Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Mathematische Statistik

Stochastic modeling in space and time with Lévy-driven random fields

Viet Son Pham

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r):		Prof. Dr. Daniel Matthes
Prüfer der Dissertation:	1.	Prof. Dr. Claudia Klüppelberg
	2.	Prof. Dr. Richard A. Davis (Columbia University)
	3.	Prof. Dr. Alexander Lindner (Universität Ulm)

Die Dissertation wurde am 17.04.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 25.06.2019 angenommen.

Summary

This dissertation investigates several aspects of stochastic modeling in space and time with Lévy-driven random fields and is organized as follows.

In Chapter 1 we propose a novel class of tempo-spatial Ornstein-Uhlenbeck processes as solutions to Lévy-driven Volterra equations with additive noise and multiplicative drift. Subsequent to formulating conditions for the existence and uniqueness of solutions, we derive an explicit solution formula and study distributional properties such as stationarity, second-order structure and short versus long memory. Furthermore, we analyze in detail the path properties of the solution process. In particular, we introduce different notions of càdlàg paths in space and time and establish conditions for the existence of versions with these regularity properties. The theoretical results are accompanied by illustrative examples.

In Chapter 2 we introduce Lévy-driven causal CARMA random fields on $\mathbb{R}^d$, extending the class of CARMA processes. The definition is based on a system of stochastic partial differential equations which generalizes the classical state-space representation of CARMA processes. The resulting CARMA model differs fundamentally from the CARMA random field of Brockwell and Matsuda. We show existence of the model under mild assumptions and examine some of its features including the second-order structure and path properties. In particular, we investigate the sampling behavior and formulate conditions for the causal CARMA random field to be an ARMA random field when sampled on an equidistant lattice. Moreover, we link the CARMA model to the processes studied in Chapter 1.

In Chapter 3 we estimate model parameters of Lévy-driven causal CARMA random fields by fitting the empirical variogram to the theoretical counterpart using a weighted least squares (WLS) approach. After deriving asymptotic results for the variogram estimator, we show strong consistency and asymptotic normality of the parameter estimator. Furthermore, we conduct a simulation study to assess the quality of the WLS estimator for finite samples. For the simulation we utilize numerical

approximation schemes based on truncation and discretization of stochastic integrals and we analyze the associated simulation errors in detail. Finally, we apply our results to real data of the cosmic microwave background.

In Chapter 4 we study the autocovariance functions of moving average random fields over the integer lattice $\mathbb{Z}^d$ from an algebraic perspective. These autocovariances are parametrized polynomially by the moving average coefficients, hence tracing out algebraic varieties. We derive dimension and degree of these varieties and we use their algebraic properties to obtain statistical consequences such as identifiability of model parameters. We connect the problem of parameter estimation to the algebraic invariants known as euclidean distance degree and maximum likelihood degree. In our simulation study we use tools from numerical algebraic geometry and compare them with classical local search methods for maximum likelihood estimation.

Zusammenfassung

Diese Dissertation untersucht einige Aspekte stochastischer Modellierung in Raum und Zeit mit Lévy-getriebenen Zufallsfeldern und ist folgendermaßen strukturiert.

In Kapitel 1 führen wir eine neuartige Klasse räumlich-zeitlicher Ornstein-Uhlenbeck-Prozesse als Lösungen von Lévy-getriebenen Volterra-Gleichungen mit additiven Rauschen und multiplikativen Drift ein. Nachdem wir Bedingungen für die Existenz und Eindeutigkeit von Lösungen formulieren, leiten wir eine explizite Lösungsformel her und studieren Verteilungseigenschaften wie Stationarität, die Struktur der Autokovarianz sowie die Gegenüberstellung von Kurzzeit- und Langzeitkorrelation. Außerdem analysieren wir im Detail die Pfadeigenschaften des Lösungsprozesses. Insbesondere formulieren wir verschiedene Càdlàg-Begriffe in Raum und Zeit und leiten Voraussetzungen für die Existenz von Versionen mit diesen Regularitätseigenschaften her. Die theoretischen Resultate werden durch illustrative Beispiele begleitet.

In Kapitel 2 stellen wir Lévy-getriebene kausale CARMA-Zufallsfelder auf $\mathbb{R}^d$ vor, welche die Klasse der CARMA-Prozesse erweitert. Die Definition basiert auf einem System von stochastischen partiellen Differentialgleichungen, welches die klassische Zustandsraumdarstellung von CARMA-Prozessen verallgemeinert. Das resultierende CARMA-Modell unterscheidet sich fundamental von dem CARMA-Modell von Brockwell und Matsuda. Wir zeigen die Existenz des Modells unter milden Voraussetzungen und ermitteln einige seiner Merkmale inklusive der Struktur der Autokovarianz und Pfadeigenschaften. Im Speziellen formulieren wir Bedingungen dafür, dass ein auf einem gleichmäßigen Gitter beobachtetes kausales CARMA-Zufallsfeld ein ARMA-Zufallsfeld ist. Weiterhin stellen wir die Verbindung zu den in Kapitel 1 studierten Prozessen her.

In Kapitel 3 schätzen wir die Modellparameter von Lévy-getriebenen kausalen CARMA-Zufallsfeldern indem wir das empirische Variogramm unter Verwendung einer gewichteten Methode kleinster Quadrate an das theoretische Variogramm an-

passen. Nachdem wir asymptotische Eigenschaften für den Variogrammschätzer herleiten, zeigen wir starke Konsistenz und asymptotische Normalität des Parameterschätzers. Außerdem, führen wir eine Simulationsstudie durch, um die Qualität des Parameterschätzers bei endlichen Stichproben zu überprüfen. Für die Simulation verwenden wir numerische Approximationsschemata, welche auf dem Abschneiden und Diskretisieren von stochastischen Integralen basieren, und analysieren deren Simulationsfehler im Detail. Zum Abschluss wenden wir unsere Resultate auf reale Daten der kosmischen Hintergrundstrahlung an.

In Kapitel 4 studieren wir die Autokovarianzfunktionen von Moving-Average-Zufallsfeldern auf $\mathbb{Z}^d$ von einem algebraischen Gesichtspunkt. Diese Autokovarianzen sind polynomiell durch die Modellparameter parametrisiert und erzeugen daher algebraische Varietäten. Wir leiten Dimension und Grad dieser Varietäten her und benutzen deren algebraischen Eigenschaften um statistische Konsequenzen zu erhalten, wie beispielsweise die Identifizierbarkeit von Modellparametern. Wir stellen eine Verbindung zwischen dem Problem der Parameterschätzung und Größen aus der algebraischen Statistik, wie euklidischer Abstandsgrad oder Maximum-Likelihood-Grad, her. In einer Simulationsstudie vergleichen wir Werkzeuge aus der numerischen algebraischen Geometrie mit klassischen lokalen Suchmethoden für die Maximum-Likelihood-Methode.

Acknowledgments

First of all, I would like to express my gratitude to my supervisor Claudia Klüppelberg. She gave me the opportunity to pursue my goal in an excellent environment and working with her has been a pleasure. I am very much indebted for countless fruitful discussions, for encouraging me to present my research to the scientific community and for continuous guidance over the last few years.

I am also immensely grateful to my co-author and mentor Carsten Chong, who gave me valuable advice, especially during the first year of my PhD studies. He introduced me to my field of study and I learned a lot from him about conducting research.

Cordial thanks also go to David Berger, Jean Jacod, Alexander Lindner and Robert Stelzer for inspiring discussions and helpful comments which led to improvements of my thesis.

Furthermore, many thanks go to my colleagues at the Chair for Mathematical Statistics for making the time at the office very pleasant. In particular, I won't forget the daily snack times with Thiago and the board game nights organized by my co-author Carlos.

I further thank the graduate program TopMath at the Technical University of Munich and the Studienstiftung des deutschen Volkes for financial support.

Finally, I want to thank my family and friends for their patience and encouragement. Most of all, I am greatly thankful for having wonderful parents and a lovely sister, who have always supported me.

Introduction

A set of data recorded repeatedly over time is called a time series. These kind of data sets appear in virtually every area of modern science, from Archaeology to Zoology. Consequently, time series analysis is an important field of research in probability theory and statistics. Theoretical studies in this field started in the beginning of the last century and nowadays its literature is comprehensive, see for instance the classical monographs Box and Jenkins [18], Brockwell and Davis [23] and Hamilton [51], or Shumway and Stoffer [81] for a more recent approach. Parallel to the advancements in time series analysis, Itô developed the theory of stochastic integration with respect to Brownian motion in his seminal work [55]. This theory opened the door to a whole new class of continuous-time models based on stochastic integral processes or stochastic differential equations and soon after, Brownian motion as the integrator was replaced by Lévy processes with jumps [6] and more general semimartingales [56]. Continuous-time processes are particularly useful for irregularly spaced or high-frequency data. In many cases however, one has to deal with data which is observed in both time and space, such as in meteorology, geology, image processing or social media. Analyzing and handling this type of data requires novel statistical models since purely temporal models are not sufficient. Unsurprisingly, there has been a growing interest in the field of tempo-spatial statistical modeling in recent years. New phenomena and hurdles emerge with the additional spatial structure and the challenge for statisticians lies in providing stochastic models which are both flexible and analytically tractable for applications.

The goal of the present thesis is to introduce two different stochastic models in space and time which satisfy the aforementioned criteria. We structure the thesis as follows: the starting point in Chapter 1 is the Lévy-driven Ornstein-Uhlenbeck (OU) process, which has applications in molecular physics [88] and stochastic volatility modeling [10]. We extend the OU process with the objective of modeling phenomena in space and time and indeed, such extensions already exist in the literature

(see [9, 47]). However, instead of modifying the process itself, in our novel approach we extend the stochastic differential equation defining the OU process and arrive at

$$X(t,x) = X(0,x) - \int_0^t \int_{\mathbb{R}^d} X(t-s,x-y)\,\mu(\mathrm{d}s,\mathrm{d}y) + \int_0^t \int_{\mathbb{R}^d} g(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y), \tag{1}$$

where Λ is a homogeneous Lévy basis, g is a function such that the integral in (1) makes sense and μ is a signed Borel measure. Equation (1) is a stochastic Volterra integral equation and we call any process solving this equation a *Volterra-type Ornstein-Uhlenbeck (VOU) process*. Employing the theory of deterministic convolutional Volterra equations [49], we derive in Theorem 1.3.1 of this chapter sufficient conditions on Λ, g and μ that imply the existence and uniqueness of solutions to (1). One advantage of the VOU model is that under additional mild conditions, we are able to find an explicit solution formula, namely

$$X(t,x) = \int_{-\infty}^t \int_{\mathbb{R}^d} (g - \rho * g)(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \tag{2}$$

which is a stochastic convolution integral, where ρ is the so-called resolvent of μ. Apart from studying solution criteria, we investigate several distributional properties such as strict stationarity or second-order structure. We find that the model is flexible enough to incorporate both tempo-spatial short-range and long-range dependency (see Proposition 1.4.5). Last but not least, we analyze the path properties of VOU processes. As in the classical one-dimensional case, Hölder-continuous sample paths can be established if the noise Λ is Gaussian (see Theorem 1.5.1). If on the other hand the noise exhibits jumps, path properties can heavily vary depending on the function g. We define two new notions of tempo-spatial càdlàg properties and give sufficient conditions for them to hold in Theorems 1.5.3 and 1.5.5.

The VOU process in Equation (2) depends on the infinite-dimensional objects g and ρ, which makes model estimation a difficult task. By contrast, the second model in this thesis, which we introduce in Chapter 2, only depends on finitely many parameters and is therefore more suited for statistical inference. In a similar fashion as in Chapter 1, we extend the state-space representation which defines the well-known class of continuous-time autoregressive moving average (CARMA) processes and obtain the system of stochastic partial differential equations

$$\begin{aligned} (I_p\partial_d - A_d)\cdots(I_p\partial_1 - A_1)Z(t) &= c\dot{L}(t), \quad t \in \mathbb{R}^d, \\ Y(t) &= b^\top Z(t), \quad t \in \mathbb{R}^d, \end{aligned} \tag{3}$$

where ∂_i denotes partial differentiation with respect to the i'th coordinate, I_p is the identity matrix in $\mathbb{R}^{p\times p}$, A_i are companion matrices and $\dot{L}$ is Lévy white noise. We show in Theorem 2.3.5 that the random field

$$Y(t) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} b^\top \mathrm{e}^{A_1(t_1-s_1)} \cdots \mathrm{e}^{A_d(t_d-s_d)} c \, \mathrm{d}L(s), \quad t = (t_1, ..., t_d) \in \mathbb{R}^d,$$

is a mild solution to (3) and call Y a *causal CARMA random field*. Due to its similar structure, the random field Y possesses many of the commonly known features of CARMA processes. This includes for instance exponentially decaying autocovariance functions, rational spectral densities and the property that equidistant sampling results in an ARMA random field. Causal CARMA random fields generate a versatile family of covariance functions which are in general anisotropic and non-separable. We also discuss path properties of Y and prove that it has càdlàg sample paths, where the càdlàg notion differs from those in Chapter 1 (see Theorem 1.5). In particular, we show that causal CARMA random fields indeed constitute a parametric submodel of VOU processes and establish the link to Chapter 1.

While Chapter 2 considers stochastic properties of causal CARMA random fields, we tackle the problem of parameter estimation in Chapter 3. We use the variogram, which is a popular dependence measure in spatial statistics, as our main tool and estimate the parameters of a causal CARMA random field in two steps. First, we compute non-parametrically an empirical version of the variogram at several lags assuming that observations of the random field are given on a regular lattice. Second, we use a weighted least squares (WLS) method in order to fit the empirical variogram to the theoretical one. Subsequent to establishing asymptotic properties of the variogram estimator, we show strong consistency and asymptotic normality of the WLS estimator under suitable identifiability conditions in Theorem 3.4.1. For our simulation study in Section 3.6 we utilize simulation algorithms which are based on truncation and discretization of stochastic integrals. Another contribution of this chapter is that we also compute the mean-squared errors for these algorithms in Section 3.5. This chapter ends with an application to cosmic microwave data.

In contrast to the previous chapters, the final Chapter 4 contains a more algebraic perspective. There, the main objects of study are the autocovariance functions of discrete-parameter moving average random fields. This type of random fields have only finitely many non-zero autocovariance values which depend polynomially on the moving average parameters, giving rise to algebraic varieties, which haven't been studied in the literature before. We call these *autocovariance varieties* and derive

their dimension and degree in Theorem 4.3.9. Moreover, we exploit their algebraic properties and study their consequences for statistical inference. For instance, we observe that it is easier to identify model parameters in higher dimensions than in one dimension (see Section 4.4). Furthermore, we find in a simulation study that maximum likelihood estimation can be facilitated by using numerical algebraic geometry methods instead of classical local search methods, since the latter may converge to a local optimum (see Section 4.5).

We remark at the end of this introduction that the following chapters are all self-contained. Necessary literature review and definitions will be given in each chapter separately. In addition, notational conventions are given in each introductory section and might differ slightly among chapters. Each chapter is based on a publication or submitted preprint:

- Chapter 1 is based on the paper [74] that is published as:

  V.S. Pham and C. Chong. Volterra-type Ornstein–Uhlenbeck processes in space and time. *Stoch. Process. Appl.*, 128(9):3082–3117, 2018.

- Chapter 2 is based on the paper [73] that is submitted for publication as:

  V.S. Pham. Lévy-driven causal CARMA random fields. 2018. Submitted. arXiv: 1805.08807.

- Chapter 3 is based on the paper [60] that is submitted for publication as:

  C. Klüppelberg and V.S. Pham. Estimation of causal CARMA random fields. 2019. Submitted. arXiv:1902.04962.

- Chapter 4 is based on the paper [2] that is submitted for publication as:

  C. Améndola and V.S. Pham. Autocovariance varieties of moving average random fields. 2019. Submitted. arXiv:1903.08611.

Contents

Chapter 1:
Volterra-type Ornstein-Uhlenbeck processes in space and time

1.1 Introduction

In 1908 Langevin [65] introduced the following equation as a model for the movement of a particle in a surrounding medium:

$$m\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\lambda v(t) + \dot{W}(t). \tag{1.1.1}$$

Here $v(t)$ denotes the velocity of the considered particle at time t, m its mass, λ a friction parameter that accounts for the friction forces acting on the particle, and $\dot{W}(t)$ the formal derivative of a Brownian motion that governs the random movement of the particle. The solution to the *Langevin equation* (1.1.1) is nowadays called *Ornstein-Uhlenbeck (OU) process*, named after the 1930 paper [88]. Employing Itô's calculus, it is well known that the stochastic differential equation (1.1.1) has a unique solution given by

$$v(t) = \mathrm{e}^{-\lambda t}v(0) + \int_0^t \mathrm{e}^{-\lambda(t-s)}\,\mathrm{d}W(s), \quad t \geq 0, \tag{1.1.2}$$

where $v(0)$ is the initial velocity. Meanwhile, OU processes have found many applications beyond molecular physics, so for example, in stochastic volatility modeling [10]. Moreover, the relatively simple model (1.1.2) has been extended in several directions, for instance, to supOU processes [8], generalized OU processes [13] or CARMA processes [20]. For all these models, the noise W in (1.1.2) no longer needs to be Gaussian, but can also be a Lévy process with jumps.

The goal of this chapter is to extend the class of OU processes to space and time and obtain a tempo-spatial statistical model that is both flexible for modeling purposes and analytically tractable. In fact, extensions of OU processes to several parameters already exist: see [9, 25, 47, 69] and the first chapter of [90] for various approaches. They all have in common that they start from (1.1.2) and generalize this formula to several parameters. In this way the main stochastic properties of the one-parameter OU process are preserved because the structure of the exponential kernel is kept. However, in these multi-parameter generalizations, the relation to the original differential equation (1.1.1) is no longer clear. And this is exactly the starting point of our present work: we consider tempo-spatial extensions of OU processes based on the differential equation (1.1.1). As we shall see, we recover some of the aforementioned extensions as particular cases of our approach.

The rationale behind our approach is based on two properties of Equation (1.1.1) which are characteristic to the Langevin equation and which we want to maintain in our generalization. First, the noise W is *additive*, that is, its effect on the process v is independent of the latter. Second, the drift term of v is a scalar *multiple* of its current value. Based on these observations, we propose the following model as a first step towards a tempo-spatial version of (1.1.1):

$$\mathrm{d}X(t,x) = -\lambda X(t,x)\,\mathrm{d}t + \int_{\mathbb{R}^d} g(x-y)\,\Lambda(\mathrm{d}t,\mathrm{d}y), \quad t \geq 0,\ x \in \mathbb{R}^d, \tag{1.1.3}$$

subjected to some initial condition $X(0,x)$ for $x \in \mathbb{R}^d$. Here Λ is a *homogeneous Lévy basis* on $[0,\infty)\times\mathbb{R}^d$ (also called an infinitely divisible independently scattered random measure or a Lévy sheet in the literature), which can be thought of as a multi-parameter analogue of a Lévy process. The function g is such that the integral in (1.1.3) makes sense, and the differential operation d on the left-hand side of (1.1.3) is taken with respect to time t. It is immediate to see that for each fixed x, the process $t \mapsto \int_{\mathbb{R}^d} g(x-y)\,\Lambda([0,t],\mathrm{d}y)$ is a Lévy process and, as a result, the process $t \mapsto X(t,x)$ an OU process in time. Therefore, (1.1.3) defines a system $(X(\cdot,x))_{x\in\mathbb{R}^d}$ of dependent OU processes. Furthermore, if g is strictly positive, every jump of Λ induces simultaneous jumps of all OU processes $X(\cdot,x)$, so innovations of Λ propagate with an infinite speed through the system. In order to include the case of finite propagation speed, we modify (1.1.3) by allowing g to depend on time as well:

$$X(t,x) = X(0,x) - \lambda\int_0^t X(s,x)\,\mathrm{d}s + \int_0^t\int_{\mathbb{R}^d} g(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y), \quad t \geq 0,\ x \in \mathbb{R}^d. \tag{1.1.4}$$

The integral form here is preferable because the kernel is time-dependent. Even more, we can allow the drift term to depend on different neighboring sites, if required, with time delay:

$$X(t,x) = X(0,x) - \int_0^t \int_{\mathbb{R}^d} \mu(t-s,x-y)X(s,y)\,\mathrm{d}(s,y) + \int_0^t \int_{\mathbb{R}^d} g(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y), \tag{1.1.5}$$

where μ is another kernel function. Unfortunately, model (1.1.5) no longer contains (1.1.4) as a special case, but we can remedy this problem by taking a measure μ instead of a function. So the final class of processes we consider in this chapter is

$$X(t,x) = X(0,x) - \int_0^t \int_{\mathbb{R}^d} X(t-s,x-y)\,\mu(\mathrm{d}s,\mathrm{d}y) + \int_0^t \int_{\mathbb{R}^d} g(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y). \tag{1.1.6}$$

If μ has a density with respect to the Lebesgue measure on $\mathbb{R}_+ \times \mathbb{R}^d$, the last model reduces to (1.1.5); if $\mu = -\lambda \mathrm{Leb}_{\mathbb{R}_+} \otimes \delta_{0,\mathbb{R}^d}$ (where $\mathrm{Leb}_{\mathbb{R}_+}$ is the Lebesgue measure on $\mathbb{R}_+$ and $\delta_{0,\mathbb{R}^d}$ the Dirac measure in the origin of $\mathbb{R}^d$), then it reduces to (1.1.4). Henceforth, we refer to (1.1.6) as the *Volterra-type Ornstein-Uhlenbeck (VOU) equation* and any solution to it as a *VOU process*. We call μ the *drift measure* and g the *noise propagation function*.

VOU processes have connections to other classes of space–time processes. They are submodels of the general class of *ambit fields* (we refer to the recent survey [12]), which have applications in several areas such as turbulence, finance and biological growth modeling. Furthermore, VOU processes are, as we shall show, generalizations of the $\mathrm{OU}_\wedge$ process considered in [9, 69], which has been applied to data of radiation anomalies in the latter reference. In contrast to the $\mathrm{OU}_\wedge$ process, the autocorrelation of a VOU process may exhibit long-range dependence and jumps may occur simultaneously for all space locations. In addition to that, VOU processes are solutions to *stochastic Volterra equations* [31, 32] with multiplicative drift and additive noise. The VOU model that we investigate in Example 1.3.7 below is a generalization of the stochastic wave equation in dimension 1, see [37].

The remaining chapter is devoted to a probabilistic analysis of the VOU model and is organized as follows. In Section 1.2 we recall the necessary background on Lévy bases and present some results on deterministic Volterra equations that we need throughout the chapter. The main theorems of Section 1.3, Theorems 1.3.1 and 1.3.3, give sufficient conditions on Λ, g and μ to ensure the existence and uniqueness of solutions to the VOU equation (1.1.6) and their convergence in distribution as time tends to infinity, respectively. The solution to (1.1.6) can be expressed as an

explicit stochastic convolution integral, revealing the interplay between the theory of deterministic Volterra equations and stochastic convolutions. Section 1.3 concludes with a detailed investigation of two Examples 1.3.6 and 1.3.7. In Section 1.4 we first summarize key distributional properties of the VOU process in Proposition 1.4.1 and Corollary 1.4.2 before we discuss conditions for the VOU model to exhibit tempo-spatial short- or long-range dependence. As Examples 1.4.6 and 1.4.7 demonstrate, long memory in (1.1.6) can arise both through a drift measure with slow decay in time and a measure with slow decay in space. Section 1.5 examines the path regularity of the VOU model. If the noise Λ is Gaussian, Hölder continuous sample paths can be obtained under mild assumptions, see Theorem 1.5.1. When the noise exhibits jumps, the path properties of the VOU process are basically dictated by the noise propagation function g. If it is sufficiently smooth, we show in Theorem 1.5.3 that the VOU process has a *t-càdlàg* version, see Definition 1.5.2. If g is discontinuous, we only have results if the spatial dimension is 1 and $g = \mathbb{1}_{-A}h$, where A is a "triangular" ambit set in space–time and h is smooth enough. In this case, by Theorem 1.5.5, the VOU process has a version which is càdlàg with respect to the triangular shape of A, see Definition 1.5.4 for a precise statement. Section 1.6 contains the proofs of the main results. Section 1.6.4 proves Proposition 1.2.2 regarding deterministic Volterra equations, Section 1.7 gives several examples for resolvent measures, and Section 1.8 lists some of their integrability properties.

1.2 Preliminaries

1.2.1 Lévy bases

We consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that supports a *homogeneous Lévy basis* Λ on $I \times \mathbb{R}^d$ where, depending on the context, $I = \mathbb{R}_+ = [0, \infty)$ or $I = \mathbb{R}$. That is, we assume that $(\Lambda(A))_{A \in \mathcal{B}_{\mathrm{b}}(I \times \mathbb{R}^d)}$ is a collection of random variables indexed by bounded Borel subsets of $I \times \mathbb{R}^d$ such that for all such A we have

$$\begin{aligned}\Lambda(A) = b\mathrm{Leb}_{I\times\mathbb{R}^d}(A) + \sigma W(A) + \int_I \int_{\mathbb{R}^d} \int_{\mathbb{R}} \mathbb{1}_A(t,x) z \mathbb{1}_{\{|z|\le 1\}} (\mathfrak{p} - \mathfrak{q})(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z) \\ + \int_I \int_{\mathbb{R}^d} \int_{\mathbb{R}} \mathbb{1}_A(t,x) z \mathbb{1}_{\{|z|>1\}} \mathfrak{p}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z), \end{aligned} \tag{1.2.1}$$

where

- $b \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ are constants,

- W is Gaussian white noise on $I \times \mathbb{R}^d$ such that $W(A)$ has variance $\mathrm{Leb}_{I\times\mathbb{R}^d}(A)$ (see e.g. Chapter I of [90]),
- $\mathfrak{p}$ is a Poisson random measure on $I\times\mathbb{R}^d\times\mathbb{R}$ with intensity measure $\mathfrak{q}(\mathrm{d}t,\mathrm{d}x,\mathrm{d}z) = \mathrm{Leb}_{I\times\mathbb{R}^d} \otimes \nu$, where ν is a Lévy measure on $\mathbb{R}$ (see e.g. Definition II.1.20 in [56]).

The triplet (b,σ^2,ν) is referred to as the *characteristics* of Λ. Analogously to the Lévy-Itô decomposition of Lévy processes, a Lévy basis is the sum of a deterministic part, a Gaussian part, a compensated sum of small jumps, and a large jumps part. For more information about the meaning of the integrals with respect to $\mathfrak{p}$ or $\mathfrak{p}-\mathfrak{q}$, we refer to Chapter II of [56]. If $\int_{\mathbb{R}} |z| \mathbb{1}_{\{|z|>1\}}\, \nu(\mathrm{d}z) < \infty$, we define $b_1 := b + \int_{\mathbb{R}} z\mathbb{1}_{\{|z|>1\}}\, \nu(\mathrm{d}z)$ as the *mean* of Λ. Similarly, if $\int_{\mathbb{R}} |z| \mathbb{1}_{\{|z|\le 1\}}\, \nu(\mathrm{d}z) < \infty$, we say that Λ has *jumps of finite variation* and define the *drift* of Λ as $b_0 := b - \int_{\mathbb{R}} z\mathbb{1}_{\{|z|\le 1\}}\, \nu(\mathrm{d}z)$. Finally, Λ is said to be *symmetric* if $b = 0$ and ν is a symmetric measure on $\mathbb{R}$.

In this chapter we only need Wiener-type stochastic integrals with respect to Lévy bases since only deterministic integrands will appear. This theory is classic [75] and we only summarize the most important results we need.

Proposition 1.2.1. *Suppose that $g\colon I\times\mathbb{R}^d \to \mathbb{R}$ is a measurable function. The* stochastic integral of g with respect to Λ*, denoted by*

$$\int_{I\times\mathbb{R}^d} g\,\mathrm{d}\Lambda = \int_{I\times\mathbb{R}^d} g(t,x)\,\Lambda(\mathrm{d}t,\mathrm{d}x) = \int_I \int_{\mathbb{R}^d} g(t,x)\,\Lambda(\mathrm{d}t,\mathrm{d}x)$$

either way, is well defined as a limit in probability of approximating simple integrals in the sense of [75] if and only if

(1) $\displaystyle\int_{I\times\mathbb{R}^d} \left| bg(t,x) + \int_{\mathbb{R}} (zg(t,x)\mathbb{1}_{\{|zg(t,x)|\le 1\}} - g(t,x)z\mathbb{1}_{\{|z|\le 1\}})\,\nu(\mathrm{d}z)\right| \mathrm{d}(t,x) < \infty,$

(2) $\displaystyle\int_{I\times\mathbb{R}^d} \sigma^2 |g(t,x)|^2\,\mathrm{d}(t,x) < \infty,$

(3) $\displaystyle\int_{I\times\mathbb{R}^d} \int_{\mathbb{R}} (1\wedge |zg(t,x)|^2)\,\nu(\mathrm{d}z)\,\mathrm{d}(t,x) < \infty.$

In this case, the stochastic integral $\int_{I\times\mathbb{R}^d} g\,\mathrm{d}\Lambda$ has an infinitely divisible distribution with characteristic triplet (b_g,σ_g^2,ν_g) given by

- $b_g = \displaystyle\int_{I\times\mathbb{R}^d} (bg(t,x) + \int_{\mathbb{R}} (zg(t,x)\mathbb{1}_{\{|zg(t,x)|\le 1\}} - g(t,x)z\mathbb{1}_{\{|z|\le 1\}})\,\nu(\mathrm{d}z))\,\mathrm{d}(t,x),$

- $\sigma_g^2 = \int_{I\times\mathbb{R}^d} \sigma^2 |g(t,x)|^2 \,\mathrm{d}(t,x)$,
- $\nu_g(B) = \int_{I\times\mathbb{R}^d} \int_{\mathbb{R}} \mathbb{1}_{\{g(t,x)z\in B\}} \,\nu(\mathrm{d}z)\,\mathrm{d}(t,x)$ *for any Borel set* $B \in \mathcal{B}(\mathbb{R})$.

A set of sufficient conditions for the integrability of g with respect to Λ, which are typically easier to check in practice, is given in Lemma 1.6.1.

1.2.2 Deterministic Volterra equations

We summarize those results on deterministic convolutional Volterra equations that will be useful in the following sections. The monograph [49] is an excellent reference for single parameter Volterra equations. In our tempo-spatial setting, the equation of interest is given by

$$X(t,x) = F(t,x) + \int_0^t \int_{\mathbb{R}^d} X(t-s, x-y)\,\mu(\mathrm{d}s,\mathrm{d}y), \quad (t,x)\in\mathbb{R}_+\times\mathbb{R}^d, \tag{1.2.2}$$

where X is the unknown function, μ is a signed Borel measure on $\mathbb{R}_+\times\mathbb{R}^d$ and $F\colon \mathbb{R}_+\times\mathbb{R}^d \to \mathbb{R}$ is a measurable *forcing function*. Actually, the VOU equation (1.1.6) is exactly of the form (1.2.2), except that the forcing function F is stochastic. Therefore, understanding the solution theory to the deterministic problem (1.2.2) is crucial to solving the VOU equation (1.1.6).

Before we proceed to Equation (1.2.2), let us fix some terminology. For any Borel subset $S\subseteq\mathbb{R}^{d+1}$ we denote by $M(S)$ the space of all signed complete Borel measures on S with finite total variation. As a matter of fact, $M(S)$ becomes a Banach space when equipped with the total variation norm $\|\mu\| := |\mu|(S)$, where $|\mu|$ is the total variation measure of μ. We also introduce the notation $M_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ for signed measures on $\mathbb{R}_+\times\mathbb{R}^d$ which belong to $M([0,T]\times\mathbb{R}^d)$ when restricted to $[0,T]\times\mathbb{R}^d$ for all positive T.

Similarly, for $p\in(0,\infty]$, the space $L^p_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ denotes the collection of all functions $\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}$ whose restrictions to $[0,T]\times\mathbb{R}^d$ belong to $L^p([0,T]\times\mathbb{R}^d) := L^p([0,T]\times\mathbb{R}^d, \mathcal{B}([0,T]\times\mathbb{R}^d), \mathrm{Leb}_{[0,T]\times\mathbb{R}^d})$ for all $T\in\mathbb{R}_+$.

Next, for two measures $\mu,\eta\in M(\mathbb{R}^{d+1})$ the *convolution* $\mu*\eta$ is the completion of the measure that assigns to each Borel set $B\subseteq\mathbb{R}^{d+1}$ the value

$$(\mu*\eta)(B) = \int_{\mathbb{R}^{d+1}} \eta(B-z)\,\mu(\mathrm{d}z), \tag{1.2.3}$$

where $B-z = \{s-z\colon s\in B\}$. Since the function $z\mapsto\eta(B-z)$ is Borel measurable and bounded, the integral (1.2.3) is always well-defined. If μ and η belong to $M(S)$ for

some $S \in \mathcal{B}(\mathbb{R}^{d+1})$, we first extend μ and η to $\mathbb{R}^{d+1}$ by setting $\bar{\mu}(B) = \mu(B \cap S)$ and $\bar{\eta}(B) = \eta(B \cap S)$, then obtain $\bar{\mu} * \bar{\eta}$ as above and finally define the convolution $\mu * \eta$ as the restriction of $\bar{\mu} * \bar{\eta}$ to S. It is customary to write $\mu^{*0} = \delta_{0,\mathbb{R}^{d+1}}$ and $\mu^{*j} = \mu * \mu^{*(j-1)}$ for $j \in \mathbb{N}$. In a similar way, if $\mu \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ and $h\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function, we define *the convolution of h with respect to μ* as the function $h * \mu = \mu * h$ that is given by

$$(h * \mu)(t,x) = (\mu * h)(t,x) := \int_0^t \int_{\mathbb{R}^d} h(t-s, x-y)\, \mu(\mathrm{d}s, \mathrm{d}y),$$

which is defined for those $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ for which the integral exists. Note that the measures considered in this chapter may have atoms. Hence we use the convention that integrals over an interval always include the endpoints.

The next theorem is the key result from the theory of convolutional Volterra equations that we need in Section 1.3. It determines conditions under which (1.2.2) has a unique solution. By a *solution* to (1.2.2) we understand a measurable function $X\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ such that (1.2.2) holds for (Lebesgue-)almost all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$. Two solutions are identified if they agree almost everywhere on $\mathbb{R}_+ \times \mathbb{R}^d$.

Proposition 1.2.2. *Let $\mu \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ be such that $\mu(\{0\} \times \mathbb{R}^d) = 0$.*

(1) There exists a unique measure $\rho \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$, called the resolvent of μ, *such that*

$$\rho + \mu = \mu * \rho.$$

(2) If $F \in L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ for some $p \in [1, \infty]$, then there exists a unique solution $X \in L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ to (1.2.2). This solution is given by

$$X(t,x) = F(t,x) - \int_0^t \int_{\mathbb{R}^d} F(t-s, x-y)\, \rho(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad (1.2.4)$$

or in short $X = F - \rho * F$, where ρ is the resolvent of μ.

(3) For every $F \in \mathcal{F}(\mu)$, where

$$\mathcal{F}(\mu) := \{F \text{ measurable}\colon |\mu| * |F|, |\rho| * |F|, (|\mu| * |\rho|) * |F|, (|\mu| * |\rho|^{*2}) * |F| < \infty \text{ a.e.}\},$$

the function X in (1.2.4) is the unique solution to (1.2.2) in the space

$$\mathcal{L}(\mu) := \{L \text{ measurable}\colon |\mu| * |L|, |\rho| * |L|, (|\mu| * |\rho|) * |L| < \infty \text{ a.e.}\}. \quad (1.2.5)$$

A proof of this theorem, together with some properties and examples of convolutions and resolvents, is given in Sections 1.6.4,1.7 and 1.8. Note that in a Banach algebra framework the resolvent is also called *quasi-inverse* (cf. Section 2.1 of [71]).

1.3 Solution to the VOU equation

In this section we prove existence and uniqueness of solutions to the VOU equation under general assumptions. In fact, we consider a slightly more general equation than (1.1.6), namely

$$X(t,x) = V(t,x) + \int_0^t \int_{\mathbb{R}^d} X(t-s,x-y)\,\mu(\mathrm{d}s,\mathrm{d}y) + \int_0^t \int_{\mathbb{R}^d} g(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y) \tag{1.3.1}$$

for $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, where μ is the drift measure, g is the noise propagation function, V is a measurable stochastic process and Λ is a homogeneous Lévy basis. As usual, we say that a process $\tilde{X}$ is a version of the process X on $\mathbb{R}_+ \times \mathbb{R}^d$ if for every $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ we have $\tilde{X}(t,x) = X(t,x)$ almost surely.

Theorem 1.3.1. *Let μ be a measure in $M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ with $\mu(\{0\} \times \mathbb{R}^d) = 0$ and resolvent ρ, $g\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function and Λ be a homogeneous Lévy basis on $\mathbb{R}_+ \times \mathbb{R}^d$ with characteristics (b, σ^2, ν). We assume that*

$$\int_{\mathbb{R}} \left(|z|^\alpha \mathbb{1}_{\{|z|>1\}} + |z|^\beta \mathbb{1}_{\{|z|\le 1\}} \right) \nu(\mathrm{d}z) < \infty \tag{1.3.2}$$

and $g \in L^\alpha_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^\beta_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ for some $\alpha \in (0,1]$, and some $\beta \in [1,2]$ if $\sigma = 0$, and $\beta = 2$ if $\sigma \neq 0$.

If $\alpha < 1$, we further suppose that there exists a submultiplicative weight function $\varphi\colon \mathbb{R}^d \to \mathbb{R}_+$ (that is, a measurable function with $\varphi(0) = 1$ and $\varphi(x+y) \le \varphi(x)\varphi(y)$ for $x, y \in \mathbb{R}^d$, such that φ is locally bounded and locally bounded away from zero) satisfying $\varphi^{-\alpha} \in L^1(\mathbb{R}^d)$, $\varphi(|\mu| * |g|) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ and $\varphi\mu \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ (where $(\varphi\mu)(\mathrm{d}t, \mathrm{d}x) := \varphi(x)\,\mu(\mathrm{d}t, \mathrm{d}x)$).

Then there exists a measurable version of the process

$$\int_0^t \int_{\mathbb{R}^d} g(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d,$$

in $\mathcal{F}(\mu)$, and for this version and every measurable process V with almost all paths in $\mathcal{F}(\mu)$, Equation (1.3.1) has a solution with almost all paths in $\mathcal{L}(\mu)$ as in (1.2.5). A version of this solution is given by

$$X(t,x) = V(t,x) - (\rho * V)(t,x) + \int_0^t \int_{\mathbb{R}^d} (g - \rho * g)(t-s,x-y)\,\Lambda(\mathrm{d}s,\mathrm{d}y),\ (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \tag{1.3.3}$$

or in short $X = V - \rho * V + (g - \rho * g) * \Lambda$.

Moreover, this solution is unique in the sense that for any other solution $\tilde{X}$ *with almost all paths in* $\mathcal{L}(\mu)$*, we have that almost surely, the paths of* X *and* $\tilde{X}$ *are equal almost everywhere on* $\mathbb{R}_+ \times \mathbb{R}^d$.

Remark 1.3.2 (1) A simple sufficient condition for V to have paths in $\mathcal{F}(\mu)$ is, for example, when the function $(t,x) \mapsto \mathbb{E}[|V(t,x)|]$ belongs to $L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$. Another would be that the paths of V almost surely belong to $L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ for some $p \in [1, \infty]$.

(2) Regarding the case $\alpha \in (0,1)$, typical examples for submultiplicative weight functions include $\varphi(x) = (1+|x|)^\eta (\log(\mathrm{e} \vee |x|))^\gamma$ and $\varphi(x) = \exp(|x|^\gamma)$ for $\eta, \gamma \geq 0$. The reason why we impose additional conditions when α is smaller than one is that in our proof we have to ensure $(|\rho| * |g|)^\alpha \in L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$. Instead of formulating conditions on ρ, which may not be known explicitly, the assumptions in Theorem 1.3.1 are solely on μ.

□

We also remark that the condition in (1.3.2) on the Lévy measure ν are not necessary in general. For instance, it is well known that the OU process is defined for all Lévy processes without any restrictions on ν. But this is different to our case because the spatial coordinate is in the non-compact space $\mathbb{R}^d$ and the noise propagation function g is not necessarily bounded. For given classes of g and μ it may be possible to relax the assumptions of Theorem 1.3.1. But given that these are already general enough to cover most practical cases, we refrain from doing so.

Having clarified the local existence of solutions to (1.3.1), our next aim is to investigate their long-term behavior and the existence of stationary solutions. A stochastic process X on $\mathbb{R}_+ \times \mathbb{R}^d$ is called *strictly stationary* if for every $n \in \mathbb{N}$ and $(\tau, \xi), (t_1, x_1), \ldots, (t_n, x_n) \in \mathbb{R}_+ \times \mathbb{R}^d$ the distributions of $(X(t_1, x_1), ..., X(t_n, x_n))$ and $(X(t_1 + \tau, x_1 + \xi), ..., X(t_n + \tau, x_n + \xi))$ are equal.

Theorem 1.3.3. *Let* Λ *be a homogeneous Lévy basis on* $\mathbb{R} \times \mathbb{R}^d$*, the conditions of Theorem 1.3.1 be valid with* $\alpha, \beta \in (0, 2]$ *and additionally* $g \in L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ *such that the process* X *as given in* (1.3.3) *is the solution to* (1.3.1)*. Moreover, we assume the following hypotheses:*

(1) For all $x_1, \ldots, x_n \in \mathbb{R}^d$ *we have that*

$$\big(V(t,x_1) - (\rho * V)(t,x_1), \ldots, V(t,x_n) - (\rho * V)(t,x_n)\big) \xrightarrow{\mathrm{d}} \big(F_\infty(x_1), \ldots, F_\infty(x_n)\big)$$

as $t \to \infty$ for some deterministic measurable function $F_\infty \colon \mathbb{R}^d \to \mathbb{R}$.

(2) We have $g - \rho * g \in L^\alpha(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^\beta(\mathbb{R}_+ \times \mathbb{R}^d)$.

(3) Conditions (1.6.1) *and* (1.6.2) *are satisfied with $K(A) := A^{1-\alpha}$ and $k(a) := a^{1-\beta}$.*

Then we have for all $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \mathbb{R}^d$ that

$$\big(X(t,x_1), \dots, X(t,x_n)\big) \xrightarrow{\mathrm{d}} \big(X_\infty(x_1), \dots, X_\infty(x_n)\big), \quad t \to \infty,$$

where X_∞ is the spatial process

$$X_\infty(x) := F_\infty(x) + \int_0^\infty \int_{\mathbb{R}^d} (g - \rho * g)(s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad x \in \mathbb{R}^d.$$

Furthermore, if g is integrable with respect to Λ and V is independent of Λ with the same finite-dimensional distributions as

$$\int_{-\infty}^0 \int_{\mathbb{R}^d} g(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \tag{1.3.4}$$

then X is a strictly stationary process on $\mathbb{R}_+ \times \mathbb{R}^d$. In particular, if V equals the process in (1.3.4)*, X can be written as the two-sided strictly stationary process*

$$X(t,x) = \int_{-\infty}^t \int_{\mathbb{R}^d} (g - \rho * g)(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d. \tag{1.3.5}$$

We give various possibilities of how to ensure the requirements in Theorem 1.3.3.

Lemma 1.3.4. *In the following cases, condition* (1.6.1) *with $K(A) := A^{1-\alpha}$ (resp.* (1.6.2) *with $k(a) := a^{1-\beta}$) is already implied by* (1.3.2)*:*

(1) $\alpha \in (0,1]$ (resp. $\beta \in [1,2]$).

(2) $\alpha \in (1,2]$ and $b_1 = 0$ (resp. $\beta \in (0,1)$ and $b_0 = 0$).

(3) Λ is symmetric.

Lemma 1.3.5. *For $p \in [1,\infty]$ we have $g - \rho * g \in L^p(\mathbb{R}_+ \times \mathbb{R}^d)$ under each of the following assumptions:*

(1) $\rho \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ and $g \in L^p(\mathbb{R}_+ \times \mathbb{R}^d)$.

(2) $\rho \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ satisfies $\rho(\mathbb{R}_+ \times \mathbb{R}^d) = 1$ and there exists a constant $g_\infty \in \mathbb{R}$ such that $g - g_\infty \in L^p(\mathbb{R}_+ \times \mathbb{R}^d)$.

(3) For some $q, s \in [1, \infty]$ *satisfying* $s^{-1} + q^{-1} = 1 + p^{-1}$ *we have* $g \in L^p(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^s(\mathbb{R}_+ \times \mathbb{R}^d)$ *and* $\rho(\mathrm{d}t, \mathrm{d}x) = r(t,x)\,\mathrm{d}(t,x)$ *with* $r \in L^q(\mathbb{R}_+ \times \mathbb{R}^d)$.

In contrast to Theorem 1.3.1, the conditions imposed in Theorem 1.3.3 (and also in Lemma 1.3.5) explicitly depend on the behavior of the resolvent measure ρ, instead merely on μ. In fact, there are no general necessary and sufficient conditions for a measure μ to have a resolvent with certain integrability properties. In Section 1.8 we present several results in this respect.

We conclude this section by investigating two variants of a VOU process sharing the same drift measure as in the classical Ornstein-Uhlenbeck process, namely

$$\mu = -\lambda \mathrm{Leb}_{\mathbb{R}_+} \otimes \delta_{0,\mathbb{R}^d}, \tag{1.3.6}$$

and different choices for the noise propagation function g.

Example 1.3.6 (VOU process with infinite speed propagation of noise) As a first example we investigate the equation

$$X(t,x) = -\lambda \int_0^t X(s,x)\,\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda'|x-y|}\,\Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \tag{1.3.7}$$

where $\lambda \in \mathbb{R}$ and $\lambda' > 0$ and $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. A closer inspection reveals two characteristic features of this model: first, the parameter λ leads to a mean-reverting behavior in time like in the classical OU case if $\lambda > 0$; second, since the noise propagation function $g(t,x) = \mathrm{e}^{-\lambda'|x|}$ does not depend on t and is strictly positive, each innovation of Λ (and jump if the Lévy measure is not zero) affects $X(\cdot, x)$ for all x simultaneously. However, as controlled by λ', the impact of an innovation decreases exponentially in the distance between the current location x and the point of origin y. For fixed x, we further observe that the second summand on the right-hand side of (1.3.7) is a Lévy process, so the solution of (1.3.7) is in fact a system $(X(\cdot,x) : x \in \mathbb{R}^d)$ of dependent classical OU processes.

Since the resolvent measure of (1.3.6) is $\rho(\mathrm{d}t, \mathrm{d}x) = \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t\,\delta_{0,\mathbb{R}^d}(\mathrm{d}x)$ (see Lemma 1.7.1), a simple calculation yields $(g - \rho * g)(t,x) = \mathrm{e}^{-\lambda t - \lambda'|x|}$. Therefore, as soon as the Lévy measure ν of Λ satisfies

$$\int_{\mathbb{R}} |z|^\alpha \mathbb{1}_{\{|z|>1\}}\,\nu(\mathrm{d}z) < \infty \tag{1.3.8}$$

for some $\alpha > 0$, we derive from Theorem 1.3.1 that the unique solution to (1.3.7) is given by

$$X(t,x) = \int_0^t \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda(t-s) - \lambda'|x-y|}\,\Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d. \tag{1.3.9}$$

Indeed, we can choose α as above and $\beta = 2$ in (1.3.2) because $g \in L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ for all $p \in (0, \infty]$. If $\alpha < 1$, we can take $\varphi(x) := (1 + |x|)^{(d+1)/\alpha}$.

Against the background that $X(\cdot, x)$ is an OU process for fixed x, it is not surprising that also Theorem 1.3.3 applies if $\lambda > 0$ (then $g - \rho * g \in L^p(\mathbb{R}_+ \times \mathbb{R}^d)$ for all $p \in (0, \infty]$). In this case the strictly stationary process (1.3.5) is given by

$$X(t,x) = \int_{-\infty}^{t} \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda(t-s)+\lambda'|x-y|}\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d. \tag{1.3.10}$$

□

In the previous example innovations of Λ at a given site have an instantaneous effect on all other sites. In contrast to this, the next model incorporates a traveling waves mechanism such that a certain amount of time is needed for the propagation of innovations from one to another point in space.

Example 1.3.7 (VOU process with finite speed propagation of noise) We consider

$$X(t,x) = -\lambda \int_0^t X(s,x)\, \mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} \mathbb{1}_{\{|x-y| \le c(t-s)\}} \mathrm{e}^{-\lambda'|x-y|}\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \tag{1.3.11}$$

with parameters $c > 0$ and $\lambda, \lambda' \in \mathbb{R}$. As a result, the time until an innovation of Λ at a site y arrives at another site x amounts to $|x-y|/c$. With $g(t,x) = \mathbb{1}_{\{|x| \le ct\}} \mathrm{e}^{-\lambda'|x|}$, an elementary computation shows that

$$(g - \rho * g)(t,x) = \mathbb{1}_{\{|x| \le ct\}} \mathrm{e}^{-\lambda t - (\lambda' - \lambda/c)|x|}, \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d.$$

Consequently, under assumption (1.3.8), Equation (1.3.11) has the unique solution

$$X(t,x) = \int_0^t \int_{\mathbb{R}^d} \mathbb{1}_{\{|x-y| \le c(t-s)\}} \mathrm{e}^{-\lambda(t-s) - (\lambda' - \lambda/c)|x-y|}\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d. \tag{1.3.12}$$

In order to determine the long-term behavior of (1.3.12), we can use Fubini's theorem to verify that only for $\lambda > 0$ and $\lambda' > 0$ the integral

$$\begin{aligned}
\int_0^\infty \int_{\mathbb{R}^d} (g - \rho * g)^p(t,x)\, \mathrm{d}(t,x) &= \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_0^\infty \mathrm{e}^{-\lambda p t} \int_0^{ct} \mathrm{e}^{-(\lambda' - \lambda/c)pr} r^{d-1}\, \mathrm{d}r\, \mathrm{d}t \\
&= \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_0^\infty \frac{\mathrm{e}^{-r\lambda p/c}}{\lambda p} \mathrm{e}^{-(\lambda' - \lambda/c)pr} r^{d-1}\, \mathrm{d}r \\
&= \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})\lambda p} \int_0^\infty \mathrm{e}^{-\lambda' p r} r^{d-1}\, \mathrm{d}r
\end{aligned}$$

is finite for $p \in (0, \infty)$. So only in this case, the finite-dimensional distributions of (1.3.12) converge to that of the process

$$X(t,x) = \int_{-\infty}^{t} \int_{\mathbb{R}^d} \mathbb{1}_{\{|x-y| \le c(t-s)\}} \mathrm{e}^{-\lambda(t-s)-(\lambda'-\lambda/c)|x-y|} \, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d. \tag{1.3.13}$$

We notice that if $\lambda' = \lambda/c$, the process X in (1.3.12) is exactly the so-called $\mathrm{OU}_\wedge$ model investigated in [9, 69]. □

Remark 1.3.8 At the end of this section we want to highlight a connection to stochastic partial differential equations as studied in [72]. For this purpose assume in addition to the conditions of Theorem 1.3.1 that Λ has mean zero and a finite second moment. Let U be a Hilbert space such that the embedding of $\mathcal{H} := L^2(\mathbb{R}^d)$ into U is dense and Hilbert-Schmidt (see Example 14.25 of [72] for an example of U). Then the process $W\colon [0,\infty) \times \mathcal{H} \to L^2(\Omega)$ defined by

$$W(t,\phi) := \int_0^t \int_{\mathbb{R}^d} \phi(y) \, \Lambda(\mathrm{d}s, \mathrm{d}y)$$

is the sum of a cylindrical Wiener process and an impulsive cylindrical process on $\mathcal{H}$ (cf. Definitions 7.11 and 7.23 of [72]). Combining Theorems 7.13 and 7.22 of [72] we obtain a U-valued square-integrable Lévy martingale L satisfying

$$L(t) = \sum_{n \in \mathbb{N}} W(t, e_n) e_n$$

for any fixed orthonormal basis $(e_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ (cf. Remark 7.14 of [72]).

If we extend the convolution operators

$$S(t)\colon \mathcal{H} \to \mathcal{H}, \quad \phi(x) \mapsto \int_{\mathbb{R}^d} g(t, x-y)\phi(y) \, \mathrm{d}y$$

onto U, we may rewrite

$$\begin{aligned}
&\int_0^t \int_{\mathbb{R}^d} g(t-s, x-y) \, \Lambda(\mathrm{d}s, \mathrm{d}y) \\
&\quad = \int_0^t \int_{\mathbb{R}^d} \sum_{n\in\mathbb{N}} \int_{\mathbb{R}^d} g(t-s, x-z) e_n(z) \, \mathrm{d}z \, e_n(y) \, \Lambda(\mathrm{d}s, \mathrm{d}y) \\
&\quad = \sum_{n\in\mathbb{N}} \int_0^t \int_{\mathbb{R}^d} g(t-s, x-z) e_n(z) \, \mathrm{d}z \int_{\mathbb{R}^d} e_n(y) \, \Lambda(\mathrm{d}s, \mathrm{d}y) \\
&\quad = \sum_{n\in\mathbb{N}} \int_0^t S(t-s) e_n \, \mathrm{d}W(s, e_n) = \int_0^t S(t-s) \, \mathrm{d}L(s).
\end{aligned}$$

From the solution formula (1.3.3) (with $V \equiv 0$) we see that $X(t,x)$ belongs to $L^2(\mathbb{R}^d, (1+|x|^r)^{-1})$ for fixed t and $r > d/2$, where a function ϕ is an element of $L^2(\mathbb{R}^d, \eta)$ if and only if $\eta\phi \in L^2(\mathbb{R}^d)$.

Further assuming that $\mu(\mathrm{d}t, \mathrm{d}x) = \nu_t(\mathrm{d}x)\,\mathrm{d}t$ for a transition kernel ν (as in Examples 1.3.6 and 1.3.7) and that the mapping

$$Q(t)\colon g(x) \mapsto \int_{\mathbb{R}^d} g(x-y)\,\nu_t(\mathrm{d}y)$$

is a linear convolution operator from $L^2(\mathbb{R}^d, (1+|x|^r)^{-1})$ into itself, we obtain

$$\int_0^t \int_{\mathbb{R}^d} X(t-s, x-y)\,\mu(\mathrm{d}s, \mathrm{d}y) = \int_0^t \int_{\mathbb{R}^d} X(s, x-y)\,\nu_{t-s}(\mathrm{d}y)\,\mathrm{d}s = \int_0^t Q(t-s)X(s)\,\mathrm{d}s, \tag{1.3.14}$$

where $X(t) := X(t,\cdot)$. In short, under the conditions above, the VOU equation

$$X(t,x) = X(0,x) + \int_0^t \int_{\mathbb{R}^d} X(t-s, x-y)\,\mu(\mathrm{d}s, \mathrm{d}y) + \int_0^t \int_{\mathbb{R}^d} g(t-s, x-y)\,\Lambda(\mathrm{d}s, \mathrm{d}y)$$

is equivalent to the infinite-dimensional equation

$$X(t) = X(0) + \int_0^t Q(t-s)X(s)\,\mathrm{d}s + \int_0^t S(t-s)\,\mathrm{d}L(s).$$

In the literature for stochastic partial differential equations several criteria are known for $t \mapsto X(t)$, viewed as a process with values in a Hilbert space, to have continuous or càdlàg sample paths (see for instance Theorem 11.8 of [72] or Theorem 4.5 and Remark 4.6 of [85]). By contrast, the random field approach to the VOU equation allows for a detailed analysis of the tempo-spatial path properties of $(t,x) \mapsto X(t,x)$ as in Section 1.5. □

1.4 Distributional properties

A convenient tool for characterizing the distribution of tempo-spatial processes is the *generalized cumulant functional* introduced in [11]. For the solution process X in (1.3.3), we obtain the following result.

Proposition 1.4.1. *Suppose that the conditions of Theorem 1.3.1 are satisfied with $V \equiv 0$ and that X is the solution to* (1.3.1) *given in* (1.3.3). *If $m \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ is supported on a compact subset of $\mathbb{R}_+ \times \mathbb{R}^d$, the integral $m[X] := \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} X(t,x)\,m(\mathrm{d}t, \mathrm{d}x)$*

is well defined and the generalized cumulant functional of X with respect to m is given by

$$\log \mathbb{E}\left[\mathrm{e}^{\mathrm{i}um[X]}\right] = \mathrm{i}b_G u - \frac{1}{2}\sigma_G^2 u^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}uz} - 1 - \mathrm{i}uz\mathbb{1}_{\{|z|\leq 1\}}\right) \nu_G(\mathrm{d}z), \quad u \in \mathbb{R},$$

where (b_G, σ_G^2, ν_G) are the characteristics as given in Proposition 1.2.1 for the function

$$G(s,y) = \int_s^\infty \int_{\mathbb{R}^d} (g - \rho * g)(t-s, x-y)\, m(\mathrm{d}t, \mathrm{d}x), \quad (s,y) \in \mathbb{R}_+ \times \mathbb{R}^d.$$

For example, if one takes $m(\mathrm{d}t, \mathrm{d}x) = \theta_1 \delta_{(t_1,x_1)} + \cdots + \theta_n \delta_{(t_n,x_n)}$, one obtains the cumulant-generating function of $(X(t_1,x_1), \ldots, X(t_n,x_n))$. Based on this, it is also possible to derive the second-order structure for X, see Proposition 2 in [11] for a proof.

Corollary 1.4.2. *Suppose that the assumptions of Theorem 1.3.1 are satisfied with $V \equiv 0$ and that X is the solution process* (1.3.3)*.*

(1) If the assumptions of Theorem 1.3.1 hold with $\alpha = 1$, $X(t,x)$ has a finite first moment for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ given by

$$\mathbb{E}[X(t,x)] = b_1 \int_0^t \int_{\mathbb{R}^d} (g - \rho * g)(s,y)\, \mathrm{d}(s,y). \tag{1.4.1}$$

(2) If the assumptions of Theorem 1.3.1 are additionally satisfied with $\alpha = 2$, $X(t,x)$ has a finite second moment for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ and

$$\mathrm{Var}[X(t,x)] = \left(\sigma^2 + \int_{\mathbb{R}} z^2\, \nu(\mathrm{d}z)\right) \int_0^t \int_{\mathbb{R}^d} (g - \rho * g)^2(s,y)\, \mathrm{d}(s,y). \tag{1.4.2}$$

Moreover, for $(t,x), (\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R}^d$ we have

$$\begin{aligned}&\mathrm{Cov}[X(t,x), X(t+\tau, x+\xi)] \\ &\quad = \left(\sigma^2 + \int_{\mathbb{R}} z^2\, \nu(\mathrm{d}z)\right) \int_0^t \int_{\mathbb{R}^d} (g - \rho * g)(s,y)(g - \rho * g)(s+\tau, y+\xi)\, \mathrm{d}(s,y).\end{aligned} \tag{1.4.3}$$

Furthermore, in the setting of Theorem 1.3.3, if X is the strictly stationary process (1.3.5)*, then the formulae* (1.4.1)*,* (1.4.2) *and* (1.4.3) *remain valid if we replace t by ∞ on the right-hand sides.*

For illustration we calculate the autocorrelation functions for the models in Examples 1.3.6 and 1.3.7.

Example 1.4.3 (Second-order structure for Example 1.3.6)
Under the moment assumptions of Corollary 1.4.2, the mean and the autocovariance function of the stationary process X in (1.3.10) are given by

$$\mathbb{E}[X(t,x)] = b_1 \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda s - \lambda'|y|} \,\mathrm{d}(s,y) = \frac{2b_1 \pi^{d/2} \Gamma(d)}{\lambda (\lambda')^d \Gamma(\frac{d}{2})}, \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d, \tag{1.4.4}$$

and, denoting $m_2 := \sigma^2 + \int_{\mathbb{R}} z^2 \,\nu(\mathrm{d}z)$,

$$\begin{aligned}\mathrm{Cov}[X(t,x), X(t+\tau, x+\xi)] &= m_2 \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda s - \lambda'|y|} \mathrm{e}^{-\lambda(s+\tau) - \lambda'|y+\xi|} \,\mathrm{d}(s,y) \\ &= \frac{m_2 \mathrm{e}^{-\lambda\tau}}{2\lambda} \int_{\mathbb{R}^d} \mathrm{e}^{-\lambda'|y| - \lambda'|y+\xi|} \,\mathrm{d}y =: \frac{m_2 \mathrm{e}^{-\lambda\tau}}{2\lambda} E(\xi).\end{aligned}$$

The integral $E(\xi)$ is the convolution of the function $f(\xi) = \mathrm{e}^{-\lambda'|\xi|}$ with itself in $\mathbb{R}^d$. Since the Fourier transform of f is known (see Theorem I.1.14 in [82]), $E(\xi)$ is the inverse Fourier transform of the function

$$\frac{c_d^2 a^2}{(a^2 + |x|^2)^{d+1}},$$

where $a := \lambda'/(2\pi)$ and $c_d := \Gamma((d+1)/2)\pi^{-(d+1)/2}$. Hence, using Theorem IV.3.3 of [82] and denoting by J_α and K_α the Bessel functions of the first kind and the modified Bessel functions of the second kind, respectively, we obtain

$$\begin{aligned}E(\xi) &= 2\pi |\xi|^{1-d/2} c_d^2 a^2 \int_0^\infty \frac{J_{d/2-1}(2\pi|\xi| r) r^{d/2}}{(a^2 + |r|^2)^{d+1}} \,\mathrm{d}r \\ &= \frac{2\Gamma(\frac{d+1}{2})^2}{\Gamma(d+1)} \left(\frac{\lambda'}{2\pi}\right)^{1-d/2} |\xi|^{1+d/2} K_{1+d/2}(\lambda'|\xi|).\end{aligned}$$

This yields for $(t,x) \in \mathbb{R} \times \mathbb{R}^d$ and $(\tau, \xi) \in \mathbb{R}_+ \times \mathbb{R}^d$

$$\mathrm{Cov}[X(t,x), X(t+\tau, x+\xi)] = \frac{m_2 \Gamma(\frac{d+1}{2})^2}{\lambda \Gamma(d+1)} \left(\frac{\lambda'}{2\pi}\right)^{1-d/2} \mathrm{e}^{-\lambda\tau} |\xi|^{1+d/2} K_{1+d/2}(\lambda'|\xi|).$$

Since $\lim_{x \downarrow 0} x^\alpha K_\alpha(x) = 2^{\alpha-1}\Gamma(\alpha)$ for $\alpha \geq 0$, the autocorrelation function reads as

$$\mathrm{corr}[X(t,x), X(t+\tau, x+\xi)] = \frac{(\lambda')^{1+d/2}}{2^{d/2}\Gamma(1+\frac{d}{2})} \mathrm{e}^{-\lambda\tau} |\xi|^{1+d/2} K_{1+d/2}(\lambda'|\xi|).$$

If $d = 1$, $d = 2$ or $d = 3$, this formula reduces to

$$\begin{aligned}
\operatorname{corr}[X(t,x), X(t+\tau, x+\xi)] &= \mathrm{e}^{-\lambda\tau}(\lambda'|\xi| + 1)\mathrm{e}^{-\lambda'|\xi|}, && (\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R},\\
\operatorname{corr}[X(t,x), X(t+\tau, x+\xi)] &= \tfrac{(\lambda')^2}{2}\mathrm{e}^{-\lambda\tau}|\xi|^2 K_2(\lambda'|\xi|), && (\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R}^2,\\
\operatorname{corr}[X(t,x), X(t+\tau, x+\xi)] &= \mathrm{e}^{-\lambda\tau}\left(\tfrac{(\lambda')^2}{3}|\xi|^2 + \lambda'|\xi| + 1\right)\mathrm{e}^{-\lambda'|\xi|}, && (\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R}^3,
\end{aligned}$$

respectively. □

Example 1.4.4 (Second-order structure for Example 1.3.7)
We obtain the same value as in (1.4.4) for the expectation of (1.3.13):

$$\mathbb{E}[X(t,x)] = b_1 \int_{\mathbb{R}_+}\int_{\mathbb{R}^d} \mathbb{1}_{\{|y|\le cs\}}\mathrm{e}^{-\lambda s - (\lambda' - \lambda/c)|y|}\,\mathrm{d}(s,y) = \frac{2b_1\pi^{d/2}\Gamma(d)}{\lambda(\lambda')^d\Gamma(\frac{d}{2})}, \quad (t,x) \in \mathbb{R}\times\mathbb{R}^d.$$

Regarding the autocovariance function, a straightforward calculation for $d = 1$ shows that

$$\begin{aligned}
&\operatorname{Cov}[X(t,x), X(t+\tau, x+\xi)]\\
&\quad= \frac{m_2}{4\lambda}\mathrm{e}^{-\lambda\tau}\mathrm{e}^{(\lambda/c-\lambda')|\xi|}\left(\frac{c}{\lambda}\left(\mathrm{e}^{-\lambda(|\xi|/c-\tau)_+} + \mathrm{e}^{-2\lambda(|\xi|/c-\tau)}\mathrm{e}^{\lambda(|\xi|/c-\tau)_+} - \mathrm{e}^{-2\lambda(|\xi|/c-\tau)}\right)\right.\\
&\qquad\left. + \frac{1}{\lambda'}\mathrm{e}^{-2\lambda(|\xi|/c-\tau)_+}\right) + \frac{m_2}{4\lambda}\mathrm{e}^{-\lambda\tau}\mathrm{e}^{-(\lambda/c+\lambda')|\xi|}\left(\frac{1}{\lambda'} - \frac{c}{\lambda}\right)
\end{aligned}$$

for all $(t,x) \in \mathbb{R}\times\mathbb{R}$ and $(\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R}$. □

The autocovariance function in the last example exhibits an exponential decay in both t and x, so the corresponding process X has a *short-range dependence* structure. However, as we shall see, under suitable choices of g and μ, it may happen that the autocovariance function is not integrable, i.e.

$$\int_0^\infty \int_{\mathbb{R}^d} |\operatorname{Cov}[X(t,x), X(t+\tau, x+\xi)]|\,\mathrm{d}(\tau,\xi) = \infty, \tag{1.4.5}$$

hence giving rise to models with *long-range dependence*. A first result concerning short- versus long-range dependence is the following.

Proposition 1.4.5. *Let $\rho \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ be the resolvent measure associated to some measure $\mu \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ with $\mu(\{0\} \times \mathbb{R}^d) = 0$ and suppose that the Lévy basis Λ has a finite second moment, that is $\int_{\{|z|>1\}} |z|^2\,\nu(\mathrm{d}z) < \infty$.*

(1) If in addition $g - \rho * g \in L^1(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^2(\mathbb{R}_+ \times \mathbb{R}^d)$*, then the process* X *in* (1.3.5) *is well defined, has a finite second moment and*

$$\int_0^\infty \int_{\mathbb{R}^d} |\mathrm{Cov}[X(t,x), X(t+\tau, x+\xi)]| \, \mathrm{d}(\tau, \xi) < \infty.$$

(2) If Λ *has zero mean,* $g - \rho * g \in L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^2(\mathbb{R}_+ \times \mathbb{R}^d)$ *but* $g - \rho * g \notin L^1(\mathbb{R}_+ \times \mathbb{R}^d)$*, and* $g - \rho * g$ *is non-negative or non-positive for all* $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$*, then the process* X *in* (1.3.5) *is well defined, has a finite second moment and is long-range dependent in the sense of* (1.4.5)*.*

Example 1.4.6 (Long-range dependence by temporal regular variation) We consider a measure μ of the form $\mu(\mathrm{d}t, \mathrm{d}x) = k(t) \, \mathrm{d}t \, \delta_{0,\mathbb{R}^d}(\mathrm{d}x)$ with some $k \in L^1_{\mathrm{loc}}(\mathbb{R}_+)$. By Lemma 1.7.1, the resolvent of μ has the form $\rho(\mathrm{d}t, \mathrm{d}x) = r(t) \, \mathrm{d}t \, \delta_{0,\mathbb{R}^d}(\mathrm{d}x)$ for some $r \in L^1_{\mathrm{loc}}(\mathbb{R}_+)$, which is the temporal resolvent of the function k (i.e., we have $r + k = r * k$ where $*$ stands for convolution on $\mathbb{R}_+$). Now suppose that the function k satisfies the following assumptions:

- $k(t) = -t^{-\alpha} L(t)$ for all $t \in \mathbb{R}_+$, some $\alpha \in (0, 1/2)$ and some function $L : \mathbb{R}_+ \to (0, \infty)$ that is slowly varying at infinity.
- k is differentiable with a continuous strictly positive derivative that belongs to $L^1(\mathbb{R}_+)$.
- The function $t \mapsto \log(-k(t))$ is convex in t.

Then, by Theorem 3.2 of [7],

$$\lim_{t \to \infty} \left(1 - \int_0^t r(s) \, \mathrm{d}s\right) t^{1-\alpha} L(t) = \frac{\sin(\alpha\pi)}{\pi}. \tag{1.4.6}$$

If now $g(t,x) = g_0(x)$ for some non-negative (or non-positive) $g_0 \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, then

$$(g - \rho * g)(t,x) = g_0(x) \left(1 - \int_0^t r(s) \, \mathrm{d}s\right), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d,$$

is, because of (1.4.6) and Corollary 8.8 of Chapter 9 in [49], non-negative (or non-positive), belongs to $L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^2(\mathbb{R}_+ \times \mathbb{R}^d)$, but not to $L^1(\mathbb{R}_+ \times \mathbb{R}^d)$. Hence we conclude from Proposition 1.4.5 that the resulting stationary process in (1.3.5) has long-range dependence. One possible choice of k is $k(t) = -1/(\alpha(1+t)^\alpha)$ with $\alpha \in (0, 1/2)$, cf. Example 3.5 in [7]. □

In the previous example, the non-integrability of the resolvent measure is essentially due to the regular variation of the function k. In the next example, long-range dependence arises through a drift measure of the form $\mu = -\lambda \mathrm{Leb}_{\mathbb{R}_+} \otimes m$ with some $\lambda > 0$ and $m \in M(\mathbb{R}^d)$.

Example 1.4.7 (Long-range dependence by spatial regular variation) Consider the measure $\mu(\mathrm{d}t, \mathrm{d}x) = -\mathrm{d}t\, f(x)\, \mathrm{d}x$ with $f(x) = 1/(\pi(1+x^2))$. In Example 1.7.2(2) the resolvent measure is found to have the Lebesgue density $r(t,x) = (2\pi x)^{-1} G(t,x)$ where G is the function given in (1.7.2). We use the software package Mathematica to check that $r \notin L^1(\mathbb{R}_+ \times \mathbb{R})$. However, we do have that $r \in L^2(\mathbb{R}_+ \times \mathbb{R})$. Indeed, if

$$\tilde{f}(x) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}xz} f(z)\, \mathrm{d}z, \quad x \in \mathbb{R},$$

denotes the Fourier transform of a function $f \in L^2(\mathbb{R})$, we can use the fact that $\widetilde{f^{*n}} = \tilde{f}^n$ and $\tilde{f}(x) = (2\pi)^{-1/2} \mathrm{e}^{-|x|}$ for $f(x) = 1/(\pi(1+x^2))$ to calculate the Fourier transform of $r(t,\cdot)$ for fixed $t \in \mathbb{R}_+$:

$$\tilde{r}(t,x) = \sum_{n=1}^{\infty} \frac{(-t)^{n-1}}{(n-1)!} (\tilde{f}(x))^n = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-|x|} \mathrm{e}^{-t(2\pi)^{-1/2} \mathrm{e}^{-|x|}}, \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}.$$

Since Fourier transformation is unitary on $L^2(\mathbb{R})$, by Plancherel's theorem, we have

$$\begin{aligned} \int_{\mathbb{R}_+} \int_{\mathbb{R}} r(t,x)^2\, \mathrm{d}(t,x) &= \int_{\mathbb{R}_+} \int_{\mathbb{R}} \tilde{r}(t,x)^2\, \mathrm{d}(t,x) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}_+} \frac{\pi - \mathrm{e}^{-\sqrt{2/\pi}t}(\pi + \sqrt{2\pi}t)}{t^2}\, \mathrm{d}t = \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

□

1.5 Path properties

The classical OU process (1.1.2) has nice path properties: if W is Gaussian, the process has a continuous version; if W has jumps, the solution has a càdlàg version. In contrast, the notion of solution in Theorem 1.3.1 basically says nothing about the paths of a VOU process (apart from being measurable and elements of the set $\mathcal{L}(\mu)$ in (1.2.5)). The goal of this section is to fill in this gap, at least partially, and to prove the existence of versions with nice regularity for the VOU process. In the

presence of jumps, it turns out that tempo-spatial path properties are much harder to establish than for processes indexed by time. But before discussing this in detail, we first consider the case where the driving noise is Gaussian and mild conditions already ensure the existence of a Hölder continuous version.

Theorem 1.5.1. *Suppose that the conditions of Theorem 1.3.1 are satisfied with* $V \equiv 0$, $\nu \equiv 0$ *and that* X *is the solution to* (1.3.1) *given in* (1.3.3). *Further assume:*

- *There exists an exponent* $u > 0$ *such that for every* $T > 0$ *there is a non-negative constant* C_T *and we have*

$$\int_0^T \int_{\mathbb{R}^d} |g(s,y) - g(s+\tau, y+\xi)|^2 \, \mathrm{d}(s,y) \leq C_T |(\tau,\xi)|^u \tag{1.5.1}$$

  whenever $|(\tau,\xi)|$ *is sufficiently small.*

- *For some* $p > 1$ *we have for every* $T > 0$ *that*

$$\int_0^T \left(\int_{\mathbb{R}^d} g(s,y)^2 \, \mathrm{d}y \right)^p \mathrm{d}s < \infty. \tag{1.5.2}$$

Then the process X *has a version which is locally Hölder continuous with any exponent in* $(0, \frac{p-1}{4p} \wedge \frac{u}{4})$.

As usual for Gaussian processes, continuity of sample paths can be established under weaker conditions than those formulated in Theorem 1.5.1. Given that $g \in L^2_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ already implies that the left-hand side of (1.5.1) converges to 0 as $|(\tau,\xi)| \to 0$ and (1.5.2) holds with $p = 1$, the assumptions of Theorem 1.5.1 are reasonably general for practical purposes, so we do not pursue this direction further and only refer to [1]

If the noise does feature jumps, we are not able to construct path properties in general. Informally speaking, if the noise propagation function g is too irregular, it is unclear how a jump at a certain time and location affects the process at other times and locations. However, the situation is different if g is smooth enough. In this case, we are able to establish versions with the following regularity property.

Definition 1.5.2 A function $\Psi\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ is *t-càdlàg* if for every $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$,

$$\lim_{\substack{(\tilde{t},\tilde{x}) \to (t,x) \\ \tilde{t} \geq t}} \Psi(\tilde{t},\tilde{x}) = \Psi(t,x) \quad \text{and} \quad \lim_{\substack{(\tilde{t},\tilde{x}) \to (t,x) \\ \tilde{t} < t}} \Psi(\tilde{t},\tilde{x}) \quad \text{exists.}$$

□

In the following $\partial^\gamma g$ denotes the partial derivative $\partial_t^{\gamma_0}\partial_{x_1}^{\gamma_1}\cdots\partial_{x_d}^{\gamma_d} g$ for a function $g\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and a multi-index $\gamma = (\gamma_0, \ldots, \gamma_d) \in \mathbb{N}_0^{d+1}$.

Theorem 1.5.3. *Suppose that the conditions of Theorem 1.3.1 are satisfied with $V \equiv 0$ and that X is the solution to (1.3.1) given in (1.3.3). We assume that g is $(d+1)$-times continuously differentiable on $\mathbb{R}_+ \times \mathbb{R}^d$ such that for all multi-indices $\gamma \in \{0,1\}^{d+1}$ the partial derivative $\partial^\gamma g$ belongs to $L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$.*

If $\alpha < 1$, we further assume that there is a non-negative decreasing function $G\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $G(|x|)$ belongs to $L^\alpha(\mathbb{R}^d)$ and we have $|g(t,x)| \le C_T G(|x|)$ for all $T \ge 0$ and $(t,x) \in [0,T] \times \mathbb{R}^d$, where C_T is a non-negative constant depending on T.

Then the process X has a t-càdlàg version. This version is continuous if g additionally satisfies $g(0,x) = 0$ for all $x \in \mathbb{R}^d$.

The solution may display a fundamentally different path behavior if the underlying noise propagation function is not smooth. Here we only have results for the spatial dimension 1 and $g(t,x) = \mathbb{1}_{\{|x| \le ct\}} h(t,x)$ with some $c > 0$ and some smooth function h. This choice for g is motivated by Example 1.3.7 and, as we shall see in the proofs, enables us to utilize maximal inequalities of multi-parameter martingales as in [89].

In order to state our result, we introduce a partial order $\preceq$ on $\mathbb{R}_+ \times \mathbb{R}$ by setting $(t,x) \preceq (\tilde t, \tilde x)$ if $t \le \tilde t$ and $|\tilde x - x| \le c(\tilde t - t)$. As usual, we write $(t,x) \prec (\tilde t, \tilde x)$ if $(t,x) \preceq (\tilde t, \tilde x)$ and $(t,x) \ne (\tilde t, \tilde x)$. The following tempo-spatial càdlàg property is weaker than the t-càdlàg property of Definition 1.5.2.

Definition 1.5.4 A function $\Psi\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ is *$\preceq$-càdlàg* if for every $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$,

$$\lim_{\substack{(\tilde t,\tilde x)\to(t,x)\\(t,x)\preceq(\tilde t,\tilde x)}} \Psi(\tilde t, \tilde x) = \Psi(t,x) \quad \text{and} \quad \lim_{\substack{(\tilde t,\tilde x)\to(t,x)\\(t,x)\succ(\tilde t,\tilde x)}} \Psi(\tilde t, \tilde x) \quad \text{exists.}$$

□

Theorem 1.5.5. *Suppose that the conditions of Theorem 1.3.1 are satisfied with $d = 1$, $V \equiv 0$ and that X is the solution to (1.3.1) given in (1.3.3). Further let $h \in L^\alpha_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}) \cap L^\beta_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R})$ be twice continuously differentiable with partial derivatives $\partial_t h(t,x), \partial_x h(t,x)$ and $\partial_t \partial_x h(t,x)$ in $L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R})$. If the noise propagation function takes the form $g(t,x) = \mathbb{1}_{\{|x| \le ct\}} h(t,x)$ for some $c > 0$, then the process X has a $\preceq$-càdlàg version.*

Remark 1.5.6 (1) The conditions of Theorem 1.5.3 and Theorem 1.5.5 imply in particular that g is bounded. It is important to notice that the assertions of these theorems are false when g has a singularity at, for example, the origin.

(2) Under the assumptions of Theorem 1.5.5 and with virtually no change of its proof, we even have a version of the process X which is not only $\preceq$-càdlàg but also has limits from the flanks, that is, both limits

$$\lim_{\substack{(\tilde{t},\tilde{x})\to(t,x)\\ \tilde{x}>x,\, c|t-\tilde{t}|\le\tilde{x}-x}} X(\tilde{t},\tilde{x}) \quad \text{and} \quad \lim_{\substack{(\tilde{t},\tilde{x})\to(t,x)\\ \tilde{x}<x,\, c|t-\tilde{t}|\le x-\tilde{x}}} X(\tilde{t},\tilde{x}) \quad \text{exist.}$$

(3) There exist other notions of càdlàg sample paths for multi-parameter stochastic processes, see e.g. [68]. In contrast to the definition in that reference, our Definitions 1.5.2 and 1.5.4 take into account that time has a natural direction, while space has none.

(4) It suffices for Theorem 1.5.3 (resp. Theorem 1.5.5) in dimension $d = 1$ that, instead of being twice continuously differentiable, g (resp. h) is continuous on $\mathbb{R}_+ \times \mathbb{R}$ and that there exist $\beta \in [1,2]$ (resp. $\beta \in (1,2]$) and functions $g_1, g_2, g_{12} \in L^\beta_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R})$ (resp. $h_1, h_2, h_{12} \in L^\beta_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R})$) such that $\int_{\mathbb{R}} |z|^\beta \mathbb{1}_{\{|z|\le 1\}}\, \nu(\mathrm{d}z) < \infty$, Equation (1.6.8) (resp. Equation (1.6.12)) holds and, if $\alpha = 1$ in Theorem 1.5.3, that also $g_1, g_2, g_{12} \in L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R})$. For higher dimensions in Theorem 1.5.3, this comment applies analogously.

□

We apply the derived theorems to the VOU model considered in Examples 1.3.6 and 1.3.7.

Example 1.5.7 (Path properties for Examples 1.3.6 and 1.3.7)

(1) By induction we can show that the partial derivatives of the noise propagation function $g(t,x) = \mathrm{e}^{-\lambda'|x|}$ can be written as

$$\partial_{x_1} \dots \partial_{x_n} g(t,x) = \sum_{j=n}^{2n-1} c_j \mathrm{e}^{-\lambda'|x|} \frac{x_1 \dots x_n}{|x|^j},$$

for all $n \le d$ and constants c_j independent of (t,x). As a consequence, every partial derivative $\partial^\gamma g$, where γ is some multi-index in $\{0,1\}^{d+1}$, belongs to $L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ for every $0 < p < \frac{d}{d-1}$. For $\alpha < 1$ we may choose the function G from

Theorem 1.5.3 as $G(u) = \mathrm{e}^{-\lambda' u}$. Therefore, as soon as the Lévy measure ν of Λ satisfies (1.3.2) for some $\alpha > 0$ and $\beta \in (0, \frac{d}{d-1})$, Theorem 1.5.3 and Remark 1.5.6 imply that the unique solution X in (1.3.9) has a t-càdlàg version.

(2) Since the partial derivatives $\partial_t h(t,x), \partial_x h(t,x)$ and $\partial_t \partial_x h(t,x)$ of the function $h(t,x) = \mathrm{e}^{-\lambda'|x|}$ satisfy $\partial_t h(t,x) = 0, \partial_x h(t,x) = -\lambda' \mathrm{e}^{-\lambda'|x|} \frac{x}{|x|}$ and $\partial_t \partial_x h(t,x) = 0$ and lie in $L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ for every $p > 0$, Theorem 1.5.5 and Remark 1.5.6 apply to Example 1.3.7. Hence, we obtain a $\preceq$-càdlàg version for the process X in (1.3.12). □

1.6 Proofs

For the rest of this chapter, C_T denotes a real constant which may depend on $T \geq 0$ and change its value from line to line.

1.6.1 Proofs for Section 1.3

For the proofs of Theorems 1.3.1 and 1.3.3 we must guarantee that the stochastic convolution of g with Λ is well defined on $[0,T] \times \mathbb{R}^d$ or on the whole $\mathbb{R}_+ \times \mathbb{R}^d$, respectively. The conditions listed in Proposition 1.2.1 are necessary and sufficient, but may be too complicated to verify in general. The following lemma provides some simpler sufficient criteria.

Lemma 1.6.1. *Suppose that Λ is a homogeneous Lévy basis on $I \times \mathbb{R}^d$ with characteristics (b, σ^2, ν) and that $g\colon I \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function.*

(1) Condition (1) of Proposition 1.2.1 is satisfied if there exist measurable functions $k, K\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\left| b + \int_{\mathbb{R}} z \mathbb{1}_{\{|z| \in (1,A]\}} \, \nu(\mathrm{d}z) \right| = \mathcal{O}(K(A)), \quad A \to \infty, \tag{1.6.1}$$

$$\left| b - \int_{\mathbb{R}} z \mathbb{1}_{\{|z| \in (a,1]\}} \, \nu(\mathrm{d}z) \right| = \mathcal{O}(k(a)), \quad a \to 0, \quad \text{and} \tag{1.6.2}$$

$$\int_I \int_{\mathbb{R}^d} |g(t,x)| \Big(k(|g(t,x)|^{-1}) \mathbb{1}_{\{|g(t,x)|>1\}} + K(|g(t,x)|^{-1}) \mathbb{1}_{\{|g(t,x)| \leq 1\}} \Big) \, \mathrm{d}(t,x) < \infty. \tag{1.6.3}$$

(2) Condition (3) of Proposition 1.2.1 is satisfied if there is an increasing function $h\colon \mathbb{R}_+ \to \mathbb{R}_+$ *such that for some constant* $C \in \mathbb{R}_+$ *we have* $u^2 h(|x|) \le C h(u|x|)$ *for all* $u \in [0,1]$ *and* $x \in \mathbb{R}$*, and*

$$\int_I \int_{\mathbb{R}^d} h(|g(t,x)|)\,\mathrm{d}(t,x) < \infty \quad \text{and} \quad \int_{\mathbb{R}} h(|z|^{-1})^{-1}\,\nu(\mathrm{d}z) < \infty.$$

Proof. (1) The left-hand side of condition (1) of Proposition 1.2.1 is bounded by

$$\int_I \int_{\mathbb{R}^d} |g(t,x)| \left| b - \int_{\mathbb{R}} z \mathbb{1}_{\{|z| \in (|g(t,x)|^{-1},1]\}}\,\nu(\mathrm{d}z) \right| \mathbb{1}_{\{|g(t,x)|>1\}}\,\mathrm{d}(t,x) + \int_I \int_{\mathbb{R}^d} |g(t,x)| \left| b + \int_{\mathbb{R}} z \mathbb{1}_{\{|z| \in (1,|g(t,x)|^{-1}]\}}\,\nu(\mathrm{d}z) \right| \mathbb{1}_{\{|g(t,x)|\le 1\}}\,\mathrm{d}(t,x),$$

which, because of (1.6.1) and (1.6.2), is in turn bounded by (1.6.3).

(2) We divide the integral term in condition (3) of Proposition 1.2.1 into

$$J_1 + J_2 := (\mathrm{Leb}_{I\times\mathbb{R}^d} \otimes \nu)(\{|zg(t,x)| > 1\}) + \int_I \int_{\mathbb{R}^d} \int_{\mathbb{R}} |zg(t,x)|^2 \mathbb{1}_{\{|zg(t,x)|\le 1\}}\,\nu(\mathrm{d}z)\,\mathrm{d}(t,x).$$

For J_1 we can now use Markov's inequality to obtain

$$J_1 \le \int_{\mathbb{R}} h(|z|^{-1})^{-1}\,\nu(\mathrm{d}z) \int_I \int_{\mathbb{R}^d} h(|g(t,x)|)\,\mathrm{d}(t,x) < \infty,$$

while for J_2 the assumption that $|zg(t,x)|^2 \le C h(|g(t,x)|) h^{-1}(|z|^{-1})$ on $\{|zg(t,x)| \le 1\}$ implies

$$J_2 \le C \int_I \int_{\mathbb{R}^d} h(|g(t,x)|)\,\mathrm{d}(t,x) \int_{\mathbb{R}} h(|z|^{-1})^{-1}\,\nu(\mathrm{d}z) < \infty.$$

□

Possible choices for the functions h, k and K are $h(x) := x^q \mathbb{1}_{[0,1]}(x) + x^p \mathbb{1}_{(1,\infty)}(x)$, $k(a) := a^{1-p}$ and $K(A) := A^{1-q}$ with some $0 < p, q \le 2$.

Proof. (of Theorem 1.3.1) We wish to apply part (3) of Proposition 1.2.2 to the stochastic forcing function defined by

$$F(t,x) := V(t,x) + \int_0^t \int_{\mathbb{R}^d} g(t-s, x-y)\,\Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d.$$

This would yield the existence and uniqueness statement of Theorem 1.3.1. Since V has paths in $\mathcal{F}(\mu)$ by hypotheses, it suffices to prove that

$$Y(t,x) := \int_0^t \int_{\mathbb{R}^d} g(t-s, x-y)\,\Lambda(\mathrm{d}s, \mathrm{d}y)$$

is well defined for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ and that Y has a version with paths in $\mathcal{F}(\mu)$. The existence of the stochastic convolution is equivalent to the integrability of g with respect to Λ. But this follows from Lemma 1.6.1 with $h(x) := x^\alpha \mathbb{1}_{[0,1]}(x) + x^\beta \mathbb{1}_{(1,\infty)}(x)$, $k(a) := a^{1-\beta}$ and $K(A) := A^{1-\alpha}$ because we have $g \in L^\alpha_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d) \cap L^\beta_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ on the one hand, and (1.3.2) on the other hand. Note that the latter also implies (1.6.1) and (1.6.2) with our choices of k and K. Furthermore, from the well-definedness of Y we can already deduce the existence of a measurable version, see Theorem 1 of [66].

Next, we prove that this measurable version of Y belongs to $\mathcal{F}(\mu)$ almost surely. We begin with the case $\alpha = 1$ and notice that applying the Jensen and the Burkholder-Davis-Gundy inequalities yields

$$\begin{aligned}
\mathbb{E}[|Y(t,x)|] &\le |b| \int_0^t \int_{\mathbb{R}^d} |g(s,y)| \,\mathrm{d}(s,y) + \left(\int_0^t \int_{\mathbb{R}^d} \sigma^2 |g(s,y)|^2 \,\mathrm{d}(s,y) \right)^{1/2} \\
&\quad + C_T \mathbb{E}\left[\left(\int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}} |g(t-s,x-y)z|^2 \mathbb{1}_{\{|z|\le 1\}} \mathfrak{p}(\mathrm{d}s,\mathrm{d}y,\mathrm{d}z) \right)^{\beta/2} \right]^{1/\beta} \\
&\quad + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}} |g(t-s,x-y)z| \mathbb{1}_{\{|z|>1\}} \mathfrak{p}(\mathrm{d}s,\mathrm{d}y,\mathrm{d}z) \right] \\
&\le |b| \int_0^t \int_{\mathbb{R}^d} |g(s,y)| \,\mathrm{d}(s,y) + \left(\int_0^t \int_{\mathbb{R}^d} \sigma^2 |g(s,y)|^2 \,\mathrm{d}(s,y) \right)^{1/2} \\
&\quad + C_T \left(\int_0^t \int_{\mathbb{R}^d} |g(s,y)|^\beta \,\mathrm{d}(s,y) \int_{\mathbb{R}} |z|^\beta \mathbb{1}_{\{|z|\le 1\}} \nu(\mathrm{d}z) \right)^{1/\beta} \\
&\quad + \int_0^t \int_{\mathbb{R}^d} |g(s,y)| \,\mathrm{d}(s,y) \int_{\mathbb{R}} |z| \mathbb{1}_{\{|z|>1\}} \nu(\mathrm{d}z). \qquad (1.6.4)
\end{aligned}$$

Therefore, the function $(t,x) \mapsto \mathbb{E}[|Y(t,x)|]$ belongs to $L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$, so Lemma 1.6.4 and Fubini's theorem imply that Y has paths in $\mathcal{F}(\mu)$ almost surely.

If $\alpha \in (0,1)$, it is enough to prove $Y^1 \in \mathcal{F}(\mu)$ almost surely where Y^1 is defined in the same way as Y but with Λ replaced by its large jumps part $\Lambda^1(\mathrm{d}t,\mathrm{d}x) := \int_{\mathbb{R}} z \mathbb{1}_{\{|z|>1\}} \mathfrak{p}(\mathrm{d}t,\mathrm{d}x,\mathrm{d}z)$. For the convolution of g with $\Lambda - \Lambda^1$ the arguments as in the case $\alpha = 1$ would apply. Letting $\eta \in \{|\mu|, |\rho|, |\mu| * |\rho|, |\mu| * |\rho|^{*2}\}$, we prove that $\eta * |Y^1|$ exists almost everywhere. Since the realizations of Λ^1 are measures on $\mathbb{R}_+ \times \mathbb{R}^d$, Fubini's theorem yields

$$\begin{aligned}
&(\eta * |Y^1|)(t,x) \\
&\quad = \int_0^t \int_{\mathbb{R}^d} |Y^1(t-s,x-y)| \,\eta(\mathrm{d}s,\mathrm{d}y)
\end{aligned}$$

$$\begin{aligned} &\leq \int_0^t \int_{\mathbb{R}^d} \int_0^{t-s} \int_{\mathbb{R}^d} \int_{\mathbb{R}} |g(t-s-r, x-y-w)z| \mathbb{1}_{\{|z|>1\}} \mathfrak{p}(\mathrm{d}r, \mathrm{d}w, \mathrm{d}z)\, \eta(\mathrm{d}s, \mathrm{d}y) \\ &= \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}} (\eta * |g|)(t-s, x-y)|z| \mathbb{1}_{\{|z|>1\}} \mathfrak{p}(\mathrm{d}s, \mathrm{d}y, \mathrm{d}z). \end{aligned}$$

Now raising the last inequality to the power α, moving the exponent into the integral and taking expectation result in

$$\begin{aligned} &\mathbb{E}\big[\big((\eta * |Y^1|)(t,x)\big)^\alpha\big] \\ &\quad \leq \int_0^t \int_{\mathbb{R}^d} (\eta * |g|)^\alpha(s,y)\, \mathrm{d}(s,y) \int_{\mathbb{R}} |z|^\alpha \mathbb{1}_{\{|z|>1\}}\, \nu(\mathrm{d}z) \\ &\quad = \int_0^t \int_{\mathbb{R}^d} \big(\varphi(y)(\eta * |g|)(s,y)\big)^\alpha \varphi^{-\alpha}(y)\, \mathrm{d}(s,y) \int_{\mathbb{R}} |z|^\alpha \mathbb{1}_{\{|z|>1\}}\, \nu(\mathrm{d}z). \end{aligned}$$

Since $\varphi^{-\alpha} \in L^1(\mathbb{R}^d)$, the assertion that $Y^1 \in \mathcal{F}(\mu)$ almost surely is proved once we can show that $\varphi(\eta * |g|) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$. For $\eta = |\mu|$ this holds by assumption. For $\eta = |\rho|$ we use the fact that $\rho = \rho * \mu - \mu$, so we only need to prove $\varphi(|\rho| * (|\mu| * |g|)) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$. Notice that $\varphi\mu \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ implies $\varphi\rho \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ by the same arguments as in Proposition 1.2.2. More precisely, one has to work in the weighted measure spaces $M([0,T] \times \mathbb{R}^d; \varphi)$, $T \in \mathbb{R}_+$, which consist of all signed complete Borel measures μ such that $\varphi\mu \in M([0,T] \times \mathbb{R}^d)$ and are complete with the weighted total variation norm $\|\cdot\|_\varphi := \|\varphi \cdot\|$ (the proof of the temporal analogue, Theorem 4.3.4 of [49], can be extended to the tempo-spatial setting in a straightforward manner). Thus, the hypothesis that $\varphi(|\mu| * |g|) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ yields $\varphi(|\rho| * (|\mu| * |g|)) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ as well (like before, one can extend Theorem 4.3.5 in [49]). Finally, for $\eta = |\mu| * |\rho|^{*2}$ the same arguments apply because we have already established $\varphi(|\rho| * (|\mu| * |g|)) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$.

It remains to demonstrate that (1.3.3) defines a version of the solution to (1.3.1). To this end, we first observe that the solution in Proposition 1.2.2 takes the form

$$\begin{aligned} X(t,x) = V(t,x) - (\rho * V)(t,x) &+ \int_0^t \int_{\mathbb{R}^d} g(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y) \\ &- \int_0^t \int_{\mathbb{R}^d} \int_0^{t-s} \int_{\mathbb{R}^d} g(t-s-r, x-y-w)\, \Lambda(\mathrm{d}r, \mathrm{d}w)\, \rho(\mathrm{d}s, \mathrm{d}y). \end{aligned}$$

Formula (1.3.3) immediately follows if we can interchange the integrals with respect to Λ and ρ, that is, if we can apply a stochastic Fubini theorem. For the large jumps part Λ^1 of Λ the ordinary Fubini theorem is sufficient because the realizations of Λ^1 are true measures and integrability has already been shown in the proof for $|\rho| * |Y^1|$ above. For the remaining part $Y^2 := Y - Y^1$ Theorem 2 in [66] is applicable because,

by the same reasoning as in (1.6.4), we have that

$$\int_0^t \int_{\mathbb{R}^d} \mathbb{E}\left[\left(\int_0^{t-s} \int_{\mathbb{R}^d} \int_{\mathbb{R}} |g(t-s-r, x-y-w)z|^2 \mathbb{1}_{\{|z|\le 1\}}\, \mathfrak{p}(\mathrm{d}r, \mathrm{d}w, \mathrm{d}z)\right)^{1/2}\right] |\rho|(\mathrm{d}s, \mathrm{d}y)$$

is finite. □

Proof. (of Theorem 1.3.3) Our first observation is that X_∞ is well defined because $g - \rho * g$ is integrable with respect to Λ on $\mathbb{R}_+ \times \mathbb{R}^d$. This in turn is a consequence of assumptions (2) and (3) together with Lemma 1.6.1 (and of course, that $\beta = 2$ if $\sigma^2 \neq 0$). Next, regarding the convergence statement, it suffices by Slutsky's theorem and hypothesis (1) to prove the convergence of the finite-dimensional distributions of the stochastic convolution in (1.3.3) when time tends to infinity. For one spatial point $x \in \mathbb{R}^d$, the claim readily follows from

$$\begin{aligned}
\int_0^t \int_{\mathbb{R}^d} (g - \rho * g)(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y) &\overset{\mathrm{d}}{=} \int_0^t \int_{\mathbb{R}^d} (g - \rho * g)(s, x-y)\, \Lambda(t - \mathrm{d}s, \mathrm{d}y) \\
&\overset{\mathrm{d}}{=} \int_0^t \int_{\mathbb{R}^d} (g - \rho * g)(s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y) \\
&\xrightarrow{\mathrm{d}} \int_0^\infty \int_{\mathbb{R}^d} (g - \rho * g)(s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y).
\end{aligned}$$

The n-dimensional case can be treated completely analogously. For the second part of the theorem, we suppose that also g is integrable with respect to Λ and that V is a process independent of Λ and with the same finite-dimensional distributions as the process given in (1.3.4). Then by the stochastic Fubini theorem (cf. the proof of Theorem 1.3.1) we obtain

$$\begin{aligned}
&V(t,x) - (\rho * V)(t,x) \\
&\quad \overset{d}{=} \int_{-\infty}^0 \int_{\mathbb{R}^d} g(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y) \\
&\qquad - \int_0^t \int_{\mathbb{R}^d} \int_{-\infty}^0 \int_{\mathbb{R}^d} g(t-s-r, x-y-w)\, \Lambda(\mathrm{d}r, \mathrm{d}w)\, \rho(\mathrm{d}s, \mathrm{d}y) \\
&\quad \overset{d}{=} \int_{-\infty}^0 \int_{\mathbb{R}^d} g(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y) - \int_{-\infty}^0 \int_{\mathbb{R}^d} (\rho * g)(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y).
\end{aligned}$$

Again, the reader can convince herself that the previous calculations also apply to n time and space points. The strict stationarity of X is now a consequence of that of the process (1.3.5). □

Proof. (of Lemma 1.3.4) (1) We have already used this tacitly in the proof of Theorem 1.3.1. In fact, if $\alpha \in (0,1]$, then

$$\begin{aligned}\left|b + \int_{\mathbb{R}} z\mathbb{1}_{\{|z|\in(1,A]\}}\,\nu(\mathrm{d}z)\right| &\le |b| + \int_{\mathbb{R}} |z|^{\alpha}|z|^{1-\alpha}\mathbb{1}_{\{|z|\in(1,A]\}}\,\nu(\mathrm{d}z)\\ &\le |b| + A^{1-\alpha}\int_{\mathbb{R}} |z|^{\alpha}\mathbb{1}_{\{|z|>1\}}\,\nu(\mathrm{d}z),\end{aligned}$$

while for $\beta \in [1,2]$ we have

$$\begin{aligned}\left|b - \int_{\mathbb{R}} z\mathbb{1}_{\{|z|\in(a,1]\}}\,\nu(\mathrm{d}z)\right| &\le |b| + \int_{\mathbb{R}} |z|^{\beta}|z|^{1-\beta}\mathbb{1}_{\{|z|\in(a,1]\}}\,\nu(\mathrm{d}z)\\ &\le |b| + a^{1-\beta}\int_{\mathbb{R}} |z|^{\beta}\mathbb{1}_{\{|z|\le 1\}}\,\nu(\mathrm{d}z).\end{aligned}$$

(2) If $\alpha \in (1,2]$ and $b_1 = 0$, then we have for $A \ge 1$ that

$$\begin{aligned}\left|b + \int_{\mathbb{R}} z\mathbb{1}_{\{|z|\in(1,A]\}}\,\nu(\mathrm{d}z)\right| &\le \int_{\mathbb{R}} |z|\mathbb{1}_{\{|z|\in(A,\infty)\}}\,\nu(\mathrm{d}z) = \int_{\mathbb{R}} |z|^{\alpha}|z|^{1-\alpha}\mathbb{1}_{\{|z|\in(A,\infty)\}}\,\nu(\mathrm{d}z)\\ &\le A^{1-\alpha}\int_{\mathbb{R}} |z|^{\alpha}\mathbb{1}_{\{|z|>1\}}\,\nu(\mathrm{d}z).\end{aligned}$$

If $\beta \in (0,1)$ and $b_0 = 0$, then

$$\begin{aligned}\left|b - \int_{\mathbb{R}} z\mathbb{1}_{\{|z|\in(a,1]\}}\,\nu(\mathrm{d}z)\right| &\le \int_{\mathbb{R}} |z|\mathbb{1}_{\{|z|\in(0,a]\}}\,\nu(\mathrm{d}z) = \int_{\mathbb{R}} |z|^{\beta}|z|^{1-\beta}\mathbb{1}_{\{|z|\in(0,a]\}}\,\nu(\mathrm{d}z)\\ &\le a^{1-\beta}\int_{\mathbb{R}} |z|^{\beta}\mathbb{1}_{\{|z|\le 1\}}\,\nu(\mathrm{d}z).\end{aligned}$$

(3) If Λ is symmetric, then the left-hand sides of (1.6.1) and (1.6.2) are identically zero. □

Proof. (of Lemma 1.3.5) (1) follows from Lemma 1.6.4, (2) holds because $\rho(\mathbb{R}_+ \times \mathbb{R}^d) = 1$ implies that $g - \rho * g = (g - g_\infty) - \rho * (g - g_\infty)$, and (3) is simply Young's inequality. □

1.6.2 Proofs for Section 1.4

Proof. (of Proposition 1.4.1) That $m[X] < \infty$ almost surely can be verified in a similar way to Theorem 1.3.1. Also by essentially the same arguments given there, the stochastic Fubini theorem is applicable for $m[X]$ and the result follows from Proposition 1.2.1. □

Proof. (of Proposition 1.4.5) (1) By assumption the function $h\colon \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}$ defined by $h(t,x) = (g - \rho * g)(t,x)\mathbb{1}_{\mathbb{R}_+}(t)$ belongs to $L^1(\mathbb{R}\times\mathbb{R}^d)$. Therefore, the claim follows from Young's inequality and the observation that, up to a multiplicative constant, $\operatorname{Cov}[X(t,x), X(t+\tau, x+\xi)]$ equals the convolution of h with h^- where $h^-(t,x) = h(-t,-x)$.

(2) That X is well defined and has a finite second moment, follows from $b_1 = 0$ and $g - \rho * g \in L^2(\mathbb{R}_+\times\mathbb{R}^d)$. Moreover, as $g - \rho * g$ does not change signs, we have

$$\begin{aligned}
&\int_0^\infty \int_{\mathbb{R}^d} \left| \int_0^\infty \int_{\mathbb{R}^d} (g-\rho*g)(s,y)(g-\rho*g)(s+\tau,y+\xi)\,\mathrm{d}(s,y) \right| \mathrm{d}(\tau,\xi) \\
&\quad = \int_0^\infty \int_{\mathbb{R}^d} \int_0^\infty \int_{\mathbb{R}^d} |g-\rho*g|(s,y)|g-\rho*g|(s+\tau,y+\xi)\,\mathrm{d}(s,y)\,\mathrm{d}(\tau,\xi) \\
&\quad = \int_0^\infty \int_{\mathbb{R}^d} |g-\rho*g|(s,y) \left(\int_0^\infty \int_{\mathbb{R}^d} |g-\rho*g|(s+\tau,y+\xi)\,\mathrm{d}(\tau,\xi) \right) \mathrm{d}(s,y).
\end{aligned}$$

Since $g - \rho * g \notin L^1(\mathbb{R}_+\times\mathbb{R}^d)$, the inner integral is infinite for all (s,y), so the whole integral is infinite as well. This shows that $(\tau,\xi) \mapsto |\operatorname{Cov}[X(t,x), X(t+\tau,x+\xi)]|$ is not an element of $L^1(\mathbb{R}_+\times\mathbb{R}^d)$. □

1.6.3 Proofs for Section 1.5

Proof. (of Theorem 1.5.1) For $\tilde{g} := g - \rho * g$ we have that

$$\int_0^t \int_{\mathbb{R}^d} \tilde{g}(t-s, x-y)\,\mathrm{d}(s,y) = \int_0^t \int_{\mathbb{R}^d} \tilde{g}(s,y)\,\mathrm{d}(s,y)$$

is continuous in (t,x). Hence, we may assume without loss of generality that $b = 0$. The additional conditions on g in Theorem 1.5.1, together with Hölder's inequality and Fubini's theorem, imply for $\rho * g$ (we extend g on the negative half space $(-\infty, 0)\times\mathbb{R}^d$ by zero):

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} |(\rho*g)(s,y) - (\rho*g)(s+\tau,y+\xi)|^2\,\mathrm{d}(s,y) \\
&\quad = \int_0^T \int_{\mathbb{R}^d} \left| \int_0^s \int_{\mathbb{R}^d} g(s-r,y-z)\,\rho(\mathrm{d}r,\mathrm{d}z) \right. \\
&\qquad \left. - \int_0^{s+\tau} \int_{\mathbb{R}^d} g(s+\tau-r,y+\xi-z)\,\rho(\mathrm{d}r,\mathrm{d}z) \right|^2 \mathrm{d}(s,y) \\
&\quad \le C_T \int_0^T \int_{\mathbb{R}^d} \int_0^T \int_{\mathbb{R}^d} |g(s-r,y-z) - g(s+\tau-r,y+\xi-z)|^2\,\mathrm{d}(s,y)\,|\rho|(\mathrm{d}r,\mathrm{d}z)
\end{aligned}$$

$$
\begin{aligned}
&+ C_T \int_0^T \int_{\mathbb{R}^d} \int_0^T \int_{\mathbb{R}^d} \left| \mathbb{1}_{[s,s+\tau]}(r) g(s+\tau-r, y+\xi-z) \right|^2 \mathrm{d}(s,y)\, |\rho|(\mathrm{d}r, \mathrm{d}z) \\
&\leq C_T \int_0^T \int_{\mathbb{R}^d} |(\tau,\xi)|^u \, |\rho|(\mathrm{d}r, \mathrm{d}z) + C_T \int_0^\tau \int_{\mathbb{R}^d} |g(s,y)|^2 \, \mathrm{d}(s,y) \\
&\leq C_T |(\tau,\xi)|^u + C_T \left(\int_0^\tau 1 \, \mathrm{d}s \right)^{\frac{p-1}{p}} \left(\int_0^\tau \left(\int_{\mathbb{R}^d} g(s,y)^2 \, \mathrm{d}y \right)^p \mathrm{d}s \right)^{\frac{1}{p}} \\
&\leq C_T |(\tau,\xi)|^u + C_T |(\tau,\xi)|^{\frac{p-1}{p}}, \qquad (1.6.5)
\end{aligned}
$$

where $|(\tau,\xi)|$ is small enough. Furthermore, we have by another application of Hölder's inequality and Fubini's theorem

$$
\begin{aligned}
&\int_0^T \left(\int_{\mathbb{R}^d} (\rho * g)(s,y)^2 \, \mathrm{d}y \right)^p \mathrm{d}s \\
&\quad \leq C_T \int_0^T \left(\int_{\mathbb{R}^d} \int_0^s \int_{\mathbb{R}^d} g(s-r, y-z)^2 \, |\rho|(\mathrm{d}r, \mathrm{d}z)\, \mathrm{d}y \right)^p \mathrm{d}s \\
&\quad \leq C_T \int_0^T \int_0^s \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(s-r,y)^2 \, \mathrm{d}y \right)^p |\rho|(\mathrm{d}r, \mathrm{d}z)\, \mathrm{d}s \\
&\quad \leq C_T \int_0^T \int_{\mathbb{R}^d} \int_0^T \left(\int_{\mathbb{R}^d} g(s,y)^2 \, \mathrm{d}y \right)^p \mathrm{d}s\, |\rho|(\mathrm{d}r, \mathrm{d}z) < \infty. \qquad (1.6.6)
\end{aligned}
$$

Next, by Corollary 1.4.2 we have for all $(t,x), (\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R}^d$ that

$$
\begin{aligned}
&\mathbb{E}\Big[|X(t,x) - X(t+\tau, x+\xi)|^2 \Big] \\
&\quad = \sigma^2 \int_0^t \int_{\mathbb{R}^d} \tilde{g}^2(s,y) \, \mathrm{d}(s,y) - 2\sigma^2 \int_0^t \int_{\mathbb{R}^d} \tilde{g}(s,y) \tilde{g}(s+\tau, y+\xi) \, \mathrm{d}(s,y) \\
&\qquad + \sigma^2 \int_0^{t+\tau} \int_{\mathbb{R}^d} \tilde{g}^2(s,y) \, \mathrm{d}(s,y) \\
&\quad = 2\sigma^2 \int_0^t \int_{\mathbb{R}^d} \tilde{g}(s,y) [\tilde{g}(s,y) - \tilde{g}(s+\tau, y+\xi)] \, \mathrm{d}(s,y) \\
&\qquad + \sigma^2 \int_t^{t+\tau} \int_{\mathbb{R}^d} \tilde{g}^2(s,y) \, \mathrm{d}(s,y).
\end{aligned}
$$

The assumptions on g and the inequalities (1.6.5) and (1.6.6) yield

$$
\int_0^t \int_{\mathbb{R}^d} \tilde{g}(s,y) [\tilde{g}(s,y) - \tilde{g}(s+\tau, y+\xi)] \, \mathrm{d}(s,y) \leq C_T |(\tau,\xi)|^{u/2} + C_T |(\tau,\xi)|^{\frac{p-1}{2p}}
$$

for small $|(\tau,\xi)|$ and

$$
\begin{aligned}
\int_t^{t+\tau} \int_{\mathbb{R}^d} \tilde{g}^2(s,y) \, \mathrm{d}(s,y) &\leq \left(\int_t^{t+\tau} 1 \, \mathrm{d}s \right)^{\frac{p-1}{p}} \left(\int_t^{t+\tau} \left(\int_{\mathbb{R}^d} \tilde{g}(s,y)^2 \, \mathrm{d}y \right)^p \mathrm{d}s \right)^{\frac{1}{p}} \\
&\leq C_T |(\tau,\xi)|^{\frac{p-1}{p}},
\end{aligned}
$$

where we have used the Cauchy-Schwarz inequality in the first and Hölder's inequality in the second step. Adding both inequalities together and using the fact that X is a Gaussian process, Kolmogorov's continuity theorem (see e.g. Theorem 3.23 of [57]) finishes the proof. □

Proof. (of Theorem 1.5.3) We prove the case when space is one-dimensional, i.e. $d = 1$. For higher dimensions the proof is similar. Clearly, it suffices to show the path property separately for the drift and Gaussian part, the compensated small jumps part, and the large jumps part.

Case 1: $\Lambda(\mathrm{d}t, \mathrm{d}x) = b\,\mathrm{d}(t,x) + \sigma W(\mathrm{d}t, \mathrm{d}x)$.

The assumptions on g imply that $g|_{[0,T]\times\mathbb{R}^d}$ belongs to the Sobolev space $W^{1,2}([0,T]\times\mathbb{R}^d)$ for all $T \in \mathbb{R}_+$ if $\sigma \neq 0$. Therefore, Theorem 3 in Section 5.8 of [43], a characterization of the Sobolev space $W^{1,2}([0,T]\times\mathbb{R}^d)$, ensures the first condition in Theorem 1.5.1. Moreover, since $g, \partial_t g \in L^2_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R})$, the fundamental theorem of calculus yields that $s \mapsto \int_{\mathbb{R}^d} g^2(s,y)\,\mathrm{d}y$ is continuous in s. Hence, by Theorem 1.5.1, X has a continuous version.

Case 2: $\Lambda(\mathrm{d}t, \mathrm{d}x) = \int_{\mathbb{R}} z\mathbb{1}_{\{|z|\leq 1\}}\,(\mathfrak{p}-\mathfrak{q})(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z)$.

We define $\Lambda_n(\mathrm{d}t, \mathrm{d}x) := \int_{\mathbb{R}} z\mathbb{1}_{\{1/n\leq|z|\leq 1\}}\mathbb{1}_{\{|x|\leq n\}}\,(\mathfrak{p}-\mathfrak{q})(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z)$. Since Λ_n has only finitely many jumps on $[0,T]\times\mathbb{R}$ almost surely, it is easy to see that the paths of $F_n := g * \Lambda_n$ are almost surely t-càdlàg and in $L^\infty_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ due to the boundedness of g. Hence, also the process $\rho * F_n$ has almost surely t-càdlàg realizations by dominated convergence. We now show that $\rho * F^n := \rho * F - \rho * F_n$ converges uniformly on compacts in probability to 0. To this end, consider for any $T > 0$ and $U = [-K, K] \subseteq \mathbb{R}$

$$\begin{aligned}
\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |(\rho * F^n)(t,x)|\right] &= \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}\left|\int_0^t\int_{\mathbb{R}} F^n(t-s, x-y)\,\rho(\mathrm{d}s, \mathrm{d}y)\right|\right] \\
&\leq \int_0^T\int_{\mathbb{R}} \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |F^n(t-s, x-y)|\right] |\rho|(\mathrm{d}s, \mathrm{d}y) \\
&= \int_0^T\int_{\mathbb{R}} \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |F^n(t, x-y)|\right] |\rho|(\mathrm{d}s, \mathrm{d}y).
\end{aligned} \tag{1.6.7}$$

Now use the fundamental theorem of calculus to decompose g as

$$\begin{aligned} g(t-s,x-\eta-y) &= g(0,x-\eta-y) + \int_s^t g_1(r-s,x-\eta-y)\,\mathrm{d}r \\ &= g(0,-\eta-y) + \int_{-\eta}^{x-\eta} g_2(0,z-y)\,\mathrm{d}z + \int_s^t g_1(r-s,-\eta-y)\,\mathrm{d}r \\ &\quad + \int_s^t \int_{-\eta}^{x-\eta} g_{12}(r-s,z-y)\,\mathrm{d}z\,\mathrm{d}r, \end{aligned} \tag{1.6.8}$$

where $\eta \in \mathbb{R}$, $g_1(t,x) = \partial_t g(t,x)$, $g_2(t,x) = \partial_x g(t,x)$ and $g_{12}(t,x) = \partial_t \partial_x g(t,x)$. With the same reasoning as in the proof of Theorem 1.3.1, the assumptions imply that the stochastic Fubini theorem is applicable and this gives us

$$\begin{aligned} F^n(t,x-\eta) &:= \int_0^t \int_{\mathbb{R}} g(t-s,x-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y) \\ &= \int_0^t \int_{\mathbb{R}} g(0,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y) + \int_0^t \int_{\mathbb{R}} \int_{-\eta}^{x-\eta} g_2(0,z-y)\,\mathrm{d}z\,\Lambda^n(\mathrm{d}s,\mathrm{d}y) \\ &\quad + \int_0^t \int_{\mathbb{R}} \int_s^t g_1(r-s,-\eta-y)\,\mathrm{d}r\,\Lambda^n(\mathrm{d}s,\mathrm{d}y) \\ &\quad + \int_0^t \int_{\mathbb{R}} \int_s^t \int_{-\eta}^{x-\eta} g_{12}(r-s,z-y)\,\mathrm{d}z\,\mathrm{d}r\,\Lambda^n(\mathrm{d}s,\mathrm{d}y) \\ &= \int_0^t \int_{\mathbb{R}} g(0,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y) + \int_{-\eta}^{x-\eta} \int_0^t \int_{\mathbb{R}} g_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}z \\ &\quad + \int_0^t \int_0^r \int_{\mathbb{R}} g_1(r-s,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}r \\ &\quad + \int_0^t \int_{-\eta}^{x-\eta} \int_0^r \int_{\mathbb{R}} g_{12}(r-s,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}z\,\mathrm{d}r \\ &=: I_{1,n}(t,x,\eta) + I_{2,n}(t,x,\eta) + I_{3,n}(t,x,\eta) + I_{4,n}(t,x,\eta), \end{aligned}$$

where $\Lambda^n := \Lambda - \Lambda_n$. Therefore, we have for fixed $\eta \in \mathbb{R}$

$$\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |F^n(t,x-\eta)|\right] \le \sum_{j=1}^4 \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |I_{j,n}(t,x,\eta)|\right].$$

Since $I_{1,n}(t,x)$ does not depend on x and is a martingale in t, we have by the Burkholder-Davis-Gundy inequalities that

$$\begin{aligned} &\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |I_{1,n}(t,x,\eta)|^2\right] \\ &\quad = \mathbb{E}\Bigg[\sup_{t\in[0,T]} \Bigg| \int_0^t \int_{\mathbb{R}} \int_{\mathbb{R}} g(0,-\eta-y) z \Big(\mathbb{1}_{\{|y|>n\}} \mathbb{1}_{\{|z|\le 1\}} \end{aligned}$$

$$
\begin{aligned}
&\quad + \mathbb{1}_{\{|y|\le n\}}\mathbb{1}_{\{|z|<1/n\}}\Big)(\mathfrak{p}-\mathfrak{q})(\mathrm{d}s,\mathrm{d}y,\mathrm{d}z)\Big|^2\Big] \\
&\le C_T\mathbb{E}\Bigg[\int_0^T\int_{\mathbb{R}}\int_{\mathbb{R}}|g|^2(0,-\eta-y)|z|^2 \\
&\qquad\qquad \times\left(\mathbb{1}_{\{|y|>n\}}\mathbb{1}_{\{|z|\le 1\}}+\mathbb{1}_{\{|y|\le n\}}\mathbb{1}_{\{|z|<1/n\}}\right)\mathfrak{p}(\mathrm{d}s,\mathrm{d}y,\mathrm{d}z)\Bigg] \\
&= C_T\int_0^T\int_{\mathbb{R}}|g|^2(0,-\eta-y)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}(s,y)\int_{\mathbb{R}}|z|^2\mathbb{1}_{\{|z|\le 1\}}\,\nu(\mathrm{d}z) \\
&\quad + C_T\int_0^T\int_{\mathbb{R}}|g|^2(0,-\eta-y)\mathbb{1}_{\{|y|\le n\}}\,\mathrm{d}(s,y)\int_{\mathbb{R}}|z|^2\mathbb{1}_{\{|z|<1/n\}}\,\nu(\mathrm{d}z)\to 0 \qquad (1.6.9)
\end{aligned}
$$

as $n\to\infty$ and that $\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}|I_{1,n}(t,x,\eta)|^2\right]$ is bounded in n and η. Here we have used that $g\in L^2_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R})$ since it is continuous on $\mathbb{R}_+\times\mathbb{R}$ and belongs to $L^\alpha_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R})$. By similar arguments, we have that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}|I_{2,n}(t,x,\eta)|^2\right] \\
&\quad = \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}\left|\int_{-\eta}^{x-\eta}\int_0^t\int_{\mathbb{R}}g_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}z\right|^2\right] \\
&\quad \le C_T\int_{-K-\eta}^{K-\eta}\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}\left|\int_0^t\int_{\mathbb{R}}g_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\right|^2\right]\mathrm{d}z \\
&\quad \le C_T\int_{-K-\eta}^{K-\eta}\int_0^T\int_{\mathbb{R}}|g_2|^2(0,z-y)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}(s,y)\int_{\mathbb{R}}|\zeta|^2\mathbb{1}_{\{|\zeta|\le 1\}}\,\nu(\mathrm{d}\zeta)\,\mathrm{d}z \\
&\qquad + C_T\int_{-K-\eta}^{K-\eta}\int_0^T\int_{\mathbb{R}}|g_2|^2(0,z-y)\mathbb{1}_{\{|y|\le n\}}\,\mathrm{d}(s,y)\int_{\mathbb{R}}|\zeta|^2\mathbb{1}_{\{|\zeta|<1/n\}}\,\nu(\mathrm{d}\zeta)\,\mathrm{d}z \\
&\quad \le C_T\int_0^T\int_{\mathbb{R}}|g_2|^2(0,y)\mathbb{1}_{\{|y|>n-K-|\eta|\}}\,\mathrm{d}(s,y)\int_{\mathbb{R}}|\zeta|^2\mathbb{1}_{\{|\zeta|\le 1\}}\,\nu(\mathrm{d}\zeta) \\
&\qquad + C_T\int_0^T\int_{\mathbb{R}}|g_2|^2(0,y)\,\mathrm{d}(s,y)\int_{\mathbb{R}}|\zeta|^2\mathbb{1}_{\{|\zeta|<1/n\}}\,\nu(\mathrm{d}\zeta)\to 0
\end{aligned}
$$

as $n\to 0$ and that $\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}|I_{2,n}(t,x,\eta)|^2\right]$ is bounded in n and η. Because $I_{3,n}$ and $I_{4,n}$ can be treated analogously to $I_{2,n}$, $\rho * F_n$ converges uniformly on compacts in probability to $\rho * F$ due to (1.6.7) and dominated convergence. This gives us a t-càdlàg version of $\rho * F$. By setting $\eta = 0$, we also obtain that F_n converges uniformly on compacts in probability to F. As a consequence, X has a t-càdlàg version.

Case 3: $\Lambda(\mathrm{d}t,\mathrm{d}x) = \int_{\mathbb{R}} z\mathbb{1}_{\{|z|>1\}}\,\mathfrak{p}(\mathrm{d}t,\mathrm{d}x,\mathrm{d}z)$.

Here we assume $\alpha < 1$ because in the situation $\alpha = 1$ we can split the Lévy basis according to

$$\Lambda(\mathrm{d}t, \mathrm{d}x) = \int_{\mathbb{R}} z \mathbb{1}_{\{|z|>1\}} (\mathfrak{p} - \mathfrak{q})(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z) + \int_{\mathbb{R}} z \mathbb{1}_{\{|z|>1\}} \mathfrak{q}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z),$$

treating the first summand as in Case 2 and the second summand as in Case 1. We consider the truncated Lévy basis $\Lambda_n(\mathrm{d}t, \mathrm{d}x) := \int_{\mathbb{R}} z \mathbb{1}_{\{|z|\geq 1\}} \mathbb{1}_{\{|x|\leq n\}} \mathfrak{p}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}z)$, which almost surely has finitely many jumps on $[0,T]\times\mathbb{R}$. As in Case 2, the processes $F_n = g * \Lambda_n$ and $\rho * F_n$ have t-càdlàg paths almost surely. It suffices therefore to prove that they converge uniformly on compacts in probability to F and $\rho * F$, respectively. We can estimate (note that we can interchange convolution with Λ and convolution with ρ as shown in Theorem 1.3.1)

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} |(\rho * F - \rho * F_n)(t,x)|^\alpha\right] \\
&\quad = \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U} \left|\int_0^t \int_{\mathbb{R}} (\rho * g)(t-s, x-y)\, (\Lambda - \Lambda_n)(\mathrm{d}s, \mathrm{d}y)\right|^\alpha\right] \\
&\quad \leq \int_0^T \int_{\mathbb{R}} \sup_{(t,x)\in[0,T]\times U} |(\rho * g)|^\alpha (t-s, x-y) \mathbb{1}_{\{|y|>n\}}\, \mathrm{d}(s,y) \int_{\mathbb{R}} |z|^\alpha \mathbb{1}_{\{|z|\geq 1\}}\, \nu(\mathrm{d}z) \\
&\quad \leq C_T \int_{\mathbb{R}} \sup_{(t,x)\in[0,T]\times U} |(\rho * g)|^\alpha (t, x-y) \mathbb{1}_{\{|y|>n\}}\, \mathrm{d}y \\
&\quad = C_T \int_{\mathbb{R}} \sup_{(t,x)\in[0,T]\times U} |\varphi(-y)(\rho * g)(t, x-y)|^\alpha\, \mathbb{1}_{\{|y|>n\}} \varphi(-y)^{-\alpha}\, \mathrm{d}y,
\end{aligned}$$

where φ is the function from Theorem 1.3.1. Since $\varphi^{-\alpha} \in L^1(\mathbb{R})$, the right-hand side of the last inequality tends to zero by dominated convergence if we can show that

$$\sup_{(t,x)\in[0,T]\times U} \varphi(-y)|\rho * g|(t, x-y)$$

is bounded in y. But this follows because for every $y \in \mathbb{R}$ we have

$$\begin{aligned}
\sup_{(t,x)\in[0,T]\times U} \varphi(-y)|\rho * g|(t, x-y) &\leq \sup_{(t,x)\in[0,T]\times U} \varphi(x-y)|\rho * g|(t, x-y)\varphi(-x) \\
&\leq C_T \sup_{x\in U} \varphi(-x) < \infty,
\end{aligned}$$

where we have used that $\varphi(|\rho| * |g|) \in L^\infty_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R})$ as shown in the proof of Theorem 1.3.1 and that φ is submultiplicative and locally bounded. Similar arguments

applied to the pair F_n and F yield

$$\begin{aligned}\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}|(F-F_n)(t,x)|^\alpha\right] &\leq C_T\int_{\mathbb{R}}\sup_{(t,x)\in[0,T]\times U}|g|^\alpha(t,x-y)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y\\ &\leq C_T\int_{\mathbb{R}}\sup_{x\in U}G^\alpha(|x-y|)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y\\ &\leq C_T\int_{\mathbb{R}}G^\alpha((|y|-K)\vee 0)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y, \qquad (1.6.10)\end{aligned}$$

where we used the monotonicity of G. Now the last line goes to 0 by dominated convergence because $G(|x|)\in L^\alpha(\mathbb{R})$. Altogether we obtain a t-càdlàg version of X. □

For the proof of Theorem 1.5.5 we need to resort to maximal inequalities for multi-parameter martingales in line with [59, 89]. To this end, let $\leq$ denote the partial order on $\mathbb{R}^2$ such that $v=(v_1,v_2)\leq w=(w_1,w_2)$ if and only if $v_1\leq w_1$ and $v_2\leq w_2$. For two subsets I_1 and I_2 of $\mathbb{R}$ set $I=I_1\times I_2$. Now a family of sub-σ-algebras $\mathbb{G}=(\mathcal{G}(v))_{v\in I}$ is called a filtration if $\mathcal{G}(v)\subseteq\mathcal{G}(w)$ for all $v\leq w$ in I. A stochastic process X indexed by I is called a *martingale* with respect $\mathbb{G}$ if X is adapted to $\mathbb{G}$, $X(v)$ is integrable for all $v\in I$ and $\mathbb{E}[X(w)\mid\mathcal{G}(v)]=X(v)$ for all $v\leq w$ in I. Furthermore, we define the marginal filtrations $\mathbb{G}^1=(\mathcal{G}^1(v_1))_{v_1\in I_1}$ and $\mathbb{G}^2=(\mathcal{G}^2(v_2))_{v_2\in I_2}$ through $\mathcal{G}^1(v_1):=\bigvee_{\xi\in I_2}\mathcal{G}(v_1,\xi)$ and $\mathcal{G}^2(v_2):=\bigvee_{\xi\in I_1}\mathcal{G}(\xi,v_2)$ and set $\mathcal{G}^*(v_1,v_2):=\mathcal{G}^1(v_1)\vee\mathcal{G}^2(v_2)$. Then a martingale X is called an *orthomartingale* if for each $(i,j)\in\{(1,2),(2,1)\}$ and each fixed $v_i\in I_i$, $v_j\mapsto X(v)$ is a one-parameter martingale with respect to $\mathbb{G}^j$. Moreover, a martingale X is called a *strong martingale* if it satisfies the condition $\mathbb{E}[X((v_1,v_2),(w_1,w_2)]\mid\mathcal{G}^*(v_1,v_2)]=0$ for all $(v_1,v_2)\leq(w_1,w_2)$ in I, where $X((v_1,v_2),(w_1,w_2)]:=X(w_1,w_2)-X(w_1,v_2)-X(v_1,w_2)+X(v_1,v_2)$ is the two-dimensional increment. Further notation includes $[v,w]_\leq:=\{u\in\mathbb{R}^2\colon v\leq u\leq w\}$ for the closed interval from v to w with respect to the partial order $\leq$. Similarly, for $(t,x),(\tilde{t},\tilde{x})\in\mathbb{R}_+\times\mathbb{R}$, $[(t,x),(\tilde{t},\tilde{x})]_\leq:=\{(s,y)\in\mathbb{R}_+\times\mathbb{R}\colon(t,x)\leq(s,y)\leq(\tilde{t},\tilde{x})\}$ denotes the closed interval from (t,x) to $(\tilde{t},\tilde{x})$ with respect to the partial order $\leq$ as defined in Section 1.5. Also, we use the abbreviations $A:=-\{(t,x)\in\mathbb{R}_+\times\mathbb{R}\colon|x|\leq ct\}$, $A(t,x):=A+(t,x)$ and $A^+(t,x):=A(t,x)\cap(\mathbb{R}_+\times\mathbb{R})$.

The following lemma extends the previously known maximal inequalities for multi-parameter martingales [59, 89] to processes that are not martingales themselves but can be seen as "rotated martingales". For later purposes, we also need the situation where Λ is a not necessarily homogeneous Lévy basis (i.e., the coefficients

b, σ and ν in (1.2.1) may depend on (t,x) in such a way that $\Lambda(A)$ is well defined for all $A \in \mathcal{B}_{\mathbf{b}}(\mathbb{R}_+ \times \mathbb{R}^d)$).

Lemma 1.6.2. *(1) If Λ is a (not necessarily homogeneous) Lévy basis with mean 0, the process*

$$X(t,x) = \int_0^t \int_{\mathbb{R}} \mathbb{1}_{A(t,x)}(s,y)\, \Lambda(\mathrm{d}s, \mathrm{d}y) = \Lambda(A^+(t,x)), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R},$$

satisfies the maximal inequality

$$\lambda \mathbb{P}\left[\sup_{(s,y)\in[(\tilde{t},\tilde{x}),(t,x)]_\preceq} |X(s,y)| \geq \lambda \right] \leq 13 \mathbb{E}\left[|X(t,x)|\right]$$

for every $\lambda > 0$ and $(\tilde{t},\tilde{x}) \preceq (t,x)$ in $\mathbb{R}_+ \times \mathbb{R}$.

(2) If Λ further has a finite p'th moment with some $p > 1$, then X satisfies

$$\mathbb{E}\left[\sup_{(s,y)\in[(\tilde{t},\tilde{x}),(t,x)]_\preceq} |X(s,y)|^p \right] \leq \left(\frac{p}{p-1}\right)^{2p} \mathbb{E}[|X(t,x)|^p]$$

for every $(\tilde{t},\tilde{x}) \preceq (t,x)$ in $\mathbb{R}_+ \times \mathbb{R}$.

Proof. Without loss of generality we assume $(\tilde{t},\tilde{x}) = (0,0)$ and define a family of sub-σ-algebras $\mathbb{F} = (\mathcal{F}(s,y))_{(s,y)\in[(0,0),(t,x)]_\preceq}$ by

$$\mathcal{F}(s,y) := \sigma(X(\tau,\xi) \colon (\tau,\xi) \in \mathbb{R}_+ \times \mathbb{R}, (\tau,\xi) \preceq (s,y)).$$

Then we have $\mathcal{F}(\tilde{s},\tilde{y}) \subseteq \mathcal{F}(s,y)$ for every $(\tilde{s},\tilde{y}) \preceq (s,y)$ and the process X is integrable and adapted to $\mathbb{F}$ on $[(0,0),(t,x)]_\preceq$. Next, we define for any $(\tilde{s},\tilde{y}) \preceq (s,y)$ in $[(0,0),(t,x)]_\preceq$

$$\mathcal{M}(\tilde{s},\tilde{y}) := \left\{ \{X(s_1,y_1) \in B_1, ..., X(s_n,y_n) \in B_n\} \colon n \in \mathbb{N}, (s_i,y_i) \preceq (\tilde{s},\tilde{y}), B_i \in \mathcal{B}(\mathbb{R}) \right\}.$$

Then the properties of a Lévy basis imply $X(s,y) - X(\tilde{s},\tilde{y}) \perp\!\!\!\perp \mathcal{M}(\tilde{s},\tilde{y})$. Since $\mathcal{M}(\tilde{s},\tilde{y})$ is intersection-stable, we get $X(s,y) - X(\tilde{s},\tilde{y}) \perp\!\!\!\perp \sigma(\mathcal{M}(\tilde{s},\tilde{y})) = \mathcal{F}(\tilde{s},\tilde{y})$ and therefore

$$\begin{aligned} \mathbb{E}[X(s,y) \mid \mathcal{F}(\tilde{s},\tilde{y})] &= \mathbb{E}[\Lambda(A^+(\tilde{s},\tilde{y})) + \Lambda(A^+(s,y) \backslash A^+(\tilde{s},\tilde{y})) \mid \mathcal{F}(\tilde{s},\tilde{y})] \\ &= \Lambda(A^+(\tilde{s},\tilde{y})) + \mathbb{E}[\Lambda(A^+(s,y) \backslash A^+(\tilde{s},\tilde{y})) \mid \mathcal{F}(\tilde{s},\tilde{y})] \\ &= \Lambda(A^+(\tilde{s},\tilde{y})) = X(\tilde{s},\tilde{y}). \end{aligned} \tag{1.6.11}$$

We now transform X into a strong martingale by considering the function

$$H\colon [(0,0),(t,x)]_{\leq} \to [(0,0),H(t,x)]_{\leq}$$

given by

$$H(s,y) = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{c} \end{pmatrix} \begin{pmatrix} s \\ y \end{pmatrix}.$$

Note that the first matrix is the rotation matrix about 45 degrees counter-clockwise and the second matrix is a rescaling in the space coordinate. It is easy to see that H is in fact order-preserving and bijective. Now let $\tilde{X}$ be the push-forward process of X through H, i.e. $\tilde{X}(v_1,v_2) \coloneqq X(H^{-1}(v_1,v_2))$, and $\tilde{\mathbb{F}}$ be the push-forward of $\mathbb{F}$ through H, i.e. $\tilde{\mathcal{F}}(v_1,v_2) \coloneqq \mathcal{F}(H^{-1}(v_1,v_2))$ for every $(v_1,v_2) \in [(0,0),H(t,x)]_{\leq}$. Then $\tilde{\mathbb{F}}$ is a filtration on $[(0,0),H(t,x)]_{\leq}$ and $\tilde{X}$ is a martingale with respect to $\tilde{\mathbb{F}}$ since the property in (1.6.11) of X is inherited through H. In fact, $\tilde{X}$ is even a strong martingale because we have for $(v_1,v_2) \leq (w_1,w_2)$ in $[(0,0),H(t,x)]_{\leq}$ that

$$\begin{aligned}
&\tilde{X}((v_1,v_2),(w_1,w_2)] \\
&\quad = \tilde{X}(w_1,w_2) - \tilde{X}(v_1,w_2) - \tilde{X}(w_1,v_2) + \tilde{X}(v_1,v_2) \\
&\quad = X(H^{-1}(w_1,w_2)) - X(H^{-1}(v_1,w_2)) - X(H^{-1}(w_1,v_2)) + X(H^{-1}(w_1,w_2)) \\
&\quad = \Lambda(A^+(H^{-1}(w_1,w_2))) - \Lambda(A^+(H^{-1}(v_1,w_2))) - \Lambda(A^+(H^{-1}(w_1,v_2))) \\
&\qquad + \Lambda(A^+(H^{-1}(w_1,w_2))) \\
&\quad = \Lambda([H^{-1}(v_1,v_2),H^{-1}(w_1,w_2)]_{\leq}),
\end{aligned}$$

where the last inequality follows from the triangular shape of A^+. Moreover, letting $(u_1,u_2) \coloneqq H(t,x)$, we have

$$\begin{aligned}
\tilde{\mathcal{F}}^*(v_1,v_2) &= \tilde{\mathcal{F}}^1(v_1) \vee \tilde{\mathcal{F}}^2(v_2) = \tilde{\mathcal{F}}(v_1,u_2) \vee \tilde{\mathcal{F}}(u_1,v_2) \\
&= \mathcal{F}(H^{-1}(v_1,u_2)) \vee \mathcal{F}(H^{-1}(u_1,v_2)) \\
&= \sigma\left(X(s,y)\colon (s,y) \in A^+(H^{-1}(v_1,u_2)) \cup A^+(H^{-1}(u_1,v_2))\right).
\end{aligned}$$

With the same argument as in (1.6.11), we can show that

$$\Lambda([H^{-1}(v_1,v_2),H^{-1}(w_1,w_2)]_{\leq})$$

is independent of $\sigma(X(s,y)\colon (s,y) \in A^+(H^{-1}(v_1,u_2)) \cup A^+(H^{-1}(u_1,v_2)))$, which implies

$$\mathbb{E}[\tilde{X}((v_1,v_2),(w_1,w_2)] \mid \tilde{\mathcal{F}}^*(v_1,v_2)] = 0.$$

Therefore $\tilde{X}$ is a strong martingale with respect to $\tilde{\mathbb{F}}$. This allows us to use Walsh's maximal inequality for strong martingales: by Corollary 3.4 in [89] we get

$$\begin{aligned}\lambda\mathbb{P}\left[\sup_{(s,y)\in[(0,0),H(t,x)]_\preceq}|\tilde{X}(u_1,u_2)|\geq\lambda\right]&\leq 13\sup_{(s,y)\in[(0,0),H(t,x)]_\preceq}\mathbb{E}[|\tilde{X}(u_1,u_2)|]\\&=13\mathbb{E}[|\tilde{X}(H(t,x))|]\end{aligned}$$

for all $\lambda>0$. By the definition of $\tilde{X}$ this is equivalent to

$$\lambda\mathbb{P}\left[\sup_{(s,y)\in[(0,0),(t,x)]_\preceq}|X(s,y)|\geq\lambda\right]\leq 13\mathbb{E}[|X(t,x)|].$$

The second part of the lemma can be proved similarly, using Cairoli's maximal inequality for orthomartingales (see e.g. Corollary 2.3.1 in [59]) and the fact that a strong martingale is always an orthomartingale (see e.g. Proposition 1.1 in [89]). □

Proof. (of Theorem 1.5.5)

The cases where Λ is equal to the drift and Gaussian part, the compensated small jumps part, or the large jumps part are considered separately.

Case 1: $\Lambda(\mathrm{d}t,\mathrm{d}x)=b\,\mathrm{d}(t,x)+\sigma W(\mathrm{d}t,\mathrm{d}x)$.

Our assertion is proved once we can show both conditions of Theorem 1.5.1. The second condition is obviously satisfied and the first condition follows similarly as in Case 1 of the proof of Theorem 1.5.3 in conjunction with the boundedness of h. We omit the details here.

Case 2: $\Lambda(\mathrm{d}t,\mathrm{d}x)=\int_{\mathbb{R}}z\mathbb{1}_{\{|z|\leq 1\}}(\mathfrak{p}-\mathfrak{q})(\mathrm{d}t,\mathrm{d}x,\mathrm{d}z)$.

The argument is similar to Case 2 of the proof of Theorem 1.5.3. Therefore, we only highlight the major differences. Again, we take advantage of the fact that the Lévy basis $\Lambda_n(\mathrm{d}t,\mathrm{d}x):=\int_{\mathbb{R}}z\mathbb{1}_{\{1/n\leq|z|\leq 1\}}\mathbb{1}_{\{|x|\leq n\}}(\mathfrak{p}-\mathfrak{q})(\mathrm{d}t,\mathrm{d}x,\mathrm{d}z)$ has only finitely many jumps on $[0,T]\times\mathbb{R}$ almost surely. Therefore, both $F_n:=g*\Lambda_n$ and $\rho*F_n$ have $\preceq$-càdlàg paths almost surely. Our claim is proved if we show that $F^n:=F-F_n$ and $\rho*F^n=\rho*F-\rho*F_n$ both converge to 0, uniformly on compacts in probability. To

this end, we estimate for $(t_1, x_1) \preceq (t_2, x_2)$ in $\mathbb{R}_+ \times \mathbb{R}$ and a sufficiently big $T > 0$

$$\begin{aligned}&\mathbb{E}\bigg[\sup_{(t,x)\in[(t_1,x_1),(t_2,x_2)]_\preceq} |(\rho * F^n)(t,x)|\bigg]\\&\quad\le \int_0^T \int_{\mathbb{R}} \mathbb{E}\bigg[\sup_{(t,x)\in[(t_1,x_1),(t_2,x_2)]_\preceq} |F^n(t-s,x-y)|\bigg]\,|\rho|(\mathrm{d}s,\mathrm{d}y)\end{aligned}$$

and use the fundamental theorem of calculus and the stochastic Fubini theorem to split the last integrand into four parts according to

$$\begin{aligned}F^n(t-u,x-\eta) &= \int_0^{t-u}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h(t-u-s,x-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\\
&= \int_0^{t-u}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h(0,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\\
&\quad+\int_{-\eta}^{x-\eta}\int_0^{t-u}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}z\\
&\quad+\int_0^{t-u}\int_0^{r}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h_1(r-s,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}r\\
&\quad+\int_0^{t-u}\int_{-\eta}^{x-\eta}\int_0^{r}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h_{12}(r-s,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}z\,\mathrm{d}r, \quad (1.6.12)\end{aligned}$$

where $(u,\eta) \in [0,t]\times\mathbb{R}$, $h_1(t,x) = \partial_t h(t,x)$, $h_2(t,x) = \partial_x h(t,x)$, $h_{12}(t,x) = \partial_t\partial_x h(t,x)$ and $\Lambda^n = \Lambda - \Lambda_n$. For the first summand, we have by Lemma 1.6.2 and a similar reasoning as in (1.6.9) that

$$\begin{aligned}&\mathbb{E}\bigg[\sup_{(t,x)\in[(t_1,x_1),(t_2,x_2)]_\preceq}\bigg|\int_0^{t-u}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h(0,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\bigg|^2\bigg]\\
&\qquad\le C_T\mathbb{E}\bigg[\bigg|\int_0^{t_2-u}\int_{\mathbb{R}} \mathbb{1}_{A(t_2-u,x_2-\eta)}(s,y)h(0,-\eta-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\bigg|^2\bigg]\end{aligned}$$

converges to 0 as $n\to\infty$ and is bounded in (u,η) and n. Note that we have used that h belongs to $L^2_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R})$ since it is a bounded function in $L^\alpha_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R})$. Regarding the second summand we have for sufficiently big $\xi > 0$ that

$$\begin{aligned}&\mathbb{E}\bigg[\sup_{(t,x)\in[(t_1,x_1),(t_2,x_2)]_\preceq}\bigg|\int_{-\eta}^{x-\eta}\int_0^{t-u}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\,\mathrm{d}z\bigg|^2\bigg]\\
&\le C_T\int_{-\xi-\eta}^{\xi-\eta}\mathbb{E}\bigg[\sup_{(t,x)\in[(t_1,x_1),(t_2,x_2)]_\preceq}\bigg|\int_0^{t-u}\int_{\mathbb{R}} \mathbb{1}_{A(t-u,x-\eta)}(s,y)h_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\bigg|^2\bigg]\,\mathrm{d}z\\
&\le C_T\int_{-\xi-\eta}^{\xi-\eta}\mathbb{E}\bigg[\bigg|\int_0^{t_2-u}\int_{\mathbb{R}} \mathbb{1}_{A(t_2-u,x_2-\eta)}(s,y)h_2(0,z-y)\,\Lambda^n(\mathrm{d}s,\mathrm{d}y)\bigg|^2\bigg]\,\mathrm{d}z\to 0,\end{aligned}$$

where we have used Lemma 1.6.2 and that h_2 also belongs to $L^2_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R})$. Analogous calculations hold for the third and fourth summand. As a consequence, both F^n and $\rho * F^n$ converge uniformly on compacts in probability to 0 by dominated convergence.

Case 3: $\Lambda(\mathrm{d}t,\mathrm{d}x) = \int_{\mathbb{R}} z\mathbb{1}_{\{|z|>1\}}\,\mathfrak{p}(\mathrm{d}t,\mathrm{d}x,\mathrm{d}z)$.

In this case the same argument as in Case 3 of the proof of Theorem 1.5.3 applies with $\mathbb{1}_{-A}h$ instead of g. Notice that, under the current setting, the first integral in (1.6.10) is actually taken for a bounded function on a compact subset of $\mathbb{R}$ (due to the indicator $\mathbb{1}_{-A}$), is therefore finite and tends to zero as $n\to\infty$. The only difference appears in the case $\alpha = 1$, where we cannot copy the proof of Theorem 1.5.3 since Lemma 1.6.2(2) requires $p > 1$. Instead, we observe for any $T > 0$ and $U = [-K, K] \subseteq \mathbb{R}$ that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}|(\rho * F - \rho * F_n)(t,x)|\right]\\
&\quad= \mathbb{E}\left[\sup_{(t,x)\in[0,T]\times U}\left|\int_0^t\int_{\mathbb{R}}(\rho * g)(t-s,x-y)\,(\Lambda-\Lambda_n)(\mathrm{d}s,\mathrm{d}y)\right|\right]\\
&\quad\le \int_0^T\int_{\mathbb{R}}\sup_{(t,x)\in[0,T]\times U}|(\rho * g)|\,(t-s,x-y)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}(s,y)\int_{\mathbb{R}}|z|\mathbb{1}_{\{|z|\ge1\}}\,\nu(\mathrm{d}z)\\
&\quad\le C_T\int_{\mathbb{R}}\sup_{(t,x)\in[0,T]\times U}|(\rho * g)|\,(t,x-y)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y\\
&\quad\le C_T\int_{\mathbb{R}}\sup_{(t,x)\in[0,T]\times U}\int_0^t\int_{\mathbb{R}}|g(t-u,x-y-v)|\,|\rho|(\mathrm{d}u,\mathrm{d}v)\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y\\
&\quad\le C_T\int_0^T\int_{\mathbb{R}}\int_{\mathbb{R}}\sup_{(t,x)\in[0,T]\times U}|g(t-u,x-y-v)|\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y\,|\rho|(\mathrm{d}u,\mathrm{d}v),\qquad(1.6.13)
\end{aligned}$$

and that $\int_{\mathbb{R}}\sup_{(t,x)\in[0,T]\times U}|g(t-u,x-y-v)|\mathbb{1}_{\{|y|>n\}}\,\mathrm{d}y$ is bounded in (n,u,v) and goes to 0 by dominated convergence since $g = \mathbb{1}_{-A}h$. Consequently also (1.6.13) tends to 0 by dominated convergence. The pair F and F_n can be treated analogously. □

1.6.4 Proof of Proposition 1.2.2

We first collect some useful properties of convolutions. The proof of the following two lemmata is analogous to the one-parameter case (see Section 4.1 of [49] or Example 10.3 of [77] for the first result, and Section 3.6 of [49] for the second result).

Lemma 1.6.3. *Let S be $\mathbb{R}^{d+1}$, $\mathbb{R}_+ \times \mathbb{R}^d$ or $[0,T] \times \mathbb{R}^d$ and μ, η and π be measures in $M(S)$. Then*

(1) $\mu * \eta \in M(S)$ and $\|\mu * \eta\| \le \|\mu\|\|\eta\|$,

(2) $(\mu * \eta) * \pi = \mu * (\eta * \pi)$,

(3) $\mu * \eta = \eta * \mu$.

The statement is still valid if $M(S)$ is replaced by $M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ and $\|\cdot\|$ in (1) is replaced by the total variation norm on $[0,T] \times \mathbb{R}^d$ for some arbitrary $T \in \mathbb{R}_+$.

Lemma 1.6.4. *Let μ and η be measures in $M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ and $h \in L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$. Then the following statements hold for all $p \in [1,\infty]$.*

(1) $h * \mu \in L^p_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$.

(2) $(h * \mu) * \eta = h * (\mu * \eta)$ and $(\mu * h) * \eta = \mu * (h * \eta)$.

(3) If additionally $\mu \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ and $h \in L^p(\mathbb{R}_+ \times \mathbb{R}^d)$, then $h * \mu$ also belongs to $L^p(\mathbb{R}_+ \times \mathbb{R}^d)$.

Proof. (of Proposition 1.2.2) (1) Our proof extends Theorem 4.1.5 of [49], and for the reader's convenience we present the details in short. Note that this part can alternatively be proven in a more abstract framework using Theorem 4.3.6(b) of [71] involving the Jacobson radical of the commutative Banachalgebra $(M([0,T] \times \mathbb{R}^d), +, *)$. First we show that for each positive T there is a unique ρ_T in $M([0,T] \times \mathbb{R}^d)$ such that

$$\rho_T + \mu_T = \mu_T * \rho_T.$$

Here μ_T is the restriction of μ on $[0,T] \times \mathbb{R}^d$. To show the existence of ρ_T we construct a geometric series and use a Banach space argument. We first consider the special case $\|\mu_T\| = |\mu_T|([0,T] \times \mathbb{R}^d) < 1$. Defining

$$\rho_m := -\sum_{j=1}^{m} \mu_T^{*j}, \quad m \in \mathbb{N},$$

we obtain

$$\rho_m + \mu_T = -\sum_{j=1}^{m} \mu_T^{*j} + \mu_T = -\sum_{j=2}^{m} \mu_T^{*j} = \mu_T * \left(-\sum_{j=1}^{m-1} \mu_T^{*j}\right) = \mu_T * \rho_{m-1}, \quad m \in \mathbb{N}\backslash\{1\}.$$

By Lemma 1.6.3, we have $\|\mu_T^{*j}\| \le \|\mu_T\|^j$, so (ρ_m) is a Cauchy sequence and converges to some $\rho_T \in M([0,T]\times\mathbb{R}^d)$ because $M([0,T]\times\mathbb{R}^d)$ is a Banach space. In addition, $\mu_T * \rho_m \to \mu_T * \rho_T$ in $M([0,T]\times\mathbb{R}^d)$ by Lemma 1.6.3, so that we get $\rho_T + \mu_T = \mu_T * \rho_T$.

In the general case where $\|\mu_T\|$ is not necessarily smaller than one, we consider the measure $\lambda_m(\mathrm{d}s,\mathrm{d}y) := \mathrm{e}^{-ms}\mu_T(\mathrm{d}s,\mathrm{d}y)$ and note that for sufficiently large m we have $\|\lambda_m\| < 1$ because $\mu(\{0\}\times\mathbb{R}^d) = 0$. In this case, by what we have already proved, there exists a measure η_m satisfying $\eta_m + \lambda_m = \lambda_m * \eta_m$. But then $\rho_T(\mathrm{d}s,\mathrm{d}y) := \mathrm{e}^{ms}\eta_m(\mathrm{d}s,\mathrm{d}y)$ satisfies

$$\begin{aligned}
&(\rho_T + \mu_T)(\mathrm{d}s,\mathrm{d}y)\\
&\quad = \mathrm{e}^{ms}\eta_m(\mathrm{d}s,\mathrm{d}y) + \mathrm{e}^{ms}\mathrm{e}^{-ms}\mu_T(\mathrm{d}s,\mathrm{d}y) = \mathrm{e}^{ms}\eta_m(\mathrm{d}s,\mathrm{d}y) + \mathrm{e}^{ms}\lambda_m(\mathrm{d}s,\mathrm{d}y)\\
&\quad = \mathrm{e}^{ms}\left(\eta_m(\mathrm{d}s,\mathrm{d}y) + \lambda_m(\mathrm{d}s,\mathrm{d}y)\right) = \mathrm{e}^{ms}\left(\lambda_m * \eta_m\right)(\mathrm{d}s,\mathrm{d}y)\\
&\quad = \left([\mathrm{e}^{ms}\lambda_m(\mathrm{d}s,\mathrm{d}y)] * [\mathrm{e}^{ms}\eta_m(\mathrm{d}s,\mathrm{d}y)]\right)(\mathrm{d}s,\mathrm{d}y) = (\mu_T * \rho_T)(\mathrm{d}s,\mathrm{d}y),
\end{aligned}$$

where the fifth equation follows from the definition of the convolution. Thus $\rho_T + \mu_T = \mu_T * \rho_T$.

In order to show the uniqueness of ρ_T, we assume that there are ρ_T and η_T in $M([0,T]\times\mathbb{R}^d)$ with $\rho_T + \mu_T = \mu_T * \rho_T$ and $\eta_T + \mu_T = \mu_T * \eta_T$. Then

$$\begin{aligned}
\rho_T &= \mu_T * \rho_t - \mu_T = (\mu_T * \eta_T - \eta_T) * \rho_T - \mu_T = \eta_T * (\mu_T * \rho_T - \rho_T) - \mu_T = \eta_T * \mu_T - \mu_T\\
&= \eta_T.
\end{aligned}$$

Now, having constructed ρ_T for every positive T and noting that for every $j \in \mathbb{N}$ the restriction of ρ_{j+1} to $[0,j]\times\mathbb{R}^d$ must be equal to ρ_j by uniqueness, we define ρ to be the unique measure on $\mathbb{R}_+\times\mathbb{R}^d$ with $\rho = \rho_T$ on $[0,T]\times\mathbb{R}^d$. We still have $\rho \in M_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ and $\rho + \mu = \mu * \rho$, so the proof of (1) is complete.

(2) Let ρ be the resolvent of μ as in part (1). Then for $F \in L^p_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ define X by (1.2.4), which is well defined by Lemma 1.6.4. Also by Lemma 1.6.4, we obtain $X \in L^p_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$ and

$$X - \mu * X = X - \mu * (F - \rho * F) = X - (\mu - \mu * \rho) * F = X + \rho * F = F,$$

thus X is a solution of (1.2.2). To show uniqueness let $\tilde{X}$ be an arbitrary solution of (1.2.2) in $L^p_{\mathrm{loc}}(\mathbb{R}_+\times\mathbb{R}^d)$. Then

$$\tilde{X} = F + \mu * \tilde{X} = F + (\rho * \mu - \rho) * \tilde{X} = F - \rho * (\tilde{X} - \mu * \tilde{X}) = F - \rho * F,$$

hence $\tilde{X} = X$.

(3) The assumptions on F guarantee that X belongs to $\mathcal{L}(\mu)$ and that all calculations in the previous part remain valid. □

1.7 Examples of resolvents

In this section we derive formulae for tempo-spatial resolvents in various examples. They are based on the following lemma whose proof follows directly from the definition of the resolvent measure.

Lemma 1.7.1. *Suppose that $\mu \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ has a resolvent measure $\rho \in M_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$.*

(1) If $\mu = m \otimes \delta_{0,\mathbb{R}^d}$ with some $m \in M_{\mathrm{loc}}(\mathbb{R}_+)$ satisfying $m(\{0\}) = 0$ and r is the temporal resolvent measure of m (i.e., the unique $r \in M_{\mathrm{loc}}(\mathbb{R}_+)$ with $r * m = r + m$), then $\rho = r \otimes \delta_{0,\mathbb{R}^d}$.

(2) If μ has a Lebesgue density $k \in L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$, then also ρ has a Lebesgue density $r \in L^1_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$, which is given by

$$r(t,x) = -\sum_{n=1}^{\infty} k^{*n}(t,x), \tag{1.7.1}$$

where the series converges absolutely for almost every $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$.

(3) If in the situation of (2) we have that $k(t,x) = -\lambda f(x)$ with some $f \in L^1(\mathbb{R}^d)$ $\lambda \in \mathbb{R}$, then (1.7.1) takes the form

$$r(t,x) = \lambda \sum_{n=1}^{\infty} \frac{(-\lambda t)^{n-1}}{(n-1)!} f^{*n}(x), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d.$$

Example 1.7.2 We present some applications of Lemma 1.7.1.

(1) If $\mu = -\lambda \mathrm{Leb}_{\mathbb{R}_+} \otimes \delta_{0,\mathbb{R}^d}$ is the measure considered in Examples 1.3.6 and 1.3.7, the resolvent measure is given by $\rho(\mathrm{d}t, \mathrm{d}x) = \lambda \mathrm{e}^{-\lambda t}\, \mathrm{d}t\, \delta_{0,\mathbb{R}^d}(\mathrm{d}x)$.

(2) Let $f(x) = 1/(\pi(1+x^2))$ be the density of the one-dimensional Cauchy$(0,1)$-distribution. Then $f^{*n}(x) = n/(\pi(x^2+n^2))$ is the density of the Cauchy$(0,n)$-distribution and

$$r(t,x) = \lambda \sum_{n=1}^{\infty} \frac{(-\lambda t)^{n-1}}{(n-1)!} \frac{n}{\pi(x^2+n^2)} = \frac{\lambda}{2\pi x} G(\lambda t, x), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R},$$

where G is the (real-valued) function given by

$$\begin{aligned} G(t,x) := \mathrm{i}\Big(& t^{-(1+\mathrm{i}x)}\big(\Gamma(1+\mathrm{i}x) - \Gamma(1+\mathrm{i}x,t) - \Gamma(2+\mathrm{i}x) + \Gamma(2+\mathrm{i}x,t)\big) \\ & - t^{-(1-\mathrm{i}x)}\big(\Gamma(1-\mathrm{i}x) - \Gamma(1-\mathrm{i}x,t) - \Gamma(2-\mathrm{i}x) + \Gamma(2-\mathrm{i}x,t)\big)\Big) \end{aligned} \tag{1.7.2}$$

and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function.

(3) If $f(x) = \mathrm{e}^{-x}\mathbb{1}_{\mathbb{R}_+}(x)$, then $f^{*n}(x) = \mathrm{e}^{-x}x^{n-1}((n-1)!)^{-1}\mathbb{1}_{\mathbb{R}_+}(x)$ and hence

$$r(t,x) = \lambda \sum_{n=1}^{\infty} \frac{(-\lambda t)^{n-1}}{(n-1)!} \frac{x^{n-1}\mathrm{e}^{-x}}{(n-1)!} \mathbb{1}_{\mathbb{R}_+}(x) = \begin{cases} \lambda \mathrm{e}^{-x} J_0(\sqrt{2\lambda t x})\mathbb{1}_{\mathbb{R}_+}(x) & \text{if } \lambda \geq 0, \\ \lambda \mathrm{e}^{-x} I_0(\sqrt{|2\lambda t x|})\mathbb{1}_{\mathbb{R}_+}(x) & \text{if } \lambda < 0, \end{cases}$$

where J_0 (I_0) is the (modified) zeroth order Bessel function of the first kind.

(4) If k is a multiple of the heat kernel, that is,

$$k(t,x) = \lambda(4\pi t)^{-d/2} \exp(-|x|^2/(4t))\mathbb{1}_{(0,\infty)}(t),$$

for some $\lambda \in \mathbb{R}$, we have $k^{*n}(t,x) = (\lambda t)^{n-1}/(n-1)!k(t,x)$ and therefore

$$r(t,x) = -\sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1}}{(n-1)!} k(t,x) = -\mathrm{e}^{\lambda t} k(t,x), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d.$$

□

1.8 Integrability properties of resolvents

In many cases integrability properties of resolvents are of interest, see for example Theorem 1.3.3. In this section we present criteria for the resolvent ρ to lie in $M(\mathbb{R}_+ \times \mathbb{R}^d)$. The conditions of the lemmata below are given in terms of the Laplace transform

of the drift measure μ which is defined as follows: for a measure μ in $M(\mathbb{R}^d)$ the Laplace transform $\hat{\mu}(z)$ is the function

$$\hat{\mu}(z) = \int_{\mathbb{R}^d} \mathrm{e}^{-z\cdot u}\, \mu(\mathrm{d}u),$$

defined for those $z \in \mathbb{C}^d$ for which the integral exists and where $z \cdot u$ denotes the standard scalar product in $\mathbb{C}^d$. Every $\mu \in M(\mathbb{R}^d)$ can be split into three parts, namely the absolutely continuous part μ_c, the discrete part μ_d and the singular continuous part μ_s. The next lemma can be proved analogously to the one-parameter case (see Theorem 4.4.3 of [49]). We use the notation $\mathfrak{Re}(z)$ for the real part of a complex number z and $\mathrm{i}\mathbb{R}^d$ for the subspace of $\mathbb{C}^d$ consisting of all vectors whose entries have all real part zero.

Lemma 1.8.1. *Let $\mu \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ satisfy*

- $\mu(\{0\} \times \mathbb{R}^d) = 0$,
- $\hat{\mu}(\tau, \xi) \neq 1$ *for all* $\mathfrak{Re}(\tau) \geq 0$, $\xi \in \mathrm{i}\mathbb{R}^d$ *and*
- $\inf_{\mathfrak{Re}(\tau)\geq 0, \xi \in \mathrm{i}\mathbb{R}^d} |\widehat{\mu_d}(\tau, \xi) - 1| > \|\mu_s\|$.

Then the resolvent ρ of μ belongs to $M(\mathbb{R}_+ \times \mathbb{R}^d)$.

In the special case where μ is absolutely continuous we can give a condition which is both sufficient and necessary. Once again, the proof is analogous to the one-parameter case, cf. Theorem 2.4.1 of [49].

Lemma 1.8.2. *Let $\mu \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ be absolutely continuous. Then the resolvent ρ of μ belongs to $M(\mathbb{R}_+ \times \mathbb{R}^d)$ if and only if $\hat{\mu}(\tau, \xi) \neq 1$ for all $\mathfrak{Re}(\tau) \geq 0$ and $\xi \in \mathrm{i}\mathbb{R}^d$.*

Integrability of ρ is sufficient, but not necessary for the convolution operator $g \mapsto \rho * g$ to map $L^2_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ into $L^2_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$. The following lemma is analogous to the one-parameter case in Theorem 2.6.2 of [49] and provides a criterion in this respect.

Lemma 1.8.3. *Let $\mu \in M(\mathbb{R}_+ \times \mathbb{R}^d)$ be absolutely continuous and satisfy the conditions*

- $\sup_{\sigma>0, (\tau,\xi)\in\mathbb{R}^d} \left| \frac{\hat{\mu}(\sigma+\mathrm{i}\tau, \mathrm{i}\xi)}{\hat{\mu}(\sigma+\mathrm{i}\tau, \mathrm{i}\xi)-1} \right| < \infty$,
- $\int_{\mathbb{R}_+\times\mathbb{R}^d} \mathrm{e}^{-\sigma t}\, |\mu|(\mathrm{d}t, \mathrm{d}x) < \infty$ *for all* $\sigma > 0$.

Then the convolution operator $g \mapsto \rho * g$ maps $L^2_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$ continuously into $L^2_{\mathrm{loc}}(\mathbb{R}_+ \times \mathbb{R}^d)$, where ρ is the resolvent of μ.

Chapter 2:
Lévy-driven causal CARMA random fields

2.1 Introduction

Originally appearing in Doob [39], continuous-time autoregressive moving average processes, or CARMA processes in short, are the continuous-time analogs of the well-known ARMA processes (cf. Brockwell and Davis [23] for details on ARMA processes). Nowadays, CARMA processes are well-studied objects due to the extensive research in recent years, which partially stems from the successful usage of these processes as stochastic models for irregularly spaced or high-frequency data (cf. the review article Brockwell [21] and the references therein). Applications can be found in turbulence modeling [28], stochastic volatility modeling [10, 27, 87] and the electricity market [14, 45], just to name a few.

Given two non-negative integers $q < p$ and real coefficients $a_1, ..., a_p, b_0, ..., b_{p-1}$ such that $b_q \neq 0$ and $b_i = 0$ for $i > q$, the CARMA(p, q) process $(Y(t))_{t\in\mathbb{R}}$ is defined as the solution to the observation and state equations

$$\begin{aligned} Y(t) &= b^\top X(t), \quad t \in \mathbb{R}, \\ \mathrm{d}X(t) &= AX(t)\,\mathrm{d}t + c\,\mathrm{d}L(t), \quad t \in \mathbb{R}, \end{aligned} \tag{2.1.1}$$

where $b = (b_0, ..., b_{p-1})^\top \in \mathbb{R}^p$, $c = (0, ..., 0, 1)^\top \in \mathbb{R}^p$ and the matrix A is given by

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{pmatrix} \in \mathbb{R}^{p\times p}$$

if $p > 1$, and $A = (-a_1)$ if $p = 1$. Here it is assumed that L is a one-dimensional Lévy process, that is, a process with independent and stationary increments, càdlàg

sample paths and $L(0) = 0$ almost surely (cf. Sato [78] for details on Lévy processes). Equations (2.1.1) can be interpreted as the pth-order stochastic differential equation

$$a(\partial_t)Y(t) = b(\partial_t)\partial_t L(t), \quad t \in \mathbb{R}, \tag{2.1.2}$$

where the polynomials $a(\cdot)$ and $b(\cdot)$ are defined as

$$a(z) = z^p + a_1 z^{p-1} + \cdots + a_p, \quad \text{and} \quad b(z) = b_0 + b_1 z + \cdots + b_{p-1} z^{p-1}. \tag{2.1.3}$$

In fact, Equation (2.1.2) constitutes the continuous-time analog of the well-known ARMA equations, which define the ARMA process in discrete time. However, since the sample paths of a Lévy process are in general not differentiable, the definition of CARMA processes is based on the state-space representation (2.1.1). Note that A is the companion matrix of the polynomial $a(\cdot)$ and therefore the eigenvalues of A are equal to the roots of $a(\cdot)$. Under the assumptions that $a(\cdot)$ and $b(\cdot)$ have no common roots, the roots of $a(\cdot)$ have strictly negative real parts and the Lévy process L has a finite logarithmic moment, it was shown in Brockwell and Lindner [24, Theorem 3.3] that the CARMA equations have a unique strictly stationary solution Y on $\mathbb{R}$ with representation

$$Y(t) = \int_{-\infty}^{t} b^\top \mathrm{e}^{A(t-s)} c \, \mathrm{d}L(s), \quad t \in \mathbb{R}. \tag{2.1.4}$$

Moreover, the CARMA process is a causal function of the driving Lévy process under the assumptions above, i.e., the value of $Y(t)$ depends only on the values of $(L(s))_{s \le t}$ and is independent of $(L(s))_{s>t}$.

The aim of this chapter is to extend CARMA processes to multiple parameters in order to obtain a tractable class of random fields indexed by $\mathbb{R}^d$, which can be used to model spatial or even tempo-spatial phenomena. A spatial extension has in fact already been introduced in Brockwell and Matsuda [25]. Their isotropic CARMA random field is defined as

$$Y(t) = \int_{\mathbb{R}^d} g(t-s) \mathrm{d}L(s), \quad t \in \mathbb{R}^d, \tag{2.1.5}$$

where the radially symmetric kernel g is given by $g(t) = \sum_{i=1}^{p} \mathrm{e}^{\lambda_i \|t\|} \theta(\lambda_i)/\phi'(\lambda_i)$, $t \in \mathbb{R}^d$, and the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ have the forms $\phi(z) = \prod_{i=1}^{p}(z^2 - \lambda_i^2)$ and $\theta(z) = \prod_{i=1}^{q}(z^2 - \xi_i^2)$ with $\lambda_i, \xi_i \in \mathbb{C}$. Furthermore, each λ_i has a strictly negative real part and L is a Lévy sheet on $\mathbb{R}^d$, which is the multi-parameter analog of a Lévy process. This procedure generates a versatile family of isotropic covariance

functions in space, which are neither necessarily monotonically decreasing nor non-negative. However, the CARMA process is classically defined through the state-space representation (2.1.1), and the connection to these defining equations is unclear in [25]. Therefore, we propose a different class of CARMA random fields based on a system of stochastic partial differential equations (SPDEs) that constitutes a generalization of (2.1.1). We will show that this system has a *mild solution* given by

$$Y(t) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} b^\top \mathrm{e}^{A_1(t_1-s_1)} \cdots \mathrm{e}^{A_d(t_d-s_d)} c \, \mathrm{d}L(s), \quad t = (t_1, ..., t_d) \in \mathbb{R}^d, \tag{2.1.6}$$

and define the random field Y in (2.1.6) as the *causal CARMA random field* on $\mathbb{R}^d$, where $A_1, ..., A_d$ are companion matrices.

It turns out that many of the commonly known features of CARMA processes can be recovered for this model, including for instance exponentially decaying autocovariance functions and rational spectral densities. Moreover, the autocovariance is in general anisotropic and non-separable. The path properties are also similar to those we have in the one-dimensional case. More precisely, there exists a Hölder continuous version under Gaussian noise, and in the presence of jumps, we may use maximal inequalities for multi-parameter martingales in order to show the existence of càdlàg sample paths (see Definition 2.4.6). Furthermore, (2.1.6) reduces to (2.1.4) if $d = 1$ and sampling on an equidistant lattice leads to an ARMA random field under mild conditions. However, the moving average part has in general infinitely many terms in contrast to the one-dimensional case. This is due to the fact that a (q_1, q_2)-dependent random field is not always a $\mathrm{MA}(q_1, q_2)$ random field (see Definition 2.4.9). We examine this issue in Examples 2.4.15 and 2.4.16.

The CARMA random field in this chapter is causal with respect to the spatial partial order $\leq$ on $\mathbb{R}^d$, which is taken componentwise. This quarter-plane-type causality can be interpreted as a directional influence and has been incorporated in several articles in the literature. For example, both Tjøstheim [86] and Drapatz [41] consider quarter-plane ARMA models and they refer to applications in econometrics, veterinary epidemiology, geography, geology and image analysis. Furthermore, [86] points out that causal representations exist for a wide class of random fields and [41] mentions that statistical inference for such representations is often easier to conduct.

The causal CARMA random field is related to some other classes of random fields. It belongs to the class of *ambit fields* (see e.g. Barndorff-Nielsen et al. [12]), which has applications in biology, finance or turbulence. In particular, causal CARMA

random fields possess ambit sets which are translation invariant and have the form of a *quadrant* if $d = 2$, an *octant* if $d = 3$, an *orthant* if $d > 3$, respectively. Additionally, we will see that they constitute a parametric submodel of the Volterra-type Ornstein-Uhlenbeck (VOU) processes studied in Pham and Chong [74] and they generalize the multi-parameter Ornstein-Uhlenbeck process in Graversen and Pedersen [47].

This chapter is organized as follows: in Section 2.2, we first recall the notions of Lévy bases, Lévy sheets and their integration theory to the extent necessary for this chapter. At the beginning of Section 2.3 we derive a system of SPDEs (cf. (2.3.2)), which lays the groundwork for extending the CARMA process. Afterwards, we define the causal CARMA(p, q) random field and the more general causal GCARMA random field, for which we drop the assumption that $A_1, ..., A_d$ are in companion form. In Theorem 2.3.5 we show that these random fields exist under mild assumptions and solve the SPDE system (2.3.2) in the mild sense. Furthermore, we investigate the multi-parameter *CARMA kernel* in more detail and present several alternative representations. This section concludes with a remark on the connection to the VOU process studied in [74]. Section 2.4 is devoted to distributional and path properties of causal CARMA random fields. Expressions for the autocovariance function (cf. Theorem 2.4.1 and Proposition 2.4.3) and the spectral density (cf. Corollary 2.4.4) are derived and Theorem 2.4.7 establishes some path properties. Finally, we investigate sampling properties of causal CARMA random fields. Under a mild spectral condition, Theorem 2.4.13 shows that sampling on an equidistant lattice leads to a spatial ARMA process, which generally has infinitely many moving average terms. By contrast, Example 2.4.16 depicts a case with finitely many moving average terms.

The following notation will be used throughout this chapter: C denotes a generic strictly positive constant which may change its value from line to line without affecting any argumentation. We use $\mathbb{1}_{\{.\}}$ for the indicator function so that the Heaviside function may be written as $\mathbb{1}_{\{t\geq 0\}}$. If A is a matrix (or a vector), then $A^\top$ denotes the transpose of A. The prime symbol $'$ stands for differentiation of a univariate function. For multivariate functions, we use ∂_z for partial differentiation with respect to the variable z, or ∂_1 for partial differentiation with respect to the first variable. Components of a d-dimensional vector u are denoted by $u_1, ..., u_d$ if not stated otherwise. Furthermore, $\|u\|$ is the Euclidean norm, $u \cdot v \in \mathbb{R}$ is the scalar product, $u \odot v \in \mathbb{R}^d$ is the componentwise product and we write $u \leq v$ if and only if $u_i \leq v_i$ for all for $u, v \in \mathbb{R}^d$ and $i \in \{1, ..., d\}$. The imaginary unit is $\boldsymbol{i}$ and we set $\mathbb{R}_+ = [0, \infty)$.

2.2 Lévy bases and Lévy sheets

Throughout this chapter we will use homogeneous Lévy bases and we define them directly through their Lévy-Itô decomposition as a sum of a deterministic drift part, a Gaussian part, a compensated small jumps part and a large jumps part. From now on, all stochastic objects live on a fixed complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.2.1

(1) A *homogeneous Lévy basis* Λ on $\mathbb{R}^d$ is a family of random variables indexed by the bounded Borel subsets of $\mathbb{R}^d$ such that for all $A \in \mathcal{B}_b(\mathbb{R}^d)$ we have

$$\begin{aligned}\Lambda(A) = \beta \mathrm{Leb}_{\mathbb{R}^d}(A) + \sigma W(A) + \int_{\mathbb{R}^d}\int_{\mathbb{R}} \mathbb{1}_A(s) z \mathbb{1}_{\{|z|\leq 1\}} (\mathfrak{p} - \mathfrak{q})(\mathrm{d}s, \mathrm{d}z) \\ + \int_{\mathbb{R}^d}\int_{\mathbb{R}} \mathbb{1}_A(s) z \mathbb{1}_{\{|z|>1\}} \mathfrak{p}(\mathrm{d}s, \mathrm{d}z),\end{aligned}$$

where

- $\mathrm{Leb}_{\mathbb{R}^d}$ is the Lebesgue measure on $\mathbb{R}^d$ and $\beta \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ are constants,
- W is Gaussian white noise on $\mathbb{R}^d$ such that $\mathrm{Var}(W(A)) = \mathrm{Leb}_{\mathbb{R}^d}(A)$ (for more details see e.g. Chapter I in Walsh [90]),
- $\mathfrak{p}$ is a Poisson random measure on $\mathbb{R}^d \times \mathbb{R}$ with intensity measure $\mathfrak{q} = \mathrm{Leb}_{\mathbb{R}^d} \otimes \nu$, where ν is a Lévy measure on $\mathbb{R}$ (see e.g. Chapter II in Jacod and Shiryaev [56] for more details on Poisson random measures and their integration theory).

(2) The triplet (β, σ^2, ν) is called the *characteristics* of Λ. If $\int_{\mathbb{R}} |z| \mathbb{1}_{\{|z|>1\}} \nu(\mathrm{d}z) < \infty$, we say that Λ has a finite first moment and define $\kappa_1 := \beta + \int_{\mathbb{R}} z \mathbb{1}_{\{|z|>1\}} \nu(\mathrm{d}z)$ as the *mean* of Λ. Likewise, we say that Λ has a finite second moment and define $\kappa_2 := \sigma^2 + \int_{\mathbb{R}} z^2 \nu(\mathrm{d}z)$ as the *variance* of Λ if $\int_{\mathbb{R}} z^2 \nu(\mathrm{d}z) < \infty$. The *cumulant generating function* (or *Lévy symbol*) $\zeta : \mathbb{R} \to \mathbb{C}$ of Λ is given by

$$\zeta(u) = \boldsymbol{i}ub - \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}} (e^{\boldsymbol{i}uz} - 1 - \boldsymbol{i}u\mathbb{1}_{\{|z|\leq 1\}})\nu(\mathrm{d}z), \quad u \in \mathbb{R}.$$

(3) We associate with each homogeneous Lévy basis Λ on $\mathbb{R}^d$ a *Lévy sheet* $(L(t))_{t\in\mathbb{R}^d}$ via the equation

$$L(t) := \Lambda(\{(s_1 t_1, ..., s_d t_d)^\top : s_1, ..., s_d \in [0,1]\}), \quad t \in \mathbb{R}^d.$$

□

The stochastic integral with respect to Lévy bases is for deterministic integrands classically defined as in Rajput and Rosiński [75, Section II] (this paper uses the term infinitely divisible independently scattered random measure for Lévy basis), i.e., it is defined as the limit in probability of stochastic integrals of an approximating sequence of simple functions, where the stochastic integral for simple functions is defined canonically. We recall an integrability characterization from [75, Theorem 2.7] in the next proposition.

Proposition 2.2.2. *Let $g\colon \mathbb{R}^d \to \mathbb{R}$ be a measurable function and Λ be a homogeneous Lévy basis on $\mathbb{R}^d$ with characteristics (β, σ^2, ν) and cumulant generating function ζ. Then the stochastic integral $\int_{\mathbb{R}^d} g(s)\,\Lambda(\mathrm{d}s)$ is well defined if and only if*

(1) $\displaystyle\int_{\mathbb{R}^d} \left| \beta g(s) + \int_{\mathbb{R}} (z g(s) \mathbb{1}_{\{|z g(s)| \le 1\}} - g(s) z \mathbb{1}_{\{|z| \le 1\}})\, \nu(\mathrm{d}z) \right| \mathrm{d}s < \infty,$

(2) $\displaystyle\int_{\mathbb{R}^d} \sigma^2 |g(s)|^2 \,\mathrm{d}s < \infty,$

(3) $\displaystyle\int_{\mathbb{R}^d} \int_{\mathbb{R}} (1 \wedge |z g(s)|^2)\, \nu(\mathrm{d}z)\,\mathrm{d}s < \infty.$

In this case, the stochastic integral is infinitely divisible with characteristic function

$$\begin{aligned}\Phi\left(\int_{\mathbb{R}^d} g(s)\,\Lambda(\mathrm{d}s)\right)(u) &= \exp\left\{\int_{\mathbb{R}^d} \zeta(u g(s))\,\mathrm{d}s\right\} \\ &= \exp\left\{iu\beta_g - \frac{1}{2}u^2\sigma_g^2 + \int_{\mathbb{R}} (e^{iuz} - 1 - iu\mathbb{1}_{\{|z|\le 1\}})\nu_g(\mathrm{d}z)\right\},\ u \in \mathbb{R},\end{aligned}$$

and characteristic triplet $(\beta_g, \sigma_g^2, \nu_g)$ given by

- $\beta_g = \displaystyle\int_{\mathbb{R}^d} (\beta g(s) + \int_{\mathbb{R}} (z g(s) \mathbb{1}_{\{|z g(s)| \le 1\}} - g(s) z \mathbb{1}_{\{|z| \le 1\}})\, \nu(\mathrm{d}z))\,\mathrm{d}s,$
- $\sigma_g^2 = \displaystyle\int_{\mathbb{R}^d} \sigma^2 |g(s)|^2 \,\mathrm{d}s,$
- $\nu_g(B) = \displaystyle\int_{\mathbb{R}^d} \int_{\mathbb{R}} \mathbb{1}_{\{g(s) z \in B\}}\, \nu(\mathrm{d}z)\,\mathrm{d}s$ *for any Borel set* $B \in \mathcal{B}(\mathbb{R})$.

The following proposition provides a sufficient integrability criterion which is easier to check and will be useful later on. For its proof, we refer to Berger [15].

Proposition 2.2.3. *Let $g\colon\mathbb{R}^d \to \mathbb{R}$ be a measurable function such that $|g(x)| \leq Ce^{-\eta\|x\|}$ for some positive constants C, η and Λ be a homogeneous Lévy basis on $\mathbb{R}^d$ with $\int_{\mathbb{R}} \log(|z|)^d \mathbb{1}_{\{|z|>1\}}\, \nu(\mathrm{d}z) < \infty$. Then g is integrable with respect to Λ.*

Finally, the stochastic integral with respect to L is defined for exactly those functions g which are integrable with respect to Λ and we set

$$\int_{\mathbb{R}^d} g(s)\,\mathrm{d}L(s) := \int_{\mathbb{R}^d} g(s)\,\Lambda(\mathrm{d}s).$$

2.3 Causal CARMA random fields as mild solutions to a system of SPDEs

Our approach to defining a CARMA random field relies on a generalization of the state-space Equations (2.1.1). As the first step, we reformulate (2.1.1) as

$$\begin{aligned} Y(t) &= b^\top X(t), \quad t \in \mathbb{R}, \\ (I_p\partial_t - A)\,X(t) &= c\dot{L}(t), \quad t \in \mathbb{R}, \end{aligned} \tag{2.3.1}$$

where I_p is the identity matrix in $\mathbb{R}^{p\times p}$, $I_p\partial_t$ is a matrix whose entries are ∂_t on the diagonal and zero otherwise, and $(I_p\partial_t - A)$ is a system of ordinary differential operators with constant coefficients acting on the state vector X. The symbol $\dot{L}$ denotes the formal partial differentiation of L in each of t's components once. In the purely temporal case, t is one-dimensional and thus $\dot{L}(t)$ coincides with $\partial_t L(t)$. However, in d dimensions we have

$$\dot{L}(t) = \partial_1 \cdots \partial_d L(t), \quad t \in \mathbb{R}^d.$$

In order to lift Equations (2.3.1) to $\mathbb{R}^d$, we iterate the system of differential operators for each of t's components, that is, we consider the system of SPDEs

$$\begin{aligned} Y(t) &= b^\top X(t), \quad t \in \mathbb{R}^d, \\ \mathfrak{D}_d X(t) &= c\dot{L}(t), \quad t \in \mathbb{R}^d, \end{aligned} \tag{2.3.2}$$

where $\mathfrak{D}_d$ is the system of linear partial differential operators

$$\mathfrak{D}_d = (I_p\partial_d - A_d)\cdots(I_p\partial_1 - A_1),$$

A_i is the companion matrix to a monic polynomial $a_i(\cdot)$ of degree p for each $i = 1, ..., d$ and L is the Lévy sheet associated to a homogeneous Lévy basis Λ on $\mathbb{R}^d$. We are interested in solutions to (2.3.2) and use the following notion, which is based on the random field approach of Walsh [90].

Definition 2.3.1

(1) If $G \in (\mathcal{D}'(\mathbb{R}^d))^{p\times p}$ is a matrix with entries in the space of real-valued distributions $\mathcal{D}'(\mathbb{R}^d)$ and the application of the system of linear partial differential operators $\mathfrak{D}_d$ on G satisfies $\mathfrak{D}_d G = I_p \delta_0$, where δ_0 is the Dirac delta function, then G is called a *fundamental solution of $\mathfrak{D}_d$ to the right* (see e.g. Section 3.8 in Hörmander [53]).

(2) If $G \in (L^1_{\mathrm{loc}}(\mathbb{R}^d))^{p\times p}$ is a fundamental solution of $\mathfrak{D}_d$ to the right with entries in the space of locally integrable functions $L^1_{\mathrm{loc}}(\mathbb{R}^d)$ and the random field

$$G * (c\Lambda)(t) := \int_{\mathbb{R}^d} G(t-s)c\,\Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d,$$

exists in the sense of Section 2.2, then $b^\top(G * (c\Lambda))$ is called a *mild solution to* (2.3.2). □

Remark 2.3.2 As in the purely temporal case, the derivative $\dot{L}$ in Equations (2.3.2) does not exist in the classical sense. Nevertheless, there is a version of L such that $\dot{L}$ exists in the distributional sense. It can then be identified with the homogeneous Lévy basis Λ in the sense that $\langle \dot{L}, \phi\rangle = \int \phi(s)\,\Lambda(\mathrm{d}s)$ for all test functions $\phi \in \mathcal{D}(\mathbb{R}^d)$, where the angle brackets denote the application of the distribution $\dot{L}$ to ϕ (see Lemma 3.6 in Dalang and Humeau [38]). If in addition the Lévy measure ν satisfies $\int_{|x|>1} |x|^\alpha\,\nu(\mathrm{d}x) < \infty$ for some $\alpha > 0$, then $\dot{L}$ is even a random element of the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ (see Theorem 3.13 in [38]). □

Motivated by the solution formula (2.1.4) for the CARMA process, we define the causal CARMA random field and we will subsequently show that it is a mild solution to (2.3.2).

Definition 2.3.3 Let q and p be two non-negative integers such that $q < p$, $b = (b_0, ..., b_{p-1})^\top \in \mathbb{R}^p$ with $b_q \neq 0$ and $b_i = 0$ for $i > q$, $c = (0, ..., 0, 1)^\top \in \mathbb{R}^p$, and A_i be the companion matrix to a monic polynomial a_i of degree p with real coefficients and roots having strictly negative real parts for $i = 1, ..., d$. A random field $(Y(t))_{t\in\mathbb{R}^d}$ is

called *(causal) CARMA*(p,q) *random field* if it satisfies the equations

$$\begin{aligned} Y(t) &= b^\top X(t), \quad t \in \mathbb{R}^d, \\ X(t) &= \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} \mathrm{e}^{A_1(t_1-s_1)} \cdots \mathrm{e}^{A_d(t_d-s_d)} c \, \Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d, \end{aligned} \tag{2.3.3}$$

where Λ is a homogeneous Lévy basis on $\mathbb{R}^d$ with $\int_{\mathbb{R}} \log(|z|)^d \mathbb{1}_{\{|z|>1\}} \nu(\mathrm{d}z) < \infty$. A (causal) CARMA$(p,0)$ random field is also called a *(causal) CAR*(p) *random field.* □

Here causality is understood in the sense that the values of $X(t)$ and $Y(t)$ at point $t \in \mathbb{R}^d$ only depend on the values of Λ on the set $(-\infty, t_1] \times \cdots \times (-\infty, t_d]$. Causal CARMA$(p,q)$ random fields belong to the following class, for which we drop the requirements that every A_i is a companion matrix and the specific choice $c = (0, ..., 0, 1)^\top$.

Definition 2.3.4 Let $p \geq 1$ be an integer, $b, c \in \mathbb{R}^p$, $A_i \in \mathbb{R}^{p\times p}$ with eigenvalues having strictly negative real parts for $i = 1, ..., d$, and Λ be a homogeneous Lévy basis on $\mathbb{R}^d$ with $\int_{\mathbb{R}} \log(|z|)^d \mathbb{1}_{\{|z|>1\}} \nu(\mathrm{d}z) < \infty$. A random field $(Y(t))_{t\in\mathbb{R}^d}$ is called *(causal) generalized CARMA (GCARMA) random field* if it satisfies Equations (2.3.3). □

Since each CARMA(p,q) random field is also a GCARMA random field, every result which applies to GCARMA random fields also applies to CARMA(p,q) random fields. On the other hand, it is easy to find GCARMA random fields which are not CARMA random fields if we fix the order p (cf. for instance Example 2.5.1). The next theorem shows existence of GCARMA random fields and establishes the connection to the system (2.3.2). In what follows, $\mu_i(\lambda_i)$ denotes the algebraic multiplicity of the eigenvalue λ_i with respect to the matrix A_i.

Theorem 2.3.5. *Under the conditions of Definition 2.3.3 (resp. Definition 2.3.4) the CARMA*(p,q) *(resp. GCARMA) random field* $(Y(t))_{t\in\mathbb{R}^d}$ *exists and it is a mild solution to* (2.3.2)*.*

Proof. For the existence we have to check that the stochastic integral in (2.3.3) exists. Let $A_i = S_i J_i S_i^{-1}$ be a Jordan decomposition of A_i for $i = 1, ..., d$. Then we have that $\mathrm{e}^{A_i(t_i-s_i)} = S_i \mathrm{e}^{J_i(t_i-s_i)} S_i^{-1}$, from which we infer that each entry of the matrix $\mathrm{e}^{A_i(t_i-s_i)}$ is a (possibly complex) linear combination of

$$\{(t_i - s_i)^{k_i} \mathrm{e}^{\lambda_i(t_i-s_i)} : \lambda_i \text{ is an eigenvalue of } A_i,\ 0 \leq k_i \leq \mu_i(\lambda_i) - 1\}$$

for $i = 1, ..., d$. Hence, each component of the integrand $e^{A_1(t_1-s_1)} \cdots e^{A_d(t_d-s_d)}c$ is a (possibly complex) linear combination of the set

$$\{(t_1-s_1)^{k_1} e^{\lambda_1(t_1-s_1)} \cdots (t_d-s_d)^{k_d} e^{\lambda_d(t_d-s_d)} : \lambda_i \text{ is an eigenvalue of } A_i,\ 0 \le k_i \le \mu_i(\lambda_i)-1\}.$$

This shows that the integrability criteria of Proposition 2.2.3 are satisfied.

In order to show that Y is a mild solution to (2.3.2), we have to show that the matrix-valued function

$$G(t) = e^{A_1 t_1} \cdots e^{A_d t_d} \mathbb{1}_{\{t \ge 0\}}, \quad t \in \mathbb{R}^d,$$

is a fundamental solution of

$$\mathfrak{D}_d = (I_p \partial_d - A_d) \cdots (I_p \partial_1 - A_1)$$

to the right. By line A.2.2 in the appendix of Ortner and Wagner [70], an application of $\mathfrak{D}_d$ on G yields

$$\begin{aligned}
&(I_p \partial_d - A_d) \cdots (I_p \partial_1 - A_1) e^{A_1 t_1} \cdots e^{A_d t_d} \mathbb{1}_{\{t_1 \ge 0\}} \cdots \mathbb{1}_{\{t_d \ge 0\}} \\
&\quad = (I_p \partial_d - A_d) \cdots (I_p \partial_2 - A_2) I_p \delta_0(t_1) e^{A_2 t_2} \cdots e^{A_d t_d} \mathbb{1}_{\{t_2 \ge 0\}} \cdots \mathbb{1}_{\{t_d \ge 0\}} \\
&\quad = I_p \delta_0(t_1) (I_p \partial_d - A_d) \cdots (I_p \partial_2 - A_2) e^{A_2 t_2} \cdots e^{A_d t_d} \mathbb{1}_{\{t_2 \ge 0\}} \cdots \mathbb{1}_{\{t_d \ge 0\}} \\
&\quad = I_p \delta_0(t_1) \cdots I_p \delta_0(t_d) = I_p \delta_0(t),
\end{aligned}$$

where the derivatives are taken in the distributional sense and we have used the tensor product of distributions in the last line (see e.g. Section 5.1 in Hörmander [54]). Since G is also locally integrable, this finishes the proof. □

Example 2.3.6 (Stable GCARMA random fields) Let $\eta > 0$, $0 < \alpha \le 2$ and Λ be a symmetric α-stable homogeneous Lévy basis with cumulant generating function

$$\zeta(u) = -\eta |u|^\alpha, \quad u \in \mathbb{R}.$$

Since Λ has moments of any order strictly smaller than α, the GCARMA random field $(Y(t))_{t \in \mathbb{R}^d}$ of Definition 2.3.4 exists and Proposition 2.2.2 shows that for each $t \in \mathbb{R}^d$ the characteristic function of $Y(t)$ is

$$\Phi\left(Y(t)\right)(u) = \exp\left\{-\eta |u|^\alpha \int_{\mathbb{R}_+^d} |b^\top e^{A_1 s_1} \cdots e^{A_d s_d} c|^\alpha \, \mathrm{d}s\right\}, \quad u \in \mathbb{R}.$$

Hence, $Y(t)$ is symmetric α-stable with the same stability index as Λ. □

The CARMA process in Equation (2.1.4) is a strictly stationary process. Similarly, we have that every GCARMA random field Y is strictly stationary, that is, for every $n \in \mathbb{N}$ and $\tau, t^{(1)}, ..., t^{(n)} \in \mathbb{R}^d$ the distributions of $(Y(t^{(1)}), ..., Y(t^{(n)}))$ and $(Y(t^{(1)} + \tau), ..., Y(t^{(n)} + \tau))$ are equal.

Corollary 2.3.7. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) (resp. GCARMA) random field. Then it has the representation*

$$Y(t) = (g * \Lambda)(t) := \int_{\mathbb{R}^d} g(t-s)\,\Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d, \tag{2.3.4}$$

where the CARMA(p,q) kernel *(resp.* GCARMA kernel*) g is given by*

$$\begin{aligned} g(s) &= b^\top \mathrm{e}^{A_1 s_1} \cdots \mathrm{e}^{A_d s_d} c \mathbb{1}_{\{s\geq 0\}} \\ &= \sum_{\lambda_1} \sum_{k_1=0}^{\mu_1(\lambda_1)-1} \cdots \sum_{\lambda_d} \sum_{k_d=0}^{\mu_d(\lambda_d)-1} d(\lambda_1, k_1, ..., \lambda_d, k_d) s_1^{k_1} \mathrm{e}^{\lambda_1 s_1} \cdots s_d^{k_d} \mathrm{e}^{\lambda_d s_d} \mathbb{1}_{\{s\geq 0\}}, \end{aligned} \tag{2.3.5}$$

$s = (s_1, ..., s_d) \in \mathbb{R}^d$, $\{d(\lambda_1, k_1, ..., \lambda_d, k_d)\}$ is a set of complex coefficients and $\sum_{\lambda_i}$ denotes the sum over distinct eigenvalues of A_i for $i = 1, ..., d$. In particular, $(Y(t))_{t\in\mathbb{R}^d}$ is strictly stationary.

Proof. By the proof of Theorem 2.3.5, each component of $\mathrm{e}^{A_1 s_1} \cdots \mathrm{e}^{A_d s_d} c$, and therefore also $b^\top \mathrm{e}^{A_1 s_1} \cdots \mathrm{e}^{A_d s_d} c$, is a (possibly complex) linear combination of the set

$$\{s_1^{k_1} \mathrm{e}^{\lambda_1 s_1} \cdots s_d^{k_d} \mathrm{e}^{\lambda_d s_d} : \lambda_i \text{ is an eigenvalue of } A_i,\ 0 \leq k_i \leq \mu_i(\lambda_i) - 1\}.$$

This fact and Equations (2.3.3) imply Equations (2.3.4) and (2.3.5). The strict stationarity follows from Equation (2.3.4). □

A direct consequence of this representation is that, under the assumption that each A_i has distinct eigenvalues, Y is the sum of p^d dependent and possibly complex valued CAR(1) random fields (cf. Proposition 2 in [27] for the temporal analog), though some of which may vanish depending on the coefficients $d(\lambda_1, k_1, ..., \lambda_d, k_d)$. Moreover, the kernel g in Equation (2.3.5) is *anisotropic* in contrast to the isotropic CARMA random field in (2.1.5). Also, g is in general *non-separable*, i.e., it cannot be written as a product of the form $g(s) = g_1(s_1) \cdots g_d(s_d)$.

In general, we do not have explicit formulae for the coefficients $d(\lambda_1, k_1, ..., \lambda_d, k_d)$ in (2.3.5) since they involve the product of d different matrix exponentials. However, explicit formulae can be derived in certain special cases. The next two results in Proposition 2.3.8 and Theorem 2.3.10 give different methods for the calculation of these coefficients provided that a CARMA(p,q) random field is given.

Proposition 2.3.8. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field on $\mathbb{R}^d$ such that $A_1 = \cdots = A_d$ and the polynomials $a_1(\cdot)$ and $b(\cdot)$ given in Definition 2.3.3 and (2.1.3) have no common roots. Then its kernel g as given in (2.3.5) can be written as*

$$g(s) = \sum_{\lambda_1} \frac{1}{(\mu_1(\lambda_1)-1)!}\Big[\partial_z^{\mu_1(\lambda_1)-1}(z-\lambda_1)^{\mu_1(\lambda_1)}\mathrm{e}^{z(s_1+\cdots+s_d)}b(z)/a_1(z)\Big]_{z=\lambda_1}\mathbb{1}_{\{s\geq 0\}}$$

for $s\in\mathbb{R}^d$. In particular, if A_1 has distinct eigenvalues, the CARMA(p,q) kernel reduces to

$$g(s) = \sum_{\lambda_1}\frac{b(\lambda_1)}{a_1'(\lambda_1)}\mathrm{e}^{\lambda_1(s_1+\cdots+s_d)}\mathbb{1}_{\{s\geq 0\}}, \quad s\in\mathbb{R}^d,$$

where $a_1'(\cdot)$ is the derivative of the polynomial $a_1(\cdot)$.

Proof. This follows directly from Lemma 2.3 in [24]. □

Lemma 2.3.9. *Suppose that $\phi(\cdot)$ and $\theta(\cdot)$ are two (complex) polynomials such that $\phi(\cdot)$ has distinct roots, which have strictly negative real parts. Furthermore, assume that ρ is a simple closed curve encircling the roots of $\phi(\cdot)$ in the complex plane. Then we have for every $s\in\mathbb{R}$ that*

$$\frac{1}{2\pi\boldsymbol{i}}\int_\rho \frac{\theta(z)}{\phi(z)}\mathrm{e}^{sz}\,\mathrm{d}z = \sum_\lambda \frac{\theta(\lambda)}{\phi'(\lambda)}\mathrm{e}^{s\lambda},$$

where $\sum_\lambda$ denotes the sum over the distinct roots of $\phi(\cdot)$.

Proof. Let $\theta(\cdot)$ has representation $\theta(z) = \sum_{k=0}^n \theta_k z^k$. Then, we observe that

$$\frac{1}{2\pi\boldsymbol{i}}\int_\rho \frac{\theta(z)}{\phi(z)}\mathrm{e}^{sz}\,\mathrm{d}z = \sum_{k=0}^n \frac{\theta_k}{2\pi\boldsymbol{i}}\int_\rho \frac{z^k}{\phi(z)}\mathrm{e}^{sz}\,\mathrm{d}z = \sum_{k=0}^n \theta_k \sum_\lambda \frac{\lambda^k}{\phi'(\lambda)}\mathrm{e}^{s\lambda} = \sum_\lambda \frac{\theta(\lambda)}{\phi'(\lambda)}\mathrm{e}^{s\lambda},$$

where in the second equation we have used the residue theorem and evaluated the residues. □

Theorem 2.3.10. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field such that the polynomial $a_i(\cdot)$ as given in Definition 2.3.3 has distinct roots for $i = 1,\ldots,d$. Let $a_i(z) = \sum_{l=0}^p \alpha_{i,l} z^{p-l}$ and define the polynomials*

$$a_{i,k}(z) := \sum_{l=0}^{p-k}\alpha_{i,l}z^{p-k-l},$$

for $k = 1, ..., p$ and $i = 1, ..., d$. Then the CARMA(p, q) kernel g of Y as given in (2.3.5) *can be written as*

$$g(s) = \sum_{\lambda_1} \cdots \sum_{\lambda_d} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_d=1}^{p} b_{k_1-1} \frac{\lambda_d^{k_d-1}}{a_d'(\lambda_d)} \prod_{i=1}^{d-1} \frac{\lambda_i^{k_i-1} a_{i,k_{i+1}}(\lambda_i)}{a_i'(\lambda_i)} \right) \mathrm{e}^{\lambda_1 s_1 + \cdots + \lambda_d s_d} \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^d. \tag{2.3.6}$$

Proof. Denoting the (k, l)-entry of the matrix $e^{A_i s_i}$ with $m_{k,l}^{(i)}$, we have by the definition of the matrix product that

$$b^\top \mathrm{e}^{A_1 s_1} \cdots \mathrm{e}^{A_d s_d} c = \sum_{k_1=1}^{p} \cdots \sum_{k_{d+1}=1}^{p} b_{k_1-1} m_{k_1,k_2}^{(1)} m_{k_2,k_3}^{(2)} \cdots m_{k_d,k_{d+1}}^{(d)} c_{k_{d+1}},$$

where we use the convention that $b = (b_0, ..., b_{p-1})^\top$. Theorem 2.1 in [42] implies that

$$m_{k,l}^{(i)} = \frac{1}{2\pi \boldsymbol{i}} \int_{\rho_i} \frac{z^{k-1} a_{i,l}(z)}{a_i(z)} \mathrm{e}^{z s_i} \, \mathrm{d}z,$$

where the contour integral is taken over a simple closed curve ρ_i encircling the eigenvalues of A_i in the open left half of the complex plane. As a consequence, we get that

$$\begin{aligned} g(s) = \sum_{k_1=1}^{p} \cdots \sum_{k_{d+1}=1}^{p} b_{k_1-1} \frac{1}{2\pi \boldsymbol{i}} \int_{\rho_1} \frac{z^{k_1-1} a_{1,k_2}(z)}{a_1(z)} \mathrm{e}^{z s_1} \, \mathrm{d}z \\ \times \cdots \times \frac{1}{2\pi \boldsymbol{i}} \int_{\rho_d} \frac{z^{k_d-1} a_{d,k_{d+1}}(z)}{a_d(z)} \mathrm{e}^{z s_d} \, \mathrm{d}z c_{k_{d+1}} \mathbb{1}_{\{s \geq 0\}} \end{aligned} \tag{2.3.7}$$

for $s \in \mathbb{R}^d$. Applying Lemma 2.3.9, we obtain that

$$g(s) = \sum_{k_1=1}^{p} \cdots \sum_{k_{d+1}=1}^{p} \sum_{\lambda_1} \cdots \sum_{\lambda_d} b_{k_1-1} \frac{\lambda_1^{k_1-1} a_{1,k_2}(\lambda_1)}{a_1'(\lambda_1)} \mathrm{e}^{\lambda_1 s_1} \cdots \frac{\lambda_d^{k_d-1} a_{d,k_{d+1}}(\lambda_d)}{a_d'(\lambda_d)} \mathrm{e}^{\lambda_d s_d} c_{k_{d+1}} \mathbb{1}_{\{s \geq 0\}},$$

which after rearranging terms and recalling that $c = (0, ..., 0, 1)^\top$ yields Equation (2.3.6). □

Remark 2.3.11 In the setting of Theorem 2.3.10, Equation (2.3.6) reduces for $d = 1$ to

$$g(s) = \sum_{\lambda_1} \frac{b(\lambda_1)}{a_1'(\lambda_1)} \mathrm{e}^{\lambda_1 s_1} \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R},$$

which is the known kernel representation of a causal CARMA process (cf. Remark 5 in [21]). For $d = 2$, Equation (2.3.6) reduces to

$$g(s) = \sum_{\lambda_1} \sum_{\lambda_2} \left(\sum_{k=1}^{p} \frac{b(\lambda_1) a_{1,k}(\lambda_1) \lambda_2^{k-1}}{a_1'(\lambda_1) a_2'(\lambda_2)} \right) \mathrm{e}^{\lambda_1 s_1 + \lambda_2 s_2} \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^2.$$

□

Remark 2.3.12 At the end of this section we reveal a connection between CARMA random fields and Volterra-type Ornstein-Uhlenbeck (VOU) processes as studied in Pham and Chong [74]. A VOU process $(W(t,x))_{(t,x)\in\mathbb{R}_+\times\mathbb{R}^d}$ is a solution to the stochastic tempo-spatial integral equation

$$\begin{aligned} W(t,x) = V(t,x) &+ \int_0^t \int_{\mathbb{R}^d} W(t-s, x-y)\, \mu_{VOU}(\mathrm{d}s, \mathrm{d}y) \\ &+ \int_0^t \int_{\mathbb{R}^d} g_{VOU}(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y), \end{aligned}$$

where μ_{VOU} is a signed measure on $\mathbb{R}_+ \times \mathbb{R}^d$, $g_{VOU}\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function, V is a stochastic process on $\mathbb{R}_+ \times \mathbb{R}^d$ and Λ is a homogeneous Lévy basis. Under the conditions of Theorem 3.3 in [74] and the particular choice of V specified therein, the unique solution to this equation is given by

$$W(t,x) = \int_{-\infty}^t \int_{\mathbb{R}^d} (g_{VOU} - \rho_{VOU} * g_{VOU})(t-s, x-y)\, \Lambda(\mathrm{d}s, \mathrm{d}y), \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d, \tag{2.3.8}$$

where $*$ denotes convolution and the resolvent ρ_{VOU} is another signed measure on $\mathbb{R}_+ \times \mathbb{R}^d$ which is uniquely determined by μ_{VOU} through $\rho_{VOU} * \rho_{VOU} = \rho_{VOU} + \rho_{VOU}$ (cf. Proposition 2.2 in [74]).

We want to show that GCARMA random fields, and thus also CARMA(p,q) random fields, are parametric examples of VOU processes. In order to do so, we consider a GCARMA random field Y on $\mathbb{R}^{d+1}$. Since VOU processes are formulated in space and time, we write Y as a function of $(t,x) = (t, x_1, ..., x_d) \in \mathbb{R}^{d+1}$ instead of $t = (t_1, ..., t_{d+1}) \in \mathbb{R}^{d+1}$ in this remark. Further, we assume for simplicity that the matrices A_1,...,A_{d+1} in the definition of Y all have distinct eigenvalues. In this case, the GCARMA random field Y satisfies

$$Y(t,x) = \int_{-\infty}^t \cdots \int_{-\infty}^{x_d} \sum_{\lambda_1} \cdots \sum_{\lambda_{d+1}} d(\lambda_1, ..., \lambda_{d+1}) \mathrm{e}^{\lambda_1 (t-s)} \cdots \mathrm{e}^{\lambda_{d+1}(x_d - y_d)}\, \Lambda(\mathrm{d}s, \mathrm{d}y) \tag{2.3.9}$$

for $(t,x) \in \mathbb{R} \times \mathbb{R}^d$, which can be seen from Equations (2.3.4) and (2.3.5). The task is now to find a suitable function g_{VOU} and a suitable measure μ_{VOU} such that the random field Y is of the form (2.3.8). For μ_{VOU} we may choose

$$\mu_{VOU} = -\lambda \mathrm{Leb}_{\mathbb{R}_+} \otimes \delta_{0,\mathbb{R}^d}$$

with some arbitrary real number $\lambda > 0$. Here, $\mathrm{Leb}_{\mathbb{R}_+}$ denotes the Lebesgue measure on $\mathbb{R}_+$ and $\delta_{0,\mathbb{R}^d}$ is the Dirac measure on $\mathbb{R}^d$. According to Example B.2 in [74], the corresponding resolvent ρ_{VOU} is then given by

$$\rho_{VOU}(\mathrm{d}s, \mathrm{d}y) = \lambda \mathrm{e}^{-\lambda s} \, \mathrm{d}s \, \delta_{0,\mathbb{R}^d}(\mathrm{d}y).$$

With this in mind, we set

$$g_{VOU}(s,y) = \sum_{\lambda_1} \left(\frac{\lambda_1 + \lambda}{\lambda_1} \mathrm{e}^{\lambda_1 s} - \frac{\lambda}{\lambda_1} \right) \left(\sum_{\lambda_2} \cdots \sum_{\lambda_{d+1}} d(\lambda_1, ..., \lambda_{d+1}) \mathrm{e}^{\lambda_2 y_1} \cdots \mathrm{e}^{\lambda_{d+1} y_d} \mathbb{1}_{\{y \geq 0\}} \right)$$

for $(s,y) \in \mathbb{R} \times \mathbb{R}^d$, and a basic calculation yields

$$\begin{aligned} &(\rho_{VOU} * g_{VOU})(s,y) \\ &\quad = \sum_{\lambda_1} \left(\frac{\lambda}{\lambda_1} \mathrm{e}^{\lambda_1 s} - \frac{\lambda}{\lambda_1} \right) \left(\sum_{\lambda_2} \cdots \sum_{\lambda_{d+1}} d(\lambda_1, ..., \lambda_{d+1}) \mathrm{e}^{\lambda_2 y_1} \cdots \mathrm{e}^{\lambda_{d+1} y_d} \mathbb{1}_{\{y \geq 0\}} \right). \end{aligned}$$

Plugging these two equations into (2.3.8) we observe that the two random fields in (2.3.8) and (2.3.9) coincide, giving us the desired result. □

2.4 Distributional and path properties

In this section we examine several features of CARMA(p,q) random fields. We investigate their autocovariance and their spectral density, followed by some path properties. Moreover, we analyze in detail the restriction on an equidistant discrete lattice.

2.4.1 Second-order structure

The first result in this section determines the autocovariance function. We use the convention that $\mathrm{Cov}[V,W]$ denotes the matrix $(\mathrm{Cov}[V_i, W_j])_{1 \leq i,j \leq d}$ for any two random vectors $V, W \in \mathbb{R}^d$.

Theorem 2.4.1. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) (resp. GCARMA) random field.*

(1) If Λ has a finite first moment, then $Y(t)$ and $X(t)$ have as well for all $t\in\mathbb{R}^d$. They are given for $t\in\mathbb{R}^d$ by

$$\mathbb{E}[X(t)] = \kappa_1 \int_{\mathbb{R}_+^d} \mathrm{e}^{A_1 s_1}\cdots\mathrm{e}^{A_d s_d} c\,\mathrm{d}s \quad \text{and} \quad \mathbb{E}[Y(t)] = \kappa_1 \int_{\mathbb{R}_+^d} b^\top \mathrm{e}^{A_1 s_1}\cdots\mathrm{e}^{A_d s_d} c\,\mathrm{d}s.$$

(2) If Λ has further a finite second moment, then $Y(t)$ and $X(t)$ have as well for all $t\in\mathbb{R}^d$. They are given for $t\in\mathbb{R}^d$ by

$$\mathrm{Var}[X(t)] = \Sigma := \kappa_2 \int_{\mathbb{R}_+^d} \mathrm{e}^{A_1 s_1}\cdots\mathrm{e}^{A_d s_d} cc^\top \mathrm{e}^{A_d^\top s_d}\cdots\mathrm{e}^{A_1^\top s_1}\,\mathrm{d}s \quad \text{and} \quad \mathrm{Var}[Y(t)] = b^\top \Sigma b.$$

In this case, the autocovariance function γ of Y has the form

$$\begin{aligned}\gamma(t) = \kappa_2 \sum_{\lambda_1} \sum_{k_1=0}^{\mu_1(\lambda_1)-1} \cdots \sum_{\lambda_d} \sum_{k_d=0}^{\mu_d(\lambda_d)-1} \sum_{v\in\{-1,1\}^d} d_v(\lambda_1,k_1,\ldots,\lambda_d,k_d)\mathbb{1}_{\{t\odot v\in\mathbb{R}_+^d\}} \\ \times t_1^{k_1}\mathrm{e}^{\lambda_1|t_1|}\cdots t_d^{k_d}\mathrm{e}^{\lambda_d|t_d|}, \quad t\in\mathbb{R}^d,\end{aligned} \tag{2.4.1}$$

where $\{d_v(\lambda_1,k_1,\ldots,\lambda_d,k_d)\}$ is a set of complex coefficients for every $v\in\{-1,1\}^d$ such that

$$d_v(\lambda_1,k_1,\ldots,\lambda_d,k_d) = d_{-v}(\lambda_1,k_1,\ldots,\lambda_d,k_d),$$

$\sum_{\lambda_i}$ denotes the sum over distinct eigenvalues of A_i for $i=1,\ldots,d$ and $\mu_i(\lambda_i)$ is the algebraic multiplicity of the eigenvalue λ_i with respect to the matrix A_i.

(3) If even further all A_i, $i=1,\ldots,d$, commute, then the autocovariance function γ of Y has representation

$$\gamma(t) = b^\top \mathrm{e}^{A_1|t_1|\mathbb{1}_{\{t_1\geq 0\}}}\cdots\mathrm{e}^{A_d|t_d|\mathbb{1}_{\{t_d\geq 0\}}}\Sigma\mathrm{e}^{A_1^\top|t_1|\mathbb{1}_{\{t_1<0\}}}\cdots\mathrm{e}^{A_d^\top|t_d|\mathbb{1}_{\{t_d<0\}}} b, \quad t\in\mathbb{R}^d. \tag{2.4.2}$$

Proof. We prove the statements for $d=2$, the proof for higher dimensions is completely analogous. The expressions for $\mathbb{E}[X(t)]$, $\mathbb{E}[Y(t)]$, $\mathrm{Var}[X(t)]$ and $\mathrm{Var}[Y(t)]$ are consequences of Corollary 4.2 in [74]. Additionally, a consideration of the involved limits of integration yields for $r\in\mathbb{R}^2$ and $t\in\mathbb{R}_+^2$ that

$$\mathrm{Cov}[X(r+t),X(r)] = \kappa_2 \int_{\mathbb{R}_+^2} \mathrm{e}^{A_1(s_1+t_1)}\mathrm{e}^{A_2(s_2+t_2)} cc^\top \mathrm{e}^{A_2^\top s_2}\mathrm{e}^{A_1^\top s_1}\,\mathrm{d}s, \tag{2.4.3}$$

and

$$\mathrm{Cov}[X(r+(t_1,-t_2)^\top),X(r)] = \kappa_2 \int_{\mathbb{R}^2_+} \mathrm{e}^{A_1(s_1+t_1)}\mathrm{e}^{A_2 s_2} cc^\top \mathrm{e}^{A_2^\top(s_2+t_2)}\mathrm{e}^{A_1^\top s_1}\,\mathrm{d}s. \quad (2.4.4)$$

Since $Y(r) = b^\top X(r)$, we have that $\mathrm{Cov}[Y(r+t),Y(r)] = b^\top \mathrm{Cov}[X(r+t),X(r)]b$ and $\mathrm{Cov}[Y(r+(t_1,-t_2)^\top),Y(r)] = b^\top \mathrm{Cov}[X(r+(t_1,-t_2)^\top),X(r)]b$. Using the symmetry of γ and a similar argument as in the proof of Theorem 2.3.5, we obtain formula (2.4.1). If the matrices A_1 and A_2 commute, then Equations (2.4.3) and (2.4.4) simplify to

$$\mathrm{Cov}[X(r+t),X(r)] = \mathrm{e}^{A_1 t_1}\mathrm{e}^{A_2 t_2}\Sigma,$$

and

$$\mathrm{Cov}[X(r+(t_1,-t_2)^\top),X(r)] = \mathrm{e}^{A_1 t_1}\Sigma\mathrm{e}^{A_2^\top t_2},$$

which proves Equation (2.4.2). □

Remark 2.4.2 (1) For $d = 2$ Equations (2.4.1) and (2.4.2) simplify to

$$\gamma(t) = \kappa_2 \sum_{\lambda_1} \sum_{k_1=0}^{\mu_1(\lambda_1)-1} \sum_{\lambda_2} \sum_{k_2=0}^{\mu_2(\lambda_2)-1} \Big(d_{(1,1)}(\lambda_1,k_1,\lambda_2,k_2)\mathbb{1}_{\{t_1t_2\geq 0\}} + d_{(1,-1)}(\lambda_1,k_1,\lambda_2,k_2)\mathbb{1}_{\{t_1t_2<0\}}\Big) t_1^{k_1}\mathrm{e}^{\lambda_1|t_1|}t_2^{k_2}\mathrm{e}^{\lambda_2|t_2|}, \quad t\in\mathbb{R}^2,$$

and

$$\gamma(t) = \begin{cases} b^\top \mathrm{e}^{A_1|t_1|}\mathrm{e}^{A_2|t_2|}\Sigma b, & \text{if } t_1t_2 \geq 0, \\ b^\top \mathrm{e}^{A_1|t_1|}\Sigma\mathrm{e}^{A_2^\top|t_2|} b, & \text{if } t_1t_2 < 0. \end{cases}$$

(2) If all matrices $A_1,\ldots,A_d$ have distinct eigenvalues, then all eigenvalues have algebraic multiplicity one. We write $d_v(\lambda_1,\ldots,\lambda_d)$ instead of $d_v(\lambda_1,0,\ldots,\lambda_d,0)$ in Equation (2.4.1) in this case.

□

Theorem 2.4.1 tells us that the value of the autocovariance function $\gamma(t)$ depends on the *quadrant* of $\mathbb{R}^2$, and more generally on the *orthant* of $\mathbb{R}^d$, in which t lies. This unique second-order structure is basically induced by the causality feature of the random field Y. It is in particular neither isotropic, in contrast to the CARMA model in [25] (see Theorem 2 in this reference), nor separable. Also, we remark that two companion matrices commute if and only if they are equal and thus the third part of Theorem 2.4.1 actually requires $A_1 = \cdots = A_d$ in the case of a CARMA(p,q) random field Y.

The next result relates the coefficients $d_v(\lambda_1, k_1, ..., \lambda_d, k_d)$ in Equation (2.4.1) to the the coefficients $d(\lambda_1, k_1, ..., \lambda_d, k_d)$ of the kernel g in Equation (2.3.5). For brevity, we only deal with the case where each matrix A_i has distinct eigenvalues and $d = 2$.

Proposition 2.4.3. *Suppose that $(Y(t))_{t\in\mathbb{R}^2}$ is a CARMA(p,q) (resp. GCARMA) random field on the plane $\mathbb{R}^2$ such that Λ has a finite second moment, both A_1 and A_2 have distinct eigenvalues and the GCARMA kernel is given by*

$$g(s) = \sum_{i_1,i_2=1}^{p} d(i_1,i_2)\mathrm{e}^{\lambda_1(i_1)s_1}\mathrm{e}^{\lambda_2(i_2)s_2}\mathbb{1}_{\{s\geq 0\}}, \quad s \in \mathbb{R}^2,$$

where $(\lambda_n(i_n))_{1\leq i_n\leq p}$ is an enumeration of the distinct eigenvalues of A_n for $n = 1, 2$. Then the autocovariance function of Y is

$$\gamma(t) = \kappa_2 \sum_{i_1,i_2=1}^{p}\left(\sum_{j_1,j_2=1}^{p}\frac{d(i_1,i_2)d(j_1,j_2)}{(\lambda_1(j_1)+\lambda_1(i_1))(\lambda_2(j_2)+\lambda_2(i_2))}\right)\mathrm{e}^{\lambda_1(i_1)t_1}\mathrm{e}^{\lambda_2(i_2)t_2},$$

if $t_1t_2 \geq 0$, and

$$\gamma(t) = \kappa_2 \sum_{i_1,i_2=1}^{p}\left(\sum_{j_1,j_2=1}^{p}\frac{d(i_1,j_2)d(j_1,i_2)}{(\lambda_1(j_1)+\lambda_1(i_1))(\lambda_2(j_2)+\lambda_2(i_2))}\right)\mathrm{e}^{\lambda_1(i_1)t_1}\mathrm{e}^{\lambda_2(i_2)t_2},$$

if $t_1t_2 < 0$. In particular, we have for all $i_1, i_2 = 1, ..., p$ that

$$\begin{aligned} d_{(1,1)}(\lambda_1(i_1),\lambda_2(i_2)) &= d_{(-1,-1)}(\lambda_1(i_1),\lambda_2(i_2)) \\ &= \sum_{j_1,j_2=1}^{p}\frac{d(i_1,i_2)d(j_1,j_2)}{(\lambda_1(j_1)+\lambda_1(i_1))(\lambda_2(j_2)+\lambda_2(i_2))}, \end{aligned}$$

and

$$\begin{aligned} d_{(1,-1)}(\lambda_1(i_1),\lambda_2(i_2)) &= d_{(-1,1)}(\lambda_1(i_1),\lambda_2(i_2)) \\ &= \sum_{j_1,j_2=1}^{p}\frac{d(i_1,j_2)d(j_1,i_2)}{(\lambda_1(j_1)+\lambda_1(i_1))(\lambda_2(j_2)+\lambda_2(i_2))}. \end{aligned}$$

Proof. It is sufficient to consider the case when $t_1 \geq 0$ and $t_2 \geq 0$ since all other

cases follow analogously. Once again, Corollary 4.2 in [74] implies that

$$\begin{aligned}
\gamma(t) &= \kappa_2 \int_{\mathbb{R}^2_+} g(s)g(s+t)\,\mathrm{d}s \\
&= \kappa_2 \int_{\mathbb{R}^2_+} \left(\sum_{i_1,i_2=1}^{p} d(i_1,i_2)\mathrm{e}^{\lambda_1(i_1)s_1}\mathrm{e}^{\lambda_2(i_2)s_2} \right) \\
&\qquad\qquad \times \left(\sum_{i_1,i_2=1}^{p} d(i_1,i_2)\mathrm{e}^{\lambda_1(i_1)(s_1+t_1)}\mathrm{e}^{\lambda_2(i_2)(s_2+t_2)} \right) \mathrm{d}s \\
&= \kappa_2 \sum_{i_1,i_2=1}^{p} d(i_1,i_2)\mathrm{e}^{\lambda_1(i_1)t_1}\mathrm{e}^{\lambda_2(i_2)t_2} \\
&\quad \times \left(\int_{\mathbb{R}^2_+} \sum_{j_1,j_2=1}^{p} d(j_1,j_2)\mathrm{e}^{(\lambda_1(j_1)+\lambda_1(i_1))s_1}\mathrm{e}^{(\lambda_2(j_2)+\lambda_2(i_2))s_2}\,\mathrm{d}s \right) \\
&= \kappa_2 \sum_{i_1,i_2=1}^{p} d(i_1,i_2)\mathrm{e}^{\lambda_1(i_1)t_1}\mathrm{e}^{\lambda_2(i_2)t_2} \left(\sum_{j_1,j_2=1}^{p} \frac{d(j_1,j_2)}{(\lambda_1(j_1)+\lambda_1(i_1))(\lambda_2(j_2)+\lambda_2(i_2))} \right) \\
&= \kappa_2 \sum_{i_1,i_2=1}^{p} \left(\sum_{j_1,j_2=1}^{p} \frac{d(i_1,i_2)d(j_1,j_2)}{(\lambda_1(j_1)+\lambda_1(i_1))(\lambda_2(j_2)+\lambda_2(i_2))} \right) \mathrm{e}^{\lambda_1(i_1)t_1}\mathrm{e}^{\lambda_2(i_2)t_2}.
\end{aligned}$$

□

From Theorem 2.4.1 we observe that the autocovariance function γ of a GCARMA random field is integrable over $\mathbb{R}^d$. This property is also called *short-range dependency* and it implies that the spectral density f of γ exists, which is defined as

$$f(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \gamma(t)\mathrm{e}^{-\boldsymbol{i}\omega\cdot t}\,\mathrm{d}t, \quad \omega \in \mathbb{R}^d.$$

We present an explicit formula in the case when all matrices A_i have distinct eigenvalues.

Corollary 2.4.4. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) (resp. GCARMA) random field such that Λ has a finite second moment and A_i has distinct eigenvalues for $i = 1,...,d$. Then the spectral density f of Y has the form*

$$f(\omega) = \frac{\kappa_2}{(2\pi)^d} \sum_{\lambda_1} \cdots \sum_{\lambda_d} \sum_{v\in\{-1,1\}^d} \frac{d_v(\lambda_1,...,\lambda_d)}{(\boldsymbol{i}v_1\omega_1 - \lambda_1)\cdots(\boldsymbol{i}v_d\omega_d - \lambda_d)}, \quad \omega \in \mathbb{R}^d. \tag{2.4.5}$$

Proof. Equation (2.4.1) shows that

$$\gamma(t) = \kappa_2 \sum_{\lambda_1} \cdots \sum_{\lambda_d} \sum_{v\in\{-1,1\}^d} d_v(\lambda_1,...,\lambda_d)\mathbb{1}_{\{t\odot v\in\mathbb{R}^d_+\}}\mathrm{e}^{\lambda_1|t_1|+\cdots+\lambda_d|t_d|}, \quad t \in \mathbb{R}^d.$$

We also have that

$$\int_{\mathbb{R}} \mathbb{1}_{\{t_i v_i \geq 0\}} \mathrm{e}^{\lambda_i |t_i|} \mathrm{e}^{-\boldsymbol{i}\omega_i t_i} \,\mathrm{d}t_i = \frac{1}{\boldsymbol{i}v_i\omega_i - \lambda_i}$$

for $i = 1, ..., d$ and $\omega_i \in \mathbb{R}$. Hence, we get (2.4.5) by applying the Fourier transform. □

Example 2.4.5 (Second-order structure of a CAR(p) random field on the plane)
Let $(Y(t))_{t\in\mathbb{R}^2}$ be a CAR(p) random field on the plane $\mathbb{R}^2$ such that Λ has a finite second moment and both A_1 and A_2 have distinct eigenvalues. Recall from Equation (2.3.7) that the CAR(p) kernel g has the representation

$$g(s) = \sum_{k=1}^{p} \frac{1}{(2\pi\boldsymbol{i})^2} \int_{\rho_1} \frac{a_{1,k}(z)}{a_1(z)} \mathrm{e}^{zs_1} \,\mathrm{d}z \int_{\rho_2} \frac{z^{k-1}}{a_2(z)} \mathrm{e}^{zs_2} \,\mathrm{d}z \mathbb{1}_{\{s\geq 0\}}, \quad s = (s_1, s_2) \in \mathbb{R}^2,$$

where we use the notation of Theorem 2.3.10. Since $|a_{1,k}(z)/a_1(z)| = \mathcal{O}(|z|^{-1})$ as $z \to \infty$, an application of Theorem 2.2 of Chapter VI in [64] yields

$$\begin{aligned} g(s) &= \sum_{k=1}^{p} \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \frac{a_{1,k}(\boldsymbol{i}\omega_1)}{a_1(\boldsymbol{i}\omega_1)} \mathrm{e}^{\boldsymbol{i}\omega_1 s_1} \,\mathrm{d}\omega_1 \int_{\mathbb{R}} \frac{(\boldsymbol{i}\omega_2)^{k-1}}{a_2(\boldsymbol{i}\omega_2)} \mathrm{e}^{\boldsymbol{i}\omega_2 s_2} \,\mathrm{d}\omega_2 \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \sum_{k=1}^{p} \frac{a_{1,k}(\boldsymbol{i}\omega_1)(\boldsymbol{i}\omega_2)^{k-1}}{a_1(\boldsymbol{i}\omega_1)a_2(\boldsymbol{i}\omega_2)} \mathrm{e}^{\boldsymbol{i}(\omega_1 s_1 + \omega_2 s_2)} \,\mathrm{d}\omega_1 \,\mathrm{d}\omega_2, \quad s = (s_1, s_2) \in \mathbb{R}^2. \end{aligned}$$

This allows us to recognize that the Fourier transform of g is equal to

$$\tilde{g}(\omega) = \sum_{k=1}^{p} \frac{a_{1,k}(\boldsymbol{i}\omega_1)(\boldsymbol{i}\omega_2)^{k-1}}{a_1(\boldsymbol{i}\omega_1)a_2(\boldsymbol{i}\omega_2)}, \quad \omega = (\omega_1, \omega_2) \in \mathbb{R}^2,$$

which immediately implies the spectral density

$$f(\omega) = \frac{\kappa_2}{(2\pi)^2} |\tilde{g}(\omega_)|^2 = \frac{\kappa_2}{(2\pi)^2} \left(\sum_{k=1}^{p} \frac{a_{1,k}(\boldsymbol{i}\omega_1)(\boldsymbol{i}\omega_2)^{k-1}}{a_1(\boldsymbol{i}\omega_1)a_2(\boldsymbol{i}\omega_2)} \right) \left(\sum_{l=1}^{p} \frac{a_{1,l}(-\boldsymbol{i}\omega_1)(-\boldsymbol{i}\omega_2)^{l-1}}{a_1(-\boldsymbol{i}\omega_1)a_2(-\boldsymbol{i}\omega_2)} \right).$$

Furthermore, we conclude for the autocovariance function that

$$\begin{aligned} \gamma(t) &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(\omega) \mathrm{e}^{\boldsymbol{i}(\omega_1 t_1 + \omega_2 t_2)} \,\mathrm{d}\omega_1 \,\mathrm{d}\omega_2 \\ &= \frac{\kappa_2}{(2\pi)^2} \sum_{k,l=1}^{p} \int_{\mathbb{R}} \frac{a_{1,k}(\boldsymbol{i}\omega_1)a_{1,l}(-\boldsymbol{i}\omega_1)}{a_1(\boldsymbol{i}\omega_1)a_1(-\boldsymbol{i}\omega_1)} \mathrm{e}^{\boldsymbol{i}\omega_1 t_1} \,\mathrm{d}\omega_1 \int_{\mathbb{R}} \frac{(\boldsymbol{i}\omega_2)^{k-1}(-\boldsymbol{i}\omega_2)^{l-1}}{a_2(\boldsymbol{i}\omega_2)a_2(-\boldsymbol{i}\omega_2)} \mathrm{e}^{\boldsymbol{i}\omega_2 t_2} \,\mathrm{d}\omega_2 \\ &= \kappa_2 \sum_{k,l=1}^{p} \left(\sum_{\lambda_1} \frac{a_{1,k}(\lambda_1)a_{1,l}(-\lambda_1)\mathbb{1}_{\{t_1\geq 0\}} + a_{1,k}(-\lambda_1)a_{1,l}(\lambda_1)\mathbb{1}_{\{t_1<0\}}}{a_1'(\lambda_1)a_1(-\lambda_1)} \mathrm{e}^{\lambda_1|t_1|} \right) \end{aligned}$$

$$\times \left(\sum_{\lambda_2} \frac{\lambda_2^{k-1}(-\lambda_2)^{l-1}\mathbb{1}_{\{t_2\geq 0\}} + (-\lambda_2)^{k-1}\lambda_2^{l-1}\mathbb{1}_{\{t_2<0\}}}{a_2'(\lambda_2)a_2(-\lambda_2)} \mathrm{e}^{\lambda_2|t_2|} \right)$$

$$= \kappa_2 \sum_{\lambda_1}\sum_{\lambda_2} \Bigg[\sum_{k,l=1}^{p} \frac{a_{1,k}(\lambda_1)a_{1,l}(-\lambda_1)\lambda_2^{k+l-2}}{a_1'(\lambda_1)a_1(-\lambda_1)a_2'(\lambda_2)a_2(-\lambda_2)}$$

$$\times \left((-1)^{l-1}\mathbb{1}_{\{t_1t_2\geq 0\}} + (-1)^{k-1}\mathbb{1}_{\{t_1t_2<0\}}\right) \Bigg] \mathrm{e}^{\lambda_1|t_1|+\lambda_2|t_2|}, \quad t = (t_1, t_2) \in \mathbb{R}^2,$$

where in the third equation we have used Lemma 2.3.9 and Theorem 2.2 of Chapter VI in [64]. We remark that the procedure in this example cannot be extended to CARMA(p, q) random fields with $q > 0$ since $|b(z)a_{1,k}(z)/a_1(z)| = \mathcal{O}(|z|^{-1})$ would not be satisfied for each $k = 1, ..., p$. □

2.4.2 Path properties

Path properties for CARMA(p, q) processes can easily be deduced from Equations (2.1.1). The sample paths of each CARMA(p, q) process are $(p - q - 2)$-times differentiable, provided $p > q + 2$, or continuous, provided $p = q + 2$, or càdlàg, provided $p = q + 1$ (see e.g. Equation (31) in [21] for more details). If additionally the driving noise is a Brownian motion, then even more regularity can be obtained.

For the class of spatial CARMA(p, q) random fields it is harder to establish path properties since the system of SPDEs (2.3.2) does not allow for the same reasoning as (2.1.1) does. However, by drawing on maximal inequalities for multi-parameter martingales, we are able to prove the existence of a version having the path property in Definition 2.4.6, which also represents a possible generalization of the classical càdlàg property. As usual, we say that a process $(\tilde{Y}(t))_{t\in\mathbb{R}^d}$ is a version of the process $(Y(t))_{t\in\mathbb{R}^d}$ if for every $t \in \mathbb{R}^d$ the equality $\tilde{Y}(t) = Y(t)$ holds almost surely.

Definition 2.4.6

(1) We write $v \not\leq w$ if and only if $v_i > w_i$ for at least one $i \in \{1, ..., d\}$. Also, for $v, w \in \mathbb{R}^d$ we define the interval $[v, w] := \{s \in \mathbb{R}^d : v \leq s \leq w\}$, which may be empty.

(2) A function $f : \mathbb{R}^d \to \mathbb{R}$ is *càdlàg* if for every $t \in \mathbb{R}^d$,

$$\lim_{\substack{s\to t \\ s\geq t}} f(s) = f(t) \quad \text{and} \quad \lim_{\substack{s\to t \\ s\not\geq t}} f(s) \quad \text{exists.}$$

□

Theorem 2.4.7. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) (resp. GCARMA) random field.*

(1) If the homogeneous Lévy basis Λ is Gaussian, then Y has a version which is Hölder continuous with any exponent in $(0,1/2)$.

(2) If the Lévy measure ν of Λ satisfies $\int_{|x|>1}|x|^\alpha\,\nu(\mathrm{d}x)<\infty$ for some $\alpha\in(0,1]$, then Y has a càdlàg version.

Proof. We only prove the assertions in two dimensions since higher dimensions can be treated completely analogously.

(1) Without loss of generality, we assume that $\beta=0$, that is, Λ has mean zero. Recall from Equation (2.4.1) that the autocovariance function is given by

$$\begin{aligned}\gamma(t)=\kappa_2\sum_{\lambda_1}\sum_{k_1=0}^{\mu_1(\lambda_1)-1}\sum_{\lambda_2}\sum_{k_2=0}^{\mu_2(\lambda_2)-1}\Big(&d_{(1,1)}(\lambda_1,\mu_1,\lambda_2,\mu_2)\mathbb{1}_{\{t_1t_2\geq 0\}}\\&+d_{(1,-1)}(\lambda_1,\mu_1,\lambda_2,\mu_2)\mathbb{1}_{\{t_1t_2<0\}}\Big)t_1^{k_1}\mathrm{e}^{\lambda_1|t_1|}t_2^{k_2}\mathrm{e}^{\lambda_2|t_2|},\quad t\in\mathbb{R}^2.\end{aligned}$$

This allows us to conclude that every $t\in\mathbb{R}^d$ and every $s\in\mathbb{R}^d$ with sufficiently small norm $\|s\|$ satisfy

$$\mathbb{E}[|Y(t)-Y(t+s)|^2]=2(\gamma(0)-\gamma(s))\leq C\|s\|.$$

Hence, applying Kolmogorov's continuity theorem (see e.g. Theorem 3.23 of [57]) and the fact that Y is a Gaussian process finishes the proof of the first part.

(2) The proof of the second part is similar to the proof of Theorem 5.5 in [74]. We therefore only sketch the main ideas. Due to the first part above, we may assume that $\sigma^2=\beta=0$. The remaining compensated small jumps part and large jumps part are considered separately.

Case 1: $\Lambda(\mathrm{d}s)=\int_{\mathbb{R}}z\mathbb{1}_{\{|z|\leq 1\}}\,(\mathfrak{p}-\mathfrak{q})(\mathrm{d}s,\mathrm{d}z)$**.**

Referring to Equation (2.3.4), we observe that the claim clearly holds true for the process $g*\Lambda_n$ with $\Lambda_n(\mathrm{d}s):=\int_{\mathbb{R}}z\mathbb{1}_{\{1/n\leq|z|\leq 1\}}\mathbb{1}_{\{|s|\leq n\}}\,(\mathfrak{p}-\mathfrak{q})(\mathrm{d}s,\mathrm{d}z)$ since the latter has only finitely many jumps. Consequently, we may close this case once we are

able to show that $g * \Lambda^n$ converges uniformly on compacts in probability to 0, where $\Lambda^n := \Lambda - \Lambda_n$. By the fundamental theorem of calculus, we have for the GCARMA kernel g that

$$\begin{aligned} g(t_1 - s_1, t_2 - s_2) = g(-s_1, -s_2) &+ \int_0^{t_1} \partial_1 g(r_1 - s_1, -s_2)\,\mathrm{d}r_1 \\ &+ \int_0^{t_2} \partial_2 g(-s_1, r_2 - s_2)\,\mathrm{d}r_2 \\ &+ \int_0^{t_1}\int_0^{t_2} \partial_1\partial_2 g(r_1 - s_1, r_2 - s_2)\,\mathrm{d}r_2\mathrm{d}r_1. \end{aligned}$$

Putting this decomposition into $g * \Lambda^n$, we end up with four processes which can be handled one by one. For instance, the last one satisfies

$$\begin{aligned} &\mathbb{E}\left[\sup_{t\in[-v,v]}\left|\int_{-\infty}^{t_1}\int_{-\infty}^{t_2}\int_0^{t_1}\int_0^{t_2} \partial_1\partial_2 g(r_1 - s_1, r_2 - s_2)\,\mathrm{d}r_2\mathrm{d}r_1\Lambda^n(\mathrm{d}s_1, \mathrm{d}s_2)\right|^2\right] \\ &\quad\le C\int_{-v_1}^{v_1}\int_{-v_2}^{v_2}\mathbb{E}\left[\sup_{t\in[-v,v]}\left|\int_{-\infty}^{t_1}\int_{-\infty}^{t_2} \partial_1\partial_2 g(r_1 - s_1, r_2 - s_2)\,\Lambda^n(\mathrm{d}s_1, \mathrm{d}s_2)\right|^2\right]\mathrm{d}r_2\mathrm{d}r_1 \\ &\quad\le C\int_{-v_1}^{v_1}\int_{-v_2}^{v_2}\mathbb{E}\left[\left|\int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \partial_1\partial_2 g(r_1 - s_1, r_2 - s_2)\,\Lambda^n(\mathrm{d}s_1, \mathrm{d}s_2)\right|^2\right]\mathrm{d}r_2\mathrm{d}r_1 \end{aligned}$$

for some $v \in \mathbb{R}^2_+$. Note that we have used a stochastic Fubini theorem (see e.g. Theorem 2 in [66]) in the second line and Cairoli's maximal inequality (see e.g. Corollary 2.3.1 of Chapter 7 in [59]) in the third line, which in turn converges to zero as n tends to infinity due to the dominated convergence theorem. The three other parts in the decomposition can be dealt with analogously.

Case 2: $\Lambda(\mathrm{d}s) = \int_{\mathbb{R}} z\mathbb{1}_{\{|z|>1\}}\,\mathfrak{p}(\mathrm{d}s, \mathrm{d}z)$**.**

The difference between this case and the previous one is that

$$\Lambda_n(\mathrm{d}s) := \int_{\mathbb{R}} z\mathbb{1}_{\{|z|>1\}}\mathbb{1}_{\{|s|\le n\}}\,\mathfrak{p}(\mathrm{d}s, \mathrm{d}z)$$

and instead of decomposing g, we directly estimate

$$\begin{aligned} &\mathbb{E}\left[\sup_{t\in[-v,v]}\left|\int_{-\infty}^{t_1}\int_{-\infty}^{t_2} g(t_1 - s_1, t_2 - s_2)\,\Lambda^n(\mathrm{d}s_1, \mathrm{d}s_2)\right|^\alpha\right] \\ &\quad\le \mathbb{E}\left[\sup_{t\in[-v,v]}\int_{-\infty}^{t_1}\int_{-\infty}^{t_2}\int_{\mathbb{R}} g(t_1 - s_1, t_2 - s_2)^\alpha|z|^\alpha\mathbb{1}_{\{|z|>1\}}\mathbb{1}_{\{|s|>n\}}\,\mathfrak{p}(\mathrm{d}s, \mathrm{d}z)\right] \\ &\quad\le \int_{-\infty}^{v_1}\int_{-\infty}^{v_2}\sup_{t\in[-v,v]} g(t_1 - s_1, t_2 - s_2)^\alpha\mathbb{1}_{\{|s|>n\}}\,\mathrm{d}s_1\mathrm{d}s_2\int_{\mathbb{R}}|z|^\alpha\mathbb{1}_{\{|z|>1\}}\,\mathrm{d}z. \end{aligned}$$

The first integral in the last line is well defined and converges to zero as $n \to \infty$ by dominated convergence. □

Remark 2.4.8 The notion of càdlàg functions in Definition 2.4.6 is slightly stronger than the notion of *lamp* functions (for *limits along monotone paths*), which is for instance defined in [83] and also in [38]. □

2.4.3 Sampling on an equidistant lattice

Real-life phenomena and data thereof are usually observed and digitally stored only for a set of discrete points even if the underlying dynamics are of a continuous nature. Therefore, it is desirable to understand the behavior of a continuous model when it is discretely sampled. If the driving Lévy process has a finite second moment, it is known that an equidistantly sampled CARMA(p,q) process is always an ARMA$(p,p-1)$ process driven by a weak white noise. This fact follows from Lemma 2.1 in [24] in conjunction with Proposition 3.2.1 in [23]. We are going to generalize these two results to higher dimensions. All results in this subsection are formulated on the plane for simplicity and we use subscripts to indicate discrete parameters.

Definition 2.4.9

(1) A random field $(Y_t)_{t\in\mathbb{Z}^2}$ is called *weakly stationary* if it has finite second moments and $\mathrm{Cov}[Y_t, Y_s] = \mathrm{Cov}[Y_{t-s}, Y_0] =: \gamma(t-s)$ for every $t, s \in \mathbb{Z}^2$. It is called a *white noise* if $\gamma(t) = 0$ for every $0 \neq t \in \mathbb{Z}^2$. Furthermore, a weakly stationary random field $(Y_t)_{t\in\mathbb{Z}^2}$ is called (q_1,q_2)-*dependent* if its autocovariance function γ satisfies $\gamma(t) = 0$ whenever $|t_1| > q_1$ or $|t_2| > q_2$, and if there are points $u, v \in \mathbb{Z}^2$ such that $|u_1| = q_1$, $|v_2| = q_2$, $\gamma(u) \neq 0$ and $\gamma(v) \neq 0$. Its *spectral density* f is then defined by

$$f(\omega) = \frac{1}{(2\pi)^2} \sum_{t\in\mathbb{Z}^2} \gamma(t) \mathrm{e}^{-i\omega\cdot t}, \quad \omega \in [-\pi,\pi]^2.$$

(2) Let p_1, p_2, q_1 and q_2 be non-negative integers and $(Z_t)_{t\in\mathbb{Z}^2}$ be a white noise on $\mathbb{Z}^2$. A random field $(Y_t)_{t\in\mathbb{Z}^2}$ is called an $ARMA((p_1,p_2),(q_1,q_2))$ *random field* if it satisfies the equation

$$\sum_{k_1=0}^{p_1}\sum_{k_2=0}^{p_2} \phi_k Y_{t-k} = \sum_{k_1=0}^{q_1}\sum_{k_2=0}^{q_2} \theta_k Z_{t-k}, \quad t \in \mathbb{Z}^2,$$

where $\phi_k, \theta_k \in \mathbb{C}$ are coefficients such that $\phi_{\mathbf{0}} \neq 0$, $\theta_{\mathbf{0}} \neq 0$, at least one of the coefficients $\phi_{(p_1,\cdot)}$ is non-zero and similarly for $\phi_{(q_1,\cdot)}$,$\phi_{(\cdot,p_2)}$ and $\phi_{(\cdot,q_2)}$. This random field is also called $AR((p_1,p_2))$ *random field* if $q_1 = q_2 = 0$ and $MA((q_1,q_2))$ *random field* if $p_1 = p_2 = 0$.

(3) Let P and Q be two non-empty subsets of $\mathbb{Z}^2$ and $(Z_t)_{t\in\mathbb{Z}^2}$ be a white noise on $\mathbb{Z}^2$. A random field $(Y_t)_{t\in\mathbb{Z}^2}$ is called an $ARMA(P,Q)$ *random field* if it satisfies the equation

$$\sum_{k\in P} \phi_k Y_{t-k} = \sum_{k\in S} \theta_k Z_{t-k}, \quad t \in \mathbb{Z}^2,$$

where $\phi_k, \theta_k \in \mathbb{C}$ are non-zero coefficients. This random field is also called $AR(P)$ *random field* if $Q = \{(0,0)\}$ and $MA(Q)$ *random field* if $P = \{(0,0)\}$. □

Before we investigate the general GCARMA random field, let us consider the special case of a CAR(1) random field $(Y(t))_{t\in\mathbb{R}^2}$ first. Assuming that $b = 1$, we have the representation

$$Y(t) = \int_{-\infty}^{t_1}\int_{-\infty}^{t_2} \mathrm{e}^{\lambda_1(t_1-s_1)+\lambda_2(t_2-s_2)}\,\Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^2,$$

where the real numbers λ_1 and λ_2 are strictly negative. This allows us to observe that

$$\begin{aligned} Y(t_1,t_2) = {} & e^{\lambda_1} Y(t_1-1,t_2) + e^{\lambda_2} Y(t_1,t_2-1) - e^{\lambda_1+\lambda_2} Y(t_1-1,t_2-1) \\ & + \int_{t_1-1}^{t_1}\int_{t_2-1}^{t_2} \mathrm{e}^{\lambda_1(t_1-s_1)+\lambda_2(t_2-s_2)}\,\Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^2. \end{aligned} \tag{2.4.6}$$

Setting $Y_t := Y(t)$ for $t \in \mathbb{Z}^2$, we conclude that the sampled random field $(Y_t)_{t\in\mathbb{Z}^2}$ is an $\mathrm{AR}((1,1))$ random field driven by the i.i.d. noise

$$Z_t := \int_{t_1-1}^{t_1}\int_{t_2-1}^{t_2} \mathrm{e}^{\lambda_1(t_1-s_1)+\lambda_2(t_2-s_2)}\,\Lambda(\mathrm{d}s).$$

With a little more effort, this procedure carries over to a more general case.

Proposition 2.4.10. *Suppose that $(Y(t))_{t\in\mathbb{R}^2}$ is a GCARMA random field on the plane $\mathbb{R}^2$ such that A_1 and A_2 commute. Then the sampled random field $(Y_t)_{t\in\mathbb{Z}^2}$ satisfies the equation*

$$\sum_{k_1,k_2=0}^{p} d_k Y_{t-k} = \sum_{k_1,k_2=0}^{p-1} \left(\sum_{l_1=0}^{k_1}\sum_{l_2=0}^{k_2} d_l b^{\top} \mathrm{e}^{(k_1-l_1)A_1+(k_2-l_2)A_2} \right) R_{t-k}, \quad t \in \mathbb{Z}^2, \tag{2.4.7}$$

where the coefficients $d_k \in \mathbb{C}$ are given by

$$d_{0,0}z^p + d_{1,0}z^{p-1} + \cdots + d_{p,0} = \chi_{e^{A_1}}(z) := \prod_{\lambda_1} \prod_{k_1=0}^{\mu_1(\lambda_1)-1} (z - e^{\lambda_1})$$

$$d_{0,0}z^p + d_{0,1}z^{p-1} + \cdots + d_{0,p} = \chi_{e^{A_2}}(z) := \prod_{\lambda_2} \prod_{k_2=0}^{\mu_2(\lambda_2)-1} (z - e^{\lambda_1})$$

$$d_{k_1,k_2} = d_{k_1,0}d_{0,k_2}, \quad k_1, k_2 = 1, ..., p,$$

and the multivariate i.i.d. noise $(R_t)_{t\in\mathbb{Z}^2}$ is given by

$$R_t = \int_{t_1-1}^{t_1} \int_{t_2-1}^{t_2} \mathrm{e}^{A_1(t_1-s_1)+A_2(t_2-s_2)} c\, \Lambda(\mathrm{d}s), \quad t \in \mathbb{Z}^2.$$

In particular, the right-hand side of (2.4.7) is a $(p-1, p-1)$-dependent random field if Λ has a finite second moment.

Proof. We extend the proof of Lemma 2.1 in Brockwell and Lindner [24] which requires several additional steps that do not appear in the one-dimensional case. To this end, we show by induction that for all $t \in \mathbb{Z}^2$, $n \in \mathbb{N} \cup \{0\}$ and coefficients $f_k \in \mathbb{C}$ with $k_1, k_2 = 0, 1, ..., n$ we have

$$\begin{aligned} \sum_{k_1,k_2=0}^{n} f_k X_{t-k} = {} & \sum_{k_1,k_2=0}^{n-1} \left(\sum_{l_1=0}^{k_1} \sum_{l_2=0}^{k_2} f_l \mathrm{e}^{(k_1-l_1)A_1+(k_2-l_2)A_2} \right) R_{t-k} \\ & + \sum_{k_2=0}^{n-1} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1} \right) X_{t_1-n,t_2-k_2} + \sum_{k_1=0}^{n-1} \left(\sum_{k_2=0}^{n} f_k \mathrm{e}^{(n-k_2)A_2} \right) X_{t_1-k_1,t_2-n} \\ & + \left(f_{n,n} I_p - \sum_{k_1,k_2=0}^{n-1} f_k \mathrm{e}^{(n-k_1)A_1+(n-k_2)A_2} \right) X_{t_1-n,t_2-n} =: S_1 + S_2 + S_3 + S_4. \end{aligned} \tag{2.4.8}$$

The case $n = 0$ is trivial. Assuming that the statement is valid for some n, we observe that

$$\begin{aligned} S_2 = {} & \left(\sum_{k_1=0}^{n} f_{k_1,0} \mathrm{e}^{(n-k_1)A_1} \right) \left(e^{A_1} X_{t_1-n-1,t_2} + e^{A_2} X_{t_1-n,t_2-1} + e^{A_1+A_2} X_{t_1-n-1,t_2-1} + R_{t_1-n,t_2} \right) \\ & + \sum_{k_2=1}^{n-1} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1} \right) X_{t_1-n,t_2-k_2} \\ = {} & \left(\sum_{k_1=0}^{n} f_{k_1,0} \mathrm{e}^{(n+1-k_1)A_1} \right) X_{t_1-n-1,t_2} + \left(\sum_{k_1=0}^{n} f_{k_1,0} \mathrm{e}^{(n-k_1)A_1} \right) R_{t_1-n,t_2} \end{aligned}$$

$$
\begin{aligned}
&- \left(\sum_{k_1=0}^{n} f_{k_1,0} \mathrm{e}^{(n+1-k_1)A_1+A_2} \right) X_{t_1-n-1,t_2-1} + \left(\sum_{k_1=0}^{n} f_{k_1,0} \mathrm{e}^{(n-k_1)A_1+A_2} \right) X_{t_1-n,t_2-1} \\
&+ \sum_{k_2=1}^{n-1} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1} \right) X_{t_1-n,t_2-k_2} \\
= &\sum_{k_2=0}^{M} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n+1-k_1)A_1} \right) X_{t_1-n-1,t_2-k_2} + \sum_{k_2=0}^{M} \left(\sum_{l_1=0}^{n} \sum_{l_2=0}^{k_2} f_l \mathrm{e}^{(n-l_1)A_1+(k_2-l_2)A_2} \right) R_{t_1-n,t_2-k_2} \\
&- \left(\sum_{k_2=0}^{M} \sum_{k_1=0}^{n} f_k \mathrm{e}^{(n+1-k_1)A_1+(M+1-k_2)A_2} \right) X_{t_1-n-1,t_2-M-1} \\
&+ \left(\sum_{k_2=0}^{M} \sum_{k_1=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1+(M+1-k_2)A_2} \right) X_{t_1-n,t_2-M-1} \\
&+ \sum_{k_2=M+1}^{n-1} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1} \right) X_{t_1-n,t_2-k_2} \\
= &\sum_{k_2=0}^{n-1} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n+1-k_1)A_1} \right) X_{t_1-n-1,t_2-k_2} + \sum_{k_2=0}^{n-1} \left(\sum_{l_1=0}^{n} \sum_{l_2=0}^{k_2} f_l \mathrm{e}^{(n-l_1)A_1+(k_2-l_2)A_2} \right) R_{t_1-n,t_2-k_2} \\
&- \left(\sum_{k_2=0}^{n-1} \sum_{k_1=0}^{n} f_k \mathrm{e}^{(n+1-k_1)A_1+(n-k_2)A_2} \right) X_{t_1-n-1,t_2-n} \\
&+ \left(\sum_{k_2=0}^{n-1} \sum_{k_1=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1+(n-k_2)A_2} \right) X_{t_1-n,t_2-n} =: U_1 + U_2 + U_3 + U_4,
\end{aligned}
$$

where M is just an induction parameter ranging from 0 to $n-1$ and we have used a similar calculation as in (2.4.6). By symmetry we have that

$$
\begin{aligned}
S_3 = &\sum_{k_1=0}^{n-1} \left(\sum_{k_2=0}^{n} f_k \mathrm{e}^{(n+1-k_2)A_2} \right) X_{t_1-k_1,t_2-n-1} + \sum_{k_1=0}^{n-1} \left(\sum_{l_1=0}^{k_1} \sum_{l_2=0}^{n} f_l \mathrm{e}^{(k_1-l_1)A_1+(n-l_2)A_2} \right) R_{t_1-k_1,t_2-n} \\
&- \left(\sum_{k_2=0}^{n} \sum_{k_1=0}^{n-1} f_k \mathrm{e}^{(n-k_1)A_1+(n+1-k_2)A_2} \right) X_{t_1-n,t_2-n-1} \\
&+ \left(\sum_{k_2=0}^{n} \sum_{k_1=0}^{n-1} f_k \mathrm{e}^{(n-k_1)A_1+(n-k_2)A_2} \right) X_{t_1-n,t_2-n} =: V_1 + V_2 + V_3 + V_4.
\end{aligned}
$$

Moreover, we may sum up

$$
S_4 + U_4 + V_4 = \left(\sum_{k_1,k_2=0}^{n} f_k \mathrm{e}^{(n-k_1)A_1+(n-k_2)A_2} \right) \left(e^{A_1} X_{t_1-n-1,t_2-n} + e^{A_2} X_{t_1-n,t_2-n-1} \right.
$$

$$- e^{A_1+A_2} X_{t_1-n-1,t_2-n-1} + R_{t_1-n,t_2-n} \Bigg) =: W_1 + W_2 + W_3 + W_4,$$

and also

$$\begin{aligned}
\sum_{k_1,k_2=0}^{n} f_k X_{t-k} &= S_1 + S_2 + S_3 + S_4 \\
&= (S_1 + U_2 + V_2 + W_4) + (U_1 + U_3 + W_1) + (V_1 + V_3 + W_2) + W_3 \\
&= \sum_{k_1,k_2=0}^{n} \left(\sum_{l_1=0}^{k_1} \sum_{l_2=0}^{k_2} f_l \mathrm{e}^{(k_1-l_1)A_1+(k_2-l_2)A_2} \right) R_{t-k} \\
&\quad + \sum_{k_2=0}^{n} \left(\sum_{k_1=0}^{n} f_k \mathrm{e}^{(n+1-k_1)A_1} \right) X_{t_1-n-1,t_2-k_2} \\
&\quad + \sum_{k_1=0}^{n} \left(\sum_{k_2=0}^{n} f_k \mathrm{e}^{(n+1-k_2)A_2} \right) X_{t_1-k_1,t_2-n-1} \\
&\quad - \sum_{k_1,k_2=0}^{n} f_k \mathrm{e}^{(n+1-k_1)A_1+(n+1-k_2)A_2} X_{t_1-n-1,t_2-n-1},
\end{aligned}$$

which is equivalent to equation (2.4.8) with $n+1$ instead of n. By choosing $n = p$ and $f_k = d_k$ for $k_1, k_2 = 0, 1, ..., p$, the Cayley-Hamilton theorem implies that S_2 and S_3 in (2.4.8) vanish since $\chi_{e^{A_1}}(z)$ and $\chi_{e^{A_2}}(z)$ are the characteristic polynomials of e^{A_1} and e^{A_2}, respectively. Finally, we have that

$$\begin{aligned}
d_{p,p} I_p - \sum_{k_1,k_2=0}^{p-1} d_k \mathrm{e}^{(p-k_1)A_1+(p-k_2)A_2} &= d_{n,n} I_p - \left(\sum_{k_1=0}^{p-1} d_{k_1,0} \mathrm{e}^{(p-k_1)A_1} \right) \left(\sum_{k_2=0}^{p-1} d_{0,k_2} \mathrm{e}^{(p-k_2)A_2} \right) \\
&= d_{p,p} I_p - (-d_{p,0} I_p)(-d_{0,p} I_p) = 0,
\end{aligned}$$

which shows that S_4 vanishes, too. By multiplying $b^\top$ to the left of (2.4.8), we arrive at equation (2.4.7). □

Equation (2.4.7) implies that $(Y_t)_{t \in \mathbb{Z}^2}$ satisfies an autoregression of order (p, p) driven by a $(p-1, p-1)$-dependent noise. Proposition 3.2.1 in [23] states that a stationary time-discrete q-dependent process is a moving average process of order q, which in turn is established by projecting the process into the past with respect to the natural order of time in order to create the white noise sequence. However, projecting on the past with respect to the partial order $\leq$ on $\mathbb{Z}^2$ does not necessarily lead to spatial white noise since this order is only a partial order and not a total order.

By contrast, the lexicographic order is total and allows us to generalize Proposition 3.2.1 in [23].

Definition 2.4.11 For $v, w \in \mathbb{Z}^2$ we define $v \preceq w$ if and only if $v_1 = w_1$ and $v_2 \leq w_2$ or $v_1 < w_1$. Furthermore, we write $[v, w]_{\preceq} := \{s \in \mathbb{Z}^2 : v \preceq s \preceq w\}$, which might be empty. □

Proposition 2.4.12. *Let $(Y_t)_{t\in\mathbb{Z}^2}$ be a weakly stationary (q_1, q_2)-dependent random field. If its spectral density f satisfies $\log f \in L^1([-\pi, \pi]^2)$, then Y is a $MA([(0,0), (q_1, q_2)]_{\preceq})$ random field.*

Proof. First of all, the (q_1, q_2)-dependency of Y implies that its spectral measure is absolutely continuous. Since $\log f \in L^1([-\pi, \pi]^2)$, Theorems 1.1.2 and 1.1.4 in Korezlioglu and Loubaton [61] imply that Y satisfies the Wold decomposition

$$Y_t = \sum_{k\succeq(0,0)} \theta_k Z_{t-k}, \quad t \in \mathbb{Z}^2,$$

where $\theta_k \in \mathbb{C}$ are such that $\theta_{(0,0)} \neq 0$ and $\sum_{k\succeq(0,0)} |\theta_k|^2 < \infty$ and the white noise Z is given by

$$Z_t = Y_t - (Y_t / H^{1+}_{t_1-1,t_2-1}), \quad t \in \mathbb{Z}^2.$$

Here $(Y_t / H^{1+}_{t_1-1,t_2-1})$ denotes the orthogonal projection of Y_t on the closed linear subspace $H^{1+}_{t_1-1,t_2-1}$ of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$, which is generated by $\{Y_{\mathbf{s}} : s_1 < t_1, s_2 \in \mathbb{Z}\}$ and $\{Y_{\mathbf{s}} : s_1 = t_1, s_2 < t_2\}$. Exploiting the (q_1, q_2)-dependency once again, we see that actually

$$Y_t = \sum_{(0,0)\preceq k\preceq(q_1,q_2)} \theta_k Z_{t-k}, \quad t \in \mathbb{Z}^2.$$

□

Theorem 2.4.13. *Suppose that $(Y(t))_{t\in\mathbb{R}^2}$ is a GCARMA random field on the plane $\mathbb{R}^2$ such that A_1 and A_2 commute and the spectral density f of the right-hand side of (2.4.7) satisfies $\log f \in L^1([-\pi, \pi]^2)$. Then the sampled random field $(Y_t)_{t\in\mathbb{Z}^2}$ is an $ARMA([(0,0), (p,p)], [(0,0), (p-1, p-1)]_{\preceq})$ random field.*

Furthermore, the driving spatial white noise of this ARMA random field is i.i.d. noise in each of the following cases:

- *$p = 1$.*
- *$b^\intercal$ is a common left eigenvector of both A_1 and A_2.*

- Λ *is Gaussian.*

Proof. The first part follows from Proposition 2.4.10 and Proposition 2.4.12. As for the second part, we have seen in Equation (2.4.6) that the driving noise is i.i.d. for $p = 1$. If $b^\top$ is a common left eigenvector of both A_1 and A_2, then the GCARMA random field $(Y(t))_{t\in\mathbb{R}^2}$ reduces to a CAR(1) random field. In the Gaussian case we have that every white noise is actually i.i.d. noise.

□

Remark 2.4.14 (1) The order defined in Definition 2.4.11 is more precisely called the column-by-column lexicographic order. By symmetry, Theorem 2.4.13 also holds for the row-by-row lexicographic order correspondingly.

(2) In respect of the second part of Theorem 2.4.13 we note that if both A_1 and A_2 have distinct eigenvalues, then they have the same left eigenvectors since we have assumed that they commute.

(3) If A_1 is a companion matrix, then the vector $v = (v_0, ..., v_{p-1})$ is a left eigenvector of A_1 to the eigenvalue λ_1 if and only if the polynomial $v(z) := v_{p-1}z^{p-1} + \cdots + v_1 z + v_0$ satisfies $v(z) = v_{p-1}a_1(z)/(z - \lambda_1)$, where $a_1(z)$ is the corresponding polynomial to A_1. In particular, $b^\top$ cannot be a left eigenvector of A_1 if we assume that $b(z)$ and $a_1(z)$ do not have common roots.

□

Every sampled GCARMA random field is an ARMA random field according to Theorem 2.4.13. However, the MA part of this random field has infinitely many terms unless $p = 1$. For instance, if we sample a CARMA$(2,1)$ random field, we obtain an ARMA$([(0,0),(2,2)],[(0,0),(1,1)]_\preceq)$ random field, where $[(0,0),(1,1)]_\preceq = \{(0,u) \in \mathbb{Z}^2 : u \geq 0\} \cup \{(1,u) \in \mathbb{Z}^2 : u \leq 1\}$. In analogy to the purely temporal case it would be desirable to have that the $(p-1,p-1)$-dependent random field on the right-hand side of Equation (2.4.7) has a MA$(p-1,p-1)$ representation such that the sampled random field is an ARMA$((p,p),(p-1,p-1))$ random field. The next two examples illustrate that unfortunately this is not always the case.

Example 2.4.15 ($(1,1)$-dependent random field with no MA$(1,1)$ representation) Let $(Y(t))_{t\in\mathbb{R}^2}$ be a GCARMA random field with parameters $b = c =$

$(1,1)^\top$,

$$A_1 = A_2 = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix},$$

and kernel

$$g(s) = b^\top \mathrm{e}^{A_1 s_1} \mathrm{e}^{A_2 s_2} c \mathbb{1}_{\{s \geq 0\}} = \left(\mathrm{e}^{-(s_1+s_2)} + \mathrm{e}^{-2(s_1+s_2)}\right) \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^2.$$

Further, we assume that the variance of Λ satisfies $\kappa_2 = 1$ and denote the $(1,1)$-dependent right-hand side of Equation (2.4.7) as $(U_t)_{t \in \mathbb{Z}^2}$. By Proposition 2.4.10 and straight forward calculations, the autocovariance $\hat{\gamma}$ of U satisfies

$$\begin{aligned}
\hat{\gamma}(0,0) &= \frac{\left(\mathrm{e}^2-1\right)^2 \left(77 + 100\mathrm{e}^2 + 222\mathrm{e}^4 + 100\mathrm{e}^6 + 77\mathrm{e}^8\right)}{144\mathrm{e}^{12}}, \\
\hat{\gamma}(1,0) &= \hat{\gamma}(0,1) = \hat{\gamma}(-1,0) = \hat{\gamma}(0,-1) \\
&= -\frac{\left(\mathrm{e}^2-1\right)^2 \left(25 + 52\mathrm{e} + 59\mathrm{e}^2 + 16\mathrm{e}^3 + 59\mathrm{e}^4 + 52\mathrm{e}^5 + 25\mathrm{e}^6\right)}{144\mathrm{e}^{11}}, \\
\hat{\gamma}(1,1) = \hat{\gamma}(-1,-1) &= \frac{25 + 52\mathrm{e}^2 - 32\mathrm{e}^3 - 90\mathrm{e}^4 - 32\mathrm{e}^5 + 52\mathrm{e}^6 + 25\mathrm{e}^8}{144\mathrm{e}^{10}}, \\
\hat{\gamma}(1,-1) = \hat{\gamma}(-1,1) &= \frac{9 + 32\mathrm{e} + 36\mathrm{e}^2 - 154\mathrm{e}^4 + 36\mathrm{e}^6 + 32\mathrm{e}^7 + 9\mathrm{e}^8}{144\mathrm{e}^{10}}.
\end{aligned}$$

All other values of $\hat{\gamma}$ are zero. Having determined the autocovariance of U explicitly, we try to match $\hat{\gamma}$ with the autocovariance of a $\mathrm{MA}(1,1)$ random field. A generic $\mathrm{MA}(1,1)$ random field is given by

$$W_t = \sum_{k_1,k_2=0}^{1} \theta_k Z_{t-k}, \quad t \in \mathbb{Z}^2,$$

with spatial white noise Z and complex coefficients θ_k. Its autocovariance γ satisfies

$$\begin{aligned}
\gamma(0,0) &= |\theta_{00}|^2 + |\theta_{10}|^2 + |\theta_{01}|^2 + |\theta_{11}|^2, \\
\gamma(1,0) = \bar{\gamma}(-1,0) &= \theta_{00}\bar{\theta}_{10} + \theta_{01}\bar{\theta}_{11}, \\
\gamma(0,1) = \bar{\gamma}(0,-1) &= \theta_{10}\bar{\theta}_{11} + \theta_{00}\bar{\theta}_{01}, \\
\gamma(1,1) = \bar{\gamma}(-1,-1) &= \theta_{00}\bar{\theta}_{11}, \\
\gamma(1,-1) = \bar{\gamma}(-1,1) &= \theta_{01}\bar{\theta}_{10}.
\end{aligned}$$

Again, all other values of γ are zero. Extracting imaginary and real parts and using Gröbner bases (see e.g. Chapter 2 of Cox et al. [34] for more details) together with a

computer algebra system such as Mathematica, we conclude that the system $\hat{\gamma} = \gamma$ has no complex solutions for $\{\theta_{00}, \theta_{10}, \theta_{01}, \theta_{11}\}$. Hence, U is not a $\mathrm{MA}(1,1)$ random field. However, the spectral density f of U has representation

$$f(\omega) = \frac{1}{(2\pi)^2}\big(\hat{\gamma}(0,0) + 2\hat{\gamma}(1,0)\cos(\omega_1) + 2\hat{\gamma}(0,1)\cos(\omega_2) + 2\hat{\gamma}(1,1)\cos(\omega_1+\omega_2) + 2\hat{\gamma}(1,-1)\cos(\omega_1-\omega_2)\big), \quad \omega \in [-\pi,\pi]^2,$$

which is strictly positive. Consequently, $\log f$ is integrable over $[-\pi,\pi]^2$ and Theorem 2.4.13 yields that $(Y_t)_{t\in\mathbb{Z}^2}$ is an $\mathrm{ARMA}([(0,0),(2,2)],[(0,0),(1,1)]_{\preceq})$ random field. □

Example 2.4.16 ($(1,1)$-dependent random field with a $\mathrm{MA}(1,1)$ representation) We keep the setting of Example 2.4.15 with the only change that

$$A_1 = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In this case, the system $\hat{\gamma} = \gamma$ has eight different real solutions for $\{\theta_{00}, \theta_{10}, \theta_{01}, \theta_{11}\}$. The exact algebraic expressions for these solutions are very lengthy and can be computed with the software Mathematica. For illustration, we present the rounded values of one of these solutions, namely

$$\theta_{00} = 0.752991, \quad \theta_{10} = -0.176944, \quad \theta_{01} = -0.277010, \quad \theta_{11} = 0.065094.$$

Since U has the second-order structure of a $\mathrm{MA}(1,1)$ random field, it also is a $\mathrm{MA}(1,1)$ random field due to Theorem 10 in Karhunen [58]. Therefore, $(Y_t)_{t\in\mathbb{Z}^2}$ is indeed an $\mathrm{ARMA}((2,2),(1,1))$ random field. □

Further research has to be done to determine explicit necessary and sufficient conditions for the right-hand side of Equation (2.4.7) to be a $\mathrm{MA}(p-1,p-1)$ random field.

2.5 Appendix

Example 2.5.1 (GCARMA but not CARMA) Let $(Y(t))_{t\in\mathbb{R}^2}$ be a GCARMA random field with parameters $b = c = (1,1)^\top$,

$$A_1 = \begin{pmatrix} -2 & 0 \\ 0 & -3 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} -5 & 0 \\ 0 & -7 \end{pmatrix},$$

and kernel

$$g(s) = b^\top e^{A_1 s_1} e^{A_2 s_2} c \mathbb{1}_{\{s \geq 0\}} = \left(e^{-2s_1 - 5s_2} + e^{-3s_1 - 7s_2}\right) \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^2.$$

In order to check whether Y has a CARMA$(2,1)$ representation, we have to find two companion matrices $\hat{A}_1, \hat{A}_2 \in \mathbb{R}^{2\times 2}$ and a vector $\hat{b} = (\hat{b}_0, \hat{b}_1)^\top \in \mathbb{R}^2$ such that

$$g(s) = \hat{b}^\top e^{\hat{A}_1 s_1} e^{\hat{A}_2 s_2} (0,1)^\top \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^2. \tag{2.5.1}$$

Observing the exponentials, we conclude that

$$\hat{A}_1 = \begin{pmatrix} 0 & 1 \\ -6 & -5 \end{pmatrix} \quad \text{and} \quad \hat{A}_2 = \begin{pmatrix} 0 & 1 \\ 35 & -12 \end{pmatrix}$$

have to hold. Plugging these into (2.5.1) implies

$$\begin{aligned} e^{-2s_1 - 5s_2} + e^{-3s_1 - 7s_2} = \frac{e^{-3s_1 - 7s_2}}{2} \Big[& \hat{b}_0 \left(-5 + 3e^{2s_2} + 4e^{s_1} - 2e^{s_1 + 2s_2}\right) \\ & + \hat{b}_1 \left(15 - 9e^{2s_2} - 8e^{s_1} + 4e^{s_1 + 2s_2}\right) \Big], \quad s \in \mathbb{R}^2, \end{aligned}$$

which has no solution for $\hat{b}$. □

Chapter 3:
Estimation of causal CARMA random fields

3.1 Introduction

Lévy-driven continuous-time autoregressive moving average (CARMA) processes are a well-studied class of stochastic processes and enjoy versatile applications in many disciplines (cf. Brockwell [21] and the references therein). By contrast, considerably less is known about CARMA random fields indexed by $\mathbb{R}^d$, which have been defined only recently. To the best of our knowledge, two different classes exist in the literature: the *isotropic CARMA random field* was introduced in Brockwell and Matsuda [25] and the *causal CARMA random field* in [73]. While Bayesian parameter estimation is included in [25], the paper Pham [73] only provides stochastic properties of causal CARMA random fields. The goal of this chapter is to provide a semiparametric method to estimate model parameters of causal CARMA random fields from discretely observed samples.

A Lévy-driven causal CARMA random field $(Y(t))_{t\in\mathbb{R}^d}$ on $\mathbb{R}^d$ is given by the equation

$$Y(t) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} b^\top \mathrm{e}^{A_1(t_1-s_1)} \cdots \mathrm{e}^{A_d(t_d-s_d)} \mathbf{e}_p \, \Lambda(\mathrm{d}s), \quad t = (t_1, ..., t_d) \in \mathbb{R}^d, \tag{3.1.1}$$

where $A_1, ..., A_d \in \mathbb{R}^{p\times p}$ are companion matrices, $\mathbf{e}_p = (0, ..., 0, 1)^\top, b \in \mathbb{R}^p$ and Λ is a homogeneous Lévy basis, i.e., the multi-parameter analog of a Lévy process (see Section 3.2 for more details). Due to its similar structure, many commonly known properties of CARMA processes also hold for Y, such as càdlàg sample paths, exponentially decreasing autocovariance functions and rational spectral densities. In fact, the random field Y reduces to a causal CARMA process if $d = 1$. Moreover, Y has an autocovariance function which is both anisotropic and non-separable in the sense of Guttorp and Schmidt [50].

Since the matrices $A_1, ..., A_d$ are in companion form, they are completely determined by their eigenvalues. These eigenvalues in conjunction with the components of the vector b will form the model parameters. As our main tool for parameter estimation we choose the variogram, which is broadly applied in spatial statistics. It is defined as

$$\psi(t) = \mathrm{Var}[Y(t+s) - Y(s)], \quad t, s \in \mathbb{R}^d,$$

for stationary random fields (cf. Section 2.2.1 of Cressie [35]). Furthermore, it is pointed out in Section 2.4.1 of [35] that variogram estimation performs better than autocovariance estimation in terms of bias and in the presence of trend contamination. Assuming that observations of Y are given on a regular lattice $L = \{\Delta, ..., N\Delta\}^d$, we estimate the model parameters by a two-step procedure. First, we calculate an empirical version of the variogram $\psi(\cdot)$ at different lags using a non-parametric estimator $\psi_N^*(\cdot)$. Second, we fit the empirical variogram to the theoretical one using a weighted least squares method. More precisely, for a given set of strictly positive weights w_j, we estimate the true vector of CARMA parameters θ_0 by means of the weighted least squares (WLS) estimator

$$\theta_N^* := \mathrm{argmin}_{\theta \in \Theta} \left\{ \sum_{j=1}^{K} w_j \left(\psi_N^*(t^{(j)}) - \psi_\theta(t^{(j)}) \right)^2 \right\},$$

where Θ is a compact parameter space containing θ_0 and K is the number of lags used (see also Equation (3.4.1)).

An important task in connection with this approach is to determine sufficiently many lags $t^{(1)}, ..., t^{(K)} \in \mathbb{R}^d$ in order to obtain identifiability of the model parameters. We tackle this problem and show that under certain conditions a small number of lags on the principal axes of the Cartesian coordinate system is already sufficient to recover the CARMA parameters. In particular, one does not need to assume the property of *invertibility* (or an analog thereof) as for CARMA processes. This fact differentiates the one-dimensional case from the higher dimensional case and we will investigate this in more detail.

Another part of this chapter is devoted to the study of different numerical simulation schemes for the causal CARMA random field. We derive approximation algorithms similar to those presented in Chen et al. [30] and Nguyen and Veraart [69] which are based on truncation or discretization of the stochastic integral in Equation (3.1.1). We show that the output converges in mean-square and almost surely to the underlying CARMA random field. The algorithms are then used to

conduct a simulation study in order to assess the quality of the WLS estimator. Subsequently, we apply the estimator to data of the cosmic microwave background.

Our chapter is organized as follows: We recall the definition and basic properties of causal CARMA random fields in Section 3.2. Therein, a new formula for the spectral density is also proven. Strong consistency and asymptotic normality of the non-parametric variogram estimator $\psi_N^*(\cdot)$ is shown in Section 3.3. Subsequently, Section 3.4 is concerned with the asymptotic properties of the WLS estimator θ_N^*. Under identifiability conditions, we show strong consistency and asymptotic normality. While it is easier to show identifiability of CAR parameters, we obtain identifiability of CARMA parameters by carefully analyzing algebraic properties of the variogram. In Section 3.5 we consider two different simulation methods and their associated algorithms. It is shown that the simulations converge pointwise both in L^2 and almost surely to the underlying true random fields as the truncation parameter tends to infinity and the discretization parameter tends to zero. The chapter concludes with a simulation study and an application to cosmic microwave background data in Section 3.6 and Section 3.7.

We use the following notation throughout this chapter: $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function such that for instance $\mathbb{1}_{\{t\geq 0\}}$ is the Heaviside function. Furthermore, $A^\top$ denotes the transpose of a matrix (or a vector) A. The components of a vector $u \in \mathbb{R}^d$ are given by $u_1, ..., u_d$ if not stated otherwise. For $u, v \in \mathbb{R}^d$, $\|u\|$ is the Euclidean norm, $u \cdot v \in \mathbb{R}$ is the scalar product, $u \odot v \in \mathbb{R}^d$ is the componentwise product, and $u \leq v$ if and only if $u_i \leq v_i$ for all $i \in \{1, ..., d\}$. The d-dimensional interval $[u, v]$ is defined as $[u, v] := \{s \in \mathbb{R}^d : u \leq s \leq v\}$ and we set $\mathbb{R}_+ = [0, \infty)$. Additionally, $\mathbf{e}_1, ..., \mathbf{e}_d$ are the unit vectors in $\mathbb{R}^d$ and $\mathbf{e} := \mathbf{e}_1 + \cdots + \mathbf{e}_d = (1, ..., 1)^\top$. Diagonal matrices are denoted by $\mathrm{diag}(\lambda_1, ..., \lambda_d) \in \mathbb{R}^{d\times d}$ and $M_d(\mathbb{R}[z])$ is the space of all matrix polynomials of dimension $d \times d$. Finally, $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ are the real and imaginary part of a complex number z, $\mathrm{Leb}(\cdot)$ is the Lebesgue measure, and $\boldsymbol{i}$ is the imaginary unit.

3.2 Preliminaries

First and foremost, we summarize some important properties of causal CARMA random fields. To this end, let us fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, supporting all stochastic objects in this chapter. As stated in the introduction, CARMA random fields are defined as stochastic integrals driven by *homogeneous Lévy bases*. These are

random measures which can be seen as a generalization of Lévy processes and their integration theory was developed in the seminal paper Rajput and Rosiński [75]. For a homogeneous Lévy basis Λ we denote its characteristic triplet by (β, σ^2, ν), where $\beta \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ and ν is a Lévy measure. We say that Λ has a finite second moment and variance $\kappa_2 := \sigma^2 + \int_{\mathbb{R}} z^2 \, \nu(\mathrm{d}z)$ if and only if $\int_{\mathbb{R}} z^2 \, \nu(\mathrm{d}z) < \infty$. The variance κ_2 always appears in conjunction with the variogram or mean squared errors throughout this chapter. For more details on Lévy bases, we refer to Section 2 in [73]. The following definition of causal CARMA random fields is taken from the same reference.

Definition 3.2.1 Let q and p be two non-negative integers such that $q < p$, $b = (b_0, ..., b_{p-1})^\top \in \mathbb{R}^p$ with $b_q \neq 0$ and $b_i = 0$ for $i > q$, $\mathbf{e}_p = (0, ..., 0, 1)^\top \in \mathbb{R}^p$, and A_i be the companion matrix to a monic polynomial a_i of degree p with real coefficients and roots having strictly negative real parts for $i = 1, ..., d$. A random field $(Y(t))_{t \in \mathbb{R}^d}$ is called *(causal) CARMA(p, q) random field* if it satisfies the equations

$$\begin{aligned} Y(t) &= b^\top X(t), \quad t \in \mathbb{R}^d, \\ X(t) &= \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} \mathrm{e}^{A_1(t_1 - s_1)} \cdots \mathrm{e}^{A_d(t_d - s_d)} \mathbf{e}_p \, \Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d, \end{aligned} \tag{3.2.1}$$

where Λ is a homogeneous Lévy basis on $\mathbb{R}^d$ with $\int_{\mathbb{R}} \log(|z|)^d \mathbb{1}_{\{|z|>1\}} \, \nu(\mathrm{d}z) < \infty$. A (causal) CARMA$(p, 0)$ random field is also called a *(causal) CAR(p) random field*. □

Under the conditions specified in this definition, it was shown in [73] that CARMA random fields exist and are well defined. Furthermore, they are by definition causal since the value of $Y(t)$ at $t \in \mathbb{R}^d$ only depends on the driving Lévy basis Λ on the set $(-\infty, t_1] \times \cdots \times (-\infty, t_d]$. This type of causality can be interpreted as a directional influence. Also, it is immediate to see that they have the moving average representation

$$Y(t) = (g * \Lambda)(t) := \int_{\mathbb{R}^d} g(t - s) \, \Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d, \tag{3.2.2}$$

where the *kernel* g is given by

$$g(s) = b^\top \mathrm{e}^{A_1 s_1} \cdots \mathrm{e}^{A_d s_d} \mathbf{e}_p \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^d. \tag{3.2.3}$$

The kernel g is anisotropic in contrast to the isotropic CARMA random field in [25]. Additionally, it is non-separable, i.e., it cannot be written as a product of the form $g(s) = g_1(s_1) \cdots g_d(s_d)$ with real-valued functions g_i except in the CAR(1) case. If $d = 1$, we recover the classical kernel of a causal CARMA process (indexed by $\mathbb{R}$).

Remark 3.2.2 Causal CARMA random fields solve a system of stochastic partial differential equations, which generalizes the classical state-space representation of CARMA processes. For more details, see Section 3 in [73]. □

In this chapter, we always impose the following additional conditions:

Assumption A

- The Lévy basis Λ has mean zero and a finite second moment.
- The companion matrix A_i has distinct eigenvalues for $i = 1, ..., d$.

□

The first part of Assumption A ensures the existence of a second-order structure of Y, which is crucial for our estimation procedure in Section 3.4. In addition, the zero mean condition facilitates some computations, however, it is neither necessary nor restrictive. The autocovariance functions of CARMA random fields are non-separable (except in the CAR(1) case), anisotropic, and integrable over $\mathbb{R}^d$ since they are exponentially decreasing (cf. [73]). This implies the existence of a spectral density, which can be shown to be rational as for CARMA processes.

The second part of Assumption A is analogous to Assumption 1 in Brockwell et al. [27], where it is also pointed out that this condition is not critical since multiple eigenvalues can be handled as a limiting case. Furthermore, this assumption implies that the kernel g from Equation (3.2.3) can alternatively be represented as

$$g(s) = \sum_{\lambda_1} \cdots \sum_{\lambda_d} d(\lambda_1, ..., \lambda_d) \mathrm{e}^{\lambda_1 s_1} \cdots \mathrm{e}^{\lambda_d s_d} \mathbb{1}_{\{s \geq 0\}}, \quad s \in \mathbb{R}^d, \tag{3.2.4}$$

where $d(\lambda_1, ..., \lambda_d)$ are (possibly complex) coefficients and $\sum_{\lambda_i}$ denotes the sum over distinct eigenvalues of A_i for $i = 1, ..., d$ (cf. Corollary 3.7 in [73]).

It is commonly known that an equidistantly sampled CARMA process is always an ARMA process. Under certain conditions, we also obtain an ARMA random field if we sample a CARMA random field on a regular lattice (see Section 4.3 in [73]). However, these conditions are rather restrictive (one of which is $A_1 = \cdots = A_d$) and it is not known whether this sampling property can be generalized to the whole class of CARMA random fields. Nevertheless, we will see in Section 3.5 that a CARMA random field sampled on a regular grid can always be approximated arbitrarily well by a discrete-parameter moving average random field (of finite order) in terms of the mean squared error and almost surely.

From Equation (3.2.2) we observe that Y is strictly stationary, which in turn implies that the variogram

$$\psi(t) = \mathrm{Var}[Y(t) - Y(0)] = \mathrm{Var}[Y(t+s) - Y(s)], \quad t, s \in \mathbb{R}^d,$$

is translation-invariant, i.e., independent of s. In Section 3.4 we will estimate the CARMA parameters b and the eigenvalues of $A_1, ..., A_d$ by fitting an empirical version of ψ to its theoretical counterpart. Therefore, it is necessary to have the variogram structure of CARMA random fields at hand, which is given in the next proposition.

Proposition 3.2.3. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field such that Assumption A holds true. Then the variogram ψ of Y has the form*

$$\psi(t) = 2\kappa_2 \sum_{\lambda_1} \cdots \sum_{\lambda_d} \sum_{v\in\{-1,1\}^d} d_v(\lambda_1, ..., \lambda_d) \mathbb{1}_{\{t\odot v \in \mathbb{R}^d_+\}} \left(1 - \mathrm{e}^{\lambda_1|t_1|} \cdots \mathrm{e}^{\lambda_d|t_d|}\right), \quad t \in \mathbb{R}^d,$$

where $\{d_v(\lambda_1, ..., \lambda_d)\}$ is a set of complex coefficients for every $v \in \{-1,1\}^d$ such that

$$d_v(\lambda_1, ..., \lambda_d) = d_{-v}(\lambda_1, ..., \lambda_d)$$

and $\sum_{\lambda_i}$ denotes the sum over distinct eigenvalues of A_i for $i = 1, ..., d$.

Proof. The statement is a combination of Theorem 4.1. in [73] and the relation

$$\psi(t) = 2(\gamma(0) - \gamma(t)), \quad t \in \mathbb{R}^d,$$

where

$$\gamma(t) = \mathrm{Cov}[Y(t), Y(0)] = \mathrm{Cov}[Y(t+s), Y(s)], \quad t, s \in \mathbb{R}^d, \tag{3.2.5}$$

is the autocovariance function of Y. □

As it was argued in [73], it is in general hard to find explicit formulae for $d_v(\lambda_1, ..., \lambda_d)$ in terms of $d, p, q, b, A_1, ..., A_d$. However, if we fix the dimension d and the orders p and q, we are able to compute the variogram explicitly. We consider the following example.

Example 3.2.4 Let $d = 2$, $p = 2$, $q = 1$ and Λ be a homogeneous Lévy basis satisfying Assumption A. We assume that the CARMA$(2,1)$ random field

$$Y(t) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} b^\top \mathrm{e}^{A_1(t_1-s_1)} \mathrm{e}^{A_2(t_2-s_2)} \mathbf{e}_p \, \Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^2,$$

has parameters $b = (b_0, b_1) \in \mathbb{R}^2$,

$$A_1 = \begin{pmatrix} 0 & 1 \\ -\lambda_{11}\lambda_{12} & \lambda_{11} + \lambda_{12} \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 0 & 1 \\ -\lambda_{21}\lambda_{22} & \lambda_{21} + \lambda_{22} \end{pmatrix},$$

such that the eigenvalues $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22} \in \mathbb{R}$ have strictly negative real parts and satisfy $\lambda_{11} \neq \lambda_{12}$ and $\lambda_{21} \neq \lambda_{22}$. In this case, the variogram ψ of Y is given by

$$\psi(t) = 2\kappa_2 \sum_{k=11}^{12} \sum_{l=21}^{22} \left(d_{(1,1)}(\lambda_k, \lambda_l) \mathbb{1}_{\{t_1 t_2 \geq 0\}} + d_{(1,-1)}(\lambda_k, \lambda_l) \mathbb{1}_{\{t_1 t_2 < 0\}} \right) \left(1 - \mathrm{e}^{\lambda_k |t_1|} \mathrm{e}^{\lambda_l |t_2|} \right),$$

where $t \in \mathbb{R}^2$,

$$\begin{aligned}
d_{(1,1)}(\lambda_{11}, \lambda_{21}) &= \frac{(\lambda_{12} - \lambda_{21})(b_0 + b_1\lambda_{11})(b_0(2\lambda_{11} + \lambda_{12} + \lambda_{21}) + b_1\lambda_{11}(\lambda_{12} - \lambda_{21}))}{4\lambda_{11}\lambda_{21}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})} \\
d_{(1,1)}(\lambda_{12}, \lambda_{21}) &= -\frac{(\lambda_{11} - \lambda_{21})(b_0 + b_1\lambda_{12})(b_0(\lambda_{11} + 2\lambda_{12} + \lambda_{21}) + b_1\lambda_{12}(\lambda_{11} - \lambda_{21}))}{4\lambda_{12}\lambda_{21}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})} \\
d_{(1,1)}(\lambda_{11}, \lambda_{22}) &= \frac{(\lambda_{12} - \lambda_{22})(b_0 + b_1\lambda_{11})(b_0(2\lambda_{11} + \lambda_{12} + \lambda_{22}) + b_1\lambda_{11}(\lambda_{12} - \lambda_{22}))}{4\lambda_{11}\lambda_{22}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{22} - \lambda_{21})(\lambda_{21} + \lambda_{22})} \\
d_{(1,1)}(\lambda_{12}, \lambda_{22}) &= \frac{(\lambda_{11} - \lambda_{22})(b_0 + b_1\lambda_{12})(b_0(\lambda_{11} + 2\lambda_{12} + \lambda_{22}) + b_1\lambda_{12}(\lambda_{11} - \lambda_{22}))}{4\lambda_{12}\lambda_{22}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})}
\end{aligned}$$

and

$$\begin{aligned}
d_{(1,-1)}(\lambda_{11}, \lambda_{21}) &= \frac{(\lambda_{12} + \lambda_{21})(b_0 + b_1\lambda_{11})(b_0(2\lambda_{11} + \lambda_{12} - \lambda_{21}) + b_1\lambda_{11}(\lambda_{12} + \lambda_{21}))}{4\lambda_{11}\lambda_{21}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})} \\
d_{(1,-1)}(\lambda_{12}, \lambda_{21}) &= -\frac{(\lambda_{11} + \lambda_{21})(b_0 + b_1\lambda_{12})(b_0(\lambda_{11} + 2\lambda_{12} - \lambda_{21}) + b_1\lambda_{12}(\lambda_{11} + \lambda_{21}))}{4\lambda_{12}\lambda_{21}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})} \\
d_{(1,-1)}(\lambda_{11}, \lambda_{22}) &= -\frac{(\lambda_{12} + \lambda_{22})(b_0 + b_1\lambda_{11})(b_0(2\lambda_{11} + \lambda_{12} - \lambda_{22}) + b_1\lambda_{11}(\lambda_{12} + \lambda_{22}))}{4\lambda_{11}\lambda_{22}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})} \\
d_{(1,-1)}(\lambda_{12}, \lambda_{22}) &= \frac{(\lambda_{11} + \lambda_{22})(b_0 + b_1\lambda_{12})(b_0(\lambda_{11} + 2\lambda_{12} - \lambda_{22}) + b_1\lambda_{12}(\lambda_{11} + \lambda_{22}))}{4\lambda_{12}\lambda_{22}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})}
\end{aligned}$$

These formulae have been computed with the computer algebra system `Mathematica`.

□

The next result, which is of theoretical interest and will be useful later on, contains a formula for the spectral density of Y which is more explicit than Equation (4.5) in [73].

Proposition 3.2.5. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field such that Λ has a finite second moment. Further, let $a_i(z) = \sum_{j=0}^{p} a_{i,j} z^{p-j}$ for $i = 1,\dots,d$ be the monic polynomials in Definition 3.2.1. Then the spectral density f of Y has the representation*

$$f(\omega) = \frac{\kappa_2}{(2\pi)^d} \left| \frac{Q(\boldsymbol{i}\omega)}{P(\boldsymbol{i}\omega)} \right|^2, \quad \omega \in \mathbb{R}^d,$$

with polynomials

$$P(z) = a_1(z_1)\cdots a_d(z_d),$$

and

$$Q(z) = b^\top Q_1(z_1)\cdots Q_d(z_d)c,$$

and matrix polynomials

$$Q_i(z) = a_i(z)(zI_p - A_i)^{-1} \in M_p(\mathbb{R}[z]).$$

For each $i = 1,\dots,d$, the (k,l)-entry of the matrix polynomial Q_i is given by

$$Q_{i,k,l}(z) = \begin{cases} z^{p-1+k-l} + \sum_{j=1}^{p-l} a_{i,j} z^{p-1-j+k-l} & k \le l, \\ -\sum_{j=p-l+1}^{p} a_{i,j} z^{p-1-j+k-l} & k > l. \end{cases}$$

Proof. Proposition 11.2.2 in Bernstein [17] and Equation (3.2.3) imply that the Fourier transform of the kernel g satisfies

$$\tilde{g}(\omega) = \int_{\mathbb{R}^d} g(s) \mathrm{e}^{-\boldsymbol{i}\omega\cdot s} \,\mathrm{d}s = b^\top (\boldsymbol{i}\omega_1 I_p - A_1)^{-1} \cdots (\boldsymbol{i}\omega_d I_p - A_d)^{-1} \mathbf{e}_p, \quad \omega \in \mathbb{R}^d.$$

Applying Lemma 3.1 in Brockwell and Schlemm [26] and the relation

$$f(\omega) = \frac{\kappa_2}{(2\pi)^d} \tilde{g}(\omega)\tilde{g}(-\omega) = \frac{\kappa_2}{(2\pi)^d} |\tilde{g}(\omega)|^2,$$

yields the claimed assertion. □

3.3 Asymptotic properties of the empirical variogram

Let $(Y(t))_{t\in\mathbb{R}^d}$ be a CARMA(p,q) random field satisfying Assumption A. If we are given observations of Y on a lattice $L = \{\Delta,\dots,N\Delta\}^d$, we can estimate the variogram

$\psi(\cdot)$ by Matheron's method-of-moment estimator (cf. Section 2.4 in Cressie [35] for more details)

$$\psi_N^*(t) := \frac{1}{|B_{N,t}|} \sum_{s \in B_{N,t}} (Y(t+s) - Y(s))^2, \quad t \in \{(1-N)\Delta, ..., (N-1)\Delta\}^d,$$

where

$$B_{N,t} := \{s \in \Delta\mathbb{Z}^d : s, s+t \in \{\Delta, ..., N\Delta\}^d\} \text{ and } |B_{N,t}| = \prod_{i=1}^{d} (N - |t_i|) \mathbb{1}_{\{|t_i| \leq N\}}.$$

We aim to show strong consistency and multivariate asymptotic normality of $\psi_N^*(\cdot)$ as N tends to infinity. To this end, we make use of the asymptotic normality of the autocovariance estimator

$$\gamma_N^*(t) := \frac{1}{|B_{N,t}|} \sum_{s \in B_{N,t}} Y(t+s)Y(s),$$

which was shown in [15] for moving average random fields by applying a blocking technique and a central limit theorem for *m-dependent* random fields.

Theorem 3.3.1. *Suppose that $(Y(t))_{t \in \mathbb{R}^d}$ is a CARMA(p,q) random field such that Assumption A holds true, Λ has a finite fourth moment κ_4 and observations of Y are given on the lattice $L = \{\Delta, ..., N\Delta\}^d$. Then we have for all $t \in \Delta\mathbb{Z}^d$ that*

$$\lim_{N \to \infty} \psi_N^*(t) = \psi(t) \quad a.s..$$

Further let $t^{(1)}, ..., t^{(K)} \in \Delta\mathbb{Z}^d$ be K distinct lags and $t^{(0)} = (0, ..., 0)^\top$. Then we have

$$N^{d/2}(\psi_N^*(t^{(1)}) - \psi(t^{(1)}), ..., \psi_N^*(t^{(K)}) - \psi(t^{(K)})) \xrightarrow{d} \mathcal{N}(0, FVF^\top), \quad \text{as } N \to \infty,$$

where the two matrices F and $V = (v_{i,j})_{i,j=0,...,K}$ are given by

$$F = 2 \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix} \in \mathbb{R}^{K \times (K+1)}$$

and

$$\begin{aligned} v_{i,j} = \sum_{l \in \Delta\mathbb{Z}^d} (\kappa_4 - 3\kappa_2^2) \int_{\mathbb{R}}^{d} & g(s)g(s+t^{(i)})g(s+l)g(s+l+t^{(j)}) \, \mathrm{d}s \\ & + \gamma(l)\gamma(l+t^{(i)}-t^{(j)}) + \gamma(l+t^{(i)})\gamma(l-t^{(j)}) \end{aligned}$$

for all $i, j = 0, ..., K$.

Proof. First of all, we show strong consistency of the variogram estimator $\psi_N^*(\cdot)$. By Corollary 3.18 in [15], we have for all $t \in \Delta\mathbb{Z}^d$ that

$$\lim_{N\to\infty} \gamma_N^*(t) = \gamma(t) \quad \text{a.s..}$$

Considering the following limit

$$\begin{aligned}
&\lim_{N\to\infty} \psi_N^*(t) - 2(\gamma_N^*(0) - \gamma_N^*(t)) \\
&= \lim_{N\to\infty} \left(\frac{1}{|B_{N,t}|} \sum_{s\in B_{N,t}} (Y(t+s)^2 + Y(s)^2) - \frac{1}{|B_{N,0}|} \sum_{s\in B_{N,0}} Y(s)^2 \right) \\
&= 2\gamma(0) - 2\gamma(0) = 0 \quad \text{a.s.},
\end{aligned}$$

we deduce that

$$\lim_{N\to\infty} \psi_N^*(t) = \lim_{N\to\infty} 2(\gamma_N^*(0) - \gamma_N^*(t)) = 2(\gamma(0) - \gamma(t)) = \psi(t) \quad \text{a.s.}$$

as desired. It remains to show asymptotic normality of $\psi_N^*(\cdot)$. Since the kernel g in Equation (3.2.4) is a sum of exponentials, we have for every $i, j = 0, ..., K$ that

$$\int_{\mathbb{R}^d} |g(s)g(s+k)g(s+t^{(i)})g(s+k+t^{(j)})| \, \mathrm{d}s = \mathcal{O}(\mathrm{e}^{2\lambda_{max,1}k_1 + ... + 2\lambda_{max,1}k_d}), \quad k \to \infty,$$

where $\lambda_{max,i} := \max\{\mathrm{Re}(\lambda_i) : \lambda_i \text{ is eigenvalue of } A_i\} < 0$ for $i = 1, ..., d$. Hence, we obtain

$$\sum_{k\in\Delta\mathbb{Z}^d} \int_{\mathbb{R}^d} |g(s)g(s+k)g(s+t^{(i)})g(s+k+t^{(j)})| \, \mathrm{d}s < \infty.$$

Moreover, by Theorem 4.1. in [73] we also have that

$$\sum_{k\in\Delta\mathbb{Z}^d} \gamma(k)^2 < \infty.$$

We conclude that the conditions of Theorem 3.8 in [16] are satisfied, which in turn shows that

$$N^{d/2}(\gamma_N^*(t^{(0)}) - \gamma(t^{(0)}), ..., \gamma_N^*(t^{(K)}) - \gamma(t^{(K)})) \xrightarrow{d} \mathcal{N}(0, V), \quad \text{as } N \to \infty. \tag{3.3.1}$$

Consider now the mapping

$$f : \mathbb{R}^{K+1} \to \mathbb{R}^K, \quad f(x_0, ..., x_K) \mapsto (2(x_0 - x_1), ..., 2(x_0 - x_K)),$$

whose Jacobian is the matrix F. The multivariate delta method (see e.g. Proposition 6.4.3 in [23]) in combination with the mapping f and (3.3.1) yields

$$N^{d/2}(2(\gamma_N^*(0) - \gamma_N^*(t^{(1)})) - \psi(t^{(1)}), ..., 2(\gamma_N^*(0) - \gamma_N^*(t^{(K)})) - \psi(t^{(K)})) \xrightarrow{d} \mathcal{N}(0, FVF^\top)$$

as $N \to \infty$, and Slutsky's theorem finishes the proof. □

3.4 Estimation of CARMA random fields

According to Definition 3.2.1, a CARMA random field is determined by the pair (p, q), the vector b, the companion matrices $A_1, ..., A_d$ and the Lévy basis Λ. To avoid redundancies in model specification one usually assumes that either b_0 or κ_2 is known. We assume the latter and thus the goal of this section is to estimate b and $A_1, ..., A_d$ when p, q and κ_2 are given. Since every companion matrix is uniquely determined by its eigenvalues, we define the *CARMA parameter vector* θ as

$$\theta = (b_0, ..., b_q, \lambda_{11}, ..., \lambda_{1p}, \lambda_{21}, ..., \lambda_{dp}) \in \mathbb{R}^{q+1} \times \mathbb{C}^{dp}, \tag{3.4.1}$$

where $\lambda_{i1}, ..., \lambda_{ip}$ are the eigenvalues of A_i for $i = 1, ..., d$. Recall that A_i is real by definition and thus its eigenvalues are real or appear in pairs of complex conjugates. In order to estimate θ, we fit the empirical variogram $\psi_N^*(\cdot)$ of the last section to the theoretical variogram $\psi_\theta(\cdot)$ using a weighted least squares approach. In other words, we consider the estimator

$$\theta_N^* := \operatorname{argmin}_{\theta \in \Theta} \left\{ \sum_{j=1}^{K} w_j \left(\psi_N^*(t^{(j)}) - \psi_\theta(t^{(j)}) \right)^2 \right\} \tag{3.4.2}$$

where $\Theta \subseteq \mathbb{R}^{q+1} \times \mathbb{C}^{dp}$ is a compact parameter space containing the true parameter vector θ_0, $w_j > 0$ are strictly positive weights and $t^{(1)}, ..., t^{(K)} \in \mathbb{R}^d$ are prescribed lags. The paper Lahiri et al. [63] determines asymptotic properties of least squares estimators for parametric variogram models subject to asymptotic properties of the underlying variogram estimators. We use these results in conjunction with Theorem 3.3.1 to show strong consistency and asymptotic normality of θ_N^*. In the following, we denote by

$$\xi_i(\theta) := ((\partial/\partial\theta_i)\psi_\theta(t^{(1)}), ..., (\partial/\partial\theta_i)\psi_\theta(t^{(K)}))$$

the vector of first order partial derivatives of $\psi_\theta(t^{(1)}), ..., \psi_\theta(t^{(K)})$ with respect to the i'th coordinate of θ and define

$$\Xi(\theta) := -(\xi_1(\theta), ..., \xi_{dp+q+1}(\theta)),$$

which is the Jacobian matrix of the mapping $\theta \mapsto (\psi_\theta(t^{(1)}), ..., \psi_\theta(t^{(K)}))$.

Theorem 3.4.1. *Suppose that $(Y(t))_{t \in \mathbb{R}^d}$ is a CARMA(p, q) random field with true parameter vector θ_0 such that Assumption A holds true, $\kappa_2 = 1$, Λ has a finite fourth moment and observations of Y are given on the lattice $L = \{\Delta, ..., N\Delta\}^d$. Further, assume that*

- *the true parameter vector θ_0 lies inside a compact parameter space $\Theta \subseteq \mathbb{R}^{q+1} \times \mathbb{C}^{dp}$,*
- *the mapping $\Theta \ni \theta \mapsto (\psi_\theta(t^{(1)}), ..., \psi_\theta(t^{(K)}))$ is injective (identifiability criterion).*

Then we have both

$$\lim_{N\to\infty} \theta_N^* = \theta_0 \quad a.s.$$

and

$$N^{d/2}(\theta_N^* - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad as\ N \to \infty,$$

where

$$\Sigma = B(\theta_0)\Xi(\theta_0)^\top W F V F^\top W \Xi(\theta_0) B(\theta_0),$$

with F and V as in Theorem 3.3.1, $W = diag(w_1, ..., w_K)$ and

$$B(\theta_0) = (\Xi(\theta_0)^\top W \Xi(\theta_0))^{-1}.$$

Proof. We only have to check conditions (C.1)-(C.3) in [63] since our assertions follow directly from Theorems 3.1 and 3.2 of this reference. Since $\psi_\theta(t) = 2(\gamma_\theta(0) - \gamma_\theta(t))$, it suffices to show that for each $t \in \mathbb{R}^d$ the autocovariance $\gamma_\theta(t)$ is continuously differentiable with respect to θ in order to check (C.2)(ii). Recall that the relation

$$A_i = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_{ip} & -a_{i(p-1)} & -a_{i(p-2)} & \cdots & -a_{i1} \end{pmatrix} = V \mathrm{diag}(\lambda_{i1}, ..., \lambda_{ip}) V^{-1}$$

is satisfied for companion matrices, where V is the Vandermonde matrix

$$V = \begin{pmatrix} 1 & \cdots & 1 \\ \lambda_{i1} & \cdots & \lambda_{ip} \\ \vdots & \ddots & \vdots \\ \lambda_{i1}^{p-1} & \cdots & \lambda_{ip}^{p-1} \end{pmatrix}.$$

Now assume $t \in \mathbb{R}_+^d$ first. Then (the proof of) Theorem 4.1. in [73] implies

$$\gamma_\theta(t) = \kappa_2 b^\top \left(\int_{\mathbb{R}_+^2} \mathrm{e}^{A_1(s_1+t_1)} ... \mathrm{e}^{A_d(s_d+t_d)} \mathbf{e}_p \mathbf{e}_p^\top \mathrm{e}^{A_d^\top s_d} ... \mathrm{e}^{A_1^\top s_1} \, \mathrm{d}s \right) b. \tag{3.4.3}$$

Owing to the exponential structure of the integrand we recognize that $\gamma_\theta(t)$ is in fact infinitely often differentiable with respect to θ and therefore condition (C.2)(ii) holds true for every $t \in \mathbb{R}^d_+$. For each $t \in \mathbb{R}^d$ an analogous argument applies with a slightly different integrand. Moreover, condition (C.2)(ii) implies both (C.2)(i) and (C.1) in light of the identifiability criterion and the fact that Θ is compact. Finally, the condition (C.3) is trivial since W does not depend on θ. □

An important task in connection with the previous theorem is to determine a sufficient set of lags $t^{(1)}, ..., t^{(K)}$ such that the identifiability condition is satisfied. For CAR(p) random fields it is enough to consider finitely many lags on the principal axes of the Cartesian coordinate system. Before examining this matter, we prepend an auxiliary lemma which presents a simplified representation of the variogram on the principal axes.

Lemma 3.4.2. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field such that Assumption A holds true. Then there exists a set of complex coefficients*

$$\{d_i^*(\lambda_i) : \lambda_i \text{ is an eigenvector of } A_i, i = 1, ..., d\}$$

such that the values of the variogram ψ is given on the principal axes by

$$\psi(\tau \boldsymbol{e}_i) = 2\kappa_2 \sum_{\lambda_i} d_i^*(\lambda_i) \left(1 - \mathrm{e}^{\lambda_i|\tau|}\right), \quad \tau \in \mathbb{R}, \quad i = 1, ..., d, \tag{3.4.4}$$

where $\sum_{\lambda_i}$ denotes the sum over distinct eigenvalues of A_i.

Proof. Proposition 3.2.3 implies for every $i = 1, ..., d$ and $\tau \in \mathbb{R}^d$ that

$$\begin{aligned}
\psi(\tau \mathbf{e}_i) &= 2\kappa_2 \sum_{\lambda_1} \cdots \sum_{\lambda_d} d_{\mathbf{e}}(\lambda_1, ..., \lambda_d) \left(1 - \mathrm{e}^0 \cdots \mathrm{e}^{\lambda_i|\tau|} \cdots \mathrm{e}^0\right) \\
&= 2\kappa_2 \sum_{\lambda_i} \left(\sum_{\lambda_j : j=1,...,i-1,i+1,...,d} d_{\mathbf{e}}(\lambda_1, ..., \lambda_d) \right) \left(1 - \mathrm{e}^{\lambda_i|\tau|}\right) \\
&= 2\kappa_2 \sum_{\lambda_i} d_i^*(\lambda_i) \left(1 - \mathrm{e}^{\lambda_i|\tau|}\right).
\end{aligned}$$

□

The next example displays more explicit formulae for $d_i^*(\lambda_i)$ in the case of a CARMA$(2,1)$ random field.

Example 3.4.3 Let $(Y(t))_{t\in\mathbb{R}^2}$ be a CARMA$(2,1)$ random field such that Assumption A holds true. Then we have for every $t_1, t_2 \in \mathbb{R}$ that

$$\begin{aligned}\psi(t_1,0) = {} & \frac{2\kappa_2(b_0+b_1\lambda_{12})\left(b_0\left(\lambda_{11}^2+2\lambda_{11}\lambda_{12}+\lambda_{21}\lambda_{22}\right)+b_1\lambda_{12}\left(\lambda_{11}^2-\lambda_{21}\lambda_{22}\right)\right)}{4\lambda_{12}\lambda_{21}\lambda_{22}(\lambda_{11}-\lambda_{12})(\lambda_{11}+\lambda_{12})(\lambda_{21}+\lambda_{22})} \\ & \times (1-e^{\lambda_{12}|t_1|}) \\ & + \frac{2\kappa_2(b_0+b_1\lambda_{11})\left(b_0\left(2\lambda_{11}\lambda_{12}+\lambda_{12}^2+\lambda_{21}\lambda_{22}\right)+b_1\lambda_{11}\left(\lambda_{12}^2-\lambda_{21}\lambda_{22}\right)\right)}{4\lambda_{11}\lambda_{21}\lambda_{22}(\lambda_{12}-\lambda_{11})(\lambda_{11}+\lambda_{12})(\lambda_{21}+\lambda_{22})} \\ & \times (1-e^{\lambda_{11}|t_1|})\end{aligned}$$

and

$$\begin{aligned}&\psi(0,t_2) \\ &= 2\kappa_2\frac{b_0^2\left(\lambda_{11}^2+3\lambda_{11}\lambda_{12}+\lambda_{12}^2-\lambda_{21}^2\right)+2b_0b_1\lambda_{11}\lambda_{12}(\lambda_{11}+\lambda_{12})+b_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12}-\lambda_{21}^2\right)}{4\lambda_{11}\lambda_{12}\lambda_{21}(\lambda_{11}+\lambda_{12})(\lambda_{22}-\lambda_{21})(\lambda_{21}+\lambda_{22})} \\ &\quad\times(1-e^{\lambda_{21}|t_2|}) \\ &\quad+2\kappa_2\frac{b_0^2\left(\lambda_{11}^2+3\lambda_{11}\lambda_{12}+\lambda_{12}^2-\lambda_{22}^2\right)+2b_0b_1\lambda_{11}\lambda_{12}(\lambda_{11}+\lambda_{12})+b_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12}-\lambda_{22}^2\right)}{4\lambda_{11}\lambda_{12}\lambda_{22}(\lambda_{11}+\lambda_{12})(\lambda_{21}-\lambda_{22})(\lambda_{21}+\lambda_{22})} \\ &\quad\times(1-e^{\lambda_{22}|t_2|}).\end{aligned}$$

□

The next theorem establishes the identifiability of CAR(p) parameters. Note that replacing the vector b by $-b$ would not change the variogram. Hence, we may assume that b_0 is non-negative.

Theorem 3.4.4. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CAR(p) random field such that Assumption A holds true, κ_2 is given, $b_0 \geq 0$, all eigenvalues λ of $A_1,\ldots,A_d$ satisfy $-\pi/\Delta \leq \mathrm{Im}(\lambda) < \pi/\Delta$ and all coefficients $d_i^*(\lambda_i)$ in Lemma 3.4.2 are nonzero. Then θ is uniquely determined by the variogram ordinates $\{\psi(j\Delta\boldsymbol{e}_i)\colon i=1,\ldots,d; j=0,\ldots,2p+1\}$.*

Proof. Assuming without loss of generality that $\Delta = 1$ and $\kappa_2 = 1/2$, and setting $\lambda_{i0} = 0$ and $d_i^*(\lambda_{i0}) = -\sum_{k=1}^p d_i^*(\lambda_{ik})$, Lemma 3.4.2 implies that

$$\psi(j\mathbf{e}_i) = -\sum_{k=0}^{p} d_i^*(\lambda_{ik})\mathrm{e}^{\lambda_{ik}j}, \quad i=1,\ldots,d, \quad j=0,\ldots,2p+1.$$

Note that $-d_i^*(\lambda_{i0}) = \psi(0)$ is twice the variance of Y and therefore nonzero. Introducing the polynomials

$$R_i(z) := \prod_{l=0}^{p}(z - \mathrm{e}^{\lambda_{il}}) =: \sum_{l=0}^{p+1} r_{il} z^l,$$

we observe for each $i = 1, ..., d$ and $j = 0, ..., 2p+1$ that

$$\sum_{l=0}^{p+1} r_{il}\psi((j+l)\mathbf{e}_i) = -\sum_{l=0}^{p+1} r_{il} \sum_{k=0}^{p} d_i^*(\lambda_{ik})\mathrm{e}^{\lambda_{ik}j}\mathrm{e}^{\lambda_{ik}l} = -\sum_{k=0}^{p} d_i^*(\lambda_{ik})\mathrm{e}^{\lambda_{ik}j}\sum_{l=0}^{p+1} r_{il}\mathrm{e}^{\lambda_{ik}l} = 0,$$

where the last equation follows from the definition of $R_i(z)$. Hence, we get the linear systems

$$\Psi_i \begin{pmatrix} r_{i0} \\ \vdots \\ r_{ip} \end{pmatrix} := \begin{pmatrix} \psi(0\mathbf{e}_i) & \cdots & \psi(p\mathbf{e}_i) \\ \vdots & \ddots & \vdots \\ \psi(p\mathbf{e}_i) & \cdots & \psi(2p\mathbf{e}_i) \end{pmatrix} \begin{pmatrix} r_{i0} \\ \vdots \\ r_{ip} \end{pmatrix} = -\begin{pmatrix} \psi((p+1)\mathbf{e}_i) \\ \vdots \\ \psi((2p+1)\mathbf{e}_i) \end{pmatrix}, \tag{3.4.5}$$

where the system matrices Ψ_i are quadratic Hankel matrices. We show that all Ψ_i are invertible. To this end, for fixed $i \in \{1, ..., d\}$, assume that there is a vector $u = (u_0, ..., u_p) \in \mathbb{R}^{p+1}$ satisfying

$$\Psi_i u = 0,$$

that is u is an element inside Ψ_i's kernel. Defining the polynomial $P(z) := \sum_{l=0}^{p} u_l z^l$, we obtain for all $j = 0, ..., p$ that

$$0 = \sum_{l=0}^{p} u_l \psi((l+j)\mathbf{e}_i) = \sum_{l=0}^{p} u_l \left(-\sum_{k=0}^{p} d_i^*(\lambda_{ik})\mathrm{e}^{\lambda_{ik}(l+j)} \right) = -\sum_{k=0}^{p} \mathrm{e}^{\lambda_{ik}j} d_i^*(\lambda_{ik}) P(\mathrm{e}^{\lambda_{ik}}).$$

This gives the linear system

$$\begin{pmatrix} 1 & \cdots & 1 \\ \mathrm{e}^{\lambda_{i0}} & \cdots & \mathrm{e}^{\lambda_{ip}} \\ \vdots & \ddots & \vdots \\ \mathrm{e}^{\lambda_{i0}p} & \cdots & \mathrm{e}^{\lambda_{ip}p} \end{pmatrix} \begin{pmatrix} d_i^*(\lambda_{i0})P(\mathrm{e}^{\lambda_{i0}}) \\ d_i^*(\lambda_{i1})P(\mathrm{e}^{\lambda_{i1}}) \\ \vdots \\ d_i^*(\lambda_{ip})P(\mathrm{e}^{\lambda_{ip}}) \end{pmatrix} = 0.$$

Since the system matrix is a regular Vandermonde matrix and the coefficients $d_i^*(\lambda_{ik})$ are nonzero, we conclude that $P(\mathrm{e}^{\lambda_{ik}}) = 0$ for $k = 0, ..., p$. The polynomial $P(\cdot)$ has $(p+1)$ different roots and is of degree p. Consequently, it has to be the zero polynomial, which means that $u = 0$ and Ψ_i is invertible. By solving the linear

systems (3.4.5) we get all r_{il}, which gives the $e^{\lambda_{il}}$ by determining the roots of $R_i(z)$. Finally, all eigenvalues λ_{il} can be obtained uniquely using the condition on the imaginary part $\mathrm{Im}(\lambda)$, and it is trivial to recover b_0 in light of Equations (3.2.5) and (3.4.3). □

Remark 3.4.5 (1) The set of parameter vectors θ of CAR random fields which have at least one vanishing coefficient $d_i^*(\lambda_i)$ is a lower dimensional algebraic variety in the parameter space $\mathbb{R} \times \mathbb{C}^{dp}$. Thus, the Lebesgue measure of this set is zero and almost all $\theta \in \mathbb{R} \times \mathbb{C}^{dp}$ satisfy the condition on the coefficients $d_i^*(\lambda_i)$ in Theorem 3.4.4. For instance, in the setting of Example 3.4.3, $d_2^*(\lambda_{21}) = 0$ if and only if $(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{21}^2) = 0$.

(2) The condition $-\pi/\Delta \leq \mathrm{Im}(\lambda) < \pi/\Delta$ is necessary due to the complex periodicity of the exponential function. In time series analysis this problem is associated with the *aliasing effect*, i.e., the emergence of redundancies when sampling the process (cf. e.g. Section 3.4 and Assumption C5 in Schlemm and Stelzer [80]). □

Having established identifiability for CAR(p) random fields, we now turn to CARMA(p, q) random fields. For classical CARMA processes on $\mathbb{R}$ it is commonly known that one needs to impose at least conditions like $b_0 \geq 0$ and *invertibility* in order to identify CARMA parameters from the second-order structure (i.e. either autocovariance, spectral density or variogram). For instance, if we consider the spectral density

$$f(\omega) = \frac{1}{2\pi} \frac{|b(\boldsymbol{i}\omega)|^2}{|a(\boldsymbol{i}\omega)|^2}, \quad \omega \in \mathbb{R},$$

of a CARMA(p, q) process with AR polynomial $a(\cdot)$ and MA polynomial $b(\cdot)$, then the numerator of f yields the polynomial $b(z)b(-z)$. For every root λ of $b(z)b(-z)$, $-\lambda$ is also a root, making it impossible to recover $b(\cdot)$ from $n(\cdot)$. Therefore, assuming invertibility, i.e., the condition that every root of $b(\cdot)$ has a negative real part, is necessary to determine the MA polynomial $b(\cdot)$ uniquely. However, this reasoning cannot be carried over to the causal CARMA random field since two additional obstacles occur: first, the spectral density f of Proposition 3.2.5 is now a multi-parameter function and, second, it is in general not separable, i.e., it cannot be written as a product of the form $f(\omega) = f_1(\omega_1) \cdots f_d(\omega_d)$. Therefore, we cannot iterate the previous argument to each dimension. Also, the roots of the numerator $Q(z)Q(-z)$ are not discrete points in $\mathbb{C}$ anymore but, more generally, form algebraic

varieties in $\mathbb{C}^d$. This makes it harder to formulate a similar condition as invertibility for the multi-parameter case. However, as we shall see by the end of this section, an invertibility condition is in fact not necessary. In order to show identifiability of CARMA random fields, we study the algebraic properties of the variogram and start with the following result.

Theorem 3.4.6. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field such that Assumption A holds true, $\kappa_2 = 1$, all eigenvalues λ of $A_1,...,A_d$ satisfy $-\pi/\Delta \leq \mathrm{Im}(\lambda) < \pi/\Delta$ and all coefficients $d_i^*(\lambda_i)$ in Lemma 3.4.2 are nonzero. Further assume that the set $S = \{\psi(j\Delta \boldsymbol{e}_i) : i = 1,...,d; j = 0,...,2p+1\}$ of variogram ordinates is given. Then there are at most 2^p different parameter values for θ which generate S.*

Proof. Analogously to Theorem 3.4.4, we can determine all eigenvalues λ of $A_1,...,A_d$ from the set S. It remains to show that only finitely many vectors b can generate S. By Lemma 3.4.2 and Assumption A, we can solve Equations (3.4.4) for all coefficients $d_i^*(\lambda_i)$. By Theorem 4.1. in [73] we have that b has to satisfy the equations

$$d_i^*(\lambda_i) = b^\top M(i,\lambda_i) b, \tag{3.4.6}$$

where $i = 1,...,d$, λ_i is an eigenvalue of A_i and $M(i,\lambda_i)$ are matrices that only depend on the (known) eigenvalues of $A_1,...,A_d$. That is, we are given pd quadratic equations in $q+1$ unknowns $b_0,...,b_q$. Assumption A and Bézout's theorem (see e.g. Theorem 18.3 in [52]) conclude the proof.

□

The previous theorem shows that every fiber of the mapping $\Theta \ni \theta \mapsto S$ is finite. This property is also called *algebraic identifiability* (cf. Section 1 in [4]). To obtain statistical identifiability, we explicitly compute the variogram coefficients $d_i^*(\lambda_i)$ in Equation (3.4.6), which yields a polynomial system in terms of the CARMA parameters. One has then to show that this system has a unique solution. We demonstrate our method for the CARMA$(2,1)$ case and show how it can be applied to higher (p,q).

Proposition 3.4.7. *Let $(Y(t))_{t\in\mathbb{R}^2}$ be a CARMA$(2,1)$ random field such that Assumption A holds true, $\kappa_2 = 1$, $b_0 \geq 0$, all eigenvalues λ of A_1, A_2 satisfy $-\pi/\Delta \leq \mathrm{Im}(\lambda) < \pi/\Delta$ and all coefficients $d_i^*(\lambda_i)$ in Lemma 3.4.2 are nonzero. Furthermore, assume the additional condition $\lambda_{11}\lambda_{12} \neq \lambda_{21}\lambda_{22}$. Then θ is uniquely determined by $\{\psi(j\Delta \boldsymbol{e}_i) : i = 1,2; j = 0,...,5\}$.* □

Proof. First of all, the eigenvalues $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}$ and coefficients $d_1^*(\lambda_{11}), d_1^*(\lambda_{12}), d_2^*(\lambda_{21})$ and $d_2^*(\lambda_{22})$ can be recovered exactly as in Theorem 3.4.4. Hence, we only have to determine the parameters b_0 and b_1. Using the formulae of Example 3.4.3, it is an easy task to verify the equations

$$
\begin{aligned}
d_1^*(\lambda_{11}) &= \frac{(b_0 + b_1\lambda_{11})\left(b_0\left(2\lambda_{11}\lambda_{12} + \lambda_{12}^2 + \lambda_{21}\lambda_{22}\right) + b_1\lambda_{11}\left(\lambda_{12}^2 - \lambda_{21}\lambda_{22}\right)\right)}{4\lambda_{11}\lambda_{21}\lambda_{22}(\lambda_{12} - \lambda_{11})(\lambda_{11} + \lambda_{12})(\lambda_{21} + \lambda_{22})},\\
d_1^*(\lambda_{12}) &= \frac{(b_0 + b_1\lambda_{12})\left(b_0\left(\lambda_{11}^2 + 2\lambda_{11}\lambda_{12} + \lambda_{21}\lambda_{22}\right) + b_1\lambda_{12}\left(\lambda_{11}^2 - \lambda_{21}\lambda_{22}\right)\right)}{4\lambda_{12}\lambda_{21}\lambda_{22}(\lambda_{11} - \lambda_{12})(\lambda_{11} + \lambda_{12})(\lambda_{21} + \lambda_{22})},\\
d_2^*(\lambda_{21}) &= \frac{b_0^2\left(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{21}^2\right) + 2b_0b_1\lambda_{11}\lambda_{12}(\lambda_{11} + \lambda_{12}) + b_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12} - \lambda_{21}^2\right)}{4\lambda_{11}\lambda_{12}\lambda_{21}(\lambda_{11} + \lambda_{12})(\lambda_{22} - \lambda_{21})(\lambda_{21} + \lambda_{22})},\\
d_2^*(\lambda_{22}) &= \frac{b_0^2\left(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{22}^2\right) + 2b_0b_1\lambda_{11}\lambda_{12}(\lambda_{11} + \lambda_{12}) + b_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12} - \lambda_{22}^2\right)}{4\lambda_{11}\lambda_{12}\lambda_{22}(\lambda_{11} + \lambda_{12})(\lambda_{21} - \lambda_{22})(\lambda_{21} + \lambda_{22})}.
\end{aligned}
$$

We have to show that b_0 and b_1 are identifiable from this system, where $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}$ and $d_1^*(\lambda_{11}), d_1^*(\lambda_{12}), d_2^*(\lambda_{21}), d_2^*(\lambda_{22})$ are known. It therefore suffices to consider all four numerators and show that the system

$$
\begin{aligned}
&(\bar{b}_0 + \bar{b}_1\lambda_{11})\left(\bar{b}_0\left(2\lambda_{11}\lambda_{12} + \lambda_{12}^2 + \lambda_{21}\lambda_{22}\right) + \bar{b}_1\lambda_{11}\left(\lambda_{12}^2 - \lambda_{21}\lambda_{22}\right)\right)\\
&\quad = (b_0 + b_1\lambda_{11})\left(b_0\left(2\lambda_{11}\lambda_{12} + \lambda_{12}^2 + \lambda_{21}\lambda_{22}\right) + b_1\lambda_{11}\left(\lambda_{12}^2 - \lambda_{21}\lambda_{22}\right)\right),\\
&(\bar{b}_0 + \bar{b}_1\lambda_{12})\left(\bar{b}_0\left(\lambda_{11}^2 + 2\lambda_{11}\lambda_{12} + \lambda_{21}\lambda_{22}\right) + \bar{b}_1\lambda_{12}\left(\lambda_{11}^2 - \lambda_{21}\lambda_{22}\right)\right)\\
&\quad = (b_0 + b_1\lambda_{12})\left(b_0\left(\lambda_{11}^2 + 2\lambda_{11}\lambda_{12} + \lambda_{21}\lambda_{22}\right) + b_1\lambda_{12}\left(\lambda_{11}^2 - \lambda_{21}\lambda_{22}\right)\right),\\
&\bar{b}_0^2\left(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{21}^2\right) + 2\bar{b}_0\bar{b}_1\lambda_{11}\lambda_{12}(\lambda_{11} + \lambda_{12}) + \bar{b}_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12} - \lambda_{21}^2\right)\\
&\quad = b_0^2\left(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{21}^2\right) + 2b_0b_1\lambda_{11}\lambda_{12}(\lambda_{11} + \lambda_{12}) + b_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12} - \lambda_{21}^2\right),\\
&\bar{b}_0^2\left(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{22}^2\right) + 2\bar{b}_0\bar{b}_1\lambda_{11}\lambda_{12}(\lambda_{11} + \lambda_{12}) + \bar{b}_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12} - \lambda_{22}^2\right)\\
&\quad = b_0^2\left(\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{22}^2\right) + 2b_0b_1\lambda_{11}\lambda_{12}(\lambda_{11} + \lambda_{12}) + b_1^2\lambda_{11}\lambda_{12}\left(\lambda_{11}\lambda_{12} - \lambda_{22}^2\right),
\end{aligned}
\tag{3.4.7}
$$

implies $\bar{b}_0 = b_0$ and $\bar{b}_1 = b_1$, where we assume that $\bar{b}_0$ is non-negative and $\bar{b}_1 \neq 0$. Defining the variables

$$
x_1 = \bar{b}_0^2 - b_0^2, \quad x_2 = \bar{b}_0\bar{b}_1 - b_0b_1, \quad x_3 = \bar{b}_1^2 - b_1^2, \tag{3.4.8}
$$

we find the equivalent linear system

$$
\begin{pmatrix}
2\lambda_{11}\lambda_{12} + \lambda_{12}^2 + \lambda_{21}\lambda_{22} & 2\lambda_{11}^2\lambda_{12} + 2\lambda_{11}\lambda_{12}^2 & \lambda_{11}^2\lambda_{12}^2 - \lambda_{11}^2\lambda_{21}\lambda_{22}\\
2\lambda_{11}\lambda_{12} + \lambda_{11}^2 + \lambda_{21}\lambda_{22} & 2\lambda_{11}^2\lambda_{12} + 2\lambda_{11}\lambda_{12}^2 & \lambda_{11}^2\lambda_{12}^2 - \lambda_{12}^2\lambda_{21}\lambda_{22}\\
\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{21}^2 & 2\lambda_{11}^2\lambda_{12} + 2\lambda_{11}\lambda_{12}^2 & \lambda_{11}^2\lambda_{12}^2 - \lambda_{11}\lambda_{12}\lambda_{21}^2\\
\lambda_{11}^2 + 3\lambda_{11}\lambda_{12} + \lambda_{12}^2 - \lambda_{22}^2 & 2\lambda_{11}^2\lambda_{12} + 2\lambda_{11}\lambda_{12}^2 & \lambda_{11}^2\lambda_{12}^2 - \lambda_{11}\lambda_{12}\lambda_{22}^2
\end{pmatrix} x = 0, \tag{3.4.9}
$$

with $x^\top = (x_1, x_2, x_3)^\top$. This system has the unique solution $x = 0$ if and only if at least one of the four 3×3-minors

$$\begin{aligned}
&2\lambda_{11}\lambda_{12}\lambda_{21}(\lambda_{11}-\lambda_{12})(\lambda_{11}+\lambda_{12})^2(\lambda_{21}+\lambda_{22})(\lambda_{11}\lambda_{12}-\lambda_{21}\lambda_{22}),\\
&2\lambda_{11}\lambda_{12}\lambda_{22}(\lambda_{11}-\lambda_{12})(\lambda_{11}+\lambda_{12})^2(\lambda_{21}+\lambda_{22})(\lambda_{11}\lambda_{12}-\lambda_{21}\lambda_{22}),\\
&-2\lambda_{11}^2\lambda_{12}(\lambda_{11}+\lambda_{12})^2(\lambda_{21}-\lambda_{22})(\lambda_{21}+\lambda_{22})(\lambda_{11}\lambda_{12}-\lambda_{21}\lambda_{22}),\\
&-2\lambda_{11}\lambda_{12}^2(\lambda_{11}+\lambda_{12})^2(\lambda_{21}-\lambda_{22})(\lambda_{21}+\lambda_{22})(\lambda_{11}\lambda_{12}-\lambda_{21}\lambda_{22}),
\end{aligned}$$

is not zero. However, this is equivalent to the condition $\lambda_{11}\lambda_{12} \neq \lambda_{21}\lambda_{22}$. Hence, by our assumptions we can indeed conclude that $x = 0$, which yields $\bar{b}_0 = b_0$ and $\bar{b}_1 = b_1$. □

Remark 3.4.8 (1) Note that in Proposition 3.4.7 we have not used the full variogram but only values on the principal axes. Working with the full variogram, we are able to dispose of the condition $\lambda_{11}\lambda_{12} \neq \lambda_{21}\lambda_{22}$. However, imposing this weak condition has the advantage that we do not have to estimate the full variogram and the set of parameters which satisfy $\lambda_{11}\lambda_{12} = \lambda_{21}\lambda_{22}$ is a Lebesgue null set in $\mathbb{C}^4$.

(2) In the setting of Proposition 3.4.7 the condition $\lambda_{11}\lambda_{12} \neq \lambda_{21}\lambda_{22}$ is not only sufficient but also necessary. For instance, if we choose $\lambda_{11} = \lambda_{21} = -2$ and $\lambda_{12} = \lambda_{22} = -6$, then both pairs $(b_0, b_1) = (2, 4)$ and $(b_0, b_1) = (20/\sqrt{7}, 9/\sqrt{7})$ will generate the same variogram on the principal axes. Hence, in this case we do not have identifiability of the model parameters.

□

In a similar fashion we can show the following result. Since all factors in (3.4.10) are nonzero, there is no extra condition like $\lambda_{11}\lambda_{12} \neq \lambda_{21}\lambda_{22}$ needed as in Proposition 3.4.7.

Proposition 3.4.9. *Let $(Y(t))_{t\in\mathbb{R}^2}$ be a CARMA$(3,1)$ random field such that Assumption A holds true, $\kappa_2 = 1$, $b_0 \geq 0$, all eigenvalues λ of A_1, A_2 satisfy $-\pi/\Delta \leq \mathrm{Im}(\lambda) < \pi/\Delta$ and all coefficients $d_i^*(\lambda_i)$ in Lemma 3.4.2 are nonzero. Then θ is uniquely determined by $\{\psi(j\Delta \boldsymbol{e}_i) : i = 1, 2; j = 0, ..., 7\}$.* □

Proof. The assertion can be proven analogously to the proof of Proposition 3.4.7. We therefore only highlight the difference. Instead of 4 we have 6 different $d_i^*(\lambda_i)$ in

this case. Defining x_1, x_2, x_3 as before, we obtain a linear system of size 6×3 similar to Equation (3.4.9). The system matrix has $\binom{6}{3} = 20$ different 3×3-minors, one of which is

$$\begin{aligned} &-6\lambda_{11}^2\lambda_{12}^2\lambda_{13}^2(\lambda_{11}+\lambda_{12})^2(\lambda_{11}+\lambda_{13})^2(\lambda_{12}+\lambda_{13})^2(\lambda_{21}-\lambda_{22})(\lambda_{21}+\lambda_{22}) \\ &\quad\times(\lambda_{21}-\lambda_{23})(\lambda_{21}+\lambda_{23})(\lambda_{22}-\lambda_{23})(\lambda_{22}+\lambda_{23}). \end{aligned} \tag{3.4.10}$$

This minor is always nonzero under our assumptions. Thus, we conclude $\bar{b}_0 = b_0$ and $\bar{b}_1 = b_1$.

□

The method used to show identifiability for CARMA$(2,1)$ and CARMA$(3,1)$ random fields on $\mathbb{R}^2$ relied on the definition of appropriate variables x_1, x_2, x_3 in Equation (3.4.8) and a system of $pd = 4$ equations in the first case and $pd = 6$ equations in the second case. Both systems have a unique solution provided that at least one of the minors of the coefficient matrix is nonzero. In the first case 4 minors of the 4×3 coefficient matrix had to be considered and in the second case 20 of the 6×3 coefficient matrix. The complexity of this method becomes too high to consider higher order models. Moreover, we have observed that for the CARMA$(3,2)$ model on $\mathbb{R}^2$ the method fails, since the determinant of the corresponding 6×6 coefficient matrix is always zero. However, this does not prevent parameter identifiability, since – as we note from Equation (3.4.8) – the components of the vector x display algebraic dependencies, that is, the variables of the corresponding linear systems are not independent.

As an alternative to the substitution (3.4.8), we can find a solution to the original system of pd quadratic equations (3.4.6) for the $q+1$ variables $b_0, \ldots, b_q$ directly taking resort to representations via Gröbner bases (see e.g. Chapter 2 of Cox et al. [34]). As a test case we have replicated Proposition 3.4.7 using the software `Mathematica`, where the $pd = 4$ quadratic equations in (3.4.7) were transformed to an equivalent system of 48 polynomial equations. From these we could read off $\bar{b}_0 = b_0$ and $\bar{b}_1 = b_1$ immediately and again obtain identifiability.

Note that in Propositions 3.4.7 and 3.4.9 we have not assumed any extra conditions on b except for $b_0 \geq 0$. In particular, it is not necessary to impose an analogous condition to invertibility in order to achieve identifiability. This illustrates a fundamental difference between CARMA processes and CARMA random fields with $d \geq 2$.

3.5 Simulation of CARMA random fields on a lattice

In this section we develop two numerical simulation schemes for the causal CARMA random field. One is designed for compound Poisson noise and the other one for general Lévy noise. In both cases, we simulate on a lattice $L = \{\Delta, ..., N\Delta\}^d$ with fixed $\Delta > 0$ and $N \in \mathbb{N}$. Techniques for simulating on more general lattices are discussed as well.

3.5.1 Compound Poisson noise

The homogeneous Lévy basis Λ is assumed to be compound Poisson in this subsection. That is, the characteristic triplet of Λ satisfies $\beta = c\int_{(-1,1)} x\, F(\mathrm{d}x)$, $\sigma = 0$ and $\nu = cF$, where $c > 0$ is the intensity parameter and F is a probability measure on $\mathbb{R}$ (cf. Section 1.2.4 in Applebaum [6]). As a consequence, the resulting CARMA random field $(Y(t))_{t\in\mathbb{R}^d}$ in Equation (3.2.2) can be represented as

$$Y(t) = \sum_{j\in\mathbb{N}} g(t - s_j)W(s_j), \quad t \in \mathbb{R}^d,$$

where s_j are the locations of the countably many Lévy jumps of Λ and the i.i.d. $W(s_j)$ are the heights of the Lévy jumps, distributed according to F. Restricted on a compact domain $D \subset \mathbb{R}^d$, there are only finitely many jumps of Λ and their number N_J follows a Poisson distribution with intensity $c\mathrm{Leb}(D)$. Conditionally on the value of N_J, the jump positions are independently and uniformly distributed on D. This motivates us to approximate Y with

$$Y_{S1}(t) = \sum_{j=1}^{N_J} g(t - s_j)W(s_j), \quad t \in \mathbb{R}^d, \quad s_j \in D.$$

The random field Y_{S1} has the alternative representation

$$Y_{S1}(t) = (g * \Lambda_{S1})(t) := \int_{\mathbb{R}^d} g(t - s)\, \Lambda_{S1}(\mathrm{d}s), \quad t \in \mathbb{R}^d, \tag{3.5.1}$$

with $\Lambda_{S1}(\mathrm{d}s) = \mathbb{1}_D(s)\Lambda(\mathrm{d}s)$, hence it arises by truncating the Lévy basis Λ. The advantage of Y_{S1} is that we can simulate it exactly. For the simulation algorithm we choose $D = [-M, M]^d$ with a sufficiently large $M > 0$ such that $L \subset D$.

Algorithm 3.5.1

(1) Input: g, F, c, M, N, Δ such that $\{\Delta, ..., N\Delta\}^d = L \subset D = [-M, M]^d$

(2) Draw N_J from a Poisson distribution with intensity $c\text{Leb}(D) = c(2M)^d$.

(3) Draw $s_1, ..., s_{N_J}$ independently and uniformly distributed on $D = [-M, M]^d$.

(4) For each s_j, $j = 1, ..., N_J$, draw $W(s_j)$ independently from the distribution F.

(5) For each $t \in L$, compute $Y_{S1}(t) = \sum_{j=1}^{N_J} g(t - s_j)W(s_j)$.

(6) Output: $Y_{S1}(t), \quad t \in L = \{\Delta, ..., N\Delta\}^d$

□

In order to assess the accuracy of this approximation algorithm, we determine its mean squared error. Note that as the simulation of Y_{S1} is exact, we only have to consider the approximation error between Y and Y_{S1}. Moreover, we show that the simulated random field $Y_{S1}(t)$ converges for fixed $t \in L$ both in L^2 and almost surely to the underlying true random field $Y(t)$ as the truncation parameter M tends to infinity.

Theorem 3.5.2. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p, q) random field such that Assumption A holds true and Λ is compound Poisson with characteristic triplet $(c\int_{(-1,1)} x\, F(\mathrm{d}x), 0, cF)$, where $c > 0$ and F is a probability distribution. Then the mean squared error of Algorithm 3.5.1 satisfies*

$$\begin{aligned}\max_{t\in L} \mathbb{E}\Big[\big(Y(t) - Y_{S1}(t)\big)^2\Big] &= \kappa_2 \sum_{\lambda_1} \cdots \sum_{\lambda_d} \sum_{\lambda_1'} \cdots \sum_{\lambda_d'} d(\lambda_1, ..., \lambda_d) d(\lambda_1', ..., \lambda_d') \\ &\quad \times \left[\frac{1}{|\lambda_1| + |\lambda_1'|} \cdots \frac{1}{|\lambda_d| + |\lambda_d'|} - \frac{1 - \mathrm{e}^{(\lambda_1+\lambda_1')M}}{|\lambda_1| + |\lambda_1'|} \cdots \frac{1 - \mathrm{e}^{(\lambda_d+\lambda_d')M}}{|\lambda_d| + |\lambda_d'|} \right] \\ &= \mathcal{O}(\mathrm{e}^{-2|\lambda_{max}|M}), \quad M \to \infty, \end{aligned} \tag{3.5.2}$$

where the coefficients $d(\cdot)$ are the same as in Equation (3.2.4), both $\sum_{\lambda_i}$ and $\sum_{\lambda_i'}$ denote the sum over distinct eigenvalues of A_i for $i = 1, ..., d$ and

$$\lambda_{max} := \max\{\mathrm{Re}(\lambda) : \lambda \text{ is eigenvalue of } A_i, i = 1, ..., d\}.$$

Furthermore, $Y_{S1}(t)$ converges to $Y(t)$ in L^2 and almost surely as $M \to \infty$ for every $t \in L = \{\Delta, ..., N\Delta\}^d$.

Proof. By the properties of Lévy bases and Equations (3.2.2) and (3.5.1) we observe that

$$\begin{aligned}\max_{t\in L}\mathbb{E}\Big[\big(Y(t)-Y_{S1}(t)\big)^2\Big] &= \max_{t\in L}\mathbb{E}\Big[\Big(\int_{[-\infty\mathbf{e},t]/[-M\mathbf{e},t]} g(t-s)\,\Lambda(\mathrm{d}s)\Big)^2\Big]\\ &= \max_{t\in L}\kappa_2\left(\int_{\mathbb{R}_+^d} g^2(s)\,\mathrm{d}s - \int_{[0,t+M\mathbf{e}]} g^2(s)\,\mathrm{d}s\right)\\ &= \kappa_2\int_{\mathbb{R}_+^d} g^2(s)\,\mathrm{d}s - \kappa_2\int_{[0,M]^d} g^2(s)\,\mathrm{d}s\\ &= \mathbb{E}\Big[\big(Y(0)-Y_{S1}(0)\big)^2\Big], \end{aligned} \tag{3.5.3}$$

where in the first equation we have taken into account that the kernel g contains the indicator function $\mathbb{1}_{\{s\geq 0\}}$. In addition, Equation (3.2.4) implies

$$g^2(s) = \sum_{\lambda_1}\cdots\sum_{\lambda_d}\sum_{\lambda_1'}\cdots\sum_{\lambda_d'} d(\lambda_1,\ldots,\lambda_d)d(\lambda_1',\ldots,\lambda_d')\mathrm{e}^{(\lambda_1+\lambda_1')s_1}\cdots\mathrm{e}^{(\lambda_d+\lambda_d')s_d}\mathbb{1}_{\{s\geq 0\}},\quad s\in\mathbb{R}^d.$$

Plugging this into (3.5.3), we arrive at Equation (3.5.2), which in turn shows that $Y_{S1}(t)$ converges to $Y(t)$ in L^2 for every $t\in L=\{\Delta,\ldots,N\Delta\}^d$. It remains to show that the convergence also holds almost surely. Owing to Chebyshev's inequality we have for each $t\in L=\{\Delta,\ldots,N\Delta\}^d$ that

$$\sum_{M=1}^{\infty}\mathbb{P}\left[|Y(t)-Y_{S1,M}(t)|\geq\frac{1}{M}\right]\leq\sum_{M=1}^{\infty}M^2\mathbb{E}[|Y(t)-Y_{S1,M}(t)|^2],$$

where we explicitly include the input parameter M into the subscript of $Y_{S1,M}(t)$. The right-hand side of the latter inequality is finite due to Equation (3.5.2). Finally, the assertion follows from the Borel-Cantelli lemma. □

Remark 3.5.3 (1) Algorithm 3.5.1 can also be applied to pure-jump Lévy bases if small jumps are truncated. This technique has been analyzed in detail in Section 3 of Chen et al. [30] for the simulation of stochastic Volterra equations in space–time. Furthermore, Section 4 of [30] considers a simulation technique which is based on series representations for Lévy bases (see also Rosiński [76]). However, we do not pursue this direction. Instead, in the next subsection we consider a method which are not restricted to pure-jump Lévy bases, easy to implement and sufficient for our simulation study in Section 3.6.

(2) One can readily replace $L=\{\Delta,\ldots,N\Delta\}^d$ in step (5) of Algorithm 3.5.1 with any finite subset of points in $\mathbb{R}^d$. Algorithm 3.5.1 is not restricted to simulation on lattices. □

3.5.2 General Lévy noise

Algorithm 3.5.1 is not suitable for CARMA random fields driven by general Lévy bases since a drift or a Gaussian part may be part of the noise. A different way to approximate a CARMA random field $(Y(t))_{t\in\mathbb{R}^d}$ is to discretize and truncate the stochastic integral in Equation (3.2.2). Introducing a truncation parameter $M \in \mathbb{N}$, we first replace the integral in (3.2.2) by

$$\int_{[t-\Delta M\mathbf{e},t]} g(t-s)\,\Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d.$$

By discretization of this integral we obtain the sum

$$\begin{aligned} Y_{S2}(t) &:= \sum_{s\in[t-\Delta M\mathbf{e},t]\cap\Delta\mathbb{Z}^d} g(t-s)Z(s) \\ &= \sum_{s\in\{0,\Delta,\ldots,M\Delta\}^d} g(s)Z(t-s), \quad t \in \mathbb{R}^d. \end{aligned} \tag{3.5.4}$$

Here, the random field Z represents spatial increments of Λ, or more precisely

$$Z(t) := \Lambda([t-\Delta, t]), \quad t \in \mathbb{R}^d.$$

This approach has also been applied in [69] to simulate the so-called $OU_\wedge$ process. Since we evaluate Y_{S2} only on the lattice $L = \{\Delta, \ldots, N\Delta\}^d$, we actually simulate a discrete-parameter moving average random field of finite order driven by i.i.d. spatial noise as given in (3.5.4). The set $\{g(s)\colon s \in \{0, \Delta, \ldots, M\Delta\}^d\}$ plays the role of the moving average coefficients and Y_{S2} can be simulated exactly. Furthermore, it is easy to check that the random field Y_{S2} also has the representation

$$Y_{S2}(t) = (g_{S2} * \Lambda)(t) := \int_{\mathbb{R}^d} g_{S2}(t-s)\,\Lambda(\mathrm{d}s), \quad t \in \mathbb{R}^d,$$

where the step function g_{S2} is given by

$$g_{S2}(s) = \sum_{j\in\{0,,\ldots,M\}^d} b^\top \mathrm{e}^{A_1 j_1 \Delta}\cdots\mathrm{e}^{A_d j_d \Delta}\mathbf{e}_p \mathbb{1}_{[j\Delta,(j+\mathbf{e})\Delta]}(s), \quad s \in \mathbb{R}^d. \tag{3.5.5}$$

This allows us to observe that truncation and discretization of the stochastic integral in Equation (3.2.2) is in fact equivalent to truncation and discretization of the kernel g, which will be useful for establishing error bounds. We sum up the simulation scheme in the following algorithm, where (β, σ^2, ν) denotes the characteristic triplet of Λ.

Algorithm 3.5.4

(1) Input: $g, (\beta, \sigma^2, \nu), M, N, \Delta$

(2) Compute $g(s)$ for $s \in \{0, \Delta, ..., M\Delta\}^d$.

(3) Draw $Z(s)$, $s \in \{(1-M)\Delta, ..., N\Delta\}^d$, independently from the infinitely divisible distribution with characteristics $(\Delta^d\beta, \Delta^d\sigma^2, \Delta^d\nu)$.

(4) For each $t \in L = \{\Delta, ..., N\Delta\}^d$, compute $Y_{S2}(t) = \sum_{s\in\{0,\Delta,...,M\Delta\}^d} g(s)Z(t-s)$.

(5) Output: $Y_{S2}(t), \quad t \in L = \{\Delta, ..., N\Delta\}^d$

□

If we collect the $g(s)$ values from the second step of Algorithm 3.5.4 in an array A_g, and the $Z(s)$ values from the third step in an array A_Z, then the $Y_{S2}(s)$ values from the fourth step can be computed as the discrete convolution of the two arrays A_g and A_Z. This can be carried out efficiently using the fast Fourier transform (FFT). In-built convolution commands using the FFT exist in computer softwares such as `R` or `Matlab`.

By approximating the CARMA random field Y by Y_{S2} we create two sources of error, one originates from the kernel truncation, the other one from the kernel discretization. A more detailed analysis yields the following result.

Theorem 3.5.5. *Suppose that $(Y(t))_{t\in\mathbb{R}^d}$ is a CARMA(p,q) random field such that Assumption A holds true. Then $Y_{S2}(t)$ converges to $Y(t)$ in L^2 as simultaneously $\Delta \to 0$ and $\Delta M \to \infty$ for every $t \in L = \{\Delta, ..., N\Delta\}^d$.*

Further, let $(\Delta_k)_{k\in\mathbb{N}}$ and $(M_k)_{k\in\mathbb{N}}$ be two sequences satisfying $\Delta_k = \mathcal{O}(k^{-1-\epsilon})$ for some $\epsilon > 0$ and $\Delta_k M_k \to \infty$ as $k \to \infty$. Then $Y_{S2}(t)$ also converges to $Y(t)$ almost surely as $k \to \infty$ for every $t \in L = \{\Delta, ..., N\Delta\}^d$.

Proof. For notational convenience we assume that all eigenvalues of $A_1, ..., A_d$ are real. The complex case can be shown analogously by similar arguments taking care of imaginary parts. The mean squared error is by stationarity for each $t \in L =$

$\{\Delta, ..., N\Delta\}^d$ the same, namely

$$\begin{aligned}\mathbb{E}\Big[\big(Y(t) - Y_{S2}(t)\big)^2\Big] &= \kappa_2 \int_{\mathbb{R}^d} (g(s) - g_{S2}(s))^2 \,\mathrm{d}s \\ &\leq \kappa_2 p^d \sum_{\lambda_1} \cdots \sum_{\lambda_d} d(\lambda_1, ..., \lambda_d)^2 \\ &\quad \times \int_{\mathbb{R}_+^d} \left(\mathrm{e}^{\lambda_1 s_1} \cdots \mathrm{e}^{\lambda_d s_d} - \sum_{j \in \{0,,...,M\}^d} \mathrm{e}^{\lambda_1 j_1 \Delta} \cdots \mathrm{e}^{\lambda_d j_d \Delta} \mathbb{1}_{[j\Delta, (j+\mathbf{e})\Delta]}(s) \right)^2 \mathrm{d}s. \end{aligned} \tag{3.5.6}$$

Here, we have used the inequality $(\sum_{j=1}^K a_j)^2 \leq K \sum_{j=1}^K a_j^2$ with $a_1, ..., a_K \in \mathbb{R}$. In order to evaluate the latter integral, we consider for fixed $\lambda_1, ..., \lambda_d$ the identities

$$\begin{aligned}&\int_{\mathbb{R}_+^d} \mathrm{e}^{2\lambda_1 s_1} \cdots \mathrm{e}^{2\lambda_d s_d} \,\mathrm{d}s = \frac{1}{-2\lambda_1} \cdots \frac{1}{-2\lambda_d}, \\ &- 2 \int_{\mathbb{R}_+^d} \mathrm{e}^{\lambda_1 s_1} \cdots \mathrm{e}^{\lambda_d s_d} \sum_{j \in \{0,,...,M\}^d} \mathrm{e}^{\lambda_1 j_1 \Delta} \cdots \mathrm{e}^{\lambda_d j_d \Delta} \mathbb{1}_{[j\Delta, (j+\mathbf{e})\Delta]}(s) \,\mathrm{d}s \\ &\quad = -2 \sum_{j \in \{0,,...,M\}^d} \frac{\mathrm{e}^{\lambda_1 (2j_1+1)\Delta} - \mathrm{e}^{2\lambda_1 j_1 \Delta}}{\lambda_1} \cdots \frac{\mathrm{e}^{\lambda_d (2j_d+1)\Delta} - \mathrm{e}^{2\lambda_d j_d \Delta}}{\lambda_d}, \end{aligned}$$

and

$$\int_{\mathbb{R}_+^d} \sum_{j \in \{0,,...,M\}^d} \mathrm{e}^{2\lambda_1 j_1 \Delta} \cdots \mathrm{e}^{2\lambda_d j_d \Delta} \mathbb{1}_{[j\Delta, (j+\mathbf{e})\Delta]}(s) \,\mathrm{d}s = \sum_{j \in \{0,,...,M\}^d} \mathrm{e}^{2\lambda_1 j_1 \Delta} \cdots \mathrm{e}^{2\lambda_d j_d \Delta} \Delta^d.$$

Summing these up, we obtain

$$\begin{aligned}&\int_{\mathbb{R}_+^d} \left(\mathrm{e}^{\lambda_1 s_1} \cdots \mathrm{e}^{\lambda_d s_d} - \sum_{j \in \{0,,...,M\}^d} \mathrm{e}^{\lambda_1 j_1 \Delta} \cdots \mathrm{e}^{\lambda_d j_d \Delta} \mathbb{1}_{[j\Delta, (j+\mathbf{e})\Delta]}(s) \right)^2 \mathrm{d}s \\ &\quad = \frac{1}{-2\lambda_1} \cdots \frac{1}{-2\lambda_d} + \sum_{j \in \{0,,...,M\}^d} \mathrm{e}^{2\lambda_1 j_1 \Delta} \cdots \mathrm{e}^{2\lambda_d j_d \Delta} \left(\Delta^d - 2 \frac{\mathrm{e}^{\lambda_1 \Delta} - 1}{\lambda_1} \cdots \frac{\mathrm{e}^{\lambda_d \Delta} - 1}{\lambda_d} \right) \\ &\quad = f_1(\Delta M, \Delta), \end{aligned}$$

where the function f_1 is defined as

$$\begin{aligned}f_1(u, v) &:= \frac{1}{-2\lambda_1} \cdots \frac{1}{-2\lambda_d} \\ &\quad + \frac{(\mathrm{e}^{2\lambda_1 (u+v)} - 1)v}{\mathrm{e}^{2\lambda_1 v} - 1} \cdots \frac{(\mathrm{e}^{2\lambda_d (u+v)} - 1)v}{\mathrm{e}^{2\lambda_d v} - 1} \left(1 - 2 \frac{\mathrm{e}^{\lambda_1 v} - 1}{\lambda_1 v} \cdots \frac{\mathrm{e}^{\lambda_d v} - 1}{\lambda_d v} \right). \end{aligned}$$

Additionally, we have the limits

$$\lim_{u\to\infty} f_1(u,v) = \frac{1}{-2\lambda_1}\cdots\frac{1}{-2\lambda_d} + \frac{-v}{\mathrm{e}^{2\lambda_1 v}-1}\cdots\frac{-v}{\mathrm{e}^{2\lambda_d v}-1}\left(1-2\frac{\mathrm{e}^{\lambda_1 v}-1}{\lambda_1 v}\cdots\frac{\mathrm{e}^{\lambda_d v}-1}{\lambda_d v}\right) \quad (3.5.7)$$
$$=: f_2(v) =: \frac{1}{-2\lambda_1}\cdots\frac{1}{-2\lambda_d} + f_3(v),$$

and

$$\lim_{v\to 0} f_2(v) = 0. \quad (3.5.8)$$

Moreover, we observe that

$$f_1(u,v) - f_2(v) = f_3(v)[(1-\mathrm{e}^{2\lambda_1(u+v)})\cdots(1-\mathrm{e}^{2\lambda_d(u+v)})-1].$$

For every $\epsilon > 0$, the function $f_3(v)$ is bounded and continuous on $(0,\epsilon)$. We therefore arrive at

$$\lim_{u\to\infty}\sup_{v\in(0,\epsilon)} |f_1(u,v) - f_2(v)| = 0,$$

which shows that the convergence in Equation (3.5.7) is actually uniform in v. Combined with (3.5.8), this implies that

$$\lim_{u\to\infty, v\to 0} f_1(u,v) = 0.$$

Hence, for every $t \in L = \{\Delta, ..., N\Delta\}^d$, $Y_{S2}(t)$ converges to $Y(t)$ in L^2 as simultaneously $\Delta \to 0$ and $\Delta M \to \infty$.

As for the second part of our assertion, we note that if $\Delta_k = \mathcal{O}(k^{-1-\epsilon})$, then all $\lambda < 0$ satisfy the inequality

$$\sum_{k=1}^{\infty}\left|\lambda - \frac{e^{\lambda\Delta_k}-1}{\Delta_k}\right| < \infty, \quad (3.5.9)$$

which can be shown with the Taylor expansion of the exponential function. Defining

$$A_{1,k} := \frac{(\mathrm{e}^{2\lambda_1(\Delta_k M_k+\Delta_k)}-1)\Delta_k}{\mathrm{e}^{2\lambda_1\Delta_k}-1}\cdots\frac{(\mathrm{e}^{2\lambda_d(\Delta_k M_k+\Delta_k)}-1)\Delta_k}{\mathrm{e}^{2\lambda_d\Delta_k}-1}\left(2-2\frac{\mathrm{e}^{\lambda_1\Delta_k}-1}{\lambda_1\Delta_k}\cdots\frac{\mathrm{e}^{\lambda_d\Delta_k}-1}{\lambda_d\Delta_k}\right)$$

and

$$A_{2,k} := \frac{1}{-2\lambda_1}\cdots\frac{1}{-2\lambda_d} - \frac{(\mathrm{e}^{2\lambda_1(\Delta_k M_k+\Delta_k)}-1)\Delta_k}{\mathrm{e}^{2\lambda_1\Delta_k}-1}\cdots\frac{(\mathrm{e}^{2\lambda_d(\Delta_k M_k+\Delta_k)}-1)\Delta_k}{\mathrm{e}^{2\lambda_d\Delta_k}-1},$$

Inequality (3.5.9) implies $\sum_{k=1}^{\infty} |A_{1,k}| < \infty$ and $\sum_{k=1}^{\infty} |A_{2,k}| < \infty$, and thus

$$\sum_{k=1}^{\infty} |f_1(\Delta_k M_k, \Delta_k)| < \infty.$$

Finally, the almost sure convergence follows similarly as in the proof of Theorem 3.5.2 by Chebyshev's inequality and the Borel-Cantelli lemma. □

Remark 3.5.6 (1) Instead of simulating on the regular lattice $L = \{\Delta, ..., N\Delta\}^d$, one can easily adjust Algorithm 3.5.4 for simulating on the more general lattice $L = \{\Delta_1, ..., N_1\Delta_1\} \times \cdots \times \{\Delta_d, ..., N_d\Delta_d\}$ with $\Delta_1, ..., \Delta_d > 0$ and $N_1, ..., N_d \in \mathbb{N}$.

(2) In Section 4 of [73] it was shown that under mild conditions every CARMA random field has a version which is càdlàg with respect to the partial order $\leq$. By inspection of Algorithm 3.5.1 and Algorithm 3.5.4 it is easy to see that both Y_{S1} and Y_{S2} are càdlàg as well. □

3.6 Simulation study

We conduct a simulation study in order to assess the empirical quality of the WLS estimator of the previous section for finite samples. We use Algorithm 3.5.4 to simulate 500 paths of a CARMA$(2,1)$ random field on a two-dimensional grid. As CARMA parameters we take the estimates from Section 3.7, which are

$$b_0 = 4.8940,\ b_1 = -1.1432,\ \lambda_{11} = -1.7776,\ \lambda_{12} = -2.0948,\ \lambda_{21} = -1.3057,\ \lambda_{22} = -2.5142.$$

We take a Gaussian Lévy basis Λ with mean zero and variance one. In accordance with the parameter estimation in Section 3.7, we first choose $\Delta = 0.04$ for the grid size of Algorithm 3.5.4, $M = 400$ for the truncation parameter and $N^2 = 1000^2$ for the number of points for each path. However, this choice results in relatively high approximation errors, yielding only poor parameter estimates. By choosing a higher truncation parameter M and a smaller grid size Δ, the step function g_{S2} in (3.5.5) approximates the CARMA kernel g in (3.2.3) better, which by (3.5.6) also reduces the approximation error of the CARMA random field. We therefore decide to simulate on a finer grid with $\Delta = 0.01$, $M = 600$ and $N^2 = 4000^2$. After simulation we save only every fourth point in each of the two axes directions of $\mathbb{R}^2$ in order to be back in the setting of Section 3.7 with $\Delta = 0.04$ and $N^2 = 1000^2$ points per path.

Having simulated the CARMA random fields on a grid, we proceed by estimating the variogram using the variogram estimator of Section 3.3. We calculate the empirical variogram at $K = 100$ different lags, namely

$$\{\psi_N^*(j\Delta\mathbf{e}_i)\colon i = 1, 2; j = 1, ..., 50\}. \tag{3.6.1}$$

These lags lie on the principal axes of $\mathbb{R}^2$ and are by Proposition 3.4.7 sufficient to identify the CARMA$(2,1)$ parameters. In the final step we estimate the CARMA parameter vector θ with the WLS estimator θ_N^* given in (3.4.2). We consider the following choices for weights and number of lags used.

Case 1:

$$\theta_N^* := \operatorname{argmin}_{\theta\in\Theta}\left\{\sum_{\substack{j=1,...,50\\ i=1,2}} w_j \left(\psi_N^*(j\Delta\mathbf{e}_i) - \psi_\theta(j\Delta\mathbf{e}_i)\right)^2\right\}, \quad w_j = \left(\frac{0.1(j-1)+50-j}{49}\right)^2. \tag{3.6.2}$$

Case 2:

$$\theta_N^* := \operatorname{argmin}_{\theta\in\Theta}\left\{\sum_{\substack{j=1,...,25\\ i=1,2}} w_j \left(\psi_N^*(j\Delta\mathbf{e}_i) - \psi_\theta(j\Delta\mathbf{e}_i)\right)^2\right\}, \quad w_j = \left(\frac{0.1(j-1)+25-j}{24}\right)^2. \tag{3.6.3}$$

Case 3:

$$\theta_N^* := \operatorname{argmin}_{\theta\in\Theta}\left\{\sum_{\substack{j=1,...,50\\ i=1,2}} w_j \left(\psi_N^*(j\Delta\mathbf{e}_i) - \psi_\theta(j\Delta\mathbf{e}_i)\right)^2\right\}, \quad w_j = \mathrm{e}^{j\Delta}. \tag{3.6.4}$$

Case 4:

$$\theta_N^* := \operatorname{argmin}_{\theta\in\Theta}\left\{\sum_{\substack{j=1,...,25\\ i=1,2}} w_j \left(\psi_N^*(j\Delta\mathbf{e}_i) - \psi_\theta(j\Delta\mathbf{e}_i)\right)^2\right\}, \quad w_j = \mathrm{e}^{j\Delta}. \tag{3.6.5}$$

Cases 1 and 2 apply quadratically decreasing weights while Cases 3 and 4 apply exponentially decreasing weights. The compact parameter space Θ is chosen to be $\Theta = [0, 10] \times [-10, 10] \times [-10, 0]^4$ which contains the true parameter vector $\theta_0 = (b_0, b_1, \lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})$. For minimization of the objective function we use the command `DEoptim` of the `R` package **DEoptim** which implements the differential evolution algorithm (for more details see [67]). This algorithm has the advantage

that we do not need an initial value for the optimization procedure. Instead, one can directly hand over the parameter space Θ as an input. The output of `DEoptim` itself is then used as an initial value for the standard `R` command `optim`. The summary of the estimation results are given in Tables 3.2 to 3.5 below.

Recall that in our parametrization b_0 actually plays the role of the white noise standard deviation. Comparing Table 3.2 with 3.3 and Table 3.4 with 3.5, we observe that using $K = 50$ instead of $K = 100$ lags generally reduces the standard deviation (Std) but increases the bias for most of the estimators. This indicates a typical variance-bias trade-off subject to the number of lags used. Moreover, we find that using exponential weights as in (3.6.4) and (3.6.5) increases the standard deviation and the root mean squared error (RMSE) for all components of θ_N^*.

According to Theorem 3.4.1, the asymptotic properties of the WLS estimator θ_N^* does not depend on the distribution of the Lévy basis Λ. To examine this statement for finite samples, we repeat the procedure above with variance gamma noise. More precisely, we simulate 500 independent CARMA$(2,1)$ paths driven by a variance gamma basis Λ with mean zero and variance one, compute the empirical variogram as in (3.6.1) and estimate the CARMA parameters as in Cases 1 to 4. The results are summarized in Tables 3.6 to 3.9. Comparing the RMSEs in Tables 3.2 to 3.9, we observe that the WLS estimation is slightly but not significantly better for the variance gamma case than for the Gaussian case.

3.7 Application to cosmic microwave background data

We apply our theory to cosmic microwave background (CMB) data from the Planck mission of the European Space Agency. The 2018 data release can be downloaded publicly from the Planck Legacy Archive `https://pla.esac.esa.int`. The CMB maps on this website cover the full sky and have been produced using four different methods. We choose the data set created by the SMICA method and refer to [33] for more information. We take data points between 50° and 70° longitude and 10° and 30° latitude, the unit is given in Kelvin. We save the data with mean -8.7316×10^{-6} and standard deviation 9.6049×10^{-5} into an $N \times N$-matrix with $N = 1000$, and plot column-wise and row-wise means. Since we do not find any deterministic trend or seasonal component in Figure 3.1, we may assume that the data is stationary. This is in line with standard assumptions for the CMB (see e.g. Section 2.1.1 of Giovannini [46]). We perform a normalization of the data to have mean zero and

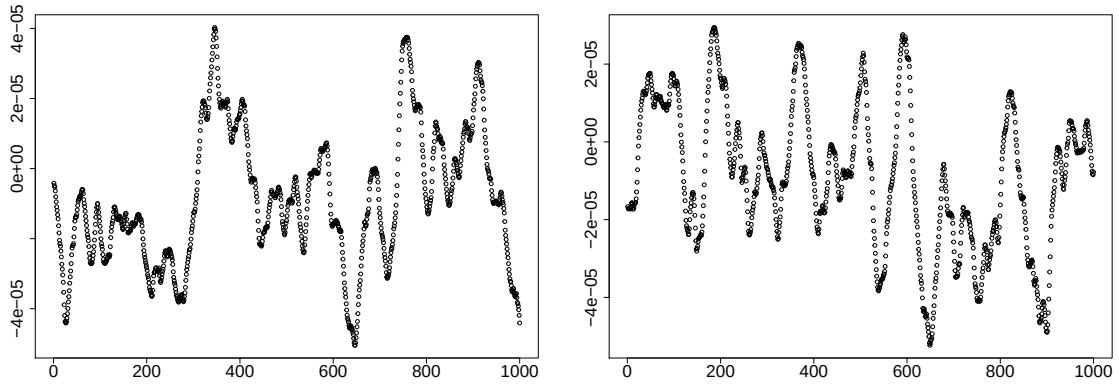

Figure 3.1: Column-wise (left) and row-wise (right) means of the CMB data.

variance one, and plot the data's empirical density against the standard normal density. An inspection of Figure 3.2 reveals that the marginal distribution of the CMB data is Gaussian. Hence, we may also assume that the Lévy basis Λ is Gaussian. We proceed as in the previous section and compute the empirical variogram at 100 different lags on the principal axes, namely $\{\psi_N^*(j\Delta\mathbf{e}_i) : i = 1, 2; j = 1, ..., 50\}$ with $\Delta = 0.04$. Assuming that the Lévy basis Λ has variance one, we estimate the parameters of CAR(1), CAR(2) and CARMA$(2,1)$ random fields with the WLS estimator in Equation (3.6.2). For the CAR(1) model we obtain:

$$b_0^* = 1.2268,\ \lambda_{11}^* = -0.4622,\ \lambda_{21}^* = -0.5159.$$

For the CAR(2) model we obtain:

$$b_0^* = 4.9991,\ \lambda_{11}^* = -1.7963,\ \lambda_{12}^* = -1.7969,\ \lambda_{21}^* = -1.2859,\ \lambda_{22}^* = -2.2212.$$

For the CARMA$(2,1)$ model we obtain:

$$b_0^* = 4.8940,\ b_1^* = -1.1432,\ \lambda_{11}^* = -1.7776,\ \lambda_{12}^* = -2.0948,\ \lambda_{21}^* = -1.3057,\ \lambda_{22}^* = -2.5142.$$

Figure 3.3 depicts the estimated variogram of the CMB data along with fitted variogram curves of our three models. Recall that b_0 is not, but plays the role of the white noise standard deviation in our parametrization (see the first paragraph of Section 3.4). The weighted sum of squares (WSS) values

$$\text{WSS} = \sum_{\substack{j=1,\dots,50 \\ i=1,2}} w_j \left(\psi_N^*(j\Delta\mathbf{e}_i) - \psi_{\theta^*}(j\Delta\mathbf{e}_i)\right)^2$$

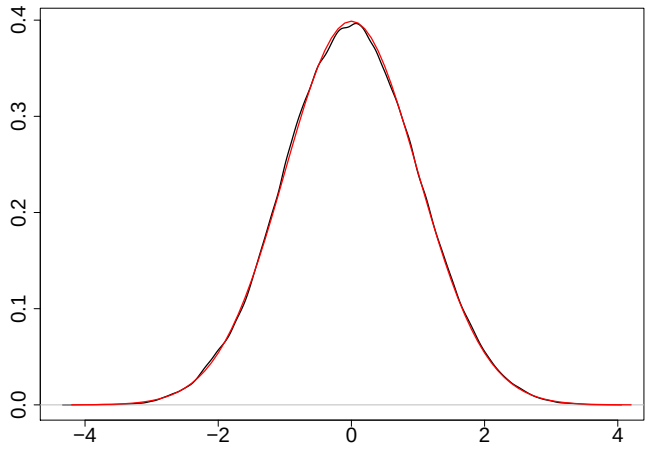

Figure 3.2: Empirical density of normalized CMB data (in black) and density of the standard normal distribution (in red).

are 7.6132×10^{-2} for CAR(1), 2.5769×10^{-2} for CAR(2) and 2.0113×10^{-2} for CARMA$(2,1)$. For model selection, we compute the Akaike information criterion (AIC)

$$\text{AIC} = 2P + K \log(WSS/K),$$

where P is the number of model parameters and K the number of lags used to calculate the WSS. The AIC values are −712.0453 for CAR(1), −816.3761 for CAR(2) and −839.1583 for CARMA$(2,1)$.

Model	WSS	P	K	AIC
CAR(1)	7.6132×10^{-2}	3	100	−712.0453
CAR(2)	2.5769×10^{-2}	5	100	−816.3761
CARMA(2,1)	2.0113×10^{-2}	6	100	−839.1583

Table 3.1: Weighted sum of squares (WSS), number of parameters (P), number of lags used (K) and Akaike Information Criterion (AIC) for the parameter estimation in Section 3.7.

These numbers are summarized in Table 3.1 and suggest that the CARMA$(2,1)$ model is optimal compared to the CAR(1) and CAR(2) models. For a visual comparison we plot the heat map of the original CMB data together with heat maps of simulated fields in Figures 3.4 to 3.7. Although we cannot draw any conclusions from a single sample path, it is possible to observe some features of the fitted models. All three models exhibit clusters of high and low values similarly to the original data. However, the cluster sizes of the CAR(1) random field are larger than those of the CMB data, whereas the CAR(2) and CARMA$(2,1)$ models display a better visual fit. Another common feature are horizontal and vertical lines, which is most visible in Figure 3.5. These lines are the consequences of the non-smoothness of the kernel function in Equation (3.2.3). One therefore can argue that the fitted CARMA random fields represent linear approximations to the spatial dependence structures of the cosmic microwave background.

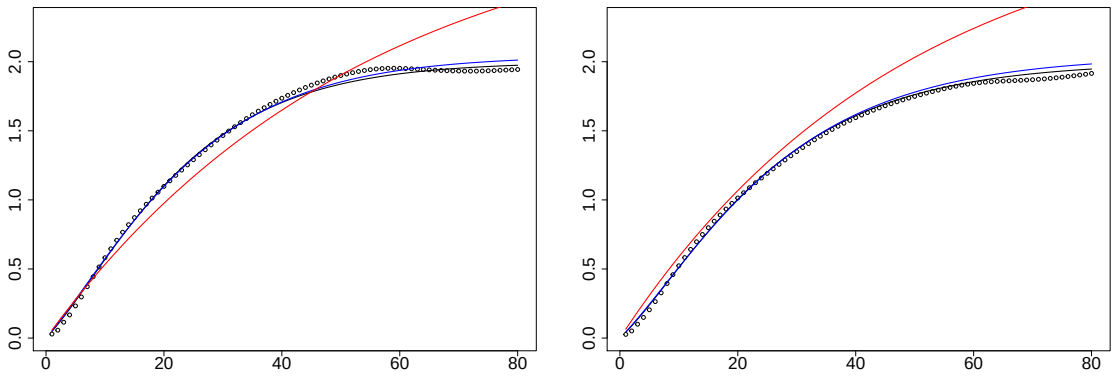

Figure 3.3: Empirical variogram ordinates of the CMB data on the horizontal axis (left) and vertical axis (right) together with fitted variogram curves of the CAR(1) model (in red), the CAR(2) model (in blue) and the CARMA$(2,1)$ model (in black).

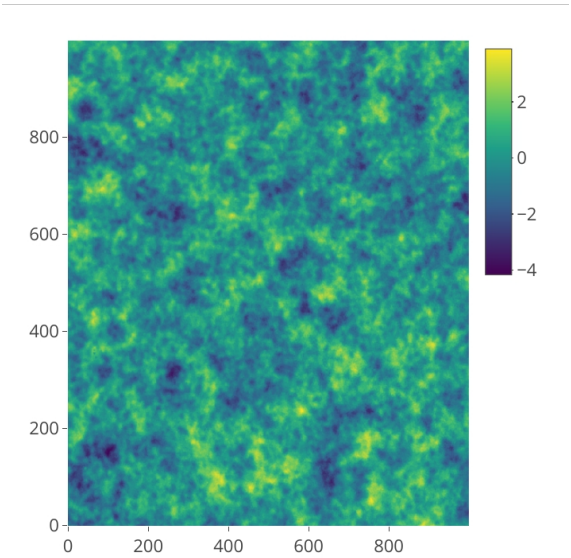

Figure 3.4: Normalized CMB data

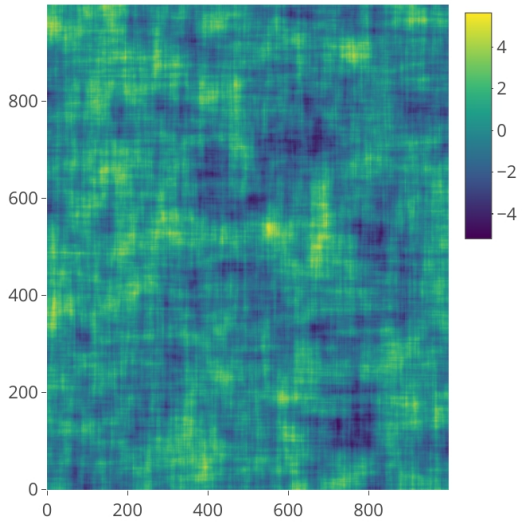

Figure 3.5: Simulated CAR(1) random field

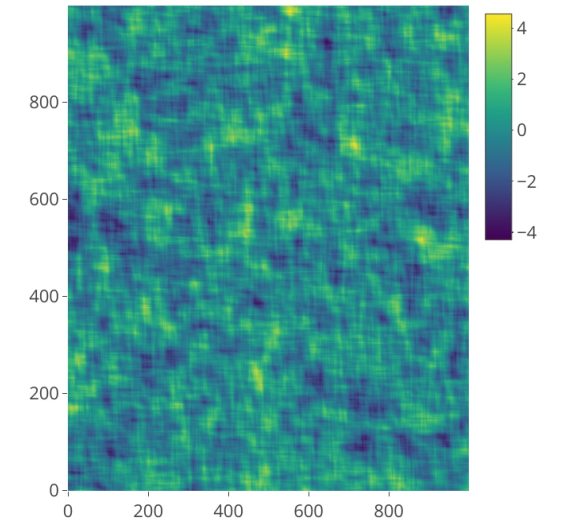

Figure 3.6: Simulated CAR(2) random field

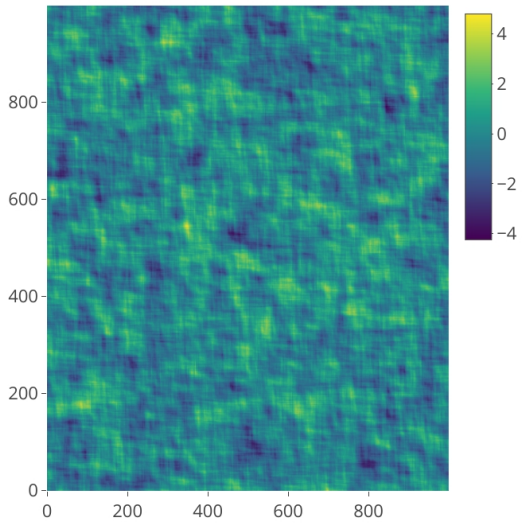

Figure 3.7: Simulated CARMA$(2,1)$ random field

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.7882	-0.1058	0.5124	0.5227
b_1	-1.1432	-1.2784	-0.1352	0.3962	0.4183
λ_{11}	-1.7776	-1.6283	0.1494	0.2377	0.2806
λ_{12}	-2.0948	-2.3193	-0.2246	0.4183	0.4744
λ_{21}	-1.3057	-1.3136	-0.0079	0.2323	0.2322
λ_{22}	-2.5142	-2.5231	-0.0089	0.4048	0.4045

Table 3.2: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 100$ lags, quadratically decreasing weights as in (3.6.2) and Gaussian basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.6929	-0.2010	0.4597	0.5013
b_1	-1.1432	-1.2252	-0.0820	0.3515	0.3606
λ_{11}	-1.7776	-1.6335	0.1442	0.2005	0.2468
λ_{12}	-2.0948	-2.2117	-0.1169	0.3246	0.3447
λ_{21}	-1.3057	-1.2947	0.0110	0.2136	0.2137
λ_{22}	-2.5142	-2.4636	0.0506	0.3065	0.3104

Table 3.3: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 50$ lags, quadratically decreasing weights as in (3.6.3) and Gaussian basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.8329	-0.0610	0.5668	0.5695
b_1	-1.1432	-1.2708	-0.1275	0.4250	0.4433
λ_{11}	-1.7776	-1.6234	0.1542	0.2473	0.2912
λ_{12}	-2.0948	-2.3569	-0.2622	0.4754	0.5425
λ_{21}	-1.3057	-1.3182	-0.0125	0.2448	0.2449
λ_{22}	-2.5142	-2.5392	-0.0250	0.4348	0.4351

Table 3.4: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 100$ lags, exponentially decreasing weights as in (3.6.4) and Gaussian basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.7525	-0.1414	0.5267	0.5448
b_1	-1.1432	-1.1995	-0.0563	0.3879	0.3915
λ_{11}	-1.7776	-1.5908	0.1868	0.2375	0.3020
λ_{12}	-2.0948	-2.2988	-0.2040	0.4240	0.4701
λ_{21}	-1.3057	-1.2765	0.0292	0.2299	0.2315
λ_{22}	-2.5142	-2.5196	-0.0054	0.3786	0.3783

Table 3.5: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 50$ lags, exponentially decreasing weights as in (3.6.5) and Gaussian basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.8407	-0.0533	0.5126	0.5148
b_1	-1.1432	-1.2383	-0.0951	0.4181	0.4283
λ_{11}	-1.7776	-1.6453	0.1323	0.2202	0.2567
λ_{12}	-2.0948	-2.3054	-0.2106	0.3828	0.4366
λ_{21}	-1.3057	-1.3149	-0.0092	0.2269	0.2269
λ_{22}	-2.5142	-2.5319	-0.0177	0.3716	0.3717

Table 3.6: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 100$ lags, quadratically decreasing weights as in (3.6.2) and variance gamma basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.7474	-0.1466	0.4635	0.4857
b_1	-1.1432	-1.2035	-0.0603	0.3578	0.3625
λ_{11}	-1.7776	-1.6325	0.1451	0.2034	0.2497
λ_{12}	-2.0948	-2.2322	-0.1375	0.3246	0.3522
λ_{21}	-1.3057	-1.2866	0.0191	0.2137	0.2144
λ_{22}	-2.5142	-2.4994	0.0148	0.3154	0.3154

Table 3.7: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 50$ lags, quadratically decreasing weights as in (3.6.3) and variance gamma basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.8696	-0.0243	0.5671	0.5671
b_1	-1.1432	-1.2780	-0.1348	0.4090	0.4302
λ_{11}	-1.7776	-1.6266	0.1511	0.2543	0.2956
λ_{12}	-2.0948	-2.3773	-0.2825	0.4731	0.5506
λ_{21}	-1.3057	-1.3086	-0.0029	0.2402	0.2400
λ_{22}	-2.5142	-2.5760	-0.0618	0.4262	0.4303

Table 3.8: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 100$ lags, exponentially decreasing weights as in (3.6.4) and variance gamma basis Λ.

	True Value	Mean	Bias	Std	RMSE
b_0	4.8940	4.7422	-0.1518	0.4806	0.5035
b_1	-1.1432	-1.2331	-0.0898	0.3942	0.4039
λ_{11}	-1.7776	-1.6102	0.1674	0.2193	0.2758
λ_{12}	-2.0948	-2.2791	-0.1843	0.3626	0.4065
λ_{21}	-1.3057	-1.2816	0.0241	0.2272	0.2283
λ_{22}	-2.5142	-2.5225	-0.0083	0.3532	0.3530

Table 3.9: Parameter estimation results for CARMA$(2,1)$ on $\mathbb{R}^2$ with $K = 50$ lags, exponentially decreasing weights as in (3.6.5) and variance gamma basis Λ.

Chapter 4:
Autocovariance varieties of moving average random fields

4.1 Introduction

Moving average random fields indexed by the integer lattice $\mathbb{Z}^d$ generalize the class of discrete-time moving average processes and constitute an important statistical spatial model. They are used to model texture images (cf. [44]), as well as in image segmentation and restoration (cf. [62]). Furthermore, they are connected to ARMA (autoregressive moving average) random fields (cf. [41] and the references therein) and the sampling problem of CARMA (continuous autoregressive moving average) random fields, in which the autocovariance functions of moving average random fields play a crucial role (cf. [73, Section 4.3]).

A *moving average random field* $(Y_t)_{t\in\mathbb{Z}^d}$ of order $q = (q_1, q_2, \ldots, q_d) \in \mathbb{N}^d$ is defined by the equation

$$Y_t = \sum_{k_1=0}^{q_1} \cdots \sum_{k_d=0}^{q_d} a_k Z_{t-k}, \quad t \in \mathbb{Z}^d,$$

where $k = (k_1, \ldots, k_d)$, a_k are real coefficients and $(Z_t)_{t\in\mathbb{Z}^d}$ is a real-valued zero-mean white noise (see Definition 4.2.1). The autocovariance function

$$\gamma(t) = \mathrm{Cov}[Y_0, Y_t], \quad t \in \mathbb{Z}^d,$$

for this type of random field is compactly supported, i.e. only finitely many values are nonzero. More precisely, we have $\gamma(t) = 0$ for every $t = (t_1, \ldots, t_d) \in \mathbb{Z}^d$ with entries satisfying $|t_i| > q_i$ for at least one $i \in \{1, \ldots, d\}$.

We study the autocovariance functions of moving average random fields from an algebraic perspective. Our motivation stems from the field of algebraic statistics [84].

Specifically, inspired by the concept of *moment varieties* [3], here we introduce *autocovariance varieties*. The *moving average variety* $\mathcal{MA}_q \subseteq \mathbb{P}^N$ (see Definition 4.3.1) is parametrized by $(q_1+1)\cdots(q_d+1)$ moving average coefficients a_k where the indices k satisfy $0 \le k_i \le q_i$ for $i = 1,\dots,d$. These coefficients induce $(2q_1+1)\cdots(2q_d+1)$ nonzero autocovariance values $\gamma(t)$. However, we only consider half of them since the relation $\gamma(-v) = \gamma(v)$ holds for all $v \in \mathbb{Z}^d$.

Example 4.1.1 Let $d = 2$ and $q = (1,1)$. Then the $Q = (1+1)(1+1) = 4$ parameters $a_{00}, a_{01}, a_{10}, a_{11}$ define the autocovariances

$$\begin{aligned}
\gamma(0,0) &= a_{00}^2 + a_{10}^2 + a_{01}^2 + a_{11}^2,\\
\gamma(1,0) &= a_{00}a_{10} + a_{11}a_{01},\\
\gamma(0,1) &= a_{00}a_{01} + a_{11}a_{10},\\
\gamma(1,-1) &= a_{10}a_{01},\\
\gamma(1,1) &= a_{00}a_{11}.
\end{aligned}$$

The moving average variety $\mathcal{MA}_{(1,1)} \subseteq \mathbb{P}^4$ is expected to be 3-dimensional. We characterize it in Theorem 4.3.4.

This chapter is organized as follows. In Section 2, we give the main definition of a moving average random field and its autocovariance function. We define our main object of study, namely autocovariance varieties, in Section 3. We contrast the properties between moving average processes (one-dimensional) from the higher dimensional moving average random fields. In Theorem 4.3.9 we establish the dimension and degree of these varieties. In Section 4 we investigate identifiability of the associated models and prove that they are algebraically identifiable. In contrast to the $d = 1$ case where the degree of the fiber grows with q, we show that for $d > 1$ there are generically only two sets of parameters that yield the same autocovariance function. Next, we study two different approaches to estimate model parameters from given samples in Section 5. First, we fit the empirical autocovariance function to the theoretical counterpart using a least squares method. Second, we consider maximum likelihood estimation. Both approaches connect nicely to concepts from algebraic statistics: respectively the ED degree and the ML degree. In Example 4.5.8, we conduct a simulation study comparing classical local optimization methods to numerical homotopy continuation, where we find that the numerical algebraic geometry (NAG) method performs slightly better.

We use the following notation and terminology in this chapter: The components of a vector $u \in \mathbb{R}^d$ are given by $u_1, ..., u_d$ if not stated otherwise. If $u, v \in \mathbb{Z}^d$, then we set $[u, v] := \{s \in \mathbb{Z}^d | u_i \leq s_i \leq v_i, 1 \leq i \leq d\}$, which may be an empty set. The symbol $\preceq$ denotes the lexicographic order and for $x \in \mathbb{R}^d$ and $t \in \mathbb{Z}^d$ we define $x^t := x_1^{t_1} \cdots x_d^{t_d}$. If $g: A \to B$ is a mapping and $y \in g(A)$, then $g^{-1}(y)$ is called fiber of y and each point inside the fiber is called preimage.

4.2 Moving average random fields

Throughout this chapter, all stochastic objects are defined on a fixed complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 4.2.1 (1) A random field $(Y_t)_{t \in \mathbb{Z}^d}$ is called *weakly stationary* if $Y_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for every $t \in \mathbb{Z}^d$ and $\gamma(t) := \mathrm{Cov}[Y_0, Y_t] = \mathrm{Cov}[Y_s, Y_{s+t}]$ for every $t, s \in \mathbb{Z}^d$. It is called a *white noise* if $\gamma(0) > 0$ and $\gamma(t) = 0$ for every $0 \neq t \in \mathbb{Z}^d$. In this case, $\sigma^2 := \gamma(0)$ is called the *white noise variance*.

(2) Let $q_1, ..., q_d$ be positive integers and $(Z_t)_{t \in \mathbb{Z}^d}$ be a real-valued zero-mean white noise on $\mathbb{Z}^d$. A random field $(Y_t)_{t \in \mathbb{Z}^d}$ is called a *moving average random field* if it satisfies the equation

$$Y_t = \sum_{k \in [0,q]} a_k Z_{t-k}, \quad t \in \mathbb{Z}^d, \tag{4.2.1}$$

where $a_k \in \mathbb{R}$ such that for each $i = 1, ..., d$ there exist at least two index vectors $l, m \in [0, q]$ satisfying $a_l \neq 0$, $a_m \neq 0$, $l_i = q_i$ and $m_i = 0$.

The last condition on the two index vectors l, m guarantees that the $MA(q)$ random field has indeed order q and not a smaller order. We associate to each $MA(q)$ random field the *moving average polynomial*

$$\theta(x) = \sum_{k \in [0,q]} a_k x^k,$$

and further, we define the formal backshift operators $B_1, ..., B_d$ which act on any random field $(X_t)_{t \in \mathbb{Z}^d}$ in the following way:

$$B_i X_t = X_{(t_1, ..., t_{i-1}, t_i - 1, t_{i+1}, ..., t_d)}, \quad t \in \mathbb{Z}^d, \quad i = 1, ..., d.$$

With this notation, (4.2.1) can be written in short as

$$Y_t = \theta(B)Z_t, \quad t \in \mathbb{Z}^d,$$

where $B = (B_1, ..., B_d)$.

The following proposition establishes the link between the moving average polynomial θ and the autocovariance function γ of a $MA(q)$ random field.

Proposition 4.2.2. *Suppose that $(Y_t)_{t\in\mathbb{Z}^d}$ is a $MA(q)$ random field driven by a white noise $(Z_t)_{t\in\mathbb{Z}^d}$ with variance σ^2. Then Y is weakly stationary, its autocovariance function γ is compactly supported and we have*

$$\sigma^2\theta(x)\theta(x^{-1}) = \sum_{t\in\mathbb{Z}^d} \gamma(t)x^t. \tag{4.2.2}$$

Proof. The facts that Y is weakly stationary and γ is compactly supported are straight-forward. Let $S = \{k \in \mathbb{Z}^d : a_k \neq 0\}$ denote the set of indexes with non-vanishing coefficient a_k. Then we have that

$$\begin{aligned}
\sigma^2\theta(x)\theta(x^{-1}) &= \sigma^2 \left(\sum_{k\in[0,q]} a_k x^k \right) \left(\sum_{k\in[0,q]} a_k x^{-k} \right) \\
&= \sigma^2 \sum_{t\in\mathbb{Z}^d} \left(\sum_{k,k+t\in S} a_k a_{k+t} \right) x^t \\
&= \sum_{t\in\mathbb{Z}^d} \operatorname{Cov}\left[\sum_{k\in[0,q]} a_k Z_{0-k}, \sum_{k+t\in[0,q]} a_{k+t} Z_{0-k} \right] x^t \\
&= \sum_{t\in\mathbb{Z}^d} \operatorname{Cov}[\theta(B)Z_0, \theta(B)Z_t] x^t \\
&= \sum_{t\in\mathbb{Z}^d} \operatorname{Cov}[Y_0, Y_t] x^t = \sum_{t\in\mathbb{Z}^d} \gamma(t)x^t.
\end{aligned}$$

□

4.3 Autocovariance varieties

We have seen that for $q = (q_1, q_2, \ldots, q_d) \in \mathbb{N}^d$, the autocovariance function of a moving average random field is only dependent on the coefficients a_k of the moving average polynomial $\theta(x) = \sum_{k\in[0,q]} a_k x^k$ and the white noise variance σ^2. In order to

avoid redundancies in model specification, one can assume without loss of generality that $\sigma^2 = 1$ and we will do so for the rest of this chapter. There are $Q+1 := \prod_{i=1}^{d}(q_i+1)$ coefficients a_k and $2N+1 := \prod_{i=1}^{d}(2q_i+1)$ non-zero autocovariances $\gamma(t)$ for $t \in [-q, q]$. Ordering them in two vectors a and γ_a, we can think of this correspondence as a polynomial map $\Gamma_q : \mathbb{R}^{Q+1} \mapsto \mathbb{R}^{2N+1}$ given by $a \mapsto \gamma_a$.

Since $\gamma(-t) = \gamma(t)$ for all $t \in \mathbb{Z}^d$, we can drop half of the autocovariances and only consider $\gamma(t)$ with $t \in [-q, q]$ and $t \succeq 0$, where $\succeq$ denotes the lexicographic order. In this way, we have a map $\mathbb{R}^{Q+1} \mapsto \mathbb{R}^{N+1}$ which we still denote by Γ_q.

The points in the image represent the set of autocovariance functions of moving average random fields. Geometrically, this is a semialgebraic set, defined by polynomial equalities and inequalities. The closure of the image of this parametrization will give a real affine algebraic variety. However, as is standard in algebraic statistics, we will first change the underlying field to be algebraically closed (so the map becomes $\Gamma_q : \mathbb{C}^{Q+1} \mapsto \mathbb{C}^{N+1}$ over the complex numbers $\mathbb{C}$) and then we pass to projective space arriving at $\Gamma_q : \mathbb{P}^Q \dashrightarrow \mathbb{P}^N$. This last step requires the polynomials $\gamma(t)$ to be homogeneous, and indeed they are in our case.

Definition 4.3.1 Let $q = (q_1, q_2, \ldots, q_d) \in \mathbb{N}^d$ and define $Q := \prod_{i=1}^{d}(q_i + 1) - 1$ and $N := (\prod_{i=1}^{d}(2q_i + 1) - 1)/2$. The *autocovariance variety* $\mathcal{MA}_q$ is the image of the autocovariance map $\Gamma_q : \mathbb{P}^Q \dashrightarrow \mathbb{P}^N$.

4.3.1 Moving average processes

If $d = 1$, moving average random fields are also called moving average processes. These processes are well-studied and belong to the important class of ARMA processes (cf. [23, Chapter 3]). Suppose that $(Y_t)_{t\in\mathbb{Z}}$ is a $MA(q)$ process given by the equation

$$Y_t = \sum_{k=0}^{q} a_k Z_{t-k}, \quad t \in \mathbb{Z}.$$

Then, the autocovariance function γ of Y has the simple expression

$$\gamma(t) = \begin{cases} \sum_{k=0}^{q-|t|} a_k a_{k+|t|}, & \text{if } |t| \le q, \\ 0, & \text{if } |t| > q. \end{cases}$$

For the class of moving average processes we have that $Q = N = q$. Thus, in this special case the autocovariance map takes the form

$$\Gamma_q : \mathbb{P}^q \dashrightarrow \mathbb{P}^q.$$

In the next subsection we will see that the map is actually defined in all of $\mathbb{P}^q$ (there are no *base points*), so we conclude the following.

Proposition 4.3.2. *If $d = 1$, then $\mathcal{MA}_q = \mathbb{P}^q$.*

While $\mathcal{MA}_q$ is not particularly interesting when $d = 1$, the parametrization coming from $\Gamma_q : \mathbb{P}^q \to \mathbb{P}^q$ has interesting fibers and computing them is important for statistical applications. This issue of identifiability will be explored in Section 4.4.

Remark 4.3.3 Going back for a moment to the real picture (over $\mathbb{R}$), the equality $\mathcal{MA}_q = \mathbb{P}^q$ is analogous to the statement that when $d = 1$, any autocovariance function with support $[-q, q]$ is an autocovariance function of a $MA(q)$ process [23, Prop 3.2.1].

4.3.2 Moving average random fields

We start by carefully analyzing the case $\mathcal{MA}_{(1,1)}$ mentioned in the introduction.

Theorem 4.3.4. *The autocovariance variety $\mathcal{MA}_{(1,1)} \subseteq \mathbb{P}^4$ is a threefold of degree 4. In the polynomial ring with variables $g_t = \gamma(t)$, it is the hypersurface defined by the quartic*

$$g_{10}^2 g_{01}^2 - g_{00} g_{10} g_{01} g_{11} + g_{10}^2 g_{11}^2 + g_{01}^2 g_{11}^2 - g_{00} g_{10} g_{01} g_{1-1} + g_{00}^2 g_{11} g_{1-1} - 2 g_{10}^2 g_{11} g_{1-1} - 2 g_{01}^2 g_{11} g_{1-1}$$

$$- 4 g_{11}^3 g_{1-1} + g_{10}^2 g_{1-1}^2 + g_{01}^2 g_{1-1}^2 + 8 g_{11}^2 g_{1-1}^2 - 4 g_{11} g_{1-1}^3 = 0. \tag{4.3.1}$$

Its singular locus is a quadratic surface, which is the union of the three irreducible components corresponding to the prime ideals

$$\left\langle g_{10} - g_{01}, g_{00} - 2g_{11} - 2g_{1-1}, 4g_{11}g_{1-1} - g_{01}^2 \right\rangle, \tag{4.3.2}$$

$$\left\langle g_{10} + g_{01}, g_{00} + 2g_{11} + 2g_{1-1}, 4g_{11}g_{1-1} - g_{01}^2 \right\rangle, \tag{4.3.3}$$

and

$$\left\langle g_{11} - g_{1-1}, g_{00} g_{1-1} - g_{10} g_{01} \right\rangle. \tag{4.3.4}$$

Proof. The proof is computational. One way to obtain the quartic (4.3.1) is through the following `Macaulay2` [48] commands:

```
R = QQ[a00,a01,a10,a11]
S = QQ[g00,g01,gm11,g10,g11]
h = map(R,S,{ a00^2  + a10^2  + a01^2  + a11^2,    a00*a01 + a10*a11,
            a10*a01, a00*a10 + a01*a11,  a00*a11} )
I = kernel h
```

For the singular locus, we compute the radical ideal of the quartic along with its vanishing gradient, and then compute its prime decomposition. □

Remark 4.3.5 Substituting the parametrization of Example 4.1.1 into (4.3.2) to (4.3.4), we find that the three irreducible components of the singular locus correspond to the three conditions

$$a_{10} = a_{01} \text{ and } a_{00} = a_{11}, \tag{4.3.5}$$

$$a_{10} = -a_{01} \text{ and } a_{00} = -a_{11}, \tag{4.3.6}$$

and

$$a_{00}a_{11} = a_{01}a_{10}. \tag{4.3.7}$$

These conditions represent submodels and we will analyze Equation (4.3.7) in more detail in Example 4.4.5.

The complexity of $\mathcal{MA}_q$ increases rapidly when $d > 1$. It is computationally challenging to obtain generators for its prime ideal even for small values of q and d. Beyond $q = (1,1)$, we were also able to do this for $q = (1,2)$ and $q = (1,1,1)$.

Proposition 4.3.6. *The autocovariance variety $\mathcal{MA}_{(1,2)} \subseteq \mathbb{P}^7$ is 5-dimensional of degree 16. Its prime ideal is cut out by 7 sextics. The autocovariance variety $\mathcal{MA}_{(1,1,1)} \subseteq \mathbb{P}^{13}$ is 7-dimensional of degree 64. Its prime ideal is cut out by 56 quartics, 90 quintics and 50 sextics.*

Table 4.1 presents the basic properties of the first autocovariance varieties $\mathcal{MA}_q$ with $q = (q_1, q_2)$, that is, for $d = 2$. The dimension appears to be the expected one, while the degree follows a clear pattern as a power of two. We will prove that this actually holds for any $\mathcal{MA}_q$. To that end, we use the next two lemmas.

Lemma 4.3.7. *The map $\Gamma_q : \mathbb{P}^Q \dashrightarrow \mathbb{P}^N$ has no base points.*

q_1	q_2	$\dim(\mathcal{MA}_q)$	N	$\deg(\mathcal{MA}_q)$	generators
1	1	3	4	4	1 quartic
1	2	5	7	16	7 sextics
1	3	7	10	64	?
2	2	8	12	128	?

Table 4.1: Summary of first autocovariance varieties for $d = 2$

Proof. Assume that $\Gamma_q(a) = 0$. We know from (4.2.2) that

$$\theta(x)\theta(x^{-1}) = \sum_{t\in\mathbb{Z}^d} \gamma(t)x^t = 0.$$

Multiplying both sides by the monomial $x^q = x_1^{q_1}\cdots x_d^{q_d}$, we obtain the product of two polynomials $\theta(x)\cdot x^q\theta(x^{-1})$ that equals the zero polynomial. Since the polynomial ring $K[x]$ is an integral domain when K is a field, we must have that either $\theta(x) = 0$ or $x^q\theta(x^{-1}) = 0$. In particular, all the coefficients $a_k = 0$, that is, $a = 0$ is the zero vector. □

Lemma 4.3.8. *The autocovariance variety $\mathcal{MA}_q$ is a linear projection of the Veronese variety. Furthermore, $\gamma(t)$ is the sum of exactly $(q_1-|t_1|+1)\cdots(q_d-|t_d|+1)$ quadratic monomials for every $t \in [-q, q]$ with $t \geq 0$, and each monomial appears exactly once.*

Proof. The quadratic Veronese embedding precisely consists of all quadratic monomials. The parametrization of $\mathcal{MA}_q$ consists of quadrics, each one is a sum of quadratic monomials. Moreover, Proposition 4.2.2 implies that

$$\gamma(t) = \sum_{k,k+t\in[0,q]} a_k a_{k+t}, \tag{4.3.8}$$

for every $t \in [-q, q]$, which shows the second part of the assertions. □

Now we state the main theorem concerning our varieties $\mathcal{MA}_q$.

Theorem 4.3.9. *Let $q \in \mathbb{N}^d$. Then*

$$\dim(\mathcal{MA}_q) = Q = \prod_{i=1}^{d}(q_i + 1) - 1$$

and if $d > 1$, then

$$\deg(\mathcal{MA}_q) = 2^{Q-1} = 2^{\prod_{i=1}^{d}(q_i+1)-2}.$$

Proof. Let $D := \dim(\mathcal{MA}_q)$ denote the dimension of $\mathcal{MA}_q$ and consider the regular map $\Gamma_q : \mathbb{P}^Q \longrightarrow \mathbb{P}^N$. Since the domain is Q-dimensional, the inequality $D \leq Q$ has to hold. However, Γ_q is not a constant map and has no base points by Lemma 4.3.7. Consequently, we actually have equality, that is $D = Q$.

By Lemmas 4.3.7 and 4.3.8, the degree of the map Γ_q has the same degree as the quadratic Veronese variety $\mathcal{V}_{Q,2}$, which is 2^Q. In addition, we know that $\deg(\Gamma_q) = \deg(\mathcal{MA}_q)\deg(\Gamma^{-1}(\gamma))$ for $\gamma \in \mathcal{MA}_q$ generic and the identifiability Theorem 4.4.4 below proves that $\deg(\Gamma^{-1}(\gamma)) = 2$. Hence, we conclude that $\deg(\mathcal{MA}_q) = 2^{Q-1}$. □

4.4 Identifiability

We show that all models $MA(q)$ are algebraically identifiable (in the sense of [4]). This means that the map from the model parameters to the autocovariances is generically finite to one. A statement that holds *generically*, or for a *generic* point $x \in \mathbb{C}^d$, can be interpreted probabilistically as holding for almost all $x \in \mathbb{C}^d$ with respect to the Lebesgue measure.

4.4.1 Moving average processes

The following result is the projective version of the known result in the moving average process literature [23].

Proposition 4.4.1. *If $d = 1$, the fibers of a generic point $\gamma \in \mathcal{MA}_q$ consist of 2^q points.*

Proof. Let $\alpha_1, ..., \alpha_q \in \mathbb{C}$ be the q roots of the moving average polynomial

$$\theta(x) = \sum_{k=0}^{q} a_k x^k = a_q(x - \alpha_1)\cdots(x - \alpha_q), \quad x \in \mathbb{C}.$$

Using Proposition 4.2.2, we see that there are exactly 2^{q+1} polynomials which generate γ as above, all of which have the form

$$\pm a_q(x - \alpha_1)_\pm \cdots (x - \alpha_q)_\pm,$$

where

$$(x - \alpha_i)_+ := (x - \alpha_i) \quad \text{and} \quad (x - \alpha_i)_- := (\alpha_i x - 1). \tag{4.4.1}$$

Hence, the fiber of any point in $\mathcal{MA}_q$ under $\Gamma_q : \mathbb{P}^q \to \mathbb{P}^q$ consists in general of 2^q points. □

Proposition 4.4.1 has two consequences. First, it implies that the map Γ_q is not injective and the moving average parameters a_i are not identifiable from a second order point of view if $d = 1$. Second, it is possible to deduce all preimage points from a single one by inverting the roots of θ as suggested in (4.4.1).

In order to obtain injectivity of Γ_q, one usually imposes the condition that all roots α_i of the polynomial θ lie strictly outside the unit disk (and $a_0 > 0$). This property is also called *invertibility* since it holds if and only if there exists coefficients $\pi_0, \pi_1, \ldots$ with $\sum_{k=0}^{\infty} |\pi_k| < \infty$ such that the white noise sequence $(Z_t)_{t\in\mathbb{Z}}$ can be expressed as

$$Z_t = \sum_{k=0}^{\infty} \pi_k Y_{t-k}, \quad t \in \mathbb{Z}.$$

Example 4.4.2 Let $q = 1$. This is the simplest moving average model $MA(1)$. We have that $\theta(x) = a_0 + a_1 x$ and $\Gamma_1 : \mathbb{P}^1 \longrightarrow \mathbb{P}^1$ is given by

$$\Gamma_1(a_0, a_1) = (a_0^2 + a_1^2, a_0 a_1).$$

The fiber of a generic point $\gamma = (\gamma_0, \gamma_1)$ consists of $2 = 2^1$ points in $\mathbb{P}^1$. They are $(\tilde{a}_0, \tilde{a}_1)$ and $(\tilde{a}_1, \tilde{a}_0)$ where

$$\tilde{a}_0 = \sqrt{\frac{\gamma_0 + \sqrt{\gamma_0^2 - 4\gamma_1^2}}{2}}, \qquad \tilde{a}_1 = \sqrt{\frac{\gamma_0 - \sqrt{\gamma_0^2 - 4\gamma_1^2}}{2}}. \tag{4.4.2}$$

The invertibility condition is equivalent to $|a_0| > |a_1|$.

The observed symmetry of the two points $(\tilde{a}_0, \tilde{a}_1)$ and $(\tilde{a}_1, \tilde{a}_0)$ above extends to higher q. In fact, it holds that

$$\Gamma_q(a_0, a_1, \ldots, a_{q-1}, a_q) = \Gamma_q(a_q, a_{q-1}, \ldots, a_1, a_0). \tag{4.4.3}$$

This can be seen from (4.2.2), where the reversal occurs by inverting all the roots in (4.4.1).

For general $q > 1$, there exist algorithms to numerically approximate the invertible solution with $a_0 = 1$. A basic one is the *innovations algorithm*, which recursively converges to the moving average parameters a_k given the autocovariance values $\gamma(t)$ under the invertibility condition (we refer to Section 2 in [22] for details). Other approaches use spectral factorization methods [79]. While we do not pursue this in this chapter, the fact remains that the desired parameters are solutions to a

polynomial system of equations, so it would be interesting to compare these with state-of-the-art algorithms in numerical algebraic geometry. See Example 4.5.8 for an illustration of such techniques.

Furthermore, the symmetry in the polynomial system means that one does not necessarily need to find a root of a polynomial of degree 2^q even when there are 2^q solutions. We illustrate this with $q = 2$.

Example 4.4.3 For $q = 2$ we have $\theta(x) = a_0 + a_1x + a_2x^2$ and $\Gamma_2 : \mathbb{P}^2 \longrightarrow \mathbb{P}^2$ given by

$$\Gamma_2(a_0, a_1, a_2) = (a_0^2 + a_1^2 + a_2^2, a_0a_1 + a_1a_2, a_0a_2).$$

The fiber of a generic point $\gamma = (\gamma_0, \gamma_1, \gamma_2) \in \mathcal{MA}_2$ consists of $2^2 = 4$ points. A Gröbner basis elimination from the system $\Gamma_2(a_0, a_1, a_2) = (\gamma_0, \gamma_1, \gamma_2)$ with order a_2, a_0, a_1 reveals a triangular system with a quadric in a_1^2:

$$\begin{aligned} a_1^4 - (\gamma_0 + 2\gamma_2)a_1^2 + \gamma_1^2 &= 0 \\ a_0^2a_1 - \gamma_1a_0 + a_1\gamma_2 &= 0 \\ a_2a_1 + a_0a_1 - \gamma_1 &= 0. \end{aligned}$$

And hence the solutions for (a_0, a_1, a_2) in terms of $(\gamma_0, \gamma_1, \gamma_2)$ can be obtained as

$$a_1 = \sqrt{\frac{\gamma_0 + 2\gamma_2 \pm \sqrt{(\gamma_0 + 2\gamma_2)^2 - 4\gamma_1^2}}{2}}, \quad a_0 = \sqrt{\frac{\gamma_1 \pm \sqrt{\gamma_1^2 - 4a_1^2\gamma_2}}{2a_1}}, \quad a_2 = \frac{\gamma_1 - a_0a_1}{a_1}.$$

4.4.2 Moving average random fields

The following result demonstrates a fundamental difference between $d = 1$ and $d > 1$ in terms of identifiability. On the other hand, it shows how the symmetry in (4.4.3) generalizes to higher dimensions.

Theorem 4.4.4. *Suppose that the moving average polynomial θ is generic. Then for $d > 1$, the fibers of a point $\gamma \in \mathcal{MA}_q$ are only two points a and a' in $\mathbb{P}^Q$. One is obtained from the other by $a'_k = a_{q-k}$ for any $k \in [0, q]$.*

Proof. Let γ be the image of the coefficients a_k of a moving average polynomial θ under the mapping Γ_q and assume that θ' is another polynomial which also generates γ and has coefficients a'_k. Due to Proposition 4.2.2, the polynomial equation

$$\theta(x)(x^q\theta(x^{-1})) = \theta'(x)(x^q\theta'(x^{-1})) \tag{4.4.4}$$

has to hold. Since generically θ is irreducible, we either have $\theta' = \theta$ or $\theta' = x^q\theta(x^{-1})$, which proves the assertion. □

Example 4.4.5 We consider again the autocovariance variety $\mathcal{MA}_{(1,1)}$ and assume that $\gamma \in \mathcal{MA}_{(1,1)}$ is generated by a generic moving average polynomial θ as in the setting of Theorem 4.4.4, that is, θ is irreducible. Then the fiber of γ is given by the equations

$$a_{00} = \sqrt{\frac{\gamma_{00}\gamma_{11} - \gamma_{01}\gamma_{10} - \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{11})^2 - 4\gamma_{11}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}},$$
$$a_{10} = \sqrt{\frac{-\gamma_{00}\gamma_{1-1} + \gamma_{01}\gamma_{10} - \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{1-1})^2 - 4\gamma_{1-1}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}},$$
$$a_{01} = \sqrt{\frac{-\gamma_{00}\gamma_{1-1} + \gamma_{01}\gamma_{10} + \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{1-1})^2 - 4\gamma_{1-1}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}},$$
$$a_{11} = \sqrt{\frac{\gamma_{00}\gamma_{11} - \gamma_{01}\gamma_{10} + \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{11})^2 - 4\gamma_{11}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}},$$

and $a'_{00} = a_{11}$, $a'_{01} = a_{10}$, $a'_{10} = a_{01}$, $a'_{11} = a_{00}$. Substituting in the formulas from Example 4.1.1, we observe that the discriminants

$$(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{11})^2 - 4\gamma_{11}^2(\gamma_{11} - \gamma_{1-1})^2,$$
$$(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{1-1})^2 - 4\gamma_{1-1}^2(\gamma_{11} - \gamma_{1-1})^2,$$

are equal to

$$\left(a_{00}^2 - a_{11}^2\right)^2 (a_{01}a_{10} - a_{00}a_{11})^2,$$
$$\left(a_{01}^2 - a_{10}^2\right)^2 (a_{01}a_{10} - a_{00}a_{11})^2,$$

which are nonnegative in the real case (as they should be when the moving average parameters are real).

If however θ is not irreducible, it has to be the product of two linear factors. Then identifiability from the above theorem fails and we have up to 4 preimages in the fiber of γ. We note that this explains the irreducible component (4.3.4) of the singular locus from Theorem 4.3.4, which is equivalent to (4.3.7). In order to see this, we first assume that Equation (4.3.7) holds. This implies that there exists a

constant $b^* \in \mathbb{R}$ such that we have $a_{01} = b^* a_{00}$ and $a_{11} = b^* a_{10}$. Thus, the polynomial θ satisfies

$$\theta(x) = (a_{00} + a_{10}x_1)(1 + b^* x_2)$$

and is therefore reducible. On the other hand, assuming that

$$\theta(x) = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 = (b_{10} + b_{11}x_1)(b_{20} + b_{21}x_2)$$

for some real-valued coefficients $b_{10}, b_{11}, b_{20}, b_{21}$, we can deduce (4.3.7). One of the four preimage points in the fiber of γ is given by the equations

$$a_{00} = \frac{1}{2}\sqrt{\gamma_{00} + \sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} - \sqrt{2\gamma_{00}\sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2)}},$$

$$a_{01} = \frac{1}{2}\sqrt{\gamma_{00} - \sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} - \sqrt{-2\gamma_{00}\sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2)}},$$

$$a_{10} = \frac{1}{2}\sqrt{\gamma_{00} - \sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + \sqrt{-2\gamma_{00}\sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2)}},$$

$$a_{11} = \frac{1}{2}\sqrt{\gamma_{00} + \sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + \sqrt{2\gamma_{00}\sqrt{\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4\,(\gamma_{01}^2 + \gamma_{10}^2)}}.$$

Remark 4.4.6 If, in contrast to the setting of Theorem 4.4.4, θ is not irreducible, then the fiber of γ under Γ_q consists of more than two preimages, as illustrated in the previous example. By (4.4.1), the maximum number of preimages is $2^{q_1+\cdots+q_d}$ and occurs exactly when θ is completely separable, that is, a product of linear forms.

4.5 Parameter estimation

In this section we go one step further and consider the problem of parameter estimation from observed sample points. We consider two methods: least squares estimation and maximum likelihood estimation. Both involve solving polynomial systems of equations. Algebraically, the computational complexity of the estimation problem is measured by the *ED degree* [40] of the associated variety in the first case and by the *ML degree* [5, 29] in the second.

4.5.1 Least squares estimation

Let $(Y_t)_{t\in\mathbb{Z}^d}$ be a $MA(q)$ random field, which by Definition 4.2.1 has mean zero. If we are given observations of Y on a lattice $L = \{1,...,n\}^d$, we can estimate the autocovariance function $\gamma(t)$ by the empirical autocovariance estimator

$$\hat{\gamma}_n(t) := \frac{1}{|B_{n,t}|}\sum_{s\in B_{n,t}} Y(t+s)Y(s), \quad t\in[-q,q], \quad t\geq 0,$$

where

$$B_{n,t} := \{s\in\mathbb{Z}^d | s, s+t\in L\} \text{ and } |B_{n,t}| = \prod_{i=1}^{d}(n-|t_i|)\mathbb{1}_{\{|t_i|\leq n\}}.$$

If $\hat{\gamma}_n(t)$ were exact values, we would be in the situation of the previous section. However, these are just numerical estimates which form a point $\hat{\gamma}_n$ that almost surely lies outside the model $\mathcal{MA}_q$. One approach is to project the estimated vector $\hat{\gamma}_n$ onto the autocovariance variety $\mathcal{MA}_q$, that is, obtaining $\gamma_n^*\in\mathcal{MA}_q$ which has the smallest Euclidean distance to $\hat{\gamma}_n$:

$$\gamma_n^* := \text{argmin}_{\gamma\in\mathcal{MA}_q}\|\gamma-\hat{\gamma}_n\|. \tag{4.5.1}$$

The number of critical points of this least squares optimization problem is counted by the Euclidean distance degree (*ED degree*).

Proposition 4.5.1. *The ED degree of $\mathcal{MA}_q$ is 1 if $d = 1$. The ED degree of $\mathcal{MA}_{(1,1)}$ is 16.*

Proof. The first part is a consequence of Proposition 4.3.2. In fact, for $d = 1$ the unique critical point for (4.5.1) is $\gamma_n^* = \hat{\gamma}_n$. For the second we use the following `M2` code:

```
R = QQ[g00,g01,gm11,g10,g11]
I = ideal(g01^2*gm11^2-g00*g01*gm11*g10+g01^2*g10^2+gm11^2*g10^2+
    g00^2*gm11*g11-2*g01^2*gm11*g11-4*gm11^3*g11-g00*g01*g10*g11
    -2*gm11*g10^2*g11+g01^2*g11^2+8*gm11^2*g11^2+g10^2*g11^2-4*gm11*g11^3)
sing = ideal singularLocus I
u = {5,7,13,11,3};
M = (matrix{apply(# gens R,i->(gens R)_i-u_i)})||(transpose(jacobian I));
time J = saturate(I + minors(2,M), sing);
dim J, degree J
```

The vector u represents a generic choice of γ and the saturation is needed to remove the critical points that lie in the singular locus. □

We illustrate with an example:

Example 4.5.2 We simulate 2500 points of a $MA(1,1)$ random field on a 50×50 grid in `R` (see Figure 4.1). As white noise we take an i.i.d. standard Gaussian random field. The moving average parameters are chosen as

$$a_{00} = 7, \qquad a_{01} = -5, \qquad a_{10} = 3, \qquad a_{11} = 1$$

and the corresponding autocovariances values are

$$\gamma = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \gamma_{1-1}) = (84, 16, -32, 7, -15).$$

After centering the sample, we compute the empirical autocovariances

$$\hat{\gamma}_n = (\hat{\gamma}_n(0,0), \hat{\gamma}_n(0,1), \hat{\gamma}_n(1,0), \hat{\gamma}_n(1,1), \hat{\gamma}_n(1,-1)).$$

By Proposition 4.5.1, we expect 16 complex critical points, and we compute them numerically. Six of them are real

$$\begin{aligned}
&(87.1147, 18.6511, -33.4739, 5.78808, -17.312),\\
&(80.8137, 30.7661, -23.1126, -3.96875, -28.7833),\\
&(61.9284, -24.7157, -16.0001, 1.76548, 19.994),\\
&(55.2165, 8.80716, 26.5528, 0.977029, 8.45708),\\
&(71.9207, -7.85594, -8.51067, 35.9693, 0.649541),\\
&(63.1632, -18.9463, -12.5151, 0.0189543, 24.6219).
\end{aligned}$$

The first line has the lowest Euclidean distance to the estimated point

$$\hat{\gamma}_n = (86.6439, 19.1877, -34.2433, 6.6726, -17.3195),$$

and therefore

$$\gamma_n^* = (87.1147, 18.6511, -33.4739, 5.78808, -17.312).$$

Moreover, we have that

$$\|\gamma - \hat{\gamma}_n\| = 5.2604 \quad \text{and} \quad \|\gamma - \gamma_n^*\| = 5.0711,$$

so that projecting onto the autocovariance variety improves the empirical estimate.

q	ED degree
(1,1)	16
(1,1)	169
(1,2)	1600
(1,4)	14641

Table 4.2: ED degree of the variety $\mathcal{MA}_q$ with $q = (1, k)$, $k = 1, 2, 3, 4$

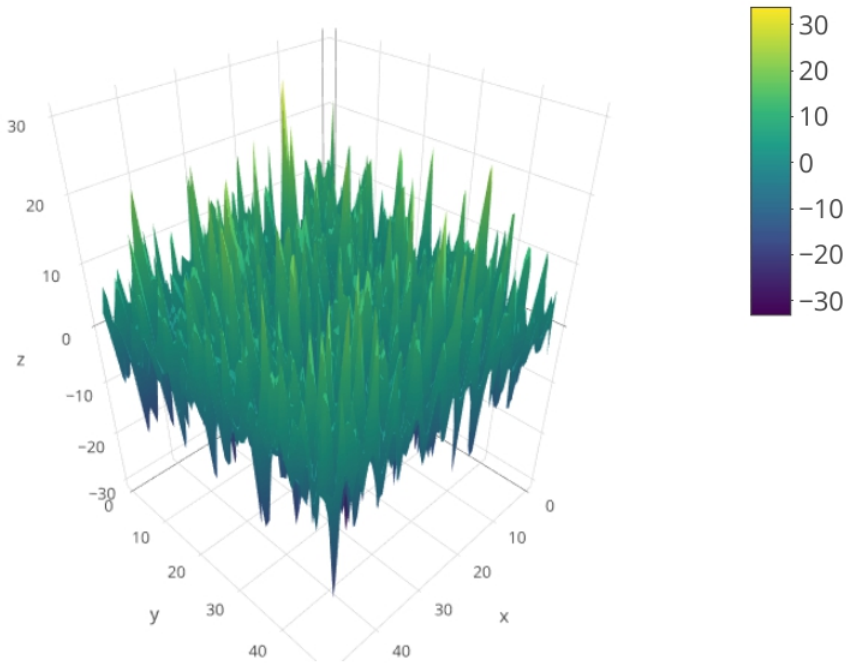

Figure 4.1: $MA(1, 1)$ random field in Example 4.5.2

The computation of the ED degree for $\mathcal{MA}_{(1,k)}$ with $k > 1$ is harder than for $\mathcal{MA}_{(1,1)}$. We therefore resort to numerical methods and obtain Table 4.2. The computations suggest the following pattern.

Conjecture 4.5.3. *The ED degree of* $\mathcal{MA}_{(1,k)}$ *equals* $\frac{(3^{k+1}-1)^2}{4}$ *for all* $k > 0$.

Note that the optimization problem (4.5.1) gives a point in $\mathcal{MA}_q$ and not a corresponding $a^* \in \mathbb{P}^Q$. Theoretically, one could apply the identifiability results of

the last section to obtain such a^* by $a_n^* := \Gamma_q^{-1}(\gamma_n^*)$. However, since γ_n^* will most often be a numerical approximation, this is not feasible in practice. Instead, one should solve the optimization problem in parametrized form:

$$a_n^* := \operatorname{argmin}_{a\in\Theta} \|\gamma_a - \hat{\gamma}_n\|,$$

where $\Theta \subseteq \mathbb{R}^{Q+1}$ is a compact parameter space. We note then that the ED degree gets multiplied by the algebraic identifiability degree of the model parametrization.

4.5.2 Maximum likelihood estimation

Suppose as before that $(Y_t)_{t\in\mathbb{Z}^d}$ is a $MA(q)$ random field with mean zero. Furthermore, we assume that n observations $Y(t^1), ..., Y(t^n)$ are given, where the vectors $t^1, ..., t^n \in \mathbb{Z}^d$ are ordered according to the lexicographic order. If the driving white noise $(Z_t)_{t\in\mathbb{Z}^d}$ is Gaussian, then the vector $Y := (Y(t^1), ..., Y(t^n))^\top$ is Gaussian as well, and its likelihood is of the form

$$L(a) \propto |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} Y^\top \Sigma^{-1} Y\right),$$

where $\Sigma = \Sigma(a)$ is the covariance matrix of Y and $|\Sigma|$ its determinant. The *maximum likelihood estimator (MLE)* is then defined as the value which maximizes the log-likelihood:

$$\hat{a}_n := \operatorname{argmax}_{a\in\Theta} -\frac{1}{2}\log(|\Sigma|) - \frac{1}{2} Y^\top \Sigma^{-1} Y, \tag{4.5.2}$$

where $\Theta \subseteq \mathbb{R}^{Q+1}$ is a compact parameter space. If Z is not Gaussian, then the latter estimator is called the *quasi maximum likelihood estimator (QMLE)*.

Remark 4.5.4 In [91] it was shown that under mild assumptions including an invertibility condition, the QMLE $\hat{a}_n$ is consistent as n tends to infinity. Furthermore, a slightly modified version of $\hat{a}_n$ (to account for the edge effect) is shown to be asymptotically normal in [91, Theorem 2].

Conveniently, the optimization problem (4.5.2) is *still* algebraic, in the sense that the critical or score equations form a system of rational functions of $a \in \Theta$. The number of critical points of the log-likelihood is invariant under generic data Y and this is known as the maximum likelihood degree (*ML degree*).

We analyze the first nontrivial case, when $q = 1$ and $n = 2$. Even this simple model is interesting. It has been observed that the MLE can sometimes correspond

to non-invertible models, which in this case is equivalent to $|a_0| = |a_1|$, and contrary to what was previously thought, this occurs with positive probability [36].

Proposition 4.5.5. *Consider the $MA(1)$ model with observed sample $Y = (Y_1, Y_2)$. The ML degree is 4, and these four critical points can be divided into three groups:*

(1) The parameters a_0 and a_1 satisfy the two equations

$$a_0 a_1 = Y_1 Y_2 \qquad \text{and} \qquad a_0^2 + a_1^2 = \frac{Y_1^2 + Y_2^2}{2}.$$

(2)

$$a_0 = a_1 = \sqrt{\frac{Y_1^2 + Y_2^2 - Y_1 Y_2}{3}}$$

(3)

$$a_0 = -a_1 = \sqrt{\frac{Y_1^2 + Y_2^2 + Y_1 Y_2}{3}}$$

If $Y_1 Y_2 = 0$, then the MLE corresponds to a degenerate model ($a_0 a_1 = 0$). Otherwise let $W = \frac{Y_1^2 + Y_2^2}{2 Y_1 Y_2}$ and the MLE is given as:

- *the point in (3) if $-2 < W < 0$*
- *the point in (2) if $0 < W < 2$*
- *the points in (1) otherwise.*

Proof. Since we have a $MA(1)$ process and $Y = (Y_1, Y_2)$, we have $\Sigma = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}$, and the log-likelihood takes the form

$$\ell(a_0, a_1) = -\frac{1}{2}\log((a_0^2 + a_1^2)^2 - a_0^2 a_1^2) - \frac{1}{2}(Y_1, Y_2)\begin{pmatrix} a_0^2 + a_1^2 & a_0 a_1 \\ a_0 a_1 & a_0^2 + a_1^2 \end{pmatrix}^{-1} (Y_1, Y_2)^\top. \quad (4.5.3)$$

There are generically four solutions to the system $\frac{\partial \ell}{\partial a_0} = \frac{\partial \ell}{\partial a_1} = 0$. This means the ML degree is 4. The critical points can be divided into the three groups (1), (2) and (3) of the statement. In order to find the MLE depending on the values of Y_1, Y_2, we evaluate the likelihood function ℓ at these 3 groups of points. In fact, substituting a_0 and a_1 from (1), (2) and (3) into the log-likelihood function, we obtain

(i) $-\frac{1}{2}\log\left((Y_1^2 - Y_2^2)^2\right) - 1 + \log(2)$,

(ii) $-\frac{1}{2}\log\left(\frac{1}{3}\left(Y_1^2 - Y_1Y_2 + Y_2^2\right)^2\right) - 1,$

(iii) $-\frac{1}{2}\log\left(\frac{1}{3}\left(Y_1^2 + Y_1Y_2 + Y_2^2\right)^2\right) - 1.$

Computing (i) - (iii) gives the expression

$$\frac{1}{2}\left(\log\left(4\left(Y_1^2 - Y_1Y_2 + Y_2^2\right)^2\right) - \log\left(3\left(Y_1^2 - Y_2^2\right)^2\right)\right),$$

which is always nonnegative since

$$4\left(Y_1^2 - Y_1Y_2 + Y_2^2\right)^2 - 3\left(Y_1^2 - Y_2^2\right)^2 = \left(Y_1^2 - 4Y_1Y_2 + Y_2^2\right)^2 \geq 0.$$

Analogously, (i) is greater than or equal to (ii) from $\left(Y_1^2 + 4Y_1Y_2 + Y_2^2\right)^2 \geq 0$.

Hence, the first value (i) is always larger than or equal to the values (ii) and (iii), independently of Y_1 and Y_2. We would conclude that the maximizers are always given by (1), but the points may not be real. Indeed, under (1), if

$$a_0a_1 \geq 0 \text{ then } a_0^2 + a_1^2 \geq 2a_0a_1 \text{ and thus } W = \frac{Y_1^2 + Y_2^2}{2Y_1Y_2} \geq 2$$

while

$$a_0a_1 \leq 0 \text{ implies } a_0^2 + a_1^2 \leq -2a_0a_1 \text{ and hence } W \leq -2.$$

Direct inspection reveals that the likelihood for (2) is larger than the one for (3) if and only if $W > 0$. Note that when $W = -2$ the points (1) and (3) coincide, while $W = 2$ means that (1) and (2) coincide. □

Compare our conditions for W with the similar ones found by [36] in their effort of computing the distribution of the MLE in this $q = 1, n = 2$ case (note the different parametrization in terms of σ, θ). Furthermore, it is gratifying to see that our computations provide a simple explanation for the 'curious' phenomenon that the MLE can belong to a non-invertible model. Algebraically, the points in (1) always maximize the likelihood, but for the specified region of Y_1, Y_2 these points are strictly complex (even though evaluating at the likelihood yields real values!), which means then that (2) or (3) becomes the MLE.

In [92], standard numerical optimization routines were used to find the MLE in samples of $MA(q)$ models with $q = 1, 2, 3, 4$. The simulations show the MLE can again lie on the non-invertible boundary.

Example 4.5.6 Consider a $MA(1)$ process with $n = 3$ sample points $Y = (Y_1, Y_2, Y_3)$. The ML degree is now 8. The expressions for the two non-invertible models $|a_0| = |a_1|$ are:

$$a_0 = a_1 = \sqrt{\frac{3x_1^2 + 4x_2^2 + 3x_3^2 - 4x_1x_2 + 2x_1x_3 - 4x_2x_3}{12}}$$
$$a_0 = -a_1 = \sqrt{\frac{3x_1^2 + 4x_2^2 + 3x_3^2 + 4x_1x_2 + 2x_1x_3 + 4x_2x_3}{12}}$$

Obtaining closed form expressions for the other 6 critical points is also possible.

For $n > 2$, the matrix Σ is tridiagonal: it has $\gamma(0)$ in the diagonal and $\gamma(1)$ in the upper and lower diagonal. Our ML degree computations of $MA(1)$ for $n = 2, 3, \ldots$ reveal the following pattern:

Conjecture 4.5.7. *The ML degree of $MA(1)$ for $n > 1$ sample points is equal to $4(n-1)$.*

In contrast, the pattern for $MA(2)$ is not as clear. The first values for $n = 2, 3, \ldots$ are recorded in Table 4.3.

n	ML degree
3	29
4	69
5	129
6	205

Table 4.3: ML degrees of the $MA(2)$ model by number of sample points

Not unusually, Gröbner basis computations quickly become prohibitive. However, this does not mean that our algebraic approach is not useful. In applied algebraic geometry, this often means one needs to go into numerical techniques. Indeed, as far as we know, the algebraic nature of the ML problem has not been exploited yet, and a numerical algebraic geometry approach brings both a fresh perspective and efficient computational tools. Knowing the ML degree beforehand helps homotopy continuation and monodromy methods find all solutions to the critical equations and thus guarantee that the MLE will be found. In contrast, classical local search

methods may only find a local maximum of the likelihood function. One way to compare these methods is to conduct simulation studies such as the one in the next example.

Example 4.5.8 We simulate 500 independent paths of a $MA(1)$ process with $n = 8$ observations for each path. In Figure 4.2 a sample path for this process is illustrated. As moving average parameters we take

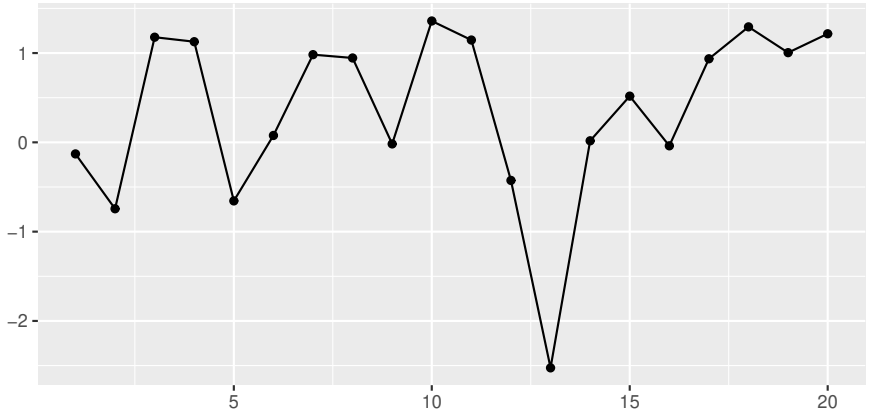

Figure 4.2: Sample path of a $MA(1)$ process.

$$a_0 = 1 \quad \text{and} \quad a_1 = 0.5.$$

The $MA(1)$ process is driven by i.i.d. standard Gaussian noise. Having simulated the $MA(1)$ process, we proceed by estimating the model parameters with the MLE in (4.5.2) in two different ways.

For our first approach we use the standard `R` command `optim` for minimization of the objective function. As it is standard in time series analysis, we take the output of the innovations algorithm as the initial value for the optimization routine (cf. [22, Section 2]).

For our second approach we differentiate the likelihood function with respect to the moving average parameters and set the derivatives to zero. In order to compute the critical points of the likelihood function we solve the resulting polynomial system using homotopy continuation. This is implemented in the `julia` package **HomotopyContinuation** [19].

Finally, we evaluate the likelihood at every critical point and choose the maximal one. The summary of the estimation results are given in Tables 4.4 and 4.5 below.

	True Value	Mean	Bias	Std
a_0	1.0000	0.8642	-0.1358	0.2459
a_1	0.5000	0.4503	-0.0497	0.4071

Table 4.4: Parameter estimation results for $MA(1)$ with $n = 8$ and `R` command `optim`.

	True Value	Mean	Bias	Std
a_0	1.0000	0.8818	-0.1182	0.2129
a_1	0.5000	0.4678	-0.0322	0.5094

Table 4.5: Parameter estimation results for $MA(1)$ with $n = 8$ and homotopy continuation.

We observe that using homotopy continuation reduces the bias for both a_0 and a_1, whereas it increases the standard deviation for a_1 and decreases the standard deviation for a_0.

Finally, we close this section by reporting the ML degree of $MA(1,1)$:

Proposition 4.5.9. *Assume that $n = 4$ sample points $Y = (Y_{11}, Y_{12}, Y_{21}, Y_{22})$ over the lattice $L = \{1,2\}^2$ of a $MA(1,1)$ random field are given. The autocovariance matrix Σ of Y is*

$$\Sigma = \begin{pmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \\ \gamma_{01} & \gamma_{00} & \gamma_{1-1} & \gamma_{10} \\ \gamma_{10} & \gamma_{1-1} & \gamma_{00} & \gamma_{01} \\ \gamma_{11} & \gamma_{10} & \gamma_{01} & \gamma_{00} \end{pmatrix}.$$

The ML degree of the model is 192 over $\mathbb{P}^3$.

Bibliography

[1] R.J. Adler. *The Geometry of Random Fields*. Wiley, Chichester, 1981.

[2] C. Améndola and V.S. Pham. Autocovariance varieties of moving average random fields. 2019. Submitted. arXiv:1903.08611.

[3] C. Améndola, J.C. Faugère, and B. Sturmfels. Moment varieties of Gaussian mixtures. *Journal of Algebraic Statistics*, 7(1):14–28, 2016.

[4] C. Améndola, K. Ranestad, and B. Sturmfels. Algebraic identifiability of Gaussian mixtures. *International Mathematics Research Notices*, 2018(21):6556–6580, 2018.

[5] C. Améndola, N. Bliss, I. Burke, C. R. Gibbons, M. Helmer, S. Hoşten, E. D. Nash, J. I. Rodriguez, and D. Smolkin. The maximum likelihood degree of toric varieties. *Journal of Symbolic Computation*, 92:222 – 242, 2019. ISSN 0747-7171. doi: https://doi.org/10.1016/j.jsc.2018.04.016.

[6] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge, 2nd edition, 2009.

[7] J.A.D. Appleby and K. Krol. Long memory in a linear stochastic Volterra differential equation. *J. Math. Anal. Appl.*, 380(2):814–830, 2011.

[8] O.E. Barndorff-Nielsen. Superposition of Ornstein-Uhlenbeck type processes. *Theory Probab. Appl.*, 45(2):175–194, 2001.

[9] O.E. Barndorff-Nielsen and J. Schmiegel. Lévy-based spatial-temporal modelling, with applications to turbulence. *Russ. Math. Surv.*, 59(1):65–90, 2004.

[10] O.E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):167–241, 2001.

[11] O.E. Barndorff-Nielsen, F.E. Benth, and A.E.D.Veraart. Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. *Banach Cent. Publ.*, 104:25–60, 2015.

[12] O.E. Barndorff-Nielsen, E. Hedevang, J. Schmiegel, and B. Szozda. Some recent developments in ambit stochastics. In F.E. Benth and G. Di Nunno, editors, *Stochastics of Environmental and Financial Economics*, pages 3–25. Springer, Cham, 2016.

[13] A. Behme. *Generalized Ornstein-Uhlenbeck Processes and Extensions.* PhD thesis, Technische Universität Braunschweig, 2011.

[14] F.E. Benth, C. Klüppelberg, G. Müller, and L. Vos. Futures pricing in electricity markets based on stable CARMA spot models. *Energy Econ.*, 44:392–406, 2014.

[15] D. Berger. CARMA Random Fields. Master's thesis, Ulm University, 2017.

[16] D. Berger. Central limit theorems for moving average random fields with non-random and random sampling on lattices. 2019. arXiv:1902.01255.

[17] D.S. Bernstein. *Matrix Mathematics.* Princeton University Press, Princeton, 2005.

[18] G.E.P. Box and G.M. Jenkins. *Time Series Analysis - Forecasting and Control.* Holden-Day, San Francisco, 1970.

[19] Paul Breiding and Sascha Timme. Homotopycontinuation. jl: A package for homotopy continuation in julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.

[20] P.J. Brockwell. Lévy-driven CARMA processes. *Ann. Inst. Stat. Math.*, 53(1): 113–124, 2001.

[21] P.J. Brockwell. Recent results in the theory and applications of CARMA processes. *Ann. Inst. Stat. Math.*, 66(4):647–685, 2014.

[22] P.J. Brockwell and R.A. Davis. Simple consistent estimation of the coefficients of a linear filter. *Stoch. Process. Appl.*, 28(1):47–59, 1988.

[23] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2nd edition, 1991.

[24] P.J. Brockwell and A. Lindner. Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stoch. Process. Appl.*, 119:2660–2681, 2009.

[25] P.J. Brockwell and Y. Matsuda. Continuous auto-regressive moving average random fields on $\mathbb{R}^n$. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 79(3):833–857, 2017.

[26] P.J. Brockwell and E. Schlemm. Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *J. Mult. Anal.*, 115:217–251, 2013.

[27] P.J. Brockwell, R.A. Davis, and Y. Yang. Estimation for non-negative Lévy-driven CARMA processes. *J. Bus. Econ. Stat.*, 29(2):250–259, 2011.

[28] P.J. Brockwell, V. Ferrazzano, and C. Klüppelberg. High-frequency sampling and kernel estimation for continuous-time moving average processes. *J. Time Ser. Anal.*, 34(3):385–404, 2013.

[29] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. *American Journal of Mathematics*, 128(3):671–697, 2006.

[30] B. Chen, C. Chong, and C. Klüppelberg. Simulation of stochastic Volterra equations driven by space–time Lévy noise. In M. Podolskij, R. Stelzer, S. Thorbjørnsen, and A.E.D. Veraart, editors, *The Fascination of Probability, Statistics and their Applications*, pages 209–229. Springer, Cham, 2016.

[31] C. Chong. Lévy-driven Volterra equations in space and time. *J. Theor. Probab.*, 30(3):1014–1058, 2017.

[32] C. Chong. Stochastic PDEs with heavy-tailed noise. *Stoch. Process. Appl.*, 127 (7):2262–2280, 2017.

[33] Planck Collaboration. Planck 2018 results. IV. Diffuse component separation, 2018. arXiv:1807.06208.

[34] D.A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer, Cham, 4th edition, 2015.

[35] N. Cressie. *Statistics for Spatial Data*. Wiley, New York, revised edition, 1993.

[36] J.D. Cryer and J. Ledolter. Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, 68(3):691–694, 1981.

[37] R.C. Dalang. The stochastic wave equation. In R. Dalang, D. Khoshnevisan, D. Nualart, and Y. Xiao, editors, *A Minicourse on Stochastic Partial Differential Equations*, pages 39–71. Springer, Berlin, 2009.

[38] R.C. Dalang and T. Humeau. Lévy processes and Lévy white noise as tempered distributions. *Ann. Probab.*, 45(6B):4389–4418, 2017.

[39] J.L. Doob. The elementary Gaussian processes. *Ann. Math. Stat.*, 15:229–282, 1944.

[40] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. Thomas. The euclidean distance degree of an algebraic variety. *Foundations of computational mathematics*, 16(1):99–149, 2016.

[41] M. Drapatz. Strictly stationary solutions of spatial ARMA equations. *Ann. Inst. Stat. Math.*, 68(2):385–412, 2016.

[42] J. Eller. On functions of companion matrices. *Linear Algebra Appl.*, 96:191–210, 1987.

[43] L.C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, 1998.

[44] J. M. Francos, A. Narasimhan, and J. W. Woods. Maximum likelihood parameter estimation of textures using a wold decomposition based model. *IEEE Trans. Image Processing*, 4:1655–1666, 1995.

[45] I. García, C. Klüppelberg, and G. Müller. Estimation of stable CARMA models with an application to electricity spot prices. *Stat. Model.*, 11(5):447–470, 2010.

[46] M. Giovannini. *A Primer on the Physics of the Cosmic Microwave Background*. World Scientific, Singapore, 2008.

[47] S.-E. Graversen and J. Pedersen. Representations of Urbanik's classes and multiparameter Ornstein-Uhlenbeck processes. *Elect. Comm. Probab.*, 16:200–212, 2011.

[48] D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at `http://www.math.uiuc.edu/Macaulay2/`.

[49] G. Gripenberg, S.-O. Londen, and O. Steffans. *Volterra Integral and Functional Equations*. Cambridge University Press, Cambridge, 1990.

[50] P. Guttorp and A.M. Schmidt. Covariance structure of spatial and spatiotemporal processes. *WIREs Comput Stat*, 5:279–287, 2013.

[51] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, 1994.

[52] J. Harris. *Algebraic Geometry*. Springer Science & Business Media, New York, 1992.

[53] L. Hörmander. *Linear Partial Differential Operators*. Springer, Berlin, 3rd edition, 1969.

[54] L. Hörmander. *The Analysis of Linear Partial Differential Operators I*. Springer, Berlin, 2nd edition, 1990.

[55] K. Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20(8):519–524, 1944.

[56] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2nd edition, 2003.

[57] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 2nd edition, 2002.

[58] K. Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung: akademische Abhandlung*. Sana, Helsinki, 1947.

[59] D. Khoshnevisan. *Multiparameter Processes*. Springer, New York, 2002.

[60] C. Klüppelberg and V.S. Pham. Estimation of causal CARMA random fields. 2019. Submitted. arXiv:1902.04962.

[61] H. Korezlioglu and P. Loubaton. Spectral factorization of wide sense stationary processes on $\mathbb{Z}^2$. *J. Mult. Anal.*, 19:24–47, 1986.

[62] R. Krishnamurthy, J. W. Woods, and J. M. Francos. Adaptive restoration of textured images with mixed spectra using a generalized Wiener filter. *IEEE Trans. Image Processing*, 5:648–652, 1996.

[63] S.N. Lahiri, Y. Lee, and N. Cressie. On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *J. Stat. Plan. Inference*, 103:65–85, 2002.

[64] S. Lang. *Complex Analysis*. Springer, New York, 4th edition, 1999.

[65] P. Langevin. Sur la théorie du mouvement brownien. *C. R. Hebd. Séances Acad. Sci.*, 146:530–533, 1908.

[66] V.A. Lebedev. The Fubini theorem for stochastic integrals with respect to L^0-valued random measures depending on a parameter. *Theory Probab. Appl.*, 40 (2):285–293, 1996.

[67] K.M. Mullen, D. Ardia, D.L. Gil, D. Windover, and J. Cline. DEoptim: An R package for global optimization by differential evolution. *J. Stat. Softw.*, 40(6): 1–26, 2011.

[68] G. Neuhaus. On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Stat.*, 42(4):1285–1295, 1971.

[69] M. Nguyen and A.E.D. Veraart. Spatio-temporal Ornstein-Uhlenbeck processes: theory, simulation and statistical inference. *Scand. J. Stat.*, 44:46–80, 2017.

[70] N. Ortner and P. Wagner. *Fundamental Solutions of Linear Partial Differential Operators*. Springer, Cham, 2015.

[71] T.W. Palmer. *Banach Algebras and The General Theory of *-Algebras Volume I*. Cambridge University Press, Cambridge, 1994.

[72] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise*. Cambridge University Press, Cambridge, 2007.

[73] V.S. Pham. Lévy-driven causal CARMA random fields. 2018. Submitted. arXiv:1805.08807.

[74] V.S. Pham and C. Chong. Volterra-type Ornstein–Uhlenbeck processes in space and time. *Stoch. Process. Appl.*, 128(9):3082–3117, 2018.

[75] B.S. Rajput and J. Rosiński. Spectral representations of infinitely divisible processes. *Probab. Theory Relat. Fields*, 82(3):451–487, 1989.

[76] J. Rosiński. Series representations of Lévy processes from the perspective of point processes. In O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick, editors, *Lévy Processes*, pages 401–415. Birkhäuser, Boston, 2001.

[77] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 2nd edition, 1991.

[78] K. Sato. *Lévy Processes and infinitely divisible distributions*. Cambridge University Press, Cambridge, 1999.

[79] A. H. Sayed and T. Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.

[80] E. Schlemm and R. Stelzer. Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electron. J. Stat.*, 6:2185–2234, 2012.

[81] R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications*. Springer, New York, 3rd edition, 2011.

[82] E.M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, Princeton, NJ, 1971.

[83] M.L. Straf. Weak convergence of stochastic processes with several parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 187–221. University of California Press, Berkeley, 1972.

[84] S. Sullivant. *Algebraic Statistics*. Graduate Studies in Mathematics. American Mathematical Society, 2018. ISBN 9781470435172. URL `https://books.google.de/books?id=EXj1uQEACAAJ`.

[85] S. Tappe. Some refinements of existence results for SPDEs driven by Wiener processes and Poisson random measures. *Int. J. Stoch. Anal.*, 2012. Article ID 236327, 24 pages.

[86] D. Tjøstheim. Statistical spatial series modelling. *Adv. Appl. Probab.*, 10(1): 130–154, 1978.

[87] V. Todorov and G. Tauchen. Simulation methods for Lévy-driven continuous-time autoregressive moving average (CARMA) stochastic volatility models. *J. Bus. Econ. Stat.*, 24(4):455–469, 2006.

[88] G.E. Uhlenbeck and L.S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36(5):823–841, 1930.

[89] J.B. Walsh. Convergence and regularity of multiparameter strong martingales. *Probab. Theory Relat. Fields*, 46(2):177–192, 1979.

[90] J.B. Walsh. An introduction to stochastic partial differential equations. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint Flour XIV – 1984*, pages 265–439. Springer, Berlin, 1986.

[91] Q. Yao and P.J. Brockwell. Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli*, 12(3):403–429, 2006.

[92] Y. Zhang and A.I. McLeod. Fitting MA(q) models in the closed invertible region. *Stat. Probab. Letters*, 76(13):1331–1334, 2016.