Inference for Stochastic Partial Differential Equations and Chaos Decomposition of the Negative Binomial Process

Ph.D. thesis

Bo Markussen
Department of Statistics and Operations Research
Institute for Mathematical Sciences
Faculty of Science
University of Copenhagen

Thesis advisor:	Michael Sørensen, University of Copenhagen
Thesis committee:	Martin Jacobsen, University of Copenhagen
	Jean Jacod, Université Paris VI Pierre et Maria Curie
	Liudas Giraitis, London School of Economics

Bo Markussen
Department of Statistics and Operations Research
University of Copenhagen
Universitetsparken 5
DK–2100 Copenhagen East
Denmark

e-mail: markusb@math.ku.dk

Preface

This thesis has been prepared in partial fulfilment of the requirements for the Ph.D. degree at the Department of Statistics and Operations Research, Institute for Mathematical Sciences at the University of Copenhagen. The work has been carried out in the period from November 1998 to November 2001 with professor Michael Sørensen as thesis advisor. The present version differs from the original one which was submitted on November 14, 2001, in that some typos and misprints have been corrected.

During the work of this thesis I visited *Institut für Mathematik* at the *Humboldt–Universität zu Berlin* from September 1999 to January 2000. This visit was partly supported by a *Forschungskurzstipendium* from the *Deutscher Akademischer Austauschdienst*, and their help is gratefully acknowledged. I want to express my gratitude to Professor Dr. Uwe Küchler for his warm welcome, and his concern to make my stay in Berlin both successful and pleasant. I also thank everyone else at the institute in *Burgstraße*, and especially Markus Reiß together with whom I read Malliavin calculus, and Ulrike Putschke who later visited Copenhagen.

Special thanks are due to my advisor Michael Sørensen for his encouragement and patience, for the numerous discussions during the last three years, and for reading all the early versions of the papers constituting this thesis. I would also like to thank docent Martin Jacobsen for his help, and everyone else at the Department of Statistics and Operations Research for the unique and always enjoyable and friendly atmosphere. Thanks are due to Helle Sørensen for providing me with the TEX-code producing the layout of this thesis.

Finally, I would like to thank my family, and especially my girlfriend Charlotte, for their support. Additional thanks are due to my brother Jan for reading the introductory essay.

Copenhagen, January 2002

Bo Markussen

Summary

The underlying theme of this thesis is statistical inference for stochastic partial differential equations observed at discrete points in time and space. The thesis consists of an introductory essay on the notion of stochastic partial differential equations and of the four enclosed papers. The first two papers are concerned with statistical inference for time series embedded in linear stochastic partial differential equations driven by Gaussian noise. The focus of the last two papers is the notion of chaos decomposition and Malliavin calculus for the underlying stochastic process. The purpose of these two papers is to investigate non-Gaussian stochastic calculus, and to compute the likelihood function given an observation at discrete points of a stochastic process defined in this framework. The prerequisite for reading this thesis is familiarity with linear partial differential equations, statistical inference based on the maximum likelihood method, spectral analysis for time series, Itô calculus, and functional analysis with emphasis on Schwartz distributions and Hilbert space theory.

In the introductory essay, models based on stochastic partial differential equations and the associated statistical inference problem is discussed. Particular partial differential equations are presented in order to exemplify the usage of equations disturbed by stochastic noise, equations with stochastic coefficients, and the usage of non-Gaussian noise. Many stochastic partial differential equations do not possess solutions in the ordinary sense due to the roughness of the underlying noise process, and a notion of a generalized solution is needed. Two different approaches employing spatial respectively stochastic distributions are presented. The latter approach allows a mathematically rigorous definition of white noise as well as the definition of the so-called Wick product. Some mathematical properties of these notions are presented, and the possibility of extending this stochastic calculus to non-Gaussian and Lévy processes is discussed briefly. From a model building and interpretation point of view the fundamental question is whether or not the Wick product should be used. It is argued that this question can actually be answered on the basis of the interpretations of the Wick product and the ordinary product. This is illustrated by two examples. Regarding statistical inference for models based on stochastic partial differential equations, the emphasis in the literature has been on spectral data for parabolic equations. These models are presented. Moreover, the four papers enclosed in the thesis are reviewed relative to the presented theory and literature.

An observation at discrete points in time and space of a stochastic partial dif-

ferential equation can be viewed as a time series. In Markussen (2001*d*), paper I of the thesis, this connection between stochastic partial differential equations and time series is exemplified for a parabolic equation. The main part of the paper is concerned with the local asymptotic properties of the likelihood function given an observation of a multivariate Gaussian time series. Especially, the well-known asymptotic properties of the maximum likelihood are proven. The required regularity conditions are given as integrability and smoothness conditions on the spectral densities. These regularity conditions are very general and include time series with long range dependence, *i.e.* spectral densities with a pole, and to some extend also spectral densities with a zero. Sequences of misspecified models, which contain the true spectral density in the limit, are also considered. The main theorem states a uniform version of the local asymptotic normality property known from Le Cam theory. The mathematical techniques used are p-norms for matrices, L^p-norms for multivariate spectral densities, and maximal inequalities for stochastic processes. Moreover, the needed properties of Fejér approximations and Toeplitz matrices are stated and proven. In the existing time series literature the p-norms and the L^p-norms are usually used only for $p = 1, 2, \infty$. The novelty of this paper probably lies in the usage of these norms for every $p \in [1, \infty]$.

In Markussen (2001*b*), paper II of the thesis, parametric inference given an observation at discrete lattice points in time and space of a hyperbolic stochastic partial differential equation is studied. An approximate state space model is proposed, and the associated likelihood function is calculated via the Kalman filter. The results developed in Markussen (2001*d*) are used to give conditions ensuring asymptotic efficiency. The parabolic equations can be obtained as limits of hyperbolic equations, and the asymptotic distribution of the likelihood ratio test statistic for a parabolic equation against the hyperbolic alternative is stated. Moreover, sample path properties of the parabolic respectively hyperbolic equations are studied.

In Markussen (2001*a*), paper III of the thesis, the chaos decomposition of the Hilbert space of quadratic integrable functionals of the negative binomial process is studied. Familiarity with functional analysis and especially firm knowledge of Hilbert space theory is probably required of the potential reader. The structure of the chaos decomposition found in the paper is different from the symmetric Fock space well-known for Gaussian and Poisson processes. Especially, the multiple integrals are not capable of generating all the quadratic integrable functionals. The developed chaos decomposition is used to introduce Malliavin derivative operators corresponding to variational derivatives w.r.t. the jump times respectively the jump heights. Although the latter derivative operator is not densely defined, this operator can be defined as a closed operator on a subspace containing the multiple integrals. The introduced Malliavin operators do not have a diagonal representation w.r.t. the developed chaos decomposition. However, the corresponding Wick product is easily expressed in terms of the chaos decomposition. The negative binomial process is investigated as a generic example of what could possibly be done for Lévy processes and especially compound Poisson processes. The developed chaos decomposition is directly connected to the characteristic functional of the negative binomial process, and does not use the identification of a compound Poisson process with the Poisson process defined on the cartesian product of the parameter space and the jump space.

If the chaos expansion of a finite dimensional random variable is known, then the corresponding Lebesgue density can be expressed as an expectation via the integration by parts setting from Malliavin calculus. In Markussen (2001*c*), paper IV of the thesis, this property is used to propose a simulation approach to compute pseudo-likelihoods given an observation at discrete points of a stochastic partial differential equation. The proposed method is developed for Gaussian spaces and requires the chaos decomposition to be known. The paper contains a short introduction to the needed Malliavin calculus, and the simulation of the involved iterated Skorohod integrals is discussed. The proposed method is seen to be numerically demanding. Moreover, the special case corresponding an observation with measurement errors exemplifies, that the method is numerically instable as well. The paper thus concludes, that the proposed method is of little practical usage.

Dansk resumé

Det underliggende tema for denne afhandling er statistisk inferens for stokastiske partielle differentialligninger observeret i diskrete punkter i tid og rum. Afhandlingen består af et indledende essay om stokastiske partielle differentialligninger og af de fire inkluderede artikler. De to første artikler omhandler statistisk inferens for tidsrækker indlejret i lineære stokastiske partielle differentialligninger drevet af Gaussisk støj. Fokus i de to sidste artikler er begreberne kaos dekomponering og Malliavin kalkyle for den underliggende stokastiske proces. Formålet med disse to artikler er at undersøge ikke-Gaussisk stokastiske kalkyle, samt i denne matematiske ramme at beregne likelihoodfunktionen givet diskrete observationer af stokastiske processer. Forudsætningerne for at læse denne afhandling er kendskab til lineære partielle differentialligninger, statistisk inferens baseret på maksimum likelihood metoden, spektral analyse for tidsrækker, Itô kalkyle, og funktional analyse med særlig vægt på Schwartz distributioner og teorien for Hilbert rum.

I det introducerende essay bliver modeller bestående af stokastiske partielle differentialligninger og det tilhørende statistiske inferens problem diskuteret. Der bliver givet konkrete eksempler på partielle differentialligninger henholdvis drevet af stokastisk støj og med stokastiske koefficienter. Videre gives der et eksempel på en ligning med ikke-Gaussisk støj. Rugheden af den underliggende stokastisk proces betyder, at mange stokastiske partielle differentialligninger ikke har en løsning i klassisk forstand. To forskellige former for generaliserede løsninger der anvender henholdvis rumlige og stokastiske distributioner bliver præsenteret. Den anden af disse løsningsmetoder gør det muligt at give en matematisk stringent definition af hvid støj, og endvidere kan det såkaldte Wick produkt defineres. Nogle matematiske egenskaber ved disse objekter bliver præsenteret, og muligheden for at overføre denne stokastiske kalkyle til ikke-Gaussiske og Lévy processer bliver kort diskuteret. I forbindelse med model bygning og fortolkninger er det fundamentale spørgsmål hvorvidt Wick produktet skal anvendes eller ej. Der bliver argumenteret for at dette spørgsmål rent faktisk kan besvares udfra en præcis fortolkning af Wick produktet og det almindelige produkt. Dette bliver illustreret med to eksempler. Statistiske modeller bestående af spektral data for parabolske ligninger, som er beskrevet meget i litteraturen, bliver præsenteret. Endelig bliver de fire artikler inkluderet i afhandlingen beskrevet i forhold til den præsenterede teoridannelse og litteratur.

Observationen af en stokastisk partiel differentialligning i diskrete punkter i

tid og rum kan betragtes som en tidsrække. I den første artikel i afhandlingen, Markussen (2001*d*), bliver denne sammenhæng eksemplificeret for en parabolsk ligning. Hovedparten af artiklen omhandler de lokal asymptotiske egenskaber for likelihoodfunktionen givet observationen af en flerdimensional Gaussisk tidsrække. Specielt fås de velkendte asymptotiske egenskaber ved maksimum likelihood estimatoren. De krævede regularitets betingelser formuleres på spektraltæthederne i form af integrabilitet og glathed. Disse regularitets betingelser er meget generelle og indeholder specielt tidsrækker med lang hukommelse, d.v.s. spektraltætheder med en pol, og til en vis grad også spektraltætheder med nulpunkter. De asymptotiske egenskaber givet en følge af misspecificerede modeller, som indeholder den sande spektraltæthed i grænsen, bliver diskuteret. Hovedresultatet er en uniform version af den lokale asymptotiske normalitets egenskab kendt fra Le Cam teori. De anvendte matematiske teknikker er p-normer for matricer, L^p-normer for flerdimensionale spektraltætheder, og maksimums uligheder for stokastiske processer. Videre bliver de anvendte egenskaber ved Fejér approksimationer og Toeplitz matricer formuleret og bevist. I den eksisterende litteratur for tidsrækker bliver p-normerne og L^p-normerne sædvanligvis kun anvendt for $p = 1, 2, \infty$. Nyskabelsen i denne artikel består formodenlig i anvendelsen af disse normer for ethvert $p \in [1, \infty]$.

Den anden artikel i afhandlingen, Markussen (2001*b*), omhandler parametrisk inferens givet en observation i diskrete gitter punkter af en hyperbolsk stokastisk partiel differentialligning. En approksimativ state space model bliver foreslået, og den tilhørende likelihoodfunktion beregnes v.h.a. Kalman filteret. Videre bliver resultaterne fra Markussen (2001*d*) brugt til at give betingelser der sikre asymptotisk efficiens. Parabolske ligninger kan fås som grænsepunkter af hyperbolske ligninger, og den asymptotiske fordeling af kvotienttestet for en parabolsk ligning mod det hyperbolske alternativ bliver undersøgt. Endelig bliver glatheden af udfaldsfunktionerne undersøgt for parabolske henholdvis hyperbolske ligninger.

Den tredje artikel i afhandlingen, Markussen (2001*a*), omhandler kaos dekomponeringen af Hilbert rummet af kvadratisk integrabel funktionaler af den negative binomial proces. Kendskab til funktional analyse og i særdeleshed godt kendskab til Hilbert rums teori er formodenlig nødvendigt for at læse denne artikel. Strukturen af den fundne kaos dekomponering er forskellig fra den symmetrisk Fock rum struktur kendt fra Gaussiske og Poisson processer. Specielt kan alle kvadratisk integrabel funktionaler ikke genereres af de multiple integraler. Den fundne kaos dekomponering bliver anvendt til at definere Malliavin differential operatore svarende til variationale aflede m.h.t. henholdvis spring tidspunkterne og spring højderne. Selvom den anden af disse differential operatore ikke er tæt defineret, kan den dog defineres som en afsluttet operator på et rum indeholdende de multiple integraler. De indførte Malliavin operatore har ikke en diagonal repræsentation m.h.t. den fundne kaos dekomponering. Derimod kan det tilhørende Wick produkt let udtrykkes via kaos dekomponeringen. Den negative binomial proces er tænkt som et generisk eksempel på hvad der muligvis kan gøres for Lévy processer og i særdeleshed sammensatte Poisson processer. Den beskrevet kaos udvikling er direkte baseret på den karakteriske funktional for den negative binomial proces, og anvender således ikke identifikationen af en sammensat Poisson proces med en Poisson proces defineret på det kartesiske produkt af parameter rummet og spring rummet.

Hvis kaos udviklingen for en endelig dimensional stokastisk variabel er kendt, så kan den tilhørende Lebesgue tæthed udtrykkes som en middelværdi via delvis integrations metoder fra Malliavin kalkylen. I den fjerde artikel i afhandlingen, Markussen (2001*c*), bliver denne egenskab udnyttet til at foreslå en simuleringstilgang til beregningen af pseudo likelihoodfunktioner givet en observation i diskrete punkter af en stokastisk partiel differentialligning. Den foreslået metode bliver beskrevet for Gaussiske rum og kræver at kaos udviklingerne er kendte. Artiklen indeholder en kort introduktion til den anvendte Malliavin kalkyle, og simuleringen af de involverede Skorohod integraler bliver diskuteret. Den foreslået metode viser sig at være numerisk krævende. Specialtilfældet svarende til en observation med målefejl viser endvidere, at metoden også er numerisk instabil. Det bliver således konkluderet i artiklen, at den foreslået metode ikke har store praktiske anvendelses muligheder.

Table of Contents

1 Solving stochastic partial differential equations

In this essay, we will discuss some of the mathematical problems related to solving *Stochastic Partial Differential Equations*[1] with the intention of building probabilistic models for real life phenomena and doing statistical inference. We will mostly write the equations in an informal manner avoiding the precise mathematical definitions, however there will be some mathematics where it is needed to make the ideas clear. It is my hope, that this essay could be helpful to anyone interested in learning about the notion of SPDEs. This essay will also serve as an introduction to the Ph.D. thesis in which it is enclosed.

Many of the probabilistic models that are employed in the world today are known not to fit the data they are supposed to model. For instance, all the models from the successful field of mathematical finance based on *Stochastic Differential Equations* can not give an exact description of the prices on the stock exchange. This is due to the mathematical fact that the Brownian motion has infinite variation. Nevertheless, these models are sensible approximations to the real world phenomena, are mathematically nice and to a large extent also tractable, and therefore provide very useful models. One might even argue that the marginal distributions of SDEs actually could describe stock prices observed at discrete points in time. We would hope to find similar features for probabilistic models based on SPDEs. Consider for instance the stochastic wave equation,

$$\begin{cases} \partial_t^2 u(t,x) = \Delta_x u(t,x) + W(t,x), & t>0,\ x\in\mathbb{R}^d, \\ u(0,x) = \partial_t u(0,x) = 0, & x\in\mathbb{R}^d, \end{cases} \tag{1.1}$$

where $W(t,x)$ denote some kind of white noise process.[2] For $d=1$ equation (1.1) models the motion of one of the strings of a *"guitar carelessly left outdoors during a sandstorm", cf.* Walsh (1986, p. 281), and the solution is given by

$$u(t,x) = \tfrac{1}{2}\int_0^t \int_{x+s-t}^{x+t-s} W(s,y)\,\mathrm{d}y\,\mathrm{d}s \overset{\mathcal{D}}{=} \tfrac{1}{2}\hat{B}\big(\tfrac{t+x}{\sqrt{2}}, \tfrac{t-x}{\sqrt{2}}\big), \quad t>0,\ x\in\mathbb{R},$$

where $\hat{B}(t,x)$ is a Brownian sheet fixed to zero along the line $t+x=0$. For $d=2$ equation (1.1) would model the motion of the drum skin of a drum carelessly

[1] In the following abbreviated by SPDEs.

[2] The nomenclature $W(t,x)$ is for *white noise* and should not be confused with a *Wiener process*. The latter will be denoted by $B(t,x)$ for *Brownian motion*.

left outdoors during a sandstorm except for the fact that there do not exist an ordinary solution to (1.1) when $d \geq 2$. We thus need some notion of generalized solutions in order to solve this physical meaningful equation, and some procedure of regularizing these generalized solutions in order to have physical meaningful solutions and possibly conduct statistical inference. Indeed, we do not believe that the sand grains hit the guitar strings or the drum skin strictly according to a *singular white noise* process, but only that this mathematical model gives a good approximation of what is happening. It is thus acceptable to arrive at generalized solutions as long as there exist physically interpretable smoothing procedures to get real life solutions.

Three different notions of SPDEs

There exist at least three different answers to the question of how to pose and solve SPDEs. It should be emphasized that the question, which of these answers is the right one, is not a strictly mathematical question but a matter of choice and personal taste, and may also depend on the particular application. Of course each approach will have different properties and consequences which could be compared, but whether a given approach will prevail also depend on the strength of the advocates in favour of this particular approach. The first answer to the question of how to handle SPDEs is to blame the roughness of the white noise for the non-solvability of these equations. Indeed, if the white noise is replaced by *coloured noise*, *i.e.* some sort of smoothed white noise, then there exist ordinary solutions to many SPDEs. In the papers Manthey & Mittmann (1998), Manthey & Mittmann (1999) existence, uniqueness and stability properties of some stochastic partial functional-differential equations, *i.e.* SPDEs in time and space where the behaviour of the equation may depend on the entire past, are proved using this approach. Although the use of coloured noise directly gives physical interpretable solutions, and we thus avoid the smoothing of a potential generalized solution, this approach seems somewhat unaesthetic. Whereas white noise in some mathematical sense is a canonical object, this is not the case for coloured noise. The solution to the stochastic version of a given PDE will depending on the coloured noise employed, and this approach is thus most sensible in the cases where there are physical reasons for choosing a particular coloured noise process, *cf.* the discussion in Holden, Øksendal, Ubøe & Zhang (1996, p. 166). The second and the third answer to the question of how to pose and solve SPDEs gives canonical albeit generalized solutions to equations containing white noise. Let us consider SPDEs similar to (1.1), *i.e.* equations depending on a random element $\omega \in \Omega$, on time $t > 0$, and on a spatial component $x \in \mathbb{R}^d$. An ordinary solution to such an equation would be a measurable function

$$u : \Omega \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$$

solving the PDE itself or more likely an integral version of the PDE. Unfortunately such solutions rarely exist when $d \geq 2$ as already mentioned. The two different answers, which we will baptize the *Walsh approach* respectively the *Norwegian approach*[3], are given by considering the solution u as a function

$$u : \Omega \times \mathbb{R}_+ \to H^{-n}(\mathbb{R}^d) \quad \text{respectively} \quad u : \mathbb{R}_+ \times \mathbb{R}^d \to (\mathcal{S})_{-1}.$$

In the Walsh approach $H^{-n}(\mathbb{R}^d)$ is a *Sobolev space* of sufficiently high order n, *i.e.* the solutions are defined pointwise in $(\omega, t) \in \Omega \times \mathbb{R}_+$ as generalized functions in space. This approach is used in the monographs Walsh (1986), Kallianpur & Xiong (1995). In the Norwegian approach $(\mathcal{S})_{-1}$ is the less known *Kondratiev space* of stochastic distributions, *i.e.* the dual of some space of smooth random variables. In this approach, which is used in the monograph Holden et al. (1996), the solutions are defined pointwise in $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ as generalized random variables. In order to enhance the discussion let us consider some other PDEs.

Pressure equation with stochastic coefficients

Suppose we want to model the pressure $p(t, x)$ and the saturation $\theta(t, x)$ of a fluid injected into a porous rock. Assume that the point $x \in \mathbb{R}^3$ of the rock is either dry at time $t > 0$, *i.e.* $\theta(t, x) = 0$, or have complete saturation $\theta_0(x) > 0$. If $f(t, x)$ is the injection rate of the fluid at time t and the point x, and $k(x) \geq 0$ is the *permeability* of the rock at the point x, then the pressure $p(t, x)$ can be modeled by the following moving boundary problem, *cf.* Holden et al. (1996),

$$\begin{cases} \operatorname{div}_x\big(k(x)\nabla_x p(t,x)\big) = -f(t,x), & x \in D_t, \\ p(t,x) = 0, & x \in \partial D_t, \\ \theta_0(x)\,\frac{\mathrm{d}}{\mathrm{d}t}(\partial D_t) = -k(x)\nabla_x p(t,x), & x \in \partial D_t, \end{cases} \tag{1.2}$$

where D_t is *the wet region* at time t, *i.e.*

$$D_t = \big\{x \in \mathbb{R}^3 : \theta(t,x) = \theta_0(x)\big\}.$$

This apparently deterministic PDE exemplifies what I believe to be the most important applications of SPDEs, namely the cases where the coefficients of deterministic PDEs from physics, biology or other research fields are unknown and replaced by stochastic processes. For instance, the permeability of the porous rock is known to be non-negative, heterogeneous and isotropic, but the actual permeability $k(x)$ at the point x is likely to be unknown. It is thus natural to replace $k(x)$ by a stochastic process with the same qualitative properties known to be satisfied by this unknown function. Similarly, the injection rate $f(t, x)$ may be replaced by a stochastic process if necessary.

[3]These names refer to the monographs, *i.e.* Walsh (1986) respectively Holden et al. (1996), from which I learned the two different approaches.

Stochastic transport equation with Poisson noise

The concentration $U(t,x)$ at the time $t>0$ and at the point $x\in\mathbb{R}^d$ of a chemical substance dispersed in a moving medium can be modeled by the PDE

$$\begin{cases} \partial_t U(t,x) = \frac{1}{2}\sigma^2\Delta_x U(t,x) + V(t,x)\cdot\nabla_x U(t,x) \\ \qquad\qquad\qquad - K(t,x)\,U(t,x) + g(t,x), \quad t>0,\ x\in\mathbb{R}^d, \\ U(0,x) = f(x), \qquad\qquad\qquad\qquad\qquad x\in\mathbb{R}^d, \end{cases} \tag{1.3}$$

where $f(x)\geq 0$ is the initial concentration of the substance, $\frac{1}{2}\sigma^2>0$ is the dispersion coefficient, $V(t,x)\in\mathbb{R}^d$ is the velocity of the medium, $K(t,x)>0$ is the relative leakage rate, and $g(t,x)\in\mathbb{R}$ is the source rate of the substance. The case $d=1$ would correspond to a river, and $d=2$ would correspond to a lake. If some of the coefficients in (1.3) are unknown and replaced by stochastic processes we get a SPDE. For instance the transport of a chemical substance dispersed in a turbulent medium could be modeled by replacing the velocity coefficient $V(t,x)$ by a white noise process, *cf.* Holden et al. (1996, p. 146). The point we would like to stress by mentioning equation (1.3) however connects to the analysis done in Kallianpur & Xiong (1995, chapter 7), where equation (1.3) especially is used to model the pollution along a river. Kallianpur and Xiong assume that all the coefficients of (1.3) are known except the source rate $g(t,x)$ of the substance, which in their chapter 7.2 is replaced by a compound Poisson process. From a modeling perspective, this assumption of the probabilistic behavior of chemical deposits is sensible, and this model exemplifies the need for employing *e.g.* compound Poisson processes. In a more general setup but still within in same mathematical framework, it would be desirable to employ general *Lévy processes*.

1.1 The Walsh approach

In the Walsh approach SPDEs depending on time $t>0$ and a spatial component $x\in\mathcal{O}\subseteq\mathbb{R}^d$ are reformulated as Itô type SDEs with values in a Hilbert and conuclear space, *e.g.* a Sobolev space. We will thus also refer to these equations as Itô type SPDEs. Besides the references already mentioned this approach is also used in the monographs Rozovskii (1990), Da Prato & Zabczyk (1992). One of the main problems with this approach arises from the lack of a multiplicative structure of these Hilbert spaces. This means that it is difficult to handle SPDEs with multiplicative noise like the non-linear cable equation

$$\begin{cases} \partial_t u(t,x) = \Delta_x u(t,x) - u(t,x) + f\big(u(t,x),t\big)W(t,x), \quad t>0,\ x\in[0,L]^d, \\ \partial_x u(t,x) = 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad x\in\partial[0,L]^d, \\ u(0,x) = u_0(x), \qquad\qquad\qquad\qquad\qquad\qquad\quad x\in[0,L]^d, \end{cases} \tag{1.4}$$

cf. Walsh (1986, p. 312), and PDEs like (1.2) with stochastic coefficients. For some equations with multiplicative noise, *e.g.* equation (1.4), this difficulty can be resolved by the aid of contraction semigroups, *cf.* Kotelenez (1992), Kallianpur & Xiong (1995, chapter 4.3). However, a general theory to handle SPDEs with non-linearity or multiplicative structures is to my knowledge not available.

All papers on statistical inference for SPDEs I have seen essentially employes the Walsh approach to tame the roughness of these equations. Suppose for instance that the water temperature of the oceans is modeled by a SPDE, and we have temperature data consisting of satellite photos taken with an infrared camera at our disposal. Since the photos are taken from a high altitude, it is then natural to assume that the data actually are spatially smoothed water temperatures. A Walsh type SPDE thus is a mathematically nice model of generalized functions with a natural smoothing procedure to arrive at ordinary functions for the observations of the water temperatures.

1.1.1 Statistical inference based on spectral data

As an example of statistical estimation problems for SPDEs let us consider the analysis done in Huebner & Rozovskii (1995) in some detail. Let $\mathcal{O}$ be a smooth bounded domain in $\mathbb{R}^d$, and for a multi index $\gamma = (\gamma_1, \dots, \gamma_d) \in \mathbb{N}_0^d$ define

$$\mathrm{D}^\gamma f(x) = \frac{\mathrm{d}^{|\gamma|}}{\mathrm{d}x_1^{\gamma_1} \cdots \mathrm{d}x_d^{\gamma_d}} f(x),$$

where $|\gamma| = \gamma_1 + \ldots + \gamma_d$. Moreover let A_0 and A_1 be formally self-adjoint partial differential operators on $\mathcal{C}^\infty(\mathcal{O})$ of orders m_0 respectively m_1, and let $m = \max\{m_1, m_2\}$.[4] Huebner and Rozovskii investigates the asymptotic properties of the maximum likelihood estimator for a scalar parameter θ belonging to some compact subset Θ of $\mathbb{R}$, in the Dirichlet problem for the parabolic SPDE

$$\begin{cases} \mathrm{d}u(t,\cdot) = (A_0 + \theta A_1) u(t,\cdot)\,\mathrm{d}t + \mathrm{d}B(t,\cdot), & t > 0, \\ \mathrm{D}^\gamma u(t,x) = 0, \quad |\gamma| \le m-1, & t > 0,\ x \in \partial\mathcal{O}, \\ u(0,x) = u_0(x), & x \in \mathcal{O}. \end{cases} \tag{1.5}$$

Here $B(t,x)$ is a so-called *cylindrical Brownian motion, i.e.* $B(t,\cdot)$, $t > 0$, is a stochastic process with values in the Schwartz space of distributions $\mathcal{D}'(\mathcal{O})$ such that $\|\phi\|_{L^2(\mathcal{O})}^{-1} \langle B(t,\cdot), \phi \rangle_{\mathcal{D}'(\mathcal{O})}$ is a one dimensional Brownian motion for every $\phi \in \mathcal{C}^\infty(\mathcal{O})$, and such that

$$\mathrm{E}\left(\langle B(t_1,\cdot), \phi_1 \rangle_{\mathcal{D}'(\mathcal{O})} \langle B(t_2,\cdot), \phi_2 \rangle_{\mathcal{D}'(\mathcal{O})} \right) = (t_1 \wedge t_2) \langle \phi_1, \phi_2 \rangle_{L^2(\mathcal{O})}.$$

Solving the SPDE using spectral analysis

If the operators A_0 and A_1 commute, and the family $A_\theta = A_0 + \theta A_1$, $\theta \in \Theta$, is uniformly strongly elliptic of order m, then the operator A_θ with boundary conditions

[4] In Huebner & Rozovskii (1995) this maximum is denoted by $2m$, and some caution is thus required when comparing this essay and Huebner & Rozovskii (1995).

$D^\gamma u(t,x) = 0$, $|\gamma| \le m-1$, can be extended to a closed, self-adjoint operator $\mathcal{L}_\theta$ on $L^2(\mathcal{O})$. Moreover the operator $\mathcal{L}_\theta$ is lower semibounded, *i.e.* there exist a constant $k_\theta \in \mathbb{R}$ such that $k_\theta I - \mathcal{L}_\theta > 0$ and the resolvent $(k_\theta I - \mathcal{L}_\theta)^{-1}$ is compact, and the spectrum of the operator $(k_\theta I - \mathcal{L}_\theta)^{1/m}$ is a discrete set consisting of eigenvalues $\lambda_n(\theta)$, $n \in \mathbb{N}$, of finite multiplicity such that

$$0 < \lambda_1(\theta) \le \lambda_2(\theta) \le \lambda_3(\theta) \le \dots, \qquad \lambda_n(\theta) \xrightarrow[n\to\infty]{} \infty,$$

with associated eigenfunctions $h_n^\theta \in \mathrm{Dom}(\mathcal{L}_\theta) \cap \mathcal{C}^\infty(\overline{\mathcal{O}})$, $n \in \mathbb{N}$, constituting an orthonormal basis for $L^2(\mathcal{O})$. All this follows from the standard theory for deterministic PDEs and unbounded operators. Huebner and Rozovskii assume that the eigenfunctions $h_n = h_n^\theta$, $n \in \mathbb{N}$, do not depend on the parameter $\theta \in \Theta$, and introduce the Sobolev spaces

$$H_\theta^s = \Big\{ u \in L^2(\mathcal{O}) : \|u\|_{s,\theta} = \Big(\sum_{n=1}^\infty \lambda_n^{2s}(\theta) \left| \langle u, h_n \rangle_{L^2(\mathcal{O})} \right|^2 \Big)^{1/2} < \infty \Big\},$$

$s \ge 0$, with dual spaces H_θ^{-s}. The functions $h_{n,\theta}^s = \lambda_n^{-s}(\theta) h_n$, $n \in \mathbb{N}$, constitute an orthonormal basis for H_θ^{-s}, and for fixed $\alpha > d/2$ the cylindrical Brownian motion $B(t,x)$ has the expansion

$$B(t,\cdot) = \sum_{n=1}^\infty B_n(t) h_n \in H_\theta^{-\alpha},$$

where $B_n(t)$, $n \in \mathbb{N}$, are pairwise independent one dimensional Brownian motions. Using this spectral decomposition it follows that if $u_0 \in H_\theta^{-\alpha}$, then the solution $u^\theta(t,x)$ to the parabolic SPDE (1.5) is given by

$$u^\theta(t,\cdot) = \sum_{n=1}^\infty u_n^\theta(t) h_{n,\theta}^{-\alpha}, \tag{1.6}$$

where the coefficient processes $u_n^\theta(t)$, $n \in \mathbb{N}$, satisfies the SDEs

$$\begin{cases} \mathrm{d}u_n^\theta(t) = -\big(\lambda_n^m(\theta) - k_\theta\big) u_n^\theta(t)\,\mathrm{d}t + \lambda_n^{-\alpha}(\theta)\,\mathrm{d}B_n(t), \\ u_n^\theta(0) = \big\langle u_0, h_{n,\theta}^{-\alpha} \big\rangle_{H_\theta^{-\alpha}}, \end{cases} \tag{1.7}$$

i.e. u_n^θ, $n \in \mathbb{N}$, are pairwise independent *Ornstein-Uhlenbeck* processes.

Statistical inference using spectral data

After having solved the SPDE Huebner and Rozovskii arrive at the crucial crossroad in deciding, precisely which statistical problem is to be considered. They first decide on considering the estimation problem in an asymptotic framework. Beside the usual frameworks of asymptotics w.r.t. the time parameter respectively w.r.t. small noise, there is also the possibility of considering asymptotics w.r.t. the spatial parameter $x \in \mathcal{O}$. The latter framework is most promising of revealing new mathematical features, and from an application point of view this framework is also very realistic, *cf.* the application in *oceanography* discussed in Piterbarg & Rozovskii (1997).

Instead of considering the estimation problem given data observed at discrete points in the fixed, bounded spatial domain $\mathcal{O}$, Huebner and Rozovskii consider data consisting of finitely many coefficients in the spectral decomposition (1.6) used to solve the SPDE. The latter is certainly most easy since it allows an explicit expression for the maximum likelihood estimator $\hat{\theta}$. Moreover considering data consisting of observations of the first coefficient processes removes, but perhaps also blurs, the problem of the generalized nature of the solution to the considered SPDE when $d \geq 2$. I believe these are the reasons why most papers on this problem take the spectral approach. In Huebner & Rozovskii (1995) the *Le Cam theory* as presented in Ibragimov & Has'minskii (1981) and the general theory for the asymptotics of the eigenvalues for partial differential operators is used to give conditions for asymptotic normality and efficiency of the maximum likelihood estimator in the spectral framework. In Huebner (1997), Lototsky & Rozovskii (1999) similar analysis are done for multi dimensional parameters and without the assumption of commutativity of the involved partial differential operators. In Huebner & Lototsky (2000) a sieve estimator is introduced for the case where the parameter θ is time dependent. A common feature of these models is that the *Fisher information* grows *e.g.* like N^3, where N is the number of observed coefficient processes, and not linearly in N as in the i.i.d. situation, *cf.* the examples Huebner & Rozovskii (1995, p. 156). This feature is due to the fact that the coefficients in the Ornstein-Uhlenbeck processes (1.7) depends more heavily on the parameter θ for large n, and might not hold when data are sampled at discrete points in space. This non-standard rate of convergence of the maximum likelihood estimator thus may only hold due to a somewhat artificial sampling scheme.

Other references to the literature

I would also like to mention the following papers. In Loges (1984) a *Girsanov theorem*, which provides the basis for doing maximum likelihood estimation, is proved for SDEs with values in a Hilbert space. In the series of papers Ibragimov & Has'minskii (1998), Ibragimov & Has'minskii (1999), Ibragimov & Has'minskii (2000) the inverse problem for parabolic SPDEs, *i.e.* estimating the functional coefficients in the PDE, is studied under small noise asymptotics. The paper Huebner (1999) also studies small noise asymptotics for the maximum likelihood estimator, and in Mohapl (1998) estimation equations for the unknown parameter given an observation at discrete points in space of a linear Gaussian SPDE are derived and analysed. The paper by Mohapl is one of the few papers that discuss estimation based on spatially discrete observations. Further references can *e.g.* be found in the proceedings of the *Workshop on SPDEs* held 4–6 January 2001 at the Department of Statistics and Operations Research, University of Copenhagen, *cf.* Sørensen & Huebner (2001).

1.1.2 Review of paper I and paper II

In this subsection we review the papers Markussen (2001*d*), Markussen (2001*b*) enclosed in this thesis. The purpose of these papers is to study asymptotic likelihood inference when $n \to \infty$ for the statistical model given by the scalar parameters $\eta_1, \eta_2, \xi_0, \xi_1, \xi_2$ and an observation of the stationary solution to the hyperbolic SPDE

$$\begin{cases} \eta_2 \partial_t^2 u(t,x) + \eta_1 \partial_t u(t,x) = \xi_0 u(t,x) + \xi_1 \partial_x u(t,x) + \xi_2 \partial_x^2 u(t,x) \\ \qquad\qquad + e^{-\frac{\xi_1}{2\xi_2}x} W(t,x), \quad t \in \mathbb{R},\ x \in (0,1), \\ u(t,0) = u(t,1), \quad t \in \mathbb{R}, \end{cases} \tag{1.8}$$

in the lattice points (t,x) in time and space given by

$$t = \Delta, 2\Delta, \dots, n\Delta, \qquad\qquad x = \frac{a_1}{b}, \dots, \frac{a_N}{b},$$

where $\Delta > 0$ and $a_1, \dots, a_N, b \in \mathbb{N}$, $a_1 < \dots < a_N < b$, are fixed. The motivation for considering this estimation problem is to investigate the possible more realistic, but also more difficult, problem of spatial discrete date instead of spectral data. We have made the analysis easier in two ways. In order to avoid having observations from a smoothed version of a generalized solution to (1.8), and thus having to decide on the smoothing procedure, we only consider one dimensional space. Moreover we have chosen the problem of asymptotics for $n \to \infty$ and not for $N \to \infty$. In Cont (1998) the SPDE (1.8) with $\eta_2 = 0$, *i.e.* the parabolic limit of (1.8), is proposed as a model for the deviations of the interest rate for bounds with different maturity from the linear interpolation between the interest rate for the bound with the shortest respectively the longest time to maturity, and in Santa-Clara & Sornette (1999) the hyperbolic equation is proposed as a model in a somewhat different framework. A similar model for the term structure of interest rates is proposed in Brace, Gąterek & Musiela (1997). Observations for such models would be spatially sparse and temporally long, and would thus fit into the analysed asymptotic framework.

Linear time series models

The statistical model described above is a multivariate time series model, where the dimension of the time series equals the spatial resolution. Moreover using a spectral decomposition similar to (1.6) it is seen, that this model can be described as an infinite dimensional state space model. It is thus natural to use the Kalman filter for a finite dimensional approximation of this infinite dimensional model in order to compute an approximate likelihood, *i.e.* to use a slightly misspecified model in order to do computations. In Markussen (2001*d*) an uniform version of the *local asymptotic normality* property is proved for a sequence of misspecified multivariate Gaussian time series models, which contains the true model in the limit. The needed regularity conditions, which are formulated in terms of integrability and smoothness conditions on the spectral densities, are more general than the regularity conditions given in *e.g.* Dahlhaus (1989). Especially multivariate spectral densities with both a pole and a zero are included. The main mathematical tools employed in Markussen (2001*d*) are *Toeplitz* matrices, *p-Schatten* matrix norms, multivariate L^p-norms, *Orlicz* norms for random variables, cumulants for bilinear forms of Gaussian vectors, *chaining inequalities* for stochastic processes, and Le Cam theory. I believe the technical novelty in the paper consists in the employment of p-norms for every $p \in [1, \infty]$, and not only for $p = 1, 2, \infty$. Unfortunately, the paper appears rather technical. This is partly explained by the fact that I have tried to solve several problems at the same time, *i.e.* asymptotics for slightly misspecified multivariate time series, which was needed for the application to SPDEs, but also for *long memory* and *almost deterministic* time series.[5]

Statistical inference for a discretely observed SPDE

In Markussen (2001*b*) the results developed in Markussen (2001*d*) are used to prove consistency, asymptotic normality and asymptotic efficiency of the maximum likelihood estimator for the parameters $\eta_1, \eta_2, \xi_0, \xi_1, \xi_2$ based on the finite dimensional approximation of the SPDE model described above. The asymptotic distribution of the likelihood ratio test for a parabolic equation, *i.e.* $\eta_2 = 0$, against a hyperbolic equation, *i.e.* $\eta_2 > 0$, is found to be a truncated χ^2-distribution with one degree of freedom. Moreover, sample path properties of the solution to (1.8) are proved both in the parabolic case and the hyperbolic case. The results given in Markussen (2001*b*) can therefore be used to test the hypothesis of parabolicity against hyperbolicity, *cf.* the discussion in Cont (1998).

1.2 The Norwegian approach

In the Norwegian approach SPDEs are interpreted in the usual strong sense w.r.t. the time parameter t and the spatial parameter $x \in \mathbb{R}^d$, and the generalized nature of the solutions is transferred to the stochastic component. It is thus necessary to introduce spaces of generalized random variables. A particular nice feature of the Norwegian approach is that a singular white noise process, *e.g.* the time derivative of the Brownian motion, can be defined as a mathematical rigorous object. SDEs can thus be solved as actual differential equations in accordance with the way they usually are interpreted, and not only as integral equations. On the Kondratiev space of generalized random variables the so-called *Wick product* is defined. The Wick product is a product in the algebraic sense, and it is thus tempting to resolve the problems of multiplicative noise and non-linear equations by replacing all products with Wick products and all non-linear functions with their Wick counterparts. Although the Wick product as a mathematical object behaves extremely

[5]Time series which either have long memory or are almost deterministic have spectral densities with a pole respectively a zero. It was clear that the employed technique using general p-norms was strong enough to include time series with these singular properties, and the temptation to do so was simply too big.

nicely it should be used with caution, *cf.* the discussion in section 1.2.4 below. Another nice feature is that the stochastic part and the PDE part in some sense are separated, this is especially so for Wick type equations. In the subsequent sections I will describe the above mentioned notions, and relate the enclosed papers Markussen (2001*a*), Markussen (2001*c*) to the Norwegian approach.

1.2.1 Chaos decomposition, Kondratiev spaces, and the Wick product

In this section we will concentrate on the classical Gaussian case, but also discuss the potential extension to Lévy processes. Let T be a σ-compact, topological space equipped with a non-atomic positive *Radon measure*. For applications to SPDEs the space T could *e.g.* be the set of temporal-spatial parameters $\mathbb{R}_+ \times \mathbb{R}^d$ equipped with the Lebesgue measure. The idea is first to construct a Wiener process indexed by T, *i.e.* a Gaussian white noise or Brownian sheet depending on how the construction is done and interpreted, and describe the associated Hilbert space (L^2) of quadratic integrable random variables measurable w.r.t. this process. Then the standard Hilbert space theory is employed to construct a *Gel'fand triplet*

$$(\mathcal{S}) \hookrightarrow (L^2) \hookrightarrow (\mathcal{S})',$$

where $(\mathcal{S})$ is some space of *smooth random variables* and $(\mathcal{S})'$ is the dual space of *generalized random variables*.[6] Finally, if the constructions are done properly, then it is possible to introduce the Wick product on these spaces. There are of course different ways of implementing this program. We will present a synthesis of the methods used in Itô (1988), Holden et al. (1996).

White noise space

Let $\mathscr{S} = \mathscr{S}(T)$ be the Schwartz space of smooth, rapidly decreasing functions on T, and let $\mathscr{S}' = \mathscr{S}'(T)$ be the dual space of tempered distributions. By the *Bochner-Minlos theorem*, *cf.* Holden et al. (1996, p. 12), there exists a unique probability measure μ, called the *white noise probability measure*, on the Borel σ-algebra on $\mathscr{S}'$ with *characteristic functional*

$$C(\eta) = \int_{\mathscr{S}'} e^{\imath\langle\omega,\eta\rangle_{\mathscr{S}'}}\,\mathrm{d}\mu(\omega) = e^{-\frac{1}{2}\|\eta\|^2_{L^2(T)}}, \quad \eta \in \mathscr{S}. \tag{1.9}$$

The random variable $\langle\cdot,\eta\rangle_{\mathscr{S}'}$ defined on the probability space $(\mathscr{S}', \mathcal{B}(\mathscr{S}'), \mu)$ thus follows a Gaussian distribution with mean zero and variance $\|\eta\|^2_{L^2(T)}$, and can be interpreted as the stochastic integral w.r.t. a Brownian sheet defined on T, *i.e.*

$$\langle\cdot,\eta\rangle_{\mathscr{S}'} = \int_T \eta(t)\,\mathrm{d}B(t), \quad \eta \in \mathscr{S}.$$

[6]The theory of Schwartz distributions or generalized functions is constructed similarly w.r.t. the spatial parameter $x \in \mathbb{R}^d$.

Chaos decomposition

The nomenclature *chaos decomposition* is unfortunate and a bit misleading. A chaos decomposition is simply an orthonormal expansion in the Hilbert space $L^2(\mathscr{S}')$ of quadratic integrable functions defined on $(\mathscr{S}', \mathcal{B}(\mathscr{S}'), \mu)$, and although it might seem chaotic in the non-mathematical sense of the word, such a decomposition gives a very firm grip on the probability space under consideration. The classical *Wiener chaos decomposition*, *cf.* Wiener (1938), can be constructed via the following commuting diagram

$$\begin{tikzcd} L^2(\mathscr{S}') \arrow[r, "\mathcal{S}"] & \mathfrak{S}. \\ \bigoplus_{n\in\mathbb{N}} \hat{L}^2(T^n) \arrow[u, "\mathcal{V}=\oplus_{n\in\mathbb{N}} I_n"] \arrow[ur, "\mathcal{U}"'] & \end{tikzcd} \tag{1.10}$$

The horizontal part of the diagram (1.10) consists of the so-called $\mathcal{S}$-transform from $L^2(\mathscr{S}')$ to a Hilbert space $\mathfrak{S}$ of non-linear, complex valued functionals defined on $\mathscr{S}$, *cf.* Hida, Kuo, Potthoff & Streit (1993). The $\mathcal{S}$-transform is the linear operator defined by normalizing the *Fourier transform* with the characteristic functional, [7] *i.e.*

$$(\mathcal{S}\phi)(\xi) = C(\xi)^{-1} \int_{\mathscr{S}'} e^{\imath\langle\omega,\xi\rangle_{\mathscr{S}'}} \phi(\omega)\,\mathrm{d}\mu(\omega), \quad \phi \in L^2(\mathscr{S}'),\ \xi \in \mathscr{S}.$$

The key to the analysis is, that $\mathfrak{S}$ is a *reproducing kernel Hilbert space*, *cf.* Aronszajn (1950), with reproducing kernel

$$K(\zeta, \eta) = C(\zeta)^{-1} C(\zeta - \overline{\eta}) C(-\overline{\eta})^{-1}. \tag{1.11}$$

The non-trivial part of this analysis is to find the Hilbert space in the lower left corner of (1.10) together with the isomorphism $\mathcal{U}$, and to describe the isomorphism $\mathcal{V} = \mathcal{S}^{-1} \circ \mathcal{U}$. In the Gaussian case the well known isomorphism $\mathcal{V}$ from the quadratic integrable, symmetric functions to $L^2(\mathscr{S}')$ via the multiple Wiener-integrals, *cf.* Itô (1951), is recovered.

Let $\mathcal{I} = (\mathbb{N}_0^{\mathbb{N}})_c$ be the set of non-negative integer sequences $a = (a_n)_{n\in\mathbb{N}}$ with only finitely many non-zero coordinates. If the functions $\xi_n \in \mathscr{S}$, $n \in \mathbb{N}$, constitute an orthonormal basis for $L^2(T)$, then an orthonormal basis for $\oplus_{n\in\mathbb{N}} \hat{L}^2(T)$ can be constructed via the tensor structure, and hence be mapped to an orthonormal basis for $L^2(\mathscr{S}')$. The resulting basis for $L^2(\mathscr{S}')$ consists of the *polynomial functionals* Φ_a, $a \in \mathcal{I}$, given by

$$\Phi_a = \prod_{n\in\mathbb{N}} (a_n!)^{-\frac{1}{2}} H_{a_n}\big(\langle \cdot, \xi_n \rangle_{\mathscr{S}'}\big), \quad a = (a_n)_{n\in\mathbb{N}} \in \mathcal{I},$$

[7] The $\mathcal{S}$-transform can also be defined via translation on the ω-space, *i.e.* by $(\mathcal{S}\phi)(\xi) = \mathrm{E}_\mu\, \phi(\cdot + \xi)$, and thus connect to *Gâteaux derivatives* and *Malliavin calculus*, *cf.* Hida et al. (1993, chapter 5).

where $H_n(x)$ denotes the n'th *Hermite polynomial*[8] defined by $H_0(x) = 1$ and

$$H_n(x) = (-1)^n \, e^{\frac{x^2}{2}} \, \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}} = \sum_{k=0}^{[\frac{n}{2}]} (-\tfrac{1}{2})^k \frac{n!}{k!(n-2k)!} \, x^{n-2k}, \quad n \in \mathbb{N}.$$

Stochastic distributions

Having described the chaos decomposition we can proceed to define the Kondratiev space of stochastic distributions as done in Holden et al. (1996, chapter 2.3). For $\rho \in [0,1]$ the Kondratiev spaces $(\mathcal{S})_\rho$ and $(\mathcal{S})_{-\rho}$ of stochastic test functions respectively stochastic distributions are defined by

$$(\mathcal{S})_\rho = \Big\{ \phi = \sum\nolimits_{a \in \mathcal{I}} c_a \Phi_a \;\Big|\; \|\phi\|_{\rho,k} < \infty \text{ for all } k \in \mathbb{N} \Big\},$$
$$(\mathcal{S})_{-\rho} = \Big\{ \phi = \sum\nolimits_{a \in \mathcal{I}} c_a \Phi_a \;\Big|\; \|\phi\|_{-\rho,-q} < \infty \text{ for some } q \in \mathbb{N} \Big\},$$

where the norms $\|\phi\|_{\rho,k}$, $\|\phi\|_{-\rho,-q}$ for $\phi = \sum_{a \in \mathcal{I}} c_a \Phi_a$ are defined by

$$\|\phi\|^2_{\rho,k} = \sum_{a \in \mathcal{I}} c_a^2 \prod_{n \in \mathbb{N}} (a_n!)^\rho \, (2n)^{k\, a_n}, \qquad \|\phi\|_{-\rho,-q} = \sum_{a \in \mathcal{I}} c_a^2 \prod_{n \in \mathbb{N}} (a_n!)^{-\rho} \, (2n)^{-q\, a_n}.$$

Whereas the elements of $(\mathcal{S})_\rho$ converge in $L^2(\mathscr{S}')$, the elements of $(\mathcal{S})_{-\rho}$ are only defined as formal sums. The space $(\mathcal{S})_{-\rho}$ can be interpreted as the dual space of $(\mathcal{S})_\rho$ via the duality

$$\langle \psi, \phi \rangle_{(\mathcal{S})_{-\rho}} = \sum_{a \in \mathcal{I}} c_a^\psi \cdot c_a^\phi,$$

given $\psi = \sum_{a \in \mathcal{I}} c_a^\psi \, \Phi_a \in (\mathcal{S})_{-\rho}$, $\phi = \sum_{a \in \mathcal{I}} c_a^\phi \, \Phi_a \in (\mathcal{S})_\rho$, and we have the following *Sobolev scale* of smooth respectively generalized random variables,

$$(\mathcal{S})_1 \subset (\mathcal{S})_\rho \subset (\mathcal{S})_0 \subset L^2(\mathscr{S}) \subset (\mathcal{S})_{-0} \subset (\mathcal{S})_{-\rho} \subset (\mathcal{S})_{-1}.$$

There are quite a few constructions of Gel'fand triplets for the Hilbert space $L^2(\mathscr{S}')$ in the literature. Especially the *Hida spaces* of test functions and distributions, *cf.* Hida et al. (1993, chapter 4), are widely used. See also Watanabe (1987) for the definition of other Gel'fand triplets and their application in Malliavin calculus. I have not been explicitly concerned with these spaces in my thesis, and will therefore not dwell further on this issue. However, it should be mentioned that the space $(\mathcal{S})_{-0}$ contains white noise. If the Brownian sheet $B(t)$ is constructed as the $L^2(\mathscr{S}')$-limit

$$B(t) = \big\langle \cdot, \chi_{[0,t]} \big\rangle_{\mathscr{S}'} = \sum_{n \in \mathbb{N}} \big\langle \chi_{[0,t]}, \xi_n \big\rangle_{L^2(T)} \, \Phi_{\varepsilon_n} \in L^2(\mathscr{S}'),$$

[8]There exist different definitions of the Hermite polynomials in the litterature. We follow the definition used in Holden et al. (1996).

where χ_A is the indicator function for the set A, and ε_n is the multi index which has a one on the n'th coordinate and zeros elsewhere, then the white noise process $W(t)$ can be defined via formal differentiation w.r.t. the parameter $t \in T$,[9] *i.e.*

$$W(t) = \frac{\mathrm{d}}{\mathrm{d}t} B(t) = \sum_{n \in \mathbb{N}} \frac{\mathrm{d}}{\mathrm{d}t} \left\langle \chi_{[0,t]}, \xi_n \right\rangle_{L^2(T)} \Phi_{\varepsilon_n} = \sum_{n \in \mathbb{N}} \xi_n(t)\, \Phi_{\varepsilon_n} \in (\mathcal{S})_{-0}. \tag{1.12}$$

Wick product

For each $\rho \in [0,1]$ the Kondratiev spaces constitutes a Gel'fand triplet for the Hilbert space $L^2(\mathscr{S}')$. Similarly, Gel'fand triplets for the two other corners of the diagram (1.10) can be constructed and the isomorphisms $\mathcal{S}$, $\mathcal{U}$, $\mathcal{V}$ be extended. In first constructing the Gel'fand triplet for the upper left corner of (1.10) we have in fact departed from the methodical correct way of proceeding. As basis for the construction lies the parameter space T and the characteristic functional $C(\eta)$. Given these objects the right corner of (1.10) is defined via the known reproducing kernel (1.11), and the lower left corner and the isomorphism $\mathcal{U}$ should be found. Thereafter, the lower left corner of the diagram hopefully can be constructed via L^2-spaces over T, and if so, then Sobolev spaces for these L^2-spaces can be constructed and mapped to Sobolev spaces for $L^2(\mathscr{S}')$ via the isomorphism $\mathcal{V} = \mathcal{S}^{-1} \circ \mathcal{U}$, *cf.* the analysis done in Itô & Kubo (1988). Suppose such a construction has been done. Since the extended isomorphism $\mathcal{S}$ is injective, the Wick product $\phi \diamond \psi$ of the generalized random variables ϕ and ψ, and the Wick version $f^\diamond$ of an analytic function f can then be defined by the equations

$$\mathcal{S}(\phi \diamond \psi) = (\mathcal{S}\phi) \cdot (\mathcal{S}\psi), \qquad \mathcal{S} f^\diamond(\phi) = f \circ (\mathcal{S}\phi). \tag{1.13}$$

In the Gaussian case it is possible to show, that the Wick product satisfies the following formal rule of multiplication,

$$\Big(\sum_{a \in \mathcal{I}} c_a^\psi\, \Phi_a\Big) \diamond \Big(\sum_{a \in \mathcal{I}} c_a^\phi\, \Phi_a\Big) = \sum_{a \in \mathcal{I}} \Bigg(\sum_{a^\psi + a^\phi = a} \Big(\prod_{n \in \mathbb{N}} \frac{a_n!}{a_n^\psi!\, a_n^\phi!} \Big)^{\frac{1}{2}} c_{a^\psi}^\psi\, c_{a^\phi}^\phi \Bigg) \Phi_a, \tag{1.14}$$

and that the Kondratiev spaces $(\mathcal{S})_1$ and $(\mathcal{S})_{-1}$ are closed w.r.t. Wick multiplication, *cf.* Holden et al. (1996, chapter 2.4). It easily follows from the defining equation (1.13), that the Wick product satisfy the usual algebraic properties of a product. However due to the functional analytic definition, the Wick product is very difficult to grasp on an intuitive level. In section 1.2.3 we will discuss some more properties of the Wick product.

Extension to Lévy processes

The methodology described above should in principle also work for general Lévy processes. All that has to be done, is to replace the Gaussian characteristic func-

[9] This derivation can be made precise in several ways, *e.g.* $W(t) = \frac{\mathrm{d}}{\mathrm{d}t} B(t)$ w.r.t. the topology on $(\mathcal{S})_{-1}$ described in Holden et al. (1996, chapter 2.8).

tional (1.9) with the characteristic functional for the Lévy process under consideration and complete the diagram (1.10). This program have been implemented for the Poisson process in Itô (1988), Itô & Kubo (1988). In paper III of this thesis, Markussen (2001*a*), the diagram corresponding to (1.10) is found for the compound Poisson process with logarithmic distributed jumps, *i.e.* the Lévy process with negative binomial distributed marginals. This process is meant as a generic example of a Lévy process, and was chosen for two reasons. Firstly, the negative binomial process is a compound Poisson process with bounded intensity and integer valued jump heights and thus a rather benign Lévy process. Secondly, the orthogonal polynomials w.r.t. the negative binomial distribution, *i.e.* the Meixner polynomials, are well known. The analysis of the negative binomial process was thus likely to be successful, which it indeed also were. However, the Meixner polynomials turned out not to play any role in the presented analysis. The proofs in Markussen (2001*a*) probably easily generalize to any given compound Poisson process. Three things obviously remain to be done in the future. Firstly, the chaos decomposition should be found for general Lévy processes, especially in the mathematical interesting case where the Lévy measure has infinite mass. Secondly, the corresponding Sobolev spaces and Wick product should be found. Thirdly, the resulting Wick calculus should be investigated.

I would like to mention the following papers. Chaos decomposition and Malliavin calculus for Poisson processes is studied in Privault (1994), Privault (1996), Dermoune (1995). A compound Poisson processes indexed by the set T and with jumps belonging to the set $N \subseteq \mathbb{R}$ is in a natural correspondence with a Poisson process indexed by the set $T \times N$. This correspondence is used in Lytvynov, Rebenko & Shchepan'uk (1997), Denis (2000) to discuss stochastic calculus for compound Poisson processes. In Kondratiev, Silva & Streit (1996) the biorthogonal *Appell systems* is used as the basis for non-Gaussian probability spaces, and the resulting stochastic calculus is investigated. For a general reference on Lévy processes see Sato (1999).

1.2.2 Solving Wick type SPDEs

Let us consider the stochastic pressure equation (1.2) once more. The Wick version of this equation arises by replacing the product with a Wick product. For a fixed instant of time we thus get the equation

$$\begin{cases} \operatorname{div}\big(k(x) \diamond \nabla p(x)\big) = -f(x), & x \in D, \\ \qquad\qquad\qquad p(x) = 0, & x \in \partial D, \end{cases} \tag{1.15}$$

where $D \subset \mathbb{R}^d$ is the wet region, and where the unknown permeability $k(x)$ is replaced by the *singular positive noise process*,

$$k(x) = \exp^{\diamond}\big(W(x)\big) \in (\mathcal{S})_{-1}, \tag{1.16}$$

cf. Holden et al. (1996, chapter 4.6). The pressure $p(x)$, the source rate $f(x)$, and the permeability $k(x)$ are stochastic distributions, whence there exists coefficient functions $c_a^p, c_a^f, c_a^k : \mathbb{R}^d \to \mathbb{R}$, $a \in \mathcal{I}$, such that

$$p(x) = \sum_{a\in\mathcal{I}} c_a^p(x)\,\Phi_a, \qquad f(x) = \sum_{a\in\mathcal{I}} c_a^f(x)\,\Phi_a, \qquad k(x) = \sum_{a\in\mathcal{I}} c_a^k(x)\,\Phi_a. \tag{1.17}$$

The coefficient functions $c_a^k(x)$ are differentiable w.r.t. $x \in \mathbb{R}^d$. Suppose the coefficient functions $c_a^f(x)$ are Hölder continuous. If the chaos expansions (1.17) are inserted in the SPDE (1.15), the partial derivations and the sum over $a \in \mathcal{I}$ are interchanged, the formal rule (1.14) for the Wick product is used, and the coefficients of the resulting chaos expansions are compared, then we get an infinite dimensional system of coupled PDEs[10] in the functions $c_a^p(x)$, $a \in \mathcal{I}$. Moreover it can be shown[11], *cf.* Holden et al. (1996, theorem 4.6.3), that there exist a unique solution $(c_a^p)_{a\in\mathcal{I}}$ to this system of PDEs, and that the stochastic distribution process $p(x) = \sum_{a\in\mathcal{I}} c_a^p(x)$ is the unique $(\mathcal{S})_{-1}$-valued solution to (1.15).

Two reasons to use the Wick product

The previous example illustrates two favourable properties of the Wick product. Firstly, many multiplicative or non-linear SPDEs are well-defined in their Wick version since $(\mathcal{S})_{-1}$ is closed w.r.t. Wick multiplication. Secondly, Wick type SPDEs are in principle easy to solve. The second property is due to the fact that a Wick type SPDE can be separated into a stochastic part and a analytic part, and hence posed as a system of deterministic PDEs in a straightforward manor. This system of PDEs can then *e.g.* be solved numerically, *cf.* Theting (2000).

This technique does not work for Itô type SPDEs. For the sake of argument suppose the pressure equation (1.15) was well-defined without the Wick product but still with the permeability $k(x)$ given by (1.16). If this SPDE were to be transformed into a system of deterministic PDEs we would use the usual product instead of the Wick product after the partial derivations and the sum over $a \in \mathcal{I}$ were interchanged. But comparing the usual product of two basis random variables Φ_a and Φ_b,

$$\Phi_a\,\Phi_b = \sum_{c\in\mathcal{I}:c_n\le a_n\wedge b_n} \left(\prod_{n\in\mathbb{N}} \frac{\sqrt{a_n!b_n!(a_n+b_n-2c_n)!}}{c_n!(a_n-c_n)!(b_n-c_n)!}\right)\Phi_{a+b-2c}, \qquad a,b\in\mathcal{I},$$

with the Wick counterpart,

$$\Phi_a \diamond \Phi_b = \left(\prod_{n\in\mathbb{N}} \frac{(a_n+b_n)!}{a_n!\,b_n!}\right)^{\frac{1}{2}} \Phi_{a+b}, \qquad a,b\in\mathcal{I},$$

[10] Usually this system will not be entangled, and can be solved recursively w.r.t. the multi indices $a \in \mathcal{I}$ of the coefficient functions.

[11] Technically this is done via the so-called *Hermite transform*, *cf.* Holden et al. (1996, chapter 2.6), which converts stochastic distributions into holomorphic functions of infinitely many variables.

it is clear that the resulting system of PDEs would be very entangled. Moreover, using formal calculations for the usual product can easily lead to undefined objects, *e.g.* the formal square of singular white noise (1.12) is given by

$$W(t)^2 = \sum_{n,m\in\mathbb{N}} \xi_n(t)\,\xi_m(t)\,\Phi_{\varepsilon_n}\Phi_{\varepsilon_m} = \underbrace{\sum_{n,m\in\mathbb{N}} \xi_n(t)\,\xi_m(t)\,\Phi_{\varepsilon_n}\diamond\Phi_{\varepsilon_m}}_{=W(t)^{\diamond 2}\in(\mathcal{S})_{-0}} + \underbrace{\sum_{n\in\mathbb{N}} \xi_n(t)^2}_{=\infty}$$

and hence not well-defined.

1.2.3 Properties of the Wick product

The analytic properties of the Wick product are reviewed in this subsection. These properties will qualify the Wick product to be interpreted as a *renormalized* product, *cf.* the terminology from physics, or as a *compensated* product.

Connection to the ordinary product

Let $X, Y \in L^2(\mathscr{S}')$ be two quadratic integrable random variables. The ordinary product $X \cdot Y \in L^1(\mathscr{S}')$ is defined pointwise w.r.t. the ω-space, *i.e.*

$$(X\cdot Y)(\omega) = X(\omega)\cdot Y(\omega), \quad \omega\in\mathscr{S}',$$

whence we also will refer to the ordinary product as the *pointwise product*. The Wick product $X\diamond Y$ is not necessarily an ordinary random variable, but is in general a Hida distribution. Moreover, the Wick product $X\diamond Y$ is *non-local* in the ω-space, *i.e.* there are in general not relations between $(X\diamond Y)(\omega)$ and the pair $X(\omega)$, $Y(\omega)$ for given fixed $\omega\in\mathscr{S}'$. I believe that this non-locality is the main culprit for the difficulty in interpreting the Wick product. The null-rule thus only holds for the Wick product in the weak sense that $X\diamond Y = 0$ if either X or Y equals zero as elements in $L^2(\mathscr{S}')$. However, the pointwise product and the Wick product coincides if the factors are *analytically independent*, *i.e.* if there exists a partition $\{A, B\}$ of $\mathscr{S}'$ such that the functions X and Y are constant on the set B respectively on the set A. Especially $X\diamond Y = X\cdot Y$ if one of the factors is deterministic. Analytic independence implies stochastic independence, but is a much stronger requirement.

Generalized expectation and commutativity properties

The notion of expectation can be generalized to Kondratiev distributions via the formal calculation

$$\mathrm{E}_\mu X = \sum_{a\in\mathcal{I}} c_a^X\,\mathrm{E}_\mu\,\Phi_a = c_0^X = (\mathcal{S}X)(0), \quad X = \sum_{a\in\mathcal{I}} c_a^X\,\Phi_a \in (\mathcal{S})_{-1},$$

where $\Phi_0 = 1$ and $\mathrm{E}_\mu \Phi_a = 0$ for every $a \in \mathcal{I} \setminus \{0\}$ have been used. It is easily seen that the generalized expectation is indeed a generalization of the ordinary expectation. Given $X, Y \in (\mathcal{S})_{-1}$ and an analytic function f this definition yields,

$$\begin{aligned} \mathrm{E}_\mu \big[X \diamond Y\big] &= \big(\mathcal{S}(X \diamond Y)\big)(0) = (\mathcal{S}X)(0) \cdot (\mathcal{S}Y)(0) = \mathrm{E}_\mu X \cdot \mathrm{E}_\mu Y, \\ \mathrm{E}_\mu f^\diamond(X) &= \big(\mathcal{S}f^\diamond(X)\big)(0) = f\big((\mathcal{S}X)(0)\big) = f(\mathrm{E}_\mu X), \end{aligned}$$

i.e. the Wick product and Wick versions of analytic functions commutes with the expectation operator. Moreover, if a generalized notion of martingale is introduced, then it can be shown, that the Wick product of two generalized martingales is again a generalized martingale, *cf.* Benth & Potthoff (1996).

Connection to Skorohod calculus and commutativity properties

An important property of the Wick calculus is the connection to *Skorohod calculus*. If the stochastic process $u(t) \in L^2(\mathscr{S}')$, $t \in \mathbb{R}$, is Skorohod integrable,[12] then

$$\int_{\mathbb{R}} u(t)\, \delta B(t) = \int_{\mathbb{R}} u(t) \diamond W(t)\, \mathrm{d}t, \tag{1.18}$$

where $\delta(u) = \int u(t)\, \delta B(t)$ denotes the Skorohod integral, and $W(t) = \frac{\mathrm{d}}{\mathrm{d}t} B(t)$ is white noise. Moreover, Itô calculus with ordinary multiplication is equivalent to ordinary calculus with Wick multiplication, *cf.* Holden et al. (1996, chapter 2.5). If $F \in L^2(\mathscr{S}')$, $u(t) \in L^2(\mathscr{S}')$, $t \in \mathbb{R}$, the stochastic processes u, $F \cdot u$, $F \diamond u$ all are Skorohod integrable, and F is Malliavin differentiable with derivative $\mathrm{D}\, F = (\mathrm{D}_t\, F)_{t \in T}$, then

$$\delta(F \cdot u) = F\, \delta(u) - \int_{\mathbb{R}} \mathrm{D}_t\, F\, u(t)\, \mathrm{d}t, \qquad \delta(F \diamond u) = F \diamond \delta(u),$$

cf. Nualart (1995, p. 40) respectively Holden et al. (1996, corollary 2.5.12), *i.e.* the Wick product also commute with the Skorohod integral.

Preservation of positivity

A Kondratiev distribution $X \in (\mathcal{S})_{-1}$ is called *positive* if $\langle X, \phi \rangle_{(\mathcal{S})_{-1}} \geq 0$ for test functions $\phi \in (\mathcal{S})_1$ that are positive almost surely. This concept naturally generalize the notion of positivity, and to some extent the Wick product of two positive Kondratiev distributions is again positive, *cf.* Holden et al. (1996, corollary 2.11.7). See also Benth (1997), where the problem of positivity of the stochastic heat equation is analysed.

[12] The Skorohod integral is an extension of the classical Itô integral to non-adapted integrands.

1.2.4 To Wick or not to Wick

The basic question is now whether the Wick product should be used instead of the ordinary product in stochastic equations, or rephrasing *Shakespeare*,

To Wick, or not to Wick, – that is the question.[13]

We have already listed several of the properties of the Wick product, which is preferable from a mathematical point since it behaves nicely in almost every sense. But whether Wick type equations gives the right mathematical models for real life phenomena is another question. This question is also addressed in Holden et al. (1996), where two different SDEs are compared in their ordinary respectively Wick version. In this section we will comment on these examples and argue that it is possible to choose between the ordinary and the Wick versions based on careful considerations of the associated interpretations.

Revisiting the stochastic pressure equation

In the first example, the stochastic pressure equation at an instant of time is considered for a one-dimensional model. Thus imagine a long, thin, porous cylinder filled with a fluid. Suppose that the pressure and the outward flux of the fluid at the left end point is equal to zero respectively equal to $a > 0$, and that no fluid is injected into the cylinder. If the unknown permeability $k(x)$, $x \geq 0$, is replaced by *smoothed positive noise, i.e.*

$$k(x) = \exp^{\diamond}\big(W_\eta(x)\big) = \exp\big(W_\eta(x) - \tfrac{1}{2}\,\|\eta\|^2_{L^2(\mathbb{R})}\big), \tag{1.19}$$

where $W_\eta(x) = \langle \cdot, \eta(\cdot - x)\rangle_{\mathscr{S}'}$, $\eta \in \mathscr{S}(\mathbb{R})$, is smoothed white noise, then the pressure $p(x)$ at the point $x > 0$ can be modeled by the SDE

$$\tfrac{\mathrm{d}}{\mathrm{d}x}\big(k(x)\,\tfrac{\mathrm{d}}{\mathrm{d}x}p(x)\big) = 0,\ x > 0, \qquad p(0) = 0, \qquad k(0)\,p'(0) = a. \tag{1.20}$$

The question is whether the quantity $k(x)\,\frac{\mathrm{d}}{\mathrm{d}x}p(x)$ should be interpreted as the ordinary product $k(x) \cdot \frac{\mathrm{d}}{\mathrm{d}x}p(x)$ or as the Wick product $k(x) \diamond \frac{\mathrm{d}}{\mathrm{d}x}p(x)$. The pressure equation can be solved pointwise in $\omega \in \mathscr{S}'$ in both cases, *cf.* Holden et al. (1996, chapter 3.5), and the mean behaviors of the solutions are given by

$$\mathrm{E}_\mu\big[p_{\text{ordinary}}(x)\big] = a\,x \cdot \exp\big(\,\|\eta\|^2_{L^2(\mathbb{R})}\big), \qquad \mathrm{E}_\mu\big[p_{\text{Wick}}(x)\big] = a\,x.$$

The mean behavior of the Wick type solution solve the equation (1.20) with the stochastic coefficient $k(x)$ replaced by its mean value $\mathrm{E}_\mu\,k(x) = \exp(\mathrm{E}_\mu\,W_\eta(x)) = 1$. This common feature of Wick type equations is due to the commutativity properties of the Wick product. The mean behavior of $p_{\text{ordinary}}(x)$ contain the factor $\exp(\|\eta\|_2^2)$, which can be interpreted as the effect of a non-constant permeability. In

[13] William Shakespeare, 1564–1616. *Hamlet*'s monologue, act 3, scene 1: *"To be, or not to be, – that is the question".*

order to answer the question of which model should be preferred the following observation can be made. The unknown permeability was modelled by the stochastic process (1.19) because of some qualitative characteristics of this process, and we do not believe this actually gives a precise description of the microscopic features of the cylinder. The additional factor $\exp(\|\eta\|_2^2)$ exactly depend on the microscopic behavior of the process (1.19), and is thus undesirable from a modeling point of view, whence the Wick type equation is to be preferred. Moreover, the Wick type equation has better stability and variance properties as demonstrated in Holden et al. (1996).

Population growth in a stochastic, crowded environment

The second example consists of stochastic versions of the classical *Verhulst* model for population growth in a crowded environment. Thus let $X(t)$ denote the size at time $t > 0$ of a population living in a environment with carrying capacity K. Assume that the relative growth rate is proportional to the free life space $K - X(t)$ with proportionality factor given by $r + a\,W(t)$, *i.e.* the white noise models unpredictable, irregular changes in the environment. If the population size at time zero equals x_0, then $X(t)$ can be modeled by the SDE

$$X(0) = x_0, \qquad \frac{\mathrm{d}X(t)}{\mathrm{d}t} = X(t)\,\big(K - X(t)\big) \diamond \big(r + a\,W(t)\big). \tag{1.21}$$

This equation is a Itô type SDE, whence the product of the factors $X(t)\,(K - X(t))$ and $r + a\,W(t)$ should be a Wick product, *cf.* the connection between Wick calculus and Skorohod integration given in equation (1.18). The question is whether the factor $X(t)\,(K - X(t))$ should be interpreted as the ordinary pointwise product $X(t) \cdot (K - X(t))$ or as the Wick product $X(t) \diamond (K - X(t))$. Holden et al. (1996, p. 111) states,

> *We emphasize that there is no reason to assume a priori that the pointwise product $X(t) \cdot (K - X(t))$ is better than the Wick product $X(t) \diamond (K - X(t))$.*[14]

This is however not the case. In the Verhulst model the factor $X(t)(K - X(t))$ is constructed in order to model two features. The growth should be proportional to the population size as well as to the free life space. This should also be the case in a stochastic environment, whence the pointwise product *a priori* is the right interpretation. If the Wick product is used instead, then the interpretation of the Verhulst model disappears completely, and the resulting model indeed features a very strange behavior. As observed in Holden et al. (1996, p. 117) the solution $X_{\text{Wick}}(t)$ to the Wick version of (1.21) with $0 < x_0 < K$ may overshoot the carrying capacity K,[15] whence the stochastic process $X_{\text{Wick}}(t)$ also must be

[14]The notation have been changed slightly.
[15]This feature is illustrated by some simulations, *cf.* Holden et al. (1996, p. 116).

non-Markovian.[16] These features are not unreasonable a priori. However, the solution $X_{\text{Wick}}(t)$ may also undershoot the lower level of a population size equal to zero, *i.e.* there is positive probability of a negative population size. This directly disqualifying feature of the Wick interpretation is due to the loss of the Verhulst interpretation, *i.e.* $X(\omega_0, t_0) = 0$ does not imply $X(\omega_0, t_0) \diamond (K - X(\omega_0, t_0)) = 0$.

1.2.5 Statistical inference for Wick type SPDEs

In this section we give a brief review of paper IV of this thesis, Markussen (2001*c*). This paper contains a suggestion on how to simulate an approximative pseudo-likelihood given an observation in discrete points in time and space of a parameterized Wick type SPDE model. The proposed procedure to compute the pseudo-likelihood for the model parameter $\theta \in \Theta$ is as follows. The coefficient functions in the chaos expansion of the SPDE corresponding to the parameter θ can be found recursively by solving deterministic PDEs, *e.g.* numerically, as explained in section 1.2.2. In order to proceed further only finitely many of the coefficient functions are determined, and the used SPDE model is approximated by the corresponding truncation of the chaos expansion. This of course introduces bias, and the approximation should somehow take into account the particular features of the model. The approximative model now consists of an observation of finitely many random variables with finite chaos expansions, whence Markussen (2001*c*) proposes to use the formulae for the Lebesgue densities of these random variables based on the *integration by parts* setting from Malliavin calculus to simulate a pseudo-likelihood. However, the employment of the integration by parts formula is numerically demanding as well as instable, and the proposed procedure is probably of little usefulness.

Choosing a good basis

Despite the disadvantages of the suggestion in Markussen (2001*c*), we will like to comment on how sensible truncations of the chaos expansion for non-linear SPDEs could be made. Instead of using the *Hermite functions*, *cf.* Holden et al. (1996, p. 18), as the orthonormal basis for $L^2(T)$, it is often possible to chose a basis $\xi_n \in \mathscr{S}(T)$, $n \in \mathbb{N}$, which relates to the SPDE under consideration. If we for instance are analyzing the non-linear cable equation (1.4), then the orthonormal basis consisting of the eigenfunctions for the linear part would have good properties, and the corresponding misspecification be less gross. Regarding the question for which multi indices $a \in \mathcal{I}$ the coefficient functions should be found see *e.g.* the discussion in Theting (2000).

[16] The process $X_{\text{Wick}}(t)$ is continuous, and if it also were strong Markov, then an overshoot of the carrying capacity would be impossible. Observe that non-linear Wick type equations typically are non-Markovian due to the non-locality of the Wick product in the ω-space.

Moment estimators

Besides the final suggestion in Markussen (2001*c*) to impose measurement errors on the observations in order to simulate the pseudo-likelihood more easily, we will point out an even easier calibration method for Wick type SPDEs. Since the Wick product commute with the expectation operator, *cf.* section 1.2.3, it is often possible to show, that the expectation of the solution of a Wick type SPDE solve the corresponding PDE, where the stochastic coefficients are replaced by their expectations. Knowing the expectation of the solution will often give the possibility of employing some kind of moment estimators.

I

Uniform convergence of approximate likelihood functions for a stationary multivariate Gaussian time series

Abstract

We study the likelihood function for a finite dimensional parameter in a misspecified Gaussian time series model. If the misspecification vanishes at rate $\sqrt{n}$, then a uniform version of the local asymptotic normality property is proved under mild regularity conditions formulated in terms of multivariate L^p-norms. Especially, time series which either are close to deterministic processes, have long range dependence or both properties are included. As examples an approximative likelihood for a stochastic partial differential equation observed at discrete points in time and space, and a multivariate time series with long range dependence, are presented. The method of proof relies on standard inequalities and the interplay between Schatten p-norms of Toeplitz matrices and L^p-norms of matrix valued functions.

Key words

Multivariate Gaussian time series, misspecified model, approximate likelihood, local asymptotic normality, long range dependence, almost deterministic processes, discretely observed stochastic partial differential equation, Schatten p-norm, multivariate L^p-norm, Toeplitz matrices.

I.1 Introduction

This paper presents a unified treatment on the asymptotic properties of the likelihood function for a finite dimensional parameter in a sequence of possible misspecified multivariate Gaussian time series models given by their spectral densities. The motivation for this study stems from a problem concerning parameter estimation given observations at discrete points in time and space of an linear stochastic partial differential equation. Such an observation will have a representation as an infinite dimensional state space model, and in practice it will be convenient to approximate this representation by a finite dimensional model. In order to achieve consistent and asymptotically efficient estimators the dimension of this approximation should increase, *i.e.* the misspecification should decrease, as the number of observations increases. For a general introduction to time series and spectral analysis see Brockwell & Davis (1991).

We use standard inequalities on multivariate L^p-norms to prove uniform convergence of the log likelihood function to its mean value, and a uniform version of the local asymptotic normality property at rate $\sqrt{n}$. As a consequence of this, asymptotic normality and efficiency of the corresponding maximum likelihood estimator follows. The needed regularity conditions, which are formulated in terms of L^p-norms on the spectral densities, cover spectral densities with pole's, *i.e.* time series with long range dependence, and zero's, *i.e.* time series close to deterministic processes. If a model without misspecification is used, then the regularity conditions specializes to very general conditions on the spectral densities. This paper strengthen earlier result for multivariate Gaussian time series, see *e.g.* Dunsmuir & Hannan (1976), and for long range depend time series, see *e.g.* Dahlhaus (1989). The required regularity conditions are weakened and an stronger conclusion is obtained in terms of the local asymptotic normality property.

This paper is organized as follows. In section I.2 we describe the multivariate L^p-norms and the Schatten matrix p-norms, which plays a crucial role in this paper. Moreover we present and discuss the regularity conditions, state the main theorem, and two examples are considered. The proof of the main theorem is given in section I.3, and in the appendix we prove results connecting L^p-norms of functions, p-norms of Toeplitz matrices and their inverses, Fejér approximations of functions, and the approximation of the product of Toeplitz matrices by the Toeplitz matrix of the product.

I.2 Main result and examples

First we will introduce some notation. In the remaining of this paper let $b \in \mathbb{N}$ be a fixed dimension, and let

$$\mathbb{M}_n = \mathbb{M}_n(\mathbb{C}^{b\times b}) \simeq \mathbb{C}^{nb\times nb}, \quad n \in \mathbb{N} \cup \{\infty\},$$

be the set of n-dimensional matrices with $\mathbb{C}^{b\times b}$-valued entries. Let $I_n \in \mathbb{M}_n$, $n \in \mathbb{N}$, denote the identity matrices, and given a matrix $A \in \mathbb{M}_n$ let A^* denote the complex conjugate transposed matrix.

Definition I.1 The singular values of $A \in \mathbb{M}_n$ are the eigenvalues of the positive semi definite matrix $|A| = (A^*A)^{1/2}$. The Schatten p-norm $\|A\|_p$, $p \in [1, \infty]$, is the l^p-norm of the singular values of A. The L^p-norm $\|f\|_p$, $p \in [1, \infty]$, of a matrix valued function

$$f : (-\tfrac{1}{2}, \tfrac{1}{2}] \to \mathbb{M}_n$$

is the usual L^p-norm of the real function $\|f(\cdot)\|_p$.

The space L^p of matrix valued functions on $(-\frac{1}{2}, \frac{1}{2}]$ with finite L^p-norm thus generalizes the usual L^p-space of complex valued functions. By periodic extension the functions in L^p will be considered as defined on the whole of $\mathbb{R}$. For $f \in L^1$ the Fourier coefficient matrices $\hat{f}(k) \in \mathbb{M}_1$, $k \in \mathbb{Z}$, and the Toeplitz matrices $T_n(f) \in \mathbb{M}_n$, $n \in \mathbb{N}$, are given by

$$\hat{f}(k) = \int \exp(-2\pi ikx) f(x) \, \mathrm{d}x, \qquad T_n(f) = \big(\hat{f}(i-j)\big)_{i,j=1,\dots,n},$$

where, as everywhere else unless explicitly stated, the integral is a Lebesgue integral over $(-\frac{1}{2}, \frac{1}{2}]$. Denote by L^p_+ the subset of functions that are positive definite almost everywhere. If $f \in L^1_+$, then the Toeplitz matrix $T_n(f)$ is positive definite and can be interpreted as a covariance matrix. This correspondence between covariance matrices and matrix valued functions will be essential for the analysis of the likelihood function for a stationary Gaussian time series given below.

In our analysis we will separate the data generating mechanism from the statistical model. Thus let $X_1, X_2, \dots$ be a stationary b-dimensional Gaussian time series with spectral density $\varphi \in L^1_+$, *i.e.*

$$\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \overset{d}{\sim} \mathcal{N}_{nb}\big(0, T_n(\varphi)\big).$$

Let $\Theta \subseteq \mathbb{R}^d$ be a d-dimensional parameter region, and assume that we for each $n \in \mathbb{N}$ have a statistical model given by a family $\varphi_{n,\theta}$, $\theta \in \Theta$, of b-dimensional spectral densities. The log likelihood function $l_n(\theta)$ given an observation of $\mathbb{X}_n$, and the associated log likelihood ratio $\chi_n(\theta)$ w.r.t. some fixed parameter $\theta_0 \in \Theta$, then are given by

$$l_n(\theta) = -\tfrac{1}{2} \log \det T_n(\varphi_{n,\theta}) - \tfrac{1}{2} \operatorname{tr}\big[T_n(\varphi_{n,\theta})^{-1} \mathbb{X}_n \mathbb{X}_n^*\big], \qquad \chi_n(\theta) = l_n(\theta) - l_n(\theta_0).$$

Let $B(\theta, \rho) = \{\eta \in \Theta : |\eta - \theta| \le \rho\}$ be the ball with center $\theta \in \Theta$ and radius $\rho > 0$, and consider the following regularity conditions.

(I.3)

Assumption I.2 The parameter space Θ is a bounded and convex subset of $\mathbb{R}^d$. The spectral densities $\varphi_{n,\theta}$ are continuously differentiable in L^1 as functions of θ with derivatives $\nabla\varphi_{n,\theta} = (\partial_i\varphi_{n,\theta})_{i=1,\dots,d}$. The Fisher information matrix

$$\mathcal{J} = \Big(\lim_{n\to\infty} \tfrac{1}{2} \int \operatorname{tr}\big(\varphi_{n,\theta_0}^{-1}(x)\partial_i\varphi_{n,\theta_0}(x)\varphi_{n,\theta_0}^{-1}(x)\partial_j\varphi_{n,\theta_0}(x)\big)\,\mathrm{d}x\Big)_{i,j=1,\dots,d}$$

exists and is positive definite. For every $\varepsilon > 0$ the norm

$$\big\|\varphi_{n,\theta_0}^{\frac{1}{2}}\varphi_{n,\theta}^{-1}\varphi_{n,\theta_0}^{\frac{1}{2}} - I_1\big\|_1$$

is bounded away from zero uniformly over $n\in\mathbb{N}$, $\theta\in\Theta\setminus B(\theta_0,\varepsilon)$.

Assumption I.2 consists of a structure condition on the parameter space, and smoothness respectively identifiability conditions on the parameterization. Only the structure condition might seem somewhat restrictive, and this condition could indeed be relaxed in many situation at the cost of additional arguments. In order to formulate the more technical regularity conditions let p, q, r be fixed numbers such that

i) $p\in[1,\infty]$ and $q, r\in(2,\infty]$,

ii) $p^{-1}+q^{-1}+r^{-1}\le\frac{1}{2}$ or $p<2$, $q=\infty$, $\frac{1}{2}p^{-1}+r^{-1}\le\frac{1}{2}$,

The integrability indices p, q, r could be allowed to depend on $\theta\in\Theta$. But in order to keep the proofs as simple as possible we will refrain from these extensions.

Assumption I.3 The norms

$$\|\varphi_{n,\theta}\|_p\,, \qquad \|\partial_i\varphi_{n,\theta}\|_p\,, \qquad \big\|\varphi_{n,\theta}^{-1}\big\|_q\,, \qquad \big\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta}\big\|_r$$

are uniformly bounded over $n\in\mathbb{N}$, $\theta\in\Theta$. For every vanishing sequence ρ_n the norms

$$\|\partial_i\varphi_{n,\theta}-\partial_i\varphi_{n,\theta_0}\|_{(1-q^{-1}-r^{-1})^{-1}}\,, \qquad \big\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta}-\varphi_{n,\theta_0}^{-1}\partial_i\varphi_{n,\theta_0}\big\|_{(1-r^{-1})^{-1}}$$

vanishes uniformly over $\theta\in B(\theta_0,\rho_n)$ as $n\to\infty$.

Assumption I.3 consists of integrability conditions on the spectral densities, and a uniform strengthening of the smoothness of the parameterization. The technique used in this paper works best when $p^{-1}+q^{-1}+r^{-1}\le\frac{1}{2}$, and the following regularity condition is constructed to encompass that situation.

Assumption I.4 Suppose $p^{-1}+q^{-1}+r^{-1}\le\frac{1}{2}$. The norms

$$n^{\frac{1}{2}}\,\|\varphi-\varphi_{n,\theta_0}\|_{(\frac{1}{2}-q^{-1})^{-1}}$$

vanish as $n\to\infty$. There exists $\alpha\in(0,1)$ such that the norms

$$|y|^{-\alpha}\,\|\varphi_{n,\theta}-\varphi_{n,\theta}(\cdot-y)\|_{(1-p^{-1}-2q^{-1})^{-1}}\,,$$
$$|y|^{-\alpha}\,\big\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta}-(\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta})(\cdot-y)\big\|_{(\frac{1}{2}-p^{-1}-q^{-1})^{-1}}$$

are uniformly bounded over $y\in(-\frac{1}{2},\frac{1}{2}]$, $n\in\mathbb{N}$, $\theta\in\Theta$.

(I.4)

Assumption I.4 consists of a rate condition on the approximation of the true spectral density φ by the sequence φ_{n,θ_0}, and smoothness conditions on the spectral densities. Assumption I.4 also allows spectral densities with a zero, *i.e.* time series that are close to be deterministic. There is a duality between the steepness of the pole's and the zero's formulated via the requirement $p^{-1}+q^{-1}+r^{-1} \leq \frac{1}{2}$. Probably conditions allowing for $p=\infty$, $q<2$ could also be constructed.

If $p < 2$, then more complicated regularity conditions are needed. The most interestingcase corresponds to long memory processes, *i.e.* spectral densities with a pole at zero, and assumption I.5 is constructed to encompass that situation. However, regularity conditions allowing for a pole away from zero could easily be constructed by using appropriate replacements for lemma I.15 and lemma I.17 in section I.4.

Assumption I.5 Suppose $p<2, q=\infty, \frac{1}{2}p^{-1}+r^{-1} \leq \frac{1}{2}$. The norms

$$n^{\frac{1}{2}} \left\|\varphi-\varphi_{n,\theta_0}\right\|_p, \qquad n^{\frac{1}{2}} \left\|x\big(\varphi(x)-\varphi_{n,\theta_0}(x)\big)\right\|_\infty$$

vanish uniformly over $x \in (-\frac{1}{2},\frac{1}{2}]$ as $n \to \infty$. There exists $\beta \in (0,1)$ such that the norms

$$\left\| \, |x|^{\beta} \, \varphi_{n,\theta}(x)\right\|_\infty, \qquad \left\| \, |x|^{-\frac{1}{2}} \, \varphi_{n,\theta}^{-1}(x)\right\|_\infty$$

are uniformly bounded over $x \in (-\frac{1}{2},\frac{1}{2}]$, $n \in \mathbb{N}$, $\theta \in \Theta$. There exists $\alpha \in (0,1)$ such that the norms

$$|y|^{-\alpha} \left\|\varphi_{n,\theta}-\varphi_{n,\theta}(\cdot-y)\right\|_1, \quad |y|^{-\alpha} \left\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta}-(\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta})(\cdot-y)\right\|_{2(1-p^{-1})^{-1}}$$

are uniformly bounded over $y \in (-\frac{1}{2},\frac{1}{2}]$, $n \in \mathbb{N}$, $\theta \in \Theta$.

Assumption I.5 consists of a rate condition on the approximation of the true spectral density φ by the sequence φ_{n,θ_0}, further integrability conditions on the spectral densities, and smoothness conditions on the spectral densities.

We will say that the regularity conditions are fulfilled if assumption I.2 and assumption I.3 hold, and if assumption I.4 holds when $p^{-1}+q^{-1}+r^{-1} \leq \frac{1}{2}$, and assumption I.5 holds when $p<2$. In the classical case, *i.e.* when the spectral densities do not depend on $n \in \mathbb{N}$, these conditions simplifies and are seen to be more general than the conditions presented in Dahlhaus (1989). A possibly extension of the regularity conditions would be to allow the parameter region Θ and the dimension d to increase with n, and thus include sieve estimation.

Theorem I.6 *Suppose that the regularity conditions are fulfilled. Then the maximum likelihood estimator* $\hat{\theta}_n = \operatorname{argmax}_{\theta\in\Theta} l_n(\theta)$ *is a* $\sqrt{n}$*-consistent estimator of* θ_0*, and there exists a sequence* G_n*,* $n \in \mathbb{N}$*, of* d*-dimensional random variables converging in distribution to* $\mathcal{N}_d(0,\mathcal{J})$ *such that*

$$\mathrm{E}\Big(\sup_{\theta\in B(\theta_0,n^{-\frac{1}{2}}\rho)} \big|\chi_n(\theta)-\big(n^{\frac{1}{2}}(\theta-\theta_0)^*G_n-\tfrac{1}{2}n(\theta-\theta_0)^*\mathcal{J}(\theta-\theta_0)\big)\big|\Big) \tag{I.5}$$

vanishes for every $\rho>0$ *as* $n\to\infty$.

Remark Theorem I.6 states that the log likelihood ratio $\chi_n(\theta)$ satisfies a uniform version of the LAN-condition, whence $\hat{\theta}_n$ especially is asymptotically efficient in sense of the minimax bound and the convolution property, see Le Cam & Yang (2000). Moreover if θ_0 is in the interior of Θ, then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to $\mathcal{N}_d(0, \mathcal{J}^{-1})$.

I.2.1 A stochastic partial differential equation

Let $W(t,x)$ be a standard Brownian sheet, see Walsh (1986), and let $X_\theta(t,x)$, $\theta \in \Theta \subseteq (0,\infty)$, be the stationary solution in time $t \geq 0$ of the parabolic stochastic partial differential equation

$$\begin{cases} \mathrm{d}X_\theta(t,x) = \theta\, \partial_x^2 X_\theta(t,x)\, \mathrm{d}t + \mathrm{d}W(t,x), & t > 0,\ 0 < x < 1, \\ X_\theta(t,0) = X_\theta(t,1) = 0, & t \geq 0. \end{cases} \tag{I.1}$$

We assume $X_{\theta_0}(t,x)$ is observed at the discrete lattice points in time and space given by

$$t = 1, \dots, n, \qquad\qquad x = x_1, \dots, x_b, \quad x_i = \frac{u_i}{v}$$

for some fixed $u_1, \dots, u_b, v \in \mathbb{N}$, $u_1 < \cdots < u_b < v$, and wish to estimate θ_0. By Walsh (1986, p. 324) the stationary solution to (I.1) is given by

$$X_\theta(t,x) = \sqrt{2} \sum_{j=1}^{\infty} \sin(\pi j x)\, Y_{j,\theta}(t),$$

where $Y_{j,\theta}(t)$, $j \in \mathbb{N}$, are pairwise independent Ornstein-Uhlenbeck processes given by the stationary solutions to the stochastic differential equations

$$\mathrm{d}Y_{j,\theta}(t) = -\theta j^2\, Y_{j,\theta}(t)\, \mathrm{d}t + \mathrm{d}W_j(t).$$

Now let the time series $X_\theta(t)$, $Y_\theta(t)$, $\tilde{Y}_\theta(t)$, $t = 1, \dots, n$, and the matrices A, $\tilde{A}$ be given by

$$\begin{aligned} X_\theta(t) &= \big(X_\theta(t,x_i)\big)_{i=1,\dots,b}, \\ Y_\theta(t) &= \big(Y_{j,\theta}(t)\big)_{j\in\mathbb{N}}, & A &= \big(\sqrt{2}\sin(\pi x_i j)\big)_{\substack{i=1,\dots,b,\\ j\in\mathbb{N}}}, \\ \tilde{Y}_\theta(t) &= \Big(\sum_{k=0}^{\infty} Y_{j+2vk,\theta}(t)\Big)_{j=1,\dots,2v}, & \tilde{A} &= \big(\sqrt{2}\sin(\pi x_i j)\big)_{\substack{i=1,\dots,b\\ j=1,\dots,2v}}. \end{aligned}$$

Since the sinus function is periodic, we find that $A = (\tilde{A}\,\tilde{A}\cdots)$. Moreover, $X_\theta(t)$ can be represented as a state space model, $X_\theta(t) = A\, Y_\theta(t) = \tilde{A}\,\tilde{Y}_\theta(t)$, where the state space vector $Y_\theta(t)$ is an infinite dimensional first order autoregressive process. The likelihood function can in principle be calculated via the infinite dimensional

Kalman filter. In order to do computations we propose to approximate the infinite dimensional model. The time series $Y_{j,\theta}(t)$ have spectral density $\psi_{j,\theta}(x)$ given by

$$\psi_{j,\theta}(x) = \frac{1}{2\theta j^2}\big(1 - 2e^{-\theta j^2}\cos(2\pi x) + e^{-2\theta j^2}\big)^{-1},$$

whence the time series $X_\theta(t)$ have spectral density $\varphi_\theta(x)$ given by

$$\varphi_\theta(x) = \tilde{A}\,\mathrm{diag}\Big(\sum_{k=0}^{\infty}\psi_{j+2vk,\theta}(x)\Big)_{j=1,\dots,2v}\tilde{A}^*.$$

We now approximate the infinite dimensional model by a finite dimensional state space model plus a white noise term, where the dimension $K(n)$ of the approximation increases with n. The approximative spectral density $\varphi_{n,\theta}$ is given by

$$\varphi_{n,\theta}(x) = \tilde{A}\,\mathrm{diag}\Big(\sum_{k\in\mathbb{N}_0:j+2vk<K(n)}\psi_{j+2vk,\theta}(x) + \tau^2_{n,j,\theta}\Big)_{j=1,\dots,2v}\tilde{A}^*, \quad n\in\mathbb{N},$$

and the white noise variances $\tau^2_{n,j,\theta}$ are given by

$$\tau^2_{n,j,\theta} = \sum_{k\in\mathbb{N}_0:j+2vk\geq K(n)}\frac{1}{2\theta(j+2vk)^2}, \quad j=1,\dots,2v,\ n\in\mathbb{N}.$$

If $K(n)$ increases sufficiently fast with n, then assumption I.4 is satisfied with $p = q = r = \infty$. How fast $K(n)$ should increase is conceived by calculating the norms $\|\varphi_\theta - \varphi_{n,\theta}\|_\infty$. Since

$$\Big\|\psi_{j,\theta} - \frac{1}{2\theta k^2}\Big\|_\infty = \frac{1}{\theta k^2}\frac{e^{-\theta k^2}}{1-e^{-\theta k^2}},$$

and the L^p-norms satisfies the Hölder inequality, *cf.* section I.3, we see that

$$\|\varphi_\theta - \varphi_{n,\theta}\|_\infty \leq \big\|\tilde{A}^*\tilde{A}\big\|_\infty \sum_{k=K(n)}^{\infty}\frac{1}{\theta k^2}\frac{e^{-\theta k^2}}{1-e^{-\theta k^2}}.$$

Thus if $e^{\theta_0 K(n)^2} = n^{\frac{1}{2}}$, *i.e.* $K(n) = \sqrt{\frac{1}{2\theta_0}\log n}$, then $n^{\frac{1}{2}}\|\varphi_{\theta_0} - \varphi_{n,\theta_0}\|_\infty$ vanishes as $n\to\infty$ and assumption I.4 is satisfied.

I.2.2 Multivariate long memory time-series

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a b-dimensional Gaussian process with spectral density φ_{θ_0}, where

$$\varphi_\theta(x) = \frac{2^{-\theta}}{1-\theta}|x|^{-\theta} I_1, \qquad \varphi_\theta^{-1}(x)\partial_\theta\varphi_\theta(x) = \Big(\frac{1}{1-\theta} - \log|2x|\Big)I_1. \tag{I.7}$$

This process has stationary variance I_1 and long memory. The family of spectral densities φ_θ, $\theta \in \Theta = [\frac{1}{2}, \theta_{\max}]$, $\theta_{\max} < 1$, satisfies the regularity conditions with

$$p = \tfrac{1}{2}(1 + \theta_{\max}^{-1}), \qquad q = \infty, \qquad r = 2(1 - p^{-1})^{-1}, \qquad \alpha = \tfrac{1}{2}(1 - \theta_{\max}),$$

and the Fisher information matrix $\mathcal{J}_\theta$ at θ is given by

$$\mathcal{J}_\theta = \int_0^{\frac{1}{2}} b\Big(\frac{1}{1-\theta} - \log(2x)\Big)^2 \,\mathrm{d}x = b\Big(1 + \frac{1}{1-\theta} + \tfrac{1}{2}\frac{1}{(1-\theta)^2}\Big).$$

The above model can easily be generalized by allowing the components to have different scale of long memory. Using $(\varepsilon_t)_{t\in\mathbb{Z}}$ as a noise process it is now straightforward to build long memory processes with *e.g.* ARMA structure. Thus let $(X_t)_{t\in\mathbb{N}}$ be a causal time invariant linear filter of $(\varepsilon_t)_{t\in\mathbb{Z}}$ with transfer function ψ_{ξ_0},

$$X_t = \sum_{j=0}^{\infty} c_{\xi_0 j}\varepsilon_{t-j}, \qquad\qquad \psi_\xi(x) = \sum_{j=0}^{\infty} c_{\xi j} e^{-2\pi i j x},$$

where $c_\xi = (c_{\xi j})_{j\in\mathbb{N}_0}$, $c_{\xi j} \in \mathbb{R}^{\tilde{b}\times b}$, is a family of coefficients parameterized by $\xi \in \Xi \subseteq \mathbb{R}^{d-1}$. If the power transfer functions $|\psi_\xi|^2$ and their inverses all belong to L_+^∞, are continuous differentiable in L_+^∞, and satisfies the smoothness condition formulated in assumption I.3, then theorem I.6 holds for the model given by the spectral densities

$$\psi_\xi(x)\varphi_\theta(x)\psi_\xi^*(x), \quad (\theta, \xi) \in \Theta \times \Xi \subseteq \mathbb{R}^d.$$

I.3 Uniform convergence of the likelihood

The purpose of this section is to prove theorem I.6. We first list some properties of the Schatten p-norms, see *e.g.* Bhatia (1997), the L^p-norms and the interplay with Toeplitz matrices. The p-norms are matrix norms, which decrease in p and satisfy the Hölder inequality, and the 1-norm bounds the trace, *i.e.*

$$\|A\|_p \overset{p>q}{\leq} \|A\|_q, \qquad \|AB\|_p \overset{p^{-1}=q^{-1}+r^{-1}}{\leq} \|A\|_q \|B\|_r, \qquad |\mathrm{tr}\, A| \leq \|A\|_1.$$

Moreover the 2-norm coincides with the Frobenius norm, and the ∞-norm coincides with the operator norm, *i.e.*

$$\|A\|_2^2 = \mathrm{tr}(A^*A), \qquad \|A\|_\infty = \max\big\{ \|Ah\|_{l_2} \;\big|\; h \in l_2(\mathbb{C}) : \|h\|_{l_2} = 1\big\}.$$

Since the p-norms satisfy the Hölder inequality, the L^p-norms also satisfy the Hölder inequality. If $f \in L_+^1$, $f^{-1} \in L_+^p$ and $g \in L^p$, then

$$\big\|T_n(f)^{-1}\big\|_p \leq \big\|T_n(f^{-1})\big\|_p, \qquad\qquad n^{-p^{-1}} \|T_n(g)\|_p \leq \|g\|_p, \tag{I.8}$$

see lemma I.12 and lemma I.16 in section I.4 below. In the rest of this section we will refer to the inequalities stated above as the standard inequalities.

Now consider the spectral densities $\varphi_{n,\theta}$ introduced in section I.2. For ease of notation we will skip the index n of the spectral densities when they appear in Toeplitz matrices, *e.g.* we abbreviate $T_n(\varphi_{n,\theta})$ by $T_n(\varphi_\theta)$ and so forth. In section I.4 we prove lemma I.8, which states that the product of Toeplitz matrices can be approximated by the Toeplitz matrix of the product.

Lemma I.8 *If the regularity conditions are fulfilled, then*

$$\begin{aligned}\sup_{\theta\in\Theta} n^{-1}\left\|T_n(\varphi_\theta)^{-1}T_n(\varphi_{\theta_0}) - T_n(\varphi_\theta^{-1}\varphi_{\theta_0})\right\|_1 &= O(n^{-\frac{\alpha}{2+4\alpha}}\log n),\\ \sup_{\theta\in\Theta} n^{-\frac{1}{2}}\left\|T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta) - T_n(\varphi_\theta^{-1}\partial_i\varphi_\theta)\right\|_2 &= O(n^{-\frac{\alpha}{1+2(\frac{1}{2}-q^{-1})^{-1}\alpha}}\sqrt{\log n}).\end{aligned}$$

Before we start the analysis of the likelihood function, we will introduce some notation. The log likelihood ratio $\chi_n(\theta)$ w.r.t. the unknown parameter θ_0 is given by

$$\chi_n(\theta) = l_n(\theta) - l_n(\theta_0) = v_n(\theta) + w_n(\theta) + Z_n(\theta), \tag{I.2}$$

where the functions $v_n(\theta)$, $w_n(\theta)$, and the mean zero stochastic process $Z_n(\theta)$ are defined by

$$\begin{aligned}v_n(\theta) &= \tfrac{1}{2}\log\det\big(T_n(\varphi_\theta)^{-1}T_n(\varphi_{\theta_0})\big) + \tfrac{1}{2}\operatorname{tr}\left[T_n(\varphi_\theta)^{-1}T_n(\varphi_\theta - \varphi_{\theta_0})\right],\\ w_n(\theta) &= \tfrac{1}{2}\operatorname{tr}\left[\big(T_n(\varphi_{\theta_0})^{-1} - T_n(\varphi_\theta)^{-1}\big)T_n(\varphi - \varphi_{\theta_0})\right],\\ Z_n(\theta) &= \tfrac{1}{2}\operatorname{tr}\left[\big(T_n(\varphi_{\theta_0})^{-1} - T_n(\varphi_\theta)^{-1}\big)\big(\mathbb{X}_n\mathbb{X}_n^* - T_n(\varphi)\big)\right].\end{aligned}$$

Moreover, let the function $\tilde{\chi}_n(\theta)$ be defined by

$$\tilde{\chi}_n(\theta) = \tfrac{1}{2}\int\Big(\log\det\big(\varphi_{n,\theta}^{-1}(x)\varphi_{n,\theta_0}(x)\big) + \operatorname{tr}\left[\varphi_{n,\theta}^{-1}(x)\big(\varphi_{n,\theta}(x) - \varphi_{n,\theta_0}(x)\big)\right]\Big)\,\mathrm{d}x.$$

The proof of theorem I.6 is divided into the following three steps

i) consistency of the maximum likelihood estimator $\hat{\theta}_n$,

ii) $\sqrt{n}$-rate of convergence of $\hat{\theta}_n$,

iii) the uniform version of the LAN-property.

In order to find the proper rate of convergence, *cf.* step (ii), we will use the following lemma formulated in terms of the decomposition (I.2) of the log likelihood ratio introduced above.

(I.9)

Lemma I.9 *If $\hat{\theta}_n$ is consistent for θ_0, $n^{\delta}(\hat{\theta}_n - \theta_0)$ is tight for some $\delta \geq 0$ and there exist constants $\varepsilon > 0$, $c_1 > 0$, $c_2 < \infty$ such that*

$$\begin{aligned} \limsup_{n\to\infty} \sup_{\theta \in B(\theta_0,\rho_n)\setminus\{\theta_0\}} |\theta-\theta_0|^{-2} n^{-1} v_n(\theta) &< -c_1, \\ \limsup_{n\to\infty} \mathrm{E}\Big(\sup_{\theta\in B(\theta_0, n^{-\delta}\rho)} n^{2\delta+\varepsilon-1}\big(w_n(\theta)+Z_n(\theta)\big)\Big) &< c_2\,\rho \end{aligned} \tag{I.3}$$

for every vanishing sequence ρ_n and every $\rho > 0$, then $n^{\delta+\varepsilon}(\hat{\theta}_n - \theta_0)$ is tight.

Proof This proof follows van der Vaart & Wellner (1996, theorem 3.2.5). Given $M \in \mathbb{N}$ the probability $\mathrm{P}\big(n^{\delta+\varepsilon}|\hat{\theta}_n - \theta_0| > 2^M\big)$ is bounded by

$$\mathrm{P}\big(|\hat{\theta}_n-\theta_0| > \rho_n\big) + \mathrm{P}\big(n^{\delta}|\hat{\theta}_n - \theta_0| > \rho\big) + \sum_{j>M: n^{-\delta-\varepsilon}2^j \leq \rho_n \wedge n^{-\delta}\rho} \mathrm{P}\big(\hat{\theta}_n \in S_{jn}\big),$$

where the shells S_{jn} are given by $S_{jn} = B(\theta_0, n^{-\delta-\varepsilon}2^j) \setminus B(\theta_0, n^{-\delta-\varepsilon}2^{j-1})$. The first term vanishes if ρ_n decreases sufficiently slowly, and the second term vanishes as $\rho\to\infty$. If $n^{-\delta-\varepsilon}2^j \leq \rho_n \wedge n^{-\delta}\rho$, then $\mathrm{P}(\hat{\theta}_n \in S_{jn})$ is bounded by

$$\mathrm{P}\Big(\sup_{\theta\in S_{jn}} n^{2\delta+\varepsilon-1}\big(w_n(\theta)+Z_n(\theta)\big) \geq -\sup_{\theta\in S_{jn}} n^{2\delta+\varepsilon-1} v_n(\theta) > c_1 n^{-\varepsilon}2^{2j-2}\Big),$$

and hence by $c_2 c_1^{-1} 2^{-j+2}$ according to Markov's inequality. The third term is thus bounded by $c_2 c_1^{-1} 2^{-M+2}$ and vanishes as $M\to\infty$. It follows that $n^{\delta+\varepsilon}(\hat{\theta}_n - \theta_0)$ is tight. □

The first step of theorem I.6 is contained in the following theorem.

Theorem I.10 *Suppose the regularity conditions are fulfilled. Then the maximum likelihood estimator $\hat{\theta}_n = \operatorname{argmax}_{\theta\in\Theta} l_n(\theta)$ is consistent for θ_0 as $n\to\infty$.*

Proof We first analyze the function $\tilde{\chi}_n(\theta)$. The function $g : (-1,\infty) \to \mathbb{R}$ defined by $g(x) = x - \log(x+1)$ is non-negative, strictly convex, has the unique minimum $g(0) = 0$, and satisfies $g(x) < g(-x)$ for $x \in (0,1)$. Thus if $0 \leq \lambda_1(x) \leq \cdots \leq \lambda_b(x)$ denote the eigenvalues of the positive semi definite matrix

$$\varphi_{n,\theta_0}^{\frac{1}{2}}(x)\varphi_{n,\theta}^{-1}(x)\varphi_{n,\theta_0}^{\frac{1}{2}}(x)$$

for fixed $n\in\mathbb{N}$ and $\theta\in\Theta$, then the Jensen inequality gives

$$\begin{aligned} -\tilde{\chi}_n(\theta) &= \tfrac{1}{2}\sum_{i=1}^{b}\int g\big(\lambda_i(x)-1\big)\,\mathrm{d}x \geq \tfrac{1}{2}\sum_{i=1}^{b}\int g\big(|\lambda_i(x)-1|\big)\,\mathrm{d}x \\ &\geq \tfrac{1}{2}b\,g\Big(b^{-1}\sum_{i=1}^{b}\int |\lambda_i(x)-1|\,\mathrm{d}x\Big) = \tfrac{1}{2}b\,g\Big(b^{-1}\big\|\varphi_{n,\theta_0}^{\frac{1}{2}}\varphi_{n,\theta}^{-1}\varphi_{n,\theta_0}^{\frac{1}{2}} - I_1\big\|_1\Big). \end{aligned} \tag{I.10}$$

Let $\varepsilon > 0$ be given. Then we by assumption I.2 have

$$-\limsup_{n\to\infty} \sup_{\theta:|\theta-\theta_0|>\varepsilon} \tilde{\chi}_n(\theta) \geq \tfrac{1}{2} b\, g\Big(b^{-1} \liminf_{n\to\infty} \inf_{\theta:|\theta-\theta_0|>\varepsilon} \big\|\varphi_{n,\theta_0}^{\frac{1}{2}} \varphi_{n,\theta}^{-1} \varphi_{n,\theta_0}^{\frac{1}{2}} - I_1\big\|_1\Big) > 0,$$

and since the maximum $\chi_n(\hat{\theta}_n)$ of the log likelihood ratio is non-negative, we have the inequality

$$\mathrm{P}\big(|\hat{\theta}_n - \theta_0| > \epsilon\big) \leq \frac{-\operatorname{E}\tilde{\chi}_n(\hat{\theta}_n)}{-\sup_{\theta:|\theta-\theta_0|>\epsilon}\tilde{\chi}_n(\theta)} \leq \frac{\operatorname{E}\big(n^{-1}\chi_n(\hat{\theta}_n) - \tilde{\chi}_n(\hat{\theta}_n)\big)}{-\sup_{\theta:|\theta-\theta_0|>\epsilon}\tilde{\chi}_n(\theta)}.$$

Thus if for every $\delta > 0$ the inequality

$$\operatorname{E}\big(n^{-1}\chi_n(\hat{\theta}_n) - \tilde{\chi}_n(\hat{\theta}_n)\big) \leq \delta$$

holds for n sufficiently large, then $\hat{\theta}_n$ converge in probability to θ_0 as $n \to \infty$. Using the decomposition (I.2) of the log likelihood ratio, we see that the random variable $n^{-1}\chi_n(\hat{\theta}_n) - \tilde{\chi}_n(\hat{\theta}_n)$ is bounded by

$$\sup_{\theta\in\Theta}\big|n^{-1}v_n(\theta) - \tilde{\chi}_n(\theta)\big| + \sup_{\theta\in\Theta} n^{-1}|w_n(\theta)| + \sup_{\theta\in\Theta} n^{-1}|Z_n(\theta)|. \tag{I.4}$$

Let $\xi_s = \theta_0 + s(\theta-\theta_0)$, $s \in [0,1]$, be the natural parameterization of the line segment from θ_0 to θ, and use the identities

$$\partial_i \log\det\big(T_n(\varphi_\theta)^{-1}T_n(\varphi_{\theta_0})\big) = -\operatorname{tr}\big[T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta)\big],$$
$$\int \partial_i \log\det\big(\varphi_{n,\theta}^{-1}(x)\varphi_{n,\theta_0}(x)\big)\,\mathrm{d}x = \operatorname{tr}\big[T_n(\varphi_\theta^{-1}\partial_i\varphi_\theta)\big]$$

to rewrite the difference $n^{-1}v_n(\theta) - \tilde{\chi}_n(\theta)$ as

$$-\tfrac{1}{2}\int_0^1 n^{-1}\operatorname{tr}\big[T_n(\varphi_{\xi_s})^{-1}T_n\big((\theta-\theta_0)^*\nabla\varphi_{\xi_s}\big) - T_n\big(\varphi_{\xi_s}^{-1}(\theta-\theta_0)^*\nabla\varphi_{\xi_s}\big)\big]\,\mathrm{d}s$$
$$+\tfrac{1}{2}n^{-1}\operatorname{tr}\big[T_n(\varphi_\theta)^{-1}T_n(\varphi_\theta - \varphi_{\theta_0}) - T_n\big(\varphi_\theta^{-1}(\varphi_\theta - \varphi_{\theta_0})\big)\big].$$

The first term in (I.4) is thus bounded by

$$\tfrac{1}{2}\sup_{\theta\in\Theta}|\theta-\theta_0| \sup_{\xi\in\Theta}\max_{i=1,\dots,d} n^{-\frac{1}{2}}\big\|T_n(\varphi_\xi)^{-1}T_n(\partial_i\varphi_\xi) - T_n(\varphi_\xi^{-1}\partial_i\varphi_\xi)\big\|_2\, n^{-\frac{1}{2}}\|I_n\|_2$$
$$+\tfrac{1}{2}\sup_{\theta\in\Theta} n^{-1}\big\|T_n(\varphi_\theta)^{-1}T_n(\varphi_{\theta_0}) - T_n(\varphi_\theta^{-1}\varphi_{\theta_0})\big\|_1,$$

and hence vanishes by lemma I.8. The remaining two terms in (I.4) also vanish as $n \to \infty$ as a consequence of lemma I.11 below. □

Before we state and prove lemma I.11, we introduce some more notation and technical tools. Firstly, we need maximal inequalities for stochastic processes.

These are formulated in terms of Orlicz norms. Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a given convex, strictly increasing function with $0 \leq \psi(0) < 1$, and let ψ^{-1} denote the inverse function of ψ. Then the Orlicz ψ-norm of a random variable Y is defined by

$$\|Y\|_\psi = \inf\big\{c > 0 : E\psi(c^{-1}|Y|) \leq 1\big\}.$$

If $\psi(x) = x^p$, $p \geq 1$, then the Orlicz ψ-norm coincides with the usual L^p-norm of random variables. Suppose a stochastic process $Y(\xi)$, $\xi \in \Xi \subseteq \mathbb{R}^d$, defined on a probability space $(\Omega, \mathscr{A}, \mathrm{P})$, and $\xi_0 \in \Xi$ are given. If $\rho = \sup_{\xi\in\Xi} |\xi - \xi_0|$ is the radius of Ξ around ξ_0, then a chaining argument gives the maximal inequality

$$\int_A \sup_{\xi\in\Xi} |Y(\xi) - Y(\xi_0)| \,\mathrm{dP} \leq \rho \sup_{\xi,\eta\in\Xi} \frac{\|Y(\xi) - Y(\eta)\|_\psi}{|\xi - \eta|} \mathrm{P}(A) \sum_{j=1}^\infty 2^{-j+1} \psi^{-1}\Big(\frac{3^d 2^{jd}}{\mathrm{P}(A)}\Big), \tag{I.5}$$

$A \in \mathscr{A}$, *cf.* Ledoux & Talagrand (1991, theorem 11.2). Below we will use the exponential function ψ_0 given by

$$\psi_0(x) = \tfrac{1}{2} \exp\big(x - \tfrac{1}{3}\big), \qquad \psi_0^{-1}(y) = \tfrac{1}{3} + \log(2y),$$

and the bound (I.5) with $\psi = \psi_0$ specializes to

$$2\rho \sup_{\xi,\eta\in\Xi} |\xi - \eta|^{-1} \|Y(\xi) - Y(\eta)\|_{\psi_0} \mathrm{P}(A)\big(\tfrac{1}{3} - \log\big(\tfrac{1}{2}\mathrm{P}(A)\big) + d\log(12)\big). \tag{I.6}$$

We thus use Schatten p-norms of matrices, L^p-norms of matrix valued functions, and Orlicz ψ-norms of random variables. This should however not give rise to confusion since which norm is used will be clear from the context. Secondly, we need a further norm inequality for the product of two Toeplitz matrices. In the remaining of this paper let the functions $\zeta_\beta \in L^1$, $\beta > -1$, be defined by

$$\zeta_\beta(x) = |x|^\beta I_1.$$

If $f \in L^p$, $g \in L^1_+$, then lemma I.17 gives the bound

$$n^{-\frac{1}{2}} \big\|T_n(g)^{-1} T_n(f)\big\|_2 \leq \big\|\zeta_{-\frac{1}{2}} g^{-1}\big\|_\infty \Big(\|\zeta_1 f\|^{\frac{1}{2}}_{(1-p^{-1})^{-1}} \|f\|^{\frac{1}{2}}_p + \sqrt{\log n}\, \|f\|_1 \Big). \tag{I.7}$$

Lemma I.11 *Suppose the regularity conditions are fulfilled, and let $\alpha > 0$ be the constant from assumption I.4 respectively assumption I.5. Given $\rho > 0$, a vanishing sequence ρ_n, and $\delta, \varepsilon \geq 0$ such that $\varepsilon < \alpha(1 + 4\alpha)^{-1} \wedge (\frac{1}{2} - r^{-1})$, $\delta + \varepsilon \leq \frac{1}{2}$, then*

$$i) \lim_{n\to\infty} \sup_{\theta\in B(\theta_0,\rho_n)} |\theta - \theta_0|^{-2} \big|n^{-1} v_n(\theta) + \tfrac{1}{2}(\theta - \theta_0)^* \mathcal{J}(\theta - \theta_0)\big| = 0,$$

$$ii) \lim_{n\to\infty} \sup_{\theta\in\Theta} |\theta - \theta_0|^{-1} n^{-\frac{1}{2}} |w_n(\theta)| = 0,$$

$$iii) \limsup_{n\to\infty} \rho^{-1}\, \mathrm{E}\Big(\sup_{\theta\in B(\theta_0, n^{-\delta}\rho)} n^{2\delta+\varepsilon-1} |Z_n(\theta)| \Big) < \infty.$$

(I.12)

Proof In order to prove (i) we rewrite the function $v_n(\theta)$ as

$$-\tfrac{1}{2}\int_0^1\int_0^1 \operatorname{tr}\left[T_n(\varphi_{\xi_s})^{-1}T_n\big((\theta-\theta_0)^*\nabla\varphi_{\xi_s}\big)T_n(\varphi_{\xi_s})^{-1}T_n\big(s(\theta-\theta_0)^*\nabla\varphi_{\xi_{su}}\big)\right]\mathrm{d}u\,\mathrm{d}s,$$

where $\xi_s=\theta_0+s(\theta-\theta_0)$. Thus if the matrix $\mathcal{J}_n$ is defined by

$$\mathcal{J}_n=\Big(\tfrac{1}{2}\int \operatorname{tr}\big(\varphi_{n,\theta_0}^{-1}(x)\partial_i\varphi_{n,\theta_0}(x)\varphi_{n,\theta_0}^{-1}(x)\partial_j\varphi_{n,\theta_0}(x)\big)\,\mathrm{d}x\Big)_{i,j=1,\dots,d},$$

then the quantity

$$4\sup_{\theta\in B(\theta_0,\rho_n)}|\theta-\theta_0|^{-2}\left|n^{-1}v_n(\theta)-\tfrac{1}{2}(\theta-\theta_0)^*\mathcal{J}_n(\theta-\theta_0)\right|$$

is bounded by the supremum over $\xi,\eta\in B(\theta_0,\rho_n)$, $i,j=1,\dots,d$ of

$$n^{-1}\left\|T_n(\varphi_\xi)^{-1}T_n(\partial_i\varphi_\xi)T_n(\varphi_\xi)^{-1}T_n(\partial_j\varphi_\eta)-T_n(\varphi_{\theta_0}^{-1}\partial_i\varphi_{\theta_0}\varphi_{\theta_0}^{-1}\partial_j\varphi_{\theta_0})\right\|_1. \tag{I.8}$$

Telescoping this difference via appropriate intermediate terms we see that (I.8) is bounded by

$$\begin{aligned}
n^{-\frac{1}{2}}&\left\|T_n(\varphi_\xi)^{-1}T_n(\partial_i\varphi_\xi)-T_n(\varphi_\xi^{-1}\partial_i\varphi_\xi)\right\|_2\, n^{-\frac{1}{2}}\left\|T_n(\varphi_\xi)^{-1}T_n(\partial_j\varphi_\eta)\right\|_2\\
&+\left\|\varphi_{n,\xi}^{-1}\partial_i\varphi_{n,\xi}\right\|_r\left\|\varphi_{n,\xi}^{-1}\right\|_q\left\|\partial_j\varphi_{n,\eta}-\partial_j\varphi_{n,\xi}\right\|_{(1-q^{-1}-r^{-1})^{-1}}\\
&+\left\|\varphi_{n,\xi}^{-1}\partial_i\varphi_{n,\xi}\right\|_2 n^{-\frac{1}{2}}\left\|T_n(\varphi_\xi)^{-1}T_n(\partial_j\varphi_\xi)-T_n(\varphi_\xi^{-1}\partial_j\varphi_\xi)\right\|_2\\
&+\left\|\varphi_{n,\xi}^{-1}\partial_i\varphi_{n,\xi}\right\|_r\left\|\varphi_{n,\xi}^{-1}\partial_j\varphi_{n,\xi}-\varphi_{n,\theta_0}^{-1}\partial_j\varphi_{n,\theta_0}\right\|_{(1-r^{-1})^{-1}}\\
&+\left\|\varphi_{n,\xi}^{-1}\partial_i\varphi_{n,\xi}-\varphi_{n,\theta_0}^{-1}\partial_i\varphi_{n,\theta_0}\right\|_{(1-r^{-1})^{-1}}\left\|\varphi_{n,\theta_0}^{-1}\partial_j\varphi_{n,\theta_0}\right\|_r.
\end{aligned}$$

The second factor in the first term requires special attention. If $p^{-1}+q^{-1}+r^{-1}\le\frac{1}{2}$, then

$$n^{-\frac{1}{2}}\left\|T_n(\varphi_\xi)^{-1}T_n(\partial_j\varphi_\eta)\right\|_2\le\left\|\varphi_{n,\xi}^{-1}\right\|_q\left\|\partial_j\varphi_{n,\eta}\right\|_{(\frac{1}{2}-q^{-1})^{-1}},$$

and if $p<2$, then $n^{-\frac{1}{2}}\|T_n(\varphi_\xi)^{-1}T_n(\partial_j\varphi_\eta)\|_2$ is bounded by

$$\left\|\zeta_{-\frac{1}{2}}\varphi_{n,\xi}^{-1}\right\|_\infty\Big(\|\zeta_1\partial_j\varphi_{n,\eta}\|_{(1-p^{-1})^{-1}}\|\partial_j\varphi_{n,\eta}\|_p+\sqrt{\log n}\,\|\partial_j\varphi_{n,\eta}\|_1\Big),$$

cf. inequality (I.7). Using these bounds and lemma I.8 it follows, that (I.8) vanishes uniformly over $\xi,\eta\in B(\theta_0,\rho_n)$, $i,j=1,\dots,d$ as $n\to\infty$. Moreover, the norms $\|\mathcal{J}-\mathcal{J}_n\|_1$ vanish as $n\to\infty$ since the entries of the matrix $\mathcal{J}_n$ converge to those of $\mathcal{J}$, and (i) easily follows. In order to prove (ii) we use the identity

$$\partial_iT_n(\varphi_\theta)^{-1}=-T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta)T_n(\varphi_\theta)^{-1}$$

to rewrite the function $w_n(\theta)$ as

$$\tfrac{1}{2}\int_0^1 \operatorname{tr}\left[T_n(\varphi_{\xi_s})^{-1}T_n\big((\theta-\theta_0)^*\nabla\varphi_{\xi_s}\big)T_n(\varphi_{\xi_s})^{-1}T_n(\varphi-\varphi_{\theta_0})\right]\mathrm{d}s,$$

where $\xi_s=\theta_0+s(\theta-\theta_0)$. Using the standard inequalities we see that the quantity

$$\sup_{\theta\in\Theta}|\theta-\theta_0|^{-1}n^{-\frac{1}{2}}|w_n(\theta)|$$

is bounded by the supremum over $\theta\in\Theta$, $i=1,\dots,d$ of

$$\begin{aligned}\tfrac{1}{2}n^{-\frac{1}{2}}&\left\|T_n(\varphi_\xi)^{-1}T_n(\partial_i\varphi_\xi)-T_n(\varphi_\xi^{-1}\partial_i\varphi_\xi)\right\|_2\left\|T_n(\varphi_\xi)^{-1}T_n(\varphi-\varphi_{\theta_0})\right\|_2\\ &+\tfrac{1}{2}\left\|\varphi_{n,\xi}^{-1}\partial_i\varphi_{n,\xi}\right\|_r\left\|\varphi_{n,\xi}^{-1}\right\|_q n^{\frac{1}{2}}\left\|\varphi-\varphi_{n,\theta_0}\right\|_{(1-q^{-1}-r^{-1})^{-1}}.\end{aligned}\tag{I.9}$$

The second term in (I.9) is immediately seen to vanish as $n\to\infty$. Moreover, if $p^{-1}+q^{-1}+r^{-1}\le\frac{1}{2}$, then the second factor in the first term is bounded by

$$\left\|\varphi_{n,\xi}^{-1}\right\|_q n^{\frac{1}{2}}\left\|\varphi-\varphi_{n,\theta_0}\right\|_{(\frac{1}{2}-q^{-1})^{-1}},$$

and if $p<2$, then this factor is bounded by

$$\begin{aligned}\left\|\zeta_{-\frac{1}{2}}\varphi_{n,\xi}^{-1}\right\|_\infty\Big(&n^{\frac{1}{4}}\left\|\zeta_1(\varphi-\varphi_{n,\theta_0})\right\|_{(1-p^{-1})^{-1}}^{\frac{1}{2}}n^{\frac{1}{4}}\left\|\varphi-\varphi_{n,\theta_0}\right\|_p^{\frac{1}{2}}\\ &+\sqrt{\log n}\,n^{\frac{1}{2}}\left\|\varphi-\varphi_{n,\theta_0}\right\|_1\Big).\end{aligned}$$

The second factor of the first term in (I.9) thus grows at most at logarithmic rate. Since the first factor decreases at polynomial rate by lemma I.8, the first term thus also vanishes as $n\to\infty$. Concerning (iii) it follows from the maximal inequality (I.6), that the quantity

$$\rho^{-1}\,\mathrm{E}\Big(\sup_{\theta\in B(\theta_0,n^{-\delta}\rho)}n^{2\delta+\varepsilon-1}|Z_n(\theta)|\Big)$$

is bounded by

$$2\big(\tfrac{1}{3}+\log(2)+d\log(12)\big)\sup_{\theta,\eta\in B(\theta_0,n^{-\delta}\rho):\theta\neq\eta}\|Y_n(\theta,\eta)\|_{\psi_0},$$

where the random variables $Y_n(\theta,\eta)$ are given by

$$\begin{aligned}Y_n(\theta,\eta)&=|\theta-\eta|^{-1}n^{\delta+\varepsilon-1}\big(Z_n(\theta)-Z_n(\eta)\big)\\ &=\tfrac{1}{2}\operatorname{tr}\Big(|\theta-\eta|^{-1}n^{\delta+\varepsilon-1}\big(T_n(\varphi_\eta)^{-1}-T_n(\varphi_\theta)^{-1}\big)\big(\mathbb{X}_n\mathbb{X}_n^*-T_n(\varphi)\big)\Big).\end{aligned}$$

We thus need a uniform bound on the ψ_0-norms of $Y_n(\theta,\eta)$. Since ψ_0 is an exponential function this can be achieved via bounds on the cumulants of $Y_n(\theta,\eta)$. The first cumulant, *i.e.* the mean, of $Y_n(\theta,\eta)$ is zero, and by the product formula

(I.14)

for cumulants, see Rosenblatt (1985, theorem 2.2), the k'th cumulant, $k \geq 2$, of $Y_n(\theta, \eta)$ is given by

$$\operatorname{cum}_k Y_n(\theta, \eta) = \frac{(k-1)!}{2} \operatorname{tr}\Big[\Big(|\theta - \eta|^{-1} n^{\delta+\varepsilon-1}\big(T_n(\varphi_\eta)^{-1} - T_n(\varphi_\theta)^{-1}\big)T_n(\varphi)\Big)^k\Big].$$

Thus if $C_{n,k}$ is a bound for

$$\sup_{\theta,\eta \in B(\theta_0, n^{-\delta}\rho)} \big\| |\theta - \eta|^{-1} n^{\delta+\varepsilon-1}\big(T_n(\varphi_\eta)^{-1} - T_n(\varphi_\theta)^{-1}\big)T_n(\varphi)\big\|_k,$$

then the k'th cumulant of $Y_n(\theta, \eta)$ is bounded by $\frac{(k-1)!}{2} C_{n,k}^k$. The matrix inside the preceding k-norm can be rewritten as

$$n^{-1+\delta+\varepsilon} \int_0^1 T_n(\varphi_{\xi_s})^{-1} T_n\big(|\eta - \theta|^{-1} (\eta - \theta)^* \nabla \varphi_{\xi_s}\big) \Big\{ T_n(\varphi_{\xi_s})^{-1} T_n(\varphi - \varphi_{\theta_0}) \\ + \int_0^1 T_n(\varphi_{\xi_s})^{-1} T_n\big((\theta_0 - \xi_s)^* \nabla \varphi_{\xi_s + u(\theta_0 - \xi_s)}\big) \, \mathrm{d}u + I_n \Big\} \, \mathrm{d}s,$$

where $\xi_s = \theta + s(\eta - \theta)$. If we decompose this representation according to the three terms in the curly parenthesis, replace $T_n(\varphi_{\xi_s})^{-1} T_n\big(|\eta - \theta|^{-1} (\eta - \theta)^* \nabla \varphi_{\xi_s}\big)$ by

$$\Big(T_n(\varphi_{\xi_s})^{-1} T_n\big(|\eta - \theta|^{-1} (\eta - \theta)^* \nabla \varphi_{\xi_s}\big) - T_n\big(\varphi_{\xi_s}^{-1} |\eta - \theta|^{-1} (\eta - \theta)^* \nabla \varphi_{\xi_s}\big)\Big) \\ + T_n\big(\varphi_{\xi_s}^{-1} |\eta - \theta|^{-1} (\eta - \theta)^* \nabla \varphi_{\xi_s}\big)$$

in these three terms, and use

$$k \geq 2, \qquad \delta + \varepsilon \leq \tfrac{1}{2}, \qquad |\xi_s - \theta_0| \leq n^{-\delta}\rho,$$

then we see that $C_{n,k}$ can be chosen as the supremum over $\xi, \eta \in B(\theta_0, n^{-\delta}\rho)$, $i, j = 1, \ldots, d$ of

$$n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_i \varphi_\xi) - T_n(\varphi_\xi^{-1} \partial_i \varphi_\xi)\big\|_2 \big\| T(\varphi_\xi)^{-1} T_n(\varphi - \varphi_{\theta_0})\big\|_\infty \\ + \big\|\varphi_{n,\xi}^{-1} \partial_i \varphi_{n,\xi}\big\|_r \big\|\varphi_{n,\xi}^{-1}\big\|_q n^{\frac{1}{2}} \|\varphi - \varphi_{n,\theta_0}\|_{(1-q^{-1}-r^{-1})^{-1}} \\ + \rho\, n^{\varepsilon} n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_i \varphi_\xi) - T_n(\varphi_\xi^{-1} \partial_i \varphi_\xi)\big\|_2 n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_j \varphi_\eta)\big\|_\infty \\ + \rho\, n^{\varepsilon + r^{-1} - \frac{1}{2}} \big\|\varphi_{n,\xi}^{-1} \partial_i \varphi_{n,\xi}\big\|_r n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_j \varphi_\eta)\big\|_{(\frac{1}{2} - r^{-1})^{-1}} \\ + n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_i \varphi_\xi) - T_n(\varphi_\xi^{-1} \partial_i \varphi_\xi)\big\|_2 \\ + n^{-\frac{1}{2} + (r \wedge k \wedge 3)^{-1}} \big\|\varphi_{n,\xi}^{-1} \partial_i \varphi_{n,\xi}\big\|_{r \wedge k \wedge 3}.$$

If we insert the crude bounds

$$\big\| T(\varphi_\xi)^{-1} T_n(\varphi - \varphi_{\theta_0})\big\|_\infty \leq \big\| T(\varphi_\xi)^{-1} T_n(\varphi - \varphi_{\theta_0})\big\|_2, \\ n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_j \varphi_\eta)\big\|_\infty \leq n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_j \varphi_\eta)\big\|_2, \\ n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_j \varphi_\eta)\big\|_{(\frac{1}{2} - r^{-1})^{-1}} \leq n^{-\frac{1}{2}} \big\| T_n(\varphi_\xi)^{-1} T_n(\partial_j \varphi_\eta)\big\|_2,$$

(I.15)

use inequality (I.7) and lemma I.8, and

$$\varepsilon < \alpha\big(1 + 2(\tfrac{1}{2} - q^{-1})^{-1}\alpha\big)^{-1}, \qquad\qquad r > 2,$$

then we see that $C_{n,2}$ is bounded away from both zero and infinity, and that $C_{n,3}$ vanish as $n \to \infty$. Moreover since $C_{n,k} = C_{n,3}$ for $k \geq 3$, we find that

$$\begin{aligned} \mathrm{E}\,\psi_0\big(C_{n,2}^{-1}\,|Y_n(\theta,\eta)|\,\big) &\leq \frac{e^{-\frac{1}{3}}}{2}\,\mathrm{E}\,\big(e^{C_{n,2}^{-1}Y_n(\theta,\eta)} + e^{-C_{n,2}^{-1}Y_n(\theta,\eta)}\big) \\ &\leq e^{-\frac{1}{3}}\exp\Big(\sum_{k=2}^{\infty}\frac{|\mathrm{cum}_k\,Y_n(\theta,\eta)|}{k!C_n^k}\Big) \\ &\leq \exp\Big(-\tfrac{1}{3} + \tfrac{1}{4} + \sum_{k=3}^{\infty}\frac{1}{2k}\frac{C_{n,3}^k}{C_{n,2}^k}\Big) \xrightarrow[n\to\infty]{} \exp\big(-\tfrac{1}{3} + \tfrac{1}{4}\big) < 1. \end{aligned}$$

This implies $\|Y_n(\theta,\eta)\|_{\psi_0} \leq C_{n,2}$ for n sufficiently large. □

Using lemma I.11 we see that the last two terms in (I.4) vanish as $n \to \infty$, whence the maximum likelihood estimator $\hat\theta_n$ is a consistent estimator of θ_0. We now use lemma I.9 to prove, that $\hat\theta_n$ is $\sqrt{n}$-consistent, *i.e.* that $n^{\frac{1}{2}}(\hat\theta_n - \theta_0)$ is tight. By lemma I.11 and since the Fisher information matrix $\mathcal{J}$ is positive definite, we see that condition (I.3) is satisfied if $\varepsilon < \alpha(1+4\alpha)^{-1} \wedge (\frac{1}{2} - r^{-1})$ and $\delta + \varepsilon \leq \frac{1}{2}$, whence it follows that $n^{\frac{1}{2}}(\hat\theta_n - \theta_0)$ is tight.

We proceed to prove the uniform version of the LAN-property. By lemma I.11,

$$\sup_{\theta\in B(\theta_0, n^{-\frac{1}{2}}\rho)} \big|v_n(\theta) + w_n(\theta) + \tfrac{1}{2}(\theta-\theta_0)^*\mathcal{J}(\theta-\theta_0)\big| \xrightarrow[n\to\infty]{} 0,$$

whence it only remains to prove that

$$\mathrm{E}\Big(\sup_{\theta\in B(\theta_0, n^{-\frac{1}{2}}\rho)} \big|Z_n(\theta) - n^{\frac{1}{2}}(\theta-\theta_0)^*G_n\big|\Big) \xrightarrow[n\to\infty]{} 0 \tag{I.10}$$

for some sequence $G_n = (G_{ni})_{i=1,\dots,d}$ of d-dimensional random variables converging in distribution to $\mathcal{N}_d(0,\mathcal{J})$. Now let G_{ni} be given by

$$G_{ni} = \tfrac{1}{2}n^{-\frac{1}{2}}\,\mathrm{tr}\,\big[T_n(\varphi_{\theta_0}^{-1}\partial_i\varphi_{\theta_0})T_n(\varphi)^{-1}\big(\mathbb{X}_n\mathbb{X}_n^* - T_n(\varphi)\big)\big].$$

By the product formula for cumulants the joint cumulant of $G_{ni_1},\dots,G_{ni_k}$ is given by

$$\frac{n^{-\frac{k}{2}}}{2k}\sum_{\sigma\in S_k}\mathrm{tr}\,\Big(\prod_{j=1}^{k}T_n(\varphi_{\theta_0}\partial_{i_{\sigma(j)}}\varphi_{\theta_0})\Big),$$

where S_k is the set of permutations of $\{1,\dots,k\}$. For $k=2$ use lemma I.18 given below, and for $k \geq 3$ use the standard inequalities, to see that these cumulants converge to the corresponding cumulants of $\mathcal{N}_d(0,\mathcal{J})$, whereby the stated convergence in distribution follows. The property (I.10) is proved by techniques similar to those used in the proof of lemma I.11. This concludes the proof of theorem I.6.

(I.16)

I.4 Matrix approximations and matrix inequalities

In this section we prove the properties of Toeplitz matrices which are used in the proof of theorem I.6. The first inequality provides an easy method of handling inverse Toeplitz matrices.

Lemma I.12 *If $f, f^{-1} \in L^1_+$, then $T_n(f)^{-1} \leq T_n(f^{-1})$ in the partial ordering of positive semi definite matrices. Especially $\|T_n(f)^{-1}A\|_p \leq \|T_n(f^{-1})A\|_p$ for every $p \in [1,\infty]$ and $A \in \mathbb{M}_1$.*

Proof The proof follows Shaman (1976, theorem 2.1). Since $\log\det f(x) \leq \|f(x) + f^{-1}(x)\|_1$, it follows that $\log\det f \in L^1$. By Hannan (1970, lemma III.5.3) there thus exist coefficient matrices $\varphi_j, \psi_j \in \mathbb{M}_1$ such that

$$\sum_{j=0}^{\infty} \|\varphi_j\|_2^2 < \infty, \qquad \sum_{j=0}^{\infty} \|\psi_j\|_2^2 < \infty,$$

and for the holomorphic functions $g(z) = \sum_{j=0}^{\infty} \varphi_j z^j$, $h(z) = \sum_{j=0}^{\infty} \psi_j z^j$ defined on the open unit disk

$$f(x) = \lim_{r\uparrow 1} g(re^{2\pi i x})^* g(re^{2\pi i x}), \qquad f^{-1}(x) = \lim_{r\uparrow 1} h(re^{2\pi i x})^* h(re^{2\pi i x})$$

holds for almost all $x \in (-\frac{1}{2}, \frac{1}{2}]$. Since

$$I_1 = f(x) f^{-1}(x) = \lim_{r\uparrow 1} g(re^{2\pi i x})^* g(re^{2\pi i x}) h(re^{2\pi i x})^* h(re^{2\pi i x}),$$

it follows by the maximum modulus principle that

$$g(z)h(z)^* = \sum_{j=0}^{\infty} \Big(\sum_{l=0}^{j} \varphi_{j-l}\psi_l^* \Big) z^j = I_1.$$

Thus if $B_n \in \mathbb{M}_n$ has identity matrices on the diagonal directly below the main diagonal and 0's elsewhere, $\Phi_n = \sum_{j=0}^{n-1} \varphi_j B_n^j$ and $\Psi_n = \sum_{j=0}^{n-1} \psi_j (B_n^*)^j$, then

$$\Phi_n \Psi_n^* = \sum_{j=0}^{n-1} \Big(\sum_{l=0}^{j} \varphi_{j-l}\psi_l^* \Big) B_n^j = I_n. \tag{I.11}$$

Now let ε_t, $t \in \mathbb{Z}$, be independent $\mathcal{N}_b(0, I_1)$ distributed random variables and define $\tilde{Y}_t = \sum_{j=0}^{t-1} \varphi_j \varepsilon_{t-j}$, $Y_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}$. Since the covariance matrices of the n-dimensional random variables $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ and $(Y_1, \ldots, Y_n)$ are given by

$$\operatorname{Var}\big(\tilde{Y}_1, \ldots, \tilde{Y}_n\big) = \Phi_n \Phi_n^*, \qquad \operatorname{Var}\big(Y_1, \ldots, Y_n\big) = T_n(f),$$

we see that $\Phi_n \Phi_n^* \leq T_n(f)$. A similar argument yields $\Psi_n \Psi_n^* \leq T_n(f^{-1})$, whence

$$T_n(f)^{-1} \leq (\Phi_n \Phi_n^*)^{-1} = \Psi_n \Psi_n^* \leq T_n(f^{-1})$$

by equation (I.11). □

To proceed further we need to approximate the involved Toeplitz matrices by band matrices. In order to do this we approximate the spectral densities by functions with finitely many Fourier coefficients.

Definition I.13 The m'th, $m \in \mathbb{N}$, Fejér approximation $F_m(f) \in L^\infty$ of the matrix valued function $f \in L^1$ is defined as the mean of the m first Fourier approximations, and can by computed by

$$F_m(f)(x) = \int f(y) K_m(x-y)\,\mathrm{d}y, \qquad K_m(y) = \frac{1}{m}\frac{\sin^2(m\pi y)}{\sin^2(\pi y)}.$$

The mathematical reason for using Fejér approximations instead of Fourier approximation is, that the Fejér kernel $K_m(y)$ is a probability density. Especially

$$\|F_m(f)\|_p \le \|f\|_p$$

by the triangular inequality. Moreover, $K_m(y)$ satisfies the bound

$$K_m(y) \le \tfrac{1}{4} m^{-1} y^{-2} \sin^2(m\pi y), \quad |y| \le \tfrac{1}{2}.$$

The following two lemmas concern the quality of Fejér approximations. Remember that $\zeta_\beta \in L^1$, $\beta > -1$, is defined by $\zeta_\beta(x) = |x|^\beta I_1$.

Lemma I.14 *If $f \in L^p$ and $\|f - f(\cdot - y)\|_p \le c_1\ |y|^\alpha$ for some $c_1 < \infty$, $\alpha \in (0,1)$, then $\|f - F_m(f)\|_p \le c_2\, m^{-\alpha}$, where the constant $c_2 < \infty$ depend on c_1 and α only.*

Proof Let the function $h(y)$ and the normalizing constant $\gamma > 0$ be given by

$$h(y) = \|f - f(\cdot - y)\|_p \le c_1\ |y|^\alpha, \qquad \gamma = \int h(y) K_m(y)\,\mathrm{d}y.$$

If $p < \infty$, then by the triangular inequality, the Jensen inequality and the Tonelli theorem $\|f - F_m(f)\|_p$ is bounded by

$$\begin{aligned}
&\Big(\int \Big(\int \|f(x) - f(x-y)\|_p\ K_m(y)\,\mathrm{d}y\Big)^p \mathrm{d}x\Big)^{p^{-1}} \\
&\quad\le \Big(\int\int \|f(x) - f(x-y)\|_p^p \Big(\frac{h(y)}{\gamma}\Big)^{1-p} K_m(y)\,\mathrm{d}y\,\mathrm{d}x\Big)^{p^{-1}} \\
&\quad= \Big(\int \gamma^p \frac{h(y)}{\gamma} K_m(y)\,\mathrm{d}y\Big)^{p^{-1}} = \int h(y) K_m(y)\,\mathrm{d}y \\
&\quad\le \tfrac{1}{2} c_1\, m^{-1} \int_0^{\frac{1}{2}} |y|^{\alpha-2} \sin^2(m\pi y)\,\mathrm{d}y \le \underbrace{\tfrac{1}{2} c_1 \int_0^\infty u^{\alpha-2}\sin^2(\pi u)\,\mathrm{d}u}_{c_2}\ m^{-\alpha}.
\end{aligned}$$

The inequality follows similarly for $p = \infty$. ☐

(I.18)

Lemma I.15 *For every $\alpha \in (0,1)$ there exists $c_\alpha < \infty$ such that*

$$\int |y|^{-\alpha} K_m(x-y)\,\mathrm{d}y \le \begin{cases} c_\alpha m^\alpha & \text{if } |x| \le m^{-1}, \\ c_\alpha m^{-1}\log(m) x^{-1-\alpha} & \text{if } m^{-1} < |x| \le \frac{1}{2}. \end{cases}$$

Especially $\|\zeta_1 F_m(\zeta_{-\beta})\|_\infty$ is bounded in $m \in \mathbb{N}$ for every $\beta < 1$.

Proof Denote the integral by $g(x)$, and observe that g is an even function and decreasing for $x \in [0, \frac{1}{2}]$. Hence,

$$g(x) \le g(0) \le \tfrac{1}{2} m^{-1} \int_0^{\frac{1}{2}} y^{-2-\alpha} \sin^2(m\pi y)\,\mathrm{d}y \le \tfrac{1}{2} m^\alpha \int_0^\infty u^{-2-\alpha}\sin^2(\pi u)\,\mathrm{d}u.$$

Moreover if $x = m^{-\gamma} > m^{-1}$, $\gamma \in (0,1)$, then $g(x)$ is bounded by

$$2\int_0^{m^{-\gamma}} (m^{-\gamma} - y)^{-\alpha} K_m(y)\,\mathrm{d}y + 2\int_0^{\frac{1}{2}-m^{-\gamma}} y^{-\alpha} K_m(y + m^{-\gamma})\,\mathrm{d}y. \qquad \text{(I.12)}$$

The first term in (I.12) is bounded by

$$\begin{aligned} &\tfrac{1}{2} m^{-1} \int_0^{m^{-\gamma}} (m^{-\gamma} - y)^{-\alpha} y^{-2} \sin^2(m\pi y)\,\mathrm{d}y \\ &\quad = \tfrac{1}{2} m^{\gamma(1+\alpha)-1} \int_0^1 (1-v)^{-\alpha} v^{-2} \sin^2(m^{1-\gamma}\pi v)\,dv \\ &\quad \le \tfrac{1}{2} m^{\gamma(1+\alpha)-1} \int_0^{\frac{1}{2}} 2^\alpha v^{-2} \sin^2(m^{1-\gamma}\pi v)\,dv + 2m^{\gamma(1+\alpha)-1} \int_{\frac{1}{2}}^1 (1-v)^{-\alpha}\,dv \\ &\quad \le 2^{\alpha-1} m^{\alpha\gamma} \int_0^\infty u^{-2}\sin^2(\pi u)\,\mathrm{d}u + 2^\alpha (1-\alpha)^{-1} m^{\gamma(1+\alpha)-1}. \end{aligned}$$

The second term in (I.12) is bounded by

$$\begin{aligned} &\tfrac{1}{2} m^{-1} \int_0^{\frac{1}{2}} y^{-\alpha} (y + m^{-\gamma})^{-2} \sin^2(m\pi(y + m^{-\gamma}))\,\mathrm{d}y \\ &\quad \le \tfrac{1}{2} m^{-1} \int_0^{m^{-\gamma}} y^{-\alpha} m^{2\gamma}\,\mathrm{d}y + \tfrac{1}{2} m^{-1} \int_{m^{-\gamma}}^{\frac{1}{2}} y^{-1} (y + m^{-\gamma})^{-1-\alpha}\,\mathrm{d}y \\ &\quad \le \tfrac{1}{2}(1-\alpha)^{-1} m^{\gamma(1+\alpha)-1} + 2^{-2-\alpha}\log(m) m^{\gamma(1+\alpha)-1}. \end{aligned}$$

Combining these bounds and defining

$$c_\alpha = 2^{\alpha-1}\int_0^\infty u^{-2}\sin^2(\pi u)\,\mathrm{d}u + 2^\alpha(1-\alpha)^{-1} + 2^{-1}(1-\alpha)^{-1} + 2^{-2-\alpha}$$

we see that $g(x) = g(m^{-\gamma}) \le c_\alpha m^{-1}\log(m) x^{-1-\alpha}$. □

In the next two lemmas we establish inequalities between p-norms of Toeplitz matrices and L^p-norms of functions. For $f \in L^1$ let

$$T(f) = \left(\hat{f}(i-j)\right)_{i,j\in\mathbb{Z}} \in \mathbb{M}_\infty$$

be the associated infinite dimensional Toeplitz matrix. For given $m \leq n$ let $P_n \in \mathbb{M}_\infty$, $Q_{nm} \in \mathbb{M}_n$ be diagonal matrices with n consecutive, respectively m not necessarily consecutive, identity matrices on the diagonal and 0's elsewhere, and let $Q_m \in \mathbb{M}_\infty$ be the extension of Q_{nm} constructed by augmenting Q_{nm} with 0's such that $Q_m \leq P_n$ in the partial ordering of positive semi definite matrices. Then the equalities

$$\|T_n(f)\|_p = \|P_n T(f) P_n\|_p\,, \qquad \|T_n(f)Q_{nm}\|_p = \|P_n T(f) Q_m\|_p$$

between p-norms on $\mathbb{M}_n$ respectively $\mathbb{M}_\infty$ hold true.

Lemma I.16 *If $f \in L^p$, $p \geq 2$, then $\|T(f)Q_m\|_p \leq m^{p^{-1}} \|f\|_p$. If $f \in L^p$, $p < 2$, then $\|T_n(f)Q_{nm}\|_p \leq n^{p^{-1}-\frac{1}{2}} m^{\frac{1}{2}} \|f\|_p$.*

Proof This proof follows Avram (1988). The statement holds for $p = 2$ and $p = \infty$ since

$$\|T(f)Q_m\|_2^2 = \operatorname{tr}\left(Q_m T(f^* f) Q_m\right) = m\,\|f\|_2^2\,, \qquad \|T(f)\|_\infty = \|f\|_\infty\,,$$

and hence also for $p \in (2,\infty)$ by the Riesz-Thorin interpolation theorem, see Bergh & Löfström (1976). If $p < 2$, let $g \in L^{(p^{-1}-\frac{1}{2})^{-1}}$, $h \in L^2$ be given by

$$g(x) = |f(x)|^{1-\frac{p}{2}}\,, \qquad h(x) = |f(x)|^{\frac{p}{2}-1} f(x).$$

Then $f(x) = g(x)h(x)$, $\|f\|_p = \|g\|_{(p^{-1}-\frac{1}{2})^{-1}} \|h\|_2$ and hence

$$\begin{aligned} \|T_n(f)Q_{nm}\|_p &= \|P_n T(g) T(h) Q_m\|_p \leq \|P_n T(g)\|_{(p^{-1}-\frac{1}{2})^{-1}} \|T(h)Q_m\|_2 \\ &\leq n^{p^{-1}-\frac{1}{2}} \|g\|_{(p^{-1}-\frac{1}{2})^{-1}}\, m^{\frac{1}{2}} \|h\|_2 = n^{p^{-1}-\frac{1}{2}} m^{\frac{1}{2}} \|f\|_p\,. \end{aligned}$$

□

Lemma I.17 *If $f \in L^p$, $p < 2$, then*

$$\left\|T_n(\zeta_{\frac{1}{2}})T_n(f)Q_{nm}\right\|_2^2 \leq n^{p^{-1}-\frac{1}{2}} m^{\frac{3}{2}-p^{-1}} \|\zeta_1 f\|_{(1-p^{-1})^{-1}} \|f\|_p + \log(n)\, m\, \|f\|_1^2\,.$$

Proof The squared norm $\|T_n(\zeta_{\frac{1}{2}})T_n(f)Q_{nm}\|_2^2$ is bounded by

$$\operatorname{tr}\left(Q_m T(f^*) P_n T(\zeta_1 f) Q_m\right) - \operatorname{tr}\left(Q_m T(f^*) P_n T(\zeta_1)(I - P_n) T(f) Q_m\right).$$

By the Hölder inequality and lemma I.16 the first term is bounded by

$$n^{p^{-1}-\frac{1}{2}} m^{\frac{3}{2}-p^{-1}} \|\zeta_1 f\|_{(1-p^{-1})^{-1}} \|f\|_p\,,$$

(I.20)

and since $\|\hat{f}(k)\|_2 \leq \|\hat{f}(k)\|_1 \leq \|f\|_1$, and $\|\hat{\zeta}_1(k)\|_\infty = -\pi^{-2}k^{-2}$ for $k \neq 0$ and even and vanish for k odd, the second term is bounded by

$$\sum_{\substack{j\in\{1,\ldots,n\}\\ k\in\mathbb{Z}\setminus\{1,\ldots,n\}}} m\left\|\hat{\zeta}_1(j-k)\right\|_\infty \|f\|_1^2 \leq 2m\sum_{k=1}^\infty \frac{n\wedge 2k}{4\pi k^2}\|f\|_1^2 \leq \log(n)m\|f\|_1^2.$$

□

If $f \in L^p$ and $g \in L^1_+$, then inequality (I.7) follows by the calculations

$$\begin{aligned} n^{-\frac{1}{2}}\left\|T_n(g)^{-1}T_n(f)\right\|_2 &\leq n^{-\frac{1}{2}}\left\|T_n(g^{-1})T_n(f)\right\|_2 \\ &\leq \left\|\zeta_{-\frac{1}{2}}g^{-1}\right\|_\infty n^{-\frac{1}{2}}\left\|T_n(\zeta_{\frac{1}{2}})T_n(f)\right\|_2 \\ &\leq \left\|\zeta_{-\frac{1}{2}}g^{-1}\right\|_\infty\Big(\|\zeta_1 f\|_{(1-p^{-1})^{-1}}^{\frac{1}{2}}\|f\|_p^{\frac{1}{2}} + \sqrt{\log n}\,\|f\|_1\Big), \end{aligned}$$

where we have used lemma I.12, lemma I.17 and the inequality $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$.

The next two lemmas concern the approximation of the product of Toeplitz matrices by the Toeplitz matrix of the product. For $2m \leq n$ let $A_{nm} \in \mathbb{M}_n$ be the diagonal matrix with identity matrices on the middle $(n-2m)$ diagonal elements and 0's elsewhere. If $f, g, gf \in L^1$, then the matrix equations

$$T(g)T(f) = T(gf), \qquad T_n\big(gF_m(f)\big)A_{nm} = T_n(g)T_n\big(F_m(f)\big)A_{nm} \tag{I.13}$$

hold. The first equation in (I.13) is well known. Since the matrix $T_n\big(F_m(f)\big)$ has 0's outside the $(m-1)$ first side diagonals the $(n-2m)$ middle columns of the matrices $T_n(g)T_n\big(F_m(f)\big)$ and $T_n\big(gF_m(f)\big)$ coincide, whence the second equation in (I.13) holds.

Lemma I.18 *If $f \in L^p$, $g \in L^q$, then $n^{-p^{-1}-q^{-1}}\|T_n(g)T_n(f) - T_n(gf)\|_{(p^{-1}+q^{-1})^{-1}}$ is bounded by*

$$\min_{m\in\mathbb{N}:2m\leq n}\big(2\|f - F_m(f)\|_p\|g\|_q + 2\,n^{-(\frac{1}{2}\wedge p^{-1})}(2m)^{\frac{1}{2}\wedge p^{-1}}\|f\|_p\|g\|_q\big).$$

Proof By equation (I.13) the matrix $T_n(g)T_n(f) - T_n(gf)$ is equal to

$$\begin{aligned} &T_n(g)T_n\big(f - F_m(f)\big)A_{nm} - T_n\big(gf - gF_m(f)\big)A_{nm} \\ &+T_n(g)T_n(f)(I_n - A_{nm}) - T_n(gf)(I_n - A_{nm}). \end{aligned} \tag{I.14}$$

The stated bound thus follows by the Hölder inequalities and lemma I.16. □

Lemma I.19 *If $f \in L^1$ and*

$$\|f - f(\cdot - y)\|_1 \leq c_1\,|y|^\alpha, \qquad \left\|\zeta_{-\frac{1}{2}}f^{-1}\right\|_\infty < \infty, \qquad \|\zeta_\beta f\|_\infty < \infty$$

for some $c_1 < \infty$, $\alpha > 0$ and $\beta < 1$, then

$$n^{-\frac{1}{2}}\left\|T_n(f^{-1})T_n(f) - I_n\right\|_2 \leq c_2\,n^{-\frac{\alpha}{2+4\alpha}},$$

where the constant $c_2 < \infty$ depend on c_1, $\left\|\zeta_{-\frac{1}{2}}f^{-1}\right\|_\infty$ and $\|\zeta_\beta f\|_\infty$ only.

(I.21)

Proof By equation (I.14) and lemma I.15 the quantity $n^{-\frac{1}{2}} \|T_n(f^{-1})T_n(f) - I_n\|_2$ is bounded by $a_1 + a_2 + a_3 + a_4$, where

$$\begin{aligned} a_1^2 &= n^{-1} \left\|T_n(f^{-1})T_n(f - F_m(f))\right\|_2^2 \\ &\leq \left\|\zeta_{-\frac{1}{2}} f^{-1}\right\|_\infty^2 \big(\|\zeta_1 f\|_\infty + \|\zeta_1 F_m(\zeta_{-\beta})\|_\infty \|\zeta_\beta f\|_\infty \big) \|f - F_m(f)\|_1 \\ &\qquad + \log(n) \left\|\zeta_{-\frac{1}{2}} f^{-1}\right\|_\infty^2 \|f - F_m(f)\|_1^2, \end{aligned}$$

$$\begin{aligned} a_2^2 &= n^{-1} \left\|T_n(I_1 - f^{-1}F_m(f))\right\|_2^2 \leq \left\|f^{-1}F_m(f) - I_1\right\|_2^2 \\ &= \left\|F_m(f^*)(f^{-1})^* f^{-1} F_m(f) - F_m(f^*)(f^{-1})^* - f^{-1}F_m(f) + I_1\right\|_1 \\ &\leq \left\|F_m(f^*)(f^{-1})^* f^{-1}\right\|_\infty \|F_m(f) - f\|_1 + \left\|f^{-1}\right\|_\infty \|F_m(f) - f\|_1 \\ &\leq \big(\left\|\zeta_{-\frac{1}{2}} f^{-1}\right\|_\infty^2 \|\zeta_\beta f\|_\infty \|\zeta_1 F_m(\zeta_{-\beta})\|_\infty + \left\|f^{-1}\right\|_\infty\big) \|f - F_m(f)\|_1, \end{aligned}$$

$$\begin{aligned} a_3^2 &= n^{-1} \left\|T_n(f^{-1})T_n(f)(I_n - A_{nm})\right\|_2^2 \\ &\leq n^{-\frac{1}{2}} (2m)^{\frac{1}{2}} \left\|\zeta_{-\frac{1}{2}} f^{-1}\right\|_\infty^2 \|\zeta_1 f\|_\infty \|f\|_1 + 2n^{-1} \log(n)\, m \left\|\zeta_{-\frac{1}{2}} f^{-1}\right\|_\infty^2 \|f\|_1, \end{aligned}$$

$$a_4^2 = n^{-1} \|I_n - A_{nm}\|_2^2 \leq 2n^{-1} m.$$

Using lemma I.14 and lemma I.15 we see that $\|f - F_m(f)\|_1 = O(m^{-\alpha})$ and $\|\zeta_1 F_m(\zeta_{-\beta})\|_\infty = O(1)$. Thus if $m \approx n^{(1+2\alpha)^{-1}}$, then a_1, a_2, a_3, a_4 all are of order $O(n^{-\frac{\alpha}{2+4\alpha}})$. □

We are now in a position to prove lemma I.8. Thus suppose assumption I.3 holds. The matrix $T_n(\varphi_\theta)^{-1}T_n(\varphi_{\theta_0}) - T_n(\varphi_\theta^{-1}\varphi_{\theta_0})$ is equal to

$$\big(I_n - T_n(\varphi_\theta^{-1})T_n(\varphi_\theta)\big)T_n(\varphi_\theta)^{-1}T_n(\varphi_{\theta_0}) + \Big(T_n(\varphi_\theta^{-1})T_n(\varphi_{\theta_0}) - T_n(\varphi_\theta^{-1}\varphi_{\theta_0})\Big). \quad \text{(I.15)}$$

Using lemma I.18 we see that the normalized 1-norm of the second term in (I.15), *i.e.* the quantity $n^{-1} \left\|T_n(\varphi_\theta^{-1})T_n(\varphi_{\theta_0}) - T_n(\varphi_\theta^{-1}\varphi_{\theta_0})\right\|_1$, is bounded by the minimum over $m \in \mathbb{N}$, $2m \leq n$, of

$$2 \|\varphi_{n,\theta_0} - F_m(\varphi_{n,\theta_0})\|_{(1-q^{-1})^{-1}} \left\|\varphi_{n,\theta}^{-1}\right\|_q + 2\, n^{-\frac{1}{2}} (2m)^{\frac{1}{2}} \|\varphi_{n,\theta_0}\|_{(1-q^{-1})^{-1}} \left\|\varphi_{n,\theta}^{-1}\right\|_q.$$

If $p^{-1} + q^{-1} + r^{-1} \leq \frac{1}{2}$, then use lemma I.18, lemma I.12 and lemma I.16 to bound the normalized 1-norm of first term in (I.15) by the minimum over $m \in \mathbb{N}$, $2m \leq n$, of

$$\begin{aligned} &2 \|\varphi_{n,\theta} - F_m(\varphi_{n,\theta})\|_{(1-p^{-1}-2q^{-1})^{-1}} \left\|\varphi_{n,\theta}^{-1}\right\|_q^2 \|\varphi_{n,\theta_0}\|_p \\ &+2\, n^{-(\frac{1}{2}\wedge(1-p^{-1}-2q^{-1}))} (2m)^{\frac{1}{2}\wedge(1-p^{-1}-2q^{-1})} \|\varphi_{n,\theta}\|_{(1-p^{-1}-2q^{-1})^{-1}} \left\|\varphi_{n,\theta}^{-1}\right\|_q^2 \|\varphi_{n,\theta_0}\|_p, \end{aligned}$$

(I.22)

and if $p < 2$, then use lemma I.19 and inequality (I.7) to bound the normalized 1-norm of first term in (I.15) by

$$c_2\, n^{-\frac{\alpha}{2+4\alpha}} \left\|\zeta_{-\frac{1}{2}}\varphi_{n,\theta}^{-1}\right\|_\infty \Big(\|\zeta_1\varphi_{n,\theta_0}\|_{(1-p^{-1})^{-1}}^{\frac{1}{2}} \; \|\varphi_{n,\theta_0}\|_p^{\frac{1}{2}} + \sqrt{\log n}\; \|\varphi_{n,\theta_0}\|_1 \Big).$$

Using lemma I.14 we see that all these bounds are of order $O(n^{-\frac{\alpha}{2+4\alpha}}\log n)$ when $m \approx n^{(1+4\alpha)^{-1}}$. To prove the second part of lemma I.8 we use equation (I.13) to rewrite the matrix

$$T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta) - T_n(\varphi_\theta^{-1}\partial_i\varphi_\theta)$$

as

$$\begin{aligned} T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta)(I_n - A_{nm}) + T_n(\varphi_\theta)^{-1}T_n\big(\partial_i\varphi_\theta - \varphi_\theta F_m(\varphi_\theta^{-1}\partial_i\varphi_\theta)\big)A_{nm} \\ -T_n\big(\varphi_\theta^{-1}\partial_i\varphi_\theta - F_m(\varphi_\theta^{-1}\partial_i\varphi_\theta)\big)A_{nm} - T_n(\varphi_\theta^{-1}\partial_i\varphi_\theta)(I_n - A_{nm}). \end{aligned} \tag{I.16}$$

The needed regularity conditions and the resulting rates of convergence depend on the bounds for the two first terms in (I.16) only. If $p^{-1}+q^{-1}+r^{-1} \le \frac{1}{2}$, then $n^{-\frac{1}{2}}\,\|T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta)(I_n - A_{nm})\|_2$ is bounded by

$$n^{-\frac{1}{2}+q^{-1}}(2m)^{\frac{1}{2}-q^{-1}}\left\|\varphi_{n,\theta}^{-1}\right\|_q \; \|\partial_i\varphi_{n,\theta}\|_{(\frac{1}{2}-q^{-1})^{-1}},$$

and $n^{-\frac{1}{2}}\left\|T_n(\varphi_\theta)^{-1}T_n\big(\partial_i\varphi_\theta - \varphi_\theta F_m(\varphi_\theta^{-1}\partial_i\varphi_\theta)\big)\right\|_2$ is bounded by

$$\left\|\varphi_{n,\theta}^{-1}\right\|_q \; \|\varphi_{n,\theta}\|_p \; \left\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta} - F_m(\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta})\right\|_{(\frac{1}{2}-p^{-1}-q^{-1})^{-1}}.$$

Using these bounds we see that for $m \approx n^{(1+(\frac{1}{2}-q^{-1})^{-1}\alpha)^{-1}}$ all four terms in (I.16) are of order $O(n^{-\alpha(1+(\frac{1}{2}-q^{-1})^{-1}\alpha)^{-1}})$. This is the exact point in the analysis where $q > 2$ and $p^{-1}+q^{-1}+r^{-1} \le \frac{1}{2}$ is needed. Otherwise if $p < 2$, then

$$n^{-1}\left\|T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta)(I_n - A_{nm})\right\|_2^2$$

is bounded by

$$\begin{gathered} n^{p^{-1}-\frac{3}{2}}\,(2m)^{\frac{3}{2}-p^{-1}}\left\|\zeta_{-\frac{1}{2}}\varphi_{n,\theta}^{-1}\right\|_\infty^2 \; \|\zeta_1\partial_i\varphi_{n,\theta}\|_{(1-p^{-1})^{-1}} \; \|\partial_i\varphi_{n,\theta}\|_p \\ +n^{-1}\,\log(n)\,(2m)\left\|\zeta_{-\frac{1}{2}}\varphi_{n,\theta}^{-1}\right\|_\infty^2 \; \|\partial_i\varphi_{n,\theta}\|_1^2, \end{gathered}$$

and $n^{-1}\left\|T_n(\varphi_\theta)^{-1}T_n(\partial_i\varphi_\theta - \varphi_\theta F_m(\varphi_\theta^{-1}\partial_i\varphi_\theta))\right\|_2^2$ is bounded by

$$\begin{gathered} \left\|\zeta_{-\frac{1}{2}}\varphi_{n,\theta}^{-1}\right\|_\infty^2 \; \|\zeta_1\varphi_{n,\theta}\|_\infty \; \|\varphi_{n,\theta}\|_p \; \left\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta} - F_m(\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta})\right\|_{2(1-p^{-1})^{-1}}^2 \\ + \log(n)\left\|\zeta_{-\frac{1}{2}}\varphi_{n,\theta}^{-1}\right\|_\infty^2 \; \|\varphi_{n,\theta}\|_p^2 \; \left\|\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta} - F_m(\varphi_{n,\theta}^{-1}\partial_i\varphi_{n,\theta})\right\|_{(1-p^{-1})^{-1}}^2. \end{gathered}$$

Thus if $m \approx n^{(1+4\alpha)^{-1}}$, then all terms in (I.16) are of order $O(n^{-\frac{\alpha}{1+4\alpha}}\sqrt{\log n})$.

II

Likelihood inference for a stochastic partial differential equation observed at discrete points in time and space

Abstract

Both parabolic and hyperbolic stochastic partial differential equations in one-dimensional space driven by Gaussian noise have been proposed as models for the term structure of interest rates. In the first part of this paper we solve these equations via the familiar separation of variables technique. The associated coefficient processes are found to be Ornstein-Uhlenbeck processes in the parabolic case, and the first components of two-dimensional Ornstein-Uhlenbeck processes in the hyperbolic case. Moreover, the sample paths properties of these equations are studied. In the parabolic case the sample paths are essentially found to be Hölder continuous of order $\frac{1}{2}$ in space and $\frac{1}{4}$ in time, whereas in the hyperbolic case the sample paths are essentially found to be Hölder continuous of order $\frac{1}{2}$ simultaneously in time and space. In the second part of the paper we consider likelihood inference for the parameters in the equation given an observation at discrete lattice points in time and space. The associated infinite dimensional state space model is described, and a finite dimensional approximation is proposed. We present conditions under which the resulting approximate maximum likelihood estimator is asymptotically efficient when the spatial resolution is fixed and the number of observations in time increases to infinity at a fixed time step. Moreover, the asymptotical distribution of the approximative likelihood ratio test for a parabolic equation against the hyperbolic alternative is found to be a truncated chi-square distribution.

Key words

Stochastic partial differential equation, parabolic equation, hyperbolic equation, sample paths properties, discrete observations, approximate likelihood inference, asymptotic efficiency, likelihood ratio test for parabolic equation.

II.1 Introduction

Motivated by applications in neurophysiology, hydrology, oceanography and finance statistical inference problems for stochastic partial differential equations has attained increasingly interest in the recent years, see Huebner & Rozovskii (1995), Piterbarg & Rozovskii (1997), Huebner (1997), Huebner & Lototsky (2000). In the cited papers an observation of the N'th Galerkin approximation of a parabolic evolution equation, *i.e.* the N first Fourier coefficient processes, is assumed to be available at either a continuous interval of time, discrete time points or just at a single time point. The asymptotic properties of the maximum likelihood estimator when the spatial resolution tends towards a continuous observation in space, *i.e.* when $N \to \infty$, is studied and shown to depend heavily on which coefficients are unknown in the stochastic partial differential equation. The purpose of the present paper is to propose an approximate likelihood and study the asymptotic properties of the associated maximum likelihood estimator for the parameter $\theta = (\eta_1, \eta_2, \xi_0, \xi_1, \xi_2)$ given observations at discrete points in time and space of the stationary solution of the parabolic ($\eta_2 = 0$, $\xi_2 > 0$) or hyperbolic ($\eta_2 > 0$, $\xi_2 > 0$) stochastic partial differential equation

$$\eta_2 \frac{\partial^2}{\partial t^2} V(t,x) + \eta_1 \frac{\partial}{\partial t} V(t,x) = \xi_0 V(t,x) + \xi_1 \frac{\partial}{\partial x} V(t,x) + \xi_2 \frac{\partial}{\partial x^2} V(t,x) + W_\xi(t,x), \tag{II.1a}$$

$t \in \mathbb{R}$, $0 < x < 1$, with Dirichlet boundary conditions

$$V(t,0) = V(t,1) = 0, \quad t \in \mathbb{R}. \tag{II.1b}$$

Here the parameters satisfies $\eta_1, \xi_2 > 0$, $\eta_2 \geq 0$ and the stochastic disturbance term $W_\xi(t,x)$ is related to Brownian white noise $W(t,x)$ via the equation

$$W_\xi(t,x) = e^{-\frac{\xi_1}{2\xi_2} x} W(t,x). \tag{II.2}$$

The motivation for studying this statistical problem is applications in mathematical finance to the modeling of the term structure for bonds of different maturity times, see Cont (1998), Santa-Clara & Sornette (1999). In these models the spatial component represents time to maturity. In Cont (1998) it is argued that the short rate ($x = 0$) and the long rate ($x = 1$) can be modeled independently of the profile from the short rate to the long rate, and that the deviation from the average profile can be modeled by the stochastic partial differential equation (II.1). Realistic data thus consist of observations at discrete points in time and space organized in a lattice. The spatial resolution is usually fairly low consisting of between 10 and 20 maturity times. Calculating the discrete Fourier transforms and using the Galerkin approximation would thus be inadequate and result in biased estimates. But since the Galerkin approximation approach yields estimates on closed forms this might be a good way to get preliminary estimates. The solutions to the parabolic and hyperbolic equations have different properties, see

(II.2)

the discussion in Cont (1998), whence it is of interest to test the hypothesis of a parabolic equation against a hyperbolic equation. Since the parabolic hypothesis lies on a one-sided boundary of the parameter space the likelihood ratio test will be asymptotically distributed as the mixture of the point measure in 0 and the chi-square distribution.

The paper is organized as follows. In section II.2 we describe for which parameters there exists a stationary solution to (II.1), give a representation of the stationary solution, and describe the sample path properties of the solution. In section II.3 we propose an approximate likelihood and give conditions under which this approximate likelihood has the same first order asymptotic properties as the exact likelihood. We show that the approximate likelihood satisfies a uniform version of the LAN-property, whence the approximate maximum likelihood estimate especially is asymptotically efficient in the sense of Hájek and Le Cam, see Le Cam & Yang (2000). Moreover we derive the likelihood ratio test for a parabolic equation against a hyperbolic equation.

Wherever possible we have tried to avoid using the explicit expression of the Greens function given by (II.4) and (II.5), and have instead used the associated differential equations (II.3). Most of the techniques used in this paper thus generalizes to stochastic partial differential equations, where higher order derivative w.r.t. time are included on the left hand side of (II.1a). Similarly, it is possible to alter the boundary conditions (II.1b) within the presented framework.

II.2 The stationary solution and its properties

The solution $V(t, x)$ to the stochastic partial differential equation (II.1) is indexed by the time parameter $t \in \mathbb{R}$ and the spatial coordinate $x \in (0, 1)$. Since the spatial coordinate is one dimensional, there is reason to believe, that $V(t, x)$ exists as an ordinary stochastic process defined on a sufficient large probability space $(\Omega, \mathscr{A}, \mathrm{P})$, *i.e.* as a measurable function

$$V : \Omega \times \mathbb{R} \times (0, 1) \to \mathbb{R}.$$

If this is the case, then the stochastic partial differential equation (II.1) can be posed and solved in both the Itô type framework of Walsh (1986) and the white noise calculus of Holden et al. (1996), and the solutions of the two different approaches will coincide. Since the equation (II.1) contains a second order derivative w.r.t. time, it is most easily solved in the white noise calculus. However, the Itô calculus is more easily interpretable and probably more familiar to most readers. We will thus use the white noise calculus to solve the equation, and then afterwards rewrite the solution as an Itô-integral w.r.t. a Brownian sheet.

Suppose the partial differential equation (II.1) is either parabolic or hyperbolic, *i.e.* $\eta_2 \geq 0$ and $\xi_2 > 0$. Then there exists a Greens function $G(t, x, y)$ for the partial

(II.3)

differential equation given by

$$G(t,x,y) = \sum_{k=1}^{\infty} T_k(t)\, X_k(x)\, X_k(y),$$

which can be found using the familiar separation of variables technique. Thus $X_k(x)$ and λ_k, $k \in \mathbb{N}$, are the eigenfunctions respectively eigenvalues of the differential operator $\xi_0 + \xi_1 \frac{\partial}{\partial x} + \xi_2 \frac{\partial^2}{\partial x^2}$ with boundary conditions $X(0) = X(1) = 0$. Moreover, in the parabolic case $T_k(t)$ is the solution of the differential equation

$$\eta_1 T'(t) = \lambda_k T(t), \qquad T(0) = 1,$$

and in the hyperbolic case $T_k(t)$ is the solution of the differential equation

$$\eta_2 T''(t) + \eta_1 T'(t) = \lambda_k T(t), \qquad T(0) = 0, \qquad T'(0) = 1. \tag{II.3}$$

Solving these differential equations we find that $X_k(x)$ and λ_k are given by

$$X_k(x) = \sqrt{2}\sin(\pi k x) e^{-\frac{\xi_1}{2\xi_2}x}, \qquad \lambda_k = \xi_0 - \frac{\xi_1^2}{4\xi_2} - \pi^2 k^2 \xi_2, \tag{II.4}$$

and $T_k(t)$ is given by

$$T_k(t) = \begin{cases} \exp\big(\frac{\lambda_k}{\eta_1}t\big) & \text{if } \eta_2 = 0 \\ \frac{\eta_2}{\sqrt{\mu_k}}\Big(\exp\big(\frac{-\eta_1+\sqrt{\mu_k}}{2\eta_2}t\big) - \exp\big(\frac{-\eta_1-\sqrt{\mu_k}}{2\eta_2}t\big)\Big) & \text{if } \eta_2 > 0 \text{ and } \mu_k > 0 \\ t\,\exp\big(\frac{-\eta_1}{2\eta_2}t\big) & \text{if } \eta_2 > 0 \text{ and } \mu_k = 0 \\ \frac{2\eta_2}{\sqrt{-\mu_k}}\sin\big(\frac{\sqrt{-\mu_k}}{2\eta_2}t\big)\,\exp\big(\frac{-\eta_1}{2\eta_2}t\big) & \text{if } \eta_2 > 0 \text{ and } \mu_k < 0, \end{cases} \tag{II.5}$$

where $\mu_k = \eta_1^2 + 4\eta_2\lambda_k$. The eigenfunctions $X_k(x)$, $k \in \mathbb{N}$, constitute an orthonormal basis for the Hilbert space

$$L^2\big([0,1], \exp(\tfrac{\xi_1}{\xi_2}x)\,\mathrm{d}x\big),$$

and if $\lambda_1 < 0 < \eta_1$, then there exists a unique stationary solution $V(t,x)$ to (II.1) given by

$$\begin{aligned} V(t,x) &= \int_{-\infty}^{t}\int_0^1 G(t-s,x,y)\, W_\xi(s,y)\, e^{\frac{\xi_1}{\xi_2}y}\,\mathrm{d}y\,\mathrm{d}s \\ &= \int_{-\infty}^{t}\int_0^1 \sum_{k=1}^{\infty} T_k(t-s)\, X_k(x)\, X_k(y)\, W(s,y)\, e^{\frac{\xi_1}{2\xi_2}y}\,\mathrm{d}y\,\mathrm{d}y \\ &= \sum_{k=1}^{\infty} U_k(t)\, X_k(x), \end{aligned} \tag{II.6}$$

where the coefficient processes $U_k(t)$ are given by

$$U_k(t) = \int_{-\infty}^{t} T_k(t-s)\, W_k(s)\,\mathrm{d}s = \int_{-\infty}^{t} T_k(t-s)\,\mathrm{d}B_k(s), \tag{II.7}$$

(II.4)

and the pairwise independent white noise processes $W_k(t)$, $k \in \mathbb{N}$, and the pairwise independent two-sided normalized Brownian motions $B_k(t)$, $k \in \mathbb{N}$, are given by

$$W_k(t) = \int_0^1 X_k(y)\, W(t,y)\, e^{\frac{\xi_1}{2\xi_2}}\, \mathrm{d}y, \qquad B_k(t) = \int_0^t W_k(s)\, \mathrm{d}s.$$

Observe that the choice (II.2) of the noise process $W_\xi(t,x)$ is made in order for the Brownian motions $B_k(t)$ to become independent, and thus facilitating the analysis. Whether this choice also is adequate from a modeling point of view will of course depend on the particular application. We summarize the above considerations in the following theorem.

Theorem II.1 *If the parameter $\theta = (\eta_1, \eta_2, \xi_0, \xi_1, \xi_2)$ belongs to the parameter space $\Theta \subset \mathbb{R}^5$ given by*

$$\eta_2 \geq 0, \qquad \eta_1, \xi_2 > 0, \qquad \xi_0, \xi_1 \in \mathbb{R}, \qquad \frac{\xi_1}{4\xi_2} + \pi^2 \xi_2 > \xi_0, \tag{II.8}$$

then there exists a unique stationary solution $V(t,x) = \sum_{k=1}^{\infty} U_k(t) X_k(x)$ to the stochastic partial differential equation (II.1)*, where the deterministic functions $X_k(x)$ and the coefficient processes $U_k(t)$ are given by* (II.4) *respectively* (II.7).

The coefficient processes are characterized by the following proposition.

Proposition II.2 *The coefficient processes $U_k(t)$, $k \in \mathbb{N}$, are pairwise independent. If $\eta_2 = 0$, then $U_k(t)$ is a stationary Ornstein-Uhlenbeck process and solves the stochastic differential equation*

$$\mathrm{d}U_k(t) = \frac{\lambda_k}{\eta_1} U_k(t)\, \mathrm{d}t + \mathrm{d}B_k(t), \tag{II.9}$$

and if $\eta_2 > 0$, then $U_k(t)$ is the first component of the two-dimensional stationary Ornstein-Uhlenbeck process $\bar{U}_k(t) = \big(U_k(t), \tilde{U}_k(t)\big)$, where

$$\tilde{U}_k(t) = \int_{-\infty}^t T_k'(t-s)\, W_k(s)\, \mathrm{d}s = \int_{-\infty}^t T_k'(t-s)\, \mathrm{d}B_k(s).$$

Moreover, $\bar{U}_k(t)$ solves the stochastic differential equation

$$\mathrm{d}\bar{U}_k(t) = \begin{pmatrix} 0 & 1 \\ \frac{\lambda_k}{\eta_2} & \frac{-\eta_1}{\eta_2} \end{pmatrix} \bar{U}_k(t)\, \mathrm{d}t + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrm{d}B_k(t).$$

Proof The statement is classical for the parabolic case, *cf.* Walsh (1986, p. 323). In the hyperbolic case we use the white noise calculus and the differential equation (II.3) to see that

$$\tfrac{\mathrm{d}}{\mathrm{d}t} U_k(t) = \int_{-\infty}^t T_k'(t-s)\, W_k(s)\, \mathrm{d}s + T_k(0)\, W_k(t) = \tilde{U}_k(t),$$

(II.5)

and

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\tilde{U}_k(t) &= \int_{-\infty}^{t} T_k''(t-s)\, W_k(s)\,\mathrm{d}s + T_k'(0)\, W_k(t) \\
&= \int_{-\infty}^{t} \Big(\frac{\lambda_k}{\eta_2} T_k(t-s) - \frac{\eta_1}{\eta_2} T_k'(t-s)\Big)\, W_k(s)\,\mathrm{d}s + W_k(t) \\
&= \frac{\lambda_k}{\eta_2} U_k(t) - \frac{\eta_1}{\eta_2}\tilde{U}_k(t) + W_k(t).
\end{aligned}$$

If these equations are rewritten as Itô stochastic differential equations, then the stated equation for $\bar{U}_k(t)$ follows. □

Let $\Delta > 0$ be some fixed time step, and let $U_{k,\Delta}(t)$, $\tilde{U}_{k,\Delta}(t)$ be the time series associated with the k'th coefficient processes at the discrete time points $t\Delta$, $t \in \mathbb{Z}$, *i.e.*

$$U_{k,\Delta}(t) = U_k(t\Delta), \qquad \tilde{U}_{k,\Delta}(t) = \tilde{U}_k(t\Delta), \qquad \bar{U}_{k,\Delta}(t) = \big(U_{k,\Delta}(t), \tilde{U}_{k,\Delta}(t)\big).$$

Proposition II.3 *If $\eta_2 = 0$, then $U_{k,\Delta}(t)$ is a first order autoregressive process, i.e.*

$$U_{k,\Delta}(t) = \rho_{k,\Delta}\, U_{k,\Delta}(t-1) + \varepsilon_{k,\Delta}(t) \sim \mathcal{N}_1\big(0, \sigma_k^2\big),$$

where the innovations $\varepsilon_{k,\Delta}(t)$, $t \in \mathbb{Z}$, are i.id. $\mathcal{N}_1(0, \sigma_k^2 - \rho_{k,\Delta}^2 \sigma_k^2)$, and where the autoregression coefficient $\rho_{k,\Delta}$ and the stationary variance σ_k^2 are given by

$$\rho_{k,\Delta} = T_k(\Delta), \qquad\qquad \sigma_k^2 = \frac{\eta_1}{-2\lambda_k}.$$

If $\eta_2 > 0$, then $\bar{U}_{k,\Delta}(t)$ is a first order autoregressive process, i.e.

$$\bar{U}_{k,\Delta}(t) = \bar{\rho}_{k,\Delta}\, \bar{U}_{k,\Delta}(t-1) + \bar{\varepsilon}_{k,\Delta}(t) \sim \mathcal{N}_2\big(0, \bar{\sigma}_k^2\big),$$

where the innovations $\bar{\varepsilon}_{k,\Delta}(t)$, $t \in \mathbb{Z}$, are i.id. $\mathcal{N}_2(0, \bar{\sigma}_k^2 - \bar{\rho}_{k,\Delta}\bar{\sigma}_k^2\bar{\rho}_{k,\Delta}^*)$, and where the autoregression coefficient $\bar{\rho}_{k,\Delta}$ and the stationary variances $\bar{\sigma}_k^2$ are given by

$$\bar{\rho}_{k,\Delta} = \begin{pmatrix} \frac{\eta_1}{\eta_2} T_k(\Delta) + T_k'(\Delta) & T_k(\Delta) \\ \frac{\lambda_k}{\eta_2} T_k(\Delta) & T_k'(\Delta) \end{pmatrix}, \qquad \bar{\sigma}_k^2 = \begin{pmatrix} \sigma_k^2 & 0 \\ 0 & \tilde{\sigma}_k^2 \end{pmatrix} = \begin{pmatrix} \frac{\eta_2^2}{-2\eta_1\lambda_k} & 0 \\ 0 & \frac{\eta_2}{2\eta_1} \end{pmatrix}.$$

Proof We will first consider the hyperbolic case, *i.e.* when $\eta_2 > 0$. Since $\bar{U}_k(t)$ is an Ornstein-Uhlenbeck processes it directly follows, that the time series $\bar{U}_{k,\Delta}(t)$ is a first order autoregressive processes. The stationary variance $\bar{\sigma}_k^2$ can be calculated by

$$\bar{\sigma}_k^2 = \mathrm{Var}\Big(\int_{-\infty}^{t} \begin{pmatrix} T_k(t-s) \\ T_k'(t-s) \end{pmatrix} \mathrm{d}B_k(s)\Big) = \int_0^{\infty} \begin{pmatrix} T_k(s) \\ T_k'(s) \end{pmatrix} \begin{pmatrix} T_k(s) \\ T_k'(s) \end{pmatrix}^* \mathrm{d}s.$$

(II.6)

The components of $\bar{U}_k(t)$ are a priori seen to be independent, *i.e.*

$$\mathrm{Cov}\left(U_k(t), \tilde{U}_k(t)\right) = \int_0^\infty T_k(s)\, T_k'(s)\, \mathrm{d}s = \tfrac{1}{2} T_k^2(s)\big|_{s=0}^{s=\infty} = 0,$$

and the marginal variances are found by direct computations of the integrals. In order to determine the autoregression coefficient we use the equation

$$\bar{U}_{k,\Delta}(t) = \underbrace{\int_{-\infty}^{(t-1)\Delta} \begin{pmatrix} T_k(t\Delta - s) \\ T_k'(t\Delta - s) \end{pmatrix} \mathrm{d}B_k(s)}_{=\bar{\rho}_{k,\Delta}\, \bar{U}_{k,\Delta}(t-1)} + \underbrace{\int_{(t-1)\Delta}^{t\Delta} \begin{pmatrix} T_k(t\Delta - s) \\ T_k'(t\Delta - s) \end{pmatrix} \mathrm{d}B_k(s)}_{\text{innovation } \bar{\varepsilon}_{k,\Delta}(t)}$$

to conclude, that $\bar{\rho}_{k,\Delta}$ satisfies the equation

$$\bar{\rho}_{k,\Delta} \begin{pmatrix} T_k(u) \\ T_k'(u) \end{pmatrix} = \begin{pmatrix} T_k(u+\Delta) \\ T_k'(u+\Delta) \end{pmatrix} \quad \text{for every } u \geq 0. \tag{II.10}$$

Inserting $u = 0$ and using the differential equation (II.3) gives the second column of $\bar{\rho}_{k,\Delta}$, *i.e.*

$$\bar{\rho}_{k,\Delta} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \bar{\rho}_{k,\Delta} \begin{pmatrix} T_k(0) \\ T_k'(0) \end{pmatrix} = \begin{pmatrix} T_k(\Delta) \\ T_k'(\Delta) \end{pmatrix}.$$

Differentiating (II.10) w.r.t. u, inserting $u = 0$ and using the differential equation (II.3) gives

$$\bar{\rho}_{k,\Delta} \begin{pmatrix} 1 \\ -\frac{\eta_1}{\eta_2} \end{pmatrix} = \bar{\rho}_{k,\Delta} \begin{pmatrix} T_k'(0) \\ T_k''(0) \end{pmatrix} = \begin{pmatrix} T_k'(\Delta) \\ T_k''(\Delta) \end{pmatrix} = \begin{pmatrix} T_k'(\Delta) \\ \frac{\lambda_k}{\eta_k} T_k(\Delta) - \frac{\eta_1}{\eta_2} T_k'(\Delta) \end{pmatrix},$$

and hence the first column of $\bar{\rho}_{k,\Delta}$, *i.e.*

$$\bar{\rho}_{k,\Delta} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \bar{\rho}_{k,\Delta} \begin{pmatrix} 1 \\ -\frac{\eta_1}{\eta_2} \end{pmatrix} + \frac{\eta_1}{\eta_2} \bar{\rho}_{k,\Delta} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\eta_1}{\eta_2} T_k(\Delta) + T_k'(\Delta) \\ \frac{\lambda_k}{\eta_2} T_k(\Delta) \end{pmatrix}.$$

The more easy parabolic case can by analyzed similarly. □

In the hyperbolic case the paths of the coefficient processes $U_k(t)$ are continuous differentiable, *cf.* proposition II.2. This fact suggests that the solution $V(t,x)$ is more smooth in the hyperbolic case than in the parabolic case. We conclude this section by considering the sample path properties of $V(t,x)$.

Lemma II.4 *There exists a constant $\alpha < \infty$ such that*

$$\mathrm{Var}\left(V(t,x) - V(t,y)\right) \leq \alpha\, |x-y|,$$
$$\mathrm{Var}\left(V(s,x) - V(t,x)\right) \leq \begin{cases} \alpha \sqrt{|t-s|} & \text{if } \eta_2 = 0 \\ \alpha\, |t-s| \log(|t-s|^{-1}) & \text{if } \eta_2 > 0 \end{cases}$$

for $|t-s|$ sufficiently small.

Proof The first estimate and the second estimate in the parabolic case are given in Walsh (1986, proposition 3.7), whence we only need to consider the second estimate in the hyperbolic case. For fixed $\delta > 0$ the representation (II.6) gives

$$\mathrm{Var}\,\big(V(t+\delta,x) - V(t,x)\big) = \sum_{k=1}^{\infty} X_k^2(x) \cdot \mathrm{Var}\,\big(U_k(t+\delta) - U_k(t)\big).$$

The squared eigenfunctions $X_k^2(x)$ are bounded by the constant $2 + 2\exp(-\frac{\xi_1}{\xi_2})$. Moreover, using the autoregression with innovations $\bar{\varepsilon}_{k,\delta} = (\varepsilon_{k,\delta}, \tilde{\varepsilon}_{k,\delta})$ and the marginal variance

$$\mathrm{Var}(\varepsilon_{k,\delta}) = \sigma_k^2 - \Big(\frac{\eta_1}{\eta_2} T_k(\delta) + T_k'(\delta)\Big)^2 \sigma_k^2 - T_k(\delta)^2\, \tilde{\sigma}_k^2,$$

cf. proposition II.3, and using the differential equation (II.3) we see that

$$\begin{aligned}
\mathrm{Var}\,\big(U_k(t+\delta) - U_k(t)\big) &= \mathrm{Var}\,\Big(\big(\tfrac{\eta_1}{\eta_2} T_k(\delta) + T_k'(\delta) - 1\big)\, U_k(t) + T_k(\delta)\, \tilde{U}_k(t) + \varepsilon_{k,\delta}\Big) \\
&= \Big(\frac{\eta_1}{\eta_2} T_k(\delta) + T_k'(\delta) - 1\Big)^2 \sigma_k^2 + T_k(\delta)^2\, \tilde{\sigma}_k^2 + \mathrm{Var}(\varepsilon_{k,\delta}) \\
&= 2\Big(1 - \frac{\eta_1}{\eta_2} T_k(\delta) + T_k'(\delta)\Big)\, \sigma_k^2 \\
&= -\frac{2\lambda_k}{\eta_2} \int_0^{\delta} T_k(u)\, \mathrm{d}u\, \sigma_k^2 \\
&= \frac{\eta_1}{\eta_2} \int_0^{\delta} T_k(u)\, \mathrm{d}u.
\end{aligned}$$

Remember that we want to bound the sum of the variances by a constant times $\delta \log(\delta^{-1})$ for $\delta > 0$ sufficiently small. Thus given $n \in \mathbb{N}$ we find

$$\begin{aligned}
\sum_{k=1}^{n} \mathrm{Var}\,\big(U_k(t+\delta) - U_k(t)\big) &= \frac{\eta_1}{\eta_2} \sum_{k=1}^{n} \int_0^{\delta} T_k(u)\, \mathrm{d}u \\
&\approx \frac{\eta_1}{\eta_2} \sum_{k=1}^{n} \int_0^{\delta} \frac{1}{k\,\pi} \sqrt{\frac{\eta_2}{\xi_2}}\, \sin\!\big(k\pi \sqrt{\tfrac{\xi_2}{\eta_2}}\, u\big)\, \exp\!\big(\tfrac{-\eta_1}{2\eta_2} u\big)\, \mathrm{d}u \\
&= \frac{\eta_1}{\pi \sqrt{\eta_2\, \xi_2}} \int_0^{\delta} f_n\big(\pi \sqrt{\tfrac{\xi_2}{\eta_2}}\, u\big)\, e^{-\frac{\eta_1}{2\eta_2} u}\, \mathrm{d}u,
\end{aligned} \tag{II.11}$$

where the function $f_n(x)$ is defined by $f_n(x) = \sum_{k=1}^{n} \frac{\sin(k\,x)}{k}$. Some comments regarding the somewhat subtle approximation in (II.11) are needed. Firstly, since the kernels $T_k(u)$ are bounded, we may replace finitely many of the integrands $T_k(u)$ by arbitrary bounded functions, whence we only need to consider the approximation of the terms for large k. Secondly, if k is large, then $\mu_k < 0$ and the replacement for $T_k(u)$ corresponds to the approximation $-\mu_k \approx 4\pi\eta_2\xi_2 k^2$, *cf.* equation (II.5). Thirdly, the approximations of $T_k(u)$ for large k are small perturbations of lower order in k and may be disregarded.

If $\delta < \frac{1}{2}\sqrt{\frac{\eta_2}{\xi_2}}$, then the function $f_n(x)$ is only employed for $0 \le x \le \frac{\pi}{2}$. Moreover, we find that

$$f_n'(x) = \sum_{k=1}^{n} \cos(k\,x) = \frac{1+\cos(x)}{2\sin(x)} \sin\big((n+1)x\big) - \frac{1+\cos((n+1)x)}{2} \ge -\frac{x+1}{x}$$

and $f(\frac{\pi}{2}) \le 1$, whence for $0 \le x \le \frac{\pi}{2}$,

$$0 \le f_n(x) \le 1 + \int_x^{\frac{\pi}{2}} \frac{u+1}{u}\,\mathrm{d}u < 4 - \log(x).$$

Using this bound we see that

$$\int_0^{\delta} f_n\big(\pi\sqrt{\tfrac{\xi_2}{\eta_2}}u\big)\, e^{-\frac{\eta_1}{2\eta_2}u}\,\mathrm{d}u \le \int_0^{\delta} \Big(4 - \log\big(\pi\sqrt{\tfrac{\xi_2}{\eta_2}}u\big)\Big)\,\mathrm{d}u = 5\delta + \delta\log\big(\tfrac{1}{\pi}\sqrt{\tfrac{\eta_2}{\xi_2}}\delta^{-1}\big),$$

whereby the desired bound for the approximation (II.11) follows. □

Theorem II.5 *The solution $V(t,x)$ to the stochastic partial differential equation* (II.1) *has a version that is continuous in (t,x). Moreover if for some fixed $t_0 < \infty$,*

$$\begin{aligned} H_t(\delta) &= \sup_{x,y\in[0,1]:|x-y|\le\delta} |V(t,x) - V(t,y)|\,, \quad t \in [0,t_0],\\ H(\delta) &= \sup_{s,t\in[0,t_0],x,y\in[0,1]:((s-t)^2+(x-y)^2)^{\frac{1}{2}}\le\delta} |V(s,x) - V(t,y)| \end{aligned}$$

are the moduli of continuity in space at time t respectively in time and space then there exists a constant $\alpha < \infty$ and random variables Y_t, $t \in [0,t_0]$, and Y with exponential moments such that for $0 \le \delta \le 1$,

$$\begin{aligned} H_t(\delta) &\le Y_t\,\delta^{\frac{1}{2}} + \alpha\,\delta^{\frac{1}{2}}\sqrt{\log(\delta^{-1})},\\ H(\delta) &\le \begin{cases} Y\,\delta^{\frac{1}{4}} + \alpha\,\delta^{\frac{1}{4}}\sqrt{\log(\delta^{-1})} & \textit{if } \eta_2 = 0\\ Y\,\delta^{\frac{1}{2}}\sqrt{\log(\delta^{-1})} + \alpha\,\delta^{\frac{1}{2}}\log(\delta^{-1}) & \textit{if } \eta_2 > 0. \end{cases} \end{aligned}$$

Proof Using the variance estimates given in lemma II.4 the proof is similar to the proof of Walsh (1986, theorem 3.8). □

Theorem II.5 states that the solution $V(t,x)$ to (II.1) has paths that essentially are Hölder continuous of order $\frac{1}{2}$ in space and $\frac{1}{4}$ in time in the parabolic case, and of order $\frac{1}{2}$ in time and space in the hyperbolic case. The paths are thus substantially more rough in time in the parabolic case. The roughness present in the parabolic case is also reflected in the property, that the solution process in that case have non-vanishing quartic variation, *cf.* Walsh (1986, theorem 3.10).

II.3 Likelihood inference given an discrete observation

In this section we give a time series representation of the statistical model given observations of $V(t,x)$ at discrete points in time and space at the lattice points (t,x) given by

$$t = \Delta, 2\Delta, \dots, n\Delta, \qquad\qquad x = \frac{a_1}{b}, \dots, \frac{a_N}{b} \tag{II.12}$$

where $\Delta > 0$ and $a_1, \dots, a_N, b \in \mathbb{N}$, $a_1 < \dots < a_N < b$ are fixed. Moreover we describe an approximate maximum likelihood estimation procedure, which is asymptotically efficient as $n \to \infty$, and we describe the associated likelihood ratio test for a parabolic equation against a hyperbolic equation.

Given observations of $V(t,x)$ in the lattice points (II.12) let the N-dimensional time series $V_\Delta(t)$, the $2b$-dimensional time series $U_\Delta(t)$, the matrices $\Xi \in \mathbb{R}^{N\times N}$ and $\Psi \in \mathbb{R}^{N\times 2b}$ be given by

$$V_\Delta(t) = V\big(t\Delta, \tfrac{a_j}{b}\big)_{j=1,\dots,N}, \qquad U_\Delta(t) = \Big(\sum\nolimits_{j=0}^{\infty} U_{k+2bj,\Delta}(t)\Big)_{k=1,\dots,2b},$$
$$\Xi = \text{diag}\big(e^{-\frac{\xi_1}{2\xi_2}\frac{a_j}{b}}\big)_{j=1,\dots,N}, \qquad \Psi = \Big(\sqrt{2}\sin\big(\pi k \tfrac{a_j}{b}\big)\Big)_{\substack{j=1,\dots,N\\ k=1,\dots,2b}}.$$

Then the observable time series $V_\Delta(t)$ has the state space representation

$$V_\Delta(t) = \Xi\, \Psi\, U_\Delta(t). \tag{II.13}$$

The components of the time series $U_\Delta(t)$ are independent and given as infinite sums of independent time series. Suppose a sequence $K(n)$, $n \in \mathbb{N}$, of positive integers, and sequences $\tau_k^2(n)$, $n \in \mathbb{N}$, $k = 1, \dots, 2b$, of variances are given. We propose to approximate the tails in the representation of $U_\Delta(t)$ by independent white noise, *i.e.* to approximate the distribution of $V_\Delta(t)$ with the distribution of $\Xi\, \Psi\, \hat{U}_\Delta(t)$, where

$$\begin{gathered}\hat{U}_\Delta(t) = \Big(\sum\nolimits_{j\in\mathbb{N}_0 : k+2bj<K(n)} U_{k+2bj,\Delta}(t) + \hat{\varepsilon}_{k,n}(t)\Big)_{k=1,\dots,2b},\\ \hat{\varepsilon}_{k,n}(t),\ t = 1, \dots, n,\ \text{i.id. } \mathcal{N}\big(0, \tau_k^2(n)\big).\end{gathered} \tag{II.14}$$

Below we will describe how to choose the cutoff points $K(n)$ and the white noise variances $\tau_k^2(n)$ such that the resulting approximate likelihood is asymptotically efficient as $n \to \infty$. Observe that we have an explicit description of the finite dimensional state space model $\Xi\, \Psi\, \hat{U}^\Delta(t)$ and can calculate the approximate likelihood via the Kalman filter.

In order to measure the quality on the proposed approximation we introduce metrics on the space of N-dimensional matrices respectively on the space of N-dimensional spectral densities. The Schatten p-norm $\|A\|_p$, $p \in [1,\infty]$, of a matrix

$A \in \mathbb{C}^{N\times N}$ is defined as the l_p-norm of the eigenvalues of the positive semi definite matrix $|A| = (A^*A)^{\frac{1}{2}}$, *i.e.* $\|A\|_\infty$ is the operator norm of A and

$$\|A\|_p = \Big(\operatorname{tr}(A^*A)^{\frac{p}{2}}\Big)^{p^{-1}}, \quad p \in [1,\infty).$$

The L^p-norm $\|\psi\|_p$, $p \in [1,\infty]$, of a matrix valued function $\psi : (-\frac{1}{2},\frac{1}{2}] \to \mathbb{C}^{N\times N}$ is defined as the usual L^p-norm of the real valued function $\|\psi(\cdot)\|_p$, *i.e.* $\|\psi\|_\infty$ is the essential supremum of $\|\psi(\cdot)\|_\infty$ and

$$\|\psi\|_p = \Big(\int_{-\frac{1}{2}}^{\frac{1}{2}} \|\psi(\omega)\|_p^p \,\mathrm{d}\omega\Big)^{p^{-1}}, \quad p \in [1,\infty).$$

These L^p-norms behave much like the usual L^p-norms and especially satisfy the Hölder inequality. We will measure the quality of the introduced approximation via the L^2-distance between the spectral densities for the exact model respectively the approximative model. The analysis relies on the following theorem proved in Markussen (2001*d*).

Theorem II.6 *Let $V(t)$, $t \in \mathbb{Z}$, be a N-dimensional Gaussian time series, i.e.*

$$\mathbb{V}_n = \big(V(1),\dots,V(T)\big)^* \sim \mathcal{N}_{n\times N}\big(0,\Sigma_n(\psi)\big),$$

where $\Sigma_n(\psi)$ is the Toeplitz matrix associated to the spectral density $\psi(\omega)$. For each $n \in \mathbb{N}$ let $\psi_{n,\theta}(\omega)$, $\theta \in \Theta \subseteq \mathbb{R}^d$, be a family of spectral densities and let $l_n(\theta)$ be the corresponding log likelihood function given by

$$l_n(\theta) = -\tfrac{1}{2}\log\det\Sigma_n(\psi_{n,\theta}) - \tfrac{1}{2}\operatorname{tr}\big(\Sigma_n(\psi_{n,\theta})^{-1}\mathbb{V}_n\mathbb{V}_n^*\big). \qquad \text{(II.15)}$$

Let $\theta_0 \in \Theta$ be some fixed and unknown parameter, and suppose that the Fisher information matrix $\mathcal{J}$ given by

$$\mathcal{J} = \lim_{n\to\infty}\Big(\tfrac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \operatorname{tr}\big(\psi_{n,\theta_0}^{-1}(\omega)\partial_i\psi_{n,\theta_0}(\omega)\psi_{n,\theta_0}^{-1}(\omega)\partial_j\psi_{n,\theta_0}(\omega)\big)\,\mathrm{d}\omega\Big)_{i,j=1,\dots,d}$$

exists and is positive definite. If $n^{\frac{1}{2}}\|\psi - \psi_{n,\theta_0}\|_2$ vanishes as $n \to \infty$, and some additional mild regularity conditions are satisfied, see Markussen (2001d), then the maximum likelihood estimator $\hat{\theta}_n = \operatorname{argmax}_{\theta\in\Theta} l_n(\theta)$ is a $\sqrt{n}$-consistent estimator for θ_0, and the localized log likelihood ratio converges uniformly to a Gaussian shift process, i.e. there exists a sequence G_n, $n \in \mathbb{N}$, of d-dimensional random variables converging in distribution to $\mathcal{N}_d(0,\mathcal{J})$ such that

$$\mathrm{E}\Big(\sup_{u\in\mathbb{R}^d:|u|\le r,\theta_0+n^{-\frac{1}{2}}u\in\Theta}\big|l_n(\theta_0+n^{-\frac{1}{2}}u) - l_n(\theta_0) - \big(u^*G_n - \tfrac{1}{2}u^*\mathcal{J}u\big)\big|\Big) \qquad \text{(II.16)}$$

vanishes for every $r > 0$ as $n \to \infty$.

As a consequence of the LAN-property (II.16), the approximative maximum likelihood estimator $\hat{\theta}_n$ is asymptotically efficient in the sense of Hájek and Le Cam, *cf.* Le Cam & Yang (2000, chapter 6.6).

Let $\Theta \subset \mathbb{R}^5$ be the parameter space described in theorem II.1, and suppose that we have an observation in the lattice points (II.12) of the stochastic partial differential equation described by the parameter $\theta_0 \in \Theta$. Then the additional regularity conditions are satisfied for the corresponding time series model, *cf.* Markussen (2001*d*), and theorem II.6 applies if

$$n^{\frac{1}{2}} \left\| \psi_{\theta_0} - \psi_{n,\theta_0} \right\|_2 \xrightarrow[n\to\infty]{} 0,$$

where $\psi_{\theta_0}(\omega)$ is the exact spectral density and $\psi_{n,\theta_0}(\omega)$ is the n'th approximative spectral density. Using the state space representation (II.13) and the approximative state vector (II.14) we see that $\psi_{\theta_0}(\omega)$ and $\psi_{n,\theta_0}(\omega)$ are given by

$$\begin{aligned} \psi_{\theta_0}(\omega) &= \Xi\, \Psi\, \mathrm{diag}\Big(\sum\nolimits_{j=0}^{\infty} \varphi_{k+2bj,\Delta}(\omega)\Big)_{k=1,\dots,2b} \Psi^*\, \Xi^*, \\ \psi_{n,\theta_0}(\omega) &= \Xi\, \Psi\, \mathrm{diag}\Big(\sum\nolimits_{j\in\mathbb{N}_0 : k+2bj<K(n)} \varphi_{k+2bj,\Delta}(\omega) + \tau_k^2(n)\Big)_{k=1,\dots,2b} \Psi^*\, \Xi^*, \end{aligned}$$

where $\varphi_{k,\Delta}(\omega)$ is the spectral densities for the coefficient process $U_{k,\Delta}(t)$. The L^2-distance between $\psi_{\theta_0}(\omega)$ and $\psi_{n,\theta_0}(\omega)$ can be estimated with the aid of the following lemma.

Lemma II.7 *In the parabolic case the spectral density $\varphi^{\Delta}_{\theta,k}(\omega)$ satisfies the bounds*

$$\frac{\eta_1}{-2\lambda_k} \frac{1 - e^{\frac{\lambda_k}{\eta_1}\Delta}}{1 + e^{\frac{\lambda_k}{\eta_1}\Delta}} < \varphi^{\Delta}_{\theta,k}(\omega) < \frac{\eta_1}{-2\lambda_k} \frac{1 + e^{\frac{\lambda_k}{\eta_1}\Delta}}{1 - e^{\frac{\lambda_k}{\eta_1}\Delta}}.$$

In the hyperbolic case and for k large enough that $\mu_k < 0$ the spectral density $\varphi^{\Delta}_{\theta,k}(\omega)$ satisfies the bounds

$$\begin{aligned} \varphi^{\Delta}_{\theta,k}(\omega) &\geq \frac{\eta_2^2}{-2\eta_1\lambda_k} \frac{e^{\frac{\eta_1}{2\eta_2}\Delta} - 1}{e^{\frac{\eta_1}{2\eta_2}\Delta} + 1} - \frac{\eta_2^2}{-\lambda_k\sqrt{-\mu_k}} \Big(e^{\frac{\eta_1}{2\eta_2}\Delta} - e^{-\frac{\eta_1}{2\eta_2}\Delta}\Big)^{-1}, \\ \varphi^{\Delta}_{\theta,k}(\omega) &\leq \frac{\eta_2^2}{-2\eta_1\lambda_k} \frac{e^{\frac{\eta_1}{2\eta_2}\Delta} + 1}{e^{\frac{\eta_1}{2\eta_2}\Delta} - 1} + \frac{\eta_2^2}{-\lambda_k\sqrt{-\mu_k}} \Big(e^{\frac{\eta_1}{2\eta_2}\Delta} - e^{-\frac{\eta_1}{2\eta_2}\Delta}\Big)^{-1}. \end{aligned}$$

Proof If $\eta_2 = 0$, then the first order autoregressive process $U_{k,\Delta}(t)$ have spectral density

$$\begin{aligned} \varphi_{k,\Delta}(\omega) &= \sigma_k^2\,(1 - \rho_{k,\Delta}^2)\,\big(1 - 2\rho_{k,\Delta}\,\cos(2\pi\omega) + \rho_{k,\Delta}^2\big) \\ &= \frac{\eta_1}{-2\lambda_k}\big(1 - e^{2\frac{\lambda_k}{\eta_1}\Delta}\big)\big(1 - 2e^{\frac{\lambda_k}{\eta_1}\Delta}\cos(2\pi\omega) + e^{2\frac{\lambda_k}{\eta_1}\Delta}\big)^{-1}, \end{aligned}$$

whereby the stated bounds immediately follow. If $\eta_2 > 0$, then we first do a linear transformation of the autoregressive process $\bar{U}_{k,\delta}(t)$. If the matrix $A \in \mathbb{R}^{2\times 2}$ is invertible, then

$$A\,\bar{U}_{k,\Delta}(t+1) = \big(A\,\bar{\rho}_{k,\Delta}A^{-1}\big)A\,\bar{U}_{k,\Delta}(t) + A\,\bar{\varepsilon}_{k,\Delta}(t+1),$$

where the variance of the innovations $A\,\bar{\varepsilon}_{k,\Delta}(t)$ equals $A\,\bar{\sigma}_k^2\,A^*$. We need to chose the matrix A such that the powers of the transformed autoregression coefficient $A\,\bar{\rho}_{k,\Delta}\,A^{-1}$ can be calculated easily. Suppose k is large, whence $\mu_k < 0$, and let the matrix A be given by

$$A = \begin{pmatrix} 1 & 0 \\ \frac{\eta_1}{\sqrt{-\mu_k}} & \frac{2\eta_2}{\sqrt{-\mu_k}} \end{pmatrix}.$$

Using lemma II.3 we see that the coefficient $A\,\bar{\rho}_{k,\Delta}\,A^{-1}$ is given by

$$A\,\bar{\rho}_{k,\Delta}\,A^{-1} = \begin{pmatrix} \frac{\eta_1}{2\eta_2}T_k(\Delta) + T_k'(\Delta) & \frac{\sqrt{-\mu_k}}{2\eta_2}T_k(\Delta) \\ \frac{-\sqrt{-\mu_k}}{2\eta_2}T_k(\Delta) & \frac{\eta_1}{2\eta_2}T_k(\Delta) + T_k'(\Delta) \end{pmatrix}.$$

Moreover, the differential equation (II.3) gives

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}\Delta}\Big[e^{\frac{\eta_1}{2\eta_2}\Delta}\Big(\frac{\eta_1}{2\eta_2}T_k(\Delta) + T_k'(\Delta)\Big)\Big] &= \frac{-\sqrt{-\mu_k}}{2\eta_2}\Big(e^{\frac{\eta_1}{2\eta_2}\Delta}\frac{-\sqrt{-\mu_k}}{2\eta_2}T_k(\Delta)\Big), \\ \tfrac{\mathrm{d}}{\mathrm{d}\Delta}\Big[e^{\frac{\eta_1}{2\eta_2}\Delta}\frac{\sqrt{-\mu_k}}{2\eta_2}T_k(\Delta)\Big] &= \frac{\sqrt{-\mu_k}}{2\eta_2}\Big(e^{\frac{\eta_1}{2\eta_2}\Delta}\Big(\frac{\eta_1}{\eta_2}T_k(\Delta) + T_k'(\Delta)\Big)\Big), \end{aligned}$$

whence it follows that

$$A\,\bar{\rho}_{k,\Delta}\,A^{-1} = e^{-\frac{\eta_1}{2\eta_2}\Delta}\begin{pmatrix} \cos\big(\frac{\sqrt{-\mu_k}}{2\eta_2}\Delta\big) & \sin\big(\frac{\sqrt{-\mu_k}}{2\eta_2}\Delta\big) \\ -\sin\big(\frac{\sqrt{-\mu_k}}{2\eta_2}\Delta\big) & \cos\big(\frac{\sqrt{-\mu_k}}{2\eta_2}\Delta\big) \end{pmatrix}.$$

Similarly, the variance $A\,\bar{\sigma}_k^2\,A^*$ is given by

$$A\,\bar{\sigma}_k^2\,A^* = \sigma_k^2\begin{pmatrix} 1 & \frac{\eta_1}{\sqrt{-\mu_k}} \\ \frac{\eta_1}{\sqrt{-\mu_k}} & \frac{\eta_1^2 - 2\eta_2\lambda_k}{-\mu_k} \end{pmatrix},$$

whence it follows, that the first component of the time series $A\,\bar{U}_{k,\Delta}(t)$, *i.e.* the coefficient process $U_{k,\Delta}(t)$, has spectral density $\varphi_{k,\Delta}(\omega)$ given by

$$\begin{aligned} &\sigma_k^2 + 2\sigma_k^2\sum_{t=1}^{\infty} e^{\frac{-\eta_1}{2\eta_2}\Delta t}\cos\big(\tfrac{\sqrt{-\mu_k}}{2\eta_2}\Delta t\big)\cos(2\pi\omega t) \\ &+2\frac{\eta_1}{\sqrt{-\mu_k}}\sigma_k^2\sum_{t=1}^{\infty} e^{-\frac{\eta_1}{2\eta_2}\Delta t}\sin\big(\tfrac{\sqrt{-\mu_k}}{2\eta_2}\Delta t\big)\cos(2\pi\omega t). \end{aligned}$$

(II.13)

The stated bounds thus follow by inserting the trigonometric relations

$$\begin{aligned} 2\cos(x)\cos(y) &= \cos(x+y)+\cos(x-y), \\ 2\sin(x)\cos(y) &= \sin(x+y)+\sin(x-y) \end{aligned}$$

and using the bounds

$$\begin{aligned} -\frac{1}{e^a+1} &\le \sum_{n=1}^{\infty} e^{-an}\cos(bn) = \frac{e^a\cos(b)-1}{1-2e^a\cos(b)+e^{2a}} \le \frac{1}{e^a-1}, \\ -\frac{1}{e^a-e^{-a}} &\le \sum_{n=1}^{\infty} e^{-an}\sin(bn) = \frac{e^a\sin(b)}{1-2e^a\cos(b)+e^{2a}} \le \frac{1}{e^a-e^{-a}}. \end{aligned}$$

□

The needed cutoff points $K(n)$ and white noise variances $\tau_k^2(n)$ are given in the following lemma.

Lemma II.8 *If the cutoff points $K(n)$ and the white noise variances $\tau_k^2(n)$ in the parabolic case are given by*

$$K(n) = \Big\lceil \frac{1}{\pi}\sqrt{\frac{\eta_1}{2\xi_2\Delta}}\sqrt{\log T}\Big\rceil, \qquad \tau_k^2(n) = \sum_{j\in\mathbb{N}_0:k+2bj\ge K(n)}^{\infty} \frac{\eta_1}{-2\lambda_{k+2bj}}, \tag{II.17a}$$

and in the hyperbolic case are given such that

$$n^{\frac{1}{2}}K(n)^{-1} \xrightarrow[n\to\infty]{} 0, \qquad \tau_k^2(n) = \sum_{j\in\mathbb{N}_0:k+2bj\ge K(n)}^{\infty} \frac{\eta_2^2}{-2\eta_1\lambda_{k+2bj}}, \tag{II.17b}$$

then $n^{\frac{1}{2}}\left\|\psi_{\theta_0}-\psi_{n,\theta_0}\right\|_2$ vanishes as $n\to\infty$.

Proof We first consider the parabolic case. Using the Hölder inequality, the triangular inequality and lemma II.7 we see that

$$\begin{aligned} \left\|\psi_{\theta_0}-\psi_{n,\theta_0}\right\|_2 &\le \left\|\Xi\right\|_\infty^2 \left\|\Psi\right\|_\infty^2 \sum_{k=K(n)}^{\infty} \left\|\varphi_{k,\Delta}-\tfrac{\eta_1}{-2\lambda_k}\right\|_2 \\ &\approx \left\|\Xi\right\|_\infty^2 \left\|\Psi\right\|_\infty^2 \sum_{k=K(n)}^{\infty} \frac{\eta_1}{-\lambda_k}\exp\big(\tfrac{\lambda_k}{\eta_1}\Delta\big), \end{aligned}$$

whence if $K(n)$ satisfies the equation

$$n^{-\frac{1}{2}} \approx \sum_{k=K(n)}^{\infty} \exp\big(-\pi^2\tfrac{\xi_2}{\eta_1}\Delta k^2\big) \lesssim \exp\big(-\pi^2\tfrac{\xi_2}{\eta_1}\Delta K(n)^2\big), \tag{II.18}$$

(II.14)

then $n^{\frac{1}{2}} \left\| \psi_{\theta_0} - \psi_{n,\theta_0} \right\|_2$ vanishes as $n \to \infty$. If we solve (II.18) w.r.t. $K(n)$, then the statement of the lemma follows. Similarly, in the hyperbolic case we use the estimates

$$n^{\frac{1}{2}} \sum_{k=K(n)}^{\infty} \left\| \varphi_{k,\Delta} - \tfrac{\eta_2^2}{-2\eta_1 \lambda_k} \right\| \lesssim n^{\frac{1}{2}} \sum_{k=K(n)}^{\infty} k^{-2} \approx n^{\frac{1}{2}} K(n)^{-1}$$

to get the statement of the lemma. □

The following theorem gives the asymptotic properties of the proposed approximative likelihood function.

Theorem II.9 *Let* $\theta_0 = (\eta_1, \eta_2, \xi_0, \xi_1, \xi_2)$ *be the true parameter, and let the cutoff points* $K(n)$ *and the white noise variances* $\tau_k^2(n)$ *be given by* (II.17)*. Then the approximate log likelihood* $l_n(\theta)$ *given in* (II.15) *can be calculated via the Kalman filter, and the approximate maximum likelihood estimator* $\hat{\theta}_n = \operatorname{argmax}_{\theta\in\Theta} l_n(\theta)$ *is an asymptotically efficient estimator for* θ_0*. Moreover,* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *converges in distribution to* $\mathcal{N}_5(0, \mathcal{J}^{-1})$*, where the Fisher information* $\mathcal{J}$ *is given by*

$$\mathcal{J} = \left(\frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \operatorname{tr}\left(\psi_{\theta_0}^{-1}(\omega) \partial_i \psi_{\theta_0}(\omega) \psi_{\theta_0}^{-1}(\omega) \partial_j \psi_{\theta_0}(\omega) \right) \mathrm{d}\omega \right)_{i,j=1,\ldots,5}.$$

If $\eta_2 = 0$*, then the approximative likelihood ratio test statistic* $\chi(n)$ *for a parabolic against a hyperbolic equation satisfies*

$$\chi(n) = 2 \sup_{\theta\in\Theta} l_n(\theta) - 2 \sup_{\theta\in\Theta:\eta_2=0} l_n(\theta) \xrightarrow[n\to\infty]{\mathcal{D}} \tfrac{1}{2}\epsilon_0 + \tfrac{1}{2}\chi_1^2,$$

where ϵ_0 *is the point measure in* 0 *and* χ_1^2 *is the chi-square distribution with one degree of freedom.*

Proof Using lemma II.8 we see that the conditions of theorem II.6 are satisfied. The properties of the estimator $\hat{\theta}_n$ then follows from the LAN-property (II.16). The asymptotic distribution of the likelihood ratio test statistic follows since the hypothesis $\eta_2 = 0$ is a hyperplane of one dimension less on the border of the parameter space, see Self & Liang (1987). □

(II.15)

III

Chaos decomposition and stochastic calculus for the negative binomial process

Abstract

In this paper stochastic analysis w.r.t. the Lévy process with negative binomial distributed marginals is investigated. The theory of reproducing kernel Hilbert spaces is used to describe the structure of the Hilbert space of quadratic integrable functionals of the negative binomial process index by an atomfree measurable space. In contrast to the well-known Gaussian and Poisson cases, the multiple integrals no longer are capable of generating the Hilbert space of quadratic integrable functionals. An orthonormal basis consisting of polynomials in the integrals w.r.t. integer powers of the increments are presented, and the chaos expansions for some polynomial functionals, especially the Meixner polynomial functionals, are calculated. The negative binomial process can be constructed as compound Poisson processes with logarithmic distributed jumps, and the constructed chaos decomposition is used to define and discuss Malliavin calculi based on differentiation w.r.t. the jump times respectively the jump heights.

Key words

Negative binomial process, reproducing kernel Hilbert space, chaos decomposition, polynomial functionals, Malliavin derivative w.r.t. jump times, Malliavin derivative w.r.t. jump heights.

III.1 Introduction

The chaos decompositions of the Hilbert spaces of quadratic integrable functionals of a Wiener respectively Poisson process have proven to be very useful. In this paper we derive a chaos decomposition of the quadratic integrable functionals of the negative binomial process. Although this process has classical applications in actuarial mathematics, it has been chosen as object of our investigations for purely mathematical reasons. The negative binomial process can be constructed as a compound Poisson process with bounded intensity and logarithmic distributed jumps, and is thus only slightly more complicated than the Poisson process. Despite the close connection to Poisson processes new interesting features are revealed. Especially in contrast to the Wiener and Poisson cases, the multiple stochastic integrals are no longer capable of generating all quadratic integrable functionals. The negative binomial process indexed by the parameter set T is in a natural correspondence with a Poisson process indexed by the parameter set $T \times \mathbb{N}$, where the second component describes the jump heights. It is thus possible to map the chaos decomposition developed for general Poisson processes to a chaos decomposition for the negative binomial process, *cf.* the papers Lytvynov et al. (1997), Denis (2000). However, to our opinion this construction is rather unnatural, and the construction presented in this paper is directly connected to the negative binomial process. After having established the chaos decomposition, which we will use as the fundamental building block, other elements of stochastic calculus are naturally considered, *e.g.* Malliavin calculus, Wick calculus, white noise and other generalized processes, solutions of stochastic differential equations, and so forth. In the present paper we will only take the first steps in developing a Malliavin calculus.

This paper is inspired by Itô (1988), where similar methods are used for the Poisson process. The author like to think of the present investigation as a generic example of what could be done in the framework of infinite divisible distributions and Lévy processes index by measure spaces, and care has been taken to make the presentation systematic and concise, and to minimize utilization of the particular form of the negative binomial distribution.

III.1.1 Notational conventions

The positive integers, the non-negative integers, the real numbers, and the complex numbers are denoted by $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$ and $\mathbb{C}$ respectively, and the imaginary unit is denoted by $\imath = \sqrt{-1}$. Much of this paper is concerned with analysis of certain topological spaces $\mathcal{F}$ of real or complex valued functions defined on certain sets E. The dual space of continuous linear functionals on $\mathcal{F}$ is denoted by $\mathcal{F}'$, and the duality between $\mathcal{F}$ and $\mathcal{F}'$ is denoted by $\langle \omega, f \rangle_{\mathcal{F}'}$, $\omega \in \mathcal{F}'$, $f \in \mathcal{F}$. If $\mathcal{F}$ consists of real valued functions, then the complexification of $\mathcal{F}$ is given by

$$\mathcal{F}_{\mathbb{C}} = \mathcal{F} \oplus \imath\mathcal{F} = \big\{ f + \imath g : f, g \in \mathcal{F} \big\},$$

(III.2)

and the duality with $\mathcal{F}'$ extends by linearity. If the set E is equipped with a σ-algebra and a positive measure λ, then we denote by $\langle\cdot\rangle_{\mathcal{F}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ the functionals

$$\langle f\rangle_{\mathcal{F}} = \int_E f\,\mathrm{d}\lambda, \quad f\in L^1(E), \qquad \langle f,g\rangle_{\mathcal{F}} = \int_E f\overline{g}\,\mathrm{d}\lambda, \quad f,g\in L^2(E),$$

where $\overline{z}$ is the complex conjugate of $z\in\mathbb{C}$. If the set E is a product, *i.e.* $E = E_0^m$, then $\hat{\mathcal{F}}$ consists of the functions $f(t_1,\dots,t_m)\in\mathcal{F}$ that are invariant under permutations of the arguments $t_1,\dots,t_m\in E_0$, and given functions $f_1,\dots,f_m$ of a single argument $t\in E_0$, the symmetric tensor product $\hat{\otimes}_{j=1}^m f_j$ is given by

$$\hat{\otimes}_{j=1}^m f_j(t_1,\dots,t_m) = (m!)^{-1}\sum_{\substack{\pi\text{ permutation}\\ \text{of }\{1,\dots,m\}}}\prod_{j=1}^m f_{\pi(j)}(t_j).$$

Moreover, vectors and multi indices will be used. The coordinates of the vector $\vec{n}$ will be denoted by n_i. The coordinates of the indexed vector $\vec{n}_i$ are thus denoted by n_{ij} and so forth. The set of multi indices $\mathcal{I}$ defined by

$$\mathcal{I} = \left\{\vec{n}\in\mathbb{N}_0^{\mathbb{N}} : n_k\neq 0 \text{ for only finitely many } k\in\mathbb{N}\right\}$$

will have a prominent role. This set is a $\mathbb{N}_0$-module with basis vectors $\vec{\varepsilon}_k$, $k\in\mathbb{N}$, having a one on the k'th coordinate and zeros elsewhere. Let $\preceq$ be the natural partial order on $\mathcal{I}$, *i.e.* $\vec{m}\preceq\vec{n}$ if $m_k\le n_k$, $k\in\mathbb{N}$. Given $\vec{n}\in\mathcal{I}$ we define

$$\vec{n}_+ = \sum_{k\in\mathbb{N}} n_k, \qquad\qquad \vec{n}_* = \sum_{k\in\mathbb{N}} k n_k.$$

The subsets $\mathcal{I}_{(n)}$ of the multi indices defined by

$$\mathcal{I}_{(n)} = \left\{\vec{n}\in\mathcal{I} : \vec{n}_* = n\right\}, \quad n\in\mathbb{N}_0,$$

will pop up repeatedly in the calculations and formulae in this paper. As a side remark we mention, that the number of elements in $\mathcal{I}_{(n)}$ equals the numbers of ways in which $n\in\mathbb{N}$ can be written as a sum of integers. A famous formula by Hardy & Ramanujan (1918) implies, that

$$\left|\mathcal{I}_{(n)}\right| \sim \tfrac{1}{4\sqrt{3}} n^{-1} e^{\pi\sqrt{\frac{2}{3}n}} \text{ asymptotically as } n\to\infty.$$

This estimate indicates, that formulae like equation (III.21) below get rather complicated as $n\in\mathbb{N}$ gets large.

III.2 The negative binomial probability space

Consider the negative binomial distribution $\mathcal{NB}_{\beta,q}$, $0<p<1$, $q=1-p$, $\beta>0$, with point probability masses

$$\mathcal{NB}_{\beta,q}\{k\} = \binom{-\beta}{k} p^\beta(-q)^k, \quad k\in\mathbb{N}_0. \tag{III.3}$$

The convolution of $\mathcal{NB}_{\beta_1,q}$ and $\mathcal{NB}_{\beta_2,q}$ is given by adding the parameters β_1 and β_2, *i.e.* $\mathcal{NB}_{\beta_1,q} * \mathcal{NB}_{\beta_2,q} = \mathcal{NB}_{\beta_1+\beta_2,q}$. Let T be a σ-compact topological space equipped with the Borel σ-algebra $\mathscr{B}_T$ and a non-atomic, σ-finite, positive Radon measure λ. Usually T will be the set of d-dimensional spatial parameters $\mathbb{R}^d$ equipped with the Lebesgue measure. Then there exists an abstractly defined Lévy process $X = (X_t)_{t\in T}$ with negative binomial distributed marginals. This process can be interpreted as a discrete random measure on T, where the positions τ_j of the point masses Y_j follows a Poisson process with intensity $-\log(1-q)$, and the point masses Y_j are pairwise independent and logarithmic distributed, *i.e.*

$$\mathrm{P}\big(Y_j = k\big) = \tfrac{1}{-\log(1-q)} k^{-1} q^k, \quad k \in \mathbb{N}.$$

For convenience we will use the terminology from the case, where T is one-dimensional time, and refer to τ_j and Y_j as the jump times and jump heights respectively. Given $\eta \in L^2(T)$ and $\alpha \in \mathbb{N}_0$ the stochastic integral of η w.r.t. the α'th power of the jumps of the process X is well defined as a L^2-limit by

$$\int_T \eta(t) \left(\tfrac{\mathrm{d}X_t}{\mathrm{d}t}\right)^\alpha \mathrm{d}t = \lim_{N\to\infty} \sum_j Y_j^\alpha \eta(\tau_j) 1_{(\tau_j \in T_N)},$$

where $T_N \in \mathscr{B}_T$ have finite measure and increase to T. Especially denote by $I(\eta)$ the stochastic integral of $\eta \in L^2(T)$ w.r.t. X. Given pairwise disjoint sets $A_1, \dots, A_n \in \mathscr{B}_T$ the random variables $I(1_{A_1}), \dots, I(1_{A_n})$ are stochastically independent, and moreover

$$I(1_A) \sim \mathcal{NB}_{\langle 1_A \rangle_{L^1(T)}, q}, \quad A \in \mathscr{B}_T.$$

Let $\mathscr{S} = \mathscr{S}(T)$ be the Schwartz space of rapidly decreasing, smooth, real valued functions defined on T, let $\mathscr{S}' = \mathscr{S}'(T)$ be the dual space of tempered Schwartz distributions, and let $\mathscr{S}_q$ be the subset of the complexification of $\mathscr{S}$ given by

$$\mathscr{S}_q = \big\{\eta \in \mathscr{S}_{\mathbb{C}} : |\mathrm{Im}(\eta)| < -\log q\big\}.$$

Using the same technique as in the proof of lemma III.1 below, it follows that the functional $C(\eta) = \mathrm{E}[\exp(\imath I(\eta))]$ is well defined and given by

$$C(\eta) = \exp\left\langle \log\left(\frac{1-q}{1-qe^{\imath\eta}}\right)\right\rangle_{L^1(T)}, \quad \eta \in \mathscr{S}_q. \tag{III.1}$$

The restriction of the functional C to $\mathscr{S}$ satisfies $C(0) = 1$, is positive definite by construction, and is easily seen to be continuous in the Fréchet topology on $\mathscr{S}$. By the Bochner-Minlos theorem, see *e.g.* Holden et al. (1996, p. 193), there thus exists a probability measure μ on the Borel σ-algebra $\mathscr{B}(\mathscr{S}')$ on $\mathscr{S}'$ with characteristic functional C, *i.e.*

$$C(\eta) = \int_{\mathscr{S}'} e^{\imath\langle\omega,\eta\rangle_{\mathscr{S}'}} \,\mathrm{d}\mu(\omega), \quad \eta \in \mathscr{S}. \tag{III.2}$$

Given $\xi, \zeta \in \mathscr{S}$, $|\zeta| \leq 1$, both sides of (III.2) with $\eta = \xi + w\zeta$ are analytic in the open stripe $w \in \mathbb{C}$, $|\mathrm{Im}(w)| < -\log q$, and coincide for $w \in \mathbb{R}$. Thus (III.2) also holds for $\eta \in \mathscr{S}_q$. The probability space $(\mathscr{S}', \mathscr{B}(\mathscr{S}'), \mu)$ will henceforth be called the negative binomial probability space. Denote by $L^2(\mathscr{S}') = L^2(\mathscr{S}', \mathscr{B}(\mathscr{S}'), \mu)$ the Hilbert space of quadratic integrable, complex valued random variables equipped with the inner product

$$\langle \phi, \psi \rangle_{L^2(\mathscr{S}')} = \int_{\mathscr{S}'} \phi(\omega) \overline{\psi(\omega)} \, \mathrm{d}\mu(\omega), \quad \phi, \psi \in L^2(\mathscr{S}').$$

Easy calculations shows that the random variable $\langle \cdot, \eta \rangle_{\mathscr{S}'}$, $\eta \in \mathscr{S}$, have mean value and variance given by

$$\mathrm{E}_\mu \langle \cdot, \eta \rangle_{\mathscr{S}'} = \frac{q}{1-q} \langle \eta \rangle_{L^1(T)}, \qquad \mathrm{Var}_\mu \langle \cdot, \eta \rangle_{\mathscr{S}'} = \frac{q}{(1-q)^2} \langle \eta^2 \rangle_{L^1(T)}.$$

The map $J : \mathscr{S} \to L^2(\mathscr{S}')$ given by

$$J(\eta) = \frac{1-q}{\sqrt{q}} \big(\langle \cdot, \eta \rangle_{\mathscr{S}'} - \mathrm{E}_\mu \langle \cdot, \eta \rangle_{\mathscr{S}'} \big) = \frac{1-q}{\sqrt{q}} \langle \cdot, \eta \rangle_{\mathscr{S}'} - \sqrt{q} \langle \eta \rangle_{L^1(T)}$$

thus extends to a linear isometry from $L^2(T)$ to $L^2(\mathscr{S}')$. Via this isometry the abstract defined stochastic integral $I(\eta)$ can be represented on the negative binomial probability space by $I(\eta) = \langle \cdot, \eta \rangle_{\mathscr{S}'}$. Similarly, the random variables

$$\langle \cdot, \eta \rangle_{\mathscr{S}', \alpha} = \lim_{N \to \infty} \sum_j Y_j^\alpha \eta(\tau_j) 1_{(\tau_j \in T_N)}, \quad \eta \in L^2(T), \alpha \in \mathbb{N}_0,$$

are well defined on $(\mathscr{S}', \mathscr{B}(\mathscr{S}'), \mu)$. The following lemma describes the fundamental probabilistic behavior of the negative binomial probability space.

Lemma III.1 *Given $\eta_1, \ldots, \eta_n \in \mathscr{S}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{N}_0$ the identity*

$$\mathrm{E}_\mu \exp\Big(\imath \sum_{l=1}^n \langle \cdot, \eta_l \rangle_{\mathscr{S}', \alpha_l} \Big) = \exp \left\langle \log(1-q) + \sum_{k=1}^\infty k^{-1} q^k e^{\imath \sum_{l=1}^n \eta_l k^{\alpha_l}} \right\rangle_{L^1(T)}$$

holds true.

Proof The expectation $\mathrm{E}_\mu \exp(\imath \sum_{l=1}^n \langle \cdot, w_l 1_A \rangle_{\mathscr{S}', \alpha_l})$, where $w_1, \ldots, w_n \in \mathbb{C}$ and $A \in \mathscr{B}_T$, is equal to

$$\sum_{j=0}^\infty e^{|A| \log(1-q)} (j!)^{-1} |A|^j \Big(\sum_{k=1}^\infty k^{-1} q^k e^{\imath \sum_{l=1}^n w_l k^{\alpha_l}} \Big)^j$$
$$= \exp\Big(|A| \log(1-q) + |A| \sum_{k=1}^\infty k^{-1} q^k e^{\imath \sum_{l=1}^n w_l k^{\alpha_l}} \Big).$$

To complete the proof approximate η_l by elementary functions, and use stochastic independence and dominated convergence. □

III.3 The Hilbert space of quadratic integrable random variables

In this section we describe the structure of the Hilbert space $L^2(\mathscr{S}')$ using the so called $\mathcal{S}$-transform, see Hida et al. (1993), which transforms $L^2(\mathscr{S}')$ into a Hilbert space $\mathfrak{S}$ of non-linear, complex valued functionals on $\mathscr{S}_{\sqrt{q}}$. The structure of the Hilbert space $\mathfrak{S}$ can be studied via the theory of reproducing kernel Hilbert spaces as described in the classical paper by Aronszajn (1950).

The $\mathcal{S}$-transform is the linear operator defined by normalizing the Fourier transform, see Hida et al. (1993, p. 317), with the characteristic functional, *i.e.*

$$(\mathcal{S}\phi)(\xi) = C(\xi)^{-1} \int_{\mathscr{S}'} e^{\imath\langle\omega,\xi\rangle_{\mathscr{S}'}} \phi(\omega)\,\mathrm{d}\mu(\omega), \quad \phi \in L^2(\mathscr{S}'),\ \xi \in \mathscr{S}_{\sqrt{q}}.$$

If $A_1, \dots, A_n \in \mathscr{B}_T$ are pairwise disjoint, and $\phi_j \in L^2(\mathscr{S}')$ is measurable w.r.t. the σ-algebra generated by the jump times and jump heights of X in A_j, $j = 1, \dots, n$, then it easily follows that

$$\mathcal{S} \prod_{j=1}^{n} \phi_j = \prod_{j=1}^{n} \mathcal{S}\phi_j.$$

In order to develop a workable calculus it is of great importance to calculate the $\mathcal{S}$-transform of stochastic monomials. Observe that the mean of the stochastic monomials are found by inserting $\xi = 0$ in lemma III.2.

Lemma III.2 *Given $\eta_1, \dots, \eta_n \in \mathscr{S}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{N}_0$ we have for $\xi \in \mathscr{S}_{\sqrt{q}}$ that*

$$\Big(\mathcal{S} \prod_{l=1}^{n} \langle \cdot, \eta_l \rangle_{\mathscr{S}',\alpha_l}\Big)(\xi) = \sum\nolimits^1 \prod_{\nu=1}^{m} \Big\langle \gamma\Big(qe^{\imath\xi}; \sum_{l\in\mathcal{F}_\nu} \alpha_l - 1\Big) \prod_{l\in\mathcal{F}_\nu} \eta_l \Big\rangle_{L^1(T)},$$

where $\sum^1$ is the sum over $m = 1, \dots, n$ and all partitions $\mathcal{F}_1, \dots, \mathcal{F}_m$ of $\{1, \dots, n\}$. Moreover, the functions $\gamma(w; m) = \sum_{k=1}^{\infty} k^m w^k$, $m \in \{-1, 0\} \cup \mathbb{N}$, are given by

$$\gamma(w; -1) = -\log(1 - w), \quad \gamma(w; 0) = \frac{w}{1 - w}, \quad \gamma(w; m) = \sum_{j=1}^{m} \gamma_{mj} \frac{w^j}{(1 - w)^{j+1}},$$

where the coefficients γ_{mj}, $j = 1, \dots, m$, are given by

$$\gamma_{m1} = 1, \qquad \gamma_{mj} = j\big(\gamma_{m-1,j-1} + \gamma_{m-1,j}\big), \qquad \gamma_{mm} = m\gamma_{m-1,m-1}. \tag{III.3}$$

Proof Using the definition of the $\mathcal{S}$-transform, interchangeability of differentiation and integration, lemma III.1, and the rules of differentiation we find that

$$\begin{aligned}
\Big(\mathcal{S}\prod_{l=1}^{n}\langle\cdot,\eta_l\rangle_{\mathscr{S}',\alpha_l}\Big)(\xi) &= C(\xi)^{-1}\int_{\mathscr{S}'} e^{\imath\langle\omega,\xi\rangle_{\mathscr{S}'}}\prod_{l=1}^{n}\langle\omega,\eta_l\rangle_{\mathscr{S}',\alpha_l}\,\mathrm{d}\mu(\omega)\\
&= C(\xi)^{-1}\imath^{-n}\tfrac{\partial^n}{\partial w_1\cdots\partial w_n}\int_{\mathscr{S}'} e^{\imath\langle\omega,\xi\rangle_{\mathscr{S}'}+\imath\sum_{l=1}^{n} w_l\langle\omega,\eta_l\rangle_{\mathscr{S}',\alpha_l}}\,\mathrm{d}\mu(\omega)\Big|_{w_1=\ldots=w_n=0}\\
&= \imath^{-n}\tfrac{\partial^n}{\partial w_1\cdots\partial w_n}\exp\Big\langle\sum_{k=1}^{\infty}k^{-1}q^k\Big(e^{\imath\xi k+\imath\sum_{l=1}^{n}w_l\eta_l k^{\alpha_l}}-e^{\imath\xi k}\Big)\Big\rangle_{L^1(T)}\Big|_{w_1=\ldots=w_n=0}\\
&= \sum{}^{1}\prod_{\nu=1}^{m}\Big\langle\gamma\Big(qe^{\imath\xi};\sum_{l\in\mathcal{F}_\nu}\alpha_l-1\Big)\prod_{l\in\mathcal{F}_\nu}\eta_l\Big\rangle_{L^1(T)},
\end{aligned}$$

where $\sum^1$ and $\gamma(w;m)$ are as defined above. The formulae for $\gamma(w;-1)$ and $\gamma(w;0)$ are well known, and the formulae for $\gamma(w;m)$, $m\in\mathbb{N}$, are found recursively using the identity

$$\tfrac{\partial}{\partial w}\gamma(w;m)=\sum_{k=1}^{\infty}k^{m+1}w^{k-1}=w^{-1}\gamma(w;m+1),\quad m\in\mathbb{N}_0.$$

□

Let $\mathbb{E}$ be the algebra of stochastic exponentials, *i.e.* the algebra generated by the random variables $\exp(\imath\langle\cdot,\eta\rangle_{\mathscr{S}'})$, $\eta\in\mathscr{S}_{\sqrt{q}}$. The $\mathcal{S}$-transform of a stochastic exponential is given by

$$\big(\mathcal{S}e^{\imath\langle\cdot,\eta\rangle_{\mathscr{S}'}}\big)(\xi)=C(\xi)^{-1}C(\eta+\xi),\quad \eta,\xi\in\mathscr{S}_{\sqrt{q}}.$$

The set of stochastic exponentials is known to be dense in $L^2(\mathscr{S}')$, see Hida et al. (1993, theorem 1.9), and as a consequence the $\mathcal{S}$-transform is injective. The $\mathcal{S}$-transform is thus an isometric isomorphism between $L^2(\mathscr{S}')$ and the Hilbert space $\mathfrak{S}$ of functionals on $\mathscr{S}_{\sqrt{q}}$ defined by

$$\mathfrak{S}=\big\{\mathcal{S}\phi:\phi\in L^2(\mathscr{S}')\big\},$$

equipped with the inner product $\langle\mathcal{S}\phi,\mathcal{S}\psi\rangle_{\mathfrak{S}}=\langle\phi,\psi\rangle_{L^2(\mathscr{S}')}$.

Theorem III.3 *The Hilbert space $\mathfrak{S}$ is a reproducing kernel Hilbert space over the set $\mathscr{S}_{\sqrt{q}}$ with reproducing kernel*

$$K(\zeta,\eta)=C(\zeta)^{-1}C(\zeta-\overline{\eta})C(-\overline{\eta})^{-1},\quad \zeta,\eta\in\mathscr{S}_{\sqrt{q}}.$$

Proof First observe

$$K(\zeta,\eta)=C(-\overline{\eta})^{-1}\big(\mathcal{S}e^{-\imath\langle\cdot,\overline{\eta}\rangle_{\mathscr{S}'}}\big)(\zeta),\qquad C(-\overline{\eta})=\overline{C(\eta)}.$$

(III.7)

Thus given $\zeta, \eta \in \mathscr{S}_{\sqrt{q}}$ the inner product $\big\langle \mathcal{S}e^{\imath\langle\cdot,\zeta\rangle_{\mathscr{S}'}}, K(\cdot,\eta)\big\rangle_{\mathfrak{G}}$ is equal to

$$\begin{aligned} C(\eta)^{-1}\left\langle \mathcal{S}e^{\imath\langle\cdot,\zeta\rangle_{\mathscr{S}'}}, \mathcal{S}e^{-\imath\langle\cdot,\overline{\eta}\rangle_{\mathscr{S}'}}\right\rangle_{\mathfrak{G}} &= C(\eta)^{-1}\left\langle e^{\imath\langle\cdot,\zeta\rangle_{\mathscr{S}'}}, e^{-\imath\langle\cdot,\overline{\eta}\rangle_{\mathscr{S}'}}\right\rangle_{L^2(\mathscr{S}')} \\ &= C(\eta)^{-1}C(\zeta+\eta) = \big(\mathcal{S}e^{\imath\langle\cdot,\zeta\rangle_{\mathscr{S}'}}\big)(\eta). \end{aligned}$$

This proves the reproducing kernel property for the dense set of stochastic exponentials, and hence the theorem. □

There is a bijective correspondence between reproducing kernel Hilbert spaces and their reproducing kernels. In order to give an orthogonal decomposition of the Hilbert space $\mathfrak{G}$ we thus rewrite the kernel $K(\zeta,\eta)$ as a sum of more simple kernels. The remaining analysis contained in this paper relies of the following particular and somewhat arbitrary decomposition

$$\begin{aligned} K(\zeta,\eta) &= \exp\Big[\int_T \log\Big(\frac{(1-qe^{\imath\zeta(t)})(1-qe^{-\imath\overline{\eta}(t)})}{(1-q)(1-qe^{\imath\zeta(t)-\imath\overline{\eta}(t)})}\Big)\,\mathrm{d}t\Big] \\ &= \exp\Big[-\int_T \log\Big(1-q\frac{1-e^{\imath\zeta(t)}}{1-qe^{\imath\zeta(t)}}\frac{1-e^{-\imath\overline{\eta}(t)}}{1-qe^{-\imath\overline{\eta}(t)}}\Big)\,\mathrm{d}t\Big] \\ &= \exp\Big[\sum_{k\in\mathbb{N}} k^{-1}\int_T\Big(q\frac{1-e^{\imath\zeta(t)}}{1-qe^{\imath\zeta(t)}}\frac{1-e^{-\imath\overline{\eta}(t)}}{1-qe^{-\imath\overline{\eta}(t)}}\Big)^k\,\mathrm{d}t\Big] \\ &= \prod_{k\in\mathbb{N}}\sum_{n\in\mathbb{N}_0}(n!)^{-1}k^{-n}\Big(\int_T\Big(q\frac{1-e^{\imath\zeta(t)}}{1-qe^{\imath\zeta(t)}}\frac{1-e^{-\imath\overline{\eta}(t)}}{1-qe^{-\imath\overline{\eta}(t)}}\Big)^k\,\mathrm{d}t\Big)^n \\ &= \sum_{(n_k)_{k\in\mathbb{N}}\in\mathbb{N}_0^{\mathbb{N}}}\prod_{k\in\mathbb{N}}(n_k!)^{-1}k^{-n_k}\Big(\int_T\Big(q\frac{1-e^{\imath\zeta(t)}}{1-qe^{\imath\zeta(t)}}\frac{1-e^{-\imath\overline{\eta}(t)}}{1-qe^{-\imath\overline{\eta}(t)}}\Big)^k\,\mathrm{d}t\Big)^{n_k} \\ &= \sum_{\vec{n}\in\mathcal{I}} K_{\vec{n}}(\zeta,\eta), \end{aligned}$$

where $\mathcal{I}$ is the set of integer valued sequences with only finitely many non-zero coordinates,

$$K_{\vec{n}}(\zeta,\eta) = \prod_{k\in\mathbb{N}} q^{kn_k}k^{-n_k}(n_k!)^{-1}\left\langle\varsigma(\zeta)^k,\varsigma(\eta)^k\right\rangle_{L^2(T)}^{n_k}, \quad \zeta,\eta\in\mathscr{S}_{\sqrt{q}}, \tag{III.4}$$

and $\varsigma : \mathbb{C}\setminus\{i\log p\}\to\mathbb{C}$ and $\varsigma^{-1}:\mathbb{C}\setminus[1,q^{-1}]\to\mathbb{C}$ are defined by

$$\varsigma(w) = \frac{1-e^{\imath w}}{1-qe^{\imath w}}, \qquad \varsigma^{-1}(w) = -\imath\log\Big(\frac{1-w}{1-qw}\Big).$$

Observe the following identities, which will be of importance later,

$$\frac{qe^{\imath w}}{1-qe^{\imath w}} = \frac{q}{1-q}\big(1-\varsigma(w)\big), \qquad \frac{1}{1-qe^{\imath w}} = \frac{1-q\varsigma(w)}{1-q}. \tag{III.5}$$

By a classical result of I. Schur, the elementwise product of two positive definite matrices is again a positive definite matrix. Using this it easily follows that the kernels $K_{\vec{n}}(\zeta,\eta)$ are positive definite. Recall that $\hat{L}^2(T^m)$ is the Hilbert space of quadratic integrable, complex valued functions $f(t_1,\dots,t_m)$ that are invariant under permutations of the arguments $t_1,\dots,t_m \in T$. We need the following two technical lemmas, which are proved in the appendix.

Lemma III.4 *Let $\mathcal{F}$ be a linear subspace of $L^2(T)$ containing an orthonormal basis $\xi_n \in \mathscr{S}_{\mathbb{C}}$, $n \in \mathbb{N}$, for $L^2(T)$. Then linear mapping $\mathcal{U}_n$ given by*

$$\mathcal{U}_n(\hat{\otimes}_{j=1}^n \xi_{\nu_j})(\eta) = (n!)^{\frac{1}{2}} \Big(\prod_{i\in\mathbb{N}} |\{j : \nu_j = i\}|! \Big)^{-\frac{1}{2}} \left\langle \eta^{\otimes n} \hat{\otimes}_{j=1}^n \xi_{\nu_j} \right\rangle_{\hat{L}^1(T^n)}, \tag{III.6}$$

$\eta \in \mathcal{F}$, $\vec{\nu} \in \mathbb{N}^n$, is an isometric isomorphism from $\hat{L}^2(T^n)$ to the reproducing kernel Hilbert space generated by the kernel $\langle \zeta,\eta\rangle^n_{L^2(T)}$, $\zeta,\eta \in \mathcal{F}$.

Lemma III.5 *Given distinct $\vec{n}_1,\vec{n}_2 \in \mathcal{I}$ and $f_i \in \otimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_{ik}})$, $i=1,2$, such that $f_1 \neq 0$, there exists $\zeta \in \mathcal{C}_0^\infty(T)$ such that*

$$\sum_{i=1,2} \left\langle \otimes_{k\in\mathbb{N}} (\zeta^k)^{\otimes n_{ik}} f_i \right\rangle_{L^1(T^{\vec{n}_{i+}})} \neq 0.$$

Let $\mathfrak{S}_{\vec{n}}$, $\vec{n} \in \mathcal{I}$, be the reproducing kernel Hilbert space over the set $\mathscr{S}_{\sqrt{q}}$ generated by the kernel $K_{\vec{n}}(\zeta,\eta)$ defined in (III.4).

Lemma III.6 *Let $\xi_n \in \mathscr{S}_{\mathbb{C}}$, $n \in \mathbb{N}$, be an orthonormal basis for $L^2(T)$. Then the linear mapping $\mathcal{U}_{\vec{n}}$ given by*

$$\begin{aligned} &\mathcal{U}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \xi_{\nu_{kj}} \big) \\ &\quad = \prod_{k\in\mathbb{N}} q^{\frac{kn_k}{2}} k^{-\frac{n_k}{2}} \Big(\prod_{i\in\mathbb{N}} |\{j : \nu_{kj} = i\}|! \Big)^{-\frac{1}{2}} \left\langle \big(\varsigma(\cdot)^k\big)^{\otimes n_k} \hat{\otimes}_{j=1}^{n_k} \xi_{\nu_{kj}} \right\rangle_{\hat{L}^1(T^{n_k})}, \end{aligned} \tag{III.7}$$

$\vec{\nu}_k \in \mathbb{N}^{n_k}$, $k \in \mathbb{N}$, is an isometric isomorphism from $\otimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_k})$ to $\mathfrak{S}_{\vec{n}}$.

Proof We first extend the kernel $K_{\vec{n}}(\zeta,\eta)$ on $\mathscr{S}_{\sqrt{q}}$ to the product set $\mathscr{S}_{\sqrt{q}}^{\mathbb{N}}$, *i.e.* let the kernel $\tilde{K}_{\vec{n}}(\vec{\zeta},\vec{\eta})$ be given by

$$\tilde{K}_{\vec{n}}(\vec{\zeta},\vec{\eta}) = \prod_{k\in\mathbb{N}} q^{kn_k} k^{-n_k} (n_k!)^{-1} \left\langle \varsigma(\zeta_k)^k, \varsigma(\eta_k)^k \right\rangle^{n_k}_{L^2(T)}, \quad \vec{\zeta},\vec{\eta} \in \mathscr{S}_{\sqrt{q}}^{\mathbb{N}}.$$

By lemma III.4 the linear mapping $\mathcal{U}_n$ defined in (III.6) is an isometric isomorphism from $\hat{L}^2(T^n)$ to the reproducing kernel Hilbert space generated by the kernel $\langle \zeta,\eta\rangle^n_{L^2(T)}$. Moreover, by Aronszajn (1950, theorem I.8.I) the linear mapping $\tilde{\mathcal{U}}_{\vec{n}}$ determined by

$$\begin{aligned} &\tilde{\mathcal{U}}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \xi_{\nu_{kj}} \big)(\vec{\eta}) \\ &\quad = \prod_{k\in\mathbb{N}} q^{\frac{kn_k}{2}} k^{-\frac{n_k}{2}} \Big(\prod_{i\in\mathbb{N}} |\{j : \nu_{kj} = i\}|! \Big)^{-\frac{1}{2}} \left\langle \big(\varsigma(\eta_k)^k\big)^{\otimes n_k} \hat{\otimes}_{j=1}^{n_k} \xi_{\nu_{kj}} \right\rangle_{\hat{L}^1(T^{n_k})}, \end{aligned}$$

(III.9)

$\vec{\nu}_k \in \mathbb{N}^{n_k}$, $k \in \mathbb{N}$, thus is an isometric isomorphism from $\otimes_{k\in\mathbb{N}}\hat{L}^2(T^{n_k})$ to the reproducing kernel Hilbert space $\tilde{\mathfrak{S}}_{\vec{n}}$ generated by the kernel $\tilde{K}_{\vec{n}}(\vec{\zeta},\vec{\eta})$. Given $f \in \otimes_{k\in\mathbb{N}}\hat{L}^2(T^{n_k})$, $f \neq 0$, there exists by lemma III.5 a vector $\vec{\eta} = (\eta)_{k\in\mathbb{N}}$, $\eta \in \mathscr{S}_{\sqrt{q}}$, such that $\tilde{\mathcal{U}}_{\vec{n}}(f)(\vec{\eta}) \neq 0$. The linear mapping from $\tilde{\mathfrak{S}}_{\vec{n}}$ to $\mathfrak{S}_{\vec{n}}$ given by restricting the functionals to the diagonal is thus injective, and (III.7) follows by Aronszajn (1950, theorem I.5). □

Remark Let two random variables $\phi, \psi \in L^2(\mathscr{S}')$ be given. If the pointwise product $\mathcal{S}(\phi)\cdot\mathcal{S}(\psi)$ belongs to the space $\mathfrak{S}$, then the Wick product $\phi \diamond \psi$ is defined by

$$\mathcal{S}(\phi \diamond \psi) = \mathcal{S}(\phi)\cdot\mathcal{S}(\psi).$$

It follows directly from lemma III.6, *cf.* also theorem III.10 below, that the Wick product is easily expressed in the introduced structure. We will however not discuss the Wick product further.

We are now in position to formulate and prove the main result of this section.

Theorem III.8 *The Hilbert space $\mathfrak{S}$ has a orthogonal sum decomposition*

$$\mathfrak{S} = \bigoplus_{\vec{n}\in\mathcal{I}} \mathfrak{S}_{\vec{n}} \simeq \bigoplus_{\vec{n}\in\mathcal{I}} \bigotimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_k}),$$

where the isomorphism between $\mathfrak{S}_{\vec{n}}$ and $\otimes_{k\in\mathbb{N}}\hat{L}^2(T^{n_k})$ is given by (III.7).

Proof It only remains to prove orthogonality of the spaces $\mathfrak{S}_{\vec{n}}$, $\vec{n} \in \mathcal{I}$. By Aronszajn (1950, theorem I.6) it suffices to prove that $\mathfrak{S}_{\vec{n}_1} \cap \mathfrak{S}_{\vec{n}_2} = \{0\}$ for given distinct $\vec{n}_1, \vec{n}_2 \in \mathcal{I}$. If $\phi \in \mathfrak{S}_{\vec{n}_1} \cap \mathfrak{S}_{\vec{n}_2}$, then by lemma III.6 there exist $f_i \in \otimes_{k\in\mathbb{N}}\hat{L}^2(T^{n_{ik}})$, $i = 1, 2$, such that

$$\phi = \left\langle \otimes_{k\in\mathbb{N}} \big(\varsigma(\cdot)^k\big)^{\otimes n_{ik}} f_i \right\rangle_{L^1(T^{\vec{n}_i+})}, \quad i = 1, 2.$$

By lemma III.5, $f_1 = 0$ since

$$0 = \phi - \phi = \sum_{i=1,2} \left\langle \otimes_{k\in\mathbb{N}} \big(\varsigma(\cdot)^k\big)^{\otimes n_{ik}} (-1)^{1+i} f_i \right\rangle_{L^1(T^{\vec{n}_i+})},$$

and hence $\phi = 0$. □

Remark Consider the subspaces $\mathfrak{S}_{(n)}$, $n \in \mathbb{N}_0$, and their corresponding kernels $K_{(n)}(\zeta, \eta)$ given by

$$\mathfrak{S}_{(n)} = \bigoplus_{\vec{n}\in\mathcal{I}_{(n)}} \mathfrak{S}_{\vec{n}}, \qquad K_{(n)}(\zeta,\eta) = \sum_{\vec{n}\in\mathcal{I}_{(n)}} K_{\vec{n}}(\zeta,\eta). \tag{III.10}$$

Given $\zeta \in \mathscr{S}_{\sqrt{q}}$ and $v \in [0,1]$ let $\zeta_v = \varsigma^{-1}(v\varsigma(\zeta)) \in \mathscr{S}_{\sqrt{q}}$. Then

$$\sum_{m=0}^{\infty} v^m K_{(m)}(\cdot,\zeta) = \sum_{m=0}^{\infty} K_{(m)}(\cdot,\zeta_v) = K(\cdot,\zeta_v).$$

Thus given $\zeta, \eta \in \mathscr{S}_{\sqrt{q}}$ and $v, w \in [0,1]$,

$$\begin{aligned}\sum_{m,n=0}^{\infty} v^m w^n \left\langle K_{(m)}(\cdot,\zeta), K_{(n)}(\cdot,\eta)\right\rangle_{\mathfrak{S}} &= \left\langle \sum_{m=0}^{\infty} v^m K_{(m)}(\cdot,\zeta), \sum_{n\in\mathbb{N}_0} w^n K_{(n)}(\cdot,\eta) \right\rangle_{\mathfrak{S}} \\ &= \langle K(\cdot,\zeta_v), K(\cdot,\eta_w)\rangle_{\mathfrak{S}} = K(\eta_w,\zeta_v) = \sum_{n\in\mathbb{N}_0} K_{(n)}(\eta_w,\zeta_v) \\ &= \sum_{n\in\mathbb{N}_0} v^n w^n K_{(n)}(\eta,\zeta).\end{aligned}$$

Comparing the coefficients of the monomials $v^m w^n$, $m, n \in \mathbb{N}_0$, it follows that

$$\left\langle K_{(m)}(\cdot,\zeta), K_{(n)}(\cdot,\eta)\right\rangle_{\mathfrak{S}} = \begin{cases} K_{(n)}(\eta,\zeta) & \text{for } m = n, \\ 0 & \text{for } m \neq n. \end{cases}$$

This reprove the orthogonality of the subspaces $\mathfrak{S}_{(n)}$, $n \in \mathbb{N}_0$, using direct calculations on the corresponding kernels. For more simple processes, *e.g.* the Poisson processes as treated in Itô (1988), this method might suffice to prove orthogonality of the component spaces, and the use of lemma III.5 can in such cases be avoided.

III.4 The chaos decomposition

In this section we derive a chaos decomposition of the negative binomial probability space, *i.e.* we find a orthonormal basis for the space $L^2(\mathscr{S}')$. The analysis done in section III.3 gives isometric isomorphisms $\mathcal{S}$ and $\mathcal{U} = \oplus_{\vec{n}\in\mathcal{I}}\mathcal{U}_{\vec{n}}$ between the Hilbert spaces $L^2(\mathscr{S}')$ and $\mathfrak{S}$ respectively $\oplus_{\vec{n}\in\mathcal{I}} \otimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_k})$ and $\mathfrak{S}$. To find the chaos decomposition of $L^2(\mathscr{S}')$ we thus need to find the isomorphism

$$\mathcal{V} = \mathcal{S}^{-1} \circ \mathcal{U} = \oplus_{\vec{n}\in\mathcal{I}}\mathcal{V}_{\vec{n}}, \quad \mathcal{V}_{\vec{n}} = \mathcal{S}^{-1} \circ \mathcal{U}_{\vec{n}},$$

that fits into the commuting diagram

$$\begin{tikzcd} L^2(\mathscr{S}') \arrow[r, "\mathcal{S}"] & \mathfrak{S}. \\ \bigoplus_{\vec{n}\in\mathcal{I}} \bigotimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_k}) \arrow[u, "\mathcal{V}"] \arrow[ur, "\mathcal{U}"'] & \end{tikzcd}$$

In this section we describe a direct approach to find the isomorphisms $\mathcal{V}_{\vec{n}}$, and in subsection III.4.1 we discuss the minor role played by the multiple integrals. In subsection III.4.2 we calculate the chaos expansion for some polynomial functionals without appealing to theorem III.10. These expansions can be inverted, and thus describe the isomorphisms $\mathcal{V}_{\vec{n}}$ indirectly.

(III.11)

Theorem III.10 *Given $\vec{n} \in \mathcal{I}_{(n)}$ and $\eta_{kj} \in \mathscr{S}$, $k \in \mathbb{N}$, $j = 1, \dots, n_k$, the random variable*

$$\mathcal{V}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \eta_{kj}\big) \tag{III.8a}$$

is equal to the sum over all $\vec{n}_1 \preceq \vec{n}$, all $\mathcal{F}_k \subseteq \{1, \dots, n_k\}$ with $|\mathcal{F}_k| = n_{1k}$, and all $\vec{n}_{kj} \in \mathcal{I}_{(k)}$, $k \in \mathbb{N}$, $j \in \complement\mathcal{F}_k = \{1, \dots, n_k\} \setminus \mathcal{F}_k$, of the terms

$$\begin{aligned}
&\bigg(\prod_{i\in\mathbb{N}} i^{\frac{n_i}{2}-n_{1i}-n_{2i}}(n_i!)^{-\frac{1}{2}}\Big(\prod_{k\in\mathbb{N}}\prod_{j\in\complement\mathcal{F}_k} n_{kji}!\Big)^{-1} q^{-\frac{in_i}{2}+in_{1i}}(q^i-1)^{n_{2i}}\bigg) \\
&\cdot\bigg(\prod_{k\in\mathbb{N}}\prod_{j\in\mathcal{F}_k}\langle\eta_{kj}\rangle_{L^1(T)}\bigg)\bigg(\prod_{k\in\mathbb{N}}\prod_{j\in\complement\mathcal{F}_k}\langle\cdot,\eta_{kj}\rangle_{\mathscr{S}',\vec{n}_{kj+}} - \sum_{\mathcal{G}_k\subseteq\complement\mathcal{F}_k:\sum_k|\mathcal{G}_k|\geq 2}(-1)^{\sum_k|\mathcal{G}_k|} \\
&\Big(\sum_{k\in\mathbb{N}}|\mathcal{G}_k|-1\Big)\Big\langle\cdot,\prod_{k\in\mathbb{N}}\prod_{j\in\mathcal{G}_k}\eta_{kj}\Big\rangle_{\mathscr{S}',\sum_{k\in\mathbb{N},j\in\mathcal{G}_k}\vec{n}_{kj+}}\prod_{k\in\mathbb{N}}\prod_{j\in(\complement\mathcal{F}_k)\setminus\mathcal{G}_k}\langle\cdot,\eta_{kj}\rangle_{\mathscr{S}',\vec{n}_{kj+}}\bigg),
\end{aligned} \tag{III.8b}$$

where $\vec{n}_2 = \sum_{k,j}\vec{n}_{kj}$.

Proof For $w \in [0, 1]$, $\eta \in \mathscr{S}_{\sqrt{q}}$ the identity

$$\mathcal{S}^{-1}K\big(\cdot, -\overline{\varsigma^{-1}(w\eta)}\big) = C\big(\varsigma^{-1}(w\eta)\big)^{-1}e^{\imath\langle\cdot,\varsigma^{-1}(w\eta)\rangle_{\mathscr{S}'}} \tag{III.9}$$

follows from theorem III.3. Since

$$\begin{aligned}
K_{\vec{n}}\big(\cdot, -\overline{\varsigma^{-1}(\eta)}\big) &= \prod_{k\in\mathbb{N}} q^{kn_k}k^{-n_k}(n_k!)^{-1}\big\langle\varsigma(\cdot)^k\eta^k\big\rangle_{L^1(T)}^{n_k} \\
&= \Big(\prod_{k\in\mathbb{N}} q^{\frac{kn_k}{2}}k^{-\frac{n_k}{2}}(n_k!)^{-\frac{1}{2}}\Big)\mathcal{U}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}}(\eta^k)^{\otimes n_k}\big),
\end{aligned}$$

the left hand side of (III.9) is given by

$$\begin{aligned}
\sum_{\vec{n}\in\mathcal{I}}\mathcal{S}^{-1}K_{\vec{n}}\big(\cdot, -\overline{\varsigma^{-1}(w\eta)}\big) &= \sum_{\vec{n}\in\mathcal{I}} w^{\vec{n}_*}\mathcal{S}^{-1}K_{\vec{n}}\big(\cdot, -\overline{\varsigma^{-1}(\eta)}\big) \\
&= \sum_{n\in\mathbb{N}_0} w^n \sum_{\vec{n}\in\mathcal{I}_{(n)}}\Big(\prod_{k\in\mathbb{N}} q^{\frac{kn_k}{2}}k^{-\frac{n_k}{2}}(n_k!)^{-\frac{1}{2}}\Big)\mathcal{V}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}}(\eta^k)^{\otimes n_k}\big).
\end{aligned}$$

The two factors on the right hand side of (III.9) are respectively given by

$$\begin{aligned}
C\big(\varsigma^{-1}(w\eta)\big)^{-1} &= \exp\Big\langle\log\Big(\frac{1-qe^{\imath\varsigma^{-1}(w\eta)}}{1-q}\Big)\Big\rangle_{L^1(T)} \\
&= \exp\langle-\log(1-qw\eta)\rangle_{L^1(T)} = \exp\Big\langle\sum_{k\in\mathbb{N}} k^{-1}q^k w^k\eta^k\Big\rangle_{L^1(T)} \\
&= \sum_{\vec{n}\in\mathcal{I}} w^{\vec{n}_*}\prod_{k\in\mathbb{N}} q^{kn_k}k^{-n_k}(n_k!)^{-1}\big\langle\eta^k\big\rangle_{L^1(T)}^{n_k},
\end{aligned}$$

(III.12)

and

$$e^{\imath\left\langle \cdot,\varsigma^{-1}(w\eta)\right\rangle_{\mathscr{S}'}} = \exp\left\langle \cdot, \log\left(\frac{1-w\eta}{1-qw\eta}\right)\right\rangle_{\mathscr{S}'} = \exp\left\langle \cdot, \sum_{k\in\mathbb{N}} k^{-1}(q^k-1)w^k\eta^k\right\rangle_{\mathscr{S}'}$$
$$= \sum_{\vec{n}\in\mathcal{I}} w^{\vec{n}_*} \prod_{k\in\mathbb{N}} (q^k-1)^{n_k} k^{-n_k}(n_k!)^{-1}\left\langle \cdot, \eta^k\right\rangle_{\mathscr{S}'}.$$

Multiplying these two factors and comparing the coefficients of w^n, $n\in\mathbb{N}_0$, on both sides of (III.9), we see that the random variable

$$\sum_{\vec{n}\in\mathcal{I}_{(n)}} \Big(\prod_{k\in\mathbb{N}} q^{\frac{kn_k}{2}} k^{-\frac{n_k}{2}}(n_k!)^{-\frac{1}{2}}\Big)\mathcal{V}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}}(\eta^k)^{\otimes n_k}\big) \tag{III.10a}$$

is equal to

$$\sum_{\vec{n}_1+\vec{n}_2\in\mathcal{I}_{(n)}} \prod_{k\in\mathbb{N}} q^{kn_{1k}}(q^k-1)^{n_{2k}} k^{-n_{1k}-n_{2k}}(n_{1k}!n_{2k}!)^{-1}\left\langle \eta^k\right\rangle_{L^1(T)}^{n_{1k}}\left\langle \cdot,\eta^k\right\rangle_{\mathscr{S}'}^{n_{2k}}. \tag{III.10b}$$

By an approximation argument it easily follows that this equation holds for functions $\eta\in L^1(T)\cap L^\infty(T)$ too.

Let $\vec{n}_0\in\mathcal{I}_{(n)}$ and $\eta_{kj}\in\mathscr{S}$ with compact support, $k\in\mathbb{N}$, $j=1,\dots,n_{0k}$, be fixed. In order to separate the term with $\vec{n}=\vec{n}_0$ in the sum (III.10a), we will use the fact that $\vec{n}_0$ is characterized in $\mathcal{I}_{(n)}$ by

$$\{\vec{n}_0\} = \Big\{\vec{n}=\sum_{k\in\mathbb{N}}\sum_{j=1}^{n_{0k}} \vec{n}_{kj}\in\mathcal{I}_{(n)} : \vec{n}_{kj}\in\mathcal{I}_{(k)}\Big\}\cap\big\{\vec{n}\in\mathcal{I}_{(n)} : \vec{n}_+\le\vec{n}_{0+}\big\}.$$

The multi indices in the first set can be isolated via a polarization argument, and the multi indices in the second set can be isolated via dimensionality.

We start with the polarization argument. Thus let C_{Nm}, $m\in\mathbb{N}$, be successively finer partitions of T, $N\in\mathbb{N}$, such that $\sup_{m\in\mathbb{N}}\langle 1_{C_{Nm}}\rangle_{L^1(T)}$ vanishes as $N\to\infty$, and let $t_{Nm}\in C_{Nm}$ be given. Moreover given $N\in\mathbb{N}$, pairwise different $m_{kj}\in\mathbb{N}$, $k\in\mathbb{N}$, $j=1,\dots,n_{0k}$, and $\vec{\sigma}\in\{-1,+1\}^n$ let

$$\eta_{N,\vec{m},\vec{\sigma}} = \sum_{i=1}^{n}\sigma_i\zeta_{N,\vec{m},i},$$

where exactly k of the functions $\zeta_{N,\vec{m},i}$, $i=1,\dots,n$, equals $\eta_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}$. Using polarization we see that

$$2^{-n}\sum_{\vec{\sigma}\in\{-1,+1\}^n}\Big(\prod_{i=1}^{n}\sigma_i\Big)\sum_{\vec{n}\in\mathcal{I}_{(n)}}\Big(\prod_{i\in\mathbb{N}} q^{\frac{in_i}{2}} i^{-\frac{n_i}{2}}(n_i!)^{-\frac{1}{2}}\Big)\mathcal{V}_{\vec{n}}\big(\otimes_{i\in\mathbb{N}}(\eta^i_{N,\vec{m},\vec{\sigma}})^{\otimes n_i}\big) \tag{III.11a}$$

(III.13)

is equal to

$$\sum_{\vec{n}_{kj}\in\mathcal{I}_{(k)},k\in\mathbb{N},j=1,\dots,n_{0k};\vec{n}=\sum_{k,j}\vec{n}_{kj}} \Gamma_{\vec{n}}\cdot\Big(\prod_{i\in\mathbb{N}} q^{\frac{in_i}{2}} i^{-\frac{n_i}{2}} (n_i!)^{-\frac{1}{2}}\Big)$$
$$\mathcal{V}_{\vec{n}}\bigg(\otimes_{i\in\mathbb{N}}\Big(\hat{\otimes}_{k\in\mathbb{N}}\hat{\otimes}_{j=1}^{n_{0k}}\big(\eta^i_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}\big)^{\otimes n_{kji}}\Big)\bigg), \quad \text{(III.11b)}$$

where $\Gamma_{\vec{n}}$ is a combinatorial constant counting the number of times the factor in (III.11b) occurs when the functions $\zeta_{N,\vec{m},l}$, $l=1,\dots,n$, are permuted. Actually $\Gamma_{\vec{n}}$ is sloppy notation since this quantity depends on all the multi indices $\vec{n}_{kj}$, but we stick to it for convenience. To compute $\Gamma_{\vec{n}}$ we first subdivide the numbers of the k functions $\zeta_{N,\vec{m},l}$ which equal $\eta_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}$ into $\vec{n}_{kj+}$ subsets A_{kjih}, $i\in\mathbb{N}$, $h=1,\dots,n_{kji}$, such that A_{kjih} has size i. For each $k\in\mathbb{N}$, $j=1,\dots,n_{0k}$, $i\in\mathbb{N}$, the subsets A_{kjih}, $h=1,\dots,n_{kji}$, have already been permuted. We thus continue by permuting for each $i\in\mathbb{N}$ the subsets A_{kjih}, $k\in\mathbb{N}$, $j=1,\dots,n_{0k}$, $h=1,\dots,n_{kji}$, subject to the constraint that subsets A_{kjih} with the same value of (k,j) are kept in a fixed order, and finally we permute the elements inside the subsets A_{kjih}. Whence,

$$\begin{aligned}\Gamma_{\vec{n}} &= \Big(\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}}\frac{k!}{\prod_{i\in\mathbb{N}}(i!)^{n_{kji}}}\Big)\Big(\prod_{i\in\mathbb{N}}\frac{n_i!}{\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}} n_{kji}!}\Big)\Big(\prod_{i\in\mathbb{N}}(i!)^{n_i}\Big)\\ &= \Big(\prod_{k\in\mathbb{N}}(k!)^{n_{0k}}\Big)\Big(\prod_{i\in\mathbb{N}}\frac{n_i!}{\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}} n_{kji}!}\Big).\end{aligned}$$

The argument of the operator $\mathcal{V}_{\vec{n}}$ is a function of $\vec{n}_+$ variables. Hence if $\vec{n}$ is one of the multi indices in the sum (III.11b) with $\vec{n}_+ > \vec{n}_{0+}$, then the sum over pairwise different $m_{kj}\in\mathbb{N}$, $k\in\mathbb{N}$, $j=1,\dots,n_{0k}$, of the term corresponding to $\vec{n}$ vanishes in $L^2(\mathscr{S}')$ as $N\to\infty$. Since $\vec{n}_0$, *i.e.* when $\vec{n}_{kj}=\vec{\varepsilon}_k$, is the only multi index in the sum (III.11b) with $\vec{n}_+=\vec{n}_{0+}$, it follows by using the identity

$$\otimes_{k\in\mathbb{N}}\hat{\otimes}_{j=1}^{n_{0k}}\eta^k_{kj} = \lim_{N\to\infty}\sum_{\substack{m_{kj}\text{ pairwise}\\ \text{different}}}\otimes_{k\in\mathbb{N}}\hat{\otimes}_{j=1}^{n_{0k}}\eta^k_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}$$

in $\otimes_{k\in\mathbb{N}}\hat{L}^2(T^{n_{0k}})$, that the limit in $L^2(\mathscr{S}')$ as $N\to\infty$ of the sum over pairwise different $m_{kj}\in\mathbb{N}$ of the random variables (III.11a) equals

$$\underbrace{\Big(\prod_{i\in\mathbb{N}}(i!)^{n_{0i}}n_{0i}!\Big)}_{\Gamma_{\vec{n}_0}}\Big(\prod_{i\in\mathbb{N}} q^{\frac{in_{0i}}{2}} i^{-\frac{n_{0i}}{2}}(n_{0i}!)^{-\frac{1}{2}}\Big)\mathcal{V}_{\vec{n}_0}\Big(\otimes_{i\in\mathbb{N}}\hat{\otimes}_{j=1}^{n_{0i}}\eta^i_{ij}\Big). \quad \text{(III.12)}$$

Now we use the same procedure on the random variable (III.10b). Polarizing (III.10b) we similarly find that

$$2^{-n}\sum_{\vec{\sigma}\in\{-1,+1\}^n}\Big(\prod_{i=1}^n\sigma_i\Big)\sum_{\vec{n}_1+\vec{n}_2\in\mathcal{I}_{(n)}}\prod_{i\in\mathbb{N}} q^{in_{1i}}(q^i-1)^{n_{2i}} i^{-n_{1i}-n_{2i}}$$
$$(n_{1i}!n_{2i}!)^{-1}\left\langle\eta^i_{N,\vec{m},\vec{\sigma}}\right\rangle^{n_{1i}}_{L^1(T)}\left\langle\cdot,\eta_{N,\vec{m},\vec{\sigma}}\right\rangle^{n_{2i}}_{\mathscr{S}',1} \quad \text{(III.13a)}$$

(III.14)

is equal to

$$\sum_{\substack{\vec{n}_{1kj}+\vec{n}_{2kj}\in\mathcal{I}_{(k)},k\in\mathbb{N},j=1,\dots,n_{0k};\\ \vec{n}_l=\sum_{k,j}\vec{n}_{lkj},l=1,2}} \Gamma_{\vec{n}_1,\vec{n}_2}\prod_{i\in\mathbb{N}} q^{in_{1i}}(q^i-1)^{n_{2i}}i^{-n_{1i}-n_{2i}}(n_{1i}!n_{2i}!)^{-1}$$
$$\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}}\left\langle \eta^i_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}\right\rangle_{L^1(T)}^{n_{1kji}}\left\langle \cdot,\eta^i_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}\right\rangle_{\mathscr{S}',1}^{n_{2kji}}, \tag{III.13b}$$

where $\Gamma_{\vec{n}_1,\vec{n}_2}$ is a combinatorial constant counting the number of times the factor in (III.13b) occurs when the functions $\zeta_{N,\vec{m},l}$, $l=1,\dots,n$, are permuted. By an argument similar to the one above, we find that

$$\begin{aligned}\Gamma_{\vec{n}_1,\vec{n}_2} &= \left(\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}}\frac{k!}{\prod_{i\in\mathbb{N}}(i!)^{n_{1kji}+n_{2kji}}}\right)\left(\prod_{l=1,2}\prod_{i\in\mathbb{N}}\frac{n_{li}!}{\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}}n_{lkji}!}(i!)^{n_{li}}\right)\\ &= \left(\prod_{k\in\mathbb{N}}(k!)^{n_{0k}}\right)\left(\prod_{l=1,2}\prod_{i\in\mathbb{N}}\frac{n_{li}!}{\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}}n_{lkji}!}\right).\end{aligned}$$

If there is at most one jump of the underlying Poisson process inside each of the sets C_{Nm} intersecting the support of the functions η_{kj}, then (III.13b) is equal to

$$\sum_{\substack{\vec{n}_{1kj}+\vec{n}_{2kj}\in\mathcal{I}_{(k)},k\in\mathbb{N},j=1,\dots,n_{0k};\\ \vec{n}_l=\sum_{k,j}\vec{n}_{lkj},l=1,2}} \Gamma_{\vec{n}_1,\vec{n}_2}\left(\prod_{i\in\mathbb{N}} q^{in_{1i}}(q^i-1)^{n_{2i}}i^{-n_{1i}-n_{2i}}(n_{1i}!n_{2i}!)^{-1}\right.$$
$$\left.\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_{0k}}\left\langle \eta^i_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}\right\rangle_{L^1(T)}^{n_{1kji}}\right)\prod_{(k,j):\vec{n}_{2kj}\neq 0}\left\langle \cdot,\eta^{\vec{n}_{2kj}*}_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}\right\rangle_{\mathscr{S}',\vec{n}_{2kj}+}. \tag{III.14}$$

Using the same reasoning as above, *i.e.* counting the dimension of the arguments, it follows that the limit in $L^2(\mathscr{S}')$ as $N\to\infty$ of the sum over pairwise different m_{kj} of the random variables (III.14) equals

$$\sum_{\substack{\vec{n}_1\preceq\vec{n}_0\\ \mathcal{F}_k\subseteq\{1,\dots,n_{0k}\},|\mathcal{F}_k|=n_{1k}\\ \vec{n}_{2kj}\in\mathcal{I}_{(k)},k\in\mathbb{N},j\in\complement\mathcal{F}_k}} \underbrace{\left(\prod_{i\in\mathbb{N}}(i!)^{n_{0i}}n_{1i}!\frac{n_{2i}!}{\prod_{k\in\mathbb{N}}\prod_{j\in\complement\mathcal{F}_k}n_{2kji}!}\right)}_{\Gamma_{\vec{n}_1,\vec{n}_2}}$$
$$\left(\prod_{i\in\mathbb{N}} q^{in_{1i}}(q^i-1)^{n_{2i}}i^{-n_{1i}-n_{2i}}(n_{1i}!n_{2i}!)^{-1}\prod_{j\in\mathcal{F}_i}\left\langle\eta^i_{ij}\right\rangle_{L^1(T)}\right) \tag{III.15}$$
$$\left(\lim_{N\to\infty}\sum_{\substack{m_{kj},k\in\mathbb{N},j\in\complement\mathcal{F}_k\\ \text{pairwise different}}}\prod_{k\in\mathbb{N}}\prod_{j\in\complement\mathcal{F}_k}\left\langle\cdot,\eta^k_{kj}(t_{Nm_{kj}})1_{C_{Nm_{kj}}}\right\rangle_{\mathscr{S}',\vec{n}_{2kj}+}\right),$$

(III.15)

where $\vec{n}_2 = \sum_{k,j} \vec{n}_{2kj}$. Using the inclusion-exclusion type algebraic identity

$$\sum_{\substack{m_k \text{ pairwise} \\ \text{different}}} \prod_{k=1}^{K} f_k(t_{m_k}) = \prod_{k=1}^{K} \sum_{m\in\mathbb{N}} f_k(t_m) - \sum_{\substack{\mathcal{G},\mathcal{H} \text{ partition of} \\ \{1,\dots,K\}, |\mathcal{G}|\geq 2}} (-1)^{|\mathcal{G}|}(|\mathcal{G}|-1)\Big(\sum_{m\in\mathbb{N}} \prod_{k\in\mathcal{G}} f_k(t_m)\Big)\Big(\prod_{k\in\mathcal{H}} \sum_{m\in\mathbb{N}} f_k(t_m)\Big),$$

we see that the last factor in (III.15) is equal to

$$\begin{aligned}\Bigg(\prod_{k\in\mathbb{N}} \prod_{j\in\complement\mathcal{F}_k} \langle\cdot,\eta_{kj}^k\rangle_{\mathscr{S}',\vec{n}_{2kj+}} - \sum_{\mathcal{G}_k\subseteq\complement\mathcal{F}_k : \sum_k|\mathcal{G}_k|\geq 2} (-1)^{\sum_k|\mathcal{G}_k|}\Big(\sum_{k\in\mathbb{N}} |\mathcal{G}_k| - 1\Big) \\ \Big\langle\cdot, \prod_{k\in\mathbb{N}}\prod_{j\in\mathcal{G}_k} \eta_{kj}^k\Big\rangle_{\mathscr{S}',\sum_{k\in\mathbb{N},j\in\mathcal{G}_k}\vec{n}_{2kj+}} \prod_{k\in\mathbb{N}} \prod_{j\in(\complement\mathcal{F}_k)\setminus\mathcal{G}_k} \langle\cdot,\eta_{kj}^k\rangle_{\mathscr{S}',\vec{n}_{2kj+}}\Bigg).\end{aligned} \tag{III.16}$$

Since the probability that for fixed $N\in\mathbb{N}$ there exist pairwise different m_{kj}, $k\in\mathbb{N}$, $j=1,\dots,n_{0k}$, such that (III.13b) and (III.14) differs is bounded by

$$\begin{aligned}&\sum_{m\in\mathbb{N}: C_{Nm}\cap(\cup_{k,j}\operatorname{supp}(\eta_{kj}))\neq\emptyset} 1-(1+\kappa)e^{-\kappa}\Big|_{\kappa=-\log(1-q)\cdot\langle 1_{C_{Nm}}\rangle_{L^1(T)}} \\ &\qquad\leq \big(\log(1-q)\big)^2 \sup_{m\in\mathbb{N}} \langle 1_{C_{Nm}}\rangle_{L^1(T)} \sum_{m\in\mathbb{N}: C_{1m}\cap(\cup_{k,j}\operatorname{supp}(\eta_{kj}))\neq\emptyset} \langle 1_{C_{1m}}\rangle_{L^1(T)},\end{aligned}$$

which vanishes as $N\to\infty$, the two $L^2(\mathscr{S}')$ limits (III.12) and (III.15) are equal. After replacing η_{kj}^k by η_{kj}, inserting (III.16) in (III.15), and simplifying and rearranging the factors, the statement of the theorem follows for $\vec{n}=\vec{n}_0$ and functions $\eta_{kj}\in\mathscr{S}$ with compact support. Since the functions $\eta\in\mathscr{S}$ with compact support are dense, the theorem follows. □

Remark The argument allowing us to approximate the random variable (III.13b) by (III.14) depends on the fact, that the Lévy process under consideration, *i.e.* the negative binomial process, is a compound Poisson process with bounded Lévy measure. If we were to consider a Lévy process with infinite Lévy measure, *e.g.* a Γ-process, then care had to be taken at exactly this point.

If we take a close look at equation (III.8b), we see that the random variable $\mathcal{V}_{\vec{n}_0}(\otimes_{k\in\mathbb{N}}\hat{\otimes}_{j=1}^{n_{0k}}\eta_{kj})$ is a linear combination of random variables of the form

$$\prod_{(k,j)\in\mathcal{F}} \langle\cdot,\eta_{kj}\rangle_{\mathscr{S}',\alpha_{kj}} \quad\text{or}\quad \Big\langle\cdot, \prod_{(k,j)\in\mathcal{G}} \eta_{kj}\Big\rangle_{\mathscr{S}',\beta} \prod_{(k,j)\in\mathcal{F}} \langle\cdot,\eta_{kj}\rangle_{\mathscr{S}',\alpha_{kj}}, \tag{III.17}$$

(III.16)

where the coefficients are polynomials in the numbers $\langle \eta_{kj} \rangle_{L^1(T)}$. If the set $\mathcal{I}_{\vec{n}_0}$ is defined by

$$\mathcal{I}_{\vec{n}_0} = \left\{ \vec{n} \in \mathcal{I} : \vec{n}_* < \vec{n}_{0*}, \vec{n}_+ \leq \vec{n}_{0+} \text{ or } \vec{n}_* \leq \vec{n}_{0*}, \vec{n}_+ < \vec{n}_{0+} \right\},$$

then it is easily seen that

$$\prod_{k \in \mathbb{N}} \prod_{j=1}^{n_{0k}} \langle \cdot, \eta_{kj} \rangle_{\mathscr{S}', (k\vec{\varepsilon}_1)_+} = \prod_{k \in \mathbb{N}} \prod_{j=1}^{n_{0k}} \langle \cdot, \eta_{kj} \rangle_{\mathscr{S}', k} \tag{III.18}$$

is the only random variable of the type (III.17) which does not occur in the formula for $\mathcal{V}_{\vec{n}}(\otimes_{k \in \mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \zeta_{kj})$ for some $\vec{n} \in \mathcal{I}_{\vec{n}_0}$ and some products ζ_{kj} of the functions η_{kj}. Consequently random variables of the form (III.18) are linearly independent for different multi indices $\vec{n}_0$ or linearly independent elements $\otimes_{k \in \mathbb{N}} \hat{\otimes}_{j=1}^{n_{0k}} \eta_{kj}$ in $\otimes_{k \in \mathbb{N}} \hat{L}^2(T^{n_{0k}})$. Thus if $\otimes_{k \in \mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \eta_{\nu_{kj}}$, $\vec{\nu} \in \mathcal{J}_{\vec{n}}$, is an orthonormal basis for $\otimes_{k \in \mathbb{N}} \hat{L}^2(T^{n_{0k}})$, then every $\phi \in L^2(\mathscr{S}')$ has a unique representation

$$\phi = \alpha_0 + \sum_{\vec{n} \in \mathcal{I} \setminus \{0\}} \sum_{\vec{\nu} \in \mathcal{J}_{\vec{n}}} \alpha_{\vec{n}, \vec{\nu}} \prod_{k \in \mathbb{N}} \prod_{j=1}^{n_k} \langle \cdot, \eta_{\nu_{kj}} \rangle_{\mathscr{S}', k}.$$

From this construction it follows that we can define an operator

$$\Delta : L^2(\mathscr{S}') \to L^2(\mathscr{S}')$$

by

$$\Delta \Big(\prod_{k \in \mathcal{F}} \langle \cdot, \eta_k \rangle_{\mathscr{S}', \alpha_k} \Big) =$$
$$\sum_{\mathcal{G} \subseteq \mathcal{F} : |\mathcal{G}| \geq 2} (-1)^{|\mathcal{G}|} \big(|\mathcal{G}| - 1 \big) \left\langle \cdot, \prod_{k \in \mathcal{G}} \eta_k \right\rangle_{\mathscr{S}', \sum_{k \in \mathcal{G}} \alpha_k} \prod_{k \in \mathcal{F} \setminus \mathcal{G}} \langle \cdot, \eta_k \rangle_{\mathscr{S}', \alpha_k}, \tag{III.19}$$

where $\mathcal{F}$ is finite, and $\eta_k \in \mathscr{S}$, $\alpha_k \in \mathbb{N}$. Since the diagonals in $T^{\vec{n}_+}$ have zero measure, the random variable $\mathcal{V}_{\vec{n}}(f)$ does not depend on the behavior of the argument $f \in \otimes_{k \in \mathbb{N}} \hat{L}^2(T^{n_k})$ on the diagonals, *i.e.* where two or more of the arguments of f are equal. This is implemented via equation (III.16), where the term subtracted is the image under Δ of the first term. The operator Δ thus describes the effect from the diagonals. Moreover given $\vec{n} \in \mathcal{I}$ and $\eta_{kj} \in \mathscr{S}$ we find that

$$\mathcal{V}_{\vec{n}}\big(\otimes_{k \in \mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \eta_{kj} \big) = \Big(\prod_{k \in \mathbb{N}} n_k! \Big)^{-\frac{1}{2}} (I - \Delta) \Big(\prod_{k \in \mathbb{N}} \prod_{j=1}^{n_k} \mathcal{V}_{\vec{\varepsilon}_k}(\eta_{kj}) \Big), \tag{III.20}$$

and specializing theorem III.10 to $\vec{n} = \vec{\varepsilon}_n$, $n \in \mathbb{N}$, and given $\eta \in \mathscr{S}$, we find that

$$\mathcal{V}_{\vec{\varepsilon}_n}(\eta) = n^{-\frac{1}{2}} q^{\frac{n}{2}} \langle \eta \rangle_{L^1(T)} + n^{\frac{1}{2}} q^{-\frac{n}{2}} \sum_{\vec{n} \in \mathcal{I}_{(n)}} \Big(\prod_{k \in \mathbb{N}} k^{-n_k} (q^k - 1)^{n_k} \Big) \langle \cdot, \eta \rangle_{\mathscr{S}', \vec{n}_+}. \tag{III.21}$$

(III.17)

Equations (III.19), (III.20) and (III.21) give an alternative description of the operators $\mathcal{V}_{\vec{n}}$, $\vec{n} \in \mathcal{I}$. In order to illustrate the structure of the basis elements

$$\mathcal{V}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \eta_{kj} \big),$$

we have calculated the random variables $\mathcal{V}_{\vec{\varepsilon}_1}(\eta_1)$, $\mathcal{V}_{\vec{\varepsilon}_2}(\eta_1)$, $\mathcal{V}_{2\vec{\varepsilon}_1}(\eta_1 \hat{\otimes} \eta_2)$, $\mathcal{V}_{\vec{\varepsilon}_3}(\eta_1)$, $\mathcal{V}_{\vec{\varepsilon}_1+\vec{\varepsilon}_2}(\eta_1 \otimes \eta_2)$ and $\mathcal{V}_{3\vec{\varepsilon}_1}(\eta_1 \hat{\otimes} \eta_2 \hat{\otimes} \eta_3)$, where $\eta_1, \eta_2, \eta_3 \in L^2(T)$. These random variables are given by

$$q^{\frac{1}{2}} \langle \eta_1 \rangle + q^{-\frac{1}{2}}(q-1) \langle \cdot, \eta_1 \rangle_1 ,$$

$$2^{-\frac{1}{2}} q \langle \eta_1 \rangle + 2^{-\frac{1}{2}} q^{-1}(q^2-1) \langle \cdot, \eta_1 \rangle_1 + 2^{-\frac{1}{2}} q^{-1}(q-1)^2 \langle \cdot, \eta_1 \rangle_2 ,$$

$$\begin{gathered} 2^{-\frac{1}{2}} q \langle \eta_1 \rangle \langle \eta_2 \rangle + 2^{-\frac{1}{2}}(q-1) \big[\langle \eta_1 \rangle \langle \cdot, \eta_2 \rangle_1 + \langle \eta_2 \rangle \langle \cdot, \eta_1 \rangle_1 \big] \\ +2^{-\frac{1}{2}} q^{-1}(q-1)^2 \big[\langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_2 \rangle_1 - \langle \cdot, \eta_1\eta_2 \rangle_2 \big], \end{gathered}$$

$$\begin{gathered} 3^{-\frac{1}{2}} q^{\frac{3}{2}} \langle \eta_1 \rangle + 3^{-\frac{1}{2}} q^{-\frac{3}{2}}(q^3-1) \langle \cdot, \eta_1 \rangle_1 \\ +2^{-1} 3^{\frac{1}{2}} q^{-\frac{3}{2}}(q-1)(q^2-1) \langle \cdot, \eta_1 \rangle_2 + 2^{-1} 3^{-\frac{1}{2}} q^{-\frac{3}{2}}(q-1)^3 \langle \cdot, \eta_1 \rangle_3 , \end{gathered}$$

$$\begin{gathered} 2^{-\frac{1}{2}} q^{\frac{3}{2}} \langle \eta_1 \rangle \langle \eta_2 \rangle + 2^{-\frac{1}{2}} q^{\frac{1}{2}}(q-1) \langle \eta_2 \rangle \langle \cdot, \eta_1 \rangle_1 \\ +2^{-\frac{1}{2}} q^{-\frac{1}{2}}(q^2-1) \langle \eta_1 \rangle \langle \cdot, \eta_2 \rangle_1 + 2^{-\frac{1}{2}} q^{-\frac{1}{2}}(q-1)^2 \langle \eta_1 \rangle \langle \cdot, \eta_2 \rangle_2 \\ +2^{-\frac{1}{2}} q^{-\frac{3}{2}}(q-1)(q^2-1) \big[\langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_2 \rangle_1 - \langle \cdot, \eta_1\eta_2 \rangle_2 \big] \\ +2^{-\frac{1}{2}} q^{-\frac{3}{2}}(q-1)^3 \big[\langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_2 \rangle_2 - \langle \cdot, \eta_1\eta_2 \rangle_3 \big], \end{gathered}$$

$$\begin{gathered} 6^{-\frac{1}{2}} q^{\frac{3}{2}} \langle \eta_1 \rangle \langle \eta_2 \rangle \langle \eta_3 \rangle \\ +6^{-\frac{1}{2}} q^{\frac{1}{2}}(q-1) \big[\langle \eta_1 \rangle \langle \eta_2 \rangle \langle \cdot, \eta_3 \rangle_1 + \langle \eta_1 \rangle \langle \eta_3 \rangle \langle \cdot, \eta_2 \rangle_1 + \langle \eta_2 \rangle \langle \eta_3 \rangle \langle \cdot, \eta_1 \rangle_1 \big] \\ +6^{-\frac{1}{2}} q^{-\frac{1}{2}}(q-1)^2 \Big[\langle \eta_1 \rangle \big(\langle \cdot, \eta_2 \rangle_1 \langle \cdot, \eta_3 \rangle_1 - \langle \cdot, \eta_2\eta_3 \rangle_2 \big) \\ + \langle \eta_2 \rangle \big(\langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_3 \rangle_1 - \langle \cdot, \eta_1\eta_3 \rangle_2 \big) + \langle \eta_3 \rangle \big(\langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_2 \rangle_1 - \langle \cdot, \eta_1\eta_2 \rangle_2 \big) \Big] \\ +6^{-\frac{1}{2}} q^{-\frac{3}{2}}(q-1)^3 \big[\langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_2 \rangle_1 \langle \cdot, \eta_3 \rangle_1 - \langle \cdot, \eta_1 \rangle_1 \langle \cdot, \eta_2\eta_3 \rangle_2 \\ - \langle \cdot, \eta_2 \rangle_1 \langle \cdot, \eta_1\eta_3 \rangle_2 - \langle \cdot, \eta_3 \rangle_1 \langle \cdot, \eta_1\eta_2 \rangle_2 + 2 \langle \cdot, \eta_1\eta_2\eta_3 \rangle_3 \big]. \end{gathered}$$

III.4.1 Multiple stochastic integrals

In this subsection the multiple integrals w.r.t. the negative binomial process are introduced, and their role is discussed. Suppose $\eta \in \hat{L}^2(T^n)$ is elementary, *i.e.*

$$\eta = \sum_{\vec{\nu}\in\mathbb{N}^n : \nu_1 \le \ldots \le \nu_n} a_{\vec{\nu}} \hat{\otimes}_{j=1}^{n} 1_{A_{\nu_j}} , \tag{III.22}$$

where $A_n \in \mathscr{B}_T$, $n \in \mathbb{N}$, is a partition of T, and the coefficient $a_{\vec{\nu}} \in \mathbb{C}$ vanishes if two coordinates of $\vec{\nu}$ coincide. Then the multiple integral $I_n(\eta)$ and the compensated multiple integral $I_n^\diamond(\eta)$ are defined by

$$
\begin{aligned}
I_n(\eta) &= \sum_{\vec{\nu}\in\mathbb{N}^n:\nu_1\leq\ldots\leq\nu_n} a_{\vec{\nu}} \prod_{j=1}^{n} \Big\langle \cdot, 1_{A_{\nu_j}} \Big\rangle_{\mathscr{S}'}, \\
I_n^\diamond(\eta) &= \sum_{\vec{\nu}\in\mathbb{N}^n:\nu_1\ldots\leq\nu_n} a_{\vec{\nu}} \prod_{j=1}^{n} \Big(\Big\langle \cdot, 1_{A_{\nu_j}} \Big\rangle_{\mathscr{S}'} - \mathrm{E}_\mu \Big\langle \cdot, 1_{A_{\nu_j}} \Big\rangle_{\mathscr{S}'} \Big).
\end{aligned}
\tag{III.23}
$$

For non-elementary functions η the multiple integrals are defined by L^2-limits.

Theorem III.12 *The closure in $L^2(\mathscr{S}')$ of the linear span of the multiple integrals and of the compensated multiple integrals both equal*

$$
\bigoplus_{n\in\mathbb{N}_0} \mathcal{V}_{n\vec{\varepsilon}_1}\big(\hat{L}^2(T^n)\big).
$$

Moreover given $\eta \in \hat{L}^2(T^n)$,

$$
\mathcal{S}\big(I_n^\diamond(\eta)\big) = \Big(\frac{-q}{1-q}\Big)^n \big\langle \varsigma(\cdot)^{\otimes n}\eta \big\rangle_{\hat{L}^1(T^n)}. \tag{III.24}
$$

Proof It is seen from (III.23) that the compensated multiple integrals of elementary functions can be expressed as sums of multiple integrals of elementary functions, and visa versa. The closure in $L^2(\mathscr{S}')$ of the linear span of these quantities thus coincide. Moreover if η is given by (III.22), then by lemma III.2

$$
\mathcal{S}\big(I_n^\diamond(\eta)\big) = \sum_{\vec{\nu}\in\mathbb{N}^n:\nu_1\leq\ldots\leq\nu_n} a_{\vec{\nu}} \Big(\frac{-q}{1-q}\Big)^n \prod_{j=1}^{n} \Big\langle \varsigma(\cdot) 1_{A_{\nu_j}} \Big\rangle_{\hat{L}^1(T)}.
$$

Equation (III.24) thus follows by taking L^2-limits. The proof is completed by combining equations (III.7) and (III.24). □

Remark The decomposition given in theorem III.10 is considerably more complicated than the orthogonal decomposition for the corresponding Hilbert spaces in the Gaussian and in the Poisson case, where

$$
L^2(\mathscr{S}') \simeq \bigoplus_{n\in\mathbb{N}} \hat{L}^2(T^n).
$$

Especially the multiple stochastic integrals do not span $L^2(\mathscr{S}')$ in the negative binomial case as they do in the Gaussian and in the Poisson case. There is an intuitive explanation for this fact. As a compound Poisson process the negative binomial process on the parameter set T is in a natural correspondence with a Poisson process on $T \times \mathbb{N}$, whence there exists a natural isomorphism between the Hilbert spaces supported by the two processes. But in contrast to the negative binomial process, the integrand in the multiple integrals w.r.t. the Poisson process on $T \times \mathbb{N}$ may depend non-linearly on the component $\mathbb{N}$, *i.e.* the jump heights.

III.4.2 Polynomial functionals

In this subsection we give the chaos expansion of some polynomial functionals on the negative binomial probability space. In general these chaos expansions can be found via lemma III.2. A typical example is the following theorem.

Theorem III.14 *Given $\eta \in \mathscr{S}$ we have that*

$$\langle \cdot, \eta \rangle_{\mathscr{S}',0} = -\log(1-q) \cdot \langle \eta \rangle_{L^1(T)} - \sum_{k=1}^{\infty} k^{-\frac{1}{2}} q^{\frac{k}{2}} \mathcal{V}_{\vec{\varepsilon}_k}(\eta),$$

$$\langle \cdot, \eta \rangle_{\mathscr{S}',1} = \frac{q}{1-q} \langle \eta \rangle_{L^1(T)} - \frac{q^{\frac{1}{2}}}{1-q} \mathcal{V}_{\vec{\varepsilon}_1}(\eta),$$

and given $n \geq 2$, $\eta \in \mathscr{S}$, the random variable $\langle \cdot, \eta \rangle_{\mathscr{S}',n}$ is equal to

$$\sum_{j=1}^{n-1} \gamma_{n-1,j} \frac{q^j}{(1-q)^{j+1}} \langle \eta \rangle_{L^1(T)} + \sum_{k=1}^{n} (-1)^k k^{\frac{1}{2}} \Bigg(1_{k \geq 2} \cdot \gamma_{n-1,k-1} \frac{q^{\frac{k}{2}}}{(1-q)^k} + \sum_{j=k}^{n-1} \left[\binom{j}{k-1} q + \binom{j}{k} \right] \gamma_{n-1,j} \frac{q^{j-\frac{k}{2}}}{(1-q)^{j+1}} \Bigg) \mathcal{V}_{\vec{\varepsilon}_k}(\eta),$$

where γ_{nj} are the coefficients defined in (III.3).

Proof Lemma III.2 and equation (III.5) gives that $(\mathcal{S} \langle \cdot, \eta \rangle_{\mathscr{S}',n})(\xi)$, where $n \geq 2$, $\eta \in \mathscr{S}$ and $\xi \in \mathscr{S}_{\sqrt{q}}$, is equal to

$$\begin{aligned} \left\langle \gamma(qe^{\imath\xi}; n-1)\eta \right\rangle_{L^1(T)} &= \sum_{j=1}^{n-1} \gamma_{n-1,j} \left\langle \frac{(qe^{\imath\xi})^j}{(1-qe^{\imath\xi})^{j+1}} \eta \right\rangle_{L^1(T)} \\ &= \sum_{j=1}^{n-1} \gamma_{n-1,j} \frac{q^j}{(1-q)^{j+1}} \left\langle (1-\varsigma(\xi))^j (1-q\varsigma(\xi)) \eta \right\rangle_{L^1(T)} \\ &= \sum_{j=1}^{n-1} \sum_{k=0}^{j} \gamma_{n-1,j} \frac{q^j}{(1-q)^{j+1}} \binom{j}{k} (-1)^k \left\langle \left(\varsigma(\xi)^k - q\varsigma(\xi)^{k+1} \right) \eta \right\rangle_{L^1(T)}, \end{aligned}$$

which is equal to

$$\sum_{j=1}^{n-1} \gamma_{n-1,j} \frac{q^j}{(1-q)^{j+1}} \langle \eta \rangle_{L^1(T)} + \sum_{k=1}^{n} (-1)^k \Bigg(1_{k \geq 2} \cdot \gamma_{n-1,k-1} \frac{q^k}{(1-q)^k} + \sum_{j=k}^{n-1} \left[\binom{j}{k-1} q + \binom{j}{k} \right] \gamma_{n-1,j} \frac{q^j}{(1-q)^{j+1}} \Bigg) \left\langle \varsigma(\xi)^k \eta \right\rangle_{L^1(T)}. \tag{III.20}$$

The third statement now follows by applying $\mathcal{S}^{-1}$ on both sides, and using lemma III.6. The first two statements follow similarly from the calculations

$$\begin{aligned}\big(\mathcal{S}\,\langle\cdot,\eta\rangle_{\mathscr{S}',0}\big)(\xi) &= \big\langle -\eta\log\big(1-qe^{\imath\xi}\big)\big\rangle_{L^1(T)} = \Big\langle \eta\log\Big(\frac{1-q\varsigma(\xi)}{1-q}\Big)\Big\rangle_{L^1(T)} \\ &= -\log(1-q)\,\langle\eta\rangle_{L^1(T)} - \sum_{k=1}^{\infty} k^{-1}q^k\,\langle\eta\varsigma(\xi)^k\rangle_{L^1(T)} \\ &= -\log(1-q)\cdot\langle\eta\rangle_{L^1(T)} - \sum_{k=1}^{\infty} k^{-\frac{1}{2}}q^{\frac{k}{2}}\mathcal{U}_{\vec{\varepsilon}_k}(\eta)(\xi),\end{aligned}$$

and

$$\big(\mathcal{S}\,\langle\cdot,\eta\rangle_{\mathscr{S}',1}\big)(\xi) = \Big\langle \frac{qe^{\imath\xi}}{1-qe^{\imath\xi}}\eta\Big\rangle_{L^1(T)} = \frac{q}{1-q}\Big(\langle\eta\rangle_{L^1(T)} - \langle\varsigma(\xi)\eta\rangle_{L^1(T)}\Big).$$

□

Equation (III.21) and theorem III.14 shows, that for every $n\in\mathbb{N}$ there are linear transformations between the random variables $\mathcal{V}_{\vec{\varepsilon}_k}(\eta)$, $k=1,\dots,n$, and $\langle\cdot,\eta\rangle_{\mathscr{S}',k}$, $k=1,\dots,n$. Alternatively the operators $\mathcal{V}_{\vec{\varepsilon}_n}$, $n\in\mathbb{N}$, can thus be found by inverting the formulae given in theorem III.14. We similarly see that the operators $\mathcal{V}_{\vec{n}}$, $\vec{n}\in\mathcal{I}$, can be found by inverting the chaos expansion of the polynomial functionals (III.18). But since we already have formulae for the operators $\mathcal{V}_{\vec{n}}$, we will not continue this line of investigation further.

We complete this subsection by considering the orthogonal polynomials w.r.t. the negative binomial distribution, *i.e.* the Meixner polynomials $m_n(x;\beta,c)$, $n\in\mathbb{N}_0$, of the first kind with parameters $0<c<1$, $\beta>0$. According to Chihara (1978, p. 176) these polynomials are defined by

$$m_n(x) = m_n(x;\beta,c) = (-1)^n n!\sum_{k=0}^{n}\binom{x}{k}\binom{-x-\beta}{n-k}c^{-k}.$$

Theorem III.15 *Given $n\in\mathbb{N}_0$ and $A\in\mathscr{B}_T$ we have that*

$$m_n\big(\langle\cdot,1_A\rangle_{\mathscr{S}'};\langle 1_A\rangle_{L^1(T)},q\big) = n!q^{-\frac{n}{2}}\sum_{\vec{n}\in\mathcal{I}_{(n)}}\Big(\prod_{k\in\mathbb{N}} k^{-\frac{n_k}{2}}(n_k!)^{-\frac{1}{2}}\Big)\,\mathcal{V}_{\vec{n}}\big(1_A^{\otimes\vec{n}_+}\big).$$

Proof Inserting $\eta=1_A$ on both sides of (III.9) gives that

$$\sum_{n\in\mathbb{N}_0} w^n\sum_{\vec{n}\in\mathcal{I}_{(n)}}\Big(\prod_{k\in\mathbb{N}} q^{\frac{kn_k}{2}}k^{-\frac{n_k}{2}}(n_k!)^{-\frac{1}{2}}\Big)\,\mathcal{V}_{\vec{n}}\big(1_A^{\otimes\vec{n}_+}\big)$$

is equal to

$$\begin{aligned}C\big(\varsigma^{-1}(w1_A)\big)e^{\imath\langle\cdot,\varsigma^{-1}(w1_A)\rangle_{\mathscr{S}'}} &= (1-qw)^{\langle 1_A\rangle_{L^1(T)}}\Big(\frac{1-w}{1-qw}\Big)^{\langle\cdot,1_A\rangle_{\mathscr{S}'}} \\ &= (1-w)^{\langle\cdot,1_A\rangle_{\mathscr{S}'}}(1-qw)^{-\langle\cdot,1_A\rangle_{\mathscr{S}'}-\langle 1_A\rangle_{L^1(T)}} \\ &= \sum_{n\in\mathbb{N}_0} m_n\big(\langle\cdot,1_A\rangle_{\mathscr{S}'};\langle 1_A\rangle_{L^1(T)},q\big)(n!)^{-1}q^n w^n.\end{aligned}$$

(III.21)

The last identity follows by Chihara (1978, p. 176, equation (3.3)). The corollary now follows by comparing the coefficients of w^n, $n \in \mathbb{N}_0$, and rearranging the factors. □

III.5 Malliavin calculus

In this section we investigate possible choices of a Malliavin calculus on the negative binomial probability space. Our analysis is based on some abstract properties of a Malliavin derivative operator D. Usually D is chosen as some kind of variational derivative with a physical interpretation, and in the subsections we describe the operators associated to derivation w.r.t. the jump times respectively the jump heights. A Malliavin derivative operator is a closed, densely defined, unbounded, linear operator

$$\mathrm{D} : L^2(\mathscr{S}') \to L^2(\mathscr{S}' \times T) \simeq L^2(\mathscr{S}') \otimes L^2(T)$$

that satisfies the product rule, *i.e.* for $\phi_1, \phi_2, \phi_1\phi_2 \in \mathrm{Dom}(\mathrm{D})$ the following identity holds

$$\mathrm{D}(\phi_1\phi_2)(\omega, t) = (\mathrm{D}\,\phi_1)(\omega, t) \cdot \phi_2(\omega) + \phi_1(\omega) \cdot (\mathrm{D}\,\phi_2)(\omega, t).$$

The tensor product $L^2(\mathscr{S}') \otimes L^2(T)$ can be identified with the space of linear operators from $L^2(T)$ to $L^2(\mathscr{S}')$. Thus given a Malliavin operator D and a function $\zeta \in L^2(T)$ let $\mathrm{D}_\zeta : L^2(\mathscr{S}') \to L^2(\mathscr{S}')$ be the associated operator.

Let $L^2_{\mathrm{ray}}(\mathscr{S}')$ be the closed linear subspace of $L^2(\mathscr{S}')$ spanned by the random variables $\mathcal{V}_{\vec{\varepsilon}_n}(\eta)$, $n \in \mathbb{N}$, $\eta \in L^2(T)$, and let $L^2_{\mathrm{fin}}(\mathscr{S}')$ be the linear subspace of $L^2(\mathscr{S}')$ consisting of the random variables with finite chaos expansion and smooth kernels, *i.e.*

$$L^2_{\mathrm{ray}}(\mathscr{S}') = \Big\{\phi = z + \sum_{n\in\mathbb{N}} \mathcal{V}_{\vec{\varepsilon}_n}(\eta_n) \in L^2(\mathscr{S}') : z \in \mathbb{C},\ \eta_n \in L^2(T)\Big\},$$

$$L^2_{\mathrm{fin}}(\mathscr{S}') = \Big\{\phi = \sum_{k=1}^{K} \mathcal{V}_{\vec{n}_k}(f_k) \in L^2(\mathscr{S}') : K \in \mathbb{N}, \vec{n}_k \in \mathcal{I}, f_k \in \otimes_{j\in\mathbb{N}} \mathscr{S}_{\mathbb{C}}^{\hat{\otimes} n_{kj}}\Big\}.$$

Observe that the elements in $L^2_{\mathrm{fin}}(\mathscr{S}')$ can be written as a finite sum of finite products of elements from $L^2_{\mathrm{ray}}(\mathscr{S}')$, and that $L^2_{\mathrm{fin}}(\mathscr{S}')$ is dense in $L^2(\mathscr{S}')$ and closed under multiplication.

Lemma III.16 *If the operator $\tilde{\mathrm{D}}_\zeta : L^2_{fin}(\mathscr{S}') \to L^2(\mathscr{S}')$ satisfies the product rule, commutes with the operator Δ defined in* (III.19)*, and acts invariantly on $L^2_{\mathrm{ray}}(\mathscr{S}')$, then*

$$\mathrm{E}_\mu\, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{n}}(f) = 0, \quad f \in \otimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_k}),$$

for every $\vec{n} \in \mathcal{I}$ with $\vec{n}_+ \geq 2$.

(III.22)

Proof Since $\tilde{\mathrm{D}}_\zeta$ acts invariantly on $L^2_{\text{ray}}(\mathscr{S}')$, there exist linear functionals $A_n : L^2(T) \to \mathbb{C}$ and linear operators $A_{nl} : L^2(T) \to L^2(T)$ such that

$$\tilde{\mathrm{D}}_\zeta \circ \mathcal{V}_{\vec{\varepsilon}_n} = A_n + \sum_{l\in\mathbb{N}} \mathcal{V}_{\vec{\varepsilon}_l} \circ A_{nl}, \quad n \in \mathbb{N}.$$

Let $\vec{n} \in \mathcal{I}$, $\vec{n}_+ \geq 2$, and $\eta_{kj} \in \mathscr{S}$, $k \in \mathbb{N}$, $j = 1, \dots, n_k$, be given. Using equation (III.20), $\tilde{\mathrm{D}}_\zeta \circ \Delta = \Delta \circ \tilde{\mathrm{D}}_\zeta$, and the product rule, we thus find that

$$\begin{aligned}
\mathrm{E}_\mu \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{j=1}^{n_k} \eta_{kj}\big) &= \Big(\prod_{k\in\mathbb{N}} n_k!\Big)^{-\frac{1}{2}} \mathrm{E}_\mu\, \tilde{\mathrm{D}}_\zeta \circ (I-\Delta)\Big(\prod_{k\in\mathbb{N}}\prod_{j=1}^{n_k} \mathcal{V}_{\vec{\varepsilon}_k}(\eta_{kj})\Big) \\
&= \Big(\prod_{k\in\mathbb{N}} n_k!\Big)^{-\frac{1}{2}} \sum_{k_0\in\mathbb{N}}\sum_{j_0=1}^{n_{k_0}} \mathrm{E}_\mu (I-\Delta)\circ \Big(\tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{\varepsilon}_{k_0}}(\eta_{k_0 j_0}) \prod_{(k,j)\neq(k_0,j_0)} \mathcal{V}_{\vec{\varepsilon}_k}(\eta_{kj})\Big) \\
&= \Big(\prod_{k\in\mathbb{N}} n_k!\Big)^{-\frac{1}{2}} \sum_{k_0\in\mathbb{N}}\sum_{j_0=1}^{n_{k_0}} \Big(A_{k_0}(\eta_{k_0 j_0})\ \mathrm{E}_\mu (I-\Delta)\Big(\prod_{(k,j)\neq(k_0,j_0)} \mathcal{V}_{\vec{\varepsilon}_k}(\eta_{kj})\Big) \\
&\qquad + \sum_{l\in\mathbb{N}} \mathrm{E}_\mu (I-\Delta)\Big(\mathcal{V}_{\vec{\varepsilon}_l}\big(A_{k_0 l}(\eta_{k_0 j_0})\big) \prod_{(k,j)\neq(k_0,j_0)} \mathcal{V}_{\vec{\varepsilon}_k}(\eta_{kj})\Big)\Big) \\
&= \Big(\prod_{k\in\mathbb{N}} n_k!\Big)^{-\frac{1}{2}} \sum_{k_0\in\mathbb{N}}\sum_{j_0=1}^{n_{k_0}} \Big(A_{k_0}(\eta_{k_0 j_0})\ \mathrm{E}_\mu\, \mathcal{V}_{\vec{n}-\vec{\varepsilon}_{k_0}}\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{(k,j)\neq(k_0,j_0)} \eta_{kj}\big) \\
&\qquad + \sum_{l\in\mathbb{N}} \mathrm{E}_\mu\, \mathcal{V}_{\vec{n}-\vec{\varepsilon}_{k_0}+\vec{\varepsilon}_l}\Big(\big(\otimes_{k\in\mathbb{N}} \hat{\otimes}_{(k,j)\neq(k_0,j_0)} \eta_{kj}\big) \hat{\otimes}^{(l)} A_{k_0 l}(\eta_{k_0 j_0})\Big)\Big) \\
&= 0,
\end{aligned}$$

where $\hat{\otimes}^{(l)}$ means that the symmetric tensor product is taken between the l'th component of the first factor and the second factor. □

The Malliavin operator D being closed is equivalent to the adjoint operator

$$\delta : L^2(\mathscr{S}') \otimes L^2(T) \to L^2(\mathscr{S}')$$

being densely defined. Among the unbounded operators are those which are closed certainly the nicest ones, and in many applications of Malliavin calculus, *e.g.* derivations of density criterions or calculations of densities, is the use of the adjoint operator crucial and if its domain were not dense, then this operator could often not be employed. Thus if we are given an operator

$$\tilde{\mathrm{D}} : L^2_{\text{fin}}(\mathscr{S}') \to L^2(\mathscr{S}') \otimes L^2(T)$$

which satisfies the product rule, then the question is whether there exists a closed extension D of $\tilde{\mathrm{D}}$. Given a vanishing sequence $\phi_N \in L^2_{\text{fin}}(\mathscr{S}')$, $N \in \mathbb{N}$, such that

(III.23)

$\tilde{\mathrm{D}}_\zeta \phi_N$ converges to ϕ_ζ in $L^2(\mathscr{S}')$ as $N \to \infty$, we thus need to show that $\phi_\zeta = 0$, $\zeta \in L^2(T)$. If for every $\zeta \in L^2(T)$ there exists $\delta(\zeta) \in L^2(\mathscr{S}')$ such that

$$\mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{n}}(f) = \mathrm{E}_\mu \big(\mathcal{V}_{\vec{n}}(f)\delta(\zeta)\big), \quad \vec{n} \in \mathcal{I},\ f \in \otimes_{k\in\mathbb{N}} \mathscr{S}_{\mathbb{C}}^{\hat{\otimes} n_k}, \tag{III.25}$$

then for every $\phi_0 \in L^2_{\mathrm{fin}}(\mathscr{S}')$,

$$\begin{aligned}\big\langle \phi_\zeta, \overline{\phi_0}\big\rangle_{L^2(\mathscr{S}')} &= \lim_{N\to\infty} \mathrm{E}_\mu \big((\tilde{\mathrm{D}}_\zeta \phi_N)\phi_0\big) = \lim_{N\to\infty} \mathrm{E}_\mu \big(\tilde{\mathrm{D}}_\zeta(\phi_N\phi_0) - \phi_N(\tilde{\mathrm{D}}_\zeta \phi_0)\big) \\ &= \lim_{N\to\infty} \mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta(\phi_N\phi_0) = \lim_{N\to\infty} \mathrm{E}_\mu \big(\phi_N\phi_0\delta(\zeta)\big) \\ &= 0.\end{aligned}$$

Since $L^2_{\mathrm{fin}}(\mathscr{S}')$ is dense in $L^2(\mathscr{S}')$, it follows that $\tilde{\mathrm{D}}$ is closable. The standard method to construct a Malliavin operator D is thus to find a operator $\tilde{\mathrm{D}}$ and random variables $\delta(\zeta) \in L^2(\mathscr{S}')$, $\zeta \in L^2(T)$, such that (III.25) holds. If $\tilde{\mathrm{D}}_\zeta$ commutes with Δ and acts invariantly on $L^2_{\mathrm{ray}}(\mathscr{S}')$, then equation (III.25) and lemma III.16 give

$$\begin{aligned}\mathrm{E}_\mu \, \delta(\zeta) &= \mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta 1 = 0, & & \\ \mathrm{E}_\mu \big(\delta(\zeta)\mathcal{V}_{\vec{n}}(f)\big) &= \mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{n}}(f) = 0, & & \vec{n} \in \mathcal{I},\ \vec{n}_+ \geq 2, \\ \mathrm{E}_\mu \big(\delta(\zeta)\mathcal{V}_{\vec{\varepsilon}_n}(\eta)\big) &= \mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{\varepsilon}_n}(\eta), & & n \in \mathbb{N}.\end{aligned}$$

Thus if $\eta_j \in \mathscr{S}$, $j \in \mathbb{N}$, is an orthonormal basis for $L^2(T)$, then the chaos decomposition of $\delta(\zeta)$ is given by

$$\delta(\zeta) = \sum_{n\in\mathbb{N}} \sum_{j\in\mathbb{N}} \big(\, \mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{\varepsilon}_n}(\eta_j)\big)\, \mathcal{V}_{\vec{\varepsilon}_n}(\eta_j). \tag{III.26}$$

Concerning the choice of the operator $\tilde{\mathrm{D}}$ observe that by the product rule $\tilde{\mathrm{D}}$ is determined by its action on $L^2_{\mathrm{ray}}(\mathscr{S}')$, and hence by the images $\tilde{\mathrm{D}}_\zeta \langle \cdot, \eta\rangle_{\mathscr{S}',n}$, $n \in \mathbb{N}$, $\zeta, \eta \in \mathscr{S}$.

III.5.1 Variational derivation w.r.t. jump times

In the case of derivation w.r.t. the jump times $\tilde{\mathrm{D}}$ is given by

$$\begin{aligned}\tilde{\mathrm{D}}_\zeta \langle \cdot, \eta\rangle_{\mathscr{S}',n} &= \tilde{\mathrm{D}}_\zeta \sum_j Y_j^n \, \eta(\tau_j) = \sum_j Y_j^n \, \zeta(\tau_j) \, \tfrac{\partial}{\partial\epsilon} \eta(\tau_j + \epsilon) \\ &= \sum_j Y_j^n \, \zeta(\tau_j) \, \eta'(\tau_j) = \langle \cdot, \zeta\eta'\rangle_{\mathscr{S}',n}, \quad \zeta, \eta \in \mathscr{S},\ n \in \mathbb{N}.\end{aligned}$$

It is clear that $\tilde{\mathrm{D}}_\zeta$ acts invariantly on $L^2_{\mathrm{ray}}(\mathscr{S}')$. Moreover by the Leibniz rule of differentiation, *i.e.*

$$\Big(\prod_{k\in\mathcal{G}} \eta_k\Big)' = \sum_{k\in\mathcal{G}} \eta_k' \prod_{j\in\mathcal{G}\setminus\{k\}} \eta_j,$$

it follows that $\tilde{\mathrm{D}}_\zeta$ commutes with Δ. Using equation (III.21) we find that

$$\tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{\varepsilon}_n}(\eta) = \mathcal{V}_{\vec{\varepsilon}_n}(\zeta\eta') - n^{-\frac{1}{2}} q^{\frac{n}{2}} \langle \zeta\eta' \rangle_{L^1(T)} = \mathcal{V}_{\vec{\varepsilon}_n}(\zeta\eta') + n^{-\frac{1}{2}} q^{\frac{n}{2}} \langle \zeta'\eta \rangle_{L^1(T)} ,$$

whence $\mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{\varepsilon}_n}(\eta) = n^{-\frac{1}{2}} q^{\frac{n}{2}} \langle \zeta', \overline{\eta} \rangle_{L^2(T)}$. Inserting an orthonormal basis $\eta_j \in \mathscr{S}$, $j \in \mathbb{N}$, for $L^2(T)$ we thus see that $\delta(\zeta)$, $\zeta \in L^2(T)$, is given by

$$\begin{aligned} \delta(\zeta) &= \sum_{n\in\mathbb{N}} n^{-\frac{1}{2}} q^{\frac{n}{2}} \mathcal{V}_{\vec{\varepsilon}_n}(\zeta') \\ &= -\log(1-q) \cdot \langle \zeta' \rangle_{L^1(T)} + \sum_{\vec{n}\in\mathcal{I}\setminus\{0\}} \Big(\prod_{k\in\mathbb{N}} k^{-n_k} (q^k - 1)^{n_k} \Big) \langle \cdot, \zeta' \rangle_{\mathscr{S}', \vec{n}_+} , \end{aligned}$$

which converges in $L^2(\mathscr{S}')$ with squared norm given by

$$\|\delta(\zeta)\|^2_{L^2(\mathscr{S}')} = \sum_{n\in\mathbb{N}} n^{-1} q^n \, \|\zeta'\|^2_{L^2(T)} = -\log(1-q) \cdot \|\zeta'\|^2_{L^2(T)} .$$

III.5.2 Variational derivation w.r.t. jump heights

In the case of derivation w.r.t. the jump heights $\tilde{\mathrm{D}}$ is given by

$$\begin{aligned} \tilde{\mathrm{D}}_\zeta \langle \cdot, \eta \rangle_{\mathscr{S}', n} &= \tilde{\mathrm{D}}_\zeta \sum_j Y_j^n \, \eta(\tau_j) = \sum_j \tfrac{\partial}{\partial\epsilon} (Y_j + \epsilon)^n \, \zeta(\tau_j) \, \eta(\tau_j) \\ &= n \sum_j Y_j^{n-1} \, \zeta(\tau_j) \, \eta(\tau_j) = n \, \langle \cdot, \zeta\eta \rangle_{\mathscr{S}', n-1} , \quad \zeta, \eta \in \mathscr{S}, \ n \in \mathbb{N}. \end{aligned}$$

It is clear that $\tilde{\mathrm{D}}_\zeta$ acts invariantly on $L^2_{\mathrm{ray}}(\mathscr{S}')$, and $\tilde{\mathrm{D}}_\zeta$ is easily shown to commute with Δ. Using equation (III.21) and lemma III.2 we find that

$$\mathrm{E}_\mu \, \tilde{\mathrm{D}}_\zeta \mathcal{V}_{\vec{\varepsilon}_n}(\eta) = n^{\frac{1}{2}} q^{-\frac{n}{2}} \Big(\sum_{\vec{n}\in\mathcal{I}_{(n)}} \vec{n}_+ \gamma(q; \vec{n}_+ - 2) \prod_{k\in\mathbb{N}} k^{-n_k} (q^k - 1)^{n_k} \Big) \langle \zeta\eta \rangle_{L^1(T)} . \tag{III.27}$$

Observe that the factor $q^{\frac{n}{2}}$ in equation (III.27) increases exponentially in n. For the sum

$$\delta(\zeta) = \sum_{n\in\mathbb{N}} n^{\frac{1}{2}} q^{-\frac{n}{2}} \Big(\sum_{\vec{n}\in\mathcal{I}_{(n)}} \vec{n}_+ \gamma(q; \vec{n}_+ - 2) \prod_{k\in\mathbb{N}} k^{-n_k} (q^k - 1)^{n_k} \Big) \mathcal{V}_{\vec{\varepsilon}_n}(\zeta) \tag{III.28}$$

to converge in $L^2(\mathscr{S}')$, the factor consisting of the sum over $\vec{n} \in \mathcal{I}_{(n)}$ must thus vanish at least exponentially fast in n. There is no immediate structure implying this to be the case, and explicit calculations of the first terms in (III.28) indicates that this sum is in fact divergent in $L^2(\mathscr{S}')$. As a consequence the Malliavin calculus given by derivation w.r.t. the jumps heights is not well-behaved.

(III.25)

Nevertheless there exists a rather artificial, but certainly implementable, way to save the Malliavin calculus based on derivation w.r.t. the jump heights. Using lemma III.16 we see that if we define the subspaces $L^2_n(\mathscr{S}')$ of $L^2(\mathscr{S}')$ by

$$L^2_n(\mathscr{S}') = \bigoplus_{\vec{n}\in\mathcal{I}:\vec{n}_* \leq n \text{ or } \vec{n}_+ \geq 2} \mathcal{V}_{\vec{n}}\big(\otimes_{k\in\mathbb{N}} \hat{L}^2(T^{n_k})\big), \quad n \in \mathbb{N},$$

then there exists a closed extension $\mathrm{D}_n : L^2_n(\mathscr{S}') \to L^2(\mathscr{S}') \otimes L^2(T)$ of the restriction of $\tilde{\mathrm{D}}$ to $L^2_{\text{fin}}(\mathscr{S}') \cap L^2_n(\mathscr{S}')$. In this case the random variable $\delta(\zeta)$, $\zeta \in L^2(T)$, is given by the finite sum

$$\delta_n(\zeta) = \sum_{j=1}^n j^{\frac{1}{2}} q^{-\frac{j}{2}} \Big(\sum_{\vec{n}\in\mathcal{I}_{(j)}} \vec{n}_+ \gamma(q; \vec{n}_+ - 2) \prod_{k\in\mathbb{N}} k^{-n_k}(q^k-1)^{n_k} \Big)\, \mathcal{V}_{\vec{\varepsilon}_j}(\zeta).$$

Although this construction has the flavour of cheating, the spaces $L^2_n(\mathscr{S}')$ are still infinite dimensional and may very well contain functionals of interest, *e.g.* all the multiple integrals as is seen from theorem III.12.

Appendix

Proof [Proof of lemma III.4] Let $\mathfrak{S}_n$ be the reproducing kernel Hilbert space over the set $\mathcal{F}$ generated by the kernel

$$\begin{aligned} K_n(\zeta,\eta) = \langle \zeta,\eta\rangle^n_{L^2(T)} &= \sum_{\vec{\nu}\in\mathbb{N}^n} \prod_{j=1}^n \big\langle \zeta, \xi_{\nu_j}\big\rangle_{L^2(T)} \overline{\big\langle \eta, \xi_{\nu_j}\big\rangle_{L^2(T)}} \\ &= \sum_{\vec{\nu}\in\mathbb{N}^n:\nu_1\leq\ldots\leq\nu_n} n!\Big(\prod_{i\in\mathbb{N}} |\{j:\nu_j=i\}|!\Big)^{-1} \prod_{j=1}^n \big\langle \zeta, \xi_{\nu_j}\big\rangle_{L^2(T)} \overline{\big\langle \eta, \xi_{\nu_j}\big\rangle_{L^2(T)}}. \end{aligned}$$

Using polarization we find that

$$\prod_{j=1}^n \big\langle \cdot, \xi_{\nu_j}\big\rangle_{L^2(T)} = (n!)^{-1} 2^{-n} \sum_{\vec{\sigma}\in\{-1,+1\}^n} \Big(\prod_{j=1}^n \sigma_j\Big) K_n\Big(\cdot, \sum_{j=1}^n \sigma_j \xi_{\nu_j}\Big).$$

Thus $K_n(\zeta,\eta) = \sum_{\vec{\nu}\in\mathbb{N}^n:\nu_1\leq\ldots\leq\nu_n} \phi_{\vec{\nu}}(\zeta)\overline{\phi_{\vec{\nu}}(\eta)}$, where $\phi_{\vec{\nu}} \in \mathfrak{S}_n$, $\vec{\nu} \in \mathbb{N}^n$, are defined by

$$\phi_{\vec{\nu}} = (n!)^{-\frac{1}{2}} \Big(\prod_{i\in\mathbb{N}} |\{j:\nu_j=i\}|!\Big)^{-1} 2^{-n} \sum_{\vec{\sigma}\in\{-1,+1\}^n} \Big(\prod_{j=1}^n \sigma_j\Big) K_n\Big(\cdot, \sum_{j=1}^n \sigma_j \xi_{\nu_j}\Big).$$

Let the linear mapping $\mathcal{U}_n : \hat{L}^2(T^n) \to \mathfrak{S}_n$ be given by $\mathcal{U}_n(\hat{\otimes}^n_{j=1}\overline{h}_{\nu_j}) = \phi_{\vec{\nu}}$. Then $\mathcal{U}_n$ is isometric and injective since the inner product $\langle \phi_{\vec{\nu}_1}, \phi_{\vec{\nu}_2}\rangle_{\mathfrak{S}_n}$ is equal to

$$\begin{aligned} &(n!)^{-1}\Big(\prod_{i\in\mathbb{N}} |\{j:\nu_{1j}=i\}|!\Big)^{-1}\Big(\prod_{i\in\mathbb{N}} |\{j:\nu_{2j}=i\}|!\Big)^{-1} \\ &\qquad 2^{-2n} \sum_{\vec{\sigma}_1,\vec{\sigma}_2} \Big(\prod_{j=1}^n \sigma_{1j}\sigma_{2j}\Big) K_n\Big(\sum_{j=1}^n \sigma_{2j}\xi_{\nu_{2j}}, \sum_{j=1}^n \sigma_{1j}\xi_{\nu_{1j}}\Big), \end{aligned} \tag{III.26}$$

which equals

$$\left\langle \hat{\otimes}_{j=1}^{n} \xi_{\nu_{2j}}, \hat{\otimes}_{j=1}^{n} \xi_{\nu_{1j}} \right\rangle_{\hat{L}^2(T^n)} = \left\langle \hat{\otimes}_{j=1}^{n} \overline{h}_{\nu_{1j}}, \hat{\otimes}_{j=1}^{n} \overline{h}_{\nu_{2j}} \right\rangle_{\hat{L}^2(T^n)}.$$

Moreover $\mathcal{U}_n$ is surjective since $\mathcal{U}_n(\overline{\eta}^{\otimes n}) = K_n(\cdot, \eta)$, $\eta \in \mathcal{F}$. □

Proof [Proof of lemma III.5] Since $f_1 \neq 0$ there exists pairwise disjoint, bounded, closed subsets $A_1, \dots, A_{\vec{n}_{1+}}$ of T such that

$$\left\langle \left(\otimes_{j=1}^{\vec{n}_{1+}} 1_{A_j} \right) f_1 \right\rangle_{L^1(T^{\vec{n}_{1+}})} \neq 0.$$

Given fixed $x_1, \dots, x_{\vec{n}_{1+}} \in \mathbb{R}$, let $\eta = \sum_{j=1}^{\vec{n}_{1+}} x_j 1_{A_j}$. Since the sets A_j are disjoint,

$$\eta^k = \sum_{j=1}^{\vec{n}_{1+}} x_j^k 1_{A_j}, \qquad \otimes_{k\in\mathbb{N}} (\eta^k)^{\otimes n_{ik}} = \sum_{\vec{\nu} \in \{1,\dots,\vec{n}_{1+}\}^{\vec{n}_{i+}}} \Big(\prod_{k\in\mathbb{N}} \prod_{j=1}^{n_{ik}} x^k_{\nu_{\phi_{\vec{n}_i}(k,j)}} \Big) \otimes_{j=1}^{\vec{n}_{i+}} 1_{A_{\nu_j}},$$

where $\phi_{\vec{n}_i}(k,j) = \sum_{l=1}^{k-1} n_{il} + j$, and

$$\begin{aligned} &\sum_{i=1,2} \left\langle \left(\otimes_{k\in\mathbb{N}} (\eta^k)^{\otimes n_{ik}} \right) f_i \right\rangle_{L^1(T^{\vec{n}_{i+}})} \\ &\quad = \sum_{i=1,2} \sum_{\vec{\nu} \in \{1,\dots,\vec{n}_{1+}\}^{\vec{n}_{i+}}} \Big(\prod_{k\in\mathbb{N}} \prod_{j=1}^{n_{ik}} x^k_{\nu_{\phi_{\vec{n}_i}(k,j)}} \Big) \left\langle \left(\otimes_{j=1}^{\vec{n}_{i+}} 1_{A_{\nu_j}} \right) f_i \right\rangle_{L^1(T^{\vec{n}_{i+}})}. \end{aligned} \tag{III.29}$$

The right hand side of (III.29) is a polynomial in the variables $x_1, \dots, x_{\vec{n}_{1+}}$, where the coefficient of the monomial $\prod_{k\in\mathbb{N}} \prod_{j=1}^{n_{1k}} x^k_{\phi_{\vec{n}_1}(k,j)}$ due to the symmetry properties of f_1 is given by

$$\begin{aligned} &\sum_{\substack{\pi_k \text{ permutation} \\ \text{of } \{1,\dots,n_{1k}\}}} \left\langle \left(\otimes_{k\in\mathbb{N}} \otimes_{j=1}^{n_{1k}} 1_{A_{\phi_{\vec{n}_1}(k,\pi_k(j))}} \right) f_1 \right\rangle_{L^1(T^{\vec{n}_{1+}})} \\ &\qquad\qquad = \Big(\prod_{k\in\mathbb{N}} n_{1k}! \Big) \left\langle \left(\otimes_{j=1}^{\vec{n}_{1+}} 1_{A_j} \right) f_1 \right\rangle_{L^1(T^{\vec{n}_{1+}})} \neq 0. \end{aligned}$$

Thus the polynomial is non-zero, and we can assume $x_1, \dots, x_{\vec{n}_{1+}} \in \mathbb{R}$ have been chosen such that the left hand side of (III.29) is non-zero. Now given a smooth function $\zeta \in \mathcal{C}_0^\infty(T)$ and $i = 1, 2$, we find that

$$\begin{aligned} &\left\| \otimes_{k\in\mathbb{N}} (\eta^k)^{\otimes n_{ik}} - \otimes_{k\in\mathbb{N}} (\zeta^k)^{\otimes n_{ik}} \right\|_{\hat{L}^2(T^{\vec{n}_{i+}})} \\ &\qquad \leq \vec{n}_{i+} \max_{k: n_{ik} \neq 0} \left\| \eta^k - \zeta^k \right\|_{L^2(T)} \Big(\max_{k: n_{ik} \neq 0} \left\| \eta^k \right\|_{L^2(T)} + \max_{k: n_{ik} \neq 0} \left\| \zeta^k \right\|_{L^2(T)} \Big)^{\vec{n}_{i+}-1}. \end{aligned}$$

Since $\mathcal{C}_0^\infty(T)$ is dense in $L^2(T)$ we see that a sufficiently good approximation $\zeta \in \mathcal{C}_0^\infty(T)$ of η satisfies the lemma. □

(III.27)

IV

Simulation of pseudo-likelihoods given discretely observed data

Abstract

In this paper we study a simulation approach for computing pseudo-likelihoods, and their partial derivatives w.r.t. the model parameters, given an observation at discrete points of a stochastic partial differential equation. The method requires that the chaos expansion of the statistical model is known, *e.g.* can be calculated numerically. The needed Malliavin calculus is included in order to make the paper self contained. The derived formulae for the pseudo-likelihood contain iterated Skorohod integrals, and a procedure for simulating these iterated integrals is presented. The complexity of the procedure is shown to grow super-exponentially in the dimension of the densities used in the pseudo-likelihood. Moreover, the special cases corresponding to additive respectively multiplicative observation noise are discussed, and these cases are used to exemplify the numerical instability of the proposed method. This instability as well as the super-exponential growths can be avoided in the cases where observation noise is present.

Key words

Stochastic partial differential equation, discrete observations, pseudo-likelihood, Malliavin calculus, iterated Skorohod integral, additive observation noise, multiplicative observation noise.

IV.1 Introduction

In this paper we will discuss an approach to calculate pseudo-likelihoods given observations at discrete points in time and/or space of a parametric model of stochastic partial differential equations. We will assume that the chaos expansion of the model under consideration is available or can be computed numerically, *e.g.* the Wick-type stochastic PDE's studied in Holden et al. (1996), Theting (2000) will fit into our framework. The statistical framework we adobe consists of a probability space $(\Omega, \mathscr{A}, \mathrm{P})$, a temporal and/or spatial parameter set T, a model parameter space Θ, and a stochastic process

$$X : \Omega \times T \times \Theta \to \mathbb{R} \tag{IV.1}$$

such that $X(\cdot, t, \theta)$ is measurable given fixed $t \in T$ and $\theta \in \Theta$. Let $\theta_0 \in \Theta$ be the *true* model parameter and assume, that the stochastic process $X(\cdot, \cdot, \theta_0)$ have been observed at the points T_0, where T_0 is some fixed finite subset of T. Thus for some random element $\omega_0 \in \Omega$ we have observed the data

$$X(\omega_0, t, \theta_0) = x(t), \quad t \in T_0. \tag{IV.2}$$

In order to draw statistical inference for the parameter θ_0 we seek to calculate the likelihood $L(\theta)$ for the parameter $\theta \in \Theta$ based on the data (IV.2), and on the model given by (IV.1) and the probability measure P. In many situations this will however not be possible and some kind of approximation or simulation schemes have to be implemented. In order to simplify the calculations even more we will be content with pseudo-likelihoods PL(θ) written on the form

$$\log \mathrm{PL}(\theta) = \sum_{S \subseteq T_0} w_S \log \rho_S\big((x(s))_{s \in S}; \theta\big), \tag{IV.3}$$

where $w_S \in \mathbb{R}$ are some fixed weights, and where $\rho_S(y; \theta)$ is the density of the S-dimensional random variable $F(\theta) = (X(\cdot, s, \theta))_{s \in S}$. In Aerts & Claeskens (1999) some general results concerning the asymptotic behavior of the pseudo-likelihoods (IV.3) are stated. If $w_{T_0} = 1$ and $w_S = 0$ for $S \neq T_0$, then the pseudo-likelihood PL(θ) is equal to the full likelihood $L(\theta)$. If $w_{T_0} = |T_0|$, $w_{T_0 \setminus \{t\}} = -1$ for $t \in T_0$, and $w_S = 0$ otherwise, then PL(θ) is the so-called full conditional likelihood. The advantage of using conditional likelihoods is that normalizing constants not depending on the observed data cancel out. The methods developed in this paper are intended for models, where it is difficult to calculate not only the normalizing constants but also the actual form of the densities. We thus aim to calculate marginal pseudo-likelihoods written on the form

$$\log \mathrm{PL}(\theta) = \sum_{j=1}^{n} \log \rho_{S_j}\big((x(s))_{s \in S_j}; \theta\big), \quad S_1, \dots, S_n \subset T_0,$$

where the location subsets S_j each contain few points. If the model and the observation locations T_0 are spatially homogeneous, *i.e.* if T is equipped with a vector space structure and there exist subsets $\mathcal{N}$ and T_1 of the parameter space T such that $\mathcal{N} + t \subset T_0$ for every $t \in T_1$, and

$$\big(X(\cdot, s, \theta)\big)_{s\in\mathcal{N}+t_1} \overset{\mathcal{D}}{=} \big(X(\cdot, s, \theta)\big)_{s\in\mathcal{N}+t_2}, \quad t_1, t_2 \in T_1, \tag{IV.4}$$

then the corresponding marginal pseudo-likelihood is given by

$$\log \mathrm{PL}(\theta) = \sum_{t\in T_1} \log \rho_{\mathcal{N}}\big((x(s))_{s\in\mathcal{N}+t}; \theta\big), \tag{IV.5}$$

where $\rho_{\mathcal{N}}(y;\theta)$ is the common $\mathcal{N}$-dimensional density of the random variables (IV.4). The purpose of this paper is to discuss the method of simulating the formulae for $\rho_{\mathcal{N}}(y;\theta)$ derived via Malliavin calculus. In the present investigation we assume, that the underlying probability space is Gaussian. But the method presented could also be developed for *e.g.* Poisson spaces.

The paper is organized as follows. In section IV.2 we review the basic elements of Malliavin calculus needed in our analysis, in subsection IV.2.1 we derive the classical formulae for Lebesgue densities known from Malliavin calculus, and in subsection IV.2.2 we give the formulae for the partial derivatives of the Lebesgue densities w.r.t. the model parameters. In section IV.3 we discuss a procedure for calculating iterated Skorohod integrals. Since the computation of general iterated Skorohod integrals is numerically very demanding, we in section IV.4 describe the special cases corresponding to independent perturbations of the observations, *i.e.* observations with measurement errors. These special cases exemplifies that the formulae for Lebesgue densities derived via the integration by parts formula from Malliavin calculus are numerically instable. Finally in section IV.5 we conclude the findings of our investigation.

IV.2 Chaos decomposition and Malliavin calculus

In this section we give a crash course on the Wiener chaos and Malliavin calculus for Gaussian processes as presented in *e.g.* Nualart (1995). We will first introduce some notation specific to the statistical problem under consideration. Observe that all functions and spaces used below are real.

Throughout this paper let $\mathcal{N}$ be a finite index set equipped with a fixed ordering and disjoint from the positive integers $\mathbb{N}$. Moreover let $\mathcal{H}_0$ be a finite dimensional abstract Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ and orthonormal basis ξ_n, $n \in \mathcal{N}$. The Gaussian process connected to $\mathcal{H}_0$ will later be interpreted as observation noise, and demanding $\mathcal{N}$ being disjoint of $\mathbb{N}$ is simply a notational convenient way of introducing stochastic independence. Assume that the spatial parameter space T is equipped with a σ-field $\mathscr{B}$ and an atomfree measure λ, and let $\mathcal{H}_1 = L^2(T, \mathscr{B}, \lambda)$ be the associated Hilbert space of quadratic integrable functions with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_1}$. Assume that $\mathcal{H}_1$ is separable and let ξ_n, $n \in \mathbb{N}$, be some fixed orthonormal basis. Define the Hilbert space $\mathcal{H}$ as the direct sum of the Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$, *i.e.* $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with orthonormal basis ξ_n, $n \in \mathcal{N} \cup \mathbb{N}$.

(IV.3)

Let $(\Omega, \mathscr{A}, \mathrm{P})$ be a complete probability space, and assume that there exists an isonormal Gaussian process $\{W(h) : h \in \mathcal{H}\}$ defined on $(\Omega, \mathscr{A}, \mathrm{P})$, *i.e.* a family of centered Gaussian random variables such that

$$\mathrm{E}_{\mathrm{P}}\, W(h)W(g) = \langle h, g\rangle_{\mathcal{H}}\,, \quad h, g \in \mathcal{H}.$$

Let $H_n(x)$ denote the n'th *Hermite polynomial*, which is defined by

$$H_n(x) = (-1)^n\, e^{\frac{x^2}{2}}\, \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}} = \sum_{k=0}^{[\frac{n}{2}]} \big(-\tfrac{1}{2}\big)^k \frac{n!}{k!(n-2k)!}\, x^{n-2k}, \quad n \in \mathbb{N}, \tag{IV.6}$$

and $H_0(x) = 1$. The following two properties of the Hermite polynomials,

$$\begin{aligned} H_n'(x) &= n\, H_{n-1}(x), && n \in \mathbb{N}, \\ H_n(x)H_m(x) &= \sum_{k=0}^{n\wedge m} \frac{n!\, m!}{k!\,(n-k)!\,(m-k)!}\, H_{n+m-2k}(x), && n, m \in \mathbb{N}_0\,, \end{aligned} \tag{IV.7}$$

will be of importance to our considerations later. Let the index set $\mathcal{I}$ be defined by

$$\mathcal{I} = \big\{a = (a_n)_{n\in\mathbb{N}\cup\mathcal{N}} \in \mathbb{N}_0^{\mathcal{N}\cup\mathbb{N}} : a_n \neq 0 \text{ for finitely many } n \in \mathcal{N} \cup \mathbb{N}\big\},$$

and let $\varepsilon_n \in \mathcal{I}$, $n \in \mathcal{N} \cup \mathbb{N}$, be the vector that have a one on the n'th coordinate and zero's elsewhere. Then $\mathcal{I}$ is a $\mathbb{N}_0$-module under coordinate wise addition and scalar multiplication, and the elements ε_n, $n \in \mathcal{N} \cup \mathbb{N}$, constitute a basis for $\mathcal{I}$. Moreover, let the random variables Φ_a be defined by

$$\Phi_a = \prod_{n\in\mathcal{N}\cup\mathbb{N}} (a_n!)^{-\frac{1}{2}}\, H_{a_n}\big(W(\xi_n)\big), \quad a = (a_n)_{n\in\mathcal{N}\cup\mathbb{N}} \in \mathcal{I}. \tag{IV.8}$$

The following theorem due to Wiener (1938) states the *Wiener chaos decomposition*, which is an orthonormal decomposition of the Hilbert space of quadratic integrable random variables. The expansion of a random variable w.r.t. this basis is called the *chaos expansion*.

Theorem IV.1 *The random variables Φ_a, $a \in \mathcal{I}$, constitute an orthonormal basis for the Hilbert space $L^2(\Omega) = L^2(\Omega, \mathscr{G}, \mathrm{P})$, where $\mathscr{G}$ is the σ-field generated by the process $W(h)$, $h \in \mathcal{H}$.*

Next we describe the associated Malliavin calculus. Let $\mathcal{C}_p^\infty(\mathbb{R}^n)$ be the space of infinitely often continuous differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f and all its derivatives have polynomial growth, and denote by $\partial_j f$ the partial derivative of f w.r.t. the j'th argument. Moreover let $\mathcal{S}$ be the set of smooth random variables, *i.e.* a random variable F belongs to $\mathcal{S}$ if F can be written on the form

$$F = f\big(W(h_1), \dots, W(h_n)\big) \tag{IV.9}$$

for some $n \in \mathbb{N}$, $f \in \mathcal{C}_p^\infty(\mathbb{R}^n)$, and $h_1, \dots, h_n \in \mathcal{H}$. Observe that $\mathcal{S}$ is dense in $L^2(\Omega)$. The following theorem, which is shown in Nualart (1995), defines the *Malliavin derivative operator* D.

Theorem IV.2 *There exists a closed, densely defined, unbounded operator*

$$\mathrm{D} : L^2(\Omega) \to L^2(\Omega) \otimes \mathcal{H}$$

such that $\mathcal{S} \subset$ *Dom*(D) *and*

$$\mathrm{D}\, F = \sum_{j=1}^{n} \partial_j f\big(W(h_1), \dots, W(h_n)\big) \otimes h_j$$

for a random variable $F \in \mathcal{S}$ *written on the form* (IV.9).

The tensor product $L^2(\Omega) \otimes L^2(T)$ can be identified with the Hilbert space $L^2(\Omega \times T)$ of quadratic integrable stochastic processes. Similarly, the elements in the Hilbert space $L^2(\Omega) \otimes \mathcal{H}$ can be interpreted as stochastic processes indexed by the disjoint union of T and $\mathcal{N}$. If $f \in \mathcal{C}_b(\mathbb{R}^{\mathcal{N}})$, *i.e.* the function f and all its derivatives are bounded, and $F = (F_i)_{i \in \mathcal{N}}$ with $F_i \in$ Dom(D) are given, then the *chain rule* holds true, *i.e.* $f(F) \in$ Dom(D) with Malliavin derivative

$$\mathrm{D}\, f(F) = \sum_{i \in \mathcal{N}} \partial_i f(F)\ \mathrm{D}\, F_i,$$

see Nualart (1995, proposition 1.2.2). For given $\xi \in \mathcal{H}$ and $n \in \mathcal{N} \cup \mathbb{N}$ we introduce the shorthand notation

$$\mathrm{D}_\xi\, F = \langle \mathrm{D}\, F, \xi \rangle_{\mathcal{H}}\,, \qquad\qquad \mathrm{D}_n = \mathrm{D}_{\xi_n}\,.$$

It immediately follows from the properties (IV.7) of the Hermite polynomials, that the random variables Φ_a, $a \in \mathcal{I}$, satisfies the relations

$$\mathrm{D}_n\, \Phi_a = \sqrt{a_n}\, \Phi_{a-\varepsilon_n}, \quad n \in \mathcal{N} \cup \mathbb{N},\ a \in \mathcal{I}, \tag{IV.10a}$$

and

$$\Phi_a\, \Phi_b = \sum_{c \in \mathcal{I} : c_n \le a_n \wedge b_n} \left(\prod_{n \in \mathbb{N}} \frac{\sqrt{a_n! b_n! (a_n + b_n - 2c_n)!}}{c_n! (a_n - c_n)! (b_n - c_n)!} \right) \Phi_{a+b-2c}, \quad a, b \in \mathcal{I}. \tag{IV.10b}$$

Since the domain Dom(D) is dense, the adjoint operator δ of the Malliavin derivative operator D,

$$\delta : L^2(\Omega) \otimes \mathcal{H} \to L^2(\Omega),$$

is uniquely defined by the duality relation

$$\mathrm{E}_{\mathrm{P}}\, \big(\delta(u) F\big) = \mathrm{E}_{\mathrm{P}}\, \langle u, \mathrm{D}\, F \rangle_{\mathcal{H}}\,, \quad u \in \text{Dom}(\delta),\ F \in \text{Dom}(\mathrm{D}). \tag{IV.11}$$

As an adjoint operator δ is closed. Moreover, since D is closed, the adjoint operator δ is densely defined. The operator δ is called the *Skorohod stochastic integral*, and can be shown to be an extension of the Itô stochastic integral. The product Fu of a random variable $F \in \mathrm{Dom}(\mathrm{D})$ and a stochastic process $u \in \mathrm{Dom}(\delta)$ is Skorohod integrable if and only if $F\delta(u) - \langle \mathrm{D}\, F, u\rangle_{\mathcal{H}} \in L^2(\Omega)$, and if so then

$$\delta(Fu) = F\delta(u) - \langle \mathrm{D}\, F, u\rangle_{\mathcal{H}}, \quad F \in \mathrm{Dom}(\mathrm{D}),\ u \in \mathrm{Dom}(\delta), \tag{IV.12}$$

see Nualart (1995, equation (1.49)). This integration by parts formula will be crucial to the analysis below, and from the properties (IV.10) we especially find that

$$\begin{aligned}\delta(\Phi_a\, \xi_n) &= \Phi_a\, \Phi_{\varepsilon_n} - \mathrm{D}_n\, \Phi_a = \sqrt{a_n + 1}\, \Phi_{a+\varepsilon_n} + \sqrt{a_n}\, \Phi_{a-\varepsilon_n} - \mathrm{D}_n\, \Phi_a \\ &= \sqrt{a_n+1}\, \Phi_{a+\varepsilon_n}.\end{aligned}$$

IV.2.1 Calculation of subdensities

In this subsection we calculate lower approximations to Lebesgue densities of random variables. The following analysis is a classical application of the Malliavin calculus introduced above. For convenience we introduce some shorthand notation for iterated Skorohod integrals, *i.e.* for $k \in \mathbb{N}$ and sufficient regular processes $v, v_1, \ldots, v_k \in L^2(\Omega) \otimes \mathcal{H}$ we define

$$\delta(v_k, \ldots, v_1) = \delta\Big(v_k \delta\big(v_{k-1} \cdots \delta(v_1) \cdots\big)\Big), \qquad \delta_k(v) = \delta(\underbrace{v, \ldots, v}_{k \text{ times}}).$$

Observe that the iterated Skorohod integral $\delta(\cdot, \ldots, \cdot)$ is linear in each argument, but it is not symmetric in the arguments. Moreover, let $L^0(\Omega)$ denote the space of random variables, *i.e.* the space $\mathscr{G}$-measurable functions defined on Ω, and let $L^0(\Omega) \otimes \mathcal{H}$ denote the space of stochastic processes u such that $\langle u, h\rangle_{\mathcal{H}}$ is $\mathscr{G}$-measurable for every $h \in \mathcal{H}$.

Proposition IV.3 *Let $k \in \mathbb{N}$, $f \in \mathcal{C}_b^k(\mathbb{R}^{\mathcal{N}})$, $i_1, \ldots, i_k \in \mathcal{N}$, and a random variable $F = (F_i)_{i\in\mathcal{N}}$ with $F_i \in$ Dom(D) be given. If there exist random variables $G_i \in L^0(\Omega)$ and stochastic processes $u_i \in L^0(\Omega) \otimes \mathcal{H}$ such that*

$$\langle \mathrm{D}\, F_i, G_j u_j\rangle_{\mathcal{H}} = 1_{(i=j)}\, G_j, \quad i, j \in \mathcal{N}, \tag{IV.13}$$

and $G_{i_1}u_{i_1}, \ldots, G_{i_k}u_{i_k}\delta(G_{i_{k-1}}u_{i_{k-1}}, \ldots, G_{i_1}u_{i_1}) \in$ Dom(δ), then the integration by parts formula

$$\mathrm{E}_{\mathrm{P}}\left[G_{i_1} \cdots G_{i_k}\, \partial^k_{i_1,\ldots,i_k} f(F)\right] = \mathrm{E}_{\mathrm{P}}\left[f(F)\, \delta(G_{i_k}u_{i_k}, \ldots, G_{i_1}u_{i_1})\right]$$

holds true.

Proof If $g \in \mathcal{C}(\mathbb{R}^{\mathcal{N}})$ and $i \in \mathcal{N}$ are given, then

$$\langle \mathrm{D}\, g(F), G_i u_i\rangle_{\mathcal{H}} = G_i \sum_{j\in\mathcal{N}} \langle \partial_j g(F) \cdot \mathrm{D}\, F_j, u_i\rangle_{\mathcal{H}} = G_i\, \partial_i g(F)$$

follows by the chain rule. Iterating this identity with g substituted by the partial derivatives $\partial^l_{i_1,\dots,i_l} f$, $l = 1,\dots,k$, we see that

$$G_{i_1}\cdots G_{i_k}\,\partial^k_{i_1,\dots,i_k} f(F) = \left\langle \mathrm{D}\left\langle \mathrm{D}\cdots\left\langle \mathrm{D}\, f(F), G_{i_k}u_{i_k}\right\rangle_{\mathcal{H}}\cdots, G_{i_2}u_{i_2}\right\rangle_{\mathcal{H}}, G_{i_1}u_{i_1}\right\rangle_{\mathcal{H}}.$$

The statement of the proposition then follows from the duality (IV.11) between D and δ. □

Theorem IV.4 *Let $F = (F_i)_{i\in\mathcal{N}}$, $(G_i)_{i\in\mathcal{N}}$ and $(u_i)_{i\in\mathcal{N}}$ be as in proposition IV.3, and moreover assume $0 \le G_i \le 1$. If $\rho(y)$ is defined by*

$$\rho(y) = \mathrm{E}_{\mathrm{P}}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\,\delta\big((G_iu_i)_{i\in\mathcal{N}}\big)\Big], \qquad y = (y_i)_{i\in\mathcal{N}},$$

then the inequality

$$0 \le \mathrm{P}\big(F\in A\big) - \int_A \rho(y)\,\mathrm{d}y \le \mathrm{P}\big((F\in A)\cap \cup_{i\in\mathcal{N}}(G_i<1)\big)$$

holds true for every $A \in \mathscr{B}(\mathbb{R}^{\mathcal{N}})$.

Proof W.l.o.g. we can assume that A is the box $\times_{i\in\mathcal{N}}[a_i,b_i]$. For each $n\in\mathbb{N}$ let $\psi^{(n)} \in \mathcal{C}^{|\mathcal{N}|-1}(\mathbb{R}^{\mathcal{N}};\mathbb{R}_+)$ be a non-negative smooth approximation of $\otimes_{i\in\mathcal{N}}1_{[a_i,b_i]}$ such that $\psi^{(n)}$ grows pointwise to this indicator function as $n\to\infty$, and let

$$\phi^{(n)}(x) = \int_{\times_{i\in\mathcal{N}}(-\infty,x_i]} \psi^{(n)}(y)\,\mathrm{d}y.$$

Since $G_i \le 1$ and by proposition IV.3 we find that

$$\begin{aligned}
\mathrm{E}_{\mathrm{P}}\left[\psi^{(n)}(F)\right] &= \mathrm{E}_{\mathrm{P}}\left[\partial^{|\mathcal{N}|}_{(x_i)_{i\in\mathcal{N}}}\phi^{(n)}(F)\right]\\
&\ge \mathrm{E}_{\mathrm{P}}\Big[\Big(\prod_{i\in\mathcal{N}} G_i\Big)\,\partial^{|\mathcal{N}|}_{(x_i)_{i\in\mathcal{N}}}\phi^{(n)}(F)\Big] \qquad\qquad (\mathrm{IV}.14)\\
&= \mathrm{E}_{\mathrm{P}}\left[\phi^{(n)}(F)\,\delta\big((G_iu_i)_{i\in\mathcal{N}}\big)\right].
\end{aligned}$$

For $n\to\infty$ dominated convergence and the Tonelli theorem then gives

$$\begin{aligned}
\mathrm{P}(F\in A) &= \mathrm{E}_{\mathrm{P}}\left[\,\otimes_{i\in\mathcal{N}} 1_{[a_i,b_i]}(F)\right]\\
&\ge \mathrm{E}_{\mathrm{P}}\Big[\int_{\times_{i\in\mathcal{N}}(-\infty,F_i]} 1_{\otimes_{i\in\mathcal{N}}[a_i,b_i]}(y)\,\mathrm{d}y\;\delta\big((G_iu_i)_{i\in\mathcal{N}}\big)\Big]\\
&= \mathrm{E}_{\mathrm{P}}\Big[\int_{\times_{i\in\mathcal{N}}[a_i,b_i]}\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\,\delta\big((G_iu_i)_{i\in\mathcal{N}}\big)\,\mathrm{d}y\Big]\\
&= \int_A \mathrm{E}_{\mathrm{P}}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\,\delta\big((G_iu_i)_{i\in\mathcal{N}}\big)\Big]\,\mathrm{d}y.
\end{aligned}$$

This proves the first inequality stated in the theorem. The second inequality follows similarly. □

Theorem IV.4 states that $\rho(y)$ is a lower approximation of the density of the absolute continuous part of the law of the $\mathcal{N}$-dimensional random variable F. Moreover, the approximation error is bounded by intersecting with the event that $G_i < 1$ for some $i \in \mathcal{N}$. Since the order in which the partial derivatives are taken in the proof of theorem IV.4 is irrelevant, we see that the mean value of the random variable

$$\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\,\delta\big((G_i u_i)_{i\in\mathcal{N}}\big)$$

is both linear and symmetric in the arguments $G_i u_i$, $i \in \mathcal{N}$. Using polarization we thus get the following corollary.

Corollary IV.5 *Let $F = (F_i)_{i\in\mathcal{N}}$, $(G_i)_{i\in\mathcal{N}}$, $(u_i)_{i\in\mathcal{N}}$ and $\rho(y)$ be as in theorem IV.4. If we put $\sigma_{i_0} = 1$ for some fixed $i_0 \in \mathcal{N}$, then $\rho(y)$ is given by*

$$(|\mathcal{N}|!)^{-1}2^{-|\mathcal{N}|+1} \sum_{\sigma_i\in\{-1,+1\}, i\in\mathcal{N}\setminus\{i_0\}} \mathrm{E}_{\mathrm{P}}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\,\delta_{|\mathcal{N}|}\Big(\sum_{i\in\mathcal{N}}\sigma_i G_i u_i\Big)\Big].$$

In order to apply theorem IV.4 or corollary IV.5 we need to find suitable processes $u_i \in L^0(\Omega)\otimes\mathcal{H}$ and variables $G_i \in L^0(\Omega)$. Let the functions $\tilde{G}_{\alpha,\beta} \in \mathcal{C}^m(\mathbb{R})$, $0 < \alpha < \beta < \infty$, be defined by

$$\tilde{G}_{\alpha,\beta}(x) = \begin{cases} 0 & \text{for } |x| \in [0,\alpha), \\ \frac{1}{B(m,m)}\int_0^{\frac{|x|-\alpha}{\beta-\alpha}} s^{m-1}(1-s)^{m-1}\,\mathrm{d}s & \text{for } |x| \in [\alpha,\beta), \\ 1 & \text{for } |x| \in [\beta,\infty), \end{cases}$$

where $B(m,n)$ denotes the Beta-function with parameters m,n.

Proposition IV.6 *For fixed $\tilde{u}_i \in L^0(\Omega)\otimes\mathcal{H}$, $i \in \mathcal{N}$, let the stochastic matrix Γ be defined by*

$$\Gamma = \big(\langle \mathrm{D}\,F_i, \tilde{u}_j\rangle_{\mathcal{H}}\big)_{i,j\in\mathcal{N}}.$$

Given a row vector $x \in \mathbb{R}^{\mathcal{N}}$ and row index $i \in \mathcal{N}$, let $\Gamma^{x\to i}$ be the matrix constructed by substituting the i'th row in Γ by x. Then the stochastic processes u_i, $i \in \mathcal{N}$, and the random variables $G_i = G_{\alpha,\beta}$, $i \in \mathcal{N}$, given by

$$u_i = \frac{1_{(\det\Gamma\neq 0)}}{\det\Gamma}\det\Gamma^{(\tilde{u}_k)_{k\in\mathcal{N}}\to i}, \qquad\qquad G_{\alpha,\beta} = \tilde{G}_{\alpha,\beta}(\det\Gamma)$$

fulfill condition (IV.13).

Proof This is essentially Cramer's rule, *i.e.* the linearity of the determinant and the inner product gives

$$\langle \mathrm{D}\,F_i, G_j u_j\rangle_{\mathcal{H}} = \frac{\tilde{G}_{\alpha,\beta}(\det\Gamma)}{\det\Gamma}\det\Gamma^{\left(\langle \mathrm{D}\,F_i,\tilde{u}_k\rangle_{\mathcal{H}}\right)_{k\in\mathcal{N}}\to j} = 1_{(i=j)}G_j.$$

□

(IV.8)

Remark In theoretical applications, *e.g.* development of density criterion, the usual choice of $\tilde{u}_i$ is $\mathrm{D}\,F_i$, and the resulting Γ matrix

$$\big(\langle \mathrm{D}\,F_i, \mathrm{D}\,F_j \rangle_{\mathcal{H}} \big)_{i,j\in\mathcal{N}},$$

which is called the *Malliavin covariance matrix*, is positive semi definite and has maximal determinant. The choice $\tilde{u}_i = \mathrm{D}\,F_i$ is thus natural in these situations. But for the purpose of simulating the subdensity $\rho(y)$ other choices of $\tilde{u}_i$ could be used in order reduce the number of terms in the iterated Skorohod integrals.

Remark Suppose the processes $\tilde{u}_i$, $i \in \mathcal{N}$, have been chosen, and that the processes $u_i \in L^0(\Omega) \otimes \mathcal{H}$, $i \in \mathcal{N}$, and the variable $G_{\alpha,\beta} \in L^\infty(\Omega)$ have been constructed via proposition IV.6. To apply theorem IV.4 or corollary IV.5 we still need to show that the employed Skorohod integrals exist. If the random variables F_i and the stochastic processes $\tilde{u}_i$ all have finite chaos expansion, then it is easily seen that we only take Skorohod integrals of processes $v \in L^0(\Omega) \otimes \mathcal{H}$ which satisfy the condition

$$\mathrm{E}_{\mathrm{P}} \langle \mathrm{D}\,v, \mathrm{D}\,v \rangle_{\mathcal{H}\otimes\mathcal{H}} < \infty$$

ensuring that $v \in \mathrm{Dom}(\delta)$, see Nualart (1995, p. 40). In the remaining of this paper we will implicitly assume, that all the Skorohod integrals under consideration exist.

Let $\rho_{\alpha,\beta}(y)$ be the subdensity defined in theorem IV.4 with u_i and $G_{\alpha,\beta}$ given by proposition IV.6 for fixed $0 < \alpha < \beta < \infty$. Inspecting the inequality (IV.14) in the proof of theorem IV.4, we see that the subdensities $\rho_{\alpha,\beta}(y)$ satisfy the following monotinicity property

$$0 < \alpha_1 \le \alpha_2 < \beta_2,\ 0 < \alpha_1 < \beta_1 \le \beta_2 \quad \Longrightarrow \quad \rho_{\alpha_2,\beta_2}(y) \le \rho_{\alpha_1,\beta_1}(y). \tag{IV.15}$$

Moreover, $\rho_\alpha(y) = \lim_{\beta\downarrow\alpha} \rho_{\alpha,\beta}(y)$ is the Lebesgue density of the measure

$$\mathrm{P}\big((F \in \cdot) \cap (|\det\Gamma| > \alpha)\big),$$

and by dominated convergence it follows that $\rho_\alpha(y)$ is given by

$$\mathrm{E}_{\mathrm{P}}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\, 1_{(|\det\Gamma|>\alpha)}\, \delta\big((u_i)_{i\in\mathcal{N}}\big)\Big], \tag{IV.16a}$$

where the iterated Skorohod integral $\delta((u_i)_{i\in\mathcal{N}})$ is defined using (IV.12) formally. Defining $u^\sigma = \sum_{i\in\mathcal{N}} \sigma_i u_i$, $\sigma_{i_0} = 1$, the subdensity $\rho_\alpha(y)$ is equivalently given by

$$(|\mathcal{N}|!)^{-1} 2^{-|\mathcal{N}|+1} \sum_{\sigma_i\in\{-1,+1\}, i\in\mathcal{N}\setminus\{i_0\}} \mathrm{E}_{\mathrm{P}}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\, 1_{(|\det\Gamma|>\alpha)}\, \delta_{|\mathcal{N}|}(u^\sigma)\Big]. \tag{IV.16b}$$

If the matrix Γ is invertible almost surely, then the law of F is absolute continuous and the $\mathcal{N}$-dimensional Lebesgue density $\rho(y)$ is given by

$$\rho(y) = \sup_{\alpha>0} \rho_\alpha(y) = \lim_{\alpha\downarrow 0} \rho_\alpha(y).$$

In order to compute the formulae for $\rho_{\alpha,\beta}(y)$, the immediate approach is to replace the mean values by the average $\hat{\rho}^{(n)}_{\alpha,\beta}(y)$ of n i.id. samples of

$$\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\,\delta\big((G_{\alpha,\beta}u_i)_{i\in\mathcal{N}}\big), \tag{IV.17}$$

cf. theorem IV.4, or n i.id. samples of

$$(|\mathcal{N}|!)^{-1}2^{-|\mathcal{N}|+1}\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\sum_{\sigma_i\in\{-1,+1\},i\in\mathcal{N}\setminus\{i_0\}}\delta_{|\mathcal{N}|}\big(G_{\alpha,\beta}u^\sigma\big), \tag{IV.18}$$

cf. corollary IV.5. The random variable (IV.18) consists of a weighted sum of $2^{|\mathcal{N}|-1}$ terms, but since the symmetric iterated Skorohod integral $\delta_{|\mathcal{N}|}(\cdot)$ is more easy to calculate than the non-symmetric iterated Skorohod integral $\delta(\cdot,\dots,\cdot)$ as demonstrated in section IV.3, it is actually more easily simulated than the random variable (IV.17). Moreover since the non-symmetric iterated Skorohod integrals $\delta(G_{\alpha,\beta}u_{i_{|\mathcal{N}|}},\dots,G_{\alpha,\beta}u_{i_1})$ are different for different permutations $(i_1,\dots,i_{|\mathcal{N}|})$ of $\mathcal{N}$, the variance of (IV.18) is smaller than the variance of (IV.17). It is thus preferable to use the random variables (IV.18) for the simulation of $\rho_{\alpha,\beta}(y)$. The following theorem states a uniform law of large numbers, which justifies the simulation approach.

Theorem IV.9 *If we defined $\rho(y)=\lim_{0<\alpha<\beta\to 0}\rho_{\alpha,\beta}(y)$, then the limit result*

$$\sup_{0<\alpha<\beta\le 1}\Big|\rho_{\alpha,\beta}(y)-\hat{\rho}^{(n)}_{\alpha,\beta}(y)\Big| \xrightarrow[n\to\infty]{a.s.} 0, \quad y\in\mathbb{R}^m \text{ such that } \rho(y)<\infty,$$

holds true.

Proof For fixed $0<\alpha<\beta\le 1$ the strong law of large numbers gives

$$\hat{\rho}^{(n)}_{\alpha,\beta}(y)\xrightarrow[n\to\infty]{\text{a.s}}\rho_{\alpha,\beta}(y).$$

A Glivenko-Cantelli argument using the monotonicity property (IV.15) easily gives, that this statement can be strengthened to hold uniformly over $0<\alpha<\beta\le 1$. □

The uniform convergence stated in theorem IV.9 permits, that the cutoff points $0<\alpha<\beta\le 1$ can be chosen after the samples have been drawn. By the chain rule $\mathrm{D}\,G_{\alpha,\beta}=0$ on the event $(|\det\Gamma|\notin[\alpha,\beta])$, whence the average $\hat{\rho}^{(n)}_{\alpha}(y)$ of n i.id. samples of

$$(|\mathcal{N}|!)^{-1}2^{-|\mathcal{N}|+1}\Big(\prod_{i\in\mathcal{N}} 1_{(F_i>y_i)}\Big)\sum_{\sigma_i\in\{-1,+1\},i\in\mathcal{N}\setminus\{i_0\}}1_{(|\det\Gamma|>\alpha)}\,\delta_{|\mathcal{N}|}\big(u^\sigma\big), \tag{IV.19}$$

converges almost surely to $\rho_\alpha(y)$ as $n\to\infty$. Moreover $\rho_\alpha(y)$ increases to $\rho(y)$ when α decreases to 0, this however at the cost that the variance of the samples (IV.19) might increase to possibly infinity. Choosing the cutoff point $\alpha>0$ thus involves a trade off.

(IV.10)

IV.2.2 Calculation of partial derivatives w.r.t. the model parameters

Suppose that the random variable $F(\theta) = (F_i(\theta))_{i\in\mathcal{N}}$ is parameterized by a model parameter $\theta \in \Theta$. If we use the setup leading to equation (IV.16), then the lower approximation $\rho_\alpha(y;\theta)$ of the Lebesgue density of the absolute continuous part of the law of $F(\theta)$ is given by

$$(|\mathcal{N}|!)^{-1}2^{-|\mathcal{N}|+1} \sum_{\substack{\sigma_i\in\{-1,+1\},\\ i\in\mathcal{N}\setminus\{i_0\}}} \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\theta)>y_i)}\Big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big].$$

In this subsection we calculate partial derivatives of $\rho_\alpha(y;\theta)$ w.r.t. the parameter θ. For ease of notation we will w.l.o.g. assume that Θ is an open subset of $\mathbb{R}$. Thus using the product rule for differentiation we find that

$$\partial_\theta\, \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\theta)>y_i)}\Big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big]$$

is equal to

$$\begin{aligned}
&\partial_\eta\, \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\eta)>y_i)}\Big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big]\Big|_{\eta=\theta}\\
&+\partial_\eta\, \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\theta)>y_i)}\Big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\eta)\big)\Big]\Big|_{\eta=\theta}\\
&+\partial_\eta\, \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\theta)>y_i)}\Big)\, 1_{(|\det\Gamma(\eta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big]\Big|_{\eta=\theta}.
\end{aligned} \tag{IV.20}$$

The last term in (IV.20) corresponds to differentiation w.r.t. the truncation variable and can be rewritten as

$$-\partial_\eta \frac{\mathrm{dP}\big((F(\theta)\in\cdot)\cap(|\det\Gamma(\eta)|\le\alpha)\big)}{\mathrm{dLebesgue}}(y)\Big|_{\eta=\theta}.$$

Thus if $\rho_\alpha(y;\theta)$ is a good approximation of the Lebesgue density $\rho(y;\theta)$, *i.e.* if the cutoff point $\alpha > 0$ is sufficiently small, then this term will be small compared to the two other terms in (IV.20). Since the last term in (IV.20) also is difficult to calculate, it is reasonable simply to ignore this term. The second term in (IV.20) is equal to

$$\mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\theta)>y_i)}\Big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \partial_\theta\delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big],$$

where $\partial_\theta\delta_{|\mathcal{N}|}(u^\sigma(\theta))$ can be found via proposition IV.12 below. To calculate the first term in (IV.20) we once more use the integration by parts formula. Thus let the functions $\psi^{(n)} \in \mathcal{C}(\mathbb{R};\mathbb{R}_+)$, $n\in\mathbb{N}$, be non-negative approximations of $1_{(0,\infty)}$ such that $\psi^{(n)}$ grows pointwise to this indicator function as $n\to\infty$, and let $\phi^{(n)} \in$

$\mathcal{C}(\mathbb{R}^{\mathcal{N}} \times \mathbb{R}^{\mathcal{N}}; \mathbb{R}_+)$ be defined by $\phi^{(n)}(x,y) = \prod_{i\in\mathcal{N}} \psi^{(n)}(x_i - y_i)$. Using the duality (IV.11), and using the identity

$$\begin{aligned}\partial_\theta \phi^{(n)}\big(F(\theta), y\big) &= \sum_{i\in\mathcal{N}} \partial_\theta F_i(\theta)\, \tfrac{\mathrm{d}}{\mathrm{d}x}\psi^{(n)}\big(F_i(\theta) - y_i\big) \prod_{j\neq i} \psi^{(n)}\big(F_j(\theta) - y_j\big) \\ &= \Big\langle \mathrm{D}\, \phi^{(n)}\big(F(\theta), y\big), \sum_{i\in\mathcal{N}} \partial_\theta F_i(\theta)\, u_i(\theta) \Big\rangle_{\mathcal{H}}\end{aligned}$$

on the event $(|\det\Gamma(\theta)| > \alpha)$, we see that

$$\begin{aligned}&\partial_\eta \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\eta) > y_i)}\Big) 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big]\Big|_{\eta=\theta} \\ &= \lim_{n\to\infty} \mathrm{E_P}\Big[\partial_\theta\phi^{(n)}\big(F(\theta), y\big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big] \\ &= \lim_{n\to\infty} \mathrm{E_P}\Big\langle \mathrm{D}\,\phi^{(n)}\big(F(\theta), y\big), \sum_{i\in\mathcal{N}} \partial_\theta F_i(\theta)\, u_i(\theta)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big\rangle_{\mathcal{H}} \\ &= \lim_{n\to\infty} \mathrm{E_P}\Big[\phi^{(n)}\big(F(\theta), y\big)\, 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta\Big(\sum_{i\in\mathcal{N}} \partial_\theta F_i(\theta)\, u_i(\theta)\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big)\Big] \\ &= \mathrm{E_P}\Big[\Big(\prod_{i\in\mathcal{N}} 1_{(F_i(\theta)>y_i)}\Big) 1_{(|\det\Gamma(\theta)|>\alpha)}\, \delta\Big(\sum_{i\in\mathcal{N}} \partial_\theta F_i(\theta)\, u_i(\theta)\, \delta_{|\mathcal{N}|}\big(u^\sigma(\theta)\big)\Big)\Big].\end{aligned}$$

IV.3 Calculation of iterated Skorohod integrals

In this section we consider the problem of calculating the non-symmetric iterated Skorohod integrals $\delta(v_k, \ldots, v_1)$ and the symmetric iterated Skorohod integrals $\delta_k(v)$ introduced and heavily used in the preceding section. Let us introduce some more notation. Let the set $\mathcal{J}$ and the subsets $\mathcal{J}_k$ be defined by

$$\begin{aligned}\mathcal{J} &= \big\{a = (a_n)_{n\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}} : a_n \neq 0 \text{ for finitely many } n \in \mathbb{N}\big\}, \\ \mathcal{J}_k &= \Big\{a = (a_n)_{n\in\mathbb{N}} \in \mathcal{J} : \sum_{n\in\mathbb{N}} n a_n = k\Big\}, \quad k \in \mathbb{N}_0,\end{aligned}$$

and let $|a| = \sum_{n\in\mathbb{N}} a_n$, $a \in \mathcal{J}$. Although the sets $\mathcal{I}$ and $\mathcal{J}$ have similar structure, these sets are used differently and should not be confused with each other. Whereas the indices of the coordinates of the elements in $\mathcal{I}$ simply are indices with no special meaning, the index set $\mathbb{N}$ of the coordinates of the elements in $\mathcal{J}$ is to be understood as the sequentially ordered integers. Moreover, given sufficient regular processes $v_1, \ldots, v_k \in L^2(\Omega) \otimes \mathcal{H}$ let the random variable $\Psi(v_1, \ldots, v_k) \in L^2(\Omega)$ be defined recursively by

$$\Psi(v_1) = \delta(v_1), \qquad \Psi(v_{n+1}, \ldots, v_1) = \langle v_{n+1}, \mathrm{D}\,\Psi(v_n, \ldots, v_1)\rangle_{\mathcal{H}}.$$

The iterated Skorohod integrals can the be calculated with the aid of the following proposition.

(IV.12)

Proposition IV.10 *Given stochastic processes $v_1, \dots, v_k \in Dom(\delta)$ the integration by parts formula* (IV.12) *formally yields that $\delta(v_k, \dots, v_1)$ is equal to*

$$\sum_{a \in \mathcal{J}_k} \sum_{\substack{A_{nj}, n\in\mathbb{N}, j=1,\dots,a_n, \\ \text{partition of } \{1,\dots,k\} \\ \text{such that } |A_{nj}|=n}} (-1)^{k-|a|} \prod_{n\in\mathbb{N}} \prod_{j=1}^{a_n} \Psi(v_{i_n}, \dots, v_{i_1})\Big|_{i_1<\dots<i_n,\ i_l \in A_{nj}}.$$

Proof The statement is proved by induction over $k \in \mathbb{N}$. For $k = 1$ the statement holds true since $\Psi(v_1) = \delta(v_1)$. If we introduce the shorthand notation

$$\Psi(A_{nj}) = \Psi(v_{i_n}, \dots, v_{i_1})\Big|_{i_1<\dots<i_n,\ i_l \in A_{nj}},$$

then the induction step follows by the calculations

$$\begin{aligned}
\delta(v_{k+1}, \dots, v_1) &= \delta\big(v_{k+1}, \delta(v_k, \dots, v_1)\big) \\
&= \delta(v_{k+1})\delta(v_k, \dots, v_1) - \langle v_{k+1}, \mathrm{D}\,\delta(v_k, \dots, v_1)\rangle_{\mathcal{H}} \\
&= \sum_{a \in \mathcal{J}_k} \sum_{\substack{A_{nj}, n\in\mathbb{N}, j=1,\dots,a_n, \\ \text{partition of } \{1,\dots,k\} \\ \text{such that } |A_{nj}|=n}} (-1)^{k-|a|} \Bigg(\underbrace{\delta(v_{k+1})}_{\Psi(\{k+1\})} \prod_{n\in\mathbb{N}} \prod_{j=1}^{a_n} \Psi(A_{nj}) \\
&\qquad\qquad - \sum_{n_0\in\mathbb{N}} \sum_{j_0=1}^{a_{n_0}} \underbrace{\langle v_{k+1}, \mathrm{D}\,\Psi(A_{n_0 j_0})\rangle_{\mathcal{H}}}_{\Psi(A_{n_0 j_0} \cup \{k+1\})} \prod_{(n,j)\neq(n_0,j_0)} \Psi(A_{nj}) \Bigg) \\
&= \sum_{a \in \mathcal{J}_{k+1}} \sum_{\substack{A_{nj}, n\in\mathbb{N}, j=1,\dots,a_n, \\ \text{partition of } \{1,\dots,k,k+1\} \\ \text{such that } |A_{nj}|=n}} (-1)^{k+1-|a|} \prod_{n\in\mathbb{N}} \prod_{j=1}^{a_n} \Psi(A_{nj}).
\end{aligned}$$

□

If we insert $v_1 = \dots = v_k = v$ in proposition IV.10 we immediately get the following corollary.

Corollary IV.11 *Given $k \in \mathbb{N}$ and a stochastic process $v \in Dom(\delta)$ the integration by parts formula* (IV.12) *formally yields*

$$\delta_k(v) = \sum_{a\in\mathcal{J}_k} (-1)^{k-|a|} \frac{k!}{\prod_{n\in\mathbb{N}} a_n!(n!)^{a_n}} \prod_{n\in\mathbb{N}} \Psi_n^{a_n}(v), \qquad \Psi_n(v) = \Psi(\underbrace{v, \dots, v}_{n \text{ times}}).$$

In order to assess the complexity of these formulae we consider the number $|\mathcal{J}_k|$ of elements in the set $\mathcal{J}_k$. The quantity $|\mathcal{J}_k|$ equals the number of ways in which k can be written as a sum of integers. The following table

k :	1	2	3	4	5	6	7	8	9	10
$\lvert\mathcal{J}_k\rvert$:	1	2	3	5	7	11	15	22	30	42

(IV.13)

contains the numbers $|\mathcal{J}_k|$ for $k = 1, \dots, 10$. A famous result proved in Hardy & Ramanujan (1918) implies, that the asymptotic behavior is given by

$$|\mathcal{J}_k| \sim \tfrac{1}{4\sqrt{3}} k^{-1} e^{\pi\sqrt{\frac{2}{3}k}} \text{ as } k \to \infty.$$

Considering the formula for $\delta_k(v)$ given in corollary IV.11 it is thus tempting to try to reduce the number of terms in the sum over $\mathcal{J}_k$ via rewriting the sum of products as a sum of products of sums. But since calculating the quantities $\Psi(v_n, \dots, v_1)$, $n = 1, \dots, k$, is a rather difficult task, and it is thus not feasible to calculate $\delta_k(v)$ for k large anyway, we will not pursue such a refinement further. Before calculating $\Psi(v_n, \dots, v_1)$ we state the following proposition, which was used in subsection IV.2.2 above.

Proposition IV.12 *Given $k \in \mathbb{N}$ and a family stochastic process v_θ parameterized by a one-dimensional real parameter θ the integration by parts formula* (IV.12) *formally yields*

$$\partial_\theta \delta_k(v_\theta) = \sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} \partial_\theta \Psi_j(v_\theta)\, \delta_{k-j}(v_\theta),$$

where $\delta_0(v_\theta) = 1$ by definition. Moreover, $\partial_\theta \Psi_n(v_\theta)$ is given recursively by

$$\partial_\theta \Psi_1(v_\theta) = \delta(\partial_\theta v_\theta), \qquad \partial_\theta \Psi_{n+1}(v_\theta) = \langle \partial_\theta v_\theta, \mathrm{D}\,\Psi_n(v_\theta)\rangle_{\mathcal{H}} + \langle v_\theta, \mathrm{D}\,\partial_\theta \Psi_n(v_\theta)\rangle_{\mathcal{H}}\,.$$

Proof Differentiating the formula given in corollary IV.11 we find that

$$\begin{aligned}
\partial_\theta \delta_k(v_\theta) &= \sum_{a \in \mathcal{J}_k} (-1)^{k-|a|} \frac{k!}{\prod_{n\in\mathbb{N}} a_n!\,(n!)^{a_n}} \sum_{j\in\mathbb{N}} a_j\, \partial_\theta \Psi_j(v_\theta) \prod_{n\in\mathbb{N}} \Psi_n^{a_n - 1_{(n=j)}}(v_\theta) \\
&= \sum_{j=1}^{k} \partial_\theta \Psi_j(v_\theta) \sum_{a\in\mathcal{J}_{k-j}} (-1)^{k-|a|+1} \frac{k!}{j!\prod_{n\in\mathbb{N}} a_n!\,(n!)^{a_n}} \prod_{n\in\mathbb{N}} \Psi_n^{a_n}(v_\theta) \\
&= \sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} \partial_\theta \Psi_j(v_\theta)\, \delta_{k-j}(v_\theta),
\end{aligned}$$

which proves the first part of the proposition. The second part follows by differentiating the identity $\Psi_{n+1}(v_\theta) = \langle v_\theta, \mathrm{D}\,\Psi_n(v_\theta)\rangle_{\mathcal{H}}$. □

Remark If the random variables $\Psi(v_{i_n}, \dots, v_{i_1})$ and the used products of these random variables all belong to $L^2(\Omega)$, then the formulae presented in the two preceding propositions are not just merely formal calculations but actually equals the sought for Skorohod integrals. This follows from the more careful interpretation of the integration by parts formula, which also include a statement regarding the existence of the Skorohod integrals.

(IV.14)

In order to make the following considerations as simple as possible we will only calculate the symmetric operator $\Psi_n(v)$ for a particular structure of the stochastic process v. But if we use proposition IV.6, then this special case will still be adequate for our purpose of finding the subdensity $\rho_{\alpha,\beta}(y)$. In the sequel let $\mathcal{M}$ be a fixed finite subset of $\mathcal{N}\cup\mathbb{N}$, and let $M, N \in \mathbb{N}$, $M \geq N$, be fixed integers. Let the subsets $\mathcal{I}^{\mathcal{M}}$ and $\mathcal{I}^{\mathcal{M}}_{\gamma}$, $\gamma \in \mathbb{N}$, of $\mathcal{I}$ be defined by

$$\mathcal{I}^{\mathcal{M}} = \big\{a \in \mathcal{I} : a_n = 0 \text{ for } n \in (\mathcal{N}\cup\mathbb{N}) \setminus \mathcal{M}\big\},$$
$$\mathcal{I}^{\mathcal{M}}_{\gamma} = \big\{a \in \mathcal{I}^{\mathcal{M}} : |a| \leq \gamma\big\}, \quad \gamma \in \mathbb{N}_0.$$

The particular structure we will assume for the stochastic process v is given by

$$\begin{aligned} v &= Y^{-1} \sum_{j\in\mathcal{M}} X_j\, \xi_j, \\ Y &= \sum_{a\in\mathcal{I}:\sum_{n\in\mathcal{M}} a_n \leq M} y(a)\, \Phi_a, \quad y(a) \in \mathbb{R} \text{ deterministic}, \\ X_j &= \sum_{a\in\mathcal{I}:\sum_{n\in\mathcal{M}} a_n \leq N} x_j(a)\, \Phi_a, \quad x_j(a) \in \mathbb{R} \text{ deterministic}, \ j \in \mathcal{M}. \end{aligned} \tag{IV.21}$$

The random variables Y and X_j, $j \in \mathcal{M}$, can be rewritten as

$$Y = \sum_{a\in\mathcal{I}^{\mathcal{M}}_M} y^{\mathcal{M}}(a)\, \Phi_a, \qquad y^{\mathcal{M}}(a) = \sum_{b\in\mathcal{I}^{(\mathcal{N}\cup\mathbb{N})\setminus\mathcal{M}}} y(a+b)\, \Phi_b, \quad a \in \mathcal{I}^{\mathcal{M}}_M,$$
$$X_j = \sum_{a\in\mathcal{I}^{\mathcal{M}}_N} x_j^{\mathcal{M}}(a)\, \Phi_a, \qquad x_j^{\mathcal{M}}(a) = \sum_{b\in\mathcal{I}^{(\mathcal{N}\cup\mathbb{N})\setminus\mathcal{M}}} x_j(a+b)\, \Phi_b, \quad a \in \mathcal{I}^{\mathcal{M}}_N.$$

Observe that the stochastic coefficients $y^{\mathcal{M}}(a)$, $x_j^{\mathcal{M}}(a)$ are measurable w.r.t. the σ-field generated by the random variables $W(\xi_j)$, $j \in (\mathcal{N}\cup\mathbb{N}) \setminus \mathcal{M}$.

Proposition IV.14 *If the stochastic process $v \in Dom(\delta)$ can be written on the form* (IV.21)*, then $\Psi_n(v)$ can be written on the form*

$$\Psi_n(v) = \sum_{m=1}^{n+1} Y^{1-n-m}\, X_{n,m}, \qquad X_{n,m} = \sum_{a\in\mathcal{I}^{\mathcal{M}}_{\gamma_{n,m}}} x^{\mathcal{M}}_{n,m}(a)\, \Phi_a, \tag{IV.22}$$

where $\gamma_{n,m} = (m-1)M + n(N-1) + 2\cdot 1_{(m\leq n)}$, and the stochastic coefficients $x^{\mathcal{M}}_{n,m}(a)$ are measurable w.r.t. the σ-field generated by $W(\xi_j)$, $j \in (\mathcal{N}\cup\mathbb{N}) \setminus \mathcal{M}$.

Proof The proof in done by induction over $n \in \mathbb{N}$. Using the integration by parts formula (IV.12) and the properties (IV.10) of the random variables Φ_a, $a \in \mathcal{I}$, the statement for $n = 1$ follows from the calculations

$$\delta(v) = Y^{-1} \sum_{j\in\mathcal{M}} \big(X_j\, \Phi_{\varepsilon_j} - \mathrm{D}_j\, X_j\big) + Y^{-2} \sum_{j\in\mathcal{M}} X_j \cdot \mathrm{D}_j\, Y = \sum_{m=1,2} Y^{-m}\, X_{1,m},$$

(IV.15)

where the random variable $X_{1,1} = \sum_{a \in \mathcal{I}_{N+1}^{\mathcal{M}}} x_{1,1}^{\mathcal{M}}(a)\, \Phi_a$ is given by

$$X_{1,1} = \sum_{j \in \mathcal{M}} \sum_{a \in \mathcal{I}_N^{\mathcal{M}}} \sqrt{a_j + 1}\, x_j^{\mathcal{M}}(a)\, \Phi_{a+\varepsilon_j},$$

and the random variable $X_{1,2} = \sum_{a \in \mathcal{I}_{M+N-1}^{\mathcal{M}}} x_{1,2}^{\mathcal{M}}(a)\, \Phi_a$ is given by

$$X_{1,2} = \sum_{j \in \mathcal{M}} \sum_{a \in \mathcal{I}_N^{\mathcal{M}}} \sum_{b \in \mathcal{I}_M^{\mathcal{M}}} \sum_{c \in \mathcal{I}_N^{\mathcal{M}} : c \le a \wedge (b - \varepsilon_j)} x_j^{\mathcal{M}}(a)\, y^{\mathcal{M}}(b) \cdot \Bigg(\prod_{n \in \mathcal{M}} \frac{\sqrt{a_n! b_n! (a_n + b_n - 1_{n=j} - 2c_n)!}}{c_n! (a_n - c_n)! (b_n - 1_{n=j} - c_n)!} \Bigg) \Phi_{a+b-\varepsilon_j - 2c}.$$

The induction step follows via the identity $\Psi_{n+1}(v) = \langle v, \mathrm{D}\, \Psi_n(v) \rangle_{\mathcal{H}}$ and the calculations

$$\begin{aligned}\Psi_{n+1}(v) &= \sum_{m=1}^{n+1} Y^{-n-m} \sum_{j \in \mathcal{M}} X_j \cdot \mathrm{D}_j\, X_{n,m} - \sum_{m=1}^{n+1} Y^{-n-m-1} \underbrace{\Big(\sum_{j \in \mathcal{M}} X_j \cdot \mathrm{D}_j\, Y \Big)}_{-X_{1,2}} X_{n,m} \\ &= \sum_{m=1}^{n+2} Y^{1-(n+1)-m}\, X_{n+1,m},\end{aligned}$$

where $X_{n+1,1} = \sum_{j \in \mathcal{M}} X_j \cdot \mathrm{D}_j\, X_{n,1}$ and $X_{n+1,n+2} = X_{1,2}\, X_{n,n+1} = X_{1,2}^{n+1}$, and where

$$X_{n+1,m} = \sum_{j \in \mathcal{M}} X_j \cdot \mathrm{D}_j\, X_{n,m} + X_{1,2}\, X_{n,m-1}, \quad m = 2, \dots, n+1,$$

The coefficients $x_{n+1,m}^{\mathcal{M}}(a)$ in the expansions $X_{n+1,m} = \sum_{a \in \mathcal{I}_{\gamma_{n+1,m}}^{\mathcal{M}}} x_{n+1,m}^{\mathcal{M}}(a)\, \Phi_a$ are found by expanding the products via equation (IV.10b). □

The simulation of the iterated Skorohod integral $\delta_k(v)$ for a stochastic process v written on the form (IV.21) can be done via the following three steps.

1) Simulate the random variables $W(\xi_j)$, $j \in (\mathcal{N} \cup \mathbb{N}) \setminus \mathcal{M}$, and compute the coefficients $y^{\mathcal{M}}(a)$, $a \in \mathcal{I}_M^{\mathcal{M}}$, and $x_j^{\mathcal{M}}(a)$, $j \in \mathcal{M}$, $a \in \mathcal{I}_N^{\mathcal{M}}$.
2) Compute the coefficients $x_{n,m}^{\mathcal{M}}(a)$, $n = 1, \dots, k$, $m = 1, \dots, n+1$, $a \in \mathcal{I}_{\gamma_{n,m}}^{\mathcal{M}}$.
3) Simulate the random variables $W(\xi_j)$, $j \in \mathcal{M}$, and compute the random variables Y and $X_{n,m}$, $n = 1, \dots, k$, $m = 1, \dots, n+1$.

If these three steps are done, then the needed variables are available in order to compute $\delta_k(v)$ via corollary IV.11 and proposition IV.14. Moreover, if the stochastic component generated by the random variables $W(\xi_j)$, $j \in \mathcal{M}$, is believed to more important than remaining stochasticity, then it is possible to repeat the third step several times every time the two first steps have been performed.

(IV.16)

Remark Often the chaos expansion of the coefficients $y^{\mathcal{M}}(a)$ and $x_j^{\mathcal{M}}(a)$ will involve infinitely many of the random variables $W(\xi_j)$, $j \in (\mathcal{N} \cup \mathbb{N}) \setminus \mathcal{M}$, whence the first step is not practically implementable on a computer. This problem can be resolved by somehow truncating the chaos expansions. If this is the case, then bias is introduced and the proper name of the presented inference method should read *"Simulation of approximative pseudo-likelihoods"*.

In order to assess the complexity of the proposed simulation procedure we estimate the number of terms in the coefficients for $X_{n,m}$. If we let $\mathcal{M}_m = \{1, \dots, m\}$, then we see that $|\mathcal{I}_0^{\mathcal{M}_m}| = |\mathcal{I}_\gamma^{\mathcal{M}_1}| = 1$ and

$$|\mathcal{I}_\gamma^{\mathcal{M}_m}| = \sum_{j=0}^{\gamma} |\mathcal{I}_{\gamma-j}^{\mathcal{M}_{m-1}}| = |\mathcal{I}_{\gamma-1}^{\mathcal{M}_m}| + |\mathcal{I}_\gamma^{\mathcal{M}_{m-1}}|,$$

whence the addition property of the binomial coefficients yields

$$|\mathcal{I}_\gamma^{\mathcal{M}}| = \binom{|\mathcal{M}| + \gamma - 1}{\gamma}.$$

If we use the identity

$$\sum_{j=0}^{J} \binom{K + j - 1}{j} = \binom{K + J}{J},$$

then we see that $\sum_{n=1}^{k} \sum_{m=1}^{n+1} |\mathcal{I}_{\gamma_{n,m}}^{\mathcal{M}}|$ can be estimated by

$$\sum_{n=1}^{k} \sum_{m=1}^{n+1} \binom{|\mathcal{M}| + \gamma_{n,m} - 1}{\gamma_{n,m}} \approx \sum_{n=1}^{k} \sum_{m=0}^{n} \binom{|\mathcal{M}| + mM + nN - 1}{mM + nN}$$
$$\approx \sum_{n=1}^{k} \frac{1}{M} \binom{|\mathcal{M}| + n(M+N)}{n(M+N)} \approx \frac{1}{M(M+N)} \binom{|\mathcal{M}|\,(M+N+1) + 1}{|\mathcal{M}|\,(M+N)}.$$

Suppose for instance that $\mathcal{N} = \{-1, \dots, -d\}$, $\mathcal{M} = \{1, \dots, d\}$, that the random variables F_i can be written on the form

$$F_i = \sum_{a \in \mathcal{I}_\kappa^{\mathcal{M}}} f_i^{\mathcal{M}}(a)\, \Phi_a, \qquad f_i^{\mathcal{M}}(a) \text{ measurable w.r.t. } W(\xi_j),\, j \in \mathbb{N} \setminus \mathcal{M},$$

and that proposition IV.6 is used with $\tilde{u}_i = \xi_{-i}$. Then $M = \kappa d$, $N = \kappa(d-1)$, and the number of coefficients $x_{n,m}^{\mathcal{M}}(a)$ is estimated by

$$\frac{1}{\kappa d(2d-1)} \binom{\kappa d(2d-1) + d - 1}{\kappa d(2d-1)}.$$

This estimate grows at a super-exponential rate, and as a consequence the iterated Skorohod integral can only be simulated for d and κ very small using this approach.

(IV.17)

IV.4 Observations with measurement errors

As demonstrated in section IV.3 simulating iterated Skorohod integrals can be a very difficult task. In this section we will discuss two cases, where the needed computations are very easy. These situations are given by perturbing the random variables with independent Gaussian noise. In a statistical framework this corresponds to observations with measurement errors.

Let $F = (F_i)_{i\in\mathcal{N}}$ be a $\mathcal{N}$-dimensional random variable such that each F_i is measurable w.r.t. the σ-field generated by $W(\xi_j)$, $j \in \mathbb{N}$. Then we have the independent random variables $W(\xi_j)$, $j \in \mathcal{N}$, at our disposal to perturb F. We will consider the perturbation of $F = (F_i)_{i\in\mathcal{N}}$ in the position respectively the scale, *i.e.* for some fixed parameter $\tau > 0$ we define

$$F_{\text{pos},i}(\tau) = F_i + \tau W(\xi_i), \qquad F_{\text{scale},i}(\tau) = \big(1 + \tau W(\xi_i)\big)F_i.$$

If we use proposition IV.6 with $\tilde{u}_i = \xi_i$, then we find in the position perturbed case

$$\Gamma_{\text{pos}}(\tau) = \big(\langle \mathrm{D}\, F_{\text{pos},i}(\tau), \xi_j\rangle_{\mathcal{H}}\big)_{i,j\in\mathcal{N}} = \tau I_{\mathcal{N}}, \qquad u_{\text{pos},i}(\tau) = \tau^{-1}\xi_i,$$

and since $\delta((\xi_i)_{i\in\mathcal{N}}) = \prod_{i\in\mathcal{N}} W(\xi_i)$ the Lebesgue density $\rho_{\text{pos}}(y;\tau)$ of random variable $F_{\text{pos}}(\tau)$ is given by

$$\rho_{\text{pos}}(y;\tau) = \tau^{-|\mathcal{N}|}\,\mathrm{E}_{\mathrm{P}}\Big[\prod_{i\in\mathcal{N}} 1_{(F_i+\tau W(\xi_i)>y_i)} \prod_{i\in\mathcal{N}} W(\xi_i)\Big]. \tag{IV.23}$$

In the scale perturbed case we find that

$$\Gamma_{\text{scale}}(\tau) = \operatorname{diag}\big((\tau F_i)_{i\in\mathcal{N}}\big), \qquad u_{\text{scale},i}(\tau) = \tau^{-1}F_i^{-1}\xi_i,$$

and since $\delta((F_i^{-1}\xi_i)_{i\in\mathcal{N}}) = \prod_{i\in\mathcal{N}} F_i^{-1}W(\xi_i)$ the Lebesgue density $\rho_{\text{scale}}(y;\tau)$ of the random variable $F_{\text{scale}}(\tau)$ is given by

$$\rho_{\text{scale}}(y;\tau) = \tau^{-|\mathcal{N}|}\,\mathrm{E}_{\mathrm{P}}\Big[\prod_{i\in\mathcal{N}} 1_{((1+\tau W(\xi_i))F_i>y_i)} \prod_{i\in\mathcal{N}} F_i^{-1}W(\xi_i)\Big]. \tag{IV.24}$$

The formulae (IV.23) and (IV.24) can be proven without appealing to Malliavin calculus, *e.g.* in the one-dimensional position perturbed case equation (IV.23) follows by the calculations

$$\begin{aligned}
\rho_{\text{pos}}(y;\tau) &= \partial_y\,\mathrm{E}_{\mathrm{P}}\big[1_{(F+\tau W(\xi)\le y)}\big] \\
&= \partial_y\,\mathrm{E}_{\mathrm{P}}\Big[\int_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} 1_{(-\infty,\tau^{-1}(y-F)]}(x)\,\mathrm{d}x\Big] \\
&= \partial_y\,\mathrm{E}_{\mathrm{P}}\Big[\int_{-\infty}^{\infty} x\tfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \min\big\{x, \tau^{-1}(y-F)\big\}\,\mathrm{d}x\Big] \\
&= \mathrm{E}_{\mathrm{P}}\Big[\int_{-\infty}^{\infty} \tau^{-1}x\tfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} 1_{(\tau^{-1}(y-F),\infty)}(x)\,\mathrm{d}x\Big] \\
&= \tau^{-1}\,\mathrm{E}_{\mathrm{P}}\big[1_{(F+\tau W(\xi)>y)}\,W(\xi)\big],
\end{aligned}$$

(IV.18)

where the third step uses the integration by parts formula from classical analysis. This should be compared to the formula avoiding the integration by parts step, *i.e.*

$$\begin{aligned}\rho_{\text{pos}}(y;\tau) &= \partial_y \operatorname{E}_{\mathrm{P}}\left[1_{(F+\tau W(\xi)\leq y)}\right] = \partial_y \operatorname{E}_{\mathrm{P}}\left[\int_{-\infty}^{\tau^{-1}(y-F)} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x\right] \\ &= \operatorname{E}_{\mathrm{P}}\left[\tfrac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(y-F)^2}\right].\end{aligned} \tag{IV.25}$$

Suppose we want to simulate the density $\rho_{\text{pos}}(y;\tau)$ by replacing the mean value in (IV.23) or (IV.25) by the average over i.id. samples of the involved random variables. In formula (IV.25) we would then take average over positive numbers, whereas in formula (IV.23) we would take average over both positive and negative numbers. The formulae derived by using the integration by parts formula, see theorem IV.4, are thus badly suited for use in simulations schemes since the procedure of taking average over large positive and negative numbers is numerically instable.

IV.5 Conclusion

The idea of calculating multivariate Lebesgue densities by simulating the mean values in the formulae derived via the integration by parts setting from Malliavin calculus is natural, but the method have two major drawbacks. Firstly, the formulae involve iterated Skorohod integrals of order equal to the dimension m of the density under consideration, and the number of terms in such an integral grows super-exponentially in m as demonstrated in section IV.3. It is thus only possible to simulate a single iterated Skorohod integral for rather small m, *i.e.* m less than five. However, for spatially homogeneous models each such simulation can be used for every observation point in marginal pseudo-likelihoods on the form (IV.5). The second and more gross drawback is the numerical instability demonstrated in section IV.4. By construction a Skorohod integral has mean equal to zero, whence the in m super-exponentially many terms of the iterated Skorohod integrals take both positive and negative values. Moreover, since there is no reason to hope that most of the terms take small absolute values, the iterated Skorohod integrals will presumably have very large variance. The simulation approach to compute the mean value of the iterated Skorohod integrals over certain events is thus bound to be numerically instable. This numerical instability is tied up with the integration by parts formula and can be avoided by avoiding the use of this formula. The immediate way to do this is to add observation noise as illustrated in equation (IV.25). Moreover, the use of measurement noise is physical meaningful, and this approach also allows any method of simulating the stochastic process to be used, *e.g.* Euler schemes and so forth, and there is no need for the chaos decomposition to be known.

Bibliography

Aerts, M. & Claeskens, G. (1999), 'Bootstrapping pseudolikelihood models for clustered binary data', *Annals of the Institute of Statistical Mathematics* **51**(3), 515–530.

Aronszajn, N. (1950), 'Theory of reproducing kernels', *Transactions American Mathematical Society* **68**, 337–404.

Avram, F. (1988), 'On bilinear forms in Gaussian random variables and Toeplitz matrices', *Probability Theory and Related Fields* **79**, 37–45.

Benth, F. E. (1997), 'On the positivity of the stochastic heat equation', *Potential Analysis* **6**, 127–148.

Benth, F. E. & Potthoff, J. (1996), 'On the martingale property for generalized stochastic processes', *Stochastics and Stochastics Reports* **58**, 349–367.

Bhatia, R. (1997), *Matrix Analysis*, Springer-Verlag.

Brace, A., Gątarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7**(2), 127–155.

Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, second edn, Springer-Verlag.

Chihara, T. S. (1978), *An Introduction to Orthogonal Polynomials*, Gordon & Breach.

Cont, R. (1998), Modeling term structure dynamics: an infinite dimensional approach. To appear in: International Journal of Theoretical and Applied Finace.

Da Prato, G. & Zabczyk, J. (1992), *Stochastic Equations in Infinite Dimensions*, Cambirdge University Press.

Dahlhaus, R. (1989), 'Efficient parameter estimation for self-similar processes', *Annals of Statistics* **17**, 1749–1766.

Denis, L. (2000), 'A criterion of density for solutions of Poisson-driven SDEs', *Probability Theory and Related Fields* **118**, 406–426.

Dermoune, A. (1995), 'Differential calculus relative to some point processes', *Statistics & Probability Letters* **24**, 233–242.

Dunsmuir, W. & Hannan, E. J. (1976), 'Vector linear time-series models', *Advances in Applied Probability* **8**, 339–364.

Hannan, E. J. (1970), *Multiple Time Series*, Wiley.

Hardy, G. H. & Ramanujan, S. (1918), 'Asymptotic formulae in combinatory analysis', *Proceedings of the London Mathematical Society* **17**(2), 75–115.

Hida, T., Kuo, H. H., Potthoff, J. & Streit, L. (1993), *White Noise - An Infinite Dimensional Calculus*, Kluwer Academic Publishers.

Holden, H., Øksendal, B., Ubøe, J. & Zhang, T. (1996), *Stochastic Partial Differential Equations*, Birkhäuser.

Huebner, M. (1997), 'A characterization of asymptotic behaviour of maximum likelihood estimators for stochastic PDEs', *Mathematical Methods of Statistics* **6**(4), 395–415.

Huebner, M. (1999), 'Asymptotic properties of the maximum likelihood estimator for stochastic PDEs disturbed by small noise', *Statistical Inference for Stochastic Processes* **2**, 57–68.

Huebner, M. & Lototsky, S. V. (2000), 'Asymptotic analysis of the sieve estimator for a class of parabolic SPDEs', *Scandinavian Journal of Statistics* **27**, 353–370.

Huebner, M. & Rozovskii, B. L. (1995), 'On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDEs', *Probability Theory and Related Fields* **103**, 143–163.

Ibragimov, I. A. & Has'minskii, R. Z. (1981), *Statistical Inference. Asymptotic Theory*, Springer-Verlag. Originally published in russian in 1979.

Ibragimov, I. A. & Has'minskii, R. Z. (1998), 'Estimation problems for coefficients of stochastic partial differential equations. Part I', *Theory of Probability and its Applications* **43**(3), 370–387.

Ibragimov, I. A. & Has'minskii, R. Z. (1999), 'Estimation problems for coefficients of stochastic partial differential equations. Part II', *Theory of Probability and its Applications* **44**(3), 469–494.

Ibragimov, I. A. & Has'minskii, R. Z. (2000), 'Estimation problems for coefficients of stochastic partial differential equations. Part III', *Theory of Probability and its Applications* **45**(2), 210–232.

Itô, K. (1951), 'Multiple Wiener integral', *Journal Mathematical Society Japan* **3**, 157–169.

Itô, Y. (1988), 'Generalized poisson functionals', *Probability Theory and Related Fields* **77**, 1–28.

Itô, Y. & Kubo, I. (1988), 'Calculus on Gaussian and Poisson white noise', *Nagoya Mathematical Journal* **111**, 41–84.

Kallianpur, G. & Xiong, J. (1995), *Stochastic Differential Equations in Infinite Dimensional Spaces*, Vol. 26 of *Lecture Notes - Monograph Series*, Institute of Mathematical Statistics.

Kondratiev, Y. G., Silva, J. L. & Streit, L. (1996), Generalized Appell systems, Preprint, Research center Bielefeld-Bonn-Stochastics. This preprint is available from http://www.physik.uni-bielefeld.de/bibos/1996.html.

Kotelenez, P. (1992), 'Existence, uniqueness and smoothness for a class of function valued stochastic partial differential equations', *Stochastics and Stochastics Reports* **41**(3), 177–199.

Le Cam, L. & Yang, G. L. (2000), *Asymptotics in Statistics*, second edn, Springer-Verlag.

Ledoux, M. & Talagrand, M. (1991), *Probability in Banach Spaces*, Springer-Verlag.

Loges, W. (1984), 'Girsanov's theorem in Hilbert space and an application to the statistics of Hilbert space-valued stochastic differential equations', *Stochastic Processes and their Applications* **17**, 243–263.

Lototsky, S. V. & Rozovskii, B. L. (1999), 'Spectral asymptotics of some functionals arising in statistical inference for SPDEs', *Stochastic Processes and their Applications* **79**, 69–94.

Lytvynov, E. W., Rebenko, A. L. & Shchepan'uk, G. V. (1997), 'Wick calculus on spaces of generalized functions of compound Poisson white noise', *Reports on Mathematical Physics* **39**(2), 219–248.

Manthey, R. & Mittmann, K. (1998), 'On a class of stochastic functional-differential equations arising in population dynamics', *Stochastics and Stochastics Reports* **64**, 75–115.

Manthey, R. & Mittmann, K. (1999), 'On the qualitative behaviour of the solution to a stochastic partial functional-differential equation arising in population dynamics', *Stochastics and Stochactics Reports* **66**, 153–166.

Markussen, B. (2001*a*), 'Chaos decomposition and stochastic calculus for the negative binomial process', Enclosed in this thesis.

Markussen, B. (2001*b*), 'Likelihood inference for a stochastic partial differential equation observed in discrete points in time and space', Enclosed in this thesis.

Markussen, B. (2001*c*), 'Simulation of pseudo-likelihoods given discretely observed data', Enclosed in this thesis.

Markussen, B. (2001*d*), 'Uniform convergence of approximate likelihood functions for a stationary multivariate Gaussian time series', Enclosed in this thesis.

Mohapl, J. (1998), 'Discrete sample estimation for Gaussian random fields generated by stochastic partial differential equations', *Communications in Statistics. Stochastic Models* **14**(4), 883–903.

Nualart, D. (1995), *The Malliavin Calculus and Related Topics*, Springer-Verlag.

Piterbarg, L. & Rozovskii, B. L. (1997), 'On asymptotic problems of parameter estimation in stochastic PDEs: Discrete time sampling', *Mathematical Methods of Statistics* **6**(2), 200–223.

Privault, N. (1994), 'Chaotic and variational calculus in discrete and continuous time for the Poisson process', *Stochastics and Stochastics Reports* **51**, 83–109.

Privault, N. (1996), 'A different quantum stochastic calculus for the Poisson process', *Probabability Theory and Related Fields* **105**, 255–278.

Rosenblatt, M. (1985), *Stationary Sequences and Random Fields*, Birkhäuser.

Rozovskii, B. L. (1990), *Stochastic Evolution Systems*, Kluwer Academic Publishers. Originally published in russian in 1983.

Santa-Clara, P. & Sornette, D. (1999), The dynamics of the forward interest rate curve with stochastic string shocks, Preprint, University of California.

Sato, K. I. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambrindge University Press.

Self, S. G. & Liang, K. Y. (1987), 'Asymptotic properties of maximum likelihood estimators and likelihood ratio test under nonstandard conditions', *Journal of the American Statistical Association* **82**, 605–610.

Shaman, P. (1976), 'Approximations for stationary covariance matrices and their inverses with applications to ARMA models', *Annals of Statistics* **4**, 292–301.

Sørensen, M. & Huebner, M., eds (2001), *Workshop on Stochastic Partial Differential Equations*, number 20 *in* 'Miscellancea', Center for Mathematical Physics and Stochastics.

Theting, T. G. (2000), 'Solving Wick-stochastic boundary value problems using a finite element method', *Stochastics and Stochastics Reports* **70**, 241–270.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag.

Walsh, J. B. (1986), An introduction to stochastic partial differential equations, *in* R. Carmona, H. Kesten & J. B. Walsh, eds, 'École d'Été de Probabilités de Saint-Flour XIV-1984', Vol. 1180 of *Springer LNM*, pp. 265–437.

Watanabe, S. (1987), 'Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels', *The Annals of Probability* **15**(1), 1–39.

Wiener, N. (1938), 'The homogeneous chaos', *American Journal of Mathematics* **60**, 879–936.