Delft University of Technology

Faculty of Electrical Engineering, Mathematics and Computer Science

Fourier Spectral Methods for Bayesian Filtering of Stochastic Partial Differential Equations

Thorben Pieper

Supervisors:
Associate Professor Frank van der Meulen
Professor Sally Cripps

Master of Science
Applied Mathematics

April 2021

Abstract

The development of computationally efficient algorithms for statistical inference of stochastic differential equations has been a long-standing research subject ever since the advent of stochastic analysis. Recently, there has been an increasing interest in extending these methods to the statistical study of stochastic partial differential equations. In this thesis, we are concerned with statistical inference for the linear stochastic convection-diffusion equation (SCDE) driven by a stochastic forcing term that is spatially a Matérn process. The approach is based on spectral methods for partial differential equations and approximates a solution to the SCDE via a Fourier spectral decomposition. The resulting spectral processes are given by a family of uncoupled Ornstein-Uhlenbeck processes, for which computationally efficient statistical inference is possible based on the Kalman filter. We give verifiable experimental results for all statistical problems - filtering, smoothing and parameter estimation - based on simulated data. We further derive a novel weak solution to the SCDE driven by spatial Matérn noise for spatially periodic boundary conditions and show that the solution is indeed approximated by the Fourier spectral decomposition, thereby validating the statistical model. For the heat equation, we generalize the spectral filtering approach on compact Riemannian manifolds. Experimental results for this generalization are given on the two-dimensional unit sphere.

Acknowledgements

First and foremost, I would like to thank my two supervisors, Frank van der Meulen and Sally Cripps, for their guidance and advice throughout this project. I am incredibly grateful for the great resilience that you have both shown to my stubborn ways of working. I want to thank you, Frank, for your patience and flexibility in supervising this work from such a great distance. Thank you for dealing with online meetings at odd times and the occasional disappearance of your student on the opposite end of the world. I would like to thank you, Sally, for the trust you have put in me by getting me here in the first place and for making this collaboration possible.

Thank you, Jessica Cornock, for always helping me out with the mountains of bureaucratic work. I now believe that any sufficiently advanced operations management is indistinguishable from magic.

Thank you to my family and friends back at home for all your support. Our video calls kept me going in a time where being an international student was not the easiest. And of course, I want to thank all newly found friends here in Sydney, who have reminded me of this place's beauty again and again. As I have learned to say here:

'Thanks heaps everyone.'

Contents

1. Introduction **1**

2. Preliminaries **4**
2.1. Review - Tools from Fourier Analysis 4
2.2. Stochastic Processes and Stochastic Integration 8
2.3. Filtering and Smoothing of Stochastic Differential Equations 28
2.4. Linear Stochastic Partial Differential Equations 40

3. A Spectral Fourier Method for the Stochastic Convection-Diffusion Equation **53**
3.1. The Stochastic Convection-Diffusion Equation 53
3.2. The Fourier Spectral Method 56
3.3. Filtering, Smoothing and Parameter Inference Setup 64
3.4. Experiments 71

4. Solving the Stochastic Convection Diffusion Equation on the Torus **80**
4.1. Matérn Processes on $\mathbb{T}^2$ 80
4.2. The Weak Solution on $\mathbb{T}^2$ 86

5. Future Work **95**
5.1. Some Immediate Generalizations 95
5.2. The Stochastic Heat Equation on Riemannian Manifolds 100
5.3. Non-Linear SPDEs 107

Bibliography **110**

A. Appendix **113**

1. Introduction

Statistical inference for dynamical systems that are subject to random forcing has been an active research area ever since the popularization of stochastic analysis in the middle of the last century. At the basis of this field lies the *filtering problem*, the task of estimating the future state of a stochastic dynamical system based on empirical observations of the system. Further tasks include the reconstruction of previous states – the *smoothing problem* – and *statistical inference* on the *model variables* that parametrize the system's dynamics.

When the system is purely time-dependent, one represents it as a *stochastic differential equation (SDE)*. In cases where its dynamics are *linear*, the filtering and smoothing problems can be solved efficiently and in closed-form using the methods developed by Kálmán [1960].

The development of efficient filtering and smoothing algorithms as well as parameter inference in the more complicated setting of *nonlinear* systems remains an open research direction. For an introduction into this field, see for example Särkkä and Solin [2019], for recent developments see Mbalawata et al. [2013], Graham et al. [2019] and Mider et al. [2020].

In this work, we are concerned with dynamical systems that are dependent on both time and space variables, i.e. represented by a *stochastic partial differential equation (SPDE)*. Compared to the case of SDEs, statistical inference for such space-time processes faces two additional challenges.

For one, even in the deterministic case, partial differential equations are notoriously harder to work with than their ordinary counterparts. In fact, with the exception of linear equations, there exists no unifying framework for solving SPDEs comparable to the *Itô calculus* for SDEs.

The second issue, commonly raised in the field of *spatio-temporal statistics*, is that space-time data sets are inherently large and computational efficiency therefore becomes a yet bigger concern (Cressie and Wikle [2011]). Even in the linear case, when closed-form solutions are available in the form of *Gaussian processes*, naive smoothing and parameter inference based on Gaussian process regression is of cubic computational complexity in regards to the size of the data set. This problem is commonly referred to as the 'big-N-problem'.

A common way of numerically solving PDEs are *spectral methods* (see e.g. Gottlieb and Orszag [1977]). The basic idea is to seek solutions given by a (finite) sum of *basis functions*, weighted by their *spectral coefficients*. In many situations, such decompositions transform the problem of solving a PDE into solving a family of ordinary differential equations for the spectral coefficients.

A recent paper by Sigrist et al. [2015] makes use of the spectral method based on the *Fourier basis functions* to construct a spatio-temporal statistical model motivated by the *stochastic convection-diffusion equation (SCDE)* driven by a stochastic forcing term that is temporally white and spatially a *Matérn process*. Statistical inference for this model can be done on the spectral processes in linear time, thereby circumventing the big-N-problem and showing promising results in an exemplary application. The model proposed by Sigrist et al. [2015] will serve as the corner stone of this thesis.

Contribution We investigate the statistical properties of the SCDE's Fourier spectral approximation as proposed by Sigrist et al. [2015]. In contrast to previous work, we derive the spectral processes in the framework of Itô calculus. In particular, it will be shown that the spectral processes are given by uncoupled complex Ornstein-Uhlenbeck processes (theorem 3.2). We prove that these have similar properties to their real-valued counterparts (lemmata 2.5 and 2.6), which results in easily obtainable temporal convergence statements of the proposed model to a stationary process (corollary 3.1).

Furthermore, we give verifiable experimental results for all statistical problems - filtering, smoothing and parameter estimation - based on simulated data. Our statistical model follows the approach given in Sigrist et al. [2015], albeit with slight differences, including our choice to retain a complex-valued formulation of the spectral processes. Other small-scale discrepancies, e.g. differing prefactors, are pointed out at appropriate times throughout the models derivation in sections 3.2 and 3.3.

Our results (section 3.4) show that the spatio-temporal process can be efficiently estimated based on noisy and discretely sampled observations using the Kalman filter and -smoother for the spectral processes. It further shows that even simple Bayesian inference schemes can recover model parameters. This fills a gap in the current literature, in which only the models predictive power has been examined.

The second part of the thesis is concerned with *solving* the stochastic convection-diffusion equation. As the equation is linear, this can readily be done in the framework of *weak solutions*. We obtain a novel weak solution of the SCDE with spatially periodic boundary conditions, driven by a stochastic forcing term that is temporally white and spatially a Matérn process (theorem 4.3). For this, we make use of a recent result by Borovitskiy et al. [2020] that generalizes the Matérn covariance function for Riemannian manifolds. Our result shows that the Fourier spectral decomposition converges in the L^2-sense to the weak solution of the SCDE as we let the number of basis functions go to infinity (corollary 4.3).

We further investigate the weak solutions long-time behavior as it converges to a stationary generalized Gaussian process (proposition 4.2) and derive the spectral density of the corresponding stationary Gaussian process (corollary 4.4).

Besides validating the statistical model, these novel results open up easily obtainable generalizations and future research directions. A noteworthy example of high practical importance is an application of the spectral filtering approach for the stochastic heat equation on the sphere (section 5.2.1).

Outline The outline of this thesis is as follows. Chapter 2 introduces all necessary theory for the later chapters. This includes reviewing some basics concepts that we assume the reader is familiar with, such as tools from Fourier analysis and the Kalman filter. More attention will be given to the specifics of complex valued stochastic processes and -Itô integrals, the spectral representation theorem for stationary processes and the framework of weak solutions to linear SPDEs.

The following chapter is concerned with the Fourier spectral approximation of the SCDE. After deriving the model based on the spectral representation theorem, the properties of its spectral processes will be investigated and experimental results will be presented. A considerable part of the chapter is devoted to the derivation of the *spectral Kalman filter*, an improvement of the original Kalman filter based on some of the advantageous properties of the discrete Fourier transform.

In chapter 4, we provide the aforementioned weak solution to the SCDE. Its first section will derive a spectral representation for stationary processes defined on periodic domains. This will allow us to reframe the problem in section 4.2 and derive its weak solution as well as investigate its convergence behavior to a stationary process.

Lastly, chapter 5 provides some details on possible generalizations of the proposed model and future search directions. In particular, it includes first experimental results for the spectral filtering method applied to the stochastic heat equation on the unit sphere.

2. Preliminaries

This chapter reviews and introduces some basic theory that we will need for later chapters. We begin by revisiting tools from Fourier analysis, in particular Fourier spectral decompositions and the discrete Fourier transform.

Next up, we take a look at complex valued processes and their Itô integrals. In particular, we introduce the spectral representation of stationary processes, which will play a key factor in motivating the filtering scheme in chapter three.

The last two sections are concerned with stochastic dynamical systems, both in the purely temporal as well as the space-time case. In the former case, we revisit the filtering problem for stochastic differential equations. For the latter case, we introduce a framework of solving linear stochastic partial differential equations, which will allow us to give a rigorous interpretation of our proposed filtering method.

2.1. Review - Tools from Fourier Analysis

Many key concepts in this thesis involve tools from Fourier analysis. For example, they play a role in representing stationary stochastic processes by their spectral decompositions or in solving differential equations by transforming differential operators to algebraic operations in the frequency domain.

We shortly summarize some of these tools and their essential properties. For a detailled introduction to the underlying theory, we refer to Stein and Shakarchi [2003]. For more details on the discrete Fourier transform, see e.g. Briggs and Henson [1995].

Definition 2.1 (Fourier Transform)**.** Let $d \in \mathbb{N}$ and $f \in L^1(\mathbb{R}^d)$ be an integrable function. Its *Fourier transform* is defined as

$$\mathscr{F}(f)(\omega) := \hat{f}(\omega) := (2\pi)^{-d} \int_{\mathbb{R}^d} f(t) \exp(-i\omega \cdot t) dt. \tag{2.1}$$

The Fourier transform is a linear operator whose inverse, when it exists, is given by

$$f(t) = \mathscr{F}^{-1}(\hat{f})(t) = \int_{\mathbb{R}^d} \hat{f}(\omega) \exp(i\omega \cdot t) d\omega. \tag{2.2}$$

The inversion formula holds whenever $\hat{f}$ is integrable. For example, this is the case when f lies in the *Schwartz space* of rapidly decreasing functions, though more general cases exist and we shall from now on assume that our functions are 'nice enough' such that the inversion formula holds.

In cases where f is a periodic function, the Fourier transform is not well defined. Instead, a decomposition likening (2.2) is given by the *Fourier series*.

Definition 2.2 (Fourier Series)**.** Let f be a periodic function on $[0, L]^d$. If existent, its Fourier series decomposition is given by

$$\begin{aligned} f(t_1, ..., t_d) &= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \alpha_{\boldsymbol{j}} \exp\left(2\pi i \left(\frac{j_1}{L} t_1 + ... + \frac{j_d}{L} t_d\right)\right), \text{ with} \\ \alpha_{\boldsymbol{j}} &= \frac{1}{L^d} \int_{[0,L]^d} f(t_1, ..., t_d) \exp\left(-2\pi i \left(\frac{j_1}{L} t_1 + ... + \frac{j_d}{L} t_d\right)\right) dt_1 ... dt_d. \end{aligned} \tag{2.3}$$

The $\boldsymbol{j}$-th *Fourier basis function*, corresponding to the $\boldsymbol{j}$-th spatial wave $\boldsymbol{\xi_j} := 2\pi(\frac{j_1}{L}, ..., \frac{j_d}{L})$, is defined by

$$\phi_{\boldsymbol{j}}(t_1, ..., t_d) := \exp\left(2\pi i \left(\frac{j_1}{L} t_1 + ... + \frac{j_d}{L} t_d\right)\right).$$

One can show that the Fourier basis functions form an orthonormal basis in the space of square-integrable functions:

Theorem 2.1. The family $(\phi_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathbb{Z}^d}$ are an orthonormal basis of the Hilbert space $\mathcal{L}^2([0, L]^d)$ of square-integrable functions equipped with the scalar product

$$\langle h, g \rangle_{L^2} := \int_{[0,L]^d} h(t_1, ..., t_d) \overline{g(t_1, ..., t_d)} dt_1 ... dt_d.$$

In particular, for any $h \in \mathcal{L}^2([0, L]^d)$, it holds (with convergence in regards to $\| \cdot \|_{\mathcal{L}^2}$)

$$h = \sum_{\boldsymbol{j} \in \mathbb{Z}^d} \alpha_{\boldsymbol{j}} \phi_{\boldsymbol{j}}.$$

The pointwise convergence of the Fourier series as defined in (2.3) is in general not trivial, but holds, for example, if f is continuous and its *Fourier coefficients* $(\alpha_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathbb{Z}^d}$ are absolutely summable. Again, we shall assume that the periodic functions of our concern are 'nice enough' such that a Fourier series representation exists.

Let us further remark that it is common to represent functions of bounded domain as periodic functions of unbounded domain by extending their domain 'mod period length'.

If f is a real-valued function, it is easy to see that its Fourier coefficient $\alpha_{0,0}$ is real-valued as well. In fact, in this case the Fourier series expansion (2.3) can be substituted by the *cosine-sine-decomposition* which makes use of real coefficients only, but we will not be concerned with that representation.

Lastly, the *discrete Fourier transform* takes these concepts to the discrete domain. It can be regarded as the numerical approximation of both the Fourier transform $\hat{f}$ and the Fourier coefficients $(\alpha_j)_j$, based on evenly gridded samples of the function f.

Definition 2.3 (Discrete Fourier Transform)**.** Let $\boldsymbol{f} = \left(f_{-N/2+1}, ..., f_{N/2}\right)$ be a vector of complex numbers for an even number $N \in \mathbb{N}$. Then its discrete Fourier transform

(DFT) $\boldsymbol{F}$ is a N-periodic complex-valued vector given by

$$(\Phi(\boldsymbol{f}))_k = F_k = \frac{1}{N} \sum_{n=-N/2+1}^{N/2} f_n \exp\big(-2\pi i \tfrac{k}{N} n\big). \tag{2.4}$$

The inverse discrete Fourier transform (IDFT) of a complex vector $\boldsymbol{F} = \big(F_{-N/2+1}, ..., F_{N/2}\big)$ is given by

$$(\Phi^{-1}(\boldsymbol{F}))_n = \sum_{k=-N/2+1}^{N/2} F_k \exp\big(2\pi i \tfrac{k}{N} n\big) \tag{2.5}$$

and satisfies the inverse relations:

$$f_n = [\Phi^{-1}(\Phi(\boldsymbol{f}))]_n,$$
$$F_k = [\Phi(\Phi^{-1}(\boldsymbol{F}))]_k.$$

Note that the DFT operator is linear and represented by the symmetric matrix

$$\boldsymbol{\Phi} = \tfrac{1}{N} \exp\big(-2\pi i \tfrac{k}{N} n\big)_{-N/2+1 \le k,n \le N/2} \in \mathbb{C}^{N \times N}. \tag{2.6}$$

It can be shown that the Fourier basis functions satisfy the following orthogonality property in n:

$$\sum_{n=-N/2+1}^{N/2} \exp\big(2\pi \tfrac{l}{N} n\big) \exp\big(-2\pi i \tfrac{k}{N} n\big) = \begin{cases} N & \text{if } (k-l) \bmod N = 0, \\ 0 & \text{else.} \end{cases}$$

From this it follows that the IDFTs matrix representation is given by

$$\boldsymbol{\Phi}^{-1} = N\overline{\boldsymbol{\Phi}}.$$

The DFTs computational costs of $\mathcal{O}(N^2)$ associated with its matrix formulation can be drastically improved upon by the *Fast Fourier Transform* (FFT), whose computation of the DFT is of complexity class $\mathcal{O}(N \log(N))$. Due to its computational advantages we shall from here on assume the FFT as a black-box DFT solver. For more details on the FFT, see Cooley and Tukey [1965], whose work popularized the FFT algorithm, or chapter 10 of Briggs and Henson [1995] for a detailled introduction.

The DFT-IDFT pair easily extends to multiple dimensions. In particular the two-dimensional case will be of interest to us. Given $\boldsymbol{f} = (f_{n,m})_{n,m}$ for $n = -\frac{N}{2}+1, ..., \frac{N}{2}, m = -\frac{M}{2}+1, ..., \frac{M}{2}$, the DFT is given by

$$\boldsymbol{\Phi} f_{k,l} = F_{k,l} = \frac{1}{NM} \sum_n \sum_m f_{n,m} \exp\big(-2\pi i (\tfrac{k}{N} n + \tfrac{l}{M} m)\big) \tag{2.7}$$

The sequence $F_{k,l}$ is N-periodic in k and M-periodic in l and it is easy to see that the DFT-IDFT properties and matrix representations extend to the multidimensional case. Of further interest is the special case where f is a real-valued vector. We then have

$$F_{k,l} = \overline{F}_{-k,-l}.$$

In that case, only the following $\frac{NM}{2} + 2$ Fourier coefficients of the DFT $\boldsymbol{F} = \boldsymbol{\Phi f}$ need be computed

$$\begin{cases} F_{0,0}, F_{N/2,0}, F_{0,M/2}, F_{N/2,M/2} \in \mathbb{R} & \\ F_{k,M/2} = \overline{F}_{-k,M/2} \in \mathbb{C}, & \text{for } k = 1, ..., \frac{N}{2} - 1 \\ F_{k,0} = \overline{F}_{-k,0} \in \mathbb{C}, & \text{for } k = 1, ..., \frac{N}{2} - 1 \\ F_{k,l} = \overline{F}_{-k,-l} \in \mathbb{C}, & \text{for } k = -\frac{N}{2} + 1, ..., \frac{N}{2}, l = 1, ..., \frac{M}{2} - 1 \end{cases} \tag{2.8}$$

and these are uniquely defined by NM real quantities.

2.2. Stochastic Processes and Stochastic Integration

We start this section off by introducing complex-valued stochastic processes, as they require special attention compared to their real-valued counterparts. Following this, we review integration in regards to stochastic processes, in particular for complex-valued Wiener processes and processes whose covariance functions are of bounded variation. Lastly, these concepts come to a use in showing proving the spectral representation theorem for stationary processes.

2.2.1. Review - Stochastic Processes

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space. A stochastic process $X = (X(t))_{t\in T}$ is a family of random variables

$$\begin{aligned} X(t) : \Omega &\longrightarrow E, \quad t \in T \\ \omega &\mapsto X(t)(\omega). \end{aligned}$$

If $X(t)$ is square-integrable for each $t \in T$, we denote the *mean-* and *covariance function* (or *kernel*) of X by

$$\begin{aligned} m(t) &:= \mathrm{E}[X(t)] \\ k(s,t) &:= \mathrm{Cov}[X(s), X(t)], \quad s,t \in T. \end{aligned}$$

If $T = \mathbb{R}_+^d$, one speaks of a X as a *random function* or a *continuous (time) stochastic process*. Likewise, if $T = \mathbb{N}$ or $T = \mathbb{Z}$, one speaks of a *discrete (time) stochastic process*. In situations where the context is clear, we shall simply write $(X(t))_t$ or X. Unless specified otherwise, we assume that $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\cdot)$ denotes the standard Borel-σ-algebra.

If $(E, \mathcal{E}) = (\mathbb{C}, \mathcal{B}(\mathbb{C}))$, we call X *complex-valued*. These processes require some additional attention, especially in regards to their second order moments. We shall therefore give them a short introduction, based on Lapidoth [2009], chapters 17 and 24.

Definition 2.4 (Complex-valued Stochastic Process)**.** A complex-valued stochastic process Z is a process

$$Z(t) = X(t) + iY(t),$$

where X, Y are two real-valued processes. Its mean function is defined as $\mathrm{E}[Z(t)] = \mathrm{E}[X(t)] + i\mathrm{E}[Y(t)]$ and we denote it by $m(t)$. Its second order momenta are given by the *covariance-* and *pseudo-covariance function*, respectively defined by

$$\begin{aligned} k(s,t) := \mathrm{Cov}[Z(s), Z(t)] &:= \mathrm{E}[(Z(s) - m(s))\overline{(Z(t) - m(t))}] \\ &= \mathrm{E}\left[Z(s)\overline{Z(t)}\right] - m(s)\overline{m(t)}, \\ k^*(s,t) := \mathrm{Cov}^*[Z(s), Z(t)] &:= \mathrm{E}[(Z(s) - m(s))(Z(t) - m(t))] \\ &= \mathrm{E}[Z(s)Z(t)] - m(s)m(t). \end{aligned} \tag{2.9}$$

Note that the covariance function is conjugate symmetric, meaning that $k(s,t) = \overline{k(t,s)}$ for all $s, t \in T$, whereas the pseudo-covariance remains symmetric as in the real-valued case.

An easy calculation shows that the *variance* and *pseudo-variance* functions can be written as

$$\begin{aligned} \mathrm{Var}[Z(t)] &:= k(t,t) = \mathrm{Var}[X(t)] + \mathrm{Var}[Y(t)] \\ \mathrm{Var}^*[Z(t)] &:= k^*(t,t) = \mathrm{Var}[X(t)] - \mathrm{Var}[Y(t)] + 2i\mathrm{Cov}[X(t), Y(t)]. \end{aligned} \tag{2.10}$$

We may also consider complex-valued random vectors, whose second-order statistics, in line with the previous definition, require special care:

Definition 2.5 (Complex-valued Random Vector)**.** A complex-valued random vector is a vector $\boldsymbol{Z} = (Z_1, ..., Z_k)^T$, where Z_i, $i = 1, ..., k$, is a complex-valued random variable. We define its mean by $\mathrm{E}[\boldsymbol{Z}] = (\mathrm{E}[Z_1], ..., \mathrm{E}[Z_k])^T$ and its covariance and pseudo-covariance matrices via

$$\begin{aligned} \boldsymbol{K}_{\boldsymbol{ZZ}} &:= \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}^H] - \mathrm{E}[\boldsymbol{Z}]\mathrm{E}[\boldsymbol{Z}]^H = [\mathrm{Cov}[Z_i, Z_j]]_{1 \le i,j \le k} \\ \boldsymbol{K}^*_{\boldsymbol{ZZ}} &:= \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}^T] - \mathrm{E}[\boldsymbol{Z}]\mathrm{E}[\boldsymbol{Z}]^T = [\mathrm{Cov}^*[Z_i, Z_j]]_{1 \le i,j \le k}. \end{aligned} \tag{2.11}$$

Let $\tilde{\boldsymbol{Z}}$ be another complex-valued random vector. Then we define the *cross-covariance* and *pseudo-cross-covariance matrices* of $\boldsymbol{Z}$ and $\tilde{\boldsymbol{Z}}$ as

$$\begin{aligned} \boldsymbol{K}_{\boldsymbol{Z}\tilde{\boldsymbol{Z}}} &:= \mathrm{E}[\boldsymbol{Z}\tilde{\boldsymbol{Z}}^H] - \mathrm{E}[\boldsymbol{Z}]\mathrm{E}[\tilde{\boldsymbol{Z}}]^H = [\mathrm{Cov}[Z_i, \tilde{Z}_j]]_{1 \le i,j \le k} \\ \boldsymbol{K}^*_{\boldsymbol{Z}\tilde{\boldsymbol{Z}}} &:= \mathrm{E}[\boldsymbol{Z}\tilde{\boldsymbol{Z}}^T] - \mathrm{E}[\boldsymbol{Z}]\mathrm{E}[\tilde{\boldsymbol{Z}}]^T = [\mathrm{Cov}^*[Z_i, \tilde{Z}_j]]_{1 \le i,j \le k}. \end{aligned} \tag{2.12}$$

Likewise to the real-valued case, we call $\boldsymbol{Z}$ and $\tilde{\boldsymbol{Z}}$ *uncorrelated* or *orthogonal* if

$$\begin{aligned} \boldsymbol{K}_{\boldsymbol{Z}\tilde{\boldsymbol{Z}}} &= 0, \\ \boldsymbol{K}^*_{\boldsymbol{Z}\tilde{\boldsymbol{Z}}} &= 0. \end{aligned}$$

The generalization of Gaussian vectors and -processes to the complex case is straightforward:

Definition 2.6 (Complex-Valued Gaussian Vector)**.** A vector $\boldsymbol{Z} = (Z_1, ..., Z_k)$, with $Z_j = X_j + iY_j$ for all j, is called a *(k-dimensional) complex Gaussian vector* if

$$(X_1, Y_1, ..., X_k, Y_k)$$

is a $2k$-dimensional Gaussian vector. If $k = 1$ and $X, Y \overset{iid}{\sim} \mathcal{N}(0, \frac{1}{2})$, we call $Z = X + iY$ *standard complex Gaussian* and denote it by $Z \sim \mathcal{CN}(0, 1)$. It is a special case of a proper complex Gaussian, which we will define below.

It can be shown that, just like in the real case, any complex linear transformation of a complex Gaussian vector or the sum of two independent complex Gaussians remains complex Gaussian.

Recall that a Gaussian process $(X(t))_t$ is a stochastic process such that $(X(t_1), ..., X(t_k))$ is a Gaussian vector for any $t_1, ..., t_k$. Accordingly, we define:

Definition 2.7 (Complex Gaussian Process)**.** A *complex Gaussian process* is a process Z such that $[Z(t_1), ..., Z(t_k)]$ is a $k-$dimensional complex Gaussian vector for any $t_1, ..., t_k$.

Whereas Gaussian processes are uniquely defined by their mean- and covariance function, this is not necessarily true in the complex case:

Example 2.1. Let X, Y be two independent, real-valued Gaussian processes with mean zero and identical covariance function $\kappa(s,t)$. Define the processes Z_1, Z_2 via

$$\begin{aligned} Z_1(t) &:= X(t) + iY(t) \\ Z_2(t) &:= X(t) + iX(t). \end{aligned}$$

It is immediate that both are complex-valued Gaussian processes with identical covariance function given by

$$\begin{aligned} \mathrm{E}[Z_1(s)\overline{Z_1(t)}] &= \mathrm{E}[X(s)X(t)] + \mathrm{E}[Y(s)Y(t)] + i\left(\mathrm{E}[Y(s)X(t)] - \mathrm{E}[X(s)Y(t)]\right) \\ &= 2\kappa(s,t) \\ &= 2\mathrm{E}[X(s)X(t)] + i\left(\mathrm{E}[X(s)X(t)] - \mathrm{E}[X(s)X(t)]\right) \\ &= \mathrm{E}[Z_2(s)\overline{Z_2(t)}]. \end{aligned}$$

On the other hand, the pseudo-covariance function of Z_1 is given by

$$\begin{aligned} \mathrm{E}[Z_1(s)Z_1(t)] &= \mathrm{E}[X(s)X(t)] - \mathrm{E}[Y(s)Y(t)] + i\left(\mathrm{E}[Y(s)X(t)] + \mathrm{E}[X(s)Y(t)]\right) \\ &= 0, \end{aligned}$$

whereas for Z_2 we have

$$\begin{aligned} \mathrm{E}[Z_2(s)Z_2(t)] &= \mathrm{E}[X(s)X(t)] - \mathrm{E}[X(s)X(t)] + i\left(\mathrm{E}[X(s)X(t)] + \mathrm{E}[X(s)X(t)]\right) \\ &= 2i\kappa(s,t) \\ &\neq 0. \end{aligned}$$

This example shows that, contrary to the real-valued cased, a complex Gaussian process is not uniquely represented by its mean and covariance function. It therefore motivates the following definiton of *properness*.

Definition 2.8 (Properness)**.** A *proper complex random vector* $\boldsymbol{Z}$ is a random vector such that $\boldsymbol{K}^*_{\boldsymbol{ZZ}} = 0$. A *proper complex process* $Z = (Z(t))_t$ is a process such that $(Z(t_1), ..., Z(t_k))$ is a proper complex random vector for all $t_1, ..., t_k$.

As hinted at before, properness comes in particularly handy in the context of Gaussian variables, as seen in the following lemma.

Lemma 2.1. Let $\boldsymbol{Z} = \boldsymbol{X} + i\boldsymbol{Y}$ be a k-dimensional, proper complex Gaussian vector. Then $\boldsymbol{Z}$ is uniquely defined by its mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{K}$ and we write $\boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{m}, \boldsymbol{K})$. Its density function is given by

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{\pi^n \det(\boldsymbol{K})} \exp\left(-(\boldsymbol{z}-\boldsymbol{m})^H \boldsymbol{K}^{-1}(\boldsymbol{z}-\boldsymbol{m})\right). \tag{2.13}$$

It further holds that

$$(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{N}_{2k}\left(\begin{pmatrix}\mathrm{Re}(\boldsymbol{m})\\ \mathrm{Im}(\boldsymbol{m})\end{pmatrix}, \frac{1}{2}\begin{pmatrix}\mathrm{Re}(\boldsymbol{K}) & -\mathrm{Im}(\boldsymbol{K})\\ \mathrm{Im}(\boldsymbol{K}) & \mathrm{Re}(\boldsymbol{K})\end{pmatrix}\right). \tag{2.14}$$

Proof. See Lapidoth [2009], proposition 24.3.12. for the first claim and the proof of proposition 24.3.7. for the second claim. □

Note that the process Z in example 2.1 is a proper Gaussian process and can therefore be defined via its mean- and covariance function. Additionally, it is a special case in the sense that its covariance function is real-valued, which does not generally have to be true - it holds if and only if $\mathrm{Cov}[X(s), Y(t)] = \mathrm{Cov}[X(t), Y(s)]$ for all s, t.

Proper complex vectors possess properties resembling those of Gaussian variables, in particular:

Lemma 2.2.

1. A $k-$dimensional complex random vector $\boldsymbol{Z}$ is proper if and only if $\boldsymbol{\alpha}^T \boldsymbol{Z}$ is proper for all $\boldsymbol{\alpha} \in \mathbb{C}^k$.
2. Let $\boldsymbol{Z}$ be a k-dimensional proper random vector. Then $\boldsymbol{AZ}$ is proper for any $\boldsymbol{A} \in \mathbb{C}^{m\times k}$.
3. Let $\boldsymbol{Z_1} \sim \mathcal{CN}(\boldsymbol{m_1}, \boldsymbol{K_1})$, $\boldsymbol{Z_2} \sim \mathcal{CN}(\boldsymbol{m_2}, \boldsymbol{K_2})$ be two independent, proper complex Gaussian vectors. Then for any $\boldsymbol{A_1}, \boldsymbol{A_2} \in \mathbb{C}^{m\times k}$, $\boldsymbol{Z} = \boldsymbol{A_1 Z_1} + \boldsymbol{A_2 Z_2}$ is proper complex Gaussian with

$$\boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{A_1 m_1} + \boldsymbol{A_2 m_2}, \boldsymbol{A_1 K_1 A_1^H} + \boldsymbol{A_2 K_2 A_2^H}). \tag{2.15}$$

Proof. See Lapidoth [2009], propositions 17.4.2 and 17.4.3. □

Remark 2.1. Let $\boldsymbol{m} \in \mathbb{C}^k$ be a complex vector and $\boldsymbol{K} \in \mathbb{C}^{k\times k}$ be a hermitian matrix with Cholesky decomposition $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^H$. We can then draw a sample $\boldsymbol{z}$ of the proper complex Gaussian vector $\boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{m}, \boldsymbol{K})$ by drawing $(\boldsymbol{x}, \boldsymbol{y})$ from the $2k$-dimensional Gaussian vector $(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{N}_{2k}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{E})$, $\boldsymbol{E}$ denoting the unit matrix, and setting

$$\boldsymbol{z} = \boldsymbol{m} + \boldsymbol{L}(\boldsymbol{x} + i\boldsymbol{y}).$$

We finish this paragraph by defining the complex counterparts of two of the most fundamental stochastic processes - discrete time white noise and the Wiener process. A *discrete time Gaussian white noise* is a real-valuled process $\dot{w} = (\dot{w}_n)_{n\in\mathbb{N}}$ such that

$$\begin{aligned} \dot{w}_n &\sim \mathcal{N}(0,1)\\ \mathrm{E}[\dot{w}_n \dot{w}_m] &= 0, \quad \text{for all } n \neq m. \end{aligned} \tag{2.16}$$

Its complex counterpart is a process $\dot{w} = (\dot{w}_n)_{n\in\mathbb{N}}$ such that

$$\dot{w}_n = \frac{\dot{w}_{1,n} + i\dot{w}_{2,n}}{\sqrt{2}}, \tag{2.17}$$

where $\dot{w}_1, \dot{w}_2$ are independent discrete time Gaussian white noise processes.

A *Wiener process* or *standard Brownian Motion* $(w(t))_{t\in[0,\infty)}$ is a centered Gaussian process with covariance function $E[w(s)w(t)] = \min(s,t)$. It has almost surely continuous trajectories $t \mapsto w(t)$ and normally distributed increments

$$w(t_{i+1}) - w(t_i) \sim \mathcal{N}(0, t_{i+1} - t_i).$$

that are independent for all $0 \leq t_1 < t_2 < ... < t_n, n \in \mathbb{N}, i \in \{0, ..., n-1\}$. Likewise to the discrete white noise case, we define *complex Wiener process (Brownian motion)* as the complex-valued process given by

$$w(t) = \frac{w_1(t) + iw_2(t)}{\sqrt{2}},$$

where w_1, w_2 are independent Wiener processes. It follows from example 2.1 that complex Brownian motion is a proper complex Gaussian process.

The issue of continuous time white noise, often times treated as the formal derivative of the Wiener process, will be addressed in sections 2.3 and 2.4.

2.2.2. Review - Stochastic Integration

This sections reviews integration in regards to stochastic processes. We first consider the case in which the integrator is a stochastic processes whose covariance function is of bounded variation, which will be of use for us in the spectral representation of stationary processes.

We also review basic properties of the Itô integral, in particular the complex-valued case and show that it defines a proper complex Gaussian process for deterministic integrands.

Let T be an $n-$dimensional interval, $X = (X(t))_t$ be a stochastic process and $f(t)$ be a function or a stochastic process itself. Recall that a stochastic integral may be constructed as the $L^2(\mathbb{P})$-limit of *Riemann-Stieltjes sums* of the form

$$\int_T f(t)dX(t) = \lim_{k\to\infty} \sum_{\Delta t\in\mathcal{P}_k} f(t')\Delta X(t). \tag{2.18}$$

Here, $(\mathcal{P}_k)_k$ denotes any sequence of sub-interval partitions of T - each a family of disjoint sub-intervals $\Delta t = [t_1 + \Delta_1] \times ... \times [t_n + \Delta_n]$ with $\cup_{\Delta t\in\mathcal{P}_k} = T$ - such that the mesh size

$$|\mathcal{P}_k| := \max_{\Delta t\in\mathcal{P}_k} \left\{ |\Delta t| := \prod_{i=1}^{n} \Delta_i \right\} \longrightarrow 0$$

and t' denotes the "lower left" corner point $t' = [t_1, ..., t_n]$ for any interval Δt. Further, $\Delta X(t)$ denotes the increment of X over any interval Δt, given by

$$\begin{aligned}\Delta X(t) := & X(t_1 + \Delta_1, ..., t_n + \Delta_n) \\ & - \sum_i X(t_1 + \Delta_1, ..., t_{i-1} + \Delta_{i-1}, t_i, t_{i+1} + \Delta_{i+1}, ...t_n + \Delta_n) \\ & + \sum_{i \neq j} X(t_1 + \Delta_1, ..., t_i, ..., t_j, ...t_n + \Delta_n) \\ & - ... + (-1)^n X(t_1, ..., t_n).\end{aligned} \tag{2.19}$$

Note that the convergence of the right-hand side in (2.18) is not trivial and requires assumptions on the integrand f and integrator X. In particular, we need the following two cases, whose assumptions we shall specify shortly:

1. The integrator is a stochastic process whose covariance function is of bounded variation. In that case it will suffice to consider determinstc integrands.

2. The integrator is a Wiener process and the integrand is itself a stochastic process. For our purpose, it will then suffice to consider $n = 1$.

The following theorem treats the first case. Its proof relies on the construction of the Lebesgue-Stieltjes integral for functions of bounded variation. For details, see appendix A.

Theorem 2.2. Let $X = (X(t))_t$ be a centered stochastic process and $g(t)$ be a measurable function, both possibly complex-valued. Denote by $k(t, s)$ the covariance function of X. Assume that $k(s, r)$ is of bounded variation on an interval $T \subset \mathbb{R}^n$ and that the Lebesgue-Stieltjes integral

$$\int_{T \times T} g(t)\overline{g(s)} dk(t, s) \tag{2.20}$$

exists. Then the stochastic integral

$$I = \int_T g(t) dX(t) = \lim_{k \to \infty} \sum_{\Delta t \in \mathcal{P}_k} g(t') \Delta X(t) \tag{2.21}$$

exists as a limit in $L^2(\mathbb{P})$ with

$$\begin{aligned}\mathrm{E}[I] &= 0 \\ \mathrm{E}[I\overline{I}] &= \int_{T \times T} g(t)\overline{g(s)} dk(t, s).\end{aligned} \tag{2.22}$$

Further, let $Y = (Y(t))_t$ be another centered stochastic process and $f(t)$ be a function such that $J = \int_{T'} f(t) dY(t)$ exists. Then

$$\mathrm{E}[I\overline{J}] = \int_{T \times T'} g(t)\overline{f(s)} dk_{XY}(t, s), \tag{2.23}$$

where k_{XY} denotes the cross-covariance of X and Y: $k_{XY}(t, s) = \mathrm{E}[X(t)\overline{Y(s)}]$.

Proof. An idea of the proof is given in Lindgren [2012], theorem 2.16, and is based on the *Loève Criterion*, which states that a sequence of random variables $(S_k)_k$ converges in $L^2(\mathbb{P})$ if and only if the limit $\lim_{k,m\longrightarrow\infty} \mathrm{E}[S_k\overline{S_m}]$ exists.

Thus, let $(\mathcal{P}_k)_k, (\mathcal{P}'_m)_m$ be two sequences of subinterval partitions of T with mesh sizes going to zero. Denote by $(S_k)_k, (S'_m)_m$ the respective Riemann-Stieltjes sums.

Note that for any intervals $\Delta t, \Delta s$, we have $\mathrm{E}[\Delta X(t)\overline{\Delta X(s)}] = \Delta k(s,t)$ where the increment $\Delta k(s,t)$ is taken over the $2n$-dimensional interval $\Delta t \times \Delta s$. We then have

$$\begin{aligned}
\lim_{k,m\to\infty} \mathrm{E}[S_k\overline{S'_m}] &= \lim_{k,m\to\infty} \mathrm{E}\left[\sum_{\Delta t\in\mathcal{P}_k} g(t')\Delta X(t) \overline{\sum_{\Delta s\in\mathcal{P}'_m} g(s')\Delta X(s)}\right] \\
&= \lim_{k,m\to\infty} \sum_{\Delta t\in\mathcal{P}_k}\sum_{\Delta s\in\mathcal{P}'_m} g(t')\overline{g(s')}\mathrm{E}\left[\Delta X(t)\overline{\Delta X(s)}\right] \\
&= \lim_{k,m\to\infty} \sum_{\Delta t\times\Delta s\in\mathcal{P}''_{k,m}} g(t')\overline{g(s')}\Delta k(s,t) \\
&= \int_{T\times T} g(t)\overline{g(s)}dk(t,s),
\end{aligned}$$

noting that $\mathcal{P}''_{k,m} := \{\Delta t\times\Delta s \mid \Delta t\in\mathcal{P}_k, \Delta s\in\mathcal{P}'_m\}$ is a sequence of partitions of $T\times T$ with mesh size going to zero. The convergence is therefore assured by the existence of the Lebesgue-Stieltjes integral $\int_{T\times T} g(t)\overline{g(s)}dk(t,s)$. We thus receive existence and uniqueness of $\int_T g(t)dX(t)$ and equation (2.22).

For the second claim, note that the Cauchy-Schwarz inequality assures the existence of both integrals in (2.23). The equality follows in the same way as the one above.

□

The case where $f(t)$ is a stochastic process itself requires more attention. A detailed construction can be found in Kallianpur and Sundar [2014], chapter 5.

Theorem 2.3. Let $(w(t))_t$ be a Wiener process and $(f(t))_t$ be an adapted process such that

$$\mathrm{E}\left[\int_0^t f^2(s)ds\right] < \infty. \tag{2.24}$$

Then, the *Itô integral* $I(t) := \int_0^t f(s)dw(s)$ exists as a $L^2(\mathbb{P})$-limit with

$$\begin{aligned}
\mathrm{E}[I(t)] &= 0 \\
\mathrm{E}[I(t)^2] &= \mathrm{E}\left[\int_0^t f^2(s)ds\right].
\end{aligned} \tag{2.25}$$

Proof. See e.g. Kallianpur and Sundar [2014], chapter 5. □

We denote by $M_w^2([0,t])$ the space of all adapted processes such that (2.24) holds and by $\mathcal{L}_X^2(T)$ the space of all complex-valued functions such that (2.20) exists. Note that the stochastic integrals above are linear mappings on the respective spaces.

The stochastic integral is easily extended to the multivariate case. Let $w_1, ..., w_m$ be independent Wiener processes such that $\boldsymbol{w} = (w_1, ..., w_m)$ is a $m-$dimensional Wiener process and let $\boldsymbol{f} = (f_{i,j})_{1\leq i\leq k, 1\leq j\leq m}$ be a matrix-valued process such that $f_{i,j} \in M_w^2$ for all i, j. The vector-valued stochastic integral $\boldsymbol{I}(t) = \int_0^t \boldsymbol{f}(s)d\boldsymbol{w}(s)$ is then defined via

$$\boldsymbol{I}(t) := (I_1(t), ..., I_k(t))^T,$$

where for all $i = 1, ..., k$

$$I_i(t) := \sum_{j=1}^{m} \int_0^t f_{i,j}(s)dw_j(s).$$

An extension for the integral defined in theorem 2.2 is defined in the same way.

We are particulary interested in the two-dimensional extension, which allows us to extend the Itô integral in regards to complex Brownian motion:

Definition 2.9 (Complex-valued Stochastic Integration)**.** Let $w = \frac{1}{\sqrt{2}}(w_1 + iw_2)$ be a complex Brownian motion and $f = f_1 + if_2$ be such that $f_1, f_2 \in M_w^2$. We define the complex-valued stochastic integral $\int_0^t f(s)dw(s)$ as

$$\begin{aligned}\int_0^t f(s)dw(s) = \frac{1}{\sqrt{2}}\Bigg[&\int_0^t f_1(s)dw_1(s) - \int_0^t f_2(s)dw_2(s) \\ &+i\left(\int_0^t f_1(s)dw_2(s) + \int_0^t f_2(s)dw_1(s)\right)\Bigg].\end{aligned} \tag{2.26}$$

Note that the complex stochastic integral can be identified with the vector-valued Itô integral

$$\frac{1}{\sqrt{2}}\int_0^t \begin{pmatrix} f_1(s) & -f_2(s) \\ f_2(s) & f_1(s)\end{pmatrix} d(w_1(s), w_2(s))^T.$$

Now, let w be a Wiener process and $f \in L_w^2$. It is a well-known fact (see e.g. Lindgren [2012], theorem 2.16) that the special case of both theorem 2.2 and 2.3 has the practical property of defining centered Gaussian variables

$$I(t) := \int_0^t f(s)dw(s), \quad t \in \mathbb{R}_{\geq 0}, \tag{2.27}$$

with covariances given by

$$\mathrm{E}[I(s)I(t)] = \int_0^{s\wedge t} f^2(u)du. \tag{2.28}$$

In fact, using the independence of Brownian motion increments, one is quick to see that any linear combination

$$\lambda_1 \int_0^{t_1} f(s)dw(s) + ... + \lambda_k \int_0^{t_k} f(s)dw(s)$$

is again a Gaussian variable, concluding that the process $(I(t))_t$ is a centered Gaussian process with covariance function given by (2.28).

The following lemma shows that this property extends to the complex case and defines a proper complex Gaussian process:

Lemma 2.3. Let $w = \frac{1}{\sqrt{2}}(w_1 + iw_2)$ be a complex Wiener process. Let $f(s) = f_1(s) + if_2(s)$ be a complex-valued function such that $f_1, f_2 \in \mathcal{L}^2_w$.

Then the process $I = (I(t))_t$, defined by

$$I(t) = \int_0^t f(s)dw(s),$$

is a proper complex Gaussian process with covariance function

$$\mathrm{E}[I(s)\overline{I(t)}] = \int_0^{s\wedge t} |f(u)|^2 du. \tag{2.29}$$

Proof. By definition, it holds

$$\begin{aligned}\int_0^t f(s)dw(s) = \frac{1}{\sqrt{2}}\Bigg[&\int_0^t f_1(s)dw_1(s) - \int_0^t f_2(s)dw_2(s) \\ &+i\left(\int_0^t f_1(s)dw_2(s) + \int_0^t f_2(s)dw_1(s)\right)\Bigg] \\ =:& \frac{1}{\sqrt{2}}\left[I_1(t) + iI_2(t)\right],\end{aligned}$$

where I_1, I_2 are both sums of independent centered Gaussian processes as given in (2.27).

To show that I is a complex Gaussian process, we need to show that $[I_1(t_1), I_2(t_1), ..., I_1(t_k), I_2(t_k)]$ is a $2k$-dimensional Gaussian vector for any $t_1, ..., t_k \in \mathbb{R}_+$. This follows from the fact that for any $\lambda_i, \mu_i \in \mathbb{R}, i = 1, ..., k$, the random variable

$$\begin{aligned}&\lambda_1 I_1(t_1) + \mu_1 I_2(t_1) + ... + \lambda_k I_1(t_k) + \mu_k I_2(t_k) \\ &= \int_0^{t_1} [(\lambda_1 + ... + \lambda_k)f_1 + (\mu_1 + ... + \mu_k)f_2]dw_1 \\ &+ \int_0^{t_1} [-(\lambda_1 + ... + \lambda_k)f_2 + (\mu_1 + ... + \mu_k)f_1]dw_2 \\ &+ .... \\ &+ \int_{t_{k-1}}^{t_k} [\lambda_k f_1 + \mu_k f_2]dw_1 + \int_{t_{k-1}}^{t_k} [(-\lambda_k)f_2 + \mu_k f_1]dw_2\end{aligned}$$

is Gaussian as a sum of independent Gaussian variables.

To show properness, following example 2.1, it suffices to show that

$$\begin{aligned}&\mathrm{Cov}[I_1(s), I_1(t)] = \mathrm{Cov}[I_2(s), I_2(t)] \\ &\mathrm{Cov}[I_1(s), I_2(t)] = 0.\end{aligned}$$

It holds

$$\begin{aligned}
\mathrm{E}[I_1(s)I_1(t)] &= \mathrm{E}\left[\left(\int_0^s f_1(u)dw_1(u) - \int_0^s f_2(u)dw_2(u)\right)\right. \\
&\quad \left.\left(\int_0^t f_1(u)dw_1(u) - \int_0^t f_2(u)dw_2(u)\right)\right] \\
&= \int_0^{s\wedge t} f_1^2(u)du + \int_0^{s\wedge t} f_2^2(u)du \\
&= \mathrm{E}\left[\left(\int_0^s f_1(u)dw_2(u) + \int_0^s f_2(u)dw_1(u)\right)\right. \\
&\quad \left.\left(\int_0^t f_1(u)dw_2(u) + \int_0^t f_2(u)dw_1(u)\right)\right] \\
&= \mathrm{E}[I_2(s)I_2(t)],
\end{aligned}$$

where we have used the independence of w_1, w_2 and equation (2.28) in steps 2 and 3. To show uncorrelatedness, the same arguments give

$$\begin{aligned}
\mathrm{E}[I_1(s)I_2(t)] &= \mathrm{E}\left[\left(\int_0^s f_1(u)dw_1(u) - \int_0^s f_2(u)dw_2(u)\right)\right. \\
&\quad \left.\left(\int_0^t f_1(u)dw_2(u) + \int_0^t f_2(u)dw_1(u)\right)\right] \\
&= \int_0^{s\wedge t} f_1(u)f_2(u)du - \int_0^{s\wedge t} f_1(u)f_2(u)du \\
&= 0.
\end{aligned}$$

Thus, $(I(t))_t$ is a proper complex Gaussian process with covariance function given by

$$\begin{aligned}
\mathrm{E}[I(s)\overline{I(t)}] &= \frac{1}{2}\left(\mathrm{E}[I_1(s)I_1(t)] + \mathrm{E}[I_2(s)I_2(t)]\right) + \frac{i}{2}(\mathrm{E}[I_2(s)I_1(t)] - \mathrm{E}[I_1(s)I_2(t)]) \\
&= \int_0^{s\wedge t} f_1^2(u) + f_2^2(u)du \\
&= \int_0^{s\wedge t} |f(u)|^2 du.
\end{aligned}$$

□

2.2.3. Spectral Representations of Stationary Stochastic Processes

The following section, based on chapter 4 of Yaglom [1987a], is concerned with stationary stochastic processes. Intuitively, any process $(X(t))_t$ whose stochastic behavior is invariant under a parameter shift $t \mapsto t + \tau$ is referred to as a stationary process.

We are particularly interested in the spectral representation of X, which enables its approximation by orthogonal random variables.

Definition 2.10 (Stationary Processes)**.** Let $X = (X(t))_{t\in\mathbb{R}^n}$ be a real-valued stochastic process with mean- and covariance functions $m(t)$ and $k(s,t)$. It is called *strictly stationary* if the distribution of any p-dimensional random vector

$$[X(t_1+\tau),...,X(t_p+\tau)], \quad p\in\mathbb{N}, t_1,...t_p,\tau\in\mathbb{R}^n,$$

is independent of the translation vector τ.

It is called *weakly stationary* (or sometimes *homogeneous* in the case of $n>1$) if its mean function is constant and its covariance function is a function of the difference τ only:

$$\begin{aligned} m(t) &= m, \quad t\in\mathbb{R}^n, \\ k(t+\tau,t) &= c(\tau), \quad t,\tau\in\mathbb{R}^n. \end{aligned}$$

We call X *isotropic* if its covariance function depends on the distance $\|\tau\|$ only.

If X is a non-proper, complex-valued process, one further requires its pseudo-covariance function to be dependent on the difference τ (distance $\|\tau\|$) only.

It is clear that a strictly stationary process with finite second-order moments is indeed stationary in the weak sense. The marginal distributions of $X(t)$ are identical and thus have constant mean, whereas the translation invariant joint distributions of $[X(t_1+\tau), X(t_2+\tau)]$ imply that the (pseudo-)covariance function depends on the distance t_2-t_1 only.

However, in the case where X is a Gaussian process, the notions of strict and weak stationarity are equivalent due to the fact that the finite dimensional distributions of a Gaussian process are uniquely defined by its mean- and covariance function. Furthermore, the main result of our interest - the spectral representation theorem - concerns stationary process in the weak sense and we shall from now on refer to those as stationary processes.

It is an imminent fact that the covariance function $c(\tau)$ of any stationary process is *positive-definite* in the sense that for any $t_1,...,t_m\in\mathbb{R}^n$, $a_1,...,a_m\in\mathbb{C}$

$$\sum_{i,j=1}^{m} a_i\overline{a_j}c(t_i-t_j)\geq 0.$$

Conversely, one may show that any positive-definite function $c(\tau)$ is the covariance function of a stationary Gaussian process (see e.g. Lindgren [2012], theorem 3.1 for the one-dimensional case).

Therefore, the class of covariance functions of stationary processes equals the class of positive-definite functions on $\mathbb{R}^n$, which opens up their characterization via a version of Bochner's theorem:

Theorem 2.4 (Bochner's Theorem)**.** Let $c(\tau)$ be a positive-definite function on $\mathbb{R}^n$. Then there exists a bounded function $F:\mathbb{R}^n\longrightarrow\mathbb{R}$ with non-negative increments

$\Delta F(\omega)$, such that $c(\tau)$ is given by the Fourier-Stieltjes transform

$$c(\tau) = \int_{\mathbb{R}^n} \exp(i\omega \cdot \tau) dF(\omega). \tag{2.30}$$

The function F is known as the *spectral distribution function* of $c(\tau)$ (or X), and the Lebesgue-Stieltjes measure μ_F as its *spectral distribution.*

Proof. See Bochner [1959], chapter 4, §20. □

Note that $c(0) = \mu_F(\mathbb{R}^n)$ and that $F(\omega)$ is unique up to an additive constant. It is common to choose F such that $\lim_{\omega_i \to \infty, i=1,\ldots,n} F(\omega) = 0$. Furthermore, in the special case that $c(\tau)$ is absolutely integrable,

$$\int_{\mathbb{R}^n} |c(\tau)| d\tau < \infty,$$

one may show that $c(\tau)$ forms a Fourier transform pair with the *spectral density*

$$f(\omega) := \tfrac{\partial^n F(\omega)}{\partial\omega_1 \ldots \partial\omega_n}.$$

Consequently, we get

$$\begin{aligned} c(\tau) &= \int_{\mathbb{R}^n} f(\omega) \exp(i\omega \cdot \tau) d\omega, \\ f(\omega) &= (2\pi)^{-n} \int_{\mathbb{R}^n} c(\tau) \exp(-i\omega \cdot \tau) d\tau. \end{aligned} \tag{2.31}$$

We come to this sections centerpiece - the spectral representation theorem. Intuitively, it takes Bochners theorem one step further and represents a stationary process itself to a Fourier-type stochastic integral, with the integrator being related to the processes spectral distribution.

Theorem 2.5 (Spectral Representation Theorem)**.** Let X be a centered stationary process with spectral distribution $F(\omega)$. Then there exists a stochastic process $Z = (Z(\omega))_{\omega \in \mathbb{R}^n}$ such that

$$X(t) = \int_{\mathbb{R}^n} \exp(i\omega \cdot t) dZ(\omega). \tag{2.32}$$

We call Z the *spectral process* of X. It has centered, orthogonal increments whose variances are given by the increments of $F(\omega)$:

$$\begin{aligned} \mathrm{E}[\Delta Z(\omega)] &= 0, \\ \mathrm{E}[\Delta Z(\omega)\overline{\Delta Z(\nu)}] &= 0, \qquad \text{if } \Delta\omega \cap \Delta\nu = \emptyset, \\ \mathrm{E}[|\Delta Z(\omega)|^2] &= \Delta F(\omega). \end{aligned} \tag{2.33}$$

We want to remark that there exists a *generalized spectral representation theorem*, of which the version above is a direct colloray from. In summary, it states that a stochastic process $(X(t))_t$ over an arbitrary index set T with covariance function $k(s,t)$ emits a spectral representation

$$X(t) = \int_A \phi(t,a) Z(da)$$

if and only if its covariance function can be written as the measure integral of a complex-valued function $\phi(t,a)$ over an arbitrary set A:

$$k(t,s) = \int_A \phi(t,a)\overline{\phi(s,a)} F(da).$$

Here, $F(\Delta a)$ is a bounded measure and $Z(\Delta a)$ an orthogonal, centered random measure, both on measurable subsets of A. These are again linked by the relation

$$\mathrm{E}[Z(\Delta a)\overline{Z(\Delta b)}] = \begin{cases} 0, & \text{if } \Delta a \cap \Delta b = \emptyset \\ F(\Delta a), & \text{if } \Delta a = \Delta b. \end{cases}$$

Theorem 2.5 then follows immediately from the fact that we may write

$$k(t,s) = c(t-s) = \int_{\mathbb{R}^n} \exp(i\omega \cdot t) \exp(-i\omega \cdot s) dF(\omega) \tag{2.34}$$

and the observation that we may represent a measure (random measure) on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ via the increments of a distribution function (random process) over intervals $\Delta\omega$. For more details and other results based on the generalized spectral representation theorem (e.g. the widely used Karhunen–Loève expansion), see Yaglom [1987a], section 26.

We shall prove theorem 2.5 in a more direct way, as the proof holds an import insight for the special case where X is a Gaussian process. It is based on theorem 3.17, Lindgren [2012], where the proof was given for the case $n = 1$.

Proof of Theorem 2.5. Denote by $L^2(\mathbb{P})$ the Hilbert space of square-integrable random variables with inner product

$$\langle U, V \rangle_{L^2(\mathbb{P})} = \mathrm{E}[U\overline{V}].$$

Let $\mathcal{H}(X)$ be the closure (in regards to the induced norm $\|\cdot\|_{L^2(\mathbb{P})}$) of random variables of the form

$$\lambda_1 X(t_1) + ... + \lambda_k X(t_k), \quad t_1, ..., t_k \in \mathbb{R}^n, \lambda_1, ..., \lambda_k \in \mathbb{C}. \tag{2.35}$$

Note that $(\mathcal{H}(X), \langle\cdot\rangle_{L^2(\mathbb{P})})$ is a closed, linear subspace of $L^2(\mathbb{P})$ and thus a Hilbert space itself.

Furthermore, denote by $\mu_F(\cdot)$ the Lebesgue-Stieltjes-measure induced by $F(\omega)$ and let $\mathcal{H}(F) = L^2(\mu_F)$ be the Hilbert space of μ_F-square-integrable functions with inner product

$$\langle f, g \rangle_{L^2(\mu_F)} := \int_{\mathbb{R}^n} f(\omega)\overline{g(\omega)} dF(\omega). \tag{2.36}$$

Our aim is to show that the spaces $\mathcal{H}(X)$ and $\mathcal{H}(F)$ are isometrically isomorphic and construct the spectral process $(Z(\omega))_\omega$ in $\mathcal{H}(X)$ via respective functions in $\mathcal{H}(F)$.

For $t \in \mathbb{R}^n$, denote by ϕ_t the function $\omega \mapsto \exp(i\omega \cdot t)$. By Bochner's theorem, we have for all $s, t \in \mathbb{R}^n$:

$$\langle X(s), X(t)\rangle_{L^2(\mathbb{P})} = \int_{\mathbb{R}^n} \exp(i\omega \cdot s)\overline{\exp(i\omega \cdot t)}dF(\omega) = \langle \phi_s, \phi_t\rangle_{L^2(\mu_F)}. \tag{2.37}$$

and in particular $\|X(t)\|_{L^2(\mathbb{P})} = \|\phi_t\|_{L^2(\mu_F)}$.

Define $\mathcal{I}(X(t)) := \phi_t$. For any random variable U of the form (2.35), let

$$\begin{aligned} \mathcal{I}(U) &= \lambda_1 \mathcal{I}(X(t_1)) + ... + \lambda_k \mathcal{I}(X(t_k)) \\ &= \lambda_1 \phi_{t_1} + ... + \lambda_k \phi_{t_k}. \end{aligned} \tag{2.38}$$

Then it is clear that (2.37) extends to

$$\langle U, V\rangle_{L^2(\mathbb{P})} = \langle \mathcal{I}(U), \mathcal{I}(V)\rangle_{L^2(\mu_F)} \tag{2.39}$$

if U, V are as in (2.35). Now, using the fact that the families $\{U \in \mathcal{H}(X) : U = \sum_{i=1}^k \lambda_i X(t_i)\}$ and $\{f \in \mathcal{H}(F) : f = \sum_{i=1}^k \lambda_i \phi_{t_i}\}$ are dense in their respective Hilbert spaces, it is a standard result that one may uniquely extend $\mathcal{I}$ to a bijective isometry $\mathcal{I} : \mathcal{H}(X) \longrightarrow \mathcal{H}(F)$.

Now, for a fixed $\omega = (\omega_1, ..., \omega_n) \in \mathbb{R}^n$, define the indicator function

$$g_\omega(\xi) := \prod_{i=1}^n \mathbb{1}_{(-\infty, \omega_i]}(\xi_i), \quad \xi \in \mathbb{R}^n. \tag{2.40}$$

Furthermore, for an interval $\Delta\omega := [\omega_1, \omega_1 + \Delta_1] \times ... \times [\omega_n, \omega_n + \Delta_n]$, define

$$\begin{aligned} g_{\Delta\omega}(\xi) := & \, g_{(\omega_1+\Delta_1, ..., \omega_n+\Delta_n)}(\xi) \\ & - \sum_i g_{(\omega_1+\Delta_1, ..., \omega_{i-1}+\Delta_{i-1}, \omega_i, \omega_{i+1}+\Delta_{i+1}, ...\omega_n+\Delta_n)}(\xi) \\ & + \sum_{i \neq j} g_{(\omega_1+\Delta_1, ..., \omega_i, ..., \omega_j, ...\omega_n+\Delta_n)}(\xi) \\ & - ... + (-1)^n g_{(\omega_1, ..., \omega_n)}(\xi). \end{aligned}$$

It holds that

$$g_{\Delta\omega}(\xi) = \prod_{i=1}^n \mathbb{1}_{[\omega_i, \omega_i+\Delta_i]}(\xi_i) = \mathbb{1}_{\Delta\omega}(\xi)$$

and we get that $g_\omega, g_{\Delta\omega} \in \mathcal{H}(F)$ with

$$\begin{aligned} \|g_\omega\|_{L^2(\mu_F)} &= \mu_F((-\infty, \omega_1] \times ... \times (-\infty, \omega_n]) = F(\omega) \\ \|g_{\Delta\omega}\|_{L^2(\mu_F)} &= \mu_F(\Delta\omega) = \Delta F(\omega). \end{aligned}$$

Define the spectral process Z via

$$Z(\omega) = \mathcal{I}^{-1}(g_\omega), \quad \omega \in \mathbb{R}^n. \tag{2.41}$$

From the linearity of $\mathcal{I}$ it follows that Z satisfies $\Delta Z(\omega) = \mathcal{I}^{-1}(g_{\Delta\omega})$ and thus

$$\mathrm{E}[|\Delta Z(\omega)|^2] = \|\Delta Z\|_{L^2(\mathbb{P})} = \|g_{\Delta\omega}\|_{L^2(\mu_F)} = \Delta F(\omega). \tag{2.42}$$

Further, Z has orthogonal increments over disjoint rectangles $\Delta\omega \cap \Delta\nu = \emptyset$:

$$\mathrm{E}[\Delta Z(\omega)\overline{\Delta Z(\nu)}] = \langle g_{\Delta\omega}, g_{\Delta\nu}\rangle_{L^2(\mu_F)} = \int_{\mathbb{R}^n} \mathbb{1}_{\Delta\omega}(\xi)\overline{\mathbb{1}_{\Delta\nu}(\xi)}dF(\xi) = 0. \tag{2.43}$$

It remains to show that

$$X(t) = \int_{\mathbb{R}^n} \exp(i\omega \cdot t)dZ(\omega).$$

Recall that for any interval T

$$\int_T \exp(i\omega \cdot t)dZ(\omega) = \lim_{k\to\infty} \sum_{\Delta\omega\in\mathcal{P}_k} \exp\big(i\omega' \cdot t\big)\Delta Z(\omega) =: \lim_{k\to\infty} S_k(t),$$

where $(\mathcal{P}_k)_k$ is a sequence of sub-interval partitions of T with mesh size going to zero and $\omega' = [\omega_1, ..., \omega_n]$ for an interval $\Delta\omega = [\omega_1, \omega_1 + \Delta_1] \times ... \times [\omega_n, \omega_n + \Delta_n]$.

Note that the limit is taken in $L^2(\mathbb{P})$ and we may thus apply $\mathcal{I}$ on the RHS and pull out the limit to get

$$\begin{aligned} \mathcal{I}\left(\lim_{k\to\infty} S_k(t)\right) &= \lim_{k\to\infty} \sum_{\Delta\omega\in\mathcal{P}_k} \exp\big(i\omega' \cdot t\big)\mathcal{I}(\Delta Z(\omega)) \\ &= \lim_{k\to\infty} \sum_{\Delta\omega\in\mathcal{P}_k} \exp\big(i\omega' \cdot t\big)\mathbb{1}_{\Delta\omega}. \end{aligned} \tag{2.44}$$

Furthermore, the RHS sum is a bounded function that, due to $|\mathcal{P}_k| \to 0$, converges pointwise for any $\xi \in \mathbb{R}^n$:

$$\lim_{k\to\infty} \sum_{\Delta\omega\in\mathcal{P}_k} \exp\big(i\omega' \cdot t\big)\mathbb{1}_{\Delta\omega}(\xi) = \exp(i\xi \cdot t) = \phi_t(\xi).$$

Thus, using dominated convergence in $L^2(\mu_F)$ for the second step, we get

$$\begin{aligned} \mathcal{I}\left(\lim_{k\to\infty} S_k(t)\right) &= \lim_{k\to\infty} \sum_{\Delta\omega\in\mathcal{P}_k} \exp\big(i\omega' \cdot t\big)\mathbb{1}_{\Delta\omega} \\ &= \phi_t \\ &= \mathcal{I}(X(t)), \end{aligned} \tag{2.45}$$

which proves the claim.

□

As mentioned before, the proof above gives an important observation. In the special case that X is a Gaussian process, any random variable $Y \in \mathcal{H}(X)$ is in fact a normal distribution as the $L^2(\mathbb{P})$-limit of Gaussian variables. Noting the vector space structure of $\mathcal{H}(X)$, it is immediate that:

Corollary 2.1. Let X be a stationary Gaussian process. Then its spectral process Z is a Gaussian process itself.

Another special case of interest occurs when X is real-valued:

Lemma 2.4. If X is real-valued, its spectral process is such that

$$\begin{cases} \Delta Z(\omega) \in \mathbb{R}, & \text{if } \Delta\omega \text{ is symmetric around 0,} \\ \Delta Z(\omega) \text{ is proper,} & \text{else.} \end{cases} \tag{2.46}$$

Furthermore, there exist two real-valued processes Z_1, Z_2 such that

$$X(t) = \int \cos(\omega \cdot t) dZ_1(\omega) + \int \sin(\omega \cdot t) dZ_2(\omega). \tag{2.47}$$

Proof. The proof of the one-sided represenation in the case $n = 1$ is given in Lindgren [2012], section 3.3.4. For completeness sake, we give the proof in the two-dimensional case, which will become relevant to us later on.

With slight abuse of notation, it holds that

$$\begin{aligned} X(t) &= \int_{\mathbb{R}^2} \exp(i\omega \cdot t) dZ(\omega) \\ &= \int_{\mathbb{R}_+ \times \mathbb{R}} \cos(\omega \cdot t) \left[dZ(\omega) + dZ(-\omega) \right] + i \int_{\mathbb{R}_+ \times \mathbb{R}} \sin(\omega \cdot t) \left[dZ(\omega) - dZ(-\omega) \right] \end{aligned}$$

Denote by $-\Delta\omega$ the two-dimensional interval gained by mirroring an interval $\Delta\omega$ with respect to $\omega_0 = 0$, and by $\Delta Z(-\omega)$ the respective increment of Z over $-\Delta\omega$. The fact that $X(t)$ is real then implies that for all $\Delta\omega$

$$\begin{aligned} \operatorname{Im}(\Delta Z(\omega) + \Delta Z(-\omega)) = 0, \\ \operatorname{Re}(\Delta Z(\omega) - \Delta Z(-\omega)) = 0. \end{aligned}$$

which gives

$$\overline{\Delta Z(\omega)} = \Delta Z(-\omega). \tag{2.48}$$

This implies $\overline{\Delta Z(\omega)} = \Delta Z(\omega) \in \mathbb{R}$ whenever $\omega_0 \in \Delta\omega$ such that $-\Delta\omega = \Delta\omega$ (or, equivalently, $\Delta\omega$ is a square centered around ω_0). Alternatively, if $\omega_0 \notin \Delta\omega$ it holds $\Delta\omega \cap -\Delta\omega = \emptyset$. Therefore, the orthogonality of increments of Z over disjoint sets gives

$$\begin{aligned} \mathrm{E}[\Delta Z(\omega) \Delta Z(\omega)] &= \mathrm{E}[\Delta Z(\omega) \overline{\Delta Z(-\omega)}] \\ &= 0. \end{aligned}$$

This proves the first claim. For the second claim, define

$$\begin{aligned}\Delta Z_1(\omega) &:= \Delta Z(\omega) + \Delta Z(-\omega) = 2\,\mathrm{Re}(\Delta Z(\omega)) \\ \Delta Z_2(\omega) &:= i(\Delta Z(\omega) - \Delta Z(-\omega)) = -2\,\mathrm{Im}(\Delta Z(\omega)).\end{aligned} \tag{2.49}$$

Returning to the limit $|\Delta\omega| \to 0$, we get

$$X(t) = \int_{\mathbb{R}_+ \times \mathbb{R}} \cos(\omega \cdot t) dZ_1(\omega) + \int_{\mathbb{R}_+ \times \mathbb{R}} \sin(\omega \cdot t) dZ_2(\omega). \tag{2.50}$$

□

Lastly, we note note that it is straight-forward to get a representation (2.32) for non-centered processes X. Let $0 \neq m \in \mathbb{R}^n$ be the mean of $X(t)$ and denote by $\tilde{X}$ the centered-process given by $\tilde{X}(t) = X(t) - m$. Then $\tilde{X}$ is stationary with identical covariance function to X. Let $\tilde{Z}$ be its spectral process and define

$$Z(\omega) := \begin{cases} m, & \text{if } \omega = 0, \\ \tilde{Z}(\omega), & \text{else.} \end{cases}$$

Then Z is a spectral process for X:

$$\begin{aligned}\int_{R^n} \exp(i\omega \cdot t) dZ(\omega) &= \int_{R^n \setminus \{0\}} \exp(i\omega \cdot t) d\tilde{Z}(\omega) + \int_{\{0\}} \exp(i\omega \cdot t) d\tilde{Z}(\omega) \\ &= \tilde{X}(t) + m \\ &= X(t).\end{aligned}$$

Note that $\Delta Z(\omega) = 0$ if and only if $0 \notin \Delta\omega$.

2.2.4. Matérn Processes

This section introduces the Matérn covariance function. It was named after Matérn [1960] by Stein [1999], who popularized its usage in statistics for spatial and spatiotemporal data. Its spectral properties have been studied by Whittle [1954]. We follow the introductions of Rasmussen and Williams [2006], chapter 4, and Lindgren et al. [2011].

The Matérn class is the class of centered Gaussian processes $X = (X(t))_{t \in \mathbb{R}^n}$ with *Matérn covariance function*

$$c(\|s - t\|) := \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|s - t\|)^\nu K_\nu(\kappa \|s - t\|). \tag{2.51}$$

Here, $K_\nu(\cdot)$ denotes the second kind modified Bessel function of order $\nu > 0$ and $\Gamma(\nu)$ the gamma function. The marginal variance is given by σ^2. For example, in the case $\nu = 1$, the fact that $\lim_{r \to 0} r K_1(r) = 1$ gives

$$\lim_{r \to 0} c(r) = \lim_{r \to 0} \sigma^2 (\kappa r) K_1(\kappa r) = \sigma^2.$$

The parameter κ is a scale parameter, often parametrized via $\kappa = \frac{1}{\rho}$, whereas ν controls the smoothness of X. It can be shown that X is k-times mean-squared differentiable[1] if and only if $\nu \geq k$. For a comparison of how these parameters influence the resulting Matérn field, see figure 2.1.

The Matérn covariance is isotropic and therefore stationary. It can be shown that its spectral density is given by

$$f(\xi) = \frac{\sigma^2}{(2\pi)^n}(\kappa^2 + \|\xi\|^2)^{-\alpha}, \quad \alpha = \nu + \tfrac{n}{2}. \tag{2.52}$$

Going on step further one can then show, based on the spectral density and the definition of the *fractional Laplace operator* [2], that the Gaussian processes with Matérn covariance function are the stationary solutions to the fractional stochastic Laplace equation

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X(s) = \dot{W}(s), \quad s \in \mathbb{R}^n, \tag{2.53}$$

where $\dot{W}(s)$ denotes spatial white noise[3]. For proofs of these facts, see Whittle [1954] and Whittle [1963].

It is noteworthy that the Matérn covariance function has interesting links to many widely used covariance functions. A full discussion on these can be found in Rasmussen and Williams [2006], whereas we shall only describe two special cases.

A reparametrization of the spectral density gives

$$f(\xi) = \frac{2^n \pi^{\frac{n}{2}} \Gamma(\nu + \frac{n}{2})(2\nu)^\nu}{\Gamma(\nu)\rho^{2\nu}} \left(\frac{2\nu}{\rho^2} + 4\pi^2 \|\xi\|^2 \right)^{-(\nu + \frac{n}{2})},$$

where $\kappa = \frac{\sqrt{2\nu}}{\rho}$. In the limit $\nu \to \infty$, this converges to the spectral density

$$f_{SE}(\xi) = \left(2\pi\rho^2\right)^{\frac{n}{2}} \exp\left(-2\pi^2\rho^2\|\xi\|^2\right)$$

of the *squared exponential kernel*

$$c_{SE}(r) = \exp\left(-\frac{r^2}{2\rho^2}\right).$$

Furthermore, in the case $\nu = \frac{1}{2}$, it can be shown that the Matérn covariance function equals the *exponential kernel*

$$c_{Exp}(r) = \sigma^2 \exp\left(-\tfrac{r}{\rho}\right). \tag{2.54}$$

[1] A process X is mean-squared differentiable if there exists a process X' such that, for all t, $\lim_{h\to 0} \frac{X(t+h)-X(t)}{h} = X'(t)$ in $L^2(\mathbb{P})$. If X is such that its paths are differentiable almost surely, the pathwise derivative equals the mean-square derivative almost surely. For more details see Lindgren [2012], chapter 2.4.

[2] The fractional Laplace operator is defined via its properties in the frequency domain: $\mathscr{F}\left((\kappa^2 - \Delta)^{\frac{\alpha}{2}}\phi\right)(\xi) = \left(\kappa^2 + \|\xi\|^2\right)^{\frac{\alpha}{2}} \mathscr{F}(\phi)(\xi)$ for any function ϕ such that the right hand side exists.

[3] See section 2.4 for a precise definition.

This kernel, parametrized with marginal variance $\sigma^2 = \frac{1}{2\gamma}$ and scale parameter $\rho = \frac{1}{\gamma}$, is a pointwise approximation of the covariance function of the Ornstein-Uhlenbeck process

$$k_{O.U.}(s,t) = \frac{1}{2\gamma}\left(\exp(-\gamma(t-s)) - \exp(-\gamma(t+s))\right), \quad s < t,$$

for sufficiently large s, t (that is, when the process "has been running for a sufficiently long time"). It is considered one of the most elementary covariance functions in one dimension, due to its property of giving rise to a Markov process.

As its two-dimensional pendant, Whittle [1954] proposed the Matérn covariance function with $\nu = 1$, noting their similar decay behavior and the fact that the discretized version of the corresponding Matérn process X, defined by equation (2.53), is a Markov field on the lattice $\mathbb{Z}^2$ given by

$$\tilde{X}_{r,t} = \tilde{\kappa}(\tilde{X}_{r+1,t} + \tilde{X}_{r-1,t} + \tilde{X}_{r,t+1} + \tilde{X}_{r,t-1}) + \epsilon(r,t), \quad (r,t) \in \mathbb{Z}^2,$$

with scaling parameter $\tilde{\kappa}$ and discrete white noise $\epsilon(r,t)$.

It will be this special case of the Matérn kernel that we shall be concerned with and we refer to it as the *Whittle covariance function*. For the sake of completeness, its covariance

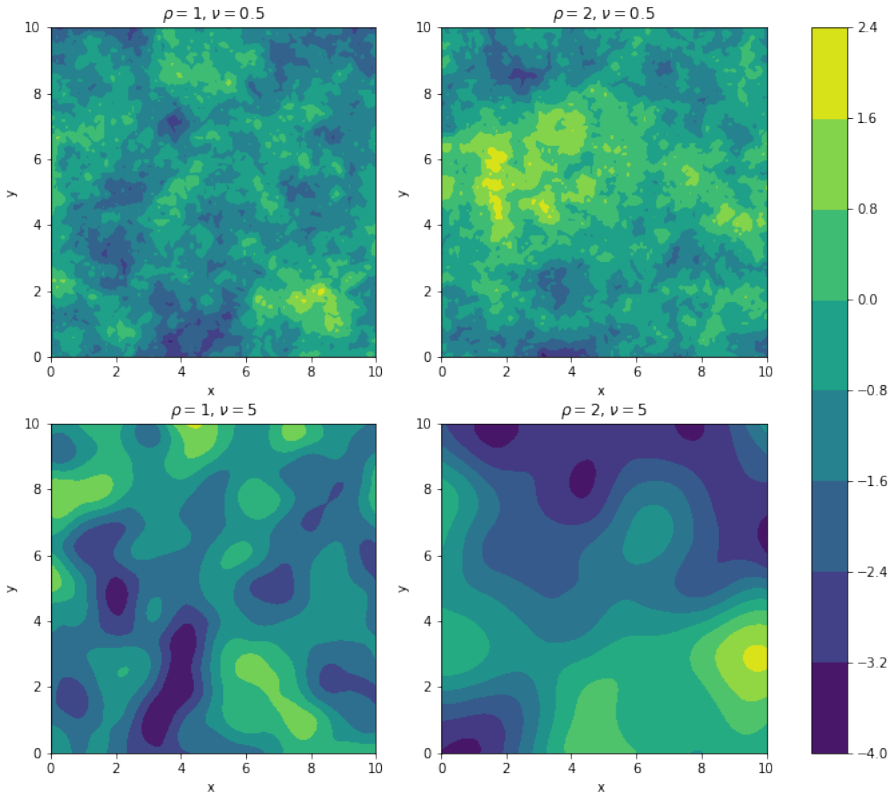

Figure 2.1.: Matérn Fields with differing scale- and smoothness parameters.

function/spectral density pair is given by

$$\begin{aligned} c_{whittle}(\|s-t\|) &= \sigma^2 \left(\kappa\|s-t\|\right) K_1(\kappa\|s-t\|) \text{ and} \\ f_{whittle}(\xi) &= \frac{\sigma^2}{(2\pi)^2}(\kappa^2 + \|\xi\|^2)^{-2} \end{aligned} \tag{2.55}$$

respectively.

2.3. Filtering and Smoothing of Stochastic Differential Equations

This section concerns the filtering and smoothing of (linear) stochastic differential equations. We start off by reviewing solutions to such equations. In the linear case, these are easily obtained in closed form. We then generalize the solutions for complex-valued processes as these will be of interest for us in chapter 3.

Section 2.3.2 then introduces the filtering and smoothing problem and derives the Kalman filter and -smoother as a solution in case of linear dynamics.

2.3.1. Review - Stochastic Differential Equations

Consider a dynamical system $x(t)$ whose behavior in continuous time $t \in \mathbb{R}_+$ is given by the ordinary differential equation

$$\frac{d}{dt}x(t) = f(x(t), t).$$

Assume that at any given moment, the system is perturbed by a stochastic forcing term $\dot{w}(t)$. The forcing term might account for model uncertainty, measurement errors or inherent stochastic nature of the modeled process.

A common assumption is that the noise process $\dot{w} = (\dot{w}(t))_t$ is stationary Gaussian and its realizations at distinct times are independent, in which case one speaks of a *(continuous-time) white noise* process. One may then be inclined to write

$$\frac{d}{dt}X(t) = f(X(t), t) + \sigma(X(t), t)\dot{w}(t), \quad t \in \mathbb{R}_+. \tag{2.56}$$

where $\sigma(x, t)$ is some function that scales noise intensity.

Note, however, that $(\dot{w}(t))_t$ does not exist in the original sense of a stochastic processes. Its proposed properties imply a covariance function that equals a Dirac delta function, which in turn suggests a constant spectral density (akin to physical white light, hence the terminology) and infinite variance for any $\dot{w}(t)$. Other undesired properties include the non-existence of continuous path realization of $\dot{w}$ and its non-measurability (see e.g. Kallianpur and Sundar [2014], 1.1.2 and 1.1.3).

The workaround for this problem is to consider white noise as the informal derivative of standard Brownian motion, formally written as $\dot{w}(t) = \frac{d}{dt}w(t)$, and to interpret (2.56) as an integral equation. A process $(X(t))_t$ is then called a *strong solution* to the *stochastic differential equation (SDE)*

$$dX(t) = f(X(t), t)dt + \sigma(X(t), t)dw(t). \tag{2.57}$$

if it satisfies [4]

$$X(t) = X(0) + \int_0^t f(X(s), s)ds + \int_0^t \sigma(X(s), s)dw(s). \tag{2.58}$$

[4] Of course, other properties it has to satisfy are adaptedness and paths that are almost surely continuous and square-integrable.

The existence of such solutions is ensured by certain growth and continuity assumptions on the drift- and diffusion functions $f(t,x)$ and $\sigma(t,x)$. For a detailed introduction see Kuo [2006], chapter 10 or Kallianpur and Sundar [2014], chapter 6.

One of the simplest stochastic differential equations is given by the *Langevin equation*

$$\begin{cases} dX(t) &= -\gamma X(t)dt + \sigma dw(t), \quad \gamma > 0, \sigma \in \mathbb{R}, \\ X(0) &= X_0. \end{cases} \tag{2.59}$$

The solution X to (2.59) is called an *Ornstein-Uhlenbeck (O.U.) process* and is given by

$$X(t) = \exp(-\gamma t)X_0 + \sigma \int_0^t \exp(-\gamma(t-s))dw(s). \tag{2.60}$$

It is a Markov process with conditional momenta for all $s < t$ given by

$$\begin{aligned} \mathrm{E}[X(t) \mid X(s) = x] &= \exp(-\gamma(t-s))x \\ \mathrm{Var}[X(t) \mid X(s) = x] &= \frac{\sigma^2}{2\gamma}[1 - \exp(-2\gamma(t-s))]. \end{aligned} \tag{2.61}$$

In particular, for any normally distributed X_0, X is a Gaussian process such that X_t converges weakly to a normally distributed random variable

$$X_\infty \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\gamma}\right) \tag{2.62}$$

and the distribution of X_∞ is an invariant measure for X. Proofs of these facts can be found in Kuo [2006], 7.4.5 - 7.4.7.

Unsurprisingly, these properties extend to the complex-valued analogue, as seen in the following two lemmata:

Lemma 2.5. Let $w = w_1 + iw_2$ be a complex Brownian motion and $Z = Z_1 + iZ_2$ be an independent complex random variable. A *complex-valued Ornstein-Uhlenbeck process* is a process $X = (X(t))_t$ that solves the *complex Langevin equation*

$$\begin{cases} dX(t) = -\gamma X(t)dt + \sigma dw(t) \\ X(0) = Z, \end{cases} \tag{2.63}$$

where $\gamma = \gamma_1 + i\gamma_2 \in \mathbb{C}, \gamma_1 \in \mathbb{R}_+, \sigma \in \mathbb{R}$.

It is given by

$$X(t) = \exp(-\gamma t)Z + \sigma \int_0^t \exp(-\gamma(t-s))dw(s). \tag{2.64}$$

Proof. Write $X(t) = X_1(t) + iX_2(t)$ and let $\tilde{\boldsymbol{X}}(t) = [X_1(t), X_2(t)]$ be its $\mathbb{R}^2$-valued pendant. Likewise, define $\tilde{\boldsymbol{w}}(t) = [w_1(t), w_2(t)]$ and $\tilde{\boldsymbol{Z}} = [Z_1, Z_2]$. Then $\tilde{\boldsymbol{X}}(t)$ solves the Langevin equation

$$\begin{cases} d\tilde{\boldsymbol{X}}(t) = -\boldsymbol{\Gamma}\tilde{\boldsymbol{X}}(t)dt + \boldsymbol{\Sigma}d\tilde{\boldsymbol{w}}(t) \\ \tilde{\boldsymbol{X}}(0) = \tilde{\boldsymbol{Z}} \end{cases} \tag{2.65}$$

with $\boldsymbol{\Sigma} := \frac{\sigma}{\sqrt{2}}\boldsymbol{E_2}$, $\boldsymbol{E_2}$ denoting the 2×2-identity matrix, and

$$\boldsymbol{\Gamma} := \begin{pmatrix} \gamma_1 & -\gamma_2 \\ \gamma_2 & \gamma_1 \end{pmatrix}.$$

Define $\boldsymbol{\phi}(t, \boldsymbol{x}) := \exp(\boldsymbol{\Gamma} t)\boldsymbol{x}$, where $\exp(\boldsymbol{\Gamma} t)$ is to be understood as the matrix exponential function. Then

$$\begin{aligned} \partial_t \boldsymbol{\phi}(t, \boldsymbol{x}) &= \boldsymbol{\Gamma} \exp(\boldsymbol{\Gamma} t)\boldsymbol{x} \\ \partial_{x_i} \boldsymbol{\phi}(t, \boldsymbol{x}) &= \exp(\boldsymbol{\Gamma} t)\boldsymbol{e}_i \\ \partial^2_{x_i, x_j} \boldsymbol{\phi}(t, \boldsymbol{x}) &= 0, \end{aligned}$$

where $i, j \in \{1, 2\}$ and $\boldsymbol{e}_i$ denotes the i-th standard euclidean basis vector.

For a matrix $\boldsymbol{A}$, denote by $\boldsymbol{A}_{i,:}$ and $\boldsymbol{A}_{:,i}$ the i-th row and column respectively. Note that $\sum_i \boldsymbol{A}_{:,i}\boldsymbol{B}_{i,:} = \boldsymbol{AB}$. Itô's lemma (see theorem A.1) gives

$$\begin{aligned} \exp(\boldsymbol{\Gamma} t)\tilde{\boldsymbol{X}}(t) &= \tilde{\boldsymbol{X}}(0) + \int_0^t \boldsymbol{\Gamma} \exp(\boldsymbol{\Gamma} s)\tilde{\boldsymbol{X}}(s) ds + \sum_{i=1}^{2} \int_0^t \exp(\boldsymbol{\Gamma} s)\boldsymbol{e}_i d\tilde{\boldsymbol{X}}_i(s) \\ &= \tilde{\boldsymbol{Z}} + \int_0^t \boldsymbol{\Gamma} \exp(\boldsymbol{\Gamma} s)\tilde{\boldsymbol{X}}(s) ds + \sum_{i=1}^{2} \int_0^t \exp(\boldsymbol{\Gamma} s)_{:,i}(-\boldsymbol{\Gamma})_{i,:}\tilde{\boldsymbol{X}}(s) ds \\ &\quad + \sum_{i=1}^{2} \int_0^t \exp(\boldsymbol{\Gamma} s)_{:,i}\boldsymbol{\Sigma}_{i,:} d\tilde{\boldsymbol{w}}(s) \\ &= \tilde{\boldsymbol{Z}} + \int_0^t \exp(\boldsymbol{\Gamma} s)\boldsymbol{\Sigma} d\tilde{\boldsymbol{w}}(s), \end{aligned} \tag{2.66}$$

where we have used the fact that $\boldsymbol{\Gamma}$ and $\exp(\boldsymbol{\Gamma} s)$ are diagonalizable with the same eigenvectors and therefore commute. Thus, the solution to 2.65 is given by

$$\tilde{\boldsymbol{X}}(t) = \exp(-\boldsymbol{\Gamma} t)\tilde{\boldsymbol{Z}} + \boldsymbol{\Sigma} \int_0^t \exp(-\boldsymbol{\Gamma}(t - s)) d\tilde{\boldsymbol{w}}(s). \tag{2.67}$$

To translate this solution back to the complex domain, note that it holds

$$\begin{aligned} \exp(-\boldsymbol{\Gamma} t) &= \exp\left(\begin{pmatrix} -\gamma_1 & \gamma_2 \\ -\gamma_2 & -\gamma_1 \end{pmatrix} t\right) \\ &= \exp\left(-\gamma_1 t \boldsymbol{E_2} + \gamma_2 t \boldsymbol{F_2}\right) \\ &= \exp\left(-\gamma_1 t \boldsymbol{E_2}\right) \exp\left(\gamma_2 t \boldsymbol{F_2}\right). \end{aligned}$$

where $\boldsymbol{F_2} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and we used that $\boldsymbol{E_2}$ and $\boldsymbol{F_2}$ commute.

Recall that for any diagonal matrix $\boldsymbol{D} = \mathbf{diag}[a, b]$, it holds

$$\exp(t\boldsymbol{D}) = \mathbf{diag}[\exp(ta), \exp(tb)] \tag{2.68}$$

and therefore

$$\exp\left(-t\gamma_1 \boldsymbol{E_2}\right) = \begin{pmatrix} \exp(-\gamma_1 t) & 0 \\ 0 & \exp(-\gamma_1 t) \end{pmatrix}.$$

To compute $\exp(t\gamma_2 \boldsymbol{F_2})$, note that the eigendecomposition of $\gamma_2 \boldsymbol{F_2}$ is given by

$$\gamma_2 \boldsymbol{F_2} = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -i\gamma_2 & 0 \\ 0 & i\gamma_2 \end{pmatrix} \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}^{-1}.$$

A simple calculation then shows that

$$\begin{aligned} \exp(t\gamma_2 \boldsymbol{F_2}) &= \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \exp(-i\gamma_2 t) & 0 \\ 0 & \exp(i\gamma_2 t) \end{pmatrix} \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \cos(\gamma_2 t) & \sin(\gamma_2 t) \\ -\sin(\gamma_2 t) & \cos(\gamma_2 t) \end{pmatrix}. \end{aligned}$$

In total, we get

$$\exp(-\boldsymbol{\Gamma} t) = \begin{pmatrix} \exp(-\gamma_1 t)\cos(\gamma_2 t) & \exp(-\gamma_1 t)\sin(\gamma_2 t) \\ -\exp(-\gamma_1 t)\sin(\gamma_2 t) & \exp(-\gamma_1 t)\cos(\gamma_2 t) \end{pmatrix},$$

whose matrix-vector multiplication in $\mathbb{R}^2$ corresponds to multiplication in $\mathbb{C}$ with the scalar

$$\exp(-(\gamma_1 + i\gamma_2)t) = \exp(-\gamma_1 t)\cos(\gamma_2 t) - i\exp(-\gamma_1 t)\sin(\gamma_2 t).$$

From this and definition 2.9, we conclude that the solution to equation (2.63) is given by

$$X(t) = \exp(-\gamma t)Z + \sigma \int_0^t \exp(-\gamma(t-s))dw(s). \tag{2.69}$$

□

Lemma 2.6. Let $X = (X(t))_t$ be a complex Ornstein-Uhlenbeck process given by (2.64), with independent initial value $Z \sim \mathcal{CN}(\mu_0, \sigma_0^2)$.

Then X is a proper complex Gaussian process with mean and covariance function given by

$$\begin{aligned} m(t) &= \exp(-\gamma t)\mu_0 \\ k(s,t) &= \exp(-\gamma s)\overline{\exp(-\gamma t)}\left[\sigma_0^2 + \tfrac{\sigma^2}{2\gamma_1}(\exp(2\gamma_1 s) - 1)\right], \quad s \le t. \end{aligned} \tag{2.70}$$

Furthermore, X is a Markov process and converges in distribution to a proper complex Gaussian variable $X_\infty \sim \mathcal{CN}(0, \frac{\sigma^2}{2\gamma_1})$, whose distribution is an invariant measure for the process.

Proof. Recall that $(I(t))_t = \left(\int_0^t \exp(\gamma u) dw(u)\right)_t$ is a centered, proper complex Gaussian process with covariance function

$$\begin{aligned} k_I(s,t) &= \int_0^s |\exp(\gamma u)|^2 du \\ &= \int_0^s \exp(2\gamma_1 u) du \\ &= \tfrac{1}{2\gamma_1}\left[\exp(2\gamma_1 s) - 1\right], \quad s \le t. \end{aligned} \tag{2.71}$$

Thus, $X(t) = \exp(-\gamma t)Z + \sigma \exp(-\gamma t) I(t)$ is both proper and complex Gaussian with mean function

$$\begin{aligned} m(t) &= \mathrm{E}[\exp(-\gamma t)Z + \sigma \exp(-\gamma t) I(t)] \\ &= \exp(-\gamma t)\mu_0 \end{aligned} \tag{2.72}$$

and covariance function

$$\begin{aligned} k(s,t) &= \mathrm{Cov}\left[\exp(-\gamma s)Z + \sigma \exp(-\gamma s) I(s), \exp(-\gamma t)Z + \sigma \exp(-\gamma t) I(t)\right] \\ &= \mathrm{Cov}\left[\exp(-\gamma s)Z, \exp(-\gamma t)Z\right] + \mathrm{Cov}\left[\sigma \exp(-\gamma s) I(s), \sigma \exp(-\gamma t) I(t)\right] \\ &= \exp(-\gamma s)\overline{\exp(-\gamma t)}\sigma_0^2 + \exp(-\gamma s)\overline{\exp(-\gamma t)}\tfrac{\sigma^2}{2\gamma_1}\left(\exp(2\gamma_1 s) - 1\right) \\ &= \exp(-\gamma s)\overline{\exp(-\gamma t)}\left[\sigma_0^2 + \tfrac{\sigma^2}{2\gamma_1}\left(\exp(2\gamma_1 s) - 1\right)\right], \quad s \le t. \end{aligned} \tag{2.73}$$

In particular, we have

$$\begin{aligned} \mathrm{Var}[X(t)] &= |\exp(-\gamma t)|^2 \left[\sigma_0^2 + \tfrac{\sigma^2}{2\gamma_1}\left(\exp(2\gamma_1 t) - 1\right)\right] \\ &= \exp(-2\gamma_1 t)\left[\sigma_0^2 + \tfrac{\sigma^2}{2\gamma_1}\left(\exp(2\gamma_1 t) - 1\right)\right] \\ &= \exp(-2\gamma_1 t)\sigma_0^2 + \tfrac{\sigma^2}{2\gamma_1}\left(1 - \exp(-2\gamma_1 t)\right). \end{aligned} \tag{2.74}$$

Note that with $\gamma_1 > 0$, it holds

$$\begin{aligned} \lim_{t\to\infty} m(t) &= \lim_{t\to\infty} \exp(-\gamma t)\mu_0 = 0 \\ \lim_{t\to\infty} \mathrm{Var}[X(t)] &= \lim_{t\to\infty} \exp(-2\gamma_1 t)\sigma_0^2 + \tfrac{\sigma^2}{2\gamma_1}\left(1 - \exp(-2\gamma_1 t)\right) \\ &= \tfrac{\sigma^2}{2\gamma_1}, \end{aligned}$$

from which we can conclude that $X(t)$ converges in distribution to a proper Gaussian variable $X_\infty \sim \mathcal{CN}(0, \frac{\sigma^2}{2\gamma_1})$. In fact, X_∞ is an invariant measure for X; if $Z \overset{d}{=} X_\infty$, then we have for all $s \le t$

$$\begin{aligned} k(s,t) &= \tfrac{\sigma^2}{2\gamma_1}\exp(-\gamma_1(t-s) + i\gamma_2(t-s)) \\ &= \tfrac{\sigma^2}{2\gamma_1}\exp\left(\overline{-\gamma(t-s)}\right). \end{aligned} \tag{2.75}$$

In particular, for all t it holds

$$\begin{aligned} m(t) &= 0, \\ k(t,t) &= \tfrac{\sigma^2}{2\gamma_1}, \end{aligned}$$

concluding that $X(t) \overset{d}{=} X_\infty$ for all t.

Lastly, the fact that X is a Markov process is clear from writing

$$\begin{aligned} X(t) &= \exp(-\gamma t)Z + \sigma \int_0^t \exp(-\gamma(t-u))dw(u) \\ &= \exp(-\gamma(t-s))X(s) + \sigma \int_s^t \exp(-\gamma(t-u))dw(u) \end{aligned} \tag{2.76}$$

for any $s \leq t$, noting the independence of $X(s), (w(u))_{u>s}$.

We get

$$\begin{aligned} \mathrm{E}[X(t) \mid X(s) = x] &= \exp(-\gamma(t-s))x \\ \mathrm{Var}[X(t) \mid X(s) = x] &= \mathrm{Var}[\exp(-\gamma(t-s))X(s) \mid X(s) = x] \\ &\quad + \mathrm{Var}\left[\sigma \int_s^t \exp(-\gamma(t-u))dw(u) \mid X(s) = x\right] \\ &= 0 + \sigma^2 \exp(-2\gamma_1 t)\mathrm{Var}\left[\int_s^t \exp(\gamma u)dw(u)\right] \\ &= \sigma^2 \exp(-2\gamma_1 t) \int_s^t \exp(2\gamma_1 u)du \\ &= \tfrac{\sigma^2}{2\gamma_1} \exp(-2\gamma_1 t)\left[\exp(2\gamma_1 t) - \exp(2\gamma_1 s)\right] \\ &= \tfrac{\sigma^2}{2\gamma_1}\left[1 - \exp(-2\gamma_1(t-s))\right], \end{aligned} \tag{2.77}$$

which, to no surprise, equals (2.72) and (2.74) with deterministic initial condition $X(s) = x$ and time lag $t - s$. □

The Langevin equation is a special case of a general linear stochastic differential equation

$$d\boldsymbol{X}(t) = \boldsymbol{F}(t)\boldsymbol{X}(t)dt + \boldsymbol{Q}(t)d\boldsymbol{w}(t), \tag{2.78}$$

where $\boldsymbol{X} \in \mathbb{R}^n$ denotes a system state, $\boldsymbol{F}(\cdot) \in \mathbb{R}^{n\times n}$ is a system dynamics matrix, $\boldsymbol{w}(\cdot)$ is a d-dimensional Brownian Motion, induced into the system via a noise input matrix $\boldsymbol{Q}(\cdot) \in \mathbb{R}^{n\times d}$ and $\boldsymbol{X}(t_0)$ denotes the initial state.

Denote by $\boldsymbol{P}(t,s)$ the *state transition matrix* of the system, defined by its properties

$$\begin{aligned} \partial_t \boldsymbol{P}(t,s) &= \boldsymbol{F}(t)\boldsymbol{P}(t,s) \\ \partial_s \boldsymbol{P}(t,s) &= -\boldsymbol{P}(t,s)\boldsymbol{F}(s) \\ \boldsymbol{P}(t,s) &= \boldsymbol{P}(t,u)\boldsymbol{P}(u,s) \\ \boldsymbol{P}(s,t)^{-1} &= \boldsymbol{P}(t,s). \end{aligned} \tag{2.79}$$

In the case of the Langevin equation, $\boldsymbol{P}(t,s)$ is given by the matrix exponential $\exp(-\boldsymbol{\gamma}(t-s))$. Analogue to the proof seen above, an application of Itô's lemma to $\phi(t,\boldsymbol{X}(t)) := \boldsymbol{P}(t_0,t)\boldsymbol{X}(t)$ shows that the solution to (2.78) is given by a Markov process

$$\boldsymbol{X}(t) = \boldsymbol{P}(t,t_0)\boldsymbol{X}(t_0) + \int_{t_0}^{t} \boldsymbol{P}(t,s)\boldsymbol{Q}(s)d\boldsymbol{w}(s). \tag{2.80}$$

For more details see Särkkä and Solin [2019], chapter 4.3.

2.3.2. Bayesian Filtering & Smoothing

Bayesian filtering concerns the problem of estimating the state of a dynamically evolving stochastic process, based on incoming noisy and possibly incomplete process measurements. Its offline pendant, the filtering problem, concerns the state estimation given all, including future, observations.

In the special case of linear dynamics and Gaussian noise terms, both problems can be solved in closed form via the *Kalman filter* and *Kalman smoother*. Their following introduction is based on Maybeck [1979], chapter 5, and Särkkä and Solin [2019], chapter 10.

Assume we are concerned with estimating the state of a process $\boldsymbol{X}$, evolving according to the linear SDE

$$d\boldsymbol{X}(t) = \boldsymbol{F}(t)\boldsymbol{X}(t)dt + \boldsymbol{Q}(t)d\boldsymbol{w}(t), \tag{2.81}$$

where $\boldsymbol{F}(\cdot) \in \mathbb{R}^{n\times n}$ is a system dynamics matrix, $\boldsymbol{w}(\cdot)$ is a d-dimensional Brownian Motion, induced into the system via a noise input matrix $\boldsymbol{Q}(\cdot) \in \mathbb{R}^{n\times d}$ and the initial state is a Gaussian variable $\boldsymbol{X}(t_0) \sim \mathcal{N}_n(\boldsymbol{m_0}, \boldsymbol{K_0})$.

We further assume the measurements $\boldsymbol{Z}(t_i) \in \mathbb{R}^m$ at times $t_1 \leq t_2 \leq .... \leq t_I$ to be given by

$$\boldsymbol{Z}(t_i) = \boldsymbol{H}(t_i)\boldsymbol{X}(t_i) + \boldsymbol{\epsilon}(t_i), \tag{2.82}$$

where $\boldsymbol{H}(\cdot) \in \mathbb{R}^{m\times n}$ is the measurement matrix and the measurement noise $\boldsymbol{\epsilon}(\cdot)$ is a discrete-time Gaussian white noise process with covariance matrix given by $\boldsymbol{R}(\cdot)$.

The Kalman Filter

The aim of any filtering algorithm is to construct an 'optimal' estimator of the true process state $\boldsymbol{X}(t)$ at any time t by incorporating knowledge of the system dynamics and previous process observations $\boldsymbol{Z}(t_1), ..., \boldsymbol{Z}(t_i)$, $t_i \leq t$. What makes an estimator 'optimal' is a matter of definition. However, given the fact that interdependencies are linear and stochastic terms are Gaussian, the Kalman filter can be fully formulated in closed-form conditional distributions.

Denote by $\boldsymbol{Z}(t_{1:i}) := [\boldsymbol{Z}(t_1), ..., \boldsymbol{Z}(t_i)]$ the vector of all available measurements up to and including $\boldsymbol{Z}(t_i)$.

In each iteration, the Kalman Filter is formulated in two steps. Step one estimates the state propagation from time t_i to t_{i+1}, based on a current estimation $\boldsymbol{X}(t_i) \mid \boldsymbol{Z}(t_{1:i})$ and the known dynamics of the system. The second step then updates the estimation $\boldsymbol{X}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})$ by including the latest measurement $\boldsymbol{Z}(t_{i+1})$.

The prediction steps assumes an estimate of the current Gaussian state $\boldsymbol{X}(t_i) \mid \boldsymbol{Z}(t_{1:i})$. Denote by $\boldsymbol{m}(t_i)$ and $\boldsymbol{K}(t_i)$ its mean and covariance matrix and by

$$\begin{aligned}\boldsymbol{m}^-(t_{i+1}) &= \mathrm{E}[\boldsymbol{X}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})] \\ \boldsymbol{K}^-(t_{i+1}) &= \mathrm{E}[\boldsymbol{X}(t_{i+1})\boldsymbol{X}^T(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})].\end{aligned}$$

the mean and covariance matrix of the propagated state prediction $\boldsymbol{X}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})$.

Recall that $\boldsymbol{X}$ is a Markov process given by

$$\boldsymbol{X}(t) = \boldsymbol{P}(t, t_0)\boldsymbol{X}(t_0) + \int_{t_0}^{t} \boldsymbol{P}(t, \tau)\boldsymbol{Q}(\tau)d\boldsymbol{w}(\tau),$$

where $\boldsymbol{P}(\cdot, \cdot)$ is the state-transition matrix defined in (2.79). As such, it follows that

$$\boldsymbol{X}(t_{i+1}) = \boldsymbol{P}(t_{i+1}, t_i)\boldsymbol{X}(t_i) + \int_{t_i}^{t_{i+1}} \boldsymbol{P}(t_{i+1}, \tau)\boldsymbol{Q}(\tau)d\boldsymbol{w}(\tau),$$

from which it is immediate that the propagation of $\boldsymbol{m}(t)$ and $\boldsymbol{K}(t)$ are given by

$$\begin{aligned}\boldsymbol{m}^-(t_{i+1}) &= \boldsymbol{P}(t_{i+1}, t_i)\boldsymbol{m}(t_i) \\ \boldsymbol{K}^-(t_{i+1}) &= \boldsymbol{P}(t_{i+1}, t_i)\boldsymbol{K}(t_i)\boldsymbol{P}^T(t_{i+1}, t_i) \\ &\quad + \int_{t_i}^{t_{i+1}} \boldsymbol{P}(t_{i+1}, \tau)\boldsymbol{Q}(\tau)\boldsymbol{Q}^T(\tau)\boldsymbol{P}^T(t_{i+1}, \tau)d\tau \\ &=: \boldsymbol{P}(t_{i+1}, t_i)\boldsymbol{K}(t_i)\boldsymbol{P}^T(t_{i+1}, t_i) + \boldsymbol{L}(t_{i+1}, t_i).\end{aligned} \tag{2.83}$$

The second step aims to update the estimations $\boldsymbol{m}^-(t_{i+1})$ and $\boldsymbol{K}^-(t_{i+1})$ based on a measurement $\boldsymbol{Z}(t_{i+1})$. The Bayes theorem gives

$$\begin{aligned}f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i+1})} &= \frac{f_{\boldsymbol{Z}(t_{i+1})|\boldsymbol{Z}(t_{1:i}),\boldsymbol{X}(t_{i+1})} f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}}{f_{\boldsymbol{Z}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}} \\ &= \frac{f_{\boldsymbol{Z}(t_{i+1})|\boldsymbol{X}(t_{i+1})} f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}}{f_{\boldsymbol{Z}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}}.\end{aligned} \tag{2.84}$$

Note that the second numerator is computed in step one. It is easy to see that the first numerator $\boldsymbol{Z}(t_{i+1}) \mid \boldsymbol{X}(t_{i+1})$ is Gaussian with mean $\boldsymbol{H}(t_{i+1})\boldsymbol{X}(t_{i+1})$ and covariance matrix $\boldsymbol{R}(t_{i+1})$.

Furthermore, the denominator is the sum of the two independent Gaussian variables $[\boldsymbol{H}(t_{i+1})\boldsymbol{X}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})]$ and $[\boldsymbol{\epsilon}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})]$ and thus Gaussian with momenta

$$\begin{aligned}\mathrm{E}[\boldsymbol{Z}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})] &= \boldsymbol{H}(t_{i+1})\boldsymbol{m}^-(t_{i+1}) \\ \mathrm{E}[\boldsymbol{Z}(t_{i+1})\boldsymbol{Z}^T(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i})] &= \boldsymbol{H}(t_{i+1})\boldsymbol{K}^-(t_{i+1})\boldsymbol{H}^T(t_{i+1}) + \boldsymbol{R}(t_{i+1}).\end{aligned} \tag{2.85}$$

From this it is not immediate that $\boldsymbol{X}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:i+1})$ is also Gaussian. However, plugging in the respective normal densities into (2.84), one can show with the help of the matrix inversion lemma (e.g. Maybeck [1979], page 213) that $f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i+1})}$ indeed has a closed Gaussian density form with

$$\begin{aligned}\boldsymbol{m}(t_{i+1}) &= \left[\boldsymbol{K}^{-}(t_{i+1})^{-1} + \boldsymbol{H}^{T}(t_{i+1})\boldsymbol{R}^{-1}(t_{i+1})\boldsymbol{H}(t_{i+1})\right]^{-1} \\ &\quad \cdot \left[\boldsymbol{K}^{-}(t_{i+1})^{-1}\boldsymbol{m}^{-}(t_{i+1}) + \boldsymbol{H}^{T}(t_{i+1})\boldsymbol{R}^{-1}(t_{i+1})\boldsymbol{Z}(t_{i+1})\right] \\ \boldsymbol{K}(t_{i+1}) &= \left[\boldsymbol{K}^{-}(t_{i+1})^{-1} + \boldsymbol{H}^{T}(t_{i+1})\boldsymbol{R}^{-1}(t_{i+1})\boldsymbol{H}(t_{i+1})\right]^{-1}.\end{aligned}$$

These can be reformulated with the help of the Kalman gain $\boldsymbol{G}(t_{i+1})$, defined as:

$$\boldsymbol{G}(t_{i+1}) := \boldsymbol{K}^{-}(t_{i+1})\boldsymbol{H}^{T}(t_{i+1})\left[\boldsymbol{H}(t_{i+1})\boldsymbol{K}^{-}(t_{i+1})\boldsymbol{H}^{T}(t_{i+1}) + \boldsymbol{R}(t_{i+1})\right]^{-1}. \tag{2.86}$$

The updates then take on the final form

$$\begin{aligned}\boldsymbol{m}(t_{i+1}) &= \boldsymbol{m}^{-}(t_{i+1}) + \boldsymbol{G}(t_{i+1})[\boldsymbol{Z}(t_{i+1}) - \boldsymbol{H}(t_{i+1})\boldsymbol{m}^{-}(t_{i+1})] \\ \boldsymbol{K}(t_{i+1}) &= \boldsymbol{K}^{-}(t_{i+1}) - \boldsymbol{G}(t_{i+1})\boldsymbol{H}(t_{i+1})\boldsymbol{K}^{-}(t_{i+1}).\end{aligned} \tag{2.87}$$

The complete Kalman filter is summarized in algorithm 2.1. Note that the final formulation of the update step only requires inversion of matrices in $\mathbb{R}^{m\times m}$ rather than $\mathbb{R}^{n\times n}$, making computation more efficient in many applications in which $m << n$. However, due to the matrix multiplications in the prediction step, the overall computational complexity class for a single Kalman filter iteration remains $\mathcal{O}(n^3)$.

Further note that the two steps inductively show that, given a Gaussian initial state, all subsequent state estimations are Gaussian. This justifies the assumption made in step one.

Algorithm 2.1: Kalman Filter

Input: data $(\boldsymbol{Z}(t_i)), i = 1, ..., I$, model matrix functions $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{L}, \boldsymbol{H}, \boldsymbol{R}$, prior state $\boldsymbol{m}_0, \boldsymbol{K}_0$.

Output: predicted and updated mean and covariance matrix estimates $\boldsymbol{m}^{-}(t_i), \boldsymbol{K}^{-}(t_i), \boldsymbol{m}(t_i), \boldsymbol{K}(t_i)$, i = 1,...,I.

Initialization: $\boldsymbol{m}(t_0) = \boldsymbol{m}_0, \boldsymbol{K}(t_0) = \boldsymbol{K}_0$.

for $i = 0, ..., I-1$ **do**

Prediction:

$\boldsymbol{K}^{-}(t_{i+1}) = \boldsymbol{P}(t_{i+1}, t_i)\boldsymbol{K}(t_i)\boldsymbol{P}^{T}(t_{i+1}, t_i) + \boldsymbol{L}(t_{i+1}, t_i)$;

$\boldsymbol{m}^{-}(t_{i+1}) = \boldsymbol{P}(t_{i+1}, t_i)\boldsymbol{m}(t_i)$;

Update:

$\boldsymbol{G}(t_{i+1}) := \boldsymbol{K}^{-}(t_{i+1})\boldsymbol{H}^{T}(t_{i+1})\left[\boldsymbol{H}(t_{i+1})\boldsymbol{K}^{-}(t_{i+1})\boldsymbol{H}^{T}(t_{i+1}) + \boldsymbol{R}(t_{i+1})\right]^{-1}$;

$\boldsymbol{K}(t_{i+1}) = \boldsymbol{K}^{-}(t_{i+1}) - \boldsymbol{G}(t_{i+1})\boldsymbol{H}(t_{i+1})\boldsymbol{K}^{-}(t_{i+1})$;

$\boldsymbol{m}(t_{i+1}) = \boldsymbol{m}^{-}(t_{i+1}) + \boldsymbol{G}(t_{i+1})[\boldsymbol{Z}(t_{i+1}) - \boldsymbol{H}(t_{i+1})\boldsymbol{m}^{-}(t_{i+1})]$;

end

The Kalman Smoother

The smoothing problem can be considered the offline pendant to the filtering problem. Instead of updating state beliefs with incoming measurements, we aim to retrospectively estimate the system states $\boldsymbol{X}(t_i) \mid \boldsymbol{Z}(t_{1:I})$, $i = 1, ..., I$, given all available "future" measurements. Similar to the Kalman filter, the state estimations of the Kalman smoother can be formulated in closed-form.

It holds, using the Markov property in the first and Bayes theorem in the second step:

$$\begin{aligned} f_{\boldsymbol{X}(t_i)|\boldsymbol{X}(t_{i+1}),\boldsymbol{Z}(t_{1:I})} &= f_{\boldsymbol{X}(t_i)|\boldsymbol{X}(t_{i+1}),\boldsymbol{Z}(t_{1:i})} \\ &= \frac{f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{X}(t_i),\boldsymbol{Z}(t_{1:i})} f_{\boldsymbol{X}(t_i)|\boldsymbol{Z}(t_{1:i})}}{f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}} \\ &= \frac{f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{X}(t_i)} f_{\boldsymbol{X}(t_i)|\boldsymbol{Z}(t_{1:i})}}{f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}}. \end{aligned}$$

Therefore, the conditional density $f_{\boldsymbol{X}(t_i)|\boldsymbol{Z}(t_{1:I})}$ can be computed by marginalizing $\boldsymbol{X}(t_{i+1}) \mid \boldsymbol{Z}(t_{1:I})$:

$$\begin{aligned} f_{\boldsymbol{X}(t_i)|\boldsymbol{Z}(t_{1:I})} &= \int f_{\boldsymbol{X}(t_i),\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:I})} d\boldsymbol{x}_{t_{i+1}} \\ &= \int f_{\boldsymbol{X}(t_i)|\boldsymbol{X}(t_{i+1}),\boldsymbol{Z}(t_{1:I})} f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:I})} d\boldsymbol{x}_{t_{i+1}} \\ &= f_{\boldsymbol{X}(t_i)|\boldsymbol{Z}(t_{1:i})} \int \frac{f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{X}(t_i)}}{f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}} f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:I})} d\boldsymbol{x}_{t_{i+1}}. \end{aligned} \tag{2.88}$$

Note that $f_{\boldsymbol{X}(t_i)|\boldsymbol{Z}(t_{1:i})}$ and $f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:i})}$ are the update and prediction steps given in the filtering problem, whereas $f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{X}(t_i)}$ is specified by the Markovian model and $f_{\boldsymbol{X}(t_{i+1})|\boldsymbol{Z}(t_{1:I})}$ is the smoothing density given by the previous step. The general smoothing algorithm is therefore initiated with the updated density $f_{\boldsymbol{X}(t_I)|\boldsymbol{Z}(t_{1:I})}$ given by the last Kalman filter step and each iteration propagates the smoothing density backwards in time from t_{i+1} to t_i.

In the case of a linear SDE above, it can be shown (see e.g. Särkkä and Solin [2019], theorem 8.2) that the smoothing densities remain Gaussian with mean $\boldsymbol{m}^s(t_i)$ and covariance matrix $\boldsymbol{K}^s(t_i)$ iteratively given by

$$\begin{aligned} \boldsymbol{m}^{\boldsymbol{s}}(t_i) &= \boldsymbol{m}(t_i) + \boldsymbol{G}^{\boldsymbol{s}}(t_{i+1}, t_i) \left[\boldsymbol{m}^{\boldsymbol{s}}(t_{i+1}) - \boldsymbol{m}^{-}(t_{i+1})\right] \\ \boldsymbol{K}^{\boldsymbol{s}}(t_i) &= \boldsymbol{K}(t_i) + \boldsymbol{G}^{\boldsymbol{s}}(t_{i+1}, t_i) \left[\boldsymbol{K}^{\boldsymbol{s}}(t_{i+1}) - \boldsymbol{K}^{-}(t_{i+1})\right] \boldsymbol{G}^{\boldsymbol{s}}(t_{i+1}, t_i)^T, \end{aligned} \tag{2.89}$$

where $\boldsymbol{m}(t_i), \boldsymbol{K}(t_i)$ and $\boldsymbol{m}^{-}(t_{i+1}), \boldsymbol{K}^{-}(t_{i+1})$ are the updated and predicted mean and covariance matrices given by the Kalman filter and the smoothing gain $\boldsymbol{G}^{\boldsymbol{s}}$ is defined via

$$\boldsymbol{G}^{\boldsymbol{s}}(t_{i+1}, t_i) := \boldsymbol{K}(t_i) \boldsymbol{P}(t_{i+1}, t_i)^T \boldsymbol{K}^{-}(t_{i+1})^{-1}.$$

The resulting *Kalman smoother* is summarized in algorithm 2.2.

Remark 2.2. Both the Kalman filter and smoother can be formulated to include estimations in continuous time, commonly referred to as 'continuous-discrete' filtering and smoothing.

This is done by including 'measurement-free steps' in the Kalman filter, which propagate forward the updated estimates $\boldsymbol{m}(t_i), \boldsymbol{K}(t_i)$ between measurements $\boldsymbol{Z}(t_i)$ and $\boldsymbol{Z}(t_{i+1})$ via

$$\begin{aligned}\boldsymbol{m}(t) &= \boldsymbol{P}(t, t_i)\boldsymbol{m}(t_i)\\ \boldsymbol{K}(t) &= \boldsymbol{P}(t, t_i)\boldsymbol{K}(t_i)\boldsymbol{P}(t, t_i)^T + \boldsymbol{L}(t, t_i), \quad t \in (t_i, t_{i+1}).\end{aligned}$$

These can then be passed on to the smoothing backpropagation to receive continuous state estimates $\boldsymbol{X}(t) \mid \boldsymbol{Z}(t_{1:I})$ with statistics

$$\begin{aligned}\boldsymbol{G^s}(t_{i+1}, t) &= \boldsymbol{K}(t)\boldsymbol{P}(t_{i+1}, t)^T\boldsymbol{K^-}(t_{i+1})^{-1}\\ \boldsymbol{m^s}(t) &= \boldsymbol{m}(t) + \boldsymbol{G^s}(t_{i+1}, t)\left[\boldsymbol{m^s}(t_{i+1}) - \boldsymbol{m^-}(t_{i+1})\right]\\ \boldsymbol{K^s}(t) &= \boldsymbol{K}(t) + \boldsymbol{G^s}(t_{i+1}, t)\left[\boldsymbol{K^s}(t_{i+1}) - \boldsymbol{K^-}(t_{i+1})\right]\boldsymbol{G^s}(t_{i+1}, t)^T,\end{aligned}$$

where t is such that $t \in (t_i, t_{i+1})$ for some i.

The Filtering Likelihood Function

We now turn our attention to the situation in which the process dynamics and observations are parametrized by unknown model parameters $\boldsymbol{\theta}$. For this case, the Kalman filter emits a closed-form likelihood function which can be passed onto any likelihood-based parameter estimation method.

Assume the dynamical model (2.81) and observation model (2.82) are parametrized by a vector $\boldsymbol{\theta}$ and denote the corresponding model matrices and Kalman filter outputs by $\boldsymbol{F_\theta}, \boldsymbol{P_\theta}, \boldsymbol{Q_\theta}, \boldsymbol{L_\theta}, \boldsymbol{H_\theta}, \boldsymbol{R_\theta}$, $\boldsymbol{G_\theta}$ and $\boldsymbol{m_\theta^-}, \boldsymbol{K_\theta^-}, \boldsymbol{m_\theta}, \boldsymbol{K_\theta}$.

Algorithm 2.2: Kalman Smoother

Input: state-transition matrix function $\boldsymbol{P}$, Kalman filter outputs $\boldsymbol{m}(t_i), \boldsymbol{K}(t_i)$, $\boldsymbol{m^-}(t_i), \boldsymbol{K^-}(t_i)$, $i = 1, ..., I$.

Output: smoothing mean and covariance matrix estimates $\boldsymbol{m^s}(t_i), \boldsymbol{K^s}(t_i)$, i = 1,...,I.

Initialization: $\boldsymbol{m^s}(t_I) = \boldsymbol{m}(t_I), \boldsymbol{K^s}(t_I) = \boldsymbol{K}(t_I)$.

for $i = I - 1, ..., 1$ **do**

$\boldsymbol{G^s}(t_{i+1}, t_i) = \boldsymbol{K}(t_i)\boldsymbol{P}(t_{i+1}, t_i)^T\boldsymbol{K^-}(t_{i+1})^{-1}$

$\boldsymbol{K^s}(t_i) = \boldsymbol{K}(t_i) + \boldsymbol{G^s}(t_{i+1}, t_i)\left[\boldsymbol{K^s}(t_{i+1}) - \boldsymbol{K^-}(t_{i+1})\right]\boldsymbol{G^s}(t_{i+1}, t_i)^T$;

$\boldsymbol{m^s}(t_i) = \boldsymbol{m}(t_i) + \boldsymbol{G^s}(t_{i+1}, t_i)\left[\boldsymbol{m^s}(t_{i+1}) - \boldsymbol{m^-}(t_{i+1})\right]$;

end

Given the observations $\boldsymbol{Z}(t_i) = \boldsymbol{z}_i, i = 1, ..., I$, we can, noting the Markov property of the underlying true process $(\boldsymbol{X}(t))_t$ and the white noise property of $\boldsymbol{\epsilon}(t_i)_i$, factorize the likelihood function in $\boldsymbol{\theta}$ as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &:= f_{\boldsymbol{Z}(t_{1:I})|\boldsymbol{\theta}}(\boldsymbol{z}_{1:I} \mid \boldsymbol{\theta}) \\ &= f_{\boldsymbol{Z}(t_1)|\boldsymbol{\theta}}(\boldsymbol{z}_1 \mid \boldsymbol{\theta}) \prod_{i=1}^{I-1} f_{\boldsymbol{Z}(t_{i+1})|\boldsymbol{Z}(t_{1:i}),\boldsymbol{\theta}}(\boldsymbol{z}_{i+1} \mid \boldsymbol{z}_{1:i}, \boldsymbol{\theta}).\end{aligned} \tag{2.90}$$

Note that the conditional densities $f_{\boldsymbol{Z}(t_{i+1})|\boldsymbol{Z}(t_{1:i}),\boldsymbol{\theta}}(\boldsymbol{z}_{i+1} \mid \boldsymbol{z}_{1:i}, \boldsymbol{\theta})$, $i = 1, ..., I$, are a byproduct of the Kalman filter. As seen in (2.85) they are Gaussian with mean and covariance matrix given by

$$\begin{aligned}\boldsymbol{m}_{\boldsymbol{Z},\boldsymbol{\theta}}(t_{i+1}) &= \boldsymbol{H}_{\boldsymbol{\theta}}(t_{i+1})\boldsymbol{m}^-_{\boldsymbol{\theta}}(t_{i+1}) \\ \boldsymbol{K}_{\boldsymbol{Z},\boldsymbol{\theta}}(t_{i+1}) &= \boldsymbol{H}_{\boldsymbol{\theta}}(t_{i+1})\boldsymbol{K}^-_{\boldsymbol{\theta}}(t_{i+1})\boldsymbol{H}^T_{\boldsymbol{\theta}}(t_{i+1}) + \boldsymbol{R}_{\boldsymbol{\theta}}(t_{i+1}),\end{aligned} \tag{2.91}$$

whereas for $i = 0$, $\boldsymbol{m}_{\boldsymbol{Z},\boldsymbol{\theta}}(t_1)$ and $\boldsymbol{K}_{\boldsymbol{Z},\boldsymbol{\theta}}(t_1)$ are the predicted statistics of the first observation, based on the Filtering initiation $\boldsymbol{X}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{K}_0)$.

Denote by $f(\cdot; \boldsymbol{m}, \boldsymbol{K})$ the density of a n-dimensional multivariate Gaussian variable with mean $\boldsymbol{m}$ and covariance $\boldsymbol{K}$. The overall likelihood function can then be written as

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{I} f_n(\boldsymbol{z}_i; \boldsymbol{m}_{\boldsymbol{Z},\boldsymbol{\theta}}(t_i), \boldsymbol{K}_{\boldsymbol{Z},\boldsymbol{\theta}}(t_i)) \tag{2.92}$$

and be passed on to any frequentist or Bayesian parameter estimation method.

Remark 2.3. All results derived in this section are easily extended to the complex-valued case, assuming the system is such that the state $(\boldsymbol{X}(t))_t$ and observations $(\boldsymbol{Z}(t_i))_i$ remain proper at all times. In the linear case, this is given when the prior state and measurement noise are assumed to be proper.

For this, recall that proper complex Gaussian vectors are uniquely defined by their mean and covariance matrix, and that properness remains under summation of independent variables. Using the density $f_{\boldsymbol{Z}}$, given by

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{\pi^n \det(\boldsymbol{K})} \exp\left(-(\boldsymbol{z} - \boldsymbol{m})^H \boldsymbol{K}^{-1}(\boldsymbol{z} - \boldsymbol{m})\right)$$

for any proper complex Gaussian $\boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{m}, \boldsymbol{K})$ and the fact that linear combinations of independent proper complex Gaussians variables are again proper complex Gaussian (see property 3 in lemma 2.2), one is quick to see that all derivations made in the section above are generalized to the complex case by simply interchanging the respective densities and the transpose- to the hermitian transpose operator.

2.4. Linear Stochastic Partial Differential Equations

This section introduces a framework of solving linear stochastic partial differential equations, formally written as

$$\partial_t U(t, \boldsymbol{x}) = \mathcal{A}U(t, \boldsymbol{x}) + \dot{W}(t, \boldsymbol{x}), \tag{2.93}$$

where $\mathcal{A}$ is a linear differential operator acting in the spatial domain and $\dot{W}(t, \boldsymbol{x})$ denotes *space-time white noise*, a stationary centered Gaussian process with random values that are independent with respect to the space-time variable. If $\dot{W}(t, \boldsymbol{x})$ is correlated over the spatial domain instead, we refer to it as temporally white and spatially coloured noise. As in the solely temporal case, $\dot{W}(t, \boldsymbol{x})$ does not exist in the original sense of a stochastic process. Previously, this issue was circumvented by understanding stochastic differential equations as integral equations through the Itô integral. We will now introduce the corresponding framework for the space-time case. This is done by considering *generalized random processes* $(U(t))_t$ and understanding (2.93) as equations of the form

$$dU(t) = \mathcal{A}U(t)dt + dW^Q(t), \tag{2.94}$$

where $(U(t))_t$ is a process taking values in a suitable Hilbert space on which $\mathcal{A}$ acts as a linear operator and $(W^Q(t))_t$ is a *cylindrical Wiener process*, likewise taking values in a suitable Hilbert space. One can then formulate the notion of a *weak solution* to equations of type (2.94) after defining a suitable way of integrating generalized processes in regards to W^Q.

We start off by introducing generalized random processes, based roughly on Lototsky and Rozovsky [2017], chapter 1.1 and 3.3. The succeeding introduction of Hilbert-space valued stochastic integration and weak solutions is based on Da Prato and Zabczyk [1992], chapter 4 and 5.

For the rest of section, denote by $D \subset \mathbb{R}^d$ some compact spatial domain and let $L^2(D)$, $L^2([0, T])$ denote the separable Hilbert spaces of square-integrable real-valued functions with scalar product $\langle h, g\rangle_D = \int_D h(x)g(x)dx$ and $\langle h, g\rangle_T = \int_0^T h(t)g(t)dt$ respectively.

2.4.1. Generalized Random Processes

Recall that one may construct a Wiener process on any interval $[0, T]$ as follows: let $(\psi_k)_k$ be an orthonormal basis of $L^2([0, T])$ and $(\eta_k)_k$ be a sequence of independent, standard Gaussian random variables. Then one can show that

$$w(t) = \sum_k \left(\int_0^t \psi_k(s)ds \right) \eta_k$$

satisfies the assumptions of Brownian motion - it is a centered Gaussian process with almost surely continuous paths and covariance function $E[w(s)w(t)] = \min(s, t)$.

This suggests the *formal* expression of white noise as the sum

$$\dot{w}(t) = \sum_k \psi_k(t)\eta_k. \tag{2.95}$$

While in general this expression diverges, it gives rise to the definition of white noise as a *generalized random function*:

Definition 2.11 (White Noise, Lototsky and Rozovsky [2017], 1.1.5)**.** Let $(\psi_k)_k$ be an orthonormal basis of $L^2([0,T])$ and $(\eta_k)_k$ be a sequence of independent, standard Gaussian variables. Then, temporal white noise on $[0,T]$ is defined as a generalized random function

$$\begin{aligned} \dot{w} : L^2([0,T]) &\longrightarrow \mathbb{R} \\ \dot{w}(h) &= \sum_k \langle h, \psi_k \rangle_T \eta_k, \end{aligned} \tag{2.96}$$

where $\langle h, \psi_k \rangle = \int_0^T h(t)\psi_k(t)$ denotes the scalar product on $L^2([0,T])$. The random variables $\dot{w}(h)$ are centered Gaussian with covariances given by

$$\mathrm{E}[\dot{w}(h)\dot{w}(g)] = \langle h, g \rangle_T.$$

Likewise, one defines *spatial white noise* on D as the generalized random function

$$\begin{aligned} \dot{W} : L^2(D) &\longrightarrow \mathbb{R} \\ \dot{W}(h) &= \sum_{j \geq 1} \langle h, \phi_j \rangle_D \eta_j, \end{aligned} \tag{2.97}$$

for some orthonormal basis $(\phi_j)_j$ of $L^2(D)$ and sequence of independent, standard Gaussian variables $(\eta_j)_j$.

It is handy to extend these considerations to the spatiotemporal case in the following way. Let $D = [0,L]^d \subset \mathbb{R}^d$ be rectangular and $(w_j)_j$ a sequene of independent Wiener processes. Then, the space-time pendant to the Wiener process is the *Brownian sheet*, which may be constructed on $[0,T] \times D$ via

$$W(t, \boldsymbol{x}) = \sum_{j \geq 1} \left(\int_0^{x_1} ... \int_0^{x_d} \phi_j(\boldsymbol{r}) d\boldsymbol{r} \right) w_j(t).$$

It is a centered Gaussian processes with almost surely continuous paths and covariance function $\mathrm{E}[W(s, \boldsymbol{x})W(t, \boldsymbol{y})] = \min(s,t) \prod_{i=1}^d \min(x_i, y_i)$. Once again, formal derivation suggests space-time white noise to take on the form

$$\dot{W}(t, \boldsymbol{x}) = \sum_{j \geq 1} \phi_j(\boldsymbol{x}) \dot{w}_j(t). \tag{2.98}$$

Note that, here, the dot in $\dot{W}$ stands for the partial derivative $\partial_t \partial_{x_1} ... \partial_{x_d}$ rather than its usual notation for the temporal derivatives. While this sum once again diverges, it gives rise to the following definition:

Definition 2.12 (Space-Time White Noise, Lototsky and Rozovsky [2017] 1.1.2)**.** Let $(\phi_j)_j$ be an orthonormal basis of $L^2(D)$ and $(w_j)_j$ be a sequene of independent Wiener processes. Then, space-time white noise on $[0,T] \times D$ is defined as the generalized random function

$$\begin{aligned} \dot{W} : L^2([0,T] \times D) &\longrightarrow \mathbb{R} \\ \dot{W}(h) &= \sum_{j \geq 1} \int_0^T \langle h(t,\cdot), \phi_j \rangle_D dw_j(t), \end{aligned} \tag{2.99}$$

with $h(t,\cdot)$ denoting the function $\boldsymbol{x} \mapsto h(t,\boldsymbol{x})$ for a fixed t. The random variables $\dot{W}(h)$ are centered Gaussian with covariances given by

$$\mathrm{E}[\dot{W}(h)\dot{W}(g)] = \int_0^T \langle h(t,\cdot) g(t,\cdot) \rangle_D dt. \tag{2.100}$$

Note that this definition does not assume D to be of rectangular form anymore.

The previous two definitions made sense of (spatio)temporal white noise as a generalized random function acting on L^2 spaces. In a much broader setting, one can define:

Definition 2.13 (Generalized Gaussian Random Fields, Lototsky and Rozovsky [2017] 3.2.10.)**.** Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space. A *generalized (Gaussian) random field* or *generalized (Gaussian) random process* is a collection of centered Gaussian random variables $(F(h))_{h \in H}$ such that there exists a bounded, linear, self-adjoint, non-negative operator $Q : H \longrightarrow H$ with

$$\mathrm{E}[F(h)F(g)] = \langle Qh, g \rangle \tag{2.101}$$

for all $h, g \in H$. We call Q the covariance operator of F. If Q is the identity operator, we call F white noise on H. If H is a function space, we also refer to F as a *generalized (Gaussian) random function.*

Clearly, the generalized white noise processes as defined above are cases of generalized Gaussian random fields on the Hilbert spaces $L^2([0,T])$ and $L^2([0,T] \times D)$ respectively. Their formal expressions given in equations (2.95) and (2.98) hint towards the existence of generalized fields that can be represented via random variables taking values in H. Indeed, these generalized random fields make up the following subclass:

Definition 2.14 (Regular Generalized Gaussian Random Field, Lototsky and Rozovsky [2017] 3.2.13)**.** We call a generalized Gaussian random field F on a Hilbert space $(H, \langle \cdot, \cdot \rangle)$ *regular* if there exists a H-valued random variable f such that $\mathrm{E}[\|f\|_H^2] < \infty$ and for all $h \in H$:

$$F(h) = \langle h, f \rangle.$$

It turns out that regular generalized fields can be identified through the properties of their covariance operators. We call a linear operator $Q : H \longrightarrow H$ on a separable Hilbert space $(H, \langle \cdot, \cdot \rangle)$ *nuclear*[5] if for any orthonormal basis $(h_j)_j$ of H it holds

$$\mathrm{tr}(Q) := \sum_{j \geq 1} \langle Q h_j, h_j \rangle < \infty.$$

In that case, $tr(Q)$ is independent of the choice of $(h_j)_j$. We then have:

Theorem 2.6 (Characterisation of Regular Generalized Fields)**.** A generalized Gaussian random field F on a Hilbert space H is regular if and only if its covariance operator Q is nuclear.

Proof. See Lototsky and Rozovsky [2017], 3.2.15. □

Note that in particular, a generalized white noise process is therefore *not* regular. However, in the space-time case, it is often useful to consider noise processes that are spatially coloured. In that case, it is a natural question to ask how one constructs a spatial noise process in the generalized sense based on covariance functions of "ordinary" Gaussian processes. The answer is readily available through Mercers theorem:

Theorem 2.7 (Mercer's Theorem)**.** Let $k : D \times D \longrightarrow \mathbb{R}$ be a continuous, symmetric, positive-definite function. Define the associated *Hilbert-Schmidt integral operator* Q_k by

$$\begin{aligned} Q_k : L^2(D) &\longrightarrow L^2(D) \\ Q_k h(\boldsymbol{x}) &:= \int_D k(\boldsymbol{x}, \boldsymbol{y}) h(\boldsymbol{y}) d\boldsymbol{y}. \end{aligned} \tag{2.102}$$

Then, there exists an orthonormal basis $(\phi_j)_j$ of $L^2(D)$ of eigenfunctions of Q_k with corresponding eigenvalues λ_j such that

$$k(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j \geq 1} \lambda_j \phi_j(\boldsymbol{x}) \phi_j(\boldsymbol{y}) \tag{2.103}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in D$, where the convergence of the sum is uniform in $\boldsymbol{x}, \boldsymbol{y}$ and absolute. In particular, Q_k is a linear, compact, self-adjoint and nuclear operator on $L^2(D)$.

Proof. A proof can be found in Jörgens and Roach [1982], theorem 8.11. □

Remark 2.4. Clearly, Mercer's theorem relates any covariance function on D to the covariance operator of a regular, generalized random field on $H = L^2(D)$. This generalized random function can be expressed in the following way: let $(\phi_j)_j$ be the eigenfunctions

[5] Note that this is actually an equivalence statement for the special case of an underlying Hilbert space, related to a broader definition of nuclear operators on Banach spaces. For a more general definition and other equivalence relations see Lototsky and Rozovsky [2017] 3.1.17 and 3.1.14.

of Q_k with eigenvalues $(\lambda_j)_j$ given by Mercer's theorem. Define the regular generalized random field f by

$$f := \sum_{j\geq 1} \sqrt{\lambda_j}\phi_j\eta_j \tag{2.104}$$

where $(\eta_j)_j$ is a sequence of independent, standard Gaussian variables. Note that the convergence of this representation sum, in the $L^2(\mathbb{P})$-sense, is ensured by

$$\mathrm{E}[\|f\|_H^2] = \sum_{j\geq 1} \lambda_j = \int_D k(\boldsymbol{x},\boldsymbol{x})d\boldsymbol{x} < \infty.$$

Indeed, noting that by Mercer's theorem

$$Q_k h = \sum_{j\geq 1} \lambda_j \langle h, \phi_j\rangle_D \phi_j$$

it follows that the covariance operator of f is given by Q_k:

$$\begin{aligned}
\mathrm{E}[\langle h, f\rangle_D \langle g, f\rangle_D] &= \mathrm{E}\left[\sum_{j\geq 1} \sqrt{\lambda_j}\langle h, \phi_j\rangle_D \eta_j \sum_{j'} \sqrt{\lambda_{j'}}\langle g, \phi_{j'}\rangle_D \eta_{j'}\right] \\
&= \sum_{j\geq 1} \lambda_j \langle h, \phi_j\rangle_D \langle g, \phi_j\rangle_D \\
&= \langle \sum_{j\geq 1} \lambda_j \langle h, \phi_j\rangle_D \phi_j, g\rangle_D \\
&= \langle Q_k h, g\rangle_D.
\end{aligned} \tag{2.105}$$

This way of representing regular generalize fields is not limited to function spaces. One can show that a nuclear, self-adjoint operator Q on any separable Hilbert space H is also compact (see e.g. Da Prato and Zabczyk [1992], App. C.3). Therefore, by the spectral theorem, there exists an orthonormal basis of H of eigenvectors $(h_j)_j$ of Q with eigenvalues $(\lambda_j)_j$. One can then construct the corresponding regular Q-Wiener process as in (2.104).

The previous considerations suggest constructing a noise process that is temporally white and spatially coloured with covariance operator Q_k via the sum

$$\dot{W}^{Q_k}(t,\boldsymbol{x}) = \sum_{j\geq 1} \sqrt{\lambda_j}\phi_j(\boldsymbol{x})\dot{w}_j(t). \tag{2.106}$$

As a generalized random function it is defined as follows.

Definition 2.15 (Temporally White, Spatially Coloured Noise)**.** Let $k : D \times D \longrightarrow \mathbb{R}$ be a covariance function and $(\phi_j)_j$, $(\lambda_j)_j$ be as given by Mercer's theorem. The corresponding temporally white, spatially coloured noise is the generalized random function

$$\begin{aligned}
\dot{W}^{Q_k} &: L^2([0,T]\times G) \longrightarrow L^2([0,T]\times G) \\
\dot{W}^{Q_k}(h) &= \sum_{j\geq 1} \sqrt{\lambda_j} \int_0^T \langle h(t,\cdot), \phi_j\rangle_D dw_j(t).
\end{aligned} \tag{2.107}$$

The random variables $\dot{W}^{Q_k}(h)$ are centered Gaussian with covariances given by

$$\mathrm{E}[\dot{W}(h)\dot{W}(g)] = \int_0^T \langle Q_k h(t,\cdot), g(t,\cdot)\rangle_D dt. \tag{2.108}$$

Remark 2.5. Note that including the temporally white property in definition 2.15 gets rid of the regularity property again. Let us for a moment ignore the regularity issues of the temporal white noise process and assume that there exists a Gaussian process $(\dot{w}(t))_t$ with some covariance function $k(s,t)$ such that for all $h \in L^2([0,T])$ (with abuse of notation)

$$\dot{w}(h) = \langle h, \dot{w}\rangle_T.$$

It turns out that the correct choice of covariance "function" is given by $k(s,t) = \delta(s-t)$, where $\delta(\cdot)$ denotes the Dirac delta function, characterized by its property

$$\int_0^T \delta(t)h(t)dt = h(0).$$

Indeed, for the corresponding integral operator on then has

$$Q_k h(t) = \int_0^T h(s)\delta(s-t)ds = h(t)$$

and therefore one recovers the covariance function of generalized white noise:

$$\mathrm{E}[\langle h, \dot{w}\rangle_T \langle g, \dot{w}\rangle_T] = \langle h, g\rangle_T.$$

Ignoring the fact that Dirac delta function acting with such properties is only a function in the generalized sense, one often defines it by

$$\delta(t) \propto \begin{cases} 1, & \text{if } t = 0 \\ 0, & \text{else .} \end{cases} \tag{2.109}$$

It is in this sense that one heuristically defines temporal white noise as the centered Gaussian process $(\dot{w}(t))_t$ with $\mathrm{E}[\dot{w}(s)\dot{w}(t)] = \delta(s-t)$.

Likewise, heuristic space-time white noise is a centered Gaussian process $(\dot{W}(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ with covariance function $\mathrm{E}[\dot{W}(s,\boldsymbol{y})\dot{W}(t,\boldsymbol{x})] = \delta(s-t)\delta(\boldsymbol{y}-\boldsymbol{x})$, whereas the representation (2.106) of spatially coloured, temporally white noise suggests it being a Gaussian process with covariance function

$$\begin{aligned}
\mathrm{E}[\dot{W}^{Q_k}(s,\boldsymbol{y})\dot{W}^{Q_k}(t,\boldsymbol{x})] &= \mathrm{E}\left[\sum_{j\geq 1}\sqrt{\lambda_j}\phi_j(\boldsymbol{y})\dot{w}_j(s)\sum_{j'}\sqrt{\lambda_{j'}}\phi_{j'}(\boldsymbol{x})\dot{w}_{j'}(t)\right] \\
&= \sum_{j\geq 1}\lambda_j\phi_j(\boldsymbol{y})\phi_j(\boldsymbol{x})\delta(s-t) \\
&= k(\boldsymbol{y},\boldsymbol{x})\delta(s-t).
\end{aligned}$$

We will refer to Gaussian processes with covariance functions such as these as *formal* or *heuristic* (temporally) white noise processes.

2.4.2. Q-cylindrical processes and Itô Integration in Hilbert Spaces

As mentioned in the beginning of this section, we can circumvent the regularity issue of white noise by understanding stochastic partial differential equations as integral equations in regards to function space valued processes. For this we need a fitting generalization of a Wiener process:

Definition 2.16 (Q-Cylindrical Processes, Lototsky and Rozovsky [2017], 3.2.35)**.** Let $(x(t))_t$ be a centered Gaussian process on $[0,T]$ with covariance function $k(s,t)$ and $(H, \langle\cdot,\cdot\rangle)$ be a separable Hilbert space.

A *(Q-)cylindrical process* on $H \times [0,T]$ is a collection of centered Gaussian random variables $(X^Q(t,h))_{t\in[0,T],h\in H}$ such that there exists a bounded, linear, self-adjoint, non-negative operator $Q : H \longrightarrow H$ with

$$\mathrm{E}[X^Q(s,h)X^Q(t,g)] = \langle Qh, g\rangle k(s,t) \tag{2.110}$$

for all $s,t \in [0,T]$ and $g,h \in H$. If $(x(t))_t$ is a Wiener process, we refer to X^Q as a *(Q-)Cylindrical Wiener process* and denote it by W^Q.

A cylindrical process is called *regular* if there exists a H-valued random process $(x^Q(t))_t$ such that for all $h \in H$, $t \in [0,T]$

$$X^Q(t,h) = \langle h, x^Q(t)\rangle. \tag{2.111}$$

Likewise to the time independent case, we get the following regularity condition for cylindrical processes:

Theorem 2.8 (Characterisation of Regular Cylindrical Processes)**.** A cylindrical process X^Q is regular if and only if Q is nuclear.

Proof. See Lototsky and Rozovsky [2017], theorem 3.2.39. □

Remark 2.6. Likewise to the construction in remark 2.4, we can represent any regular $Q-$cylindrical process as

$$X^Q(t) = \sum_{j\geq 1} \sqrt{\lambda_j} h_j x_j(t) \tag{2.112}$$

where $(x_j)_j$ are independent copies of the process $(x(t))_t$ and $(h_j)_j, (\lambda_j)_j$ are the eigenvectors and eigenvalues of Q on H. A similar computation as before shows that indeed

$$\mathrm{E}\left[\langle h, W^Q(s)\rangle_H \langle g, W^Q(t)\rangle_H\right] = \langle Qh, g\rangle_H \min(s,t). \tag{2.113}$$

We need two more definitions in order to define a correct way of integrating with respect to a Q-Wiener process:

Theorem 2.9. Let $Q : H \longrightarrow H$ be a bounded, linear, non-negative and self-adjoint operator. Then there exists a unique non-negative, self-adjoint operator R such that $RR = Q$. We call R the *(positive) square root operator* of Q and denote it by $\sqrt{Q}$.

Furthermore, if Q is such that there exists an orthonormal basis $(h_j)_j$ of H of eigenvectors of Q with eigenvalues $(\lambda_j)_j$, then these are eigenfunctions of $\sqrt{Q}$ with eigenvalues $\sqrt{\lambda_j}$.

Proof. For the proof of existence, see Lototsky and Rozovsky [2017], theorem 3.1.28. The second claim follows easily by showing that

$$h \mapsto \sum_{j\geq 1} \sqrt{\lambda_j}\langle h, h_j\rangle_H h_j$$

satisfies all necessary properties of $\sqrt{Q}$. □

Definition 2.17 (Hilbert Schmidt Operators, Da Prato and Zabczyk [1992] App. C.5)**.** Let $(H, \langle\cdot,\cdot\rangle_H)$ and $(G, \langle\cdot,\cdot\rangle_G)$ be separable Hilbert spaces and $A : H \longrightarrow G$ be a bounded linear operator. A is called a *Hilbert-Schmidt operator* if

$$\|A\|_{HS}^2 := \sum_{j\geq 1} \|Ah_j\|_G^2 < \infty \tag{2.114}$$

for any orthonormal basis $(h_j)_j$ of H. Then, $\|A\|_{HS}$ is independent of the choice of $(h_j)_j$. The space of Hilbert-Schmidt operators $L_{HS}(H, G)$ is a separable Hilbert space in regards to

$$\langle A, B\rangle_{HS} := \sum_{j\geq 1} \langle Ah_j, Bh_j\rangle_G. \tag{2.115}$$

We can now define the Itô integral in regards to cylindrical Wiener processes. For us, it suffices to consider the deterministic case where the integrand is a function taking values in the space L_{HS}. The construction we use was given without a proof of existence in Bréhier [2014]. A more general construction for L_{HS}-valued stochastic processes is given in chapter 4.2 of Da Prato and Zabczyk [1992].

Theorem 2.10 (The Hilbert Space Valued Itô Integral)**.** Let $(H, \langle\cdot,\cdot\rangle_H)$ be a separable Hilbert space and W^Q be a regular $Q-$Wiener process on H with eigenvectors and eigenvalues $(h_j)_j$ and $(\lambda_j)_j$ of Q. Further, let $(G, \langle\cdot,\cdot\rangle_G)$ be a separable Hilbert space with orthonormal basis $(g_i)_i$ and let $S : [0, T] \longrightarrow L_{HS}(H, G)$ be a Hilbert-Schmidt operator valued function such that

$$\int_0^T \|S(t)\sqrt{Q}\|_{HS}^2 dt < \infty. \tag{2.116}$$

Then, the integral

$$\int_0^T S(t)dW^Q(t) := \sum_{i,j} \int_0^T \sqrt{\lambda_j}\langle S(t)h_j, g_i\rangle_G dw_j(t)g_i \tag{2.117}$$

is well defined as a $L^2(\mathbb{P})$-limit in G and it holds the Itô symmetry in G:

$$\mathrm{E}\left[\left\|\int_0^T S(t)dW^Q(t)\right\|_G^2\right] = \int_0^T \|S(t)\sqrt{Q}\|_{HS}^2 dt. \tag{2.118}$$

Proof. Note that the eigenvectors $(h_j)_j$ of $\sqrt{Q}$ build an orthonormal basis of H and by the spectral theorem for compact self-adjoint operators, there exists an ordering of the corresponding eigenvalues $(\sqrt{\lambda_j})_j$ such that $\sqrt{\lambda_j} \longrightarrow 0$. In particular $\sup_j \lambda_j < \infty$ and we have that

$$\begin{aligned}
\|S(t)\sqrt{Q}\|_{HS}^2 &= \sum_{j\geq 1} \langle S(t)\sqrt{Q}h_j, S(t)\sqrt{Q}h_j\rangle_G \\
&= \sum_j \lambda_j \langle S(t)h_j, S(t)h_j\rangle_G \\
&= \sum_j \lambda_j \|S(t)h_j\|_G^2 \\
&\leq \sup_j \lambda_j \|S(t)\|_{HS}^2
\end{aligned}$$

and therefore $S(t)\sqrt{Q}$ is again a Hilbert-Schmidt operator. For $n, m \in \mathbb{N}$, define

$$I_{n,m} := \sum_{i=1}^n \sum_{j=1}^m \int_0^T \sqrt{\lambda_j}\langle S(t)h_j, g_i\rangle_G dw_j(t) g_i.$$

We show that $(I_{n,m})_{n,m}$ is Cauchy in $L^2(\mathbb{P})$. It holds,

$$\begin{aligned}
\mathrm{E}[\|I_{n,m} - I_{n',m'}\|_G^2] &= \mathrm{E}\left[\left\|\sum_{i=n}^{n'} \sum_{j=m}^{m'} \int_0^T \sqrt{\lambda_j}\langle S(t)h_j, g_i\rangle_G dw_j(t) g_i\right\|_G^2\right] \\
&= \mathrm{E}\left[\sum_{i=n}^{n'} \left(\sum_{j=m}^{m'} \int_0^T \sqrt{\lambda_j}\langle S(t)h_j, g_i\rangle_G dw_j(t)\right)^2\right] \\
&= \sum_{i=n}^{n'} \sum_{j=m}^{m'} \int_0^T \lambda_j |\langle S(t)h_j, g_i\rangle_G|^2 dt \\
&\xrightarrow[n,m,n',m'\to\infty]{} 0,
\end{aligned}$$

where we have used orthonormality of $(g_i)_i$ in the second equality, independence of $(w_j)_j$ and Itô symmetry in the third equality and lastly the fact that by Parseval's identity

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^m \int_0^T \lambda_j |\langle S(t)h_j, g_i\rangle_G|^2 dt &\xrightarrow[n,m\to\infty]{} \int_0^T \sum_{j\geq 1} \lambda_j \|S(t)h_j\|_G^2 dt \\
&= \int_0^T \|S(t)\sqrt{Q}\|_{HS}^2 dt < \infty.
\end{aligned}$$

In particular, this shows the Hilbert space Itô symmetry. □

Let us define the infinite dimensional pendant to a Gaussian process:

Definition 2.18 (Da Prato and Zabczyk [1992], section 3.6 [6])**.** Let $X(t), t \in [0,\infty)$, be a family of regular generalized Gaussian random fields on a Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$. Then, $(X(t))_t$ is called a *(centered) $H-$valued Gaussian process* if for any $t_1, ..., t_n \in [0,\infty)$, the vector $(X(t_1), ..., X(t_n))$ is such that

$$(\langle h_1, X(t_1)\rangle, ..., \langle h_n, X(t_n)\rangle)$$

is a Gaussian vector in $\mathbb{R}^n$ for all $h_1, ..., h_n \in H$. It is called stationary if its *covariance operator function* $K_{s,t}$, defined by

$$\mathrm{E}[\langle h, X(s)\rangle_H \langle g, X(t)\rangle_H] = \langle K_{s,t}h, g\rangle_H \tag{2.119}$$

for all $g, h \in H$, is such that $K_{s,t} = K_{s+\tau,t+\tau}$ for all $s, t, \tau \in [0,\infty)$.

Likewise to the finite dimensional case, the Hilbert space valued Itô integral is a Gaussian process in H.

Lemma 2.7. In the situation of theorem 2.10, denote by $(I(t))_t$ the process

$$I(t) = \int_0^t S(r)dW^Q(r)$$

Then $(I(t))_t$ is a $G-$valued Gaussian process.

Proof. For reasons of simplicity, we assume that $S : [0,T] \longrightarrow L_{HS}(H,H)$, i.e. $H = G$. Then, $I(t)$ reduces to

$$I(t) = \sum_{j\geq 1} \int_0^t \sqrt{\lambda_j}\langle S(r)h_j, h_j\rangle_H dw_j(r)h_j.$$

The conditions in theorem 2.10 assure that

$$\langle g, I(t)\rangle_H = \sum_{j\geq 1} \int_0^t \sqrt{\lambda_j}\langle S(r)h_j, h_j\rangle_H dw_j(r)\langle g, h_j\rangle_H.$$

is indeed Gaussian as the weak limit of Gaussian variables.

Now let $\mu_1, \mu_2 \in \mathbb{R}$, $t_1 \leq t_2 \in [0,\infty)$ and $g_1, g_2 \in H$. We show that $\mu_1\langle g_1, I(t_1)\rangle_H + \mu_2\langle g_2, I(t_2)\rangle_H$ is again Gaussian.

We may write

$$\begin{aligned} I(t_2) &= I(t_1) + \sum_{j\geq 1} \int_{t_1}^{t_2} \sqrt{\lambda_j}\langle S(r)h_j, h_j\rangle_H dw_j(r)h_j \\ &:= I(t_1) + J \end{aligned}$$

[6] The source uses tensor product notation to define this. A couple of careful manipulations show that the original definition is equivalent to this definition our notation

Then

$$\mu_1\langle g_1, I(t_1)\rangle_H + \mu_2\langle g_2, I(t_2)\rangle_H = \langle \mu_1 g_1 + \mu_2 g_2, I(t_1)\rangle_H + \langle \mu_2 g_2, J\rangle_H.$$

Note that by the independence of the Wiener integrals for disjoint time intervals, this is again Gaussian as the sum of weak limits of independent Gaussin variables.

By the same argument, one shows that

$$\mu_1\langle g_1, I(t_1)\rangle_H + ... + \mu_n\langle g_n, I(t_n)\rangle_H$$

is Gaussian for any $\mu_1, ..., \mu_n \in \mathbb{R}$, $t_1 \leq ... \leq t_2 \in [0, \infty)$ and $g_1, ..., g_n \in H$. □

We close this subsection with two remarks:

Remark 2.7. For the construction of the Hilbert space valued integral, we assumed W^Q to be regular. Of course one wishes to define an integral for the more general case when Q is any valid covariance operator, in particular when $Q = \mathrm{Id}$ is the identity and therefore relates to white noise on H. The trick is to "turn" W^Q into a regular process by embedding H into a larger space $\tilde{H}$, we give a rough idea on how to do this in the following:

Define $H_0 := \sqrt{Q}(H)$. This is again a separable Hilbert space in regards to the inner product

$$\langle h, g\rangle_0 := \langle \sqrt{Q}^{-1} h, \sqrt{Q}^{-1} g\rangle_H.$$

Denote by $(h_j)_j$ a corresponding orthonormal basis of H_0. Further, assume $\tilde{H}$ is a separable Hilbert space such that $H \subset \tilde{H}$ with the embedding being continuous and the embedding $H_0 \subset \tilde{H}$ being a Hilbert-Schmidt operator. Under these assumptions, one can show (Da Prato and Zabczyk [1992], proposition 4.11) that

$$W^{\tilde{Q}}(t) := \sum_{j\geq 1} h_j w_j(t)$$

defines a regular Q-Wiener process on $\tilde{H}$ that, when restricted to H, acts with covariance operator Q. Like above, one can define the stochastic integral as the $L^2(\mathbb{P})$ limit

$$\int_0^T S(t) dW^{\tilde{Q}}(t) = \sum_{i,j} \langle S(t)h_j, g_i\rangle_G dw_j(t) g_i$$

for any $S : [0, T] \longrightarrow L_{HS}(\tilde{H}, G)$ such that

$$\int_0^T \sum_{j\geq 1} \|S(t)h_j\|_G^2 dt < \infty.$$

Note that condition (2.120) is equivalent to the one given in theorem 2.10 and this workaround does therefore not change the class of eligible integrands.

In particular, when $Q = \mathrm{Id}$, we define

$$\int_0^T S(t)dW^Q(t) = \sum_{i,j} \int_0^T \langle S(t)h_j, g_i\rangle_G dw_j(t)g_i \tag{2.120}$$

for any orthonormal basis $(h_j)_j$ of H and $S : [0,T] \longrightarrow L_{HS}(H,G)$ with

$$\int_0^T \|S(t)\|_{HS}^2 dt < \infty.$$

Remark 2.8 (About the Complex-Valued Case)**.** One can easily extend all previous definitions and results to the complex valued case. The respective complex-valued generalized fields, cylindrical processes and Itô integrals are defined on $\mathbb{C}$-Hilbert spaces by simply replacing the Gaussian variables and Wiener processes used above by their complex counterparts as defined in chapter 2.2. In regards to space-time processes, it is natural to then consider the $\mathcal{L}^2$-Hilbert spaces of square-integrable, complex-valued functions with scalar products on the respective domains defined as

$$\langle h, g\rangle = \int h(\boldsymbol{x})\overline{g(\boldsymbol{x})}d\boldsymbol{x}.$$

Furthermore, the complex valued formulation of Mercer's theorem links conjugate symmetric positive definite functions $k(s,t)$ to the representation

$$k(\boldsymbol{x},\boldsymbol{y}) = \sum_{j\geq 1} \sqrt{\lambda_j}\phi_j(\boldsymbol{x})\overline{\phi_j(\boldsymbol{y})},$$

with $(\phi_j)_j, (\lambda_j)_j$ again being the eigenfunctions and eigenvalues of the integral operator Q_k on the respective $\mathcal{L}^2$-space. In future applications we will make use of the complex valued case more often then not and, unless it is not clear in the given context, will not give it any further mention.

2.4.3. Weak Solutions

Recall that we seek a definition of solutions to equations of the form

$$\begin{cases} dU(t) = \mathcal{A}U(t)dt + BdW^Q(t), \\ U(0) = U_0, \end{cases}$$

where U_0 and $(U(t))_t$ take values in a Hilbert space H on which $\mathcal{A}$ acts as a linear operator and $(W^Q(t))_t$ is a Q-Wiener process, likewise taking values in a suitable Hilbert space G and $B : G \longrightarrow H$ is a bounded linear operator.

This representation suggests that we can formulate a solution by integrating over the Q-Wiener process. To make this final, we need to define one last tool:

Definition 2.19 (C_0-Semigroup, Da Prato and Zabczyk [1992] A.4)**.** Let E be a Banach space. A *C_0-Semigroup* is a mapping $(S(t))_{t\in[0,T]}$ into the space of bounded linear operators $L(E,E)$ such that

1. $S(0) = \mathrm{Id}$
2. $S(t+s) = S(t)S(s)$ for all $s, t \in [0,T]$
3. $\lim_{t\to 0} \|S(t)x - x\|_E = 0$ for all $x \in E$ fixed.

Its *infinitesimal generator* A is the unique mapping $A : D(A) \longrightarrow E$ with dense domain $D(A) \subset E$ given by

$$D(A) := \left\{ x \in E : Ax := \lim_{t\to 0} \frac{S(t)x - x}{t} \text{ exists} \right\}.$$

We finally have all the necessary tools to give a proper meaning to the solution of a linear stochastic partial differential equation:

Definition 2.20 (Weak Solution, Da Prato and Zabczyk [1992])**.** Let H and G be separable Hilbert spaces and W^Q be a regular Q-Wiener process on G. Further, let $\mathcal{A} : D(\mathcal{A}) \subset H \longrightarrow H$ be the generator of a C_0-semigroup on H and $B : G \longrightarrow H$ be a bounded linear operator. We call the $H-$valued process $(U(t))_{t\in[0,T]}$ a *weak solution* to the linear equation

$$\begin{cases} dU(t) &= \mathcal{A}U(t)dt + BdW^Q(t), \quad t \in [0,T], \\ U(0) &= U_0, \end{cases} \tag{2.121}$$

if the trajectories $t \mapsto U(t)$ are almost surely integrable in H and for all $V \in D(\mathcal{A}^*)$ we have, almost surely:

$$\langle V, U(t)\rangle_H = \langle V, U_0\rangle_H + \int_0^T \langle \mathcal{A}^*V, U(s)\rangle_H ds + \langle V, BW^Q(t)\rangle_H. \tag{2.122}$$

The existence of a weak solution is ensured by a simple integration criterion:

Theorem 2.11 (Existence of a Weak Solution, Da Prato and Zabczyk [1992])**.** In the situation of definition 2.20, assume that the C_0-semigroup $(S(t))_t$ generated by $\mathcal{A}$ is such that

$$\int_0^T \|S(t)\sqrt{Q}B\|_{HS}^2 dt < \infty.$$

Then there exists a unique weak solution to equation 2.121. For all $t \in [0,T]$, it is given by

$$U(t) = S(t)U_0 + \int_0^t S(t-s)BdW^Q(s). \tag{2.123}$$

Proof. For a proof see Da Prato and Zabczyk [1992], theorem 5.4. □

3. A Spectral Fourier Method for the Stochastic Convection-Diffusion Equation

This chapter aims to derive a spectral method for computationally efficient filtering, smoothing and parameter estimation of processes that evolve according to the stochastic linear convection-diffusion equation driven by a spatially coloured, temporally white Gaussian forcing term.

We approximate the solution by a Fourier series decomposition, which translates the temporal evolution of the approximation into the evolution of a family of uncoupled spectral processes. The spectral processes are given as strong solutions to a linear stochastic differential equation and filtering, smoothing and model parameter estimation for these spectral processes are therefore shown to be available in linear time complexity.

We remark that this chapter deals with SPDEs and their solutions in a *formal* way. For a treatment in the framework of section 2.4, we refer to chapter 3, where many of the practical results derived in this chapter are validated in a rigorous way.

The stochastic convection-diffusion equation and the problem statement are introduced in section 3.1. The following section 3.2 establishes the spectral method based on the spatial Fourier decomposition and spectral representation of the stochastic forcing term.

Section 3.3 derives the Kalman filter in the spectral domain, which allows for computationally efficient filtering- and smoothing as well as likelihood-based inference on the model parameters. Experimental performance of these methods are presented on toy data in section 3.4

3.1. The Stochastic Convection-Diffusion Equation

In section 2.3 we have introduced the Bayesian filtering and smoothing problem for time-dependent processes $(X(t))_t$ that arise as strong solutions of linear stochastic differential equations. Consider now the space-time pendant of estimating the state of some stochastic process $(U(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ governed by a *linear stochastic evolution equation*

$$\partial_t U(t,\boldsymbol{x}) - \mathcal{A}U(t,\boldsymbol{x}) = \dot{W}^Q(t,\boldsymbol{x}), \tag{3.1}$$

based on discretely observed, possibly noise-corrupted observations of U. Here, $\mathcal{A} = \mathcal{A}(\partial_{\boldsymbol{x}}, \partial^2_{\boldsymbol{xx}}, ...)$ is a linear differential operator acting in the spatial domain and $\dot{W}^Q(t,\boldsymbol{x})$ represents a Gaussian noise process, specified to be either space-time white noise or temporally white and spatially coloured.

Likewise to the original filtering problem, a subsequent objective that we will be concerned with is model parameter inference in cases where the differential operator $\mathcal{A}_{\boldsymbol{\theta}}$ as well as the stochastic forcing term $\dot{W}^Q_{\boldsymbol{\theta}}(t, \boldsymbol{x})$ are parametrized by some unknown parameter $\boldsymbol{\theta}$.

The next two sections are based on the work of Sigrist et al. [2015], who derived an approximated solution of (3.1) in the case where $\mathcal{A}_{\boldsymbol{\theta}}$ equals the linear convection-diffusion operator

$$\mathcal{A}_{\boldsymbol{\theta}} = -\boldsymbol{\mu} \cdot \nabla + \nabla \cdot \boldsymbol{\Sigma} \nabla - \zeta,$$

parametrized by the model parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \zeta)$.

In detail, we are concerned with an underlying process $(U(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ governed by the *stochastic convection-diffusion equation*

$$\begin{cases} \frac{\partial}{\partial t} U(t, \boldsymbol{x}) & = [-\boldsymbol{\mu} \cdot \nabla + \nabla \cdot \boldsymbol{\Sigma} \nabla - \zeta] \, U(t, \boldsymbol{x}) + \dot{W}^{Q_k}(t, \boldsymbol{x}), \quad t \in [0, \infty), \ \boldsymbol{x} \in D, \\ U(0, \boldsymbol{x}) & = U_0(\boldsymbol{x}). \end{cases} \tag{3.2}$$

The Nabla operator $\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2} \right)$ acts in the spatial domain $D \subset \mathbb{R}^2$. The parameters $\boldsymbol{\mu} \in \mathbb{R}^2$, $\boldsymbol{\Sigma} \in \mathbb{R}^{2 \times 2}$ and $\zeta > 0$ respectively denote the drift, diffusion and damping parameter. The diffusion matrix $\boldsymbol{\Sigma}$ is assumed to symmetric positive definite. In the special case where $\boldsymbol{\Sigma}$ is diagonal, one speaks of isotropic diffusion.

The stochastic forcing term $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ is assumed to be a temporally white, spatially Matérn noise process. We write its formal covariance function, in the sense of remark 2.5, as

$$k((t, \boldsymbol{x}), (s, \boldsymbol{y})) = \begin{cases} 0, & \text{if } t \neq s \\ c(\|\boldsymbol{x} - \boldsymbol{y}\|), & \text{if } t = s, \end{cases} \tag{3.3}$$

where $c(r)$ is the Whittle covariance function given by

$$c(\boldsymbol{r}) = \frac{\sigma_0^2 \boldsymbol{r}}{\rho_0} K_1 \left(\frac{\boldsymbol{r}}{\rho_0} \right) \tag{3.4}$$

with marginal variance σ_0^2 and scale parameter ρ_0.

The motivation behind this choice of the stochastic forcing term is to build on the "elementary properties" of the Whittle covariance function for spatial processes as discussed in section 2.2.4. Sigrist et al. [2015] argue that a process $(U(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ governed by the stochastic convection-diffusion equation with the stochastic forcing term defined by (3.6) extends these properties to the space-time domain due to the ubiquitous appearance of convective and diffusive behavior in natural processes and therefore makes for a "good candidate for an elementary spatiotemporal process".

Recall that the Matérn process with Whittle covariance function is stationary with spectral density given by

$$f(\boldsymbol{\xi}) = \frac{\sigma_0^2}{(2\pi)^2} (\rho_0^{-2} + \|\boldsymbol{\xi}\|^2)^{-2}, \quad \boldsymbol{\xi} \in \mathbb{R}^2. \tag{3.5}$$

A key result, proven by Whittle [1963], links the spectral denstiy $f(\boldsymbol{\xi})$ to the spectral density of the stationary solution of the stochastic convection-diffusion equation (3.2):

Theorem 3.1. Let $(U(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ be the stationary solution of the stochastic convection-diffusion equation (3.2), defined on the spatial domain $D = \mathbb{R}^2$ and driven by the Gaussian noise process $(\dot{W}^{Q_k}(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ with covariance function

$$k((t,\boldsymbol{x}),(s,\boldsymbol{y})) = \begin{cases} 0, & \text{if } t \neq s \\ \dfrac{\sigma_0^2 r}{\rho_0} K_1\left(\dfrac{\|\boldsymbol{x}-\boldsymbol{y}\|}{\rho_0}\right), & \text{if } t = s. \end{cases} \tag{3.6}$$

Then, $(U(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ is itself a centered Gaussian process with spectral density

$$g(\omega,\boldsymbol{\xi}) = f(\boldsymbol{\xi})\frac{1}{2\pi}[(\boldsymbol{\xi}^T\boldsymbol{\Sigma}\boldsymbol{\xi}+\zeta)^2 + (\omega+\boldsymbol{\mu}^T\boldsymbol{\xi})^2]^{-1}. \tag{3.7}$$

Proof. See Whittle [1963]. □

Remark 3.1. Following Bochners theorem, the covariance function of the stationary solution of (3.2) can be computed by integration of its spectral density $g(\omega,\boldsymbol{\xi})$. Sigrist et al. [2015] showed that

$$\begin{aligned} C(\tau,\boldsymbol{r}) &:= \mathrm{E}[U(t,\boldsymbol{x})U(t+\tau,\boldsymbol{x}+\boldsymbol{r})] \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}^2} g(\omega,\boldsymbol{\xi})\exp(i\omega\tau)\exp(i\boldsymbol{\xi}\cdot\boldsymbol{r})d\boldsymbol{\xi}d\omega \\ &= \int_{\mathbb{R}^2} f(\boldsymbol{\xi})\frac{\exp((-i\boldsymbol{\mu}^T\boldsymbol{\xi}-\boldsymbol{\xi}^T\boldsymbol{\Sigma}\boldsymbol{\xi}-\zeta)\tau)}{2(\boldsymbol{\xi}^T\boldsymbol{\Sigma}\boldsymbol{\xi}+\zeta)}\exp(i\boldsymbol{\xi}\cdot\boldsymbol{r})d\boldsymbol{\xi}, \end{aligned} \tag{3.8}$$

and claimed that the spatial integral has no closed form solution but can be numerically approximated.

Note that, even if the computation of $C(\tau,\boldsymbol{r})$ can be done efficiently, a likelihood evaluation using the covariance function remains costly. Assume $(U(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ is the stationary solution of (3.2) with covariance function $C(\tau,\boldsymbol{r})$. Then, for any set of space-time locations $(t_1,\boldsymbol{x}_1),(t_1,\boldsymbol{x}_2),...,(t_I,\boldsymbol{x}_{N-1}),(t_I,\boldsymbol{x}_N)$, the vector $\boldsymbol{U} = (U(t_i,\boldsymbol{x}_n))_{1\leq i\leq I,1\leq n\leq N} \in \mathbb{R}^{IN}$ is Gaussian with mean zero and covariance matrix

$$\boldsymbol{C} = (C(t_i-t_j,\boldsymbol{x_n}-\boldsymbol{x_m}))_{1\leq i,j\leq I,1\leq n,m\leq N} \in \mathbb{R}^{IN\times IN}.$$

The computation of the Gaussian likelihood function of $\boldsymbol{U}$ involves the inversion of the matrix $\boldsymbol{C}$ and is therefore of complexity class $\mathcal{O}((IN)^3)$. The following sections show that these cost can be dramatically reduced by translating the problem into the spectral domain.

3.2. The Fourier Spectral Method

This section derives an approximation of the stochastic convection-diffusion equation using a Fourier spectral decomposition. Spectral methods aim to approximately solve partial differential equations by constructing solutions U as (finite) linear combinations of basis functions. In the case of evolution equations, the decompositions take on the form

$$U(t, \boldsymbol{x}) = \sum_{i \in \mathcal{I}} \alpha_i(t) \phi_i(\boldsymbol{x}),$$

where $\mathcal{I}$ denotes some index set, $\{\phi_i(x), i \in \mathcal{I}\}$ is a family of orthogonal basis functions [1] and $\{\alpha_i(t), i \in I\}$ are the *spectral processes* (or *spectral coefficients* in time independent systems). A detailed introduction can for example be found in Gottlieb and Orszag [1977] or Canuto et al. [1988].

A common choice of basis functions $\{\phi_i(x), i \in \mathcal{I}\}$ are the *Fourier basis functions* due to their property of transforming differential operations in the spatial domain to algebraic operations in the spectral domain. As we will explain shortly, the consideration of the stochastic forcing term $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ in the stochastic convection-diffusion equation (3.2) motivates a decomposition based on the Fourier basis functions.

In that case, one usually restricts the problem at hand to a bounded spatial domain $D = [0, L]^2$ and seeks to find spatially periodic solutions on D. This reduces our problem to the approximation of a solution to the stochastic convection-diffusion equation with periodic boundary conditions, given by

$$\begin{cases} \dfrac{\partial}{\partial t} U(t, \boldsymbol{x}) = \left[-\boldsymbol{\mu} \cdot \nabla + \nabla \cdot \boldsymbol{\Sigma} \nabla - \zeta \right] U(t, \boldsymbol{x}) + \dot{W}^{Q_k}(t, \boldsymbol{x}), & \boldsymbol{x} \in [0, L]^2, t \in [0, \infty), \\ \quad U(t, \boldsymbol{x}) = U(t, \boldsymbol{x} + \boldsymbol{k}L), & \boldsymbol{k} \in \mathbb{Z}^2, \\ \quad U(0, \boldsymbol{x}) = U_0(\boldsymbol{x}). \end{cases}$$

The ansatz is to construct a solution $(U(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ in the form

$$U(t, \boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathcal{J}} \alpha_{\boldsymbol{j}}(t) \phi_{\boldsymbol{j}}(\boldsymbol{x}), \tag{3.9}$$

where $\mathcal{J} \subset \mathbb{Z}^2$ denotes a set of spatial wave indices and, for every $\boldsymbol{j} \in \mathcal{J}$, the function $\phi_{\boldsymbol{j}}(\boldsymbol{x})$ denotes the Fourier basis function corresponding to the $\boldsymbol{j}$-th spatial wave vector $\boldsymbol{\xi}_{\boldsymbol{j}}$, respectively defined as

$$\begin{aligned} \phi_{\boldsymbol{j}}(\boldsymbol{x}) &:= \exp(i \boldsymbol{\xi}_{\boldsymbol{j}} \cdot \boldsymbol{x}) \\ \boldsymbol{\xi}_{\boldsymbol{j}} &:= 2\pi (\tfrac{j_1}{L}, \tfrac{j_2}{L}). \end{aligned}$$

[1] Of course, the definition of *orthogonality* is dependent on the underlying function space in which a solution is approximated. The most common choice is the space $\mathcal{L}^2(D)$ of square-integrable functions over a bounded spatial domain D, equipped with the $\mathcal{L}^2$-scalar product $\langle f, g \rangle_{\mathcal{L}^2} := \int_D f(x) \overline{g(x)} dx$.

Note that since $\{\phi_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$ are deterministic functions, the stochastic nature of U must be given by the spectral processes $\{\alpha_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$. If $\mathcal{J} = \mathbb{Z}^2$, the expansion (3.9) is simply the spatial Fourier series expansion with temporally varying, stochastic Fourier coefficients.

In practice, one approximates a solution using a finite number of basis functions. It is common to limit the choice of Fourier basis functions to the ones corresponding to the low spatial waves centered around $\mathbf{0}$, a practice also referred to as *low pass filtering*. In that case, we denote the set of spatial wave indices by $\mathcal{J}_K := \{\boldsymbol{j} \in \mathbb{Z}^2 : -\frac{K}{2} + 1 \leq j_i \leq \frac{K}{2},\ i = 1, 2\}$ for some even $K \in \mathbb{N}$.

As mentioned earlier, the usage of the Fourier spectral decomposition is motivated by the consideration of the stochastic forcing term $\dot{W}^{Q_k}$. This motivation is based on the spectral representation theorem for stationary processes. As we will see now, it allows us to approximate the stochastic forcing term via a linear combination of Fourier basis functions.

Recall that we assume $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ to be white in time and spatially a Matérn process with formal covariance function

$$k((t, \boldsymbol{x}), (s, \boldsymbol{y})) = \begin{cases} 0, & \text{if } t \neq s, \\ \dfrac{\sigma_0^2 r}{\rho_0} K_1\left(\dfrac{\|\boldsymbol{x} - \boldsymbol{y}\|}{\rho_0}\right), & \text{if } t = s. \end{cases}$$

Note that for any fixed time t, $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{\boldsymbol{x}}$ is a Matérn process and thus stationary in the spatial domain with spectral density

$$f(\boldsymbol{\xi}) = \frac{\sigma_0^2}{(2\pi)^2}(\rho_0^{-2} + \|\boldsymbol{\xi}\|^2)^{-2}, \quad \boldsymbol{\xi} \in \mathbb{R}^2.$$

Therefore, the spectral representation theorem (theorem 2.5) states that, for any fixed t, there exists a complex-valued, centered Gaussian process $(\eta(t, \boldsymbol{\xi}))_{\boldsymbol{\xi} \in \mathbb{R}^2}$ such that

$$\begin{aligned} \dot{W}^{Q_k}(t, \boldsymbol{x}) &= \int_{\mathbb{R}^2} \exp(i\boldsymbol{\xi} \cdot \boldsymbol{x}) d\eta(t, \boldsymbol{\xi}), \\ &= \lim_{k \to \infty} \sum_{\Delta\boldsymbol{\xi} \in \mathcal{P}_k} \exp\big(i\boldsymbol{\xi}' \cdot \boldsymbol{x}\big) \Delta\eta(t, \boldsymbol{\xi}), \quad (t, \boldsymbol{x}) \in [0, \infty) \times D \text{ fixed.} \end{aligned} \tag{3.10}$$

Here, $(\mathcal{P}_k)_k$ is a sequence of interval partitions with vanishing mesh-size and $\Delta\eta(t, \boldsymbol{\xi})$ denotes the increment of $(\eta(t, \boldsymbol{\xi}))_{\boldsymbol{\xi}}$ over the interval $\Delta\boldsymbol{\xi}$, whereas $\boldsymbol{\xi}' \in \Delta\boldsymbol{\xi}$ for each $\Delta\boldsymbol{\xi}$. For more details on the notation, see section 2.2. The process $(\eta(t, \boldsymbol{\xi}))_{\boldsymbol{\xi}}$ has orthogonal increments $\Delta\eta(t, \boldsymbol{\xi})$ with variances

$$\mathrm{E}[|\Delta\eta(t, \boldsymbol{\xi})|^2] \approx f(\boldsymbol{\xi}')|\Delta\boldsymbol{\xi}|.$$

This motivates a Fourier-type decomposition of the stochastic forcing term δ as follows: consider the partition intervals $\Delta\boldsymbol{\xi}_{\boldsymbol{j}} \subset \mathbb{R}^2$ defined by

$$\Delta\boldsymbol{\xi}_{\boldsymbol{j}} := \left[2\pi\left(\tfrac{j_1}{L} - \tfrac{1}{2L}\right), 2\pi\left(\tfrac{j_1}{L} + \tfrac{1}{2L}\right)\right] \times \left[2\pi\left(\tfrac{j_2}{L} - \tfrac{1}{2L}\right), 2\pi\left(\tfrac{j_2}{L} + \tfrac{1}{2L}\right)\right], \quad \boldsymbol{j} \in \mathbb{Z}^2.$$

These correspond to the even squares with volume $|\Delta\boldsymbol{\xi}_{\boldsymbol{j}}| = \left(\frac{2\pi}{L}\right)^2$, centered around the spatial wave numbers $\boldsymbol{\xi}_{\boldsymbol{j}} = 2\pi(\frac{j_1}{L}, \frac{j_2}{L})$, $\boldsymbol{j} \in \mathbb{Z}^2$. Then, denoting by $\eta_{\boldsymbol{j}}(t) = \Delta\eta(t, \boldsymbol{\xi}_{\boldsymbol{j}})$ the increment of $\eta(t, \boldsymbol{\xi}_{\boldsymbol{j}})$ over $\Delta\boldsymbol{\xi}_{\boldsymbol{j}}$, we approximate the stochastic integral in (3.10) by the sum

$$\dot{W}^{Q_k}(t, \boldsymbol{x}) \approx \dot{W}^{\tilde{Q}}(t, \boldsymbol{x}) := \sum_{\boldsymbol{j} \in \mathcal{J}} \exp(i\boldsymbol{\xi}_{\boldsymbol{j}} \cdot \boldsymbol{x})\eta_{\boldsymbol{j}}(t), \tag{3.11}$$

where $\mathcal{J} \subset \mathbb{Z}^2$ is a set of spatial wave indices and for every fixed t, $\{\eta_{\boldsymbol{j}}(t)), \boldsymbol{j} \in \mathcal{J}\}$ is a family of independent, complex Gaussian variables with

$$\begin{aligned} \mathrm{E}[\eta_{\boldsymbol{j}}(t)] &= 0 \\ \mathrm{E}[\eta_{\boldsymbol{j}}(t)\overline{\eta_{\boldsymbol{j}}(t)}] &= f(\boldsymbol{\xi}_{\boldsymbol{j}})|\Delta\boldsymbol{\xi}_{\boldsymbol{j}}| \\ &=: \tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}}) \end{aligned}$$

Further, since $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{t,\boldsymbol{x}}$ is a real-valued process, a direct consequence of lemma 2.4 is that for all t

$$\begin{cases} \eta_{\boldsymbol{j}}(t) \text{ is real-valued}, & \text{if } \boldsymbol{j} = \boldsymbol{0} \\ \eta_{\boldsymbol{j}}(t) \text{ is proper}, & \text{else.} \end{cases}$$

When letting time t vary again, we assume that for each $\boldsymbol{j} \in \mathcal{J}$, the process $(\eta_{\boldsymbol{j}}(t))_t$ is white in time, in order to maintain the temporally white nature of the stochastic forcing term δ. In total, we asssume that $\{\eta_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$ is an independent family of scaled white noise processes, formally defined by

$$\eta_{\boldsymbol{j}}(t) := \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})}\dot{w}_{\boldsymbol{j}}(t) \tag{3.12}$$

where $(\dot{w}_{\boldsymbol{j}}(t))_t$ denotes standard complex (or real, in the case of $\boldsymbol{j} = \boldsymbol{0}$) white noise process in the sense of remark 2.5.

Ultimately, these observations allow us now to construct a solution to the stochastic convection-diffusion equation with spatially periodic boundary conditions and driven by the approximated forcing term $\dot{W}^{\tilde{Q}}(t, \boldsymbol{x})$:

Theorem 3.2. Let $\mathcal{J} \subset \mathbb{Z}^2$ be a finite set of spatial wave indices and let $\{\eta_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$ be a family of independent Gaussian white noise processes as defined in 3.12. Further, let $\{\alpha_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$ be a family of independent, complex-valued Ornstein-Uhlenbeck processes, given as solutions to the Langevin equations

$$\begin{cases} d\alpha_{\boldsymbol{j}}(t) &= \lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t)dt + \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})}dw_{\boldsymbol{j}}(t), \quad t \in [0, \infty), \\ \alpha_{\boldsymbol{j}}(0) &= \alpha_{\boldsymbol{j},0}, \end{cases} \tag{3.13}$$

where $\{w_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$ is a family of independent complex (or real, in the case of $\boldsymbol{j} = 0$) Wiener processes and $\lambda_{\boldsymbol{j}} := -i\boldsymbol{\mu}^T\boldsymbol{\xi}_{\boldsymbol{j}} - \boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}} - \zeta$.

Then, the process $(\tilde{U}(t,\boldsymbol{x}))_{(t,\boldsymbol{x})\in[0,\infty)\times[0,L]^2}$, given by

$$\tilde{U}(t,\boldsymbol{x}) := \sum_{\boldsymbol{j}\in\mathcal{J}} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x}), \tag{3.14}$$

solves the approximated stochastic convection-diffusion equation

$$\begin{cases} \frac{\partial}{\partial t}\tilde{U}(t,\boldsymbol{x}) & = -\boldsymbol{\mu}\cdot\nabla\tilde{U}(t,x) + \nabla\cdot\boldsymbol{\Sigma}\nabla\tilde{U}(t,\boldsymbol{x}) - \zeta\tilde{U}(t,\boldsymbol{x}) + \dot{W}^{\tilde{Q}}(t,\boldsymbol{x}), \\ \tilde{U}(0,\boldsymbol{x}) & = \sum_{\boldsymbol{j}\in\mathcal{J}} \phi_{\boldsymbol{j}}(\boldsymbol{x})\alpha_{\boldsymbol{j}}(0) \end{cases} \tag{3.15}$$

with spatially periodic boundary conditions and stochastic forcing term

$$\dot{W}^{\tilde{Q}}(t,\boldsymbol{x}) = \sum_{\boldsymbol{j}\in\mathcal{J}} \eta_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x}). \tag{3.16}$$

Proof. The proof is based on Sigrist et al. [2015], proposition 1. Note that the result in Sigrist et al. [2015] is lacking the volume scaling factor in $\tilde{f}(\boldsymbol{\xi_j})$.

For $\phi_{\boldsymbol{j}}(\boldsymbol{x}) = \exp(i\boldsymbol{\xi_j}\cdot\boldsymbol{x})$, it holds:

$$\begin{aligned} \nabla\phi_{\boldsymbol{j}}(\boldsymbol{x}) &= i\boldsymbol{\xi}_{\boldsymbol{j}}\phi_{\boldsymbol{j}}(\boldsymbol{x}) \\ \nabla\cdot\nabla\phi_{\boldsymbol{j}}(\boldsymbol{x}) &= -\boldsymbol{\xi_j}\cdot\boldsymbol{\xi_j}\phi_{\boldsymbol{j}}(\boldsymbol{x}). \end{aligned}$$

Thus, the right-hand side of (3.15) becomes:

$$\begin{aligned} &\sum_{\boldsymbol{j}\in\mathcal{J}} \alpha_{\boldsymbol{j}}(t)\left[-\boldsymbol{\mu}^T\nabla\phi_{\boldsymbol{j}}(\boldsymbol{x}) + \nabla^T\boldsymbol{\Sigma}\ \nabla\phi_{\boldsymbol{j}}(\boldsymbol{x}) - \zeta\phi_{\boldsymbol{j}}(\boldsymbol{x})\right] + \dot{W}^{\tilde{Q}}(t,\boldsymbol{x}) \\ &= \sum_{\boldsymbol{j}\in\mathcal{J}} \alpha_{\boldsymbol{j}}(t)\left[-i\boldsymbol{\mu}^T\boldsymbol{\xi_j}\phi_{\boldsymbol{j}}(\boldsymbol{x}) - \boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\ \boldsymbol{\xi_j}\phi_{\boldsymbol{j}}(\boldsymbol{x}) - \zeta\phi_{\boldsymbol{j}}(\boldsymbol{x})\right] + \dot{W}^{\tilde{Q}}(t,\boldsymbol{x}) \\ &= \sum_{\boldsymbol{j}\in\mathcal{J}} \left[\lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x})\right] + \dot{W}^{\tilde{Q}}(t,\boldsymbol{x}). \end{aligned}$$

where $\lambda_{\boldsymbol{j}} = -i\boldsymbol{\mu}^T\boldsymbol{\xi_j} - \boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\ \boldsymbol{\xi_j} - \zeta$.

In regards to the left-hand side, we formally write:

$$\begin{aligned} \frac{d\alpha_{\boldsymbol{j}}(t)}{dt} &= \lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t) + \sqrt{\tilde{f}(\boldsymbol{\xi_j})}\dot{w}_{\boldsymbol{j}}(t) \\ &= \lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t) + \eta_{\boldsymbol{j}}(t). \end{aligned}$$

We thus get

$$\begin{aligned} \frac{\partial}{\partial t}\sum_{\boldsymbol{j}\in\mathcal{J}} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x}) &= \sum_{\boldsymbol{j}\in\mathcal{J}} (\lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t) + \eta_{\boldsymbol{j}}(t))\phi_{\boldsymbol{j}}(\boldsymbol{x}) \\ &= \sum_{\boldsymbol{j}\in\mathcal{J}} \left[\lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x})\right] + \dot{W}^{\tilde{Q}}(t,\boldsymbol{x}), \end{aligned}$$

which proves the claim. □

Note that theorem 3.2 transforms the stochastic partial differential equation into a family of ordinary stochastic differential equation, namely the complex Langevin equations

$$\begin{cases} d\alpha_{\boldsymbol{j}}(t) &= \lambda_{\boldsymbol{j}}\alpha_{\boldsymbol{j}}(t)dt + \sqrt{\tilde{f}(\boldsymbol{\xi_j})}dw_{\boldsymbol{j}}(t), \quad t \in (0,T], \\ \alpha_{\boldsymbol{j}}(0) &= \alpha_{\boldsymbol{j},0}. \end{cases}$$

It follows from lemma 2.5 and 2.6 that for each $\boldsymbol{j} \in \mathcal{J}, \boldsymbol{j} \neq \mathbf{0}$, $\alpha_{\boldsymbol{j}}$ is a proper complex Gaussian process given by

$$\begin{aligned} \alpha_{\boldsymbol{j}}(t) &= \exp(\lambda_{\boldsymbol{j}}t)\alpha_{\boldsymbol{j},0} + \sqrt{\tilde{f}(\boldsymbol{\xi_j})}\int_0^t \exp(\lambda_{\boldsymbol{j}}(t-s))dw_{\boldsymbol{j}}(s) \\ \lambda_{\boldsymbol{j}} &= -(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} + \zeta) - i\boldsymbol{\mu}^T\boldsymbol{\xi_j} \end{aligned} \tag{3.17}$$

that converges weakly to its invariant distribution

$$\lim_{t\to\infty}\alpha_{\boldsymbol{j}}(t) \overset{d}{=} \alpha_{\boldsymbol{j},\infty} \sim \mathcal{CN}\left(0, \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} + \zeta)}\right). \tag{3.18}$$

Likewise, for the special case $\boldsymbol{j} = \mathbf{0}$, we have that $\alpha_{\mathbf{0}}$ is a real-valued process with

$$\begin{aligned} \alpha_{\mathbf{0}}(t) &= \exp(-\zeta t)\alpha_{\mathbf{0},0} + \sqrt{\tilde{f}(\mathbf{0})}\int_0^t \exp(-\zeta(t-s))dw_{\mathbf{0}}(s) \\ \lim_{t\to\infty}\alpha_{\mathbf{0}}(t) &\overset{d}{=} \alpha_{\mathbf{0},\infty} \sim \mathcal{N}\left(0, \frac{1}{2\zeta}\tilde{f}(\mathbf{0})\right). \end{aligned} \tag{3.19}$$

Remark 3.2. The convergence rate of the spectral process $(\alpha_{\boldsymbol{j}}(t))_t$ to its invariant distribution $\alpha_{\boldsymbol{j},\infty}$ is influenced by its corresponding spatial wave $\boldsymbol{\xi_j}$ in the following way: By symmetric positive definiteness of $\boldsymbol{\Sigma}$, we have

$$\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} \geq \boldsymbol{\Sigma}_{\min}\|\boldsymbol{\xi_j}\|^2$$

where $\boldsymbol{\Sigma}_{\min} > 0$ is the smallest eigenvalue of $\boldsymbol{\Sigma}$. Therefore, we get

$$\begin{aligned} |\mathrm{E}[\alpha_{\boldsymbol{j}}(t)]| &= |\exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} + \zeta)t\big)| \\ &\leq \exp\big(-\boldsymbol{\Sigma}_{\min}\|\boldsymbol{\xi_j}\|^2 t\big)\exp(-\zeta t) \end{aligned}$$

and similarly

$$\begin{aligned} |\mathrm{Var}[\alpha_{\boldsymbol{j}}(t)] - \mathrm{Var}[\alpha_{\boldsymbol{j},\infty}]| &\leq \exp\big(-2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} + \zeta)t\big)\left[\sigma_{\boldsymbol{j},0}^2 + \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} + \zeta)}\right] \\ &\leq \exp\big(-2\boldsymbol{\Sigma}_{\min}\|\boldsymbol{\xi_j}\|^2 t\big)\exp(-2\zeta t)\left[\sigma_{\boldsymbol{j},0}^2 + \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi_j} + \zeta)}\right]. \end{aligned}$$

Note further that $\tilde{f}(\boldsymbol{\xi_j}) \in \mathcal{O}(\|\boldsymbol{\xi_j}\|^{-4})$ and therefore

$$\mathrm{Var}[\alpha_{\boldsymbol{j},\infty}] \in \mathcal{O}(\|\boldsymbol{\xi_j}\|^{-4}).$$

From this we can conclude that for high frequency spatial waves $\boldsymbol{\xi_j}$, the spectral process $(\alpha_{\boldsymbol{j}}(t))_t$ rapidly converges to its invariant distribution with variance only marginally larger than 0, therefore essentially letting such coefficients vanish in practical applications.

The weak convergence of the spectral processes implies the following temporal convergence property of the spatiotemporal process $\tilde{U}$:

Corollary 3.1. Let $\mathcal{J} \subset \mathbb{Z}^2$ be finite and $\tilde{U}$ be the process $\tilde{U}(t,\boldsymbol{x}) = \sum_{\boldsymbol{j}\in\mathcal{J}} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x})$ as given in theorem 3.2. Then, for $t \longrightarrow \infty$, $\tilde{U}(t,\boldsymbol{x})$ converges weakly to a spatially stationary Gaussian process $\tilde{U}_\infty(\boldsymbol{x})$ with covariance given by

$$\mathrm{Cov}[\tilde{U}_\infty(\boldsymbol{x}), \tilde{U}_\infty(\boldsymbol{y})] = \sum_{\boldsymbol{j}\in\mathcal{J}} \phi_{\boldsymbol{j}}(\boldsymbol{x}-\boldsymbol{y}) \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)}. \tag{3.20}$$

Further, if $\alpha_{\boldsymbol{j},0} \overset{d}{=} \alpha_{\boldsymbol{j},\infty}$ for all $\boldsymbol{j} \in \mathcal{J}$, the approximated solution $(\tilde{U}(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ is stationary both in space and time with covariance function

$$\mathrm{Cov}[\tilde{U}(t,\boldsymbol{x}), \tilde{U}(s,\boldsymbol{y})] = \sum_{\boldsymbol{j}\in\mathcal{J}} \phi_{\boldsymbol{j}}(\boldsymbol{x}-\boldsymbol{y}) \exp(\lambda_j(t-s)) \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \tag{3.21}$$

for all $s \le t, \boldsymbol{x}, \boldsymbol{y} \in [0,L]^2$. In particular $\mathrm{Cov}[\tilde{U}(t,\boldsymbol{x}), \tilde{U}(t,\boldsymbol{y})] = \mathrm{Cov}[\tilde{U}_\infty(\boldsymbol{x}), \tilde{U}_\infty(\boldsymbol{y})]$ for all t and $\boldsymbol{x}, \boldsymbol{y} \in [0,L]^2$.

Proof. Let $\tilde{U}$ be given as stated. It is clear that $\tilde{U}$ is a Gaussian process as the finite sum of independent Gaussian processes. It is centered and, using the independence of $\alpha_{\boldsymbol{j}}$, its covariance is given by

$$\begin{aligned}\mathrm{Cov}[\tilde{U}(t,\boldsymbol{x}), \tilde{U}(s,\boldsymbol{y})] &= \sum_{\boldsymbol{j}\in\mathcal{J}} \mathrm{E}[\alpha_{\boldsymbol{j}}(t)\overline{\alpha_{\boldsymbol{j}}(s)}]\phi_{\boldsymbol{j}}(\boldsymbol{x})\overline{\phi_{\boldsymbol{j}}(\boldsymbol{y})} \\ &= \sum_{\boldsymbol{j}\in\mathcal{J}} \mathrm{E}[\alpha_{\boldsymbol{j}}(t)\overline{\alpha_{\boldsymbol{j}}(s)}]\phi_{\boldsymbol{j}}(\boldsymbol{x}-\boldsymbol{y}).\end{aligned}$$

Then, the first claim follows directly from the weak convergence of the Ornstein-Uhlenbeck processes $\{\alpha_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}\}$ with $\lim_{t\to\infty} \mathrm{E}[\alpha_{\boldsymbol{j}}(t)\overline{\alpha_{\boldsymbol{j}}(t)}] = \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)}$ for all $\boldsymbol{j} \in \mathcal{J}$.

The second claim follows from the fact that, if $\alpha_{\boldsymbol{j},0} \overset{d}{=} \alpha_{\boldsymbol{j},\infty}$, the process $(\alpha_{\boldsymbol{j}}(t))_t$ is stationary with covariance function

$$\mathrm{E}[\alpha_{\boldsymbol{j}}(t)\overline{\alpha_{\boldsymbol{j}}(s)}] = \exp(\lambda_{\boldsymbol{j}}(t-s)) \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)}$$

for all $s \le t$. This is a direct consequence from equation (2.75) in the proof of lemma 2.6. □

The previous lemma showed that the process $\tilde{U}$ converges to a spatially stationary process, independently of the initial condition, whereas in the particular case that the spectral processes $\alpha_{\boldsymbol{j}}$ are initiated with their invariant distribution, the process $\tilde{U}$ is stationary both in space and time.

We end this section showing that in this stationary case, the approximation $\tilde{U}$ indeed converges (weakly) to the true stationary solution $(U(t,\boldsymbol{x}))_{\boldsymbol{x}}$ on the spatial domain $[0,L]^2$ with periodic boundary conditions. For this, we consider the true stationary solution as given in theorem 3.1, defined on $\mathbb{R}^2$, as a periodic process on $[0,L]^2$ by mapping $\mathbb{R}^2$ onto the 2-Torus [2] via

$$\begin{aligned} \psi : \mathbb{R}^2 &\longrightarrow \mathbb{T}^2 \cong [0,L]^2 \\ (x_1,x_2) &\mapsto (\exp\big(2\pi i\tfrac{x_1}{L}\big), \exp\big(2\pi i\tfrac{x_2}{L}\big)) \end{aligned} \tag{3.22}$$

In that case, the Fourier series expansion of the periodic processes covariance function gives the following result:

Lemma 3.1. Let $\tilde{U}_K$ be the approximated process $\tilde{U}_K(t,\boldsymbol{x}) = \sum_{\boldsymbol{j}\in\mathcal{J}_K} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x})$ as given in theorem 3.2. Further, let the initial states $\alpha_{\boldsymbol{j}}(0)$ be distributed as per the invariant distributions

$$\alpha_{\boldsymbol{j}}(0) \sim \begin{cases} \mathcal{CN}\left(0, \frac{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}}+\zeta)}\right), & \text{if } \boldsymbol{j}\neq\boldsymbol{0} \\ \mathcal{N}\left(0, \frac{1}{2\zeta}\tilde{f}(\boldsymbol{0})\right), & \text{if } \boldsymbol{j}=\boldsymbol{0}, \end{cases} \tag{3.23}$$

such that the processes $(\alpha_{\boldsymbol{j}}(t))_t$ are stationary as given in equations (3.18) and (3.19). Assume the stationary solution to the stochastic-convection diffusion equation $(U(t,\boldsymbol{x}))_{t,\boldsymbol{x}\in\mathbb{R}^2}$ is $[0,L]^2$-periodic. Then we have

$$|C(s,\boldsymbol{r}) - \tilde{C}_K(s,\boldsymbol{r})| \xrightarrow{N\to\infty} 0,$$

where C and $\tilde{C}_K$ denote the covariance functions of U and $\tilde{U}_K$ respectively. Thus, the approximated solution converges weakly to the true stationary solution:

$$\lim_{N\to\infty} \tilde{U}_K(t,\boldsymbol{x}) \overset{d}{=} U(t,\boldsymbol{x}), \tag{3.24}$$

for all $(t,\boldsymbol{x}) \in [0,T]\times[0,L]^2$.

Proof. The proof is based on Sigrist et al. [2015], proposition 2.

Identify $U(t,\boldsymbol{x})$ as a $[0,L]^2$-periodic function. Then it suffices to consider its covariance function $C(s,\boldsymbol{r})$ on the bounded spatial domain $\boldsymbol{r}\in[0,L]^2$. For any fixed $s\in\mathbb{R}_{\geq 0}$ we may represent $C(s,\boldsymbol{r})$ via a Fourier series expansion in the spatial domain, given by

$$C(s,\boldsymbol{r}) = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} a_{\boldsymbol{j}}(s)\exp(i\boldsymbol{\xi}_{\boldsymbol{j}}\cdot\boldsymbol{r})$$

[2] The rationale behind this is that any bounded square $[0,L]^2$ is homeomorphic to the 2-Torus $\mathbb{T}^2 := \{(\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}^4 : \|\boldsymbol{x}\|,\|\boldsymbol{y}\| = 1\}$ with the respective homeomorphism given by $\psi_{|[0,L]^2}$ defined as in (3.22) restriced on $[0,L]^2$. More details on this are given in section 4.1.

where

$$a_{\boldsymbol{j}}(s) = \frac{1}{L^2} \int_{[0,L]^2} C(s, \boldsymbol{r}) \exp(-i\boldsymbol{\xi_j} \cdot \boldsymbol{r}) d\boldsymbol{r}.$$

Note that the spectral density $g(\omega, \boldsymbol{\xi})$ of $C(s, \boldsymbol{r})$, evaluated at $\boldsymbol{\xi} = \boldsymbol{\xi_j}$, is given by

$$g(\omega, \boldsymbol{\xi_j}) = \frac{1}{2\pi} \left(\frac{2\pi}{L}\right)^2 \int_{\mathbb{R}} a_{\boldsymbol{j}}(s) \exp(-i\omega s) ds.$$

Thus, applying the inverse Fourier transform in the temporal domain gives

$$\begin{aligned} a_{\boldsymbol{j}}(s) &= \left(\frac{2\pi}{L}\right)^2 \int_{\mathbb{R}} g(\omega, \boldsymbol{\xi_j}) \exp(i\omega s) d\omega \\ &= |\Delta \boldsymbol{\xi_j}| \frac{f(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \exp(\lambda_{\boldsymbol{j}} s) \\ &= \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \exp(\lambda_{\boldsymbol{j}} s) \end{aligned} \tag{3.25}$$

where the integral term equals the integration in remark 3.1 and $\left(\frac{2\pi}{L}\right)^2 = |\Delta \boldsymbol{\xi_j}|$ is the volume element corresponding to the spatial wave interval decomposition as introduced in the beginning of this section. In total, we have

$$C(s, \boldsymbol{r}) = \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \frac{\tilde{f}(\boldsymbol{\xi_j})}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \exp(\lambda_{\boldsymbol{j}} s) \exp(i\boldsymbol{\xi_j} \cdot \boldsymbol{r}). \tag{3.26}$$

Note that $\tilde{C}_K(s, \boldsymbol{r})$ is given in (3.21) and we receive:

$$\begin{aligned} |C(s, \boldsymbol{r}) - \tilde{C}_K(s, \boldsymbol{r})| &\leq \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \left| \left(\tilde{f}(\boldsymbol{\xi_j}) \frac{\exp(\lambda_{\boldsymbol{j}} s)}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \exp(i\boldsymbol{\xi_j} \cdot r) \right) (1 - \mathbb{1}_{\mathcal{J}_K}) \right| \\ &\leq \sum_{\boldsymbol{j} \in \mathbb{Z}^2 \setminus \mathcal{J}_K} \left(\tilde{f}(\boldsymbol{\xi_j}) \frac{1}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \right) \xrightarrow{N \to \infty} 0, \end{aligned}$$

where we used that $|\exp(\lambda_{\boldsymbol{j}} s)| \leq 1$ due to the fact that $\mathrm{Re}(\lambda_{\boldsymbol{j}}) = -(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta) < 0$ and that

$$\sum_{\boldsymbol{j} \in \mathbb{Z}^2} \left(\tilde{f}(\boldsymbol{\xi_j}) \frac{1}{2(\boldsymbol{\xi_j}^T \boldsymbol{\Sigma} \boldsymbol{\xi_j} + \zeta)} \right) = C(0, 0)$$

is the marginal variance of $U(t, \boldsymbol{x})$ and thus a convergent series. □

3.3. Filtering, Smoothing and Parameter Inference Setup

We now turn our attention to the filtering and smoothing problem, and, consequently, to the problem of model parameter inference. Assume we are given noise-corrupted observations at discrete times of the spectral approximation to the stochastic convection-diffusion equation as derived in the previous section. As usual, we aim to estimate the underlying process state based on these measurements and subsequently, infer unknown model parameters.

The key ideas in this section are the following two: firstly, the temporal evolution as well as the stochastic nature of the Fourier approximated solution is solely based on the linearly evolving spectral processes. Therefore, any state estimation or model parameter inference can be done efficiently using the Kalman filter once we translate the problem to the spectral domain.

Secondly, the numerical integration involved in the transformation to the spectral domain can be formulated as the DFT and therefore be efficiently computed using the FFT. This demands an even discretization of the spatial domain, whereas the temporal evolution of the process can be kept continuous and filtering can therefore be done in the continuous-discrete formulation. Overall, this results in the formulation of the specral Kalman filter as derived in the following.

Let $(\tilde{U}(t,\boldsymbol{x}))_{t,\boldsymbol{x}}$ be the approximated solution to the stochastic convection-diffusion equation, given by

$$\tilde{U}(t,\boldsymbol{x}) = \sum_{\boldsymbol{j}\in\mathcal{J}_K} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x}) \tag{3.27}$$

with $\{\phi_{\boldsymbol{j}} : \boldsymbol{j} \in \mathcal{J}\}, \{\alpha_{\boldsymbol{j}} : \boldsymbol{j} \in \mathcal{J}\}$ denoting the Fourier basis functions and spectral processes as thoroughly discussed in the previous section and $\mathcal{J}_K = \{\boldsymbol{j} \in \mathbb{Z}^2 : -\frac{K}{2}+1 \leq j_i \leq \frac{K}{2},\ i = 1,2\}$ denoting the set of the first K^2 spatial waves centered around $\mathbf{0}$ for an even K. Assume $\tilde{U}$ is evaluated on an even grid

$$\mathcal{D} := \{\boldsymbol{x}_{n,m} = L(\tfrac{n}{N}, \tfrac{m}{N}) : n, m = 0, ..., N-1\} \tag{3.28}$$

of the spatial domain $D = [0, L]^2$. For simplicity, we assume that N is even.

For every $t \in [0, T]$, denote by $\tilde{\boldsymbol{U}}(t)$ the stacked vector of the spatially discretized process

$$\tilde{\boldsymbol{U}}(t) := (\tilde{U}(t, \boldsymbol{x}_{n,m}))_{n,m} \in \mathbb{R}^{N^2}. \tag{3.29}$$

Evaluating the spatially discretized process $\tilde{U}(t, \boldsymbol{x}_{n,m})$, we receive

$$\begin{aligned}\tilde{U}(t, \boldsymbol{x}_{n,m}) &= \sum_{\boldsymbol{j}\in\mathcal{J}_K} \alpha_{\boldsymbol{j}}(t) \exp(i\boldsymbol{\xi}_{\boldsymbol{j}} \cdot \boldsymbol{x}_{n,m}) \\ &= \sum_{\boldsymbol{j}\in\mathcal{J}_K} \alpha_{\boldsymbol{j}}(t) \exp\Big(2\pi i(\tfrac{j_1}{N}n + \tfrac{j_2}{N}m)\Big),\end{aligned}$$

which for $N = K$ equals the inverse discrete Fourier transform. In other words, $\tilde{\boldsymbol{U}}(t)$ and the stacked vector of spectral coefficients

$$\boldsymbol{\alpha}(t) := (\alpha_{\boldsymbol{j}}(t_i))_{\boldsymbol{j}\in\mathcal{J}_N} \in \mathbb{C}^{N^2}$$

form a DFT-IDFT pair and we write

$$\tilde{\boldsymbol{U}}(t) = \boldsymbol{\Phi}^{-1}\boldsymbol{\alpha}(t) \in \mathbb{R}^{N^2}, \tag{3.30}$$

where $\boldsymbol{\Phi}$ denotes the discrete Fourier transform operator. As noted in the introduction, the temporal evolution of the process $\tilde{\boldsymbol{U}}$ is dependent on the spectral processes $\boldsymbol{\alpha}$ only. Recall that $(\alpha_{\boldsymbol{j}}(t))_t$ is a Markov process such that, for any time lag $\Delta \in \mathbb{R}_+$, we have

$$\begin{aligned}\alpha_{\boldsymbol{j}}(t+\Delta) &= \exp(\lambda_{\boldsymbol{j}}\Delta)\alpha_{\boldsymbol{j}}(t) + \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})}\int_t^{t+\Delta}\exp(\lambda_{\boldsymbol{j}}(t+\Delta-s))dw_{\boldsymbol{j}}(s)\\ &=: \exp(\lambda_{\boldsymbol{j}}\Delta)\alpha_{\boldsymbol{j}}(t) + \nu_{\boldsymbol{j}}(\Delta).\end{aligned} \tag{3.31}$$

The family $\{\nu_{\boldsymbol{j}}(\Delta) : \boldsymbol{j} \in \mathcal{J}_N\}$ is independent and from lemma 2.3 it follows that, for every $\boldsymbol{j} \in \mathcal{J}_N$, $\nu_{\boldsymbol{j}}(\Delta)$ is a proper complex (or real, in the case of $\boldsymbol{j} = 0$) Gaussian variable with with mean zero and variance

$$\begin{aligned}q_{\boldsymbol{j}}(\Delta) &:= \mathrm{E}[\nu_{\boldsymbol{j}}(\Delta)\overline{\nu_{\boldsymbol{j}}(\Delta)}]\\ &= \tfrac{\tilde{f}(\boldsymbol{\xi}_j)}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}}+\zeta)}\left[1-\exp\big(-2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}}+\zeta)\Delta\big)\right].\end{aligned} \tag{3.32}$$

Let $(\boldsymbol{V}(t_i))_i$ denote the noise-corrupted observations of the spatially discretized process $\tilde{\boldsymbol{U}}$ at discrete observation times $t_i \le t_{i+1}, i = 1, ..., I$, defined by

$$\boldsymbol{V}(t_i) := \tilde{\boldsymbol{U}}(t_i) + \boldsymbol{\epsilon}(t_i) \in \mathbb{R}^{N^2},$$

where $(\boldsymbol{\epsilon}(t_i))_{i=1,...,I}$ denotes temporally and spatially white Gaussian measurement noise such that

$$\boldsymbol{\epsilon}(t_i) \sim \mathcal{N}(0, \sigma_\epsilon^2\boldsymbol{E}_{N^2})$$

on the spatial grid $\mathcal{D}$ for each t_i. Denote by $\Delta_i = |t_{i+1} - t_i|$ the time lags between observation times. Then, we can formulate the overall discretized process system on $\{t_1, ..., t_I\} \times \mathcal{D}$ as a linear state space model as described in section 2.3.2, given by

$$\begin{aligned}\textbf{Process Level: } &\boldsymbol{\alpha}(t_{i+1}) = \boldsymbol{F}(\Delta_i)\boldsymbol{\alpha}(t_i) + \boldsymbol{\nu}(\Delta_i)\\ \textbf{Observation Level: } &\boldsymbol{V}(t_{i+1}) = \boldsymbol{\Phi}^{-1}\boldsymbol{\alpha}(t_{i+1}) + \boldsymbol{\epsilon}(t_{i+1}), \quad i \in \{1, ..., I\}.\end{aligned} \tag{3.33}$$

The matrix $\boldsymbol{\Phi}^{-1} \in \mathbb{C}^{N^2\times N^2}$ is the inverse DFT operator, transforming the spatiotemporal process $\tilde{\boldsymbol{U}}$ corrupted by measurement noise $\boldsymbol{\epsilon}(t_{i+1})$ into the spectral domain, and for each $i \in \{1, ..., I\}$:

- $\boldsymbol{\alpha}(t_i) = (\alpha_{\boldsymbol{j}}(t_i) : \boldsymbol{j} \in \mathcal{J}_N) \in \mathbb{C}^{N^2}$ is the stacked vector of Fourier coefficients,
- $\boldsymbol{F}(\Delta_i) = \mathbf{diag}\,(\exp(\lambda_{\boldsymbol{j}}\Delta_i) : \boldsymbol{j} \in \mathcal{J}_N) \in \mathbb{C}^{N^2 \times N^2}$ is the discretized propagation matrix with $\lambda_{\boldsymbol{j}} = -i\boldsymbol{\mu}^T\boldsymbol{\xi}_{\boldsymbol{j}} - \boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}} - \zeta$ parametrized by the continuous time dynamical process parameters and the corresponding spatial wave $\boldsymbol{\xi}_{\boldsymbol{j}}$,
- $\boldsymbol{\nu}(\Delta_i) = (\nu_{\boldsymbol{j}}(\Delta_i) : \boldsymbol{j} \in \mathcal{J}_N) \in \mathbb{C}^{N^2}$ is the discrete Gaussian forcing term with elements

$$\nu_{\boldsymbol{j}}(\Delta_i) := \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})} \int_{t_i}^{t_{i+1}} \exp(\lambda_{\boldsymbol{j}}(t_{i+1} - s))dw_{\boldsymbol{j}}(s)$$

  that arises from the discretization of the independent Ornstein-Uhlenbeck processes. It has diagonal covariance matrix $\boldsymbol{Q}(\Delta_i) = \mathbf{diag}\,(q_{\boldsymbol{j}}(\Delta_i) : \boldsymbol{j} \in \mathcal{J}_N) \in \mathbb{R}_+^{N^2 \times N^2}$ and is temporally white due to the martingale property of the stochastic integral: $E[\boldsymbol{\nu}(\Delta_i)\overline{\boldsymbol{\nu}(\Delta_k)}] = 0$ for all $i \neq k$.

The filtering and smoothing problems can therefore be solved using the complex-valued Kalman filter in algorithm 2.1 and Kalman smoother in algorithm 2.2, though the next section provides a computationally more efficient formulation of the Kalman filter for this specific model setup. Note that this state-space model aligns with the real-valued one as given in proposition 3 of Sigrist et al. [2015], who derived it based on the real DFT. For notational simplicity, we have decided to give the more general, complex-valued formulation.

3.3.1. The Spectral Kalman Filter

Recall that a single iteration of the Kalman filter is of cubic computational complexity $\mathcal{O}((N^2)^3)$ in regards to the number of spatial grid points N^2. In our case however, we can optimize these costs, abusing the fact that the spectal processes $(\alpha_{\boldsymbol{j}})_{\boldsymbol{j}}$ are assumed to be independent and what is previously the measurement matrix in the Kalman filter specification is now given by the IDFT matrix. We take advantage of these facts to derive the *spectral Kalman filter* which was presented in Sigrist et al. [2015] without further detail. Note that our version differs by a prefactor of N^2 in the measurement updates.

Applying the DFT, rewrite the observation level as

$$\textbf{Observation Level: } \hat{\boldsymbol{V}}(t_{i+1}) = \boldsymbol{\alpha}(t_{i+1}) + \hat{\boldsymbol{\epsilon}}(t_{i+1}) \in \mathbb{C}^{N^2}, \tag{3.34}$$

where $\hat{\boldsymbol{V}}(t_{i+1}) = \boldsymbol{\Phi}\boldsymbol{V}(t_{i+1})$ and likewise $\hat{\boldsymbol{\epsilon}}(t_{i+1}) = \boldsymbol{\Phi}\boldsymbol{\epsilon}(t_{i+1})$. Note that $\hat{\boldsymbol{\epsilon}}(t_{i+1})$ is a centered Gaussian noise vector, whose elements $\hat{\epsilon}_{\boldsymbol{j}}(t_{i+1})$ are given by

$$\hat{\epsilon}_{\boldsymbol{j}}(t_{i+1}) = \tfrac{1}{N^2} \sum_{n,m} \epsilon_{n,m}(t_{i+1}) \exp\Big(-2\pi i(\tfrac{j_1}{N}n + \tfrac{j_2}{N}m)\Big).$$

Using the orthogonality of $(\epsilon_{n,m}(t_{i+1}))_{n,m}$ and the discrete Fourier basis functions, their (co-)variance reduces to

$$\begin{aligned}\mathrm{E}[\hat{\epsilon}_{\boldsymbol{j}}(t_{i+1})\overline{\hat{\epsilon}_{\boldsymbol{j}'}(t_{i+1})}] &= \tfrac{1}{(N)^4}\sigma_\epsilon^2 \sum_{n,m} \exp\Big(-2\pi i(\tfrac{j_1-j_1'}{N}n + \tfrac{j_2-j_2'}{N}m)\Big) \\ &= \begin{cases} \frac{\sigma_\epsilon^2}{N^2}, & \text{if } (j_1,j_2) = (j_1',j_2'), \\ 0, & \text{else.} \end{cases}\end{aligned} \tag{3.35}$$

A likewise computation shows that the pseudo-(co-)variance $\mathrm{E}[\hat{\epsilon}_{\boldsymbol{j}}(t_{i+1})\hat{\epsilon}_{\boldsymbol{j}'}(t_{i+1})]$ vanishes for all $\boldsymbol{j}, \boldsymbol{j}' \in \mathcal{J}_N$, $\boldsymbol{j} \neq \mathbf{0}$, whereas in the special case $\boldsymbol{j} = \mathbf{0}$, $\hat{\epsilon}_{\mathbf{0}}(t_{i+1})$ remains real-valued Gaussian with variance

$$\mathrm{E}[\hat{\epsilon}_{\mathbf{0}}(t_{i+1})\hat{\epsilon}_{\mathbf{0}}(t_{i+1})] = \tfrac{\sigma_\epsilon^2}{N^2}. \tag{3.36}$$

In total we can conclude that the measurement noise $\hat{\boldsymbol{\epsilon}}(t_{i+1})$ remains a Gaussian white noise vector in the spectral domain with uncorrelated components $(\epsilon_{\boldsymbol{j}}(t_{i+1}))_{\boldsymbol{j}}$, which are respectively proper complex Gaussian for $\boldsymbol{j} \neq \mathbf{0}$ and real Gaussian for $\boldsymbol{j} = \mathbf{0}$. In particular, this shows that the filtering of the spectral coefficients $(\alpha_{\boldsymbol{j}})$ corresponding to differing spatial wave numbers can be done independently and we now turn our attention to their individual filtering formulation.

For any t_i, let $m_{\boldsymbol{j}}^-(t_i), K_{\boldsymbol{j}}^-(t_i), m_{\boldsymbol{j}}(t_i), K_{\boldsymbol{j}}(t_i)$ respectively denote the predicted and updated mean and variance of $\alpha_{\boldsymbol{j}}(t_i)$ as defined in the Kalman filter formulation in subsection 2.3.2. A simple computation shows that the prediction step is then given by

$$\begin{aligned} m_{\boldsymbol{j}}^-(t_{i+1}) &= \exp(\lambda_{\boldsymbol{j}}\Delta_i) m_{\boldsymbol{j}}(t_i) \\ &=: f_{\boldsymbol{j}}(\Delta_i) m_{\boldsymbol{j}}(t_i) \\ K_{\boldsymbol{j}}^-(t_{i+1}) &= \exp\big(-2\Delta_i(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)\big) K_{\boldsymbol{j}}(t_i) + q_{\boldsymbol{j}}(\Delta_i) \\ &=: h_{\boldsymbol{j}}(\Delta_i) K_{\boldsymbol{j}}(t_i) + q_{\boldsymbol{j}}(\Delta_i). \end{aligned} \tag{3.37}$$

Further, the Kalman gain reduces to

$$G_{\boldsymbol{j}}(t_{i+1}) = K_{\boldsymbol{j}}^-(t_{i+1}) \left(K_{\boldsymbol{j}}^-(t_{i+1}) + \tfrac{\sigma_\epsilon^2}{N^2}\right)^{-1},$$

from which we get the variance update

$$\begin{aligned} K_{\boldsymbol{j}}(t_{i+1}) &= K_{\boldsymbol{j}}^-(t_{i+1}) - K_{\boldsymbol{j}}^-(t_{i+1}) \left(K_{\boldsymbol{j}}^-(t_{i+1}) + \tfrac{\sigma_\epsilon^2}{N^2}\right)^{-1} K_{\boldsymbol{j}}^-(t_{i+1}) \\ &= \left(K_{\boldsymbol{j}}^-(t_{i+1})^{-1} + \tfrac{N^2}{\sigma_\epsilon^2}\right)^{-1} \end{aligned} \tag{3.38}$$

where we used the identiy $a - \frac{a^2}{a+b} = (a^{-1} + b^{-1})^{-1}$ for any $a > 0, b > 0$. Lastly, with

$b(a+b)^{-1} = a^{-1}(a^{-1}+b^{-1})^{-1}$, the adjusted mean update becomes

$$\begin{aligned} m_{\boldsymbol{j}}(t_{i+1}) &= m^-_{\boldsymbol{j}}(t_{i+1}) + K^-_{\boldsymbol{j}}(t_{i+1})\left(K^-_{\boldsymbol{j}}(t_{i+1}) + \tfrac{\sigma^2_\epsilon}{N^2}\right)^{-1}\left(\hat{V}_{\boldsymbol{j}}(t_{i+1}) - m^-_{\boldsymbol{j}}(t_{i+1})\right) \\ &= m^-_{\boldsymbol{j}}(t_{i+1}) + \tfrac{N^2}{\sigma^2_\epsilon}\left(K^-_{\boldsymbol{j}}(t_{i+1})^{-1} + \tfrac{N^2}{\sigma^2_\epsilon}\right)^{-1}\left(\hat{V}_{\boldsymbol{j}}(t_{i+1}) - m^-_{\boldsymbol{j}}(t_{i+1})\right) \\ &= m^-_{\boldsymbol{j}}(t_{i+1}) + \tfrac{N^2}{\sigma^2_\epsilon}K_{\boldsymbol{j}}(t_{i+1})\left(\hat{V}_{\boldsymbol{j}}(t_{i+1}) - m^-_{\boldsymbol{j}}(t_{i+1})\right). \end{aligned} \tag{3.39}$$

The *spectral Kalman filter* in its basic form is summarized in algorithm 3.1, with $\odot$ denoting the Hadamard product, $\circ^{-1}$ the element-wise inverse operator and $\boldsymbol{f}(\Delta_i), \boldsymbol{h}(\Delta_i)$ and $\boldsymbol{q}(\Delta_i)$ the stacked vectors with elements $f_{\boldsymbol{j}}(\Delta_i), h_{\boldsymbol{j}}(\Delta_i)$ and $q_{\boldsymbol{j}}(\Delta_i)$, $\boldsymbol{j} \in \mathcal{J}_N$, respectively. Note that following remark (2.3) in section 2.3.2, this formulation includes the real-valued case $\alpha_{\mathbf{0}}$ without any further complications.

The spectral formulation gets rid of any previously needed matrix calculations and a single iteration can therefore be done in linear time $\mathcal{O}(N^2)$ as a function of the number of spatial grid points N^2. As filtering of the spectral processes can be done independently, computational costs can further be controlled by omitting the spectral processes corresponding to high frequency spatial waves, which comes in especially useful in regards to likelihood computations based on the filtering estimations. In that case the complexity costs becomes $\mathcal{O}(K)$ where K is the number of considered spectral processes. In particular, since $\tilde{\boldsymbol{U}}$ is a real-valued process, even when one wishes the 'complete' spectral process, this can be done by only including $\frac{N^2}{2}+2$ Fourier modes using the real DFT in the Filtering computations.

Algorithm 3.1: Spectral Kalman Filter

Input: Discrete Fourier transformed data $(\hat{\boldsymbol{v}}(t_i)), i = 1, ..., I$, model parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \zeta, \sigma_0^2, \rho_0, \sigma_\epsilon^2$, prior state $\boldsymbol{m}_0, \boldsymbol{K}_0$.

Output: mean and variance predictions and updates $\boldsymbol{m}^-(t_i), \boldsymbol{K}^-(t_i), \boldsymbol{m}(t_i), \boldsymbol{K}(t_i)$, i =1,...,I.

Initialization: $\boldsymbol{m}(t_0^+) = \boldsymbol{m}_0, \boldsymbol{K}(t_0^+) = \boldsymbol{K}_0$.

for $i = 0, ..., I-1$ **do**

Prediction:

$\boldsymbol{K}^-(t_{i+1}) = \boldsymbol{h}(\Delta_i) \odot \boldsymbol{K}(t_i) + \boldsymbol{q}(\Delta_i)$;

$\boldsymbol{m}^-(t_{i+1}) = \boldsymbol{f}(\Delta_i) \odot \boldsymbol{m}(t_i)$;

Update:

$\boldsymbol{K}(t_{i+1}) = \left(\boldsymbol{K}^-(t_{i+1})^{-1} + \frac{N^2}{\sigma^2_\epsilon}\mathbf{1}\right)^{\circ^{-1}}$;

$\boldsymbol{m}(t_{i+1}) = \boldsymbol{m}^-(t_{i+1}) + \frac{N^2}{\sigma^2_\epsilon}\boldsymbol{K}(t_{i+1}) \odot (\hat{\boldsymbol{v}}(t_{i+1}) - \boldsymbol{m}^-(t_{i+1}))$;

end

Note that the parallel filtering of the spectal processes carries over the smoothing problem, allowing one to replace all matrix operations with element-wise ones and carry out a single iteration of the Kalman smoother in $\mathcal{O}(N^2)$ time as well. Since the computations liken the one in the original Kalman smoother we will omit further details at this point in time.

3.3.2. Likelihood Expression for Parameter Inference

We end this section by formulating the likelihood function that results from the state-space model given in (3.33) and (3.34).

Assume the model parameters $\boldsymbol{\theta} := [\boldsymbol{\mu}, \boldsymbol{\Sigma}, \zeta, \rho_0, \sigma_0, \sigma_\epsilon]$ are unknown and therefore to be estimated, based on the available observations $\boldsymbol{V}(t_i) = \boldsymbol{v}(t_i)$, $i = 1, ..., I$, of the spatially discretized spatiotemporal process $\tilde{\boldsymbol{U}}$. Denote by

$$\hat{\boldsymbol{v}}(t_{1:i}) := [\hat{\boldsymbol{v}}(t_1), ..., \hat{\boldsymbol{v}}(t_i)] \tag{3.40}$$

the vector of the discrete Fourier transformed observations $\hat{\boldsymbol{v}}(t_i) = \Phi \boldsymbol{v}(t_i)$ up to time t_i. Due to the Markov property of the spectral process $(\boldsymbol{\alpha}(t))_t$ and measurement noise being temporally white, the likelihood function $f_{\hat{\boldsymbol{V}}(t_1:I)|\boldsymbol{\theta}}(\hat{\boldsymbol{v}}(t_{1:I}) \mid \boldsymbol{\theta})$ factorizes to

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= f_{\hat{\boldsymbol{V}}(t_{1:I})|\boldsymbol{\theta}}(\hat{\boldsymbol{v}}(t_{1:I}) \mid \boldsymbol{\theta}) \\ &= f_{\hat{\boldsymbol{V}}(t_1)|\boldsymbol{\theta}}(\hat{\boldsymbol{v}}(t_1) \mid \boldsymbol{\theta}) \prod_{i=1}^{I-1} f_{\hat{\boldsymbol{V}}(t_{i+1})|\hat{\boldsymbol{V}}(t_{1:i}),\boldsymbol{\theta}}(\hat{\boldsymbol{v}}(t_{1:i+1}) \mid \hat{\boldsymbol{v}}(t_{1:i}), \boldsymbol{\theta}).\end{aligned}$$

Following equation (3.34), we have for the components of $\hat{\boldsymbol{V}}(t_{i+1}) \mid \hat{\boldsymbol{V}}(t_{1:i}), \boldsymbol{\theta}$:

$$\begin{aligned}\hat{V}_{\boldsymbol{j}}(t_{i+1}) \mid \hat{V}_{\boldsymbol{j}}(t_{1:i}), \boldsymbol{\theta} &= \alpha_{\boldsymbol{j}}(t_{i+1}) + \hat{\epsilon}_{\boldsymbol{j}}(t_{i+1}) \mid \hat{V}_{\boldsymbol{j}}(t_{1:i}), \boldsymbol{\theta} \\ &\sim \begin{cases} \mathcal{CN}(m^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1}), r_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1})), & \text{if } \boldsymbol{j} \in \mathcal{J}_N, \boldsymbol{j} \neq \boldsymbol{0}, \\ \mathcal{N}(m^-_{\boldsymbol{0},\boldsymbol{\theta}}(t_{i+1}), r_{\boldsymbol{0},\boldsymbol{\theta}}(t_{i+1})), & \text{if } \boldsymbol{j} = \boldsymbol{0}, \end{cases}\end{aligned}$$

where $r_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1}) := K^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1}) + \frac{\sigma_\epsilon{}^2}{NM}$ and $m^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1})$, $K^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1})$ denote the Kalman filtering mean and variance outputs

$$\begin{aligned}m^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1}) &= \mathrm{E}[\alpha_{\boldsymbol{j}}(t_{i+1}) \mid \hat{\boldsymbol{v}}(t_{1:i}), \boldsymbol{\theta}] \\ K^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_{i+1}) &= \mathrm{E}[\alpha_{\boldsymbol{j}}(t_{i+1})\overline{\alpha_{\boldsymbol{j}}(t_{i+1})} \mid \hat{\boldsymbol{v}}(t_{1:i}), \boldsymbol{\theta}]\end{aligned}$$

for each $\boldsymbol{j} \in \mathcal{J}_N$ under model parameters $\boldsymbol{\theta}$.

Denote by $f_{\mathcal{N}}(\cdot; m, K)$ and $f_{\mathcal{CN}}(\cdot; m, K)$ respectively the density of a real and proper complex Gaussian with mean m and variance K. The likelihood function then takes on the form (with $\times$ denoting arithmetic multiplication)

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \prod_{i=1}^{I} f_{\mathcal{N}}\left(\hat{v}_{\boldsymbol{0}}(t_i); m^-_{\boldsymbol{0},\boldsymbol{\theta}}(t_i), r_{\boldsymbol{0},\boldsymbol{\theta}}(t_i)\right) \\ &\quad \times \prod_{i=1}^{I} \prod_{\boldsymbol{j} \in \mathcal{J}_K, \boldsymbol{j} \neq \boldsymbol{0}} f_{\mathcal{CN}}\left(\hat{v}_{\boldsymbol{j}}(t_i); m^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_i), r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\right) \\ &= \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)}} \exp\left(-\frac{1}{2r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)} \|\hat{v}_{\boldsymbol{j}}(t_i) - m^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\|^2\right) \\ &\quad \times \prod_{i=1}^{I} \prod_{\boldsymbol{j} \in \mathcal{J}_K, \boldsymbol{j} \neq \boldsymbol{0}} \frac{1}{\pi r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)} \exp\left(-\frac{1}{r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)} \|\hat{v}_{\boldsymbol{j}}(t_i) - m^-_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\|^2\right).\end{aligned} \tag{3.41}$$

In practice, to prevent computational overflow, it is useful to consider the log-likelihood instead, given by

$$\begin{aligned}\ell(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{I} &- \log(2\pi) - \log(r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)) - \frac{1}{r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)}\|\hat{v}_{\boldsymbol{j}}(t_i) - m^{-}_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\|^2 \\ + \sum_{i=1}^{I}\sum_{\boldsymbol{j}\in\mathcal{J}_K,\boldsymbol{j}\neq\mathbf{0}} &- \log(\pi) - \log(r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)) - \frac{1}{r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)}\|\hat{v}_{\boldsymbol{j}}(t_i) - m^{-}_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\|^2.\end{aligned} \tag{3.42}$$

Note that after applying the discrete Fourier transform, a single evaluation of the likelihood, including the filtering, can therefore be done in $\mathcal{O}(I|\mathcal{J}_K|)$ time, where $|\mathcal{J}_K|$ is the number of Fourier coefficients taken into account and I the number of observation times. This is a remarkable improvement compared to the issue of cubic computational complexity raised in remark 3.1.

3.4. Experiments

The following section presents experimental results of the models performance in terms of both the process state estimation and model parameter inference. The experiments are conducted on simulated data of the spatiotemporal process $\tilde{\boldsymbol{U}}$ using the linear state space model as formulated in equation (3.33):

$$\begin{aligned} \textbf{Process Level: } \boldsymbol{\alpha}(t_{i+1}) &= \boldsymbol{F}(\Delta_i)\boldsymbol{\alpha}(t_i) + \boldsymbol{\nu}(\Delta_i) \\ \tilde{\boldsymbol{U}}(t_{i+1}) &= \boldsymbol{\Phi}^{-1}\boldsymbol{\alpha}(t_{i+1}) \\ \textbf{Observation Level: } \boldsymbol{V}(t_{i+1}) &= \tilde{\boldsymbol{U}}(t_{i+1}) + \boldsymbol{\epsilon}(t_{i+1}), \quad i \in \{1, ..., I\}. \end{aligned}$$

We assume a spatial domain that equals the rectangle $D = [0, 2\pi]^2$, evenly gridded by the mesh $\mathcal{D} = \{\boldsymbol{x}_{n,m} = 2\pi(\frac{n}{N}, \frac{m}{N}) : n = 0, ..., N-1, m = 0, ..., N-1\}$. In our particular

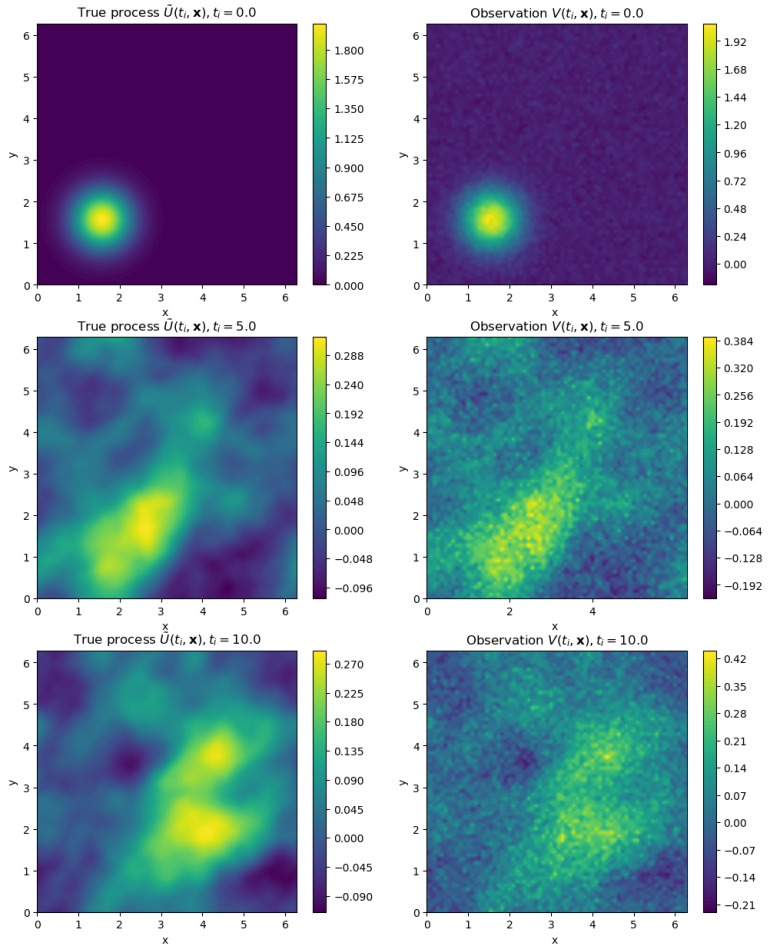

Figure 3.1.: Examples of the simulated process $\tilde{\boldsymbol{U}}$ and observations $\boldsymbol{V}$ for three different observation times. Note the differing color scales.

simulation we choose $N = 70$. For the initial condition we assume a bell curve

$$U(0, \boldsymbol{x}) = 2 \exp\big(-2\|\boldsymbol{x} - \boldsymbol{x}_0\|^2\big)$$

centered at $\boldsymbol{x}_0 = (\frac{\pi}{2}, \frac{\pi}{2})$. From the initial condition, the process is simulated forward in time in the spectral domain with time steps $\Delta t = 0.02$ over the time span $t \in [0, 10]$. The spatiotemporal, 'true' process $\tilde{\boldsymbol{U}}(t)$ is respectively computed by the inverse discrete Fourier transform at each time step. Note that the equally distanced time steps are chosen for experimental simplicity only and more general setups are possible.

The observations $\boldsymbol{V}(t_i) = (V(t_i, \boldsymbol{x}_{n,m}))_{n,m}$ are made available at every tenth time step, corrupted by spatial Gaussian white noise with standard deviation $\sigma_\epsilon = 0.05$.

The model parameters are chosen as follows: The drift parameter $\boldsymbol{\mu} = [0.12, 0.1]^T$ resembles a slightly stronger drift in the $x-$direction, whereas $\boldsymbol{\Sigma} = \begin{pmatrix} 0.12 & 0.05 \\ 0.05 & 0.12 \end{pmatrix}$ is set such that diffusion is anisotropic. We choose a weak damping parameter $\zeta = 0.005$ and the stochastic Matérn forcing term to be parametrized by $\rho_0 = 0.3$ and $\sigma_0 = 0.9$.

Examples of the sampled process and observations are given in figure 3.1. Note that color scales are differing inbetween times.

3.4.1. Filtering & Smoothing Results

We run the spectral Kalman filter specified in algorithm 3.1 and its smoothing pendant on the given discrete-time observations $\{\boldsymbol{V}(t_i) : t_i = 0, 0.2, ..., 10\}$ of the simulated process above. To analyze the models estimation perfomance inbetween measurements, we choose to run the continuous-discrete Kalman filter and -smoother by including 'measurement-free' steps as described in remark 2.2.

Figure 3.2 shows the true and estimated paths for four different spatial wave numbers $\boldsymbol{\xi_j}$, with shaded areas resembling the distance of one standard deviation. The spatial waves have been picked to represent a range of frequency magnitudes. We observe:

- As expected, the Kalman smoother performs better than the filter. Whereas the smoothing estimates recover a path with small-scale fluctuations, the Kalman filters measurement-free prediction steps are dominated by the filters linear mean propagation. Further recognizable is the reduced uncertainty of the smoothing estimates.

- Furthermore, the spectal processes' convergence behavior described in remark 3.2 is well visible. For large spatial frequencies, $\alpha_{\boldsymbol{j}}(t)$ rapidly converges to its centered stationary distribution, whereas low frequencies remain non-stationary in the given time-frame. Note also the difference in the stationary distributions variances for larger wave numbers. This behavior again suggests that the impact of high frequency components in the Fourier decomposition of the true process $U(t, \boldsymbol{x})$ vanishes rapidly.

Both observations are also recognizable in figure 3.3, which shows the mean absolute estimation error between the true simulated process $\tilde{\boldsymbol{U}}(t, \boldsymbol{x})$ and its reconstruction based on the filtering and smoothing estimates of the spectral process $\boldsymbol{\alpha}(t)$. The latter reconstruction is shown once using all Fourier coefficients $\alpha_{\boldsymbol{j}}, \boldsymbol{j} \in \mathcal{J}_N$, and additionally for two different low pass filters with cutoff frequencies $\|\boldsymbol{\xi_j}\| < 5$ and $\|\boldsymbol{\xi_j}\| < 3$. Taking into account that we can make use of the real DFT, the respective number of Fourier coefficients for these three reconstructions are $2520, 39$ and 15.

Both algorithms manage to improve the process estimation beyond observation quality, even inbetween measurement steps. Unsurprisingly, the smoothing estimate outperforms the Kalman filter in all three frequency filter configurations.

What is slightly suprising however, is that in this particular parameter setting, the low pass filtered reconstruction with cutoff frequency $\|\boldsymbol{\xi_j}\| < 5$ manages to almost match estimation accuracy of the fully reconstructed process. This again indicates that, spatially, the process is dominated by its low frequency components.

In our particular parametrization, this cutoff frequency seems to be close to the sweet spot, as already the low pass frequency cutoff $\|\boldsymbol{\xi_j}\| < 3$ starts out with a noticeable worse estimation performance. It improves as time progresses and spectral processes relating to frequencies in the intermediate range converge to their respective stationary distributions.

Figure (3.4) visualises the smoothing reconstruction using the different low pass filter settings at the two observation times $t_i = 5, 10$.

Figure 3.2.: Filtering and smoothing paths for differing spectral processes $\alpha_{\boldsymbol{j}}(t)$. Different colors represent the true and estimated process for the real and imaginary part respectively. The shaded areas represent the mean estimate $\pm$ the estimated standard deviation $m(t) \pm \sqrt{K(t)}$.

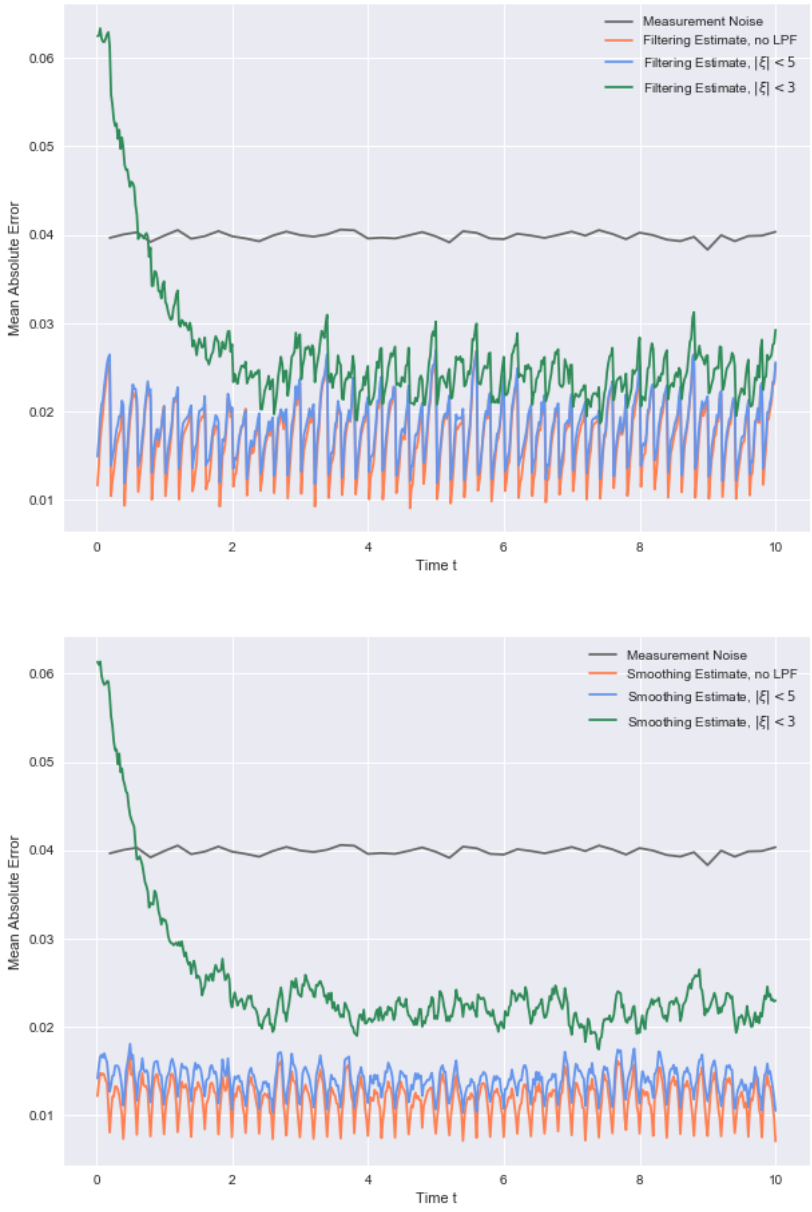

Figure 3.3.: Mean absolute error $\|\tilde{\boldsymbol{U}}(t_i) - \boldsymbol{\Phi}^{-1}\boldsymbol{\alpha}(t_i)\|$ of the spectral filtering (upper panel) and smoothing estimates (lower panel) of $\boldsymbol{\alpha}(t_i)$ over the spatial domain as a function of time. The grey line represents the averaged measurement error $\boldsymbol{\epsilon}(t_i)$. The other colors represent different numbers of Fourier modes considered in $\boldsymbol{\alpha}(t_i)$.

3.4.2. Parameter Estimation Results

We are now concerned with the setting in which the model parameters

$$\boldsymbol{\theta} := [\boldsymbol{\mu}, \boldsymbol{\Sigma}, \zeta, \rho_0, \sigma_0, \sigma_\epsilon]$$

are unknown and therefore to be estimated based on the available observations $\boldsymbol{V}(t_i) = \boldsymbol{v}(t_i) \in \mathbb{R}^{N^2}, i = 1, ..., 50$. Recall that the model log-likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{I} &- \log(2\pi) - \log(r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)) - \frac{1}{r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)} \|\hat{v}_{\boldsymbol{j}}(t_i) - m^{-}_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\|^2 \\ + \sum_{i=1}^{I} \sum_{\boldsymbol{j} \in \mathcal{J}_K, \boldsymbol{j} \neq \boldsymbol{0}} &- \log(\pi) - \log(r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)) - \frac{1}{r_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)} \|\hat{v}_{\boldsymbol{j}}(t_i) - m^{-}_{\boldsymbol{j},\boldsymbol{\theta}}(t_i)\|^2,\end{aligned}$$

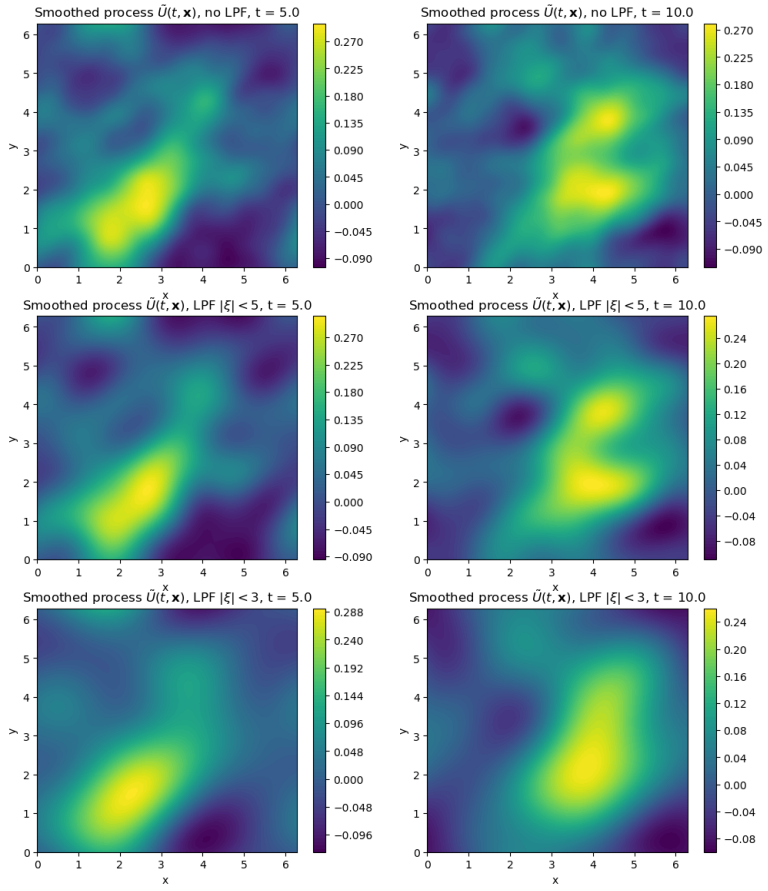

Figure 3.4.: Process reconstructions of $\tilde{U}(t, \boldsymbol{x})$ based on the Kalman smoother. The columns correspond to the second and third time step in figure 3.1. Different rows correspond to different numbers of Fourier modes used in the reconstruction: LPF $|\boldsymbol{\xi}| < K$ denotes a filter with cutoff frequency K.

where $\hat{v}_{\boldsymbol{j}}(t_i) = (\Phi \boldsymbol{v}(t_i))_{\boldsymbol{j}}$ denotes the $\boldsymbol{j}$-th discrete Fourier coefficient of $\boldsymbol{V}(t_i)$. For the parameter estimation, we set $K = 5$.

The (log-)likelihood can be passed onto any likelihood-based parameter estimation method. We choose to adapt the Bayesian setting and treat the model parameters $\boldsymbol{\Theta}$ as a random variable whose posterior $\boldsymbol{\Theta} \mid \hat{\boldsymbol{V}}(t_{1:I})$ is now to be estimated, based on the Bayes theorem

$$f_{\boldsymbol{\Theta}|\hat{\boldsymbol{V}}(t_1:I)}(\boldsymbol{\theta} \mid \hat{\boldsymbol{v}}(t_{1:I})) \propto \mathcal{L}(\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}). \tag{3.43}$$

Here, $f_{\boldsymbol{\Theta}}$ denotes the prior density function of $\boldsymbol{\Theta}$, whose exact form we shall specify shortly.

Estimation of (3.43) is done by sampling form a Markov chain generated by the *Metropolis-Hastings* (MH) algorithm, a general *Markov chain Monte Carlo* scheme that ensures convergence of the Markov chain to its stationary distribution $f_{\boldsymbol{\Theta}|\hat{\boldsymbol{V}}(t_1:I)}$ under very flexible conditions. We shall treat is as a simple 'off-the-shelf' solution for Bayesian inference. For more detail, see e.g. Brooks et al. [2011]. For completeness sake, its general formulation is summarized in algorithm 3.2.

Besides the likelihood function and prior distributions, the Metropolis-Hastings requires a proposal density $q(\cdot \mid \boldsymbol{\theta})$ which determines the rate at which the chain explores the parameter space. The diffusion matrix requires special attention due to its symmetric positive definite property. We parametrize it via a Cholesky decomposition

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T,$$

where $\boldsymbol{L} = \begin{bmatrix} L_1 & 0 \\ L_2 & L_3 \end{bmatrix}$ is some lower triangular matrix. For the choices of proposal and prior densities for the rest of the model parameters, we follow Sigrist et al. [2015].

We choose Gaussian proposals in $\boldsymbol{\mu}$ and $\boldsymbol{L}$, centered at the current chain state

$$q(\boldsymbol{\mu_p} \mid \boldsymbol{\mu_i}) \sim \mathcal{N}_2(\boldsymbol{\mu_i}, \sigma^2_{\boldsymbol{\mu}})$$
$$q(\boldsymbol{L_p} \mid \boldsymbol{L_i}) \sim \mathcal{N}_3(\boldsymbol{L_i}, \sigma^2_{\boldsymbol{L}}).$$

To keep parameters positive, all other proposals are drawn from log-posterior normal distributions, e.g.

$$q(\zeta_p \mid \zeta_i) \sim \mathcal{LN}(\ln(\zeta_i), \sigma^2_\zeta)$$

and likewise for $\rho_0, \sigma_0, \sigma_\epsilon$. The proposal densities hyperparameters $\sigma^2_\bullet, \bullet = \boldsymbol{\mu}, ..., \sigma_\epsilon$, are subject to tuning. The proposal term in the acceptance ratio takes on the form

$$\frac{q(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_p)}{q(\boldsymbol{\theta}_p \mid \boldsymbol{\theta}_i)} = \frac{\zeta_p \rho_{0,p} \sigma_{0,p} \sigma_{\epsilon,p}}{\zeta_i \rho_{0,i} \sigma_{0,i} \sigma_{\epsilon,i}}, \tag{3.44}$$

where the proposal terms in $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ vanish due to the symmetry of the normal distribution.

The priors are chosen to be flat on either the unit interval or $[-1, 1]$:

$$\begin{aligned} \zeta, \rho_0, \sigma_0, \sigma_\epsilon &\sim \text{Unif}([0,1]) \\ \mu_1, \mu_2 &\sim \text{Unif}([-1,1]) \\ L_1, L_2, L_3 &\sim \text{Unif}([-1,1]). \end{aligned}$$

Algorithm 3.2: Metropolis-Hastings Algorithm

Input: proposal density $q(\cdot \mid \cdot)$, likelihood function $\mathcal{L}(\boldsymbol{\theta})$, priors $f(\boldsymbol{\theta})$, number of iterations n.

Output: Markov chain samples $\boldsymbol{\theta}_{i=0,\ldots,n}$

Initialization: $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

for $i = 0, ..., n$ **do**

Propose:

Draw $\boldsymbol{\theta}_p \sim q(\cdot \mid \boldsymbol{\theta}_i)$;

Acceptance Ratio:

Calculate $\alpha(\boldsymbol{\theta}_p, \boldsymbol{\theta}_{i+1}) := \min\left(1, \frac{\mathcal{L}(\boldsymbol{\theta}_p) f(\boldsymbol{\theta}_p)}{\mathcal{L}(\boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i)} \frac{q(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_p)}{q(\boldsymbol{\theta}_p \mid \boldsymbol{\theta}_i)}\right)$;

Accept/Reject:

Draw $u \sim \text{Unif}([0,1])$;

if $u < \alpha(\boldsymbol{\theta}_p, \boldsymbol{\theta}_{i+1})$ **then**

$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_p$

end

else

$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$

end

end

Figure 3.5 shows the samples of the Metropolis-Hastings algorithm run with $50,000$ iterations, of which the first $20,000$ have been discarded as 'burn-in'. The proposal variance hyperparameters have been fine tuned such that the acceptance ratio of the chain over the last $30,00$ iterations lies between 30 to 50 percent. The figure shows that indeed the simple MH Markov chain is able to recover the model parameters, even given a likelihood function based on a low number of Fourier modes.

Noteworthy is the behavior of the chain in the parameters σ_0^2 and ρ_0 - the parameters that solely define the variance term in the forward propagation of the spectral processes. Both their chain samples show large deviations and are respectively over- and under-estimated. Noting that the spectral processes variance is a monotone function of both parameters, this suggests that the misguided estimation in these parameters seems to even out in the likelihood function.

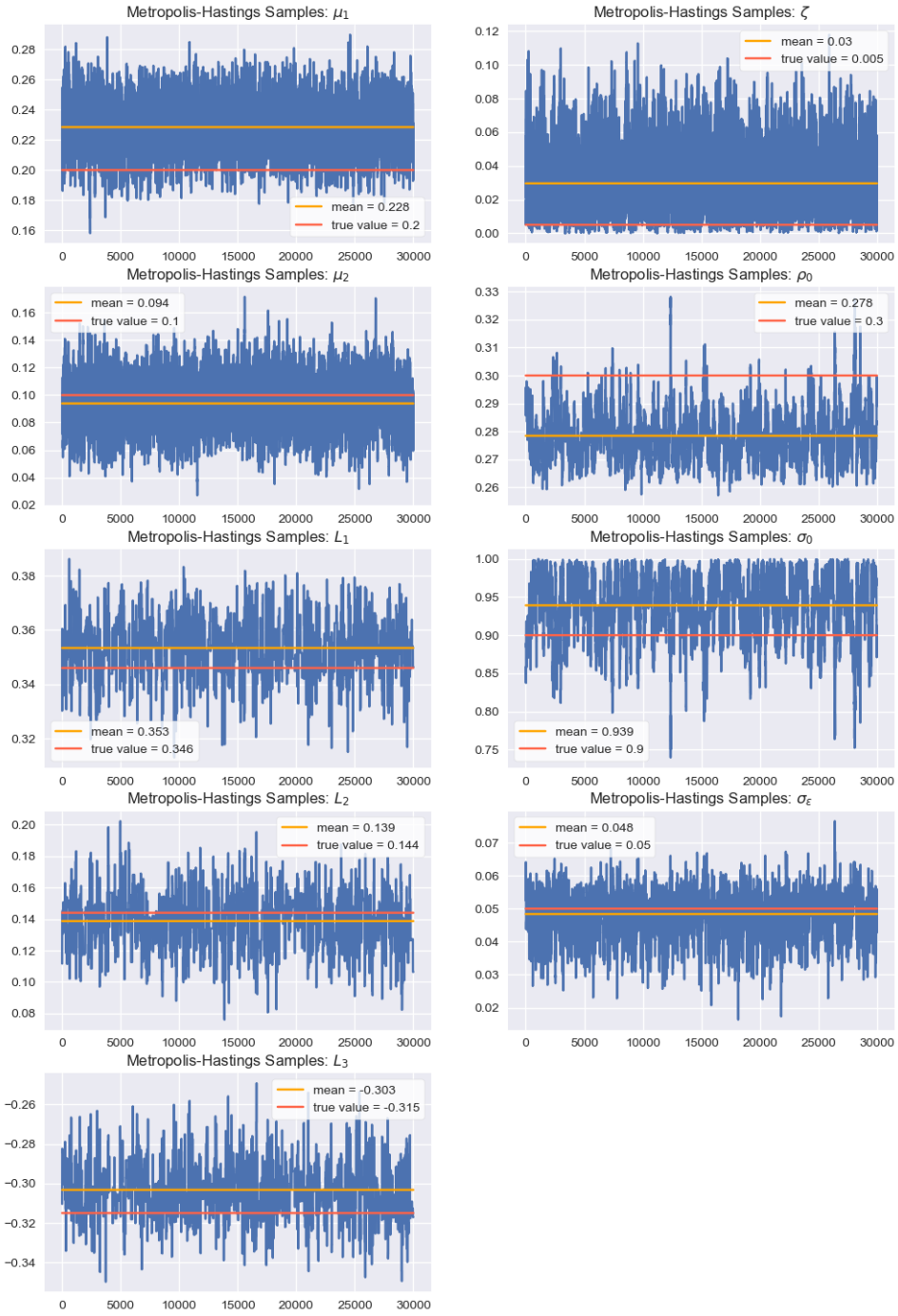

Figure 3.5.: Metropolis-Hastings samples for the last 30,000 out of 50,000 iterations. Different panels show samples for different parameters. The orange line represents the true parameter value, whereas the yellow line represents the sample mean.

4. Solving the Stochastic Convection Diffusion Equation on the Torus

The previous chapter derived a method for computationally efficient filtering, smoothing and parameter estimation of the stochastic convection-diffusion equation. For this purpose, it was assumed that the stochastic forcing term can be represented by a Fourier spectral decomposition. This led to a "formal solution" of the approximated equation that was again given by a linear combination of Fourier basis functions.

As we will show in section 4.1, the approximation of the forcing term is only correct under the assumption of spatially periodic boundary conditions, i.e. when the problem is defined on the two-dimensional torus.

However, reformulating the problem under this assumption allows us to derive a true solution to the equation. This will be done in section 4.2 within the framework of weak solutions introduced in chapter 2.4. In particular, we will show that under these assumptions, the previously obtained formal solution indeed approximates the equations weak solution, leading to stronger convergence results.

4.1. Matérn Processes on $\mathbb{T}^2$

For the rest of the chapter, denote once again by $\boldsymbol{\xi_j} = 2\pi \left(\frac{j_1}{L}, \frac{j_2}{L}\right)$, $\boldsymbol{j} \in \mathbb{Z}^2$, the $\boldsymbol{j}$-th elementary spatial wave vector on the domain $[0, L]^2$ with corresponding Fourier basis function $\phi_{\boldsymbol{j}}(\boldsymbol{x}) = \exp(i\boldsymbol{\xi_j} \cdot \boldsymbol{x})$ and by $\mathcal{J}_K := \{\boldsymbol{j} \in \mathbb{Z}^2 : -\frac{K}{2} + 1 \leq j_i \leq \frac{K}{2},\ i = 1, 2\}$ the first K^2 spatial wave vectors, centered at the origin.

Recall that in section 3.1, for any fixed time $t \in [0, T]$, we approximated the Matérn process $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{\boldsymbol{x}}$ on $[0, L]^2$ via a finite sum

$$\begin{aligned} \dot{W}^{Q_k}(t, \boldsymbol{x}) &= \int_{\mathbb{R}^2} \exp(i\boldsymbol{\xi} \cdot \boldsymbol{x}) d\eta(t, \boldsymbol{\xi}) \\ &\approx \sum_{\boldsymbol{j} \in \mathcal{J}_K} \phi_{\boldsymbol{j}}(\boldsymbol{x}) \eta_{\boldsymbol{j}}, \end{aligned} \tag{4.1}$$

where $(\eta_{\boldsymbol{j}})_{\boldsymbol{j}}$ are independent, complex Gaussian variables with variance $\mathrm{E}[\eta_{\boldsymbol{j}} \overline{\eta_{\boldsymbol{j}}}] = \left(\frac{2\pi}{L}\right)^2 f(\boldsymbol{\xi_j})$, $f(\boldsymbol{\xi})$ denoting the spectral density of the Matérn covariance function.

Note that this approximation is not "exact in $\mathcal{J}$", meaning that the right hand-side in (4.1) does not converge to the Matern field $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{\boldsymbol{x}}$ as we let $K \longrightarrow \infty$. This is due to the fact that the stochastic integral - in order to take advantage of the Fourier

basis functions on $[0, L]^2$ - is approximated via a partition with intervals

$$\Delta\boldsymbol{\xi}_{\boldsymbol{j}} := \left[2\pi\left(\tfrac{j_1}{L} - \tfrac{1}{2L}\right), 2\pi\left(\tfrac{j_1}{L} + \tfrac{1}{2L}\right)\right] \times \left[2\pi\left(\tfrac{j_2}{L} - \tfrac{1}{2L}\right), 2\pi\left(\tfrac{j_2}{L} + \tfrac{1}{2L}\right)\right], \quad \boldsymbol{j} \in \mathbb{Z}^2.$$

Therefore, the approximation converges only when one takes the limit in both L and K simultaneously, a disadvantageous fact as the spatial domain is assumed to be constant and one naturally wants the quality of an approximation to be a function of the number of basis functions only.

However, the fact that this sum resembles a Fourier series decomposition indicates its ability of approximating a spatially periodic process. Indeed, in the following section we we will show:

Theorem (Informal)**.** In the limit $K \to \infty$, the sum $\sum_{\boldsymbol{j} \in \mathcal{J}_K} \phi_{\boldsymbol{j}}(\boldsymbol{x})\eta_{\boldsymbol{j}}$ as defined in (4.1) converges in the L^2-sense to a Matérn process defined on the torus $\mathbb{T}^2$.

We begin by reviewing some basic properties of the torus which will allow us to apply the generalized Bochner's and spectral representation theorem for stationary processes on the torus. A result by Borovitskiy et al. [2020], which generalized Matérn processes on compact Riemannian manifolds, then provides us with an explicit expression for their covariance function and spectral density on the torus.

Definition 4.1 (The d-dimensional Torus)**.** Let $\mathbb{S} := \{z \in \mathbb{C} : \|z\| = 1\}$ be the unit cirlce in $\mathbb{C}$. The *$d-$dimensional torus* is defined as the product

$$\mathbb{T}^d := \underbrace{\mathbb{S} \times ... \times \mathbb{S}}_{\text{d-times}}.$$

We need the following properties of the torus:

Lemma 4.1. The $d-$dimensional Torus is a compact, abelian group in regards to component-wise multiplication in $\mathbb{C}$. Its dual group $\hat{\mathbb{T}}^d$ [1] is given by functions of the form

$$\phi(\boldsymbol{z}) = \prod_{i=1}^{d} z_i^{j_i}, \quad \boldsymbol{j} \in \mathbb{Z}^d.$$

and is therefore commonly identified with $\mathbb{Z}^d$.

Proof. These are standard results. Proofs can for example be found (or are direct consequences of the results in) in section 1.2 of Rudin [1962]. □

[1] The dual group $(\hat{G}, \cdot)$ of any locally compact, abelian group G is defined as the set of functions $\hat{G} = \{\phi : G \longrightarrow \mathbb{T} : \phi \text{ is a continuous group homomorphism}\}$ with operation $(\phi \cdot \psi)(x) := \phi(x)\psi(x)$.

Remark 4.1. Let $\boldsymbol{x} \sim \boldsymbol{y}$ be the equivalence relation on $\mathbb{R}^d$ defined by $\boldsymbol{x} \sim \boldsymbol{y} \iff \boldsymbol{x} - \boldsymbol{y} \in \mathbb{Z}^d$. Recall that there exists a canonical homeomorphism $q(\cdot)$ between the d-dimensional torus $\mathbb{T}^d$ and the quotient space $(\mathbb{R}/\mathbb{Z})^d := (\mathbb{R}^d/\sim)$ given by

$$\begin{aligned} q : (\mathbb{R}/\mathbb{Z})^d &\longrightarrow \mathbb{T}^d \\ q(\boldsymbol{x}) &= (\exp(2\pi i x_1), ..., \exp(2\pi i x_d)). \end{aligned}$$

It is therefore common to identify the torus with $(\mathbb{R}/\mathbb{Z})^d$. In that case, the corresponding dual group $\widehat{(\mathbb{R}/\mathbb{Z})^d}$ is the group of functions of the form

$$\begin{aligned} \phi : (\mathbb{R}/\mathbb{Z})^d &\longrightarrow \mathbb{T} \\ \phi(\boldsymbol{x}) &= \exp(2\pi i \boldsymbol{x} \cdot \boldsymbol{j}). \end{aligned}$$

Note that these are the Fourier basis functions $(\phi_{\boldsymbol{j}})_{\boldsymbol{j}}$ that wrap the unit interval $[0,1]^d$ around the torus $\mathbb{T}^d$.

Of course, one may likewise consider the quotient space $(\mathbb{R}/L\mathbb{Z})^d$ with the canonical homeomorphism $\tilde{q} : \boldsymbol{x} \mapsto \big(\exp\big(\frac{2\pi}{L} i x_1\big), ..., \exp\big(\frac{2\pi}{L} i x_d\big)\big) \in \mathbb{T}^d$ and dual group of Fourier basis functions $\tilde{\phi}_{\boldsymbol{j}} : \boldsymbol{x} \mapsto \exp\big(\frac{2\pi}{L} i \boldsymbol{j} \cdot \boldsymbol{x}\big)$ on $[0, L]^2$. In that case, all results derived in this chapter hold true and we can therefore assume without loss of generality that $L = 1$.

We now state the generalized versions of Bochner's and the spectral representation theorem.

Theorem 4.1 (Bochner's Theorem (Generalized))**.** Let G be a locally compact, abelian group with dual group $\hat{G}$. A continuous function $h : G \longrightarrow \mathbb{C}$ is positive-definite, in the sense that

$$\sum_{i,j=1}^{n} \lambda_i \overline{\lambda_j} h(x_i - x_j) \geq 0$$

for all $\lambda_1, ..., \lambda_n \in \mathbb{C}, x_1, ..., x_n \in G, n \in \mathbb{N}$, if and only if there exists a unique, finite, non-negative measure μ on $\hat{G}$ such that

$$h(x) = \int_{\hat{G}} \phi(x) d\mu(\phi) \tag{4.2}$$

for all $x \in G$.

Proof. See Rudin [1962], section 1.4.3. □

Theorem 4.2 (Spectral Representation Theorem (Generalized))**.** Let T be some index set and $(X(t))_{t \in T}$ be a stochastic process. Further let $(A, \mathcal{A})$ be some measurable space. There exists a spectral representation, such that for all $t \in T$

$$X(t) = \int_A \phi(t, a) dZ(a), \tag{4.3}$$

where $\phi(\cdot,\cdot)$ is some complex-valued function and $Z(\cdot)$ is an orthogonal random measure on $(A,\mathcal{A})$ if and only if there exists a measure μ on $(A,\mathcal{A})$ such that

$$\mathrm{E}[X(s)\overline{X(t)}] = \int_A \phi(s,a)\overline{\phi(t,a)}d\mu(a) \tag{4.4}$$

for all $s,t \in T$. Then, it holds

$$\mathrm{E}[Z(E_1)\overline{Z(E_2)}] = \begin{cases} 0, & \text{if } E_1 \cap E_2 = \emptyset \\ \mu(E_1), & \text{if } E_1 = E_2 \end{cases}$$

for all $E_1, E_2 \in \mathcal{A}$.

Proof. See Yaglom [1987b], chapter 2 note 17. □

As usual, we say that a stochastic process U on the torus is weakly stationary if it has constant mean and a covariance function given by

$$k(\boldsymbol{x}+\boldsymbol{r},\boldsymbol{x}) = c(\boldsymbol{r})$$

for all $\boldsymbol{x},\boldsymbol{r} \in (\mathbb{R}/\mathbb{Z})^d$ and some positive definite function $c : (\mathbb{R}/\mathbb{Z})^d \longrightarrow \mathbb{C}$. The special cases of the previous two theorems lets us now obtain a spectral representation for these processes:

Corollary 4.1 (Spectral Representation on the Torus)**.** Let $(U(\boldsymbol{x}))_{\boldsymbol{x}\in(\mathbb{R}/\mathbb{Z})^d}$ be a stationary process with covariance function $c(\boldsymbol{r})$. Then there exists a unique measure μ on $\mathbb{Z}^d$ such that for all $\boldsymbol{r} \in (\mathbb{R}/\mathbb{Z})^d$

$$c(\boldsymbol{r}) = \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \phi_{\boldsymbol{j}}(\boldsymbol{r})\mu(\boldsymbol{j}). \tag{4.5}$$

Furthermore, there exists a stochastic process $(Z_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{Z}^d}$ such that for all $\boldsymbol{x} \in (\mathbb{R}/\mathbb{Z})^d$

$$U(\boldsymbol{x}) = \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \phi_{\boldsymbol{j}}(\boldsymbol{x})Z_{\boldsymbol{j}} \tag{4.6}$$

and it holds

$$\mathrm{E}[Z_{\boldsymbol{j}}\overline{Z_{\boldsymbol{k}}}] = \begin{cases} 0, & \text{if } \boldsymbol{j} \neq \boldsymbol{k} \\ \mu(\boldsymbol{j}), & \text{if } \boldsymbol{j} = \boldsymbol{k}. \end{cases} \tag{4.7}$$

We again refer to μ and Z as the *spectral measure* and *spectral process* of U respectively.

Proof. Recall that $\psi \in \widehat{(\mathbb{R}/\mathbb{Z})^d}$ if and only if $\psi = \phi_{\boldsymbol{j}}$ for some $\boldsymbol{j} \in \mathbb{Z}^d$. The first claim then follows from Bochners theorem by identifying $\widehat{(\mathbb{R}/\mathbb{Z})^d}$ with $\mathbb{Z}^d$, equipped with the canonical σ-Algebra $\mathcal{P}(\mathbb{Z}^d)$.

Likewise, the second claim follows directly from the generalized spectral representation theorem with $(A,\mathcal{A}) = (\mathbb{Z}^d, \mathcal{P}(\mathbb{Z}^d))$ and $\phi(\boldsymbol{x},\boldsymbol{j}) = \phi_{\boldsymbol{j}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in (\mathbb{R}/\mathbb{Z})^d, \boldsymbol{j} \in \mathbb{Z}^d$. □

We now wish to find a spectral representation for the Matérn processes on the torus. Recall from section 2.2.4 that the Matérn processes can be defined as solutions to the fractional stochastic Laplace equation

$$(\kappa^2 - \Delta)^{\nu/2+d/4} U(\boldsymbol{x}) = \dot{W}(\boldsymbol{x}).$$

Recent work by Borovitskiy et al. [2020] generalized this approach by replacing the Laplace operator with the Laplace-Beltrami operator, the generalization of the Laplace operator on manifolds, thereby introducing a framework which allows one to define Matérn processes on compact Riemannian manifolds by solving the arising fractional stochastic Laplace-Beltrami equation.

We summarize the results for the case that is of our interest - the torus - in the following lemma. For the general results, we refer to Borovitskiy et al. [2020], appendix D.

Lemma 4.2 (Matérn Processes on the Torus)**.** Let U be a solution to the fractional stochastic Laplace-Beltrami equation

$$\left(\rho_0^{-2} - \Delta_g\right)^{\frac{\alpha}{2}} U(\boldsymbol{x}) = \dot{W}(\boldsymbol{x}), \quad \alpha = \nu + \tfrac{d}{2}.$$

Here, Δ_g denotes the Laplace-Beltrami operator and $\dot{W}$ is spatial white noise, both defined on the $d-$dimensional torus. Then, U is a stationary, centered Gaussian process with covariance function

$$c(\boldsymbol{r}) = \sigma_0^2 \sum_{\boldsymbol{j} \in \mathbb{Z}^d} (\rho_0^{-2} + \|2\pi \boldsymbol{j}\|^2)^{-\alpha} \phi_{\boldsymbol{j}}(\boldsymbol{r}). \tag{4.8}$$

We refer to U as a Matérn process on the torus. In particular, its spectral measure on $\mathbb{Z}^d$ is given by

$$\mu(\boldsymbol{j}) = \sigma_0^2 (\rho_0^{-2} + \|2\pi \boldsymbol{j}\|^2)^{-\alpha}. \tag{4.9}$$

Proof. This follows from theorem 4, Borovitskiy et al. [2020]. □

Let us now return to the case treated in section 3.1, where we considered the stochastic forcing term $\dot{W}^{Q_k}(t, \boldsymbol{x})$ to be temporally white and spatially coloured with Whittle covariance function; the Matérn covariance function with $\nu = 1$ and $d = 2$. Assume now that $\dot{W}^{Q_k}(t, \boldsymbol{x})$ is spatially periodic, i.e. $\boldsymbol{x} \in (\mathbb{R}/\mathbb{Z})^2$. Then, for any fixed time t, $(\dot{W}^{Q_k}(t, \boldsymbol{x}))_{\boldsymbol{x}}$ has covariance function and spectral measure given by

$$\begin{aligned} c(\boldsymbol{r}) &= \sum_{\boldsymbol{j} \in \mathbb{Z}^d} f_{\boldsymbol{j}} \phi_{\boldsymbol{j}}(\boldsymbol{r}) \quad \text{and} \\ f_{\boldsymbol{j}} &= \sigma_0^2 (\rho_0^{-2} + \|\boldsymbol{\xi}_{\boldsymbol{j}}\|^2)^{-2} \end{aligned} \tag{4.10}$$

respectively. In particular, by the generalized spectral representation theorem, there exists a proper complex Gaussian process $(\eta_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathbb{Z}^d}$ such that

$$\dot{W}^{Q_k}(t, \boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathbb{Z}^d} \phi_{\boldsymbol{j}}(\boldsymbol{x}) \eta_{\boldsymbol{j}}, \tag{4.11}$$

with covariance given by

$$\mathrm{E}[\eta_{\boldsymbol{j}}\overline{\eta_{\boldsymbol{k}}}] = \begin{cases} 0, & \text{if } \boldsymbol{j} \neq \boldsymbol{k} \\ f_{\boldsymbol{j}}, & \text{if } \boldsymbol{j} = \boldsymbol{k}. \end{cases} \tag{4.12}$$

Note that $f_{\boldsymbol{j}} = (2\pi)^2 f(\boldsymbol{\xi}_{\boldsymbol{j}})$, where f denotes the spectral density of the Matérn covariance function on $\mathbb{R}^2$. We therefore recover the approximation as introduced in 3.1 and this approximation indeed converges in the number of Fourier basis functions, i.e.

$$\lim_{K\to\infty} \sum_{\boldsymbol{j}\in\mathcal{J}_k} \phi_{\boldsymbol{j}}(\boldsymbol{x})\eta_{\boldsymbol{j}} = \dot{W}^{Q_k}(t,\boldsymbol{x})$$

in the $L^2(\mathbb{P})$-sense, for all $\boldsymbol{x} \in (\mathbb{R}/\mathbb{Z})^2, t \in [0,T]$ fixed. When letting time vary, in order to retain the temporally white nature of $\dot{W}^{Q_k}(t,\boldsymbol{x})$, we again assume that $(\eta_{\boldsymbol{j}}(t))_{\boldsymbol{j}}$ are independent, temporally white Gaussian noise processes. However, the next section will deal with this rigorously within the framework introduced in section 2.4.

4.2. The Weak Solution on $\mathbb{T}^2$

The previous section showed that we can approximate the temporally white, spatially Matérn process $\dot{W}^{Q_k}(t, \boldsymbol{x})$ defined on the torus by a linear combination of Fourier basis functions. We now derive a solution to the corresponding stochastic convection diffusion equation driven by $\dot{W}^{Q_k}(t, \boldsymbol{x})$ on $\mathbb{T}^2$. This will be done in the framework of generalized random processes and weak solutions introduced in section 2.4.

In this sense, let us define the stochastic convection diffusion equation as

$$\begin{cases} dU(t) = \mathcal{A}U(t)dt + dW^{Q_k}(t), \quad t \in (0, \infty), \\ U(0) = U_0, \end{cases} \tag{4.13}$$

where $\mathcal{A}$ is the convection-diffusion operator, U_0 is a generalized Gaussian random field and W^{Q_k} is a Q-Wiener process, all acting on a Hilbert space H.

We therefore need to identify a suitable underlying Hilbert space H, construct W^{Q_k} such that it corresponds to the Whittle covariance function in the spatial domain and show that the convection-diffusion operator $\mathcal{A}$ is the generator of a C_0-semigroup on H.

The underlying Hilbert space To take advantage of the Fourier spectral decomposition, we assumed our formal solution in chapter 3 to be spatially periodic on $[0, 1]^2$. The natural choice of the underlying Hilbert space is therefore the $\mathbb{C}$-Hilbert space $\mathcal{L}^2(\mathbb{T}^2)$ of square-integrable, complex valued functions on the torus, equipped with the inner product

$$\langle h, g \rangle_{\mathcal{L}^2} := \int_{\mathbb{T}^2} h(\boldsymbol{z})\overline{g(\boldsymbol{z})}dz,$$

where we define integration over $\mathbb{T}^2$ as

$$\int_{\mathbb{T}^2} h(\boldsymbol{z})d\boldsymbol{z} := \int_{[0,1]^2} h(\exp(2\pi i \boldsymbol{x}))d\boldsymbol{x}.$$

By definition of the integral and the canonical homeomorphism between $\mathbb{T}^2$ and $[0, 1]^2$, we have that $\mathcal{L}^2(\mathbb{T}^2)$ is isometrically isomorphic to $\mathcal{L}^2([0, 1]^2)$ and it is useful to identify these with each other and simply denote them by $\mathcal{L}^2$.

The Q-Wiener process Lemma 4.2 showed that the Whittle covariance function for a periodic process on $[0, 1]^2$ is given by

$$k(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{j} \in \mathbb{Z}^2} f_{\boldsymbol{j}} \phi_{\boldsymbol{j}}(\boldsymbol{x})\overline{\phi_{\boldsymbol{j}}(\boldsymbol{y})}, \tag{4.14}$$

where $f_{\boldsymbol{j}} = \sigma_0^2(\rho_0^{-2} + \|\boldsymbol{\xi}_{\boldsymbol{j}}\|^2)^{-2}$. Recall that the Fourier basis functions $(\phi_{\boldsymbol{j}}(\boldsymbol{x}))_{\boldsymbol{j}}$ form an orthonormal basis function in $\mathcal{L}^2$. By Mercer's theorem, the corresponding Hilbert-Schmidt integral operator $Q_k(h) := \int_{[0,1]^2} h(\boldsymbol{y})k(\boldsymbol{x}, \boldsymbol{y})d\boldsymbol{y}$ is a nuclear covariance operator on $\mathcal{L}^2$ with eigenfunctions and eigenvalues $(\phi_{\boldsymbol{j}})_{\boldsymbol{j}}, (f_{\boldsymbol{j}})_{\boldsymbol{j}}$.

To construct the corresponding regular Q-Wiener process W^{Q_k}, let $(w_{\boldsymbol{j}})_{\boldsymbol{j}}$ be a family of independent complex valued Wiener processes. Following remark 2.6 and 2.8 we can represent W^{Q_k} by

$$W^{Q_k}(t) := \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \sqrt{f_{\boldsymbol{j}}}\phi_{\boldsymbol{j}} w_{\boldsymbol{j}}(t) \tag{4.15}$$

such that $\mathrm{E}\left[\langle h, W^{Q_k}(s)\rangle_{\mathcal{L}^2}\langle h, W^{Q_k}(t)\rangle_{\mathcal{L}^2}\right] = \langle Q_k h, g\rangle_{\mathcal{L}^2} \min(s,t)$.

The convection-diffusion C_0-semigroup It remains to show that $\mathcal{A}$ generates a C_0-semigroup on $\mathcal{L}^2$. To do this, we take advantage of the fact that the Fourier basis functions are eigenfunctions of $\mathcal{A}$ by constructing a corresponding diagonal C_0-semigroup:

Lemma 4.3 (Diagonal C_0-semigroups)**.** Let H be a separable $\mathbb{C}$-Hilbert space with orthonormal basis $(h_j)_{j\geq 1}$ and let $(\lambda_j)_{j\geq 1}$ be a sequence in $\mathbb{C}$. Then, the operator $A : D(A) \subset H \longrightarrow H$ defined by

$$Ah := \sum_{j\geq 1} \lambda_j \langle h, h_j\rangle h_j$$

with domain $D(A) = \{h \in H : \sum_{j\geq 1} |\lambda_j\langle h, h_j\rangle|^2 < \infty\}$ defines a C_0-semigroup $(S(t)_t)$ on H if and only if $\sup_{j\geq 1} \mathrm{Re}(\lambda_j) < \infty$. In that case, $(S(t))_t$ is given by

$$S(t)h = \sum_{j\geq 1} \exp(\lambda_j t)\langle h, h_j\rangle h_j.$$

Any C_0-semigroup of such form is called a *diagonal C_0-semigroup.*

Proof. A proof for the real valued case can be found Curtain and Zwart [1995], 2.1.13. The complex case follows likewise. □

We can now show that, in the sense of *weak derivatives*, the convection-diffusion operator $\mathcal{A}$ defines a diagonal C_0-semigroup on $\mathcal{L}^2$. Recall that for any $g \in \mathcal{L}^2$, the weak derivative $D^\alpha g \in \mathcal{L}^2$, if existent, is uniquely defined by the property [2]

$$\langle D^\alpha g, v\rangle_{\mathcal{L}^2} = (-1)^{|\alpha|}\langle g, D^\alpha v\rangle_{\mathcal{L}^2} \tag{4.16}$$

for all smooth functions $v \in C^\infty(\mathbb{T}^2)$, where D^α denotes the differential operator $D^\alpha = \frac{\partial^{\alpha_1+\alpha_2}}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}}$ for some $\alpha \in \mathbb{N}^2$. Following this, one defines the *formal adjoint* $\mathcal{A}^*$ of $\mathcal{A}$ by

$$\mathcal{A}^* := \boldsymbol{\mu}\cdot\nabla + \nabla\cdot\boldsymbol{\Sigma}\nabla - \zeta, \tag{4.17}$$

such that $\mathcal{A}g$, if existent, is uniquely defined by

$$\langle \mathcal{A}g, v\rangle_{\mathcal{L}^2} = \langle g, \mathcal{A}^* v\rangle_{\mathcal{L}^2} \tag{4.18}$$

for all $v \in C^\infty(\mathbb{T}^2)$.

[2] The definition of a weak derivative is motivated by the fact that equation (4.16) indeed holds for any $|\alpha|-$times differentiable g. This can be shown by integration of parts plus the fact that v is periodic.

Corollary 4.2. The convection-diffusion operator generates the diagonal C_0-semigroup

$$S(t)g = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \exp(\lambda_{\boldsymbol{j}} t)\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}} \tag{4.19}$$

on $\mathcal{L}^2$, where $\lambda_{\boldsymbol{j}} = -i\boldsymbol{\mu}^T\boldsymbol{\xi}_{\boldsymbol{j}} - \boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}} - \zeta$ are the eigenvalues of $\phi_{\boldsymbol{j}}$ in regards to $\mathcal{A}$.

Proof. Define the operator $A : D(A) \subset \mathcal{L}^2 \longrightarrow \mathcal{L}^2$ by

$$Ag := \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \lambda_{\boldsymbol{j}}\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}.$$

Note that $\sup_{\boldsymbol{j}} \mathrm{Re}(\lambda_{\boldsymbol{j}}) = -\zeta$ by positive-definiteness of $\boldsymbol{\Sigma}$. Following lemma 4.3, A generates the diagonal C_0-semigroup

$$S(t)g = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \exp(\lambda_{\boldsymbol{j}} t)\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}$$

on $\mathcal{L}^2$ and its domain is given by

$$D(A) = \left\{ g \in \mathcal{L}^2 : \sum_{\boldsymbol{j}\in\mathbb{Z}^2} |\lambda_{\boldsymbol{j}}\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}|^2 < \infty \right\}.$$

By uniqueness of generators of C_0-semigroups, it suffices to show that $Ag = \mathcal{A}g$ whenever existent. Using Riesz-Fischer's theorem and the fact that $\mathcal{A}\phi_{\boldsymbol{j}} = \lambda_{\boldsymbol{j}}\phi_{\boldsymbol{j}}$, it holds for all $v \in C^\infty(\mathbb{T}^2)$:

$$\begin{aligned}
\langle Ag, v\rangle_{\mathcal{L}^2} &= \left\langle \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \lambda_{\boldsymbol{j}}\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}, v \right\rangle_{\mathcal{L}^2} \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\langle \lambda_{\boldsymbol{j}}\phi_{\boldsymbol{j}}, v\rangle_{\mathcal{L}^2} \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\langle \phi_{\boldsymbol{j}}, \mathcal{A}^* v\rangle_{\mathcal{L}^2} \\
&= \left\langle \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}, \mathcal{A}^* v \right\rangle_{\mathcal{L}^2} \\
&= \langle g, \mathcal{A}^* v\rangle_{\mathcal{L}^2}.
\end{aligned}$$

This shows that $Ag = \mathcal{A}g$ for any $g \in D(A)$. Conversely, when $g \in \mathcal{D}(\mathcal{A})$, there exists some $h \in \mathcal{L}^2$ such that $\langle g, \mathcal{A}^*\phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2} = \langle h, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}$ for all $\boldsymbol{j} \in \mathbb{Z}^2$. Therefore, it holds by Parseval's identity

$$\begin{aligned}
\|h\|_{\mathcal{L}^2}^2 &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} |\langle h, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}|^2 \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} |\langle g, \mathcal{A}^*\phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}|^2 \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} |\lambda_{\boldsymbol{j}}\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}|^2 < \infty,
\end{aligned}$$

where we used that $\mathcal{A}^*\phi_{\boldsymbol{j}} = \overline{\lambda_{\boldsymbol{j}}}\phi_{\boldsymbol{j}}$. In particular $D(\mathcal{A}) = D(A)$ and this proves the claim. □

With all the groundwork done, we can finally derive a weak solution to the problem at hand:

Theorem 4.3. There exists a unique weak solution $(U(t))_t$ to the stochastic convection-diffusion equation (4.13).

Further, let the initial condition $U_0 = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \alpha_{\boldsymbol{j},0}\phi_{\boldsymbol{j}}$ be a regular generalized Gaussian random field on $\mathcal{L}^2$, where $\alpha_{\boldsymbol{j},0} \sim \mathcal{CN}(0, \sigma^2_{\boldsymbol{j},0})$, $\boldsymbol{j} \in \mathbb{Z}^2$, are independent complex (real if $\boldsymbol{j} = \boldsymbol{0}$) Gaussian variables. Then solution is given by

$$U(t) = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}} \tag{4.20}$$

where $(\alpha_{\boldsymbol{j}})_{\boldsymbol{j}}$ are the complex Ornstein-Uhlenbeck processes given by

$$\alpha_{\boldsymbol{j}}(t) = \exp(\lambda_{\boldsymbol{j}}t)\alpha_{\boldsymbol{j},0} + \sqrt{f_{\boldsymbol{j}}}\int_0^t \exp(\lambda_{\boldsymbol{j}}(t-s))dw_{\boldsymbol{j}}(s).$$

Proof. By uniqueness, it suffices to show the existence on any domain $[0,T]$, $T > 0$. For this we need to show that the C_0-semigroup $(S(t))_{t\in[0,T]}$ generated by $\mathcal{A}$ is such that

$$\int_0^T \|S(t)\sqrt{Q_k}\|^2_{HS}dt < \infty.$$

Recall that $S(t)g = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \exp(\lambda_{\boldsymbol{j}}t)\langle g, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}$ and therefore $S(t)\phi_{\boldsymbol{j}} = \exp(\lambda_{\boldsymbol{j}}t)\phi_{\boldsymbol{j}}$ for all $\boldsymbol{j} \in \mathbb{Z}^2$. With this we get

$$\begin{aligned}
\|S(t)\sqrt{Q_k}\|^2_{HS} &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \langle S(t)\sqrt{Q_k}\phi_{\boldsymbol{j}}, S(t)\sqrt{Q_k}\phi_{\boldsymbol{j}}\rangle_{HS} \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} f_{\boldsymbol{j}}\langle S(t)\phi_{\boldsymbol{j}}, S(t)\phi_{\boldsymbol{j}}\rangle_{HS} \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} f_{\boldsymbol{j}} \exp(\lambda_{\boldsymbol{j}}t)\overline{\exp(\lambda_{\boldsymbol{j}}t)} \\
&= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} f_{\boldsymbol{j}} \exp\big(-2(\boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}\boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)t\big) \\
&\leq \sum_{\boldsymbol{j}\in\mathbb{Z}^2} f_{\boldsymbol{j}} \\
&< \infty,
\end{aligned}$$

where $\sum_{\boldsymbol{j}\in\mathbb{Z}^2} f_{\boldsymbol{j}} = c(\boldsymbol{0})$ is the marginal Whittle covariance on the torus. It thus holds:

$$\int_0^T \|S(t)\sqrt{Q_k}\|^2_{HS}dt \leq Tc(\boldsymbol{0}) < \infty.$$

By theorem 2.11, this shows the existence of a unique weak solution on any interval $[0, T]$. It is given by

$$U(t) = S(t)U_0 + \int_0^t S(t-s)dW^{Q_k}(s).$$

Now, let U_0 be a regular generalized random function on $\mathcal{L}^2$ with representation $U_0 = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \alpha_{\boldsymbol{j},0}\phi_{\boldsymbol{j}}$ for a sequence of independent complex Gaussian variables $\alpha_{\boldsymbol{j},0}, \boldsymbol{j} \in \mathbb{Z}^2$. We then have

$$\begin{aligned} S(t)U_0 &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \alpha_{\boldsymbol{j},0} S(t)\phi_{\boldsymbol{j}} \\ &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \alpha_{\boldsymbol{j},0} \exp(\lambda_{\boldsymbol{j}} t)\phi_{\boldsymbol{j}} \end{aligned}$$

for any $t \in [0, T]$. Furthermore, by definition of the Ito integral in $\mathcal{L}^2$ it holds

$$\begin{aligned} \int_0^t S(t-s)dW^{Q_k}(s) &= \sum_{\boldsymbol{i},\boldsymbol{j}\in\mathbb{Z}^2} \int_0^t \sqrt{f_{\boldsymbol{j}}}\langle S(t-s)\phi_{\boldsymbol{j}}, \phi_{\boldsymbol{i}}\rangle_{\mathcal{L}^2} dw_{\boldsymbol{j}}(s)\phi_{\boldsymbol{i}} \\ &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \sqrt{f_{\boldsymbol{j}}} \int_0^t \exp(\lambda_{\boldsymbol{j}}(t-s))dw_{\boldsymbol{j}}(s)\phi_{\boldsymbol{j}}. \end{aligned}$$

Setting $\alpha_{\boldsymbol{j}}(t) = \exp(\lambda_{\boldsymbol{j}} t)\alpha_{\boldsymbol{j},0} + \sqrt{f_{\boldsymbol{j}}} \int_0^t \exp(\lambda_{\boldsymbol{j}}(t-s))dw_{\boldsymbol{j}}(s)$ gives the second claim. □

Corollary 4.3. In the sense of a $\mathcal{L}^2$-valued random process, for any fixed $t \in [0, T]$, the spectral approximation

$$\tilde{U}_K(t, \boldsymbol{x}) := \sum_{\boldsymbol{j}\in\mathcal{J}_K} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x})$$

as derived in proposition 3.2 converges in $L^2(\mathbb{P})$ to the weak solution $(U(t))_t$ of the stochastic convection diffusion equation:

$$\lim_{K\to\infty} \mathrm{E}[\|\tilde{U}_K(t) - U(t)\|^2_{\mathcal{L}^2}] = 0.$$

Proof. This follows immediately from the previous theorem and the definition of the Itô integral in $\mathcal{L}^2$. □

This can be seen as an improvement of the result in lemma 3.1, where we proved the approximations *pointwise weak convergence* to the *stationary solution*, under the assumption that the initial condition is given by the processes invariant measure.

We now turn our attention to proving the processes convergence to its invariant measure in the infinite dimensional case. We start by showing the weak solution indeed defines a generalized random field on $\mathcal{L}^2$ for any fixed time t. For this, recall that for

any $\boldsymbol{j} \in \mathbb{Z}^2$, $(\alpha_{\boldsymbol{j}}(t))_t$ is a proper complex (or real for $\boldsymbol{j} = \mathbf{0}$) Gaussian process with mean zero and covariance function

$$\kappa_{\boldsymbol{j}}(s,t) := \exp(\lambda_{\boldsymbol{j}} s)\overline{\exp(\lambda_{\boldsymbol{j}} t)} \left[\sigma^2_{\boldsymbol{j},0} + \tfrac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \left(\exp\big(2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)s\big) - 1\right)\right]$$

for $s \leq t$. Denote by $\kappa_{\boldsymbol{j}}(t)$ the variance of $\alpha_{\boldsymbol{j}}(t)$; $\kappa_{\boldsymbol{j}}(t) = \kappa_{\boldsymbol{j}}(t,t)$.

Proposition 4.1. In the situation of theorem 4.3, the process $(U(t))_t$ is a $\mathcal{L}^2$-valued Gaussian process with covariance operator function

$$K_{s,t} h := \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \kappa_{\boldsymbol{j}}(s,t) \langle h, \phi_{\boldsymbol{j}} \rangle_{\mathcal{L}^2} \phi_{\boldsymbol{j}}. \tag{4.21}$$

In particular, any $U(t)$ is a generalized Gaussian random field on $\mathcal{L}^2$ with covariance operator

$$Q_t h := \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \kappa_{\boldsymbol{j}}(t) \langle h, \phi_{\boldsymbol{j}} \rangle_{\mathcal{L}^2} \phi_{\boldsymbol{j}}. \tag{4.22}$$

Proof. The fact that $U = (U(t))_t$ is indeed a $\mathcal{L}^2$-valued Gaussian process follows from its integral representation and lemma 2.7. To derive $K_{s,t}$, define for any $s \leq t$ fixed

$$k_{s,t}(\boldsymbol{x}, \boldsymbol{y}) := \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \kappa_{\boldsymbol{j}}(s,t) \phi_{\boldsymbol{j}}(\boldsymbol{x}) \overline{\phi_{\boldsymbol{j}}(\boldsymbol{y})}. \tag{4.23}$$

To show that $k_{s,t}$ is well-defined, note that by the symmetric positive-definiteness of $\boldsymbol{\Sigma}$ we have

$$\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} \geq \boldsymbol{\Sigma}_{\min} \|\boldsymbol{\xi}_{\boldsymbol{j}}\|_2^2$$

where $\boldsymbol{\Sigma}_{\min} > 0$ is the smallest eigenvalue of $\boldsymbol{\Sigma}$. Therefore, we can choose some $M > 0$ such that for all $\boldsymbol{\xi}_{\boldsymbol{j}}$ with $\|\boldsymbol{\xi}_{\boldsymbol{j}}\|_2^2 > M$:

$$\begin{aligned}
&\exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t+s)\big) \leq 1, \\
&0 \leq \big[\exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t-s)\big) - \exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t+s)\big)\big] \leq 1, \\
&2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta) \geq 1.
\end{aligned}$$

For all such $\boldsymbol{\xi}_{\boldsymbol{j}}$ it then holds:

$$\begin{aligned}
|\kappa_{\boldsymbol{j}}(s,t)| &= |\exp(\lambda_{\boldsymbol{j}} s)\overline{\exp(\lambda_{\boldsymbol{j}} t)}| \left|\sigma^2_{\boldsymbol{j},0} + \tfrac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \left(\exp\big(2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)s\big) - 1\right)\right| \\
&\leq \exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t+s)\big)\sigma^2_{\boldsymbol{j},0} \\
&\quad + \exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t+s)\big) \left|\tfrac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \left(\exp\big(2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)s\big) - 1\right)\right| \\
&= \exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t+s)\big)\sigma^2_{\boldsymbol{j},0} \\
&\quad + \tfrac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \left[\exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t-s)\big) - \exp\big(-(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)(t+s)\big)\right] \\
&\leq \sigma^2_{\boldsymbol{j},0} + f_{\boldsymbol{j}}.
\end{aligned}$$

With this, it follows that

$$\begin{aligned}\sum_{\boldsymbol{j}\in\mathbb{Z}^2} |\kappa_{\boldsymbol{j}}(s,t)| &= \sum_{\substack{\boldsymbol{j}\in\mathbb{Z}^2,\\ \|\boldsymbol{\xi}_{\boldsymbol{j}}\|_2^2\leq M}} |\kappa_{\boldsymbol{j}}(s,t)| + \sum_{\substack{\boldsymbol{j}\in\mathbb{Z}^2,\\ \|\boldsymbol{\xi}_{\boldsymbol{j}}\|_2^2> M}} |\kappa_j(t)| \\ &\leq \sum_{\substack{\boldsymbol{j}\in\mathbb{Z}^2,\\ \|\boldsymbol{\xi}_{\boldsymbol{j}}\|_2^2\leq M}} |\kappa_{\boldsymbol{j}}(s,t)| + \sum_{\substack{\boldsymbol{j}\in\mathbb{Z}^2,\\ \|\boldsymbol{\xi}_{\boldsymbol{j}}\|_2^2> M}} (\sigma_{\boldsymbol{j},0}^2 + f_{\boldsymbol{j}}) \\ &< \infty,\end{aligned}$$

where we have used the fact that $\sum_{\boldsymbol{j}} \sigma_{\boldsymbol{j},0}^2 < \infty$ by regularity of the initial condition U_0 and that $\sum_{\boldsymbol{j}} f_{\boldsymbol{j}} < \infty$ is the marginal variance of the Whittle covariance function.

This assures the convergence of $k_{s,t}(\boldsymbol{x},\boldsymbol{y})$. The fact that it is a continuous, conjugate symmetric and positive definite function follows from Bochner's theorem. Note that

$$K_{s,t}h = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \kappa_{\boldsymbol{j}}(s,t)\langle h, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}.$$

is the integral operator corresponding to $k_t(\boldsymbol{x},\boldsymbol{y})$ on $\mathcal{L}^2$. By Mercer's theorem, it is therefore a well defined, nuclear covariance operator.

Lastly, $K_{s,t}$ is indeed the covariance operator function of U. For all h, g we have:

$$\begin{aligned}\mathrm{E}\left[\langle h, U(s)\rangle_{\mathcal{L}^2}\overline{\langle g, U(t)\rangle_{\mathcal{L}^2}}\right] &= \mathrm{E}\left[\sum_{\boldsymbol{j}\in\mathbb{Z}^2}\langle h,\phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\alpha_{\boldsymbol{j}}(s)\sum_{\boldsymbol{j}\in\mathbb{Z}^2}\overline{\langle g,\phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\alpha_{\boldsymbol{j}}(t)}\right] \\ &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2}\mathrm{E}[\alpha_{\boldsymbol{j}}(s)\overline{\alpha_{\boldsymbol{j}}(t)}]\langle h,\phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\langle\phi_{\boldsymbol{j}},g\rangle_{\mathcal{L}^2} \\ &= \left\langle\sum_{\boldsymbol{j}\in\mathbb{Z}^2}\kappa_{\boldsymbol{j}}(s,t)\langle h,\phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2}\phi_{\boldsymbol{j}}, g\right\rangle_{\mathcal{L}^2} \\ &= \langle K_{s,t}h, g\rangle_{\mathcal{L}^2}.\end{aligned}$$

□

In the finite dimensional case, we have seen that the solutions to linear SDEs converge weakly against their stationary measure. It is therefore only natural to consider the convergence behavior in the Hilbert space valued case as well. For this we define:

Definition 4.2. Let $X(t), t\in[0,\infty)$ be a process of regular generalized Gaussian random fields with covariance operators Q_t. We say that $(X(t))_t$ *converges weakly* to a generalized Gaussian random field X_∞ with covariance operator Q_∞ if

$$\lim_{t\to\infty}\mathrm{tr}(Q_t - Q_\infty) = \lim_{t\to\infty}\sum_{\boldsymbol{j}\in\mathbb{Z}^2}\langle(Q_t - Q)h_{\boldsymbol{j}}, h_{\boldsymbol{j}}\rangle_H = 0. \tag{4.24}$$

Then, in particular

$$\langle h, X(t)\rangle_H \xrightarrow[t\to\infty]{} \langle h, X_\infty\rangle_H \tag{4.25}$$

weakly for all $h \in H$.

Proposition 4.2. In the situation of theorem 4.3, the process $U = (U(t))_t$ converges weakly to the regular generalized Gaussian random field U_∞ with covariance operator

$$Q_\infty h = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \langle h, \phi_{\boldsymbol{j}}\rangle_{\mathcal{L}^2} \phi_{\boldsymbol{j}}. \tag{4.26}$$

Furthermore, U_∞ is invariant for U in the sense that, if $U_0 \overset{\mathrm{d}}{=} U_\infty$ then $(U(t))_t$ is a stationary $\mathcal{L}^2$ valued Gaussian process.

Proof. The fact that Q_∞ is a well defined, nuclear covariance operator follows by the same arguments as in the proof of proposition 4.1 .

Recall that $\mathrm{tr}(Q_t - Q_\infty)$ is independent of the choice of orthonormal basis. Therefore, by the fact that $(\phi_{\boldsymbol{j}})_{\boldsymbol{j}}$ are eigenfunctions of Q_∞ and Q_t for any t, we have by absolute convergence of the series

$$\begin{aligned}\mathrm{tr}(Q_t - Q_\infty) &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \langle (Q_t - Q_\infty)\phi_{\boldsymbol{j}}, \phi_{\boldsymbol{j}}\rangle_H \\ &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \left(\kappa_{\boldsymbol{j}}(t) - \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \right) \xrightarrow[t\to\infty]{} 0.\end{aligned}$$

For the second claim, let $U_0 \overset{\mathrm{d}}{=} U_\infty$. Then there exists a representation

$$U_0 = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \alpha_{\boldsymbol{j},0} \phi_{\boldsymbol{j}}$$

with independent Gaussian variables $\alpha_{\boldsymbol{j},0} \sim \mathcal{CN}(0, \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)})$.

Recall that $\alpha_{\boldsymbol{j},0}$ is invariant for the Gaussian process $\alpha_{\boldsymbol{j}}$. Therefore, under $U_0 \overset{\mathrm{d}}{=} U_\infty$, we have that $\alpha_{\boldsymbol{j}}$ is stationary for all $\boldsymbol{j} \in \mathbb{Z}^2$ with covariance function

$$\kappa_{\boldsymbol{j}}(s,t) = \overline{\exp(\lambda_{\boldsymbol{j}}(t-s))} \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)}$$

for all $s \le t$. By proposition 4.1 it follows that $K_{s,t} = K_{s+r,t+r}$ for all r. □

We end this chapter with the following corollary that recovers the spectral density of the "ordinary"" Gaussian process corresponding to the weak solution.

Corollary 4.4. The stationary solution of the stochastic convection-diffusion equation on $\mathbb{T}^2$ has spectral spectral given by

$$g(\omega, \boldsymbol{j}) = f_{\boldsymbol{j}} \frac{1}{2\pi} [(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)^2 + (\omega + \boldsymbol{\mu}^T \boldsymbol{\xi}_{\boldsymbol{j}})^2]^{-1}, \quad \omega \in \mathbb{R}, \boldsymbol{j} \in \mathbb{Z}^2. \tag{4.27}$$

Proof. Note that

$$K_{s,t} h = \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \overline{\exp(\lambda_{\boldsymbol{j}}(t-s))} \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \langle h, \phi_{\boldsymbol{j}} \rangle_{\mathcal{L}^2} \phi_{\boldsymbol{j}}.$$

is the integral operator on $\mathcal{L}^2$ corresponding to a covariance function on the torus defined by

$$k_{s,t}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \overline{\exp(\lambda_{\boldsymbol{j}}(t-s))} \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \phi_{\boldsymbol{j}}(\boldsymbol{x}) \overline{\phi_{\boldsymbol{j}}(\boldsymbol{y})}, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{T}^2.$$

To recover previous notation of corollary 4.1, write $\tau = t - s, \boldsymbol{r} = \boldsymbol{y} - \boldsymbol{x}$. Then

$$\begin{aligned}
k_{s,t}(\boldsymbol{y}, \boldsymbol{x}) &= \overline{k_{s,t}(\boldsymbol{x}, \boldsymbol{y})} \\
&= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \exp(\lambda_{\boldsymbol{j}}(t-s)) \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \overline{\phi_{\boldsymbol{j}}(\boldsymbol{x})} \phi_{\boldsymbol{j}}(\boldsymbol{y}) \\
&= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \exp(\lambda_{\boldsymbol{j}}(t-s)) \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \phi_{\boldsymbol{j}}(\boldsymbol{y} - \boldsymbol{x}) \\
&= k_{s,s+\tau}(\boldsymbol{x} + \boldsymbol{r}, \boldsymbol{x}) \\
&=: c(\tau, \boldsymbol{r}).
\end{aligned}$$

Clearly, $c(\tau, \boldsymbol{r})$ is the covariance function of a stationary Gaussian process defined on $[0, T] \times \mathbb{T}^2$. By the generalized spectral representation theorem, there exists a unique spectral density $g(\omega, \boldsymbol{j})$ on $\mathbb{R} \times \mathbb{Z}^2$ such that

$$c(\tau, \boldsymbol{r}) = \int_{\mathbb{R}} \sum_{\boldsymbol{j} \in \mathbb{Z}^2} g(\omega, \boldsymbol{j}) \exp(i\omega\tau) \phi_{\boldsymbol{j}}(\boldsymbol{r}).$$

Now, the same computation of the frequency domain integral as in remark 3.1 gives:

$$\begin{aligned}
&\int_{\mathbb{R}} \sum_{\boldsymbol{j} \in \mathbb{Z}^2} f_{\boldsymbol{j}} \frac{1}{2\pi} [(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)^2 + (\omega + \boldsymbol{\mu}^T \boldsymbol{\xi}_{\boldsymbol{j}})^2]^{-1} \exp(i\omega\tau) \phi_{\boldsymbol{j}}(\boldsymbol{r}) d\omega \\
&= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \exp\big((-i\boldsymbol{\mu}^T \boldsymbol{\xi}_{\boldsymbol{j}} - \boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} - \zeta)\tau\big) \phi_{\boldsymbol{j}}(\boldsymbol{r}) \\
&= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \frac{f_{\boldsymbol{j}}}{2(\boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma} \boldsymbol{\xi}_{\boldsymbol{j}} + \zeta)} \exp(\lambda_{\boldsymbol{j}} \tau) \phi_{\boldsymbol{j}}(\boldsymbol{r}) \\
&= c(\tau, \boldsymbol{r})
\end{aligned}$$

and this proves the claim. □

5. Future Work

In chapter 3, we have introduced a spectral method for the linear stochastic convection-diffusion equation, motivated by the need of computationally efficient statistical inference on its approximated solutions and model parameters. Chapter 4 verified the spectral method in its approximation properties to the weak solution of the SCDE. The results in these chapters open up a number of further research questions:

- How flexible is the spectral filtering method in regards to more complex statistical applications, e.g. when one cannot assume constant model parameters?
- The derivation of the weak solution to the SCDE showed that the spectral approximation becomes exact when the solution is defined the torus. Does this open up opportunities to consider processes defined on more complicated manifolds?
- The usage of Fourier basis functions is particularly efficient due to their property of being eigenfunctions for any linear differential operator. How can one make use of this when confronted with non-linear differential equations?

We give first considerations to each of these questions in the following chapter.

5.1. Some Immediate Generalizations

This section covers some immediate generalizations of the spectral filtering method for the SCDE. Each are formalized based on the conditions introduced in chapter 3. Details that are left out in the following are therefore assumed to be as specified previously. At large these extensions don't interfere with each other and can therefore be employed 'parallelly'.

5.1.1. Temporally Varying Convection-Diffusion Operator

Filtering, Smoothing & Parameter Inference Assume that the PDE model parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \zeta$ of the stochastic convection-diffusion equation are not constant but rather functions of time, i.e. we are concerned with the equation

$$\begin{cases} \frac{\partial}{\partial t} U(t, \boldsymbol{x}) & = \mathcal{A}(t) U(t, \boldsymbol{x}) + \dot{W}^{Q_k}(t, \boldsymbol{x}), \\ U(0) & = U_0, \end{cases} \tag{5.1}$$

where the differential operator $\mathcal{A}(t)$ is defined as

$$\mathcal{A}(t) := -\boldsymbol{\mu}(t) \cdot \nabla + \nabla \cdot \boldsymbol{\Sigma}(t) \nabla - \zeta(t). \tag{5.2}$$

Assume for now that $\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t), \zeta(t)$ are known. The same considerations as in theorem 3.2 show that we can construct an approximated solution $\tilde{U}(t, \boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathcal{J}} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}}(\boldsymbol{x})$ by deriving its spectral processes as strong solutions to the Langevin equation

$$\begin{cases} d\alpha_{\boldsymbol{j}}(t) &= \lambda_{\boldsymbol{j}}(t)\alpha_{\boldsymbol{j}}(t)dt + \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})}dw_{\boldsymbol{j}}(t), \\ \alpha_{\boldsymbol{j}}(0) &= \alpha_{\boldsymbol{j},0}, \end{cases} \tag{5.3}$$

with $\lambda_{\boldsymbol{j}}(t) := -i\boldsymbol{\mu}(t)^T\boldsymbol{\xi}_{\boldsymbol{j}} - \boldsymbol{\xi}_{\boldsymbol{j}}^T\boldsymbol{\Sigma}(t)\boldsymbol{\xi}_{\boldsymbol{j}} - \zeta(t)$.

Let $p_{\boldsymbol{j}}(t, s)$ be the corresponding state transition function defined via its properties in equation (2.79), i.e. it is the solution to the ODE

$$\begin{cases} \partial_t p_{\boldsymbol{j}}(t, s) &= \lambda_{\boldsymbol{j}}(t)p_{\boldsymbol{j}}(t, s) \\ p_{\boldsymbol{j}}(s, s) &= 1. \end{cases} \tag{5.4}$$

Assume that $\lambda_{\boldsymbol{j}}$ is such that the corresponding state transition function exists and is unique (this is for example guaranteed by piecewise continuity of $\lambda_{\boldsymbol{j}}$). The strong solution to equation (5.3) is then given by

$$\alpha_{\boldsymbol{j}}(t) = p_{\boldsymbol{j}}(t, 0)\alpha_{\boldsymbol{j}}(0) + \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})} \int_0^t p_{\boldsymbol{j}}(t, s)dw_{\boldsymbol{j}}(s). \tag{5.5}$$

Filtering and smoothing for the spectral processes $\alpha_{\boldsymbol{j}}$ can therefore be done via the Kalman filter and -smoother as introduced for time-varying linear systems in section 2.3.2.

In cases where the functions $\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t), \zeta(t)$ are unknown but dependent on some parameter $\boldsymbol{\theta} \in \mathbb{R}^k$, it is evident that parameter estimation can be done based on the filtering likelihood function derived in (2.92). Note that, depending on the parametrization of $\boldsymbol{\mu}_{\boldsymbol{\theta}}(t), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(t), \zeta_{\boldsymbol{\theta}}(t)$, sampling from the posterior $\boldsymbol{\theta} \mid \boldsymbol{V}$ might become more complex than previously seen in the experiments in section 3.4.2 and more sophisticated sampling techniques will become necessary.

Existence of a Weak Solution A natural question that arises is whether a Fourier spectral decomposition based on the processes $(\alpha_{\boldsymbol{j}})_{\boldsymbol{j}}$ does indeed approximate a weak solution to the SPDE

$$\begin{cases} dU(t) = \mathcal{A}(t)U(t)dt + dW^{Q_k}(t), \\ U(0) = U_0. \end{cases} \tag{5.6}$$

Equations of this kind are known in the literature as *non-autonomous Ornstein-Uhlenbeck equations* (see e.g. Seidler [1993], Veraar [2010]). Under rightful assumptions, one can show the existence and uniqueness of a weak solution to (5.6). Without going into too much detail here, the objective is to show that the family of operators $(\mathcal{A}(t))_t$ is such that there exists a solution $P = (P(t, s)_{s \leq t}$ to the operator valued ODE

$$\begin{cases} \partial_t P(t, s) &= \mathcal{A}(t)P(t, s) \\ P(s, s) &= \mathrm{Id} \end{cases} \tag{5.7}$$

and that the solution P is an *exponentially bounded evolution family* on $\mathcal{L}^2$, i.e. fulfills the requirements that

1. for all $s \leq r \leq t \in [0,T]$: $P(t,s)P(s,r) = P(t,r)$,
2. for all $h \in \mathcal{L}^2$ fixed: $(t,s) \mapsto P(t,s)h$ is continuous,
3. there exist constants $M > 0, c > 0$: $\|P(t,s)\|_{op} \leq M \exp(-c(t-s))$.

The weak solution to (5.6) is then again given by a stochastic convolution

$$U(t) = P(t,0)U_0 + \int_0^t P(t,s)dW^{Q_k}(s). \tag{5.8}$$

Our previous considerations for the case of constant PDE parameters suggest that the right operator family $(\mathcal{A}(t))_t$ for our application is defined on $\mathcal{L}^2(\mathbb{T}^2)$ by

$$\mathcal{A}(t)h := \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \lambda_{\boldsymbol{j}}(t) \langle h, \phi_{\boldsymbol{j}} \rangle_{\mathcal{L}^2} \phi_{\boldsymbol{j}}. \tag{5.9}$$

Based on this, we propose: [1]

Proposition 5.1. The operators $P(t,s) : [0,T] \times [0,t] \longrightarrow \mathcal{L}^2$ defined by

$$P(t,s)h := \sum_{\boldsymbol{j} \in \mathbb{Z}^2} p_{\boldsymbol{j}}(t,s) \langle h, \phi_{\boldsymbol{j}} \rangle_{\mathcal{L}^2} \phi_{\boldsymbol{j}} \tag{5.10}$$

form an exponentially bounded evolution family such that P solves equation (5.7).

Under the assumption that proposition 5.1 holds, we get a weak solution U given by (5.8) with

$$\begin{aligned} \int_0^t P(t,s)dW^{Q_k}(s) &= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \sqrt{f_j} \int_0^t \langle P(t,s)\phi_{\boldsymbol{j}}, \phi_{\boldsymbol{j}} \rangle_{\mathcal{L}^2} dw_{\boldsymbol{j}}(s) \phi_{\boldsymbol{j}} \\ &= \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \sqrt{f_j} \int_0^t p_{\boldsymbol{j}}(t,s) dw_{\boldsymbol{j}}(s) \phi_{\boldsymbol{j}}. \end{aligned}$$

Together with a similar spectral decomposition for $P(t,s)U_0$ one gets the generalization of corollary 4.3:

$$\lim_{K \to \infty} \mathrm{E}[\| \sum_{\boldsymbol{j} \in \mathcal{J}_K} \alpha_{\boldsymbol{j}}(t)\phi_{\boldsymbol{j}} - U(t)\|_{\mathcal{L}^2}^2] = 0,$$

where $\alpha_{\boldsymbol{j}}$ denote the spectral processes as defined in (5.5).

[1] Proving this is not as trivial as one would hope. In fact, it is not clear if the exponential boundedness of $P(t,s)$ is the right assumption to base the existence of a weak solution on. Thankfully this is a future work chapter.

5.1.2. Spatially Varying Convection-Diffusion Operator

Filtering, Smoothing & Parameter Inference Newly published work by Liu et al. [2021] extends the spectral filtering method as introduced in section 3.2 to the case of spatially varying PDE parameters, i.e. when the convection-diffusion operator is given by

$$\mathcal{A}(\boldsymbol{x}) := -\boldsymbol{\mu}(\boldsymbol{x}) \cdot \nabla + \nabla \cdot \boldsymbol{\Sigma}(\boldsymbol{x}) \nabla - \zeta(\boldsymbol{x}).$$

Again, one seeks to find a finite dimensional solution $\tilde{U}(t, \boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathcal{J}} \alpha_{\boldsymbol{j}}(t) \phi_{\boldsymbol{j}}(\boldsymbol{x})$ to the approximated SCDE equation

$$\begin{cases} \frac{\partial}{\partial t} \tilde{U}(t, \boldsymbol{x}) & = \mathcal{A}(\boldsymbol{x}) \tilde{U}(t, \boldsymbol{x}) + \dot{W}^{\tilde{Q}}(t, \boldsymbol{x}), \\ \tilde{U}(0, \boldsymbol{x}) & = \sum_{\boldsymbol{j} \in \mathcal{J}} \phi_{\boldsymbol{j}}(\boldsymbol{x}) \alpha_{\boldsymbol{j}}(0), \end{cases} \tag{5.11}$$

with spatially periodic boundary conditions and approximated stochastic forcing term

$$\dot{W}^{\tilde{Q}}(t, \boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathcal{J}} \phi_{\boldsymbol{j}}(\boldsymbol{x}) \eta_{\boldsymbol{j}}(t).$$

The key idea applied by Liu et al. [2021] is to construct a solution in the *Fourier-Galerkin* sense, i.e. a solution $\tilde{U}(t, \boldsymbol{x})$ that satisfies

$$\left\langle \frac{\partial}{\partial t} \tilde{U}(t, \boldsymbol{x}), \phi_{\boldsymbol{k}} \right\rangle_{L^2(\mathbb{T}^2)} = \left\langle \mathcal{A}(\boldsymbol{x}) \tilde{U}(t, \boldsymbol{x}) + \dot{W}^{\tilde{Q}}(t, \boldsymbol{x}), \phi_{\boldsymbol{k}} \right\rangle_{L^2(\mathbb{T}^2)}$$

for any $\boldsymbol{k} \in \mathbb{Z}^2$. In particular, $\tilde{U}$ then solves the approximated SCDE in the weak sense on any finite dimensional subspace $V = \text{span}(\{\phi_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{J}\}) \subset \mathcal{L}^2(\mathbb{T}^2)$. With this approach, one obtains the following result:

Theorem 5.1. Let $\mathcal{J} \subset \mathbb{Z}^2$ be finite with $|\mathcal{J}| = K$. Let $\boldsymbol{\alpha}(t)$ be the K-dimensional stochastic differential equation satisfying

$$d\boldsymbol{\alpha}(t) = \boldsymbol{G}\boldsymbol{\alpha}(t) + \boldsymbol{Q} d\boldsymbol{w}(t),$$

where $\boldsymbol{Q} = \mathbf{diag}[\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})]$, $(\boldsymbol{w}(t))_t$ is a K-dimensional complex Wiener process and $\boldsymbol{G}$ is the matrix given by

$$\boldsymbol{G}_{k,j} = \int_{\mathbb{T}^2} \left[i \left(-\boldsymbol{\mu}^T(\boldsymbol{x}) \boldsymbol{\xi}_{\boldsymbol{j}} + (\nabla \cdot \boldsymbol{\Sigma}(\boldsymbol{x}))^T \boldsymbol{\xi}_{\boldsymbol{j}} \right) - \boldsymbol{\xi}_{\boldsymbol{j}}^T \boldsymbol{\Sigma}(\boldsymbol{x}) \boldsymbol{\xi}_{\boldsymbol{j}} - \zeta(\boldsymbol{x}) \right] \phi_{\boldsymbol{j}}(\boldsymbol{x}) \overline{\phi_{\boldsymbol{k}}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{5.12}$$

Then $\tilde{U}(t, \boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathcal{J}} \alpha_{\boldsymbol{j}}(t) \phi_{\boldsymbol{j}}(\boldsymbol{x})$ solves the approximated SCDE (5.11) in the Fourier-Galerkin sense.

Proof. See proposition 1, Liu et al. [2021]. □

This result derives a system of coupled spectral processes, given by a multivariate complex Ornstein-Uhlenbeck process.

Assuming that $\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x}), \boldsymbol{\zeta}$ are known and such that the integral in (5.12) can be solved in closed-form or efficiently numerically approximated, filtering and smoothing can be done via the Kalman filter- and smoother. If one furthers assumes that $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{x}), \zeta_{\boldsymbol{\theta}}(\boldsymbol{x})$ are parametrized by some $\boldsymbol{\theta}$, estimation of $\boldsymbol{\theta}$ can again be done based on the filtering likelihood function derived in (2.92).

Existence of a Solution Assuming that $\boldsymbol{\Sigma}(\boldsymbol{x})$ is symmetric positive definite for all $\boldsymbol{x}$, the SCDE

$$\begin{cases} \frac{\partial}{\partial t} U(t, \boldsymbol{x}) & = \mathcal{A}(\boldsymbol{x}) U(t, \boldsymbol{x}) + \dot{W}^Q(t, \boldsymbol{x}) \\ U(0, \boldsymbol{x}) & = U_0(\boldsymbol{x}) \end{cases} \tag{5.13}$$

is a linear second-order parabolic SPDE. The existence of a solution is then ensured (under further assumptions of operator-norm boundedness and measurability of $\mathcal{A}$) by e.g. Lototsky and Rozovsky [2017], theorem 4.4.3.

However, deriving the solution in closed-form is not as straight-forward as in our previous example. Therefore, ensuring that the solution is indeed approximated by the spectral decomposition of the previous theorem is subject to further research.

5.1.3. General Linear PDEs

Denote by $\boldsymbol{\beta} \in \mathbb{N}^2$ a multi-index with $|\boldsymbol{\beta}| = \beta_1 + \beta_2$. For any $\boldsymbol{x} \in \mathbb{R}^2$ let $\boldsymbol{x}^{\boldsymbol{\beta}} = x_1^{\beta_1} x_2^{\beta_2}$. Further, let $P(\boldsymbol{x}) = \sum_{|\boldsymbol{\beta}| \leq k} c_\beta \ \boldsymbol{x}^{\boldsymbol{\beta}}$ be some polynomial of order $k \in \mathbb{N}$ with constant coefficients $c_{\boldsymbol{\beta}} \in \mathbb{R}$. We can then define a corresponding differential operator

$$\mathcal{A} = P(\nabla_{\boldsymbol{x}}) := \sum_{|\boldsymbol{\beta}| \leq k} c_{\boldsymbol{\beta}} \frac{\partial^{|\boldsymbol{\beta}|}}{\partial x_1^{\beta_1} \partial x_2^{\beta_2}}. \tag{5.14}$$

By the differential properties of the Fourier basis functions $(\phi_{\boldsymbol{j}})_{\boldsymbol{j}}$ we have:

$$\mathcal{A}\phi_{\boldsymbol{j}} = P(i\boldsymbol{\xi}_{\boldsymbol{j}})\phi_{\boldsymbol{j}}. \tag{5.15}$$

In other words, the Fourier basis functions are eigenfunctions for any linear differential operator $\mathcal{A}$ of type (5.14) with eigenvalues $\lambda_{\boldsymbol{j}} := P(i\boldsymbol{\xi}_{\boldsymbol{j}})$. Further assume that the polynomial P is such that

$$\mathrm{Re}(\lambda_{\boldsymbol{j}}) = \mathrm{Re}(P(i\boldsymbol{\xi}_{\boldsymbol{j}})) < 0 \tag{5.16}$$

for all $\boldsymbol{j} \in \mathbb{Z}^2$. In that case, it is easy to see that the spectral Filtering model derived in chapter 3 can be generalized to the linear SPDE

$$\frac{\partial}{\partial t} U(t, \boldsymbol{x}) = \mathcal{A} U(t, \boldsymbol{x}) + \dot{W}^{Q_k}(t, \boldsymbol{x}).$$

The spectral approximation is again given by a Fourier series decomposition with its spectral processes solving the complex Langevin equation

$$d\alpha_{\boldsymbol{j}}(t) = \lambda_{\boldsymbol{j}} \alpha_{\boldsymbol{j}}(t) dt + \sqrt{\tilde{f}(\boldsymbol{\xi}_{\boldsymbol{j}})} dw_{\boldsymbol{j}}(t), \quad t \in (0, T].$$

Note that the assumption (5.16) guarantees both that $\alpha_{\boldsymbol{j}}$ are well-defined complex Ornstein-Uhlenbeck processes as well as the existence of a weak solution that the spectral approximation converges to. The proof of the second fact follows identically to the one given in theorem 4.3.

5.2. The Stochastic Heat Equation on Riemannian Manifolds

In chapter 3, the need for computationally efficient filtering and parameter estimation for the stochastic convection-diffusion equation motivated a spectral decomposition based on the Fourier basis functions. The following chapter verified that this spectral decomposition indeed approximates the unique weak solution when the problem is defined on a domain with spatially periodic boundaries, i.e. the torus.

This raises the question of how to deal with situations in which the spatial domain is a more complicated manifold. We now give first considerations of how one could go about solving this problem, motivated by the recent work of Borovitskiy et al. [2020] who have generalized Matérn processes on Riemannian manifolds.

Let (M, g) be a compact Riemannian manifold without boundary and Δ_g be the corresponding *Laplace-Beltrami* operator. Consider the stochastic heat equation

$$\begin{cases} dU(t) = \Sigma \Delta_g U(t) dt + dW^{Q_k}(t), \quad t \in (0, \infty), \\ U(0) = U_0, \end{cases} \tag{5.17}$$

where $\Sigma > 0$ is a diffusion constant and W^{Q_k} is a regular Q-cylindrical Wiener process on $L^2(M)$. We assume that the covariance operator of W^{Q_k} relates to a Matérn covariance function, not necessarily restricted to the Whittle case as before.

For now, we restrict ourselves to the heat equation due to the Laplace-Beltrami operators property of inducing an orthonormal basis of eigenfunctions on the space $L^2(M)$ of square-integrable functions on M:

Theorem 5.2. There exists an orthornomal basis of $L^2(M)$ of eigenfunctions $(\phi_j)_j$ of $-\Delta_g$ with non-negative eigenvalues $(\eta_j)_j$ such that $\eta_j \longrightarrow \infty$.

Proof. See e.g. Canzani [2013], theorem 44. □

This fact was also used by Borovitskiy et al. [2020] to show:

Theorem 5.3. Let $(\phi_j)_j$ be the orthornormal basis of $L^2(M)$ of eigenfunctions of $-\Delta_g$ with non-negative eigenvalues $(\eta_j)_j$. Then, the Matérn covariance function on (M, g) is given by

$$k(x, y) = \sum_{j \geq 1} \sigma_0^2 \left(\rho_0^{-2} + \eta_j\right)^{-\nu - \frac{d}{2}} \phi_j(x) \phi_j(y) \tag{5.18}$$

for all $x, y \in M$.

Proof. See Borovitskiy et al. [2020], theorem 5. □

With these two results, we are able to show:

Theorem 5.4. There exists a unique weak solution $(U(t))_t$ to the stochastic heat equation (5.17) on (M, g). If the initial condition $U_0 = \sum_{j\geq 1} \alpha_{j,0}\phi_j$ is a regular generalized Gaussian random field on $L^2(M)$, the weak solution is given by

$$U(t) = \sum_{j\geq 1} \alpha_j(t)\phi_j \tag{5.19}$$

where $(\alpha_j)_j$ are the Ornstein-Uhlenbeck processes given by

$$\alpha_j(t) = \exp(\lambda_j t)\alpha_{j,0} + \sqrt{f_j}\int_0^t \exp(\lambda_j(t-s))dw_j(s), \tag{5.20}$$

with parameters

$$\begin{aligned} \lambda_j &:= -\Sigma\eta_j, \\ f_j &:= \sigma_0^2\left(\rho_0^{-2} + \eta_j\right)^{-\nu-\frac{d}{2}}. \end{aligned} \tag{5.21}$$

Proof. The proof follows the same manner as in theorem 4.3 and we therefore give the main steps only.

By theorem 5.3, we can construct a $Q-$Wiener process on $L^2(M)$ that corresponds to the Matérn kernel by

$$W^{Q_k}(t) = \sum_{j\geq 1} \sqrt{f_j}\phi_j w_j(t).$$

As there exists an orthonormal basis $(\phi_j)_j$ of $-\Delta_g$ on $L^2(M)$ with eigenvalues $\eta_j \geq 0$, $\Sigma\Delta_g$ generates a diagonal C_0-semigroup on $L^2(M)$ defined by $S(t)h = \sum_{j\geq 1}\exp(\lambda_j t)$, $\lambda_j = -\Sigma\eta_j \leq 0$, and one shows that

$$\|S(t)\sqrt{Q_k}\|_2^2 \leq \sum_{j\geq 1} f_j < \infty.$$

Following theorem 2.11, there exists a unique weak solution

$$U(t) = S(t)U_0 + \int_0^t S(t-s)dW^{Q_k}(s).$$

By the definition of the Itô integral plus the fact that $(\phi_j)_j$ are eigenfunctions of $-\Delta_g$ we receive

$$\int_0^t S(t-s)dW^{Q_k}(s) = \sum_{j\geq 1}\sqrt{f_j}\int_0^t \exp(\lambda_j(t-s))dw_j(s).$$

Lastly, under the assumption that $U_0 = \sum_{j\geq 1}\alpha_j(0)\phi_j$ is a regular generalized Gaussian random field on $L^2(M)$, we have

$$S(t)U_0 = \sum_{j\geq 1}\exp(\lambda_j t)\alpha_{\boldsymbol{j}}(0)\phi_j.$$

Setting $(\alpha_j)_j$ as in the proposition ends the proof. □

As in the previous chapters, this transforms any statistical task concerning the infinite dimensional SPDE solution into a task on the one-dimensional spectral processes $(\alpha_j)_j$. These are again uncoupled Ornstein-Uhlenbeck processes for which computationally efficient algorithms are readily available.

Of course, the non-trivial task in this situation is finding the eigenfunctions and eigenvalues of Δ_g. In the following subsection, we provide some insight into an example of high practical importance - the case where the underlying manifold is the unit sphere S^{d-1}.

5.2.1. The Stochastic Heat Equation on S^{d-1}

Let S^{d-1} be the d-dimensional unit sphere

$$S^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}.$$

Denote by Δ_S the Laplace-Beltrami operator on the unit sphere. For any $j \in \mathbb{N}$, let $\mathcal{H}^j$ be the space of homogeneous, harmonic polynomials ϕ on $\mathbb{R}^d$ of degree j, meaning that ϕ is of the form $\phi(\boldsymbol{x}) = \sum_{|\boldsymbol{\beta}|=j} c_\beta \, \boldsymbol{x}^{\boldsymbol{\beta}}, c_\beta \in \mathbb{C}, \boldsymbol{x} \in \mathbb{R}^d$, such that

$$\Delta\phi = 0,$$

where Δ is the usual Laplacian on $\mathbb{R}^d$. Further denote by H^j the homogeneous, harmonic polynomials ϕ restricted onto the unit sphere S^{d-1}:

$$H^j := \{\phi_{|S^{d-1}} : \phi \in \mathcal{H}^j\} \subset L^2(S^{d-1}).$$

Then one can show that:

Theorem 5.5. Any $\phi \in H^j, j \in \mathbb{N}$, is an eigenfunction of $-\Delta_S$ with eigenvalue $\eta_j = j(j+d-2)$:

$$-\Delta_S\phi = j(j+d-2)\phi.$$

Furthermore, it holds that

$$L^2(S^{d-1}) = \bigoplus_{j\geq 0} H^j, \tag{5.22}$$

where $\bigoplus$ denotes the orthogonal sum of Hilbert spaces in regards to $\langle\cdot,\cdot\rangle_{L^2}$.

Proof. These results are derived in e.g. Canzani [2013], section 5.4. □

A direct consequence of the previous theorem is that for any $h \in L^2(S^{d-1})$, we have

$$h = \sum_{j\geq 0} \left(\sum_{k=1}^{\dim(H^j)} \langle h, \phi_k^j \rangle_{L^2} \phi_k^j \right)$$

where the convergence is in the L^2-sense and ϕ_k^j, $k = 0, ..., \dim(H^j)$, are the orthonormal bases of H^j respectively, also referred to as *spherical harmonics of degree j and order k*. Note that (see Canzani [2013], cor. 28)

$$\dim(H^j) = (2j + d - 2)\frac{(j + d - 3)!}{j!(d - 2)!} \tag{5.23}$$

is the geometric multiplicity of the j-th eigenvalue η_j of Δ_S.

By theorem 5.4, the weak solution to the stochastic heat equation (5.17) on the sphere S^{d-1} is therefore given by

$$U(t) = \sum_{j \geq 0} \left(\sum_{k=1}^{\dim(H^j)} \alpha_k^j(t)\phi_k^j \right),$$

where for every $j \in \mathbb{N}$, α_k^j, $k = 1, ..., \dim(H^j)$, are independent Ornstein-Uhlenbeck processes as given in (5.20) and (5.21) with $\eta_j = j(j + d - 2)$.

This result has the practical side effect that the computation of the propagation- and covariance matrices for filtering and smoothing of the spectral processes α_j^k must only be done once for every $j \in \mathbb{N}$, therefore making any statistical inference based on the spectral approximation even more computationally efficient. We give first experimental results for the spectral filtering and smoothing of the stochastic heat equation in the following subsection.

5.2.2. Spectral Filtering and Smoothing for the Stochastic Heat Equation on S^2

Let $d = 3$. For the unit sphere S^2 in $\mathbb{R}^3$ we get from (5.23) that for any $H^j, j \in \mathbb{N}$:

$$\dim(H^j) = 2j + 1.$$

For any $j \in \mathbb{N}$, an orthonormal basis of H^j is given by the *real-valued spherical harmonics* $\phi_k^j, k \in \{-j, ..., 0, ...j\}$, defined as

$$\phi_k^j(\theta, \iota) = \begin{cases} \sqrt{2}C_k^j \cos(k\iota)P_k^j(\cos(\theta)), & \text{if } k > 0, \\ C_0^j P_0^j(\cos(\theta)), & \text{if } k = 0, \\ \sqrt{2}C_k^j \sin(-k\iota)P_{-k}^j(\cos(\theta)), & \text{if } k < 0, \end{cases}$$

where $(\theta, \iota) \in [0, \pi] \times [0, 2\pi]$ are the spherical coordinates, C_k^j are constants such that $\|\phi_k^j\|_{L^2} = 1$ and P_k^j are the *associated Legendre polynomials* (see Canzani [2013], chapter 5.4 for more details). For any degree $j \in \mathbb{N}$, the eigenvalue η_j of $-\Delta_S$ for the spherical harmonics of degree j is given by

$$\eta_j = j(j + 1).$$

Data Simulation In the same manner as previously done in chapter 3.4, we simulate the weak solution to the stochastic heat equation as given in theorem 5.4 on S^2 by means of simulating the forward propagation of the spectral processes $(\alpha_k^j(t))_t, k \in -j, ..., j, j \in 1, ..., J$.

We skip the exact details, as the procedure follows our previous one with only two exceptions. First, for the sake of simplicity, we assume that observations of the spatiotemporal process are not corrupted by measurement noise.

Secondly, the transformation between the spectral- and space-time domain is not given the DFT $\mathbf{\Phi}$ anymore, but rather must be done by direct numerical integration of the spherical harmonics coefficients $\alpha_k^j(t) = \langle \boldsymbol{U}(t), \phi_k^j \rangle_{L^2}$. For our simulation, this is done using the Gauss–Legendre quadrature method implemented in the *SHtools* python library (Wieczorek and Meschede [2018]).

Figure 5.1 shows a number of samples for the simulated process with the following model parameters

$$\Sigma = 0.005,\ \sigma_0^2 = 3.5,\ \rho_0 = 0.4.$$

The initial condition $\boldsymbol{U}_0$ is chosen to be a spherical Matérn field with parameters

$$\sigma_{init}^2 = 8,\ \rho_{init} = 0.5.$$

From the initial condition, the process is simulated forward over the time domain $T = [0, 10]$ with discretized time steps of length $\Delta t = 0.02$. The spherical harmonics of up to degree $J = 100$ have been used for this simulation.

Experimental Results: Filtering and Smoothing For the filtering and smoothing experiments, we assume that full observations of the simulated process $\boldsymbol{U}(t)$ - and therefore of the spectral processes $(\alpha_k^j(t))_{j,k}$ - are available at every 25-th time step (or, in other words with time steps of length $\Delta t_{obs} = 0.5$). We run the Kalman filter and -smoother as specified in algorithm 2.1 and algorithm 2.2. The results are presented in figures 5.2 and 5.3.

Unsurprisingly, we obtain results likening the ones in section 3.4, with the Kalman smoother outperforming the filter by a noticeable larger degree than before. This is due to the observations being more sparse in time compared to our previous experiments.

Likewise to the Fourier spectral decomposition, the spherical harmonics decomposition is dominated by its low degree coefficients, with spectral processes corresponding to high degrees rapidly converging to their stationary distribution centered around 0 (see fig. 5.2).

As a result, for both the Kalman filter- and smoother, the reconstruction based on only the first 5 degrees almost matches the reconstruction performance using the full 100 degrees that the simulation is based on (see fig. 5.3).

For a lower degree cutoff (here, $j <= 2$), the reconstruction becomes considerably worse. Note also that in this case, the mean squared error in the reconstruction differs only marginally between the Kalman filter and -smoother, due to the fact that its error

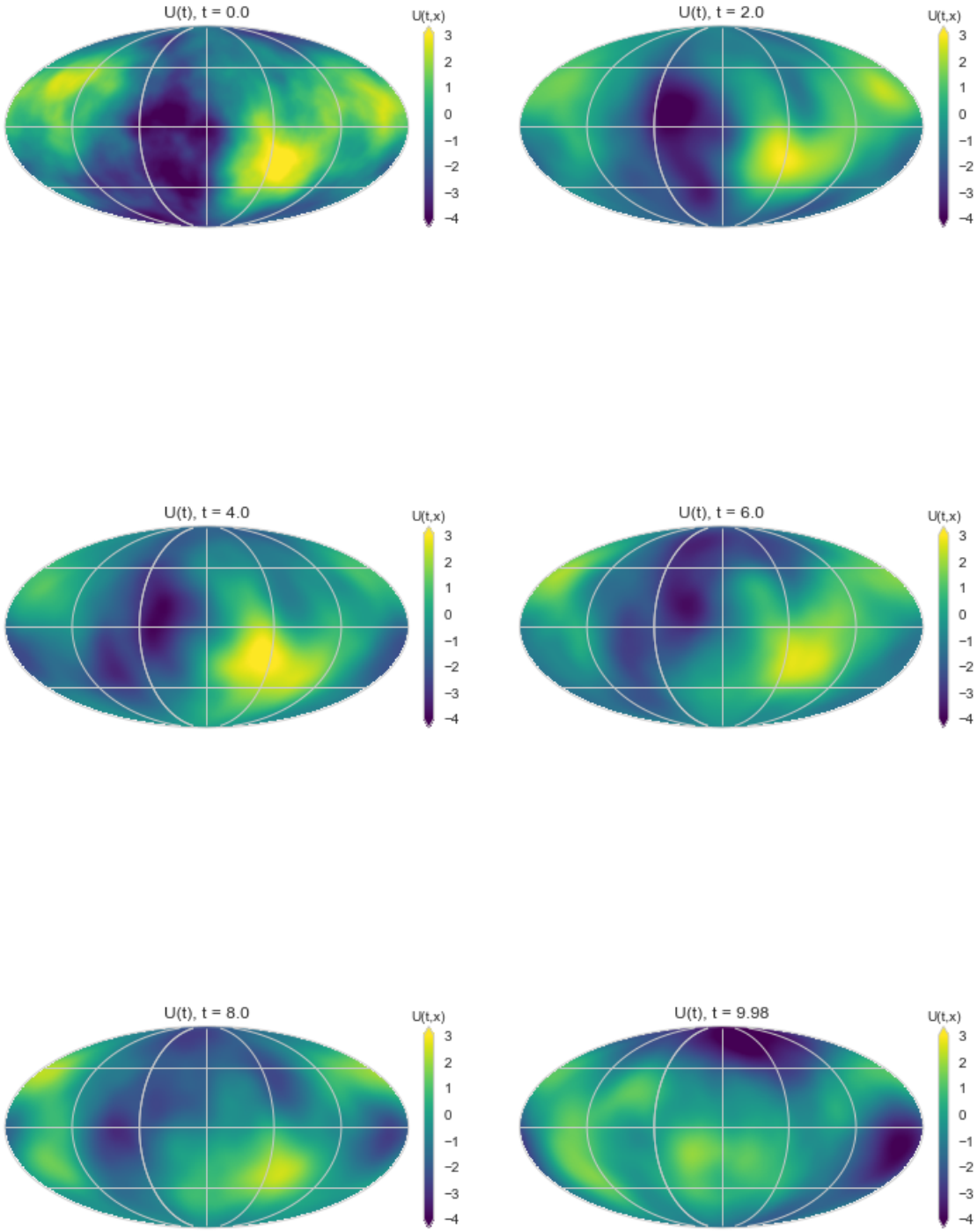

Figure 5.1.: Examples of the simulated solution $\boldsymbol{U}(t)$ of the stochastic heat equation on the sphere S^2. Different panels show different simulation times.

stems mostly from the low number of spherical harmonics used in the decomposition, not from the wrong estimation of their respective coefficients.

We leave the task of inferring the model parameters for future projects. However, given that this task is merely dependent on the filtering likelihood of the spectral processes, we expect similarly good results as obtained in section 3.4.

5.3. Non-Linear SPDEs

We finish this thesis with a first heuristic consideration of how to approach non-linear SPDEs with Fourier spectral filtering method. We make use of notation as introduced in chapter 3.2.

As a first example, consider the nonlinear *stochastic Burger's equation*

$$\begin{cases} \partial_t U(t,x) & = U(t,x)\partial_x U(t,x) + \Sigma\, \partial_x^2 U(t,x) + \dot{W}^{Q_k}(t,x), \quad (t,x) \in [0,\infty) \times [0,1] \\ U(0,x) & = U_0(x), \end{cases} \tag{5.24}$$

with spatially periodic boundary conditions, diffusion constant $\Sigma > 0$ and stochastic forcing term $\dot{W}^{Q_k}(t,x)$. Assume $\dot{W}^{Q_k}(t,x)$ is temporally white and spatially colored and is approximated by the finite sum

$$\dot{W}^{Q_k}(t,x) \approx \dot{W}^{\tilde{Q}}(t,x) := \sum_{j \in \mathcal{J}} \eta_j(t)\phi_j(x),$$

with $\eta_{\boldsymbol{j}}(t) \sim \mathcal{CN}(0, \tilde{f}(\xi_j))$. Once again, the ansatz is to look for a solution of the form

$$U(t,x) = \sum_{j \in J} \alpha_j(t)\phi_j(x)$$

with temporally evolving spectral processes $\alpha_j(t)$. Substituting the ansatz and approximated forcing term $\dot{W}^{\tilde{Q}}$ into (5.24), we receive

$$\begin{aligned} \sum_{j \in J} \partial_t \alpha_j(t)\phi_j(x) = &\left(\sum_{j \in J} \alpha_j(t)\phi_j(x) \right) \left(\sum_{j' \in \mathcal{J}} i\xi_{j'} \alpha_{j'}(t)\phi_{j'}(x) \right) \\ &- \sum_{j \in \mathcal{J}} \Sigma \xi_j^2 \alpha_j(t)\phi_j(x) + \sum_{j \in \mathcal{J}} \eta_j(t)\phi_j(x). \end{aligned}$$

Noting that for all $j, j' \in \mathbb{Z}$, we have $\phi_j(x)\phi_{j'}(x) = \phi_{j+j'}(x)$, we get

$$\left(\sum_{j \in J} \alpha_j(t)\phi_j(x) \right) \left(\sum_{j' \in \mathcal{J}} i\xi_{j'} \alpha_{j'}(t)\phi_{j'}(x) \right) = \sum_{j,j' \in \mathcal{J}} i\xi_{j'} \alpha_j(t)\alpha_{j'}(t)\phi_{j+j'}(x).$$

Therefore, by matching spectral processes, we receive a system of coupled stochastic differential equations

$$d\alpha_j(t) = -\Sigma \xi_j^2 \alpha_j(t)dt + \left(\sum_{k \in \mathcal{J}} i\xi_k \alpha_{j-k}(t)\alpha_k(t) \right) + \sqrt{f(\tilde{\xi}_j)} dw_j(t).$$

The corresponding multivariate SDE $\boldsymbol{\alpha}(t) = [\alpha_j(t), j \in \mathcal{J}]$ is highly nonlinear. A possible solution to the smoothing problem and parameter inference could be obtained using *Backward Filtering Forward Guiding* (Mider et al. [2020]).

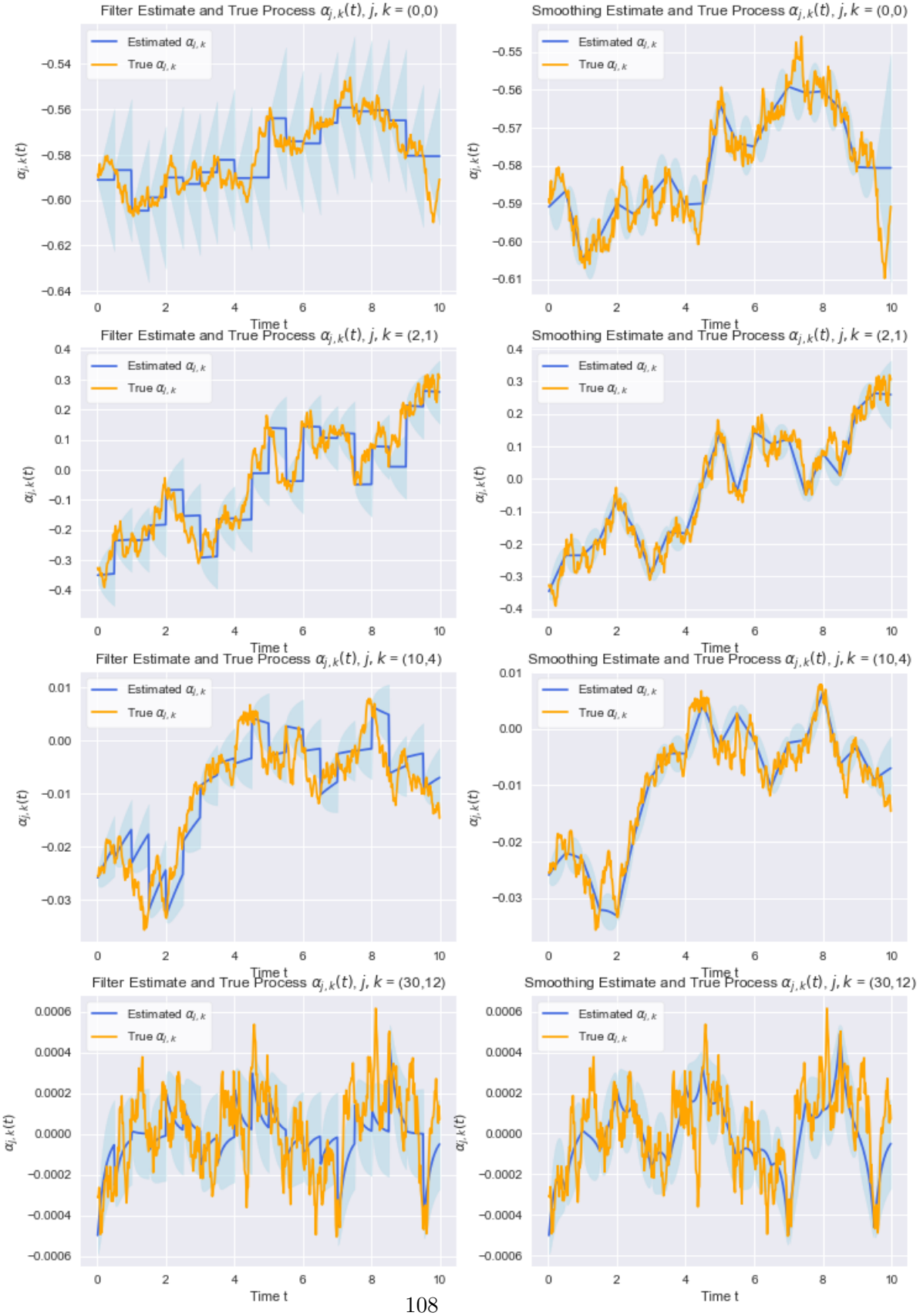

Figure 5.2.: Filtering and smoothing paths for spectral processes $\alpha_k^j(t)$ of different degrees j and orders k. Different colors represent the true and estimated process. The shaded areas represent the mean estimate $\pm$ the estimated standard deviation $m(t) \pm \sqrt{K(t)}$.

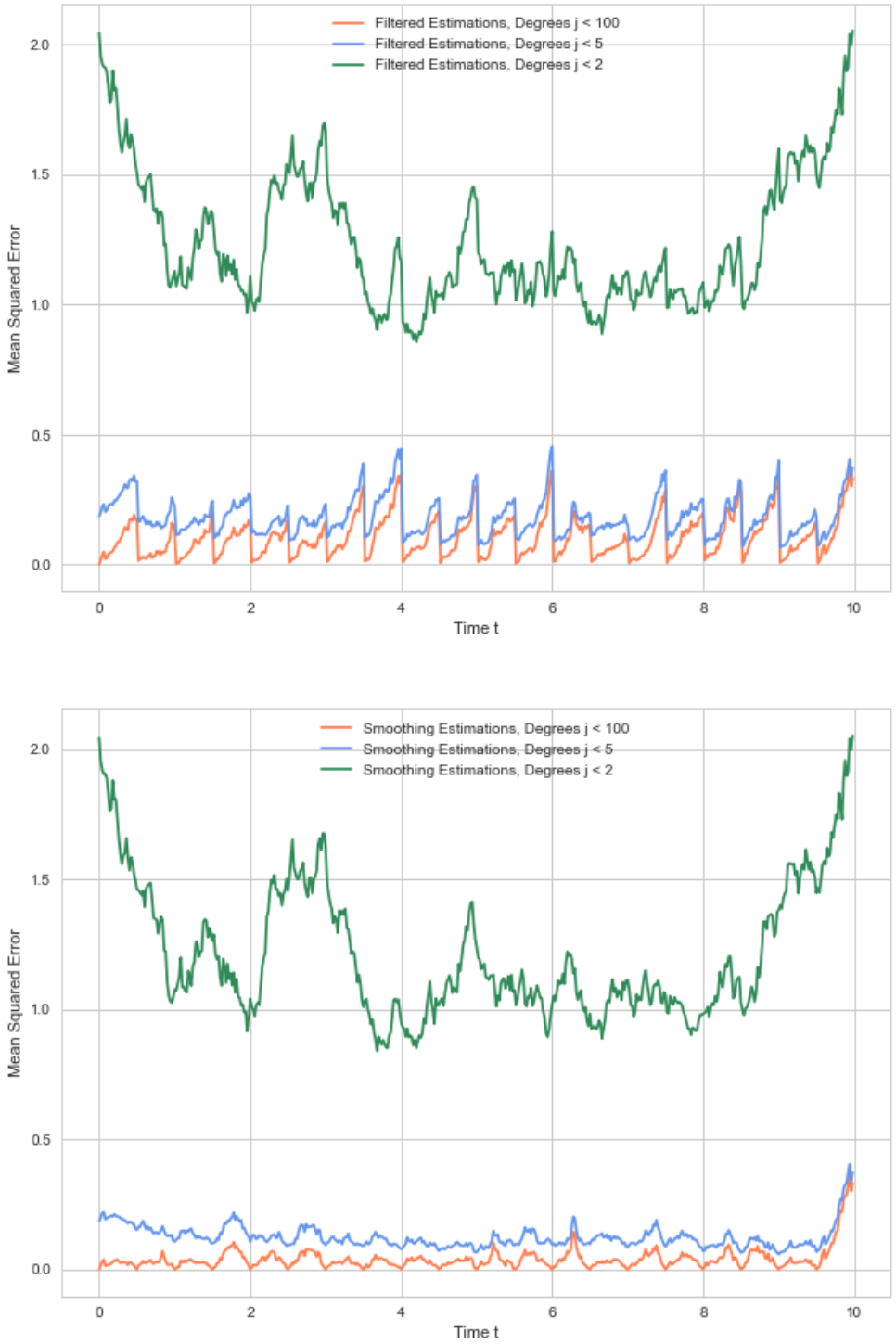

Figure 5.3.: Mean squared error $\|\boldsymbol{U}(t) - \tilde{\boldsymbol{U}}_J(t)\|^2$ between the true simulated process $\boldsymbol{U}(t)$ and the reconstructed process $\tilde{\boldsymbol{U}}_J(t)$ based on the filtering (upper panel) and smoothing estimates (lower panel) of the spectral processes $\alpha_k^j(t)$ of degrees $j = 0, ..., J$. Different colors represent different numbers of spherical harmonics degrees considered in the reconstruction.

Bibliography

Salomon Bochner. *Lectures on Fourier Integrals*. Annals of Mathematics Studies. Princeton University Press, 1959.

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc Peter Deisenroth. Matérn gaussian processes on riemannian manifolds. *34th Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

Charles-Edouard Bréhier. A short introduction to stochastic pdes. art. hal-00973887v1f, 2014.

William L. Briggs and Van Emden Henson. *The DFT - An Owner's Manual for the Discrete Fourier Transform*. Society for Industrial and Applied Mathematics, 1995.

Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.

Claudio Canuto, M.Yousuff Hussaini, Alfio Maria Quarteroni, and Thomas A.Jr. Zang. *Spectral Methods in Fluid Dynamics*. Springer-Verlag Berlin Heidelberg, 1 edition, 1988.

Yaiza Canzani. *Analysis on Manifolds via the Laplacian*. Harvard University, 2013.

James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19:297 – 301, 1965.

Noel Cressie and Christopher K. Wikle. *Statistics for Spatio-Temporal Data*, volume 1. Wiley, 2011.

Ruth F. Curtain and Hans Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer-Verlag, 1995.

Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1 edition, 1992.

David Gottlieb and Steven A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications*. Society for Industrial and Applied Mathematics, 1 edition, 1977.

Matthew M. Graham, Alexandre H. Thiery, and Alexandros Beskos. Manifold Markov chain Monte Carlo methods for Bayesian inference in a wide class of diffusion models. *arXiv e-prints*, art. arXiv:1912.02982, December 2019.

K. Jörgens and G.F. Roach. *Linear Integral Operators*. Pitman Advanced Pub. Program, 1982.

Gopinath Kallianpur and P. Sundar. *Stochastic Analysis and Diffusion Processes*. Oxford Graduate Texts in Mathematics, 2014.

Rudolf Kálmán. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering (Series D)*, 82:35–45, 1960.

Hui-Hsiung Kuo. *Introduction to Stochastic Integration*. Springer-Verlag New York, 2006.

Amos Lapidoth. *A Foundation in Digital Communication*. Cambridge University Press, 2009.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73:423 – 498, 2011.

Georg Lindgren. *Stationary Stochastic Processes - Theory and Application*. Chapman and Hall/CRC, 2012.

Xiao Liu, Kyongmin Yeo, and Siyuan Lu. Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association*, 0(0):1–18, 2021.

Sergey V. Lototsky and Boris L. Rozovsky. *Stochastic Partial Differential Equations*. Springer International Publishing, 2017.

Bertil Matérn. *Spatial Variation*. Springer-Verlag New York, 1960.

Peter S. Maybeck. *Stochastic Models, Optimization and Control: Volume 1*. Academic Press, 1979.

Isambi Mbalawata, Simo Särkkä, and Heikki Haario. Parameter estimation in stochastic differential equations with markov chain monte carlo and non-linear kalman filtering. *Computational Statistics*, 28:1195–1223, 2013.

Edward James McShane. *Integration*. Princeton University Press, 1947.

Marcin Mider, Moritz Schauer, and Frank van der Meulen. Continuous-discrete smoothing of diffusions. *arXiv e-prints*, art. arXiv:1712.03807, 2020.

CE. Rasmussen and CKI. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 2006.

Walter Rudin. *Fourier Analysis on Groups*. Wiley Online Library, 1962.

Jan Seidler. Da prato-zabczyk's maximal inequality revisited. *Mathematica Bohemica*, 118:67–105, 1993.

Fabio Sigrist, Hans R. Künsch, and Werner A. Stahel. An spde based spatio-temporal model for large data sets with an application to postprocessing precipitation forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:3–33, 2015.

Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2019.

Elias M. Stein and Rami Shakarchi. *Fourier Analysis*. Princeton University Press, 2003.

Michael L. Stein. *Interpolation of Spatial Data*. Springer Series in Statistics. Springer-Verlag, 1999.

Mark C. Veraar. Non-autonomous stochastic evolution equations and applications to stochastic partial differential equations. *Journal of Evolution Equations*, 10:85–127, 2010.

Alan J. Weir. *General Integration and Measure*. Cambridge University Press, 1974.

Peter Whittle. On stationary processes in the plane. *Biometrika*, 41:434 – 449, 1954.

Peter Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40:974–994, 1963.

Mark A. Wieczorek and Matthias Meschede. Shtools: Tools for working with spherical harmonics. *Geochemistry, Geophysics, Geosystems*, 19:2574–2592, 2018.

Akiwa M. Yaglom. *Correlation Theory of Stationary and Related Random Functions, Volume 1: Basic Resuls*. Springer-Verlag New York, 1987a.

Akiwa M. Yaglom. *Correlation Theory of Stationary and Related Random Functions, Volume 2: Supplementary Notes and References*. Springer-Verlag New York, 1987b.

A. Appendix

A.1 Lebesgue-Stieltjes Integral

We summarize the construction of the Lebesgue-Stieltjes integral for functions of several variables. For a detailed introduction, see e.g. McShane [1947], chapter 7, or Weir [1974], chapter 9.

Asssume $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ to be a bounded function with non-negative increments $\Delta F(t)$. We may then define a pre-measure μ_F on the intervals Δt via

$$\mu_F(\Delta t) := \Delta F(t).$$

The Carathéodory extension theorem extends μ_F to a unique measure on $\mathcal{B}(\mathbb{R}^n)$, referred to as the *Lebesgue-Stieltjes measure* of F. We call the integral in regards to μ_F the *Lebesgue-Stieltjes integral* and denote it by

$$\int_{\mathbb{R}^n} g(t)dF(t) := \int_{\mathbb{R}^n} g(t)\mu_F(dt). \tag{A.1}$$

Note that the Lebesgue-Stieltjes integral agrees with - if existent - the *Riemann-Stieltjes integral*, defined over any interval $T \subset \mathbb{R}^n$ as the limit

$$\int_T g(t)dF(t) := \lim_{k\to\infty} \sum_{\Delta t \in \mathcal{P}_k} g(t')\Delta F(t).$$

Over $\mathbb{R}^n$, this integral is defined as an improper one in the usual way.

The Riemann sum representation also suggests - and indeed it is true that - when F is smooth enough with the partial derivatives

$$f(t) := \frac{\partial^n}{\partial t_1 ... \partial t_n} F(t) = \lim_{|\Delta t|\to 0} \frac{\Delta F(t)}{|\Delta t|}$$

existing, the Lebesgue-Stieltjes integral reduces to the Lebesgue integral

$$\int_{\mathbb{R}^n} g(t)dF(t) = \int_{\mathbb{R}^n} g(t)f(t)dt.$$

Furthermore, for any function $G : \mathbb{R}^n \longrightarrow \mathbb{R}$, let the *total variation* of G over an interval $T \subset \mathbb{R}^n$ be defined as

$$V_G(T) := \sup_{\mathcal{P}} \left\{ \sum_{\Delta t \in \mathcal{P}} \Delta G(t) \ : \ \mathcal{P} \text{ is a sub-interval partition of } T \right\}.$$

If $V_G(T) < \infty$, we call G of bounded variation on I.

It can be shown that a function of bounded variation is of the form $G = G_1 - G_2$, where both G_1, G_2 are bounded and have non-negative increments. One then defines the Lebesgue-Stieltjes integral in regards to G via

$$\int g dG = \int g dG_1 - \int g dG_2.$$

A.2 Itô's Lemma

Theorem A.1 (Itô's Lemma - Multidimensional Case)**.** Let $(\boldsymbol{X}(t))_t$ be a n-dimensional *Itô process* of the form

$$\boldsymbol{X}(t) = \boldsymbol{X}_0 + \int_0^t \boldsymbol{f}(\tau) d\boldsymbol{w}(\tau) + \int_0^t \boldsymbol{g}(\tau) d\tau, \quad 0 \leq t \leq T \tag{A.2}$$

where $\boldsymbol{w}$ is an m-dimensional Brownian motion, $\boldsymbol{f}(\cdot)$ is a $(n \times m)-$dimensional process such as above, and $\boldsymbol{g}(\cdot)$ is a n-dimensional process such that g_i is a.s. integrable on $[0, T]$ for all $1 \leq i \leq n$.

Let $\phi(t, x_1, ..., x_n)$ be a continuous function with continuous partial derivatives $\partial_t \phi$, $\partial_{x_i} \phi$ and $\partial^2_{x_i x_j} \phi$ for all $1 \leq i, j \leq n$. Then

$$\begin{aligned} \phi(t, \boldsymbol{X}(t)) = \phi(0, \boldsymbol{X}(0)) &+ \int_0^t \partial_t \phi(\tau, \boldsymbol{X}(\tau)) d\tau \\ &+ \sum_{i=1}^n \int_0^t \partial_{x_i} \phi(\tau, \boldsymbol{X}(\tau)) dX_i(\tau) \\ &+ \frac{1}{2} \sum_{i,j=1}^n \int_0^t \partial^2_{x_i x_j} \phi(\tau, \boldsymbol{X}(\tau)) dX_i(\tau) dX_j(\tau). \end{aligned} \tag{A.3}$$

where $dX_i(\tau) dX_j(\tau)$ is computated via $dw_j(\tau) dw_i(\tau) = \delta_{i,j} d\tau$, $dw_j(\tau) d\tau = 0$ and $d\tau d\tau = 0$.

Proof. See Kuo [2006], theorems 7.4.2 and 7.4.3. □