p$ dimensional random matrix $(\bar{W}(t_1), \ldots, \bar{W}(t_p))$ is $\mathbb{P}$-almost surely not zero for any pairwise different time points $t_1, \ldots, t_p$, because $\bar{W}$ has independent increments. ∎

Proof of Lemma A.2. It is enough to prove the claim for $A_i^* \in \{1, D_j, D_{jk}\}$ with $n_i \in \{0, 1, 2\}$. Let $u_i = \delta^{-n_i}(v_i)_{\delta,x}$ for $v_i = A_i^* K$. Suppose first $X_0 \in L^p(\Lambda)$. The scaling in Lemma A.13, the Hölder inequality and Lemma A.16 applied to $\delta = 1$, $s = 0$, yield for $1/p + 1/q = 1$

$$\sup_{x\in\mathcal{J}} \langle X_0, S_\vartheta^*(t)u_i\rangle^2 \lesssim \|S_\vartheta(t)X_0\|_{L^p(\Lambda)}^2 \sup_{x\in\mathcal{J}} \|u_i\|_{L^q(\Lambda)}^2 \lesssim \delta^{d(1-2/p)-2n_i} \|X_0\|_{L^p(\Lambda)}^2,$$

The same Lemmas applied to $s = n_i$ also show for $\varepsilon, \varepsilon' > 0$

$$\begin{aligned} \sup_{x\in\mathcal{J}} \int_{\varepsilon'}^T \langle X_0, S_\vartheta^*(t)u_i\rangle^2 \,\mathrm{d}t &\leq \|X_0\|^2 \sup_{x\in\mathcal{J}} \int_{\varepsilon'}^T \|S_\vartheta^*(t)u_i\|^2 \,\mathrm{d}t \\ &\lesssim \delta^{-2n_i} \sup_{x\in\mathcal{J}} \int_{\varepsilon'}^T \|S_{\vartheta,\delta,x}^*(t\delta^{-2})v_i\|_{L^2(\Lambda_{\delta,x})}^2 \,\mathrm{d}t \lesssim \delta^{-2n_i} \int_{\varepsilon'}^T (t\delta^{-2})^{-n_i - d/2+\varepsilon} \,\mathrm{d}t. \end{aligned}$$

Assumption H(ii) implies $1 - n_k - d/2 < 0$, and so the last line is of order $O((\varepsilon')^{1-n_i-d/2+\varepsilon}\delta^{d-2\varepsilon})$. After splitting up the integral we conclude

$$\sup_{x\in\mathcal{J}} \int_0^T \langle X_0, S_\vartheta^*(t)u_i\rangle^2 \,\mathrm{d}t \lesssim \varepsilon' \delta^{d(1-2/p)-2n_i} + (\varepsilon')^{1-n_i-d/2+\varepsilon}\delta^{d-2\varepsilon}.$$

Choosing $\varepsilon' = \delta^{\frac{2n_i+2d/p-2\varepsilon}{n_i+d/2-\varepsilon}}$ yields the order $O(\delta^{h(p)-\varepsilon''})$ with the function $h(p) = d(1-2/p) + 2(n_i + d/p)/(n_i + d/2) - 2n_i$ for any $\varepsilon'' > 0$. We get $h(2) = 2 - 2n_i$ and $h'(p) > 0$. From this obtain the claim when $X_0 \in L^p(\Lambda)$.

Let now $X_0 = \int_{-\infty}^0 S_\vartheta(-t')\,\mathrm{d}W(t')$ and $c_\vartheta \leq 0$. By Itô's isometry, the δ-scaling and changing variables we get

$$\mathbb{E}[\langle X_0, S_\vartheta^*(t)u_i\rangle^2] = \int_0^\infty \|S_\vartheta^*(t'+t)u_i\|^2 \,\mathrm{d}t' = \delta^{2-2n_i} \int_0^\infty \|S_{\vartheta,\delta,x}^*(t' + t\delta^{-2})v_i\|_{L^2(\Lambda_{\delta,x})}^2 \,\mathrm{d}t'.$$

By Lemma A.16 and $\widetilde{c}_\vartheta \leq 0$ the integral is uniformly bounded in $x \in \mathcal{J}$ and $0 \leq t \leq T$ and converges to zero by dominated convergence, because the integrand does so as $\delta \to 0$. From this obtain the claim in the stationary case. ∎

THEOREM A.25. *Let $M_\delta = (M_\delta(t))_{t\geq 0}$ be a family of continuous p-dimensional square integrable martingales with respect to the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, with $M_\delta(0) = 0$ and with quadratic covariation processes $([M_\delta]_t)_{t\geq 0}$. If $T > 0$ is such that*

$$[M_\delta]_T \xrightarrow{\mathbb{P}} I_{p\times p}, \quad \delta \to 0,$$

then we have the convergence in distribution

$$M_\delta(T) \xrightarrow{d} \mathcal{N}(0, I_{p\times p}), \quad \delta \to 0.$$

Proof. For $x \in \mathbb{R}^p$ the process $Y_\delta(t) = x^\top M_\delta(t) x$ defines a one dimensional continuous martingale with respect to $(\mathcal{F}_t)$ with $Y_\delta(0) = 0$ and with quadratic variation

$$[Y_\delta]_T \xrightarrow{\mathbb{P}} x^\top x, \quad \delta \to 0.$$

An application of the Dambis–Dubins–Schwarz theorem ([30, Theorem 3.4.6]) shows $Y_\delta(t) = w_\delta([Y_\delta]_t)$ with scalar Brownian motions $(w_\delta(t))_{t\geq 0}$, which are possibly defined on an extension of the underlying probability space. From the last display Slutsky's lemma implies the joint weak convergence $(w_\delta, [Y_\delta]_T) \xrightarrow{d} (w_0, x^\top x)$ on the product Borel sigma algebra of $C([0,\infty)) \times \mathbb{R}$, where $C([0,\infty))$ is endowed with the uniform topology on compact subsets of $[0,\infty)$, and where w_0 is another scalar Brownian motion. The continuous mapping theorem with respect to $(f,t) \mapsto \phi(f,t) = f(t)$ yields then the result, noting that $w_0(x^\top x)$ has distribution $\mathcal{N}(0, x^\top x)$. ■

A.7.2 *Proof of Proposition A.12*

Proof. Note first that $A_\vartheta = \Delta + \vartheta$ corresponds to $A_1 = 1$, $A_0 = \Delta$ and with observed Fisher information $\mathcal{I}_\delta = \sum_{k=1}^M \int_0^T \langle X(t), K_{\delta,x_k}\rangle^2 \, dt$. In particular, Assumptions H(i), (iii) and (iv) hold.

As in the proof of Theorem A.3, we can suppose that $\|K\|_{L^2(\mathbb{R}^d)} = 1$. Recall from the basic decomposition (A.11), $\widehat{\vartheta}_\delta = \vartheta + \mathcal{I}_\delta^{-1}\mathcal{M}_\delta$ and from the proof of Theorem A.3 that $\mathcal{M}_\delta = \mathcal{M}_\delta(T)$ for a square integrable martingale $(\mathcal{M}_\delta(t))_{0\leq t\leq T}$, whose quadratic variation at $t = T$ coincides with $\mathcal{I}_\delta$. We show below

$$\log(\delta^{-1})^{-1}\mathcal{I}_\delta = O_\mathbb{P}(1), \quad (\log(\delta^{-1})^{-1}\mathcal{I}_\delta)^{-1} = O_\mathbb{P}(1). \tag{A.35}$$

A well-known result about tail properties of square integrable martingales (e.g., [61, p. 3.8]) therefore implies $\mathcal{M}_\delta = O_\mathbb{P}(\log(\delta^{-1})^{1/2})$, and we conclude from the basic decomposition that $\widehat{\vartheta}_\delta = \vartheta + O_\mathbb{P}(\log(\delta^{-1})^{-1/2})$ as claimed.

For (A.35) it is enough to show that $\mathcal{I}_\delta / \mathbb{E}[\mathcal{I}_\delta] \xrightarrow{\mathbb{P}} 1$ and $\log(\delta^{-1})\mathbb{E}[\mathcal{I}_\delta] \asymp 1$, which in turn holds if for some $c, C > 0$, independent of δ,

$$c \leq \log(\delta^{-1})^{-1}\mathbb{E}[\mathcal{I}_\delta] \leq C, \quad \log(\delta^{-1})^{-2}\operatorname{Var}(\mathcal{I}_\delta) = o(1). \tag{A.36}$$

As in the proofs of Lemmas A.18, A.19 and using their notation we compute

$$\begin{aligned}
\mathbb{E}[\mathcal{I}_\delta] &\leq M\delta^2 \int_0^T \int_0^{t\delta^{-2}} \sup_{x\in\mathcal{J}} \|S^*_{\vartheta,\delta,x}(t')K\|^2_{L^2(\Lambda_{\delta,x})} \, dt' \, dt, \\
\operatorname{Var}(\mathcal{I}_\delta) &\lesssim M^2 \sup_{x\in\mathcal{J}} \operatorname{Var}\left(\int_0^T \langle X(t), K_{\delta,x}\rangle^2 \, dt\right) \\
&= M^2 \sup_{x\in\mathcal{J}} 4\delta^6 \int_0^T \int_0^{t\delta^{-2}} \left(\int_0^{t\delta^{-2}-s} f_{1,1}(s,s') \, ds'\right)^2 ds \, dt.
\end{aligned}$$

By the supplement in Lemma A.16 we find in $d = 2$ that

$$\sup_{x\in\mathcal{J}} \|S^*_{\vartheta,\delta,x}(t)K\|_{L^2(\Lambda_{\delta,x})} \lesssim 1 \wedge t^{-1/2}, \quad t \geq 0, \delta \geq 0. \tag{A.37}$$

Plugging this into the last display and using $M\delta^2 \lesssim 1$ provides us with

$$\mathbb{E}[\mathcal{I}_\delta] \lesssim \int_0^T \int_0^{t\delta^{-2}} (1 \wedge s^{-1})\,\mathrm{d}s\,\mathrm{d}t \lesssim \int_0^T \log(t\delta^{-2})\,\mathrm{d}t \lesssim \log(\delta^{-1}),$$
$$\operatorname{Var}(\mathcal{I}_\delta) \lesssim M^2\delta^6 \left(\int_0^{T\delta^{-2}} (1 \wedge (t')^{-1})\,\mathrm{d}t'\right)\left(\int_0^{T\delta^{-2}} (1 \wedge t^{-1/2})\,\mathrm{d}t\right)^2 \lesssim \log(\delta^{-1}).$$

We are thus left with showing $\mathbb{E}[\mathcal{I}_\delta] \gtrsim \log(\delta^{-1})$. First, note that

$$\|S^*_{\vartheta,\delta,x}(t)K\|_{L^2(\mathbb{R}^2)} \geq e^{-T|\vartheta|}\|\widetilde{S}_{\vartheta,\delta,x}(t)K\|_{L^2(\mathbb{R}^2)}$$

and decompose

$$\begin{aligned}\|\widetilde{S}_{\vartheta,\delta,x}(t)K\|^2_{L^2(\mathbb{R}^2)} &= \langle \widetilde{S}_{\vartheta,\delta,x}(t)K, \widetilde{S}_{\vartheta,\delta,x}(t)K\rangle_{L^2(\mathbb{R}^2)} \\ &= \|e^{t\Delta_0}K\|^2_{L^2(\mathbb{R}^2)} + \langle e^{t\Delta_0}K + \widetilde{S}_{\vartheta,\delta,x}(t)K, \widetilde{S}_{\vartheta,\delta,x}(t)K - e^{t\Delta_0}K\rangle_{L^2(\mathbb{R}^2)}.\end{aligned}$$

Recalling $K \geq 0$, the inner product here is uniformly in $x \in \mathcal{J}$ up to a universal constant upper bounded by

$$\langle e^{t\Delta_0}K, (e^{t\Delta_0}K^2)^{1/2}\rangle_{L^2(\mathbb{R}^2)}(\delta t^{1/2}e^{-\delta^{-2}t^{-1}})^{1/2} \leq \|e^{t\Delta_0}K\|_{L^2(\mathbb{R}^2)}\|e^{t\Delta_0}K^2\|_{L^1(\mathbb{R}^2)}\delta t^{1/2},$$

concluding by the Cauchy–Schwarz inequality and $\sup_{x\geq 0} xe^{-x} \lesssim 1$ in the last inequality. Since $\|e^{t\Delta_0}K^2\|_{L^1(\mathbb{R}^2)} = \|K\|^2_{L^2(\mathbb{R}^2)}$ and using (A.37), it thus follows for some $C > 0$ that

$$\|\widetilde{S}_{\vartheta,\delta,x}(t)K\|^2_{L^2(\mathbb{R}^2)} \geq \|e^{t\Delta_0}K\|^2_{L^2(\mathbb{R}^2)} - C(1 \wedge t^{-1/2})\delta t^{1/2}.$$

Hence, using $M\delta^2 \gtrsim 1$,

$$\mathbb{E}[\mathcal{I}_\delta] \quad \geq \int_0^{\delta^{-1}} \|e^{t\Delta_0}K\|^2_{L^2(\mathbb{R}^2)}\,\mathrm{d}t - \int_0^{\delta^{-1}} C\delta\,\mathrm{d}t = \int_0^{\delta^{-1}} \|e^{t\Delta_0}K\|^2_{L^2(\mathbb{R}^2)}\,\mathrm{d}t - C.$$

Suppose without loss of generality that the support of K is contained in the unit ball $B_1(0)$. Writing $e^{t\Delta_0}K = q_t * K$ as convolution with the heat kernel $q_t(x) = (4\pi t)^{-1}\exp(-|x|^2/(4t))$ we have

$$\|e^{t\Delta_0}K\|^2_{L^2(\mathbb{R}^2)} = \int_{\mathbb{R}^2}\left(\int_{B_1(0)} q_t(y-x)K(x)\,\mathrm{d}x\right)^2\mathrm{d}y.$$

The heat kernel $q_t(x)$ is decreasing as $|x| \to \infty$. Hence, for $x \in B_1(0)$, we bound $q_t(y-x) \geq q_t(y + y/|y|)$ for any $y \in \mathbb{R}^2 \setminus \{0\}$. Plugging this into the preceding display yields by $K \geq 0$

$$\begin{aligned}\|e^{t\Delta_0}K\|^2_{L^2(\mathbb{R}^2)} &\geq \int_{\mathbb{R}^2}\left(\int_{B_1(0)} q_t(y + y/|y|)K(x)\,\mathrm{d}x\right)^2\mathrm{d}y \\ &= \|K\|^2_{L^1(\mathbb{R}^2)}\int_{\mathbb{R}^2} q_t(y + y/|y|)^2\,\mathrm{d}y \gtrsim t^{-1}.\end{aligned}$$

In all, we conclude that $\mathbb{E}[\mathcal{I}_\delta] \gtrsim C'\int_0^{\delta^{-1}} t^{-1}\,\mathrm{d}t - C$ for $C' > 0$, implying the wanted lower bound in (A.36). ∎

A.7.3 *Proof of Lemma A.22*

By definition, we have

$$\sum_{j\geq 1}\sigma_j^{-2}\|(C_{\vartheta^1}-C_{\vartheta^0})u_j\|_{H_{\vartheta^0}}^2=\sum_{j,k\geq 1}\sigma_j^{-2}\sigma_k^{-2}\langle u_j,(C_{\vartheta^1}-C_{\vartheta^0})u_k\rangle_{\mathcal{Z}}^2.$$

Combining this with (A.34) and the fact that $(u_j)_{j\geq 1}$ is an orthonormal basis of $\mathcal{Z}$, the infinite matrix $(\langle u_j,(C_{\vartheta^1}-C_{\vartheta^0})u_k\rangle_{\mathcal{Z}}/(\sigma_j\sigma_k))_{j,k=1}^{\infty}$ defines an Hilbert–Schmidt operator S on $\mathcal{Z}$. Let $(v_j,\mu_j)_{j\geq 1}$ be an eigensequence of S with $(v_j)_{j\geq 1}$ being an orthonormal basis of $\mathcal{Z}$. Since $\|S\|_{\mathrm{HS}(\mathcal{Z})}^2\leq 1/2$ by (A.34), we have $\tau_j:=\mu_j+1\in[1-2^{-1/2},1+2^{-1/2}]$ for all $j\geq 1$. Now, let $\{\xi_j\}_{j\geq 1}$ be a sequence of independent standard Gaussian random variables. Then the series

$$\sum_{j\geq 1}\xi_j C_{\vartheta^0}^{1/2}v_j \qquad \text{and} \qquad \sum_{j\geq 1}\sqrt{\tau_j}\xi_j C_{\vartheta^0}^{1/2}v_j$$

converge a.s. and their laws coincide with those of $\mathbb{P}_{\vartheta^0}$ and $\mathbb{P}_{\vartheta^1}$, respectively (see, e.g., [15, Proof of Theorem 2.25] or [34, Pages 166-167]). By standard properties of the Hellinger distance (see, e.g., Equation (A.4) in [49]), we have

$$\begin{aligned}H^2\Big(\bigotimes_{j\geq 1}\mathcal{N}(0,1),\bigotimes_{j\geq 1}\mathcal{N}(0,\tau_j)\Big)&\leq\sum_{j\geq 1}H^2(\mathcal{N}(0,1),\mathcal{N}(0,\tau_j))\\&\leq 2\sum_{j\geq 1}(\tau_j-1)^2=2\|S\|_{\mathrm{HS}(\mathcal{Z})}^2\leq 1.\end{aligned}\tag{A.38}$$

Moreover, defining $Q_{\vartheta^0}=\bigotimes_{j\geq 1}\mathcal{N}(0,1)$, $Q_{\vartheta^1}=\bigotimes_{j\geq 1}\mathcal{N}(0,\tau_j)$ and the measurable map $\mathcal{T}:\mathbb{R}^\infty\to\mathcal{Z}$ by $\mathcal{T}(\{\alpha_j\})=\sum_{j\geq 1}\alpha_j C_{\vartheta^0}^{1/2}v_j$ if the limit exists and $T(\{\alpha_j\})=0$ otherwise, the image measures satisfy $Q_{\vartheta^0}\circ\mathcal{T}^{-1}=\mathbb{P}_{\vartheta^0}$ and $Q_{\vartheta_1}\circ\mathcal{T}^{-1}=\mathbb{P}_{\vartheta_1}$. Finally, by the transformation formula, the minimax risk in (A.33) can be written as $\inf_{\widehat{\vartheta}}\max_{\vartheta\in\{\vartheta^0,\vartheta^1\}}Q_\vartheta(|\widehat{\vartheta}\circ\mathcal{T}-\vartheta|\geq|\vartheta^0-\vartheta^1|/2)$, where the infimum is taken over all measurable functions from $\mathcal{Z}$ to $\mathbb{R}^p$. Allowing for general estimators depending on the whole coefficent vector in $\mathbb{R}^\infty$, the claim follows from (A.38) and (A.33) applied to the product measures Q_{ϑ^0} and Q_{ϑ^1}. ■

A.7.4 *Proof of Lemma A.23*

Let us recall some simple facts on the space $\mathcal{Z}=L^2([0,T])^M$ and a bounded linear operator $I:\mathcal{Z}\to\mathcal{Z}$. First, $\mathcal{Z}$ is a Hilbert space equipped with the inner product $\langle(f_k)_{k=1}^M,(g_k)_{k=1}^M\rangle=\sum_{j=1}^M\langle f_j,g_j\rangle_{L^2([0,T])}$. Second, I can be represented by linear operators $I_{k,l}:L^2([0,T])\to L^2([0,T])$ such that $I(f_k)_{k=1}^M=(\sum_{l=1}^M I_{k,l}f_l)_{k=1}^M$. Finally, I is a Hilbert–Schmidt operator if and only if all I_{jk} are Hilbert–Schmidt operators and we have

$$\|I\|_{\mathrm{HS}(\mathcal{Z})}^2=\sum_{k,l=1}^M\|I_{k,l}\|_{\mathrm{HS}(L^2([0,T]))}^2,$$

where $\|\cdot\|_{\mathrm{HS}(L^2([0,T]))}$ denotes the Hilbert–Schmidt norm on $L^2([0,T])$. Recall also that $(\sigma_j^2)_{j\geq 1}$ are the strictly positive eigenvalues of $C_{\vartheta^0,\delta}$ and that $(u_j)_{j\geq 1}$ with $u_j=(u_{j,k})_{k=1}^m\in\mathcal{Z}$ is a corresponding orthonormal basis of eigenvectors. We first prove a more general version of Lemma A.23.

LEMMA A.26. *Grant Assumption L. Consider an integral operator* $I = (I_{k,l})_{k,l=1}^M : \mathcal{Z} \to \mathcal{Z}$, $I_{k,l}f(t) = \int_0^t \kappa_{k,l}(t-t')f(t')\,dt' + \int_t^T \kappa_{l,k}(t'-t)f(t')\,dt'$ *with square integrable and twice continuously differentiable functions* $\kappa_{k,l}$ *satisfying* $\kappa_{k,l}(0) = \kappa_{l,k}(0)$ *and* $\kappa'_{k,l}(0) = -\kappa'_{l,k}(0)$ *for all* $1 \le k,l \le M$. *Then we have*

$$\sum_{j=1}^{\infty} \sigma_j^{-2} \|Iu_j\|_{H_{\vartheta 0,\delta}}^2 \le \sum_{k,l=1}^{M} \left(240T \frac{\|\Delta K\|_{L^2(\mathbb{R}^d)}^4}{\delta^8} \|\kappa_{k,l}\|_{L^2([0,T])}^2 + 200T \|\kappa''_{k,l}\|_{L^2([0,T])}^2 \right)$$

for all $\delta^2 \le \|\Delta K\|_{L^2(\mathbb{R}^d)}$ *and all* $T \ge 1$.

Proof of Lemma A.26. We divide the proof into the cases of single and multiple measurements.

Case $M = 1$ If $M = 1$, then we consider an integral operator $I : L^2([0,T]) \to L^2([0,T])$, $If(t) = \int_0^T \kappa(|t-t'|)f(t')\,dt'$ with some square integrable and twice continuously differentiable function κ satisfying $\kappa'(0) = 0$. Define the operators

$$I'f(t) = \int_0^T \operatorname{sign}(t-t')\kappa'(|t-t'|)f(t')\,dt',$$
$$I''f(t) = \int_0^T \kappa''(|t-t'|)f(t')\,dt'.$$

We show first

$$(If)'(t) = I'f(t), \quad (If)''(t) = I''f(t). \tag{A.39}$$

Indeed, after splitting up the integral defining $If(t)$ it follows from the chain rule that

$$\begin{aligned}
(If)'(t) &= \left(\int_0^t \kappa(t-t')f(t')\,dt' + \int_t^T \kappa(t'-t)f(t')\,dt' \right)' \\
&= \kappa(0)f(t) + \int_0^t \kappa'(t-t')f(t')\,dt' - \kappa(0)f(t) - \int_t^T \kappa'(t'-t)f(t')\,dt', \\
(If)''(t) &= \kappa'(0)f(t) + \int_0^t \kappa''(t-t')f(t')\,dt' + \kappa'(0)f(t) + \int_t^T \kappa''(t'-t)f(t')\,dt',
\end{aligned}$$

from which (A.39) follows by inserting the assumption $\kappa'(0) = 0$. Thus, Corollary A.6 (applied with $M = 1$) and (A.39) yield for all $j \ge 1$

$$\begin{aligned}
\|Iu_j\|_{H_{\vartheta 0,\delta}}^2 &\le 4\delta^{-4} \|\Delta K\|_{L^2(\mathbb{R}^d)}^2 \|Iu_j\|_{L^2([0,T])}^2 + 2\|(Iu_j)'\|_{L^2([0,T])}^2 \\
&= 4\delta^{-4} \|\Delta K\|_{L^2(\mathbb{R}^d)}^2 \|Iu_j\|_{L^2([0,T])}^2 + 2\|I'u_j\|_{L^2([0,T])}^2
\end{aligned}$$

for all $\delta^2 \le \|\Delta K\|_{L^2(\mathbb{R}^d)}$ and all $T \ge 1$. By construction, I is symmetric, while I' is anti-symmetric, implying that $\langle Iu_j, u_{j'}\rangle_{L^2([0,T])}^2 = \langle u_j, Iu_{j'}\rangle_{L^2([0,T])}^2$ and $\langle I'u_j, u_{j'}\rangle_{L^2([0,T])}^2 = \langle u_j, I'u_{j'}\rangle_{L^2([0,T])}^2$ for all $j, j' \ge 1$. Combining this with Parseval's identity, we get

$$\|Iu_j\|_{H_{\vartheta 0,\delta}}^2 \le \sum_{j'=1}^{\infty} \left(4\delta^{-4} \|\Delta K\|_{L^2(\mathbb{R}^d)}^2 \langle u_j, Iu_{j'}\rangle_{L^2([0,T])}^2 + 2\langle u_j, I'u_{j'}\rangle_{L^2([0,T])}^2 \right).$$

Multiplying the right-hand side with σ_j^{-2} and summing over $j \geq 1$ yields

$$\sum_{j'=1}^{\infty} \left(4\delta^{-4}\|\Delta K\|^2_{L^2(\mathbb{R}^d)}\|Iu_{j'}\|^2_{H_{\vartheta 0,\delta}} + 2\|I'u_{j'}\|^2_{H_{\vartheta 0,\delta}}\right),$$

as can be seen from (A.23). Applying again Corollary A.6 and the definition of the Hilbert–Schmidt norm, we arrive at

$$\begin{aligned}\sum_{j=1}^{\infty} \sigma_j^{-2}\|Iu_j\|^2_{H_{\vartheta 0,\delta}} \leq 16\frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8}\|I\|^2_{\mathrm{HS}(L^2([0,T]))} \\ + 16\frac{\|\Delta K\|^2_{L^2(\mathbb{R}^d)}}{\delta^4}\|I'\|^2_{\mathrm{HS}(L^2([0,T]))} + 4\|I''\|^2_{\mathrm{HS}(L^2([0,T]))}\end{aligned}$$

for all $\delta^2 \leq \|\Delta K\|_{L^2(\mathbb{R}^d)}$ and all $T \geq 1$. Inserting

$$\begin{aligned}&\|I\|^2_{\mathrm{HS}(L^2([0,T]))} = \int_0^T\int_0^T \kappa^2(|t-t'|)\,\mathrm{d}t\,\mathrm{d}t' \leq 2T\|\kappa\|^2_{L^2([0,T])},\\ &\|I'\|^2_{\mathrm{HS}(L^2([0,T]))} \leq 2T\|\kappa'\|^2_{L^2([0,T])}, \quad \|I''\|^2_{\mathrm{HS}(L^2([0,T]))} \leq 2T\|\kappa''\|^2_{L^2([0,T])},\end{aligned} \tag{A.40}$$

we get

$$\begin{aligned}\sum_{j} \sigma_j^{-2}\|Iu_j\|^2_{H_{\vartheta 0,\delta}} \leq 32T\frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8}\|\kappa\|^2_{L^2([0,T])} \\ + 32T\frac{\|\Delta K\|^2_{L^2(\mathbb{R}^d)}}{\delta^4}\|\kappa'\|^2_{L^2([0,T])} + 8T\|\kappa''\|^2_{L^2([0,T])}\end{aligned} \tag{A.41}$$

for all $\delta^2 \leq \|\Delta K\|_{L^2(\mathbb{R}^d)}$ and all $T \geq 1$. The claim now follows from an interpolation inequality (see, e.g., [10]). To get precise constants with respect to T, we give a self-contained argument. By partial integration and the fact that $\kappa'(0) = 0$, we have

$$\int_0^T (\kappa'(t))^2\ \mathrm{d}t = -\int_0^T \kappa''(t)\kappa(t)\ \mathrm{d}t + \kappa'(T)\kappa(T).$$

Let $t_0 \in [0,T]$ such that $\kappa(t_0) = T^{-1}\int_0^T \kappa(t)\ \mathrm{d}t$. Then, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\kappa^2(T) &= 2\int_{t_0}^T \kappa'(t)\kappa(t)\ \mathrm{d}t + \left(T^{-1}\int_0^T \kappa(t)\ \mathrm{d}t\right)^2 \\ &\leq 2\|\kappa\|_{L^2([0,T])}\|\kappa'\|_{L^2([0,T])} + T^{-1}\|\kappa\|^2_{L^2([0,T])}\end{aligned}$$

and similarly

$$\begin{aligned}(\kappa'(T))^2 &= (\kappa'(T))^2 - (\kappa'(0))^2 \\ &= \int_0^T 2\kappa''(t)\kappa'(t)\ \mathrm{d}t \leq 2\|\kappa'\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])}.\end{aligned}$$

Combining these estimates, using also the Cauchy–Schwarz inequality, the fact that $T \geq 1$ and the inequality $2ab \leq \epsilon a^2 + \epsilon^{-1} b^2$, $\epsilon > 0$, $a, b \in \mathbb{R}$ consecutively with $\epsilon \in \{1/4, 1/\sqrt{2}, 1\}$, we get

$$\begin{aligned}
\|\kappa'\|^2_{L^2([0,T])} &\leq \|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])} \\
&\quad + 2\sqrt{\|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])}}\|\kappa'\|_{L^2([0,T])} \\
&\quad + \sqrt{2}\sqrt{\|\kappa'\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])}}\|\kappa\|_{L^2([0,T])} \\
&\leq \|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])} \\
&\quad + 4\|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])} + \|\kappa'\|^2_{L^2([0,T])}/4 \\
&\quad + \|\kappa'\|_{L^2([0,T])}\|\kappa\|_{L^2([0,T])}/2 + \|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])} \\
&\leq 6\|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])} + \|\kappa'\|^2_{L^2([0,T])}/2 + \|\kappa\|^2_{L^2([0,T])}/4
\end{aligned}$$

and thus

$$\|\kappa'\|^2_{L^2([0,T])} \leq 12\|\kappa\|_{L^2([0,T])}\|\kappa''\|_{L^2([0,T])} + \|\kappa\|^2_{L^2([0,T])}/2.$$

Using again the inequality $2ab \leq a^2 + b^2$, $a, b \in \mathbb{R}$, we conclude that

$$\begin{aligned}
&32T\frac{\|\Delta K\|^2_{L^2(\mathbb{R}^d)}}{\delta^4}\|\kappa'\|^2_{L^2([0,T])} \\
&\leq 192T\frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8}\|\kappa\|^2_{L^2([0,T])} + 192T\|\kappa''\|^2_{L^2([0,T])} + 16T\frac{\|\Delta K\|^2_{L^2(\mathbb{R}^d)}}{\delta^4}\|\kappa\|^2_{L^2([0,T])} \\
&\leq 208T\frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8}\|\kappa\|^2_{L^2([0,T])} + 192T\|\kappa''\|^2_{L^2([0,T])}
\end{aligned}$$

where we used the inequality $\delta^2 \leq \|\Delta K\|_{L^2(\mathbb{R}^d)}$ in the last step. Inserting this into (A.41), the claim follows.

Case $M > 1$ We now extend the result to the general case $M > 1$. Define the operators $I' = (I'_{k,l})^M_{k,l=1}$ and $I'' = (I''_{k,l})^M_{k,l=1}$ by

$$\begin{aligned}
I'_{k,l}f(t) &= \int_0^t \kappa'_{k,l}(t-t')f(t')\,\mathrm{d}t' - \int_t^T \kappa'_{l,k}(t'-t)f(t')\,\mathrm{d}t', \\
I''_{k,l}f(t) &= \int_0^t \kappa''_{k,l}(t-t')f(t')\,\mathrm{d}t' + \int_t^T \kappa''_{l,k}(t'-t)f(t')\,\mathrm{d}t'.
\end{aligned}$$

Using that $\kappa_{k,l}(0) = \kappa_{l,k}(0)$ and $\kappa'_{k,l}(0) = -\kappa'_{l,k}(0)$, we have

$$(I_{k,l}f_l)'(t) = I'_{k,l}f_l(t), \quad (I_{k,l}f_l)''(t) = I''_{k,l}f_l(t),$$

as can be seen by proceeding similarly as in the case $M = 1$. Hence, we get

$$(I(f_k)^M_{k=1})' = I'(f_k)^M_{k=1} \quad \text{and} \quad (I(f_k)^M_{k=1})'' = I''(f_k)^M_{k=1}.$$

Thus, Corollary A.6 again yields for all $j \geq 1$

$$\|I(u_{j,k})_{k=1}^M\|_{H_{\vartheta 0,\delta}}^2 \leq 4\delta^{-4}\|\Delta K\|_{L^2(\mathbb{R}^d)}^2 \|I(u_{j,k})_{k=1}^M\|_{L^2([0,T])^M}^2 + 2\|I'(u_{j,k})_{k=1}^M\|_{L^2([0,T])^M}^2.$$

Next, by construction, we have $(I_{k,l})^* = I_{l,k}$ and $(I'_{k,l})^* = -I'_{l,k}$, implying that I is symmetric, while I' is anti-symmetric. Combining this with Parseval's identity, we get

$$\begin{aligned}\|I(u_{j,k})_{k=1}^M\|_{H_{\vartheta 0,\delta}}^2 \leq 4\delta^{-4}\|\Delta K\|_{L^2(\mathbb{R}^d)}^2 &\sum_{j'=1}^\infty \langle (u_{j,k})_{k=1}^M, I(u_{j',k})_{k=1}^M\rangle_{L^2([0,T])^M}^2 \\ &+ 2\sum_{j'=1}^\infty \langle (u_{j,k})_{k=1}^M, I'(u_{j',k})_{k=1}^M\rangle_{L^2([0,T])^M}^2.\end{aligned}$$

Multiplying this with σ_j^{-2} and summing over j yields

$$\begin{aligned}&\sum_{j=1}^\infty \sigma_j^{-2}\|I(u_{j,k})_{k=1}^M\|_{H_{\vartheta 0,\delta}}^2 \\ &\leq \sum_{j'=1}^\infty \left(4\delta^{-4}\|\Delta K\|_{L^2(\mathbb{R}^d)}^2 \|I(u_{j',k})_{k=1}^M\|_{H_{\vartheta 0,\delta}}^2 + 2\|I'(u_{j',k})_{k=1}^M\|_{H_{\vartheta 0,\delta}}^2\right).\end{aligned}$$

Applying again Theorem A.5, we arrive at

$$\begin{aligned}\sum_{j=1}^\infty \sigma_j^{-2}\|I(u_{j,k})_{k=1}^M\|_{H_{\vartheta 0,\delta}}^2 \leq 16\frac{\|\Delta K\|_{L^2(\mathbb{R}^d)}^4}{\delta^8}&\|I\|_{\mathrm{HS}(L^2([0,T])^M)}^2 \\ + 16\frac{\|\Delta K\|_{L^2(\mathbb{R}^d)}^2}{\delta^4}&\|I'\|_{\mathrm{HS}(L^2([0,T])^M)}^2 + 4\|I''\|_{\mathrm{HS}(L^2([0,T])^M)}^2.\end{aligned}$$

Inserting

$$\|I^{(j)}\|_{\mathrm{HS}(L^2([0,T])^M)}^2 = \sum_{k,l=1}^M \|I_{k,l}^{(j)}\|_{\mathrm{HS}(L^2([0,T]))}^2, \qquad j = 0, 1, 2,$$

with

$$\|I_{k,k}^{(j)}\|_{\mathrm{HS}(L^2([0,T]))}^2 = \int_0^T\int_0^T (\kappa_{k,k}^{(j)}(|t-t'|))^2 \,\mathrm{d}t\,\mathrm{d}t' \leq 2T\|\kappa_{k,k}^{(j)}\|_{L^2([0,T])}^2$$

for all $k = 1, \dots, M$ and

$$\begin{aligned}&\|I_{k,l}^{(j)}\|_{\mathrm{HS}(L^2([0,T]))}^2 + \|I_{l,k}^{(j)}\|_{\mathrm{HS}(L^2([0,T]))}^2 \\ &= \int_0^T\int_0^t (\kappa_{k,l}^{(j)}(t-t'))^2 \,\mathrm{d}t'\,\mathrm{d}t + \int_0^T\int_t^T (\kappa_{l,k}^{(j)}(t'-t))^2 \,\mathrm{d}t'\,\mathrm{d}t \\ &+ \int_0^T\int_0^t (\kappa_{l,k}^{(j)}(t-t'))^2 \,\mathrm{d}t'\,\mathrm{d}t + \int_0^T\int_t^T (\kappa_{k,l}^{(j)}(t'-t))^2 \,\mathrm{d}t'\,\mathrm{d}t \\ &= \int_0^T\int_0^T (\kappa_{k,l}^{(j)}(|t-t'|))^2 \,\mathrm{d}t'\,\mathrm{d}t + \int_0^T\int_0^T (\kappa_{l,k}^{(j)}(|t'-t|))^2 \,\mathrm{d}t'\,\mathrm{d}t\end{aligned}$$

$$\leq 2T\|\kappa_{k,l}^{(j)}\|^2_{L^2([0,T])} + 2T\|\kappa_{l,k}^{(j)}\|^2_{L^2([0,T])}$$

for all $1 \leq k \neq l \leq M$ (here, we used $I_{l,k}^{(0)} = I_{l,k}$, $I_{l,k}^{(1)} = I'_{l,k}$ and $I_{l,k}^{(2)} = I''_{l,k}$ and similar notation for the derivatives of κ), we arrive at

$$\sum_{j=1}^{\infty} \frac{\|Iu_j\|^2_{H_{\vartheta^0,\delta}}}{\sigma_j^2} \leq 32T \frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8} \sum_{k,l=1}^{M} \|\kappa_{k,l}\|^2_{L^2([0,T])}$$
$$+ 32T \frac{\|\Delta K\|^2_{L^2(\mathbb{R}^d)}}{\delta^4} \sum_{k,l=1}^{M} \|\kappa'_{k,l}\|^2_{L^2([0,T])} + 8T \sum_{k,l=1}^{M} \|\kappa''_{k,l}\|^2_{L^2([0,T])}$$

for all $\delta^2 \leq \|\Delta K\|_{L^2(\mathbb{R}^d)}$ and all $T \geq 1$. The claim now follows as in the case $M = 1$ by an interpolation argument. ■

Proof of Lemma A.23. We apply Lemma A.26 to $I = C_{\vartheta^0,\delta} - C_{\vartheta^1,\delta} : \mathcal{Z} \to \mathcal{Z}$. This means that $I = (I_{k,l})_{k,l=1}^M$ with $I_{k,l} f(t) = \int_0^t \kappa_{k,l}(t'-t) f(t')\,\mathrm{d}t' + \int_t^T \kappa_{l,k}(t-t') f(t')\,\mathrm{d}t'$ for the integral kernels $\kappa_{k,l}(t) = c_{\vartheta^0,\delta,k,l}(t) - c_{\vartheta^1,\delta,k,l}(t)$, $0 \leq t \leq T$, with

$$c_{\vartheta,\delta,k,l}(t) = \mathbb{E}_\vartheta[X_{\delta,k}(t)X_{\delta,l}(0)] = \int_0^\infty \langle S^*_\vartheta(t+t')K_{\delta,x_k}, S^*_\vartheta(t')K_{\delta,x_l}\rangle\,\mathrm{d}t', \tag{A.42}$$

where the last equality follows from Lemma A.17(ii). From (A.42), we immediately infer $\kappa_{k,l}(0) = \kappa_{l,k}(0)$. The first two derivatives of the cross-covariance integral kernels for $\vartheta \in \{\vartheta^0, \vartheta^1\}$ are

$$c'_{\vartheta,\delta,k,l}(t) = \int_0^\infty \langle A^*_\vartheta S^*_\vartheta(t+t')K_{\delta,x_k}, S^*_\vartheta(t')K_{\delta,x_l}\rangle\,\mathrm{d}t', \tag{A.43}$$
$$c''_{\vartheta,\delta,k,l}(t) = \int_0^\infty \langle (A^*_\vartheta)^2 S^*_\vartheta(t+t')K_{\delta,x_k}, S^*_\vartheta(t')K_{\delta,x_l}\rangle\,\mathrm{d}t'. \tag{A.44}$$

We note that

$$c'_{\vartheta,\delta,k,l}(0) + c'_{\vartheta,\delta,l,k}(0)$$
$$= \int_0^\infty (\mathrm{d}/\mathrm{d}t')\langle S^*_\vartheta(t')K_{\delta,x_k}, S^*_\vartheta(t')K_{\delta,x_l}\rangle\,\mathrm{d}t' = -\langle K_{\delta,x_k}, K_{\delta,x_l}\rangle$$

is independent of ϑ, and hence, $\kappa'_{k,l}(0) + \kappa'_{l,k}(0) = 0$. ■

A.7.5 *Proof of Lemma A.24*

It is sufficient to upper bound the L^2-norms of the $\kappa_{k,l}$. Indeed, the proof below for this remains valid if K_{δ,x_k} is replaced by $\delta^{-4}(A^2_{\vartheta,\delta,x_k}K)_{\delta,x_k}$ and so it yields also the wanted bound on the L^2-norm of $\kappa''_{k,l}$, cf. (A.44). As in (A.17) we have

$$S^*_{\vartheta^0}(t) = e^{t\Delta}, \quad S^*_{\vartheta^1}(t) = U^{-1}_{\vartheta^1} e^{t\vartheta_1\Delta} U_{\vartheta^1} e^{t\widetilde{c}_{\vartheta^1}}$$

with $U_{\vartheta^1}(x) = e^{-(2\vartheta_1)^{-1}\vartheta_2 b\cdot x}$ and $\widetilde{c}_{\vartheta^1} = \vartheta_3 - (4\vartheta_1)^{-1}\vartheta_2^2 \leq 0$. From Lemma A.13, we also have

$$S_{\vartheta^0,\delta,x_k}(t) = e^{t\Delta_{\delta,x_k}}, \quad S^*_{\vartheta^1,\delta,x_k}(t) = U^{-1}_{\vartheta^1,\delta,x_k} e^{t\vartheta_1\Delta_{\delta,x_k}} U_{\vartheta^1,\delta,x_k} e^{t\delta^2\widetilde{c}_{\vartheta^1}}$$

with $U_{\vartheta^1,\delta,x_k}(x)=U_{\vartheta^1}(x_k+\delta x)$. Note that

$$e^{t\vartheta_1\Delta}=U_{\vartheta^1}(x_k)^{-1}e^{t\vartheta_1\Delta}U_{\vartheta^1}(x_k).$$

To get started, let $1\leq k,l\leq M$ and decompose $\kappa_{k,l}=\sum_{j=1}^6\kappa_{k,l}^{(j)}$ with

$$\begin{aligned}
\kappa_{k,l}^{(1)}(t)&=\int_0^\infty\langle(e^{(t+t')\Delta}-e^{(t+t')\vartheta_1\Delta})K_{\delta,x_k},e^{t'\Delta}K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa_{k,l}^{(2)}(t)&=\int_0^\infty\langle e^{(t+t')\vartheta_1\Delta}K_{\delta,x_k},(e^{t'\Delta}-e^{t'\vartheta_1\Delta})K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa_{k,l}^{(3)}(t)&=\int_0^\infty\langle U_{\vartheta^1}(x_k)^{-1}e^{(t+t')\vartheta_1\Delta}(U_{\vartheta^1}(x_k)-U_{\vartheta^1}e^{\widetilde{c}_{\vartheta^1}(t+t')})K_{\delta,x_k},e^{t'\vartheta_1\Delta}K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa_{k,l}^{(4)}(t)&=\int_0^\infty\langle(U_{\vartheta^1}(x_k)^{-1}-U_{\vartheta^1}^{-1})U_{\vartheta^1}S_{\vartheta^1}^*(t+t')K_{\delta,x_k},e^{t'\vartheta_1\Delta}K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa_{k,l}^{(5)}(t)&=\int_0^\infty\langle S_{\vartheta^1}^*(t+t')K_{\delta,x_k},U_{\vartheta^1}(x_k)^{-1}e^{t'\vartheta_1\Delta}(U_{\vartheta^1}(x_k)-U_{\vartheta^1}e^{\widetilde{c}_{\vartheta^1}t'})K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa_{k,l}^{(6)}(t)&=\int_0^\infty\langle S_{\vartheta^1}^*(t+t')K_{\delta,x_k},(U_{\vartheta^1}(x_k)^{-1}-U_{\vartheta^1}^{-1})e^{t'\vartheta_1\Delta}U_{\vartheta^1}e^{\widetilde{c}_{\vartheta^1}t'}K_{\delta,x_l}\rangle\,\mathrm{d}t'.
\end{aligned}$$

We only show $\sum_{1\leq k,l\leq M}\|\kappa_{k,l}^{(j)}\|_{L^2([0,T])}^2\leq c\delta^8M(\delta^{-2}(1-\vartheta_1)^2+\vartheta_2^2+\delta^2\vartheta_3^2)$ for $j=1,3,4$. The proof that the same bound holds for $j=2,5,6$ follows from similar arguments and is therefore skipped. Diagonal (i.e., $k=l$) and off-diagonal (i.e., $k\neq l$) terms are treated separately. Set $K_{k,l}=K(\cdot+\delta^{-1}(x_k-x_l))$.

Case $j=1$ The scaling in Lemma A.13 and changing variables yield

$$\begin{aligned}
\kappa_{k,l}^{(1)}(t\delta^2)&=\delta^2\int_0^\infty\langle(e^{(t+t')\Delta_{\delta,x_k}}-e^{(t+t')\vartheta_1\Delta_{\delta,x_k}})K,e^{t'\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'\\
&=\delta^2\langle\int_0^\infty(e^{(t+2t')\Delta_{\delta,x_k}}-e^{(t\vartheta_1+(t'(1+\vartheta_1))\Delta_{\delta,x_k}})\,\mathrm{d}t'\,K,K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\\
&=\frac{\delta^2}{2}\langle(e^{t\Delta_{\delta,x_k}}-2(1+\vartheta_1)^{-1}e^{t\vartheta_1\Delta_{\delta,x_k}})(-\Delta_{\delta,x_k})^{-1}K,K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\\
&=\frac{\delta^2}{2}\langle(e^{t\Delta_{\delta,x_k}}-e^{t\vartheta_1\Delta_{\delta,x_k}})(-\Delta_{\delta,x_k})^{-1}K,K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\\
&\quad-\frac{\delta^2(1-\vartheta_1)}{2(1+\vartheta_1)}\langle e^{t\vartheta_1\Delta_{\delta,x_k}}(-\Delta_{\delta,x_k})^{-1}K,K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}.
\end{aligned}$$

With

$$e^{t\vartheta_1\Delta_{\delta,x_k}}-e^{t\Delta_{\delta,x_k}}=(1-\vartheta_1)\int_0^t e^{s(\vartheta_1-1)\Delta_{\delta,x_k}}\,\mathrm{d}s\,e^{t\Delta_{\delta,x_k}}(-\Delta_{\delta,x_k}),$$

as follows from the variation of parameters formula, see [18, p. 162], the identity $e^{t\Delta_{\delta,x_k}}=e^{(t/2)\Delta_{\delta,x_k}}e^{(t/2)\Delta_{\delta,x_k}}$ and the Cauchy–Schwarz inequality, we get

$$\left|\langle(e^{t\Delta_{\delta,x_k}}-e^{t\vartheta_1\Delta_{\delta,x_k}})(-\Delta_{\delta,x_k})^{-1}K,K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\right|$$

$$= |1-\vartheta_1|\left|\int_0^t \langle e^{s(\vartheta_1-1)\Delta_{\delta,x_k}}e^{(t/2)\Delta_{\delta,x_k}}K, e^{(t/2)\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\right| \tag{A.45}$$
$$\lesssim |1-\vartheta_1|t\|e^{(t/2)\Delta_{\delta,x_k}}K\|_{L^2(\Lambda_{\delta,x_k})}\|e^{(t/2)\Delta_{\delta,x_k}}K_{k,l}\|_{L^2(\Lambda_{\delta,x_k})}.$$

In the same way, and using $K=\Delta^2\widetilde{K}$,

$$\left|\langle e^{t\vartheta_1\Delta_{\delta,x_k}}(-\Delta_{\delta,x_k})^{-1}K, K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\right| \tag{A.46}$$
$$\le \|e^{(t/2)\vartheta_1\Delta_{\delta,x_k}}\Delta\widetilde{K}\|_{L^2(\Lambda_{\delta,x_k})}\|e^{(t/2)\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\|_{L^2(\Lambda_{\delta,x_k})}.$$

Lemma A.16 therefore gives $\kappa^{(1)}_{k,l}(t\delta^2)\lesssim\delta^2|1-\vartheta_1|(1\wedge t^{-1-d/2+\varepsilon})$ for any $\varepsilon>0$. Changing variables one more time and recalling that $\vartheta_1\ge 1$ already proves for the sum of diagonal terms that $\sum_{1\le k\le M}\|\kappa^{(1)}_{k,k}\|^2_{L^2([0,T])}\lesssim\delta^6(1-\vartheta_1)^2M$, and the implied constant depends only on K.

With respect to the off-diagonal terms, by exploring the different supports of K and $K(\cdot+\delta^{-1}(x_k-x_l))$, we can obtain a second bound for $\kappa^{(1)}_{k,l}$. First, Lemma A.15 gives

$$\sup_{y\in\operatorname{supp}K}\left|(e^{t\Delta_{\delta,x_k}}\Delta^2\widetilde{K}_{k,l})(y)\right|\le\|e^{t\Delta_{\delta,x_k}}\Delta^2\widetilde{K}_{k,l}\|_{L^\infty(\Lambda_{\delta,x_k})}$$
$$\lesssim\|\Delta^2\widetilde{K}_{k,l}\|_{L^\infty(\mathbb{R}^d)}\wedge(t^{-2}\|\widetilde{K}_{k,l}\|_{L^\infty(\mathbb{R}^d)})\lesssim 1\wedge t^{-2}, \tag{A.47}$$

while on the other hand Lemma A.14(i) shows

$$\sup_{y\in\operatorname{supp}K}|(e^{t\Delta_{\delta,x_k}}K_{k,l})(y)|\lesssim\sup_{y\in\operatorname{supp}K}|(e^{t\Delta_0}|K_{k,l}|)(y)|$$
$$=\sup_{y\in\operatorname{supp}K}\int_{\mathbb{R}^d}(4\pi t)^{-d/2}\exp(-|x-y|^2/(4t))|K_{k,l}(x)|\,\mathrm{d}x$$
$$\le(4\pi t)^{-d/2}e^{-c'\frac{|x_k-x_l|^2}{\delta^2 t}}\|K\|_{L^1(\mathbb{R}^d)}\lesssim t^{-d/2}e^{-c'\frac{|x_k-x_l|^2}{\delta^2 t}}, \tag{A.48}$$

for some $c'>0$. By applying the Hölder inequality and using the results from the last two displays we thus obtain for (A.45) the upper bound

$$(\vartheta_1-1)t\|K\|_{L^1(\mathbb{R})}\sup_{0\le s\le t,y\in\operatorname{supp}K}|(e^{(t+s(\vartheta_1-1))\Delta_{\delta,x_k}}K_{k,l})(y)|^{1/2+1/2}$$
$$\lesssim(\vartheta_1-1)\sup_{y\in\operatorname{supp}K}|(e^{t\Delta_0}|K_{k,l}|)(y)|^{1/2}\lesssim(\vartheta_1-1)t^{-d/4}e^{-(c'/2)\frac{|x_k-x_l|^2}{\delta^2 t}}.$$

The same upper bound (up to the factor ϑ_1-1 and with c' instead of $c'/2$) holds in (A.46). Hence, together with the bound $\kappa^{(1)}_{k,l}(t\delta^2)\lesssim\delta^2(\vartheta_1-1)t^{-(1+d/2-\epsilon d/4)(1-\epsilon)^{-1}})$ from above (for sufficiently small ϵ) we get

$$|\kappa^{(1)}_{k,l}(t\delta^2)|\lesssim\delta^2(\vartheta_1-1)\min\left(t^{-(1+d/2-\epsilon d/4)(1-\epsilon)^{-1}},t^{-d/4}e^{-(c'/2)\frac{|x_k-x_l|^2}{\delta^2 t}}\right)$$
$$\le\delta^2(\vartheta_1-1)t^{-1-d/2}e^{-\epsilon'\frac{|x_k-x_l|^2}{\delta^2 t}} \tag{A.49}$$

for $\epsilon'=c'\epsilon/2$, where we have used the inequality $\min(a,b)\le a^{1-\epsilon}b^{\epsilon}$ valid for $a,b\ge0$. Applying the bound

$$\int_0^\infty t^{-p-1}e^{-a/t}\,\mathrm{d}t=a^{-p}\int_0^\infty t^{-p-1}e^{-1/t}\,\mathrm{d}t\lesssim a^{-p}$$

to $p = 1 + d > 0$ and $a = 2\epsilon'\delta^{-2}|x_k - x_l|^2$ this means

$$\int_0^T \kappa_{k,l}^{(1)}(t)^2 \,\mathrm{d}t \lesssim \delta^6(1-\vartheta_1)^2 \int_0^\infty t^{-2-d} e^{-2\epsilon' \frac{|x_k-x_l|^2}{\delta^2 t}} \,\mathrm{d}t$$
$$\lesssim \delta^6(1-\vartheta_1)^2 \frac{\delta^{2+2d}}{|x_k - x_l|^{2+2d}}. \tag{A.50}$$

Recalling that the x_k are δ-separated we obtain from Lemma A.27 below that

$$\sum_{1\le k\neq l\le M} \|\kappa_{k,l}^{(1)}\|_{L^2([0,T])}^2 \lesssim \delta^6(1-\vartheta_1)^2 \sum_{k=1}^M \sum_{l=1, l\neq k}^M \frac{\delta^{2+2d}}{|x_k - x_l|^{2+2d}} \lesssim \delta^6(1-\vartheta_1)^2 M.$$

Together with the bounds for the diagonal terms this yields in all $\sum_{1\le k,l\le M}\|\kappa_{k,l}^{(1)}\|_{L^2([0,T])}^2 \le c\delta^6 M(1-\vartheta_1)^2$ for a constant c depending only on K.

Case $j = 3$ We begin again with the scaling from Lemma A.13 and changing variables such that with the multiplication operators $V_{t,t',\delta}(x) = 1 - e^{\widetilde{c}_{\vartheta^1}\delta^2(t+t') - (2\vartheta_1)^{-1}\vartheta_2\delta b\cdot x}$

$$\kappa_{k,l}^{(3)}(t\delta^2) = \delta^2 \int_0^\infty \langle e^{(t+t')\vartheta_1\Delta_{\delta,x_k}} V_{t,t',\delta}K, e^{t'\vartheta_1\Delta_{\delta,x_k}} K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})} \,\mathrm{d}t'$$
$$= \delta^2 \int_0^\infty \langle e^{(t/2+t')\vartheta_1\Delta_{\delta,x_k}} V_{t,t',\delta}K, e^{(t/2+t')\vartheta_1\Delta_{\delta,x_k}} K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})} \,\mathrm{d}t'. \tag{A.51}$$

Since K is compactly supported and $\widetilde{c}_{\vartheta^1} \le 0$, $V_{t,t',\delta}$ can be extended to smooth multiplication operators with operator norms bounded by $v_{t,t',\delta} = -\widetilde{c}_{\vartheta^1}\delta^2(t + t') + (2\vartheta_1)^{-1}\delta\vartheta_2$. Recalling $K = \Delta^2\widetilde{K}$, Lemma A.16 gives for any $\epsilon > 0$

$$|\kappa_{k,l}^{(3)}(t\delta^2)| \le \delta^2 \int_0^\infty \|e^{(t/2+t')\vartheta_1\Delta_{\delta,x_k}} V_{t,t',\delta}K\|_{L^2(\Lambda_{\delta,x_k})} \|e^{(t/2+t')\vartheta_1\Delta_{\delta,x_k}} K_{k,l}\|_{L^2(\Lambda_{\delta,x_k})} \,\mathrm{d}t'$$
$$\le \delta^2 \int_0^\infty v_{t,t',\delta}(1 \wedge (t+t')^{-4-d/2+\epsilon}) \,\mathrm{d}t' \lesssim \delta^3(-\widetilde{c}_{\vartheta^1}\delta + (2\vartheta_1)^{-1}\vartheta_2)(1 \wedge t^{-1-d/2})$$
$$\le \delta^3(\delta|\vartheta_3| + \vartheta_2)(1 \wedge t^{-1-d/2}), \tag{A.52}$$

recalling in the last line that $\vartheta_1 \ge 1$ and $\vartheta_2 \le 1$. Changing variables therefore proves for the sum of diagonal terms $\sum_{1\le k\le M}\|\kappa_{k,k}^{(3)}\|_{L^2([0,T])}^2 \lesssim \delta^8(\delta|\vartheta_3| + \vartheta_2)^2 M$.

With respect to the off-diagonal terms we have

$$\kappa_{k,l}^{(3)}(t\delta^2) = \delta^2 \int_0^\infty \langle V_{t,t',\delta}K, e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}} K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})} \,\mathrm{d}t'.$$

Write $K = \Delta\bar{K}$ for some compactly supported $\bar{K}$ and note that

$$V_{t,t',\delta}K = V_{t,t',\delta}\Delta\bar{K} = \Delta(V_{t,t',\delta}\bar{K}) - (\Delta V_{t,t',\delta})\bar{K} - 2\nabla V_{t,t',\delta}\cdot\nabla\bar{K}.$$

Similar to (A.47) we find from Lemma A.15

$$\sup_{y\in\operatorname{supp}\bar{K}} \left|(\Delta e^{t\Delta_{\delta,x_k}} K_{k,l})(y)\right| \lesssim \|\Delta K_{k,l}\|_{L^\infty(\mathbb{R}^d)} \wedge (t^{-3}\|\widetilde{K}_{k,l}\|_{L^\infty(\mathbb{R}^d)}) \lesssim 1 \wedge t^{-3}.$$

Together with the Hölder inequality and (A.48) this provides us for sufficiently small $\epsilon > 0$ with

$$\begin{aligned}
&\delta^2\left|\int_0^\infty \langle \Delta(V_{t,t',\delta}\bar{K}), e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'\right| \\
&= \delta^2\left|\int_0^\infty \langle V_{t,t',\delta}\bar{K}, \Delta e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'\right| \\
&\lesssim \delta^2\int_0^\infty v_{t,t',\delta}\|\bar{K}\|_{L^1(\mathbb{R}^d)}\sup_{y\in\operatorname{supp}K}\left|(\Delta e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}}K_{k,l})(y)\right|\mathrm{d}t' \\
&\lesssim \delta^2\int_0^\infty v_{t,t',\delta}(1\wedge(t+2t'))^{-3(1-\epsilon)}\sup_{y\in\operatorname{supp}K}\left|(e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}}\Delta K_{k,l})(y)\right|^{\epsilon}\mathrm{d}t' \\
&\lesssim \delta^3(\delta|\vartheta_3|+\vartheta_2)e^{-\epsilon c'\frac{|x_k-x_l|^2}{\delta^2 t}}\int_0^\infty (1\wedge(t')^{-1-\epsilon d/2})\,\mathrm{d}t' \lesssim \delta^3(\delta|\vartheta_3|+\vartheta_2)e^{-\epsilon c'\frac{|x_k-x_l|^2}{\delta^2 t}}.
\end{aligned}$$

Next, using $\vartheta_2 \le 1$, we have $\|\Delta V_{t,t',\delta}\|_{L^\infty(\Lambda_{\delta,x_k})} + \||\nabla V_{t,t',\delta}|\|_{L^\infty(\Lambda_{\delta,x_k})} \lesssim \delta\vartheta_2$, and so analogously to the computations in the last display

$$\begin{aligned}
&\delta^2\left|\int_0^\infty \langle (\Delta V_{t,t',\delta})\bar{K} + 2\nabla V_{t,t',\delta}\cdot\nabla\bar{K}, e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'\right| \\
&\lesssim \delta^3\vartheta_2\int_0^\infty (\|\bar{K}\|_{L^1(\mathbb{R}^d)} + \||\nabla\bar{K}|\|_{L^1(\mathbb{R}^d)})\sup_{y\in\operatorname{supp}K}\left|(e^{(t+2t')\vartheta_1\Delta_{\delta,x_k}}K_{k,l})(y)\right|\mathrm{d}t' \\
&\lesssim \delta^3\vartheta_2 e^{-\epsilon c'\frac{|x_k-x_l|^2}{\delta^2 t}}.
\end{aligned}$$

In all, this means $|\kappa^{(3)}_{k,l}(t\delta^2)| \lesssim \delta^3(\delta|\vartheta_3|+\vartheta_2)e^{-\epsilon c'\frac{|x_k-x_l|^2}{\delta^2 t}}$. Arguing as for (A.49) and (A.50), as well as using (A.52) we conclude that $|\kappa^{(3)}_{k,l}(t\delta^2)| \lesssim \delta^3(\delta|\vartheta_3|+\vartheta_2)t^{-1/2-d/2}e^{-\epsilon'\frac{|x_k-x_l|^2}{\delta^2 t}}$ for some $\epsilon' > 0$ and

$$\int_0^T \kappa^{(3)}_{k,l}(t)^2\,\mathrm{d}t \lesssim \frac{\delta^{8+2d}(\delta|\vartheta_3|+\vartheta_2)^2}{|x_k-x_l|^{2d}}.$$

We thus get for diagonal and off-diagonal terms that $\sum_{1\le k,l\le M}\|\kappa^{(3)}_{k,l}\|^2_{L^2([0,T])} \le c\delta^8(\delta|\vartheta_3|+\vartheta_2)^2 M$ for a constant c depending only on K.

Case $j = 4$ As in the previous cases we have

$$\kappa^{(4)}_{k,l}(t\delta^2) = \delta^2\int_0^\infty \langle (e^{-\delta(\vartheta_2/\vartheta_1)b\cdot x}-1)S^*_{\vartheta_1,\delta,x_k}(t+t')K, e^{t'\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'.$$

Using the Cauchy–Schwarz inequality, Lemma A.14(i) and Lemma A.16 with $K = \Delta^2\widetilde{K}$ we get for any $\epsilon > 0$, and recalling that $\vartheta_1 \ge 1$,

$$\begin{aligned}
&\langle (e^{-\delta(\vartheta_2/\vartheta_1)b\cdot x}-1)S^*_{\vartheta_1,\delta,x_k}(t+t')K, e^{t'\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})} \\
&\qquad\lesssim \delta\vartheta_2(1\wedge(t+t')^{-2-d/4+\epsilon})\||x|e^{t'\vartheta_1\Delta_0}|K_{k,l}|\|_{L^2(\mathbb{R}^d)}.
\end{aligned}\tag{A.53}$$

Note that $K_{k,l} \in C^1_c(\mathbb{R}^d)$ such that $|K_{k,l}| \in H^{1,\infty}(\mathbb{R}^d)$ and $\nabla|K_{k,l}| \in L^\infty(\mathbb{R}^d)$ with compact support. Using now [4, Lemma A.2(ii)] to the extent that

$$x(e^{t'\vartheta_1\Delta_0}|K_{k,l}|)(x) = (e^{t'\vartheta_1\Delta_0}(-2t'\vartheta_1\nabla|K_{k,l}| + x|K_{k,l}|))(x), \tag{A.54}$$

we find that the $L^2(\mathbb{R}^d)$-norm in (A.53) is uniformly bounded in $t' > 0$. Hence, $|\kappa^{(4)}_{k,l}(t\delta^2)| \lesssim \delta^3\vartheta_2(1 \wedge t^{-1/2-d/4-\epsilon})$ and changing variables shows for the sum of diagonal terms

$$\sum_{1\leq k\leq M} \|\kappa^{(4)}_{k,k}\|^2_{L^2([0,T])} \lesssim \delta^8\vartheta_2^2 M.$$

Regarding the off-diagonal terms we have similarly for some $\bar{K} \in L^\infty(\mathbb{R}^d)$ having compact support

$$\begin{aligned}
&\left|\langle (e^{-\delta(\vartheta_2/\vartheta_1)b\cdot x} - 1)S^*_{\vartheta^1,\delta,x_k}(t+t')K, e^{t'\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\right| \\
&= \left|\langle K, S_{\vartheta^1,\delta,x_k}(t+t')(e^{-\delta(\vartheta_2/\vartheta_1)b\cdot x} - 1)e^{t'\vartheta_1\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\right| \\
&\lesssim \delta\vartheta_2\|K\|_{L^1(\mathbb{R}^d)} \sup_{y\in\operatorname{supp} K}\left|\left(e^{(t+t')\vartheta_1\Delta_0}|x|e^{t'\vartheta_1\Delta_0}|K_{k,l}|\right)(y)\right| \\
&\lesssim \delta\vartheta_2(1\vee t') \sup_{y\in\operatorname{supp} K}\left|\left(e^{(t+2t')\vartheta_1\Delta_0}|\bar{K}_{k,l}|\right)(y)\right| \\
&\lesssim \delta\vartheta_2(1\vee t')t^{-d/2}e^{-c'\frac{|x_k-x_l|^2}{\delta^2 t}},
\end{aligned}$$

using (A.48). Arguing as for (A.49) and (A.50) we then find from combining the last display with (A.53) that $|\kappa^{(4)}_{k,l}(t\delta^2)| \lesssim \delta^3\vartheta_2 t^{-1/2-d/4-\epsilon'}e^{-\epsilon'\frac{|x_k-x_l|^2}{\delta^2 t}}$ for some $\epsilon' > 0$ and

$$\int_0^T \kappa^{(4)}_{k,l}(t)^2\,\mathrm{d}t \lesssim \frac{\delta^{8+d+4\epsilon'}\vartheta_2^2}{|x_k-x_l|^{4\epsilon'+d}},$$

and so in all, for diagonal and off-diagonal terms, $\sum_{1\leq k,l\leq M}\|\kappa^{(4)}_{k,l}\|^2_{L^2([0,T])} \leq c\delta^8\vartheta_2^2 M$ for a constant c depending only on K. ∎

Lemma A.27. *Let $x_1,\dots,x_M$ be δ-separated points in $\mathbb{R}^d$, and let $p > d$. Then we have*

$$\sum_{k=2}^M \frac{1}{|x_1-x_k|^p} \leq C\delta^{-p},$$

where C is a constant depending only on d and p.

Proof. Since $x_1,\dots,x_M$ are δ-separated, the Euclidean balls $B(x_k,\delta/2) = \{y \in \mathbb{R}^d : |y-x_k| \leq \delta/2\}$ around the x_k of radius $\delta/2$ are disjoint. Moreover, for $y \in B(x_k,\delta/2)$ and $k > 1$, we have

$$|y-x_1| \leq |y-x_k| + |x_k-x_1| \leq \frac{\delta}{2} + |x_k-x_1| \leq \frac{3}{2}|x_k-x_1|,$$

implying that

$$\frac{1}{|x_k-x_1|} \leq \frac{3}{2}\frac{1}{|y-x_1|}.$$

We conclude that

$$\sum_{k=2}^M \frac{1}{|x_1-x_k|^p} = \sum_{k=2}^M \frac{1}{\operatorname{vol}(B(x_k,\delta/2))}\int_{B(x_k,\delta/2)} \frac{1}{|x_1-x_k|^p}\,\mathrm{d}y$$

$$
\begin{aligned}
&\leq \frac{(3/2)^p}{(\delta/2)^d \operatorname{vol}(B(0,1))} \sum_{k=2}^{M} \int_{B(x_k,\delta/2)} \frac{1}{|y-x_1|^p}\,\mathrm{d}y \\
&\leq \frac{(3/2)^p}{(\delta/2)^d \operatorname{vol}(B(0,1))} \int_{B(x_1,\delta/2)^c} \frac{1}{|y-x_1|^p}\,\mathrm{d}y \\
&= \frac{(3/2)^p}{(\delta/2)^d \operatorname{vol}(B(0,1))} \int_{B(0,\delta/2)^c} \frac{1}{|y|^p}\,\mathrm{d}y.
\end{aligned}
$$

Changing to polar coordinates, we arrive at

$$
\sum_{k=2}^{M} \frac{1}{|x_1-x_k|^p} \leq \frac{d(3/2)^p}{(\delta/2)^d} \int_{\delta/2}^{\infty} \frac{1}{t^p} t^{d-1}\,\mathrm{d}t = \frac{d(3/2)^p}{(\delta/2)^p} \int_{1}^{\infty} s^{d-p-1}\,\mathrm{d}s.
$$

Since $p > d$, the latter integral is finite, and the claim follows. ∎

A.7.6 *Proof of Theorem A.9*

Argue as in the proof of Theorem A.7, using slight modifications of Lemmas A.23 and A.24. The key additional ingredient is an appropriate extension of Corollary A.6. For this, let $A_1, \dots, A_p$ be as in Section A.2.1. We assume that

$$
(A_i^* K)_{i=1}^p \text{ are linearly independent in } L^2(\Lambda). \tag{A.55}
$$

Define

$$
H = \left(\left\langle \frac{A_i^* K}{\|A_i^* K\|_{L^2(\Lambda)}}, \frac{A_j^* K}{\|A_j^* K\|_{L^2(\Lambda)}} \right\rangle \right)_{i,j=1}^p,
$$

and let $\lambda_{\min} = \lambda_{\min}(H)$ be the smallest eigenvalue of H. By (A.55), H is non-singular, meaning that $\lambda_{\min}(H) > 0$. Finally, let

$$
v = \sum_{i=1}^{p} \frac{\|\Delta A_i^* K\|_{L^2(\Lambda)}^2}{\|A_i^* K\|_{L^2(\Lambda)}^2}.
$$

COROLLARY A.28. *Let $(H_{X_\delta}, \|\cdot\|_{X_\delta})$ be the RKHS of the measurements X_δ with differential operator $A_\vartheta = \Delta$, where $X_\delta(t) = (\langle X(t), K_{1k}\rangle, \dots, \langle X(t), K_{pk}\rangle)_{k=1}^M$ and $K_{ik} = A_i^* K_{\delta,x_k} / \|A_i^* K_{\delta,x_k}\|_{L^2(\Lambda)}$. Then we have $H_{X_\delta} = (H^p)^M$ and*

$$
\|((h_{ip})_{i=1}^p)_{k=1}^M\|_{X_\delta}^2 \leq \frac{4vp}{\delta^4 \lambda_{\min}^2} \sum_{k=1}^{M} \sum_{i=1}^{p} \|h_{ik}\|_{L^2([0,T])}^2 + \frac{2}{\lambda_{\min}^2} \sum_{k=1}^{M} \sum_{i=1}^{p} \|h_{ik}'\|_{L^2([0,T])}^2
$$

for all $((h_{ip})_{i=1}^p)_{k=1}^M \in (H^p)^M$, $\delta^2 \leq \sqrt{v}$ and $T \geq 1$.

Proof of Corollary A.28. First, let $M = 1$. Additionally to H, define

$$
H_\Delta = \left(\left\langle \frac{\Delta A_i^* K}{\|A_i^* K\|_{L^2(\Lambda)}}, \frac{\Delta A_j^* K}{\|A_j^* K\|_{L^2(\Lambda)}} \right\rangle \right)_{i,j=1}^p.
$$

By the Cauchy–Schwarz inequality, we have $\|H_\Delta\|_{\mathrm{op}} \le \upsilon$. Moreover, we have $G = H$ and $G_\Delta = \delta^{-4} H_\Delta$. Inserting these bounds into Theorem A.5, we obtain that

$$\|(h_i)_{i=1}^p\|_{X_K}^2 \le \left(\frac{3\upsilon}{\delta^4 \lambda_{\min}^2} + \frac{1}{\lambda_{\min}}\right) \sum_{i=1}^p \|h_i\|_{L^2([0,T])}^2 + \frac{2}{\lambda_{\min}} \sum_{i=1}^p \|h_i'\|_{L^2([0,T])}^2.$$

Next, let $M > 1$. Then G and G_Δ are block-diagonal with M equal $p \times p$-blocks all of the above form and we get

$$\|((h_i)_{i=1}^p)_{k=1}^M\|_{X_K}^2 \le \frac{4\upsilon}{\delta^4 \lambda_{\min}^2} \sum_{i=1}^p \sum_{k=1}^M \|h_{ik}\|_{L^2([0,T])}^2 + \frac{2}{\lambda_{\min}^2} \sum_{i=1}^p \sum_{k=1}^M \|h_{ik}'\|_{L^2([0,T])}^2,$$

where we also used that $\lambda_{\min} \in (0,1]$ and $\delta^2 \le \sqrt{\upsilon}$. ∎

A.7.7 *Second proof of Theorem A.5*

In this Appendix, we prove Theorem A.5 under the weaker assumption $K_1, \dots, K_M \in \mathcal{D}(A)$. This is achieved by an additional approximation argument. Let $X_m(t) = \sum_{j \le m} Y_j(t) e_j$, $0 \le t \le T$, be the projection of $X(t)$ onto $V_m = \mathrm{span}\{e_1, \dots, e_m\}$, and taking values in $L^2([0,T]; V_m)$. We start with the following consequence of Lemma A.20.

Lemma A.29. *For every $m \ge 1$, the RKHS $(H_{X_m}, \|\cdot\|_{X_m})$ of X_m satisfies*

$$H_{X_m} = \left\{ h = \sum_{j=1}^m h_j e_j : h_j \in H, 1 \le j \le m \right\} \quad \text{and} \quad \|h\|_{X_m}^2 = \sum_{j=1}^m \|h_j\|_{Y_j}^2.$$

Moreover, we have $\|h\|_{X_m} = \|h\|_X$ with the latter norm defined in Theorem A.4.

Proof. Since $L^2([0,T]; V_m)$ is isomorphic to $L^2([0,T])^m$, it suffices to compute the RKHS of the coefficient vector $Y = (Y_1, \dots, Y_m)$. Using that $Y_1, \dots, Y_m$ are independent stationary Ornstein–Uhlenbeck processes, the vector Y is a Gaussian process in $L^2([0,T])^m$ with expectation zero and covariance operator $\bigoplus_{j=1}^m C_{Y_j}$ with $C_{Y_j} : L^2([0,T]) \to L^2([0,T])$ being the covariance operator of Y_j. Combining this with (A.23) and Lemma A.20, we conclude that the RKHS of Y is equal to H^m with norm $\sum_{j=1}^m \|h_j\|_{Y_j}^2$. Translating this back to X_m, the first claim follows. The second one follows from $(\lambda_j, e_j)_{j=1}^\infty$ being an eigensystem of $-A$. ∎

Proof of Theorem A.5. The first step will be to compute the RKHS $(H_{X_{K,m}}, \|\cdot\|_{X_{K,m}})$ of $X_{K,m} = (\langle X_m, K_k \rangle_{\mathcal{H}})_{k=1}^M$. To this end define the bounded linear map

$$L : L^2([0,T]; V_m) \to L^2([0,T])^M, f \mapsto (\langle f, K_k \rangle_{\mathcal{H}})_{k=1}^M.$$

Combining the fact that $LX_m = X_{K,m}$ in distribution with Proposition 4.1 in [38] and Lemma A.29, we obtain that $H_{X_{K,m}} = L(\{h : h = \sum_{j=1}^m h_j e_j : h_j \in H\})$. This implies $H_{X_{K,m}} \subset H^M$. To see the reverse inclusion, let P_m be the orthogonal projection of $L^2(\Lambda)$ onto $V_m = \mathrm{span}\{e_1, \dots, e_m\}$, and let $G_m = (\langle P_m K_k, P_m K_l \rangle_{\mathcal{H}})_{k,l=1}^M$. Since $(e_j)_{j \ge 1}$ is an orthonormal basis of $\mathcal{H}$, G_m tends (e.g., in operator norm) to G as $m \to \infty$. Since G is non-singular, we deduce that G_m is non-singular for all m large enough (which we assume from now on). Hence, for $(h_k)_{k=1}^M \in H^M$, we have that

$$f = \sum_{k,l=1}^{M} (G_m)^{-1}_{k,l} P_m K_k h_l \in H_{X_m} \quad \text{satisfies} \quad Lf = (h_k)_{k=1}^M, \tag{A.56}$$

where we also used that $P_m K_k \in V_m$ for all $1 \le k \le M$. Hence, $H^M \subset H_{X_{K,m}}$ and therefore $H_{X_{K,m}} = H^M$. Moreover, combining (A.56) with Proposition 4.1 in [38] and the fact that $AP_m = P_m A$, we get from Lemma A.29

$$\begin{aligned}
&\|(h_k)_{k=1}^M\|^2_{X_{K,m}} \le \|\sum_{k,l=1}^{M} (G_m)^{-1}_{k,l} P_m K_k h_l\|^2_X \\
&\le 3\|P_m \sum_{k,l=1}^{M} (G_m)^{-1}_{k,l} A K_k h_l\|^2_{L^2([0,T];\mathcal{H})} + \|P_m \sum_{k,l=1}^{M} (G_m)^{-1}_{k,l} K_k h_l\|^2_{L^2([0,T];\mathcal{H})} \\
&+ 2\|P_m \sum_{k,l=1}^{M} (G_m)^{-1}_{k,l} K_k h_l'\|^2_{L^2([0,T];\mathcal{H})}.
\end{aligned}$$

Letting m go to infinity, in which case $(G_m)^{-1}$ converges to G^{-1}, and so by definition of G_A, the last display becomes

$$\begin{aligned}
&\limsup_{m\to\infty} \|(h_k)_{k=1}^M\|^2_{X_{K,m}} \\
&\le 3\int_0^T \sum_{k,l=1}^{M} (G^{-1}G_A G^{-1})_{kl} h_k(t) h_l(t)\ \mathrm{d}t + \int_0^T \sum_{k,l=1}^{M} (G^{-1})_{kl} h_k(t) h_l(t)\ \mathrm{d}t \\
&+ 2\int_0^T \sum_{k,l=1}^{M} (G^{-1})_{kl} h_k'(t) h_l'(t)\ \mathrm{d}t.
\end{aligned}$$

Using standard results for the operator norm of symmetric matrices yields thus the upper bound claimed for $\limsup_{m\to\infty}\|(h_k)_{k=1}^M\|^2_{X_{K,m}}$ in the statement of the theorem to hold for $\|h\|^2_{X_K}$.

Next, we use the above results to compute the RKHS of $X_K = (\langle X, K_k\rangle_{\mathcal{H}})_{k=1}^M$. First, let us argue that the RKHS of a single measurement $\langle X, K_k\rangle_{\mathcal{H}}$ (as a set) equals H. Combining Girsanov's theorem for the Itô process $\langle X, K_k\rangle_{\mathcal{H}}$ in (A.6) with Feldman–Hájek's theorem, the RKHS of $\langle X, K_k\rangle_{\mathcal{H}}$ starting in zero is H_β. Adding an independent Gaussian random variable with variance greater zero, we obtain that in the stationary case $\langle X, K_k\rangle_{\mathcal{H}}$ has RKHS $H = H_{\bar\beta}$ (see also the proof of Lemma A.20). Now, consider the case $M > 1$. By Proposition 4.1 in [38], each coordinate projection maps the RKHS of X_K to the RKHS of a single measurement, thus to H by the first step. Hence, we have $H_{X_K} \subset H^M$. It remains to show the reverse inclusion $H^M = H_{X_{K,m}} \subset H_{X_K}$. To see this, note that

$$\langle X_m, K_k\rangle_{\mathcal{H}} = \sum_{j=1}^{m} \langle K_k, e_j\rangle_{\mathcal{H}} Y_j \quad \text{and} \quad \langle X, K_k\rangle_{\mathcal{H}} = \sum_{j=1}^{\infty} \langle K_k, e_j\rangle_{\mathcal{H}} Y_j,$$

so that $X_K = X_{K,m} + (X_K - X_{K,m})$ can be written as a sum of two independent processes taking values in the Hilbert space $L^2([0,T])^M$. Letting C_K and $C_{K,m}$ be the covariance operators of

X_K and $X_{K,m}$, respectively, this implies that $C_K = C_{K,m} + \widetilde{C}$ with $\widetilde{C}$ self-adjoint and positive. Combining this with the characterisation of the RKHS norm in Proposition 2.6.8 of [22], we get $\|(h_k)_{k=1}^M\|_{X_K}^2 \leq \|(h_k)_{k=1}^M\|_{X_{K,m}}^2$ and $H_{X_{K,m}} \subset H_{X_K}$ for all $m \geq 1$. Finally, inserting the upper bound on $\|(h_k)_{k=1}^M\|_{X_{K,m}}^2$ derived above, the proof is complete. ■

Acknowledgements We are grateful for the helpful comments from three anonymous referees. The research of AT and MW has been partially funded by the Deutsche Forschungsgemeinschaft (DFG)- Project-ID 318763901 - SFB1294. AT further acknowledges financial support of Carlsberg Foundation Young Researcher Fellowship grant CF20-0604. RA gratefully acknowledges support by the European Research Council, ERC grant agreement 647812 (UQMSI).

References

[1] S. Alonso, M. Stange, and C. Beta. “Modeling random crawling, membrane deformation and intracellular polarity of motile amoeboid cells”. In: *PloS one* 13.8 (2018), e0201977.

[2] R. Altmeyer, T. Bretschneider, J. Janák, and M. Reiß. “Parameter Estimation in an SPDE Model for Cell Repolarisation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1 (2022), pp. 179–199.

[3] R. Altmeyer, I. Cialenco, and G. Pasemann. “Parameter estimation for semilinear SPDEs from local measurements”. In: *Bernoulli* 29.3 (2023), pp. 2035–2061.

[4] R. Altmeyer and M. Reiß. “Nonparametric estimation for linear SPDEs from local measurements”. In: *Annals of Applied Probability* 31.1 (2021), pp. 1–38.

[5] H. Amann. *Linear and quasilinear parabolic problems*. Volumne I: Abstract Linear Theory. Birkhäuser, 1995.

[6] T. Aspelmeier, A. Egner, and A. Munk. “Modern statistical challenges in high-resolution fluorescence microscopy”. In: *Annual Reviews of Statistics and Its Applications* 2 (2015), pp. 163–202.

[7] A. S. Backer and W. E. Moerner. “Extending Single-Molecule Microscopy Using Optical Fourier Processing”. In: *The Journal of Physical Chemistry B* 118.28 (2014), pp. 8313–8329.

[8] F. E. Benth, D. Schroers, and A. E. Veraart. “A weak law of large numbers for realised covariation in a Hilbert space setting”. In: *Stochastic Processes and their Applications* 145 (2022), pp. 241–268.

[9] M. Bibinger and M. Trabs. “Volatility estimation for stochastic PDEs using high-frequency observations”. In: *Stochastic Processes and their Applications* 130.5 (2020), pp. 3005–3052.

[10] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2011.

[11] F. Catania, M. Massabò, and O. Paladino. “Estimation of Transport and Kinetic Parameters Using Analytical Solutions of the 2D Advection-Dispersion-Reaction Model”. In: *Environmetrics* 17.2 (2006), pp. 199–216.

[12] C. Chong. “High-frequency analysis of parabolic stochastic PDEs”. In: *Annals of Statistics* 48.2 (2020), pp. 1143–1167.

[13] I. Cialenco, F. Delgado-Vences, and H.-J. Kim. “Drift estimation for discretely sampled SPDEs”. In: *Stochastics and Partial Differential Equations: Analysis and Computations* 8 (2020), pp. 895–920.

[14] I. Cialenco, H.-J. Kim, and G. Pasemann. “Statistical analysis of discretely sampled semilinear SPDEs: a power variation approach”. In: *Stochastics and Partial Differential Equations: Analysis and Computations* (2023).

[15] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.

[16] A. Debussche, S. de Moor, and M. Hofmanová. “A Regularity Result for Quasilinear Stochastic Partial Differential Equations of Parabolic Type”. In: *SIAM Journal on Mathematical Analysis* 47.2 (2015), pp. 1590–1614.

[17] A. Egner, C. Geisler, and R. Siegmund. “STED Nanoscopy”. In: *Nanoscale Photonic Imaging*. Ed. by T. Salditt, A. Egner, and D. R. Luke. Topics in Applied Physics. Springer, 2020, pp. 3–34.

[18] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer, 2000.

[19] L. C. Evans. *Partial Differential Equations*. American Mathematical Soc., 2010.

[20] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. CRC Press, 2015.

[21] S. Gaudlitz and M. Reiß. “Estimation for the reaction term in semi-linear SPDEs under small diffusivity”. In: *Bernoulli* 29.4 (2023), pp. 3033–3058.

[22] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Vol. 40. Cambridge University Press, 2016.

[23] S. Gugushvili, A. Van Der Vaart, and D. Yan. “Bayesian linear inverse problems in regularity scales”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 56.3 (2020), pp. 2081–2107.

[24] M. Hairer. *An Introduction to Stochastic PDEs*. 2023. arXiv: 0907.4178 [math.PR].

[25] F. Hildebrandt and M. Trabs. “Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations”. In: *Stochastic Processes and their Applications* 162 (2023), pp. 171–217.

[26] F. Hildebrandt and M. Trabs. “Parameter estimation for SPDEs based on discrete observations in time and space.” In: *Electronic Journal of Statistics* 15 (2021), pp. 2716–2776.

[27] M. Huebner and B. Rozovskii. “On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE’s”. In: *Probability Theory and Related Fields* 103.2 (1995), pp. 143–163.

[28] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[29] M. Karalashvili, S. Groß, W. Marquardt, A. Mhamdi, and A. Reusken. “Identification of Transport Coefficient Models in Convection-Diffusion Equations”. In: *SIAM Journal on Scientific Computing* 33.1 (2011), pp. 303–327.

[30] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.

[31] T. Kato. *Perturbation theory for linear operators*. Springer, 1995.

[32] P. Kriz and B. Maslowski. “Central Limit Theorems and Minimum-Contrast Estimators for Linear Stochastic Evolution Equations”. In: *Stochastics* 91.8 (2019), pp. 1109–1140.

[33] U. Küchler and M. Sørensen. *Exponential families of stochastic processes*. Springer, 1997.

[34] A. Kukush. *Gaussian Measures in Hilbert Space: Construction and Properties*. Wiley, 2020.

[35] G. Kulaitis, A. Munk, and F. Werner. *What is resolution? A statistical minimax testing perspective on super-resolution microscopy*. 2020. arXiv: 2005.07450 [math.ST].

[36] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2013.

[37] M. Kwaśnicki. “Ten equivalent definitions of the fractional Laplace operator”. In: *Fractional Calculus and Applied Analysis* 20.1 (2017), pp. 7–51.

[38] M. Lifshits. *Lectures on Gaussian processes*. Springer, 2012.

[39] R. Liptser and A. Shiryaev. *Statistics of Random Processes I*. General Theory. Springer, 2001.

[40] W Liu and M Röckner. *Stochastic Partial Differential Equations: An Introduction*. Springer, 2015.

[41] X. Liu, K. Yeo, and S. Lu. “Statistical Modeling for Spatio-Temporal Data From Stochastic Convection-Diffusion Processes”. In: *Journal of the American Statistical Association* 117.539 (2021), pp. 1482–1499.

[42] F. P. Llopis, N. Kantas, A. Beskos, and A. Jasra. “Particle filtering for stochastic Navier-Stokes signal observed with linear additive noise”. In: *SIAM Journal on Scientific Computing* 40.3 (2018), A1544–A1565.

[43] R. Lockley. “Image-based Modelling of Cell Reorientation”. PhD Thesis. University of Warwick, 2017.

[44] S. V. Lototsky. “Parameter Estimation for Stochastic Parabolic Equations: Asymptotic Properties of a Two-Dimensional Projection-Based Estimator”. In: *Statistical Inference for Stochastic Processes* 6.1 (2003), pp. 65–87.

[45] C. H. Luce, D. Tonina, F. Gariglio, and R. Applebee. “Solutions for the Diurnally Forced Advection-Diffusion Equation to Estimate Bulk Fluid Velocity and Diffusivity in Streambeds from Temperature Time Series”. In: *Water Resources Research* 49.1 (2013), pp. 488–506.

[46] A. Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, 1995.

[47] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Vol. 44. Springer, 1983.

[48] S. C. Reddy and L. N. Trefethen. “Pseudospectra of the convection-diffusion operator”. In: *SIAM Journal on Applied Mathematics* 54.6 (1994), pp. 1634–1649.

[49] M. Reiß. “Asymptotic equivalence for inference on the volatility from noisy observations”. English. In: *Ann. Stat.* 39.2 (Apr. 2011), pp. 772–802.

[50] M. Sauer and W. Stannat. “Analysis and approximation of stochastic nerve axon equations”. English. In: *Mathematics of Computation* 85.301 (2016), pp. 2457–2481.

[51] S.-J. Sheu. “Some estimates of the transition density of a nondegenerate diffusion markov process”. In: *The Annals of Probability* 19 (1991), pp. 538–561.

[52] N. Shimakura. *Partial Differential Operators of Elliptic Type*. American Mathematical Soc., 1992.

[53] F. Sigrist, H. R. Künsch, and W. A. Stahel. "Stochastic partial differential equation based modelling of large space–time data sets". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.1 (2015), pp. 3–33.

[54] Y. Tonaki, Y. Kaino, and M. Uchida. "Parameter estimation for linear parabolic SPDEs in two space dimensions based on high frequency data". In: *Scandinavian Journal of Statistics* 50.4 (2023), pp. 1568–1589.

[55] H. Triebel. *Theory of function spaces*. Birkhäuser, 1983.

[56] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

[57] H. C. Tuckwell. "Stochastic partial differential equations in neurobiology: Linear and nonlinear models for spiking neurons". In: *Stochastic biomathematical models: with applications to neuronal modeling*. Ed. by M. Bachar, J. Batzel, and S. Ditlevsen. Springer, 2013, pp. 149–173.

[58] A. W. van der Vaart, J. H. van Zanten, and others. "Rates of contraction of posterior distributions based on Gaussian process priors". In: *The Annals of Statistics* 36.3 (2008). Publisher: Institute of Mathematical Statistics, pp. 1435–1463.

[59] M. J. Wainwright. *High-dimensional statistics*. Cambridge University Press, 2019.

[60] J. B. Walsh. "A stochastic model of neural response". In: *Advances in Applied Probability* 13.2 (1981), pp. 231–281.

[61] W. Whitt. "Proofs of the martingale FCLT". In: *Probability Surveys* 4 (2007), pp. 268–302.

[62] A. Yagi. *Abstract Parabolic Evolution Equations and their Applications*. Springer Science & Business Media, 2009.

Nonparametric velocity estimation in stochastic convection-diffusion equations from multiple local measurements

B

Claudia Strauch and Anton Tiepner

Abstract

We investigate pointwise estimation of the function-valued velocity field of a second-order linear SPDE. Based on multiple spatially localised measurements, we construct a weighted augmented MLE and study its convergence properties as the spatial resolution of the observations tends to zero and the number of measurements increases. By imposing Hölder smoothness conditions, we recover the pointwise convergence rate known to be minimax-optimal in the linear regression framework. The optimality of the rate in the current setting is verified by adapting the lower bound ansatz based on the RKHS of local measurements to the nonparametric situation.

B.1 Introduction

Stochastic partial differential equations (SPDEs) are an appealing tool to model spatio-temporal data. They describe the evolution of dynamical systems and can be utilised in almost all areas of natural sciences, finance, economics, and many more applied disciplines. By including random forcing terms, SPDEs also account for microscopic scaling limits or model misspecification. We will focus on the important subclass of *stochastic convection-diffusion* or *advection-diffusion equations* which can also serve as a basis for more complex models. They describe the movement of quantities (such as particles, heat, energy, etc.) in a physical system and find applications in, but are not limited to, weather forecasts [31, 39, 40], neuronal responses [44, 45], solar radiation [10], air quality [30], sediment concentrations [41], biomass distributions [13], groundwater flows [38], and term structure movements [11].

More specifically, for a finite time horizon T, we consider the solution $X = (X(t))_{0\leq t\leq T}$ to the linear parabolic SPDE

$$\begin{cases} \mathrm{d}X(t) = A_\vartheta X(t)\,\mathrm{d}t + \mathrm{d}W(t), & t \in (0,T], \\ X(0) = X_0 \in L^2(\Lambda), & \\ X(t)|_{\partial\Lambda} = 0, & t \in (0,T], \end{cases} \tag{B.1}$$

on a bounded open domain $\Lambda \subset \mathbb{R}^d$ with C^2-boundary $\partial\Lambda$, Dirichlet boundary conditions, driven by a cylindrical Brownian motion $W = (W(t))_{0\leq t\leq T}$. The second-order elliptic operator A_ϑ appearing in (B.1) is specified as

$$A_\vartheta z = a\Delta z + \vartheta \cdot \nabla z + cz, \tag{B.2}$$

where $a > 0$, ϑ and c represent the (constant) diffusivity, the velocity field and the reaction coefficient, inducing spatial diffusion, transportation and damping, respectively. While the analytical theory of SPDEs is well understood and established, see, e.g., [12, 20, 29, 33], the literature on their statistical aspects is somewhat limited, and many research questions are still open. As a concrete example, to the best of our knowledge, estimation of a function-valued velocity has not yet been investigated. We want to fill in this gap by estimating the function-valued velocity field ϑ by means of nonparametric methods based on local measurements.

Parameter estimation for SPDEs is widely studied in the literature, but primarily devoted to a scalar parameter in $A_\vartheta = \vartheta A + B$ for some (non-) linear operators A and B. When A and B share a common system of eigenvectors and are self-adjoint, [23] constructed a maximum likelihood estimator (MLE) for ϑ, relying on N spectral measurements $(\langle X(t), e_i\rangle)_{0\le t\le T}$, $i = 1,\dots,N$, with an eigenbasis $(e_i)_{i\ge 1}$. Given some relation between the differential order of A and B and the dimension d, the derived MLE was shown to be consistent and asymptotically normal. This so-called spectral approach was subsequently adapted and extended to different settings, such as nonlinear SPDEs [8, 36], fractional noise [9], or joint parameter estimation [34]. However, the majority of these studies considered the case where ϑ specifies the highest order operator, i.e., $\mathrm{ord}(A) > \mathrm{ord}(B)$, and there is no known estimator for a constant transport coefficient ϑ in (B.2) in the spectral approach setting. Based on discrete observations $X(t_j, x_i)$ in time and space, [22, 26, 42] analysed power variations and contrast estimators in dimension one and two for all occurring quantities in the parametric version of (B.2). Reaction or source-sink terms have been studied, for example, by [17, 21, 24]. We refer to [7] for a detailed overview of further related literature.

In this paper, we construct a pointwise estimator $\widehat{\vartheta}(x)$ for the velocity field ϑ, evaluated at the spatial location $x \in \Lambda$ from local measurement processes for multiple locations, i.e., our data are given by observing the solution to (B.1) locally in space and continuously in time. Given some fixed function $K \in H^2(\mathbb{R}^d)$ with compact support, we consider points $x_1,\dots,x_N \in \Lambda$ and a resolution level $\delta > 0$ small enough such that the localised functions K_{δ,x_k}, defined through

$$K_{\delta,x_k}(x) := \delta^{-d/2} K(\delta^{-1}(x - x_k)), \quad x \in \Lambda, \quad k = 1,\dots,N,$$

are supported in Λ. In optical systems, they are known as *point spread functions* [5, 6], and they describe the physical limitation that $X(t_j, x_i)$ can only be measured up to some locally blurred average, i.e., a convolution with the point spread function. Specifically, the local measurements of X are given as the continuously observed processes $X_\delta, X_\delta^\Delta \in L^2([0,T];\mathbb{R}^N)$ and $X_\delta^\nabla \in L^2([0,T];\mathbb{R}^{d\times N})$, where, for $k = 1,\dots,N$,

$$\begin{aligned} (X_\delta)_k &= X_{\delta,k} = (\langle X(t), K_{\delta,x_k}\rangle)_{0\le t\le T}, \\ (X_\delta^\nabla)_k &= X_{\delta,k}^\nabla = (\langle X(t), \nabla K_{\delta,x_k}\rangle)_{0\le t\le T}, \\ (X_\delta^\Delta)_k &= X_{\delta,k}^\Delta = (\langle X(t), \Delta K_{\delta,x_k}\rangle)_{0\le t\le T}. \end{aligned} \tag{B.3}$$

Local measurements were introduced by [4]. There, the authors investigated the estimation of a nonparametric diffusivity $a(x)$ and demonstrated that it can already be estimated with the parametric minimax rate δ upon observing the local information $X_{\delta,x}$. The method proved to be robust to low-order nonlinearities, cf. [2, 3], and was used in a direct application to cell repolarisation, estimating the diffusivity of the activator in the stochastic Meinhardt model [2]. Adapting the extended MLE approach of [4], Paper A considered the fully anisotropic parametric version of (B.2), addressing joint estimation of diffusivity, velocity, and reaction components. In particular, it has been shown that transport and damping coefficients cannot be estimated consistently in finite time if the number $N = N(\delta)$ of local measurements remains finite. If the number of measurements is chosen to be maximal, i.e., $N \asymp \delta^{-d}$, the derived convergence rates agree with the convergence rates obtained with the spectral approach of [23]. In the case of the transport coefficient ϑ, the convergence rate $N^{-1/2}$ has been proven to be optimal.

Let us briefly describe our main findings. We combine the approach of Paper A with techniques from nonparametric regression and local likelihood estimation. The contribution of each measurement is individually weighted and controlled by a bandwidth $h = h(\delta)$ to account for bias reduction. The obtained *weighted augmented MLE* $\widehat{\vartheta}_\delta(x)$ is consistent, and under appropriate Hölder smoothness assumptions, it satisfies, for $x \in \Lambda$,

$$\widehat{\vartheta}_\delta(x) - \vartheta(x) = O_{\mathbb{P}}(h^\beta + (Nh^d)^{-1/2}), \quad \beta \in (1, 2]. \tag{B.4}$$

Optimising (B.4) with respect to h, we obtain the convergence rate $N^{-\beta/(2\beta+d)}$ known from local linear regression estimation. This convergence rate turns out to be optimal, as we demonstrate by adapting the lower bound ansatz of Paper A, which is based on the RKHS of our local measurements and its relation to the Hellinger distance, to the nonparametric framework.

The paper is structured as follows. We specify the model and construct the estimator in Section B.2. Section B.3 provides upper bounds on the pointwise risk of the estimator, along with a discussion of the involved assumptions and a number of examples and applications. Lower bounds are stated in Section B.4. All proofs are deferred to Section B.5.

Notation Throughout this paper, the time horizon $T < \infty$ is fixed, and we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$. We write $a \lesssim b$ if $a \le Mb$ holds for a universal constant M not depending on δ, N, h, or a spatial location $x \in \Lambda$, and $a \lesssim_s b$ if $a \le Cb$ with a constant C explicitly depending on the quantity s. Unless otherwise stated, all limits are to be understood as the spatial resolution level $\delta \to 0$. For an open set $U \subset \mathbb{R}^d$, $L^2(U)$ is the usual L^2 space with inner product $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{L^2(U)}$. The Euclidean inner product and distance of two vectors $a, b \in \mathbb{R}^p$ are denoted by $a \cdot b$ and $|a - b|$, respectively. Let $H^k(U)$ denote the usual Sobolev spaces, and denote by $H_0^1(U)$ the completion of $C_c^\infty(U)$, the space of smooth compactly supported functions, relative to the $H^1(U)$ norm. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$, let D^α be the α-th weak partial derivative operator of order $|\alpha| = \alpha_1 + \cdots + \alpha_d$. The gradient, divergence and Laplace operator are denoted by ∇, $\nabla\cdot$ and Δ, respectively. For $\beta > 0$, denote by $\mathcal{H}(\beta)$ the space of functions $f \colon \Lambda \to \mathbb{R}$ with continuous derivatives up to order $\lfloor \beta \rfloor$ such that their $\lfloor \beta \rfloor$-th partial derivatives are Hölder continuous with exponent $\beta - \lfloor \beta \rfloor \le 1$.

B.2 Pointwise estimation approach

Our interest is in estimating the velocity coefficient appearing in the second-order linear elliptic differential operator A_ϑ as introduced in (B.2) with domain $H_0^1(\Lambda) \cap H^2(\Lambda)$. For $z \in H_0^1(\Lambda) \cap H^2(\Lambda)$, its adjoint A_ϑ^* is defined by

$$A_\vartheta^* z := a\Delta z - \nabla \cdot \vartheta z + cz = a\Delta z - \vartheta \cdot \nabla z + (c - \nabla \cdot \vartheta)z.$$

Both A_ϑ and A_ϑ^* generate analytic semigroups, denoted by $(S_\vartheta(t))_{t\ge0}$ and $(S_\vartheta^*(t))_{t\ge0}$, respectively, on $L^2(\Lambda)$. The weak solution to (B.1) is given by

$$X(t) = S_\vartheta(t)X_0 + \int_0^t S_\vartheta(t - t')\, dW(t').$$

As discussed in [4, Proposition 2.1], it only takes values in negative-order Sobolev spaces, but still allows the definition of real-valued centred Gaussian processes $(\langle X(t), z\rangle)_{0\le t\le T, z\in L^2(\Lambda)}$, satisfying for $z \in H_0^1(\Lambda)\cap H^2(\Lambda)$

$$\langle X(t), z\rangle = \langle X_0, z\rangle + \int_0^t \langle X(t'), A_\vartheta^* z\rangle \,\mathrm{d}t' + \langle W(t), z\rangle, \quad 0 \le t \le T. \tag{B.5}$$

Our nonparametric analysis relies on Hölder smoothness conditions. Let $\beta \in (1, 2]$. We assume that the (possibly unknown) diffusion coefficient $a > 0$ is constant, each component $\vartheta_i \colon \Lambda \to \mathbb{R}$, $i = 1, \dots, d$, of the velocity field ϑ is contained in $\mathcal{H}(\beta)$, and the (nuisance) reaction function $c\colon \Lambda \to \mathbb{R}$ belongs to $\mathcal{H}(\beta - 1)$. Since the differential operator A_ϑ^* contains the first-order derivative of ϑ, we require for existence reasons, cf. also [4, Proposition 3.5], that ϑ is continuously differentiable, i.e., $\beta > 1$.

Recall that we work in a local measurements framework, i.e., we construct an estimator based on the observations (B.3). Let $W_k(t) := \langle W(t), K_{\delta,x_k}\rangle \|K\|^{-1}_{L^2(\mathbb{R}^d)}$ be scalar Brownian motions. Each local measurement forms an Itô process with initial condition $X_{\delta,k}(0) = \langle X_0, K_{\delta,x_k}\rangle$ and, using (B.5),

$$\mathrm{d}X_{\delta,k}(t) = \langle X(t), A_\vartheta^* K_{\delta,x_k}\rangle \,\mathrm{d}t + \|K\|_{L^2(\mathbb{R}^d)} \,\mathrm{d}W_k(t), \quad k = 1, \dots, N. \tag{B.6}$$

Before constructing an estimator for $\vartheta(x)$, we give a brief recap on the construction of local polynomial log-likelihood estimators. The following is based on [32]. We also refer to [15, 16, 37] for further discussion of the local likelihood approach and to the monograph [46] for an overview of general nonparametric techniques. Suppose we observe response variables

$$Y_i \sim f(\cdot, \mu(x_i)), \quad i = 1, \dots, n,$$

with density f depending on the design points $x_i \in \mathbb{R}$ via an unknown function μ. Simple examples are given by nonparametric regression, where $Y_i = \mu(x_i) + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, or logistic regression, where $P(Y_i = 1) = p(x_i)$, $P(Y_i = 0) = 1 - p(x_i)$, and we consider the link function $\mu(x_i) = \log(p(x_i)/(1 - p(x_i))$. Assuming that $\mu(x)$ has a polynomial fit of degree $p \in \mathbb{N}_0$, i.e., by a p-th order Taylor approximation,

$$\mu(x_i) \approx \sum_{j=0}^{p} \frac{a_j (x_i - x)^j}{j!} = \langle a, A(x_i - x)\rangle_{\mathbb{R}^{p+1}},$$

where $a = (a_0, \dots, a_p)^\top$, $A(y) = (1, y, \dots, y^p/p!)^\top$, the basic idea is to maximise the local polynomial log-likelihood

$$l(\mu, x) = \sum_{i=1}^{n} w_i(x) \log(f(Y_i, \langle a, A(x_i - x)\rangle_{\mathbb{R}^{p+1}})) \tag{B.7}$$

over $a \in \mathbb{R}^{p+1}$ and with weight functions w_i, $i = 1, \dots, n$. The estimator for $\mu(x)$ is then given by $\widehat{\mu}(x) = \widehat{a}_0$. For a smoothing parameter h (bandwidth), only observations within a given window $(x - h, x + h)$ are used, and each observation in (B.7) is weighted by $w_i(x) = \mathcal{W}\big(\frac{x_i - x}{h}\big)$. Often, $\mathcal{W}$ is chosen as a positive kernel function, but in principle it can be more general. This approach can be extended to the multivariate case (cf., amongst others, [1, 16, 37]), and is also close to local polynomial regression (cf. [19, 43]) as a generalisation of it. We will adapt the above method in the next section to construct a nonparametric estimator for $\vartheta(x)$.

The weighted augmented MLE The local observation processes X_δ, X_δ^∇ and X_δ^Δ as introduced in (B.3) are no longer Markovian, as the time evolution at the point x_k, $k \in \{1, \ldots, N\}$, depends on the entire spatial structure of X. Therefore, a general Girsanov theorem for multivariate Itô processes, cf. [28, Section 7.6], results in the modified log-likelihood

$$\|K\|_{L^2(\mathbb{R}^d)}^{-1} \sum_{k=1}^N \left(\int_0^T \langle X(t), A_\vartheta^* K_{\delta,x_k} \rangle \, \mathrm{d}X_{\delta,k}(t) - \frac{1}{2} \int_0^T \langle X(t), A_\vartheta^* K_{\delta,x_k} \rangle^2 \, \mathrm{d}t \right), \tag{B.8}$$

provided the driving Brownian motions W_k in (B.6) are independent. For parametric ϑ, it is straightforward to derive an estimate based on the observed processes by maximising (B.8), as shown in Paper A. In our set-up, we assume instead that we can approximate ϑ locally by a constant, i.e., for some $\gamma \in \mathbb{R}^d$,

$$\langle X(t), A_\vartheta^* K_{\delta,x_k} \rangle \approx a X_{\delta,k}^\Delta(t) + \gamma^\top X_{\delta,k}^\nabla(t), \quad 0 \le t \le T.$$

Note that approximations by a polynomial of degree $p \ge 1$ result in additional observations, which we do not have access to and which cannot be recovered by convolution and a finite difference scheme, see Remark B.1 below. Therefore, we restrict our investigations to the local constant approximation.

Note further that we cannot use the local likelihood approach introduced before directly since ϑ is incorporated in $\langle X(t), A_\vartheta^* K_{\delta,x_k} \rangle$ via $\langle X(t), \nabla \cdot \vartheta K_{\delta,x_k} \rangle \neq \vartheta(x_k)^\top X_{\delta,k}^\nabla(t)$. Instead, we adapt the underlying idea by weighting the contribution of the k-th summand individually. Hence, we maximise

$$l_\delta(\gamma, x) = \|K\|_{L^2(\mathbb{R}^d)}^{-1} \sum_{k=1}^N w_k(x) \bigg(\int_0^T \Big(a X_{\delta,k}^\Delta(t) - \gamma^\top X_{\delta,k}^\nabla(t) \Big) \, \mathrm{d}X_{\delta,k}(t) - \frac{1}{2} \int_0^T \Big(a X_{\delta,k}^\Delta(t) - \gamma^\top X_{\delta,k}^\nabla(t) \Big)^2 \, \mathrm{d}t \bigg)$$

over $\gamma \in \mathbb{R}^d$ to derive the *weighted augmented MLE*, given by

$$\widehat{\vartheta}_\delta(x) = -(\mathcal{I}_\delta^x)^{-1} \sum_{k=1}^N w_k(x) \left(\int_0^T X_{\delta,k}^\nabla(t) \, \mathrm{d}X_{\delta,k}(t) - \int_0^T a X_{\delta,k}^\Delta(t) X_{\delta,k}^\nabla(t) \, \mathrm{d}t \right) \tag{B.9}$$

with the *weighted observed Fisher information*

$$\mathcal{I}_\delta^x = \sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\nabla(t) X_{\delta,k}^\nabla(t)^\top \, \mathrm{d}t.$$

Remark B.1 (higher order approximations). Intuitively, approximating ϑ by a polynomial of higher order should perform an automatic bias correction and should thus improve the quality of the estimation. However, due to the spatial influence of ϑ in $\langle X(t), A_\vartheta^* K_{\delta,x_k} \rangle$, i.e., in

$$\langle X(t), \nabla \cdot \vartheta K_{\delta,x_k} \rangle,$$

the log-likelihood (B.8) depends not only on the pointwise evaluations $\vartheta(x_k)$, $k = 1, \ldots, N$, but rather on ϑ and $\nabla \cdot \vartheta$ in a neighbourhood around x_k. While the processes $X_{\delta,k}^\nabla$ and $X_{\delta,k}^\Delta$ can, in principle, be obtained by observing $X_{\delta,y}(t)$ in a neighbourhood of x_k, cf. [4], this fails to hold true for the additional observations required for higher order approximations. Even in the simplest case, i.e., a linear approximation of the form $\vartheta(y) = \gamma + \Gamma(y - x)$, one obtains

$$\langle X(t), \nabla \cdot (\gamma + \Gamma(\cdot - x))K_{\delta,x_k}\rangle = \gamma^\top X^\nabla_{\delta,k}(t) + \langle X(t), (\Gamma(\cdot - x))^\top \nabla K_{\delta,x_k}\rangle + \mathrm{tr}(\Gamma) X_{\delta,k}(t).$$

Since K_{δ,x_k} takes only non-zero values in a neighbourhood around x_k, we could instead approximate the non-observable term on the right hand side of the last display, while also ignoring the lower order perturbation, by

$$\langle X(t), \nabla \cdot (\gamma + \Gamma(\cdot - x))K_{\delta,x_k}\rangle \approx \gamma^\top X^\nabla_{\delta,k}(t) + (\Gamma(x_k - x))^\top X^\nabla_{\delta,k}(t).$$

Extending this idea to arbitrary polynomial approximations yields an estimate of ϑ and its partial derivatives at point x. The analysis of this estimator, however, is nonstandard and seems to provide only limited, if any, improvement over $\widehat{\vartheta}_\delta(x)$ in (B.9) as its resulting bias component also depends on the approximation error within the accessible and inaccessible observations which restricts the usage of arbitrary Hölder regularity.

B.3 Convergence in probability: Upper bound results

Plugging (B.6) into (B.9) yields the decomposition

$$\widehat{\vartheta}_\delta(x) = \vartheta(x) + (\mathcal{I}^x_\delta)^{-1}\mathcal{R}^x_\delta - (\mathcal{I}^x_\delta)^{-1}\mathcal{M}^x_\delta \|K\|_{L^2(\mathbb{R}^d)}, \tag{B.10}$$

where the martingale term and the remainder, respectively, are specified as

$$\mathcal{M}^x_\delta = \sum_{k=1}^N w_k(x) \int_0^T X^\nabla_{\delta,k}(t)\, \mathrm{d}W_k(t),$$
$$\mathcal{R}^x_\delta = \sum_{k=1}^N w_k(x) \int_0^T X^\nabla_{\delta,k}(t) \langle X(t), ((\vartheta - \vartheta(x)) \cdot \nabla + \varphi_\vartheta)K_{\delta,x_k}\rangle\, \mathrm{d}t,$$

with $\varphi_\vartheta := \nabla \cdot \vartheta - c \in \mathcal{H}(\beta - 1)$. The following assumption gathers technical conditions required for our statistical analysis.

Assumption L. Assume that the following conditions are satisfied:

(i) The locations x and $x_1, \dots, x_N$ belong to a fixed compact set $\mathcal{J} \subset \Lambda$, which is independent of the resolution δ and N. There exists $\delta' > 0$ such that $\mathrm{supp}(K_{\delta,x_k}) \cap \mathrm{supp}(K_{\delta,x_l}) = \emptyset$ for $k \neq l$, $k, l \leq N$, and all $\delta \leq \delta'$.

(ii) There exists a compactly supported function $\bar{K} \in H^4(\mathbb{R}^d)$, which is either even or odd, such that $K = (-\Delta)\bar{K}$.

(iii) Given $h > 0$, there exist weight functions $w_k = w_k(N, h, x_1, \dots, x_N) \colon \mathcal{J} \to \mathbb{R}$, depending only on $N, h, x_1, \dots, x_N$, fulfilling the following conditions for a universal constant C_* not depending on N, δ, h and x:

  (1) $\sup_{k \in [N], x \in \mathcal{J}} |w_k(x)| \leq C_*(Nh^d)^{-1}$;

(2) $\sum_{k=1}^{N} |w_k(x)| \le C_*$;

(3) $w_k(x) = 0$ if $|x_k - x| > h$;

(4) $\sum_{k=1}^{N} w_k(x) = 1$, and for any α with $|\alpha| = 1$, it holds $\sum_{k=1}^{N}(x_k - x)^\alpha w_k(x) = 0$.

(iv) The initial condition X_0 is such that either $X_0 \in L^p(\Lambda) \cap \mathcal{D}(A_\vartheta)$, $p > 2$, or, if in addition there exists a constant $\gamma < 0$ such that $c - \nabla \cdot \vartheta \le \gamma$, $X_0 = \int_{-\infty}^{0} S_\vartheta(t')\,\mathrm{d}W(t')$.

A few comments on the above conditions are in order. The support condition in Assumption L(i) guarantees that the Brownian motions W_k in (B.6) are independent as $\delta \to 0$, while the processes $X_{\delta,k}$ defined in (B.3) do not inherit independence. It requires the measurement locations x_k to be separated by a Euclidean distance of at least $C\delta$ for a fixed constant C, which means that N grows at most as $N = O(\delta^{-d})$. Existence of weight functions w_k in Assumption L(iii) holds true under standard structural assumptions on the locations x_k (see Lemma B.7 and the subsequent remark below). Since the partial derivatives $\partial_i K$, $i = 1, \ldots, d$, are mutually independent, condition (4) also implies that $\mathcal{I}_\delta^x$ is $\mathbb{P}$-a.s. invertible, which can be deduced from Lemma A.1 in Paper A. The weights can be constructed similarly to weights in local polynomial regression, cf. for instance [43, Chapter 1.6], such that they are reproducing of order one. Assumption L(iv) guarantees that a general initial condition is asymptotically neglectable. If $\vartheta = 0$, it can be further relaxed such that $\gamma = 0$ is allowed, i.e., $X_0 = \int_{-\infty}^{0} S_\vartheta(t')\,\mathrm{d}W(t')$ is, for instance, also valid for $A_\vartheta = a\Delta$. Despite using the local constant (LP(0)) approach, we will show that $\widehat{\vartheta}_\delta(x)$ achieves the convergence rate of an LP(1)-estimator. While in local polynomial regression, this is known to happen for the Nadaraya–Watson estimator if, for instance, one works with equidistant design points x_k and estimates at one of those locations, see Example B.8, we only rely on a first-order multivariate Taylor expansion and use the reproducing property of the weights as well as (anti-) symmetry of ∇K, implied by Assumption L(ii). Depending on more information about the initial condition and the dimension d, Assumption L(ii) can also be softened.

A precise control of the error decomposition (B.10) results in consistency of the estimate $\widehat{\vartheta}_\delta(x)$ as the resolution level δ tends to zero. As known from the parametric case, cf. Paper A, consistent estimation in finite time T of the velocity ϑ naturally requires $N = N(\delta) \to \infty$ as $\delta \to 0$. On the other hand, the bandwidth $h \to 0$ is usually chosen in dependence on the number N of observations to balance between bias and variance terms. In that case, h is implicitly also related to δ.

THEOREM B.2. *Under Assumption L, the weighted augmented MLE satisfies*

$$\widehat{\vartheta}_\delta(x) - \vartheta(x) = O_{\mathbb{P}}(h^\beta + (Nh^d)^{-1/2}), \quad \beta \in (1, 2]. \tag{B.11}$$

In particular, this bound is independent of the spatial location $x \in \mathcal{J}$ in the sense that, for any $\varepsilon > 0$, there exist some $M > 0$, $\delta' > 0$ such that, for any $x \in \mathcal{J}$ and for any $\delta \le \delta'$, we have

$$\mathbb{P}\Big(\big|\widehat{\vartheta}_\delta(x) - \vartheta(x)\big|\big(h^\beta + (Nh^d)^{-1/2}\big)^{-1} > M\Big) \le \varepsilon. \tag{B.12}$$

To achieve consistency in the first place, the above result implies that $Nh^d \to \infty$ is required. Hence, it can only hold if $\delta \ll h$, since Assumption L(i) imposes at most $N \asymp \delta^{-d}$ measurement locations.

Remark B.3 (convergence rate). Optimising the upper bound stated in (B.11) with respect to the bandwidth h yields

$$h \asymp N^{-1/(2\beta+d)}, \quad \text{that is,} \quad h^\beta \asymp N^{-\beta/(2\beta+d)}, \quad \beta \in (1,2], \tag{B.13}$$

thus matching the standard rates for the mean-squared error in nonparametric regression. The usual bias-variance trade-off, resulting from choosing suboptimal h, is illustrated in Figure B.1. For a maximal choice $N \asymp \delta^{-d}$, the optimal bandwidth specification gives

$$h \asymp \delta^{d/(2\beta+d)}, \quad \text{that is,} \quad h^\beta \asymp \delta^{\beta d/(2\beta+d)}, \quad \beta \in (1,2]. \tag{B.14}$$

A graphical illustration in $d = 1$ for $\beta = 2$, i.e., $h^\beta \asymp \delta^{2/5}$, is given in Figure B.1. As demonstrated in Section B.4, the rates in (B.13) and (B.14) are optimal.

Naturally, one may ask if the rate in (B.13) also holds true under higher order Hölder regularity assumptions. Indeed, Theorem B.2 might, in principle, be extended to arbitrary $\beta > 2$, using reproducing weights functions w_k of order $\lfloor \beta \rfloor$ instead. The analysis of the remainder $\mathcal{R}_\delta^x$ in Section B.5.3, however, indicates that its order is not determined by the bandwidth h and smoothness parameter β alone, yet also dependent on the resolution level δ. In particular,

$$\mathcal{R}_\delta^x = O_{\mathbb{P}}(h^\beta + \delta h + \delta^2).$$

Thus, the dominating term varies, depending on the dimension d and the assumed smoothness β. If $\beta \le 2$, the remainder is always of order h^β whilst the parametric order δ^2 can be achieved for $d \to \infty$ and arbitrary $\beta \ge 2$. This matches the observations made in [4] or Paper A, where the bias term does neither depend on the time horizon T, the diffusivity a, nor the number of spatial observations N. As a consequence, arbitrary $\beta > 1$ allow for the dimension-improving convergence rates $\delta^{\beta d/(2\beta+d)} \vee \delta^2$. This phenomenon, however, is no contradiction to the curse of dimensionality stated in (B.13) as it results by reparametrisation of N. Nonetheless, it is in harmony with the CLT A.3 in Paper A which also yields a better rate if N is chosen maximal.

A second extension of Theorem B.2 involves the diffusivity a. While the estimator $\widehat{\vartheta}_\delta(x)$ in (B.9) requires knowledge of this parameter, in general it may be unknown. Replacing thus a by a reasonable estimate $\widehat{a}_\delta$ yields another estimator $\widetilde{\vartheta}_\delta(x)$ which achieves the same convergence rate.

COROLLARY B.4. *Grant Assumption L, and suppose that the diffusivity $a > 0$ is unknown. Define the estimator $\widetilde{\vartheta}_\delta(x)$ similarly to (B.9), replacing a with an estimate $\widehat{a}_\delta$. If $\widehat{a}_\delta$ satisfies*

$$\widehat{a}_\delta - a \in O_{\mathbb{P}}(h^\beta + (Nh^d)^{-1/2}), \tag{B.15}$$

then $\widetilde{\vartheta}_\delta(x) - \vartheta(x) \in O_{\mathbb{P}}(h^\beta + (Nh^d)^{-1/2})$.

Estimators which fulfill (B.15) are, for instance, given by

$$\widehat{a}_\delta = \frac{\sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\Delta(t)\,\mathrm{d}X_{\delta,k}(t)}{\sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\Delta(t)^2\,\mathrm{d}t} \quad \text{or} \quad \widehat{a}_\delta = \frac{\sum_{k=1}^N \int_0^T X_{\delta,k}^\Delta(t)\,\mathrm{d}X_{\delta,k}(t)}{\sum_{k=1}^N \int_0^T X_{\delta,k}^\Delta(t)^2\,\mathrm{d}t}. \tag{B.16}$$

Finally, we can also extend Theorem B.2 beyond the pointwise risk and quantify the quality of $\widehat{\vartheta}_\delta$ on the whole domain Λ. Since the estimator $\widehat{\vartheta}_\delta(x)$ in (B.9) was only defined for $x \in \mathcal{J}$, we

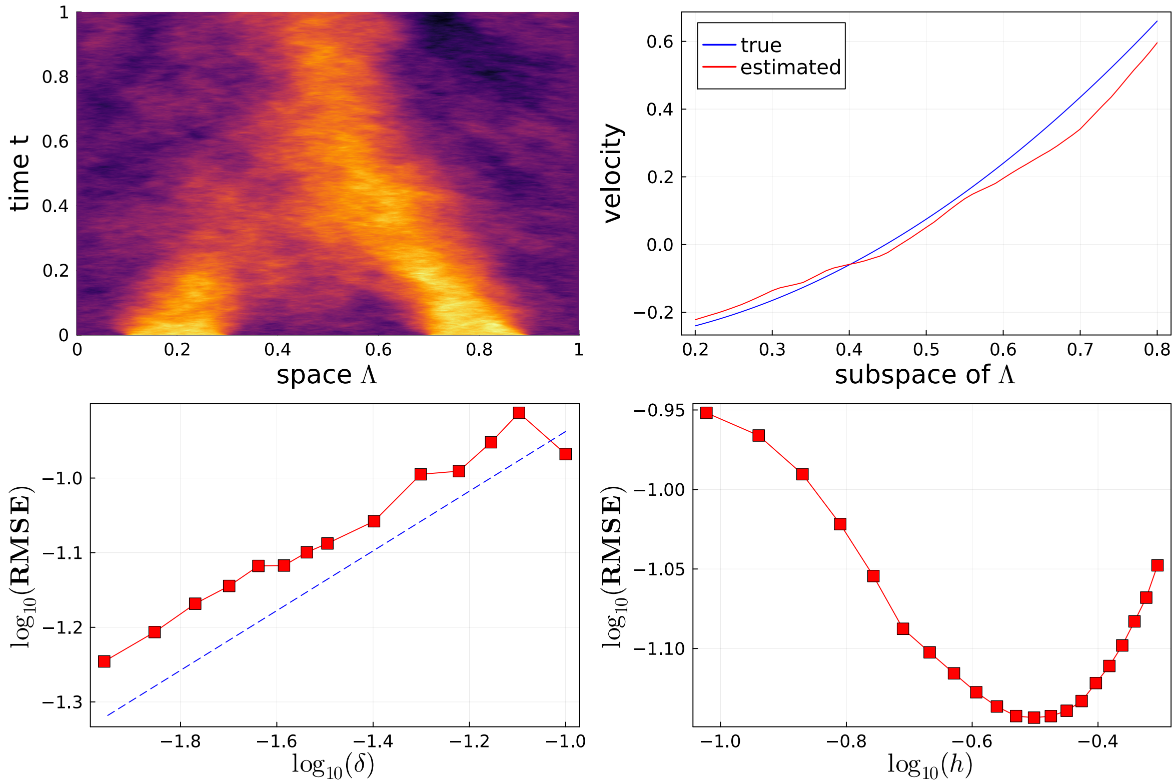


Figure B.1: (top-left) typical realisation of the solution $X(t,x)$ in $d=1$ with domain $\Lambda=(0,1)$; (top-right) trajectory of $\widehat{\vartheta}_\delta(x)$ compared to $\vartheta(x)=-0.3+1.5x^2$ in the interval $[0.2,0.8]\subset\Lambda$ with weights $w_k(x)$ based on the Epanechnikov kernel; (bottom) log-log plot of the root mean squared error for estimating ϑ at $x=0.5$ with $\delta\to 0$, $h\asymp\delta^{2/5}$ (left); δ fix, $h\to 0$ (right).

start by expanding its definition to Λ. Its value at $x\in\Lambda\setminus\mathcal{J}$ is set to a value $\widehat{\vartheta}_\delta(x_0)$, whereas $x_0\in\mathcal{J}$ is closest to x, that is,

$$\widehat{\vartheta}_\delta(x) := \inf_{x_0}\widehat{\vartheta}_\delta(x_0), \tag{B.17}$$

with $x_0\in\{x\in\mathcal{J}:|x-x_0|=\operatorname{dist}(x,\mathcal{J})\}$. Hence, we take the estimate at the closest point $x_0\in\mathcal{J}$ to further exploit Hölder continuity. The infimum over all possible x_0 is taken to obtain a unique estimate. Alternatively, one could also consider polynomial interpolation outside of $\mathcal{J}$.

Corollary B.5. *Grant Assumption L, and define $\widehat{\vartheta}_\delta$ outside of $\mathcal{J}$ via (B.17). Then,*

$$\int_\Lambda\left(\widehat{\vartheta}_\delta(x)-\vartheta(x)\right)^2\mathrm{d}x = O_{\mathbb{P}}\left(h^{2\beta}+\frac{1}{Nh^d}\right)+O(d_{\max}^2\lambda(\Lambda\setminus\mathcal{J})), \tag{B.18}$$

where λ denotes the Lebesgue measure on $\mathbb{R}^d$ and $d_{\max}^2 := \sup_{x\in\Lambda\setminus\mathcal{J}}\operatorname{dist}^2(x,\mathcal{J})$ is the maximal squared distance of $\mathcal{J}$ to the boundary $\partial\Lambda$.

Remark B.6 (discussion of Corollary B.5). Equation (B.18) splits the squared integrated error into a term of stochastic order, similar to the pointwise risk in Theorem B.2, and a deterministic part which is entirely dependent on the compact set $\mathcal{J}\subset\Lambda$. While the question of consistency thus is not immediately clear, it still can be achieved with a (possibly) slower rate. The supports of $K_{\delta,x}$ are contained in $\bar{\Lambda}$ for all $x \in \mathcal{J}$ and any $\delta \leq \delta'$ for δ' small enough due to the compactness of $\mathcal{J}$. This means that the distance between the boundary $\partial\Lambda$ and $\mathcal{J}$ behaves at best like δ', i.e., $d_{\max}^2 \asymp (\delta')^2$. On the other hand, $\lambda(\Lambda \setminus \mathcal{J})$ becomes small if $d_{\max}^2$ decreases. In fact, $d_{\max}^2 \asymp (\delta')^2$ implies $\lambda(\Lambda \setminus \mathcal{J}) = O(\delta')$. Hence, under a maximal choice of N and optimisation in h, (B.18) yields the order

$$O_{\mathbb{P}}(\delta^{2\beta d/(2\beta+d)}) + O((\delta')^3).$$

Cases where $d_{\max}^2 \asymp \delta'^2$ are given, for instance, if

- Λ is an d-dimensional open ball of radius r, and $\mathcal{J}$ is the closed ball with radius $r - \delta'$ and the same centre point;
- Λ is a rectangular cuboid of the form $(a_1, b_1) \times \cdots \times (a_d, b_d)$, and $\mathcal{J}$ is chosen as $[a_1 + \delta', b_1 - \delta'] \times \cdots \times [a_d + \delta', b_d - \delta']$.

Let us finish this section with a closer inspection of the weight functions $w_k(x)$ from Assumption L(iii). Their existence holds under general design assumptions, cf. also [43, Lemma 1.4 and Lemma 1.5].

LEMMA B.7. *Let $h > 0$ and $V\colon \mathbb{R}^d \to \mathbb{R}$ be a kernel function. Consider the $\mathbb{R}^{d+1}$-valued function U given by $U(u) = (1, u_1, \ldots, u_d)^\top$, and define the matrix*

$$B_{Nx} = \frac{1}{Nh^d} \sum_{k=1}^{N} U\Big(\frac{x_k - x}{h}\Big) U^\top\Big(\frac{x_k - x}{h}\Big) V\Big(\frac{x_k - x}{h}\Big).$$

Assume that the following conditions hold:

(LP1) *There exist a real number $\lambda_0 > 0$ and a positive integer n_0 such that the smallest eigenvalue $\lambda_{\min}(B_{Nx}) \geq \lambda_0$ for all $n \geq n_0$ and any $x \in \mathcal{J}$.*

(LP2) *There exists a real number $a_0 > 0$ such that, for any $A \subset \mathcal{J}$ and all $N \geq 1$,*

$$\frac{1}{N} \sum_{k=1}^{N} \mathbf{1}(x_k \in A) \leq a_0 \max(\lambda(A), 1/N),$$

with λ denoting the Lebesgue measure.

(LP3) *The kernel V has compact support in $[-1, 1]^d$, and there exists a number $V_{\max} < \infty$ such that $V(u) \leq V_{\max}$ for all $u \in \mathbb{R}^d$.*

Then, the weights defined by

$$w_k(x) := \frac{1}{Nh^d} U^\top(0) B_{Nx}^{-1} U\Big(\frac{x_k - x}{h}\Big) V\Big(\frac{x_k - x}{h}\Big) \tag{B.19}$$

satisfy Assumption L(iii).

Assumptions (LP1)-(LP3) in Lemma B.7 are satisfied under reasonable constraints on the design points $x_1, \ldots, x_N$. (LP2) means that they are densely enough distributed over $\mathcal{J}$. This holds true, for instance, under equidistant design, noting that at most $N \asymp \delta^{-d}$. (LP1) is satisfied if $V(u) > V_{\min} > 0$ in a neighbourhood around 0 and if additionally $x_1, \ldots, x_N$ are sufficiently dense in $\mathcal{J}$, cf. [43, Lemma 1.4 and Lemma 1.5]. (LP3) presents no restriction since the kernel V can be chosen according to need.

Example B.8. Let us give a concrete example of the weights w_k in (B.19). Assume $d = 1, \mathcal{J} = [0, 1]$, and choose the rectangular kernel $V(y) = \mathbf{1}(-1/2 \le y \le 1/2)$. Define $I_h = \{k : |x_k - x| \le h/2\}$. Then,

$$B_{Nx} = \frac{1}{Nh}\begin{pmatrix} \sum_{k\in I_h} 1 & \sum_{k\in I_h} \frac{x_k-x}{h} \\ \sum_{k\in I_h} \frac{x_k-x}{h} & \sum_{k\in I_h} \left(\frac{x_k-x}{h}\right)^2 \end{pmatrix}$$

has strictly positive determinant if there are at least two different points x_i, x_j in an $h/2$-neighbourhood around x. If the measurement points x_k are equidistantly distributed on $\mathcal{J}$, that is, if $x_k = (k-1)/(N-1)$, $k = 1, \ldots, N$, and we estimate at the location $x = r/(N-1)$, $1 \le r \le N-2$, with $h < \min(x, 1-x)/2$, then $\sum_{k\in I_h}(x_k - x)/h = 0$ by symmetry. The weights $w_k(x)$ in (B.19) are given by

$$w_k(x) = (\#I_h)^{-1}\mathbf{1}(k \in I_h).$$

In that case, the weights correspond to the weight function of the Nadaraya–Watson estimator with rectangular kernel.

An estimated trajectory based on the weights in (B.19) with Epanechnikov kernel $V(y) = 0.75(1-y^2)\mathbf{1}(|y| \le 1)$ is given in Figure B.1.

B.4 Lower bounds

The convergence rate $N^{-\beta/(2\beta+d)}$ established for the weighted augmented MLE in Theorem B.2 is optimal and cannot be improved in our general setup, as will be shown in this section. We will only consider submodels $\mathbb{P}_\vartheta$ such that A_ϑ involves a negative reaction term, assuming a sufficiently regular kernel function K and a stationary initial condition.

Assumption O. Suppose that $\mathbb{P}_\vartheta$ corresponds to the law of the stationary solution X to the SPDE (B.1), and assume that the following conditions hold:

(i) The kernel function satisfies $K = \Delta^2\widetilde{K}$ with $\widetilde{K} \in C_c^\infty(\mathbb{R}^d)$.

(ii) The model is $A_\vartheta = \Delta + \vartheta\cdot\nabla + c$ with a nonpositive reaction function $c\colon \Lambda \to \mathbb{R}$ and such that $\vartheta\colon \Lambda \to \mathbb{R}$ lies in the class Θ of β-Hölder continuous functions with the properties that there exists a constant $\gamma \le 0$ such that the $(\beta-1)$ Hölder-continuous function $c - \nabla\cdot\vartheta$ is smaller or equal than γ and that ϑ is a conservative vector field.

(iii) Let $x_1, \ldots, x_N$ be δ-separated points in Λ, that is, $|x_k - x_l| > \delta$ for all $1 \le k \ne l \le N$. Moreover, suppose that $\operatorname{supp}(K_{\delta,x_k}) \subset \Lambda$ for all $k = 1, \ldots, N$, and that $\operatorname{supp}(K_{\delta,x_k}) \cap \operatorname{supp}(K_{\delta,x_l}) = \emptyset$ for all $1 \le k \ne l \le N$.

We consider the null model to be $A_\vartheta = \Delta$, i.e., $\vartheta = 0$, $c = \gamma = 0$, and we test against alternatives where $\vartheta \ne 0$ and c is strictly negative such that $c - \nabla\cdot\vartheta \le \gamma < 0$.

Theorem B.9. *Grant Assumption O. Then, there exist $c_1 > 0$, depending only on K and d, and an absolute constant $c_2 > 0$ such that, for any $x \in \Lambda$, the following assertion holds:*

$$\inf_{\widehat{\vartheta}} \sup_{\vartheta\in\Theta} \mathbb{P}_\vartheta\left(|\widehat{\vartheta}(x) - \vartheta(x)| \ge \frac{c_1}{2}N^{-\beta/(2\beta+d)}\right) > c_2,$$

where the infimum is taken over all real-valued estimators $\widehat{\vartheta}_i = \widehat{\vartheta}_i(X_\delta)$.

As the weighted augmented MLE is not only based on the observations of X_δ, but also on X_δ^Δ and X_δ^∇, Theorem B.9 can be furthermore extended to estimators $\widehat{\vartheta}$ using those additional observations.

THEOREM B.10. *Theorem B.9 remains valid when the infimum is taken over all real-valued estimators $\widehat{\vartheta}_i = \widehat{\vartheta}_i(X_\delta, X_\delta^\Delta, X_\delta^\nabla)$, provided that K, ΔK and $\partial_i K$ are independent and Assumption O(i) holds for K, ΔK and $\partial_i K$, $1 \leq i \leq d$.*

Theorem B.9 is proven in Section B.5.4 below. The proof of Theorem B.10 is skipped as it relies only on minor modifications, see also Theorem A.9 in Paper A.

B.5 TECHNICAL SUPPLEMENT: AUXILIARY RESULTS AND PROOFS

We start with a few initial notations and remarks. Write $\Lambda_{\delta,y} = \{\delta^{-1}(u-y) : u \in \Lambda\}$, $\Lambda_{0,y} = \mathbb{R}^d$, and introduce the rescaled operators $A_{\vartheta,\delta,y}$ and $\bar{A}_{\delta,y}$ with domain $H_0^1(\Lambda_{\delta,y}) \cap H^2(\Lambda_{\delta,y})$ by setting

$$A_{\vartheta,\delta,y} := a\Delta + \delta\vartheta(y+\delta\cdot)\cdot\nabla + \delta^2 c(y+\delta\cdot), \quad \bar{A}_{\delta,y} := a\Delta.$$

The associated analytic semigroups on $L^2(\Lambda_{\delta,y})$ are denoted by $(S_{\vartheta,\delta,y}(t))_{t\geq 0}$ and $(\bar{S}_{\delta,y}(t))_{t\geq 0}$, respectively. Write $\mathrm{e}^{ta\Delta}$ for the semigroup on $L^2(\mathbb{R}^d)$ generated by $a\Delta$ on $H^2(\mathbb{R}^d)$. Define the heat kernel $q_t(u) = (4\pi t)^{-d/2}\exp(-|u|^2/(4t))$, and notice that, for $(\mathrm{e}^{ta\Delta})z = q_{at} * z$, by Young's inequality,

$$\|\mathrm{e}^{ta\Delta}z\|_{L^2(\mathbb{R}^d)} \lesssim (1 \wedge t^{-d/4})(\|z\|_{L^1(\mathbb{R}^d)} + \|z\|_{L^2(\mathbb{R}^d)}).$$

We denote $\varphi_\vartheta = \nabla\cdot\vartheta - c$, and we want to estimate ϑ at the (fixed) location $x \in \mathcal{J}$. The stochastic order $O_{\mathbb{P}}(h^\beta + \delta h + \delta^2)$ of $\mathcal{R}_\delta^x$, which can, in principle, be found from the proofs in Section B.5.3 below, will always be dominated by h^β as $\beta \leq 2$. This is clear since our methodology is only applicable if $\delta \ll h$. Indeed, if $h \leq \delta$, then consistency cannot be achieved as the number of observations used to construct the estimator in (B.9) remains finite. Optimising (B.11) with respect to the bandwidth h yields that $h \asymp N^{-1/(2\beta+d)}$. Furthermore, Assumption L implies that there exist at most $N \asymp \delta^{-d}$ spatial observation locations. Together, this gives for any dimension $d \geq 1$ and $\beta \in (1,2]$,

$$\delta^2 \ll \delta^{\beta d/(2\beta+d)} = O(h^\beta), \quad \delta^{d/2} \ll h^\beta, \tag{B.20}$$

which we will frequently use in Sections B.5.2 and B.5.3 down below.

B.5.1 *The rescaled semigroup*

In this section, we present properties of the rescaled semigroup $(S_{\vartheta,\delta,y}^*(t))_{t\geq 0}$ and its infinitesimal generator $A_{\vartheta,\delta,y}^*$.

LEMMA B.11 (Lemma 3.1 of [4]). *For $\delta > 0$ and $y \in \Lambda$, it holds:*

(i) *If $z \in H_0^1(\Lambda_{\delta,y}) \cap H^2(\Lambda_{\delta,y})$, then $A_\vartheta^* z_{\delta,y} = \delta^{-2}(A_{\vartheta,\delta,y}^* z)_{\delta,y}$;*

(ii) *if $z \in L^2(\Lambda_{\delta,y})$, then $S_\vartheta^*(t)z_{\delta,y} = (S_{\vartheta,\delta,y}^*(t\delta^{-2})z)_{\delta,y}$, $t \geq 0$.*

The following lemma is a classical result for sectorial operators and corresponding analytic semigroups. Our version holds for growing domains $\Lambda_{\delta,y}$, uniformly in $y \in \mathcal{J}$.

Lemma B.12. *There exist universal constants $M_0, M_1, C > 0$ such that, for $\delta \geq 0$, $t > 0$,*

$$\sup_{y\in\mathcal{J}} \|S^*_{\vartheta,\delta,y}(t)\|_{L^2(\Lambda_{\delta,y})} \leq M_0 \mathrm{e}^{C\delta^2 t},$$
$$\sup_{y\in\mathcal{J}} \|t(C\delta^2 I - A^*_{\vartheta,\delta,y})S^*_{\vartheta,\delta,y}(t)\|_{L^2(\Lambda_{\delta,y})} \leq M_1 \mathrm{e}^{C\delta^2 t}.$$

This lemma shows that the shifted semigroup $\mathrm{e}^{-2C\delta^2 t} S^*_{\vartheta,\delta,y}(t)$ decays exponentially,

$$\|\mathrm{e}^{-2C\delta^2 t} S^*_{\vartheta,\delta,y}(t)\|_{L^2(\Lambda_{\delta,y})} \leq \mathrm{e}^{-C\delta^2 t},$$

and so the resolvent set of the correspondingly shifted infinitesimal generator $2C\delta^2 - A^*_{\vartheta,\delta,y}$ contains the right half of the complex plane. This allows for defining the fractional powers $(2C\delta^2 - A^*_{\vartheta,\delta,y})^s$ for $s \in \mathbb{R}$, see [20, Section 4.4], and we obtain by [20, Proposition 4.37] the usual smoothing property of analytic semigroups.

Lemma B.13. *There exists a universal constant M_2 such that, for $\delta \geq 0$, $t > 0$ and $s \geq 0$,*

$$\sup_{y\in\mathcal{J}} \|t^s (2C\delta^2 - A^*_{\vartheta,\delta,y})^s S^*_{\vartheta,\delta,y}(t)\|_{L^2(\Lambda_{\delta,y})} \leq M_2 \mathrm{e}^{C\delta^2 t}.$$

Intuitively, letting $\delta \to 0$, the semigroup on $\Lambda_{\delta,y}$ will be close to the semigroup on $\mathbb{R}^d$. The following auxiliary result states this more precisely.

Lemma B.14. *Let $t > 0$, and grant Assumption L.*

(i) *There exist universal constants c_1, c_2, c_3 such that, if $z \in C_c(\mathbb{R}^d)$ is supported in $\bigcap_{y\in\mathcal{J}} \Lambda_{\delta,y}$ for some $\delta \geq 0$, then*

$$\sup_{y\in\mathcal{J}} \left|(S^*_{\vartheta,\delta,y}(t)z)(u)\right| \leq c_3 \mathrm{e}^{c_1\delta^2 t}(q_{c_2 t} * |z|)(u), \quad u \in \mathbb{R}^d.$$

(ii) *If $z \in L^2(\mathbb{R}^d)$, then, as $\delta \to 0$,*

$$\sup_{y\in\mathcal{J}} \|S^*_{\vartheta,\delta,y}(t)(z|_{\Lambda_{\delta,y}}) - \mathrm{e}^{ta\Delta} z\|_{L^2(\mathbb{R}^d)} \to 0.$$

(iii) *If $z \in L^2(\mathbb{R}^d)$, then, for any $t \geq 0$,*

$$\sup_{y\in\mathcal{J}} \|\bar{S}_{\delta,y}(t)z - \mathrm{e}^{ta\Delta} z\|_{L^2(\mathbb{R}^d)} \lesssim \delta^{1/2} t^{1/4} \mathrm{e}^{-\delta^{-2}t^{-1}/2}.$$

The action of the semigroup operators $S^*_{\vartheta,\delta,y}(t)$ applied to functions of a certain smoothness and integrability is given in the next lemma. The proof relies on the Bessel potential spaces $H^{s,p}_0(\Lambda_{\delta,y})$, $1 < p < \infty$, $s \in \mathbb{R}$, defined for $\delta > 0$ as the domains of the fractional weighted Dirichlet–Laplacian $(-\bar{A}_{\vartheta,\delta,y})^{s/2}$ of order $s/2$ on $\Lambda_{\delta,y}$ with norms $\|\cdot\|_{H^{s,p}(\Lambda_{\delta,y})} = \|(-\bar{A}_{\vartheta,\delta,y})^{s/2}\cdot\|_{L^p(\Lambda_{\delta,y})}$.

LEMMA B.15 (Lemma A.16 in Paper A). *Let $\delta \in [0,1]$, $t > 0$, and grant Assumption L. Let $z \in H_0^s(\mathbb{R}^d)$, $s \geq 0$, be compactly supported in $\bigcap_{y \in \mathcal{J}} \Lambda_{\delta,y}$, and suppose that $V_{\delta,y} \colon L^p(\Lambda_{\delta,y}) \to H_0^{-s,p}(\Lambda_{\delta,y})$ are bounded linear operators with $\|V_{\delta,y} z\|_{H^{-s,p}(\Lambda_{\delta,y})} \leq V_{\mathrm{op}} \|z\|_{L^p(\Lambda_{\delta,y})}$, for some V_{op} independent of δ and y. Then, there exists a universal constant $C > 0$ such that, for $1 < p \leq 2$ and $\gamma = (1/p - 1/2)d/2$,*

$$\sup_{y \in \mathcal{J}} \|S^*_{\vartheta,\delta,y}(t) V_{\delta,y} z\|_{L^2(\Lambda_{\delta,y})} \leq C e^{c_1 t \delta^2} \sup_{y \in \mathcal{J}} \Big(\|V_{\delta,y} z\|_{L^2(\Lambda_{\delta,y})} \wedge (V_{\mathrm{op}} t^{-s/2-\gamma} \|z\|_{L^p(\Lambda_{\delta,y})}) \Big),$$

where c_1 is the constant described in Lemma B.14(i). If $s = 0$, the inequality holds also for $p = 1$.

B.5.2 *Properties of multiple local measurements*

For the reader's convenience, we give the result of Paper A, specifying the covariance function of the Gaussian process defined in (B.5).

LEMMA B.16 (Lemma A.17 in Paper A).

(i) *If $X_0 = 0$, then the Gaussian process from (B.5) has mean zero and covariance function*

$$\mathrm{Cov}(\langle X(t), z\rangle, \langle X(t'), z'\rangle) = \int_0^{t \wedge t'} \langle S^*_\vartheta(t-s) z, S^*_\vartheta(t'-s) z'\rangle \, \mathrm{d}s.$$

(ii) *If X_0 is the stationary initial condition from Assumption L(iv), then the Gaussian process from (B.5) has mean zero and covariance function*

$$\mathrm{Cov}(\langle X(t), z\rangle, \langle X(t'), z'\rangle) = \int_0^{\infty} \langle S^*_\vartheta(t+s) z, S^*_\vartheta(t'+s) z'\rangle \, \mathrm{d}s.$$

LEMMA B.17. *Grant Assumption L, and consider $u, w \in \{D^\alpha K : |\alpha| \leq 2\} = \{-D^\alpha \Delta \bar{K} : |\alpha| \leq 2\}$. Let $X_0 = 0$, and set $f_0(t) := \langle e^{ta\Delta} u, e^{ta\Delta} w\rangle_{L^2(\mathbb{R}^d)}$. Then, the following properties hold true:*

(i) *$\psi(u,w) = \int_0^\infty f_0(t)\,\mathrm{d}t$ is well-defined, i.e., $f_0 \in L^1([0,\infty))$.*

(ii) *For $\delta \to 0$,*

$$\sup_{y \in \mathcal{J}} \left| \delta^{-2} \int_0^T \mathrm{Cov}(\langle X(t), u_{\delta,y}\rangle, \langle X(t), w_{\delta,y}\rangle)\,\mathrm{d}t - T\psi(u,w) \right| \to 0.$$

(iii) *If, additionally, $\psi(u,w) = 0$, then*

$$\sup_{y \in \mathcal{J}} \left| \delta^{-3} \int_0^T \mathrm{Cov}(\langle X(t), u_{\delta,y}\rangle, \langle X(t), w_{\delta,y}\rangle)\,\mathrm{d}t \right| \lesssim 1.$$

LEMMA B.18. *Grant Assumption L, and let $X_0 = 0$.*

(i) *For $u, w \in \{D^\alpha K : |\alpha| \leq 2\}$,*

$$\sup_{y \in \mathcal{J}} \mathrm{Var}\left(\int_0^T \langle X(t), u_{\delta,y}\rangle \langle X(t), w_{\delta,y}\rangle \,\mathrm{d}t \right) = O(\delta^6).$$

(ii) *For* $u \in \{\partial_i K : 1 \le i \le d\}$ *and* $w := g^{(\vartheta,y,\delta)} \cdot \nabla K$, *with* $g^{(\vartheta,y,\delta)}$ *defined in* (B.22) *below, it holds*

$$\sup_{x\in\mathcal{J}} \sup_{y\in\mathcal{J},|y-x|\le h} \mathrm{Var}\left(\int_0^T \langle X(t), u_{\delta,y}\rangle\langle X(t), w_{\delta,y}\rangle \,\mathrm{d}t\right) = O(\delta^4 h^{2\beta}).$$

(iii) *For* $u \in \{\partial_i K : 1 \le i \le d\}$ *and* $w := \varphi_\vartheta(y + \delta\cdot)K$, *we have*

$$\sup_{x\in\mathcal{J}} \sup_{y\in\mathcal{J},|y-x|\le h} \mathrm{Var}\left(\int_0^T \langle X(t), u_{\delta,y}\rangle\langle X(t), w_{\delta,y}\rangle \,\mathrm{d}t\right) = O(\delta^2 h^{2\beta}).$$

B.5.3 *Proof of the upper bound*

Before proving Theorem B.2, we carefully inspect the observed Fisher information $\mathcal{I}_\delta^x$ and the remainder $\mathcal{R}_\delta^x$ appearing in the decomposition (B.10).

The Fisher information and the martingale part

Proposition B.19. *Grant Assumption L. Then,*

$$\mathcal{I}_\delta^x \xrightarrow{\mathbb{P}} \Sigma, \qquad \textit{where} \qquad \Sigma_{ij} = \frac{T}{2a}\langle(-\Delta)^{-1}\partial_i K, \partial_j K\rangle, \; i, j \in \{1, \dots, d\},$$

and Σ *thus defined is invertible.*

Proof. We only consider the case where $X_0 = 0$. Note that Assumption L(iv) implies the assumed structure in Paper A, cf. Lemma A.2. We hence refer to Theorem A.3 for the invertibility of Σ and the generalisation of the initial condition. Thus, it suffices to show that, for $1 \le i, j \le d$,

$$\mathbb{E}[\mathcal{I}_\delta^x]_{ij} \to \Sigma_{ij}, \quad \mathrm{Var}((\mathcal{I}_\delta^x)_{ij}) \to 0. \tag{B.21}$$

Recall that, for $z, z' \in L^2(\mathbb{R}^d)$, the function $\psi(\cdot,\cdot)$ introduced in Lemma B.17 is defined as

$$\psi(z, z') = \int_0^\infty \langle \mathrm{e}^{ta\Delta} z, \mathrm{e}^{ta\Delta} z'\rangle \,\mathrm{d}t = \frac{1}{2a}\langle(-\Delta)^{-1} z, z'\rangle.$$

In view of $\sum_{k=1}^N w_k(x) = 1$, $\sum_{k=1}^N |w_k(x)| \lesssim 1$, the first part of (B.21) follows by

$$\begin{aligned}
&\sup_{x\in\mathcal{J}} |\mathbb{E}[\mathcal{I}_\delta^x]_{ij} - \Sigma_{ij}| \\
&\le \sup_{x\in\mathcal{J}} \sum_{k=1}^N |w_k(x)| \left|\delta^{-2}\int_0^T \mathrm{Cov}(\langle X(t), (\partial_i K)_{\delta,x_k}\rangle, \langle X(t), (\partial_j K)_{\delta,x_k}\rangle)\,\mathrm{d}t - T\psi(\partial_i K, \partial_j K)\right| \\
&\le C_* \sup_{y\in\mathcal{J}} \left|\delta^{-2}\int_0^T \mathrm{Cov}(\langle X(t), (\partial_i K)_{\delta,y}\rangle, \langle X(t), (\partial_j K)_{\delta,y}\rangle)\,\mathrm{d}t - T\psi(\partial_i K, \partial_j K)\right| \\
&\to 0,
\end{aligned}$$

where the convergence statement in the last line is a consequence of Lemma B.17(ii). By the Cauchy–Schwarz inequality and Lemma B.18(i), we obtain

$$\sup_{x\in\mathcal{J}} \mathrm{Var}((\mathcal{I}_\delta^x)_{ij})^{1/2} \le \sup_{x\in\mathcal{J}} \sum_{k=1}^N |w_k(x)|\delta^{-2}\mathrm{Var}\left(\int_0^T \langle X(t), (\partial_i K)_{\delta,x_k}\rangle\langle X(t), (\partial_j K)_{\delta,x_k}\rangle\,\mathrm{d}t\right)^{1/2}$$

$$\leq C_* \sup_{y \in \mathcal{J}} \delta^{-2} \mathrm{Var}\bigg(\int_0^T \langle X(t), (\partial_i K)_{\delta,y} \rangle \langle X(t), (\partial_j K)_{\delta,y} \rangle \,\mathrm{d}t \bigg)^{1/2} \longrightarrow 0,$$

concluding the proof. ∎

PROPOSITION B.20. *Grant Assumption L. Then,*

$$[\mathcal{M}_\delta^x]_T := \sum_{k=1}^N w_k(x)^2 \int_0^T X_{\delta,k}^\nabla(t) X_{\delta,k}^\nabla(t)^\top \,\mathrm{d}t = O_{\mathbb{P}}((Nh^d)^{-1}),$$

where the stochastic order of the right hand side is indepenent of $x \in \mathcal{J}$.

Proof. Again, it suffices to verify the claim with initial condition $X_0 = 0$. We show $\mathbb{E}[[\mathcal{M}_\delta^x]_T] = O((Nh^d)^{-1})$ and $\mathrm{Var}(([\mathcal{M}_\delta^x]_T)_{ij}) = o((Nh^d)^{-2})$. Using that $\sum_{k=1}^N w_k(x)^2 \lesssim (Nh^d)^{-1}$, we get similarly to the proof of Proposition B.19 that

$$\sup_{x\in\mathcal{J}} \mathbb{E}[[\mathcal{M}_\delta^x]_T]_{ij} \leq \sup_{x\in\mathcal{J}} \sum_{k=1}^N w_k(x)^2 \bigg| \delta^{-2} \int_0^T \mathrm{Cov}(\langle X(t), (\partial_i K)_{\delta,x_k} \rangle, \langle X(t), (\partial_j K)_{\delta,x_k} \rangle) \,\mathrm{d}t \bigg| \lesssim (Nh^d)^{-1}$$

as well as

$$\begin{aligned} \sup_{x\in\mathcal{J}} \mathrm{Var}(([\mathcal{M}_\delta^x]_T)_{ij})^{1/2} &\leq \sup_{x\in\mathcal{J}} \sum_{k=1}^N w_k(x)^2 \sup_{y\in\mathcal{J}} \delta^{-2} \mathrm{Var}\bigg(\int_0^T \langle X(t), (\partial_i K)_{\delta,y} \rangle \langle X(t), (\partial_j K)_{\delta,y} \rangle \,\mathrm{d}t \bigg)^{1/2} \\ &\leq C_* (Nh^d)^{-1} \sup_{y\in\mathcal{J}} \delta^{-2} \mathrm{Var}\bigg(\int_0^T \langle X(t), (\partial_i K)_{\delta,y} \rangle \langle X(t), (\partial_j K)_{\delta,y} \rangle \,\mathrm{d}t \bigg)^{1/2} \\ &= o((Nh^d)^{-1}). \end{aligned}$$

∎

The remainder term

In this subsection, we will study the expected value and variance of the remainder term $\mathcal{R}_\delta^x$, given by

$$\mathcal{R}_\delta^x = \sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\nabla(t) \langle X(t), ((\vartheta - \vartheta(x)) \cdot \nabla + \varphi_\vartheta) K_{\delta,x_k} \rangle \,\mathrm{d}t.$$

We start by exploring the connection between the weight functions $w_k(x)$ and the multivariate Taylor expansion. Define the difference

$$g^{(\vartheta,x_k,\delta)}(y) := \vartheta(x_k + \delta y) - \vartheta(x). \tag{B.22}$$

For $1 \leq i \leq d$, its i-th entry is given by the first order multivariate Taylor expansion with Peano remainder $P_{1,i,x_k}(y)$

$$\begin{aligned} g_i^{(\vartheta,x_k,\delta)}(y) &:= \sum_{|\alpha|=1} \frac{D^\alpha \vartheta_i(x)}{\alpha!} (x_k + \delta y - x)^\alpha + \sum_{|\alpha|=1} \frac{D^\alpha(\vartheta_i(\xi) - \vartheta_i(x))}{\alpha!} (x_k + \delta y - x)^\alpha \\ &= \sum_{|\alpha|=1} \frac{D^\alpha \vartheta_i(x)}{\alpha!} (x_k + \delta y - x)^\alpha + P_{1,i,x_k}(y), \end{aligned} \tag{B.23}$$

for some value $\xi \in B_{x_k+\delta y}(x)$.

Corollary B.21. *Grant Assumption L. Then, for any $0 \leq s \leq T\delta^{-2}$,*

$$\sum_{k=1}^{N} w_k(x)\langle e^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)} \lesssim h^{\beta}(1 \wedge s^{-3/2-d/4}). \tag{B.24}$$

Proof. By Assumption L(iii),

$$\sum_{k=1}^{N} w_k(x) g_i^{(\vartheta,x_k,\delta)}(y) = \sum_{k=1}^{N} w_k(x)\left(\sum_{|\alpha|=1} \frac{D^{\alpha}\vartheta_i(x)}{\alpha!}(\delta y)^{\alpha} + P_{1,i,x_k}(y)\right).$$

Thus,

$$\begin{aligned}
&\sum_{k=1}^{N} w_k(x)\langle e^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)} \\
&\quad = \sum_{k=1}^{N} w_k(x)\langle e^{2sa\Delta}\nabla K, \sum_{i=1}^{d}\left(\sum_{|\alpha|=1} \frac{D^{\alpha}\vartheta_i(x)}{\alpha!}(\delta\cdot)^{\alpha} + P_{1,i,x_k}\right)\partial_i K\rangle_{L^2(\mathbb{R}^d)}.
\end{aligned}$$

By the symmetry of the heat kernel, $e^{2as\Delta}\partial_j K(y) = (q_{2as} * \partial_j K)(y)$ is even if $\partial_j K$ is even and odd if $\partial_j K$ is odd, respectively, and Assumption L guarantees that one of these cases always holds true. Moreover, the identity in $\mathbb{R}^d$ is an odd function which implies that $y_l \partial_i K(y)$ is odd if $\partial_i K$ is even and even if $\partial_i K$ is odd, respectively. Hence, for all $1 \leq j, l, i \leq d$,

$$\langle e^{2sa\Delta}\partial_j K(y), \delta y_l \partial_i K(y)\rangle_{L^2(\mathbb{R}^d)} = 0$$

as an integral over an odd function. Note that $\|P_{1,i,x_k}\partial_i K\|_{L^2(\mathbb{R}^d)} \lesssim h^{\beta}$ whenever $|x_k - x| \leq h$ due to $\delta \ll h$, the Hölder assumption on ϑ and $\partial_i K$ having compact support. Indeed,

$$\begin{aligned}
\|P_{1,i,x_k}\partial_i K\|_{L^2(\mathbb{R}^d)} &\lesssim \||x_k + \delta y - x|^{\beta}\partial_i K(y)\|_{L^2(\mathbb{R}^d)} \\
&\lesssim |x_k - x|^{\beta}\|\partial_i K\|_{L^2(\mathbb{R}^d)} \lesssim h^{\beta}.
\end{aligned}$$

(B.24) hence follows by the Cauchy–Schwarz inequality, $K = (-\Delta)\bar{K}$ and Lemma B.15, since

$$\begin{aligned}
\sum_{k=1}^{N} w_k(x)\sum_{i=1}^{d}\langle e^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)} &= \sum_{k:|x_k-x|\leq h} w_k(x)\sum_{i=1}^{d}\langle e^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)} \\
&= \sum_{k:|x_k-x|\leq h} w_k(x)\langle e^{2sa\Delta}\nabla K, \sum_{i=1}^{d} P_{1,i,x_k}\partial_i K\rangle_{L^2(\mathbb{R}^d)} \\
&\lesssim h^{\beta}(1 \wedge s^{-3/2-d/4}).
\end{aligned}$$

■

Proposition B.22. *Grant Assumption L, and assume that $X_0 = 0$. Then,*

$$\sup_{x \in \mathcal{J}} \mathbb{E}[\mathcal{R}_{\delta}^{x}] = O(h^{\beta}).$$

Proof. Using the covariance structure in Lemma B.16 and the rescaling Lemma B.11, we obtain

$$\begin{aligned}\mathbb{E}[\mathcal{R}_\delta^x] &= \sum_{k=1}^N w_k(x)\int_0^T \mathbb{E}\big[X_{\delta,k}^\nabla(t)\langle X(t),((\vartheta-\vartheta(x))\cdot\nabla+\varphi_\vartheta)K_{\delta,x_k}\rangle\big]\,\mathrm{d}t\\ &= \sum_{k=1}^N w_k(x)\int_0^T \mathrm{Cov}(X_{\delta,k}^\nabla,\langle X(t),((\vartheta-\vartheta(x))\cdot\nabla+\varphi_\vartheta)K_{\delta,x_k}\rangle)\,\mathrm{d}t\\ &= \int_0^T\int_0^{t\delta^{-2}} A(s)\,\mathrm{d}s\,\mathrm{d}t+\delta\int_0^T\int_0^{t\delta^{-2}} B(s)\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}\tag{B.25}$$

with

$$\begin{aligned}A(s) &:= \sum_{k=1}^N w_k(x)\langle S^*_{\vartheta,\delta,x_k}(s)\nabla K, S^*_{\vartheta,\delta,x_k}(s)(\vartheta(x_k+\delta\cdot)-\vartheta(x))\cdot\nabla K\rangle_{L^2(\Lambda_{\delta,x_k})},\\ B(s) &:= \sum_{k=1}^N w_k(x)\langle S^*_{\vartheta,\delta,x_k}(s)\nabla K, S^*_{\vartheta,\delta,x_k}(s)\varphi_\vartheta(x_k+\delta\cdot)K\rangle_{L^2(\Lambda_{\delta,x_k})}.\end{aligned}$$

Noting that $S^*_{\vartheta,\delta,x_k}(s)u(x)=0$ for $x\notin\Lambda_{\delta,x_k}$ and using multivariate Taylor expansion for ϑ_i, we can write

$$A(s)=\sum_{k=1}^N w_k(x)\langle S^*_{\vartheta,\delta,x_k}(s)\nabla K, S^*_{\vartheta,\delta,x_k}(s)g^{(\vartheta,x_k,\delta)}\cdot\nabla K\rangle_{L^2(\mathbb{R}^d)},$$

with $g^{(\vartheta,x_k,\delta)}$ given by (B.22). Corollary B.21 already implies

$$\sum_{k=1}^N w_k(x)\langle \mathrm{e}^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)}\cdot\nabla K\rangle_{L^2(\mathbb{R}^d)}\lesssim h^\beta(1\wedge s^{-3/2-d/4}).$$

Hence,

$$\int_0^T\int_0^{t\delta^{-2}}\sum_{k=1}^N w_k(x)\langle \mathrm{e}^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)}\cdot\nabla K\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}s\,\mathrm{d}t\lesssim h^\beta.\tag{B.26}$$

Thus, it remains to control the error terms resulting from the switch of semigroups. This is given in the next lemma. The proof relies on the L^2-distance of $\mathrm{e}^{sa\Delta}$ to $\bar{S}_{\delta,y}(s)$ (pointed out in Lemma B.14(iii)), the L^2-distance of $\bar{S}_{\delta,y}(s)$ to $S^*_{\vartheta,\delta,y}(s)$ (which can be controlled via the variation of parameters formula) and a sufficiently sharp upper bound for $\|S^*_{\vartheta,\delta,y}(s)g^{(\vartheta,\delta,y)}\cdot\nabla K\|_{L^2(\Lambda_{\delta,y})}$.

LEMMA B.23. *It holds*

$$\int_0^T\int_0^{t\delta^{-2}} A(s)\,\mathrm{d}s\,\mathrm{d}t=\int_0^T\int_0^{t\delta^{-2}}\sum_{k=1}^N w_k(x)\langle \mathrm{e}^{2sa\Delta}\nabla K, g^{(\vartheta,x_k,\delta)}\cdot\nabla K\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}s\,\mathrm{d}t+o(h^\beta),$$

where the o-term is independent of $x\in\mathfrak{J}$.

Lemma B.23 combined with (B.26) already yields the desired rate h^β for the leading order term $\int_0^T\int_0^{t\delta^{-2}}A(s)\,\mathrm{d}s\,\mathrm{d}t$ in (B.25). The lower order term $\delta\int_0^T\int_0^{t\delta^{-2}}B(s)\,\mathrm{d}s\,\mathrm{d}t$ is bounded in the same manner. Expand the right-hand side of the scalar product by adding and subtracting $\varphi_\vartheta(x)$.

Following the same structure as above, i.e., switching from the semigroup on $L^2(\Lambda_{\delta,x_k})$ to the heat kernel on $L^2(\mathbb{R}^d)$, we similarly obtain

$$\delta\int_0^T\int_0^{t\delta^{-2}}\sum_{k=1}^N w_k(x)\langle S^*_{\vartheta,\delta,x_k}(s)\nabla K, S^*_{\vartheta,\delta,x_k}(s)(\varphi_\vartheta(x_k+\delta\cdot)-\varphi_\vartheta(x))K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t = o(h^\beta).$$

On the other hand, using $\psi(\nabla K,\varphi_\vartheta(x)K)=0$ (due to integration by parts), Lemma B.17(iii), Lemma B.18(i) and (B.20), we derive

$$\delta\int_0^T\int_0^{t\delta^{-2}}\sum_{k=1}^N w_k(x)\langle S^*_{\vartheta,\delta,x_k}(s)\nabla K, S^*_{\vartheta,\delta,x_k}(s)\varphi_\vartheta(x)K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t \lesssim \delta^2 = o(h^\beta).$$

■

PROPOSITION B.24. *Grant Assumption L, and suppose that $X_0=0$. Then,*

$$\sup_{x\in\mathcal{J}}\mathrm{Var}(\mathcal{R}^x_\delta) = O(h^{2\beta}).$$

Proof. We will show that each entry of the covariance matrix of $\mathcal{R}^x_\delta$ satisfies the required order, which then directly implies the order for the entire covariance matrix. Note that $\mathrm{Cov}(\mathcal{R}^x_\delta)_{ij}$ is given by

$$\sum_{k=1}^N\sum_{l=1}^N w_k(x)w_l(x)\delta^{-2}\mathrm{Cov}\Bigg(\int_0^T\langle X(t),(\partial_i K)_{\delta,x_k}\rangle\langle X(t),(\varphi_\vartheta+(\vartheta-\vartheta(x))\cdot\nabla)K_{\delta,x_k}\rangle\,\mathrm{d}t,$$
$$\int_0^T\langle X(t),(\partial_j K)_{\delta,x_l}\rangle\langle X(t),(\varphi_\vartheta+(\vartheta-\vartheta(x))\cdot\nabla)K_{\delta,x_l}\rangle\,\mathrm{d}t\Bigg).$$

The Cauchy–Schwarz inequality and $(a+b)^2\le 2a^2+2b^2$ imply that, up to constants independent of $x\in\mathcal{J}$, this last quantity is upper bounded by

$$\delta^{-2}\sup_{y\in\mathcal{J},|y-x|\le h,k\le d}\mathrm{Var}\Bigg(\int_0^T\langle X(t),(\partial_k K)_{\delta,y}\rangle\langle X(t),(\varphi_\vartheta(y+\delta\cdot)K)_{\delta,y}\rangle\,\mathrm{d}t\Bigg)$$
$$+\,\delta^{-4}\sup_{y\in\mathcal{J},|y-x|\le h,k\le d}\mathrm{Var}\Bigg(\int_0^T\langle X(t),(\partial_k K)_{\delta,y}\rangle\langle X(t),g^{(\vartheta,y,\delta)}\cdot\nabla K)_{\delta,y}\rangle\,\mathrm{d}t\Bigg),$$

with $g^{(\vartheta,y,\delta)}$ from (B.22). The result follows then immediately by Lemma B.18(ii) and (iii). ■

Proposition B.22 and B.24 already imply that $\mathcal{R}^x_\delta$ is of stochastic order $O_{\mathbb{P}}(h^\beta)$ whenever $X_0=0$. Under Assumption L(iv), this can furthermore be extended to general initial conditions.

PROPOSITION B.25. *Grant Assumption L. Define $\bar{\mathcal{R}}^x_\delta$ analogous to $\mathcal{R}^x_\delta$, but with respect to $\bar{X}$ satisfying (B.1) with initial condition $\bar{X}(0)=0$. Then,*

$$\mathcal{R}^x_\delta = \bar{\mathcal{R}}^x_\delta + o_{\mathbb{P}}(h^\beta),$$

where the $o_{\mathbb{P}}$-term does not depend on $x\in\mathcal{J}$.

Proof of the upper bound statement

Proof of Theorem B.2. We use the error decomposition (B.10). To prove (B.11), it suffices to show for $\delta \to 0$ that $\mathcal{I}_\delta^x \xrightarrow{\mathbb{P}} \Sigma$ for some invertible, deterministic matrix Σ, while $\mathcal{M}_\delta^x = O_{\mathbb{P}}((Nh^d)^{-1/2})$ and $\mathcal{R}_\delta^x = O_{\mathbb{P}}(h^\beta)$. Proposition B.19 gives that $\mathcal{I}_\delta^x \xrightarrow{\mathbb{P}} \Sigma$ for some invertible Σ. Define a sequence of martingales via

$$\mathcal{M}_\delta^x(t) = \sum_{k=1}^N w_k(x) \int_0^t X_{\delta,k}^\nabla(s)\, \mathrm{d}W_k(s).$$

In particular, due to the independence of the Brownian motions W_k guaranteed by Assumption L, the quadratic variation of $\mathcal{M}_\delta^x = \mathcal{M}_\delta^x(T)$ is given by

$$[\mathcal{M}_\delta^x]_T = \sum_{k=1}^N w_k(x)^2 \int_0^T X_{\delta,k}^\nabla(t) X_{\delta,k}^\nabla(t)^\top \, \mathrm{d}t.$$

A standard argument, cf. [47, Lemma 3.6 or Lemma 3.8], shows that $\mathcal{M}_\delta^x$ behaves like the squared root of its quadratic variation, i.e., using Proposition B.20, $\mathcal{M}_\delta^x = O_{\mathbb{P}}((Nh^d)^{-1/2})$. Combining Proposition B.22, Proposition B.24 and Proposition B.25 yields the rate $O_{\mathbb{P}}(h^\beta)$ for $\mathcal{R}_\delta^x$.

To prove the supplement (B.12), it is enough to show that

$$\mathbb{P}\Big(\big|(\mathcal{I}_\delta^x)^{-1}\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) \le \frac{\varepsilon}{2}, \tag{B.27}$$

$$\mathbb{P}\Big(\big|(\mathcal{I}_\delta^x)^{-1}\mathcal{M}_\delta^x\big| (Nh^d)^{1/2} > M \|K\|_{L^2(\mathbb{R}^d)}^{-1}\Big) \le \frac{\varepsilon}{2}. \tag{B.28}$$

We only show the statement (B.27), as the arguments for (B.28) are similar. Now,

$$\begin{aligned}\mathbb{P}\Big(\big|(\mathcal{I}_\delta^x)^{-1}\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) &\le \mathbb{P}\Big(\big|((\mathcal{I}_\delta^x)^{-1} - \Sigma^{-1})\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) + \mathbb{P}\Big(\big|\Sigma^{-1}\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) \\ &\le \mathbb{P}\Big(\|(\mathcal{I}_\delta^x)^{-1} - \Sigma^{-1}\| \big|\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) + P\Big(\big|\mathcal{R}_\delta^x\big| h^{-\beta} > M\|\Sigma^{-1}\|^{-1}\Big) \end{aligned} \tag{B.29}$$

with arbitrary matrix norm $\|\cdot\|$ on $\mathbb{R}^{d\times d}$. Due to Proposition B.22, Proposition B.24, Chebyshev's inequality, and for δ sufficiently small and M sufficiently large, the term $\mathbb{P}\Big(\big|\mathcal{R}_\delta^x\big| h^{-\beta} > M\|\Sigma^{-1}\|^{-1}\Big)$ is uniformly bounded in $x \in \mathcal{J}$ by $\varepsilon/4$. On the other hand,

$$P\Big(\|(\mathcal{I}_\delta^x)^{-1} - \Sigma^{-1}\| \big|\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) \le P\Big(\big|\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) + P\Big(\|(\mathcal{I}_\delta^x)^{-1} - \Sigma^{-1}\| > 1\Big).$$

Again, δ and M can be chosen such that $P\Big(\big|\mathcal{R}_\delta^x\big| h^{-\beta} > M\Big) \le \varepsilon/8$. Moreover, there exists a value $\eta > 0$ with the property that $\|((\mathcal{I}_\delta^x)^{-1} - \Sigma^{-1})\| \le 1$ whenever $\|(\mathcal{I}_\delta^x - \Sigma)\| \le \eta$, due to the continuity of the function $y \mapsto y^{-1}$ and the fact that both Σ and $\mathcal{I}_\delta^x$ are (a.s.) invertible. Hence, for sufficiently small δ,

$$\mathbb{P}\Big(\|(\mathcal{I}_\delta^x)^{-1} - \Sigma^{-1}\| > 1\Big) \le P\big(\|\mathcal{I}_\delta^x - \Sigma\| > \eta\big) \le \varepsilon/8$$

due to Proposition B.19, thus showing the assertion. ∎

B.5.4 Proof of the lower bound

The proof of Theorem B.9 relies on the general reduction scheme in [43, Section 2.2] and the RKHS machinery described in detail in Section A.6.3 in Paper A. In what follows, we will therefore summarise the key components until the nonparametric setup requires a different reasoning.

Let $\mathbb{P}_{\vartheta^0}$ and $\mathbb{P}_{\vartheta^1}$ be two Gaussian measures defined on a separable Hilbert space $\mathcal{H}$ with expectation zero and positive self-adjoint trace-class covariance operators C_{ϑ^0} and C_{ϑ^1}, respectively. ϑ^0 and ϑ^1 belong to a set of functions Θ. By the spectral theorem, there exist (strictly) positive eigenvalues $(\sigma_j^2)_{j\geq 1}$ and an associated orthonormal system of eigenvectors $(u_j)_{j\geq 1}$ such that $C_{\vartheta^0} = \sum_{j\geq 1}\sigma_j^2(u_j\otimes u_j)$. The reproducing kernel Hilbert space (RKHS) associated to $\mathbb{P}_{\vartheta^0}$ is given by

$$H_{\vartheta^0} = \{h\in\mathcal{H} : \|h\|_{H_{\vartheta^0}} < \infty\}, \qquad \|h\|^2_{H_{\vartheta^0}} = \sum_{j\geq 1}\frac{\langle u_j, h\rangle^2_{\mathcal{H}}}{\sigma_j^2}.$$

Instead of Lemma A.22 in Paper A, we rely on its nonparametric equivalent. The proof is identical and therefore skipped.

Lemma B.26. *In the above Gaussian setting, suppose that $(u_j)_{j\geq 1}$ is an orthonormal basis of $\mathcal{H}$ and that*

$$\sum_{j\geq 1}\sigma_j^{-2}\|(C_{\vartheta^1}-C_{\vartheta^0})u_j\|^2_{H_{\vartheta^0}} \leq \frac{1}{2}. \tag{B.30}$$

Then, the squared Hellinger distance satisfies the bound $H^2(\mathbb{P}_{\vartheta^0},\mathbb{P}_{\vartheta^1}) \leq 1$. Therefore, for any $x\in\Lambda$ and a generic constant $c_1>0$,

$$\inf_{\widehat{\vartheta}}\max_{\vartheta\in\{\vartheta^0,\vartheta^1\}}\mathbb{P}_{\vartheta}\left(|\widehat{\vartheta}(x)-\vartheta(x)| \geq \frac{c_1 N^{-\beta/(2\beta+d)}}{2}\right) \geq \frac{1}{4}\cdot\frac{2-\sqrt{3}}{4} =: c_2.$$

We assume without loss of generality that $\|K\|_{L^2(\mathbb{R}^d)} = 1$. Choose ϑ^0 such that the null model is $A_{\vartheta^0} = \Delta$, i.e., $\vartheta = 0$, $c = 0$, and choose ϑ^1 such that the alternatives are $A_{\vartheta^1} = \Delta + \vartheta\cdot\nabla + c$, where $c - \nabla\cdot\vartheta \leq \gamma < 0$ and ϑ is componentwise β-Hölder continuous and a conservative vector field. For $\vartheta\in\{\vartheta^0,\vartheta^1\}$, let $\mathbb{P}_{\vartheta,\delta}$ be the law of X_δ on $\mathcal{H} = L^2([0,T])^M$, let $C_{\vartheta,\delta}$ be its covariance operator, and let $(H_{\vartheta,\delta}, \|\cdot\|_{H_{\vartheta,\delta}})$ be the associated RKHS. For $(f_k)_{k=1}^M\in\mathcal{H}$, we have $C_{\vartheta,\delta}(f_k)_{k=1}^M = (\sum_{l=1}^M C_{\vartheta,\delta,k,l}f_l)_{k=1}^M$ with (cross-) covariance operators $C_{\vartheta,\delta,k,l}\colon L^2([0,T])\to L^2([0,T])$ defined by

$$C_{\vartheta,\delta,k,l}f_l(t) = \mathbb{E}_\vartheta[\langle X_{\delta,l}, f_l\rangle_{L^2([0,T])}X_{\delta,k}(t)], \qquad 0\leq t\leq T.$$

Due to stationarity of X_δ (cf. Assumption O), we have, for $0\leq t\leq T$,

$$C_{\vartheta,\delta,k,l}f_l(t) = \int_0^t c_{\vartheta,\delta,k,l}(t-t')f_l(t')\,\mathrm{d}t' + \int_t^T c_{\vartheta,\delta,l,k}(t'-t)f_l(t')\,\mathrm{d}t',$$

with covariance kernels $c_{\vartheta,\delta,k,l}(t) = \mathbb{E}_\vartheta[X_{\delta,k}(t)X_{\delta,l}(0)]$, $0\leq t\leq T$.

Let $(\sigma_j^2)_{j\geq 1}$ be the strictly positive eigenvalues of $C_{\vartheta^0,\delta}$, and let $(u_j)_{j\geq 1}$ with $u_j = (u_{j,k})_{k=1}^M\in\mathcal{H}$ be a corresponding orthonormal system of eigenvectors. We want to verify the assumption in (B.30), for which we require the following lemma.

LEMMA B.27 (Lemma A.23 in Paper A). *In the above setting, we have*

$$\sum_{j=1}^{\infty} \sigma_j^{-2} \|(C_{\vartheta^0,\delta} - C_{\vartheta^1,\delta})u_j\|^2_{H_{\vartheta^0,\delta}}$$
$$\leq CT \sum_{k,l=1}^{N} \Big(\frac{\|\Delta K\|^4_{L^2(\mathbb{R}^d)}}{\delta^8} \|c_{\vartheta^0,\delta,k,l} - c_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])} + \|c''_{\vartheta^0,\delta,k,l} - c''_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])} \Big)$$

for all $\delta^2 \leq \|\Delta K\|_{L^2(\mathbb{R}^d)}$ and all $T \geq 1$, where $C > 0$ is an absolute constant.

Adapting Lemma A.24 in Paper A to our setting results in another upper bound.

LEMMA B.28. *In the above setting, let $\vartheta^1 \in \Theta$ with $N \geq 1$. Then, there exists a constant $c_3 > 0$, depending only on K and d, such that*

$$\sum_{k,l=1}^{N} \Big(\delta^{-8} \|c_{\vartheta^0,\delta,k,l} - c_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])} + \|c''_{\vartheta^0,\delta,k,l} - c''_{\vartheta^1,\delta,k,l}\|^2_{L^2([0,T])} \Big)$$
$$\leq c_3 \sum_{k=1}^{N} \Big(|\vartheta(x_k)|^2 + \delta^2 \widetilde{c}_\vartheta(x_k)^2 \Big),$$

with $\widetilde{c}_\vartheta = c - \nabla \cdot \vartheta/2 - |\vartheta|^2/4$.

Let $c_4, c_5 > 0$ be constants independent of N and h. Consider a kernel function $V \in C_c^\infty(\mathbb{R}^d; [0,\infty])$ with compact support in $[-1/2, 1/2]^d$. Define the potential

$$\xi(y) = c_4 h^{\beta+1} V((y-x)/h),$$

and let $\vartheta = \nabla \xi$. We consider hence the alternative

$$\vartheta(y) = c_4 h^\beta (\nabla V)\Big(\frac{y-x}{h}\Big)$$

and a reaction function $c \colon \Lambda \to \mathbb{R}_-$ small enough. For $h = c_5 N^{-1/(2\beta+d)}$, we have that

$$c_3 \sum_{k=1}^{N} \Big(|\vartheta(x_k)|^2 + \delta^2 \widetilde{c}_\vartheta(x_k)^2 \Big) \lesssim \sum_{k=1}^{N} |\vartheta(x_k)|^2 \lesssim N h^d h^{2\beta} \lesssim 1.$$

The claim of Theorem B.9 follows now from Lemma B.26 in combination with Lemmas B.27 and B.28 and sufficiently small constants c_4, c_5. ∎

B.5.5 *Remaining proofs*

Remaining proofs for Section B.3

Proof of Corollary B.4. We decompose

$$\widetilde{\vartheta}_\delta(x) = \vartheta(x) - (\mathcal{I}^x_\delta)^{-1} \mathcal{M}^x_\delta \|K\|_{L^2(\mathbb{R}^d)} + (\mathcal{I}^x_\delta)^{-1} \mathcal{R}^x_\delta + (\widehat{a}_\delta - a)(\mathcal{I}^x_\delta)^{-1} \mathcal{U}^x_\delta$$

with

$$\mathcal{U}_\delta^x = \sum_{k=1}^N w_k(x) \int_0^T X_{\delta,k}^\nabla(t) X_{\delta,k}^\Delta(t)\,\mathrm{d}t.$$

Combining Lemma B.17(iii) and Lemma B.18(i), it follows by the arguments given in Section B.5.3 that $\mathcal{U}_\delta^x = O_\mathbb{P}(1)$. Thus, the claim hold once $\widehat{a}_\delta$ satisfies (B.15). Just as the estimator $\widehat{\vartheta}_\delta(x)$ described in (B.10), the estimates in (B.16) can again be decomposed into a bias and martingale part. While the orders of the appearing coefficients differ due to a different scaling in δ, all terms can be controlled with the techniques used in Section B.5.3 and B.5.3. It is therefore straightforward to verify that both given candidates for $\widehat{a}_\delta$ satisfy

$$\widehat{a}_\delta - a \in O_\mathbb{P}(\delta h + \delta^2 + \delta(Nh^d)^{-1/2})$$

and thus fulfill (B.15). ■

Proof of Corollary B.5. By decomposing the integral and using (B.12) from Theorem B.2, we obtain

$$\begin{aligned}\int_\Lambda \Big(\widehat{\vartheta}_\delta(x) - \vartheta(x)\Big)^2 \mathrm{d}x &= \int_{\mathcal{J}} \Big(\widehat{\vartheta}_\delta(x) - \vartheta(x)\Big)^2 \mathrm{d}x + \int_{\Lambda\setminus\mathcal{J}} \Big(\widehat{\vartheta}_\delta(x) - \vartheta(x)\Big)^2 \mathrm{d}x \\ &= O_\mathbb{P}\left(h^{2\beta} + \frac{1}{Nh^d}\right) + \int_{\Lambda\setminus\mathcal{J}} \Big(\widehat{\vartheta}_\delta(x) - \vartheta(x)\Big)^2 \mathrm{d}x.\end{aligned}$$

Due to the decomposition (B.10) and (B.17), it holds for $x \notin \mathcal{J}$ and appropriate $x_0 = x_0(x) \in \mathcal{J}$ that

$$\widehat{\vartheta}_\delta(x) = \widehat{\vartheta}_\delta(x_0) = \vartheta(x) + (\vartheta(x_0) - \vartheta(x)) + O_\mathbb{P}(h^\beta + (Nh^d)^{-1/2}).$$

Thus, plugging this into the previous display yields by the Hölder regularity of ϑ,

$$\begin{aligned}\int_{\Lambda\setminus\mathcal{J}} \Big(\widehat{\vartheta}_\delta(x) - \vartheta(x)\Big)^2 \mathrm{d}x &\lesssim O_\mathbb{P}\left(h^{2\beta} + \frac{1}{Nh^d}\right) + \int_{\Lambda\setminus\mathcal{J}} (\vartheta(x) - \vartheta(x_0))^2\,\mathrm{d}x \\ &\lesssim O_\mathbb{P}\left(h^{2\beta} + \frac{1}{Nh^d}\right) + \int_{\Lambda\setminus\mathcal{J}} \mathrm{dist}^2(x,\mathcal{J})\,\mathrm{d}x \\ &\lesssim O_\mathbb{P}\left(h^{2\beta} + \frac{1}{Nh^d}\right) + d_{\max}^2 \lambda(\Lambda\setminus\mathcal{J}).\end{aligned}$$

■

Proof of Lemma B.7. We use the well-known theory for local polynomial estimators, more specifically, for the local linear case. The one-dimensional case in [43, Chapter 1.6] can be easily extended to the general d-dimensional version. By a first order multivariate Taylor expansion for a function $f\colon \mathbb{R}^d \to \mathbb{R}$, we can write for $y, z \in \mathbb{R}^d$, a multiindex α, and any $h > 0$,

$$f(z) \approx \sum_{0\le|\alpha|\le 1} \frac{D^\alpha f(y)}{\alpha!}(z-y)^\alpha = \xi^\top(y) U\Big(\frac{z-y}{h}\Big),$$

where

$$U(u) = \big((u^\alpha/\alpha!)_{0\le|\alpha|\le 1}\big)^\top, \quad \xi(x) = \Big((D^\alpha f(x) h^{|\alpha|})_{0\le|\alpha|\le 1}\Big)^\top.$$

Modifying [43, Proposition 1.12] and [43, Lemma 1.3] to their multivariate counterparts, it follows that the weights $w_k(x)$ are reproducing of order 1 and satisfy Assumption L(iii) if (LP1)-(LP3) hold true. ■

Remaining proofs for Section B.5.1

Proof of Lemma B.12. Since A_ϑ is elliptic, it follows as in the proof of [4, Proposition A.4], after formally replacing $\Delta_{\vartheta(\delta\cdot)}$ and $\min_x \vartheta(x)$ contained there by $a\Delta$ and the lower bound on the spectrum of a, respectively, that $A^*_{\vartheta,\delta,y}$ is a sectorial operator on $L^2(\Lambda_{\delta,y})$, that is, there exists a constant M, independent of δ and $y \in \mathcal{J}$, such that

$$\|(\lambda I - A^*_{\vartheta,\delta,y})^{-1}\|_{L^2(\Lambda_{\delta,y})} \leq \frac{M}{|\lambda - C\delta^2|}$$

for all $\lambda \in \Sigma_\eta = \{\rho \in \mathbb{C} : |\arg(\rho - C\delta^2)| < \eta\} \setminus \{C\delta^2\}$ with some $\eta \in (\pi/2, \pi)$ or, equivalently, for all $\lambda \in \Sigma_\eta + C\delta^2$,

$$\|(\lambda I + (C\delta^2 - A^*_{\vartheta,\delta,y}))^{-1}\|_{L^2(\Lambda_{\delta,y})} \leq \frac{M}{|\lambda|}.$$

The shifted operator $C\delta^2 - A^*_{\vartheta,\delta,y}$ generates the semigroup $\mathrm{e}^{-C\delta^2 t} S^*_{\vartheta,\delta,y}(t)$, and so the result follows from [35, Proposition 2.1.1]. ∎

Proof of Lemma B.14. The proof is a combination of [4, Proposition 3.5] and Lemma A.14 in Paper A. For fixed $y \in \mathcal{J}$, $u \in \mathbb{R}^d$, it holds by a Feynman–Kac representation that

$$S^*_{\vartheta,\delta,y}(t)z(u) = \widetilde{\mathbb{E}}_u\left[z(Y_t^{(\delta,y)}) \exp\left(\int_0^t \widetilde{c}_{\delta,y}(Y_s^{(\delta,y)})\,\mathrm{d}s\right)\mathbf{1}\left(t < \tau_{\delta,y}(Y^{(\delta,y)})\right)\right],$$

where the process $Y^{(\delta,y)}$ takes the form

$$\mathrm{d}Y_t^{(\delta,y)} = \widetilde{b}_{\delta,y}(Y_t^{(\delta,y)})\,\mathrm{d}t + \sqrt{2}a^{1/2}\,\mathrm{d}\widetilde{W}_t, \quad Y_0^{(\delta,y)} = u \in \mathbb{R}^d,$$

with $\widetilde{b}_{\delta,y}(\cdot) = -\delta\vartheta(y+\delta\cdot)$, $\widetilde{c}_{\delta,y}(\cdot) = \delta^2(c(y+\delta\cdot) - \nabla\cdot\vartheta(y+\delta\cdot))$, a scalar Brownian motion $\widetilde{W}$, and with the stopping times $\tau_{\delta,y} := \inf\{t \geq 0 : Y_t^{(\delta,y)} \notin \Lambda_{\delta,y}\}$.

(i). By upper bounding the transition densities of $Y^{(\delta,y)}$ as in [4, Proposition 3.5(i)], we get

$$\sup_{y\in\mathcal{J}}(S^*_{\vartheta,\delta,y}(t)|z|)(u) \leq c_3 \mathrm{e}^{c_1 t\delta^2}(\mathrm{e}^{c_2 t\Delta}|z|)(u),$$

where the right hand side is in $L^2(\mathbb{R}^d)$.

(ii). By dense approximation, it is enough to consider $z \in C_c(\bar{\Lambda})$ and such that z is supported in $\Lambda_{\delta,y}$ for δ small enough, hence, $z|_{\Lambda_{\delta,y}} = z$. With $(\mathrm{e}^{ta\Delta}z)(u) = \widetilde{\mathbb{E}}_u[z(Y_t^{(0)})]$, decompose

$$S^*_{\vartheta,\delta,y}(t)z(u) - \mathrm{e}^{ta\Delta}z(u) = T_1(y,u) + T_2(y,u) + T_3(y,u)$$

with

$$T_1(y,u) := \widetilde{\mathbb{E}}_u\left[z(Y_t^{(\delta,y)}) - z(Y_t^{(0)})\right],$$
$$T_2(y,u) := \widetilde{\mathbb{E}}_u\left[z(Y_t^{(\delta,y)})\left(\exp\left(\int_0^t \widetilde{c}_{\delta,y}(Y_s^{(\delta,y)})\,\mathrm{d}s\right) - 1\right)\mathbf{1}(t < \tau_{\delta,y}(Y^{(\delta,y)}))\right],$$
$$T_3(y,u) := -\widetilde{\mathbb{E}}_u\left[z(Y_t^{(\delta,y)})\mathbf{1}(t \geq \tau_{\delta,y}(Y^{(\delta,y)}))\right].$$

The arguments in [4, Proposition 3.5(ii)] yield

$$\sup_{y\in\mathcal{J}}|T_1(y,u)|\to 0 \quad \text{and} \quad \sup_{y\in\mathcal{J}}|T_2(y,u)|\to 0,$$

while compactness of $\mathcal{J}$ guarantees for sufficiently small δ the existence of a ball $B_{\rho\delta^{-1}}\subset\bigcap_{y\in\mathcal{J}}\Lambda_{\delta,y}$ with centre 0 and radius $\rho\delta^{-1}$ for some $\rho>0$. Using that the running maximum of a Brownian motion decays exponentially, see, for instance [27, Problem 2.8.3], we conclude similarly to Lemma A.14(ii) in Paper A that

$$\begin{aligned}\sup_{y\in\mathcal{J}}|T_3(y,u)| &= \sup_{y\in\mathcal{J}}|\widetilde{\mathbb{E}}_u\big[z(Y_t)\mathbf{1}(t\geq\tau_{\delta,y}(Y))\big]|\\ &\lesssim \sup_{y\in\mathcal{J}}\widetilde{\mathbb{P}}_u(\tau_{\delta,y}(Y)\leq t)\leq\widetilde{\mathbb{P}}_u(\max_{0\leq s\leq t}|Y_s|\geq\rho\delta^{-1})\\ &\leq\widetilde{\mathbb{P}}_u(\max_{0\leq s\leq t}|\widetilde{W}_s|\geq\widetilde{\rho}\delta^{-1})\leq\delta t^{1/2}C\mathrm{e}^{-C\delta^{-2}t^{-1}}\to 0,\end{aligned}$$

for a modified constant $\widetilde{\rho}$. This implies pointwise, for all $u\in\mathbb{R}^d$,

$$\sup_{y\in\mathcal{J}}|S^*_{\vartheta,\delta,y}(t)z(u)-\mathrm{e}^{ta\Delta}z(u)|\to 0,\quad \delta\to 0.$$

By (i), we know $\sup_{y\in\mathcal{J}}|(S^*_{\vartheta,\delta,y}(t)z)(u)|\in L^2(\mathbb{R}^d)$. Dominated convergence yields the claim.

(iii). We use the decomposition in (ii). The process $Y^{(\delta,y)}$ is independent of δ and $\widetilde{b}_{\delta,y}=0$, $\widetilde{c}_{\delta,y}=0$. This implies $T_1(y,u)=T_2(y,u)=0$ for all $y\in\mathcal{J}$ and $u\in\mathbb{R}^d$. Hölder's inequality thus yields

$$\begin{aligned}\sup_{y\in\mathcal{J}}\|(\bar{S}_{\delta,y}(t)-\mathrm{e}^{ta\Delta})z\|_{L^2(\mathbb{R}^d)} &\leq\sup_{y\in\mathcal{J}}\Big(\|(\bar{S}_{\delta,y}(t)-\mathrm{e}^{ta\Delta})z\|_{L^1(\mathbb{R}^d)}\|(\bar{S}_{\delta,y}(t)-\mathrm{e}^{ta\Delta})z\|_{L^\infty(\mathbb{R}^d)}\Big)^{1/2}\\ &\lesssim\delta^{1/2}t^{1/4}\mathrm{e}^{-\delta^{-2}t^{-1}/2}.\end{aligned}$$

■

Proof of Lemma B.15. While the result matches Lemma A.16 in Paper A, the proof differs as we cannot rely on diagonalisability of $S^*_{\vartheta,\delta,y}(t)$ in the nonparametric framework.

We write $u=V_{\delta,y}z$. Let first $s=0$ such that $H_0^{-s,p}(\mathbb{R}^d)=L^p(\mathbb{R}^d)$. Approximating u by continuous and compactly supported functions, we obtain by Lemma B.14(i) and hypercontractivity of the heat kernel on $\mathbb{R}^d$ uniformly in $y\in\mathcal{J}$

$$\begin{aligned}\|S^*_{\vartheta,\delta,y}(t)u\|_{L^2(\Lambda_{\delta,y})} &\lesssim\mathrm{e}^{c_1t\delta^2}\|\mathrm{e}^{Ct\Delta}|u|\|_{L^2(\mathbb{R}^d)}\\ &\lesssim\mathrm{e}^{c_1t\delta^2}t^{-\gamma}\|u\|_{L^p(\mathbb{R}^d)}\lesssim\mathrm{e}^{c_1t\delta^2}t^{-\gamma}\|z\|_{L^p(\mathbb{R}^d)}.\end{aligned}$$

This yields the result for $s=0$. These inequalities hold also for $p=1$, thus proving the supplement of the statement. For $s>0$ and $p>0$, we apply first Lemma B.13 and then the inequality from the last display to $(2C\delta^2-A^*_{\vartheta,\delta,y})^{-s/2}u$ instead of u. Thus, uniformly in $y\in\mathcal{J}$,

$$\begin{aligned}\|S^*_{\vartheta,\delta,y}(t)u\|_{L^2(\Lambda_{\delta,y})} &=\|(2C\delta^2-A^*_{\vartheta,\delta,y})^{s/2}S^*_{\vartheta,\delta,y}(t)(2C\delta^2-A^*_{\vartheta,\delta,y})^{-s/2}u\|_{L^2(\Lambda_{\delta,y})}\\ &\lesssim\mathrm{e}^{c_1t\delta^{-2}}t^{-s/2}\|S^*_{\vartheta,\delta,y}(t)(2C\delta^2-A^*_{\vartheta,\delta,y})^{-s/2}u\|_{L^2(\Lambda_{\delta,y})}\end{aligned}$$

$$
\begin{aligned}
&\lesssim e^{c_1 t\delta^{-2}} t^{-s/2-\gamma} \|(2C\delta^2 - A^*_{\vartheta,\delta,y})^{-s/2} u\|\|_{\|} L^p(\Lambda_{\delta,y}) \\
&\lesssim e^{c_1 t\delta^{-2}} t^{-s/2-\gamma} \|(-\bar{A}_{\vartheta,\delta,y})^{-s/2} u\|_{L^p(\Lambda_{\delta,y})} \\
&\lesssim e^{c_1 t\delta^{-2}} t^{-s/2-\gamma} \|u\|_{H^{-s,p}(\Lambda_{\delta,y})} \\
&\lesssim e^{c_1 t\delta^{-2}} t^{-s/2-\gamma} V_{\mathrm{op}} \|z\|_{L^p(\Lambda_{\delta,y})}.
\end{aligned}
$$

∎

Remaining proofs for Section B.5.2

Proof of Lemma B.17. Lemma B.15 applied for $s = 2$ shows that, for $v \in \{u, w\}$ and any $\varepsilon > 0$,

$$
\sup_{y \in \mathcal{J}} \|S^*_{\vartheta,\delta,y}(t) v\|_{L^2(\Lambda_{\delta,y})} \lesssim_\varepsilon 1 \wedge t^{-1-d/4+\varepsilon}. \tag{B.31}
$$

(i). Applying (B.31) to u and w, the Cauchy–Schwarz inequality gives for all dimensions $d \geq 1$ that

$$
|f_0(t)| \lesssim \|e^{ta\Delta} u\|_{L^2(\mathbb{R}^d)} \|e^{ta\Delta} w\|_{L^2(\mathbb{R}^d)} \lesssim 1 \wedge t^{-2}.
$$

This yields $f_0 \in L^1([0,\infty))$, proving the claim.

(ii). Lemma B.16 and Lemma B.11(ii) imply that

$$
\delta^{-2} \int_0^T \mathrm{Cov}(\langle X(t), u_{\delta,x}\rangle, \langle X(t), w_{\delta,x}\rangle)\, dt = \int_0^T \int_0^{t\delta^{-2}} f_{t,\delta,y}(t')\, dt'\, dt,
$$

with

$$
f_{t,\delta,y}(t') = \langle S^*_{\vartheta,\delta,y}(t') u, S^*_{\vartheta,\delta,y}(t') w\rangle_{L^2(\Lambda_{\delta,y})} \mathbf{1}(0 \leq t' \leq t\delta^{-2}). \tag{B.32}
$$

Note that $\int_0^T \int_0^\infty f_0(t')\, dt'\, dt = T\psi(u, w)$, and write

$$
\begin{aligned}
&\sup_{y \in \mathcal{J}} \left| \int_0^T \int_0^{t\delta^{-2}} f_{t,\delta,y}(t')\, dt'\, dt - \int_0^T \int_0^\infty f_0(s)\, dt'\, dt \right| \\
&\leq \int_0^T \int_0^{t\delta^{-2}} \sup_{y \in \mathcal{J}} \left| f_{t,\delta,y}(t') - f_0(t') \right| dt'\, dt + \int_0^T \int_{t\delta^{-2}}^\infty |f_0(t')|\, dt'\, dt.
\end{aligned}
$$

Lemma B.14(ii) readily yields the pointwise convergence $|f_{t,\delta,y}(t') - f_0(t')| \to 0$ as $\delta \to 0$, uniformly in $y \in \mathcal{J}$ and for any fixed $t, t' > 0$. Dominated convergence, i.e., (B.31), implies convergence to zero.

(iii). Define $\bar{f}_{t,\delta,y}$ analogously to $f_{t,\delta,y}$ from (B.32), now with respect to the semigroup $\bar{S}_{\delta,y}(t)$. The first step of the proof is to reduce the argument to $\bar{f}_{t,\delta,y}$. More specifically, we will show that

$$
\sup_{y \in \mathcal{J}} \int_0^T \int_0^{t\delta^{-2}} f_{t,\delta,y}(t')\, dt'\, dt = \sup_{y \in \mathcal{J}} \int_0^T \int_0^{t\delta^{-2}} \bar{f}_{t,\delta,y}(t')\, dt'\, dt + O(\delta).
$$

For doing so, consider the decomposition $f_{t,\delta,y}(t') - \bar{f}_{t,\delta,y}(t') = f^{(1)}_{t,\delta,y}(t') + f^{(2)}_{t,\delta,y}(t')$ with

$$
f^{(1)}_{t,\delta,y}(t') := \langle (S^*_{\vartheta,\delta,y}(t') - \bar{S}_{\delta,y}(t')) u, S^*_{\vartheta,\delta,y}(t') w\rangle_{L^2(\Lambda_{\delta,y})} \mathbf{1}(0 \leq t' \leq t\delta^{-2}),
$$

$$f^{(2)}_{t,\delta,y}(t') := \langle \bar{S}_{\delta,y}(t')u, (S^*_{\vartheta,\delta,y}(t') - \bar{S}_{\delta,y}(t'))w\rangle_{L^2(\Lambda_{\delta,y})}\mathbf{1}(0 \le t' \le t\delta^{-2}).$$

The variation of parameters formula, see p. 162 in [14], shows

$$\begin{aligned} S^*_{\vartheta,\delta,y}(t') - \bar{S}_{\delta,y}(t') &= \int_0^{t'} \bar{S}_{\delta,y}(s)\Big(A^*_{\vartheta,\delta,y} - \bar{A}_{\delta,y}\Big)S^*_{\vartheta,\delta,y}(t'-s)\,\mathrm{d}s \\ &= -\delta\int_0^{t'} \bar{S}_{\delta,y}(s)(\vartheta(y+\delta\cdot)\cdot\nabla + \delta\varphi_\vartheta(y+\delta\cdot))S^*_{\vartheta,\delta,y}(t'-s)\,\mathrm{d}s. \end{aligned}$$

Letting $\widetilde{w} = \bar{S}_{\delta,y}(s)S^*_{\vartheta,\delta,y}(t')w$, Lemma B.13 applied for $s = 1/2$ gives

$$\|\nabla\widetilde{w}\|_{L^2(\Lambda_{\delta,y})} \lesssim \|(-\bar{A}_{\delta,y})^{1/2}\bar{S}_{\delta,y}(s)S^*_{\vartheta,\delta,y}(t')w\|_{L^2(\Lambda_{\delta,y})} \lesssim (s)^{-1/2}\|S^*_{\vartheta,\delta,y}(t')w\|_{L^2(\Lambda_{\delta,y})}.$$

Note furthermore that the adjoint of $\vartheta(y+\delta\cdot)\cdot\nabla$ is given by

$$-\vartheta(y+\delta\cdot)\cdot\nabla - \delta\varphi_\vartheta(y+\delta\cdot) - \delta c(y+\delta\cdot).$$

Consequently, integration by parts, the Cauchy–Schwarz inequality and (B.31) show that, for any sufficiently small $\varepsilon > 0$, $s \le T\delta^{-2}$, uniformly in $y \in \mathcal{J}$,

$$\begin{aligned} &\left|\delta^{-1}\int_0^{t\delta^{-2}} f^{(1)}_{t,\delta,y}(t')\,\mathrm{d}t'\right| \qquad \text{(B.33)} \\ &= \left|\int_0^{t\delta^{-2}}\int_0^{t'} \langle \bar{S}_{\delta,y}(s)(\vartheta(y+\delta\cdot)\cdot\nabla + \delta\varphi_\vartheta(y+\delta\cdot))S^*_{\vartheta,\delta,y}(t'-s)u,\, S^*_{\vartheta,\delta,y}(t')w\rangle_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}s\,\mathrm{d}t'\right| \\ &= \left|\int_0^{t\delta^{-2}}\int_s^{t\delta^{-2}} \langle \bar{S}_{\delta,y}(s)(\vartheta(y+\delta\cdot)\cdot\nabla + \delta\varphi_\vartheta(y+\delta\cdot))S^*_{\vartheta,\delta,y}(t'-s)u,\, S^*_{\vartheta,\delta,y}(t')w\rangle_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t'\,\mathrm{d}s\right| \\ &= \left|\int_0^{t\delta^{-2}}\int_s^{t\delta^{-2}} \langle S^*_{\vartheta,\delta,y}(t'-s)u,\, (\vartheta(y+\delta\cdot)\cdot\nabla - \delta c(y+\delta\cdot))\bar{S}_{\delta,y}(s)S^*_{\vartheta,\delta,y}(t')w\rangle_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t'\,\mathrm{d}s\right| \\ &\lesssim \int_0^{t\delta^{-2}}\int_0^{t\delta^{-2}} \|S^*_{\vartheta,\delta,y}(t')u\|_{L^2(\Lambda_{\delta,y})}s^{-1/2}\|S^*_{\vartheta,\delta,y}(t'+s)w\|_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t'\,\mathrm{d}s \lesssim 1. \end{aligned}$$

The bound for $f^{(2)}_{t,\delta,y}$ is obtained similarly. We will conclude by proving that

$$\sup_{y\in\mathcal{J}}\int_0^T\int_0^{t\delta^{-2}} \bar{f}_{t,\delta,y}(t')\,\mathrm{d}t'\,\mathrm{d}t = o(\delta). \tag{B.34}$$

By Assumption L, there exists a compactly supported function z, given by $z = (D^\alpha\bar{K})/a$, such that $u = (-\bar{A})z = (-\bar{A}_{\delta,y})z$ for sufficiently small δ. As $\bar{S}_{\delta,y}(t')$ is self-adjoint,

$$\begin{aligned} \int_0^{t\delta^{-2}} \bar{f}_{t,\delta,y}(t')\,\mathrm{d}t' &= \int_0^{t\delta^{-2}} \langle \bar{S}_{\delta,y}(2t')u, w\rangle_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t' \\ &= \frac{1}{2}\langle (I - \bar{S}_{\delta,y}(2t\delta^{-2}))(-\bar{A}_{\delta,y})^{-1}u, w\rangle_{L^2(\Lambda_{\delta,y})} \\ &= \frac{1}{2}\langle z, w\rangle_{L^2(\Lambda_{\delta,y})} - \frac{1}{2}\langle \bar{S}_{\delta,y}(2t\delta^{-2})z, w\rangle_{L^2(\Lambda_{\delta,y})}. \end{aligned}$$

The first summand vanishes, as can be seen from

$$\begin{aligned}\frac{1}{2}\langle z, w\rangle_{L^2(\Lambda_{\delta,y})} = \frac{1}{2}\langle z, w\rangle_{L^2(\mathbb{R}^d)} &= \int_0^\infty \langle \mathrm{e}^{2ta\Delta}(-a\Delta)z, w\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}t \\ &= \int_0^\infty \langle \mathrm{e}^{ta\Delta}u, \mathrm{e}^{ta\Delta}w\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}t = \psi(u,w) = 0.\end{aligned}$$

Consequently, (B.34) follows from Lemma B.15 such that, uniformly in $y \in \mathcal{J}$,

$$\begin{aligned}\left|\int_0^T \int_0^{t\delta^{-2}} \bar{f}_{t,\delta,y}(t')\,\mathrm{d}t'\,\mathrm{d}t\right| &\leq \int_0^T \frac{1}{2}|\langle \bar{S}_{\delta,y}(2t\delta^{-2})z, w\rangle_{L^2(\Lambda_{\delta,y})}|\,\mathrm{d}t \\ &\lesssim \delta^2 \int_0^{T\delta^{-2}} \|\bar{S}_{\delta,y}(t)z\|_{L^2(\Lambda_{\delta,y})}\|\bar{S}_{\delta,y}(t)w\|_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t \\ &\lesssim \delta^2 \int_0^{T\delta^{-2}} (1\wedge t^{-d/2-1+2\varepsilon})\,\mathrm{d}t = O(\delta^2).\end{aligned}$$

■

Proof of Lemma B.18. Using Wick's theorem (see [25, Theorem 1.28]), write

$$\delta^{-6}\mathrm{Var}\left(\int_0^T \langle X(t), u_{\delta,y}\rangle\langle X(t), w_{\delta,y}\rangle\,\mathrm{d}t\right) = 2V_1 + 2V_2,$$

where $V_1 = V(u,u,w,w)$, $V_2 = V(u,w,w,u)$, and, for $v, v', z, z' \in L^2(\Lambda_{\delta,y})$,

$$\begin{aligned}V(v,v',z,z') &= \delta^{-6}\int_0^T\int_0^t \mathrm{Cov}(\langle X(t), v_{\delta,y}\rangle, \langle X(s), v'_{\delta,y}\rangle)\mathrm{Cov}(\langle X(t), z_{\delta,y}\rangle, \langle X(s), z'_{\delta,y}\rangle)\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_0^T\int_0^{t\delta^{-2}}\int_0^{t\delta^{-2}-s} f_{\delta,y}((s+r),v),(r,v')\,\mathrm{d}r\int_0^{t\delta^{-2}-s} f_{\delta,y}((s+r',z),(r',z'))\,\mathrm{d}r'\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

with

$$f_{\delta,y}((l,v),(l',z)) = \langle S^*_{\vartheta,\delta,y}(l)v, S^*_{\vartheta,\delta,y}(l')z\rangle_{L^2(\Lambda_{\delta,y})}, \qquad \text{for } 0 \leq l, l' \leq T\delta^{-2}.$$

Since the arguments for treating both terms are similar, we restrict ourselves to the upper bound for V_1.

(i). By the Cauchy–Schwarz inequality and (B.31), we find for any $\varepsilon > 0$ that

$$\begin{aligned}\sup_{y\in\mathcal{J}}|f_{\delta,y}((s+r,u),(r,u))| &\lesssim \sup_{y\in\mathcal{J}}\|S^*_{\vartheta,\delta,y}(s+r)u\|_{L^2(\Lambda_{\delta,y})}\sup_{y\in\mathcal{J}}\|S^*_{\vartheta,\delta,y}(r)u\|_{L^2(\Lambda_{\delta,y})} \\ &\lesssim_\varepsilon (1\wedge(s+r)^{-1-d/4+\varepsilon})(1\wedge r^{-1-d/4+\varepsilon}).\end{aligned} \tag{B.35}$$

Similar results are obtained for w. Hence,

$$\begin{aligned}\sup_{y\in\mathcal{J}}|V_1| &\lesssim \int_0^{T\delta^{-2}}(1\wedge s^{-2-d/2+2\varepsilon})\,\mathrm{d}s\int_0^{T\delta^{-2}}(1\wedge r^{-1-d/4+\varepsilon})\,\mathrm{d}r\int_0^{T\delta^{-2}}(1\wedge r'^{-1-d/4+\varepsilon})\,\mathrm{d}r' \\ &\lesssim 1.\end{aligned}$$

(ii). Note that

$$\sup_{y\in\mathcal{J},|y-x|\leq h} \|S^*_{\vartheta,\delta,y}(t)g^{(\vartheta,y,\delta)}\cdot\nabla K\|_{L^2(\Lambda_{\delta,y})} \lesssim h(1\wedge t^{-d/4+\varepsilon}),$$

implying that

$$\sup_{y\in\mathcal{J},|y-x|\leq h} |f_{\delta,y}((s+r,w),(r,w))| \lesssim h^2(1\wedge(s+r)^{-d/4+\varepsilon})(1\wedge r^{-d/4+\varepsilon}).$$

Combining this with (B.20) and (B.35) gives

$$\sup_{y\in\mathcal{J},|y-x|\leq h} |V_1| \lesssim \int_0^{T\delta^{-2}} h^2(1\wedge r^{-d/2+2\varepsilon})\,\mathrm{d}r \lesssim h^2(1\vee\delta^{-2+d-4\varepsilon}) \lesssim h^{2\beta}\delta^{-2}.$$

(iii). The result follows similarly to part (ii), noting now that

$$\sup_{y\in\mathcal{J},|y-x|\leq h} \|S^*_{\vartheta,\delta,y}(t)\varphi_\vartheta(y+\delta\cdot)K\|_{L^2(\Lambda_{\delta,y})} \lesssim (1\wedge t^{-d/4+\varepsilon}),$$

and thus

$$\sup_{y\in\mathcal{J},|y-x|\leq h} |V_1| \lesssim \int_0^{T\delta^{-2}} (1\wedge r^{-d/2+2\varepsilon})\,\mathrm{d}r \lesssim (1\vee\delta^{-2+d-4\varepsilon}) \lesssim h^{2\beta}\delta^{-4}.$$

■

Remaining proofs for Section B.5.3

Proof of Lemma B.23. We start with deriving the following useful upper bound, which holds for any $\varepsilon>0$, and which will be applied several times: Lemma B.15 yields

$$\begin{aligned}\sup_{y\in\mathcal{J},|y-x|\leq h} \|S^*_{\vartheta,\delta,y}(t)g^{(\vartheta,\delta,y)}\cdot\nabla K\|_{L^2(\Lambda_{\delta,y})} &\lesssim_\varepsilon \left(h(1\wedge t^{-1-d/4+\varepsilon})+(h^\beta+\delta)(1\wedge t^{-d/4})\right) \\ &\leq h(1\wedge t^{-d/4}).\end{aligned} \tag{B.36}$$

Indeed, by the Minkowski inequality and (B.23) with the identity function id on $\mathbb{R}^d$,

$$\begin{aligned}&\sup_{y\in\mathcal{J},|y-x|\leq h} \|S^*_{\vartheta,\delta,y}(t)g^{(\vartheta,\delta,y)}\cdot\nabla K\|_{L^2(\Lambda_{\delta,y})} \\ &\lesssim \sup_{y\in\mathcal{J},|y-x|\leq h}\sum_{i=1}^d\sum_{|\alpha|=1}\Big(|(y-x)^\alpha|\|S^*_{\vartheta,\delta,y}(t)\partial_iK\|_{L^2(\mathbb{R}^d)}+\delta\|S^*_{\vartheta,\delta,y}(t)\,\mathrm{id}\,\partial_iK\|_{L^2(\mathbb{R}^d)} \\ &\qquad\qquad + \|S^*_{\vartheta,\delta,y}(t)(D^\alpha(\vartheta_i(\xi)-\vartheta(x))(y+\delta\,\mathrm{id}-x)^\alpha)\partial_iK\|_{L^2(\mathbb{R}^d)}\Big) \\ &\lesssim h(1\wedge t^{-1-d/4+\varepsilon})+(h^\beta+\delta)(1\wedge t^{-d/4}).\end{aligned}$$

The last bound holds by three applications of Lemma B.15, noting that both the functions id ∂_iK and $(D^\alpha(\vartheta_i(\xi)-\vartheta(x))(y+\delta\,\mathrm{id}-x)^\alpha)\partial_iK$ are compactly supported with

$$\|(D^\alpha(\vartheta_i(\xi)-\vartheta(x))(y+\delta\,\mathrm{id}-x)^\alpha)\partial_iK\|_{L^2(\mathbb{R}^d)} \lesssim h^\beta$$

by the Hölder assumption on ϑ. Next, we study the shift from $\bar{S}_{\vartheta,\delta,y}(t)$ to $\mathrm{e}^{ta\Delta}$. Assumption L, the triangle inequality and (B.36) imply

$$\begin{aligned}
&\sum_{k=1}^{N} w_k(x)\langle(\bar{S}_{\delta,x_k}(2s) - \mathrm{e}^{2sa\Delta})\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)} \\
&\quad \leq \sum_{k=1}^{N} |w_k(x)||\langle \bar{S}_{\delta,x_k}(s)\nabla K, \bar{S}_{\delta,x_k}(s) g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)}| \\
&\qquad + \sum_{k=1}^{N} |w_k(x)||\langle \mathrm{e}^{sa\Delta}\nabla K, \mathrm{e}^{sa\Delta} g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)}| \\
&\quad \lesssim h s^{-1-d/2+2\varepsilon}.
\end{aligned}$$

On the other hand, by Lemma B.14(iii), we have for $z \in L^2(\mathbb{R}^d)$

$$\sup_{y\in\mathcal{J}} \|\bar{S}_{\delta,y}(s) z - \mathrm{e}^{sa\Delta} z\|_{L^2(\mathbb{R}^d)} \lesssim \delta^{1/2} s^{1/4} \mathrm{e}^{-\delta^{-2}s^{-1}/2} \lesssim \delta^{6+1/2} s^{3+1/4},$$

using that $\mathrm{e}^{-x} \leq x^{-3}$ for $x > 0$. Thus, by splitting the integral at some $r \in [0, t\delta^{-2}]$, we obtain

$$\begin{aligned}
&\int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^{N} w_k(x)\langle(\bar{S}_{\delta,x_k}(2s) - \mathrm{e}^{2sa\Delta})\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\mathbb{R}^d)} \,\mathrm{d}s\,\mathrm{d}t \\
&\quad \lesssim \int_0^T \int_0^r \delta^{6+1/2} s^{3+1/4} \,\mathrm{d}s\,\mathrm{d}t + \int_0^T \int_r^{t\delta^{-2}} h s^{-1-d/2+2\varepsilon} \,\mathrm{d}s\,\mathrm{d}t \\
&\quad \lesssim \delta^{6+1/2} \int_0^r s^{3+1/4} \,\mathrm{d}s + h \int_r^{T\delta^{-2}} s^{-1-d/2+2\varepsilon} \,\mathrm{d}s \\
&\quad \lesssim \delta^{6+1/2} r^{4+1/4} + h r^{-d/2+2\varepsilon}
\end{aligned} \tag{B.37}$$

for any $\varepsilon > 0$. The choice $r = \delta^{-1}$ yields that the last display is of order $o(\delta^2 + \delta^{d/2})$. We are left with the shift from $S^*_{\vartheta,\delta,x_k}(t)$ to $\bar{S}_{\delta,x_k}(t)$. By the variation of parameters formula, cf. [14, p. 161], we have for $y \in \Lambda$

$$\begin{aligned}
G_{\vartheta,\delta,y}(s) &:= S^*_{\vartheta,\delta,y}(s) - \bar{S}_{\delta,y}(s) = \int_0^s \bar{S}_{\delta,y}(r)(A^*_{\vartheta,\delta,y} - \bar{A}_{\delta,y}) S^*_{\vartheta,\delta,y}(s-r) \,\mathrm{d}r \\
&= -\delta \int_0^s \bar{S}_{\delta,y}(r)(\vartheta(y+\delta\cdot) \cdot \nabla + \delta\varphi_\vartheta(y+\delta\cdot)) S^*_{\vartheta,\delta,y}(s-r) \,\mathrm{d}r.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^N w_k(x) \langle S^*_{\vartheta,\delta,x_k}(s)\nabla K, S^*_{\vartheta,\delta,x_k}(s) g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t \\
&= \int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^N w_k(x) \langle \bar{S}_{\delta,x_k}(2s)\nabla K, g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t \\
&\quad + \int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^N w_k(x) \langle \bar{S}_{\delta,x_k}(s)\nabla K, G_{\vartheta,\delta,x_k}(s) g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t \\
&\quad + \int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^N w_k(x) \langle G_{\vartheta,\delta,x_k}(s)\nabla K, \bar{S}_{\delta,x_k}(s) g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t \\
&\quad + \int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^N w_k(x) \langle G_{\vartheta,\delta,x_k}(s)\nabla K, G_{\vartheta,\delta,x_k}(s) g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t.
\end{aligned}$$

The first summand in the last display has already been examined. We show the desired rate for the second summand. The bound for the other ones is obtained analogously. Arguing as for (B.33), we get

$$\begin{aligned}
&\left| \int_0^T \int_0^{t\delta^{-2}} \sum_{k=1}^N w_k(x) \langle \bar{S}_{\delta,x_k}(s)\nabla K, G_{\vartheta,\delta,x_k}(s) g^{(\vartheta,x_k,\delta)} \cdot \nabla K\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}t \right| \\
&\lesssim \delta \int_0^T \int_0^{t\delta^{-2}} \int_0^{t\delta^{-2}} \sum_{k=1}^N |w_k(x)| \|\bar{S}_{\delta,x_k}(s+s')\nabla K\|_{L^2(\Lambda_{\delta,x_k})} s'^{-1/2} \\
&\qquad\qquad \cdot \|S^*_{\vartheta,\delta,x_k}(s') g^{(\vartheta,x_k,\delta)}\|_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}t \\
&\lesssim \delta \int_0^T \int_0^{t\delta^{-2}} \int_0^{t\delta^{-2}} (1 \wedge (s+s')^{-1-d/4+\varepsilon}) s'^{-1/2} (1 \wedge s'^{-d/4+\varepsilon}) h\,\mathrm{d}s'\,\mathrm{d}s\,\mathrm{d}t \\
&= O(h\delta(1 \vee \delta^{-1/2+d/2-6\varepsilon}))
\end{aligned} \tag{B.38}$$

for any $\varepsilon > 0$. Combining (B.37), (B.38) and (B.20) yields the assertion. ∎

Proof of Proposition B.25. Writing for $u \in L^2(\Lambda)$

$$\langle X(t), u\rangle = \langle S_\vartheta(t) X_0, u\rangle + \langle \bar{X}(t), u\rangle,$$

we obtain the decomposition

$$\begin{aligned}
\mathcal{R}^x_\delta = \bar{\mathcal{R}}^x_\delta &+ \sum_{k=1}^N w_k(x) \int_0^T \langle \bar{X}(t), \nabla K_{\delta,x_k}\rangle \langle S_\vartheta(t) X_0, (\varphi_\vartheta + (\vartheta - \vartheta(x)) \cdot \nabla) K_{\delta,x_k}\rangle\,\mathrm{d}t \\
&+ \sum_{k=1}^N w_k(x) \int_0^T \langle S_\vartheta(t) X_0, \nabla K_{\delta,x_k}\rangle \langle \bar{X}(t), (\varphi_\vartheta + (\vartheta - \vartheta(x)) \cdot \nabla) K_{\delta,x_k}\rangle\,\mathrm{d}t \\
&+ \sum_{k=1}^N w_k(x) \int_0^T \langle S_\vartheta(t) X_0, \nabla K_{\delta,x_k}\rangle \langle S_\vartheta(t) X_0, (\varphi_\vartheta + (\vartheta - \vartheta(x)) \cdot \nabla) K_{\delta,x_k}\rangle\,\mathrm{d}t.
\end{aligned}$$

We only show that the higher order terms are of the desired order. The arguments for the lower order ones, i.e., terms containing φ_ϑ are similar and thus skipped. We hence have to show for all $1 \le i \le d$, using the definition of $g^{(\vartheta,x_k,\delta)}$ in (B.22), that

$$\delta^{-2}\sum_{k=1}^{N} w_k(x)\int_0^T \langle \bar{X}(t), (\partial_i K)_{\delta,x_k}\rangle\langle S_\vartheta(t)X_0, (g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t \tag{B.39}$$

$$+\,\delta^{-2}\sum_{k=1}^{N} w_k(x)\int_0^T \langle S_\vartheta(t)X_0, (\partial_i K)_{\delta,x_k}\rangle\langle \bar{X}(t), (g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t \tag{B.40}$$

$$+\,\delta^{-2}\sum_{k=1}^{N} w_k(x)\int_0^T \langle S_\vartheta(t)X_0, (\partial_i K)_{\delta,x_k}\rangle\langle S_\vartheta(t)X_0, (g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t \tag{B.41}$$

$$= o_{\mathbb{P}}(h^\beta)$$

which is done by controlling the expectations and standard deviations of (B.39), (B.40) and (B.41) separately for a deterministic initial condition $X_0 \in L^p(\Lambda)\cap\mathcal{D}(A_\vartheta)$, $p > 2$, and for the stationary case $X_0 = \int_{-\infty}^0 S_\vartheta(-t')\,\mathrm{d}W(t')$ under the extra constraint that $c - \nabla\cdot\vartheta \le \gamma < 0$.

Case 1: X_0 is deterministic Recalling (B.20), the definition (B.23) and the upper bound (B.36), it holds for the deterministic term (B.41) by Lemma B.15, noting furthermore $K = (-\Delta)\bar{K}$ for some $\bar{K} \in H^4(\mathbb{R}^d)$ with compact support, that

$$\begin{aligned}
&\delta^{-2}\sum_{k=1}^{N} w_k(x)\int_0^T \langle S_\vartheta(t)X_0, (\partial_i K)_{\delta,x_k}\rangle\langle S_\vartheta(t)X_0, (g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t\\
&= \delta^{-2}\sum_{k=1}^{N} w_k(x)\int_0^T \langle (-A_\vartheta)X_0, S_\vartheta^*(t)(-A_\vartheta^*)^{-1}(\partial_i K)_{\delta,x_k}\rangle\langle X_0, S_\vartheta^*(t)(g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t\\
&\lesssim \delta^2\int_0^{T\delta^{-2}} \|A_\vartheta X_0\| \sup_{y\in\mathcal{J},|y-x|\le h} \|S_{\vartheta,\delta,y}^*(t)(A_{\vartheta,\delta,y}^*)^{-1}\partial_i K\|_{L^2(\Lambda_{\delta,y})}\|S_{\vartheta,\delta,y}^*(t)g^{(\vartheta,y,\delta)}\cdot\nabla K\|_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t\\
&\lesssim \delta^2\int_0^{T\delta^{-2}} (1\wedge t^{-1/2-d/4+\varepsilon})\Big(h(1\wedge t^{-1-d/4+\varepsilon}) + (\delta+h^\beta)(1\wedge t^{-d/4})\Big)\,\mathrm{d}t\\
&= o(h^\beta).
\end{aligned}\tag{B.42}$$

The expectations of (B.39) and (B.40) are zero. For its standard deviations, note first that, for any $y \in \mathcal{J}$ with $|y-x| \le h$, $u, v \in L^2(\mathbb{R}^d)$, it holds

$$\begin{aligned}
&\mathrm{Var}\Big(\int_0^T \langle \bar{X}(t), u_{\delta,y}\rangle\langle X_0, S_\vartheta^*(t)v_{\delta,y}\rangle\,\mathrm{d}t\Big)\\
&= 2\int_0^T\int_0^t \langle X_0, S_\vartheta^*(t)v_{\delta,y}\rangle\langle X_0, S_\vartheta^*(s)v_{\delta,y}\rangle\,\mathrm{Cov}\big(\langle\bar{X}(t),u_{\delta,y}\rangle, \langle\bar{X}(s),u_{\delta,y}\rangle\big)\,\mathrm{d}s\,\mathrm{d}t\\
&= 2\int_0^T\int_0^t \langle X_0, S_\vartheta^*(t)v_{\delta,y}\rangle\langle X_0, S_\vartheta^*(s)v_{\delta,y}\rangle\int_0^s \langle S_\vartheta^*(t-r)u_{\delta,y}, S_\vartheta^*(s-r)u_{\delta,y}\rangle\,\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}t\\
&= 2\int_0^T\int_0^t \langle X_0, S_\vartheta^*(t)v_{\delta,y}\rangle\langle X_0, S_\vartheta^*(s)v_{\delta,y}\rangle\int_0^s \langle S_\vartheta^*(t-s+r)u_{\delta,y}, S_\vartheta^*(r)u_{\delta,y}\rangle\,\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}t
\end{aligned}\tag{B.43}$$

$$= 2\int_0^T\int_0^t\int_0^{t-s}\langle X_0, S_\vartheta^*(t)v_{\delta,y}\rangle\langle X_0, S_\vartheta^*(t-s)v_{\delta,y}\rangle\langle S_\vartheta^*(s+r)u_{\delta,y}, S_\vartheta^*(r)u_{\delta,y}\rangle \,\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}t. \quad \text{(B.44)}$$

Applying the scaling Lemma B.11 to (B.43) with $v = \partial_i K$ and $u = g^{(\vartheta,y,\delta)}\cdot\nabla K$, followed by multiple applications of the Cauchy–Schwarz inequality and Lemma A.16, thus yields by (B.20)

$$\begin{aligned}&\sup_{y\in\mathcal{J},|y-x|\le h}\operatorname{Var}\left(\int_0^T\langle S_\vartheta(t)X_0,(\partial_i K)_{\delta,y}\rangle\langle\bar{X}(t), g^{(\vartheta,y,\delta)}\cdot\nabla K)_{\delta,y}\rangle\,\mathrm{d}t\right)\\&\quad\lesssim \delta^6\int_0^{T\delta^{-2}}\left(h^2(1\wedge r^{-2-d/2+2\varepsilon})+(h^{2\beta}+\delta^2)(1\wedge r^{-d/2})\right)\mathrm{d}r = o(\delta^4h^{2\beta}).\end{aligned}$$

Hence,

$$\begin{aligned}&\operatorname{Var}\left(\delta^{-2}\sum_{k=1}^N w_k(x)\int_0^T\langle S_\vartheta(t)X_0,(\partial_i K)_{\delta,x_k}\rangle\langle\bar{X}(t),(g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t\right)\\&\quad\lesssim\delta^{-4}\sum_{k=1}^N|w_k(x)|\operatorname{Var}\left(\int_0^T\langle S_\vartheta(t)X_0,(\partial_i K)_{\delta,x_k}\rangle\langle\bar{X}(t),g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t\right)\\&\quad= o(h^{2\beta}).\end{aligned}\quad\text{(B.45)}$$

Analogue calculations with $u = \partial_i K$, $v = g^{(\vartheta,y,\delta)}\cdot\nabla K$ applied to (B.44) also imply

$$\operatorname{Var}\left(\delta^{-2}\sum_{k=1}^N w_k(x)\int_0^T\langle\bar{X}(t),(\partial_i K)_{\delta,x_k}\rangle\langle S_\vartheta(t)X_0,(g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t\right) = o(h^{2\beta}).\quad\text{(B.46)}$$

Combining (B.42), (B.45) and (B.46) yields the claim.

Case 2: X is stationary Itô's isometry implies again that the expectations of (B.39) and (B.40) are zero, while the expected value of (B.41) is bounded by

$$\begin{aligned}&\delta^{-2}\sum_{k=1}^N w_k(x)\int_0^T\mathbb{E}[\langle S_\vartheta(t)X_0,(\partial_i K)_{\delta,x_k}\rangle\langle S_\vartheta(t)X_0,(g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle]\,\mathrm{d}t\\&= \delta^{-2}\sum_{k=1}^N w_k(x)\int_0^T\int_0^\infty\langle S_\vartheta^*(t+t')(\partial_i K)_{\delta,x_k}, S_\vartheta^*(t+t')(g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t'\,\mathrm{d}t\\&\lesssim\delta^2\int_0^{T\delta^{-2}}\int_0^\infty\sup_{y\in\mathcal{J},|y-x|\le h}\|S_{\vartheta,\delta,y}^*(t+t')\partial_i K\|_{L^2(\Lambda_{\delta,y})}\|S_{\vartheta,\delta,y}^*(t+t')g^{(\vartheta,y\delta)\cdot\nabla K}\|_{L^2(\Lambda_{\delta,y})}\,\mathrm{d}t'\,\mathrm{d}t\\&\lesssim\delta^2\int_0^{T\delta^{-2}}\left(h(1\wedge t^{-1-d/4+\varepsilon})+(\delta+h^\beta)(1\wedge t^{-d/4})\right)\mathrm{d}t = o(h^\beta).\end{aligned}$$

We can bound the variance of (B.39) again by

$$\operatorname{Var}\left(\delta^{-2}\sum_{k=1}^N w_k(x)\int_0^T\langle\bar{X}(t),(\partial_i K)_{\delta,x_k}\rangle\langle S_\vartheta(t)X_0,(g^{(\vartheta,x_k,\delta)}\cdot\nabla K)_{\delta,x_k}\rangle\,\mathrm{d}t\right)$$

$$\lesssim \delta^{-4} \sup_{y\in\mathcal{J},|y-x|\le h} \operatorname{Var}\left(\int_0^T \langle \bar{X}(t), (\partial_i K)_{\delta,y}\rangle \langle S_\vartheta(t)X_0, (g^{(\vartheta,y,\delta)}\cdot\nabla K)_{\delta,y}\rangle \,\mathrm{d}t\right)$$

similar to the deterministic case. Since $\varphi_\vartheta = c - \nabla\cdot\vartheta \le \gamma$ for some $\gamma < 0$ as assumed, the upper bound in Lemma B.15 holds with $\mathrm{e}^{-\gamma t\delta^2}$, i.e., $c_1 = -\gamma$. By similar calculations as in Lemma B.18, i.e., by Wick's Theorem, and using again Itô's isometry we get

$$\begin{aligned}
&\sup_{y\in\mathcal{J},|y-x|\le h} \operatorname{Var}\left(\int_0^T \langle \bar{X}(t), (\partial_i K)_{\delta,y}\rangle \langle S_\vartheta(t)X_0, (g^{(\vartheta,y,\delta)}\cdot\nabla K)_{\delta,y}\rangle \,\mathrm{d}t\right)\\
&= 2\int_0^T\int_0^t \operatorname{Cov}\big(\langle \bar{X}(t), (\partial_i K)_{\delta,y}\rangle, \langle \bar{X}(s), (\partial_i K)_{\delta,y}\rangle\big)\\
&\qquad\cdot \operatorname{Cov}\Big(\langle S_\vartheta(t)X_0, (g^{(\vartheta,y,\delta)\cdot K)_{\delta,y}}\rangle, \langle S_\vartheta(s)X_0, (g^{(\vartheta,y,\delta)\cdot K)_{\delta,y}}\rangle\Big)\,\mathrm{d}s\,\mathrm{d}t\\
&= 2\int_0^T\int_0^t\int_0^s \langle S_\vartheta^*(t-r)(\partial_i K)_\delta, y, S_\vartheta^*(s-r)(\partial_i K)_{\delta,y}\rangle\,\mathrm{d}r\\
&\qquad\cdot\int_0^\infty \langle S_\vartheta^*(t+r')(g^{(\vartheta,y\delta)}\cdot\nabla K)_{\delta,y}, S_\vartheta^*(s+r')(g^{(\vartheta,y\delta)}\cdot\nabla K)_{\delta,y}\rangle\,\mathrm{d}r'\,\mathrm{d}s\,\mathrm{d}t\\
&\lesssim \delta^6\int_0^{T\delta^{-2}} (1\wedge t^{-1-d/4+\varepsilon})\,\mathrm{d}t\int_0^{T\delta^{-2}}(1\wedge s^{-1-d/4+\varepsilon})\,\mathrm{d}s\\
&\qquad\cdot\int_0^\infty \Big(h^2(1\wedge r^{-2-d/2+2\varepsilon}) + (h^{2\beta}+\delta^2)e^{-\gamma r\delta^2}(1\wedge r^{-d/2})\Big)\,\mathrm{d}r.
\end{aligned}$$

If $d\ge 3$, the last display is already of order $o(\delta^4 h^{2\beta})$. For $d\le 2$, we bound $\mathrm{e}^{-\gamma r\delta^2}\lesssim r^{-1/2-\varepsilon}\delta^{-1-2\varepsilon}$, and hence

$$\delta^6\Big(h^2 + (h^{2\beta}+\delta^2)\Big)\delta^{-1-2\varepsilon} = o(\delta^4 h^{2\beta})$$

by (B.20). Similar calculations also hold for the standard deviations of (B.40) and (B.41), implying the claim. ■

Remaining proofs for Section B.5.4

Proof of Lemma B.28. Define the integral kernels

$$\kappa_{k,l}(t) = c_{\vartheta^0,\delta,k,l}(t) - c_{\vartheta^1,\delta,k,l}(t).$$

It suffices to derive the upper bound for the L^2-norm of $\kappa_{k,l}$, as the proof remains valid if one replaces K_{δ,x_k} by $\delta^{-4}(A^2_{\vartheta,\delta,x_k}K)_{\delta,x_k}$. This also gives the desired upper bound on the L^2-norm of $\kappa''_{k,l}(t)$. Following the structure as in the proof of Lemma A.24 in Paper A, we start by some initial notation and the diagonalizability of the semigroup $S^*_{\vartheta,\delta,x_k}(t)$. We write Δ and $\mathrm{e}^{t\Delta}$ for the Laplacian and its generated semigroup on $L^2(\Lambda)$, as well as $\Delta_{\delta,x}$ and $\mathrm{e}^{t\Delta_{\delta,x}}$ on $L^2(\Lambda_{\delta,x})$, and Δ_0 and $\mathrm{e}^{t\Delta_0}$ on $L^2(\mathbb{R}^d)$. We have that

$$A^*_{\vartheta^1} = \Delta - \vartheta\cdot\nabla + (c - \nabla\cdot\vartheta).$$

Given that ϑ is a conservative vector field, we choose a potential ξ such that $\nabla\xi(x) = \vartheta(x)/2$ for some function ξ. By [18, Example 10], $A^*_{\vartheta^1}$ is diagonalizable, i.e.,

$$U^{-1}_{\vartheta^1}A^*_{\vartheta^1}U_{\vartheta^1}z = \Delta z + \widetilde{c}_\vartheta z$$

with the multiplication operator $(U_{\vartheta^1}z)(x) = \mathrm{e}^{\nabla\xi(x)}z(x)$ and $\widetilde{c}_\vartheta = c - \frac{\nabla\cdot\vartheta}{2} - \frac{|\vartheta|^2}{4} \le 0$ due to the choice of ϑ^1. [14, Example 2.1 in Section II.2] and the rescaling Lemma B.11 furthermore imply that

$$S^*_{\vartheta^1,\delta,x_k}(t) = U^{-1}_{\vartheta^1,\delta,x_k}\mathrm{e}^{t\Delta_{\delta,x_k}}U_{\vartheta^1,\delta,x_k}\mathrm{e}^{t\delta^2\widetilde{c}_\vartheta(x_k+\delta x)}, \quad S_{\vartheta^0,\delta,x_k}(t) = \mathrm{e}^{t\Delta_{\vartheta,\delta,x_k}}$$

with $U_{\vartheta^1,\delta,x_k}(x) = U_{\vartheta^1}(x_k+\delta x)$. Note that

$$\mathrm{e}^{t\Delta} = U_{\vartheta^1}(x_k)^{-1}\mathrm{e}^{t\Delta}U_{\vartheta^1}(x_k).$$

We decompose $\kappa_{k,l} = \sum_{j=1}^4 \kappa^{(j)}_{k,l}$, with

$$\begin{aligned}
\kappa^{(1)}_{k,l}(t) &= \int_0^\infty \langle U_{\vartheta^1}(x_k)^{-1}\mathrm{e}^{(t+t')\Delta}(U_{\vartheta^1}(x_k) - U_{\vartheta^1}\mathrm{e}^{\widetilde{c}_{\vartheta^1}(t+t')})K_{\delta,x_k}, \mathrm{e}^{t'\Delta}K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa^{(2)}_{k,l}(t) &= \int_0^\infty \langle (U_{\vartheta^1}(x_k)^{-1} - U^{-1}_{\vartheta^1})U_{\vartheta^1}S^*_{\vartheta^1}(t+t')K_{\delta,x_k}, \mathrm{e}^{t'\Delta}K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa^{(3)}_{k,l}(t) &= \int_0^\infty \langle S^*_{\vartheta^1}(t+t')K_{\delta,x_k}, U_{\vartheta^1}(x_k)^{-1}\mathrm{e}^{t'\Delta}(U_{\vartheta^1}(x_k) - U_{\vartheta^1}\mathrm{e}^{\widetilde{c}_{\vartheta^1}t'})K_{\delta,x_l}\rangle\,\mathrm{d}t',\\
\kappa^{(4)}_{k,l}(t) &= \int_0^\infty \langle S^*_{\vartheta^1}(t+t')K_{\delta,x_k}, (U_{\vartheta^1}(x_k)^{-1} - U^{-1}_{\vartheta^1})\mathrm{e}^{t'\Delta}U_{\vartheta^1}\mathrm{e}^{\widetilde{c}_{\vartheta^1}t'}K_{\delta,x_l}\rangle\,\mathrm{d}t'.
\end{aligned}$$

It suffices to show that $\sum_{1\le k,l\le N}\|\kappa^{(j)}_{k,l}\|^2_{L^2([0,T])} \le c_3\delta^8\sum_{1\le k\le N}(|\vartheta(x_k)|^2 + \delta^2\widetilde{c}_\vartheta(x_k)^2)$ for $j = 1,2$. The arguments for $j = 3,4$ are similar and therefore skipped. Diagonal (i.e., $k = l$) and off-diagonal (i.e., $k \ne l$) terms are treated separately. Set $K_{k,l} = K(\cdot + \delta^{-1}(x_k - x_l))$. Lemma B.14 yields

$$\begin{aligned}
\sup_{y\in\operatorname{supp}K}|(\mathrm{e}^{t\Delta_{\delta,x_k}}K_{k,l})(y)| &\lesssim \sup_{y\in\operatorname{supp}K}|(\mathrm{e}^{t\Delta_0}|K_{k,l}|)(y)|\\
&= \sup_{y\in\operatorname{supp}K}\int_{\mathbb{R}^d}(4\pi t)^{-d/2}\exp(-|x-y|^2/(4t))|K_{k,l}(x)|\,\mathrm{d}x\\
&\le (4\pi t)^{-d/2}\mathrm{e}^{-c'\frac{|x_k-x_l|^2}{\delta^2 t}}\|K\|_{L^1(\mathbb{R}^d)} \lesssim t^{-d/2}\mathrm{e}^{-c'\frac{|x_k-x_l|^2}{\delta^2 t}}, \qquad \text{(B.47)}
\end{aligned}$$

for some $c' > 0$.

Case $j = 1$. We start with scaling as in Lemma B.11 and changing variables such that, using the multiplication operators

$$V_{t,t',\delta,k}(x) = 1 - \mathrm{e}^{\widetilde{c}_{\vartheta^1}(x_k+\delta x)\delta^2(t+t') - \xi(x_k) + \xi(x_k+\delta x)},$$

$$\begin{aligned}
\kappa^{(1)}_{k,l}(t\delta^2) &= \delta^2\int_0^\infty \langle \mathrm{e}^{(t+t')\Delta_{\delta,x_k}}V_{t,t',\delta,k}K, \mathrm{e}^{t'\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'\\
&= \delta^2\int_0^\infty \langle \mathrm{e}^{(t/2+t')\Delta_{\delta,x_k}}V_{t,t',\delta,k}K, \mathrm{e}^{(t/2+t')\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'. \qquad \text{(B.48)}
\end{aligned}$$

Since K is compactly supported and $\widetilde{c}_{\vartheta^1} \le 0$, $V_{t,t',\delta,k}$ can be extended to smooth multiplication operators with operator norms bounded by $v_{t,t',\delta,k} = -\widetilde{c}_{\vartheta^1}(x_k)\delta^2(t+t') + |\vartheta(x_k)|\delta$. (This can be seen from a Taylor expansion, using the Hölder smoothness assumptions for the higher order Taylor terms.) Recalling $K = \Delta^2\widetilde{K}$, Lemma B.15 gives, for any $\epsilon' > 0$,

$$\begin{aligned}
|\kappa_{k,l}^{(1)}(t\delta^2)| &\le \delta^2 \int_0^\infty \|e^{(t/2+t')\Delta_{\delta,x_k}} V_{t,t',\delta,k} K\|_{L^2(\Lambda_{\delta,x_k})} \|e^{(t/2+t')\Delta_{\delta,x_k}} K_{k,l}\|_{L^2(\Lambda_{\delta,x_k})} \, dt' \\
&\le \delta^2 \int_0^\infty v_{t,t',\delta,k} (1 \wedge (t+t')^{-4-d/2+\epsilon'}) \, dt' \\
&\lesssim \delta^3 (-\widetilde{c}_{\vartheta^1}(x_k)\delta + |\vartheta(x_k))| (1 \wedge t^{-1-d/2}) \\
&\le \delta^3 (|\vartheta(x_k)| + |\delta c_\vartheta(x_k)|)(1 \wedge t^{-1-d/2}).
\end{aligned}$$

Changing variables therefore proves for the sum of diagonal terms

$$\sum_{1\le k\le N} \|\kappa_{k,k}^{(1)}\|^2_{L^2([0,T])} \lesssim \sum_{1\le k\le M} \delta^8 |\vartheta(x_k)|^2 + \delta^{10} |c_\vartheta(x_k)|^2.$$

Using Lemma B.15, the integrand in (B.48) can be bounded as follows,

$$\langle e^{(t/2+t')\Delta_{\delta,x_k}} V_{t,t',\delta,k} K, e^{(t/2+t')\Delta_{\delta,x_k}} K_{k,l} \rangle_{L^2(\Lambda_{\delta,x_k})} \lesssim v_{t,t',\delta,k} (1 \wedge (t+t')^{-4-d/2+\epsilon'}).$$

On the other hand, using (B.47), it also satisfies the bound

$$\begin{aligned}
\langle e^{(t/2+t')\Delta_{\delta,x_k}} V_{t,t',\delta,k} K, e^{(t/2+t')\Delta_{\delta,x_k}} K_{k,l} \rangle_{L^2(\Lambda_{\delta,x_k})} &= \langle V_{t,t',\delta,k} K, e^{(t+2t')\Delta_{\delta,x_k}} K_{k,l} \rangle_{L^2(\Lambda_{\delta,x_k})} \\
&\lesssim \|V_{t,t',\delta,k} K\|_{L^1(\mathbb{R}^d)} \sup_{y\in \operatorname{supp} K} \left| (e^{(t+2t')\Delta_{\delta,x_k}} K_{k,l})(y) \right| \\
&\lesssim v_{t,t',\delta,k} (t')^{-d/2} \exp\left(-c' \frac{|x_k - x_l|^2}{\delta^2 t} \right).
\end{aligned}$$

With respect to the off-diagonal terms, we therefore have, using the inequality $\min(a,b) \le a^{1-\varepsilon} b^\varepsilon$ for $a, b \ge 0$,

$$\begin{aligned}
\kappa_{k,l}^{(1)}(t\delta^2) &= \delta^2 \int_0^\infty \langle V_{t,t',\delta,k} K, e^{(t+2t')\Delta_{\delta,x_k}} K_{k,l} \rangle_{L^2(\Lambda_{\delta,x_k})} \, dt' \\
&\lesssim \delta^2 \int_0^\infty v_{t,t',\delta,k}^{1-\varepsilon} (1 \wedge (t+t')^{-4-d/2+\epsilon'})^{1-\varepsilon} \sup_{y\in \operatorname{supp} K} \left| (e^{(t+2t')\Delta_{\delta,x_k}} K_{k,l})(y) \right|^\varepsilon dt' \\
&\lesssim \delta^3 (|\vartheta(x_k)| + \delta |\widetilde{c}_\vartheta(x_k)|)(1 \wedge t^{-1-d/2}) e^{-\varepsilon c' \frac{|x_k - x_l|^2}{\delta^2 t}}.
\end{aligned} \tag{B.49}$$

Applying the bound

$$\int_0^\infty t^{-p-1} e^{-a/t} \, dt = a^{-p} \int_0^\infty t^{-p-1} e^{-1/t} \, dt \lesssim a^{-p}$$

to $p = 1 + d > d$ and $a = c'\epsilon\delta^{-2}|x_k - x_l|^2$, we obtain

$$\begin{aligned}
\int_0^T \kappa_{k,l}^{(1)}(t)^2 \, dt &\lesssim \delta^8 (|\vartheta(x_k)| + \delta|\widetilde{c}_\vartheta(x_k)|)^2 \int_0^\infty t^{-2-d} e^{-c'\epsilon \frac{|x_k - x_l|^2}{\delta^2 t}} \, dt \\
&\lesssim \frac{\delta^{10+2d} (|\vartheta(x_k)|^2 + \delta^2 |\widetilde{c}_\vartheta(x_k)|^2)}{|x_k - x_l|^{2+2d}}.
\end{aligned} \tag{B.50}$$

Recalling that the x_k are δ-separated, we get from Lemma B.29 below that

$$\sum_{1\le k\neq l\le N}\|\kappa^{(1)}_{k,l}\|^2_{L^2([0,T])}\lesssim \delta^{10+2d}\sum_{k=1}^N\Big(|\vartheta(x_k)|^2+\delta^2|\widetilde{c}_\vartheta(x_k)|^2\Big)\sum_{l=1,l\neq k}^N\frac{1}{|x_k-x_l|^{2+2d}}$$
$$\lesssim \delta^8\sum_{k=1}^N|\vartheta(x_k)|^2+\delta^2|\widetilde{c}_\vartheta(x_k)|^2.$$

Together with the bounds for the diagonal terms, this yields, for a constant C depending only on K,

$$\sum_{1\le k,l\le N}\|\kappa^{(1)}_{k,l}\|^2_{L^2([0,T])}\le C\delta^8\sum_{1\le k\le N}|\vartheta(x_k)|^2+\delta^2|\widetilde{c}_\vartheta(x_k)|^2.$$

Case $j=2$. As in the previous case, we have

$$\kappa^{(2)}_{k,l}(t\delta^2)=\delta^2\int_0^\infty\langle(\mathrm{e}^{\xi(x_k+\delta x)-\xi(x_k)}-1)S^*_{\vartheta^1,\delta,x_k}(t+t')K,\mathrm{e}^{t'\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\,\mathrm{d}t'.$$

Using the Cauchy–Schwarz inequality, Lemma B.14(i) and Lemma B.15 with $K=\Delta^2\widetilde{K}$, we get for any $\epsilon>0$

$$\begin{aligned}&\langle(\mathrm{e}^{\xi(x_k+\delta x)-\xi(x_k)}-1)S^*_{\vartheta^1,\delta,x_k}(t+t')K,\mathrm{e}^{t'\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\\&\qquad\lesssim\delta|\vartheta(x_k)|(1\wedge(t+t')^{-2-d/4+\epsilon})\||x|\mathrm{e}^{t'\Delta_0}|K_{k,l}|\|_{L^2(\mathbb{R}^d)}.\end{aligned}\tag{B.51}$$

Note that $K_{k,l}\in C^1_c(\mathbb{R}^d)$ such that $|K_{k,l}|\in H^{1,\infty}(\mathbb{R}^d)$ and $\nabla|K_{k,l}|\in L^\infty(\mathbb{R}^d)$ with compact support. Using now [4, Lemma A.2(ii)] to the extent that

$$x(\mathrm{e}^{t'\Delta_0}|K_{k,l}|)(x)=(\mathrm{e}^{t'\Delta_0}(-2t'\nabla|K_{k,l}|+x|K_{k,l}|))(x),$$

we find that the $L^2(\mathbb{R}^d)$-norm in (B.51) is uniformly bounded in $t'>0$. Hence, $|\kappa^{(2)}_{k,l}(t\delta^2)|\lesssim\delta^3|\vartheta(x_k)|(1\wedge t^{-1/2-d/4-\epsilon})$, and changing variables shows for the sum of diagonal terms

$$\sum_{1\le k\le N}\|\kappa^{(2)}_{k,k}\|^2_{L^2([0,T])}\lesssim\delta^8\sum_{1\le k\le N}|\vartheta(x_k)|^2.$$

Regarding the off-diagonal terms, we have similarly for some $\bar{K}\in L^\infty(\mathbb{R}^d)$ having compact support

$$\begin{aligned}&\Big|\langle(\mathrm{e}^{\xi(x_k+\delta x)-\xi(x_k)}-1)S^*_{\vartheta^1,\delta,x_k}(t+t')K,\mathrm{e}^{t'\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\Big|\\&\quad=\Big|\langle K,S_{\vartheta^1,\delta,x_k}(t+t')(\mathrm{e}^{\xi(x_k+\delta x)-\xi(x_k)}-1)\mathrm{e}^{t'\Delta_{\delta,x_k}}K_{k,l}\rangle_{L^2(\Lambda_{\delta,x_k})}\Big|\\&\quad\lesssim\delta|\vartheta(x_k)|\|K\|_{L^1(\mathbb{R}^d)}\sup_{y\in\operatorname{supp}K}\Big|\Big(\mathrm{e}^{(t+t')\Delta_0}|x|\mathrm{e}^{t'\Delta_0}|K_{k,l}|\Big)(y)\Big|\\&\quad\lesssim\delta|\vartheta(x_k)|(1\vee t')\sup_{y\in\operatorname{supp}K}\Big|\Big(\mathrm{e}^{(t+2t')\Delta_0}|\bar{K}_{k,l}|\Big)(y)\Big|\\&\quad\lesssim\delta|\vartheta(x_k)|(1\vee t')t^{-d/2}\mathrm{e}^{-c'\frac{|x_k-x_l|^2}{\delta^2 t}},\end{aligned}$$

using (B.47). Arguing as for (B.49) and (B.50), we then find from combining the last display with (B.51) that $|\kappa_{k,l}^{(4)}(t\delta^2)| \lesssim \delta^3|\vartheta(x_k)|t^{-1/2-d/4-\epsilon'}\mathrm{e}^{-\epsilon c'\frac{|x_k-x_l|^2}{\delta^2 t}}$ for some $\epsilon, \epsilon' > 0$ and

$$\int_0^T \kappa_{k,l}^{(2)}(t)^2\,\mathrm{d}t \lesssim \frac{\delta^{8+d+4\epsilon}|\vartheta(x_k)|}{|x_k - x_l|^{4\epsilon+d}}.$$

So, all in all, for diagonal and off-diagonal terms,

$$\sum_{1\le k,l\le N} \|\kappa_{k,l}^{(2)}\|_{L^2([0,T])}^2 \le C\delta^8 \sum_{1\le k\le N} |\vartheta(x_k)|^2,$$

for a constant C depending only on K. ■

LEMMA B.29 (Lemma A.27 in Paper A). *Let $x_1, \ldots, x_N$ be δ-separated points in $\mathbb{R}^d$, and let $p > d$. Then, for a constant $C = C(d, p)$,*

$$\sum_{k=2}^N \frac{1}{|x_1 - x_k|^p} \le C\delta^{-p}.$$

Acknowledgements The authors gratefully acknowledge financial support provided by the Carlsberg Foundation Young Researcher Fellowship grant CF20-0604 “Exploring the potential of nonparametric modelling of complex systems via SPDEs”.

References

[1] M. Aerts and G. Claeskens. “Local Polynomial Estimation in Multiparameter Likelihood Models”. In: *Journal of the American Statistical Association* 92.440 (1997), pp. 1536–1545.

[2] R. Altmeyer, T. Bretschneider, J. Janák, and M. Reiß. “Parameter Estimation in an SPDE Model for Cell Repolarisation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1 (2022), pp. 179–199.

[3] R. Altmeyer, I. Cialenco, and G. Pasemann. “Parameter estimation for semilinear SPDEs from local measurements”. In: *Bernoulli* 29.3 (2023), pp. 2035–2061.

[4] R. Altmeyer and M. Reiß. “Nonparametric estimation for linear SPDEs from local measurements”. In: *Annals of Applied Probability* 31.1 (2021), pp. 1–38.

[5] T. Aspelmeier, A. Egner, and A. Munk. “Modern statistical challenges in high-resolution fluorescence microscopy”. In: *Annual Reviews of Statistics and Its Applications* 2 (2015), pp. 163–202.

[6] A. S. Backer and W. E. Moerner. “Extending Single-Molecule Microscopy Using Optical Fourier Processing”. In: *The Journal of Physical Chemistry B* 118.28 (2014), pp. 8313–8329.

[7] I. Cialenco. “Statistical inference for SPDEs: an overview”. In: *Statistical Inference for Stochastic Processes* 21.2 (2018), pp. 309–329.

[8] I. Cialenco and N. Glatt-Holtz. “Parameter estimation for the stochastically perturbed Navier–Stokes equations”. In: *Stochastic Processes and their Applications* 121.4 (2011), pp. 701–724.

[9] I. Cialenco, S. V. Lototsky, and J. Pospíšil. “Asymptotic properties of the maximum likelihood estimator for stochastic parabolic equations with additive fractional Brownian motion”. In: *Stochastics and Dynamics* 9.02 (2009), pp. 169–185.

[10] L. Clarotto, D. Allard, T. Romary, and N. Desassis. “The SPDE approach for spatio-temporal datasets with advection and diffusion”. In: *Spatial Statistics* (2024), Paper No. 100847.

[11] R. Cont. “Modeling Term Structure Dynamics: An Infinite Dimensional Approach”. In: *International Journal of Theoretical and Applied Finance* 08.03 (2005), pp. 357–380.

[12] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.

[13] G. Denaro, D. Valenti, A. La Cognata, B. Spagnolo, A. Bonanno, G. Basilone, S. Mazolla, S. W. Zgozi, S. Aronica, and C. Brunet. “Spatio-temporal behaviour of the deep chlorophyll maximum in Mediteran Sea: Development of a stochastic model for picophytoplankton dynamics”. In: *Ecological Complexity* 13 (2013), pp. 21–34.

[14] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer, 2000.

[15] J. Fan, M. Farmen, and I. Gijbels. “Local Maximum Likelihood Estimation and Inference”. In: *Journal of the Royal Statistical Society* 60.3 (1998), pp. 591–608.

[16] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, 1996.

[17] S. Gaudlitz and M. Reiß. "Estimation for the reaction term in semi-linear SPDEs under small diffusivity". In: *Bernoulli* 29.4 (2023), pp. 3033–3058.

[18] S. Giani, L. Grubišić, A. Międlar, and J. S. Ovall. "Robust error estimates for approximations of non-self-adjoint eigenvalue problems". In: *Numerische Mathematik* 133 (2016), pp. 471–495.

[19] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002.

[20] M. Hairer. *An Introduction to Stochastic PDEs*. 2023. arXiv: 0907.4178 [math.PR].

[21] F. Hildebrandt and M. Trabs. "Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations". In: *Stochastic Processes and their Applications* 162 (2023), pp. 171–217.

[22] F. Hildebrandt and M. Trabs. "Parameter estimation for SPDEs based on discrete observations in time and space." In: *Electronic Journal of Statistics* 15 (2021), pp. 2716–2776.

[23] M. Huebner and B. Rozovskii. "On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's". In: *Probability Theory and Related Fields* 103.2 (1995), pp. 143–163.

[24] I. A. Ibragimov and R. Z. Khas'minskii. "Estimation Problems for Coefficients of Stochastic Partial Differential Equations. Part III". In: *Theory of Probability and Its Applications* 45.2 (2001), pp. 210–232.

[25] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[26] Y. Kaino and M. Uchida. "Parametric estimation for a parabolic linear SPDE model based on discrete observations". In: *Journal of Statistical Planning and Inference* 211 (2021), pp. 190–220.

[27] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.

[28] R. Liptser and A. Shiryaev. *Statistics of Random Processes I*. General Theory. Springer, 2001.

[29] W Liu and M Röckner. *Stochastic Partial Differential Equations: An Introduction*. Springer, 2015.

[30] X. Liu, K. Yeo, Y. Hwang, J. Singh, and J. Kalagnanam. "A statistical modeling approach for air quality data based on physical dispersion processes and its application to ozone modeling". In: *Annals of Applied Statistics* 10.2 (2016), pp. 756–785.

[31] X. Liu, K. Yeo, and S. Lu. "Statistical Modeling for Spatio-Temporal Data From Stochastic Convection-Diffusion Processes". In: *Journal of the American Statistical Association* 117.539 (2021), pp. 1482–1499.

[32] C. Loader. *Local Regression and Likelihood*. Springer, 1999.

[33] S. Lototsky and B. Rozovsky. *Stochastic partial differential equations*. Springer, 2017.

[34] S. V. Lototsky. "Parameter Estimation for Stochastic Parabolic Equations: Asymptotic Properties of a Two-Dimensional Projection-Based Estimator". In: *Statistical Inference for Stochastic Processes* 6.1 (2003), pp. 65–87.

[35] A. Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, 1995.

[36] G. Pasemann and W. Stannat. “Drift estimation for stochastic reaction-diffusion systems”. In: *Electronic Journal of Statistics* 14.1 (2020), pp. 547–579.

[37] D. Ruppert and M. P. Wand. “Multivariate locally weighted least squares regression”. In: *Ann. Statist.* 22.3 (1994), pp. 1346–1370.

[38] S. E. Serrano. “Random Evolution Equations in Hydrology”. In: *Applied Mathematics and Computation* 38 (1990), pp. 201–226.

[39] F. Sigrist, H. R. Künsch, and W. A. Stahel. “A dynamic nonstationary spatio-temporal model for short term prediction of precipitation”. In: *Annals of Applied Statistics* 6.4 (2012), pp. 1452–1477.

[40] F. Sigrist, H. R. Künsch, and W. A. Stahel. “Stochastic partial differential equation based modelling of large space–time data sets”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.1 (2015), pp. 3–33.

[41] J. R. Stroud, M. L. Stein, B. M. Lesht, D. J. Schwab, and D. Beletsky. “An Ensemble Kalman Filter and Smoother for Satellite Data Assimilation”. In: *Journal of the American Statistical Association* 105.491 (2010), pp. 978–990.

[42] Y. Tonaki, Y. Kaino, and M. Uchida. “Parameter estimation for linear parabolic SPDEs in two space dimensions based on high frequency data”. In: *Scandinavian Journal of Statistics* 50.4 (2023), pp. 1568–1589.

[43] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

[44] H. C. Tuckwell. “Stochastic partial differential equations in neurobiology: Linear and nonlinear models for spiking neurons”. In: *Stochastic biomathematical models: with applications to neuronal modeling*. Ed. by M. Bachar, J. Batzel, and S. Ditlevsen. Springer, 2013, pp. 149–173.

[45] J. B. Walsh. “A stochastic model of neural response”. In: *Advances in Applied Probability* 13.2 (1981), pp. 231–281.

[46] L. Wasserman. *All of nonparametric Statistics*. Springer, 2006.

[47] W. Whitt. “Proofs of the martingale FCLT”. In: *Probability Surveys* 4 (2007), pp. 268–302.

Multivariate change estimation for a stochastic heat equation from local measurements

C

Anton Tiepner and Lukas Trottner

Abstract

We study a stochastic heat equation with piecewise constant diffusivity ϑ having a jump at a hypersurface Γ that splits the underlying space $[0,1]^d$, $d \geq 2$, into two disjoint sets $\Lambda_- \cup \Lambda_+$. Based on multiple spatially localized measurement observations on a regular δ-grid of $[0,1]^d$, we propose a joint M-estimator for the diffusivity values and the set Λ_+ that is inspired by statistical image reconstruction methods. We study convergence of the domain estimator $\widehat{\Lambda}_+$ in the vanishing resolution level regime $\delta \to 0$ and with respect to the expected symmetric difference pseudometric. Our main finding is a characterization of the convergence rate for $\widehat{\Lambda}_+$ in terms of the complexity of Γ measured by the number of intersecting hypercubes from the regular δ-grid. Implications of our general result are discussed under two specific structural assumptions on Λ_+. For a β-Hölder smooth boundary fragment Γ, the set Λ_+ is estimated with rate δ^β. If we assume Λ_+ to be convex, we obtain a δ-rate. While our approach only aims at optimal domain estimation rates, we also demonstrate consistency of our diffusivity estimators.

C.1 Introduction

Over the last decades interest in statistics for stochastic partial differential equations (SPDEs) has continuously increased for several reasons. Not only is it advantageous to model many natural space-time phenomena by SPDEs as they automatically account for model uncertainty by including random forcing terms that describe a more accurate picture of data dynamics, but also the general surge in data volume combined with enlarged computational power of modern computers makes it more appealing to investigate statistical problems for SPDEs.

In this paper we study a multivariate change estimation model for a stochastic heat equation on $\Lambda = (0,1)^d$, $d \geq 2$, given by

$$\mathrm{d}X(t) = \Delta_\vartheta X(t)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 \leq t \leq T, \tag{C.1}$$

with discontinuous diffusivity ϑ. The driving force is space-time white noise $\dot{W}(t)$ and the weighted Laplace operator $\Delta_\vartheta = \nabla \cdot \vartheta \nabla$ is characterized by a jump in the diffusivity

$$\vartheta(x) = \vartheta_- \mathbf{1}_{\Lambda_-}(x) + \vartheta_+ \mathbf{1}_{\Lambda_+}(x), \quad x \in (0,1)^d, \tag{C.2}$$

where the sets $\Lambda_\pm$ form a partition of $\overline{\Lambda} = [0,1]^d$. Our primary interest lies in the construction of a nonparametric estimator of the change domain Λ_+, which is equivalently characterized by the hypersurface

$$\Gamma := \partial\Lambda_- \cap \partial\Lambda_+ \subset [0,1]^d. \tag{C.3}$$

The SPDE (C.1) can, for instance, be used to describe the heat flow through two distinct materials with different heat conductivity, colliding in Γ. Structurally, the statistical problem of estimating

Γ is closely related to image reconstruction problems where one usually considers a regression model with (possibly random) design points X_k and observational noise ε_k given by

$$Y_k = f(X_k) + \varepsilon_k, \quad 1 \leq k \leq N,$$

where the Y_k correspond to the observed color of a pixel centred around the spatial point $X_k \in [0,1]^d$, and the otherwise continuous function $f\colon [0,1]^d \to [0,1]$ has a discontinuity along the hypersurface Γ, that is

$$f(x) = f_-(x)\mathbb{1}_{\Lambda_-}(x) + f_+(x)\mathbb{1}_{\Lambda_+}(x), \quad x \in [0,1]^d.$$

Such problems are, for instance, studied in [17–20, 24–26, 28, 31, 34]. Assuming specific structures such as boundary fragments [34] or star-shapes [30, 31], nonparametric regression methods are employed to consistently estimate both the image function f as well as the edge Γ, which in the boundary fragment case is characterized as the epigraph of a function $\tau\colon [0,1]^{d-1} \to [0,1]$. The rates of convergence depend on the smoothness of f and τ, the dimension d as well as the imposed distance function. Moreover, optimal convergence rates for higher-order Hölder smoothness $\beta > 1$ can in general not be achieved under equidistant, deterministic design, cf. [34, Chapter 3-5].

Estimation of scalar parameters in SPDEs is well-studied in the literature. When observing spectral measurements $(\langle X(t), e_k\rangle)_{0\leq t\leq T, k\leq N}$ for an eigenbasis $(e_k)_{k\in\mathbb{N}}$ of a parameterized differential operator A_ϑ, [14] derive criteria for identifiability of ϑ depending on ord A_ϑ and the dimension. This approach was subsequently adapted to joint parameter estimation [23], hyperbolic equations [22], lower-order nonlinearities [27], temporal discretization [8] or fractional noise [9]. If only discrete points $X(x_k, t_i)$ on a space-time grid are available, then estimation procedures relying on power variation approaches and minimum-contrast estimators are analyzed, amongst others, in [6, 13, 16, 33]. For a comprehensive overview of statistics for SPDEs we refer to the survey paper [7] and the website [3].

Our estimation approach is based on *local measurements*, as first introduced in [4], which are continuous in time and localized in space around δ-separated grid center points $x_\alpha \in (0,1)^d$, $\alpha \in \{1,\dots,\delta^{-1}\}^d$. More precisely, for a compactly supported and sufficiently smooth kernel function K and a resolution level $\delta \in 1/\mathbb{N}$, we observe for $\alpha \in \{1,\dots,\delta^{-1}\}^d$

$$(X_{\delta,\alpha}(t))_{0\leq t\leq T} = (\langle X(t), K_{\delta,\alpha}\rangle)_{0\leq t\leq T}, \quad (X^{\Delta}_{\delta,\alpha}(t))_{0\leq t\leq T} = (\langle X(t), \Delta K_{\delta,\alpha}\rangle)_{0\leq t\leq T},$$

for the localized functions $K_{\delta,\alpha}(\cdot) = \delta^{-d/2}K(\delta^{-1}(\cdot - x_\alpha))$. The rescaled function $\delta^{-d/2}K(\delta^{-1}\cdot)$ is also referred to as *point-spread function*, which is motivated from applications in optical systems, and the local measurement $X_{\delta,\alpha}$ represents a blurred image—typically owing to physical measurement limitations—that is obtained from convoluting the solution with the point spread function at the measurement location x_α. The asymptotic regime $\delta \to 0$ therefore allows for higher resolution images of the heat flow at the chosen measurement locations. Given such local measurements, we employ a CUSUM approach leading to an M-estimator for the quantities $(\vartheta_-, \vartheta_+, \Lambda_+)$.

Since their introduction in [4], local measurements have been used in numerous statistical applications. In [4] it was shown that a continuously differentiable diffusivity ϑ can be identified at location x_α from the observation of a single local measurement $(X_{\delta,\alpha}(t))_{0\leq t\leq T}$. Subsequently, their approach has been extended to semilinear equations [2], convection-diffusion equations

(Paper A and Paper B), multiplicative noise [15] and wave equations [35]. In [1] the practical revelance of the method has been demonstrated in a biological application to cell repolarization.

Closely related to this paper is the one-dimensional change point estimation problem for a stochastic heat equation studied in [29], which should be understood as the one-dimensional analogue to our problem setting. Indeed, in $d = 1$, the estimation of (C.3) boils down to the estimation of a single spatial change point at the jump location of the diffusivity. In [29] two different jump height regimes are analyzed, where the absolute jump height is given by $\eta := |\vartheta_+ - \vartheta_-|$. In the vanishing jump height regime $\eta \to 0$ as $\delta \to 0$, the authors demonstrate distributional convergence of the centralized change point estimator, where the asymptotic distribution is given by the law of the minimizer of a two-sided Brownian motion with drift, cf. [29, Theorem 4.2]. In contrast, if η is uniformly bounded away from 0, it is shown in [29, Theorem 3.12] that the change point can be identified with rate δ while the estimators for $(\vartheta_-, \vartheta_+)$ achieve the optimal rate $\delta^{3/2}$ in one dimension, cf. Paper A regarding optimality for parameter estimation, also in higher dimensions.

Coming back to our multivariate model (C.1), change estimation is no longer a parametric problem but becomes a nonparametric one, and we may either target Λ_+ directly or indirectly via estimation of the change interface (C.3). In this paper, we will first discuss the estimation problem for general sets Λ_+ and then specialize our estimation strategy and result to specific domain shapes.

Let us briefly describe our estimation approach in non-technical terms. For simplicity and to underline the correspondence to image reconstruction problems, let us consider for the moment only the case $d = 2$. We may then interpret the regular δ-grid as pixels, indexed by $\alpha \in \{1, \dots, \delta^{-1}\}^2$. By the nature of local observations that give only aggregated information on the heat flow on each of these pixels, the best we can hope for is a good approximation of a pixelated version of the true "foreground image" Λ_+^0 that we wish to distinguish from the true "background image" Λ_-^0. The pixelated version $\Lambda_+^{\updownarrow}$ is defined as the union of pixels that have a non-zero area intersection with Λ_+^0 as illustrated in Figure C.1. Based on a generalized

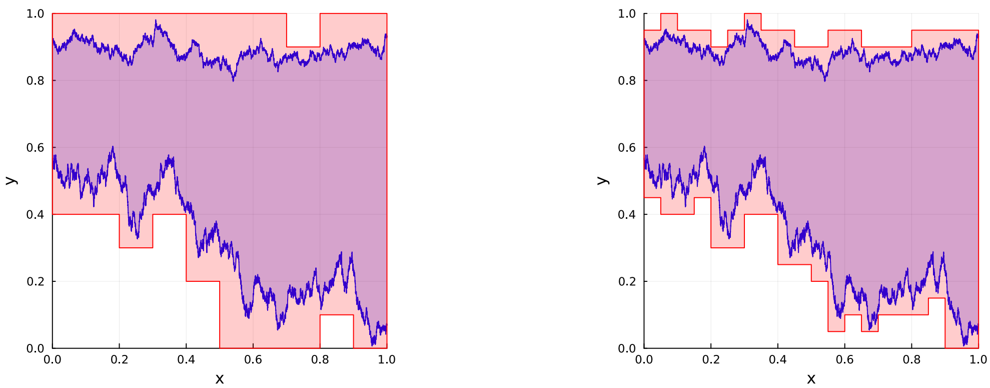


Figure C.1: Λ_+^0 (blue) is approximated by the pixelated version $\Lambda_+^{\updownarrow}$ (red); left: $\delta = 0.1$; right: $\delta = 0.05$.

Girsanov theorem for Itô processes, we can assign a modified local log-likelihood $\ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+)$ to each α-pixel for all pixelated candidate sets $\Lambda_+ \in \mathcal{A}_+$ that assigns the diffusivity value $\vartheta_\pm$ to the α-pixel if and only if $x_\alpha \in \Lambda_\pm$. An estimator $(\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+)$ is then obtained as the maximizer of the aggregated contrast function

$$(\vartheta_-, \vartheta_+, \Lambda_+) \mapsto \sum_{\alpha} \ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+),$$

which may be referred to as a CUSUM approach in analogy to change point estimation problems. Let us emphasize that we only require $\Lambda_+^\updownarrow \in \mathcal{A}_+$ of the pixelated candidate sets $\mathcal{A}_+$, which, given specific information on the shape of the true domain Λ_+^0, allows for much more parsimonious choices than the canonical choice of all possible black and white $\delta^{-1} \times \delta^{-1}$-images. On a more technical note, to establish the convergence bound, we reformulate our estimator as an M-estimator based on an appropriate empirical process $\chi \mapsto Z_\delta(\chi)$, so that quite naturally, concentration analysis of Z_δ becomes key. The basic idea of taking $\widehat{\Lambda}_+$ as union of best explanatory pixels by optimizing over a given family of candidate sets originates from classical statistical image reconstruction methods [19, 24, 25, 34].

The convergence rate of our estimator $\widehat{\Lambda}_+$ is entirely characterized by the complexity of the separating hypersurface Γ that induces a bias between the true domain Λ_+^0 and its pixelated version Λ_+^0. In particular, assuming that the set $\mathcal{B}$, describing the number of pixels that are sliced by Γ into two parts of non-zero volume, is of size

$$|\mathcal{B}| \lesssim \delta^{-d+\beta}, \quad \beta \in (0,1], \tag{C.4}$$

we show in Theorem C.7 that

$$\mathbb{E}[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \vartriangle \Lambda_+^0)] \lesssim \delta^\beta,$$

with the symmetric set difference $\vartriangle$. This result immediately entails estimation rates for Λ_+^0 in terms of the Minkowski dimension of its boundary. Furthermore, the estimation procedure results in the diffusivity parameter estimation rates $|\widehat{\vartheta}_\pm - \vartheta_\pm| = O_{\mathbb{P}}(\delta^{\beta/2})$, which yields the same estimation rate for the diffusivity or "image" estimator $\widehat{\vartheta} := \widehat{\vartheta}_+ \mathbb{1}_{\widehat{\Lambda}_+} + \widehat{\vartheta}_- \mathbb{1}_{\widehat{\Lambda}_-}$.

To make this general estimation strategy and result concrete, we apply it to two specific shape constraints on Λ_+^0. Assuming that Γ is a boundary fragment that is described by a change interface with graph representation $\tau^0\colon [0,1]^{d-1} \to [0,1]$, that is, $\Lambda_+^0 = \{(x,y) \in [0,1]^d : y > \tau^0(x)\}$, we choose closed epigraphs of piecewise constant grid functions $\tau\colon [0,1]^{d-1} \to [0,1]$ as candidate sets $\mathcal{A}_+$. The boundary of the estimator $\widehat{\Lambda}_+$ may then be interpreted as the epigraph of a random function $\widehat{\tau}\colon [0,1]^{d-1} \to [0,1]$ that gives a nonparametric estimator of the true change interface τ^0. Then, given β-Hölder smoothness assumptions on τ^0, where $\beta \in (0,1]$, we can verify (C.4) and obtain

$$\mathbb{E}[\|\widehat{\tau} - \tau^0\|_{L^1([0,1]^{d-1})}] \lesssim \delta^\beta \quad \text{or equivalently} \quad \mathbb{E}[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \vartriangle \Lambda_+^0)] \lesssim \delta^\beta.$$

In a second model, we assume that Λ_+^0 is a convex set with boundary Γ. Based on the idea that any ray that intersects the interior of a convex set does so in exactly two points, we construct a family of candidate sets $\mathcal{A}_+$ of size $|\mathcal{A}_+| \asymp \delta^{-(d+1)}$ and show that

$$\mathbb{E}[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \vartriangle \Lambda_+^0)] \lesssim \delta.$$

Optimality of the obtained rates in both models is discussed in the related image reconstruction problem.

Outline The paper is structured as follows. In Section C.2 we formalize the model and discuss fundamental properties of the solution to (C.1). The general estimation strategy and our main result is given in Section C.3. In Section C.4, those findings are applied to the two explicitly studied change domain structures outlined above. Lastly, we summarize our results and discuss potential extensions in future work in Section C.5.

Notation Throughout this paper, we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ with fixed time horizon $T < \infty$. The resolution level δ is such that $n := \delta^{-1} \in \mathbb{N}$ and $N := n^d$. For a set $A \subset \mathbb{R}^d$, the notation A° is exclusively reserved for its interior in $\mathbb{R}^d$ endowed with the standard Euclidean topology. For a general topological space $\mathcal{X}$ and a subset $A \subset \mathcal{X}$, we denote its interior by $\operatorname{int} A$, let $\overline{A}$ be its closure and ∂A be its boundary in $\mathcal{X}$. If not mentioned explicitly otherwise, we always understand the topological space in this paper to be $\mathcal{X} = [0,1]^d$ endowed with the standard subspace topology. For two numbers $a, b \in \mathbb{R}$, we write $a \lesssim b$ if $a \le Cb$ holds for a constant C that does not depend on δ. For an open set $U \subset \mathbb{R}^d$, $L^2(U)$ is the usual L^2-space with inner product $\langle \cdot, \cdot \rangle_{L^2(U)}$ and we set $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{L^2(\Lambda)}$. The Euclidean norm of a vector $a \in \mathbb{R}^p$ is denoted by $|a|$. If $f\colon \mathbb{R}^d \to \mathbb{R}^d$ is a vector valued function, we write $\|f\|_{L^2(\mathbb{R}^d)} := \||f|\|_{L^2(\mathbb{R}^d)}$. By $H^k(U)$ we denote the usual Sobolev spaces, and let $H_0^1(U)$ be the completion of $C_c^\infty(U)$, the space of smooth compactly supported functions, relative to the $H^1(U)$ norm. The gradient, divergence and Laplace operator are denoted by ∇, $\nabla\cdot$ and Δ, respectively.

C.2 Setup

We start by formally introducing the SPDE model, discussing existence of solutions and introducing the local measurement observation scheme that we will be working with in our statistical analysis.

C.2.1 *The SPDE model*

In the following we consider a stochastic partial differential equation on $\Lambda := (0,1)^d$, where $d \ge 2$, with Dirichlet boundary condition, which is specified by

$$\begin{cases} \mathrm{d}X(t) = \Delta_\vartheta X(t)\,\mathrm{d}t + \mathrm{d}W(t), & 0 \le t \le T, \\ X(0) \equiv 0, & \\ X(t)|_{\partial\Lambda} = 0, & 0 \le t \le T, \end{cases} \tag{C.5}$$

for driving space-time white noise $(\dot{W}(t))_{t\in[0,T]}$ on $L^2(\Lambda)$. The operator Δ_ϑ with domain $\mathcal{D}(\Delta_\vartheta)$ is given formally by

$$\Delta_\vartheta u = \nabla \cdot \vartheta \nabla u = \sum_{i=1}^d \partial_i(\vartheta \partial_i u), \quad u \in \mathcal{D}(\Delta_\vartheta), \tag{C.6}$$

where ϑ is piecewise constant in space, given by

$$\vartheta(x) = \vartheta_- \mathbf{1}_{\Lambda_-}(x) + \vartheta_+ \mathbf{1}_{\Lambda_+}(x), \quad x \in (0,1)^d,$$

for two measurable and disjoint sets $\Lambda_\pm$ s.t. $\Lambda_- \cup \Lambda_+ = [0,1]^d$ and $\vartheta_-, \vartheta_+ \in [\underline{\vartheta}, \overline{\vartheta}] \subset (0,\infty)$. Equivalently, we may rewrite the diffusivity in terms of the jump height $\eta := \vartheta_+ - \vartheta_-$ as

$$\vartheta(x) = \vartheta_- + \eta \mathbb{1}_{\Lambda_+}(x), \quad x \in (0,1)^d.$$

Under these assumptions, Δ_ϑ is a uniformly elliptic divergence-form operator. We assume that Λ_- and Λ_+ are separated by a hypersurface Γ parameterizing the set of points in the intersection of the boundaries

$$\Gamma := \partial\Lambda_+ = \partial\Lambda_- \subset [0,1]^d, \tag{C.7}$$

where $\partial\Lambda_\pm$ denotes the boundary of $\Lambda_\pm$ as a subset of the topological space $[0,1]^d$. We are mainly interested in estimating the domain Λ_+ which is intrinsically related to Γ. We first propose a general estimator based on local measurements of the solution on a uniform grid of hypercubes, whose convergence properties are determined by the complexity of the boundary Γ measured in terms of the number of hypercubes that are required to cover it.

The more structural information we are given on the set Λ_+, the better we can fine-tune the family of candidate sets underlying the estimator in order to increase the feasibility of implementation. Specifically, we will consider two different models for $\Lambda_\pm$.

Model A: Graph representation

Γ forms a boundary fragment that has a graph representation, denoted by a change interface $\tau\colon [0,1]^{d-1} \to [0,1]$, i.e.,

$$\Gamma = \left\{(x,\tau(x)) : x \in [0,1]^{d-1}\right\} \tag{C.8}$$

and the set Λ_+ takes the form

$$\Lambda_+ = \{(x,y) \in [0,1]^d : y > \tau(x)\}.$$

Accordingly, the estimation problem of identifying $\Lambda_\pm$ can equivalently be broken down to

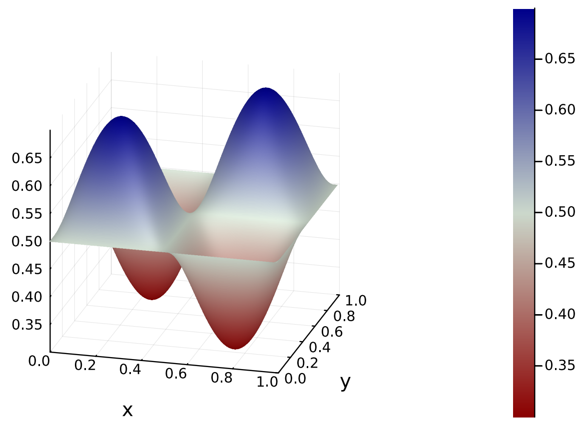


Figure C.2: Change interface τ in dimension $d = 3$.

the nonparametric estimation of the function τ. Specifically, for an estimator $\widehat{\tau}$ of τ we let $\widehat{\Lambda}_+ := \overline{\{(x,y) \in [0,1]^d : y > \widehat{\tau}(x)\}}$ be the closure of the epigraph of $\widehat{\tau}$ and $\widehat{\Lambda}_- := [0,1]^d \setminus \widehat{\Lambda}_+$ be its complement. This gives

$$\mathbb{E}\left[\lambda\left(\widehat{\Lambda}_\pm \triangle \Lambda_\pm\right)\right] = \mathbb{E}\left[\|\widehat{\tau} - \tau\|_{L^1([0,1]^{d-1})}\right], \tag{C.9}$$

such that evaluating the quality of the domain estimators $\widehat{\Lambda}_\pm$ measured in terms of the expected Lebesgue measure of the symmetric differences $\widehat{\Lambda}_\pm \triangle \Lambda_\pm$, is equivalent to studying the L^1-risk of the nonparametric estimator $\widehat{\tau}$ of the change interface. An exemplary illustration of τ in three spatial dimensions is present in Figure C.2.

Model B: Convex set

Λ_+ is convex. By convexity, for any $x \in \operatorname{int}\Lambda_+^0$, the vertical ray $y \mapsto x + ye_d$ intersects $\partial\Lambda_+$ in exactly two points and those intersection points can be modeled by a lower convex function f_1 and an upper concave function f_2. Estimation of Λ_+ is then heuristically speaking equivalent to the estimation of the upper and lower function, taking the closure of $\{(x,y) \in [0,1]^d : \widehat{f_1}(x) \le y \le \widehat{f_2}(x)\}$ as an estimator for Λ_+.

C.2.2 *Characterization of the solution*

We shall first discuss properties of the operator Δ_ϑ based on general theory of elliptic divergence form operators with measurable coefficients from [11]. Let the closed quadratic form $\mathcal{E}_\vartheta$ with domain $\mathcal{D}(\mathcal{E}_\vartheta)$ be given by

$$\begin{cases} \mathcal{D}(\mathcal{E}_\vartheta) = H_0^1(\Lambda), \\ \mathcal{E}_\vartheta(u,v) = \int_\Lambda \vartheta \nabla u \cdot \nabla v. \end{cases}$$

By [11, Theorem 1.2.1], $\mathcal{E}_\vartheta$ is the form of a positive self-adjoint operator $-\Delta_\vartheta$ on $L^2(\Lambda)$ in the sense that $\|(-\Delta_\vartheta)^{1/2}u\|^2 = \mathcal{E}_\vartheta(u,u)$ for $\mathcal{D}((-\Delta_\vartheta)^{1/2}) = \mathcal{D}(\mathcal{E}_\vartheta)$ and according to [11, Theorem 1.2.7], we have $u \in \mathcal{D}(-\Delta_\vartheta) \subset H_0^1(\Lambda)$ if there exists $g \in L^2(\Lambda)$ such that for any $v \in C_c^\infty(\Lambda)$,

$$\mathcal{E}_\vartheta(u,v) = \int_\Lambda gv,$$

in which case $-\Delta_\vartheta u = g$. Thus, (C.6) can be interpreted in a distributional sense and we have the relation

$$\mathcal{E}_\vartheta(u,v) = -\langle \Delta_\vartheta u, v\rangle, \quad (u,v) \in \mathcal{D}(\Delta_\vartheta) \times \mathcal{D}(\mathcal{E}_\vartheta).$$

Moreover, [11, Theorem 1.3.5] shows that $\mathcal{E}_\vartheta$ is a Dirichlet form, whence Δ_ϑ generates a strongly continuous, symmetric semigroup $(S_\vartheta(t))_{t\ge 0} := (\exp(\Delta_\vartheta t))_{t\ge 0}$. The spectrum of Δ_ϑ is discrete and the minimal eigenvalue, denoted by $\underline{\lambda}$, is strictly positive, cf. [12, Theorem 6.3.1]. Thus, $(-\Delta_\vartheta)^{-1}$ exists as a bounded linear operator with domain $L^2(\Lambda)$ and we may fix an orthonormal basis $\{e_k, k \in \mathbb{N}\}$ consisting of eigenvectors corresponding to the eigenvalues $\{\lambda_k, k \in \mathbb{N}\} = \sigma(-\Delta_\vartheta)$ that we denote in increasing order. Using the heat kernel bounds for the transition density of $S_\vartheta(t)$ given in [11, Corollary 3.2.8], it follows that for any $t > 0$, $S_\vartheta(t)$ is a Hilbert–Schmidt operator, but no weak or mild solution to (C.5) in the sense of [10, Theorem 5.4] exists in $L^2((0,1)^d)$ since $d \ge 2$ implies that $\int_0^T \|S_\vartheta(t)\|_{HS}^2 \,\mathrm{d}t = \infty$.

However, following the discussion in [4, Section 2.1] and Section A.6.2 in Paper A, taking into account that by [12, Theorem 6.3.1] we have $\lambda_k \asymp k^{2/d}$ for any $k \in \mathbb{N}$, the stochastic convolution

$$X(t) := \int_0^t S_\vartheta(t-s)\,\mathrm{d}W(s), \quad t \in [0,T],$$

is well-defined as a stochastic process on the embedding space $\mathcal{H}_1 \supset L^2((0,1)^d)$, where $\mathcal{H}_1$ can be chosen as a Sobolev space of negative order $-s < -d/2 + 1$ that is induced by the eigenbasis

$(\lambda_k, e_k)_{k\in\mathbb{N}}$. Extending the dual pairings $\langle X(t), z\rangle_{\mathcal{H}_1\times\mathcal{H}_1'}$ then allows us to obtain a Gaussian process $(\langle X(t), z\rangle)_{t\in[0,T],z\in L^2((0,1)^d)}$ given by

$$\langle X(t), z\rangle := \int_0^t \langle S_\vartheta(t-s)z, \mathrm{d}W(s)\rangle, \quad t\in[0,T], \quad z\in L^2((0,1)^d),$$

that solves the SPDE in the sense that for any $z\in\mathcal{D}(\Delta_\vartheta)$,

$$\langle X(t), z\rangle = \int_0^t \langle X(s), \Delta_\vartheta z\rangle\,\mathrm{d}s + \langle W(t), z\rangle, \quad t\in[0,T], \quad z\in L^2((0,1)^d). \tag{C.10}$$

C.2.3 *Local measurements*

We decompose $[0,1]^d$ into n^d closed d-dimensional hypercubes $(\mathrm{Sq}(\alpha))_{\alpha\in[n]^{d-1}}$, where $\mathrm{Sq}(\alpha)$ has edge length δ and is centered at $x_\alpha := \delta(\alpha - \frac{1}{2}\mathbb{1})$ for any $\alpha\in[n]^d$. By $\mathrm{Sq}(\alpha)^\circ$ we denote the interior of $\mathrm{Sq}(\alpha)$ in $\mathbb{R}^d$. Let also $\mathfrak{P} := 2^{\{\mathrm{Sq}(\alpha):\alpha\in[n]^d\}}$ be the power set of $\{\mathrm{Sq}(\alpha) : \alpha\in[n]^d\}$ and let $\mathcal{P} := \{\bigcup_{C\in\mathfrak{C}} C : \mathfrak{C}\in\mathfrak{P}\}$ be the family of sets that can be built from taking unions of hypercubes in $\{\mathrm{Sq}(\alpha) : \alpha\in[n]^d\}$. We refer to the hypercubes $\mathrm{Sq}(\alpha)$ as tiles and for any set $A\subset[0,1]^d$ we call a set $C\in\mathcal{P}$ such that $A\subset C$ a tiling of A.

Our estimation procedure is based on continuous-time observations of

$$X_{\delta,\alpha}(t) = \langle X(t), K_{\delta,\alpha}\rangle, \quad X^\Delta_{\delta,\alpha}(t) = \langle X(t), \Delta K_{\delta,\alpha}\rangle, \quad \alpha\in[n]^d, \quad 0\le t\le T,$$

with $K_{\delta,\alpha}(y) = \delta^{-d/2}K(\delta(y-x_\alpha))$, and $K\colon\mathbb{R}^d\to\mathbb{R}$ a kernel function such that $\operatorname{supp}K\subset[-1/2,1/2]^d$ and $K\in H^2(\mathbb{R}^d)$. The measurement points x_α, $\alpha\in[n]^d$, are separated by an Euclidean distance of order δ such that the supports of the $K_{\delta,\alpha}$ are non-overlapping. In other words, we have $N=\delta^{-d}$ measurement locations. Note that $K_{\delta,\alpha}\in\mathcal{D}(\Delta_\vartheta)$ whenever $\operatorname{supp}K_{\delta,\alpha}\cap\partial\Lambda_+=\varnothing$, since then ϑ is constant on the support of $K_{\delta,\alpha}$. Thus, for $g=-\vartheta(x_\alpha)\Delta K_{\delta,\alpha}$, integration by parts reveals $\int_\Lambda gv = \int_\Lambda \vartheta\nabla K_{\delta,\alpha}\cdot\nabla v$ for any $v\in C_c^\infty(\Lambda)$, i.e., $g=-\Delta_\vartheta K_{\delta,\alpha}$.

C.3 Estimation strategy and main result

From here on, we denote the truth, i.e., the true values of the diffusivity and the true set partition of Λ, by an additional superscript 0 for statistical purposes. To avoid some technicalities, we impose from now on the following assumption on Λ_+^0.

Assumption C.1. *Λ_+^0 is open in $[0,1]^d$.*

Since we are primarily interested in recovering the change domain Λ_+^0 and therefore treat the diffusivity parameters $\vartheta_\pm^0$ as nuisance parameters, we also make the following, slightly restricting assumption throughout the remainder of the paper:

Assumption C.2. *We have access to two compact sets $\Theta_-, \Theta_+\subset[\underline{\vartheta},\overline{\vartheta}]$ such that*

(i) $\vartheta_-^0\in\Theta_-$ and $\vartheta_+^0\in\Theta_+$, and

(ii) Θ_- and Θ_+ are separated by $\underline{\eta}>0$, i.e., for any $\vartheta_-\in\Theta_-$ and $\vartheta_+\in\Theta_+$, it holds $|\vartheta_+-\vartheta_-|\ge\underline{\eta}$.

In particular, this assumption implies that we have access to a lower bound $\underline{\eta} > 0$ on the absolute diffusivity jump height $|\eta^0| = |\vartheta_+^0 - \vartheta_-^0|$. For some set $C \subset [0,1]^d$ define

$$C^+ := \bigcup_{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap C \neq \varnothing} \mathrm{Sq}(\alpha),$$

which, if C is open in $[0,1]^d$, is the minimal tiling of C, so that in particular $C \subset C^+ \in \mathcal{P}$. Let us also set

$$\Lambda_+^\updownarrow := (\Lambda_+^0)^+,$$

which is the minimal tiling of Λ_+^0 by our assumption that Λ_+^0 is open in $[0,1]^d$. Let $\mathcal{A}_+ \subset \mathcal{P}$ be a family of candidate sets for a tiling of Λ_+^0 such that $\Lambda_+^\updownarrow \in \mathcal{A}_+$. Note that $\mathcal{A}_+ = \mathcal{P}$ is always a valid choice. However, as we shall see later, much more parsimonious choices are possible if we can assume some structure on the set Λ_+^0. We now introduce the modified local log-likelihood

$$\ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+) = \vartheta_{\delta,\alpha}(\Lambda_+) \int_0^T X_{\delta,\alpha}^\Delta(t)\, \mathrm{d}X_{\delta,\alpha}(t) - \frac{\vartheta_{\delta,\alpha}(\Lambda_+)^2}{2} \int_0^T X_{\delta,\alpha}^\Delta(t)^2\, \mathrm{d}t, \tag{C.11}$$

for the decision rule

$$\vartheta_{\delta,\alpha}(\Lambda_+) := \begin{cases} \vartheta_+, & \mathrm{Sq}(\alpha) \subset \Lambda_+, \\ \vartheta_-, & \text{else}, \end{cases} = \begin{cases} \vartheta_+, & x_\alpha \in \Lambda_+, \\ \vartheta_-, & \text{else}, \end{cases} \tag{C.12}$$

where $\vartheta_\pm \in \Theta_\pm$ and the candidate sets $\Lambda_+ \in \mathcal{A}_+$ are anchored on the grid $\mathcal{P}$ spanned by the hypercubes $\mathrm{Sq}(\alpha)$. The interpretation of the stochastic integral in (C.11) is provided by the following result that characterizes the tested processes $X_{\delta,\alpha}$ as semimartingales whose dynamics are determined by the location of the hypercube $\mathrm{Sq}(\alpha)$ relative to $\Lambda_\pm^0$.

Proposition C.3 (Modification of Proposition 2.1 in [29]). *For any $\alpha \in [n]^d$ and $t \in [0,T]$ we have*

$$X_{\delta,\alpha}(t) = \begin{cases} \vartheta_-^0 \int_0^t X_{\delta,\alpha}^\Delta(s)\, \mathrm{d}s + B_{\delta,\alpha}(t), & \mathrm{Sq}(\alpha) \subset \overline{\Lambda_-^0}, \\ \vartheta_+^0 \int_0^t X_{\delta,\alpha}^\Delta(s)\, \mathrm{d}s + B_{\delta,\alpha}(t), & \mathrm{Sq}(\alpha) \subset \overline{\Lambda_+^0}, \\ \int_0^t \int_0^s \langle \Delta_{\vartheta^0} S_{\vartheta^0}(s-u) K_{\delta,\alpha}, \mathrm{d}W(u) \rangle\, \mathrm{d}s + B_{\delta,\alpha}(t), & \textit{else}, \end{cases}$$

where $(B_{\delta,\alpha})_{\alpha \in [n]^d}$ is an n^d-dimensional vector of independent scalar Brownian motions.

Proof. The first two lines follow from (C.10) using $K_{\delta,\alpha} \in \mathcal{D}(\Delta_{\vartheta^0})$ if $\operatorname{supp} K_{\delta,\alpha} \cap \partial\Lambda_- \cap \partial\Lambda_+ = \varnothing$, in which cases $\Delta_{\vartheta^0} K_{\delta,\alpha} = \vartheta_-^0 \Delta K_{\delta,\alpha}$ and $\Delta_\vartheta K_{\delta,\alpha} = \vartheta_+^0 \Delta K_{\delta,\alpha}$, respectively. The expression on the change areas, where generally $K_{\delta,\alpha} \notin \mathcal{D}(\Delta_{\vartheta^0})$, is proven in complete analogy to [29, Lemma A.1] using spectral calculus and the stochastic Fubini theorem. ∎

We employ a CUSUM estimation approach based on the local log-likelihoods in (C.11) and therefore specify an estimator $(\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+)$ via

$$(\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+) \in \operatorname*{arg\,max}_{(\vartheta_-, \vartheta_+, \Lambda_+) \in \Theta_- \times \Theta_+ \times \mathcal{A}_+} \sum_{\alpha \in [n]^d} \ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+). \tag{C.13}$$

Here, the set of maximizers is well defined and we can make a measurable choice for a maximizer since $\Theta_\pm$ are compact and $\mathcal{A}_+$ is a finite set. Setting

$$\widehat{\Lambda}_- := [0,1]^d \setminus \widehat{\Lambda}_+,$$

as the corresponding estimator of Λ_-^0 we obtain the nonparametric diffusivity estimator

$$\widehat{\vartheta}(x) := \widehat{\vartheta}_- \mathbb{1}_{\widehat{\Lambda}_-}(x) + \widehat{\vartheta}_+ \mathbb{1}_{\widehat{\Lambda}_+}(x), \quad x \in (0,1)^{d-1}.$$

Introduce further

$$\vartheta_{\delta,\alpha}^0 := \begin{cases} \vartheta_+^0, & \mathrm{Sq}(\alpha) \subset \Lambda_+^{\updownarrow}, \\ \vartheta_-^0, & \text{else}, \end{cases} = \begin{cases} \vartheta_+^0, & x_\alpha \in \Lambda_+^{\updownarrow} \\ \vartheta_-^0, & \text{else}, \end{cases} \tag{C.14}$$

and define by

$$\mathcal{B} = \{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap \partial\Lambda_+^0 \neq \emptyset\},$$

the indices α of hypercubes whose interiors intersect the boundary of Λ_+^0. It is important to observe that the boundary tiles $\mathcal{B}$ may equivalently be expressed as follows.

LEMMA C.4. *It holds that*

$$\mathcal{B} = \{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap (\Lambda_+^{\updownarrow} \vartriangle \Lambda_+^0) \neq \emptyset\}.$$

Proof. Since Λ_+^0 is open in $[0,1]^d$, it holds that $\Lambda_+^0 \subset \Lambda_+^{\updownarrow}$ and therefore $\Lambda_+^{\updownarrow} \vartriangle \Lambda_+^0 = \Lambda_+^{\updownarrow} \setminus \Lambda_+^0$. Since $\Lambda_+^{\updownarrow}$ is closed and $\partial\Lambda_+^0 \cap \Lambda_+^0 = \emptyset$, it follows that $\partial\Lambda_+^0 \subset \Lambda_+^{\updownarrow} \vartriangle \Lambda_+^0$, which gives the inclusion

$$\mathcal{B} \subset \{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap (\Lambda_+^{\updownarrow} \vartriangle \Lambda_+^0) \neq \emptyset\}.$$

Conversely, if $\emptyset \neq \mathrm{Sq}(\alpha)^\circ \cap (\Lambda_+^{\updownarrow} \vartriangle \Lambda_+^0) = \mathrm{Sq}(\alpha)^\circ \cap (\Lambda_+^{\updownarrow} \setminus \Lambda_+^0)$, it follows that $\mathrm{Sq}(\alpha)^\circ \not\subset \Lambda_+^0$, but $\mathrm{Sq}(\alpha) \subset \Lambda_+^{\updownarrow}$ since the latter belongs to $\mathcal{P}$. Thus, by definition of $\Lambda_+^{\updownarrow}$, we have $\mathrm{Sq}(\alpha)^\circ \cap \Lambda_+^0 \neq \emptyset$, which because $\mathrm{Sq}(\alpha)^\circ$ is connected and $\mathrm{Sq}(\alpha)^\circ \not\subset \Lambda_+^0$ implies that $\mathrm{Sq}(\alpha)^\circ \cap \partial\Lambda_+^0 \neq \emptyset$. This now also yields

$$\mathcal{B} \supset \{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap (\Lambda_+^{\updownarrow} \vartriangle \Lambda_+^0) \neq \emptyset\}.$$

■

Plugging the given representation of Proposition C.3 into (C.11), we may now express $\ell_{\delta,\alpha}$ in the following way,

$$\begin{aligned}\ell_{\delta,\alpha}(\vartheta_-, \vartheta_+, \Lambda_+) = &\left(\vartheta_{\delta,\alpha}(\Lambda_+)\vartheta_{\delta,\alpha}^0 - \vartheta_{\delta,\alpha}(\Lambda_+)^2/2\right) I_{\delta,\alpha} + \vartheta_{\delta,\alpha}(\Lambda_+) M_{\delta,\alpha} \\ &+ \mathbb{1}_{\mathcal{B}}(\alpha)\vartheta_{\delta,\alpha}(\Lambda_+) R_{\delta,\alpha},\end{aligned}$$

where we denote

$$M_{\delta,\alpha} := \int_0^T X_{\delta,\alpha}^\Delta(t)\,\mathrm{d}B_{\delta,\alpha}(t), \quad I_{\delta,\alpha} := \int_0^T X_{\delta,\alpha}^\Delta(t)^2\,\mathrm{d}t,$$

and

$$R_{\delta,\alpha} := \int_0^T X_{\delta,\alpha}^\Delta(t)\left(\int_0^t \langle \Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha} - \vartheta_+^0 S_{\vartheta^0}(t-s)\Delta K_{\delta,\alpha}, \mathrm{d}W(s)\rangle\right)\mathrm{d}t.$$

The estimator (C.13) therefore can be represented as

$$(\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+) \in \underset{(\vartheta_-,\vartheta_+,\Lambda_+)\in\Theta_-\times\Theta_+\times\mathcal{A}_+}{\arg\max} \left\{ \sum_{\alpha\in[n]^d} \left((\vartheta_{\delta,\alpha}(\Lambda_+) - \vartheta_{\delta,\alpha}^0) M_{\delta,\alpha} - \frac{1}{2}(\vartheta_{\delta,\alpha}(\Lambda_+) - \vartheta_{\delta,\alpha}^0)^2 I_{\delta,\alpha} \right) \right.$$

$$+ \sum_{\alpha\in\mathcal{B}} \vartheta_{\delta,\alpha}(\Lambda_+)R_{\delta,\alpha} + \sum_{\alpha\in[n]^d}\Big(\vartheta^0_{\delta,\alpha}M_{\delta,\alpha} + \frac{(\vartheta^0_{\delta,\alpha})^2}{2}I_{\delta,\alpha}\Big)\Big\}$$

$$= \underset{(\vartheta_-,\vartheta_+,\Lambda_+)\in\Theta_-\times\Theta_+\times\mathcal{A}_+}{\arg\min}\Big\{Z_\delta(\vartheta_-,\vartheta_+,\Lambda_+) - \sum_{\alpha,\in\mathcal{B}}\vartheta_{\delta,\alpha}(\Lambda_+)R_{\delta,\alpha}\Big\}, \tag{C.15}$$

where the empirical process $Z_\delta(\cdot)$ is given by

$$\begin{aligned} Z_\delta(\vartheta_-,\vartheta_+,\Lambda_+) &:= \sum_{\alpha\in[n]^d}\frac{1}{2}(\vartheta_{\delta,\alpha}(\Lambda_+)-\vartheta^0_{\delta,\alpha})^2 I_{\delta,\alpha} - \sum_{\alpha\in[n]^d}(\vartheta_{\delta,\alpha}(\Lambda_+)-\vartheta^0_{\delta,\alpha})M_{\delta,\alpha} \\ &= \frac{1}{2}I_{T,\delta}(\vartheta_-,\vartheta_+,\Lambda_+) - M_{T,\delta}(\vartheta_-,\vartheta_+,\Lambda_+), \end{aligned} \tag{C.16}$$

where

$$\begin{aligned} I_{T,\delta}(\vartheta_-,\vartheta_+,\Lambda_+) &:= \sum_{\alpha,\in[n]^d}(\vartheta_{\delta,\alpha}(\Lambda_+)-\vartheta^0_{\delta,\alpha})^2 I_{\delta,\alpha}, \\ M_{T,\delta}(\vartheta_-,\vartheta_+,\Lambda_+) &:= \sum_{\alpha\in[n]^d}(\vartheta_{\delta,\alpha}(\Lambda_+)-\vartheta^0_{\delta,\alpha})M_{\delta,\alpha}. \end{aligned}$$

Note in particular that $Z_\delta(\vartheta^0_-,\vartheta^0_+,\Lambda^\updownarrow_+)=0$, so that $Z_\delta(\cdot)$ is centered around the truth $(\vartheta^0_-,\vartheta^0_+,\Lambda^\updownarrow_+)$. It will be crucial to have good control on the empirical process Z_δ, which is provided by the following lemma that specifies the order of the observed Fisher informations $I_{\delta,\alpha}$. The proof is a straightforward extension of the corresponding one-dimensional result [29, Lemma 3.3] using the analogous spectral arguments in higher dimensions and can therefore be omitted.

Lemma C.5 (Modification of Lemma 3.3 in [29]).

(i) For any $\alpha\in[n]^d$ with $\alpha\notin\mathcal{B}$, it holds

$$\mathbb{E}[I_{\delta,\alpha}] = \frac{T}{2\vartheta^0_{\delta,\alpha}}\|\nabla K\|^2_{L^2(\mathbb{R}^d)}\delta^{-2}+O(1).$$

(ii) For any $\alpha\in\mathcal{B}$, it holds

$$\mathbb{E}[I_{\delta,\alpha}] \in \left[\frac{2\underline{\lambda}T-1+\mathrm{e}^{-2\underline{\lambda}T}}{4\underline{\lambda}\overline{\vartheta}}\|\nabla K\|^2_{L^2(\mathbb{R}^d)}\delta^{-2}, \frac{T}{2\underline{\vartheta}}\|\nabla K\|^2_{L^2(\mathbb{R}^d)}\delta^{-2}\right].$$

(iii) For any vector $\beta\in(\mathbb{R}^n)^d$ with $\beta_\alpha=0$ if $\alpha\in\mathcal{B}$, we have

$$\mathrm{Var}\Big(\sum_{\alpha\in[n]^d}\beta_\alpha I_{\delta,\alpha}\Big) \le \frac{T}{2\underline{\vartheta}^3}\delta^{-2}\|\beta\|^2_{l^2}\|\nabla K\|^2_{L^2(\mathbb{R}^d)}.$$

Furthermore, convergence of the estimator $(\widehat{\vartheta}_-,\widehat{\vartheta}_+,\widehat{\Lambda}_+)$ requires insight on the order of the remainders $R_{\delta,\alpha}$ in the representation (C.15), given by the following lemma.

LEMMA C.6. *For any $\alpha \in \mathcal{B}$ we have*

$$\mathbb{E}\Big[|R_{\delta,\alpha}|\Big] \lesssim \delta^{-2}.$$

Proof. By Lemma C.5 we know that $\mathbb{E}[I_{\delta,\alpha}] \lesssim \delta^{-2}$ for any $\alpha \in [n]^d$. Since

$$R_{\delta,\alpha} = \int_0^T X^{\Delta}_{\delta,\alpha}(t) \int_0^t \langle \Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}, \mathrm{d}W(s)\rangle \,\mathrm{d}t - \vartheta^0_+ I_{\delta,\alpha},$$

it is enough to show

$$\mathbb{E}\Big[\Big|\int_0^T X^{\Delta}_{\delta,\alpha}(t) \int_0^t \langle \Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}, \mathrm{d}W(s)\rangle \,\mathrm{d}t\Big|\Big] \lesssim \delta^{-2}. \tag{C.17}$$

We have

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\int_0^T X^{\Delta}_{\delta,\alpha}(t) \int_0^t \langle \Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}, \mathrm{d}W(s)\rangle \,\mathrm{d}t\Big|\Big] \\
&\quad\le \mathbb{E}\Big[I_{\delta,\alpha}\Big]^{1/2} \mathbb{E}\Big[\int_0^T \Big(\int_0^t \langle \Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}, \mathrm{d}W(s)\rangle\Big)^2 \mathrm{d}t\Big]^{1/2} \\
&\quad\lesssim \delta^{-1}\Big(\mathbb{E}\Big[\int_0^T \Big(\int_0^t \langle \Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}, \mathrm{d}W(s)\rangle\Big)^2 \mathrm{d}t\Big]\Big)^{1/2} \\
&\quad= \delta^{-1}\Big(\int_0^T \int_0^t \big\|\Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}\big\|^2 \,\mathrm{d}s\,\mathrm{d}t\Big)^{1/2},
\end{aligned}$$

where we used the Cauchy–Schwarz inequality for the first two inequalities and Fubini's theorem together with the Itô-isometry for the last line. Since

$$\begin{aligned}
\int_0^T \int_0^t \big\|\Delta_{\vartheta^0} S_{\vartheta^0}(t-s) K_{\delta,\alpha}\big\|^2 \,\mathrm{d}s\,\mathrm{d}t &\le T \int_0^T \big\|\Delta_{\vartheta^0} S_{\vartheta^0}(t) K_{\delta,\alpha}\big\|^2 \,\mathrm{d}t \\
&= T \sum_{k\in\mathbb{N}} \int_0^T \lambda_k^2 \mathrm{e}^{-2\lambda_k t} \,\mathrm{d}t \langle e_k, K_{\delta,\alpha}\rangle^2 \\
&\le \frac{T}{2} \sum_{k\in\mathbb{N}} \lambda_k \langle e_k, K_{\delta,\alpha}\rangle^2 \\
&= \frac{T}{2} \big\|(-\Delta_{\vartheta^0})^{1/2} K_{\delta,\alpha}\big\|^2 \\
&= \frac{T}{2} \int_\Lambda \vartheta^0(x) |\nabla K_{\delta,\alpha}(x)|^2 \,\mathrm{d}x \\
&\le \frac{T\overline{\vartheta}}{2} \|\nabla K_{\delta,\alpha}\|^2_{L^2(\mathbb{R}^d)} \\
&= \delta^{-2} \frac{T\overline{\vartheta} \|\nabla K\|^2_{L^2(\mathbb{R}^d)}}{2},
\end{aligned}$$

the bound (C.17) follows, proving the assertion. ■

We are now ready to prove our main result.

THEOREM C.7. *Suppose that for some $\beta \in (0,1]$ and some constant $c > 0$ it holds that*

$$|\mathcal{B}| \leq c\delta^{-d+\beta}. \tag{C.18}$$

Then, for some absolute constant C depending only on $c, d, \underline{\vartheta}, \overline{\vartheta}, T$ and $\underline{\eta}$ it holds that

$$\mathbb{E}\Big[\lambda(\widehat{\Lambda}_+ \vartriangle \Lambda_+^0)\Big] \leq C\delta^{\beta}.$$

Proof. Let $\chi^0 := (\vartheta_-^0, \vartheta_+^0, (\Lambda_+^0)^+) = (\vartheta_-^0, \vartheta_+^0, \Lambda_+^{\updownarrow})$ and set for $\chi = (\vartheta_-, \vartheta_+, \Lambda_+) \in \Theta_- \times \Theta_+ \times \mathcal{A}_+$,

$$L_\delta(\chi) = \delta^{d+2} Z_\delta(\vartheta_-, \vartheta_+, \Lambda_+), \quad \widetilde{L}_\delta(\chi) = \mathbb{E}[L_\delta(\chi)].$$

We first observe that (C.15) implies

$$L_\delta(\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+) \leq \min_{\chi \in \Theta_- \times \Theta_+ \times \mathcal{A}_+} L_\delta(\chi) + 2\delta^{2+d}\overline{\vartheta} \sum_{\alpha \in \mathcal{B}} |R_{\delta,\alpha}|. \tag{C.19}$$

Furthermore, since $L_\delta(\chi^0) = 0$ and $\mathbb{E}[M_{\delta,\alpha}] = 0$, we obtain with Lemma C.5 that for any $\chi = (\vartheta_-, \vartheta_+, \Lambda_+) \in \Theta_- \times \Theta_+ \times \mathcal{A}_+$,

$$\begin{aligned}
\widetilde{L}_\delta(\chi) - \widetilde{L}_\delta(\chi^0) &= \mathbb{E}[L_\delta(\chi)] \asymp \delta^d \sum_{\alpha \in [n]^d} (\vartheta_{\delta,\alpha}(\Lambda_+) - \vartheta_{\delta,\alpha}^0)^2 \\
&= \delta^d \sum_{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \subset (\Lambda_+ \cap \Lambda_+^{\updownarrow})} (\vartheta_+ - \vartheta_+^0)^2 + \delta^d \sum_{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \subset (\Lambda_+ \cup \Lambda_+^{\updownarrow})^c} (\vartheta_- - \vartheta_-^0)^2 \\
&\quad + \delta^d \sum_{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \subset (\Lambda_+ \vartriangle \Lambda_+^{\updownarrow})} (\vartheta_{\delta,\alpha}(\Lambda_+) - \vartheta_{\delta,\alpha}^0)^2 \\
&\geq \underline{\eta}^2 \delta^d \Big|\{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \subset (\Lambda_+ \vartriangle \Lambda_+^{\updownarrow})\}\Big| \\
&= \underline{\eta}^2 \lambda(\Lambda_+ \vartriangle \Lambda_+^{\updownarrow}),
\end{aligned} \tag{C.20}$$

where for the penultimate line we observe that for $\mathrm{Sq}(\alpha)^\circ \subset (\Lambda_+ \vartriangle \Lambda_+^{\updownarrow})$ we have $\vartheta_{\delta,\alpha}(\Lambda_+) = \vartheta_\pm$ iff $\vartheta_{\delta,\alpha}^0 = \vartheta_\mp^0$, and conclude with $\vartheta_\pm, \vartheta_\pm^0 \in \Theta_\pm$ and the fact that Θ_+ and Θ_- are $\underline{\eta}$-separated. Now, using the characterization from Lemma (C.4) and the assumption (C.18), it follows that

$$\lambda(\Lambda_+^0 \vartriangle \Lambda_+^{\updownarrow}) \leq \delta^d |\mathcal{B}| \leq c\delta^{\beta}. \tag{C.21}$$

Combined with (C.20), triangle inequality for the symmetric difference pseudometric therefore yields

$$\lambda(\Lambda_+ \vartriangle \Lambda_+^0) \lesssim \widetilde{L}_\delta(\chi) - \widetilde{L}_\delta(\chi^0) + \delta^{\beta}, \quad \chi = (\vartheta_-, \vartheta_+, \Lambda_+) \in \Theta_- \times \Theta_+ \times \mathcal{A}_+,$$

whence the assertion follows once we have verified that

$$\mathbb{E}[\widetilde{L}_\delta(\widehat{\chi}) - \widetilde{L}_\delta(\chi^0)] \lesssim \delta^{\beta}, \quad \widehat{\chi} := (\widehat{\vartheta}_-, \widehat{\vartheta}_+, \widehat{\Lambda}_+), \tag{C.22}$$

where by definition $\widetilde{L}_\delta(\widehat{\chi}) = \mathbb{E}[L_\delta(\chi)]|_{\chi = \widehat{\chi}}$. Taking into account (C.18), (C.19) and Lemma C.6, we arrive at

$$\begin{aligned}
\mathbb{E}[\widetilde{L}_\delta(\widehat{\chi}) - \widetilde{L}_\delta(\chi^0)] &\leq \mathbb{E}\Big[\widetilde{L}_\delta(\widehat{\chi}) - \widetilde{L}_\delta(\chi^0) + L_\delta(\chi^0) - L_\delta(\widehat{\chi})\Big] \\
&\quad + 2\delta^{d+2} c\delta^{\beta-d} \max_{\alpha \in \mathcal{B}} \mathbb{E}[|R_{\delta,\alpha}|] \\
&\lesssim \mathbb{E}\Big[\sup_{\chi \in \Theta_- \times \Theta_+ \times \mathcal{A}_+} |L_\delta(\chi) - \widetilde{L}_\delta(\chi)|\Big] + \delta^{\beta}.
\end{aligned} \tag{C.23}$$

To prove (C.22) it therefore remains to show

$$\mathbb{E}\Big[\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|L_\delta(\chi)-\widetilde{L}_\delta(\chi)|\Big]\lesssim\delta^\beta,$$

which, recalling the decomposition (C.16) and using $\mathbb{E}[M_{T,\delta}(\chi)]=0$, boils down to show

$$\mathbb{E}\Big[\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|I_{T,\delta}(\chi)-\mathbb{E}[I_{T,\delta}(\chi)]|\Big]\lesssim\delta^{\beta-d-2},\tag{C.24}$$

$$\mathbb{E}\Big[\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|M_{T,\delta}(\chi)|\Big]\lesssim\delta^{\beta-d-2}.\tag{C.25}$$

Clearly, by triangle inequality, we get the rough bounds

$$\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|I_{T,\delta}(\chi)-\mathbb{E}[I_{T,\delta}(\chi)]|\le(\overline{\vartheta}-\underline{\vartheta})^2\sum_{\alpha\in[n]^d}|I_{\delta,\alpha}-\mathbb{E}[I_{\delta,\alpha}]|,$$
$$\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|M_{T,\delta}(\chi)|\le|\overline{\vartheta}-\underline{\vartheta}|\sum_{\alpha\in[n]^d}|M_{\delta,\alpha}|.$$

Thus, applying once more the assumption (C.18) and the bounds from Lemma C.5,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|I_{T,\delta}(\chi)-\mathbb{E}[I_{T,\delta}(\chi)]|\Big]\\
&\lesssim\delta^{\beta-d}\max_{\alpha\in\mathcal{B}}\mathbb{E}[|I_{\delta,\alpha}-\mathbb{E}[I_{\delta,\alpha}]|]+\delta^{-d}\max_{\alpha\in[n]^d\setminus\mathcal{B}}\mathbb{E}[|I_{\delta,\alpha}-\mathbb{E}[I_{\delta,\alpha}]|]\\
&\le 2\delta^{\beta-d}\max_{\alpha\in\mathcal{B}}\mathbb{E}[|I_{\delta,\alpha}|]+\delta^{-d}\max_{\alpha\in[n]^d\setminus\mathcal{B}}\operatorname{Var}(I_{\delta,\alpha})^{1/2}\\
&\lesssim\delta^{\beta-d-2}+\delta^{-d-1},
\end{aligned}$$

which establishes (C.24). For the martingale part, using $\mathbb{E}[|M_{\delta,\alpha}|^2]=\mathbb{E}[I_{\delta,\alpha}]$, the Cauchy–Schwarz inequality and Lemma C.5, we have

$$\begin{aligned}
\mathbb{E}\Big[\sup_{\chi\in\Theta_-\times\Theta_+\times\mathcal{A}_+}|M_{T,\delta}(\chi)|\Big]&\lesssim\delta^{-d}\max_{\alpha\in[n]^d}\mathbb{E}[|M_{\delta,\alpha}|]\\
&\le\delta^{-d}\max_{\alpha\in[n]^d}(\mathbb{E}[I_{\delta,\alpha}])^{1/2}\\
&\lesssim\delta^{-d-1},
\end{aligned}$$

showing (C.25). This finishes the proof. ∎

This result entails convergence rates for the estimation of domains Λ_+^0 with boundary of Minkowski dimension (also called *box counting dimension*) $d-\beta$. Recall that if for a set $A\subset[0,1]^d$ we let $N(A,\delta)$ be the minimal number of hypercubes $\mathrm{Sq}(\alpha)$ needed to cover A and set

$$\underline{\dim}_{\mathcal{M}}(A)=\liminf_{\delta\to0}\frac{\log N(A,\delta)}{\log 1/\delta},\quad\overline{\dim}_{\mathcal{M}}(A)=\limsup_{\delta\to0}\frac{\log N(A,\delta)}{\log 1/\delta},$$

then, if $\underline{\dim}_{\mathcal{M}}(A)=\overline{\dim}_{\mathcal{M}}(A)$, we call

$$\dim_{\mathcal{M}}(A)=\lim_{\delta\to0}\frac{\log N(A,\delta)}{\log 1/\delta},$$

the Minkowski dimension of A (see [5, p.2] for the fact that this is an equivalent characterization of the Minkowski dimension in the Euclidean space $[0,1]^d$). Clearly,

$$|\mathcal{B}| \leq N(\partial\Lambda_+^0, \delta),$$

so that Theorem C.7 yields the following corollary.

Corollary C.8. *Suppose that for some $\beta \in (0,1]$ it holds that*

$$\dim_{\mathcal{M}}(\partial\Lambda_+^0) \leq d - \beta.$$

Then, for any $\varepsilon > 0$,

$$\mathbb{E}\big[\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0)\big] = o(\delta^{\beta-\varepsilon}).$$

Remark C.9. The Minkowski dimension $\dim_{\mathcal{M}}$ always dominates the Hausdorff dimension $\dim_{\mathcal{H}}$. For many reasonable sets they coincide and in these cases the condition on $\dim_{\mathcal{M}}(\partial\Lambda_+^0)$ may be replaced by the same one on $\dim_{\mathcal{H}}(\partial\Lambda_+^0)$. In most cases, $\dim_{\mathcal{M}}(\partial\Lambda_+^0) = d-\beta$ is verified explicitly by establishing that $N(\partial\Lambda_+^0, \delta) \asymp c\delta^{-d+\beta}$, which then improves the result to $\mathbb{E}\big[\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0)\big] = \mathcal{O}(\delta^\beta)$.

As outlined before, our estimator only aims at rate optimality for inference on Λ_+^0. Given a necessary identifiability assumption, the nuisance parameters $\vartheta_\pm^0$ are still consistently estimated under the assumptions of Theorem C.7.

Corollary C.10. *Suppose that $|\mathcal{B}| \leq c\delta^{-d+\beta}$ for some $\beta \in (0,1]$ and some constant $c > 0$. If $\lambda(\Lambda_\pm^0) > 0$, then $\widehat{\vartheta}_\pm$ is a consistent estimator satisfying $|\widehat{\vartheta}_\pm - \vartheta_\pm^0| = \mathcal{O}_{\mathbb{P}}(\delta^{\beta/2})$. In particular, if both Λ_+^0 and Λ_-^0 have positive Lebesgue measure, it holds that $\|\widehat{\vartheta} - \vartheta^0\|_{L^1((0,1)^d)} \in \mathcal{O}_{\mathbb{P}}(\delta^{\beta/2})$.*

Proof. We only prove the assertion on $\widehat{\vartheta}_+$ given $\lambda(\Lambda_+^0) \neq 0$; the case for $\widehat{\vartheta}_-$ under the assumption $\lambda(\Lambda_-^0) \neq 0$ is analogous and the final statement on the convergence rate of $\widehat{\vartheta}$ then follows from combining the first statement and Theorem C.7 based on the inequality

$$|\widehat{\vartheta}(x) - \vartheta^0(x)| \leq |\widehat{\vartheta}_- - \vartheta_-^0| + |\widehat{\vartheta}_+ - \vartheta_+^0| + 2\overline{\vartheta}|\mathbb{1}_{\widehat{\Lambda}_+}(x) - \mathbb{1}_{\Lambda_+^0}(x)|, \quad x \in (0,1)^d.$$

Let $\kappa = \lambda(\Lambda_+^0) > 0$. From (C.21) it follows that on the event $\{\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0) \leq \delta^{\beta/2}\}$ we have for δ small enough

$$\frac{\kappa}{2} \leq \lambda(\widehat{\Lambda}_+ \cap \Lambda_+^{\updownarrow}) = \delta^d\big|\{\alpha \in [n]^d : \mathrm{Sq}^\circ(\alpha) \subset (\widehat{\Lambda}_+ \cap \Lambda_+^{\updownarrow})\}\big|,$$

where the equality follows from the fact that for $A, B \in \mathcal{P}$ we also have $A \cap B \in \mathcal{P}$. Thus, for δ small enough, it follows from the second line of the calculation in (C.20) that on the event $\{\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0) \leq \delta^{\beta/2}\}$ we have

$$\widetilde{L}_\delta(\widehat{\chi}) - \widetilde{L}_\delta(\chi^0) \geq \frac{\kappa}{2}(\widehat{\vartheta}_+ - \vartheta_+^0)^2.$$

Consequently, there exists $C' > 0$ such that for any $\delta^{-1} \geq C'$ and $z > 0$,

$$\begin{aligned}
&\mathbb{P}\big(\delta^{-\beta}(\widehat{\vartheta}_+ - \vartheta_+^0)^2 \geq z\big)\\
&\leq \mathbb{P}\Big(\delta^{-\beta}(\widehat{\vartheta}_+ - \vartheta_+^0)^2 \geq z, \lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0) \leq \delta^{\beta/2}\Big) + \mathbb{P}\Big(\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0) > \delta^{\beta/2}\Big)\\
&\leq \mathbb{P}\big(\widetilde{L}_\delta(\widehat{\chi}) - \widetilde{L}_\delta(\chi^0) \geq \tfrac{\kappa}{2}\delta^\beta z\big) + \mathbb{P}\Big(\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0) > \delta^{\beta/2}\Big)
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\frac{\mathbb{E}[\widetilde{L}_\delta(\widehat{\chi}) - \widetilde{L}_\delta(\chi^0)]}{\kappa\delta^\beta z} + \frac{\mathbb{E}\big[\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+^0)\big]}{\delta^{\beta/2}} \\
&\leq \widetilde{C}\big(\tfrac{1}{z} + \delta^{\beta/2}\big),
\end{aligned}
$$

for some finite constant $\widetilde{C} > 0$ independent of δ where the last line follows from (C.22) and Theorem C.7. Thus, if for given $\varepsilon > 0$ we choose $z = 2\widetilde{C}/\varepsilon$, it follows that for any $\delta^{-1} \geq (2\widetilde{C}/\varepsilon)^{-2/\beta} \vee C'$ we have

$$
\mathbb{P}\big(\delta^{-\beta}(\widehat{\vartheta}_+ - \vartheta_+^0)^2 \geq z\big) \leq \varepsilon,
$$

which establishes $(\widehat{\vartheta}_+ - \vartheta_+^0)^2 = \mathcal{O}_{\mathbb{P}}(\delta^\beta)$ as $\delta^{-1} \to \infty$. ■

Given the results from Paper A, which can be applied to a stochastic heat equation with constant diffusivity, this rate is not expected to be optimal. As argued in [29], a more careful approach in the design of the estimator that accounts for the irregularities of the diffusivity in a more elaborate way would be needed to also achieve rate-optimality for the diffusivity parameters. Since the paper focuses on the estimation of the domain Λ_+^0, these issues will not be discussed further.

C.4 Results for specific models

In this section we give explicit constructions of the candidate sets $\mathcal{A}_+$ and estimator convergence rates for two specific shape restrictions.

C.4.1 *Estimation of change interfaces with graph representation*

Consider model A from Section C.2, for which Λ_+^0 is fully determined by the continuous change interface $\tau^0\colon [0,1]^{d-1} \to [0,1]$. For any function $\tau\colon [0,1]^{d-1} \to [0,1]$ let us define the open epigraph

$$
\operatorname{epi}\tau := \{(x,y) \in [0,1]^{d-1} \times [0,1] : \tau(x) > y\}.
$$

For $\gamma \in [n]^{d-1}$ let $\widetilde{\mathrm{Sq}}_{d-1}(\gamma)$ be a $(d-1)$-dimensional hypercube with edge length δ that is centered at $z_\gamma := \delta(\gamma - \frac{1}{2}\mathbb{1})$. The hypercubes $\widetilde{\mathrm{Sq}}_{d-1}(\gamma)$ are chosen such that $(\widetilde{\mathrm{Sq}}_{d-1}(\gamma))_{\gamma\in[n]^{d-1}}$ forms a partition of $[0,1]^{d-1}$, i.e., $[0,1]^{d-1} = \dot{\bigcup}_{\gamma\in[n]^{d-1}} \widetilde{\mathrm{Sq}}_{d-1}(\gamma)$. Let us also denote $\mathrm{Sq}_{d-1}(\gamma) := \overline{\widetilde{\mathrm{Sq}}_{d-1}(\gamma)}$. With

$$
\mathcal{G} := \{i\delta : i = 0, 1, \dots, n\}^{n^{d-1}},
$$

we now define the grid functions

$$
\tau_\zeta(x) = \sum_{\gamma\in[n]^{d-1}} \zeta_\gamma \mathbf{1}_{\widetilde{\mathrm{Sq}}_{d-1}(\gamma)}(x), \quad x \in [0,1]^{d-1}, \quad \zeta = (\zeta_\gamma)_{\gamma\in[n]^{d-1}} \in \mathcal{G},
$$

and set

$$
\Lambda_+(\zeta) := (\operatorname{epi}\tau_\zeta)^+ \in \mathcal{P}, \quad \zeta \in \mathcal{G}.
$$

In other words, $\Lambda_+(\zeta)$ can be written as

$$
\Lambda_+(\zeta) = \bigcup_{\gamma\in[n]^{d-1}:\zeta_\gamma<1} \mathrm{Sq}_{d-1}(\gamma) \times [\zeta_\gamma, 1].
$$

We then choose our candidate tiling sets as

$$\mathcal{A}_+ = \{\Lambda_+(\zeta) : \zeta \in \mathcal{G}\}.$$

Note that the size $|\mathcal{A}_+| = n^{d-1}(n+1)$ of this family of candidate sets if significantly smaller than that of the uninformed choice $\mathcal{A}_+ = \mathcal{P}$, which is 2^{n^d}. To see that $\mathcal{A}_+$ is a valid choice, note that for

$$\zeta_\gamma^\uparrow := \delta\lceil \delta^{-1} \sup\{\tau^0(x) : x \in \widetilde{\mathrm{Sq}}_{d-1}(\gamma)\}\rceil, \quad \zeta_\gamma^\downarrow := \delta\lfloor \delta^{-1} \inf\{\tau^0(x) : x \in \widetilde{\mathrm{Sq}}_{d-1}(\gamma)\}\rfloor,$$

it holds that $\tau_{\zeta^\downarrow}$ is the maximal grid function dominated by τ^0 and therefore

$$\Lambda_+^\updownarrow = (\operatorname{epi}\tau^0)^+ = (\operatorname{epi}\tau_{\zeta^\downarrow})^+ = \Lambda_+(\zeta^\downarrow),$$

whence in particular $\Lambda_+^\updownarrow \in \mathcal{A}_+$ as required. Furthermore, the function

$$\varphi \colon \mathcal{A}_+ \to \mathcal{G}, \quad \Lambda_+(\zeta) \to \zeta,$$

is a bijection. Thus, for the estimator $\widehat{\Lambda}_+$ from the previous section, if we set

$$\widehat{\zeta} := \varphi(\widehat{\Lambda}_+),$$

it follows that

$$\widehat{\Lambda}_+ = \Lambda_+(\widehat{\zeta}) = (\operatorname{epi}\tau_{\widehat{\zeta}})^+.$$

Consequently, if we define the change interface estimator

$$\widehat{\tau} = \tau_{\widehat{\zeta}},$$

then we have the identity

$$\|\widehat{\tau} - \tau^0\|_{L^1([0,1]^{d-1})} = \boldsymbol{\lambda}(\widehat{\Lambda}_+ \vartriangle \Lambda_+^0).$$

An example of a change interface and its grid function approximation can be seen in Figure C.3. Under Hölder smoothness assumptions on the change interface τ^0, Theorem C.7 allows us to obtain a convergence rate for the change domain estimator $\widehat{\Lambda}_+$ in the symmetric difference pseudometric, or equivalently, for the change interface estimator $\widehat{\tau}$ in the L^1-metric. For $\beta \in (0,1]$, $L > 0$, let $\mathcal{H}(\beta, \mathsf{L})$ be the (β, L)-Hölder class on $[0,1]^{d-1}$ w.r.t. the maximum metric, i.e.,

$$\mathcal{H}(\beta, \mathsf{L}) := \{f \colon [0,1]^{d-1} \to \mathbb{R} : |f(x) - f(y)| \le \mathsf{L}\|x - y\|_\infty^\beta \text{ for all } x, y \in [0,1]^{d-1}\}.$$

Proposition C.11. *If $\tau^0 \in \mathcal{H}(\beta, \mathsf{L})$ for $\beta \in (0,1]$ and $\mathsf{L} > 0$, then there exists a constant C depending only on $\beta, \mathsf{L}, d, \underline{\vartheta}, \overline{\vartheta}, T$ and $\underline{\eta}$ such that*

$$\mathbb{E}\big[\|\widehat{\tau} - \tau^0\|_{L^1([0,1]^{d-1})}\big] = \mathbb{E}\big[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \vartriangle \Lambda_+^0)\big] \le C\delta^\beta.$$

Moreover, if τ^0 is not identically 1 (resp. not identically 0), then $|\widehat{\vartheta}_+ - \vartheta_+^0| = \mathcal{O}_{\mathbb{P}}(\delta^{\beta/2})$ (resp. $|\widehat{\vartheta}_- - \vartheta_-^0| = \mathcal{O}_{\mathbb{P}}(\delta^{\beta/2})$). In particular, if $\tau^0(x) \in (0,1)$ for some $x \in [0,1]^{d-1}$, it holds that $\|\widehat{\vartheta} - \vartheta^0\|_{L^1((0,1)^d)} \in \mathcal{O}_{\mathbb{P}}(\delta^{\beta/2})$.

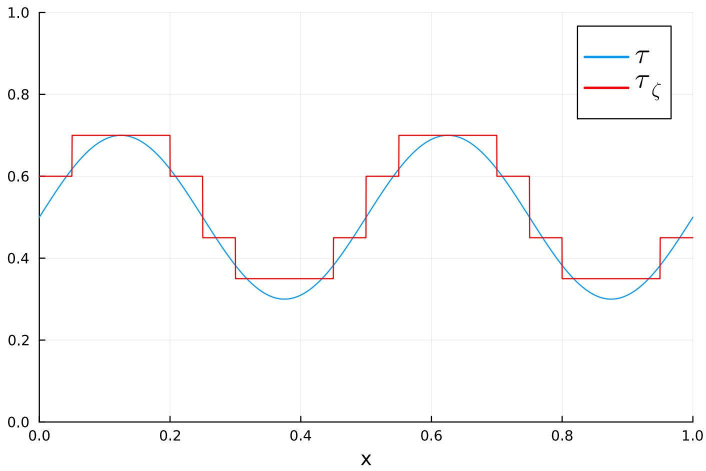


Figure C.3: Example of a change interface τ and a piecewise constant approximation τ_ζ in dimension $d = 2$ with $\delta = 0.05$.

Proof. By Theorem C.7 and Corollary C.10, it suffices to show that $|\mathcal{B}| \leq \mathsf{L}\delta^{d-\beta}$. To this end, we first observe that in the change interface model it holds that

$$\begin{aligned}
\mathcal{B} &= \left\{\alpha \in [n]^d : \mathrm{Sq}(\alpha) \cap \partial\Lambda_+^0 \neq \varnothing\right\} \\
&= \left\{(\gamma, j) \in [n]^{d-1} \times [n] : (\mathrm{Sq}_{d-1}(\gamma)^\circ \times ((j-1)\delta, j\delta)) \cap \tau^0(\mathrm{Sq}_{d-1}(\gamma)) \neq \varnothing\right\} \\
&\subset \left\{(\gamma, j) \in [n]^{d-1} \times [n] : j \in \delta^{-1}(\zeta_\gamma^\downarrow, \zeta_\gamma^\uparrow]\right\}.
\end{aligned}$$

Because $\tau^0 \in \mathcal{H}(\beta, \mathsf{L})$ implies that $|\zeta_\gamma^\uparrow - \zeta_\gamma^\downarrow| \leq \mathsf{L}\delta^\beta$ for any $\gamma \in [n]^{d-1}$, we obtain

$$|\{j \in \delta^{-1}(\zeta_\gamma^\downarrow, \zeta_\gamma^\uparrow] \cap [n]\}| \leq \mathsf{L}\delta^{\beta-1}.$$

Thus, from above,

$$|\mathcal{B}| \leq \delta^{-(d-1)}\mathsf{L}\delta^{\beta-1} = \mathsf{L}\delta^{-d+\beta},$$

as desired. ■

Remark C.12. The domain estimation rate δ^β translates to $N^{-\beta/d}$ in terms of the number of observations $N = \delta^{-d}$. As pointed out in [19, Chapter 3-5], for appropriately designed random measurement locations, the minimax rate for estimating a boundary fragment in image reconstruction is given by $N^{-\beta/(\beta+d-1)}$ for arbitrary $\beta > 0$. Unless we have Lipschitz regularity $\beta = 1$, this rate is, however, not achievable for an equidistant deterministic design, which is usually referred to as *regular design*. In fact, it can be shown that with regular design, the rate $N^{-\beta/d}$ is the optimal rate for $\beta \in (0, 1]$ in the edge estimation problem by adapting [19, Theorem 3.3.1]. Indeed, consider the regular design image reconstruction problem

$$Y_\alpha = \vartheta(x_\alpha) + \varepsilon_\alpha, \quad \alpha \in [n]^d,$$

where

$$\vartheta(x) = \vartheta_- \mathbb{1}_{\Lambda_-}(x) + \vartheta_+ \mathbb{1}_{\Lambda_+}(x), \quad x \in [0, 1]^d,$$

for known values $\vartheta_- \neq \vartheta_+$ and $\Lambda_- \dot{\bigcup} \Lambda_+ = [0,1]^d$ and denote by $\mathbb{P}_{\Lambda_+}$ the law generated by the observations for fixed Λ_+. Let $\tau^0 \equiv 0$ and further let

$$\tau^1(x) := \sum_{\gamma \in [n]^{d-1}} \|x - z_\gamma\|_\infty^\beta \mathbb{1}_{\{\|x - z_\gamma\|_\infty \leq \delta/2\}}, \quad x \in [0,1]^{d-1}.$$

Then, for $\Lambda_+^i := \{(x,y) \in [0,1]^d : \tau^i(x) > y\}$, $i = 0,1$, we have

$$\{x_\alpha : \alpha \in [n]^d\} \subset \Lambda_+^0 \cap \Lambda_+^1 = \Lambda_+^1,$$

and therefore $\mathbb{P}_{\Lambda_+^1} = \mathbb{P}_{\Lambda_+^0}$. Since furthermore $\boldsymbol{\lambda}(\Lambda_+^0 \vartriangle \Lambda_+^1) = \int \tau^1(x)\,\mathrm{d}x \asymp \delta^\beta$ and $\tau_i \in \mathcal{H}(\beta,1)$ for $i = 0,1$, we obtain the minimax lower bound

$$\inf_{\widehat{\Lambda}_+} \sup_{\Lambda_+ \in \Xi(\beta,\mathsf{L})} \mathbb{E}_{\Lambda_+}\big[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \vartriangle \Lambda_+)\big] \gtrsim \delta^\beta = N^{-\beta/d},$$

for the class $\Xi(\beta,\mathsf{L})$ of epigraphs of functions $\tau \in \mathcal{H}(\beta,1)$.

Remark C.13. Note that the underlying assumption on the graph orientation, that is Λ_+^0 is an epigraph w.r.t. the d-th coordinate, can be easily circumvented by adapting the candidate set $\mathcal{A}_+$ to represent any graph structure that may only become visible after rotation of the cube $[0,1]^d$. Since a d-dimensional hypercube admits $2d$ faces, the size of the adapted candidate set would increase from $n^{d-1}(n+1)$ to $2dn^{d-1}(n+1)$.

C.4.2 *Estimation of convex sets*

Suppose that model B from Section C.2 holds, i.e., $\Lambda_+^0 \subset \overline{\Lambda}$ is convex. A simple choice for $\mathcal{A}_+$ is given by $\mathcal{A}_+ = \{C^+ : C \subset [0,1]^d \text{ convex}\}$. This choice of candidate sets is however not particularly constructive. Let us therefore propose another family of candidate sets, whose construction follows a similar principle as the one for the graph representation model from the previous subsection.

The basic observation is that by convexity, for any $x \in \operatorname{int} \Lambda_+^0$, the vertical ray $y \mapsto x + ye_d$ intersects $\partial\Lambda_+$ in exactly two points. The natural idea is therefore to build candidate sets from hypercuboids $\mathrm{Sq}_{d-1}(\gamma) \times [\underline{\zeta}_\gamma, \overline{\zeta}_\gamma]$ for $\underline{\zeta}_\gamma, \overline{\zeta}_\gamma$ living on the grid $\mathcal{G}$. Heuristically speaking, as the upper and lower intersection points of the vertical ray can be described by a concave and convex function, respectively, we aim to approximate those by piecewise constant functions on $\widetilde{\mathrm{Sq}}_{d-1}(\gamma)$ in analogy to the graph representation of Section C.4.1. In similarity to Section C.4.1, let

$$\zeta_\gamma^\uparrow := \delta\lceil \delta^{-1} \sup\{x_d : x \in \partial\Lambda_+^0 \cap (\mathrm{Sq}_{d-1}(\gamma)^\circ \times [0,1])\}\rceil, \quad \gamma \in [n]^{d-1},$$
$$\zeta_\gamma^\downarrow := \delta\lfloor \delta^{-1} \inf\{x_d : x \in \partial\Lambda_+^0 \cap (\mathrm{Sq}_{d-1}(\gamma)^\circ \times [0,1])\}\rfloor, \quad \gamma \in [n]^{d-1},$$

be the grid projections of upper and lower limits of the intersection of the convex set Λ_0^+ with the strip $\mathrm{Sq}_{d-1}(\gamma)^\circ \times [0,1]$. Here, the supremum and the infimum of the empty is set to 0. Consider candidate sets

$$\Lambda_+(\zeta) := \bigcup_{\gamma \in [n]^{d-1} : \underline{\zeta}_\gamma < \overline{\zeta}_\gamma} \mathrm{Sq}_{d-1}(\gamma) \times [\underline{\zeta}_\gamma, \overline{\zeta}_\gamma], \quad \underline{\zeta}_\gamma, \overline{\zeta}_\gamma \in \mathcal{G},$$

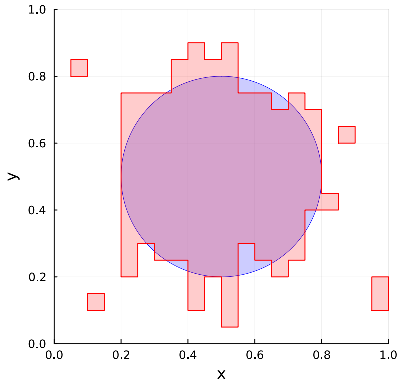
and the minimal tiling

$$\Lambda_+^\updownarrow := \bigcup_{\gamma \in [n]^{d-1} : \zeta_\gamma^\downarrow < \zeta_\gamma^\uparrow} \mathrm{Sq}_{d-1}(\gamma) \times [\zeta_\gamma^\downarrow, \zeta_\gamma^\uparrow]$$

both belonging to

$$\mathcal{A}_+ = \left\{\Lambda_+(\zeta) : \underline{\zeta}_\gamma, \overline{\zeta}_\gamma \in \mathcal{G}, \gamma \in [n]^{d-1}\right\}. \tag{C.26}$$

Then, it holds that $|\mathcal{A}_+| = n^{d-1}((n+n^2)/2+1) \asymp n^{d+1}$. An exemplary illustration in $d = 2$ can be found in Figure C.4.



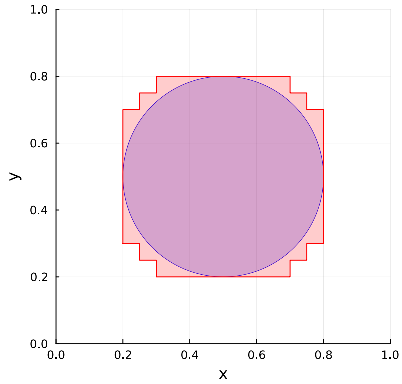


Figure C.4: Approximation (red) of Λ_+^0 (blue) for $\delta = 0.05$.; left: by some $\Lambda_+(\zeta)$ from (C.26); right: by $\Lambda_+^{\updownarrow}$.

In order to apply Theorem C.7, it remains to control the size of the boundary tiling indices $\mathcal{B}$, which can be done with a classical result from convex geometry. By [21, Corollary 2], the boundary of any convex set $C \subset [0,1]^d$ can be covered by at most

$$n^d - (n-2)^d < 2dn^{d-1} = 2d\delta^{-d+1}$$

hypercubes from the tiling $\{\mathrm{Sq}(\alpha) : \alpha \in [n]^d\}$ of $[0,1]^d$. This entails the bound

$$|\mathcal{B}| = |\{\alpha \in [n]^d : \mathrm{Sq}(\alpha)^\circ \cap \partial\Lambda_+^0 \neq \emptyset\}| < 2d\delta^{-d+1}.$$

Consequently, Theorem C.7 and Corollary C.10 yield the following convergence result.

Proposition C.14. *Suppose that Λ_+^0 is convex and $\mathcal{A}_+$ be given by (C.26). Then, for some absolute constant C depending only on $d, \underline{\vartheta}, \overline{\vartheta}, T$ and $\underline{\eta}$ it holds that*

$$\mathbb{E}\left[\boldsymbol{\lambda}(\widehat{\Lambda}_+ \triangle \Lambda_+^0)\right] \leq C\delta.$$

Moreover, if $\boldsymbol{\lambda}(\Lambda_\pm^0) \neq 0$, it holds that $|\widehat{\vartheta}_\pm - \vartheta_\pm^0| \in \mathcal{O}_{\mathbb{P}}(\delta^{1/2})$. In particular, if $\boldsymbol{\lambda}(\Lambda_+^0) \notin \{0,1\}$, we have $\|\widehat{\vartheta} - \vartheta^0\|_{L^1((0,1)^d)} \in \mathcal{O}_{\mathbb{P}}(\delta^{1/2})$.

Remark C.15. For an indication of optimality of the domain convergence rate, let us again consider the regular design image reconstruction problem

$$Y_{\alpha}, = \vartheta(x_\alpha) + \varepsilon_\alpha, \quad \alpha \in [n]^d,$$

from Remark C.12. Let $\Lambda_+^0 \subset (0,1)^d$ be an open hypercube with volume $v_0 \in (0,1)$ and edge length $l_0 = v_0^{1/d}$ such that the corners of the hypercube lie on $\{x_\alpha\}_{\alpha \in [n]^d}$ and let Λ_+^1 be the open hypercube containing Λ_+^0 such that $d_\infty(\Lambda_+^0, \Lambda_+^1) = \delta/2$, where δ is chosen small enough s.t. $\Lambda_+ \subset (0,1)^d$. Then $x_\alpha \in \Lambda_0^+$ iff $x_\alpha \in \Lambda_+^0$ and therefore $\mathbb{P}_{\Lambda_+^0} = \mathbb{P}_{\Lambda_+^1}$. Moreover,

$$\lambda(\Lambda_+^0 \triangle \Lambda_+^1) = (l_0 + \delta/2)^d - l_0^d \geq \frac{d}{2} l_0^{d-1}\delta = \delta \frac{d}{2} v_0^{\frac{d-1}{d}},$$

yielding the minimax lower bound

$$\inf_{\widehat{\Lambda}_+} \sup_{\Lambda_+ \in \mathcal{C}(v_0)} \mathbb{E}_{\Lambda_+}\left[\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+)\right] \gtrsim C(v_0)\delta,$$

for the class $\mathcal{C}(v_0)$ of convex sets in $[0,1]^d$ with volume at least $v_0 \in (0,1)$.

The constructed estimator $\widehat{\Lambda}_+$ has the drawback of generally not being a convex or even connected set. However, we can easily transform our estimator $\widehat{\Lambda}_+$ into a convex estimator $\widehat{\Lambda}_+^{\text{con}}$ that converges at the same rate. To this end we employ a minimum distance fit by choosing an estimator $\widehat{\Lambda}_+^{\text{con}}$ such that

$$\lambda\big(\widehat{\Lambda}_+^{\text{con}} \triangle \widehat{\Lambda}_+\big) \leq \inf_{C \subset [0,1]^d : C \text{ is convex}} \lambda\big(C \triangle \widehat{\Lambda}_+\big) + \delta.$$

Up to a δ-margin, $\widehat{\Lambda}_+^{\text{con}}$ is therefore a convex set with maximal overlap in volume with $\widehat{\Lambda}_+$.

Corollary C.16. *The convex estimator $\widehat{\Lambda}_+^{\text{con}}$ satisfies*

$$\mathbb{E}\left[\lambda(\widehat{\Lambda}_+^{\text{con}} \triangle \Lambda_+^0)\right] \leq (2C+1)\delta,$$

where C is the constant from Proposition C.14.

Proof. By triangle inequality for the symmetric difference pseudometric, it follows that

$$\begin{aligned}\mathbb{E}\left[\lambda\big(\Lambda_+^0 \triangle \widehat{\Lambda}_+^{\text{con}}\big)\right] &\leq \mathbb{E}\left[\lambda\big(\widehat{\Lambda}_+^{\text{con}} \triangle \widehat{\Lambda}_+\big)\right] + \mathbb{E}\left[\lambda\big(\Lambda_+^0 \triangle \widehat{\Lambda}_+\big)\right] \\ &\leq 2\mathbb{E}\left[\lambda\big(\Lambda_+^0 \triangle \widehat{\Lambda}_+\big)\right] + \delta,\end{aligned}$$

where we used that since Λ_+^0 is convex we have

$$\lambda\big(\widehat{\Lambda}_+^{\text{con}} \triangle \widehat{\Lambda}_+\big) \leq \lambda\big(\Lambda_+^0 \triangle \widehat{\Lambda}_+\big) + \delta.$$

The assertion therefore follows from Proposition C.14. ■

Given the estimator $\widehat{\Lambda}_+$, numerical implementation of $\widehat{\Lambda}_+^{\text{con}}$ can be conducted with methods for convexity constrained image segmentation based on the binary image input $\{(x_\alpha, \mathbb{1}_{\widehat{\Lambda}_+}(x_\alpha)) : \alpha \in [n]^d\}$, see, e.g., the recent implicit representation approach in [32].

C.5 Conclusion and outlook

Before discussing limitations and possible extensions of our work, let us briefly summarize our results. We have studied a change estimation problem for a stochastic heat equation (C.1) in $d \geq 2$. The underlying space is partitioned into $\Lambda_- \cup \Lambda_+$ by a separating hypersurface $\Gamma = \partial\Lambda_+$, where the piecewise constant diffusivity ϑ exhibits a jump. Following a CUSUM approach, we have constructed an M-estimator $\widehat{\Lambda}_+$ based on local measurements on a fixed uniform δ-grid that exhibits certain analogies to regular design estimators in statistical image reconstruction. Our main result, Theorem C.7, shows how the convergence properties of our estimator are determined by the number of tiles that are sliced by Γ. The estimation principle and rates are made concrete for two specific models that impose shape restrictions on the change domain Λ_+: (A) a graph representation of Γ with Hölder smoothness $\beta \in (0,1]$, and (B) a convex shape. We have established the rates of convergences δ^β for model A and δ for model B with respect to the symmetric difference risk $\mathbb{E}[\lambda(\widehat{\Lambda}_+ \triangle \Lambda_+)]$, which are the optimal rates of convergence in the corresponding image reconstruction problems with regular design. Furthermore, the diffusivity parameters can be recovered with rate $\delta^{\beta/2}$ and $\delta^{1/2}$. To conclude the paper, let us now give an outlook on potential future work that can build on our results.

Based on Theorem C.7, an extension of Proposition C.11 to a known number m of change interfaces $\tau_1, \ldots, \tau_m$ that yields a partition $\overline{\Lambda} = \bigcup_{0 \le i \le m} \Lambda_i$ of layers Λ_i with alternating diffusivities $\vartheta_\pm$, is straightforward. Assuming that each τ_m belongs to $\mathcal{H}(\beta, L)$, the same rate of convergence can be established. For the canonical choice $\mathcal{A}_+ = \mathcal{P}$, our general estimator from Theorem C.7 would also adapt to an *unknown* number of layers m, but it might be interesting in this scenario to develop a more implementation friendly procedure.

Having an unknown number of layers, or, more generally, an unknown number of "impurities" is also particularly interesting from a testing perspective. In this spirit, a further model extension would allow a partition $\overline{\Lambda} = \bigcup_{0 \le i \le m} \Lambda_i$ with $\vartheta \equiv \vartheta_i$ on Λ_i and $\vartheta_i \ne \vartheta_j$ for $j \ne i$, which can be used, for instance, to model sediment layers. This model extension is unproblematic from an SPDE perspective, but poses additional statistical challenges.

More generally, besides spatial change areas, future work could contain temporal change points, i.e., the diffusivity $\vartheta = \vartheta(t,x)$ is also discontinuous in time, thereby allowing for the modeling of thermal spikes. In this case, tools from change point detection in time series have to be incorporated into the estimation procedure and online estimation becomes an intriguing question.

In this paper, we have fixed the parameters $\vartheta_\pm$ and therefore also the absolute jump height $\eta = |\vartheta_+ - \vartheta_-|$. Extensions of the result to a δ-dependent, but non-vanishing jump height, i.e., $\eta = \eta(\delta) \ge \underline{\eta}$ for some fixed $\underline{\eta} > 0$, are straightforward. In contrast, the vanishing jump height regime $\eta \to 0$ as $\delta \to 0$ that has been considered for the one-dimensional change point problem in [29], introduces significant technical challenges that require a sharper concentration analysis. Similarly to how the limit result from [29] in this regime draws analogies to classical change point limit theorems, in our multivariate case one would expect asymptotics that are comparable to [24].

As mentioned in Remark C.12 and Remark C.15, the convergence rates for $N^{-\beta/d}$ and $N^{-1/d}$, respectively, are optimal in the related image reconstruction problem when working with a regular design. However, as alluded to before, it is shown in [19] that the minimax optimal rate for irregular measurement designs that introduce a certain level of randomness is given by $N^{-\beta/(\beta+d-1)}$ for arbitrary $\beta > 0$, which is not only substantially faster for $\beta \in (0,1)$, but also allows to exploit higher-order smoothness of the change interface. Appropriately introducing such randomness in the measurement locations x_α of the local observation scheme while preserving favorable probabilistic properties such as independence of the associated Brownian motions $B_{\delta,\alpha}$ is a conceptually challenging task that could contribute to improve change domain estimation performance in the given heat equation model.

Finally, let us reiterate that we have focused on optimal change domain estimation for regular local measurements and have not attempted to optimize estimation rates for the diffusivity parameters $\vartheta_\pm$. For the binary image reconstruction model with regeression function $\vartheta(x) = \vartheta_+\mathbf{1}(x \in \Lambda_+) + \vartheta_-\mathbf{1}(x \in \Lambda_-)$, [19, Theorem 5.1.2] establishes the typical parametric rate $N^{-1/2} = \delta^{d/2}$ for $\vartheta_\pm$. On the other hand, the minimax rate $\delta^{d/2+1}$ for a constant diffusivity is proven in Paper A, which can also be obtained in the one-dimensional change point estimation problem for a stochastic heat equation by introducing an additional nuisance parameter $\vartheta_\circ$ in the estimation procedure that reduces the bias from a constant approximation $\vartheta \equiv \vartheta_\circ$ on a proposed spatial change interval, cf. [29, Theorem 3.12]. This demonstrates a significant difference between heat diffusivity and image estimation. Optimal diffusivity estimation in the here considered change estimation problem is therefore an especially relevant task for future work, which may also open the door to the investigation of fully nonparametric diffusivities $\vartheta(x) = \vartheta_-(x)\mathbb{1}_{\Lambda_-}(x) + \vartheta_+(x)\mathbb{1}_{\Lambda_+}(x)$.

Acknowledgements AT and LT gratefully acknowledge financial support of Carlsberg Foundation Young Researcher Fellowship grant CF20-0640 “Exploring the potential of nonparametric modelling of complex systems via SPDEs”.

REFERENCES

[1] R. Altmeyer, T. Bretschneider, J. Janák, and M. Reiß. "Parameter Estimation in an SPDE Model for Cell Repolarisation". In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1 (2022), pp. 179–199.

[2] R. Altmeyer, I. Cialenco, and G. Pasemann. "Parameter estimation for semilinear SPDEs from local measurements". In: *Bernoulli* 29.3 (2023), pp. 2035–2061.

[3] R. Altmeyer, I. Cialenco, and M. Reiß. *Statistics for SPDEs*. https://sites.google.com/view/stats4spdes. Accessed: 2024-07-30.

[4] R. Altmeyer and M. Reiß. "Nonparametric estimation for linear SPDEs from local measurements". In: *Annals of Applied Probability* 31.1 (2021), pp. 1–38.

[5] C. J. Bishop and Y. Peres. *Fractals in probability and analysis*. Vol. 162. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2017.

[6] C. Chong. "High-frequency analysis of parabolic stochastic PDEs". In: *Annals of Statistics* 48.2 (2020), pp. 1143–1167.

[7] I. Cialenco. "Statistical inference for SPDEs: an overview". In: *Statistical Inference for Stochastic Processes* 21.2 (2018), pp. 309–329.

[8] I. Cialenco, F. Delgado-Vences, and H.-J. Kim. "Drift estimation for discretely sampled SPDEs". In: *Stochastics and Partial Differential Equations: Analysis and Computations* 8 (2020), pp. 895–920.

[9] I. Cialenco, S. V. Lototsky, and J. Pospíšil. "Asymptotic properties of the maximum likelihood estimator for stochastic parabolic equations with additive fractional Brownian motion". In: *Stochastics and Dynamics* 9.02 (2009), pp. 169–185.

[10] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.

[11] E. B. Davies. *Heat kernels and spectral theory*. Vol. 92. Cambridge Tracts in Mathematics. Cambridge University Press, 1990.

[12] E. B. Davies. *Spectral theory and differential operators*. Vol. 42. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1995.

[13] F. Hildebrandt and M. Trabs. "Parameter estimation for SPDEs based on discrete observations in time and space." In: *Electronic Journal of Statistics* 15 (2021), pp. 2716–2776.

[14] M. Huebner and B. Rozovskii. "On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's". In: *Probability Theory and Related Fields* 103.2 (1995), pp. 143–163.

[15] J. Janák and M. Reiß. "Parameter estimation for the stochastic heat equation with multiplicative noise from local measurements". In: *Stochastic Process. Appl.* 175 (2024), Paper No. 104385.

[16] Y. Kaino and M. Uchida. "Parametric estimation for a parabolic linear SPDE model based on discrete observations". In: *Journal of Statistical Planning and Inference* 211 (2021), pp. 190–220.

[17] A. P. Korostelëv, L. Simar, and A. B. Tsybakov. "Efficient estimation of monotone boundaries". In: *Ann. Statist.* 23.2 (1995), pp. 476–489.

[18] A. P. Korostelev, L. Simar, and A. B. Tsybakov. "On estimation of monotone and convex boundaries". In: *Publ. Inst. Statist. Univ. Paris* 39.1 (1995), pp. 3–18.

[19] A. P. Korostelëv and A. B. Tsybakov. *Minimax theory of image reconstruction*. Vol. 82. Lecture Notes in Statistics. Springer, 1993.

[20] C. Krishnamoorthy and E. Carlstein. "Practical Considerations in Boundary Estimation: Model-Robustness, Efficient Computation, and Bootstrapping". In: *Lecture Notes-Monograph Series* 23 (1994), pp. 177–193.

[21] M. Lassak. "Covering the Boundary of a Convex Set by Tiles". In: *Proceedings of the American Mathematical Society* 104.1 (1988), pp. 269–272.

[22] W Liu and S. Lototsky. "Parameter estimation in hyperbolic multichannel models". In: *Asymptotic Analysis* 68 (2010), pp. 223–248.

[23] S. V. Lototsky. "Parameter Estimation for Stochastic Parabolic Equations: Asymptotic Properties of a Two-Dimensional Projection-Based Estimator". In: *Statistical Inference for Stochastic Processes* 6.1 (2003), pp. 65–87.

[24] H. G. Müller and K.-S. Song. "A set-indexed process in a two-region image". In: *Stochastic Processes and their Applications* 62 (1996), pp. 87–101.

[25] H. G. Müller and K.-S. Song. "Cube Splitting in Multidimensional Edge Estimation". In: *Lecture Notes-Monograph Series* 23 (1994), pp. 210–223.

[26] H. G. Müller and K.-S. Song. "Maximin estimation of multidimensional boundaries". In: *Journal of Multivariate Analysis* 50.2 (1994), pp. 265–281.

[27] G. Pasemann and W. Stannat. "Drift estimation for stochastic reaction-diffusion systems". In: *Electronic Journal of Statistics* 14.1 (2020), pp. 547–579.

[28] P. Qiu. "Jump Surface Estimation, Edge Detection, and Image Restoration". In: *Journal of the American Statistical Association* 102.478 (2007), pp. 745–756.

[29] M. Reiß, C. Strauch, and L. Trottner. *Change point estimation for a stochastic heat equation*. 2023. arXiv: 2307.10960 [math.ST].

[30] M. Rudemo and H. Stryhn. "Approximating the Distribution of Maximum Likelihood Contour Estimators in Two-Region Images". In: *Scandinavian Journal of Statistics* 21.1 (1994), pp. 41–55.

[31] M. Rudemo and H. Stryhn. "Boundary Estimation for Star-Shaped Objects". In: *Lecture Notes-Monograph Series* 23 (1994), pp. 276–283.

[32] J. P. Schneider, M. Fatima, J. Lukasik, A. Kolb, M. Keuper, and M. Moeller. "Implicit Representations for Constrained Image Segmentation". In: *International Conference on Machine Learning*. PMLR. Forthcoming.

[33] Y. Tonaki, Y. Kaino, and M. Uchida. "Parameter estimation for linear parabolic SPDEs in two space dimensions based on high frequency data". In: *Scandinavian Journal of Statistics* 50.4 (2023), pp. 1568–1589.

[34] A. B. Tsybakov. "Multidimensional change-point problems and boundary estimation". In: *Change-point problems (South Hadley, MA, 1992)*. Vol. 23. IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Hayward, CA, 1994, pp. 317–329.

[35] E. Ziebell. *Non-parametric estimation for the stochastic wave equation*. 2024. arXiv: 2404.18823 [math.ST].

Parameter estimation in hyperbolic linear SPDEs from multiple measurements

D

Anton Tiepner and Eric Ziebell

Abstract

The coefficients of elastic and dissipative operators in a linear hyperbolic SPDE are jointly estimated using multiple spatially localised measurements. As the resolution level of the observations tends to zero, we establish the asymptotic normality of an augmented maximum likelihood estimator. The rate of convergence for the dissipative coefficients matches rates in related parabolic problems, whereas the rate for the elastic parameters also depends on the magnitude of the damping. The analysis of the observed Fisher information matrix relies upon the asymptotic behaviour of rescaled M, N-functions generalising the operator sine and cosine families appearing in the undamped wave equation. In contrast to the energetically stable undamped wave equation, the M, N-functions emerging within the covariance structure of the local measurements have additional smoothing properties similar to the heat kernel, and their asymptotic behaviour is analysed using functional calculus.

D.1 Introduction

We study parameter estimation for a general second-order stochastic Cauchy problem

$$\ddot{u}(t) = A_\vartheta u(t) + B_\eta \dot{u}(t) + \dot{W}(t), \quad 0 < t \leq T, \tag{D.1}$$

driven by space-time white noise $\dot{W}$ on an open, bounded spatial domain $\Lambda \subset \mathbb{R}^d$. The differential operators A_ϑ and B_η defined through

$$A_\vartheta = \sum_{i=1}^{p} \vartheta_i (-\Delta)^{\alpha_i}, \quad \alpha_1 > \cdots > \alpha_p \geq 0,$$
$$B_\eta = \sum_{j=1}^{q} \eta_j (-\Delta)^{\beta_j}, \quad \beta_1 > \cdots > \beta_q \geq 0,$$

are parameterised by unknown constants $\vartheta \in \mathbb{R}^p$, $\eta \in \mathbb{R}^q$. In general, such equations model elastic systems, and we refer to A_ϑ as the elastic operator while B_η is called the dissipation (or damping) operator.

In the absence of any damping ($B_\eta = 0$) and noise, a prototypical example of (D.1) is the isotropic plate equation (without any in-plane forces, thermal loads or elastic foundation)

$$\rho h \ddot{u}(t) = -D\Delta^2 u(t), \tag{D.2}$$

modelling the bending of elastic plates over time. The parameters governing (D.2) are the material density ρ and the bending stiffness $D = \frac{h^3 E}{12(1-\nu^2)}$. The bending stiffness D depends on the plate thickness h, the material-specific Poisson-ratio ν and Young's modulus E. Numerous extensions and applications of such equations can, for instance, be found in [22, 31]. To account for the system's energy loss, damping is added to the equation, where a higher differential order of B_η describes stronger damping. In fact, a parabolic behaviour of (D.1) is obtained for $\beta_1 > 0$ due to the smoothing effects within the related C_0-semigroup [9, 13]. In closely related situations, i.e. when both A_ϑ and B_η are negative operators, the C_0-semigroup has been shown to become analytic if and only if $2\beta_1 \geq \alpha_1$, where a borderline case occurs under equality, cf. [10].

While parameter estimation for SPDEs is well-studied in second-order parabolic equations, e.g. in Paper A and [11, 14, 15, 26, 30, 37] and the references therein, the literature on higher-order hyperbolic equations is limited. We refer to [39] and the references mentioned there for studies of the (non-)parametric wave equation. In [1, 2], the authors considered a *weakly damped* system, i.e. $\beta_1 = 0$, and developed a first approach for identifying coefficients of the elastic operator in a Kalman filtering problem based on the methods of sieves. [27] studied equations driven by a fractional cylindrical Brownian motion and derived a consistent estimator of a scalar drift coefficient using the ergodicity of the underlying system. Based on spectral measurements $(\langle u(t), e_j\rangle)_{0\leq t\leq T}$, $j = 1, \dots, N$, where $(e_j)_{j\geq 1}$ forms an orthonormal basis of $L^2(\Lambda)$ composed of eigenvectors for A_ϑ and B_η, [23] constructed maximum-likelihood estimators and established the asymptotic normality for diagonalisable hyperbolic equations given that the number N of observed Fourier-modes tends to infinity.

In contrast, our estimator is based on continuous observations of local measurement processes

$$u_{\delta,k} = (\langle u(t), K_{\delta,x_k}\rangle)_{0\leq t\leq T}, \quad u^{\gamma}_{\delta,k} = (\langle u(t), (-\Delta)^{\gamma} K_{\delta,x_k}\rangle)_{0\leq t\leq T},$$

for locations $x_1, \dots, x_N \in \Lambda$ and $\gamma \in \{\alpha_i, \beta_j | 1 \leq i \leq p; 1 \leq j \leq q\}$. The *point spread functions* [7, 8] K_{δ,x_k} are compactly supported functions taking non-zero values in an area centred around x_k with radius δ. Local measurements emerge naturally as they describe the physical limitation of measuring $u(t, x_k)$, which, in general, is only possible up to a convolution with a point spread function. Local observations were introduced to the field of statistics for SPDEs in [5], where the authors investigated a stochastic heat equation with a spatially varying diffusivity. It was shown that the diffusivity at location $x_k \in \Lambda$ can be estimated based on a single local measurement process at x_k as the resolution level δ tends to zero. The local observation scheme turned out to be robust under semilinearities [3, 4], multiplicative noise [17], discontinuities [32] or lower-order perturbation terms (Paper A and B). In contrast to the estimation of the diffusivity, the identifiability of transport or reaction coefficients necessarily requires an increasing amount $N \to \infty$ of measurements. In the recent contribution [39], the local measurement approach was extended to hyperbolic problems and the non-parametric wave speed in the undamped stochastic wave equation was estimated by relating the observed Fisher information to the energetic behaviour of an associated deterministic wave equation.

Based on the local measurement approach, we construct the *augmented* maximum likelihood estimator (MLE) $(\widehat{\vartheta}_\delta, \widehat{\eta}_\delta)^\top \in \mathbb{R}^{p+q}$ and prove the asymptotic normality of

$$\begin{pmatrix} N^{1/2}\delta^{-2\alpha_i+\alpha_1+\beta_1}(\widehat{\vartheta}_{\delta,i} - \vartheta_i) \\ N^{1/2}\delta^{-2\beta_j+\beta_1}(\widehat{\eta}_{\delta,j} - \eta_j) \end{pmatrix}_{i\leq p, j\leq q}, \quad \delta \to 0,$$

with $N = N(\delta)$ measurements. The consistent estimation for ϑ_i holds if $N^{1/2}\delta^{-2\alpha_i+\alpha_1+\beta_1} \to \infty$, whereas η_j can be estimated in the asymptotic regime $N^{1/2}\delta^{-2\beta_j+\beta_1} \to \infty$. In particular, estimating elastic coefficients is more difficult under higher dissipation, while damping coefficients are unaffected by the order of the elastic operator A_ϑ and their convergence rates reflect the rates obtained in advection-diffusion equations, cf. Paper A. For the maximal number of non-overlapping observations $N \asymp \delta^{-d}$, our convergence rates match the rates obtained in the spectral approach up to specific boundary cases. In the weakly damped case, i.e. $\beta_1 = 0$, we confirm the results in [24]. That is, the dependence of the time horizon T of the asymptotic variance resembles the explosive, stable and ergodic cases of the maximum likelihood drift estimator for an Ornstein–Uhlenbeck process, cf. [21, Proposition 3.46]. In the structural damped case ($\beta_1 > 0$), the asymptotic variance is of order T^{-1} instead.

We begin this paper by specifying the model and discussing properties of the local measurements in Section D.2. The augmented MLE is constructed and analysed in Section D.3, and the CLT is established. The section additionally contains various remarks and examples, complemented by a numerical study underpinning and illustrating the main result. All proofs are deferred to Section D.4.

D.2 Setup

D.2.1 *Notation*

Throughout this paper, we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ with a fixed time horizon $T < \infty$. We write $a \lesssim b$ if $a \le Mb$ holds for a universal constant M, independent of the resolution level $\delta > 0$ and the number of spatial points N. Unless stated otherwise, all limits are to be understood as the spatial resolution level tending to zero, i.e. for $\delta \to 0$. For an open set $\Lambda \subset \mathbb{R}^d$, $L^2(\Lambda)$ is the usual L^2-space with the inner product $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_{L^2(\Lambda)}$. The Euclidean inner product and distance of two vectors $a, b \in \mathbb{R}^p$ are denoted by $a^\top b$ and $|a - b|$, respectively. We abbreviate the Laplace operator with Dirichlet boundary conditions on the bounded spatial domain Λ by Δ and on the unbounded spatial domain $\mathbb{R}^d$ by Δ_0. Let $H^k(\Lambda)$ denote the usual Sobolev spaces, and denote by $H_0^1(\Lambda)$ the completion of $C_c^\infty(\Lambda)$, the space of smooth compactly supported functions, relative to the $H^1(\Lambda)$ norm. As in [19], let $\dot{H}^{2s}(\Lambda) := \mathcal{D}((-\Delta)^s)$ for $s > 0$ be the domain of the fractional Laplace operator on $L^2(\Lambda)$ with Dirichlet boundary conditions. The order of a differential operator D is denoted by $\operatorname{ord}(D)$.

D.2.2 *The model*

Consider the second-order stochastic Cauchy problem

$$\begin{cases} \mathrm{d}v(t) = \big(A_\vartheta u(t) + B_\eta v(t)\big)\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \le T, \\ \mathrm{d}u(t) = v(t)\,\mathrm{d}t, \\ u(0) = u_0 \in L^2(\Lambda), \\ v(0) = v_0 \in L^2(\Lambda), \\ u(t,x) = v(t,x) = 0, \quad 0 \le t \le T, \quad x \in \Lambda|_{\partial\Lambda} \end{cases} \tag{D.3}$$

on an open, bounded domain $\Lambda \subset \mathbb{R}^d$ having C^2-boundary $\partial\Lambda$. We assume Dirichlet boundary conditions and a driving space-time white noise $\mathrm{d}W$ in (D.3). The elasticity and damping operators A_ϑ and B_η are parameterised by $\vartheta \in \mathbb{R}^p$ and $\eta \in \mathbb{R}^q$ and given by

$$\begin{aligned} A_\vartheta &= \sum_{i=1}^{p} \vartheta_i(-\Delta)^{\alpha_i}, \quad D(A_\vartheta) = D((-\Delta)^{\alpha_1}) = \dot{H}^{2\alpha_1}(\Lambda), \\ B_\eta &= \sum_{j=1}^{q} \eta_j(-\Delta)^{\beta_j}, \quad D(B_\eta) = D((-\Delta)^{\beta_1}) = \dot{H}^{2\beta_1}(\Lambda), \end{aligned} \tag{D.4}$$

with $\alpha_1 > 0$ and $0 \le \alpha_i, \beta_j < \infty$ satisfying $\alpha_1 > \alpha_2 > \cdots > \alpha_p$ and $\beta_1 > \beta_2 > \cdots > \beta_q$.

Example D.1.

(a) Weakly damped wave equation ($\beta_1 = 0$): $A_\vartheta = -\vartheta_1\Delta$, $B_\eta = \eta_1$.

(b) Clamped plate equation:

1) Weakly damped ($\beta_1 = 0$) : $A_\vartheta = \vartheta_1\Delta^2$, $B_\eta = \eta_1$.
2) Structurally damped ($0 < \beta_1 < \alpha_1$): $A_\vartheta = \vartheta_1\Delta^2$, $B_\eta = -\eta_1\Delta$.
3) Strongly damped ($\beta_1 = \alpha_1$): $A_\vartheta = \vartheta_1\Delta^2$, $B_\eta = \eta_1\Delta^2$.

Figure D.1 displays a heatmap illustrating both the weakly and structurally damped plate equation in one spatial dimension. The solution of the SPDEs were approximated on a fine time-space grid using the finite difference scheme associated with the semi-implicit Euler–Maruyama method, see [25, Chapter 10]. Additional smoothing properties in the structurally damped case due to the dissipative operator B_η result in an accelerated energetic decay in comparison to the weakly damped case.

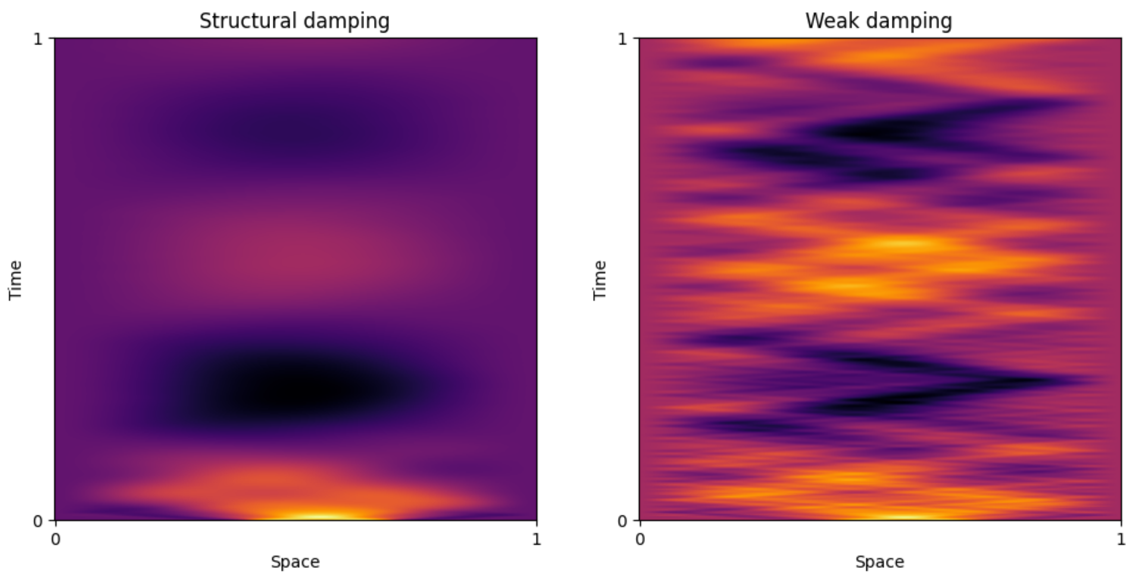


Figure D.1: Realisation of the solution $u(t,x)$ to the clamped plate equation on $(0,1)$; (left) $\ddot{u}(t) = -0.3\Delta^2 u(t) + 0.3\Delta\dot{u}(t) + \dot{W}(t)$; (right) $\ddot{u}(t) = -0.3\Delta^2 u(t) - 0.3\dot{u}(t) + \dot{W}(t)$.

Throughout the rest of the paper, we impose the following assumptions on the parameters.

Assumption D.2 (Assumption on the parameters).

(i) $\vartheta_1 < 0$;

(ii) *If* $\beta_1 > 0$, *then* $\eta_1 < 0$;

(iii) $\alpha_1 \geq 2\beta_1$ *and if* $\alpha_1 = 2\beta_1$ *then* $\vartheta_1 + \eta_1^2/4 < 0$.

Assumption D.2 does not guarantee that either A_ϑ or B_η are, in general, negative operators. Instead, it implies that at least all but finitely many of their eigenvalues are negative. Moreover, conditions (i) and (iii) are necessary to ensure that the difference $L_{\vartheta,\eta} := -A_\vartheta - B_\eta^2/4$ is a positive operator, which itself is not required for the proofs and is just assumed for technical reasons in Section D.4 as all arguments also carry over to the non-positive case, resulting in a complex-valued operator, cf. Remark D.10.

D.2.3 *Local measurements*

For $\delta > 0$, $y \in \Lambda$ and $z \in L^2(\mathbb{R}^d)$ we define the rescaling

$$\begin{aligned} \Lambda_{\delta,y} &= \{\delta^{-1}(x-y) : x \in \Lambda\}, \\ z_{\delta,y}(x) &= \delta^{-d/2} z(\delta^{-1}(x-y)), \quad x \in \mathbb{R}^d. \end{aligned} \tag{D.5}$$

Fix a function $K \in H^{\lceil 2\alpha_1 \rceil}(\mathbb{R}^d)$ with compact support. By a slight abuse of notation, we define local measurements at the location $x \in \Lambda$ with resolution level δ as the continuously observed processes $u_{\delta,x}, u_{\delta,x}^{\Delta_i}, v_{\delta,x}, v_{\delta,x}^{\Delta_j}$ where for $i = 1, \dots, p$ and $j = 1, \dots, q$:

$$\begin{aligned} u_{\delta,x} &= (\langle u(t), K_{\delta,x}\rangle)_{0\leq t\leq T}, \quad & u_{\delta,x}^{\Delta_i} &= (\langle u(t), (-\Delta)^{\alpha_i} K_{\delta,x}\rangle)_{0\leq t\leq T}, \\ v_{\delta,x} &= (\langle v(t), K_{\delta,x}\rangle)_{0\leq t\leq T}, \quad & v_{\delta,x}^{\Delta_j} &= (\langle v(t), (-\Delta)^{\beta_j} K_{\delta,x}\rangle)_{0\leq t\leq T}. \end{aligned}$$

Analogously to [39], these local measurements satisfy the following Itô-dynamics

$$\mathrm{d}u_{\delta,x}(t) = v_{\delta,x}(t)\,\mathrm{d}t, \quad \mathrm{d}v_{\delta,x}(t) = \left(\sum_{i=1}^{p} \vartheta_i u_{\delta,x}^{\Delta_i}(t) + \sum_{j=1}^{q} \eta_j v_{\delta,x}^{\Delta_j}(t)\right)\mathrm{d}t + \|K\|_{L^2(\mathbb{R}^d)}\,\mathrm{d}W_x(t), \tag{D.6}$$

with scalar Brownian motions $(W_x(t))_{0\leq t\leq T} = (\|K\|_{L^2(\mathbb{R}^d)}^{-1} \langle W(t), K_{\delta,x}\rangle)_{0\leq t\leq T}$, which become mutually independent provided that

$$\langle K_{\delta,x}, K_{\delta,x'}\rangle = \|K\|_{L^2(\mathbb{R}^d)}^2 \delta_{x,x'} = 0, \quad x, x' \in \mathbb{R}^d,$$

where the Kronecker-delta $\delta_{x,x'}$ evaluates to zero for $x \neq x'$.

For locations $x_1, \dots, x_N \in \Lambda$, we define the vector process $Y_\delta \in L^2([0,T]; \mathbb{R}^{(p+q)\times N})$ of observations through

$$Y_{\delta,k} = \begin{pmatrix} u_{\delta,x_k}^{\Delta_1} & \dots & u_{\delta,x_k}^{\Delta_p} & v_{\delta,x_k}^{\Delta_1} & \dots & v_{\delta,x_k}^{\Delta_q} \end{pmatrix}^\top \in \mathbb{R}^{p+q}, \quad k = 1, \dots, N. \tag{D.7}$$

Remark D.3 (Accessibility of the measurements). The measurements $(u_{\delta,x}^{\Delta_i}, i = 1, \dots, p)$, can be approximated by observing $u_{\delta,y}$, $y \in \Lambda$, on a fine spatial grid in Λ. Moreover, all the measurements wrt. v, i.e. $v_{\delta,x}$ and $v_{\delta,x}^{\Delta_j}$, $j \leq q$, can be obtained by differentiating $u_{\delta,x}$ and $u_{\delta,x}^{\Delta_j} := \langle u(\cdot), (-\Delta)^{\beta_j} K_{\delta,x}\rangle$ in time.

D.3 The estimator

Motivated by a general Girsanov theorem, as described in detail in [5] or Paper A, the augmented MLE $(\widehat{\vartheta}_\delta, \widehat{\eta}_\delta)^\top \in \mathbb{R}^{p+q}$ is given by

$$\begin{pmatrix} \widehat{\vartheta}_\delta \\ \widehat{\eta}_\delta \end{pmatrix} = \mathcal{I}_\delta^{-1} \sum_{k=1}^N \int_0^T Y_{\delta,k}(t)\, \mathrm{d}v_{\delta,x_k}(t) \tag{D.8}$$

with the observed Fisher information matrix

$$\mathcal{I}_\delta = \sum_{k=1}^N \int_0^T Y_{\delta,k}(t) Y_{\delta,k}(t)^\top \, \mathrm{d}t. \tag{D.9}$$

Clearly, the matrix $\mathcal{I}_\delta$ is symmetric and positive semidefinite. By plugging the Itô-dynamics (D.6) into the definition of the estimator (D.8), we obtain the decomposition

$$\begin{pmatrix} \widehat{\vartheta}_\delta \\ \widehat{\eta}_\delta \end{pmatrix} = \begin{pmatrix} \vartheta_\delta \\ \eta_\delta \end{pmatrix} + \|K\|_{L^2(\mathbb{R}^d)} \mathcal{I}_\delta^{-1} \mathcal{M}_\delta \tag{D.10}$$

on the event $\{\det(\mathcal{I}_\delta) > 0\}$ with the martingale part

$$\mathcal{M}_\delta = \sum_{k=1}^N \int_0^T Y_{\delta,k}(t)\, \mathrm{d}W_{x_k}(t).$$

As the limiting object of the rescaled observed Fisher information is deterministic and invertible (see Theorem D.5 below), $\mathcal{I}_\delta$ will itself be invertible for sufficiently small δ, cf. [20, Theorem A.7.7, Corollary A.7.8].

Assumption D.4 (Regularity of the kernel and the initial condition).

1. *The locations x_k, $k = 1, \ldots, N$, belong to a fixed compact set $\mathcal{J} \subset \Lambda$, which is independent of δ and N. There exists $\delta' > 0$ such that $\operatorname{supp}(K_{\delta,x_k}) \cap \operatorname{supp}(K_{\delta,x_l}) = \emptyset$ for $k \neq l$ and all $\delta \leq \delta'$.*

2. *There exists a compactly supported function $\widetilde{K} \in H^{\lceil 2\alpha_1 \rceil + 2\lceil \alpha_1 \rceil}(\mathbb{R}^d)$ such that $K = \Delta_0^{\lceil \alpha_1 \rceil} \widetilde{K}$.*

3. *The functions $(-\Delta)^{\alpha_i - (\alpha_1+\beta_1)/2} K$ are linearly independent for all $i = 1, \ldots, p$, and the functions $(-\Delta)^{\beta_j - \beta_1/2} K$ are linearly independent for all $j = 1, \ldots, q$.*

4. *The initial condition $(u_0, v_0)^\top$ in (D.3) takes values in $\dot{H}^{2\alpha_1}(\Lambda) \times \dot{H}^{\alpha_1}(\Lambda)$.*

Assumption D.4 (i) ensures that $\langle K_{\delta,x_k}, K_{\delta,x_l} \rangle = \|K\|^2_{L^2(\mathbb{R}^d)} \delta_{k,l}$ with the Kronecker-delta $\delta_{k,l}$. Consequently, the Brownian motions W_{x_k} become mutually independent if δ is sufficiently small. Thus, $\mathcal{I}_\delta$ forms the quadratic variation process of the time-martingale $\mathcal{M}_\delta$, and we expect $\mathcal{I}_\delta^{-1/2}\mathcal{M}_\delta$ to be asymptotically normally distributed. Both (ii) and (iii) guarantee that the limiting object of the observed Fisher information is well-defined and invertible, while (iv) ensures that the initial condition is asymptotically negligible. In principle, the required smoothness in (ii) can be relaxed, depending on the dimension d and the identifiability of the appearing parameters in (D.3), but is kept for the simplification of the proofs.

We define a diagonal matrix of scaling coefficients $\rho_\delta \in \mathbb{R}^{(p+q)\times(p+q)}$ via

$$(\rho_\delta)_{ii} := \begin{cases} N^{-1/2}\delta^{2\alpha_i-\alpha_1-\beta_1}, & 1 \le i \le p, \\ N^{-1/2}\delta^{2\beta_{i-p}-\beta_1}, & p < i \le p+q, \end{cases} \tag{D.11}$$

and the constant $C(\eta_1, T)$ through

$$C(\eta_1, T) := \begin{cases} \frac{e^{T\eta_1}-T\eta_1-1}{2\eta_1^2}, & \eta_1 \neq 0, \\ \frac{T^2}{4}, & \eta_1 = 0. \end{cases} \tag{D.12}$$

The following result shows the asymptotic normality of the estimator (D.8).

Theorem D.5 (Asymptotic behaviour of the joint estimator). *Grant Assumption D.4.*

(i) The matrix $\Sigma_{\vartheta,\eta} \in \mathbb{R}^{(p+q)\times(p+q)}$, given by

$$\Sigma_{\vartheta,\eta} := \begin{pmatrix} \Sigma_{1,\vartheta,\eta} & 0 \\ 0 & \Sigma_{2,\vartheta,\eta} \end{pmatrix}$$

with

$$(\Sigma_{1,\vartheta,\eta})_{ij} = \begin{cases} -\frac{C(\eta_1,T)}{\vartheta_1}\|(-\Delta_0)^{(\alpha_i+\alpha_j-\alpha_1)/2}K\|^2_{L^2(\mathbb{R}^d)}, & \beta_1 = 0, \\ \frac{T}{2\vartheta_1\eta_1}\|(-\Delta_0)^{(\alpha_i+\alpha_j-\alpha_1-\beta_1)/2}K\|^2_{L^2(\mathbb{R}^d)}, & \beta_1 > 0, \end{cases}$$

$$(\Sigma_{2,\vartheta,\eta})_{kl} = \begin{cases} C(\eta_1,T)\|K\|^2_{L^2(\mathbb{R}^d)}, & \beta_1 = 0, \\ -\frac{T}{2\eta_1}\|(-\Delta_0)^{(\beta_k+\beta_l-\beta_1)/2}K\|^2_{L^2(\mathbb{R}^d)}, & \beta_1 > 0, \end{cases}$$

for $1 \le i,j \le p$ and $1 \le k,l \le q$, is well-defined and invertible. In particular, the observed Fisher information matrix admits the convergence

$$\rho_\delta \mathcal{I}_\delta \rho_\delta \xrightarrow{\mathbb{P}} \Sigma_{\vartheta,\eta}, \quad \delta \to 0.$$

(ii) The estimator $(\widehat{\vartheta}_\delta, \widehat{\eta}_\delta)^\top$ is consistent and asymptotically normal, i.e

$$\rho_\delta^{-1}\begin{pmatrix} \widehat{\vartheta}_\delta - \vartheta \\ \widehat{\eta}_\delta - \eta \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \|K\|^2_{L^2(\mathbb{R}^d)}\Sigma^{-1}_{\vartheta,\eta}), \quad \delta \to 0.$$

The convergence rates among the different parameters are given by (D.11). As the number of observation points cannot exceed $N \asymp \delta^{-d}$ due to the disjoint support condition of Assumption D.4, not all coefficients can, in general, be consistently estimated in all dimensions, see Example D.7. In contrast to parameter estimation in convection-diffusion equations based on local measurements in Paper A, the convergence rates for speed parameters are influenced not only by the order α_1 of A_ϑ, but also by β_1, the order of the damping operator B_η. Unsurprisingly, higher-order damping results in worse convergence rates as the parameters are harder to identify due to the associated dissipation of energy within the system. On the other hand, the rates for the damping coefficients are not influenced by the order of A_ϑ, and their rates mirror the rates known from parabolic equations, cf. [16] or Paper A. Similar effects were already observed under the full observation scheme $N \asymp \delta^{-d}$ in the spectral approach, cf. [23], leading to identical convergence rates.

In addition to the joint asymptotic normality of the augmented MLE $(\widehat{\vartheta}_\delta, \widehat{\eta}_\delta)^\top$, Theorem D.5 further yields the asymptotic independence of its components, i.e. the marginal estimators for elastic and damping parameters are asymptotically independent.

If the equation is weakly damped, i.e. if $\beta_1 = 0$ and $\eta_1 < 0$, then the term $-T\eta_1$ dominates the expression $(e^{T\eta_1} - T\eta_1 - 1)/2\eta_1^2$ in the asymptotic variance within (D.12) as $T \to \infty$. The converse is true in the amplified case with $\eta_1 > 0$. If $\eta_1 = 0$, the asymptotic variance of the augmented MLE for ϑ and η_1 depends on the time horizon through T^{-2} as discussed in [39, Remark 5.8]. It mirrors the rate of convergence of the MLE in the ergodic, stable, and explosive case of the standard Ornstein–Uhlenbeck process as described in [21, Proposition 3.46].

If, on the other hand, $\beta_1 > 0$, then the asymptotic variance of the MLE is of order T^{-1} in time. In other words, any dissipation decelerates the temporal convergence rate to the rate T^{-1} associated with parabolic equations.

Remark D.6 (Parameter estimation under higher-order damping). For simplicity, we did not consider cases where the damping dominates (D.3), i.e. where $2\beta_1 > \alpha_1$. Nonetheless, studying parameter estimation in those situations is neither impossible nor does it require new approaches. It solely relies on a careful analysis of underlying terms within the asymptotic analysis of the observed Fisher information, which may potentially become complex-valued, cf. also Remark D.10. Taking this into account, similar convergence rates may be established.

Example D.7.

(a) Weakly damped (or amplified) wave equation: Consider the weakly damped (or amplified) wave equation ($\vartheta_1 > 0$, $\eta_1 \in \mathbb{R}$):

$$\mathrm{d}v(t) = (\vartheta_1 \Delta u(t) + \eta_1 v(t))\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \leq T.$$

Then, Theorem D.5 implies

$$\begin{pmatrix} N^{-1/2}\delta(\widehat{\vartheta}_\delta - \vartheta_1) \\ N^{-1/2}(\widehat{\eta}_\delta - \eta_1) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\vartheta_1 \|K\|^2_{L^2(\mathbb{R}^d)}}{C(\eta_1,T)\|(-\Delta_0)^{1/2}K\|^2_{L^2(\mathbb{R}^d)}} & 0 \\ 0 & \frac{1}{C(\eta_1,T)} \end{pmatrix} \right).$$

Thus, the augmented MLE attains the convergence rate known from the spectral approach, see [23, 24] if it is provided with the maximal number of spatial observations $N \asymp \delta^{-d}$. Interestingly, the limiting variance of $\widehat{\eta}_\delta$ is independent of the kernel function K similar to the augmented MLE for the first order transport coefficient in Paper A.

(b) Clamped plate equation: Consider the clamped plate equation with

1) Weak damping ($\vartheta_1 > 0$, $\eta_1 \in \mathbb{R}$):

$$\mathrm{d}v(t) = (-\vartheta_1 \Delta^2 u(t) + \eta_1 v(t))\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \leq T. \tag{D.13}$$

2) Structural damping ($\vartheta_1 > 0$, $\eta_1 > 0$):

$$\mathrm{d}v(t) = (-\vartheta_1 \Delta^2 u(t) + \eta_1 \Delta v(t))\,\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \leq T. \tag{D.14}$$

A realisation of the solution can be seen in Figure D.1. Depending on the type of damping, the convergence rate for both ϑ_1 and η_1 changes. In the case of the weakly damped plate equation (D.13), the CLT yields

$$\begin{pmatrix} N^{1/2}\delta^{-2}(\widehat{\vartheta}_\delta - \vartheta_1) \\ N^{1/2}(\widehat{\eta}_\delta - \eta_1) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} \frac{\vartheta_1 \|K\|^2_{L^2(\mathbb{R}^d)}}{C(\eta_1,T)\|\Delta_0 K\|^2_{L^2(\mathbb{R}^d)}} & 0 \\ 0 & \frac{1}{C(\eta_1,T)} \end{pmatrix}\right),$$

while

$$\begin{pmatrix} N^{1/2}\delta^{-1}(\widehat{\vartheta}_\delta - \vartheta_1) \\ N^{1/2}\delta^{-1}(\widehat{\eta}_\delta - \eta_1) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} \frac{2\vartheta_1\eta_1 \|K\|^2_{L^2(\mathbb{R}^d)}}{T\|(-\Delta_0)^{1/2} K\|^2_{L^2(\mathbb{R}^d)}} & 0 \\ 0 & \frac{2\eta_1\|K\|^2_{L^2(\mathbb{R}^d)}}{T\|(-\Delta_0)^{1/2} K\|^2_{L^2(\mathbb{R}^d)}} \end{pmatrix}\right).$$

holds under the structural damping given in (D.14). The asymptotic variances between $\widehat{\vartheta}_\delta$ and $\widehat{\eta}_\delta$ coincide in the cases $\vartheta_1 = \|\Delta_0 K\|^2_{L^2(\mathbb{R}^d)}\|K\|^{-2}_{L^2(\mathbb{R}^d)}$ or $\vartheta_1 = 1$, respectively. The consistency and the varying convergence rates of the estimators are visualised in Figure D.2. Based on the finite difference scheme within the semi-implicit Euler–Maruyama method [25, Chapter 10] (with 10000000×2000 time-space grid points), we computed the root mean squared error (RMSE) for decreasing resolution level δ from 100 Monte Carlo runs, $N \asymp \delta^{-1}$ measurement locations and the kernel function $K(x) = \exp(-5/(1-x^2))\mathbf{1}(|x| < 1)$. In the weakly damped case, it can be seen that the estimator for the elastic coefficient ϑ_1 achieves a much quicker convergence rate than the estimator of the damping coefficient η_1. On the other hand, their rates are equal under structural damping. The asymptotic variances are attained in both cases.

(c) General hyperbolic equation: Consider the hyperbolic equation

$$\mathrm{d}v(t) = \left(\sum_{i=1}^{p} \vartheta_i(-\Delta)^{\alpha_i}u(t) + \sum_{j=1}^{q} \eta_j(-\Delta)^{\beta_j}v(t)\right)\mathrm{d}t + \mathrm{d}W(t), \quad 0 < t \le T,$$

with $p+q$ unknown parameters. Then, the convergence rates for ϑ_i and η_j, respectively, are given by

$$\begin{cases} N^{-1/2}\delta^{2\alpha_i-\alpha_1-\beta_1}, & 1 \le i \le p, \\ N^{-1/2}\delta^{2\beta_j-\beta_1}, & 1 \le j \le q. \end{cases}$$

Given the maximal number of local measurements $N \asymp \delta^{-d}$, these rates translate to

$$\begin{cases} \delta^{d/2+2\alpha_i-\alpha_1-\beta_1}, & 1 \le i \le p, \\ \delta^{d/2+2\beta_j-\beta_1}, & 1 \le j \le q. \end{cases} \tag{D.15}$$

Thus, our method provides a consistent estimator for a parameter ϑ_i or η_j, respectively, if and only if the conditions

$$\alpha_i > (\alpha_1 + \beta_1 - d/2)/2, \quad i \le p, \tag{D.16}$$

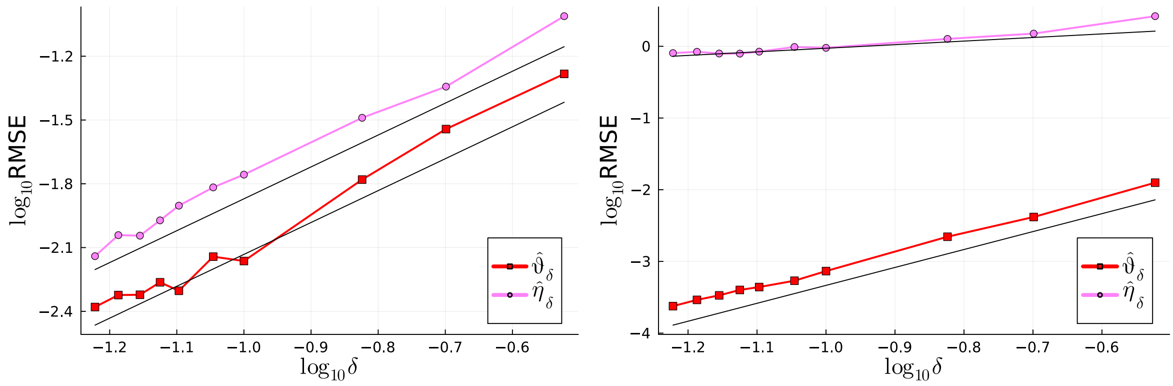


Figure D.2: log-log plot of the RMSE for $\delta \to 0$ and a maximal number of measurement locations in $d = 1$ compared with the theoretical rate in black; (left) structurally damped (D.14) with $\eta_1 = 0.3$, $\vartheta_1 = 0.3$; (right) weak damping (D.13) with $\eta_1 = -0.3$, $\vartheta_1 = 0.3$.

$$\beta_j > (\beta_1 - d/2)/2, \quad j \leq q, \tag{D.17}$$

hold. Similar results were found in the spectral regime, cf. [23, Theorem 1.1 and Theorem 4.1]. Interestingly, the authors verified a slightly stronger consistency condition, resulting in a logarithmic rate under equality in (D.16) and (D.17). Otherwise, they also obtain the rates in (D.15). We believe that the logarithmic rates in the boundary cases are also valid in the local measurement approach given that less restrictive assumptions on the kernel K are imposed, similar to Proposition A.12 in Paper A in a related parabolic problem.

D.4 Proofs

For $\delta > 0$ and $x \in \Lambda$ denote by $\Delta_{\delta,x}$ the Laplace operator with Dirichlet boundary conditions on $\Lambda_{\delta,x}$ and define the following differential operators with domain $\dot{H}^{2\alpha_1}(\Lambda)$ and $\dot{H}^{2\alpha_1}(\Lambda_{\delta,x})$, respectively:

$$L_{\vartheta,\eta} z := (-A_\vartheta - B_\eta^2/4)z = -\sum_{i=1}^{p} \vartheta_i(-\Delta)^{\alpha_i} z - \frac{1}{4}\sum_{k,l=1}^{q} \eta_k\eta_l(-\Delta)^{\beta_k+\beta_l} z,$$

$$L_{\vartheta,\eta,\delta,x} z := -\sum_{i=1}^{p} \delta^{2\alpha_1-2\alpha_i}\vartheta_i(-\Delta_{\delta,x})^{\alpha_i} z - \frac{1}{4}\sum_{k,l=1}^{q} \delta^{2\alpha_1-2\beta_k-2\beta_l}\eta_k\eta_l(-\Delta_{\delta,x})^{\beta_k+\beta_l} z.$$

Introduce further the rescaled versions of B_η and A_ϑ, defined through

$$B_{\eta,\delta,x} := \sum_{j=1}^{q} \delta^{2\beta_1-2\beta_j}\eta_j(-\Delta_{\delta,x})^{\beta_j}, \quad D(B_{\eta,\delta,x}) = D((-\Delta_{\delta,x})^{\beta_1}) = \dot{H}^{2\beta_1}(\Lambda_{\delta,x}),$$

$$A_{\vartheta,\delta,x} := \sum_{i=1}^{p} \delta^{2\alpha_1-2\alpha_i}\vartheta_i(-\Delta_{\delta,x})^{\alpha_i}, \quad D(A_{\vartheta,\delta,x}) = D((-\Delta_{\delta,x})^{\alpha_1}) = \dot{H}^{2\alpha_1}(\Lambda_{\delta,x}),$$

and the limiting objects

$$\bar{L}_{\vartheta,\eta} z := \begin{cases} -\vartheta_1(-\Delta_0)^{\alpha_1} z, & \alpha_1 > 2\beta_1, \\ -(\vartheta_1 + \frac{\eta_1^2}{4})(-\Delta_0)^{\alpha_1} z, & \alpha_1 = 2\beta_1, \end{cases} \qquad D(\bar{L}_{\vartheta,\eta}) = D((-\Delta_0)^{\alpha_1}) = \dot{H}^{2\alpha_1}(\mathbb{R}^d),$$

$$\bar{B}_\eta := \eta_1(-\Delta_0)^{\beta_1}, \quad D(\bar{B}_\eta) = D((-\Delta_0)^{\beta_1}) = \dot{H}^{2\beta_1}(\mathbb{R}^d),$$
$$\bar{A}_\vartheta := \vartheta_1(-\Delta_0)^{\alpha_1}, \quad D(\bar{A}_\vartheta) = D((-\Delta_0)^{\alpha_1}) = \dot{H}^{2\alpha_1}(\mathbb{R}^d).$$

We will frequently use that $L_{\vartheta,\eta}$ is an (unbounded) normal operator with spectrum $\sigma(L_{\vartheta,\eta})$ and the resolution of identity E (cf. [33, Chapter 13]). By the functional calculus for normal unbounded operators, we can define the operator $f(L_{\vartheta,\eta}) := \int_{\sigma(L_{\vartheta,\eta})} f(\lambda)\,\mathrm{d}E(\lambda)$ on the domain

$$\mathcal{D}_f := \mathcal{D}(f(L_{\vartheta,\eta})) = \left\{ z \in L^2(\Lambda) : \int_{\sigma(L_{\vartheta,\eta})} |f(\lambda)|^2 \,\mathrm{d}E_{z,z}(\lambda) < \infty \right\},$$

for any measurable function $f : \mathbb{C} \to \mathbb{C}$. Analogous statements also apply to A_ϑ, B_η and the rescaled differential operators.

Lemma D.8 (Rescaling of operators). *Let $\delta > 0, x \in \Lambda$ and $f : \mathbb{C} \to \mathbb{C} \cup \{\pm\infty\}$ be measurable. If $z_{\delta,x} \in \mathcal{D}_f$, then*

$$f(L_{\vartheta,\eta})z_{\delta,x} = (f(\delta^{-2\alpha_1}L_{\vartheta,\eta,\delta,x})z)_{\delta,x},$$
$$f(B_\eta)z_{\delta,x} = (f(\delta^{-2\beta_1}B_{\eta,\delta,x})z)_{\delta,x},$$
$$f(A_\vartheta)z_{\delta,x} = (f(\delta^{-2\alpha_1}A_{\vartheta,\delta,x})z)_{\delta,x}.$$

Proof of Lemma D.8. Suppose $z \in \dot{H}^{2\alpha_1}(\Lambda_{\delta,x})$ such that $z_{\delta,x} \in \dot{H}^{2\alpha_1}(\Lambda) \subset \mathcal{D}_f$. Then, the claim follows immediately for $f(x) = x$ by differentiating $z_{\delta,x}$ and from the definition of $L_{\vartheta,\eta,\delta,x}$, see also [4, Lemma 16]. Using (i) and (iii) of [34, Proposition 5.15], the result can be extended to measurable $f : \mathbb{C} \to \mathbb{C} \cup \{\pm\infty\}$ by first passing to the associated resolution of the identities of $L_{\vartheta,\eta}, B_\eta, A_\vartheta$ and $L_{\vartheta,\eta,\delta,x}, B_{\eta,\delta,x}, A_{\vartheta,\delta,x}$ respectively, and interpreting the localisation as a bounded linear operator from $L^2(\Lambda_{\delta,x})$ to $L^2(\Lambda)$. ∎

Throughout the remainder of the paper, we will assume that $L_{\vartheta,\eta}$, and thus also $L_{\vartheta,\eta,\delta,x}$, is a positive operator. A sufficient condition for this is given in the next lemma.

Lemma D.9 (Sufficient condition for positivity). *Let $(e_k)_{k\in\mathbb{N}}$ form an orthonormal basis of $(-\Delta)$ with eigenvalues $\lambda_k \geq c(\Lambda)$ for some constant $c(\Lambda) > 0$. Then, $L_{\vartheta,\eta}$ is a positive operator if one of the following conditions is satisfied:*

(i) Assumption D.2 holds, $c(\Lambda) \geq 1$ and

$$|\vartheta_1| > \sum_{i=2}^{p} |\vartheta_i| + \frac{1}{4}\sum_{k,l=1}^{q} |\eta_k\eta_l|;$$

(ii) Assumption D.2 holds, $c(\Lambda) < 1$ and

$$|\vartheta_1| > \sum_{i=2}^{p} |\vartheta_i| c(\Lambda)^{\alpha_i-\alpha_1} + \frac{1}{4}\sum_{k,l=1}^{q} |\eta_k\eta_l| c(\Lambda)^{\beta_k+\beta_l-\alpha_1}.$$

Proof of Lemma D.9. Assumption D.2 states in particular that $\vartheta_1 < 0$ and, additionally, $\vartheta_1 + \eta_1^2/4 < 0$ in case $\alpha_1 = 2\beta_1$. It is now enough to show that all eigenvalues of $L_{\vartheta,\eta}$ are positive, which holds if for all $x \geq c(\Lambda)$:

$$|\vartheta_1|x^{\alpha_1} - \sum_{i=2}^{p} |\vartheta_i|x^{\alpha_i} - \frac{1}{4}\sum_{k,l=1}^{q} |\eta_k\eta_l|x^{\beta_k+\beta_l} > 0. \tag{D.18}$$

(i) If $c(\Lambda) \geq 1$, then both x^{α_i} and $x^{\beta_k+\beta_l}$ are bounded by x^{α_1} for any $i \leq p$ and $k, l \leq q$, thus (D.18) is satisfied.

(ii) If $c(\Lambda) < 1$, then $x^{\alpha_i-\alpha_1} \leq c(\Lambda)^{\alpha_i-\alpha_1}$ and $x^{\beta_k+\beta_l-\alpha_1} \leq c(\Lambda)^{\beta_k+\beta_l-\alpha_1}$. ■

Remark D.10. If $L_{\vartheta,\eta}$ is not a positive operator and has non-positive eigenvalues, any choice of the operator root is a complex-valued operator. Consequently, the associated family of M, N-functions in the following subsection is again complex-valued. Thus, inner products hereinafter are associated with complex Hilbert spaces. However, this does not influence the asymptotic results of Section D.4.2 due to the convergence $L_{\vartheta,\eta,\delta,x} \to \bar{L}_{\vartheta,\eta}$ to a positive limiting operator $\bar{L}_{\vartheta,\eta}$.

D.4.1 *Properties of generalised cosine and sine operator functions*

LEMMA D.11 (Representations M, N-functions). *The operators A_ϑ and B_η defined in* (D.4) *generate a family of M, N-functions $(M(t), N(t), t \geq 0)$ given by*

$$M(t) := \mathbf{m}_t(L_{\vartheta,\eta}, B_\eta) := e^{B_\eta t/2}\Big(\cos(L_{\vartheta,\eta}^{1/2}t) - \frac{B_\eta}{2}\sin(L_{\vartheta,\eta}^{1/2}t)L_{\vartheta,\eta}^{-1/2}\Big), \quad L^2(\Lambda) \subset \mathcal{D}(N(t)), \tag{D.19}$$

$$N(t) := \mathbf{n}_t(L_{\vartheta,\eta}, B_\eta) := e^{B_\eta t/2}\sin(L_{\vartheta,\eta}^{1/2}t)L_{\vartheta,\eta}^{-1/2}, \quad L^2(\Lambda) \subset \mathcal{D}(N(t)). \tag{D.20}$$

Proof of Lemma D.11. Note that by [34, Theorem 5.9] all of the appearing operators $e^{B_\eta t/2}$, $\cos(tL_{\vartheta,\eta}^{1/2})$, $\sin(tL_{\vartheta,\eta}^{1/2})$, B_η and $L_{\vartheta,\eta}^{-1/2}$ in (D.19) and (D.20) are well-defined and even commute on the smallest occurring domain as they are all based on the same underlying Laplace operator. By direct computation, one can now verify that the conditions **(M1)-(M4)** in [28, Definition 1.7.2] are satisfied by $M(t)$ and $N(t)$ from (D.19) and (D.20) using the functional calculus. ■

LEMMA D.12 (Self-adjointness of M, N-functions). *Assume that $L_{\vartheta,\eta}$ is a positive operator. Then, the M, N-functions defined through* (D.19) *and* (D.20) *are self-adjoint.*

Proof of Lemma D.12. The unique positive self-adjoint operator root of the positive self-adjoint operator $L_{\vartheta,\eta}$ is well-defined and exists by [34, Proposition 5.13]. Thus, in view of Lemma D.11, the M, N-functions can each be interpreted as the applications of a real-valued function to the underlying Laplace operator on a bounded spatial domain. In particular, by [34, Theorem 5.9] the M, N-functions are self-adjoint. ■

As we are interested in the effect of M, N-functions applied to localised functions, we further define the rescaled M, N-functions:

$$M_{\delta,x}(t) := \mathbf{m}_t(\delta^{-2\alpha_1}L_{\vartheta,\eta,\delta,x}, \delta^{-2\beta_1}B_{\eta,\delta,x}), \tag{D.21}$$

$$N_{\delta,x}(t) := \mathbf{n}_t(\delta^{-2\alpha_1}L_{\vartheta,\eta,\delta,x}, \delta^{-2\beta_1}B_{\eta,\delta,x}). \tag{D.22}$$

An application of Lemma D.8 yields the scaling properties of the M, N-functions in analogy to [5, Lemma 3.1] and [39, Lemma 3.1]:

$$M(t)z_{\delta,x} = (M_{\delta,x}(t)z)_{\delta,x}, \quad N(t)z_{\delta,x} = (N_{\delta,x}(t)z)_{\delta,x}, \quad z \in L^2(\Lambda_{\delta,x}).$$

Lemma D.13 (Semigroup upper bounds). *Let $0 \leq t \leq T\delta^{-2\beta_1}$, $\gamma \geq 0$ and $z \in H^{2\lceil\gamma\rceil}(\mathbb{R}^d)$ with compact support in $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$ and such that there exists a compactly supported function $\widetilde{z} \in H^{2\lceil\gamma\rceil+\lceil\alpha_1\rceil}(\mathbb{R}^d)$ with $z = \Delta_0^{\lceil\alpha_1\rceil}\widetilde{z}$. Then, if $\beta_1 > 0$, we have*

$$\sup_{x\in\mathcal{J}} \|e^{tB_{\eta,\delta,x}} L_{\vartheta,\eta,\delta,x}^{-1/2}(-\Delta_{\delta,x})^{\gamma} z\|_{L^2(\Lambda_{\delta,x})} \lesssim 1 \wedge t^{-(\gamma+\lceil\alpha_1\rceil-\alpha_1/2)/\beta_1}; \tag{D.23}$$

$$\sup_{x\in\mathcal{J}} \|e^{tB_{\eta,\delta,x}} (-\Delta_{\delta,x})^{\gamma} z\|_{L^2(\Lambda_{\delta,x})} \lesssim 1 \wedge t^{-(\gamma+\lceil\alpha_1\rceil)/\beta_1}. \tag{D.24}$$

Moreover, in case $\beta_1 = 0$, the left-hand sides in (D.23) and (D.24) are bounded by a constant independent of δ and t.

Proof of Lemma D.13. The key idea of the proof is that all involved operators emerge as an application of the functional calculus applied to the same Laplace operator. In particular, they are simultaneously diagonalisable through the same eigenfunctions. Note that in contrast to the eigenfunctions, the associated eigenvalues do not depend themselves on the shift, i.e. $x \in \Lambda$, within the rescaling of the Laplace operator.

Let $\beta_1 > 0$. We only prove (D.24), since the argument for (D.23) is similar, using additionally that $L_{\vartheta,\eta,\delta,x}$ commutes with $(-\Delta_{\delta,x})^{\gamma}$, $\mathrm{ord}(L_{\vartheta,\eta,\delta,x}) = 2\alpha_1$ and a bound of $L_{\vartheta,\eta,\delta,x}$ in terms of its leading term $(-\Delta_{\delta,x})^{\alpha_1}$. Let $(e_k)_{k\in\mathbb{N}}$ form an orthonormal basis of $(-\Delta)$ in $L^2(\Lambda)$ with eigenvalues $\lambda_k > 0$. Then, there exists a constant $c(\Lambda)$ such that $\lambda_k \geq c(\Lambda)$ for all $k \geq 1$, see [35, Proposition 5.2 and Corollary 5.3]. We consider the most involved case, that is, $\eta_1 < 0$ and $\eta_2, \dots, \eta_q > 0$. Consequently, B_η will, in general, not be a negative operator, but there exists $y_0 > 0$ such that for all $y \geq y_0$, we have

$$\eta_1 y^{\beta_1} + \sum_{j=2}^{q} \eta_j y^{\beta_j} \leq \frac{\eta_1 y^{\beta_1}}{2},$$

and all but finitely many eigenvalues of B_η will be negative due to Assumption D.2 (ii). Consider the polynomial

$$P_\eta(y) := \frac{\eta_1}{2} y^{\beta_1} + \sum_{j=2}^{q} \eta_j y^{\beta_j}$$

and define $C_1 := \max_{y\in[c(\Lambda),y_0]} |P_\eta(y)|$. Then

$$\eta_1 y^{\beta_1} + \sum_{j=2}^{q} \eta_j y^{\beta_j} - C_1 \leq \frac{\eta_1 y^{\beta_1}}{2}$$

holds for all $y \geq c(\Lambda)$, and all eigenvalues of the operator $B_\eta - C_1$ are negative and upper bounded by $\eta_1 c(\Lambda)/2$. Analogously, $(e_{k,\delta,x})_{k\in\mathbb{N}}$ forms an orthonormal basis of $(-\Delta_{\delta,x})$ in $L^2(\Lambda_{\delta,x})$ with eigenvalues $\lambda_{k,\delta,x} = \delta^2\lambda_k \geq \delta^2 c(\Lambda)$. Similar calculations imply that

$$\eta_1 y^{\beta_1} + \sum_{j=2}^{q} \eta_j \delta^{2(\beta_1-\beta_j)} y^{\beta_j} - \delta^{2\beta_1} C_1 \leq \frac{\eta_1 y^{\beta_1}}{2}, \quad y \geq \delta^2 c(\Lambda). \tag{D.25}$$

Thus, all eigenvalues of the operator difference $B_{\eta,\delta,x} - \delta^{2\beta_1}C_1$ are negative and the difference is bounded by $\eta_1(-\Delta_{\delta,x})^{\beta_1}/2$ in the sense that

$$\|e^{t(B_{\eta,\delta,x}-\delta^{2\beta_1}C_1)}w\|_{L^2(\Lambda_{\delta,x})} \leq \|e^{t\eta_1(-\Delta_{\delta,x})^{\beta_1}/2}w\|_{L^2(\Lambda_{\delta,x})}, \quad w \in L^2(\Lambda_{\delta,x}),$$

independent of $x \in \mathcal{J}$. Note further that $z = \Delta_0^{\lceil\alpha_1\rceil}\tilde{z} = \Delta_{\delta,x}^{\lceil\alpha_1\rceil}\tilde{z}$ since z is compactly supported in $\bigcap_{x\in\mathcal{J}}\Lambda_{\delta,x}$. With that, we have all the ingredients to prove (D.24). For $0 \leq t \leq T\delta^{-2\beta_1}$, we obtain

$$\begin{aligned}\sup_{x\in\mathcal{J}}\|e^{tB_{\eta,\delta,x}}(-\Delta_{\delta,x})^{\gamma}z\|_{L^2(\Lambda_{\delta,x})} &= \sup_{x\in\mathcal{J}}\|e^{t\delta^{2\beta_1}C_1}e^{t(B_{\eta,\delta,x}-\delta^{2\beta_1}C_1)}(-\Delta_{\delta,x})^{\gamma}z\|_{L^2(\Lambda_{\delta,x})}\\ &\leq e^{C_1T}\sup_{x\in\mathcal{J}}\|e^{t(B_{\eta,\delta,x}-\delta^{2\beta_1}C_1)}(-\Delta_{\delta,x})^{\gamma}(-\Delta_{\delta,x})^{\lceil\alpha_1\rceil}\tilde{z}\|_{L^2(\Lambda_{\delta,x})}\\ &\leq e^{C_1T}\sup_{x\in\mathcal{J}}\|e^{t\eta_1(-\Delta_{\delta,x})^{\beta_1}/2}(-\Delta_{\delta,x})^{\gamma+\lceil\alpha_1\rceil}\tilde{z}\|_{L^2(\Lambda_{\delta,x})}\\ &\lesssim (1 \wedge t^{-(\gamma+\lceil\alpha_1\rceil)/\beta_1})\sup_{x\in\mathcal{J}}\left(\|\tilde{z}\|_{L^2(\Lambda_{\delta,x})} + \|(-\Delta_{\delta,x})^{\gamma}z\|_{L^2(\Lambda_{\delta,x})}\right),\end{aligned}$$

where the last line follows from the fact that $(-\Delta_{\delta,x})^{\beta_1}\eta_1/2$ generates a contraction semigroup and the smoothing property of semigroups. As $\Delta_{\delta,x}^n z = \Delta_0^n z$ holds for any $x \in \mathcal{J}$ and $1 \leq n \leq \lceil\gamma\rceil$, an application of the functional calculus yields

$$\begin{aligned}\sup_{x\in\mathcal{J}}\|(-\Delta_{\delta,x})^{\gamma}z\|^2_{L^2(\Lambda_{\delta,x})} &\leq \sup_{x\in\mathcal{J}}\|(-\Delta_{\delta,x})^{\lceil\gamma\rceil}z\|^2_{L^2(\mathbb{R}^d)} + \sup_{x\in\mathcal{J}}\|(-\Delta_{\delta,x})^{\lfloor\gamma\rfloor}z\|^2_{L^2(\mathbb{R}^d)}\\ &= \|(-\Delta_0)^{\lceil\gamma\rceil}z\|^2_{L^2(\mathbb{R}^d)} + \|(-\Delta_0)^{\lfloor\gamma\rfloor}z\|^2_{L^2(\mathbb{R}^d)}\\ &< \infty,\end{aligned}$$

proving the assertion. The claim for $\beta_1 = 0$ follows directly by bounding

$$|e^{tB_{\eta,\delta,x}}| = |e^{t\eta_1}| \leq C, \quad 0 \leq t \leq T,$$

for some constant C only depending on η_1 and T. ∎

The theory of cosine operator functions was developed by Sova [36] and led to general deterministic solution theory for undamped second-order abstract Cauchy problems. By substituting the time derivative as its own variable, it is possible to rewrite a second-order abstract Cauchy problem as a first-order abstract Cauchy problem in two components. The associated strongly continuous semigroup then lives on a product of Hilbert spaces called the phase-space; see [6, Chapter 3.14], [29] and [28, Chapter 0.3]. The same remains true under suitable assumptions on the elastic and dissipation operator within a damped abstract second-order Cauchy problem [28, Chapter 1.7].

LEMMA D.14 (Semigroup on the phase-space). *The operator $\mathcal{A}_{\vartheta,\eta}$, defined through*

$$\mathcal{A}_{\vartheta,\eta} := \begin{pmatrix} 0 & I \\ A_\vartheta & B_\eta \end{pmatrix}, \quad D(\mathcal{A}_{\vartheta,\eta}) = \dot{H}^{2\alpha_1}(\Lambda) \times \dot{H}^{2\beta_1}(\Lambda),$$

generates a C_0-semigroup $(J_{\vartheta,\eta}(t))_{t\geq0}$ on the phase-space $\dot{H}^{\alpha_1}(\Lambda) \times L^2(\Lambda)$ given by

$$J_{\vartheta,\eta}(t) := \begin{pmatrix} M(t) & N(t) \\ A_\vartheta N(t) & M(t) + B_\eta N(t) \end{pmatrix} = \begin{pmatrix} M(t) & N(t) \\ M'(t) & N'(t) \end{pmatrix}, \quad t \geq 0, \tag{D.26}$$

with $M(t)$ and $N(t)$ given in (D.19) and (D.20), respectively.

Proof of Lemma D.14. It is well-known that in the special case where both A_ϑ and B_η are strictly negative operators $\mathcal{A}_{\vartheta,\eta}$ generates a C_0-semigroup, which is even analytic if and only if $\alpha_1/2 \le \beta_1 \le \alpha_1$, cf. [10]. On the other hand, $(J_{\vartheta,\eta}(t))_{t\ge 0}$ given by (D.26) is indeed a semigroup generated by $\mathcal{A}_{\vartheta,\eta}$ which follows by direct verification of the differential properties of M, N-functions in [28, p. 131] using the functional calculus. ∎

The coupled second-order system (D.3) can also be written as a first-order system

$$\mathrm{d}X(t) = \mathcal{A}_{\vartheta,\eta} X(t)\,\mathrm{d}t + \begin{pmatrix} 0 \\ I \end{pmatrix} \mathrm{d}W(t), \quad 0 < t \le T,$$

for $X(t) = (u(t), v(t))^\top$ and the matrix-valued differential operator $\mathcal{A}_{\vartheta,\eta}$ generating the strongly continuous semigroup $(J_{\vartheta,\eta}(t))_{t\ge 0}$ constituted by the M, N-functions defined in Lemma D.11. The M, N-functions correspond to the cosine and sine functions in the undamped wave equation, see [6, Chapter 3]. Naturally, they appear in the solution to the stochastic partial differential equation (D.3):

$$\begin{aligned} \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} &= J_{\vartheta,\eta}(t) \begin{pmatrix} u(0) \\ v(0) \end{pmatrix} + \int_0^t J_{\vartheta,\eta}(t-s) \begin{pmatrix} 0 \\ I \end{pmatrix} \mathrm{d}W(s) \\ &= \begin{pmatrix} M(t)u_0 + N(t)v_0 \\ M'(t)u_0 + N'(t)v_0 \end{pmatrix} + \begin{pmatrix} \int_0^t N(t-s)\,\mathrm{d}W(s) \\ \int_0^t N'(t-s)\,\mathrm{d}W(s) \end{pmatrix}. \end{aligned} \tag{D.27}$$

D.4.2 *Asymptotic properties of local measurements*

In this section, we study the asymptotic covariance structure of the local measurements, which is crucial in showing the convergence of the observed Fisher information matrix $\mathcal{I}_\delta$.

Lemma D.15 (Covariance structure). *Assume that $(u_0, v_0)^\top = (0,0)^\top$. For any $t, s \in [0,T]$, $x \in \Lambda$, $1 \le i, j \le p$, $1 \le k, l \le q$, the covariance between local measurements is given by*

$$\begin{aligned} &\mathrm{Cov}(u^{\Delta_i}_{\delta,x}(t), u^{\Delta_j}_{\delta,x}(s)) \\ &\quad = \delta^{-2\alpha_i - 2\alpha_j} \int_0^{t\wedge s} \langle N_{\delta,x}(t-r)(-\Delta_{\delta,x})^{\alpha_i} K, N_{\delta,x}(s-r)(-\Delta_{\delta,x})^{\alpha_j} K \rangle_{L^2(\Lambda_{\delta,x})} \mathrm{d}r, \\ &\mathrm{Cov}(v^{\Delta_k}_{\delta,x}(t), v^{\Delta_l}_{\delta,x}(s)) \\ &\quad = \delta^{-2\beta_k - 2\beta_l} \int_0^{t\wedge s} \langle N'_{\delta,x}(t-r)(-\Delta_{\delta,x})^{\beta_k} K, N'_{\delta,x}(s-r)(-\Delta_{\delta,x})^{\beta_l} K \rangle_{L^2(\Lambda_{\delta,x})} \,\mathrm{d}r, \\ &\mathrm{Cov}(u^{\Delta_i}_{\delta,x}(t), v^{\Delta_k}_{\delta,x}(s)) \\ &\quad = \delta^{-2\alpha_i - 2\beta_k} \int_0^{t\wedge s} \langle N_{\delta,x}(t-r)(-\Delta_{\delta,x})^{\alpha_i} K, N'_{\delta,x}(s-r)(-\Delta_{\delta,x})^{\beta_k} K \rangle_{L^2(\Lambda_{\delta,x})} \mathrm{d}r. \end{aligned}$$

Proof of Lemma D.15. Using (D.27) and [12, Proposition 4.28], we observe

$$\begin{aligned} \mathrm{Cov}(u^{\Delta_i}_{\delta,x}(t), u^{\Delta_j}_{\delta,x}(s)) &= \int_0^{t\wedge s} \langle N^*(t-r)(-\Delta)^{\alpha_i} K_{\delta,x}, N^*(s-r)(-\Delta)^{\alpha_j} K_{\delta,x} \rangle \mathrm{d}r, \\ \mathrm{Cov}(v^{\Delta_k}_{\delta,x}(t), v^{\Delta_l}_{\delta,x}(s)) &= \int_0^{t\wedge s} \langle (N'(t-r))^*(-\Delta)^{\beta_k} K_{\delta,x}, (N'(s-r))^*(-\Delta)^{\beta_l} K_{\delta,x} \rangle \mathrm{d}r, \end{aligned}$$

$$\mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), v_{\delta,x}^{\Delta_k}(s)) = \int_0^{t\wedge s} \langle N^*(t-r)(-\Delta)^{\alpha_i} K_{\delta,x}, (N'(s-r))^*(-\Delta)^{\beta_k} K_{\delta,x}\rangle \mathrm{d}r.$$

We can rewrite the last equations through the functional calculus by using Lemma D.8, the representations (D.21) and (D.22), self-adjointness by Lemma D.12 as well as [34, Theorem 5.9]:

$$\begin{aligned}
&\mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), u_{\delta,x}^{\Delta_j}(s)) \\
&\quad = \delta^{-2\alpha_i-2\alpha_j} \int_0^{t\wedge s} \langle N_{\delta,x}(t-r)(-\Delta_{\delta,x})^{\alpha_i} K, N_{\delta,x}(s-r)(-\Delta_{\delta,x})^{\alpha_j} K\rangle_{L^2(\Lambda_{\delta,x})} \mathrm{d}r, \\
&\mathrm{Cov}(v_{\delta,x}^{\Delta_k}(t), v_{\delta,x}^{\Delta_l}(s)) \\
&\quad = \delta^{-2\beta_k-2\beta_l} \int_0^{t\wedge s} \langle N'_{\delta,x}(t-r)(-\Delta_{\delta,x})^{\beta_k} K, N'_{\delta,x}(s-r)(-\Delta_{\delta,x})^{\beta_l} K\rangle_{L^2(\Lambda_{\delta,x})} \,\mathrm{d}r, \\
&\mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), v_{\delta,x}^{\Delta_k}(s)) \\
&\quad = \delta^{-2\alpha_i-2\beta_k} \int_0^{t\wedge s} \langle N_{\delta,x}(t-r)(-\Delta_{\delta,x})^{\alpha_i} K, N'_{\delta,x}(s-r)(-\Delta_{\delta,x})^{\beta_k} K\rangle_{L^2(\Lambda_{\delta,x})} \mathrm{d}r.
\end{aligned}$$

■

LEMMA D.16 (Scaling limits for M, N-functions). *Let $\delta > 0$. Let $z_1, z_2 \in L^2(\mathbb{R}^d)$ with compact support in $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$ such that there exist compactly supported functions $\bar{z}_1, \bar{z}_2 \in H^{2\lceil\alpha_1\rceil}(\mathbb{R}^d)$ with $z_i = \Delta_0^{\lceil\alpha_1\rceil}\bar{z}_i$, $i = 1, 2$. As $\delta \to 0$, we obtain the following convergences.*

1. *Let $t \geq 0$. Let $\beta_1 = 0$, i.e. $B_\eta = \eta_1$. Then, uniformly in $x \in \mathcal{J}$,*

$$\begin{aligned}
\delta^{-2\alpha_1}\langle N_{\delta,x}(t)z_1, N_{\delta,x}(t)z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to -e^{\eta_1 t}\frac{1}{2}\langle \bar{A}_\vartheta^{-1} z_1, z_2\rangle_{L^2(\mathbb{R}^d)}, \\
\langle M_{\delta,x}(t)z_1, M_{\delta,x}(t)z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to e^{\eta_1 t}\frac{1}{2}\langle z_1, z_2\rangle_{L^2(\mathbb{R}^d)}, \\
\delta^{-\alpha_1}\langle N_{\delta,x}(t)z_1, M_{\delta,x}(t)z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to 0.
\end{aligned}$$

2. *Let $r_1 \neq r_2$. Let $\beta_1 = 0$. Then, uniformly in $x \in \mathcal{J}$,*

$$\begin{aligned}
\delta^{-2\alpha_1}\langle N_{\delta,x}(r_1)z_1, N_{\delta,x}(r_2)z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to 0, \\
\langle M_{\delta,x}(r_1)z_1, M_{\delta,x}(r_2)z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to 0, \\
\delta^{-\alpha_1}\langle N_{\delta,x}(r_1)z_1, M_{\delta,x}(r_2)z_2\rangle_{L^2(\Lambda_{\Lambda_{\delta,x}})} &\to 0.
\end{aligned}$$

3. *Let $0 < 2\beta_1 \leq \alpha_1$, $t \in (0, T]$. Then, uniformly in $x \in \mathcal{J}$,*

$$\begin{aligned}
\delta^{-2\alpha_1-2\beta_1}\langle \int_0^t N_{\delta,x}(r)^2 \,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to \frac{1}{2}\langle \bar{B}_\eta^{-1}\bar{A}_\vartheta^{-1} z_1, z_2\rangle_{L^2(\mathbb{R}^d)}, \\
\delta^{-2\beta_1}\langle \int_0^t (N'_{\delta,x}(r))^2 \,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to -\frac{1}{2}\langle \bar{B}_\eta^{-1} z_1, z_2\rangle_{L^2(\mathbb{R}^d)}, \\
\delta^{-2\beta_1-\alpha_1}\langle \int_0^t N_{\delta,x}(r)N'_{\delta,x}(r) \,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} &\to 0.
\end{aligned}$$

Proof of Lemma D.16.

1. Using (D.20) we have

$$N_{\delta,x}(t) = \delta^{\alpha_1} e^{\eta_1 t/2} \sin(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) L_{\vartheta,\eta,\delta,x}^{-1/2}, \quad t \in [0,T],$$

and thus

$$\begin{aligned}&\delta^{-2\alpha_1} \langle N_{\delta,x}(t) z_1, N_{\delta,x}(t) z_2 \rangle_{L^2(\Lambda_{\delta,x})} \\ &\quad = e^{\eta_1 t} \langle \sin(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) L_{\vartheta,\eta,\delta,x}^{-1/2} z_1, \sin(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) L_{\vartheta,\eta,\delta,x}^{-1/2} z_2 \rangle_{L^2(\Lambda_{\delta,x})}.\end{aligned}$$

Since $S_{\vartheta,\eta,\delta,x}(t) := \sin(tL_{\vartheta,\eta,\delta,x}) L_{\vartheta,\eta,\delta,x}^{-1/2}$ is the operator sine function, which is generated by $-L_{\vartheta,\eta,\delta,x}$, and $L_{\vartheta,\eta,\delta,x} \to \bar{L}_{\vartheta,\eta}$ as $\delta \to 0$, the desired convergence follows by repeating the steps of [39, Proposition A.10 (i)] regarding asymptotic equipartition of energy. Note that the assumptions in [39, Proposition A.10] can be relaxed, as we are not considering the non-parametric case. The employed strong resolvent convergence and the involved convergence of the spectral measures then follow from [38, Theorem 1 and 2] by choosing the core $C_c^\infty(\mathbb{R}^d)$ as described in [39, Lemma A.6]. The convergences for the functional calculus associated with the respective spectral measures are then immediate, see [39, Proposition 3.3]. Similarly, we observe

$$\begin{aligned}&\langle M_{\delta,x}(t) z_1, M_{\delta,x}(t) z_2 \rangle_{L^2(\Lambda_{\delta,x})} \\ &= e^{\eta_1 t} \langle (\cos(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) - \delta^{\alpha_1} \eta_1/2 \sin(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) L_{\vartheta,\eta,\delta,x}^{-1/2}) z_1, \\ &\qquad (\cos(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) - \delta^{\alpha_1} \eta_1/2 \sin(t\delta^{-\alpha_1} L_{\vartheta,\eta,\delta,x}^{1/2}) L_{\vartheta,\eta,\delta,x}^{-1/2}) z_2 \rangle_{L^2(\Lambda_{\delta,x})}.\end{aligned}$$

Likewise, $C_{\vartheta,\eta,\delta,x}(t) := \cos(tL_{\vartheta,\eta,\delta,x}^{1/2})$ is the cosine operator function associated with the operator $-L_{\vartheta,\eta,\delta,x}$. [6, Example 3.14.15] yields the representation

$$C_{\vartheta,\eta,\delta,x}(t\delta^{-\alpha_1}) = \frac{1}{2}\big(U_{\vartheta,\eta,\delta,x}(t\delta^{-\alpha_1}) + U_{\vartheta,\eta,\delta,x}(-t\delta^{-\alpha_1})\big)$$

with the unitary group $(U_{\vartheta,\delta,x}(t))_{t\in\mathbb{R}}$ generated by $i(L_{\vartheta,\eta,\delta,x}^{1/2})$ on $L^2(\Lambda_{\delta,x})$ and the steps of [39, Proposition A.10 (i)] can be repeated to verify convergence. Analogous calculations show

$$\delta^{-\alpha_1} \langle N_{\delta,x}(t) z_1, M_{\delta,x}(t) z_2 \rangle_{L^2(\Lambda_{\delta,x})} \longrightarrow 0.$$

All the above convergences hold uniformly in $x \in \mathcal{J}$ since in the parametric case the convergences in [39, Proposition 4.5 and Lemma A.6] are uniform in $x \in \mathcal{J}$ when applied to functions with support in $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$. In fact, restricted to $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$, the Laplacian $\Delta_{\delta,x}$ is identical to $\Delta_{\delta,y}$ for $y \in \mathcal{J}$ and the associated spectral measures become independent of the spatial point $y \in \mathcal{J}$, when applied to functions with support in $\bigcap_{x\in\mathcal{J}} \Lambda_{\delta,x}$.

2. The convergences follow similarly to (i) by using the slow-fast orthogonality as presented in [39, Proposition A.10 (ii)].

(iii) For readability, we suppress various indices throughout the remainder of the proof. Thus, we introduce the following notation:

$$A := A_{\vartheta,\delta,x}; \quad B := B_{\eta,\delta,x}; \quad L := L_{\vartheta,\eta,\delta,x}; \quad \alpha = \delta^{\alpha_1}; \quad \beta = \delta^{\beta_1}. \tag{D.28}$$

By definition of M, N-functions, substitution and the fundamental theorem of calculus, we then obtain

$$\begin{aligned}
&\delta^{-2\alpha_1-2\beta_1}\langle\int_0^t N_{\delta,x}(r)^2\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}\\
&= \alpha^{-2}\beta^{-2}\langle\int_0^t N_{\delta,x}(r)^2\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}\\
&= \langle\int_0^{t\beta^{-2}} e^{rB}\sin^2(r\alpha^{-1}\beta^2L^{1/2})L^{-1}\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}\\
&= \langle\Big(e^{t\beta^{-2}B}\sin^2(t\alpha^{-1}L^{1/2})B^2 - 2\alpha^{-1}\beta^2e^{t\beta^{-2}B}BL^{1/2}\cos(t\alpha^{-1}L^{1/2})\sin(t\alpha^{-1}L^{1/2})\\
&\quad +2\alpha^{-2}\beta^4(e^{t\beta^{-2}B}-I)L\Big)B^{-1}L^{-1}(4\alpha^{-2}\beta^4L+B^2)^{-1}z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}.
\end{aligned} \tag{D.29}$$

We can rewrite the last display as

$$\frac{1}{2}\langle B^{-1}A^{-1}z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} \tag{D.30}$$

$$-\frac{\alpha^2\beta^{-4}}{4}\langle e^{t\beta^{-2}B}\sin^2(t\alpha^{-1}L^{1/2})L^{-1}BA^{-1}z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} \tag{D.31}$$

$$+\frac{\alpha\beta^{-2}}{2}\langle e^{t\beta^{-2}B}\cos(t\alpha^{-1}L^{1/2})\sin(t\alpha^{-1}L^{1/2})L^{-1/2}A^{-1}z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} \tag{D.32}$$

$$-\frac{1}{2}\langle e^{t\beta^{-2}B}B^{-1}A^{-1}z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}. \tag{D.33}$$

Since $A = A_{\vartheta,\delta,x}$ converges to $\bar{A}_\vartheta$, (D.30) converges to $\frac{1}{2}\langle\bar{B}_\eta^{-1}\bar{A}_\vartheta^{-1}z_1, z_2\rangle_{L^2(\mathbb{R}^d)}$, while (D.31), (D.32) and (D.33) tend to zero by the Cauchy–Schwarz inequality and Lemma D.13.

As we will integrate (D.29) on the time interval $[0, T]$ in Lemma D.17, we will already compute a uniform upper bound of (D.29), enabling the usage of the dominated convergence theorem. By the Cauchy–Schwarz inequality and Lemma D.13 we obtain for a constant C independent of the spatial point x and the resolution level $\delta > 0$:

$$\begin{aligned}
&\langle\int_0^{t\beta^{-2}} e^{rB}\sin^2(r\alpha^{-1}\beta^2L^{1/2})L^{-1}\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}\\
&\qquad\le \int_0^{T\delta^{-2\beta_1}} \|e^{rB_{\eta,\delta,x}/2}L_{\vartheta,\eta,\delta,x}^{-1/2}z_1\|_{L^2(\Lambda_{\delta,x})}\|e^{rB_{\eta,\delta,x}/2}L_{\vartheta,\eta,\delta,x}^{-1/2}z_2\|_{L^2(\Lambda_{\delta,x})}\,\mathrm{d}r\\
&\qquad\le \int_0^\infty C\Big(1\wedge r^{-\alpha_1/(2\beta_1)}\Big)^2\,\mathrm{d}r =: V < \infty.
\end{aligned} \tag{D.34}$$

Similarly to (D.29), as $\delta \to 0$, we obtain

$$\delta^{-2\beta_1}\langle\int_0^t (N'_{\delta,x}(r))^2\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$

$$= \delta^{-2\beta_1}\langle \int_0^t \Big(M_{\delta,x}(r) + \delta^{-2\beta_1} B_{\eta,\delta,x} N_{\delta,x}(r)\Big)^2 \,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$
$$= \langle \int_0^{t\beta^{-2}} e^{rB}\Big(\cos(r\alpha^{-1}\beta^2 L^{1/2}) + \frac{\alpha\beta^{-2}}{2} B \sin(r\alpha^{-1}\beta^2 L^{1/2})\Big)^2 \,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$
$$= \langle\Big(\alpha^2\beta^{-4} B e^{t\beta^{-2}B} \sin^2(t\alpha^{-1}L^{1/2})L^{-1}/4$$
$$+ \alpha\beta^{-2} e^{t\beta^{-2}B}\cos(t\alpha^{-1}L^{1/2})\sin(t\alpha^{-1}L^{1/2})L^{-1/2}/2$$
$$+(e^{t\beta^{-2}B} - I)B^{-1}/2\Big) z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$
$$\longrightarrow -\frac{1}{2}\langle \bar{B}_\eta^{-1} z_1, z_2\rangle_{L^2(\mathbb{R}^d)},$$

and

$$\delta^{-2\beta_1-\alpha_1}\langle \int_0^t N_{\delta,x}(r) N'_{\delta,x}(r)\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$
$$= \delta^{-2\beta_1-\alpha_1}\langle \int_0^t N_{\delta,x}(r)\Big(M_{\delta,x}(r) + \delta^{-2\beta_1} B_{\eta,\delta,x} N_{\delta,x}(r)\Big)\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$
$$= \langle \int_0^{t\beta^{-2}} e^{rB}\sin(r\alpha^{-1}\beta^2 L^{1/2})L^{-1/2}$$
$$\cdot\Big(\cos(r\alpha^{-1}\beta^2 L^{1/2}) + \frac{\alpha\beta^{-2}}{2} B\sin(r\alpha^{-1}\beta^2 L^{1/2})L^{-1/2}\Big)\,\mathrm{d}r z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})}$$
$$= \frac{1}{2}\langle \alpha\beta^{-2} e^{t\beta^{-2}B}\sin^2(t\alpha^{-1}L^{1/2})L^{-1} z_1, z_2\rangle_{L^2(\Lambda_{\delta,x})} \longrightarrow 0,$$

again having a uniform upper bound in analogy to (D.34).

■

LEMMA D.17. *Grant Assumption D.4 (i)-(iii) and suppose* $(u_0, v_0)^\top = (0,0)^\top$. *Recall the definition of* $C(\eta_1, T)$ *given by* (D.12) *and let* $\beta_1 = 0$. *Then, for* $1 \le i, j \le p$ *and* $1 \le k, l \le q$, *we obtain, as* $\delta \longrightarrow 0$, *the convergences*

$$\sup_{x\in\mathcal{J}}\left|\delta^{2(\alpha_i+\alpha_j-\alpha_1)}\int_0^T \mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), u_{\delta,x}^{\Delta_j})\,\mathrm{d}t + \frac{C(\eta_1,T)}{\vartheta_1}\|(-\Delta)^{(\alpha_i+\alpha_j-\alpha_1)/2}K\|^2_{L^2(\mathbb{R}^d)}\right| \longrightarrow 0; \tag{D.35}$$
$$\sup_{x\in\mathcal{J}}\left|\int_0^T \mathrm{Var}(v_{\delta,x}(t))\,\mathrm{d}t - C(\eta_1,T)\|K\|^2_{L^2(\mathbb{R}^d)}\right| \longrightarrow 0; \tag{D.36}$$
$$\sup_{x\in\mathcal{J}}\left|\delta^{2(\alpha_i-\alpha_1/2)}\int_0^T \mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), v_{\delta,x}(t))\,\mathrm{d}t\right| \longrightarrow 0. \tag{D.37}$$

If, $0 < 2\beta_1 \le \alpha_1$ *we obtain the convergences*

$$\sup_{x\in\mathcal{J}}\left|\delta^{2(\alpha_i+\alpha_j-\alpha_1-\beta_1)}\int_0^T \mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), u_{\delta,x}^{\Delta_j})\,\mathrm{d}t - \frac{T}{2\eta_1\vartheta_1}\|(-\Delta)^{(\alpha_i+\alpha_j-\alpha_1-\beta_1)/2}K\|^2_{L^2(\mathbb{R}^d)}\right| \longrightarrow 0; \tag{D.38}$$
$$\sup_{x\in\mathcal{J}}\left|\delta^{2(\beta_k+\beta_l-\beta_1)}\int_0^T \mathrm{Cov}(v_{\delta,x}^{\Delta_k}(t), v_{\delta,x}^{\Delta_l}(t))\,\mathrm{d}t + \frac{T}{2\eta_1}\|(-\Delta)^{(\beta_k+\beta_l-\beta_1)/2}K\|^2_{L^2(\mathbb{R}^d)}\right| \longrightarrow 0; \tag{D.39}$$

$$\sup_{x\in\mathcal{J}}\left|\delta^{2(\alpha_i+\beta_k-\beta_1-\alpha_1/2)}\int_0^T \mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), v_{\delta,x}^{\Delta_k}(t))\,\mathrm{d}t\right| \to 0. \tag{D.40}$$

Proof of Lemma D.17. With the majorant constructed in (D.34) in case of (D.38) or directly by Lemma D.13 for (D.35), we obtain uniformly in $x \in \mathcal{J}$ by Lemma D.15, Lemma D.16(i),(iii) and the dominated convergence theorem:

$$\begin{aligned}
&\delta^{2(\alpha_i+\alpha_j-\alpha_1-\beta_1)}\int_0^T \mathrm{Cov}(u_{\delta,x}^{\Delta_i}(t), u_{\delta,x}^{\Delta_j}(t))\,\mathrm{d}t\\
&= \delta^{-2\alpha_1-2\beta_1}\int_0^T\int_0^t \langle N_{\delta,x}(r)^2(-\Delta_0)^{\alpha_i}K, (-\Delta_0)^{\alpha_j}K\rangle_{L^2(\Lambda_{\delta,x})}\,\mathrm{d}r\,\mathrm{d}t\\
&= \begin{cases} -\int_0^T\int_0^t e^{\eta_1 r}\frac{1}{2}\langle \bar{A}_\vartheta^{-1}(-\Delta_0)^{\alpha_i}K, (-\Delta_0)^{\alpha_j}K\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}r\,\mathrm{d}t + o(1), & \beta_1 = 0,\\ \int_0^T \frac{1}{2}\langle \bar{B}_\eta^{-1}\bar{A}_\vartheta^{-1}(-\Delta_0)^{\alpha_i}K, (-\Delta_0)^{\alpha_j}K\rangle_{L^2(\mathbb{R}^d)}\,\mathrm{d}t + o(1), & \beta_1 > 0,\end{cases}\\
&= \begin{cases} -\frac{C(\eta_1,T)}{\vartheta_1}\|(-\Delta_0)^{(\alpha_i+\alpha_j-\alpha_1)/2}K\|_{L^2(\mathbb{R}^d)}^2 + o(1), & \beta_1 = 0,\\ \frac{T}{2\eta_1\vartheta_1}\|(-\Delta_0)^{(\alpha_i+\alpha_j-\alpha_1-\beta_1)/2}K\|_{L^2(\mathbb{R}^d)}^2 + o(1), & \beta_1 > 0.\end{cases}
\end{aligned}$$

This proves (D.35) and (D.38). Analogously, (D.36) and (D.39) as well as (D.37) and (D.40) follow using the remaining convergences in Lemma D.16. ∎

LEMMA D.18. *Grant Assumption D.4 (i)-(iii) and let $(u_0, v_0)^\top = (0,0)^\top$. Then, for $1 \le i,j \le p$, $1 \le k,l \le q$, we observe*

$$\sup_{x\in\mathcal{J}} \mathrm{Var}\left(\int_0^T u_{\delta,x}^{\Delta_i}(t)u_{\delta,x}^{\Delta_j}(t)\,\mathrm{d}t\right) = o(\delta^{-4(\alpha_i+\alpha_j-\alpha_1-\beta_1)}); \tag{D.41}$$

$$\sup_{x\in\mathcal{J}} \mathrm{Var}\left(\int_0^T v_{\delta,x}^{\Delta_k}(t)v_{\delta,x}^{\Delta_l}(t)\,\mathrm{d}t\right) = o(\delta^{-4(\beta_k+\beta_l-\beta_1)}); \tag{D.42}$$

$$\sup_{x\in\mathcal{J}} \mathrm{Var}\left(\int_0^T u_{\delta,x}^{\Delta_i}(t)v_{\delta,x}^{\Delta_k}(t)\,\mathrm{d}t\right) = o(\delta^{-4(\alpha_i+\beta_k-\beta_1-\alpha_1/2)}). \tag{D.43}$$

Proof of Lemma D.18. We only show the assertion for (D.41). The other two statements (D.42) and (D.43) follow in the same way. By Wick's formula [18, Theorem 1.28] it holds

$$\delta^{4(\alpha_i+\alpha_j-\alpha_1-\beta_1)}\,\mathrm{Var}\left(\int_0^T u_{\delta,x}^{\Delta_i}(t)u_{\delta,x}^{\Delta_j}(t)\,\mathrm{d}t\right) = \delta^{4(\alpha_1+\alpha_j-\alpha_1-\beta_1)}(V_1+V_2) \tag{D.44}$$

with

$$\begin{aligned}
V_1 &:= V((-\Delta)^{\alpha_i}K_{\delta,x}, (-\Delta)^{\alpha_i}K_{\delta,x}, (-\Delta)^{\alpha_j}K_{\delta,x}, (-\Delta)^{\alpha_j}K_{\delta,x}),\\
V_2 &:= V((-\Delta)^{\alpha_i}K_{\delta,x}, (-\Delta)^{\alpha_j}K_{\delta,x}, (-\Delta)^{\alpha_j}K_{\delta,x}, (-\Delta)^{\alpha_i}K_{\delta,x}),
\end{aligned}$$

and

$$V(w,w',z,z') := \int_0^T\int_0^T \mathrm{Cov}(\langle u(t),w\rangle, \langle u(s),w'\rangle)\,\mathrm{Cov}(\langle u(t),z\rangle, \langle u(s),z'\rangle)\,\mathrm{d}s\,\mathrm{d}t,$$

for $w,w',z,z' \in L^2(\Lambda)$. By Lemma D.15 and rescaling, we obtain the representation

$$\delta^{4(\alpha_i+\alpha_j-\alpha_1-\beta_1)}V_1$$

$$= \int_0^T \int_0^T \operatorname{Cov}(u^{\Delta_i}_{\delta,x}(t), u^{\Delta_i}_{\delta,x}(s)) \operatorname{Cov}(u^{\Delta_j}_{\delta,x}(t), u^{\Delta_j}_{\delta,x}(s)) \mathrm{d}s\mathrm{d}t \tag{D.45}$$

$$= 2\delta^{2\beta_1} \int_0^T \int_0^{t\delta^{-2\beta_1}} \left(\int_0^{t\delta^{-2\beta_1}-s} f_{i,i}(s,s')\,\mathrm{d}s' \right) \left(\int_0^{t\delta^{-2\beta_1}-s} f_{j,j}(s,s'')\,\mathrm{d}s'' \right) \mathrm{d}s\,\mathrm{d}t, \tag{D.46}$$

where, for $s, s' \in [0, T\delta^{-2\beta_1}]$, we have set

$$f_{i,j}(s,s') := \langle e^{(s+s')B_{\eta,\delta,x}/2} \sin(\delta^{-\alpha_1+2\beta_1}(s+s') L^{1/2}_{\vartheta,\eta,\delta,x}) L^{-1/2}_{\vartheta,\eta,\delta,x} (-\Delta_{\delta,x})^{\alpha_i} K,$$
$$e^{s' B_{\eta,\delta,x}/2} \sin(\delta^{-\alpha_1+2\beta_1}(s') L^{1/2}_{\vartheta,\eta,\delta,x}) L^{-1/2}_{\vartheta,\eta,\delta,x} (-\Delta_{\delta,x})^{\alpha_j} K\rangle_{L^2(\Lambda_{\delta,x})}.$$

In case that $\beta_1 = 0$, we use the pointwise convergences $f_{i,i}(s,s') \to 0$ and $f_{j,j}(s,s'') \to 0$, given by the slow-fast orthogonality in Lemma D.16(ii), and dominated convergence over fixed finite time intervals to prove the claim directly from the representation (D.45). If, however, $\beta_1 > 0$, we use Assumption D.4 (ii), i.e. $K = \Delta_0^{\lceil \alpha_1 \rceil} \widetilde{K}$, and Lemma D.13 such that

$$\sup_{x\in\mathcal{J}} |f_{i,i}(s',s)| \lesssim (1 \wedge (s+s')^{-(\alpha_i+\alpha_1/2)/\beta_1})(1 \wedge s^{-(\alpha_i+\alpha_1/2)/\beta_1})$$
$$\lesssim (1 \wedge s'^{-1})(1 \wedge s^{-1}).$$

Thus implies $\sup_{x\in\mathcal{J}} |V_1| = O(\delta^{-4(\alpha_i+\alpha_j-\alpha_1-\beta_1)} \delta^{2\beta_1} \log(\delta^{-2\beta_1})) = o(\delta^{-4(\alpha_i+\alpha_j-\alpha_1-\beta_1)})$. The arguments for V_2 follow in the same way be replacing $f_{i,i}$ and $f_{j,j}$ with $f_{i,j}$ and $f_{j,i}$ in (D.46), respectively. The assertion follows in view of (D.44). ■

Lemma D.19 (Bounds on the initial condition). *Grant Assumption D.4 (i) (ii) and (iv). Then, for $1 \le i \le p$, $1 \le j \le q$, we have*

(i) $\sup_{x\in\mathcal{J}} \delta^{4\alpha_i-2\alpha_1-2\beta_1} \left(\int_0^T \langle M(t)u_0 + N(t)v_0, (-\Delta)^{\alpha_i} K_{\delta,x}\rangle^2 \,\mathrm{d}t \right) = o(1);$

(ii) $\sup_{x\in\mathcal{J}} \delta^{4\beta_j-2\beta_1} \left(\int_0^T \langle A_\vartheta N(t)u_0 + (M(t) + B_\eta(N(t))v_0, (-\Delta)^{\beta_j} K_{\delta,x}\rangle^2 \,\mathrm{d}t \right) = o(1).$

Proof of Lemma D.19.

(i) Define the reverse scaling operation for $z \in L^2(\mathbb{R}^d)$ via

$$z_{(\delta,x)^{-1}}(y) := \delta^{d/2} z(x+\delta y), \quad y \in \mathbb{R}^d.$$

The rescaling Lemma D.8, self-adjointness and the commutativity of operators imply

$$\langle M(t)u_0, (-\Delta)^{\alpha_i} K_{\delta,x}\rangle^2$$
$$= \delta^{-4\alpha_i} \langle (u_0)_{(\delta,x)^{-1}}, M_{\delta,x}(t)(-\Delta_{\delta,x})^{\alpha_i} K\rangle^2_{L^2(\Lambda_{\delta,x})}$$
$$= \delta^{-4\alpha_i+4\alpha_1} \langle ((-\Delta)^{\alpha_1} u_0)_{(\delta,x)^{-1}}, M_{\delta,x}(t)(-\Delta_{\delta,x})^{\alpha_i-\alpha_1} K\rangle^2_{L^2(\Lambda_{\delta,x})}$$
$$\lesssim \delta^{-4\alpha_i+4\alpha_1} \|(-\Delta)^{\alpha_1} u_0\|^2_{L^2(\Lambda)} \|e^{t\delta^{-2\beta_1} B_{\eta,\delta,x}} (-\Delta_{\delta,x})^{\alpha_i-\alpha_1} K\|^2_{L^2(\Lambda_{\delta,x})}$$
$$\lesssim \delta^{-4\alpha_i+4\alpha_1} \|e^{t\delta^{-2\beta_1} B_{\eta,\delta,x}} (-\Delta_{\delta,x})^{\alpha_i-\alpha_1} K\|^2_{L^2(\Lambda_{\delta,x})}.$$

Thus, using Lemma D.13 and that $K = \Delta_0^{\lceil \alpha_1 \rceil}\widetilde{K}$, we obtain the upper bound

$$\begin{aligned}
&\sup_{x\in\mathcal{J}} \delta^{4\alpha_i-2\alpha_1-2\beta_1} \int_0^T \langle M(t)u_0, (-\Delta)^{\alpha_i} K_{\delta,x}\rangle^2 \,\mathrm{d}t \\
&\lesssim \delta^{2\alpha_1} \int_0^{T\delta^{-2\beta_1}} \sup_{x\in\mathcal{J}} \| e^{tB_{\eta,\delta,x}} (-\Delta_{\delta,x})^{\alpha_i-\alpha_1} K\|^2_{L^2(\Lambda_{\delta,x})} \,\mathrm{d}t \\
&\lesssim \delta^{2\alpha_1} \int_0^{T\delta^{-2\beta_1}} 1 \wedge t^{-\alpha_i/\beta_1} \,\mathrm{d}t \\
&= O(\delta^{2(\alpha_1-\beta_1)}) = o(1).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\langle N(t)v_0, (-\Delta)^{\alpha_i} K_{\delta,x}\rangle^2 \\
&\quad = \delta^{-4\alpha_i} \langle (v_0)_{(\delta,x)^{-1}}, N_{\delta,x}(t)(-\Delta_{\delta,x})^{\alpha_i} K\rangle^2_{L^2(\Lambda_{\delta,x})} \\
&\quad = \delta^{-4\alpha_i+2\alpha_1} \langle ((-\Delta)^{\alpha_1/2} v_0)_{(\delta,x)^{-1}}, N_{\delta,x}(t)(-\Delta_{\delta,x})^{\alpha_i-\alpha_1/2} \Delta K\rangle^2_{L^2(\Lambda_{\delta,x})} \\
&\quad \lesssim \delta^{-4\alpha_i+4\alpha_1} \| e^{t\delta^{-2\beta_1} B_{\eta,\delta,x}} L^{-1/2}_{\vartheta,\eta,\delta,x} (-\Delta_{\delta,x})^{\alpha_i-\alpha_1/2} K\|^2_{L^2(\Lambda_{\delta,x})}
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sup_{x\in\mathcal{J}} \delta^{4\alpha_i-2\alpha_1-2\beta_1} \int_0^T \langle N(t)v_0, \Delta K_{\delta,x}\rangle^2 \,\mathrm{d}t \\
&\lesssim \delta^{2\alpha_1} \int_0^{T\delta^{-2\beta_1}} \sup_{x\in\mathcal{J}} \| e^{tB_{\eta,\delta,x}} L^{-1/2}_{\vartheta,\eta,\delta,x} (-\Delta_{\delta,x})^{\alpha_i-\alpha_1/2} K\|^2_{L^2(\Lambda_{\delta,x})} \,\mathrm{d}t \\
&= O(\delta^{2(\alpha_1-\beta_1)}) = o(1),
\end{aligned}$$

proving the assertion.

(ii) The steps from (i) can be repeated, resulting in

$$\begin{aligned}
&\sup_{x\in\mathcal{J}} \delta^{4\beta_j-2\beta_1} \left(\int_0^T \langle A_\vartheta N(t)u_0 + (M(t) + B_\eta(N(t))v_0, (-\Delta)^{\beta_j} K_{\delta,x}\rangle^2 \,\mathrm{d}t \right) \\
&\quad = O(\delta^{2(\alpha_1-\beta_1)}) = o(1). \qquad \blacksquare
\end{aligned}$$

D.4.3 *Proof of the CLT*

Proof of Theorem D.5.

1. Assume first that $(u_0, v_0)^\top = (0,0)^\top$. For any $1 \le i, j \le p+q$, we obtain from Lemma D.17 and Lemma D.18 that

$$\begin{aligned}
(\rho_\delta \mathcal{I}_\delta \rho_\delta)_{ij} &= \rho_{ii}\rho_{jj} \sum_{k=1}^N \int_0^T (Y_{\delta,k}(t))_i (Y_{\delta,k}(t))_j \,\mathrm{d}t \\
&= (\Sigma_{\vartheta,\eta})_{ij} + o_{\mathbb{P}}(1).
\end{aligned}$$

This yields for zero initial conditions the convergence

$$(\rho_\delta \mathfrak{I}_\delta \rho_\delta) \xrightarrow{\mathbb{P}} \Sigma_\vartheta, \quad \delta \to 0. \tag{D.47}$$

In order to extend this result to a general initial condition $(u_0, v_0)^\top$ satisfying Assumption D.4 (iii), let $(\bar{u}(t), \bar{v}(t))^\top$ be defined as $(u(t), v(t))^\top$, but starting in $(0,0)^\top$ such that for $z \in L^2(\Lambda)$

$$\begin{aligned}\langle u(t), z\rangle &= \langle \bar{u}(t), z\rangle + \langle M(t)u_0, z\rangle + \langle N(t)v_0, z\rangle,\\ \langle v(t), z\rangle &= \langle \bar{v}(t), z\rangle + \langle A_\vartheta N(t)u_0, z\rangle + \langle (M(t) + B_\eta N(t))v_0, z\rangle.\end{aligned}$$

If $\bar{\mathfrak{I}}_\delta$ corresponds to the observed Fisher information with zero initial condition, then by the Cauchy–Schwarz inequality

$$\left|(\rho_\delta \mathfrak{I}_\delta \rho_\delta)_{ij} - (\rho_\delta \bar{\mathfrak{I}}_\delta \rho_\delta)_{ij}\right| \lesssim (\rho_\delta \bar{\mathfrak{I}}_\delta \rho_\delta)_{ii}^{1/2} w_j^{1/2} + (\rho_\delta \bar{\mathfrak{I}}_\delta \rho_\delta)_{jj}^{1/2} w_i^{1/2} + w_i^{1/2} w_j^{1/2},$$

for all $1 \le i, j \le p+q$, where

$$w_i = \begin{cases} \sup_{x\in\mathcal{J}} N\rho_{ii}^2 \left(\int_0^T \langle M(t)u_0 + N(t)v_0, (-\Delta)^{\alpha_i} K_{\delta,x}\rangle^2 \,\mathrm{d}t\right), & 1 \le i \le p,\\ \sup_{x\in\mathcal{J}} N\rho_{ii}^2 \left(\int_0^T \langle A_\vartheta N(t)u_0 + (M(t) + B_\eta(N(t))v_0, (-\Delta)^{\beta_i} K_{\delta,x}\rangle^2 \,\mathrm{d}t\right), & \text{else.}\end{cases}$$

By the first part, $(\rho_\delta \bar{\mathfrak{I}}_\delta \rho_\delta)_{ii}$ is bounded in probability and Lemma D.19 shows $w_i = o(1)$. Hence, we obtain (D.47) also in the case of non-zero initial conditions.

Due to Assumption D.4 (i)-(iii), $\Sigma_{\vartheta,\eta}$ is well-defined as all entries are finite. Regarding invertibility, note first that $\Sigma_{\vartheta,\eta}$ is invertible if and only if both $\Sigma_{1,\vartheta,\eta}$ and $\Sigma_{2,\vartheta,\eta}$ are invertible. We only argue that $\Sigma_{1,\vartheta,\eta}$ is invertible as the argument for $\Sigma_{2,\vartheta,\eta}$ is identical. Let $\lambda \in \mathbb{R}^p$ such that

$$0 = \sum_{i,j=1}^p \lambda_i\lambda_j (\Sigma_{1,\vartheta,\eta})_{ij} \iff 0 = \sum_{i,j=1}^p \lambda_i\lambda_j \|(-\Delta)^{(\alpha_i+\alpha_j-\alpha_1-\beta_1)} K\|^2_{L^2(\mathbb{R}^d)}.$$

Now,

$$\begin{aligned}0 &= \sum_{i,j=1}^p \lambda_i\lambda_j \|(-\Delta)^{(\alpha_i+\alpha_j-\alpha_1-\beta_1)} K\|^2_{L^2(\mathbb{R}^d)}\\ &= \langle \sum_{i=1}^p \lambda_i (-\Delta)^{\alpha_i-(\alpha_1+\beta_1)/2} K, \sum_{i=1}^p \lambda_i (-\Delta)^{\alpha_i-(\alpha_1+\beta_1)/2} K\rangle_{L^2(\mathbb{R}^d)},\end{aligned}$$

hence $\sum_{i=1}^p \lambda_i (-\Delta)^{\alpha_i-(\alpha_1+\beta_1)/2} K = 0$. Since the functions $(-\Delta)^{\alpha_i-(\alpha_1+\beta_1)/2} K$,$1 \le i \le p$, are linearly independent by Assumption D.4 (iii), $\Sigma_{1,\vartheta,\eta}$ is invertible.

2. We refer to Theorem A.3 in Paper A for a detailed proof of the CLT in the case of the perturbed stochastic heat equation, which relies on a general multivariate martingale central limit theorem. All steps translate directly into our setting due to the stochastic convergence $\rho_\delta \mathfrak{I}_\delta \rho_\delta \xrightarrow{\mathbb{P}} \Sigma_{\vartheta,\eta}$ from (i).

■

Acknowledgements AT gratefully acknowledges the financial support of Carlsberg Foundation Young Researcher Fellowship grant CF20-0604. The research of EZ has been partially funded by Deutsche Forschungsgemeinschaft (DFG)—SFB1294/1-318763901.

References

[1] I. Aihara S. and A. Bagchi. “Parameter identification for hyperbolic stochastic systems”. In: *Journal of mathematical analysis and applications* 160.2 (1991), pp. 485–499.

[2] S. I. Aihara. “Identification of Discontinuous Parameter in Stochastic Hyperbolic Systems”. In: *IFAC Proceedings Volumes* 27.8 (1994), pp. 191–196.

[3] R. Altmeyer, T. Bretschneider, J. Janák, and M. Reiß. “Parameter Estimation in an SPDE Model for Cell Repolarisation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1 (2022), pp. 179–199.

[4] R. Altmeyer, I. Cialenco, and G. Pasemann. “Parameter estimation for semilinear SPDEs from local measurements”. In: *Bernoulli* 29.3 (2023), pp. 2035–2061.

[5] R. Altmeyer and M. Reiß. “Nonparametric estimation for linear SPDEs from local measurements”. In: *Annals of Applied Probability* 31.1 (2021), pp. 1–38.

[6] W. Arendt, C. J. K. Batty, M. Hieber, and F. Neubrander. *Vector-Valued Laplace Transforms and Cauchy Problems*. Springer Basel, 2001.

[7] T. Aspelmeier, A. Egner, and A. Munk. “Modern statistical challenges in high-resolution fluorescence microscopy”. In: *Annual Reviews of Statistics and Its Applications* 2 (2015), pp. 163–202.

[8] A. S. Backer and W. E. Moerner. “Extending Single-Molecule Microscopy Using Optical Fourier Processing”. In: *The Journal of Physical Chemistry B* 118.28 (2014), pp. 8313–8329.

[9] G. CHEN and D. L. RUSSELL. “A mathematical model for linear elastic systems with structural damping”. In: *Quarterly of Applied Mathematics* 39.4 (1982), pp. 433–454.

[10] S. P. Chen and R. Triggiani. “Proof of extensions of two conjectures on structural damping for elastic systems.” In: *Pacific Journal of Mathematics* 136.1 (1989), pp. 15 –55.

[11] C. Chong. “High-frequency analysis of parabolic stochastic PDEs”. In: *Annals of Statistics* 48.2 (2020), pp. 1143–1167.

[12] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.

[13] A. Favini and E. Obrecht. “Conditions for parabolicity of second order abstract differential equations”. In: *Differential and Integral Equations* 4.5 (1991), pp. 1005 –1022.

[14] S. Gaudlitz and M. Reiß. “Estimation for the reaction term in semi-linear SPDEs under small diffusivity”. In: *Bernoulli* 29.4 (2023), pp. 3033 –3058.

[15] F. Hildebrandt and M. Trabs. “Parameter estimation for SPDEs based on discrete observations in time and space.” In: *Electronic Journal of Statistics* 15 (2021), pp. 2716–2776.

[16] M. Huebner and B. Rozovskii. “On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE’s”. In: *Probability Theory and Related Fields* 103.2 (1995), pp. 143–163.

[17] J. Janák and M. Reiß. "Parameter estimation for the stochastic heat equation with multiplicative noise from local measurements". In: *Stochastic Process. Appl.* 175 (2024), Paper No. 104385.

[18] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[19] M. Kovács, S. Larsson, and F. Saedpanah. "Finite Element Approximation of the Linear Stochastic Wave Equation with Additive Noise". In: *SIAM Journal on Numerical Analysis* 48.2 (2010), pp. 408–427.

[20] U. Küchler and M. Sørensen. *Exponential families of stochastic processes*. Springer, 1997.

[21] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2013.

[22] W. Leissa A. *Vibration of Plates*. National Aeronautics and Space Administration, 1969.

[23] W Liu and S. Lototsky. "Parameter estimation in hyperbolic multichannel models". In: *Asymptotic Analysis* 68 (2010), pp. 223–248.

[24] W. Liu and S. V. Lototsky. *Estimating Speed and Damping in the Stochastic Wave Equation*. 2008. arXiv: 0810.0046 [math.PR].

[25] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. English. Cambridge University Press, 2014.

[26] S. V. Lototsky. "Statistical inference for stochastic parabolic equations: a spectral approach". In: *Publicacions Matemàtiques* 53.1 (2009), pp. 3–45.

[27] B. Maslowski and J. Pospíšil. "Ergodicity and Parameter Estimates for Infinite-Dimensional Fractional Ornstein-Uhlenbeck Process". In: *Applied Mathematics and Optimization* 57 (2008), pp. 401–429.

[28] V. Melnikova Irina and A. Filinkov. *Abstract Cauchy Problems: Three Approaches*. Chapman and Hall, 2001.

[29] L. Pandolfi. *Systems with Persistent Memory: Controllability, Stability, Identification*. Vol. 54. Interdisciplinary Applied Mathematics. Springer International Publishing, 2021.

[30] G. Pasemann and W. Stannat. "Drift estimation for stochastic reaction-diffusion systems". In: *Electronic Journal of Statistics* 14.1 (2020), pp. 547–579.

[31] N. Reddy J. *Theory and Analysis of Elastic Plates and Shells (2nd. ed.)* CRC Press, 2006.

[32] M. Reiß, C. Strauch, and L. Trottner. *Change point estimation for a stochastic heat equation*. 2023. arXiv: 2307.10960 [math.ST].

[33] W. Rudin. *Functional analysis*. McGraw-Hill, 1991.

[34] K. Schmüdgen. *Unbounded Self-adjoint Operators on Hilbert Space*. Vol. 265. Graduate Texts in Mathematics. Springer, 2012.

[35] N. Shimakura. *Partial Differential Operators of Elliptic Type*. American Mathematical Soc., 1992.

[36] M. Sova. *Cosine operator functions*. Instytut Matematyczny Polskiej Akademi Nauk, 1966.

[37] Y. Tonaki, Y. Kaino, and M. Uchida. "Parameter estimation for linear parabolic SPDEs in two space dimensions based on high frequency data". In: *Scandinavian Journal of Statistics* 50.4 (2023), pp. 1568–1589.

[38] J. Weidmann. “Strong Operator Convergence and Spectral Theory of Ordinary Differential Operators”. In: *ZESZYTY NAUKOWE-UNIWERSYTETU JAGIELLONSKIEGO-ALL SERIES* 1208 (1997), pp. 153–163.

[39] E. Ziebell. *Non-parametric estimation for the stochastic wave equation*. 2024. arXiv: 2404.18823 [math.ST].