University College London
Department of Statistical Science

Tensor Approximation of Generalized Correlated Diffusions and Applications

Antonio Dalessandro

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy
of the
University College London

September 12, 2017

Contents

Abstract

This thesis documents my research activity conducted in the past three years at the Department of Statistical Science at the University College London. My investigation is focused on functional-analytic methods applied to the characterization of generalized correlated Markov processes. The main objective of the research is to formalize the properties of such a class of stochastic processes when approximated in a tensor space. This lead to the development of a new interpretation of the correlation among processes that is exploited for the analysis of copula functions and their statistical properties.

London, August 2017
Antonio Dalessandro

Impact Statement

The thesis is concerned with a new way to construct continuous-time Markov chains (CTMC) which are employed to mimic high dimensional correlated Ito process. The research has been carried out over three years from both theoretical and numerical perspectives, with extensive computational examples. The range of applications of CTMC in both academic research and Industry is vast. Continuous-time Markov chains (CTMC) are used to model various random phenomena occurring in finance, economics, genetics, queueing theory, demography, epidemiology, and social science. High dimensional CTMC is specifically adopted for their properties in data search engines, most notably by the Google search engine, and also by big data applications.

Therefore not only the chosen thesis' topic is very popular for a wide range of applications but offers practical perspectives to improve them.

In a very simple and high-level way, the scope of this research was to mimic the realized (terminal) density of a multidimensional Ito process. In doing so the original process, which exhibits non-trivial correlations, is decomposed in orthogonal independent mimicking Birth-Death processes, each characterized by its terminal density. Then the mimicking joint terminal density is put together through suitably constructed continuous-time Markov chains spanning orthogonal state spaces. This methodology proved to have two main advantages: one consists in the ability to simplify the complexity of the original Ito processes while preserving their dependence structure and the other advantage is the capacity to tackle the curse of dimensionality when computing the approximated weak solution of a high dimensional Ito process.

The research also showed how to construct CTMCs which accurately mimic the correlation structure among Ito processes, and how to specify and calculate concordance measures. Numerical examples showed that the methodology is computationally efficient to evaluate complex measures of concordance among processes. In this respect, the thesis opens new research perspectives on Copula functions, dependence modelling and mimicking non-Gaussian processes.

Therefore the proposed methodology has a wide range of applicability both in Academic Research and Industry. The content of the thesis can be valuable from a business perspective, especially for financial modelling, market execution of financial contracts and improvement of the accuracy and response times of web search engines.

Acknowledgements

I am immensely grateful to Prof. Gareth W. Peters and to Dr. Codina Cotar for the time we spent together discussing ideas and the way it is possible sometimes to make them work and happen. A remarkable thank to Prof. Mark Girolami who gave me support at beginning of my PhD studies, encouragements and stimulating discussions. I thank also Mark Davis, Damiano Brigo, Luca Dalessandro, Chris Oduneye, Matthew Whihelm, Andreas Eckner for their friendship, help and inspiration.

Declaration

I, the undersigned, hereby declare that the work presented in this thesis is my own. When material has been sourced from other books, journal papers, notes and authors, these have been acknowledged. This thesis has not previously been presented for this or any other PhD examinations. Furthermore the content of this thesis is by no means related to any opinions that might be held or practices that might be followed at the institutions I worked for, I work for or I am associated with.

London, August 2017
Antonio Dalessandro

Notation and Assumptions

General

X_t, S_t	Markov process, with $t \geq 0$.
M_t	Martingale process.
$X_t^{(n)}$	Markov chain, a collection of discrete random variables with values in a finite set, $\mathcal{X} = \{x_0, x_1, x_2, \ldots x_n\}$.
$\mathbb{E}$	Expectation.
$b(X_t)$	Drift function, continuous vector valued function $(b_j) = b(t, X_t) : \mathbb{R}^d \to \mathbb{R}^d$.
$\Sigma(X_t)$	Continuous, symmetric, positive definite, $d \times d$ matrix valued function.
W_t or $W(t)$	Standard Wiener process.
$<,>$	Quadratic variation.
ρ_{ij}	Instantaneous correlation between Brownian B_t^i and B_t^j, $\rho_{ij} = d{<}B_t^i$ and $B_t^j{>}/dt$.
$o(x)$	The symbol $o(x)$, is one of the Landau symbols and is used to symbolically express the asymptotic behavior of a given function. Let f and g be two functions defined on $\mathbb{R}$. $f(x) = o(g(x))$ means for all $c > 0$, there exists some $k > 0$ such that $0 \leq f(x) < cg(x)$ for all $n \geq k$.
$O(x)$	The symbol $O(x)$, is one of the Landau symbols and is used to symbolically express the asymptotic behavior of a given function. Let f and g be two real functions and we write $f(x) = O(g(x))$ if and only if $\lvert f(x) \rvert \leq M \lvert g(x) \rvert$ for $x > \bar{x}$ where $M > 0$, $\bar{x} \in \mathbb{R}$.
$'$ or ^{T}	Symbol for transpose.
$\otimes$	Kronecker product symbol.
$\oplus$	Kronecker sum symbol.
$\oplus_S$	Kronecker sum of a sequence symbol.
$\lim_{x \to a^+} f(x)$, $\lim_{x \downarrow a} f(x)$	The limit as x decreases in value approaching a.
$\lim_{x \to a^-} f(x)$, $\lim_{x \uparrow a} f(x)$	The limit as x increases in value approaching a.

Time, discretisations, Time Steps

s, t, u	Three time instants, $0 \leq s \leq t \leq u$.
Δt, h	Time step in discretisation.
Δx, h	Space step in discretisation.
n	Variable usually used to denote the number of space discretisation points.
$x_0, x_1, \ldots$	A sequence of state points.

Functions, Spaces, Function Spaces

$\mathbb{R}^d$	d-dimensional space of real numbers.
$\mathcal{X} \subset \mathbb{R}$	Countable set.
$C^k(S)$	Set of functions with k continuous derivatives on the domain S.
C^0	Space of continuous functions.
C^1	Space of continuously differentiable functions.
C^∞	Space of functions for which all orders of derivatives are continuous.
$\|f\|_\infty$	The uniform norm assigns to real valued bounded functions f defined on a set S the non-negative number $\sup_{x\in S} \|f(x)\|$.

Operator Semigroups

T_t	Markov semigroup.
A, L	Infinitesimal generators of the semigroup T_t.
$A^{(n)}_{X\|Y}$	Sequence of approximated infinitesimal generator matrices $\{A^{(n)}_{X\|Y}\} \in \mathbb{R}^{n\times n}$. The length of the sequence is equivalent to the number of states $y \in \mathcal{Y}$.

Assumptions and Research Setting

A.1 The analysis takes place on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, with P positive measure.

A.2 The problem formulations and associated mathematical assumptions were specifically selected to ensure that all problems were formulated in the thesis to correspond to cases where it is known and previously proven in papers and books cited that a unique solution exists and is well defined. Any more general contexts outside of these assumptions not specifically stated are not of relevance to the work in this thesis.

A.3 We assume that the diffusion covariance $\Sigma(X_t)$ is always a positive definite matrix valued function.

S.1 In this dissertation we do not consider ODEs or investigate SDEs with time dependence.

S.2 In this dissertation we do consider time-homogeneous diffusions only.

S.3 In this dissertation we do not deal with time dependent infinitesimal generators for diffusions or with time discretization techniques and we are not proposing an Alternating Direction Implicit (ADI) method.

S.4 In this dissertation we do not require the considered diffusions to have an invariant measure.

S.5 The proposed numerical examples and convergence results reported in the thesis are confined to Gaussian transition densities with constant coefficients, and convergence results are obtained against the corresponding closed form expression of the probability density function.

S.6 This is neither a finite difference scheme for SDEs nor an SDE discretization scheme like Euler.

S.7 Numerical evidence of convergence is reported in the Gaussian case.

S.8 There is no time discretization involved, because we exploit the infinitesimal transition densities and we use the same instantaneous law to compute terminal densities.

Chapter 1

Introduction and Literature Review

In this thesis we investigate aspects of the semimartingale decomposition and martingale representation for multi-dimensional correlated Markov processes. The aim of our research is to construct a mimicking continuous time Markov chain that approximates such processes and their dependence structure. This work is motivated by the problems of finding the expression of a multi-dimensional CTMC that both closely follows the dynamics of the corresponding correlated Ito processes and also can effectively deal with the representation and simulation of large dimensional processes that exhibit various correlation structures. This proposed methodology proves to be very useful to understand and explain dependence and concordance among processes.

Although the literature on Markov processes and Markov chains is very rich and mature, see Rogers & Williams [2000]; Ethier & Kurtz [2009]; Karatzas & Shreve [2000], we find that there is still room to further investigate and characterize multi-dimensional chains and the relationship between the correlation structure among marginal Markov chains and dependence concepts like copula functions Nelsen [1999] and concordance measures of dependence McNeil *et al.* [2015]; Scarsini [1984]. In fact these concepts have always been treated separately, and there is lack in literature of a theory that reconcile them.

Our findings and results show that our approach, based on linear and tensor algebra, is a powerful way to produce accurate solutions of multidimensional correlated SDEs that exhibit a correlation that can be fully modelled through copula functions. Specifically given a multi-dimensional Ito processes whose drift and diffusion terms are adapted processes, we show how to construct the approximated infinitesimal generator and how to characterize the process properties by its associated continuous time Markov chain (CTMC). We construct an approximated weak solution to the stochastic differential equation that weakly converges to the distribution of the multi-dimensional Ito processes.

We give a new interpretation of the correlation among processes using the martingale approach to the study of diffusions. The novelty is that it is possible to represent, in both continuous and discrete space, that a multidimensional correlated generalized diffusion is a linear combination of processes that originate from the decomposition of the starting multidimensional semimartingale. The only assumption required by our approximation approach is that the martingale problem for the associated generator of the multidimensional Markov process is well posed. Stroock & Varadhan [2007] formulated the martingale problem as a means of studying Markov processes, especially multidimensional diffusions. This approach is deemed to be more powerful and more intrinsic than the alternative approaches represented by the Markov process approach and the Ito approach.

Our result reconciles with the existing theory of diffusion approximations and decompositions in the probability literature and is more closely related to the work Gyöngy [1986] and more recently to Brunick & Shreve [2013]. In the seminal manuscript Gyöngy [1986], the author considers a multidimensional Ito process, and construct a weak solution to a stochastic differential equation which mimics the marginals of the original Ito process at each fixed time. The drift and covariance coefficients for the mimicking process can be interpreted as the expected value of the instantaneous drift and covariance of the original Ito process, conditional on its terminal value. In Brunick & Shreve [2013] the authors extend the result of Gyöngy [1986], proving that they can match the joint distribution at each fixed time for various functionals of the Ito process. The mimicking process takes the form of a stochastic functional differential equation and the diffusion coefficient is given by the so-called Markovian projection. In our framework we further generalize findings from Brunick & Shreve [2013] and the mimicking process takes the form of a sequence of conditional continuous time Markov chains with instantaneous drift and diffusion coefficients given by projected instantaneous local moments.

The results reported in this thesis define the general representation of the approximated infinitesimal generators for both multidimensional generalized diffusions and for copula densities. In fact our proposed framework is also a tractable generic model for copula functions along the lines of Dalessandro & Peters [2017] and we mainly look at two aspects of copulas' characterization: firstly we aim to represent any copula within a local Gaussian correlation framework based on tensor algebra that allows us to visualize and quantify their linear and non linear dependence structures. Secondly we calculate within the proposed framework concordance measures of dependence generated by the approximated copula function.

We create an algebraic framework based on CTMC which allows us to represent and characterize copula functions, which not only constitutes a novel contribution to the existing literature, but allows gaining deeper understanding of copula functions parameters.

A copula distribution is simply a multivariate probability distribution for which the marginal probability distribution of each variable is uniform, see Nelsen [1999]. Copulas are often used in high-dimensional statistical applications as they allow one to separate out the modelling and estimation of the distribution of dependent random variables by estimating first the marginals and then capturing the dependence structure through estimation of a copula function. Their range of applications is very broad, see a recent review of many aspects and applications of this modelling approach in Cruz *et al.* [2015a] and works such as Demarta & McNeil [2005], Aas *et al.* [2009], Genest *et al.* [2011], Haff *et al.* [2013] and Patton [2009]. Furthermore, there is an increasing number of copula models aimed at modelling different forms of dependence, both parametrically and non-parametrically, see discussions for instance in Nelsen [1999], Genest & Favre [2007], Cherubini *et al.* [2004], Joe [1997], Embrechts *et al.* [2001] and Brigo *et al.* [2010a].

However, in many cases there is no clear and intuitive correspondence between the magnitude of a copula's parameters and the dependence structure they create. This has led to numerous studies of asymptotic relationships between concordance measures such as tail dependence and the copula functions tail behaviours, as characterized by the copula parameters. In some cases this has yielded simple relationships that are closed form and analytic. However, more often than not there is still a need to better understand how dependence manifests locally in the support of the target copula distribution, and how it changes with changes in the copula parameters or the dimension of the copula since it is often difficult to characterise in closed form.

In our modelling approach we rely on the tensor approximation of the joint multidimensional density for local copula calculation, so that any copula can be approximated on the tensor product space. The representation of joint Markov processes through approximated correlated continuous time Markov chains is not new in literature. See for example Di Graziano & Rogers [2009] and Kushner & Dupuis [2001]. However our model is new in the sense that the copula approximation we propose converges in a weak sense to the copula density function, see Dalessandro & Peters [2017].

A tractable generic copula modelling framework should be of interest to a variety of researchers and practitioners. In fact copulas are often used in high-dimensional statistical applications as they allow one to easily model and estimate the distribution of correlated random variables by estimating marginals and copulae separately Nelsen [1999]. Their range of applications is very broad. In reliability engineering, copulas are being used for reliability analysis of complex systems of machine components with competing failure modes Wu [2014]. Copulas are used in modelling turbulent partially premixed combustion, which is common in practical combustors Ruan *et al.* [2014] and they have also been successfully applied to the analysis of neuronal dependencies Eban *et al.* [2013]. Copulas are useful in portfolio risk management McNeil *et al.* [2015]; Meucci [2011] because they help to analyse the effects of financial market regimes and such functions are also very popular in derivatives pricing where dependence modelling with copulas is widely used for example in the pricing of collateralized debt obligations (CDOs) Brigo *et al.* [2010b]. Although copula functions are simple to use their parametric expression is neither intuitive nor reflecting the dependence structure that they create, most of the times very non linear Nelsen [1999]. This leads to some challenging problems like the difficulty of copula parameter(s) interpretation, the evaluation of concordance measures Scarsini [1984] and the understanding of the relationship between these concordance measures and the copula parameter(s) Joe [1990, 1997].

For these reasons, various authors have proposed new ways to analyze the dependence structure of extreme observations like investigating the asymptotic shape of the level sets of the joint density Balkema *et al.* [2012, 2013], tail density approach Li & Wu [2013] or through a transformation-based nonparametric estimation of multivariate densities Chang & Wu [2015]. Nevertheless all these models provide an approximate estimation of dependence and are therefore unsuitable analyzing concurrence measures.

In Taylor [2007] they provided a representation of the axioms of a concordance measure explicitly in terms of copula models, which extends previous work by Scarsini [1984]. This provides a direct link between these measures of dependence and the copula model. However, a good understanding of the strength or significance of a concordance measure as a function of the copula model parameters is not well understood and difficult to study as they often involve complicated intractable integrations of copula functions, not easy to achieve in a computationally efficient and accurate manner in arbitrary dimensions. In this thesis we aim to address this challenge in a general approach that is numerically accurate

and computationally efficient.
The thesis is organized as follows. Chapter 2 introduces some notation and the mathematical background and settings for our analysis that is centered on aspects of birth-death processes, continuous time Markov chains and multi-dimensional diffusions and also includes some definitions and properties of tensor algebra and Kronecker products. These notions will be recurrently used throughout the thesis. In Chapter 3 we explain how it is possible to mimic a diffusion through a birth-death process and we present additional properties of the continuous-time Markov chain and its semigroup and infinitesimal generator matrix. Here we focus on the approximate solutions of stochastic differential equations (SDEs) by means of mimicking continuous time Markov chains (CTMC). We provide in this chapter some basic and illustrative examples on how it is possible to calculate numerical solutions of diffusions using CTMC.

In Chapter 4 we formalize the approximation of a multidimensional diffusion by means of a mimicking CTMC. The material of this chapter is based on Dalessandro & Peters [2017]. The extent of the proposed approximation is novel in the fact that we approximate the multidimensional correlated SDEs by means of conditional independent marginal Markov chains, as shown in Proposition 5. The solution produced by this proposed scheme not only converges to the original joint diffusion processes, as demonstrated in Section 4.3, but can be exploited, by means of basic tensor algebra, to compute solutions in high dimensionality with parsimonious computational cost compared to a full CTMC, as illustrated by the numerical results presented in Section 4.3.4.

The material in Chapter 5 and Chapter 6 is based on Dalessandro & Peters [2016b] and Dalessandro & Peters [2016a] respectively. In Chapter 5 we give more insights about the approximated local correlated Gaussian model for dependent Markov processes and we introduce and characterize the local functional Gaussian copula. The key results produced are propositions 9 and 10, where we characterize the conditional CTMC operator decomposition and we formulate the CTMC functional copula infinitesimal operator respectively. In this chapter we also show how it is possible to map our model to different families of copula functions. The copula mapping procedure is described in Section 5.2.3 and we prove that there is always existence of a solution to the mapping problem in Proposition 11.

In Chapter 6 we describe and perform tail dependence analyses and concordance measure calculations. We apply our proposed methodology to capture and measure different concordance measures and quantify dependence in stochastic processes. We demonstrate through numerical examples how our technique is accurate and computationally efficient to evaluate concordance measures for a given copula. Chapter 7 concludes.

1.1 List of publications and pre-prints

The proposed thesis is the collections of results which have been published or about to be published in the following peer-reviewed journals:

1. Chapter 4 is based on the published paper 'A. Dalessandro and G.W. Peters. Tensor Approximation of Generalized Correlated Diffusions and Functional Copula Operators. Methodology and Computing in Applied Probability (MCAP). Springer US. 2017'. DOI: https://doi.org/10.1007/s11009-017-9545-8.

2. Chapter 5 is based on the submitted paper, currently under review 'A. Dalessandro

and G.W. Peters. Mimicking Copulas in a Tensor Space and Generalized Copula Mapping'. Bernoulli Journal, 2018.

3. Chapter 6 is based on the paper 'A. Dalessandro and G.W. Peters. Efficient and Accurate Evaluation Methods for Concordance Measures via Functional Tensor Characterizations of Copulas. Methodology and Computing in Applied Probability (MCAP). Springer US. 2019'. The paper has been accepted for publication.

4. 'A. Dalessandro and G.W. Peters. Approximate Solutions to SDEs using CTMC with applications to Quantile Dynamics'. Working paper, Department of Statistical Science, University College London. 2017.

1.2 List of invited presentations

C.1 Tensor Approximation of Generalized Correlated Diffusions and Functional Copula Operators. MCQMC2014, April 611, 2014, KU Leuven, Belgium

C.2 Options Arbitrage Bounds and Volatility Smile Dynamics: a Tensor Approach. 9th World Congress of the Bachelier finance Society, 15-19 July 2016, New York.

C.3 Functional Tensor Characterizations of Copulas and Generalized Copula Mappings. MCQMC2016 August 14-19, 2016, Stanford, California

C.4 Functional Tensor Characterizations of Copulas and Application to Machine Learning. NIPS 2017 Monday December 04 – Saturday December 09, 2017, Long Beach Convention Center, Long Beach, California.

C.5 On the relationship between Copula functions and Radon-Nikodym derivative of discrete measures. Constructing multivariate probability measures by local projections. Bayes Centre Colloquium 2019, February 2019, Edinburgh.

Chapter 2

Mathematical Background

This chapter is based on standard probability references which include Ethier & Kurtz [2009]; Grimmett & Stirzaker [2001] and Brzeniak & Zastawniak [1999]; Daruich [2014]; Kuo [2006] among others and provides basic background material and introduces notations needed throughout the dissertation. It briefly summarizes concepts and results about Markov processes, Markov semigroups, Markov generators, continuous time Markov chains (CTMC) with particular attention to Birth-Death process (BDP), Kolmogorov equations and diffusions. It also includes a section on tensor algebra and tensor algebraic operators which are relevant in all subsequent chapters. Among the objective of this dissertation one is to show how it is possible to mimic joint diffusions by means of multivariate Birth-Death processes and another is to introduce a numerical tractable and efficient way to compute the approximate weak solution of a multidimensional diffusion based on tensor algebra.

2.1 Preliminaries

In this dissertation we assume the existence of a probability space $(\Omega, \mathcal{F}, P)$, where the pair $(\Omega, \mathcal{F})$ defines a measurable space. In particular Ω denotes a set, $\mathcal{F}$ is a σ-algebra on Ω namely a collection of subsets of Ω which is closed under complement and countable union. $\mathcal{F}$ is also referred as the set of all events of a random variable. P is a probability measure on $(\Omega, \mathcal{F})$ which is a map $P : \mathcal{F} \to [0, 1]$, normalized $P(\Omega) = 1$ and σ-additive. If $A \in \mathcal{F}$ is an event, we denote by $P(A)$ the probability of such event to occur. We denote with the pair (S, d) a locally compact metric space S with metric d. The notation $\mathcal{B}(S)$ indicates the Borel σ-algebra on S, while $\mathcal{P}(S)$ indicates the set of all the probability measure on S. We also consider the functions $f : \Omega \to \mathbb{S}$ which are measurable with respect to the σ-algebra $\mathcal{F}$, and denote the set of such functions by $M(S)$, while the set $B(S) \subset M(S)$ is the Banach space of bounded functions with $\|f\| := \sup_{x\in S} |f(x)| < \infty$. $C(S) \subset B(S)$ denotes the subspace of bounded continuous functions. We also denote the space of $\mathbb{R}^d$-valued continuous functions defined on $[0, T]$ by $C([0, T], \mathbb{R}^d)$, and the space of functions that are right continuous with left-hand limits endowed with the Skorohod topology by $D([0, T], \mathbb{R}^d)$.

2.2 Markov Processes

In this section we introduce basic concepts, definitions and properties concerning Markov processes.

Definition 1. *(Stochastic process, see Brzeniak & Zastawniak [1999] pag. 139) A stochastic process in* $\mathbb{R}^d, d \geq 1$*, with* $\mathbb{R}^d$ *denoting the d-dimensional Euclidean space, is a family of random variables* $\mathbf{X}(\cdot) = \{\mathbf{X}(t) \in \mathbb{R}^d, t \geq 0\}$*, parametrized by time* $t \in T$*, where* $T \subset \mathbb{R}$*. For each* $t \in T$*,* $\mathbf{X}(t)$ *is an* $\mathbb{R}^d$ *random vector, and when* $T = \{1, 2, \ldots\}$*, we shall say that* $\{\mathbf{X}(t), t \geq 0\}$ *is a stochastic process in discrete time. When* T *is an interval in* $\mathbb{R}$*, typically* $T = [0, \infty)$*, we shall say that* $\{\mathbf{X}(t), t \geq 0\}$ *is a stochastic process in continuous time. For every* $\omega \in \Omega$ *the function*

$$t \mapsto \mathbf{X}(t, \omega), \quad t \in T \tag{2.1}$$

is called sample path of $\{\mathbf{X}(t), t \geq 0\}$*. We write it as* $\{\mathbf{X}(t), t \geq 0\}$*, or simply* $\mathbf{X}(t)$ *or* $\mathbf{X}_t$ *if there is no confusion with* $\mathbf{X}_t$*,* $t \geq 0$ *denoting a d-dimensional Markov process on the probability space* $(\Omega, \mathcal{F}, P)$ *with* $\mathbf{X}_t := (X_t^1, \ldots, X_t^d) \in \mathbb{R}^d$ *denoting its components vector at time* t*.*

Definition 2. *(Markov Process) If a stochastic process* $\{\mathbf{X}(t), t \geq 0\}$ *taking values in* $\mathbb{R}^d$ *and adapted to a filtration* $\{\mathcal{F}(t); t \geq 0\}$ *satisfies the condition*

$$P(\mathbf{X}(t+s) \in A | \mathcal{F}(t)) = P(\mathbf{X}(t+s) \in A | \mathbf{X}(t)), \tag{2.2}$$

for every $s, t \geq 0$ *and* $A \in \mathcal{B}(\mathbb{R}^d)$*, then* $\mathbf{X}(t)$ *is a Markov process. Property (2.2) is referred as the Markov property.*

We denote by D the subset of the space $L^1 = L^1(\Omega, \mathcal{F}, P)$ which contains all the density functions which have the following properties:

$$D = \{f \in L^1 : f \geq 0, \|f\| = 1\} \tag{2.3}$$

.

Definition 3 (Markov Operator)**.** *A linear mapping* $M : L^1 \mapsto L^1$ *is called a Markov operator if* $M(D) \subset D$*.*

Definition 4 (Transition Probability Function)**.** *A function* $P_t(x, A)$ *is a time-homogeneous transition probability function on* $[0, \infty) \times S \times \mathcal{B}$ *if*

$$P_t(x, \cdot) \in \mathcal{P}(S), \qquad \text{is a measure on } [0,\infty) \times S \text{ for all } x \in S, t \geq 0, \tag{2.4}$$

$$P_.(\cdot, A) \in B([0, \infty) \times S), \qquad \text{for all } A \in \mathcal{B}(S, .) \tag{2.5}$$

$$P_0(x, \cdot) = \delta_x, \qquad \text{is the unit mass at } x, x \in S. \tag{2.6}$$

and

$$P_{t+s}(x, A) = \int P_s(x, A) P_t(y, dx) \quad y \in S, A \in \mathcal{B}(S) \text{ and } s, t \geq 0. \tag{2.7}$$

Furthermore $P_t(x, A)$ is a conditional probability function, which is the probability of moving from state x at time $t = 0$ to a state $y \in A$ in the next period at time t. That is

$$P_t(x, A) := P(X(t) \in A | X(0) = x). \tag{2.8}$$

The dynamics of a Markov process is entirely described by its transition probability function, which applies to all periods.

Definition 5 (Transition function for a time-homogeneous Markov process, see Ethier & Kurtz [2009]). *A function $P_t(x, A)$ is a transition function for a Markov process X_t if*

$$P_t(x, A) = P(X(t+s) \in A|\mathcal{F}_t) \quad \text{for all } s, t \geq 0 \text{ and } A \in \mathcal{B}(S). \tag{2.9}$$

Equivalently if

$$\mathbb{E}[X(t+s) \in A|\mathcal{F}_t] = \int f(x) P_s(X(t), dx) \tag{2.10}$$

for all $s, t \geq 0$, and $f \in B(S)$.

Definition 6. *(Markov semigroup) A Markov semigroup $\{T_t(x, dy)\}$ is a family of probability transition functions on S[1] depending on the parameter $t \in [0, \infty)$, such that*

$$\int_{y \in S} T_s(x, dy) P_t(y, dz) = P_{s+t}(x, dz), \quad \text{for } s, t \in [0, \infty) \tag{2.11}$$

and satisfies the following anxioms:

- *$T_0 = I$, with I the identity operator,*
- *$T_t\ T_s = T_{t+s}$, for $s, t \geq 0$,*
- *$\lim_{t \downarrow 0} \|T_t f - f\| = 0$, or equivalently for each $f \in L^1$ the function $t \mapsto T_t f$ is continuous.*

and its action on bounded and positive functions $f \in B(S)$ is given by

$$T_t f(x) = \int f(y) P_t(x, dy). \tag{2.12}$$

Markov processes are naturally related to Markov semigroups because the probability measure $P_t(x, dy)$ with $T_t f(x) = \mathbb{E}_x[f(X_t)]$ is the law of $f(X_t)$ starting from the value x at time 0.

Definition 7. *(Feller semigroup) A semigroup $\{T_t(x, dy)\}$ is Feller if*

- *It is strongly continuous, which means $\lim_{t \downarrow 0} \|T_t f - f\| = 0$*
- *It is bounded, so there is $M > 0$ such that $\|T_t\| \leq M$, for all $t > 0$ and $M = 1$,*

for every function $f \in C_0(S)$, where $C_0(S)$ denotes the space continuous functions vanishing at infinity.

Definition 8. *(Stationary, see Kuo [2006]) A Markov process X_t is said to be stationary if its transition probabilities $P(X(t) \in A|X(s) = x)$ depends on x and the time difference $t - s$, therefore $P(X(t+h) \in A|X(h) = x)$ does not depend on h. In this case we write*

$$P_t(x, A) = P(X(t+h) \in A|X(h) = x), \quad t > 0, x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R}). \tag{2.13}$$

[1] In all our application $S = \mathbb{R}^d$, with $d \geq 1$.

Definition 9 (Invariant probability Measure)**.** *The measure μ is an invariant for the semigroup T_t, and this means that for all $f \in L^1(\mu)$*

$$\int T_t f(x)\mu(dx) = \int f(x)\mu(dx)$$

and

$$\mu(A) = \int P_t(x, A)\mu(dx) \tag{2.14}$$

Furthermore T_t is a contraction semigroup in $L^1(\mu)$ for all t.

Remark 1. *We do not assume the existence of an invariant measure for the diffusion in the proposed examples of this thesis.*

Definition 10. *(Infinitesimal generator) The infinitesimal generator of a strongly continuous contraction semigroup is defined by the map*

$$Af := \lim_{t \downarrow 0} \frac{T_t f - f}{t} \tag{2.15}$$

for all $f \in D$.

The properties of A and D entirely specify the semigroup T_t. In fact given $f \in D$ the function $U(x,t) = T_t f(x)$ is the unique solution of the equation

$$\frac{\partial U(x,t)}{\partial t} = AU(x,t) \tag{2.16}$$

defined in D for all $t > 0$ and with $U(x, 0) = \nu_0$, the initial probability distribution.

2.3 Continuous Time Markov Chains

In this section we introduce some background definitions on continuous-time Markov chain, its transition probabilities and the associated infinitesimal generator.

Definition 11. *(Jump Process) A jump process $\{X(t), t \geq 0\}$ is a right-continuous stochastic process with piecewise constant sample paths.*

Definition 12. *(Continuous Time Markov Chain (CTMC)) Suppose that $S \subset \mathbb{R}$ is a finite or countable set. The jump process $X(t)$ defined on $(\Omega, \mathcal{F}, P)$ and taking values on S, is called a continuous time Markov chain if it satisfies the Markov property, i.e. for all the states $x, x_0, \ldots, x_n \in S$ and for any sequence of times $t_0 < t_1 < \ldots < t_n$:*

$$\begin{aligned} P(X(t_{n+1}) = x | X(t_0) = x_0, \ldots, X(t_n) = x_n) \\ = P(X(t_{n+1}) = x | X(t_n) = x_n). \end{aligned} \tag{2.17}$$

Hence the conditional probability $P(X(t_{n+1}) = x|X(t_n) = x_n)$ with respect to the random variable $X(t_n)$ is equivalent to the conditional probability $P(X(t_{n+1}) = x|X(t_0) = x_0, \ldots, X(t_n) = x_n)$ with respect to the σ-field $\sigma(X(t_0), X(t_1), \ldots, X(t_n))$. The set S, in the context of Markov chains is referred as state space of the chain and the elements of S are called states.

Definition 13. *(Transition Probabilities) A relation between the random variable $X(s)$ and the random variable $X(t), s < t$, is defined by the transition probabilities which are given by*

$$p_{ij}(s,t) = P\big(X(t) = x_j | X(s) = x_i\big), \quad s < t, \quad \text{for } i,j = 0,1,2,\dots \tag{2.18}$$

If the transition probabilities do not depend explicitly on the times s or t but depend only on the length of the time interval, $t-s$, they are called stationary or homogeneous transition probabilities; otherwise the transition probabilities are referred to as nonstationary or non-homogeneous, see Kushner & Dupuis [2001]. Assuming stationary transition probabilities we have,

$$p_{ij}(t-s) = P\big(X(t) = x_j | X(s) = x_i\big) = P\big(X(t-s) = x_j | X(0) = x_i\big), \quad s < t. \tag{2.19}$$

We denote the matrix of time-homogeneous transition probabilities or the transition matrix as

$$P(t) = (p_{ij}(t)) = P(X_t = x_j | X_0 = x_i) \tag{2.20}$$

and we will use also the equivalent notation P_t.

Definition 14. *(Stochastic Matrix) A matrix $P(t)$ is called a stochastic matrix for all $t \geq 0$, if it satisfies $\sum_j p_{ij}(t) = 1, \quad t \geq 0$ and $p_{ij} \geq 0$ for all i,j.*

Definition 15 (Transition Semigroup)**.** *The family $\{P(t), t \geq 0\}$ is called the transition semigroup of the continuous-time Markov chain and satisfies:*

1. *$P(t)$ is a stochastic matrix at all t.*
2. *The transition probabilities are solutions of the Chapman-Kolmogorov equations*

$$\sum_k p_{ik}(s) p_{kj}(t) = p_{ij}(s+t). \tag{2.21}$$

In matrix form, $P(s)P(t) = P(s+t), \quad$ for all $s,t \geq 0$.

3. *$\lim_{t \to s^+} p_{ij}(s,t) = \delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.*

We consider only stationary Markov chain, when the transition probability $P\big(X(t) = x_j | X(s) = x_i\big)$ depends only on $t - s$.

In this dissertation we are interested in the behaviour of the transition probabilities $p_{i,j}(h)$ when h is very small, which is approximatively linear in h when h is small.
This implies that we can characterize the transitions probabilities as:

$$\begin{cases} p_{ij}(h) \simeq q_{ij}h + o(h), & \text{if } i \neq j \\ p_{ii}(h) \simeq 1 + q_{ii}h + o(h), & \text{if } i = j \end{cases} \tag{2.22}$$

where $q_{ij} \geq 0$ for $i \neq j$.
The transition probabilities p_{ij} are used to derive transition rates q_{ij}. The transition rates form a matrix known as the infinitesimal generator matrix $Q = (q_{ij})$. Matrix Q defines a relationship between the rates of change of the transition probabilities.

Let $P(h) = (p_{ij}(h))$, with $h \to 0$, be the infinitesimal transition matrix and let I be the matrix of the same dimension as $P(h)$ but with ones along the diagonal and zeros elsewhere (identity matrix in the finite case). Then matrix Q equals:

$$Q = \lim_{h\to 0} \frac{P(h) - I}{h} \tag{2.23}$$

2.3.1 Computing the Solution of Kolmogorov Differential Equations by CTMC

The forward and backward Kolmogorov differential equations are expressions for the rate of change of the transition probabilities. The transition probability $p_{ij}(t + \Delta t)$ can be expanded by applying the Chapman-Kolmogorov equations. In fact, given the Chapman-Kolmogorov equation, as introduced in definition 15,

$$p_{ij}(s + t) = \sum_k p_{ik}(s)p_{kj}(t) \tag{2.24}$$

we can differentiate with respect to s and obtain

$$\frac{\partial}{\partial s} p_{ij}(s + t) = \sum_k \frac{\partial}{\partial s} p_{ik}(s)p_{kj}(t) \tag{2.25}$$

Setting $s = 0$ gives

$$\frac{\partial}{\partial s} p_{ij}(t) = \sum_k \frac{\partial}{\partial s} p_{ik}(0)p_{kj}(t) \tag{2.26}$$

equivalent to

$$\frac{\partial}{\partial s} P(t) = QP(t) \tag{2.27}$$

which defines a Kolmogorov backward equation (BKE). On the other hand, the forward Kolmogorov differential equation describes the probability distribution of a state in time t keeping the initial point fixed. It decomposes the time interval $(0, t + s)$ into $(0, t)$ and $(t, t + s)$. We now differentiate with respect to t and obtain

$$\frac{\partial}{\partial t} p_{ij}(s + t) = \sum_k p_{ik}(s) \frac{\partial}{\partial t} p_{kj}(t) \tag{2.28}$$

Setting $s = 0$ gives

$$\frac{\partial}{\partial t} p_{ij}(t) = \sum_k \frac{\partial}{\partial t} p_{ik}(0)p_{kj}(t) = \sum_k p_{ik}(s) \frac{\partial}{\partial t} p_{kj}(t) \tag{2.29}$$

equivalent to

$$\frac{\partial}{\partial t} P(s) = P(s)Q \tag{2.30}$$

The solution of the backward Kolmogorov equation is $P(t) = e^{Qt}$, given $P(0) = I$.
It is easy to show that $P_t = e^{Qt}$ is the only solution of the forward and backward equations. Let us assume that $\hat{P}_t$ satisfies the forward Kolmogorov equation, then

$$\frac{\partial}{\partial t}\Big(\hat{P}_t e^{-At}\Big) = \frac{\partial}{\partial t}\Big(\hat{P}_t\Big)e^{-At} + \frac{\partial}{\partial t}\hat{P}_t\Big(e^{-At}\Big) = \hat{P}_t A e^{-At} + \hat{P}_t(-A)e^{-At} = 0 \tag{2.31}$$

which implies that $\hat{P}_t = P_t$. The derivation in eq. (2.31) can be done in an analogous way for the BKE.

2.3.2 Birth-Death process

In this thesis we focus on a particular case of CTMC which is called birth-death process (BDP) and formally introduced below.

Definition 16. *(See Kuo [2006] pag. 76) A Poisson process with parameter $\lambda > 0$ is a stochastic process $\{N(t, \omega), t \geq 0\}$ satisfying the following four properties:*

1. $P(\omega; N(0, \omega) = 0) = 1$.
2. *For any $0 \leq s < t$ the random variable $N(t) - N(s)$ is a Poisson random variable with parameter $\lambda(t - s)$, i.e.,*
$$P(N(t) - N(s) = k) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}, \quad k = 1, 2, \ldots. \tag{2.32}$$
3. *$N(t, \omega)$ has independent increments, i.e., for any $0 \leq t_1 < t_2 < \ldots, < t_n$, the random variables*
$$N(t_1), N(t_2) - N(t_1), \ldots, N(t) - N(t_{n-1}),$$
are independent.
4. *Almost all sample paths of $N(t, \omega)$ are right continuous functions with left-hand limits, i.e.,*
$$P(\omega; N(\cdot, \omega) \text{ is right continuous with left-hand limits }) = 1.$$

Condition (2) in the above definition can be replaced by the following condition: For any $0 \leq s < t$,

$$\begin{aligned} P(N(t) - N(s) = 1) &= \lambda(t - s) + o(t - s), \\ P(N(t) - N(s) = 2) &= o(t - s), \end{aligned} \tag{2.33}$$

where $o(\delta)$ denotes a quantity such that $\lim_{\delta \to 0} \frac{o(\delta)}{\delta} = 0$.

Definition 17. *(Birth Process) A pure birth process X_t on the state space $S = \mathcal{X} = \{x_1, \ldots, x_n\}$ is a Markov chain characterized by a birth rate with Poisson distribution with parameter $\lambda(x_i)$, for all $x_i \in \mathcal{X}$ and can be specified by the following transition probabilities:*

$$p_{ij}(h) = \begin{cases} \lambda(x_i)h + o(h), & \text{for } j = i + 1 \\ 1 - \lambda(x_i)h + o(h), & \text{for } i = j \\ o(h), & \text{otherwise.} \end{cases}$$

Its generator matrix Q is given by:

$$Q = \begin{pmatrix} -\lambda(x_1) & \lambda(x_1) & 0 & & & \\ 0 & -\lambda(x_2) & \lambda(x_2) & 0 & & \\ & 0 & -\lambda(x_3) & \lambda(x_3) & 0 & \\ & & 0 & -\lambda(x_4) & \lambda(x_4) & 0 \\ & & & \ddots & \ddots & \end{pmatrix}$$

Therefore a pure birth process with birth rate $\lambda(x_i)$ when in state $X_t = x_i$ is a Poisson process with parameter $\lambda(x_i)h$ so that $\lambda(x_i)$ is the expected number of birth events that occur per unit time. The probability of a birth over a short interval h is $\lambda(x_i)h + o(h)$. Similarly, if the process is in state $X_t = x_i$ and a death rate is $\mu()i$, then the probability of a death event in a very small time interval of length h is $\mu_i h + o(h)$.

Definition 18. *(Birth and Death Process (BDP)) A birth and death process X_t on the state space S is a Markov chain characterized by a birth rate and an death rate, which are independent and with Poisson distribution with parameter $\lambda(x_i)$ and $\mu(x_i)$ respectively, for all $x_i \in S$. The BDP is specified by the following transition probabilities*

$$p_{ij}(h) = \begin{cases} \lambda(x_i)h + o(h), & \textit{for } j = i+1 \\ \mu(x_i)h + o(h), & \textit{for } j = i-1 \\ 1 - \lambda(x_i) - \mu(x_i)h + o(h), & \textit{for } i = j \\ o(h), & \textit{otherwise.} \end{cases}$$

In the case of birth-and-death process, we have both birth and death events possible, with rates $\lambda(x_i)$ and $\mu(x_i)$ respectively for all . Since birth and death processes are independent and have Poisson distribution with parameters $\lambda(x_i)h$ and $\mu(x_i)h$, their sum is a Poisson distribution with parameter $h(\lambda(x_i) + \mu(x_i))$.

Equivalently $p_{ij}(h) = \delta_{ij} + hq_{ij} + o(h)$ where δ_{ij} is the Kronecker's delta, such that $\delta_{ij} = \lim_{t\downarrow 0} p_{ij}(t)$ and

$$q_{ij} = \begin{cases} \lambda(x_i) & \text{if } j = i+1 \\ \mu(x_i) & \text{if } j = i-1 \\ -(\lambda(x_i) + \mu(x_i)) & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

In matrix notation we have

$$\lim_{h\downarrow 0} P(h) = I \tag{2.34}$$

where I is the identity matrix, with $P(0) = I$.

The q_{ij} are called transition rates, and they define the birth and death process infinitesimal generator matrix $Q = (q_{ij})$, which is

$$Q = \begin{bmatrix} -\lambda(x_1) & \lambda(x_1) & 0 & 0 & \dots \\ \mu(x_2) & -(\mu(x_2) + \lambda(x_2)) & \lambda(x_2) & 0 & \dots \\ 0 & \mu(x_3) & -(\mu(x_3) + \lambda(x_3)) & \lambda(x_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}. \tag{2.35}$$

The matrix Q has the property $q_{ii} = \sum_{i \neq j} q_{ij}$.

In what follows, we introduce diffusions and basic concepts on martingales in which will be recurrent throughout the thesis.

Definition 19. *(Diffusion) The Markov processes $\mathbf{X}_t$ has stochastic differential equation (SDE) of the form,*

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Psi(\mathbf{X}_t)d\mathbf{W}_t \tag{2.36}$$

where $(d\mathbf{W}_t, \mathcal{F}_t)$ is a d-dimensional Wiener process, $\mathbf{b}$ and Ψ are bounded $\mathcal{F}_t$-adapted processes such that $\Sigma = \Psi\Psi'$ is positive semidefinite.

We indicate with $\frac{\partial}{\partial x_i}$ or ∂_i the first derivative with respect x_i, for $i = 1, \ldots, d$.
The infinitesimal generator of the above diffusion has the expression:

$$A = \sum_{i}^{d} b_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j}^{d} \Sigma_{i,j}(x) \frac{\partial^2}{\partial x_i \partial x_j} \tag{2.37}$$

where $\mathbf{b} = (b_i(x))$, $i = 1, \ldots, d$ is a drift vector and $\mathbf{\Sigma} = (\Sigma_{i,j})$, $i, j = 1, \ldots, d$ is a dispersion matrix with $\Sigma = \Psi\Psi'$.
The dynamics of $\mathbf{X}_t$ are characterized completely by the infinitesimal operator and therefore by the laws of the drift and diffusion coefficients, including the conditional probability law. In particular the infinitesimal operator in eq. (2.37) is specified also by such coefficient with an explicit expression and represents a very convenient tool to analyse the properties of any diffusion process and to investigate the solution of its associated SDE. In this respect the connection between the generator of $\mathbf{X}_t$ and the solution of the SDE for $\mathbf{X}_t$ has a rigorous formulation given by the Martingale problem of Stroock & Varadhan [2007].

Theorem 1 (Uniqueness of SDE solution, see Stroock & Varadhan [2007], Theorem 3.2.1)**.** *The SDE*

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Psi(\mathbf{X}_t)d\mathbf{W}_t, \quad 0 \le t \le s, \ \mathbf{X}_0 = \nu \tag{2.38}$$

with ν a random variable independent from $\mathbf{W}_t$, $t \ge 0$ and $\mathbb{E}[|\nu|^2] < \infty$, has unique solution adapted to the filtration generated by $\mathbf{W}_t$ and ν, for constants $L > 0$ and $K > 0$, if the measurable functions $\mathbf{b}(\mathbf{x})$ and $\Psi(\mathbf{x})$ for $t \le s$ satisfy the following two conditions:

1. *Lipschitz continuity*

$$|\mathbf{b}(\mathbf{x}) - \mathbf{b}(\mathbf{y})| + |\Psi(\mathbf{x}) - \Psi(\mathbf{y})| \le L(|\mathbf{x} - \mathbf{y}|), \quad \textit{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d. \tag{2.39}$$

2. *Linear growth*

$$|\mathbf{b}(\mathbf{x})| + |\Psi(\mathbf{x})| \le K(1 + |\mathbf{x}|), \quad \textit{for all } \mathbf{x} \in \mathbb{R}^d. \tag{2.40}$$

It is straightforward to show the connection between the differential operator A and the probabilistic interpretation of the solution to the corresponding SDE.
If $\mathbf{X}_t$ satisfies the SDE of eq. (2.36) and f is a twice differentiable function and $f \in D$, the stochastic differential of $f(\mathbf{X}_t)$ is by Itô's Lemma is

$$df(\mathbf{X}_t) = \sum_{i}^{d} \frac{\partial f(\mathbf{X}_t)}{\partial x_i} dx_i + \frac{1}{2} \sum_{i,j}^{d} \frac{\partial^2 f(\mathbf{X}_t)}{\partial x_i \partial x_j} \Sigma_{ij} dt \tag{2.41}$$

with $\Sigma_{ij} = \sum_{k=1}^{d} \Psi_{ik}\Psi_{kj}$ and therefore it is possible to write

$$df(\mathbf{X}_t) = Af(\mathbf{X}_t)dt + \sum_{i}^{d} \frac{\partial f(\mathbf{X}_t)}{\partial x_i} \Psi_{ij}(\mathbf{X}_t) d\mathbf{W}_t^j \tag{2.42}$$

where A is the infinitesimal generator of eq. (2.37).

Proposition 1. *(Local Martingales) If $f \in C^2(\mathbb{R}^d)$, Ito's Lemma yields*

$$f(\mathbf{X}_t) = f(\mathbf{X}_0) + \int_0^t Af(\mathbf{X}_s)ds + \int_0^t \nabla f(\mathbf{X}_s)'\Psi(X_s)dW_s \tag{2.43}$$

where ∇ is the gradient operator. This means that

$$f(\mathbf{X}_t) - f(\mathbf{X}_0) - \int_0^t Af(\mathbf{X}_s)ds \tag{2.44}$$

is a local martingale. In particular the processes

$$\mathbf{M}_t = \mathbf{X}_t - \mathbf{X}_0 - \int_0^t b(\mathbf{X}_s)ds = \int_0^t \Psi(\mathbf{X}_s)d\mathbf{W}_s \tag{2.45}$$

$$M_t^i M_t^j - \int_0^t \Sigma_{ij}(\mathbf{X}_s)ds \tag{2.46}$$

are local martingales for all $t \in [0, T]$.

Definition 20. *A process $\mathbf{X}_t$ defined on the space $(\Omega, \mathcal{F}, P)$ and taking values in S is said to be a solution for the martingale problem for (A, ν) with respect to the filtration $\{\mathcal{F}_t\}$, $t \in [0, T]$ if*

1. *ν is the initial distribution of $\mathbf{X}_t$,*
2. *$\mathbf{X}_t$ is $\mathcal{F}_t$-measurable, for all t,*
3. *$\mathbb{E}[\int_0^T |Af(\mathbf{X}_u)|du] < \infty$, for all functions $f \in D$*
4. *$f(\mathbf{X}_t) - \int_0^t Af(\mathbf{X}_u)du$, is an $\mathcal{F}_t$-measurable martingale for all $t \in [0, T]$.*

2.4 Tensor algebra

In this section we briefly recall some relevant notations and definitions that will aid in the developments in future sections of the manuscript. Next, we introduce some basic tensor algebraic operators that are extended for sequences of matrices, presented in [Dalessandro & Peters, 2017, Theorem 3] and [Dalessandro & Peters, 2017, Theorem 4]. To understand the matrix sequence case, we first recall some key properties of Kronecker products, Kronecker sums and exponentiation.

Theorem 2. *Let $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$; then*

$$e^{A \oplus B} = e^A \otimes e^B. \tag{2.47}$$

Proof. The Kronecker sum of the two matrices is a matrix $C \in \mathbb{R}^{mn \times mn}$ given by:

$$M = A \oplus B = A \otimes I_n + I_m \otimes B, \tag{2.48}$$

where $I_n \in \mathbb{R}^{n \times n}$ denotes the identity matrix. The exponential of the matrix M is:

$$e^C = e^{A \oplus B} = e^{(A \otimes I_n + I_m \otimes B)} = e^{A \otimes I_n} e^{I_m \otimes B}. \tag{2.49}$$

Eq. (2.49) is true if $(A \otimes I_n)(I_m \otimes B) = (I_m \otimes B)(A \otimes I_n)$. This a property of the matrix exponential where

$$e^{(A+B)t} = e^{At}e^{Bt} \quad \text{if} \quad AB = BA. \tag{2.50}$$

We have by theorem 1 of Zhang & Ding [2013] that

$$A \otimes B = (A \otimes I_n)(I_m \otimes B) = (I_m \otimes B)(A \otimes I_n). \tag{2.51}$$

Therefore the matrices $(A \otimes I_n)$ and $(I_m \otimes B)$ in eq. (2.49) always commute. Furthermore by theorem 3 of Zhang & Ding [2013] we have that, if $A, C \in \mathbb{R}^{m\times m}$ and $B, D \in \mathbb{R}^{n\times n}$, then

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{2.52}$$

By theorem 18 of Zhang & Ding [2013] we have that, if $A \in \mathbb{R}^{m\times m}$ and $f(z)$ is analytic and $f(A)$ exists, then

$$f(I_n \otimes A) = I_n \otimes f(A) \quad \text{and} \quad f(A \otimes I_n) = f(A) \otimes I_n. \tag{2.53}$$

By combining eq. (2.48), eq. (2.49) and eq. (2.53), we have

$$e^{A\oplus B} = e^{(A\otimes I_n + I_m \otimes B)} = e^{A\otimes I_n}e^{I_m\otimes B} \tag{2.54}$$

$$= (e^A I_m) \otimes (I_n e^B) = e^A \otimes e^B. \tag{2.55}$$

□

We may then extend this result to the Matrix sequence case as follows.

Theorem 3 (Kronecker Product of Matrix Sequence, $\otimes_S$)**.** *If A is an $\mathbb{R}^{m\times n}$ matrix and $\{B_j\}, j = 1, \ldots, m$ is a sequence of m $\mathbb{R}^{p\times q}$ matrices, then the Kronecker product of a matrix sequence $A \otimes_S \{B_j\}$ is the $mp \times nq$ block matrix:*

$$A \otimes_S \{B_j\} = \begin{bmatrix} a_{11}B_1 & \cdots & a_{1n}B_1 \\ \vdots & \ddots & \vdots \\ a_{m1}B_m & \cdots & a_{mn}B_m \end{bmatrix} \tag{2.56}$$

We have that:

$$A \otimes_S \{B_j\}^T = \left(A \otimes I^{(p)}\right)\left(I^{(n)} \otimes_S \{B_j\}\right)^T = \left(I^{(n)} \otimes_S \{B_j\}\right)\left(A \otimes I^{(p)}\right)^T, \tag{2.57}$$

and also

$$A \oplus_S \{B_j\}^T = \left(A \otimes I^{(p)}\right) + \left(I^{(n)} \otimes_S \{B_j\}\right)^T = \left(I^{(n)} \otimes_S \{B_j\}\right) + \left(A \otimes I^{(p)}\right)^T. \tag{2.58}$$

Proof. Results in eq. (2.57) and eq. (2.58) follow from standard tensor algebra, see the reference Zhang & Ding [2013]. □

Theorem 4. *If A is an $\mathbb{R}^{m\times n}$ matrix and $\{B_j\}, j = 1, \ldots, m$ is a sequence of m $\mathbb{R}^{p\times q}$ matrices, then*

$$\exp\left(A \oplus_S \{B_j\}\right) \approx \exp(A) \otimes_S \exp(\{B_j\}) + O([A, \{B_j\}]). \tag{2.59}$$

Proof. See Zhang & Ding [2013]. □

Chapter 3

Mimicking Correlated Diffusions Using CTMC

In this chapter we introduce Markov chains which is, together with their properties, the main theme of this dissertation. A Markov chain is a Markov process in either discrete or continuous time with a countable or continuous state space, see Asmussen [2003]. The chapter contains facts and results on the calculation of the probabilistic solution to stochastic differential equations (SDEs) by means of a continuous time Markov chain (CTMC) approximation.

Specifically, we show how a CTMC can be used to approximate the weak solution of a diffusion though practical and illustrative examples.

3.1 Introduction

Calculating the numerical solution of stochastic differential equations (SDEs) has been the subject of research for many years and it is still a dominant area of research in probability, stochastic control, see Kushner & Dupuis [2001], physics, see Gardiner [2004], mathematical physics, see Milstein & Tretyakov [2004], and mathematics, see Øksendal [2003].

A comprehensive reference describing many different algorithms is Kloeden & Platen [2011]. Methods for the computational solution of stochastic differential equations are based on similar techniques for ordinary differential equations, but generalized to provide support for stochastic dynamics, Asmussen & Glynn [2007]. The Euler-Maruyama method, Milstein method and Runge-Kutta method (SDE), first and high order approximation schemes are popular ones, see Kloeden & Platen [2011]. Among the criteria to assess the goodness of an approximation scheme are properties and rates of convergence, accuracy in weak and strong senses, computational costs, properties when dimensionality increases. For example, methods such as the Euler method and linear multi-step methods, see Kloeden & Platen [2011], that are used for the solution of ordinary differential equations, will work very poorly for SDEs, having very poor numerical convergence. Several extension to the discretization approximations of numerical SDEs have been developed from computational perspectives, including higher order schemes which produce better weak and strong convergence rates, see for instance Glasserman [2004], Kushner & Dupuis [2001] and Marchuk [2011]. However the scope of this chapter is not to compare features and properties of different numerical schemes for SDEs, rather provide an introduction and a reference on how to compute solution of stochastic differential equations (SDE) associated to diffusions, using approximation schemes based on continuous time Markov chains, see Kushner & Dupuis [2001]. We show how it is possible to mimic the diffusion using a birth-death process and how then compute the approximate solution. This means that we seek for another process which is not a diffusion, that enjoys some certain local properties and able to mimic specific features of diffusions. The main advantage in doing so consists in the ability of characterize the problem of finding the approximate solution of diffusions using the new features of the mimicking process, which in our specific case will be tensor algebra and the ability to exploit all the algebraic properties of a multidimensional tensor space.

We aim to convey to the readers sufficient background on this topic and allow them to more comfortably access the content of the following Chapter 4, based on Dalessandro & Peters [2017], where on the other hand, we will give a full characterization of multidimensional approximation of correlated diffusion by CTMCs and investigate rates of weak convergence.

The background content of this chapter will be also a useful reference to chapters 5 and 6, based on the associated papers of Dalessandro & Peters [2016b] and of Dalessandro & Peters [2016a], where the authors develop a numerical procedure to map any copula function to a generalized Gaussian copula function. Such procedure has its root in the content of this chapter and it is used to visualize and interpret the copula functions characteristics and understanding the relationships between copula model parameters and the strength of induced concordance structures captured by particular copula types.
The content of this chapter aims to explain how to compute the solution of an SDE by CTMC, and we propose the following list of core building blocks to explain this numerical approximation:

1. The aim is then to understand how CTMC are related to diffusions, which are introduced in 3.2. In this section we give more details of a specific Markov chain, called a Birth-Death process and we explain why it can closely resemble the dynamics of a diffusion.

2. In section 3.2 we introduce the fundamental definitions and properties for a birth-death process (BDP) and why we identify this process as a suitable candidate to mimic a diffusion. We show how to construct the continuous time Markov chain and present properties of the semigroup for the approximating BDP.

3. In section 3.3, we introduce some numerical aspects of the computation of CTMC and we show how it is possible to compute the approximate solution of popular diffusions.

3.2 Mimicking a diffusion with a BDP

This section introduces the fundamental idea of mimicking a diffusion by BDP.

It is instructive to see how a Brownian motion can be derived as the limiting process of two suitably chosen linear combination Poisson processes.

Proposition 2. *Let the process M_t be specified by the stochastic differential equation (SDE)*

$$dM_t = \frac{1}{\sqrt{\lambda}}(dN_t^+ - dN_t^+) \tag{3.1}$$

with $M_0 = 0$ *where* N_t^+ *and* dN_t^- *are independent Poisson processes with rate* $\frac{\lambda}{2}$. *Then* M_t *has the following properties*

a) $\mathbb{E}[dM_t] = 0$,

b) dM_t *is memoryless,*

c) $\lim_{\lambda\to\infty} dM_t = dB_t$, *where* dB_t *is a Brownian motion.*

Proof. It is evident that $\mathbb{E}[dM_t] = 0$ from the specification of the intensities of the Poisson processes. In fact

$$\mathbb{E}[dM_t] = \mathbb{E}[dN_t^+ - dN_t^-] = 0.$$

The memoryless properties directly comes from the Poisson processes. In order to show that the limiting behaviour of dM_t is a Brownian motion we proceed by calculating the moments of dM_t. By Ito's rule we obtain

$$dM_t^p = \Big((M_t + \frac{1}{\sqrt{\lambda}})^p - M_t^p\Big)dN_t^+ + \Big((M_t - \frac{1}{\sqrt{\lambda}})^p - M_t^p\Big)dN_t^-$$

and then calculate the moments by means of the following binomial expansion

$$\frac{\partial}{\delta}\mathbb{E}[M_t^p] = \begin{cases} \sum_{i=1}^{p/2-1} \frac{1}{\lambda^{i-1}}\binom{p}{2i}\mathbb{E}[M_t^{p-2i}] & \text{if } p \text{ is even,} \\ 0 & \text{if } p \text{ is odd.} \end{cases} \tag{3.2}$$

If we solve the recursion for p and take the limit for the Poisson rate $\lambda \to \infty$ we obtain the moments of a Gaussian distribution:

$$\lim_{\lambda\to\infty} \mathbb{E}[M_t^p] = (p-1)(p-3)\cdots 3\Big(\frac{t}{2}\Big)^{\frac{p}{2}}. \tag{3.3}$$

Therefore

$$\lim_{\lambda\to\infty} dM_t = dB_t.$$

□

Proposition 3. *Let us consider the continuous d-dimensional Markov processes with SDE of the form,*

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Psi(\mathbf{X}_t)d\mathbf{W}_t. \tag{3.4}$$

Its solution can be interpreted as the limit to the solutions of the SDE

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + g^+(\mathbf{X}_t)d\mathbf{N}^+(t) - g^-(\mathbf{X}_t)d\mathbf{N}_t^-. \tag{3.5}$$

Proof. The result is analogue to the limiting procedure explained above in proposition (2). The vector valued coefficients $g^-(\mathbf{X}_t)$ and $g^+(\mathbf{X}_t)$ have local intensities Ψ_{ij}/λ for all $i, j = 1, \dots, d$. □

Therefore from the propositions above it is clear that a diffusion can be seen as the limiting process of the linear combination of two Poisson processes. The diffusions approximation scheme proposed in this dissertation is fundamentally based on the above propositions, but based on the following further settings:

1. Approximating a diffusion using a linear combination of two Poisson processes but on a discrete state space and therefore using a birth death process (BDP) which is a specific case of continuous time Markov chain (CTMC).

2. We specify the local instantaneous BDP intensities in order to match the first two instantaneous local moments of the target diffusion. In this way we can build the infinitesimal generator matrix associated to the BDP.

3. We use the generator matrix to compute the approximate weak solution to the target diffusion.

3.2.1 Construction of the Approximating BDP CTMC for a Diffusion

Let us consider the following one dimensional SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{3.6}$$

where the drift function $b(X_t)$ and diffusion function $\sigma(X_t)$ are such that there is a unique solution, and we construct the continuous Markov chain for the birth-death process $dX_t^{(n)} = dN_t^+ - dN_t^-$ over a discretized state space $\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}$ with uniform discretization [1] unit $h = x_i - x_{i-1}$ for all i, where the birth and death rates are calculated in order to simultaneously match the instantaneous first two local moments of eq. (3.6). Therefore given the instantaneous moments for all $x_i \in S = \mathbb{R}$

$$\begin{aligned} \frac{\partial}{\partial t}\mathbb{E}[dX_t|X_t = x_i] &= b(x_i) \\ \frac{\partial}{\partial t}\mathbb{E}[dX_t^2|X_t = x_i] &= \sigma^2(x_i) \end{aligned}$$

we construct the approximating DBP over the set $\mathcal{X} \in S$ with intensities given by the solution of the following system

$$\begin{cases} \frac{d}{dt}\mathbb{E}[dX_t^{(n)}|X_t^{(n)} = x_i] = -ha(x_i, x_{i-1}) + ha(x_i, x_{i+1}) = b(x_i) \\ \frac{d}{dt}\mathbb{E}[(dX_t^{(n)})^2|X_t^{(n)} = x_i] = h^2a(x_i, x_{i-1}) + h^2a(x_i, x_{i+1}) = \sigma^2(x_i). \end{cases} \tag{3.7}$$

In this way we can calculate the BDP rates and construct its infinitesimal generator matrix $A^{(n)}$ which entries are computed by locally solving the following system of equations:

$$A^{(n)} = (a_{ij}) = \begin{cases} a(x_1, x_2) = a(x_n, x_{n-1}) = 0, \\ a(x_i, x_{i+1}) = \frac{1}{2}\left(\frac{b(x_i)}{h} + \frac{\sigma(x_i)^2}{h^2}\right), \\ a(x_i, x_{i-1}) = \frac{1}{2}\left(\frac{\sigma(x_i)^2}{h^2} - \frac{b(x_i)}{h}\right), \\ a(x_i, x_i) = -a(x_i, x_{i-1}) - a(x_i, x_{i+1})). \end{cases} \tag{3.8}$$

where $\sigma(\cdot) > 0$ and $a_{ij} > 0$, for all $i \neq j$.

[1] The discretiziation unit of state space $\mathcal{X}$ does not need to be uniform.

3.2.2 Properties of the Semigroup for the Approximating BDP CTMC $X_t^{(n)}$

We denote the semigroup for the CTMC $X_t^{(n)} : \mathcal{X} \to \mathbb{R}$ as $\{T_t^{(n)}\}$ defined on the set $M_b(\mathcal{X})$ of all bounded measurable functions $f : \mathcal{X} \to \mathbb{R}$ with

$$T_t^{(n)} f(x) = \mathbb{E}[f(X_t^{(n)})|X_0^{(n)} = x],\ x \in \mathcal{X}. \tag{3.9}$$

The semigroup satisfies the standard properties which are:

1. $T_t^{(n)} T_s^{(n)} = T_{s+t}^{(n)}, \quad$ for $s, t \geq 0$ [Chapman-Kolmogorov]
2. $T_0^{(n)} f(x) = \mathbb{E}[f(X_0^{(n)})|X_0^{(n)} = x] = f(x)$
3. $T_t^{(n)} f(x) = \mathbb{E}[f(X_t^{(n)})|X_0^{(n)} = x] \leq \sup_{z\in\mathcal{X}} |f(z)|$, [contraction semigroup].

Another important property of the semigroup $\{T_t^{(n)}\}$ is the strong continuity, which we will use in the next chapter to highlight desirable convergence properties of the BDP $X_t^{(n)}$ to the diffusion X_t.

At this purpose we define $c^{(n)} = \max_{x\in\mathcal{X}} |f(x+h) - f(x)|$, $c^{(br)} = \max_{x\in\mathcal{X}} |a(x, x+h)|$, $c^{(dr)} = \max_{x\in\mathcal{X}} |a(x, x-h)|$. Then it is possible to show that the semigroup is continuous as following:

$$\begin{aligned}
\mathbb{E}[f(X_t^{(n)}) - f(x)|X_0^{(n)} = x] &\leq \left|(a(x, x+h))t + o(t)\Big(f(x+h) - f(x)\Big)\right| \\
&\quad \left|(a(x, x-h))t + o(t)\Big(f(x-h) - f(x)\Big) + o(t)\right| \\
&\leq c^{(n)} t\Big(c^{(br)} + c^{(dr)} + \frac{o(t)}{t}\Big).
\end{aligned} \tag{3.10}$$

Therefore the semigroup $T_t^{(n)}$ is strongly continuous,

$$\lim_{t\to 0^+} \sup_{x\in\mathcal{X}} \left|T_t^{(n)} f(x) - f(x)\right| = 0 \tag{3.11}$$

In particular $T_t^{(n)}$ is a Feller semigroup.

3.3 Numerical Introduction to CTMC and Examples

We provide some illustrative examples of the application of CTMC local approximation to diffusions illustrated above.

Solution to Popular SDE's

We calculate the solution of the following popular SDE's, see Brigo & Mercurio [2008].

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{3.12}$$

Model	drift term: $b(X_t)$	diffusion coefficient: $\sigma(X_t)$
Vasicek	$K(\theta - X_t)$	σ
Cox Ingersoll Ross	$K(\theta - X_t)$	$\sigma\sqrt{X_t}$
Dothan	KX_t	σX_t
Exponential Vasicek	$X_t(\eta - K\ln(X_t)$	σX_t
Hull & White	$K(\theta_t - X_t)$	σ

Table 3.1: Summary of popular SDEs coefficients specifications.

We construct the BDP Markov chain mimicking the dynamics of the SDE in eq. (3.12) and calculate the approximated generator matrix $A^{(n)} \in \mathbb{R}^{n\times n}$ of eq. (3.8). Then, in line with Section 2.3.1, we calculate the approximated solution $U(x)$ of the SDE in eq. (3.12) by solving

$$\frac{\partial U(x)}{\partial t} = A^{(n)}U(x). \tag{3.13}$$

given an initial distribution ν_0.

Black-Scholes Solution

We calculate the solution to the popular Black-Scholes SDE, see Black & Scholes [1973], by means of CTMC approximation. Assume that the asset price S follows the geometric Brownian motion,

$$dS_t = \mu S_t + \sigma S_t dW_t, \quad t \geq 0 \tag{3.14}$$

where, μ and σ are constant, and W_t is a Wiener process. Hereafter, we will drop the subscript t for both a better understanding and simplicity. Let $V = V(S,t)$ denote the value of an option (or a contingent claim) that is sufficiently smooth, namely, its second-order derivatives with respect to S and first-order derivative with respect to t are continuous in the domain $D_V = \{(S,t) : S \geq 0, 0 \leq t \leq T\}$.
Then, for European options the Black-Scholes partial differential equality, under the assumptions of no arbitrage, is,

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r-d)S\frac{\partial V}{\partial S} - rV = 0 \tag{3.15}$$

in the domain D_V, with r the constant risk free rate, and $d \geq 0$ the constant dividend rate.

Pricing an European Call Option with a CTMC approximation

In this example we compare the value of an European call option value $C(S,t)$ computed using the Back-Scholes formula and an approximated value $C^{(n)}$ based on the CTMC approximation. Specifically we compare

$$C(S,t) = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \tag{3.16}$$

where N is the $\mathcal{N}(0,1)$ distribution function and

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right] \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned} \tag{3.17}$$

which

$$C^{(n)} = e^{-r(T-t)} P^{(n)}_{T-t} 1_{\{\mathcal{X}=S_t\}} \cdot \max(\mathcal{X} - K, 0) \tag{3.18}$$

with $P^{(n)}_{T-t} = e^{(T-t)A^{(n)}}$, and $A^{(n)}$ the generator matrix of the CTMC approximating the diffusion in eq. (3.14) over the discrete state space $\mathcal{X} \subset (0, \infty)$.

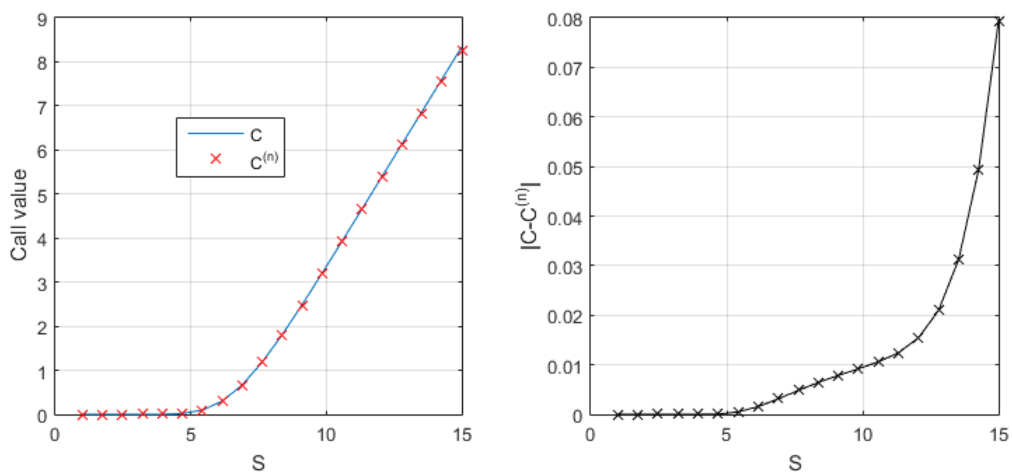

Figure 3.1: The left plot reports the values for an European call option, where C is calculated using the Black-Scholes formula of eq. (3.16) and $C^{(n)}$ is the approximated value based on the CTMC approximation of eq. (3.18). In this example we used the values $r = 5\%$, $\sigma = 20\%$, $t = 0$, $T = 1$, and the strike $K = 7$. The generator matrix $A^{(n)}$ has been constructed using eq. (3.8) and uniformly discretized state space $\mathcal{X} = \{0, \ldots, 20\}$ with n = 300.

In what follows a $T = 5$ years European call option is priced using a five states state space. A constant interest rate of $r = 5\%$, dividend yield of $\delta = 3\%$, flat volatility of $\sigma = 10\%$ are assumed. We specify the CTMC states space as:

$$\mathcal{X} = \{x_0, x_1, x_2, x_3, x_4\} = \{1, 100, 105, 110, 1000\}$$

and underlying spot price of $x_1 = x_{spot}) = 100$ and strike of $K = 100$. We use the eq. (3.8) to construct the generator $A^{(n)}$ which is

$$A^{(n)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.9713 & -20.2020 & 19.2308 & 0 & 0 \\ 0 & 243.1012 & -486.2025 & 243.1012 & 0 \\ 0 & 0 & 3.2717 & -3.2901 & 0.0184 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The right eigenvector and eigenvalues of $A^{(n)}$ are respectively:

$$V = \begin{pmatrix} 0 & 0 & 0 & 0.5417 & 0 \\ -0.0402 & -0.5534 & -0.9088 & 0.4880 & 0.0975 \\ -0.9992 & -0.5771 & -0.3902 & 0.4853 & 0.1024 \\ 0.0066 & -0.6005 & 0.1475 & 0.4826 & 0.1074 \\ 0 & 0 & 0 & 0 & 0.9841 \end{pmatrix},$$

and

$$\Lambda = \begin{pmatrix} -497.6042 & 0 & 0 & 0 & 0 \\ 0 & -0.1459 & 0 & 0 & 0 \\ 0 & 0 & -11.9446 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The inverse of right eigenvector is:

$$V^{-1} = \begin{pmatrix} -0.0010 & 0.4978 & -0.9776 & 0.4808 & -0.000 \\ 1.5022 & -0.2256 & -0.0186 & -1.4393 & 0.1813 \\ 0.0765 & -0.9409 & -0.0320 & 0.8977 & -0.0014 \\ 1.8461 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0162 \end{pmatrix}.$$

The three matrices are the elements of the eigenvector and eigenvalue decomposition which we use to calculate the transition probability matrix P_t,

$$P_t = e^{tA^{(n)}} = e^{V\Lambda V^{-1}t} = Ve^{t\Lambda}V^{-1}. \tag{3.19}$$

We obtain:

$$e^{\Lambda 5} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4822 & 0 & 0 & 0 \\ 0 & 0 & 0.0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$Ve^{\Lambda 5} = \begin{pmatrix} 0 & 0 & 0 & 0.5417 & 0 \\ 0 & -0.2668 & -0.000 & 0.4880 & 0.0975 \\ 0 & -0.2783 & -0.000 & 0.4853 & 0.1024 \\ 0 & -0.2896 & 0.0054 & 0.4168 & 0.0566 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P_t = Ve^{\Lambda 5}V^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5000 & 0.0602 & 0.0050 & 0.3841 & 0.0507 \\ 0.4778 & 0.0628 & 0.0052 & 0.4005 & 0.0536 \\ 0.4559 & 0.0653 & 0.0054 & 0.4168 & 0.0566 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The European call option $C^{(n)}$ including a dividend yield can be priced using the following equation

$$C^{(n)} = P^{(n)}_{T-t} 1_{\{\mathcal{X}=S_t\}} \cdot \max(e^{-\delta(T-t)}\mathcal{X} - e^{-r(T-t)}K, 0) \tag{3.20}$$

which yields the value:

$$C^{(n)} = 0.0000 \cdot 0.5000 + 8.1907 \cdot 0.0602 + 12.4943 \cdot 0.0050 + 16.7978 \cdot 0.3841 + 782.8279 \cdot 0.0507 = 46.7041,$$

with input quantities as follows:

$$\begin{aligned}
&e^{-\delta t} = e^{-0.03 \cdot 5} = 0.8607 \\
&e^{-rt} = e^{-0.06 \cdot 5} = 0.7788 \\
&\max(e^{-\delta t}\mathcal{X} - e^{-rt}K, 0) = (0.0, 8.1907, 12.4943, 16.7978, 782.8279) \\
&P_t 1_{S_0=100} = (0.5, 0.0602, 0.0050, 0.3841, 0.0507).
\end{aligned} \tag{3.21}$$

We included the ©MATLAB code referring to this section in Appendix A. Specifically we include in listing A.1 the code for the infinitesimal generator matrix of the 1D Black-Scholes operator. Listing A.2 contains code for the closed form Black-Scholes option pricing formula, while listing A.3 contains code for the Black-Scholes option pricing through the approximated infinitesimal generator matrix.

Useful state space transformations

In this section we look at simplifying the expression of the instantaneous local moments of the diffusion through a state space transformation. If we consider the following SDE

$$dX_t = bX_t dt + \sigma X_t dW_t \tag{3.22}$$

we can rewrite it as

$$d\log X_t = (b - \frac{\sigma^2}{2})dt + \sigma dW_t \tag{3.23}$$

or as

$$d\Big\{\Big(\log X_t - (b - \frac{\sigma^2}{2})t\Big)/\sigma\Big\} = dW_t. \tag{3.24}$$

Therefore $\Big(\log X_t - (b - \frac{\sigma^2}{2})t\Big)/\sigma \sim \mathcal{N}(0, t)$. Appendix A contains the ©MATLAB code referring to this section. Specifically we include in listing A.4 the code that illustrates how it is possible to build equivalent infinitesimal generators of a diffusion. In this case the process dynamics is lognormal and we apply state space transformations.

3.3.1 Two-dimensional Examples

In this section we anticipate what will be the topic of the next Chapter 4, namely the approximation by CTMCs of solutions to multidimensional paired diffusions, and all the findings and results are based on Dalessandro & Peters [2017]. For example, below we consider two correlated diffusions, and in the code listing A.5 we propose a way to construct the corresponding approximated infinitesimal generator.

$$\begin{cases} dX_t = b_1 dt + \sigma_1 dW_t \\ dY_t = b_2 dt + \sigma_2 dZ_t \\ \mathbb{E}[dW_t dZ_t] = \rho dt \end{cases} \tag{3.25}$$

We can decompose the correlated Brownian motions into their orthogonal components

$$\begin{cases} dX_t = b_1 dt + \sigma_1(\rho dZ_t + \sqrt{1-\rho^2} dW_t^1) \\ dY_t = b_2 dt + \sigma_2 dZ_t \\ \mathbb{E}[dW_t^1 dZ_t] = 0 \end{cases} . \tag{3.26}$$

Let us rewrite the dynamic of X_t conditionally to the process $dZ_t = z$.

$$(dX_t | dZ_t = z) = b_1 dt + \sigma_1(\rho z + \sqrt{1-\rho^2} dW_t^1). \tag{3.27}$$

However dZ_t can be expressed as a normalized dY_t, namely $dZ_t = \frac{1}{\sigma_2}(dY_t - b_2 dt)$ which yields the equivalent expression

$$(dX_t | dY_t = y - y_0) = b_1 dt + \sigma_1 \Big(\rho(\frac{1}{\sigma_2}((y-y_0) - b_2 dt)) + \sqrt{1-\rho^2} dW_t^1 \Big), \tag{3.28}$$

with y_0 denoting the starting value of the process Y_t at $t = 0$.
Therefore we have that the conditional random process $(dX_t | dY_t = y - y_0)$ follows the distribution with moments given by:

$$(dX_t | dY_t = y - y_0) \sim \mathcal{N}\Big(b_1 + \rho \frac{\sigma_1}{\sigma_2}((y-y_0)/dt - b_2) dt, \sigma_1^2(1-\rho^2) dt \Big). \tag{3.29}$$

Appendix A contains the ©MATLAB code referring to this section. Specifically we include in listing A.5 the code for the operator matrix associated to paired lognormal processes with instantaneous conditional moments as per eq. (3.29). In this case the joint processes are lognormal and we apply state space transformations as per Section 3.3 .

3.4 Conclusions

We presented introductory concepts about continuous time Markov chains and the way they can be used to construct an approximate solution to an SDE. The theory is accompanied by illustrative examples and Matlab code listings. We aimed to convey to the readers sufficient background on this topic and allow them to more comfortably access the content of the following Chapter 4, based on Dalessandro & Peters [2017], where on the other hand, we will give a full characterization of multidimensional approximation of correlated diffusion by CTMCs and investigate rates of weak convergence.

Chapter 4

Tensor Approximation of Correlated Diffusions

In this chapter we develop a class of applied probabilistic continuous time but discretized state space decompositions of a multivariate generalized diffusion process. This decomposition is novel and in particular it allows one to construct families of mimicking classes of processes for such continuous state and continuous time diffusions in the form of a discrete state space but continuous time Markov chain representations. In this chapter we present this novel decomposition and study its discretization properties from several perspectives. This class of decomposition both brings insight into understanding locally in the state space the induced dependence structures from the generalized diffusion process as well as admitting computationally efficient representations for evaluating functionals of generalized multivariate diffusion processes.

In particular, we investigate aspects of semimartingale decompositions, approximation and the martingale representation for multidimensional correlated Markov processes. A new interpretation of the dependence among processes is given using the martingale approach. We show that it is possible to represent, in both continuous and discrete space, that a multidimensional correlated generalized diffusion is a linear combination of processes that originate from the decomposition of the starting multidimensional semimartingale. This result not only reconciles with the existing theory of diffusion approximations and decompositions, but defines the general representation of infinitesimal generators for both multidimensional generalized diffusions and, as we will demonstrate, also for the specification of copula density dependence structures. This new result provides immediate representation of the approximate weak solution for correlated stochastic differential equations. We demonstrate desirable convergence results for the proposed multidimensional semimartingales decomposition approximations.

4.1 Introduction

In this chapter we develop a class of applied probabilistic continuous time but discretized state space decompositions of the characterization of a multivariate generalized diffusion process. This decomposition is novel and in particular it allows one to construct families of mimicking classes of processes for such continuous state and continuous time diffusions in the form of a discrete state space but continuous time Markov chain representations.

We present and develop different algebraic tensor representations of this novel decomposition approach that are beneficial for different numerical approximation settings and in particular we study the approximation behaviour and discretization properties from several perspectives. This class of decomposition both brings insight into understanding locally in the state space the induced dependence structures from the generalized diffusion process as well as admitting computationally efficient representations for evaluating functionals of generalized multivariate diffusion processes.

We consider multidimensional correlated diffusions, introduced in definition 19, with stochastic differential equation (SDE) of the form

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Psi(\mathbf{X}_t)d\mathbf{W}_t, \tag{4.1}$$

where $(d\mathbf{W}_t, \mathcal{F}_t)$ is a d-dimensional Wiener process, $\mathbf{b}$ and Ψ are bounded $\mathcal{F}_t$-adapted processes such that $\Sigma = \Psi\Psi'$ is positive semidefinite.
We focus our analysis to time-homogeneous diffusions. The infinitesimal generator associated to the SDE of eq. (4.1) is,

$$A = \sum_i^d b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j}^d \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{4.2}$$

We aim to mimic the diffusion $\mathbf{X}_t$ with a suitably constructed birth-dead process (BDP) and its associated continuous time Markov chain $\mathbf{X}_t^{(n)}$, with $\mathbf{X}_t^{(n)} := (X_t^{(n_1)}, \dots, X_t^{(n_d)}) \in \mathcal{X}$ denoting its components vector at time t. In particular the Markov chain is constructed as

$$\mathbf{X}_t^{(n)} : \mathcal{X} \to \mathbb{R}^d, \tag{4.3}$$

with $\mathcal{X} = \bigotimes_{i=1}^d \mathcal{X}^{\mathbf{i}} \in \mathbb{R}^d$ denoting the d-dimensional tensor space and $n = n_1 \cdots n_d$. Specifically we denote a countably finite set by $\mathcal{X}^k := \{x_0^{(k)}, \dots, x_{n_k}^{(k)}\} \in \mathbb{R}, \;\; k = 1, \dots, d$, with n_k the number of discretization points for the k-dimension, and $\mathcal{X}^k$ represents the state space for the Markov chain $X_t^{(n_k)}$ approximating X_t^k.

We shall see in this chapter that the chain $\mathbf{X}_t^{(n)}$ has transition probability at time t equal to $P_t^{(n)}$ which weakly converges to P_t, being P_t the distribution of the admitted weak solution $\mathbf{X}_t$ of the SDE in eq. (4.1).

The aim of our work is then to calculate the approximate solution for the multivariate correlated diffusion processes of eq. (4.1), in settings which may involve potentially very high dimensional state spaces which display non-trivial dependence structures. We propose two ways to calculate the weak solution of the SDE in eq. (4.1) in an approximate manner:

1. Direct approximation of the infinitesimal generator A with particular emphasis on the structure of its mixed derivative terms.

2. Decomposition of the infinitesimal generator A into orthogonal components.

The approximation schemes we propose are based on tensor algebra decompositions such as those considered in Hackbusch [2012]. The novelty of our work consists in the introduction of new concepts like, correlated tensor representation, conditional infinitesimal generator and the framework developed to perform a parametric copula function mapping as will be detailed in the remainder of the chapter. In general the proposed results aim to develop a new characterization of the cross space among dimensions using tensor algebra. Furthermore, our schemes are supported by approximation and convergence results that constitute a little utilised perspective to look at solutions of multidimensional SDEs.

Our investigation is focused on aspects of the semimartingale decomposition and martingale representation for multi-dimensional correlated Markov processes. The objective is to construct a continuous time Markov chain (CTMC) that approximates or mimics such processes and their dependence structures induced throughout the state space, which are only implicitly defined by the joint structure of the marginal process volatility functional forms and the joint coupling of the correlation structures in the driving noise processes. Once such a mimicking process is obtained we may then transform its structure to produce a copula function mapping theorem that allows one to obtain structural characterizations of general dependence frameworks, through the specification of the mimicking multivariate diffusion process.

This work is motivated by the problems of finding the expression of a multi-dimensional CTMC that both closely follows the dynamics of the corresponding correlated Ito-processes and also can effectively deal with the representation and simulation of large dimensional processes that exhibit various correlation structures. Although the literature on Markov processes and Markov chains is very rich and mature, see Rogers & Williams [2000]; Ethier & Kurtz [2009]; Karatzas & Shreve [2000]; Kushner & Dupuis [2001], we find that there is still room for further investigation and characterization of multi-dimensional chains and the relationship between the correlation structure among marginal Markov chains and dependence concepts like copula functions Nelsen [1999] and concordance measures of dependence McNeil *et al.* [2015]; Scarsini [1984]. In fact these concepts have always been treated separately, and there is lack in literature of a theory that begins to reconcile them.

Our findings and results show that our approach, based on linear and tensor algebra, is a powerful way to produce accurate solutions of multidimensional correlated SDEs that exhibit a correlation that can be fully modelled through copula functions. Specifically given a multi-dimensional Ito processes whose drift and diffusion terms are adapted processes, we show how to construct the approximated infinitesimal generator and how to characterize the process properties by its associated continuous time Markov chain (CTMC). We construct an approximated weak solution to the stochastic differential equation that weakly converges to the distribution of the multi-dimensional Ito process.

We develop an interpretation for the correlation among processes using the martingale approach applied to the study of diffusions. The novelty is that it is possible to represent, in both continuous and discrete space, that a multidimensional correlated generalized diffusion is a linear combination of processes that originate from the decomposition of the starting multidimensional semimartingale.

The only assumption required by our approximation approach is that the martingale problem for the associated generator of the multidimensional Markov process is well posed. Stroock & Varadhan [2007] formulated the martingale problem as a means of studying Markov processes, especially multidimensional diffusions. This approach is deemed to be more powerful and more intrinsic than the alternative approaches represented by the Markov process approach and the Ito approach. More recently some authors Brunick [2013] extended the study of the martingale problem associated to operators of the same type as in eq. (4.2) and showed that this is well posed when the covariance matrix takes a particular lower-diagonal block form.

Our result reconciles with the existing theory of diffusion approximations and decompositions existing in the probability literature and is more closely related to the work of Gyöngy [1986] and more recently to Brunick & Shreve [2013]. In the seminal manuscript Gyöngy [1986] considers a multi-dimensional Ito process, and constructs a weak solution to a stochastic differential equation which mimics the marginals of the original Ito process at each fixed time instant. The drift and covariance coefficients for the mimicking process can be interpreted as the expected value of the instantaneous drift and covariance of the original Ito process, conditional on its terminal value. In Brunick & Shreve [2013] the authors extend the result of Gyöngy [1986], proving that they can match the joint distribution at each fixed time for various functionals of the Ito process. The mimicking process takes the form of a stochastic functional differential equation and the diffusion coefficient is given by the so-called Markovian projection. In our framework we further generalize findings from Brunick & Shreve [2013] and the mimicking process takes the form of a sequence of conditional continuous time Markov chains with instantaneous drift and diffusion coefficients given by projected instantaneous local moments. We use generalized diffusion approximations, similarly to the work of E. Ekström & Tysk [2013], in order to produce desired target multivariate distributions. However the results presented in E. Ekström & Tysk [2013] are limited to the univariate case and the authors propose an approach that involves the speed measure of a diffusion and time-changes of a Brownian motion allowing for target distributions with arbitrary support. The results reported in this manuscript define the general representation of the approximated infinitesimal generators for both multidimensional generalized diffusions and for what we will define as the function copula specification.

The chapter is organized as follows. The general theoretical construct for the processes we will be working with are along the lines of Chapter 2, and in the section 4.2 we introduce and characterize the approximation schemes for the infinitesimal generator of correlated Markov processes while in section 4.3 we show some desirable convergence results of our approximations. In section 4.4 we apply our results to calculate the approximate solution of a multi-dimensional SDE.

4.2 Approximation of correlated Markov processes

In this section we illustrate two new approximation schemes based on tensor algebra for correlated Markov processes where dynamics are expressed by the SDE in eq. (4.1). The novelty we introduce in these approximations involves the characterization of the cross space among dimensions using tensor algebra and in the introduction of new concepts like correlated tensor representation, conditional infinitesimal generator and copula operators. The proposed schemes we develop are based on tensor algebra which we will demonstrate makes them highly amenable to address problems in high dimensional state spaces for correlated processes. As an overview, the approximations we develop involve two aspects:

1. Direct approximation of the infinitesimal generator A with particular emphasis on its mixed derivatives terms;

2. Decomposition of the infinitesimal generator A into orthogonal components.

In what follows we first illustrate the SDE approximation for $d = 1$, in order to establish some useful notation and the building blocks of the approximation schemes. We then present the details of approximations of correlated processes when $d \geq 2$.

4.2.1 Univariate diffusion approximation by CMTC

In order to mimic a one dimensional process X_t with a birth-death process (BDP) $X_t^{(n)}$, we construct a state space $\mathcal{X} := \{x_1, \dots, x_n\} \in \mathbb{R}, \quad k = 1, \dots, d$, which is a countably finite set. We denote with $\mathcal{X}^o := \mathcal{X} \backslash \partial\mathcal{X}$ and $h = \frac{x_n - x_1}{n}$ a positive constant that represents the homogeneous state space discretization unit, and where the boundary $\partial\mathcal{X}$ consist of the smallest (i.e. x_1) and largest (i.e. x_n) elements in $\mathcal{X}$,and the interior $\mathcal{X}^o$ is the complement of the boundary. We denote by $\pi_n : \mathbb{R} \to \mathcal{X}$ the bounded transformation from the continuous state space to the discretized one.

It is possible to construct a BDP and its continuous time Markov chain $X_t^{(n)}$ as the mimicking process of the diffusion X_t on $\mathcal{X}$ by building a generator matrix $A^{(n)} := (a_{ij}) = (a(x_i, x_j))$ for all $i, j = 1, \dots, n$, that is the discretized approximation of the diffusion's generator A in eq. (4.2) and each entry can be calculated by solving the following system of local moment matching equations:

$$\begin{cases} a(x_1, x_2) = a(x_n, x_{n-1}) = 0, \\ a(x_i, x_{i+1}) = \frac{1}{2}\left(\frac{b(x_i)}{h} + \frac{\sigma^2(x_i)}{h^2}\right), \\ a(x_i, x_{i-1}) = \frac{1}{2}\left(\frac{\sigma^2(x_i)}{h^2} - \frac{b(x_i)}{h}\right), \\ a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})), \end{cases} \tag{4.4}$$

for all $i = 1, \dots, n-1$, with $-\frac{\sigma^2(x_i)}{h} \leq b(x_i) \leq \frac{\sigma^2(x_i)}{h}$. However, the discrete state space $\mathcal{X}$ does not need to be uniform and alternative discretization routines are presented in Tavella & Randall [2000].

Remark 2. *In the following we will denote by $A^{(n)} = A_X^{(n)} := A_X^{(n)}(b(x_i), \sigma(x_i))$, for all i, the approximated infinitesimal generator for the Markov process X_t with local parameters $b(\cdot)$ and $\sigma(\cdot)$. In particular $A_{X^k}^{(n)}$ is the approximated infinitesimal generator for the Markov process X_t^k. Furthermore, given a process X_t^k we will denote by $\mathcal{X}^k$ the set corresponding to its discrete state space.*

Assumption 1. *Note that for $x \in \partial\mathcal{X}$, for computational aspects, we impose an absorbing boundary condition. However, it is important to choose the boundary states sufficiently in the extreme of the state space that the laws of the processes $X_t^{(n)}$ and X_t are close to each other during the finite time interval of interest in the approximation.*

The resulting matrix $A^{(n)}$ is a tridiagonal matrix in $\mathbb{R}^{n \times n}$, with always positive extra-diagonal elements. The previous system calculates the entries of this generator by specifying the first and second instantaneous moments of the process $X_t^{(n)}$ that have to coincide with those of X_t on the set $\mathcal{X}^o$. This is equivalent to satisfying the following conditions:

$$\mathbb{E}_{X_t}[(X_{t+\Delta t} - X_t)^z] = \mathbb{E}_{X_t}[(X_{t+\Delta t}^{(n)} - X_t^{(n)})^z] + o(\Delta t), \qquad z \in \{1, 2\} \text{ and } X_t^{(n)} \in \mathcal{X}^o. \tag{4.5}$$

Furthermore, one could in principle produce more accurate results by matching higher instantaneous moments of the process and in general there will be a trade off between the number of local moments matched and the coarsity of the grid/stencil h. We also observe that the numerical problem we face is of the same type as in eq. (2.16) and its analytic solution is $U_t = e^{tA}\nu$ which represents the transition probability of a Markov chain with n states.

4.2.2 Multivariate diffusion approximation by CMTC

We are now in a position to introduce the approximation schemes for multivariate generalized correlated diffusions $\mathbf{X}_t$ when $d \geq 2$. Let $A_{X^k}^{(n_k)} \in \mathbb{R}^{n_k \times n_k}$ denote the approximated infinitesimal generator for the continuous Markov process X_t^k, with $k = 1, 2, \ldots, d$. This notation is useful when describing d correlated processes and the unique approximated generator for the multidimensional process. Each matrix $A_{X^k}^{(n_k)}$ is tridiagonal and its entries calculated using instantaneous local moment matching as described in eq. (4.5).

Under the local moment matching formulation the representation of the infinitesimal generator for correlated Markov processes given in eq. (2.37) can be rewritten as follows:

$$A = \sum_{i}^{d} b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{\substack{i,j\\ i=j}}^{d} \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}}^{d} \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{4.6}$$

Now denote this operator in two components,

$$A_X^{\perp} := A_{X^1,\ldots,X^d} = \sum_{i}^{d} b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{\substack{i,j\\ i=j}}^{d} \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j}, \tag{4.7}$$

and

$$A_X^{(c)} := A_{X^1,\ldots,X^d}^{(c)} = \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}}^{d} \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j}, \tag{4.8}$$

then we can rewrite A in eq. (4.6) as the sum of two linear operators

$$A = A_X^{\perp} + A_X^{(c)} = A_{X^1,\ldots,X^d} + A_{X^1,\ldots,X^d}^{(c)}. \tag{4.9}$$

In particular $A_{X^1,\ldots,X^d}$ is the continuous operator for the independent Markov processes $\mathbf{X}_t = (X_t^1, \ldots, X_t^d)$, while $A_{X^1,\ldots,X^d}^{(c)}$ is the continuous operator just for the dependence structure of such processes. We denote by $\mathcal{X} := \bigotimes_{i=1}^{d} \mathcal{X}^i$ the state space for the mimicking process $X_t^{(n)}$, where $\mathcal{X}^k := \{x_0^{(k)}, \ldots, x_{n_k}^{(k)}\} \in \mathbb{R}, \quad k = 1, \ldots, d$, with n_k the number of discretization points for the k-dimension, and $\mathcal{X}^k$ represents the state space for the Markov chain $X_t^{(n_k)}$ approximating X_t^k. Next we develop two ways to approximate eq. (4.6) on a discrete multidimensional space $\mathcal{X} := \bigotimes_{i=1}^{d} \mathcal{X}^i$, namely:

1. With direct approximation of the processes $\mathbf{X}_t = (X_t^1, \ldots, X_t^d)$ and therefore calculating operator matrices $A_{X^1}^{(n_1)}, \ldots, A_{X^d}^{(n_d)}$ within the orthogonal dimensions and the operator $A_{X^1,\ldots,X^d}^{(c)}$ defined on the cross spaces;

2. The operator is approximated only over the orthogonal spaces. This is possible through the introduction of the notion of conditional operator $A^{(n_i)}_{X^i|X^j=a}$ with $a \in \mathcal{X}^j$ and $i \neq j, j = 1, \dots, d$.

Definition 21 (Multidimensional approximated generator (independent processes))**.** *The multidimensional approximated generator for d independent Markov process $\mathbf{X}_t := (X^i_t)$, for $i = 1, \dots, d$ with approximated generators $A^{(n_k)}_{X^k} \in \mathbb{R}^{n_k \times n_k}$, $k = 1, 2, \dots, d$ is the $A^{(n)}_X \in \mathbb{R}^{n \times n}$ matrix with $n = n_1 \cdots n_d$*

$$A^{(n)}_{\mathbf{X}} := A^{(n_1 \cdots n_d)}_{X^1, \dots, X^d} = A^{(n_1)}_{X^1} \oplus \cdots \oplus A^{(n_d)}_{X^d}, \tag{4.10}$$

where $\oplus$ denotes the standard Kronecker sum. The operator matrix $A^{(n)}_{\mathbf{X}}$ approximates $A^{\perp}_{\mathbf{X}}$ over $\mathcal{X}$.

Definition 22. *Let the matrix $I^{(n)} \in \mathbb{R}^{n \times n}$ be the identity matrix of size $n > 0, n \in \mathbb{N}$, $I^{(n)}_l \in \mathbb{R}^{n \times n}$ a matrix of all zeros and with lower diagonal equal to ones, and $I^{(n)}_u \in \mathbb{R}^{n \times n}$ a matrix of all zeros and with upper diagonal equal to ones, namely:*

$$I^{(n)}_l = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \quad I^{(n)}_u = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{4.11}$$

In the following proposition we propose a first way to characterize the infinitesimal generator matrix for the mimicking chain $\mathbf{X}^{(n)}_t$ when there is correlation among the marginal processes. Namely we exploit the linearity of the generator operator, and we construct the joint generator matrix as the sum of a generator which does not include correlation plus a generator characterizing only the correlation part.

Proposition 4 (Joint generator matrix)**.** *Let $\mathbf{X}_t = (X^1_t, X^2_t)$ be two correlated Markov processes and $\mathbf{X}^{(n)}_t = (X^{(n_1)}_t, X^{(n_2)}_t)$ their mimicking BDPs, where $\mathbf{X}^{(n)}_t : \mathcal{X} \to \mathbb{R}^2$ and $\mathcal{X} = \bigotimes_{i=1}^2 \mathcal{X}^i$ with properties as described in Chapter 2 and with associated approximated infinitesimal generators $A^{(n_1)}_{X^1}$ and $A^{(n_2)}_{X^2}$ respectively. It is possible to define a DBP $\mathbf{Z}^{(n)}_t : \mathcal{X} \to \mathbb{R}^2$ with approximated generator the tridiagonal matrix $A^{(n)}_Z$. Then the infinitesimal approximated generator of the correlated processes $\mathbf{X}_t$ with local correlation parameter $\rho_{ij} := \rho(x^{(1)}_i, x^{(2)}_j)$ can be written as*

$$A^{(n)}_{\mathbf{X}} = A^{(n_1)}_{X^1} \oplus A^{(n_2)}_{X^2} + A^{(c)(n)}_{\mathbf{X}} \tag{4.12}$$

with

$$\begin{aligned} A^{(c)(n)}_X &= -\Big(A^{(n_1)}_Z \oplus A^{(n_2)}_Z\Big) - diag\Big(I^{(n_1)} \otimes diag(A^{(n_2)}_Z)\Big) \\ &+ 1_{\{\rho>0\}}\Big(I^{(n_2)}_m \otimes A^{(m)(n_1)}_Z + A^{(p)(n_2)}_Z \otimes I^{(n_1)}_p\Big) \\ &+ 1_{\{\rho<0\}}\Big(I^{(n_2)}_m \oplus A^{(p)(n_1)}_Z + A^{(p)(n_2)}_Z \otimes I^{(n_1)}_p\Big) \end{aligned} \tag{4.13}$$

and where $A_Z^{(n_1)}$ *and* $A_Z^{(n_2)}$ *denote the component generator matrices of* $A_Z^{(n)}$ *that act on* $\mathcal{X}^1$ *and* $\mathcal{X}^2$ *respectively. The superscript* (p) *means that the* $A_Z^{(p)(n_1)}$ *has same upper triangular entries as the matrix* $A_Z^{(n_1)}$*, while* $A_Z^{(m)(n_1)}$ *has same lower triangular entries as* $A_Z^{(n_1)}$*. Furthermore eq. (4.12) can be interpreted and rewritten as*

$$A_{\mathbf{X}}^{(n)} = A_{X^1|Z}^{(n_1)} \oplus A_{X^2|Z}^{(n_2)} + A_Z^{(n)}. \tag{4.14}$$

Proof. We start with the analysis of the mixed derivative terms in eq. (4.8) when $d = 2$. For positive correlation the following approximations hold for finite difference approximations for the partial derivative operators,

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_1 \partial x_2} &= \frac{\left(\frac{\partial f}{\partial x_2}\right)_{i+1,j} - \left(\frac{\partial f}{\partial x_2}\right)_{i,j}}{h_1} + O(h_1) \\
&= \frac{f_{i+1,j+1} - f_{i+1,j} - f_{i,j+1} + f_{i,j}}{h_1 h_2} + O(h_1) + O(h_2), \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} &= \frac{\left(\frac{\partial f}{\partial x_2}\right)_{i,j-1} - \left(\frac{\partial f}{\partial x_2}\right)_{i-1,j-1}}{h_1} + O(h_1) \\
&= \frac{f_{i,j} - f_{i,j-1} - f_{i-1,j} + f_{i-1,j-1}}{h_1 h_2} + O(h_1) + O(h_2).
\end{aligned} \tag{4.15}$$

The above equations (4.15) can be combined together to yield,

$$\begin{aligned}
\frac{\partial f_{ij}}{\partial x_1 \partial x_2} = &\frac{f_{i+1,j+1} - f_{i+1,j} - \left(f_{i,j-1} - 2f_{i,j} + f_{i,j+1}\right) - f_{i-1,j} + f_{i-1,j-1}}{2h_1 h_2} \\
&+ O(h_1^2) + O(h_2^2) + O(h_1 h_2).
\end{aligned} \tag{4.16}$$

Here we denoted by h_1 and h_2 the discretization unit for the state space $\mathcal{X}^1$ and $\mathcal{X}^2$ respectively. Note that the same discretization scheme applies for negative correlation. We observe that eq. (4.16) can be decomposed into the following three terms

$$\frac{\partial f_{ij}}{\partial x_1 \partial x_2} = T3 - (T1 + T2),$$

where

$$\begin{aligned}
T1 &= \frac{f_{i+1,j} - 2f_{i,j} + f_{i-1,j}}{2h_1 h_2}, \\
T2 &= \frac{f_{i,j+1} - 2f_{i,j} + f_{i,j-1}}{2h_1 h_2}, \\
T3 &= \frac{f_{i+1,j+1} - 2f_{i,j} + f_{i-1,j-1}}{2h_1 h_2}.
\end{aligned} \tag{4.17}$$

Next we observe that the operator $A_{\mathbf{X}}^{(c),(n)}$ is a 'correlation' operator acting on the joint discretized product space $\mathcal{X} = \bigotimes_{k=1}^2 \mathcal{X}^k$. The term T1 acts only along $\mathcal{X}^1$, the term T2 acts only along $\mathcal{X}^2$, while T3 acts only along the cross-space of $\mathcal{X}$. We then use these finite difference operators to calculate the entries of the operators in eq. (4.13). In fact we can give to the finite difference operators the interpretation of instantaneous BDP rates. In particular, we use the scheme T1 for $A_Z^{(n_1)}$, which entries are

$$A_Z^{(n_1)} = \begin{cases} a(x_1,x_2) = a(x_n,x_{n-1}) = 0, \\ a(x_i,x_{i+1}) = \frac{1}{2}\frac{\sigma_1\sigma_2-\rho_{12}^2(x_i,y_j)}{h_1h_2}, \\ a(x_i,x_{i-1}) = \frac{1}{2}\frac{\sigma_1\sigma_2-\rho_{12}^2(x_i,y_j)}{h_1h_2}, \\ a(x_i,x_i) = -(a(x_i,x_{i-1})+a(x_i,x_{i+1})), \end{cases} \tag{4.18}$$

for $x_i \in \mathcal{X}^1$. We use the scheme T2 for $A_Z^{(n_2)}$ which entries are

$$A_Z^{(n_2)} = \begin{cases} a(x_1,x_2) = a(x_n,x_{n-1}) = 0, \\ a(x_i,x_{i+1}) = \frac{1}{2}\frac{\sigma_1\sigma_2-\rho_{12}^2(x_i,y_j)}{h_1h_2}, \\ a(x_i,x_{i-1}) = \frac{1}{2}\frac{\sigma_1\sigma_2-\rho_{12}^2(x_i,y_j)}{h_1h_2}, \\ a(x_i,x_i) = -(a(x_i,x_{i-1})+a(x_i,x_{i+1})), \end{cases} \tag{4.19}$$

and T3 for $1_{\{\rho>0\}}\left(I_l^{(n_2)} \otimes A_Z^{(n_1),u} + A_Z^{(n_2),l} \otimes I_u^{(n_1)}\right)$, where $A_Z^{(n_1),u}$ is a upper diagonal matrix with entries equal to the upper diagonal of the operator matrix $A_Z^{(n_1)}$ and $A_Z^{(n_2),l}$ is a lower diagonal matrix with entries equal to the upper lower of $A_Z^{(n_1)}$. The magnitude of the local instantaneous intensities is $\rho(x_i,x_j)\sigma(x_i),\sigma(x_j)$. We can therefore rewrite eq. (4.12) as

$$\begin{aligned} A_{\mathbf{X}}^{(n)} &= (A_{X_1}^{(n_1)} - A_Z^{(n_1)}) \oplus (A_{X_2}^{(n_2)} - A_Z^{(n_2)}) && (4.20)\\ &- \Big[\mathrm{diag}(I^{(n_1)} \otimes diag(A_Z^{(n_2)})) && (4.21)\\ &+ 1_{\{\rho>0\}}\left(I_l^{(n_2)} \otimes A_Z^{(n_1),l} + A_Z^{(n_2),u} \otimes I_u^{(n_1)}\right) && (4.22)\\ &+ 1_{\{\rho<0\}}\left(I_l^{(n_2)} \oplus A_Z^{(n_1),u} + A_Z^{(n_2),u} \otimes I_u^{(n_1)}\right)\Big]. && (4.23) \end{aligned}$$

Note that when correlation is zero $A_{\mathbf{X}}^{(n)} = A_{X_1}^{(n_1)} \oplus (A_{X_2}^{(n_2)}$, therefore all the terms related to the action of the BDP vanish. In particular we notice in eq. (4.20) how the generator matrices $A_{X_1}^{(n_1)}$ and $A_{X_2}^{(n_2)}$ are compensated in the orthogonal dimensions by the action of Z, which acts on the cross-space of $\mathcal{X}$. If we define the following operators,

$$\begin{aligned} A_{X_1|Z}^{(n_1)} &= (A_{X_1}^{(n_1)} - A_Z^{(n_1)}), \\ A_{X_2|Z}^{(n_2)} &= (A_{X_2}^{(n_2)} - A_Z^{(n_2)}), \\ A_Z^{(n)} &= -\mathrm{diag}(I^{(n_1)} \otimes diag(A_Z^{(n_2)})) + 1_{\{\rho>0\}}\left(I_l^{(n_2)} \otimes A_Z^{(n_1),l} + A_Z^{(n_2),u} \otimes I_p^{(n_1)}\right) \\ &\quad +1_{\{\rho<0\}}\left(I_l^{(n_2)} \oplus A_Z^{(n_1),u} + A_Z^{(n_2),u} \otimes I_u^{(n_1)}\right). \end{aligned} \tag{4.24}$$

This proves eq. (4.14). □

Proposition 5 (Conditional Infinitesimal Generator)**.** *Let $\mathbf{X}_t$ a 2-dimensional diffusion with infinitesimal generator*

$$A = \sum_{i=1}^{2} b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{2} \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} \tag{4.25}$$

as per definition 4.2. We assume that the diffusions $\mathbf{X}_t = \{X_t^1, X_t^2\}_{t\geq 0}$ *are locally correlated with instantaneous local covariation given by*

$$\left\langle \sigma_1(x)dW_t^1, \sigma_2(y)dW_t^2 \right\rangle = \sigma_i(x)\sigma_j(x)\rho_{ij}(x,y)dt, \qquad \text{for all } (x,y) \in \mathbb{R}^2. \tag{4.26}$$

Let $\mathbf{X}_t^{(n)} : \mathcal{X} \to \mathbb{R}$ *the BDP mimicking* $\mathbf{X}_t$*, with* $\mathcal{X} = \bigotimes_{i=1}^2 \mathcal{X}^i$*. The conditional approximated infinitesimal generator* $A_{X^1|X^2}^{(n_1)}$ *is defined by the sequence of operator matrices* $\left\{A_{X^1|X^2=x_i^{(2)}}^{(n_1)}\right\} \in \mathbb{R}^{n_1\times n_1}$, $x_i^{(2)} \in \mathcal{X}^2$ *each of whose entries are obtained according to local moment matching by:*

$$A_{X^1|X^2=x_i^{(2)}}^{(n_1)} = \begin{cases} a(x_1^{(1)}, x_2^{(1)}) = a(x_{n_1}^{(1)}, x_{n_1-1}^{(1)}) = 0, \\ a(x_i^{(1)}, x_{i+1}^{(1)}) = \frac{1}{2}\Big(\frac{b_1(x_i^{(1)})+\rho_{12}(x_i^{(1)},x_i^{(2)})\frac{\sigma_1(x_i^{(1)})}{\sigma_2(x_i^{(2)})}((x_{i+1}^{(2)}-x_i^{(2)})/dt-b_2(x_i^{(2)}))}{h} + \frac{\sigma_1^2(x_i^{(1)})(1-\rho_{12}^2(x_i^{(1)},x_i^{(2)}))}{h^2}\Big), \\ a(x_i^{(1)}, x_{i-1}^{(1)}) = \frac{1}{2}\Big(\frac{\sigma_1^2(x_i^{(1)})(1-\rho_{12}^2(x_i^{(1)},x_i^{(2)}))}{h^2} - \frac{b_1(x_i^{(1)})+\rho_{12}(x_i^{(1)},x_i^{(2)})\frac{\sigma_1(x_i^{(1)})}{\sigma_2(x_i^{(2)})}((x_{i+1}^{(2)}-x_i^{(2)})/dt-b_2(x_i^{(2)}))}{h}\Big), \\ a(x_i^{(1)}, x_i^{(1)}) = -(a(x_i^{(1)}, x_{i-1}^{(1)}) + a(x_i^{(1)}, x_{i+1}^{(1)})), \end{cases} \tag{4.27}$$

for all $x_i^{(k)} \in \mathcal{X}^k$*, with* $-\frac{\sigma_k^2(x_i^{(k)})}{h} \leq b_k(x_i^{(k)}) \leq \frac{\sigma_k^2(x_i^{(k)})}{h}$ *for* $k = 1, 2$, $\rho_{12}(x_i^{(1)}, x_i^{(2)}) \in [-1, 1]$*, and* $a(x_i^{(k)}, x_{i+1}^{(i)}) \geq 0$ *and* $a(x_i^{(k)}, x_{i-1}^{(i)}) \geq 0$ *for all* i*.*

Proof. Let's consider the two-dimensional Markov process $\mathbf{X}_t = (X_t^1, X_t^2)$ and how to derive its corresponding infinitesimal generator approximation in matrix form. The local instantaneous intensities are calculated in the same way as described in Section 4.2 for the one dimensional case, namely using local moment matching as reported in eq. (4.5). For the 2-dimensional case the instantaneous intensities are calculated using the local transition kernel

$$p(X_{t+\Delta t}^1, X_{t+\Delta t}^2 | X_t^1, X_t^2), \quad \text{as } \Delta t \to 0. \tag{4.28}$$

In particular considering the transition from a local state $(x_i^{(1)}, x_j^{(2)}) \in \bigotimes_{k=1}^2 \mathcal{X}^k$ to the state $(x_{i+1}^{(1)}, x_{j+1}^{(2)})$ the transition probability in eq. (4.28) can be rewritten in local form as

$$p_t(x_{i+1}^{(1)}, x_{j+1}^{(2)} | x_i^{(1)}, x_j^{(2)}) = p_t(x_{i+1}^{(1)} | x_{j+1}^{(2)}, x_j^{(2)}, x_i^{(1)}) p_t(x_{j+1}^{(2)} | x_j^{(1)}), \tag{4.29}$$

where we denote by

$$p_t(x_{j+1}^{(2)} | x_j^{(2)}) = p\Big(X_{t+\Delta t}^2 = x_{j+1}^{(2)} | X_t^2 = x_j^{(2)}\Big) \sim \mathcal{N}\Big(b(y_j), \sigma^2(y_j)\Big) \tag{4.30}$$

$$p_t(x_{i+1}^{(1)} | x_{j+1}^{(2)}, x_j^{(2)}, x_i^{(1)}) = p\Big(X_{t+\Delta t}^{(1)} = x_{i+1}^{(1)} | X_{t+\Delta t}^{(2)} = x_{j+1}^{(2)}, X_t^{(2)} = x_j^{(2)}, X_t^1 = x_i^{(1)}\Big) \sim$$
$$= \mathcal{N}\Big(b_1(x_i^{(1)}) + \frac{\sigma_1(x_i^{(1)})}{\sigma_2(x_j^{(2)})}\rho(x_i^{(1)}, x_j^{(2)})((x_{j+1}^{(2)} - x_j^{(2)})/\Delta t - b_2(x_j^{(2)})), (1 - \rho^2(x_i^{(1)}, x_j^{(2)})\sigma_1^2(x_j^{(1)}))\Big)$$

Therefore the generator of $(X^1|X^2)$ which we denote as $A_{X^1|X^2}$ can be approximated as a sequence of conditional generators of operator matrices $\left\{A_{X^1|X^2=x_i^{(2)}}^{(n_1)}\right\} \in \mathbb{R}^{n_1\times n_1}$, $x_i^{(2)} \in$

$\mathcal{X}^2$ each of whose entries are obtained by matching the first two local conditional moments

$$\begin{cases} \frac{d}{dt}\mathbb{E}[dX_t^{(n_1)}|X_t^{(n_1)} = x_i^{(1)}, X_t^{(n_2)} = x_j^{(2)}] = b(x_i) + \frac{\sigma_1(x_i^{(1)})}{\sigma_2(x_j^{(2)})}\rho(x_i^{(1)}, x_j^{(2)})(dx_j^{(2)}/dt - b_2(x_j^{(2)})) \\ \frac{d}{dt}\mathbb{E}[(dX_t^{(n_1)})^2|X_t^{(n_1)} = x_i^{(1)}, X_t^{(n_2)} = x_j^{(2)}] = \sigma^2(x_i)(1 - \rho^2(x_i^{(1)}, x_j^{(2)}). \end{cases} \tag{4.31}$$

for all $x_j^{(2)} \in \mathcal{X}^2$ and in the same way as described in Section 4.2, and this completes our proof. Extension to dimensions larger than two is straightforward due to independence of each conditional operator. □

Note that the instantaneous conditional drift in eq. (4.31) includes a temporal average of the conditioning state with respect to the starting state, see example 3.3.1. This is not an approximation but an exact quantity for the mean of a conditional normal distribution.

4.2.3 Exponential of the Multidimensional Infinitesimal Generator Matrix

In this section we provide details on how to exponentiate a multivariate infinitesimal generator matrix of the type of eq. (4.43),

$$A_X^{(n)} = A_{X^1}^{(n_1)} \oplus_S \{A_{X^2|X^1}^{(n_2)}\} \oplus_S \cdots \oplus_S \{A_{X^d|X^1,\dots,X^{d-1}}^{(n_d)}\}. \tag{4.32}$$

where each term of type $\{A_{X^j|X^k}^{(n_j)}\}$, $j = 1, \dots, d$ and $j \neq k$ represents the conditional infinitesimal generator introduced in Proposition 5 or equivalently a series of infinitesimal generator matrices,

$$\{A_{X^j|X^k}^{(n_j)}\} := \{A_{X^j|X^k=x_0^{(k)}}^{(n_j)}, A_{X^j|X^k=x_1^{(k)}}^{(n_j)}, \dots, A_{X^j|X^k=x_{n_k}^{(k)}}^{(n_j)}\} \tag{4.33}$$

where

$$A_{X^j|X^k=x_z^{(k)}}^{(n_j)} = \begin{cases} a(x_0^{(j)}, x_1^{(j)}) = a(x_{n_1}^{(j)}, x_{n_1-1}^{(j)}) = 0, \\ a(x_i^{(j)}, x_{i+1}^{(j)}) = \frac{1}{2}\Big(\frac{b_j(x_i^{(j)}) + \rho(x_i^{(j)}, x_j^{(k)})\frac{\sigma_1(x_i^{(j)})}{\sigma_k(x_z^{(k)})}(x_z^{(k)} - b_k)}{h_j} + \frac{\sigma_1^2(x_i^{(j)})(1-\rho^2(x_i^{(j)}, x_z^{(k)}))}{h_j^2}\Big), \\ a(x_i^{(j)}, x_{i-1}^{(j)}) = \frac{1}{2}\Big(\frac{\sigma_j^2(x_i^{(j)})(1-\rho^2(x_i^{(j)}, x_z^{(k)}))}{h_j^2} - \frac{b_j(x_i^{(j)}) + \rho(x_i^{(j)}, x_z^{(k)})\frac{\sigma_j(x_i^{(j)})}{\sigma_2(x_z^{(k)})}(x_z^{(k)} - b_k)}{h_j}\Big), \\ a(x_i^{(j)}, x_i^{(j)}) = -(a(x_i^{(j)}, x_{i-1}^{(j)}) + a(x_i^{(j)}, x_{i+1}^{(j)})), \end{cases} \tag{4.34}$$

for $\mathcal{X}^k = \{x_0^{(k)}, \dots, x_{n_k}^{(k)}\}$, $k = 1, \dots, d$. Note that, by Proposition 9, the matrix sequence can be decomposed into

$$\{A_{X_j|X_k}^{(n_j)}\} = A_{X_j}^{(n_j)} - \{\hat{A}_{X_j|X_k}^{(n_j)}\}. \tag{4.35}$$

In the following we formulate a decomposition of the conditional operator introduced in Proposition 5. Let us consider a coupled diffusion process $\mathbf{X}_t = (X_t^1, X_t^2)$ and its mimicking CTMC $X_t^{(n)}$ on $\mathcal{X} = \bigotimes_{i=1}^2 \mathcal{X}^i$. It is almost straightforward to see from eq. (4.27) that the sequence of conditional generator matrices $\{A_{X^1|X^2}^{(n_1)}\}$ can be decomposed into

$$A_{X^2|X^1}^{(n_2)}\} = A_{X^2}^{(n_2)} - \{A_{X^2|X^1}^{(c)(n_2}\} \tag{4.36}$$

We now give more details of the decomposition in eq. (4.36). We saw in eq. (2.58) that the expression of the Kronecker sum of a matrix $A \in \rho^{q\times q}$ with a matrix sequence $\{B_j\}, B_j \in \rho^{p\times p}$ can be written as:

$$A \oplus_S \{B_j\}^T = \left(A \otimes I^{(p)}\right) + \left(I^{(q)} \otimes_S \{B_j\}\right)^T = \left(I^{(q)} \otimes_S \{B_j\}\right) + \left(A \otimes I^{(p)}\right)^T \quad (4.37)$$

The expression of the approximated infinitesimal generator for the mimicking BDP $\mathbf{X}_t^{(n)}$ on the state space $\mathcal{X}^1 \otimes \mathcal{X}^2$ reads

$$\begin{aligned} A^{(n)} &= A_{X^1}^{(n_1)} \oplus_S \{A_{X^2|X^1}^{(n_2)}\} \\ &= \left(A_{X^1}^{(n_1)} \otimes I^{(n_2)}\right) + \left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(n_2)}\}\right) \\ &= \left(A_{X^1}^{(n_1)} \otimes I^{(n_2)}\right) + \left(I^{(n_1)} \otimes \{A_{X^2}^{(n_2)}\}\right) + \left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(c)(n_2)}\}\right) \\ &= \left(A_{X^1}^{(n_1)} \oplus \{A_{X^2}^{(n_2)}\}\right) + \left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(c)(n_2)}\}\right) \\ &= A_X^{(n)\perp} + \left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(c)(n_2)}\}\right) \end{aligned} \quad (4.38)$$

In eq. (4.38) we then express the generator of $X_t^{(n)}$ as the sum of $A_X^{(n)\perp}$, which is the the generator matrix when $X_t^{(n_1)}$ and $X_t^{(n_2)}$ are independent, plus the matrix $\left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(c)(n_2)}\}\right)$ defined on $\mathcal{X}^1 \otimes \mathcal{X}^2$ which holds all the correlation structure of the joint process. In particular the matrix sequence $\{A_{X^2|X^1}^{(c)(n_2)}\}$ is defined on $\mathcal{X}^2$ and represents the instantaneous projection of $X_t^{(n_1)}$ onto $X_t^{(n_2)}$.

We will investigate this decomposition in greater detail in Chapter 5, where we shall see that $\left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(c)(n_2)}\}\right)$ can specify an operator, that we name copula operator, which is an approximated infinitesimal generator matrix sequence calculated as the difference between the CTMC joint operator $A^{(n)}$ and the independent one $A^{(n)\perp}$. In fact with some algebra we have that:

$$\begin{aligned} \{A_{X^2|X^1}^{(c)}\} &= A_{X^1}^{(n_1)} \oplus_S \{A_{X^2|X^1}^{(n_2)}\} - (A_{X^1}^{(n_1)} \oplus A_{X^2}^{(n_2)}) \\ &= \left(A_{X^1}^{(n_1)} \otimes I^{(n_2)}\right) + \left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(n_2)}\}\right) - \left[\left(A_{X^1}^{(n_1)} \otimes I^{(n_2)}\right) + \left(I^{(n_1)} \otimes A_{X^2}^{(n_2)}\right)\right] \\ &= \left(I^{(n_1)} \otimes_S \{A_{X^2|X^1}^{(n_2)}\}\right) - \left(I^{(n_1)} \otimes A_{X^2}^{(n_2)}\right) \end{aligned} \quad (4.39)$$

with generator matrix entries given by:

$$\{A_{X^2|X^1}^{(c)}\} = \begin{cases} a(x_1^{(2)}, x_2^{(2)}) = a(x_{n_2}^{(2)}, x_{n_2-1}^{(2)}) = 0, \\ a(x_i^{(2)}, x_{i+1}^{(2)}) = \frac{1}{2}\Big(\frac{\sigma_1^2(x_i^{(2)})(1-\rho^2(x_i^{(2)},x_i^{(1)}))}{h_2^2} + \frac{\rho(x_i^{(2)},x_i^{(1)}))\frac{\sigma_1(x_i^{(1)})}{\sigma_2(x_i^{(2)})}(x_j-b_1(x_i^{(1)}))}{h_2}\Big), \\ a(x_i^{(2)}, x_{i-1}^{(2)}) = \frac{1}{2}\Big(\frac{\sigma_1^2(x_i^{(2)})(1-\rho^2(x_i^{(2)},x_i^{(1)}))}{h_2^2} - \frac{\rho(x_i^{(2)},x_i^{(1)}))\frac{\sigma_1(x_i^{(1)})}{\sigma_2(x_i^{(2)})}(x_j-b_1(x_i^{(1)}))}{h_2}\Big), \\ a(x_i^{(2)}, x_i^{(2)}) = -(a(x_i^{(2)}, x_{i-1}^{(2)}) + a(x_i^{(2)}, x_{i+1}^{(2)})), \end{cases} \quad (4.40)$$

for all $x_i^{(1)} \in \mathcal{X}^1$. The representation of the copula transition probability in tensor form is:

$$P_{X^2}^{(n)(c)} = \alpha \frac{P_{X^1,X^2}^{(n)}}{P_{X^1}^{(n_1)} \otimes P_{X^2}^{(n_2)}} = \alpha \frac{I^{(n_1)} \otimes_S P_{X^2|X^1}^{(n_2)}}{I^{(n_1)} \otimes P_{X^2}^{(n_2)}} = \alpha \frac{I^{(n_1)} \otimes_S e^{\{A_{X^2|X^1}^{(n_2)}\}}}{I^{(n_1)} \otimes e^{A_{X^2}^{(n_2)}}} \quad (4.41)$$

where α is just a scaling factor. Analogously,

$$P_{X^1}^{(n)(c)} = \alpha' \frac{P_{X^1,X^2}^{(n)}}{P_{X^1}^{(n_1)} \otimes P_{X^2}^{(n_2)}} = \alpha' \frac{I^{(n_2)} \otimes_S P_{X^1|X^2}^{(n_1)}}{I^{(n_2)} \otimes P_{X^1}^{(n_1)}} = \alpha' \frac{I^{(n_2)} \otimes_S e^{\{A_{X^1|X^2}^{(n_1)}\}}}{I^{(n_2)} \otimes e^{A_{X^1}^{(n_1)}}} \tag{4.42}$$

with α' just a scaling factor. Note that $P_{X^2}^{(n)(c)} = P_{X^1}^{(n)(c)}$ for standardized uniform marginals, or equivalently $\mathcal{X}^2 = \mathcal{X}^1$.

Proposition 6 (Multidimensional approximated generator (correlated processes))**.** *The multidimensional approximated generator for d correlated Markov process $\mathbf{X}_t = (X_t^i)$, for $i = 1, \ldots, d$ with approximated generators $A_{X^k}^{(n_k)} \in \mathbb{R}^{n_k \times n_k}$, $k = 1, 2, \ldots, d$ is the $n_1 n_2 \cdots n_d \times n_1 n_2 \cdots n_d$ matrix*

$$A^{(n)} = A_{X^1,\ldots,X^d}^{(n_1 \cdots n_d)} = \left\{ A_{X^1|X^2,\ldots,X^d}^{(n_1)} \right\} \oplus_S \cdots \oplus_S A_{X^d}^{(n_d)}, \tag{4.43}$$

where $\oplus_S$ denotes the Kronecker sum over a matrix sequence.

From proposition 5 it is clear that we can represent the approximated multidimensional generator of eq. (4.10) as a decomposition of independent conditional generators. Furthermore it is also possible to exploit standard results of conditional probability partitioning in order to facilitate the local characterization of the independent multidimensional conditional generators. Given a multivariate Gaussian variable $\mathbf{X} \sim \mathcal{N}(\mathbf{b}, \mathbf{\Sigma})$ and consider the partition of $\mathbf{X}$ and equivalently of $\mathbf{b}$ and $\mathbf{\Sigma}$ into

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_{1,1}} & \mathbf{\Sigma_{1,2}} \\ \mathbf{\Sigma_{2,1}} & \mathbf{\Sigma_{2,2}} \end{bmatrix} \tag{4.44}$$

Then $(\mathbf{x_1}|\mathbf{x_2}) = \mathbf{y} = \mathbf{x_1} + \mathbf{C}\mathbf{x_2}$, where $\mathbf{C} = -\mathbf{\Sigma_{1,2}}\mathbf{\Sigma_{2,2}}^{-1}$, the conditional distribution of the first partition given the second, is $\mathcal{N}(\overline{\mathbf{b}}, \overline{\mathbf{\Sigma}})$, with mean

$$\overline{\mathbf{b}} = \mathbf{b_1} + \mathbf{\Sigma_{1,2}}\mathbf{\Sigma_{2,2}}^{-1}(\mathbf{x_2} - \mathbf{b_2}), \tag{4.45}$$

and covariance matrix

$$\overline{\mathbf{\Sigma}} = \mathbf{\Sigma_{1,1}} - \mathbf{\Sigma_{1,2}}\mathbf{\Sigma_{2,2}}^{-1}\mathbf{\Sigma_{2,1}}. \tag{4.46}$$

More generally we denote by $\Sigma = (\Sigma_{i,j}(x)) \in \mathbb{R}^{d \times d}$ the positive covariance matrix introduced within the definition of the generator A in eq. (2.37) and eq. (4.6),

$$\mathbf{\Sigma} = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,d} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{d,1} & \Sigma_{d,2} & \cdots & \Sigma_{d,d} \end{pmatrix}. \tag{4.47}$$

Conditional probability partitioning is a very important property when creating the sequence of conditional approximated generators of eq. (4.27) because large multivariate Gaussian vectors can be easily partitioned as the combination of sets of independent sub-multivariate Gaussian vectors. Each of the sub-multivariate Gaussian vectors can be further characterized and locally approximated through a principal component analysis (PCA). Therefore it is possible to construct a reduced dimensionality infinitesimal generator of a large dimension process without an aggregate PCA of the global process covariance structure.

Example 1 (Three-dimensional approximated generator)**.** *Given the results in Propositions 5 and 6, we show how to calculate* $A^{(n)}_{\mathbf{X}} = A^{(n_1 n_2 n_3)}_{X^1 X^2 X^3}$. *We can express the approximated three-dimensional generator under a conditional decomposition according to*

$$A^{(n)}_{\mathbf{X}} = A^{(n_1)}_{X^1|X^2X^3} \oplus_S A^{(n_2)}_{X^2|X^3} \oplus_S A^{(n_3)}_{X^3}. \tag{4.48}$$

Note that $A^{(n_1)}_{X^1|X^2X^3} = \left\{A^{(n_1)}_{X^1|X^2=x^{(2)},X^3=x^{(3)}}\right\} \in \mathbb{R}^{n_1 \times n_1}$ *with* $x^{(2)} \in \mathcal{X}^2$ *and* $x^{(3)} \in \mathcal{X}^3$ *and* $A^{(n_2)}_{X^2|X^3} = \left\{A^{(n_2)}_{X^2|X^3=x^{(3)}}\right\} \in \mathbb{R}^{n_2 \times n_2}$ *with* $x^{(3)} \in \mathcal{X}^3$.

Remark 3. *It is important to note that the joint infinitesimal generator approximation in the Proposition 5 is not equivalent to the joint representation of Proposition 4. The approximations do not produce the same instantaneous local correlations in the cross space, but they produce equivalent 'terminal' joint distribution. We would like to introduce the explicit expressions of the quadratic variation and quadratic covariation among CTMC with generators approximated using the method introduced in Proposition 5. This is formalized in the following Lemmas.*

Lemma 1. *[Orthogonal CTMCs and zero instantaneous covariation] Let* $\overline{\mathbf{X}} = (X_1^{(n_1)}, \ldots, X_d^{(n_d)})'$ *be a d-dimensional CTMC and define its partition as* $\overline{\mathbf{X}} = [\mathbf{x_1}, \mathbf{x_2}]'$. *Let's define* $\mathbf{y} = \mathbf{x_1} + \mathbf{C}\mathbf{x_2}$, *where* $\mathbf{C} = -\mathbf{\Sigma_{1,2}}\mathbf{\Sigma_{2,2}}^{-1}$. *Then the chains* $\mathbf{x_2}$ *and the conditional chains* $\mathbf{y}$ *are orthogonal and var*$(\mathbf{y}) =$ *var*$(\mathbf{x_1}|\mathbf{x_2})$.

Proof. We want to show that the instantaneous covariation of the chains $\mathbf{x_2}$ and $\mathbf{y}$ is zero. Therefore we compute

$$\begin{aligned} \text{cov}(\mathbf{x_2}, \mathbf{y}) &= \text{cov}(\mathbf{x_2}, \mathbf{x_1}) + \text{cov}(\mathbf{x_2}, \mathbf{C}\mathbf{x_2}) \\ &= \mathbf{\Sigma_{1,2}} + \mathbf{C}\text{cov}(\mathbf{x_2}, \mathbf{x_2}) \\ &= \mathbf{\Sigma_{1,2}} - \mathbf{\Sigma_{1,2}}\mathbf{\Sigma_{2,2}}^{-1}\mathbf{\Sigma_{2,2}} = 0 \end{aligned} \tag{4.49}$$

Furthermore we have

$$\begin{aligned} \text{var}(\mathbf{y}) = &= \text{var}(\mathbf{x_1} + \mathbf{C}\mathbf{x_2}) \\ &= \text{var}(\mathbf{x_1}) + \mathbf{C}\text{var}(\mathbf{x_2})\mathbf{C}' + \mathbf{C}\text{cov}(\mathbf{x_1}, \mathbf{x_2}) + \text{cov}(\mathbf{x_2}, \mathbf{x_1})\mathbf{C}' \\ &= \text{var}(\mathbf{y}|\mathbf{x_2}) = \text{var}(\mathbf{x_1}|\mathbf{x_2}). \end{aligned} \tag{4.50}$$

$$\begin{aligned} \text{var}(\mathbf{x_1}|\mathbf{x_2}) &= \text{var}(\mathbf{x_1} + \mathbf{C}\mathbf{x_2}) \\ &= \text{var}(\mathbf{x_1}) + \mathbf{C}\text{var}(\mathbf{x_2})\mathbf{C}' + \mathbf{C}\text{cov}(\mathbf{x_1}, \mathbf{x_2}) + \text{cov}(\mathbf{x_2}, \mathbf{x_1})\mathbf{C}' \\ &= \text{var}(\mathbf{y}|\mathbf{x_2}) = \text{var}(\mathbf{y}). \end{aligned} \tag{4.51}$$

□

Lemma 2 ('Terminal' covariance of orthogonal CTMCs)**.** *Given the settings and results from proposition 5 and Lemma 1 we can obtain the covariance matrix* $\mathbf{\Sigma}$ *which is:*

$$[\mathbf{y}, \mathbf{x_2}] \begin{bmatrix} \mathbf{\Sigma} & 0 \\ 0 & \mathbf{\Sigma_{2,2}} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x_2} \end{bmatrix} = [\mathbf{x_1}, \mathbf{x_2}] \begin{bmatrix} \mathbf{\Sigma_{1,1}} & \mathbf{\Sigma_{1,2}} \\ \mathbf{\Sigma_{2,1}} & \mathbf{\Sigma_{2,2}} \end{bmatrix} \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix}. \tag{4.52}$$

Theorem 5 (Equivalence of the joint 'terminal' representations). *Given two correlated Markov processes $\mathbf{X}_t = (X_t^1, X_t^2)$, with properties as described in Chapter 2, the approximated 'terminal' joint transition probability kernel $P_t^{(n)} := P_{\mathbf{X}_t}^{(n)} = P_{X^1,X^2}^{(n_1 n_2)}$ is the solution of the Cauchy problem of eq. (2.16) where the infinitesimal operator can be expressed in either cross space and marginals decomposition or as conditional decomposition given by*

$$\bar{A}_{\mathbf{X}}^{(n)} = A_{X_1}^{(n_1)} \oplus A_{X_2}^{(n_2)} + A_{X_1,X_2}^{(c)(n)}, \tag{4.53}$$

or

$$A_{\mathbf{X}}^{(n)} = A_{X^1|X^2}^{(n_1)} \oplus_S A_{X^2}^{(n_2)}. \tag{4.54}$$

The approximations in eq. (4.53) and eq. (4.54) produce equivalent 'terminal' joint transition probabilities, although while the generator matrix $\bar{A}_{\mathbf{X}}^{(n)}$ spans also the cross space of $\mathcal{X}$, $A_{\mathbf{X}}^{(n)}$ is confined only on its orthogonal dimensions.

Proof. Refer to Lemma 1 and Lemma 2. □

Proposition 7 (Correlated kernels tensor representation). *Given d-correlated Markov processes $\mathbf{X}_t$, with properties as described in Section 4.1, the approximated weak of the SDE in eq. (4.1), is given by the product measure representation involving the generator decomposition according to,*

$$P_t^{(n)} = e^{tA^{(n)}} \approx e^{t(A_{X^1,\dots,X^d}^{(n_1\cdots n_d)} = A_{X^1|X^2,\dots,X^d}^{(n_1)} \oplus_S \cdots \oplus_S A_{X^d}^{(n_d)})} \tag{4.55}$$

$$= e^{tA_{X^1|X^2,\dots,X^d}^{(n_1)}} \otimes_S e^{tA_{X^2|X^3,\dots,X^d}^{(n_2)}} \otimes_S \cdots \otimes_S e^{tA_{X^d}^{(n_d)}} \tag{4.56}$$

Proof. In order to calculate the d-dimensional tensor product of eq. (4.55) we exploit the orthogonality of the conditional approximated infinitesimal operators $A_{X^i|\dots,X^d}^{(n_i)}$ for all i. This follows from Proposition 5. Then the transition density of eq. (4.55) is computed along the same lines as in example 1. □

Remark 4. *We would like to remark that in eq. (4.55) we compute the approximate solution $P_t^{(n)}$ of a correlated diffusion $\mathbf{X}_t$, which spans only the orthogonal dimension of the space $\mathcal{X} = \bigotimes_{k=1}^{d} \mathcal{X}^k$, and not the cross-spaces of $\mathcal{X}$, which is usually the space where the approximation of correlation takes place. This has the remarkable numerical advantage that a n-dimensional problem can be dealt as independent one-dimensional problems. The solution $P_t^{(n)}$ of eq. (4.55) is subject to an approximation error which has two sources: the first one is due to the discretization of each support $\mathcal{X}^k$, while the other error contribution occurs because the generator matrices do not commute. We elaborate about this last error in the next section 4.2.4, while we discuss about weak convergence of the solution $P_t^{(n)}$ in section 4.3.*

The use of orthogonal tensor spaces to overcome the curse of dimensionality is a very well known practice, and used many different type of applications from statistics to big data, see Kuang *et al.* [2014].

4.2.4 Commutators in Exponential Operators

In this section we want to give more insights of the matrix exponential of the Kronecker sum of a sequence of matrices, namely

$$\exp\Big(A \oplus_S \{B_j\}\Big) \approx \exp(A) \otimes_S \exp(\{B_j\}) + O([A, \{B_j\}]), \tag{4.57}$$

where the notation $\{B_j\}$, $j = 1, 2, \ldots$ denotes a sequence of matrices $\{B_1, B_2, \ldots\}$. We also give details of exponential of matrix operators, a topic very well studied in mathematical physics and extensively used in quantum physics, see Wilcox [1967].

Remark 5. *We would like to remark that the matrices $\Big(A \otimes I^{(p)}\Big)$ and $\Big(I^{(n)} \otimes_S \{B_j\}\Big)$ in eq. (2.58) do not commute. However the key result in our proposed multidimensional diffusion approximation is based on theorem 4, namely*

$$\exp\Big(A \oplus_S \{B_j\}\Big) \approx \exp(A) \otimes_S \exp(\{B_j\}) + O([A, \{B_j\}]) \tag{4.58}$$

While the left side of eq. (4.58) represents the expression of the diffusion approximation that will be used in the study of weak convergence in section 4.3.4, the right hand side is instead a further algebraic approximation that is key to a fast computation of the proposed numerical scheme illustrated in section 4.3.4. We have that, while the left side of eq. (4.58) can be interpreted as a multidimensional approximation of the transition probability at time 1, the right hand side represents the same multidimensional approximation made by one dimensional approximations. The error term of order $O([A, \{B_j\}])$, is due to the fact that the matrices A and the matrix sequence $\{B_j\}$ do not commute.

In eq. (4.57) we utilise the definition that if A is an $\mathbb{R}^{m\times m}$ matrix and $\{B_j\}, j = 1, \ldots, m$ is a sequence of m matrices in $\mathbb{R}^{n\times n}$, then the Kronecker product of a matrix sequence $A \otimes_S \{B_j\}$ is the $nm \times nm$ block matrix:

$$A \otimes_S \{B_j\} = \begin{bmatrix} a_{11}B_1 & \cdots & a_{1m}B_m \\ \vdots & \ddots & \vdots \\ a_{m1}B_1 & \cdots & a_{mm}B_m \end{bmatrix} = (A \otimes I_n)(I_m \otimes_S \{B_j\}), \tag{4.59}$$

where $I_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix, and

$$I_m \otimes_S \{B_j\} = \begin{bmatrix} B_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_m \end{bmatrix}. \tag{4.60}$$

Therefore

$$A \otimes_S \{B_j\} = \begin{bmatrix} a_{11}I_n & \cdots & a_{1m}I_n \\ \vdots & \ddots & \vdots \\ a_{m1}I_n & \cdots & a_{mm}I_n \end{bmatrix} \begin{bmatrix} B_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_m \end{bmatrix}. \tag{4.61}$$

We would like to give some more details about the approximation in eq. (4.57). According to the Baker-Campbell-Hausdorff (BCH) formula:

$$Z(X,Y) = \log(\exp(X)\exp(Y)) = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}([X,[X,Y]] + [Y,[Y,X]]) + \ldots \tag{4.62}$$

where Z, X and Y are linear operators associated to semigroups, and $[X,Y] := XY - YX$. Note that when the groups X and Y commute then $[X,Y] = 0$. In our specific case the application of the above BCH formula up to the first commutator yields

$$\exp\big((A\otimes I_n)+(I_m\otimes_S\{B_j\})\big) = \exp\big((A\otimes I_n)\big)\exp\big((I_m\otimes_S\{B_j\})\big)\exp\big(-\frac{1}{2}[(A\otimes I_n),(I_m\otimes_S\{B_j\})]\big). \tag{4.63}$$

In particular we would like to closely understand the expression of the commutator

$$[(A\otimes I_n),(I_m\otimes_S\{B_j\})]. \tag{4.64}$$

We note that the matrix operator sequence $\{B_j\}$ can be decomposed into a constant matrix B and a state dependent matrix sequence $\{\hat{B}_j\}$, and therefore we can write,

$$I_m\otimes_S\{B_j\} = \begin{bmatrix} B_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_m \end{bmatrix} \tag{4.65}$$

$$= \begin{bmatrix} B & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B \end{bmatrix} - \begin{bmatrix} \hat{B}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{B}_m \end{bmatrix} \tag{4.66}$$

$$= I_m\otimes_S(\{B\}-\{\hat{B}_j\}) = I_m\otimes B - I_m\otimes_S\{\hat{B}_j\}. \tag{4.67}$$

Note that this decomposition is equivalent to the decomposition in eq. (4.35). Then eq. (4.64) simplifies to

$$[(A\otimes I_n),(I_m\otimes_S\{B_j\})] = [(A\otimes I_n),(I_m\otimes_S\{\hat{B}_j\})] \quad \text{[because } [(A\otimes I_n),(I_m\otimes B)] = 0\text{]} \tag{4.68}$$

Therefore the commutator $C = [(A\otimes I_n),(I_m\otimes_S\{\hat{B}_j\})]$ reads

$$C = \begin{bmatrix} a_{11}I_n & \cdots & a_{1m}I_n \\ \vdots & \ddots & \vdots \\ a_{m1}I_n & \cdots & a_{mm}I_n \end{bmatrix}\begin{bmatrix} \hat{B}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{B}_m \end{bmatrix} - \begin{bmatrix} \hat{B}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{B}_m \end{bmatrix}\begin{bmatrix} a_{11}I_n & \cdots & a_{1m}I_n \\ \vdots & \ddots & \vdots \\ a_{m1}I_n & \cdots & a_{mm}I_n \end{bmatrix} \tag{4.69}$$

$$= \begin{bmatrix} 0 & a_{12}(\hat{B}_2-\hat{B}_1) & 0 & \cdots & 0 \\ a_{21}(\hat{B}_1-\hat{B}_2) & 0 & a_{23}(\hat{B}_3-\hat{B}_2) & \cdots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & a_{m-1m-2}(\hat{B}_{m-2}-\hat{B}_{m-1}) & 0 & a_{m-1m}(\hat{B}_m-\hat{B}_{m-1}) \\ 0 & \cdots & 0 & a_{mm-1}(\hat{B}_{m-1}-\hat{B}_m) & 0 \end{bmatrix}.$$

Therefore eq. (4.63) can be rewritten as

$$\exp\big((A\otimes I_n)+(I_m\otimes_S\{B_j\})\big) = \exp\big((A\otimes I_n)\big)\exp\big((I_m\otimes_S\{B_j\})\big)\exp\big(-\frac{1}{2}C\big). \tag{4.70}$$

Note that in all the applications presented in this thesis the terms $(\hat{B}_i - \hat{B}_{i-1}) \approx 0$ for all i.

4.3 Convergence of the approximated CTMC

In this section we study the convergence of the proposed CTMC approximation towards diffusion processes. The main references used for this investigation are Ethier & Kurtz [2009]; Stroock & Varadhan [2007]; Billingsley [2013]; Baldi [2017]; Brzeniak & Zastawniak [1999].

Specifically if the mimicking CTMC $\mathbf{X}_t^{(n)}$ has measure $P^{(n)}$ mimicking a diffusion process $\mathbf{X_t}$ with infinitesimal operator

$$A = \sum_i^d b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j}^d \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j}$$

and measure P.

Definition 23 (Weak convergence of probability measures)**.** *A sequence of of finite measures $\{P^{(n)}\}$, on the measurable space $(S, \mathcal{B}(S))$, converges to the measure P weakly if for every contiuous bounded function $f : S \to \mathbb{R}$*

$$\lim_{n\to\infty} \int f dP^{(n)} = \int f dP. \tag{4.71}$$

In Section 4.2 we defined the Markov chain $\mathbf{X}_t^{(n)}$ approximating the multidimensional generalized diffusion $\mathbf{X}_t$. Here we present weak convergence results for $\mathbf{X}_t^{(n)}$ to the solution of the SDE for $\mathbf{X}_t$, introduced in eq. (4.1). The convergence is studied through different methods:

1. From a semigroup point of view,
2. In a spectral way through a Fourier unitary transformation of the approximated generator providing desired rate of convergence results,
3. Through the martingale problem for the associated infinitesimal generator.

4.3.1 Semigroup approach

We want that the continuous Markov chain $\mathbf{X}_t^{(n)}$ has a dynamics as close as possible to the corresponding approximated process $\mathbf{X}_t$. To this purpose we can define an error

$$\epsilon_n(f) := \sup_{x\in S} \|A^{(n)} f(x) - Af(x)\|. \tag{4.72}$$

Using the semigroup approach to weak convergence we can state that, if $\epsilon_n(f)$ tends to zero as n tends to infinity for f in $D(S)$, then the sequences of processes $\{\mathbf{X}_t^{(n)}\}$, converges weakly to $\mathbf{X}_t$, in the space $D(S)$. The following theorem states that if the error $\epsilon_n(f)$ goes to zero as n tends to infinity, implying norm convergence of the approximated generator $A^{(n)}f(x)$ to $Af(x)$, this would imply also convergence of the corresponding approximated semigroup $P_t^{(n)}$ to P_t and of the chain $X^{(n)}$ to X for $t \geq 0$.

Definition 24. *(Feller property, Feller Process, see Baldi [2017]) A transition function p is said to enjoy the Feller property if, for every fixed $h \geq 0$ and for every bounded continuous function $f : S \to \mathbb{R}$ the map*

$$(t, x) \mapsto \int_S f(y)p(t, t+h, x, dy) \tag{4.73}$$

is continuous , or in the homogeneous case

$$x \mapsto \int_S f(y)p(h, x, dy) \tag{4.74}$$

is a continuous function of x. A Markov process X_t is said to be a Feller process if its transition function enjoys Feller's property.

The Feller property introduced above is relevant to study the weak convergence of the transition probability. In fact it is equivalent of stating that if the sequences $t_n \to t$ and $x_n \to x$ then

$$p(t_n, x_n, \cdot) \underset{n\to\infty}{\to} p(t, x, \cdot) \tag{4.75}$$

Examples of Feller processes include solutions to SDEs with Lipschitz continuous coefficients, the Brownian motion and the Poisson process. Therefore also a birth-death process is a Feller process.

Theorem 6. *Let X_t be a Feller process on S with infinitesimal generator A, and $X_t^{(n)} : \mathcal{X} \to \mathbb{R}$ be a mimicking BDP with generator matrix $A^{(n)}$, and with $\pi^{(n)} : S \to \mathcal{X}$ a bounded linear transformation. Then it holds that*

$$\lim_{n\to\infty} \epsilon_n(f) = 0 \tag{4.76}$$

for every function $f \in D(S)$, the set of $f \in M_b(S)$ for which A exists, which is equivalent to the statement that

$$T_t^{(n)}\pi_n f \to T_t f, \qquad \textit{for all } f \in D(S),\ t \geq 0.$$

Then $X_t^{(n)}$, converges weakly to X_t, in $D(S)$.

Proof. From Theorem 6.1 in Chapter 1 pag. 28 of Ethier & Kurtz [2009], we have that if $\{T_t^{(n)}\},\ n = 1, 2, \ldots$, and $\{T_t\}$ are strongly continuous contraction semigroups on S with generators $A^{(n)}$ and A respectively, and let $D(S)$ be a core for A, then the following are equivalent:

(i) For each $f \in S$

$$P_t^{(n)}\pi_n f \to P_t f$$

uniformly on bounded intervals.

(ii) For each $f \in S$

$$P_t^{(n)}\pi^{(n)} f \to P_t f$$

for all $t \geq 0$.

(iii) For each $f \in D(S)$, there exists $f^{(n)} \in \mathcal{D}(A^{(n)})$[1] for each $n \geq 1$ such that $f^{(n)} \to f$ and $A^{(n)} f^{(n)} \to Af$.

Furthermore, following Ethier & Kurtz [2009], Chapter 4 pag. 172 Theorem 2.11, we have that if $\{T_t\}$ is a Feller semigroup on $M_b(S)$ and that for each $t \geq 0$ and $M_b(S)$

$$P_t^{(n)} \pi^{(n)} f \to P_t f.$$

If $X_0^{(n)}$ has limiting distribution ν, then there is a Markov Process X_t corresponding to P_t with initial distribution ν and sample paths in S, and

$$X_t^{(n)} \to X_t \quad \text{in } S.$$

□

Therefore, in order to be able to use the result of the above Theorem 6, we need to prove the convergence of the proposed approximated generator $A^{(n)}$ to the generator A and this is done in the following theorem, using the argument that the operator A is local.

Proposition 8. *(Local Infinitesimal Generator) If $B_r(x)$ denotes a sphere of radius $r > 0$ centered in $x \in D$ and if*

$$\lim_{t \to 0^+} (p(t, x, B_r^c(x))) = 0, \tag{4.77}$$

where $B_r^c(x)$ denotes the complement set of the sphere $B_r(x)$, then A is local.

Theorem 7. *For all $f \in D(S) \subset M_b(S)$, with $D(S)$ domain of the generator A as previously defined,*

$$\lim_{n \to \infty} \sup_{x \in S} |A^{(n)} f(x) - Af(x)| = 0,$$

with a rate at most of $O(h^2)$, where h denotes the discretization unit of $\mathcal{X}$.

Proof. Let $f \in D$ and $x, y \in S$, such that $y = x + \Delta x \in B_r(x)$. We recall that the action of the generator A on a function $f(x) \in D$ is equal to

$$\lim_{t \to 0^+} \frac{1}{t} \Big(T_t f(x) - f(x) \Big) = \lim_{t \to 0^+} \frac{1}{t} \int (f(y) - f(x)) p(t, x, dy) \tag{4.78}$$

and equivalently in a neighbourhood of the point x to

$$\lim_{t \to 0^+} \frac{1}{t} \int_{B_r(x)} (f(y) - f(x)) p(t, x, dy). \tag{4.79}$$

We now compute the Taylor expansion of the function $f(y)$ about x which is

$$f(y) = f(x) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x)(y_i - x_i) + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x_i x_j}(x)(y_i - x_i)(y_j - x_j) + o(|x - y|^2). \tag{4.80}$$

Therefore we have that

$$\lim_{t \to 0^+} \frac{1}{t} \Big(T_t f(x) - f(x) \Big) = Af(x) + \lim_{t \to 0^+} \frac{1}{t} \int_{B_r(x)} o(|x - y|^2) p(t, x, dy). \tag{4.81}$$

[1] Where $\mathcal{D}(A^{(n)})$ denotes the domain of $A^{(n)}$.

We now characterize the limit $\lim_{t\to 0^+} \frac{1}{t}\int_{B_r(x)} o(|x-y|^2)p(t,x,dy)$ using the notation developed in the thesis for the approximated generator $A^{(n)}$. By Taylor approximation

$$f_n(x) = f(y) \approx f(x+h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + o(h^2)$$

we obtain for all $x \in \mathcal{X}$,

$$\begin{aligned}
A^{(n)}f_n(x) &= a(x,x+h)\Big(f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + o(h^2) - f(x)\Big) \\
&+ a(x,x-h)\Big(f(x) - f'(x)h + \frac{1}{2}f''(x)h^2 + o(h^2) - f(x)\Big) \\
&= f'(x)h\Big(a(x,x+h) - a(x,x-h)\Big) \\
&+ \frac{1}{2}f''(x)h^2\Big(a(x,x+h) - a(x,x-h)\Big) \\
&+ \Big(a(x,x+h) - a(x,x-h)\Big)o(h^2).
\end{aligned}$$

Due to the fact that $f \in D$, the error term $o(h^2)$ is uniform in x. We have that

$$\begin{aligned}
a(x_i,x_{i+1}) + a(x_i,x_{i-1}) &= \frac{\sigma^2(x)}{h^2}, \\
a(x_i,x_{i+1}) - a(x_i,x_{i-1}) &= \frac{\mu(x)}{h}.
\end{aligned}$$

In section 4.2 we made precise assumptions, see the assumption 1 on the behaviour at the boundary of the process. Without loss of generality we can assume that the boundary is at a point of infinity, i.e. not attainable in finite time and it is furthermore absorbing. It is possible to express $A^{(n)}f_n(x)$ for all $x \in D$, as

$$\begin{aligned}
A^{(n)}f_n(x) &= f'(x)h\Big(\frac{b(x)}{h}\Big) + \frac{1}{2}f''(x)h^2\Big(\frac{\sigma^2(x)}{h^2}\Big) + \Big(\frac{\sigma^2(x)}{h^2}\Big)o(h^2) \\
&= Af(x) + \Big(\frac{\sigma^2(x)}{h^2}\Big)o(h^2).
\end{aligned}$$

We obtain

$$\sup_{x\in D}|A^{(n)}f(x) - Af(x)| = \sup_{x\in D}|\frac{\sigma^2(x)}{h^2}o(h^2)| \le \frac{\sigma^2(x)}{h^2}KO(h^3) \xrightarrow{n\to\infty} 0. \tag{4.82}$$

where K is a constant. In case of $x \in \partial\mathcal{X}$, consisting of the smallest (i.e. $\underline{x}$) and largest (i.e. $\overline{x}$) elements in $\mathcal{X}$, are absorbing states of the Markov chain we can continue the above analysis with a further investigation of the weak convergence. Furthermore the behaviour on the boundary of the functions $f \in D$ is $f'(\underline{x}) = f'(\overline{x}) = 0$. For $x = \underline{x}$ we have,

$$\begin{aligned}
|A^{(n)}f(\underline{x}) - Af(\underline{x})| &\le |A^{(n)}f(\underline{x}) - A^{(n)}f(\underline{x}+h)| \\
&+ |A^{(n)}f(\underline{x}+h) - Af(\underline{x}+h)| \\
&+ |Af(\underline{x}+h) - Af(\underline{x})|
\end{aligned}$$

The second term tends to 0 as shown above, the third term by continuity of Af in $\underline{x}$. For the first term we have

$$\begin{aligned}|A^{(n)}f(\underline{x}) - A^{(n)}f(\underline{x}+h)| &= |a(\underline{x},\underline{x}+h)(f(\underline{x}+h)-f(\underline{x})) \\ &- a(\underline{x}+h,\underline{x})(f(\underline{x}+2h)-f(\underline{x}+h)) \\ &- a(\underline{x}+h,\underline{x}+2h)(f(\underline{x}+h)-f(\underline{x}))| \\ &= O(h)o(h) \xrightarrow{n\to\infty} 0.\end{aligned}$$

since $a(\underline{x},\underline{x}+h)$, $a(\underline{x}+h,\underline{x})$, $a(\underline{x}+h,\underline{x}+2h)$ are on order $o(h)$ and $(f(\underline{x}+h)-f(\underline{x}))$, $(f(\underline{x}+2h)-f(\underline{x}+h))$, $(f(\underline{x}+h)-f(\underline{x}))$ are of order $O(h)$ because $f'(\underline{x})=0$ for $f \in D$. The result for $\overline{x}$ follows in the same way. □

4.3.2 Weak Convergence through Spectral approach

For this purpose we need to introduce multiplication operators that are considered as an infinite-dimensional generalization of diagonal matrices and they are extremely simple to construct. Furthermore, they appear naturally in the context of the Fourier transform or when one applies the spectral theorem and deals with spectral representation of operators on Hilbert spaces. It is an equivalent way to represent the same operator and can be useful for calculations and further analysis.

We recall that the definitions of discrete Fourier transform matrix and its inverse as a unitary operator are given by,

$$\begin{aligned}F_{s,k} &= \frac{1}{\sqrt{n}} e^{-\mathrm{i}\frac{2\pi}{n}sk}, \\ F^{-1}_{k,s} &= \frac{1}{\sqrt{n}} e^{\mathrm{i}\frac{2\pi}{n}sk}, \\ \sum_k F_{l,k}F^{-1}_{k,j} &= \delta_{l,j},\end{aligned}$$

where $\delta_{l,j}$ is the Kronecker delta, so these matrices give the resolution of the identity matrix and define a unitary transformation. Also, if $f(x)$ is a function belonging to the space L^2 and B_n the Brillouin zone defined as $B_n = \left\{ -\frac{\pi}{h} + kh, \quad k = 0, \dots, \frac{2\pi}{h^2} = \frac{n}{h} \right\}$ the transformation $\mathcal{F}_n : L^2(\mathcal{X}) \mapsto L^2(B_n)$

$$\mathcal{F}_n(f)(s) = \sum_{x\in\mathcal{X}} F_{s,k} f(x), \tag{4.83}$$

is the discrete Fourier transform of f, with $\mathcal{X} := h\mathbb{Z}$ [2] . In fact,

$$\begin{aligned}\mathcal{F}(f)(s) &= \frac{1}{\sqrt{n}} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-\mathrm{i}sx} f(x)dx & (4.84)\\ &\approx \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} e^{\mathrm{i}s\left(-\frac{\pi}{h}+kh\right)} \underbrace{f(-\frac{\pi}{h}+kh)}_{f_k}, & (4.85)\end{aligned}$$

[2]In all our practical applications we consider $\mathcal{X} \subset D := [-K, K] \subset \mathbb{R}$, $K \in (0, \infty)$, with D the operator domain and assuming for simplicity also periodic boundary conditions.

where $h = \frac{2\pi}{n}$. We now extend the above transformation to the d-dimensional case, and the following results set the notation for our subsequent theorems and proofs.

Theorem 8. *(See Jacod & Protter [2004], Theorem 16.1, pag. 126.) $\mathbf{X}$ is an $\mathbb{R}^d$-valued Gaussian random variable if and only if its characteristic function has the form*

$$\varphi_{\mathbf{X}}(\mathbf{s}) = \exp\left(\mathbf{i}\mathbf{b}\mathbf{s} - \frac{1}{2}\mathbf{s}'\mathbf{\Sigma}\mathbf{s}\right) \tag{4.86}$$

where $\mathbf{b} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ is a symmetric semipositive definite matrix. Σ is then the covariance of $\mathbf{X}$ and $\mathbf{b}$ is the mean of $\mathbf{X}$, that is $b_k = \mathbb{E}[X^k]$ for all k.

Theorem 9. *(See Lukacs [1958], Theorem 3.2.3) Let $f(s)$ be an arbitrary characteristic function. For every real x the limit*

$$p(x) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-\mathrm{i}sx} f(s) ds \tag{4.87}$$

exists and is equal to the saltus of the distribution function of $f(s)$ at the point x.

We can obtain a spectral representation of the operator $A_{\mathbf{X}}^{(n)}$ by applying the above unitary transformation leading to the following diagonal operator,

$$q_{n_1\cdots n_d}(\mathbf{s}) = \mathcal{F}(A_{\mathbf{X}}^{(n)}(f))\mathcal{F}^{-1}(\mathbf{s}) \tag{4.88}$$

where the approximated generator $A^{(n)}$ can be written as

$$A^{(n)} = \mu'\nabla + \frac{1}{2}\sigma'\tilde{\mathbf{H}}\sigma \tag{4.89}$$

with ∇ the discrete d-dimensional gradient operator and $\tilde{\mathbf{H}}$ the discrete d-dimensional Hessian operator.

In order to derive some convergence properties of the operator (4.10) we do this by comparing the spectral representations of the probability density functions for the continuous infinitesimal generator and its approximated counterpart, namely

$$p_t(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_{\mathbf{X}} e^{\mathbf{i}\mathbf{s}(\mathbf{y}-\mathbf{x})} d\mathbf{s}.$$

and

$$P_t^{(n)}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^d} \int_{[-\frac{\pi}{h},\frac{\pi}{h}]^d} e^{q_{n_1\cdots n_d}(\mathbf{s})t} e^{\mathbf{i}\mathbf{s}(\mathbf{y}-\mathbf{x})} d\mathbf{s},$$

In this way we are able to assess the order of convergence of the error

$$\epsilon_n := \left| p_t(\mathbf{x}, \mathbf{y}) - P_t^{(n)}(\mathbf{x}, \mathbf{y}) \right|. \tag{4.90}$$

To assess the rate of convergence of eq. (4.90), we exploit the relationship between the distribution function and its corresponding characteristic function and in particular we refer to the Continuity Theorem.

Theorem 10 (Continuity Theorem, see Lukacs [1958], Theorem 3.6.1.)**.** *Let $\{F_n(x)\}$ be a sequence of distribution functions and denote by $\{f_n(s)\}$ the sequence of the corresponding characteristic functions. The sequence $\{F_n(x)\}$ converges weakly to a distribution function $F(x)$ if, and only if, the sequence $\{f_n(s)\}$ converges for every s to a function $f(s)$ which is continuous at $s = 0$. The limiting function is then the characteristic function of $F(x)$.*

We can therefore focus on the analysis of the passage to the limit $\mathbf{h} \to 0$ of the following spectral representation, conditional on a time t,

$$\lim_{\mathbf{h}\downarrow 0} \frac{1}{(2\pi)^d} \int_{[-\frac{\pi}{h},\frac{\pi}{h}]^d} e^{q_{n_1\cdots n_d}(\mathbf{s})t} e^{\mathrm{i}\mathbf{s}(\mathbf{y}-\mathbf{x})} d\mathbf{s} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_{\mathbf{X}} e^{\mathrm{i}\mathbf{s}(\mathbf{y}-\mathbf{x})} d\mathbf{s}. \tag{4.91}$$

Theorem 11. *(Convergence of the d-dimensional approximated operator) For all $\mathbf{x}$ we consider the sequence of density functions $P_t^{(n_1\cdots n_d)}(\mathbf{x},\mathbf{y})$ and the sequence of the corresponding characteristic functions $\{e^{q_{n_1\cdots n_d}(\mathbf{s})t}\}$. The sequence $\{F_n(\mathbf{x})\}$ converges weakly to a distribution function $F(\mathbf{x})$ if, and only if, the sequence $\{e^{q_{n_1\cdots n_d}(\mathbf{s})t}\}$ converges for every $\mathbf{s}$ to a function $\varphi_{\mathbf{X}}$ which is continuous at $\mathbf{s} = 0$. The limiting function is then the characteristic function of $F(\mathbf{x})$. Furthermore if h denotes the uniform discretization unit of the space $\mathcal{X}$, we found that the attainable rate of weak convergence is at most $O(h^2)$.*

Proof. We prove convergence and characterization of the rate of convergence for $d = 1$ and $d = 2$ being the proof in higher dimensions just an algebraic extension of the case of $d = 2$. The calculation of $q_{n_1\cdots n_d}(\mathbf{s})t$ is straightforward and it is just an application of the shift theorem. For $d = 1$

$$\mathcal{F}_n(A_{X_1}^{(n_1)})(f)(s) = \sum_{x\in\mathbf{H}} F_{s,k}\Big(\mu\nabla_{h_1}(f)(x) + \frac{\sigma^2}{2}\triangle_{h_1}(f)(x)\Big)$$

$$\begin{aligned}
\mathcal{F}_n(\mu\nabla_{h_1})(f)(s) &= \sum_{x\in\mathbf{H}} F_{s,k}\mu\nabla_{h_1}(f)(x) \\
&= \mu\sum_{k=0}^{n-1} F_{s,k}\frac{f(kh+h_1) - f(kh-h_1)}{2h_1} \\
&= \frac{\mu}{2h_1}\Big(e^{\mathrm{i}h_1 s}\sum_{k=0}^{n-1} F_{s,k}f(kh) - e^{-\mathrm{i}h_1 s}\sum_{k=0}^{n-1} F_{s,k}f(kh)\Big) \\
&= \mu\frac{e^{\mathrm{i}h_1 s} - e^{-\mathrm{i}h_1 s}}{2h_1}\mathcal{F}(f)(s) = -\mathrm{i}\mu\frac{\sin h_1 s}{h_1}\mathcal{F}(f)(s).
\end{aligned}$$

Doing a similar calculation for $\mathcal{F}(\frac{\sigma^2}{2}\triangle_{h_1})(f)(s)$ we obtain

$$q_{n_1}(s) = \Big(-\mathrm{i}\mu\frac{\sin h_1 s}{h_1} + \sigma^2\frac{\cos(h_1 s) - 1}{h_1^2}\Big)(s). \tag{4.92}$$

Let us consider the integral on the left-hand side of eq. (4.91) for $d = 1$ and it can be rewritten as

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\frac{\pi}{h_1}}^{\frac{\pi}{h_1}} e^{\left(-\mathrm{i}\mu\frac{\sin h_1 s}{h_1}+\sigma^2\frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{\mathrm{i}s(y-x)}ds \\
&= \frac{1}{2\pi}\int_{-\frac{\pi}{h_1}}^{-K} e^{\left(-\mathrm{i}\mu\frac{\sin h_1 s}{h_1}+\sigma^2\frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{\mathrm{i}s(y-x)}ds \\
&+ \frac{1}{2\pi}\int_{-K}^{K} e^{\left(-\mathrm{i}\mu\frac{\sin h_1 s}{h_1}+\sigma^2\frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{\mathrm{i}s(y-x)}ds \\
&+ \frac{1}{2\pi}\int_{K}^{\frac{\pi}{h_1}} e^{\left(-\mathrm{i}\mu\frac{\sin h_1 s}{h_1}+\sigma^2\frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{\mathrm{i}s(y-x)}ds
\end{aligned}$$

and it is possible to make the first and the third integral on the right end side arbitrary small by choosing $K = K(h_1)$ and by selecting $h_1 > 0$ sufficiently small, such that $K(h_1) \to \infty$ when $h_1 \to 0$. If we consider the second integral we can analyse the behaviour as $h_1 \to 0$. We notice that for all s the function

$$\begin{aligned}
\lim_{h_1\to 0}\left(\mathrm{i}\mu\frac{\sin h_1 s}{h_1}\right) &= \lim_{h_1\to 0}\mathrm{i}\mu\ \frac{s}{s}\frac{\sin h_1 s}{h_1} \\
&= \lim_{h_1\to 0}\mathrm{i}s\mu\frac{\sin h_1 s}{h_1 s} = \mathrm{i}s\mu
\end{aligned}$$

and

$$\begin{aligned}
\lim_{h_1\to 0}\left(\sigma^2\frac{\cos h_1 s-1}{h_1^2}\right) &= \lim_{h_1\to 0}\sigma^2\frac{s^2}{s^2}\frac{\cos h_1 s-1}{h_1^2} \\
&= \lim_{h_1\to 0}\sigma^2 s^2\frac{\cos h_1 s-1}{(h_1 s)^2} = -\frac{1}{2}\sigma^2 s^2.
\end{aligned}$$

We would like to examine in more details the order of convergence of the above functions as $h_1 \to 0$. For the limit

$$\lim_{h_1\to 0}\frac{\sin h_1}{h_1} = 1,$$

using $\sin h_1 = h_1 - \frac{h_1^3}{6} + \ldots$, we get

$$\frac{\sin h_1}{h_1} - 1 = \frac{\sin h_1 - h_1}{h_1} = -\frac{h_1^3}{6h} + \ldots = -\frac{{h_1}^2}{6} + \ldots$$

we find an order of $O(h_1^2)$. In the same way using $\cos h_1 = 1 - \frac{h_1^2}{2} + \ldots$, we can assess the order of convergence of the limit

$$\lim_{h_1\to 0}\frac{1-\cos h_1}{h_1^2} = \frac{1}{2},$$

$$\frac{1-\cos h}{h_1^2} = \frac{1-\left(1-\frac{h_1^2}{2}+\ldots\right)}{h_1^2} = \frac{1}{2} + \ldots$$

and convergence order of $O(1)$. Therefore the order of convergence is at most $O(h_1^2)$. This results can be extended to all the marginals of a d-dimensional approximated operator in case of independent marginals. In the presence of correlation we have the presence of mixed derivative terms. For $d = 2$ the calculation of $q_{n_1 n_2}$ is as follows:

$$\begin{aligned}
&\mathcal{F}_{\mathbf{n}}\Big(A^{(n_1 n_2)}_{X^1,X^2}\Big)(f)(\mathbf{s}) \\
&= \mathcal{F}_{\mathbf{n}}\Big(A^{(n_1)}_{X^1} \oplus A^{(n_2)}_{X^2} + A^{(c)(n_1 n_2)}_{X^1,X^2}\Big)(f)(\mathbf{s}) \\
&= \sum_{x_1\in\mathcal{X}^1}\sum_{x_2\in\mathcal{X}^2} F_{s,k}\Big(\big(\mu_1\nabla_{h_1} + \frac{\sigma_2^2}{2}\triangle_{h_1} + \mu_2\nabla_{h_2} + \frac{\sigma_2^2}{2}\triangle_{h_2} + \rho\sigma_1\sigma_2\nabla_{h_1}\nabla_{h_2}\big)(f)(\mathbf{x})\Big).
\end{aligned} \tag{4.93}$$

The above eq. (4.93) is equivalent to

$$\begin{aligned}
&\mathcal{F}_{\mathbf{n}}\Big(A^{(n_1)}_{X^1} \oplus A^{(n_2)}_{X^2} + A^{(c)(n_1 n_2)}_{X^1,X^2}\Big)(f)(\mathbf{s}) \\
&= \mathcal{F}_{n}\Big(A^{(n_1)}_{X^1}\Big)(f)(\mathbf{s}) \oplus \mathcal{F}_{n}\Big(A^{(n_2)}_{X^2}\Big)(f)(\mathbf{s}) + \mathcal{F}_{\mathbf{n}}\Big(A^{(c)(n_1 n_2)}_{X^1,X^2}\Big)(f)(\mathbf{s}).
\end{aligned}$$

Therefore it is sufficient to analyze the term

$$\mathcal{F}_{\mathbf{n}}\Big(A^{(c)(n_1 n_2)}_{X^1,X^2}\Big)(f)(\mathbf{s}), \tag{4.94}$$

where the operator $A^{(c)(n_1 n_2)}_{X^1,X^2}$ was introduced in proposition 4 as the correlation operator.

$$\begin{aligned}
&\mathcal{F}_{\mathbf{n}}\Big(A^{(c)(n_1 n_2)}_{X^1,X^2}\Big)(f)(\mathbf{s}) \\
&= \sum_{x_1\in\mathcal{X}^1}\sum_{x_2\in\mathcal{X}^2} F_{s,k}\Big(\rho\sigma_1\sigma_2\nabla_{h_1}\nabla_{h_2}\big)(f)(\mathbf{x})\Big) \\
&= \rho_{12}\sigma_1\sigma_2\frac{1}{h_1 h_2}\Big(\cos(h_1 s_1 + h_2 s_2) - \cos(h_1 s_1) - \cos(h_2 s_2) + 1\Big),
\end{aligned}$$

because of the mixed derivative term approximation of Proposition 4. The following approximations hold:

$$\begin{aligned}
\cos(h_1 s_1 + h_2 s_2) - 1 &= -\frac{h_2^2 s_2^2}{2} - s_1 s_2 h_1 h_2 - \frac{h_1^2 s_1^2}{2} + \frac{h_1^2 s_1^2 h_2^2 s_2^2}{4} \\
&\quad + (1 + h_1 + h_1^2)O(h_2)^3 + O(h)^3, \\
\cos(h_1 s_1) - 1 &= -\frac{s_1^2 h_1^2}{2} + O(h_1)^3, \\
\cos(h_2 s_2) - 1 &= -\frac{s_2^2 h_2^2}{2} + O(h_2)^3.
\end{aligned}$$

Therefore,

$$\lim_{h_1,h_2\to 0}\rho_{12}\sigma_1\sigma_2\frac{1}{h_1 h_2}\Big(\cos(h_1 s_1 + h_2 s_2) - \cos(h_1 s_1) - \cos(h_2 s_2) + 1\Big) = -\rho_{12}\sigma_1\sigma_2 s_1 s_2, \tag{4.95}$$

that is exactly the characteristic function of the covariance term for the bivariate normal distribution at time t. Given this proof, extension to higher dimensions is algebraically straightforward. □

4.3.3 Weak Convergence Using the Martingale Central Limit Theorem

Here we present a weak convergence result for the multivariate diffusion approximation introduced in Section 4.2, along the lines of the general diffusion convergence Theorem 4.1 at pag. 354 of Ethier & Kurtz [2009]. The convergence result we propose is based on the arguments belonging to the formulation of diffusion theory in terms of the martingale problem, see Stroock & Varadhan [2007], which requires minimal assumptions about the smoothness of the coefficients of the SDE and can be seen as an extension of the martingale central limit theorem [Martingale CLT, Theorem 1.4, Ethier & Kurtz [2009], pag. 339]. The results are mainly obtained by compactness arguments which do not require a priori regularity. These arguments are the same as those devised to provide an existence theory for the multidimensional SDE under consideration, and they just refer to properties of the SDE coefficients as per Chapter 2.

Definition 25. *The function space $D([0,T],\mathbb{R}^d) = D_{\mathbb{R}^d}[0,T]$ if the set of all $\mathbb{R}^d$-valued functions $f = (f^1,\dots,f^d)$ on $[0,T]$ that are right continuous at all $t \in [0,T)$ and have left limits at all $t \in (0,T]$. Therefore if a function f belongs to the set $D_{\mathbb{R}^d}[0,T]$, then*

$$\begin{aligned} &\lim_{s\to t^+} f(s) = f(t^+) \qquad \textit{exists and } \; f(t^+) = f(t), \; \textit{for } t \in [0,T) \\ &\lim_{s\to t^-} f(s) = f(t^-) \qquad \textit{exists, for } t \in (0,T]. \end{aligned} \tag{4.96}$$

In particular for each function $f \in D_{\mathbb{R}^d}[0,T]$ the number of discontinuities of f is finite.

Theorem 12. *Let us consider the infinitesimal generator of a diffusion $\mathbf{X}_t$ which has the expression*

$$A = \sum_i^d b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j}^d \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} \tag{4.97}$$

where $\mathbf{b} := (b_i(s,x)) : \mathbb{R}^d \to \mathbb{R}^d$, $i = 1,\dots,d$ is a continuous drift vector function and $\mathbf{\Sigma} = (\Sigma_{ij}), i,j = 1,\dots,d$ be a continuous, symmetric, positive definite, $d\times d$ matrix valued function with $\Sigma = \Psi\Psi' \in \mathbb{R}^d$ and we assume that the martingale problem for (A,ν) is well posed, with ν the initial distribution. For $n = 1,2,\dots$ let $X_t^{(n)}$ and $B_t^{(n)}$ be processes with sample paths in $D_{\mathbb{R}^d}[0,\infty)$ and let $\Sigma_t^{(n)} = (\Sigma_{ij}^{(n)}(t))$ be a symmetric $d\times d$ matrix-valued process such that $\Sigma_{ij}^{(n)}(t)$ has sample paths in $D_R[0,\infty)$ and $(\Sigma_t^{(n)} - \Sigma_s^{(n)})$ is positive for $t > s \geq 0$. Set the filtration of these processes equal to $\mathcal{F}_t^n = \sigma\Big(X_s^{(n)}, B_s^{(n)}, \Sigma_s^{(n)} : s \leq t\Big)$. Let $\tau_n^r = \inf\{t : |X_t^{(n)}| \geq r \;\; \text{or} \;\; |X_{t-}^{(n)}| \geq r\}$, and suppose that

$$M_t^{(n)} = X_t^{(n)} - B_t^{(n)} \tag{4.98}$$

and

$$M_i^{(n)}(t)M_j^{(n)}(t) - \Sigma_{ij}^{(n)}(t) \tag{4.99}$$

are $\mathcal{F}_t^n$-local martingales, and for each $r > 0$, $T > 0$, and $i,j = 1,2,\dots,d$

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{t\leq T\wedge\tau_n^r}\Big|X_t^{(n)} - X_{t-}^{(n)}\Big|^2\Big] = 0, \tag{4.100}$$

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{t\leq T\wedge\tau_n^r}\Big|B_t^{(n)} - B_{t-}^{(n)}\Big|^2\Big] = 0, \tag{4.101}$$

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{t\leq T\wedge\tau_n^r}\Big|\Sigma_{ij}^{(n)}(t) - \Sigma_{ij}^{(n)}(t-)\Big|^2\Big] = 0, \tag{4.102}$$

$$\sup_{t\leq T\wedge\theta_n^r}\Big|B_i^{(n)}(t) - \int_0^t b_i(X_s^{(n)})ds\Big| \xrightarrow{p} 0 \tag{4.103}$$

and

$$\sup_{t\leq T\wedge\theta_n^r}\Big|\Sigma_{ij}^{(n)}(t) - \int_0^t \Sigma_{ij}(X_s^{(n)})ds\Big| \xrightarrow{p} 0. \tag{4.104}$$

Suppose that $P(X_0^{(n)}) \Rightarrow \nu$, with ν the starting distribution. Then $X_t^{(n)} \Rightarrow X_t$, with X_t the solution of the martingale problem for (A, ν).

Proof. We aim to prove the validity of the above theorem when the approximating process $X_t^{(n)}$ is a BDP with rates satisfying local moment matching conditions as per eq. (4.31) or more generally as eq. (4.27). Due to the fact that the martingale problem for (A, ν) is well posed, the process $M_t = X_t - B_t$ is a martingale, with $B_t = \int_0^t b_s ds$. By the optional stopping theorem, see Rogers & Williams [2000], if τ is a stopping time also the stopped martingale $M(\tau)$ is a martingale. In particular, if $r > 0$ is the radius of a sphere, this is valid also for the stopping time

$$\tau_n^r = \inf\{t : |X_t^{(n)}| \geq r \ \text{ or } \ |X_{t-}^{(n)}| \geq r\}.$$

The choice of this stopping time is not casual, as we have $\tau_n^r \to \infty$ as $r \to \infty$. This extends also to the stopped martingale $M_n(t \wedge \tau_n^r \wedge \tau_a)$ where

$$\tau_a = \inf\{t : \Sigma_{ii}^{(n)}(t) - t\sup_{|x|\leq r} a_{ii}(x) > 1, \text{for some i}\},$$

which existence derives from eq. (4.103) and implies that

$$P(\tau_a \leq T \wedge \tau_n^r) \to 0, \tag{4.105}$$

for all $r > 0$ and for all $T > 0$. Eq. (4.105) is a direct consequence of the Doob martingale inequality theorem (see Varadhan [2007]) which also imply that the martingale

$$\Sigma_{ij}^{(n)}(t) - \int_0^t \Sigma_{ij}(X_s^{(n)})ds \tag{4.106}$$

has zero quadratic variation, and therefore eq. (4.103) holds. Furthermore relative compactness in a set C is a condition equivalent to the condition that each sequence in C contains a convergent subsequence, see for example Billingsley [2013]. By eq. (4.98) the process $M_t^{(n)} = X_t^{(n)} - B_t^{(n)}$ is a martingale. Relative compactness properties of the process $M_t^{(n)}$ implies relative compactness of $X_t^{(n)}$ and $B_t^{(n)}$ and therefore for $X^{(n)}(t\wedge\tau_n)$ and $B^{(n)}(t\wedge\tau_n)$ as in eq. (4.100) and eq. (4.101) respectively. In particular eq. (4.100) is the definition of a Poisson jump magnitude which in our settings is identical to the jump sized of the BDP jumps and it goes to zero as $n \to \infty$. In fact the mimicking process $X_t^{(n)}$ is a birth-death process with jump sizes equal to local state space discretization which goes to zero as $n \to \infty$, see Varadhan [2007]. For this reason also eq. (4.102) holds. $B_t^{(n)}$ is a continuous function of t and $X_t^{(n)}$ and therefore also eq. (4.101) holds. This means that every subsequence $X^{(n_k)}(t \wedge \tau_n^r) \Rightarrow X^{\bar{r}}(t \wedge \tau^r)$, where $\tau^r = \inf\{t : |X^{\bar{r}}(t)| \geq r\}$, for all $\bar{r} \geq r > 0$. Therefore the stopped processes

$$M^{\bar{r}}(t \wedge \tau^r) = X^{\bar{r}}(t \wedge \tau^r) - \int_0^{t\wedge\tau^r} b(X^{\bar{r}}(s))ds \tag{4.107}$$

$$M_i^{\bar{r}}(t \wedge \tau^r) M_j^{\bar{r}}(t \wedge \tau^r) - \int_0^{t\wedge\tau^r} \Sigma_{ij}(X^{\bar{r}}(s))ds \tag{4.108}$$

are martingales and by Ito's lemma

$$f(X^{\bar{r}}(t \wedge \tau^r)) - \int_0^{t\wedge\tau^r} Af(X^{\bar{r}}(s))ds \tag{4.109}$$

is a martingale for each $f \in D$, and $Af(X^{\bar{r}}(s))$ is the approximated infinitesimal generator applied to the function f. In particular if the martingale problem is well posed uniqueness argument for the solution hold and hold also for the solution for the stopped problem, hence

$$X^{(n)}(t \wedge \tau^r) \Rightarrow X(t \wedge \tau^r) \tag{4.110}$$

for all r. Also $r \to \infty$ implies $\tau^r \to \infty$. Therefore $X^{(n)} \to X$. □

4.3.4 Numerical Study of Weak Convergence

In this section we present some numerical weak convergence results for the proposed CTMC approximation of diffusions. Specifically we study numerically the behaviour of the L_1-norm error

$$e^{(n)} = |P_t - P_t^{(n)}| \tag{4.111}$$

as $n \to \infty$, with n the number of discretization points in each dimension where P_t is the exact density at time t, $P_t^{(n)}$ is the approximated one.
In particular with this analysis we aim to validate the convergence findings from previous sections where we found that we can at most reach the following rate of convergence

$$e^{(n)} = |P_t - P_t^{(n)}| \approx O(1/n^2) \leq C\frac{1}{n^2} \tag{4.112}$$

with C a suitable constant.

We produce numerical evidence that shows convergence for the case of Gaussian transition densities with constant coefficients, and convergence results are obtained against the corresponding closed form expression of the probability density function.

One-dimensional case

We consider a one-dimensional diffusion of eq. (4.1) with $b = 0$ and $\sigma = 1$. We denote its exact transition density solution as P_t at time $t = 1$. We approximate the diffusion with a BDP over the support $\mathcal{X} = \{-K, K\}$, with $K = \{4, 5, 6, 7, 8, 10\}$ and corresponding discretization points $n = 2^K$. This means that the interval $[-K, K]$ is larger as $n \to \infty$. This is illustrated in Figure 4.1, where we report all the density kernels for $n = 2^K, K = 4, 5, 6, 7, 8, 10$.

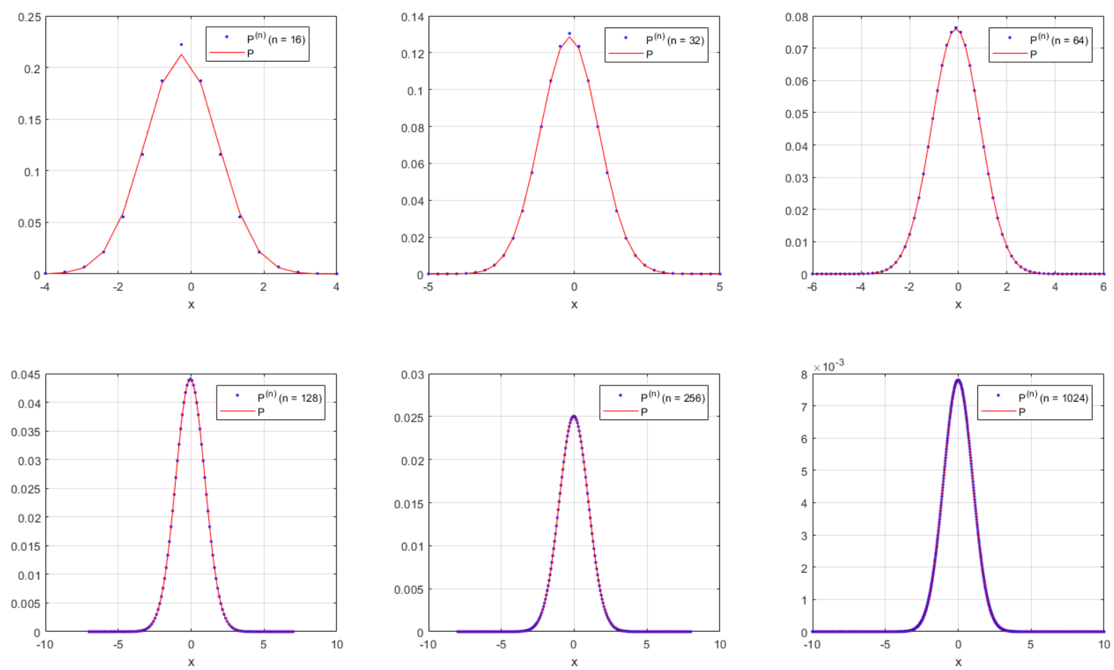

Figure 4.1: The plot illustrates the transition densities as the number n of the state space discretization points increases. The exact density P is plotted in red while the blue dots refer to the approximate density $P^{(n)}$, both computed at $t = 1$, and weak solutions to a one-dimensional dimensional diffusion with $b = 0$ and $\sigma = 1$.

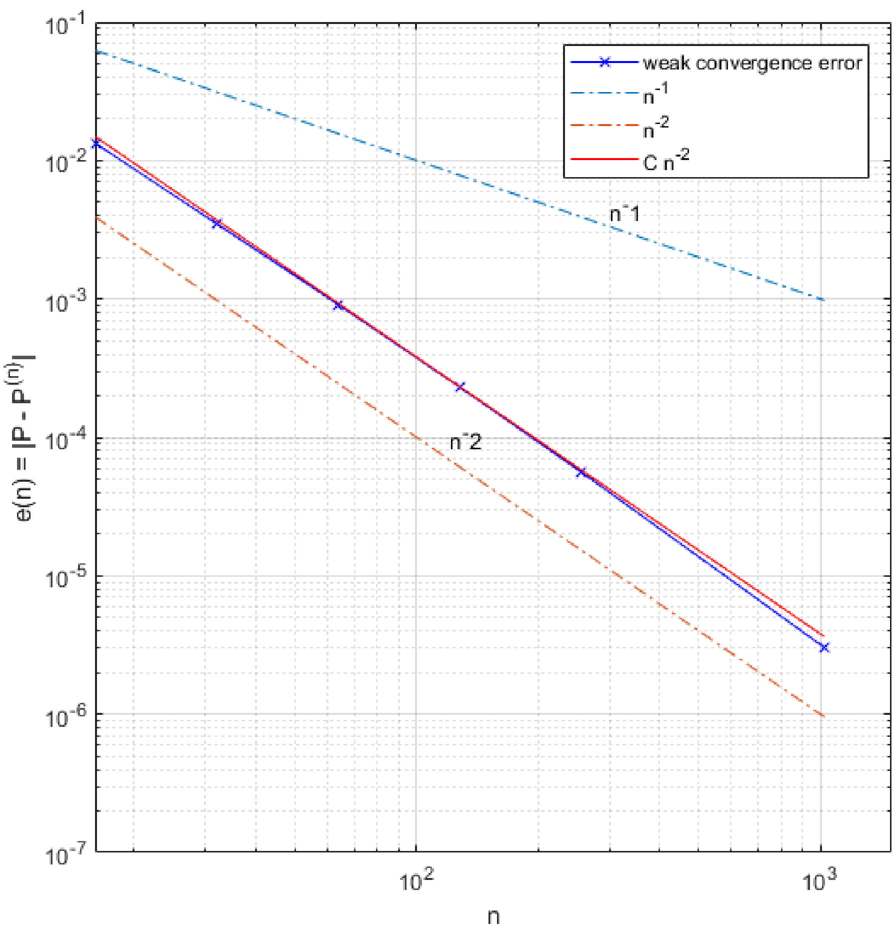

Figure 4.2: The plot illustrates the rate of convergence of the error $e^{(n)}$ in eq. (4.112) as $n \to \infty$. The rate is clearly of order $O(1/n^2)$, and in particular $C\frac{1}{n^2}$, $C \approx 3.2$. This result confirms and validates our theoretical rate of convergence investigation.

N-dimensional case

We produce some numerical weak convergence results for the proposed CTMC approximation of diffusions when the dimensionality of the diffusion of eq. (4.1) is $d > 1$, and specifically $d = 5$. We study numerically the behaviour of the L1-norm error

$$e^{(n)} = |P_t - P_t^{(n)}| \tag{4.113}$$

using the same settings as described above. However in this case we investigate the convergence rate for different values of the correlation ρ and specifically $\rho = -0.9, -0.5, -0.3, 0, 0.3, 0.5, 0.9$. Furthermore we compare the convergence rates of the CTMC approximation based on the full generator matrix exponential, see proposition 4, and the rates obtained using a CTMC approximation based on conditional generator matrices, see proposition 5.

Results are reported in Figure 4.3, where the left plot refers to the convergence rates for the full generator matrix exponential (Setting 1), and the right plot the rates obtained using a CTMC approximation based on conditional generator matrices (Setting 2).

It is important to remark few important facts regarding this convergence study:

1. Setting 1 is numerically more expensive than Setting 2, due to the fact that requires the full generator matrix exponential and its computation becomes almost impossible when $d > 5$. On the contrary computation within Setting 2 does not have any numerical limitation because the generator matrix is always calculated as decomposed across the orthogonal dimensions.

2. The convergence rate is the same when correlation is zero.

3. Setting 2 shows slower convergence rates than Setting 1.

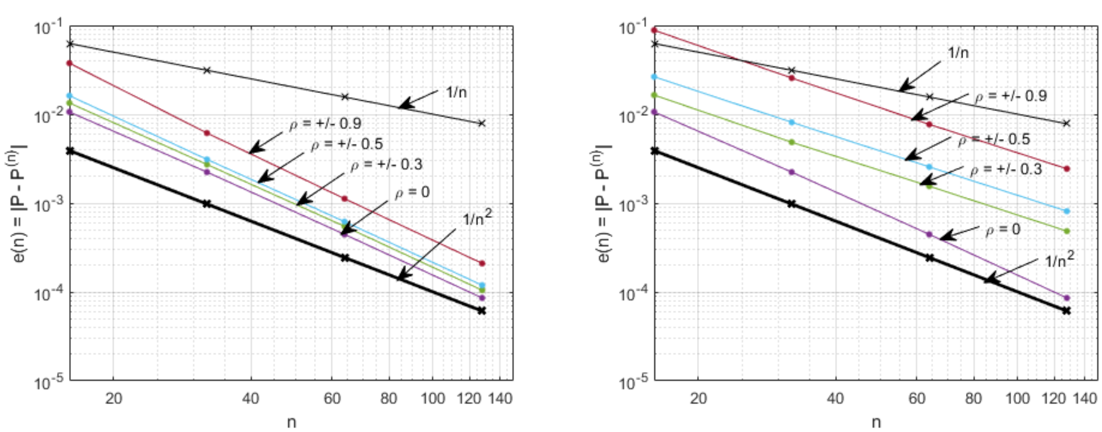

Figure 4.3: Weak convergence rates for the CTMC approximation of a d-dimensional diffusions, with $d = 5$. The left plot illustrates the convergence rates obtained for the full generator matrix exponential, while the right plot reports the rates obtained using a CTMC approximation based on conditional generator matrices.

4.4 Application to the solution of multi-dimensional SDEs

In this section we are going to apply the results developed in the previous sections to calculate the approximate solution of a multi-dimensional SDE, as per eq. (4.1).
Let us consider the following pair of stochastic differential equations,

$$dX_t = (\theta - X_t)dt + \sigma_1 dW_t^{(1)} \tag{4.114}$$

$$dY_t = bdt + \sigma_2 dW_t^{(2)} \tag{4.115}$$

where $\theta \geq 0$, $\sigma_1 > 0$, $b \geq 0$, $\sigma_2 > 0$ are real-valued parameters, the initial conditions $X_0 = x_0$, $Y_0 = y_0$ hold, and $W^{(1)}$ and $W^{(2)}$ are correlated Wiener processes with correlation coefficient ρ, i.e. $\mathbb{E}[W_t^{(1)} W_t^{(2)}] = \rho t$.
In this case of time-homogeneous correlated diffusion processes of eq. (4.114) and function $f = f(x, y)$ that are twice differentiable with compact support, the Markov generator A is given by

$$Af = (\theta - x)\frac{\partial f}{\partial x} + \frac{\sigma_1^2}{2}\frac{\partial^2 f}{\partial x^2} + b\frac{\partial f}{\partial y} + \frac{\sigma_2^2}{2}\frac{\partial^2 f}{\partial y^2} + \rho\sigma_1\sigma_2\frac{\partial^2 f}{\partial x \partial y} \tag{4.116}$$

We are interested in calculating the transition probability $P_t^{(n)}$ which approximates P_t, the density which satisfies the partial differential equation

$$\frac{\partial P_t}{\partial t} + AP_t = 0 \tag{4.117}$$

with starting conditions $\nu_0 = \delta(z)$, with $z = (\bar{x}, \bar{y})$, where δ is the Dirac delta function.
In order to approximate the two-dimensional diffusion process $Z_t = (X_t, Y_t)$ with a BDP $Z_t^{(n)}$, we construct the infinitesimal generator matrix $A^{(n)}$ following the algorithm below based on the theory and results of the previous sections.

Description of the algorithm

- Define the state-space $\mathcal{Z} = \mathcal{X} \otimes \mathcal{Y}$ for the mimicking BDP $Z_t^{(n)}$. This implies that we need to define a discrete state space for each dimension. A grid $\mathcal{X} := \{x_1, \dots, x_{n_x}\} \subset \mathbb{R}$ for the chain $X_t^{(n_x)}$ with $n_x \in \mathbb{N}$ elements, such that $x_i < x_j$ for any integers $0 \leq i < j \leq n_x$. Likewise, we create a grid $\mathcal{Y} := \{y_1, \dots, y_{n_y}\} \subset \mathbb{R}$ for the chain $Y_t^{(n_y)}$ with $n_y \in \mathbb{N}$ elements.
- We then construct the infinitesimal generator matrix $A^{(n)}$ of eq. (4.116) as

$$A^{(n)} = A^{(n_y)} \oplus_S \{A_{X|Y}^{(n_x)}\}^T \tag{4.118}$$

  where:

  1. The entries of the generator matrix $A^{(n_y)}$, associated to the second of eq. (4.115), are computed by locally solving the system of equations (4.4) which is

$$\begin{cases} a(y_1, y_2) = a(y_{n_y}, y_{n_y-1}) = 0, \\ a(y_i, y_{i+1}) = \frac{1}{2}\left(\frac{b}{h_y} + \frac{\sigma_2^2}{h_y^2}\right), \\ a(y_i, y_{i-1}) = \frac{1}{2}\left(\frac{\sigma_2^2}{h_y^2} - \frac{b}{h_y}\right), \\ a(y_i, y_i) = -(a(y_i, y_{i-1}) + a(y_i, y_{i+1})), \end{cases} \tag{4.119}$$

with $h_y = \frac{y_{n_y} - y_1}{n_y}$ and $-\sigma_2^2 \leq \mu h_y \leq \sigma_2^2$. However, the discrete state space $\mathcal{Y}$ does not need to be uniform.

2. We need to construct the generator matrix sequence $\{A^{(n_x)}_{X|Y}\}$. Each conditional infinitesimal generator matrix of the sequence $\{A^{n_x}_{X|Y}\}$ can be computed by solving eq. (4.27). The conditional approximated infinitesimal generator matrix $A^{(n_x)}_{X|Y}$ is defined by the sequence of matrices $\left\{A^{(n_x)}_{X|Y=y_j}\right\} \in \mathbb{R}^{n_x \times n_x}$, $y_j \in \mathcal{Y}$ each of whose entries are obtained according to local moment matching by:

$$A^{(n_x)}_{X|Y=y_j} = \begin{cases} a(x_1, x_2) = a(x_m, x_{m-1}) = 0, \\ a(x_i, x_{i+1}) = \frac{1}{2}\Big(\frac{b_1(x_i) + \rho(x_i, y_j)\frac{\sigma_1(x_i)}{\sigma_2(y_j)}(y_j - b(y_j))}{h_x} + \frac{\sigma_1^2(x_i)(1-\rho^2(x_i, y_j))}{h_x^2}\Big), \\ a(x_i, x_{i-1}) = \frac{1}{2}\Big(\frac{\sigma_1^2(x_i)(1-\rho^2(x_i, y_j))}{h_x^2} - \frac{b_1(x_i) + \rho(x_i, y_j)\frac{\sigma_1(x_i)}{\sigma_2(y_j)}(y_j - b(y_j))}{h_x}\Big), \\ a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})), \end{cases}$$

for all $y_j \in \mathcal{Y}$, $x_i \in \mathcal{X}$, $-1 \leq \rho(x_i, y_j) \leq 1$, and birth rates $a(x_i, x_{i+1}) > 0$ and death rates $a(x_i, x_{i-1}) > 0$ for all i, where $b_1(x) = (\theta - x)$.

- Compute the transition probability using the generator matrix of eq. (4.118)

$$P_t^{(n)} = e^{tA^{(n)}} = e^{tA^{(n_y)} \oplus_S \{A^{(n_x)}_{X|Y}\}^T} \tag{4.120}$$

We can approximate the computation of the transtion probability of eq. (4.120) further

$$P_t^{(n)} \approx e^{tA^{(n_y)}} \otimes_S e^{t\{A^{(n_x)}_{X|Y}\}^T} \tag{4.121}$$

Although the transition probability obtained from eq. (4.121) is approximated, its computation is faster than the one calculated using eq. (4.120) because the problem of matrix exponentiation becomes one-dimensional.

Two-dimensional Example and Analysis of Results

In this example we apply the algorithm described in th previous section and compute the joint transition density $P_t^{(n)}$ for the coupled process $\mathbf{X}_t = (X_t^1, X_t^2)$ using the following parameters for the first marginal process $(b_1, \sigma_1) = (0.02, 0.4)$, and these parameters $(b_2, \sigma_2) = (0.03, 0.4)$ for the second process, with $\rho_{1,2} = 0.6$, $t = 1$, and with initial condition given by $(X_0^1, X_0^2) = (0, 0)$. The computation results are reported below in Figure 4.4.
We propose in listing A.6 the MATLAB code which closely follows the steps of the above two-dimensional CTMC approximation algorithm.

4.5 Conclusion

In this chapter we investigated aspects of correlated diffusions approximation by means of a multidimensional birth-death process. We formulated a new interpretation of the dependence among Markov processes using the martingale approach. We showed that it is possible to represent, in both continuous and discrete space, that a multidimensional

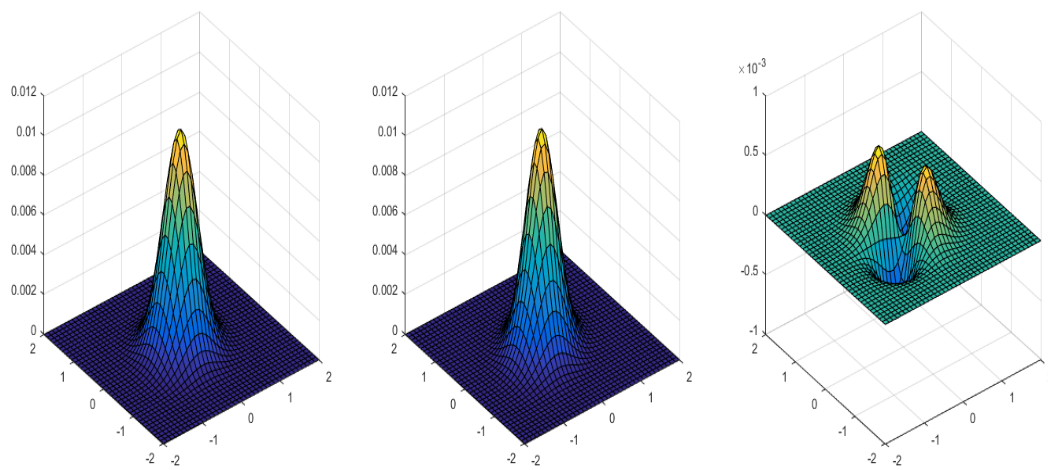

Figure 4.4: Approximated transition probabilities computed using the above algorithms with the suggested diffusion parameters with time $t = 1$. On the left hand side the transition probability corresponding to $e^{tA^{(n_y)} \oplus_S \{A^{(n_x)}_{X|Y}\}^T}$. In the middle we plot the transition probability corresponding to $e^{tA^{(n_y)}} \otimes_S e^{t\{A^{(n_x)}_{X|Y}\}^T}$, therefore simplifying the matrix exponential problem from a two-dimensional one into two one-dimensional problems, and such approximation is numerically faster than the computation of the exponential of the corresponding two-dimensional matrix. On the right hand side we plot the difference of the two obtained transition probability kernels.

correlated generalized diffusion is a linear combination of processes that originate from the decomposition of the starting multidimensional semimartingale. This result not only reconciles with the existing theory of diffusion approximations and decompositions, but defines the general representation of approximated infinitesimal generators for both multidimensional generalized diffusions and, as we demonstrated, also for the specification of copula density dependence structures. This new result provides immediate representation of the approximate solution for correlated stochastic differential equations. We showed convergence results for the proposed approximations.

Chapter 5

Mimicking Copulas in a Tensor Space and Generalized Copula Mapping

In this chapter we propose a methodology to mimic any copula function in a tensor space through the construction of a generalized local Gaussian copula function, which is a novel class of functional copula representations for dependence based on the continuous time Markov chain (CTMC) approximation of a generalized local Gaussian copula function introduced in Chapter 4.

This can be viewed as a reinterpretation of Sklar's well known copula representation theorem of multivariate dependence and is achieved in a manner that allows one to characterize any copula dependence function via a unique map to a generalized local Gaussian copula function.

This unique mapping is obtained through the quantification of local Gaussian dependence properties of the reference copula model over a discretized tensor space. Such a representation is proven to be exact as the discretization interval of the target copulas support diminishes, with known convergence rate.

In particular, in this chapter we study the tensor approximation and its numerical characteristics for a wide variety of copula models, demonstrating how the functional copula characterization proposed can accurately reconstruct any target copula by means of local Gaussian copula projections. Developing such representations then has many advantages for evaluation of integral functionals taken with respect to a variety of copula functions. Importantly, it also allows one to visualize and interpret the target copula functions characteristics. This is particularly useful in understanding the relationships between copula model parameters and the strength of induced concordance structures captured by particular copula types.

5.1 Introduction

The subject of this chapter is the construction of Gaussian copula distribution functions, approximated by continuous time Markov chains over a tensor space, whose values mimic the behaviour of more complicated parametric copula distribution functions.

The main motivation of our work is the necessity to better understand and interpret the parameters of some complex copula functions and how their value contributes in creating elaborated dependence structures. The method proposed to represent complex copula functions can be classified as indirect method because it does not aim to just approximate the copula with a numerical scheme, rather to mimic its local properties and functional behaviour with a local equivalent Gaussian copula, which yields an identical correlation structure but with a simpler interpretation.

Another more general motivation is to develop a functional approximating framework for copula distributions which would make the computation of functional or integral of copulas simpler.

Copulas are among the most important and used functions in statistics to characterize and study the dependence between the components of a random vector. A copula distribution is simply a multivariate probability distribution for which the marginal probability distribution of each variable is uniform Nelsen [1999]. Copulas are often used in high-dimensional statistical applications as they allow one to separate out the modelling and estimation of the distribution of dependent random variables by estimating first the marginals and then capturing the dependence structure through estimation of a copula function. Their range of applications is very broad, see a recent review of many aspects and applications of this modelling approach in Cruz *et al.* [2015a] and works such as Demarta & McNeil [2005], Aas *et al.* [2009],Genest *et al.* [2011], Haff *et al.* [2013] and Patton [2009]. Furthermore, there is an increasing number of copula models aimed at modelling different forms of dependence, both parametrically and non-parametrically, see discussions for instance in Nelsen [1999], Genest & Favre [2007], Cherubini *et al.* [2004], Joe [1997], Embrechts *et al.* [2001] and Brigo *et al.* [2010a].

However, in many cases there is no clear and intuitive correspondence between the magnitude of a copula's parameters and the dependence structure they create.

This has led to numerous studies of asymptotic relationships between concordance measures such as tail dependence and the copula functions tail behaviours, as characterized by the copula parameters. In some cases this has yielded simple relationships that are closed form and analytic. However, more often than not there is still a need to better understand how dependence manifests locally in the support of the target copula distribution, and how it changes with changes in the copula parameters or the dimension of the copula since it is often difficult to characterise in closed form.

In Taylor [2007] they provided a representation of the axioms of a concordance measure explicitly in terms of copula models, which extends the work previously by Scarsini [1984]. This provides a direct link between these measures of dependence and the copula model. However, a good understanding of the strength or significance of a concordance measure as a function of the copula model parameters is not well understood and difficult to study as they often involve complicated intractable integrations of copula functions, not easy to achieve in a computationally efficient and accurate manner in arbitrary dimensions. In this chapter we aim to address this challenge in a general approach that is numerically accurate and computationally efficient.

5.1.1 Summary of the proposed methodology

Let us consider a d-dimensional copula distribution function, denoted by $C_{\text{target}}(\mathbf{u};\theta_c) := C_{\text{target}}(\mathbf{u}) : [0,1]^d \rightarrow [0,1]$, with $d > 1$ and $\mathbf{u} := (u_1, \ldots, u_d) \in [0,1]^d$ a random vector, and θ_c the set of copula parameters. The set θ_c can be large, its elements may not exhibit direct correspondence with the statistical properties of C_{target}, which is a non-trivial copula function. We aim to simplify the analysis of the statistical properties of $C_{\text{target}}(\mathbf{u}, \theta_c)$ and also its computation, by mimicking the distribution $C_{\text{target}}(\mathbf{u}, \theta_c)$ with $C^{(n)}$, a discrete local Gaussian copula distribution originated by a continuous time Markov chains (CTMC) approximation of a correlated diffusion on a tensor space.

We report below a summary of the copula mimicking result, although we refer the reader to the next sections of this chapter for a more rigorous introduction to the problem and to our copula mapping methodology.

Let $\mathbf{X}_t$, $t \geq 0$ a d-dimensional Markov process on the probability space $(\Omega, \mathcal{F}, P)$ with $\mathbf{X}_t := (X_t^1, \ldots, X_t^d) \in \mathbb{R}^d$ denoting its components vector at time t, with $\mathbb{R}^d$ denoting the d-dimensional Euclidean space. Our modelling framework deals with the local approximation of X_t by CTMC.

The Markov processes $\mathbf{X}_t$ has stochastic differential equation (SDE) of the form,

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Psi(\mathbf{X}_t)d\mathbf{W}_t \tag{5.1}$$

where $(d\mathbf{W}_t, \mathcal{F}_t)$ is a d-dimensional Wiener process, $\mathbf{b}$ and Ψ are bounded $\mathcal{F}_t$-adapted processes such that $\Sigma = \Psi\Psi'$ is positive definite. Then we saw in Chapter 4 that there exists a continuous time Markov chain $\mathbf{X}_t^{(n)}$ approximating $\mathbf{X}_t$, with $\mathbf{X}_t^{(n)} := (X_t^{(n_1)}, \ldots, X_t^{(n_d)}) \in \mathcal{X}$ denoting its components vector at time t. In particular the Markov chain is constructed as

$$\mathbf{X}_t^{(n)} : \mathcal{X} \to \mathbb{R}^d, \tag{5.2}$$

with $\mathcal{X} = \bigotimes_{i=1}^d \mathcal{X}^{\mathbf{i}} \in \mathbb{R}^d$ denoting the d-dimensional tensor space and $n = n_1 \cdots n_d$. Specifically we denote a countably finite set by $\mathcal{X}^k := \{x_0^{(k)}, \ldots, x_{n_k}^{(k)}\} \in \mathbb{R},\ \ k = 1, \ldots, d$, with n_k the number of discretization points for the k-dimension, and $\mathcal{X}^k$ represents the state space for the Markov chain $X_t^{(n_k)}$ approximating X_t^k.

The chain $\mathbf{X}_t^{(n)}$ has transition probability at time t equal to $P_t^{(n)}$ which weakly converges to P_t (see Chapter 4, sections 4.3 and 4.3.4), being P_t the distribution of the admitted weak solution $\mathbf{X}_t$ of the SDE 5.1.
We then investigate properties for the Markov chain joint distribution function which is defined as

$$F_t^{(n)}(\mathbf{x}) = P_t^{(n)}(\mathbf{X}_t^{(n)} \leq \mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d, \tag{5.3}$$

and we specify the approximated copula function $C^{(n)}$, as the joint distribution function of the Markov chain

$$\mathbf{U}_t^{(n)} = F_t^{(n)}(\mathbf{X}_t^{(n)}) \tag{5.4}$$

with $C^{(n)} : \mathbf{U} \to [0,1]$, and $\mathbf{U} := \bigotimes_{i=1}^d \mathbf{U}^i \in [0,1]^d$ is the discretized unit hypercube, with $\mathbf{U}^k := \{u_0^{(k)}, \ldots, u_{n_k}^{(k)}\} \in [0,1],\ \ k = 1, \ldots, d$, such that

$$F_t^{(n)}(\mathbf{x}) = C^{(n)}(F_t^{(n_1)}(x_1), \ldots, F_t^{(n_d)}(x_d)), \quad \text{for all } \mathbf{x} \in \mathcal{X} \in \mathbb{R}^d. \tag{5.5}$$

Moreover for $\mathbf{u} \in \mathbf{U} \subset [0,1]^d$ we have that

$$C^{(n)}(\mathbf{u}; \rho(\mathbf{u})) = F_t^{(n)}\Big(F_t^{-1,(n_1)}(u_1), \ldots, F_t^{-1,(n_d)}(u_d); \rho(\mathbf{u})\Big) \tag{5.6}$$

We then compute the local likelihood with respect to each local Gaussian correlation parameter $\rho(\mathbf{u})$ based on the following local minimizer

$$\min_{\rho(\mathbf{u})} \left\| \log \left(C^{(n)}_{\text{target}}(\mathbf{u}; \theta_c) - C^{(n)}(\mathbf{u}; \{\theta, \rho(\mathbf{u})\}) \right) \right\|_2, \quad \text{for} \quad \mathbf{u} \in \mathbf{U} \tag{5.7}$$

where C_{target} can be any target copula function and denote by $C^{(n)}_{\text{target}}(\mathbf{u}, \theta_c)$ the evaluation of the copula distribution $C_{\text{target}}(\mathbf{u}, \theta_c)$ on a point $\mathbf{u} \in \mathbf{U}$. Therefore starting from the correlated diffusion of eq. (5.1) we construct through a specific methodology detailed in this chapter, the local Gaussian copula distribution function $C^{(n)}(\mathbf{u}; \rho(\mathbf{u}))$ in eq. (5.6). This is the main result of this chapter and in particular we focus on how it is possible with this local approximated Gaussian copula to mimic more complex distributions, namely $C_{\text{target}}(\mathbf{u}, \theta_c) \approx C^{(n)}(\mathbf{u}; \rho(\mathbf{u}))$ on the discretized hypercube $\mathbf{U}$, and then understand in a simpler way and also visualize the dependence properties of each C_{target} through the inferred Gaussian local correlation coefficients $\rho(\mathbf{u}) : \mathbf{U} \to [-1, 1]$.

5.1.2 Outline of Contributions

In this chapter, we aim to develop an alternative approach to representing any copula distributional model that would facilitate three key objectives: the first is to exactly and uniquely create a functional copula characterization of Sklar's theorem, which is based on a locally adapted generalized Gaussian copula representation, that will be applicable to all copula distributions in arbitrary dimension. This functional copula is constructed from a continuous time discrete state space mimicking Markov chain that we term a (CTMC) functional generalized Gaussian copula; secondly, we aimed to develop such a representation so that it may be numerically approximated efficiently in arbitrary dimensions, in any subspaces of the support of the target copula distribution; thirdly, the characterization should easily facilitate visualization, evaluation of functionals of the copula over local regions of the support of the target copula, and greater understanding of the relationship between the target copula model parameters and the concordance/dependence features present in the copula model. Examples of such functionals of interest include as illustration, the concordance measures characterized in Taylor [2007].

The approach we adopt involves extending the theoretical functional copula characterizations first presented in Chapter 4 and summarized in a self-contained manner in the next section 5.1.3. In in Chapter 4 we developed a class of applied probabilistic continuous time but discretized state space decompositions of the characterization of a multivariate generalized diffusion process. This decomposition is novel and, in particular, it allows one to construct families of mimicking classes of processes for such continuous state and continuous time diffusions in the form of a discrete state space but continuous time Markov chain representation. At first sight, this doesn't seem relevant to addressing the three objectives targeted in this chapter, however it turns out to be the key to addressing this issue through a functional copula representation. We demonstrate that the result they developed defines the general representation of infinitesimal generators for multidimensional generalized diffusions and, as demonstrated in their work, can also be used directly for the specification of copula density dependence structures in what they term the functional copula characterization. The work in Chapter 4 was of a purely theoretical nature, and didn't provide methodological or algorithmic approaches to demonstrate how such characterizations can be adopted by statisticians, this chapter aims to develop these aspects.

In this chapter, we develop the methodological aspects of this theoretical characterization to allow statisticians to implement and develop copula functions' approximations based on this characterization. The functional copula characterization developed allows one to build a unique class of approximations replete with a variety of resulting numerical implementations for working with and characterizing general copula distributions in arbitrary dimensions. In particular, in this chapter we study the tensor approximation and its numerical characteristics for a wide variety of copula models, demonstrating how the functional copula characterization proposed can accurately reconstruct any target copula.

The paper is structured as follow. In section 5.2, we develop a functional copula representation of Sklar's Theorem based on a continuous time, discrete state-space Markov chain characterization of a multivariate generalized diffusion which aims to characterize Gaussian copula projections locally in a tensor space. In section 5.3 we assess the accuracy of the proposed functional local Gaussian copula approximations though a series of numerical examples and we carry out a convergence analysis. In this section we illustrate how the local Gaussian copula mapping performs when the target copula distributions belong to either the Elliptical family, Archimedean or distorted Archimedean class. Section 5.4 concludes.

5.1.3 Background Results on Reconstruction of a Copula via a Tensor Decomposition

In this section we briefly recall some relevant notations and definitions and results, derived in details in Chapter 4, that will aid in the developments in future sections of this chapter. The main idea of these background results is to illustrate a methodology on how to mimic 'locally' a multivariate correlated diffusion by means of a multivariate orthogonal birth-death process (DBP). By 'locally' we mean on each coordinate of a suitable constructed tensor space, of which we exploit both its local and tensor algebraic properties.

In each dimension of this tensor space, each mimicking BDP approximates the dynamics of the diffusion projected in that dimension, and this is done through suitably constructing the BDP associated infinitesimal generator matrix. This leads to the approximation of the whole diffusion onto a space fully described by the tensor sum of generator matrices.

We remark 'tensor' sum instead of 'Kronecker' sum. In fact in our proposed methodology to mimic correlated diffusion in a tensor space, the terminology we use to indicate operations on this space depends on how we look at generator matrices of the mimicking BDP's. The notions of Kronecker product on matrices and tensor product on linear maps between vector spaces technically represent operations on different objects. However a generator matrix is linear map between vector spaces equipped with a chosen basis. Therefore the Kronecker product of two generator matrices then represents the tensor product of the two associated linear maps. And this is the way we have chosen to look at the algebra within our proposed methodology.

We will build on these background results to propose the methodology for mimicking a copula function in a tensor space and to map this mimicking copula onto any copula function.

Below we provide a self contained summary of the key results and algorithms from Chapter 4, which include:

1. Construction of a process mimicking a diffusion through local moment matching. At

this purpose the mimicking process is a birth-death process (BDP), and we show how to calculate its birth and death rate in a way to exactly replicate some target moments. The BDP is represented through a continuous time Markov chain (CTMC). See Section 4.2 and specifically 4.2.1.

2. Construction of a process mimicking a conditional diffusion. At this purpose we introduce the concept of conditional infinitesimal generator matrix, which is a sequence of generator matrices. We also introduce the concept of orthogonal projections in a tensor space , see Section 4.2.1.

3. Approximation of the weak solution of a diffusion through the matrix exponentiation of the infinitesimal generator matrix. See Proposition 4.

4. Approximation of the weak solution of a multidimensional diffusion through the matrix exponentiation of an infinitesimal generator matrix which spans a tensor Kronecker product space. The weak solution is approximated only over orthogonal dimensions and this means that the multivariate weak solution is not directly constructed rather calculated from elementary one dimensional conditional weak solutions. See Proposition 7.

5.2 Developing a Functional Copula Representation of Sklar's Theorem

In the approach proposed, we will be required to work with a continuous time, discrete state-space Markov chain characterization of a multivariate generalized diffusion. The reason that one is interested in the generators of continuous time discrete state space multivariate Markov processes is that these may then be seen, after a functional transformation, as decompositions of the multivariate distribution characterizing the process. Furthermore, they may then be used to match or represent or mimic any target multivariate distribution.

This will be the key to the results presented in this manuscript when developing the functional representation of Sklar's theorem.

We are interested in considering new representations for a variety of statistical dependence models, known as copula models. Copula models parametrically characterize different structural dependence features, either explicitly or implicitly depending on the class of copula model under consideration. However, understanding the relationships between copula parameters and their concordance (dependence) structures induced by the parametrization considered is a challenge for statisticians.

We begin by recalling the formal definition of such dependence copula models, followed by the unique representation theorem developed in Sklar's theorem. This represents a fundamental result about the relationship between marginals and joint distribution for multivariate concordant (dependent) random variables.

Definition 26. *[d-dimensional Copula Function] An d-dimensional copula is a function* $C : [0, 1]^d \to [0, 1]$ *which satisfies the following conditions:*

1. $C_i(u^{(i)}) = C(1, \ldots, 1, u^{(i)}, 1, \ldots, 1) = u^{(i)}$ *for every* $i \leq d$ *all* $u^{(i)} \in [0, 1]$*. This means that a copula on* $\mathbb{R}^d$ *is a cumulative distribution function (cdf) C whose marginal cdf's are all equal to the [0,1]-uniform cdf.*

2. *$C(\mathbf{u}) = 0$ if $u^{(i)} = 0$ for $i \leq d$. It states that the joint probability of all outcomes is zero if one marginal probability is zero.*

3. *C is d-increasing. This means that the C-volume of any d-dimensional interval is non-negative. C assigns non-negative volumes to cuboids of $[0,1]^d$.*

Properties 2 and 3 are general properties of multivariate distribution functions. It follows that an d-dimensional copula function can be defined as an d-dimensional cdf whose support is contained in $[0,1]^d$ and whose one-dimensional marginals are uniform on $[0,1]$. Therefore a d-dimensional copula is an d-dimensional distribution function with all d univariate marginals being $U(0,1)$.

The relationship between distribution functions and copulas is given by the following well known result first developed by Sklar [1959] and further overviews in for instance Sklar [1996].

Theorem 13 (Sklar's Theorem. d-Dimendional Case.)**.** *Let $F \in \mathcal{F}(F_1,\ldots,F_d)$ be an d-dimensional distribution function with marginals $F_1,\ldots,F_d$, where the set $\mathcal{F}(F_1,\ldots,F_i,\ldots)$ is the Fréchet class of the F_i's which collects all multivariate joint distribution functions that have the same marginals. Members of a Fréchet class only differ with respect to the interdependence between their marginals. Then there exist a copula $C \in \mathcal{F}(U_1,\ldots,U_d)$ with uniform marginals such that*

$$F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots,F_d(x_d)) \tag{5.8}$$

which is uniquely determined on $[0,1]^d$ for distributions F with absolutely continuous marginals. Conversely any copula function can be used together a set of univariate distribution functions $(F_1,\ldots,F_d)$ and produce a joint distribution function F using eq. (5.8).

5.2.1 Bivariate Functional Copula Constructions by CTMC approximation

In order to construct the framework of a functional copula representation by means of CTMC, we first deal with some new decompositions of the infinitesimal generator of a process that will be used to obtain such a representation.

The mimicking of the target copula model is achieve through local, in the state space, moment matching conditions. To understand these conditions, we first recall a global (holds over the entire support) decomposition result that will be then extended to a local framework. We begin by providing a very simple motivating example in two dimensions to motivate what will be presented in a very general class of results in proposition 10 and proposition 12 which characterize the tensor copula representation of Sklar's theorem in this simple bivariate case.

Example 2. *Consider the problem of estimating a random variable X given the value of the variable Y. Let's assume that X and Y are normal variables with positive variance and define*

$$\hat{X} = \rho \frac{\sigma_X}{\sigma_Y} Y \tag{5.9}$$

with ρ the correlation coefficient between X and Y. The variable $\hat{X}$ is the linear least square estimator of X. If we define by

$$\tilde{X} = X - \hat{X} \tag{5.10}$$

the estimation error we have that the variables $\tilde{X}$ and Y are jointly normal, uncorrelated and therefore independent, due to

$$\mathbb{E}[Y\tilde{X}] = \mathbb{E}[YX] - \mathbb{E}[Y\hat{X}] = \rho\sigma_X\sigma_Y - \rho\frac{\sigma_X}{\sigma_Y}\sigma_Y^2 = 0. \tag{5.11}$$

Since $\hat{X}$ is a scalar multiple of Y, it follows that $\hat{X}$ and $\tilde{X}$ are independent. We have therefore decomposed X into the sum of two independent normal random variables, namely,

$$X = \hat{X} + \tilde{X} = \rho\frac{\sigma_X}{\sigma_Y}Y + \tilde{X} \tag{5.12}$$

Taking the condition expectation of both sides of eq.(5.12) given Y, we obtain

$$\mathbb{E}[X|Y] = \rho\frac{\sigma_X}{\sigma_Y}\mathbb{E}[Y|Y] + \mathbb{E}[\tilde{X}|Y] = \rho\frac{\sigma_X}{\sigma_Y}Y = \hat{X} \tag{5.13}$$

where $\mathbb{E}[\tilde{X}|Y] = 0$ due to independence between Y and $\tilde{X}$. Therefore the conditional expectation $\mathbb{E}[X|Y]$ is a linear function of the random variable Y.
We also note that the first two moments of $\hat{X}$ are

$$\mathbb{E}[\hat{X}] = \rho\frac{\sigma_X}{\sigma_Y}\mathbb{E}[Y] \quad \text{and} \quad \mathbb{V}ar[\hat{X}] = \rho^2\sigma_X^2 \tag{5.14}$$

This illustrative bivariate example is a general characterization of the relationships between random variable X and random variable Y that holds over the entire support of the two random variables (X, Y). However, one can also define such dependence relationships locally in the state space and in arbitrary dimensions. Therefore, this bivariate global result is useful to serve as a familiar motivation of the following generalizations to arbitrary dimensions. In this generalization, we will utilise such results for local state space representations of the structure of a random vectors dependence. Note, local here refers to the fact that we consider a local area of the support of the reference joint distribution under study.

We will now take the concept from the bivariate projection above, to utilise to create a mimicking BDP which can locally match the first two conditional moments of a diffusion. The aim is to be able to retrieve the approximate solution of a multidimensional diffusion from conditional mimicking birth death processes which approximate the diffusion only along orthogonal dimensions. This means that the proposed approximation of a correlated diffusion does not involve cross-spaces and this will be extremely beneficial for computational speed.

The following generalization of the above motivating example can be developed to propose a new decomposition of a bivariate distribution, locally in its state space (distribution support), for Markov processes with correlated operators.

Proposition 9 (Conditional CTMC Operator Decomposition)**.** *Let $\mathbf{X}_t^{(n)} := (X_t^{(n_1)}, X_t^{(n_2)})$ be a bivariate continuous time Markov chain, a bivariate BDP, mimicking the diffusion $\mathbf{X}_t := (X_t^1, X_t^2)$ with generator $A(x) = \sum_{i=1}^2 b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^2 a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}$ over the space $\mathcal{X} := \bigotimes_{i=1}^2 \mathcal{X}^{\mathbf{i}} \in \mathbb{R}^2$. The processes $\mathbf{X}_t$ exhibit non zero correlation. We then consider the conditional process $(X_t^1|X_t^2)$ and construct the mimicking BDP $(X_t^{(n_1)}|X_t^{(n_2)})$, with its associated conditional infinitesimal generator matrix sequence $\{A_{X^1|X^2}^{(n_1)}\}$ by matching the* *first two conditional moments along the lines of Proposition 5. It is possible to show that the generator sequence can be decomposed as the sum of the following operators*

$$\{A^{(n_1)}_{X^1|X^2}\} = A^{(n_1)}_{X^1} - \{\hat{A}^{(n_1)}_{X^1|X^2}\} \tag{5.15}$$

where $A^{(n_1)}_{X^1}$ is the marginal infinitesimal generator matrix of the Markov chain $X^{(n_1)}_t$ and $\{\hat{A}^{(n_1)}_{X^1|X^2}\}$ is the generator matrix sequence associated to the Markov chain sequence $\hat{X}^{(n_1)}_t$, the linear least square estimator.

Proof. We observe that each local conditional infinitesimal generator matrix $A^{(n_1)}_{X^1|X^2=y_j} \in \mathbb{R}^{n_1\times n_1}$ for the Markov chain process $X^{(n_1)}_t$ conditional on the value of the chain $X^{(n_2)}_t = x^{(2)}_j$ for all $x^{(2)}_j \in \mathcal{X}^2$, is constructed, as per Proposition 5, by matching the following instantaneous local moments:

$$\begin{cases} \mathbb{E}[X^{(n_1)}_{t+\Delta t} - X^{(n_1)}_t | X^{(n_1)}_t] = b_1(x^{(1)}_i) + \rho(x^{(1)}_i, x^{(2)}_j)\frac{\sigma_1(x^{(1)}_i)}{\sigma_2(x^{(2)}_j)}(x^{(2)}_j - b_2) \\ \mathbb{V}ar[X^{(n_1)}_{t+\Delta t} - X^{(n_1)}_t | X^{(n_1)}_t] = \sigma^2_1(x^{(1)}_i)(1-\rho^2(x^{(1)}_i, x^{(2)}_j)) \end{cases} \text{for all } i,j. \tag{5.16}$$

In this way we calculate the conditional birth and death rates entries for the BDP infinitesimal generator matrix sequence. Then if we denote by $\{\hat{A}^{(n_1)}_{X^1|X^2}\} \in \mathbb{R}^{n_1\times n_1}$ the generator matrix sequence associated to the Markov chain $\hat{X}^{(n_1)}_t$ constructed by specifying the following local instantaneous moments

$$\begin{cases} \mathbb{E}[\hat{X}^{(n_1)}_{t+\Delta t} - \hat{X}^{(n_1)}_t | X^{(n_1)}_t] = -\rho(x^{(1)}_i, x^{(2)}_j)\frac{\sigma_1(x^{(1)}_i)}{\sigma_2(x^{(2)}_j)}(x^{(2)}_j - b_2) \\ \mathbb{V}ar[\hat{X}^{(n_1)}_{t+\Delta t} - \hat{X}^{(n_1)}_t | X^{(n_1)}_t] = \sigma^2_1(x^{(1)}_i)\rho^2(x^{(1)}_i, x^{(2)}_j) \end{cases} \text{for all } i,j, \tag{5.17}$$

we have that $A^{(n1)}_{X^1} = \{A^{(n_1)}_{X^1|X^2}\} + \{\hat{A}^{(n_1)}_{X^1|X^2}\}$. □

Remark 1 *We note that one can interpret this sequence of operator matrices $\{\hat{A}^{(n_1)}_{X^1|X^2}\}$ as representing the operator associated to the linear least squares estimator of the conditional Markov chain process $X^{(n_1)}_t$ given local information on the marginal Markov chain process $X^{(n_2)}_t$, namely the process $\hat{X}^{(n_1)}_t = \rho^2(x_i, y_j)\frac{\sigma(x_i)}{\sigma(y_j)}\hat{X}^{(n_2)}_t$.*

Given the results in the above proposition 9 we can formulate the following proposition 10 which is the first result we present and denote as a CTMC functional Copula infinitesimal operator, in this case for an arbitrary bivariate distribution of random vector $\mathbf{X}_t := (X^1_t, X^2_t)$.

Proposition 10 (CTMC Functional Copula Infinitesimal Operator)**.** *Let $\mathbf{X}^{(n)}_t := (X^{(n_1)}_t, X^{(n_2)}_t)$ be a bivariate continuous time Markov chain with approximate weak solution $P^{(n)}_t$, mimicking the correlated diffusion $\mathbf{X}_t := (X^1_t, X^2_t)$. Then the CTMC mimicking process can be characterized by the bivariate functional copula density operator representation given by*

$$c^{(n)} = c^{(n_1 n_2)} := I^{(n_1)} \otimes_S \{\exp(t\hat{A}^{(n_2)}_{X^2|X^1})\} \tag{5.18}$$

where $I^{(n_1)} \in \mathbb{R}^{n_1\times n_1}$ is the unitary matrix, and $\{\hat{A}^{(n_2)}_{X^2|X^1}\}$ is the operator matrix sequence of the linear least square estimator of the conditional Markov chain process $(X^{(n_1)}_t | X^{(n_2)}_t)$ as per proposition 9.

Proof. In order to calculate the approximated expression for the mimicking copula density $c^{(n)}$, we first look at how we can rewrite the approximate solution $P^{(n)}$. We have:

$$
\begin{aligned}
P_t^{(n)} := P_{X_t^1,X_t^2}^{(n_1 n_2)} &\approx P_{X_t^1}^{(n_1)} \otimes_S \{P_{X_t^2|X_t^1}^{(n_2)}\} \\
&= \exp(tA_{X^1}^{(n_1)}) \otimes_S \Big(\{\exp(tA_{X^2|X^1}^{(n_2)})\}\Big) \\
&= \exp(tA_{X^1}^{(n_1)}) \otimes_S \Big(\exp(tA_{X^2}^{(n_2)}) \times \{\exp(t\hat{A}_{X^2|X^1}^{(n_2)})\}\Big) \\
&= \Big(\exp(tA_{X^1}^{(n_1)}) \otimes \exp(tA_{X^2}^{(n_2)})\Big) \times \Big(I^{(n_1)} \otimes_S \{\exp(t\hat{A}_{X^2|X^1}^{(n_2)})\}\Big)
\end{aligned}
\tag{5.19}
$$

Then we can rewrite

$$
c^{(n)} = I^{(n_1)} \otimes_S \{\exp(t\hat{A}_{X^2|X^1}^{(n_2)})\} = \frac{P_t^{(n)}}{\exp(tA_{X^1}^{(n_1)}) \otimes \exp(tA_{X^2}^{(n_2)})} \tag{5.20}
$$

□

The result of the above Proposition 10 is remarkable, because it allows us to directly compute the approximated operator of a generic copula function together with its density. For instance, given this result we may now extend it to the representation of joint distribution function of any desired target bivariate distribution by using tensor algebra results from the ancillary results section. Without loss of generality we present the result in two dimensions. Extension to higher dimensional case is straightforward. To achieve this we first create a conditional generator in each orthogonal dimension of the partitioned state space $\mathcal{X}$, as per eq. (4.43) which we report below

$$
A^{(n)} = A_{X^1}^{(n_1)} \oplus_S \{A_{X^2|X^1}^{(n_2)}\} \oplus_S \cdots \oplus_S \{A_{X^d|X^1,\ldots,X^{d-1}}^{(n_d)}\} \tag{5.21}
$$

This consists in creating sequences of generator matrices each conditional on a point or a set of points of the state space. The complexity of the algorithm is at most equal to the total number of state space points in $\mathcal{X}$, which is n^2 with $n = n_1 \cdots n_d$. The second step is the calculation of $P^{(n)}$, which is straightforward given $A^{(n)}$ and using the reduced rank approximation of the Kronecker product for sequence of matrices as per eq. (4.43).

5.2.2 Approximation Scheme for a Copula via a Tensor Decomposition

With these results we can now explain how to perform the representation of a target copula distribution through this framework.

We denote by $\mathbf{x} := (x_1, \ldots, x_d) \in \mathbb{R}^d$ a d-dimensional vector, and inequalities $\mathbf{x} \leq \mathbf{z}$ are intended componentwise, i.e., $x_i \leq z_i$ for all $i = 1, 2, \ldots, d$. Let $\mathbf{X}_t^{(n)}$ a d-dimensional Markov chain with $d \geq 2$, specifically with $X_t^{(n_1)}, X_t^{(n_2)}, \ldots, X_t^{(n_d)}$ marginal chains, each with support $\mathcal{X}^k = (x_1^{(k)}, x_2^{(k)}, \ldots, x_{n_k}^{(k)}) \in \mathbb{R}$, where $n_1, n_2, \ldots$ denotes the number of corresponding chains' states. Therefore

$$
\mathbf{X}_t^{(n)} : \mathcal{X} \to \mathbb{R}^d, \quad \text{with } \mathcal{X} = \bigotimes_{i=1}^{d} \mathcal{X}^i \in \mathbb{R}^d \tag{5.22}
$$

We will use the equivalent notation $\mathbf{X}_t^{(n_1,n_2,\ldots,n_d)}$ for the Markov chain $\mathbf{X}_t^n$, with $n = n_1 \cdots n_d$. We denote the Markov chain joint distribution function by

$$F_t^{(n)}(\mathbf{x}) = P_t^{(n)}(\mathbf{X}_t^{(n)} \leq \mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d, \tag{5.23}$$

with marginal distribution functions

$$F_t^{(n_k)}(x) = P_t^{(n_k)}(X_t^{(n_k)} \leq x), \quad x \in \mathcal{X}^{\mathbf{k}} \subset \mathbb{R},\ k = 1, \ldots, d. \tag{5.24}$$

We specify the approximated copula function as the joint distribution function of the Markov chain

$$\mathbf{U}_t^{(n)} = F_t^{(n)}(\mathbf{X}_t^{(n)}) \tag{5.25}$$

with marginal chains $U_t^{(n_k)} = F_t^{(n_k)}(X_t^{(n_k)})$, for $k = 1, \ldots, d$, each with support $\mathbf{U}^k = F_t^{(n_k)}(\mathbf{X}^k) = (u_1^{(k)}, u_2^{(k)}, \ldots, u_{n_k}^{(k)}) \in [0,1]$, where $F_t^{(n_k)}$ denotes the application of the distribution function of the Markov chain $X_t^{(n_k)}$ to each point, for all $k = 1, \ldots, d$.

We denote the approximated copula function by $C^{(n)} : \mathbf{U} \to [0,1]$, where $\mathbf{U} := \bigotimes_{i=1}^d \mathbf{U}^i \in [0,1]^d$ is the discretized unit hypercube, such that

$$F_t^{(n)}(\mathbf{x}) = C^{(n)}(F_t^{(n_1)}(x_1), \ldots, F_t^{(n_d)}(x_d)), \quad \text{for all } \mathbf{x} \in \mathcal{X} \in \mathbb{R}^d. \tag{5.26}$$

Moreover if $\mathbf{u} := (u_1, \ldots, u_d) \in \mathbf{U} \subset [0,1]^d$ is a d-dimensional vector in the unit hypercube, we have that

$$C^{(n)}(\mathbf{u}) = F_t^{(n)}(F_t^{-1,(n_1)}(u_1), \ldots, F_t^{-1,(n_d)}(u_d)) \tag{5.27}$$

We denote by $c^{(n)}(\mathbf{u};\theta_c)$ and $C^{(n)}(\mathbf{u},\theta_c)$ the copula density and the copula distribution function respectively, for all vectors $\mathbf{u} \in \mathbf{U} \subset [0,1]^d$ with copula parameters' set θ_c.
Then we calculate the copula density as

$$c^{(n)}(\mathbf{u};\theta_c) = \frac{P_t^{(n)}(\mathbf{x};\{\theta,\rho(\mathbf{u})\})}{P_t^{(n)}(\mathbf{x};\{\theta,0\})}, \quad \text{for } \mathbf{u} \in \mathbf{U},\ \text{for } \mathbf{x} = F^{-1,(n)}(\mathbf{u}). \tag{5.28}$$

with $t = 1$, and $\{\theta, \rho(\mathbf{U})\}$ denoting the set of the Markov chain parameters with $\rho(\mathbf{u}) : \mathbf{U} \to [-1,1]$ the local correlation function, and $P_t^{(n)}(\mathbf{x};\{\theta,0\}) = P_t^{(n)\perp}$.

Note that every coordinate vector $\mathbf{u} := (u_1, \ldots, u_d) \in \mathbf{U} \subset [0,1]^d$ may exhibit a specific mapped local correlation, $\rho(\mathbf{u}) = \rho(u_1, \ldots, u_d) \in [-1,1]$.

In particular the parameters of copula distribution function in eq. (5.27) are linked to the Markov chain distribution function by the relation

$$C^{(n)}(\mathbf{u};\theta_c) = F_1^{(n)}(\mathbf{x};\{\theta,\rho(\mathbf{u})\}). \tag{5.29}$$

Eq. (5.29) is of key importance in our approach because it allows us to link the often complex and not so intuitive copula parameters set θ_c to the local correlation parameter set $\rho(\mathbf{u})$. In this way the copula parametric dependence structure is explained and displayed through an equivalent representation by means of local Gaussian correlation coefficients.

5.2.3 Functional Copula Mapping to a Generalised Local Gaussian Copula

In this section we develop a class of general mappings that will allow us to characterise any desired copula density locally in its support according to what we denote as a generalized Gaussian copula which is achieved via a general mapping that will be detailed in this section. In addition, we will also provide an algorithm to allow for efficient implementation of this mapping result. The aim of this section is effectively to take a target copula distribution $C^{(n)}_{\text{target}}(\mathbf{u};\theta_c)$ with copula parameter θ_c and to develop a local discrete mapping from this copula over the discretized support $\mathbf{U}$, to a new generalized Guassian distribution with implicit copula that we term the generalized Gaussian copula and denote by $C^{(n)}(\mathbf{u};\{\theta,\rho(\mathbf{u})\})$. This result is important as it allows us to provide a unique characterization of the target copula at local regions of the support of the target distribution state space not in terms of θ_c but instead in terms of local sufficient statistics characterizing the generalized Gaussian copula in that local region of support, denoted by parameter vector in each local region by $\rho(\mathbf{u}) \in [-1,1]$. This automatically provides the local equivalent Gaussian correlation information for the target copula such as local correlation structure, which can characterize the copula and explain the local effect of the copula parameter in inducing dependence locally in the state space.

Given a discretized target copula distribution function $C^{(n)}_{\text{target}}(\mathbf{u};\theta_c)$ over the unit hypercube $\mathbf{U}$ its equivalent representation in a tensor space is given by

$$\min_{\rho(\mathbf{u})} \left\| C^{(n)}_{\text{target}}(\mathbf{u};\theta_c) - F^{(n)}_1(\mathbf{x};\{\theta,\rho(\mathbf{u})\}) \right\|_2, \quad \text{for } \mathbf{u} \in \mathbf{U}, \ \text{for } \mathbf{x} = F^{-1,(n)}(\mathbf{u}). \tag{5.30}$$

Eq. (5.30) means that for a given target copula distribution function C_{target} with parameters set θ_c, belonging to any copula family it is possible to find a set of local Gaussian correlations $\rho(\mathbf{u}) \in [-1,1]$, that would produce a joint distribution function F that has minimal Euclidean distance from C_{target} over all the states $\mathbf{u}$ of the unit hypercube $\mathbf{U}$. A practical way to solve eq. (5.30) is to compute the local likelihood with respect to each local Gaussian correlation parameter $\rho(\mathbf{u})$

$$\min_{\rho(\mathbf{u})} \left\| C^{(n)}_{\text{target}}(\mathbf{u};\theta_c) - C^{(n)}(\mathbf{u};\{\theta,\rho(\mathbf{u})\}) \right\|_2, \quad \text{for all } \mathbf{u} \in \mathbf{U}, \tag{5.31}$$

where C_{target} can be any target copula function. The mimicking copula distribution function $C^{(n)} = C^{(n)}(\mathbf{u};\{\theta,\rho(\mathbf{u})\})$ is locally evaluated in all vector points $\mathbf{u} = (u_1,\ldots,u_d) \in \mathbf{U}$ and its value is of a Gaussian copula function with parameter $\rho(\mathbf{u})$. This means that in a two dimensional case the local minimization problem of eq. (5.31) needs to be solved for all the grid points $\mathbf{u} \in \mathbf{U} = \bigotimes_{i=1}^{2} \mathbf{U}^i = \mathbf{U}^1 \otimes \mathbf{U}^2$, as shown in the illustrative Example 3 below.

Example 3. *Below we report in Figures 5.1 and 5.2 a first illustrative example of the copula mapping presented above.*

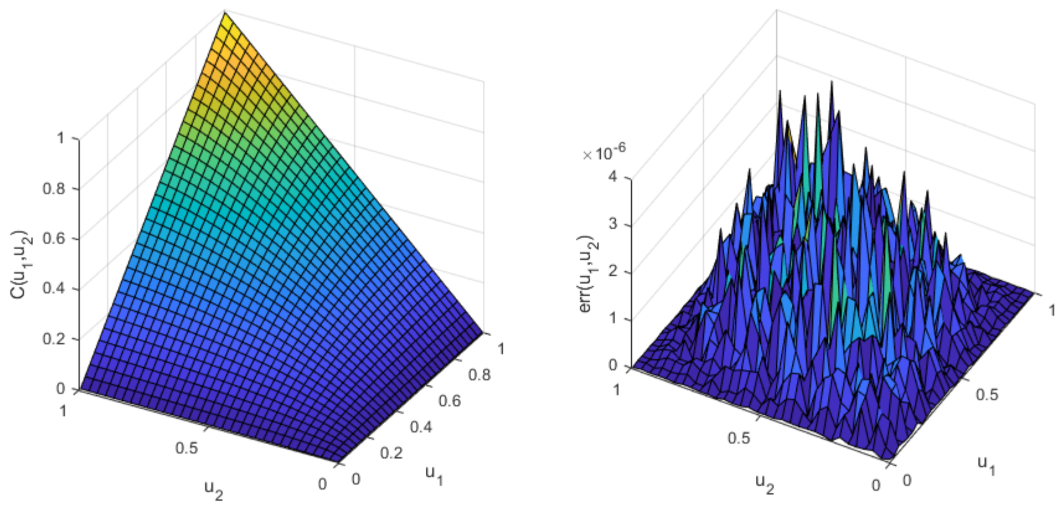

Figure 5.1: The left plot reports the copula distribution function $C^{(n)}_{\text{target}}$ belonging to Student's t-copula, evaluated on the unit square $\mathbf{U} := \bigotimes_{i=1}^{2} \mathbf{U}^i \in [0,1]^2$ and with parameters $\theta_c = \{\rho = 0.5, \nu = 5\}$. The number of states used in this example is $n_1 = n_2 = 35$. This represents the discretized target copula distribution, or the copula distribution we seek to approximate. The right end side plot reports the error from solving eq. (5.31) across the unit square $\mathbf{U}$, namely $\min_{\rho(\mathbf{u})} \left\| C^{(n)}_{\text{target}}(\mathbf{u};\theta_c) - C^{(n)}(\mathbf{u};\{\theta,\rho(\mathbf{u})\}) \right\|_2$ for all $\mathbf{u} \in \mathbf{U}$.

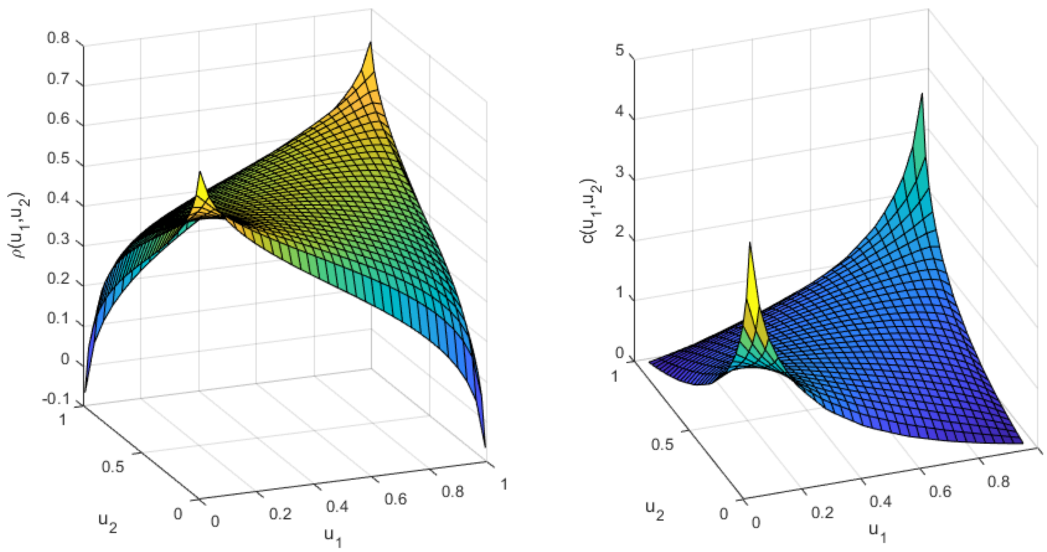

Figure 5.2: The plot on the left end side reports $\rho(\mathbf{u})$ for all $\mathbf{u} \in \mathbf{U}$ from solving eq. (5.31) which represent the local Gaussian correlations equivalent to the Student's t-copula parameters $\theta_c = \{\rho = 0.5, \nu = 5\}$. This plot is the visualization of the dependence structure of the target Student's t-copula by means of local Gaussian correlation parameters $\rho(\mathbf{u})$ for all $\mathbf{u} \in \mathbf{U}$. Each point coordinate $\mathbf{u} = (u_1, u_2)$ on the discretized copula support $\mathbf{U}$ is uniquely characterized by a Gaussian correlation value $\rho(u_1, u_2)$ which gives the magnitude of the dependency between the marginals. The plot on the right side displays the calculated Student's t-copula density $c(\mathbf{u}, \{\theta, \rho(\mathbf{u})\})$ function of the Gaussian local correlations $\rho(\mathbf{u})$ for all $\mathbf{u} \in \mathbf{U}$.

Example 4. *Below we report in Figure 5.3 a second illustrative example of the copula mapping presented above. The target distribution is the simplest possible case, a bivariate Gaussian distribution. In this case the local generalize Gaussian copula will locally approximate the target Gaussian copula and we characterize the outcome of the application of this representation in Figure 5.3.*

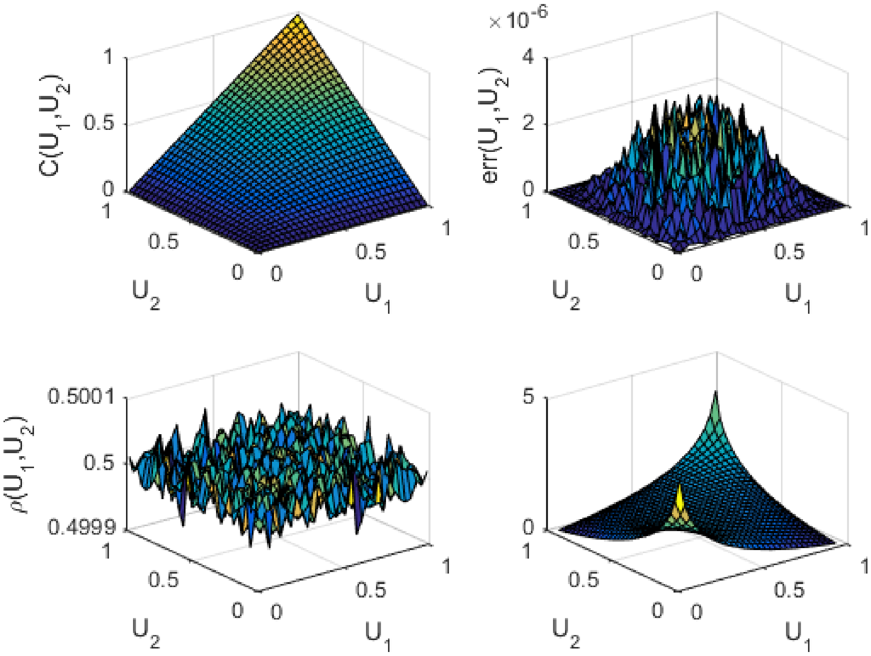

Figure 5.3: Plots of a 2-D example of local copula approximation results for a Gaussian copula with parameter $\theta_c = 0.5$. Top left plot approximated cdf of the target Gaussian copula $C^{(n)}_{\text{target}}(\mathbf{u};\theta_c)$ in eq. (5.31) with correlation coefficient $\theta = 0.5$; Top right plot is the local approximation error solving eq. (5.31) for all the pairs $\mathbf{u} = (u_1, u_2)$; Bottom left plot illustrates the obtained local Gaussian copula coefficients $\rho(\mathbf{u})$; Bottom right plot shows the calculated local Gaussian copula density using eq. (5.28). The number of states used in this example is $n_1 = n_2 = 50$ and this means that we sought for a total of $n = n_1 \times n_2 = 2500$ local correlation coefficients using eq. (5.31).

Theorem 14. *[Fréchet-Hoeffding bounds, see Embrechts et al. [2001]] If C is a d-dimesional copula, then for every $\mathbf{u} \in [0,1]^d$, the Fréchet-Hoeffding inequality is*

$$W^d(\mathbf{u}) \leq C(\mathbf{u}) \leq M^d(\mathbf{u}) \tag{5.32}$$

where the functions $W^d(\mathbf{u})$ and $M^d(\mathbf{u})$ defined on $[0,1]^d$ are as following:

$$\begin{aligned} W^d(\mathbf{u}) &= \max(u_1 + \ldots + u_d - d + 1, 0) \\ M^d(\mathbf{u}) &= \min(u_1, \ldots, u_d). \end{aligned}$$

In particular the function M^d is a d-dimensional copula for all d, however W^d is not a copula for $d > 3$.

Proposition 11 (Existence of the Minimizer of eq. (5.31))**.** *Given a target d-dimensional copula distribution C with parameters set θ_c, which we denote by $C_{target}(\mathbf{u};\theta_c)$ and a generalized local Gaussian copula $C^{(n)}$ defined on the discretized hypecube $\mathbf{U}$ we have that the local minimizer*

$$\min_{\rho(\mathbf{u})} \left\| C^{(n)}_{target}(\mathbf{u};\theta_c) - C^{(n)}(\mathbf{u};\{\theta,\rho(\mathbf{u})\}) \right\|_2 \tag{5.33}$$

always exists for all $\mathbf{u} \in \mathbf{U}$.

Proof. We observe that by the Fréchet-Hoeffding bounds theorem 14 we have that for all $\mathbf{u} \in \mathbf{U}$

$$W^d(\mathbf{u}) = C^{(n)}(\mathbf{u}; \{\theta, -1\}) \leq C^{(n)}(\mathbf{u}; \{\theta, \rho(\mathbf{u})\}) \leq C^{(n)}(\mathbf{u}; \{\theta, 1\}) = M^d(\mathbf{u}) \tag{5.34}$$

However we have that such bounds are valid for any copula and in particular

$$W^d(\mathbf{u}) \leq C_{\text{target}}(\mathbf{u}; \theta_c) \leq M^d(\mathbf{u}) \tag{5.35}$$

Therefore for any given coordinate point $\mathbf{u} \in \mathbf{U}$ it is always possible to find a correlation value $\rho(\mathbf{u}) \in [-1, 1]$ which minimizes the local minimizer of eq. (5.33). □

5.2.4 Sklar's Theorem for joint Diffusion processes approximated by CTMCs

We would like to propose the Sklar's Theorem using the CTMC approximation for correlated diffusions introduced in the previous section

Definition 27 (Sklar's Theorem in Generator Space for Bivariate CTMC Processes)**.** *Let* $P^{(n)} = P^{(n_1 n_2)}_{X^1_t, X^2_t}$ *be the transition probability which is the approximated weak solution of the diffusion*

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \Psi(\mathbf{X}_t)d\mathbf{W}_t$$

with infinitesimal generator

$$A = \sum_i^2 b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j}^2 \Sigma_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{5.36}$$

Then the approximated local copula function associated to the correlated approximated diffusion is:

$$c^{(n)} = I^{(n_1)} \otimes_S \{\exp(t\hat{A}^{(n_2)}_{X^2|X^1})\} = \frac{P^{(n)}_t}{\exp(tA^{(n_1)}_{X^1}) \otimes \exp(tA^{(n_2)}_{X^2})} \tag{5.37}$$

while the local corresponding copula approximated distribution function is given by:

$$C^{(n)} := F^{(n)}_t = F^{(n_1 n_2)}_{X^1_t, X^2_t} \approx F^{(n_2)}_{X^2_t} \otimes_S \{F^{(n_1)}_{X^1_t|X^2_t}\}. \tag{5.38}$$

Remark 6 (Orthogonality of the tensor basis)**.** *By construction the approximated local copula density function* $c^{(n)}$ *is defined on a orthogonal basis, which is the basis resulting from the tensor product of the constituent operators. In fact each operator matrix used in its calculation is orthogonal by construction, see section 5.1.3. Furthermore we construct the copula over the support* $\mathcal{X}$ *which may be represented as union of disjoint subsets* $\{B_j\}$, *i.e.* $\mathcal{X} = \bigotimes_{i=1}^d \mathcal{X}^i = \bigcup_{j=1}^n B_j$, *where each set* B_j *as coordinate vector point* $\mathbf{x}$ *in the space* $\mathcal{X}$. *If we then set* $B_0 = \{\mathbf{x} = (x_0^{(1)}, \ldots, x_0^{(d)})\}, B_1 = \{\mathbf{x} = (x_1^{(1)}, \ldots, x_0^{(d)})\}, \ldots, B_n = \{\mathbf{x} = (x_{n_1}^{(1)}, \ldots, x_{n_d}^{(d)})\}$, *therefore creating a countable and ordered sequence of sets spanning* *the whole discretized support $\mathcal{X}$ then it is straightforward to obtain the copula distribution functions as the cumulative sum over all the sets, i.e.*

$$C^{(n)}(B_z) = \sum_{j \leq z} c^{(n)}(B_j) \tag{5.39}$$

Remark 7 (Copula distribution integration in the tensor space)**.** *Integration of the mimicking local Gaussian copula $C^{(n)}$ over the unit hypercube $\mathbf{U}$ as per eq. (5.39) satisfies the third condition of the copula definition 26. Note that the space $\mathbf{U}$ is constructed from the state space $\mathcal{X}$ through the pointwise application of the marginal approximated distributions, see eq. (5.25).*

Remark 8 (Sklar's Theorem in Tensor Space)**.** *We must now complete the picture, in the bivariate setting, by linking the result just presented for Sklar's Theorem in Tensor Space for Bivariate CTMC Processes, that we denote as the functional Sklar representation for a generalized Gaussian copula, with a general bivariate copula distribution.*

One can show that this discretized state space, ie. the discrete approximation of a continuous distribution obtained via a mimicking CTMC process characterized by Sklar's representation in generator space can be made exact for a continuous distribution as the discretization interval shrinks to zero.

Proposition 12 (Convergence of CTMC Sklar's Theorem for Bivariate Copula)**.** *Let $\mathbf{X}_t := (X_t^1, X_t^2)$ be correlated Markov processes with joint approximated distribution function $F_t^{(n)} := F_{X_t^1, X_t^2}^{(n_1 n_2)}(\mathbf{x})$ as in eq. (5.38) and with marginal distributions $F_{X_t^1}$ and $F_{X_t^2}$. Let $C(u_1, u_2) : [0,1]^2 \mapsto [0,1]$ a continuous copula function. Then the following convergence result holds:*

$$\lim_{n_1, n_2 \to \infty} F_{X_t^1, X_t^2}^{(n_1 n_2)}(\mathbf{x}) = C(F_{X_t^1}(x_1), F_{X_t^2}(x_2)) \tag{5.40}$$

Definition 28 (Weak Convergence, see Varadhan [2007])**.** *A sequence $\{\alpha_n\}_{n \in \mathbb{N}}$ of probability measures on the real line $\mathbb{R}$ with distribution functions $\{F_n(x)\}_{n \in \mathbb{N}}$ is said to converge weekly to a limiting probability measure α with distribution function $F(x)$ (in symbols $\alpha_n \Rightarrow \alpha$ or $F_n \Rightarrow F$) if*

$$\lim_{n \to \infty} F_n(x) = F(x) \tag{5.41}$$

In order to prove the desired weak convergence properties of proposition 12 we recall a fundamental theorem, which is the Lévy-Cramér continuity theorem (see Theorem 2.3, pag. 25 from Varadhan [2007]) reported below.

Theorem 15 (Lévy-Cramér Continuity Theorem on $\mathbb{R}$)**.** *The following are equivalent:*

(i) $\alpha_n \Rightarrow \alpha$ *or* $F_n \Rightarrow F$.

(ii) For every bounded continuous function $f(x)$ on $\mathbb{R}$.

$$\lim_{n\to\infty}\int_{\mathbb{R}} f(x)d\alpha_n = \int_{\mathbb{R}} f(x)d\alpha$$

(iii) If $\phi_n(s)$ and $\phi(s)$ are, respectively, the characteristic functions of α_n and α for every real s,

$$\lim_{n\to\infty}\phi_n(s) = \phi(s). \tag{5.42}$$

.

Proof. See Varadhan [2007], pag 25. □

Note that the extension of the Lévy-Cramér continuity theorem to $\mathbb{R}^d$ can be proved exactly as in the one dimensional case and it suffices to show that the sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ of probability measures is tight, see Varadhan [2007]. So we limit our convergence argument to the one dimensional case.
However given a sequence of probability measures $\{\alpha_n\}_{n\in\mathbb{N}}$ on $\mathbb{R}$ it is easier to show that the moments

$$\int x^z\alpha_n(dx) \to m_z \in \mathbb{R} \ \text{ for each } z\in\mathbb{N} \tag{5.43}$$

rather than to show that the characteristic functions of $\{\alpha_n\}_{n\in\mathbb{N}}$ converge. It remains to find the conditions on the moments sequence $\{m_z\}_{z\in\mathbb{N}}$ such that there exists a unique probability measure α with $\int x^z\alpha_n(dx) = m_z \in \mathbb{R}$ for each $z\in\mathbb{N}$. Furthermore the condition on the moment sequence $\{m_z\}_{z\in\mathbb{N}}$ which would allow us to conclude that $\alpha_n \Rightarrow \alpha$ for some probability measure α can be formulated as following. If $\int x^2\alpha_n(dx) \to m_2 < \infty$ implies that $\sup_n \int x^2\alpha_n(dx) = C < \infty$, $C > 0$ and therefore for any $k > 0$,

$$\sup_n \alpha_n(|x| > k) \le \frac{C}{k^2} \tag{5.44}$$

which implies that $\{\alpha_n\}_{n\in\mathbb{N}}$ is a tight family of probability measures and therefore subsequential weak limits are guaranteed to exist.

In our specific case we want to prove that the sequence of the first two moments of the BDP approximated probability measure sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ converge to those of a Normal, and this is referred as methods of moments for proving weak convergence.

Proof. [of proposition 12] In order to prove and show some properties of convergence as presented in proposition 12 we need to first introduce the characteristic function of the mimicking local BDP

$$dY_t = h(dN_t^+ - dN_t^-) \tag{5.45}$$

We recall that if a random variable Y has moment generating function $m_Y(s)$, then $X = aY+b$ has moment generating function $m_X(s) = e^{sb}m_Y(as)$. Therefore if the characteristic function (CF) of a Poisson distribution with intensity parameter λ is:

$$\phi_{\text{Poisson}}(s) = \exp(\lambda(e^{\text{i}s}-1)) \tag{5.46}$$

consequently the CF of the BDP dY_t of eq. (5.45) with birth and death rate respectively λ_b and λ_d at time $t=1$ is

$$\phi_{\text{BDP}}(s) = \exp(\lambda_b(e^{\text{i}sh}-1))\exp(\lambda_d(e^{-\text{i}sh}-1)) \tag{5.47}$$

with moment generating function

$$m_{\text{BDP}}(s) = \exp(\lambda_b(e^{sh}-1))\exp(\lambda_d(e^{sh}-1)) \tag{5.48}$$

with

$$\begin{aligned} m'_{\text{BDP}}(s)|_{s=0} &= \lambda_b h - \lambda_d h \\ m''_{\text{BDP}}(s)|_{s=0} &= \lambda_b h^2 + \lambda_d h^2 - (\lambda_b h - \lambda_d h)^2 \end{aligned}$$

which yields to the first two moments equal to respectively

$$\begin{aligned} E[Y] &= \lambda_b h - \lambda_d h \\ Var[Y] &= \lambda_b h^2 + \lambda_d h^2 \end{aligned} \tag{5.49}$$

The moment generating function of a random variable $Y \sim \mathcal{N}(\mu, \sigma^2)$ is $m_Y(s) = \exp(\mu s + \frac{\sigma^2 s^2}{2})$ Therefore if we want the BDP to produce a distribution with first two moments identical to those of a normal random variable $\mathcal{N}(\mu, \sigma^2)$ then the expression of the intensities needs to be:

$$\lambda_b = \frac{1}{2}\Big(\frac{\mu}{h} + \frac{\sigma^2}{h^2}\Big), \qquad \lambda_d = \frac{1}{2}\Big(\frac{\sigma^2}{h^2} - \frac{\mu}{h}\Big), \tag{5.50}$$

which are in line with the expression of the BDP intensities in eq. (4.4). Now we would like to study numerically and analytically the following convergence:

$$\text{BDP CF} \rightarrow \text{Normal CF} \quad \text{for } h \rightarrow 0,$$

and specifically how the following CF

$$\phi_{\text{BDP}}(s) = \exp\Big(\frac{1}{2}(\frac{\mu}{h} + \frac{\sigma^2}{h^2})(e^{\mathrm{i}sh}-1)\Big)\exp\Big(\frac{1}{2}(\frac{\sigma^2}{h^2} - \frac{\mu}{h})(e^{-\mathrm{i}sh}-1)\Big) \tag{5.51}$$

will converge to

$$\phi_{\text{Normal}}(s) = \exp\Big(\mathrm{i}\mu s - \frac{\sigma^2 s^2}{2}\Big) \tag{5.52}$$

In order to show the convergence of the characteristic function we notice that since

$$\begin{aligned} e^{\mathrm{i}sh} - 1 &= \mathrm{i}hs - \frac{1}{2}s^2h^2 - \frac{1}{6}\mathrm{i}h^3s^3 + h^4O(s^4) \\ e^{-\mathrm{i}sh} - 1 &= -\mathrm{i}hs - \frac{1}{2}s^2h^2 + \frac{1}{6}\mathrm{i}h^3s^3 + h^4O(s^4) \end{aligned}$$

we can rewrite eq. (5.51) as

$$\begin{aligned} \phi_{\text{BDP}}(s) &= \exp\Big(\lambda_b(\mathrm{i}hs - \frac{1}{2}s^2h^2 - \frac{1}{6}\mathrm{i}h^3s^3 + h^4O(s^4)) + \lambda_d(-\mathrm{i}hs - \frac{1}{2}s^2h^2 + \frac{1}{6}\mathrm{i}h^3s^3 + h^4O(s^4))\Big) \\ &= \exp\Big((\lambda_b - \lambda_d)\mathrm{i}hs - (\lambda_b + \lambda_d)\frac{1}{2}s^2h^2 - (\lambda_b - \lambda_d)\frac{1}{6}\mathrm{i}h^3s^3 + (\lambda_b + \lambda_d)h^4O(s^4)\Big) \\ &= \exp\Big(\mathrm{i}\mu s - \sigma^2\frac{1}{2}s^2 - \mathrm{i}\frac{1}{6}\mu h^2s^3 + \sigma^2h^2O(s^4)\Big) \\ &= \exp\Big(\mathrm{i}\mu s - \sigma^2\frac{1}{2}s^2 + O(h^2)O(s^3)\Big) \end{aligned} \tag{5.53}$$

□

Therefore the BDP CF of eq. (5.51) will converge to a Normal CF as $h \to 0$ at a rate at most $O(h^2)$. Note that the first two moments will always match by construction and the error comes from the higher moments, and ultimately will converge to zero as $h \to 0$. In Figure 5.4 we show how the BDP density function obtained through fast Fourier inversion converges to the Normal density function with same parameters as $h \to 0$.

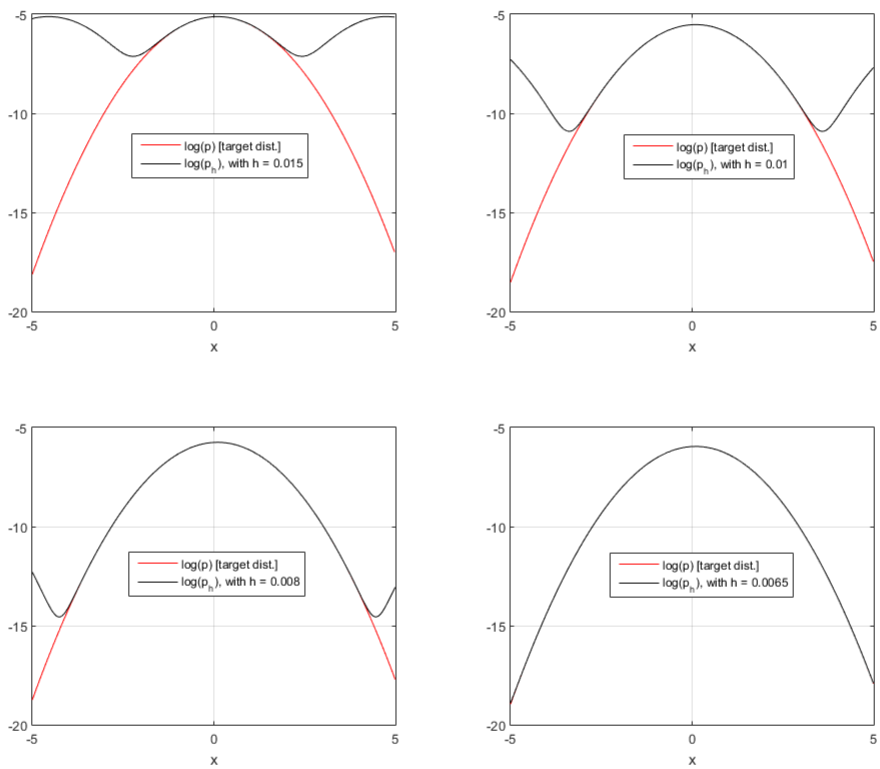

Figure 5.4: Convergence of BDP density function obtained through fast Fourier inversion to the Normal density function with same parameters as $h \to 0$.

5.2.5 Copula Marginals and Projections

In this section we discuss how to utilise the copula decompositions developed to consider marginalization and conditioning. Consider the joint probability as $P(A \cap B)$, where A and B are events on the same probability space, and express it in the most general way as a function of the marginal densities, copula function, input correlation matrix. According to Kolmogorov's definition of probability spaces every set A with non-zero probability, that is $P(A) > 0$ defines another probability measure $Q(A) = P(B|A) = \frac{P(A \cap B)}{P(A)}$ on the space. For example, if we denote the joint density of the variables pair (X, Y) by $f_{X,Y}$ with marginal densities $f(x)$ and $g(y)$ then it is possible to express the conditional density

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f(x)} = g(y)c(F(x), G(y)) \tag{5.54}$$

More generally we consider the case where each marginal distribution F_i is continuous and differentiable. If the copula C and marginals $F_1, \ldots, F_k$ are differentiable, then the join density $f(x_1, \ldots, x_k)$ corresponding to the join distribution function $F(x_1, \ldots, x_k)$ can be written by canonical represenatation as a product of the marginal densities and the copula density

$$f(x_1, \ldots, x_k) = c(F_1, \ldots, F_k) \prod_{i=1}^{k} f_i(x_i) \tag{5.55}$$

which is exactly the result of proposition 10. Taking into account the univariate marginals and joint density defined by the copula we can write the conditional density in the following form:

$$f(x_k|x_1, \ldots, x_{k-1}) = f_k(x_k) \frac{c(F_1, \ldots, F_k)}{c(F_1, \ldots, F_{k-1})} \tag{5.56}$$

From an algebraic point of view each calculated infinitesimal generator matrix $A^{(n_i)}_{X^i}$, $i = 1, \ldots, d$ defines an n_i-orthogonal basis B_i spanning the associated vector space which we denote with V_i.

The multidimensional approximated generator $A^{(n)} = A^{(n_1 \cdots n_d)}_{X^1, \ldots, X^d}$ is an element of V, the tensor product $V = V_1 \otimes \ldots \otimes V_d$. $A^{(n)}$ is a large very sparse matrix and corresponds to a linear mapping.

It is useful to be able to identify the properties of $A^{(n)}$ over local subspaces of V. In the following we provide some results on projection operator P on the space V.

Proposition 13. *Let us consider a two dimensional approximated infinitesimal generator* $A^{(n)} = A^{(n_1 n_2)}_{X^1, X^2}$ *with associated transition kernel* $P^{(n_1 n_2)}_{X^1_t, X^2_t}$, $t \geq 0$. *If we define the operator* $P_2 = (e^{A^{(n_1)}_{X^1}})^{-1} \otimes I_{n_2}$, $t = 1$, *then the projected kernel on the space of* V_2 *is*

$$P^{n_1 n_2}_{X^2_t | X^2_t} = P_2 P^{(n_1 n_2)}_{X^1_t, X^2_t} \tag{5.57}$$

We call P_2 *projection operator on* V_2 *for* $P^{(n_1 n_2)}_{X^1_t, X^2_t}$.

Proof. The generator $A^{(n_1 n_2)}_{X^1, X^2}$ can be written as

$$\begin{aligned} A^{(n_1 n_2)}_{X^1, X^2} &= A^{(n_1)}_{X^1} \oplus A^{(n_2)}_{X^2} \\ &= A^{(n_1)}_{X^1} \otimes I_{n_2} + I_{n_1} \otimes A^{(n_2)}_{X^2} \end{aligned} \tag{5.58}$$

In order to obtain the projection on V_2 it is necessary to subtract the component $A^{(n_1)}_{X^1} \otimes I_{n_2}$. The corresponding projection operator is therefore given by $e^{-\left(A^{(n_1)}_{X^1} \otimes I_{n_2}\right)}$, that is equivalent to $(e^{A^{(n_1)}_{X^1}})^{-1} \otimes I_{n_2}$. □

5.2.6 Numerical Algorithms for Evaluation of the Functional Copula Characterization

A practical way to solve eq. (5.31) is to compute the local likelihood with respect to each local Gaussian correlation $\rho(\mathbf{u}) \in [-1, 1]$ for all $\mathbf{u} \in \mathbf{U}$, and therefore obtain a full set off correlation coefficients which gives a representation of the dependency structure of the target copula as a function of Gaussian correlation. By doing this we are mimicking a target copula $C^{(n)}_{\text{target}}(\mathbf{u};\theta_c)$ which properties are specified by a set of parameters θ_c, to a local Gaussian copula characterized by unique local parameter $\rho(\mathbf{u})$ which is very intuitive and gives immediate information about the magnitude of the local dependency.

The solution to this local mapping representation can be obtained via the following algorithm.
Below we present a three steps algorithm for the computation of

$$\min_{\rho(\mathbf{u})} \left\| \log\left(C^{(n)}_{\text{target}}(\mathbf{u};\theta_c) - C^{(n)}(\mathbf{u};\{\theta,\rho(\mathbf{u})\}) \right) \right\|_2, \tag{5.59}$$

where C_{target} can be any target copula function.

1. **Step 1:** Construction of 2-D infinitesimal generator matrix by means of a conditional infinitesimal generator matrix sequence as described in Proposition 7. In this way we can calculate the approximated correlated joint distribution $P^{(n)}(\mathbf{x};\{\theta,\rho(\mathbf{u})\})$, with θ being the set of marginals' parameters and $\rho(\mathbf{u})$ the local Gaussian correlation coefficient.

2. **Step 2:** 2-D copula density computation through the copula infinitesimal generator matrix as per Proposition 10.

3. **Step 3:** Local Gaussian copula operator calibration as described in section 5.2.3 and calculation of the local Gaussian copula coefficient projection $\rho(\mathbf{u})$ using eq. (5.31).

The numerical complexity of this algorithm is at most n^2. In fact is possible to associate at most one value of local Gaussian copula correlation $\rho(\mathbf{u})$ per each coordinate point $\mathbf{u} \in \mathbf{U}$.

5.3 Assessing the Accuracy of the Functional Copula Approximations and Convergence Analysis

In this section we illustrate through examples how the local approximation of correlated diffusions by a mimicking BDP, introduced in the previous sections, can be used in practice to approximate the distribution and density functions of a desired target copula. We demonstrate the ability of our modeling framework to characterize the dependence structure of known copula functions in terms of the local correlation. Furthermore the proposed examples give evidence to the findings and converge properties presented in proposition 12. The mapping procedure, which corresponds to eq. (5.31), is norm based and seeks for the level of local correlation coefficients that can make the generalised Gaussian copula locally reproduce the dependence structure of other copula functions globally over the entire discretized unit hypercube.

Firstly we illustrate how to build the local copula generator and visualize various correlation structures that realize specific tail dependencies among the marginals. Secondly we give details of the local Gaussian correlation produced by our methodology and show how to map such local correlation to standard copula functions known in the literature.

We provide examples using both categories of copula functions: those derived from distributions, in the sense that typical multivariate distributions describe important dependence structures, like elliptical copulas. And the copulas which functional expression can be stated directly and have a quite simple form, like Archimedean copulas.

5.3.1 Functional Copula Characterizations of Elliptical Copula Families

We start this section discussing a widely used family of copula models known as the elliptical copula models. We will present the Student's t-copula model and the grouped, generalized and skewed t-copula models. Several parameterizations are available for the skewed t-copula models, see discussions in Demarta & McNeil [2005] and Cruz *et al.* [2015b]. We note the following scale mixture relationship between these multivariate distributions. They are all sub-families of copula models that characterize the dependence in the following generalized hyperbolic family of d-dimensional multivariate distributions given by the relationship:

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\gamma} W + \sqrt{W}\boldsymbol{Z}, \tag{5.60}$$

with $\boldsymbol{Z} \sim N(0, \Sigma)$ independent of $W \sim Inv.Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ and $\boldsymbol{\gamma} \in \mathbb{R}^d$ denoting a skewing vector of parameters. If $\boldsymbol{\gamma} = \mathbf{0}$ and $W = 0$ then one has the joint distribution implying a Gaussian copula case; if one has $\boldsymbol{\gamma} = \mathbf{0}$ and W included one has the Student's t-copula case; and if $\boldsymbol{\gamma} \neq \mathbf{0}$ then one has the model with an implicit skewed t-copula.

Constructing the Generalized Gaussian Copula Functional Copula for a Target Student's t-Copula

In practice, one of the most popular copulas in modeling multivariate data is perhaps the t-copula implied by the multivariate Student's t-distribution (hereafter referred to as *standard-t copula*); see Embrechts *et al.* [2001] and Demarta & McNeil [2005]. This is due to its simplicity in terms of simulation and calibration, combined with its ability to model tail dependence.

It will be useful at this stage to introduce the following notation:

- $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ is a random vector from the multivariate normal distribution $\Phi_{\mathbf{\Sigma}}(\mathbf{z})$ with zero mean vector, unit variances and correlation matrix $\mathbf{\Sigma}$.
- $\mathbf{U} = (U_1, \ldots, U_n)'$ is defined on $[0,1]^n$ domain.
- V is a random variable from the uniform (0,1) distribution independent of $\mathbf{Z}$.
- $W = G_\nu^{-1}(V)$, where $G_\nu(\cdot)$ is the distribution function of $\sqrt{\nu/S}$ with S distributed from the chi-square distribution with ν dof, i.e. W and $\mathbf{Z}$ are independent.
- $t_\nu(\cdot)$ is the standard univariate Student's t-distribution and $t_\nu^{-1}(\cdot)$ is its inverse.

We may now define the d-dimensional Student's t-copula as follows

Definition 29 (Standard Student's t-copula)**.** *The d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is said to have a non-singular multivariate Student's t-distribution with ν*

degrees of freedom, mean vector $\boldsymbol{\mu}$ *and positive definite dispersion or scatter matrix* Σ*, denoted by* $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$*, if its density is given by*

$$f(\boldsymbol{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d|\Sigma|}}\left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}} \tag{5.61}$$

Where, under this parameterization one has $\mathbb{C}\mathrm{ov}(\boldsymbol{X}) = \frac{\nu}{\nu-2}\Sigma$*. Then the random vector*

$$\mathbf{X} = W \times \mathbf{Z} \tag{5.62}$$

is distributed from a multivariate t-distribution and random vector

$$\mathbf{U} = (t_\nu(X_1), \ldots, t_\nu(X_d))' \tag{5.63}$$

is distributed from the standard t-copula where the distribution is implicitly given by

$$C^t_{\nu,P}(\boldsymbol{u}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_d)} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d|P|}}\left(1 + \frac{\boldsymbol{x}^T P^{-1}\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+d}{2}} d\boldsymbol{x} \tag{5.64}$$

where P *is the correlation matrix implied by the dispersion matrix* Σ*.*

In the case of $d = 2$ the bivariate t-distribution with ν degrees of freedom and correlation ρ,

$$C^t(u_1, u_2; \nu, \rho) = \int_{-\infty}^{\theta_\nu^{-1}(u_1)} \int_{-\infty}^{\theta_\nu^{-1}(u_2)} \frac{1}{2\pi(1-\rho)^{\frac{1}{2}}}\left(1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho)}\right)^{-\frac{\nu+2}{2}} ds dt \tag{5.65}$$

where $\theta_\nu^{-1}(u_1)$ denotes the inverse of the cdf of the standard univariate t-distribution with ν degrees of freedom. The two dependence parameters are (ν, ρ). The parameter ν controls the heaviness of the tails. For $\nu < 3$, the variance does not exist and for $\nu < 5$, the fourth moment does not exist. For $\nu \to \infty$, $C^t(u_1, u_2; \nu, \rho) \to \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)$.

In the following results we demonstrate the representation obtained for the generalized Gaussian functional copula representation of a Student's t-copula model for a range of different parameter settings. In particular we focus on the following two sets of results:

Case Study 1: in this first study we consider both a bivariate and trivariate Student's t-copula model. We fix the parameters in the bivariate case to $\theta = (\nu = 5, \theta = 0.5)$ for the degrees of freedom parameter and correlation parameter and in the trivariate case to $\theta = (\nu = 5, \theta_1 = 0.4, \theta_2 = 0.4)$. We perform a reconstruction of the resulting Student's t-copula model in terms of the generalized Gaussian functional copula representation developed in this chapter. In the process we aim to study the influence of the stencil mesh spacing, so we consider two illustrative settings one in which we use just a small number of points $n_1 = n_2 = 50$ and then a second in the trivariate case with $n_1 = n_2 = n_3 = 200$ points. The aim of the study is to demonstrate that a very accurate reconstruction can be achieve with minimal computational effort that yields a detailed reconstruction of the local correlation in the state-space from the functional copula mapping undertaken. The results of this analysis are presented in Figure 5.5.

Case Study 2: in this second study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the Student's t-copula model. In particular, we will construct the local correlation surface obtained from the generalized Gaussian functional copula mapping and plot a range of such surfaces over the support of the bivariate Student's t-copula as we vary the copula parameters θ. The results of this study are demonstrated in Figure 5.6.

In Figure 5.5 we show the results of the implementation of the approximation for the copula generator representation given by proposition 10 which we used to obtain the approximated density function $c^{(n_1,n_2)}$.

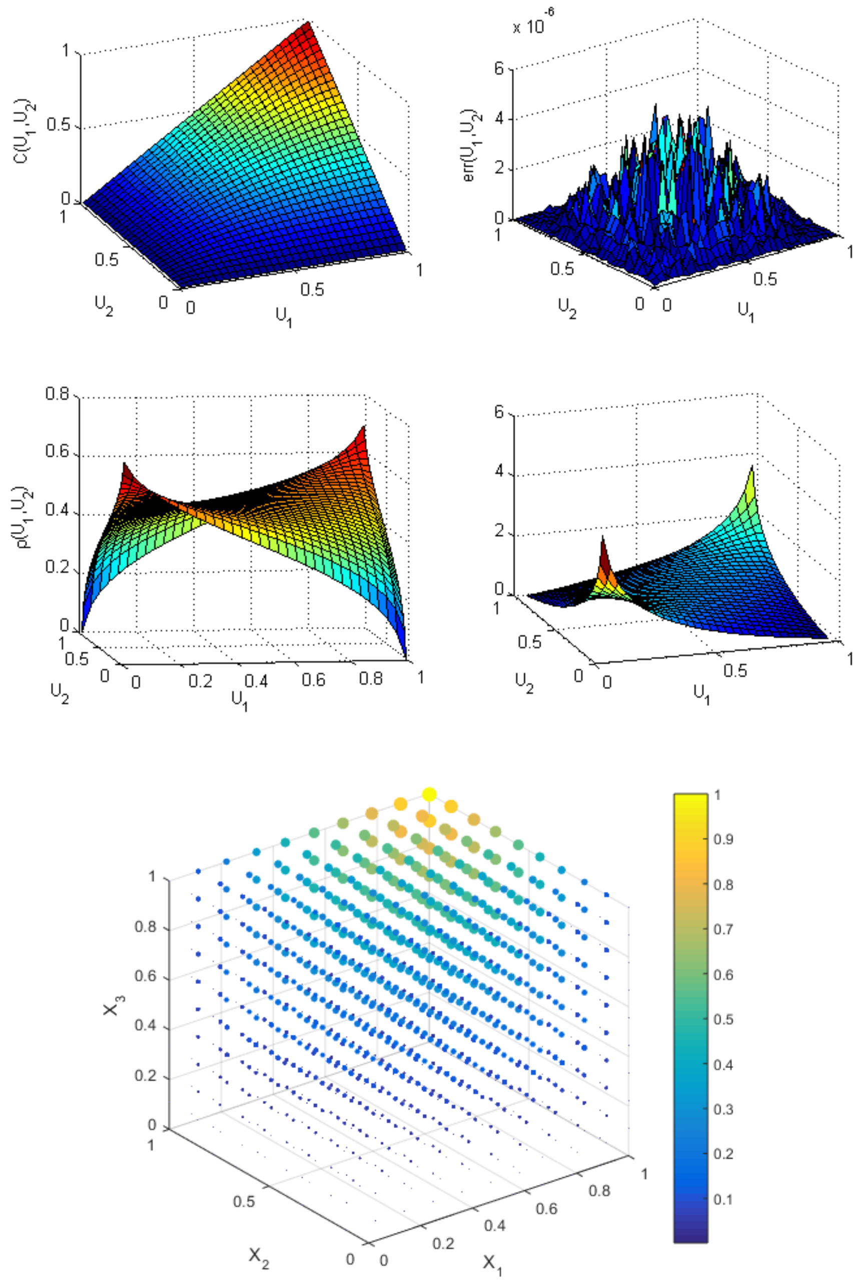

Figure 5.5: Top plots 2-D example of local approximation results for a Student's t-copula with parameters $\theta = 0.5$ and $\nu = 5$. Top left plot approximated cdf of the target Student's t-copula with correlation coefficient $\theta = 0.5$ and $\nu = 5$; Top right plot is the local approximation error of the decomposition method; Bottom left plot is the local Gaussian copula coefficients $\rho(u_1, u_2)$; Bottom right plot is approximated t-copula pdf. The number of discretization points is $n_1 = n_2 = 50$.
Bottom plot displays the approximated 3-D Student's t-copula cumulative density function with parameters $\rho = 0.4$ and $\nu = 0.5$ and number of states $n_1 = n_2 = n_3 = 200$.

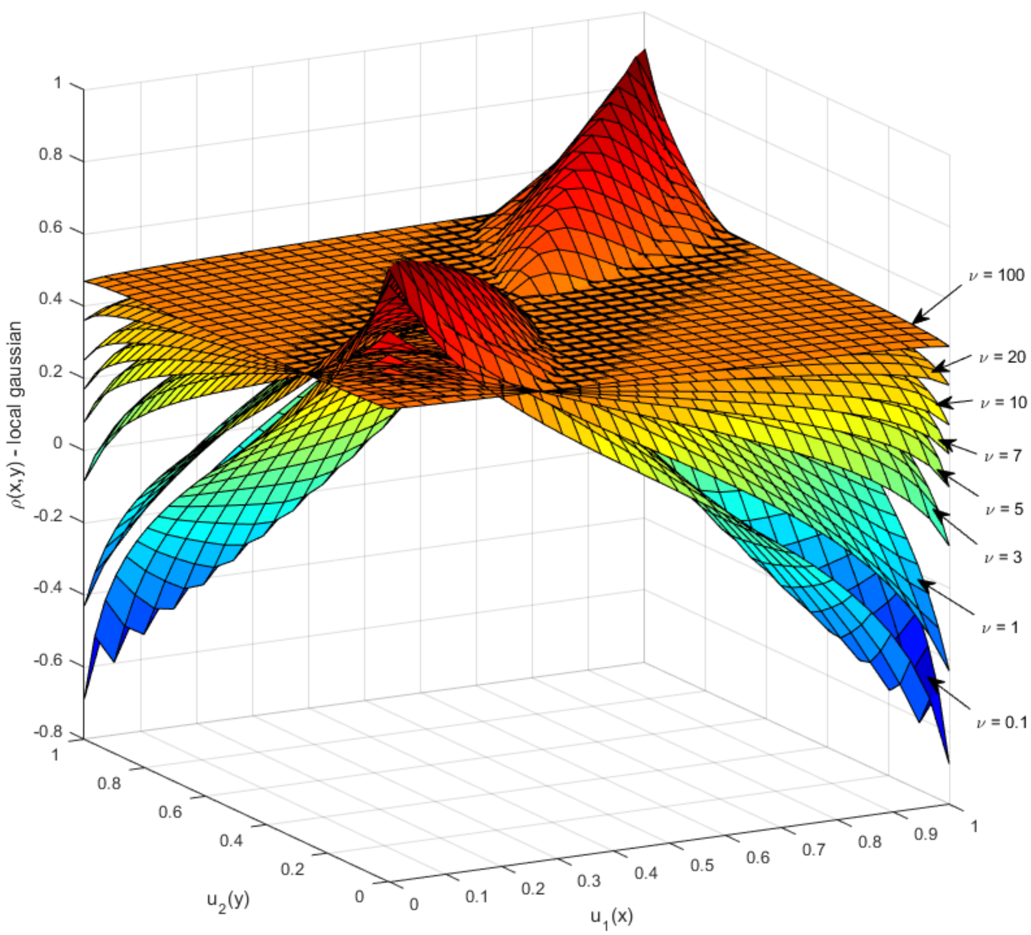

Figure 5.6: This set of local Gaussian copula correlation coefficients $\rho(x, y) := \rho(u_1, u_2)$ surfaces is the result of the copula mapping where the target copula is a 2-D Student's t-copula with parameter $\theta = 0.5$ and degree of freedom parameter $\nu = [0.1, 1, 3, 5, 7, 10, 20, 100]$. With the copula mapping it is possible to visualize though the local value of the Gaussian correlation $\rho(u_1, u_2)$ how the parameter ν controls the heaviness of the Student's t-distribution tails . When $\nu \to 0$ the Student's t-distribution exhibits the strongest tail dependence, and this is visible in the above plot where the local Gaussian correlation coefficient $\rho(u_1, u_2)$ diverges from $\theta = 0.5$ to almost $\rho(u_1, u_2) = 0.9$. For $\nu = 20$ the tail dependence starts to exhibit and increases for smaller values of ν. On the contrary when $\nu \to \infty$ the Student's t-distribution converges to a Gaussian distribution and this is illustrated by the fact that the local Gaussian correlation resulting from the copula mapping procedure is equal to $\rho(u_1, u_2) = 0.5$ across all the unit square when $\nu = 100$. The number of discretization points used in this example is $n_1 = n_2 = 50$.

Constructing the Generalized Gaussian Copula Functional Copula for Grouped and Generalized Target Student's t-Copulas

However, the standard t-copula is often criticized due to the restriction of having only one parameter for the degrees of freedom (dof), which may limit its ability to model tail dependence in multivariate case. To overcome this problem, Daul *et al.* [2003] proposed the use of the *grouped t-copula*, where each of the variables are grouped into classes and each class has its own standard t-copula with a specific dof. As a natural extension to address this concern, recently, the grouped t-copula was developed and can be specified below.

Definition 30 (Grouped Student's t-copula)**.** *Partition $\{1, 2, \ldots, n\}$ into m non-overlapping sub-groups of sizes $n_1, \ldots, n_m$. Then the copula of the random vector*

$$\mathbf{X} = (W_1 Z_1, \ldots, W_1 Z_{n_1}, W_2 Z_{n_1+1}, \ldots, W_2 Z_{n_1+n_2}, \ldots, W_m Z_n)', \tag{5.66}$$

where $W_k = G^{-1}_{\nu_k}(V)$, $k = 1, \ldots, m$, is the grouped t-copula. That is,

$$\mathbf{U} = (t_{\nu_1}(X_1), \ldots, t_{\nu_1}(X_{n_1}), t_{\nu_2}(X_{n_1+1}), \ldots, t_{\nu_2}(X_{n_1+n_2}), \ldots, t_{\nu_m}(X_n))' \tag{5.67}$$

is a random vector from the grouped t-copula. Here, the copula for each group is a standard t-copula with its own dof parameter (i.e. ν_k is dof parameter of the standard t-copula for the k-th group).

Recent extensions to the grouped t-copula have also been developed, where the generalized t-copula was specified with multiple dof parameters (hereafter referred to as *generalized Student's t-copula*); see Luo & Shevchenko [2010] and Venter *et al.* [2007]. This copula can be viewed as a grouped t-copula with each group having only one member. It has the advantages of a grouped t-copula with flexible modelling of multivariate dependences, yet at the same time it overcomes the difficulties with *a priori* choice of groups.

For convenience, denote the new copula as $\tilde{t}_{\boldsymbol{\nu}}$-copula, where $\boldsymbol{\nu} = (\nu_1, ..., \nu_d)$ denotes the vector of dof parameters and d is the number of dimensions. Luo and Shevchenko (2010) demonstrated that some characteristics of this new copula in the bivariate case are quite different from those of the standard t-copula. For example, the copula is not exchangeable if $\nu_1 \neq \nu_2$ and tail dependence implied by the $\tilde{t}_{\boldsymbol{\nu}}$-copula depends on both dof parameters.

Definition 31 (Generalized t-copula with multiple dof ($\tilde{t}_{\boldsymbol{\nu}}$-copula))**.** *Consider the grouped t-copula where each group has a single member. In this case the copula of the random vector*

$$\mathbf{X} = (W_1 Z_1,\ W_2 Z_2, \ldots, W_n Z_n)' \tag{5.68}$$

is said to have a t-copula with multiple dof parameters $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$, which we denote as $\tilde{t}_{\boldsymbol{\nu}}$-copula. That is,

$$\mathbf{U} = (t_{\nu_1}(X_1), t_{\nu_2}(X_2), \ldots, t_{\nu_n}(X_n))' \tag{5.69}$$

is a random vector distributed according to $\tilde{t}_{\boldsymbol{\nu}}$-copula. Note, all W_i are perfectly dependent.

From the above definitions, it is easy to show that the $\tilde{t}_{\boldsymbol{\nu}}$-copula distribution has the following explicit integral expression

$$C^{\boldsymbol{\Sigma}}_{\boldsymbol{\nu}}(\mathbf{u}) = \int_0^1 \Phi_{\boldsymbol{\Sigma}}(z_1(u_1, s), \ldots, z_n(u_n, s))ds \tag{5.70}$$

and its density is

$$c_{\boldsymbol{\nu}}^{\Sigma}(\mathbf{u}) = \frac{\partial^n C_{\boldsymbol{\nu}}^{\boldsymbol{\Sigma}}(\mathbf{u})}{\partial u_1 \dots \partial u_n} = \int_0^1 \varphi_{\boldsymbol{\Sigma}}\left(z_1(u_1,s),\dots,z_n(u_n,s)\right) \prod_{i=1}^{n} [w_i(s)]^{-1} ds \Big/ \prod_{i=1}^{n} f_{\nu_i}(x_i). \quad (5.71)$$

Here:

- $z_i(u_i,s) = t_{\nu_i}^{-1}(u_i)/w_i(s),\ i = 1,2,\dots,n;$
- $w_i(s) = G_{\nu_i}^{-1}(s);$
- $\varphi_{\boldsymbol{\Sigma}}(z_1,\dots,z_n) = \exp(-\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z})/[(2\pi)^{n/2}(\det\boldsymbol{\Sigma})^{1/2}]$ is the multivariate normal density;
- $x_i = t_{\nu_i}^{-1}(u_i),\ i = 1,2,\dots,n;$
- $f_\nu(x) = \left(1 + x^2/\nu\right)^{-(\nu+1)/2}\Gamma((\nu+1)/2)/[\Gamma(\nu/2)\sqrt{\nu\pi}]$ is the univariate t-density, where $\Gamma(\cdot)$ is a gamma function.

If all the dof parameters are equal, i.e. $\nu_1 = \dots = \nu_n = \nu$, then it is easy to show that the copula defined by eq. (5.70) becomes the standard t-copula; see Luo and Shevchenko (2010) for a proof.

In the following results we demonstrate the representation obtained for the generalized Gaussian functional copula representation of a generalized Student's t-copula model for a range of different parameter settings.

Case Study 1: in this study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the generalize Student's t-copula model. In particular, we will construct the local correlation surface obtained from the generalized Gaussian functional copula mapping and plot a range of such surfaces over the support of the bivariate generalized Student's t-copula as we vary the copula parameters θ. The results of this study are demonstrated in Figure 5.7.

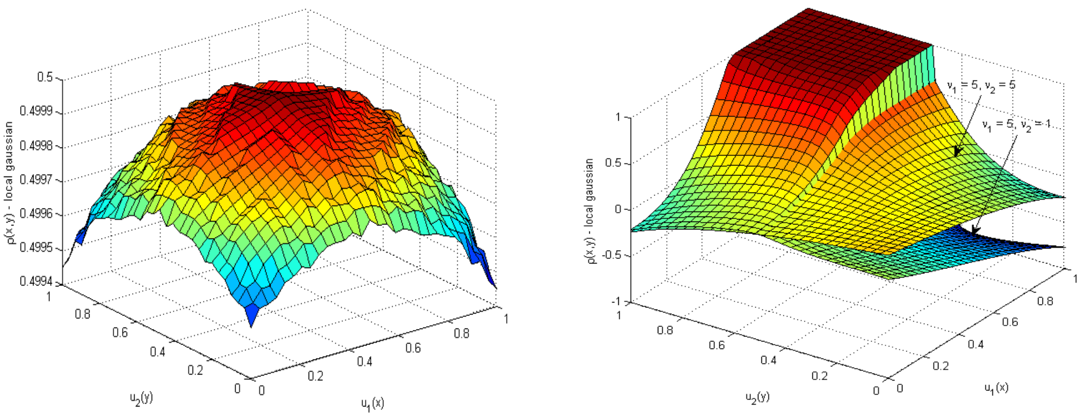

Figure 5.7: Local Gaussian copula coefficients $\rho(x,y)$ for the local approximation of a Generalized Student's t-copula. Right plot has parameters $\theta = 0.5$ and $\nu_1 = \nu_2 = 200$; Left plot has parameters $\theta = 0.5$ and $\nu_1 = 5$ and $\nu_2 = \{5, 1\}$. The number of discretization points is $n_1 = n_2 = 50$.

Constructing the Generalized Gaussian Copula Functional Copula for a Target Skewed Student's t-Copula

The skewed t-copula model is characterized relative to the Student's t-copula model specified above. As observed it corresponds to the skewed infinite mixture between a transform of a Gaussian random vector and an independent Inverse Gamma random variable, as discussed in eq. (5.60) above. The covariance matrix for this parameterization of the skewed t-copula, denoted by $\widetilde{\Sigma}$, can be obtained relative to the Student's t-copula covariance matrix, denoted by Σ, as follows (for $\nu/2 > 2$):

$$\begin{aligned}\widetilde{\Sigma} &= \boldsymbol{\gamma}\boldsymbol{\gamma}^T \mathbb{V}\mathrm{ar}\,[W] + \mathbb{E}[W]\Sigma \\ &= \boldsymbol{\gamma}\boldsymbol{\gamma}^T \left(\frac{(\nu/2)^2}{(\nu/2-1)^2(\nu/2-2)}\right) + \left(\frac{(\nu/2)}{(\nu/2-1)}\right)\Sigma\end{aligned} \tag{5.72}$$

where the shape and scale of the mixing Inverse Gamma distribution are given by $\nu/2$. In this parameterization one can obtain the d-dimensional multivariate skewed t-distribution according to the density, see ,

$$f_X(\boldsymbol{x}) = \frac{cK_{(\nu+2)/2}\left(\sqrt{(\nu + Q(\boldsymbol{x}))\boldsymbol{\gamma}^T\widetilde{\Sigma}^{-1}\boldsymbol{\gamma}}\right)\exp\left((\boldsymbol{x}-\boldsymbol{\mu})^T\widetilde{\Sigma}^{-1}\boldsymbol{\gamma}\right)}{\left(\sqrt{(\nu + Q(\boldsymbol{x}))\boldsymbol{\gamma}^T\widetilde{\Sigma}^{-1}\boldsymbol{\gamma}}\right)^{-(\nu+d)/2}(1+Q(\boldsymbol{x})/\nu)^{(\nu+2)/2}} \tag{5.73}$$

with the following functions

$$\begin{aligned}Q(\boldsymbol{x}) &= (\boldsymbol{x}-\boldsymbol{\mu})^T\widetilde{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}), \\ c &= \frac{2^{1-(\nu+d)/2}}{\Gamma(\nu/2)(\pi\nu)^{d/2}|\widetilde{\Sigma}|^{1/2}}.\end{aligned} \tag{5.74}$$

Hence, the resulting implicit copula density function for this skewed t-distribution is given by the ratio of this joint density over the marginal univariate skewed t-densities which are given by

$$f_{X_i}(x) = \frac{cK_{(\nu+1)/2}\left(\sqrt{\left(\nu + \frac{(x-\mu_i)^2}{\widetilde{\sigma}_i^2}\right)\frac{\gamma_i^2}{\widetilde{\sigma}_i^2}}\right)\exp\left((x-\mu_i)\frac{\gamma_i}{\widetilde{\sigma}_i^2}\right)}{\left(\sqrt{\left(\nu + \frac{(x-\mu_i)^2}{\widetilde{\sigma}_i^2}\right)\frac{\gamma_i^2}{\widetilde{\sigma}_i^2}}\right)^{-(\nu+1)/2}(1+\frac{(x-\mu_i)^2}{\nu\widetilde{\sigma}_i^2})^{(\nu+1)/2}}. \tag{5.75}$$

In the following results we show the representation obtained for the generalized local Gaussian copula mapping of a skewed Student's t-copula model for a range of different parameters settings.

Case Study 1: in this study we demonstrate how the marginal distribution of the skewed Student's t-copula can be obtained from the reconstructions developed. In particular, we can study how accurate the marginal distributions are captured by the bivariate generalized Gaussian functional copula reconstruction developed. The results are demonstrated for a very efficient case with just a small mesh where $n_1 = n_2 = 50$. It is clear that the marginals are very accurately reconstructed, we show this for a range of different parameter settings as we vary the skew parameters, for the skewed Student's t-copula model. The results of this analysis are presented in Figure 5.8.

Case Study 2: in this second study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the skewed Student's t-copula model. In particular, we will construct the copula density obtained from the generalized Gaussian functional copula mapping and plot a range of such surfaces over the support of the bivariate skewed Student's t-copula as we vary the copula parameters θ. The results of this study are demonstrated in Figure 5.9.

Case Study 3: in this third study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the skewed Student's t-copula model. In particular, we will construct the local correlation surfaces obtained from the generalized Gaussian functional copula mapping and plot a range of such surfaces over the support of the bivariate skewed Student's t-copula as we vary the copula parameters θ. The results of this study are demonstrated in Figure 5.10.

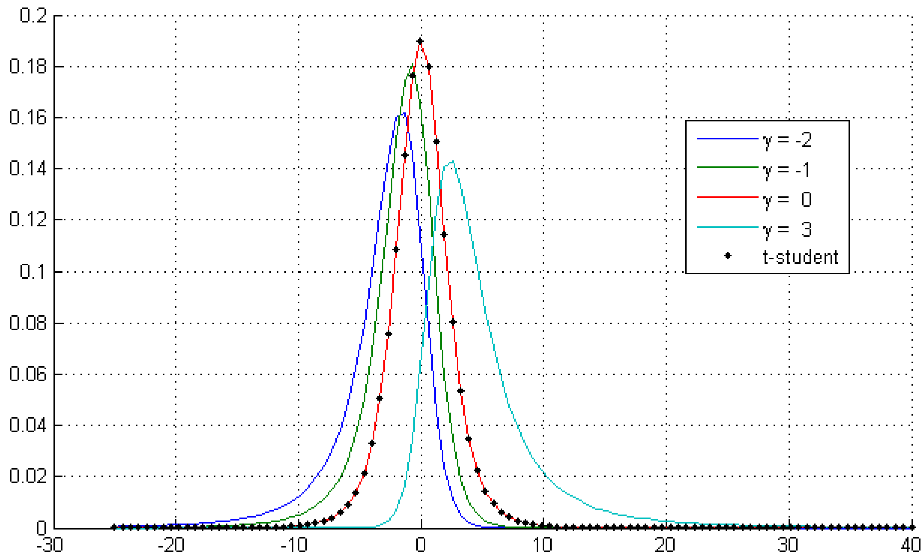

Figure 5.8: Approximation of the marginal 1-D density for a skewed Student's t-distribution with parameters $\mu = 0, \sigma = 3, \nu = 5$. The parameter γ takes values $(-2, -1, 0, 3)$. When $\gamma = 0$ then the skewed Student's t-distribution simplifies to the Student's t-distribution. The number of discretization points is $n_1 = n_2 = 50$.

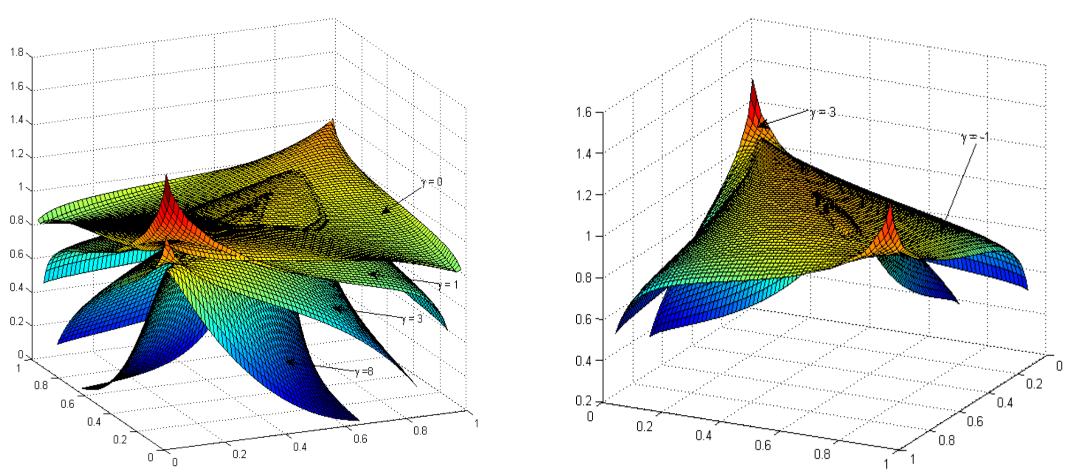

Figure 5.9: Approximated skewed Student's t-copula density with parameters $\mu = 0, \sigma = 1, \nu = 2, \rho = 0.1$. Left plot has parameter γ taking values $(0, 1, 3, 8)$; Right plot has parameter γ taking values $(-1, 3)$. When $\gamma = 0$ then the skewed Student's t-copula simplifies to the Student's t-copula density function. The number of discretization points is $n_1 = n_2 = 50$.

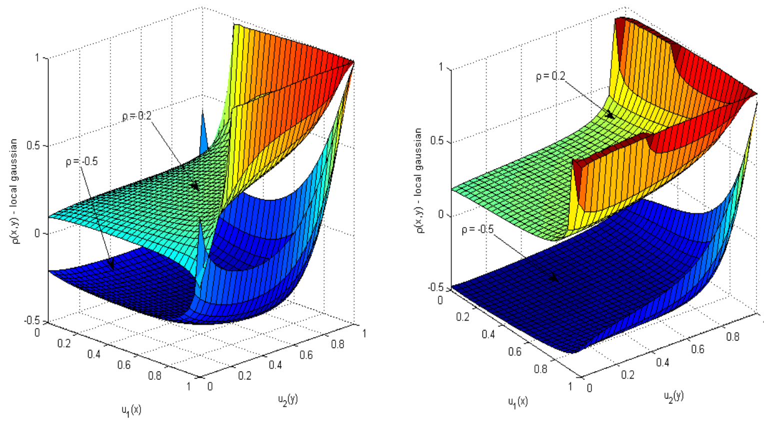

Figure 5.10: Local Gaussian copula coefficients $\rho(x, y)$ for the local approximation of a skewed Student's t-copula. Left plot has parameters $\{\rho = 0.2, g = -1, \nu = 19\}$ and $\{\rho = -0.5, g = -1, \nu = 19\}$ respectively; Right plot has parameters $\{\rho = 0.2, g = 1, \nu = 19\}$ and $\{\rho = -0.5, g = 1, \nu = 19\}$ respectively. The Gaussian local correlation values reflect the tail dependence structure of the target skewed Student's t-copula. The number of discretization points used is $n_1 = n_2 = 50$.

5.3.2 Functional Copula Characterizations of Archimedean and Distorted Archimedean Copula Families

Another family of copula model that is widely used in practice is the family of Archimedean copula models and the distorted extensions. We will consider the two most popular subfamilies of such copula models for practitioners, the Clayton and Gumbel copula models before presenting some interesting recent developments based on Archimax copula involving combinations of extreme value copula with Archimedean copula.

The first introduction to Archimedean copulas involves a specification in general via generator functions where a d-dimensional Archimedean copula model is given by

$$C(\mathbf{u}) = C\left(u_1, \ldots, u_d\right) = \psi\left(\psi^{-1}\left(u_1\right) + \cdots + \psi^{-1}\left(u_d\right)\right), \tag{5.76}$$

where ψ is a decreasing function known as the generator for the given copula; see Frees & Valdez [1998]. To fully define an Archimedean copula one considers a generator function whcih is at least d-monotone.

Definition 32. *[Completely Monotone Generators and Existence of Archimedean Copulae] If a generator ψ that is a mapping $\psi : [0, \infty] \mapsto [0, 1]$ is continuous and strictly decreasing such that $\psi(0) = 1$ and $\psi(\infty) = 0$, that is, $\psi \in C^{\infty}(0, \infty)$ and one has that $(-1)^k \psi^{(k)}(x) \geq 1$ for $k = 1, \ldots$ then this class of generators can create Archimedean copulae models in any dimension. This class of completely monotone generators for Archimedean copula in any dimension are denoted by ψ_{∞}.*

The requirement for complete monotonicity is only required to create copula of any dimension, so this was then further relaxed for d-variate Archimedean copula in further studies to include only the positivity of derivatives for $k = 1, 2, \ldots, d$ for a d-variate Archimedean copula; see discussion in McNeil & Nešlehová [2009], where it was shown that one only requires the necessary and sufficient conditions on the generator function to be a d-monotone function as given in Definition 33 in order to create Archimedean copulae models up to dimension d.

Definition 33. *[D-Monotone Functions] A real function $g(\cdot)$ is d-monotone in a range (a, b) for $a, b \in \mathbb{R}$ and $d \geq 2$ if it is differentiable on this range up to order $d-2$ and the derivatives satisfy the condition that*

$$(-1)^k g^{(k)}(x) \geq 0, \quad k = 0, 1, \ldots, d-2 \tag{5.77}$$

for any $x \in (a, b)$ and $(-1)^{d-2} g^{(d-2)}$ is nonincreasing and convex in (a, b).

One can then conclude that a function φ is said to generate an Archimedean copula if it satisfies the following properties.

Definition 34. *[Archimedean Generator] An Archimedean generator is a continuous, decreasing function $\varphi : [0, \infty] \to [0, 1]$ that satisfies the following conditions:*

1. *$\varphi : [0, \infty) \mapsto [0, 1]$ with $\varphi(0) = 1$ and $\lim_{t\to\infty} \varphi(t) = 0$;*
2. *φ is a continuous function;*
3. *φ^{-1} is given by $\varphi^{-1}(t) = \inf\{u : \varphi(u) \leq t\}$;*
4. *φ is strictly decreasing on $[0, \inf\{t : \varphi(t) = 0\}] = \left[0, \varphi^{-1}(0)\right]$.*

Constructing the Generalized Gaussian Copula Functional Copula for Target Clayton and Gumbel Copulas

The Clayton copula has a generator given by $\varphi_\theta(u) = \frac{1}{\theta}(u^{-\theta} - 1)$ and the density in the $d = 2$ case therefore takes the form,

$$C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{5.78}$$

with the dependence parameter θ restricted on the region $(0, \infty)$. As θ approaches zero, the marginals become independent. As θ approaches infinity, the copula attains the Fréchet upper bound, but for no value does it attain the Fréchet lower bound. The Clayton copula cannot account for negative dependence. It has been used to study correlated risks because it exhibits strong left tail dependence and relatively weak right tail dependence.

The density of this copula is given by

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2; \theta)}{\partial u_1 \partial u_2} = (1 + \theta)(u_1 u_2)^{-(1+\theta)}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}-2} \tag{5.79}$$

For the bidimensional Clayton copula there exist a relationship between the copula parameter θ and the concordance coefficient Kendall's tau ρ_τ, given by the following estimator for θ

$$\hat{\theta} = \frac{2\hat{\rho}_\tau}{1 - \hat{\rho}_\tau} \tag{5.80}$$

The d-dimensional Clayton copula can be defined as following (see McNeil *et al.* [2015])

$$C(u_1, \ldots, u_d; \theta) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1\right)^{-\frac{1}{\theta}} \tag{5.81}$$

The Gumbel copula has a generator given by $[-\log(u)]^{\theta}$ and the density then takes the form,

$$C(u_1, u_2; \theta) = \exp\left(-\left[(-\log(u_1))^{-\theta} + (-\log(u_2))^{-\theta}\right]\right)^{\frac{1}{\theta}} \tag{5.82}$$

The dependence parameter is restricted to the interval $[1, \infty)$. Values of 1 and ∞ correspond to independence and the Fréchet upper bound, but this copula does not attain the Fréchet lower bound for any value of θ. Similar to the Clayton copula, Gumbel does not allow negative dependence, but it contrast to Clayton, Gumbel exhibits strong right tail dependence and relatively weak left tail dependence. If outcomes are known to be strongly correlated at high values but less correlated at low values, then the Gumbel copula is an appropriate choice. For the bidimensional Gumbel copula there exist a relationship between the copula parameter θ and the concordance coefficient Kendall's tau ρ_{τ}, given by the following estimator for θ

$$\hat{\theta} = \frac{1}{1 - \hat{\rho}_{\tau}} \tag{5.83}$$

The d-dimensional Clayton copula can be defined as following (see McNeil *et al.* [2015])

$$C(u_1, \ldots, u_d; \theta) = \exp\left\{\left(\sum_{i=1}^{d}(-\log u_i)^{\theta}\right)^{-\frac{1}{\theta}}\right\}. \tag{5.84}$$

Case Study 1: in this first study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the Clayton copula model. In particular, we will provide four summary results: the reconstruction of the copula density and distribution obtained from the generalized Gaussian functional copula mapping; the accuracy of the reconstruction; and the local correlation surface obtained from the generalized functional copula mapping. The results of this study are demonstrated in left panel of sub-plots in Figure 5.11.

Case Study 2: in this second study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the Gumbel copula model. In particular, we will provide four summary results: the reconstruction of the copula density and distribution obtained from the generalized Gaussian functional copula mapping; the accuracy of the reconstruction; and the local correlation surface obtained from the generalized functional copula mapping. The results of this study are demonstrated in right panel of sub-plots in Figure 5.11.

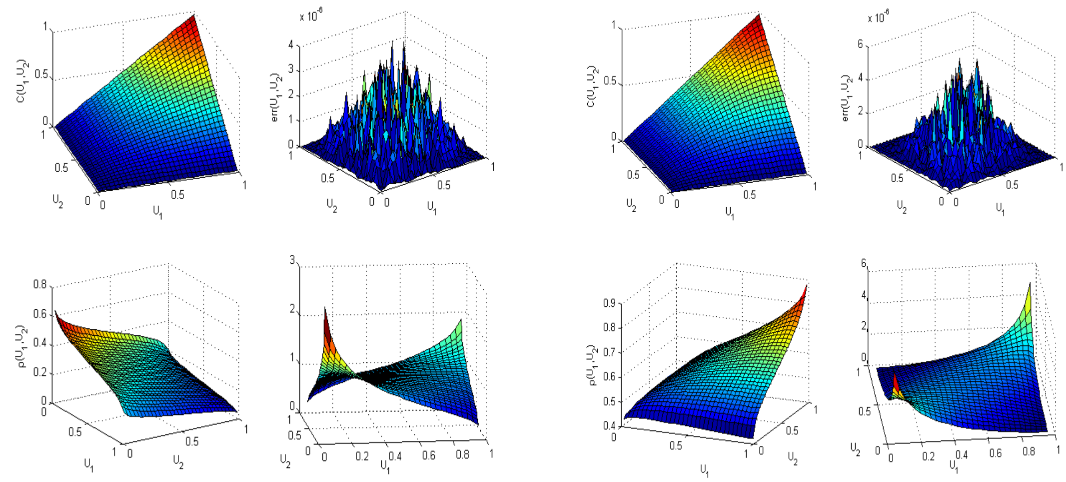

Figure 5.11: Local approximations for a Clayton copula with parameters $\theta = 0.4$ (left plots) and Gumbel copula with parameters $\theta = 1.4$ (right plots). Top left subplots are approximated cdf of the target copula; Top right subplots are the local approximation error; Bottom left subplots are the local Gaussian copula coefficients $\rho(x, y)$; Bottom right subplots are the approximated density. The number of discretization points is $n_1 = n_2 = 50$.

Constructing the Generalized Gaussian Copula Functional Copula for Target Archimax Copula

Recently there has been a growing interest in developing archimedean copula models with distortion features, based on the works of Genest & Rivest [2001] and Morillas [2005] which explore ways of distorting a given copula to obtain a new copula with additional features. For instance, they explored the multivariate probability integral transform and its application in distorting existing copula models to obtain new copula models.

For instance in Morillas [2005] they study under what conditions the following distortion copula transform produces a valid copula where $g(\cdot)$ is assumed to be a strictly increasing and continous function from $[0, 1]$ to $[0, 1]$ such that

$$C_g\left(u_1, \ldots, u_d\right) = g^{-1}\left(C\left(g(u_1), \ldots, g(u_d)\right)\right) \tag{5.85}$$

is a valid distorted copula.

Definition 35. *Distorted Copula Define the function g to be some distortion function, such that $g : [0, 1] \mapsto [0, 1]$ and is defined according to*

$$g(t) = \exp[-\varphi(t)], \tag{5.86}$$

where φ is for instance an Archimedean generator function. Now denoted C as a based copula that is to be distorted to create a new copula, then

$$C_g(u_1, ..., u_d) = g^{-1}\left(C(g(u_1), , g(u_d))\right) \tag{5.87}$$

is a copula known as the distortion of C.

Several examples of bivariate and multivariate distorted copula models have begun to be studied. Though the current emphasis has focused on their specification and little is know about their properties such as tail dependence features and other concordance measure features they may exhibit.

Here we consider two examples based on ideas developed in Charpentier *et al.* [2014]. We begin with a bivariate archimax copula given by a parameteric model of the distributional form

$$C_{\phi,\mathcal{A}}\left(u_1,u_2\right)=\varphi\left[\left\{\varphi^{-1}\left(u_1\right)+\varphi^{-1}\left(u_2\right)\right\}\mathcal{A}\left\{\frac{\varphi^{-1}\left(u_1\right)}{\varphi^{-1}\left(u_1\right)+\varphi^{-1}\left(u_2\right)}\right\}\right] \tag{5.88}$$

where $\mathcal{A}:[0,1]\to[1/2,1]$ and $\varphi:[0,\infty)\to[0,1]$ such that

1. $\mathcal{A}$ is convex and for all $t\in[0,1]$ one has $\max(t,1-t)\leq\mathcal{A}(t)\leq 1$.
2. φ is convex, decreasing and such that $\varphi(0)=1$ and $\lim_{x\to\infty}\varphi(x)=0$ with convention that $\varphi^{-1}(0)=\inf\{x\geq 0:\varphi(x)=0\}$.

Two special cases arise from this model:

- If $\mathcal{A}=1$ one recovers the well known family of Archimedean copula dependence models.
- If $\varphi(t)=\exp(-t)$ then one recovers the extreme-value copula.

As in Capéraà *et al.* [2000] we utilise in this study the choice of function

$$\mathcal{A}(t)=\left\{t^{1/\alpha}+(1-t)^{1/\alpha}\right\}^{\alpha}A^{\alpha}\left\{\frac{t^{1/\alpha}}{t^{1/\alpha}+(1-t)^{1/\alpha}}\right\}. \tag{5.89}$$

where $A(\cdot)$ is the Pickands EVT dependence function given by

$$A(t)=1-\min\{\beta t,\alpha(1-t)\} \tag{5.90}$$

for some parameters $\alpha,\beta\in[0,1]$.

Then we consider a d-variate distortion copula in the Archimax family, examples of such extensions include the works of Bacigál & Mesiar [2012], Mesiar & Jágr [2013] and Charpentier *et al.* [2014]. One possible version of such a d-dimensional Archimax copula is defined through the use of a distortion function based on the stable tail function that must satisfy certain properties described in detail in Charpentier *et al.* [2014]. In general a multivariate stable tail function is obtained via the multivariate Generalized Extreme Value distribution G via

$$-\log G(x_1,\ldots,x_d)=\mu([\mathbf{0},\infty)[\mathbf{0},\mathbf{x}]),\qquad\forall\mathbf{x}\in\mathbb{R}_+^d \tag{5.91}$$

such that G is the limiting distribution (max domain of attraction) of the normalized component wise maxima of

$$\boldsymbol{X}_{n:n}=\left(\max\{X_{1,i}\},\ldots,\max\{X_{d,i}\}\right) \tag{5.92}$$

and then the stable tail function is obtained via measure μ or distribution G according to the following

$$l(x_1,\ldots,x_d)\quad=\mu([\mathbf{0},\infty)[\mathbf{0},\mathbf{x}^{-1}]),\qquad\forall\mathbf{x}\in\mathbb{R}_+^d \tag{5.93}$$

$$-\log G(x_1,\ldots,x_d)\quad=l\left(-\log G_1(x_1),\ldots,-\log G_d(x_d)\right) \tag{5.94}$$

Definition 36. *Stable Tail Function l A function $l : [0,\infty)^d \mapsto [0,\infty)$ is a d-dimensional stable tail dependence function if and only if it satisfies the following properties:*

1. *function l is homogeneous of degree $\lambda = 1$ which means that*

$$l(\lambda x_1, \ldots, \lambda x_d) = \lambda l(x_1, \ldots, x_d), \;\; \forall \lambda \in [0,\infty). \tag{5.95}$$

2. *The function l must produce for all $x_1, \ldots, x_d \in [0,\infty)$ that*

$$\overline{G}_l(x_1, \ldots, x_d) = \left[\max\left\{0, 1 - l(x_1, \ldots, x_d)\right\}\right]^{d-1} \tag{5.96}$$

defines a d-dimensional survival function with $\mathcal{B}(1, d-1)$ margins.

An example of such a stable tail function involves the transformation of a d-variate extreme value copula C^{EVT} given by

$$l(x_1, x_2, \ldots, x_d) = -\ln\left\{C^{EVT}\left(\exp(-x_1), \exp(-x_2), \ldots, \exp(-x_d)\right)\right\} \tag{5.97}$$

and we have for instance the Gumbel extreme-value copula (symmetric logistic model) producing for parameter $\theta \geq 1$ the function

$$l_\theta(x_1, x_2, \ldots, x_d) = \left(x_1^\theta + \ldots + x_d^\theta\right)^{1/\theta}. \tag{5.98}$$

One can then combine this with an Archimedean generator $\varphi(x)$ of an Archimedean copula to produce the resulting family of d-dimensional Archimax copula, given by

$$C_{\varphi,l}(u_1, \ldots, u_d) - \varphi \circ l\left(\varphi^{-1}(u_1), \varphi^{-1}(u_2), \ldots, \varphi^{-1}(u_d)\right). \tag{5.99}$$

One can use this type of representation to develop the d-dimensional extension of the bivariate example above, for instance using the Archimax copula structure of [Charpentier *et al.*, 2014, Corollary 6.3], which is defined as follows:

$$C_{l^*,\mathcal{A}^*}(u_1, u_2, \ldots, u_d) = \exp\left[\mathcal{A}^*(t_1, \ldots, t_{d-1}) \sum_{i=1}^{d} \ln(u_i)\right] \tag{5.100}$$

where $\mathcal{A}^*$ is given by the function

$$\mathcal{A}^* = l^\alpha\left(t_1^{1/\alpha}, \ldots, 1 - \sum_{k=1}^{d-1} t_k^{1/\alpha}\right) \tag{5.101}$$

where $t_k = |\ln(u_k)| / \left\{\sum_{i=1}^{d} \ln(u_i)\right\}$.

Case Study 1: in this study we demonstrate how one can utilise the copula mapping to study the local affect of dependence induced in different regions of the state space, arising as one varies the parameters of the Archimax copula model. In particular, we will construct the local correlation surfaces obtained from the generalized Gaussian functional copula mapping and plot a range of such surfaces over the support of the bivariate Archimax copula as we vary the copula parameters θ. The results of this study are demonstrated in Figure 5.12.

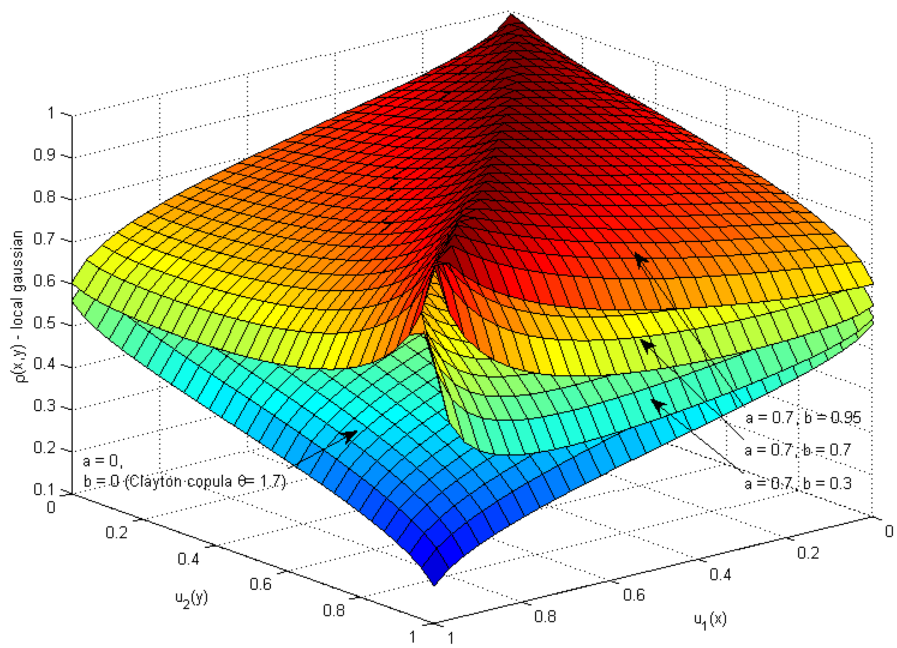

Figure 5.12: Local Gaussian copula coefficients $\rho(u_1, u_2)$ for a 2-D Archimax Copula with Clayton generating function $\varphi(x, \theta)$ and Pickands function $\mathcal{A}(t)$. The parameters are $\theta = 1.7$ and $(a, b) = [(0, 0), (0.7, 0.3), (0.7, 0.7), (0.7, 0.95)]$. Note that when $(a, b) = (0, 0)$ then the Archimax copula is equivalent to the Clayton one with the same θ parameter which is equal to 1.7 in this specific example.

5.4 Conclusions

In this chapter we presented a numerical procedure to map any copula function to a generalized local Gaussian copula function. This allows us to visualize a given copulas' characteristics and better quantify the local correlation generated by the copula function. In fact copulas' parametric expressions are neither intuitive nor do they reflect the dependence structure that they create. The proposed copula mapping is entirely based on tensor algebra and is part of a new modelling framework we also introduced consisting in mimicking multidimensional correlated diffusions by means of continuous time Markov chains. We demonstrated the simplicity and intuition behind the copula mapping through illustrative examples and all the copula functions mapping results are exact up to the tensor space local discretization error and the error due to approximating the generator matrix exponential.

Chapter 6

Evaluating concordance measures by tensor approximation

There is now an increasingly large number of proposed concordance measures available to capture, measure and quantify different notions of dependence in stochastic processes. However, evaluation of concordance measures to quantify such types of dependence for different copula models can be challenging. In this work, we propose a class of new methods that involves a highly accurate and computationally efficient procedure to evaluate concordance measures for a given copula, applicable even when sampling from the copula is not easily achieved. In addition, this then allows us to reconstruct maps of concordance measures locally in all regions of the state space for any range of copula parameters. We believe this technique will be a valuable tool for practitioners to understand better the behaviour of copula models and associated concordance measures expressed in terms of these copula models.

6.1 Introduction

In Chapters 4 and 5 we developed a novel class of functional copula representations for dependence that can be viewed as a reinterpretation of Sklars' well known copula representation theorem of multivariate dependence. This is achieved in a manner that allows one to characterize any copula dependence function via a unique map to a generalized Gaussian copula function. This unique mapping is obtained through the quantification of local dependence properties of the reference copula model over a discretized representation. Furthermore, we also demonstrated that such a representation is proven to be exact as the discretization interval of the target copulas support diminishes, with known convergence rate as studied in detail in 4.3.4. In this chapter we extend and develop this representation for the purpose of calculating and evaluation concordance measures, which can be written as functionals of copula models. These can include concepts such as multivariate upper negative (positive) dependence, lower negative (positive) dependence, negative (positive) dependence; multivariate negative (positive) quadrant dependence; multivariate association, co-monotinicity, stochastic ordering; regression dependence negative (positive) and extreme dependence, asymptotic tail dependence and intermediate tail dependence; as well as other concepts such as directional dependence, see Joe [1990, 1997]; Nelsen [2002]. Measuring the dependence between random variables has long been of interest to statisticians and practitioners alike. A history of the development of dependency measures can be found in Mari & Kotz [2001]. One should realise that in general, the dependence structure between two random variables can *only* be captured in full by their joint probability distribution, and thus any scalar quantity extracted from this structure must be viewed as a representation of some feature of the dependence.

Scarsini [1984] gives the following intuitive definition of dependence:

> Dependence is a matter of association between X and Y along any measurable function, i.e. the more X and Y tend to cluster around the graph of a function, either $y = f(x)$ or $x = g(y)$, the more they are dependent.

The choice of dependence measure is influenced by the type of dependence one seeks to study, such as lower left quadrant, upper right quadrant, etc. However, in nontrivial multivariate distributions, it is not possible to capture all of the possible combinations of dependence patterns within a single dependence measure. For this reason there is now an increasingly large number of proposed concordance measures available to capture, measure and quantify different notions of dependence in stochastic processes. The study of such measures of dependence began in pairwise constructions in the works of Cambanis *et al.* [1976], Tchen [1980] and Yanagimoto & Okamoto [1969]. Where they demonstrated that an ordering on discrete bivariate distributions, which formalized the notion of concordance, was shown to be equivalent to stochastic ordering of distribution functions with identical marginals. These notions were then generalised in multivariate settings by for instance Joe [1990] and Scarsini [1984] with the latter introducing a formal axiomatic representation of multivariate concordance measures given in Proposition 14. A detailed overview of these concepts is provided in Cruz *et al.* [2015a].

Definition 37. *Consider the following basic definitions for permutation and symmetry, used throughout.*

- ***Symmetries:*** *a symmetry of* $[0,1]^d$ *is a one-to-one, onto map* $\phi : [0,1]^d \mapsto [0,1]^d$ *of form* $\phi(x_1, \ldots, x_d) = (u_1, \ldots, u_d)$ *where for each* i *one has* $u_i = x_{k_i}$ *or* $1 - x_{k_i}$ *and where* $(k_1, \ldots, k_d)$ *is a permutation of* $(1, \ldots, n)$*;*
- ***Permutation:*** *the map* ϕ *is a permutation if for each* i *one has* $u_i = x_{k_i}$*;*
- ***Reflection:*** *the map* ϕ *is a reflection if for each* i *one has* $u_i = x_i$ *or* $u_i = 1 - x_i$*.*
  - ***Elementary reflections:*** *an elementary reflection of the* i*-th component, denoted* σ_i *is given by*

$$\sigma_i(x_1, \ldots, x_d) = (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_d)$$

- ***Symmetry Length:*** *the length of a symmetry is denoted by* $|\phi|$ *and corresponds to the number elementary reflections required to obtain it.*

We can now proceed to provide the general form of axioms for specification of a multivariate concordance measure as detailed in Scarsini [1984].

Proposition 14 (Multivariate Concordance Measures)**.** *A general concordance measures* κ *is a function attaching to all* d*-tuples of continuous r.v.'s* $(X_1, X_2, \ldots, X_d)$ *defined on a common probability space, when* $d \geq 2$*, a real number* $\kappa(X_1, X_2, \ldots, X_d)$ *satisfying:*

- ***Normalization:*** *$\kappa(X_1, X_2, \ldots, X_d) = 1$ if each X_i is a.s. an increasing function of every other X_j and $\kappa(X_1, X_2, \ldots, X_d) = 0$ if $X_1, \ldots, X_d$ are independent;*

- ***Monotonicity:*** *If $X_1, \ldots, X_d$ is less concordent than $Y_1, \ldots, Y_d$ then $\kappa(X_1, X_2, \ldots, X_d) < \kappa(Y_1, Y_2, \ldots, Y_d)$;*

- ***Continuity:*** *If F_k is the joint distribution of $(X_{k1}, \ldots, X_{kd})$ and F the distribution of $(X_1, \ldots, X_d)$ and one has convergence in the sequence $F_k \to F$ as $k \to \infty$, then $\kappa(X_{k1}, \ldots, X_{kd}) \to \kappa(X_1, \ldots, X_d)$;*

- ***Permutation Invariance:*** *If $(i_1, \ldots, i_d)$ is a permutation of $(1, \ldots, d)$ then $\kappa(X_{i_1}, \ldots, X_{i_d}) = \kappa(X_1, \ldots, X_n)$;*

- ***Duality:*** *$\kappa(-X_1, \ldots, -X_n) = \kappa(X_1, \ldots, X_n)$;*

- ***Reflection Symmetry:*** *$\sum_{\epsilon_1, \ldots, \epsilon_d = \pm 1} \kappa(\epsilon_1 X_1, \ldots, \epsilon_d X_d) = 0$ where the sum is over 2^d vectors of the form $(\epsilon_1 X_1, \ldots, \epsilon_d X_d)$ with $\epsilon_i \in \{-1, 1\}$;*

- ***Transition:*** *There exists a sequence $\{r_d\}$ for $d \geq 2$ such that every d-tuple of continuous r.v.'s $(X_1, \ldots, X_d)$ satisfies*

$$r_{d-1}\kappa(X_2, \ldots, X_d) = \kappa(X_1, \ldots, X_d) + \kappa(-X_1, X_2, \ldots, X_d)$$

In addition, there is an increasing number of copula models aimed at modelling dependence and in many cases there is no clear and intuitive correspondence between the magnitude of a copula's parameters and the dependence structure they create. Furthermore the form of concordance and the strength of that concordance measure in different dimensions as a function of the value of the parameters of the copula model. We view the work created in this manuscript as a useful class of representations and tools that can help study such features in non-trivial copula models.

In Taylor [2007] they provided a representation of the axioms of a concordance measure from Scarsini [1984] early work, rewritten explicitly in terms of copula models. This provides a link between these measures of dependence and the copula model, however a good understanding of the strength or significance of a concordance measure as a function of the copula model parameters is not well understood and difficult to study at present. The main reason for this is that often the evaluation of these concordance measures for different copula models can be very challenging and typically does not admit simple closed form solutions except in some special well known cases. There are many copula families in which a framework such as the one we propose in this manuscript will facilitate efficient and computationally accurate methods to gain an understanding of the relationships between different notions of concordance and the copula models parameters.

Therefore, the aim of this chapter is to develop a class of numerical approximations that practitioners can utilise to study locally in the state space of the multivariate random vector two important features: firstly the effects of parametric specifications of dependence in the form of different copula models and their local induced concordance structures; and secondly the role of each parameter in the copula model specifications in varying the concordance measure strength locally in the state space. This should be achievable for any copula model and any desired concordance measure. To achieve these goals we extend the class of methods developed in Chapters 4 and 5 by developing a highly accurate and computationally efficient procedure to evaluate concordance measures for a given copula.

In Chapter 4 a theoretical framework for tensor representation is developed and theoretical convergence properties are proven along with computational rates of convergence and complexity. Then in Chapter 5 we proposed a generic model for approximating any target copula function which will allow us to develop a clear and efficient characterization of their properties locally in any desired region of the state space (support of the target copula model). This characterization admits a accurate reconstruction of a given copula that we will demonstrate in this chapter allows one to then utilise to accurately calculate a wide range of concordance measures of dependence both locally and globally in the support of the given target copula model.

We recall that a copula is simply a multivariate probability distribution for which the marginal probability distribution of each variable is uniform Nelsen [1999]. Copulas are often used in high-dimensional statistical applications as they allow one to separate out the modelling and estimation of the distribution of dependent random variables by estimating first the marginals and then capturing the dependence structure through estimation of a copula function.

Outline of Contributions

In section 6.2 we will investigate further the properties of the local Gaussian copula $c^{(n)}$ developed in the previous Chapter 5 which make our proposed methodology to mimic approximated target copulas extremely appealing when dealing with calculations involving copulas and specifically the concordance measures.

In this chapter we extend these results to utilise the copula model decompositions to evaluate efficiently and locally in the state space of the random vector a wide range of multivariate concordance measures of dependence. To achieve this in Section 6.3 we present the relationship between concordance measures and copula models and then introduce our tensor algebraic approximations for a range of multivariate concordance measures for different copula families. This allows us to study these concordance measures as a function of the copula parameters to better understand how the copula parameters induce dependence and what type of dependence is present for these copula models. This is far from trivial to understand in the grouped, generalized and skewed Student's t-copula cases and the Archimax cases. We finish with illustrations of approximations of concordance measures for multivariate rank correlations, directional dependences, intermediate tail dependence and asymptotic tail dependence measures.

6.2 Properties of the Approximated Local Gaussian Copula $c^{(n)}$

In this section we give another explanation to the way we construct the copula approximation. We prove that the approximating function $c^{(n)}$ is interpretable the approximating density or Radon-Nikodym derivative of the approximated measure $P^{(n)}$ with respect to $P^{(n)\perp}$, where the notation for the approximated density $P^{(n)}$ and approximated uncorrelated density $P^{(n)\perp}$ is as described in Section 5.2.2.

Remark 9 (Orthogonality of the tensor basis)**.** *By construction the approximated local copula density function $c^{(n)}$ is defined on a orthogonal basis, which is the basis resulting from the tensor product of the constituent operators. In fact each operator matrix used in its calculation is orthogonal by construction, see Appendix 5.1.3. Furthermore we construct the copula over the support $\mathcal{X}$ which may be represented as union of disjoint subsets $\{B_j\}$, i.e. $\mathcal{X} = \bigotimes_{i=1}^{d} \mathcal{X}^i = \bigcup_{j=1}^{n} B_j$, where each set B_j as coordinate vector point $\mathbf{x}$ in the space $\mathbf{X}$. If we then set $B_0 = \{\mathbf{x} = (x_0^{(1)}, \ldots, x_0^{(d)})\}, B_1 = \{\mathbf{x} = (x_1^{(1)}, \ldots, x_0^{(d)})\}, \ldots, B_n = \{\mathbf{x} = (x_{n_1}^{(1)}, \ldots, x_{n_d}^{(d)})\}$, therefore creating a countable and ordered sequence of sets spanning the whole discretized support $\mathcal{X}$ then it is straightforward to obtain the copula distribution functions as the cumulative sum over all the sets, i.e.*

$$C^{(n)}(B_z) = \sum_{j \leq z} c^{(n)}(B_j) \tag{6.1}$$

In order to prove the above properties we first need to remind some basic definitions of the density and conditional distribution of a copula.

Definition 38. *Let $F_1, \ldots, F_d$ be continuous marginal distributions. Let C be a copula distribution*

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \tag{6.2}$$

When the copula distribution C and the joint distribution F are both differentiable, the joint density function P satisfies

$$c(\mathbf{u}) = \frac{P(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{P_1(F_1^{-1}(u_1)) \cdots P_d F_d^{-1}(u_d))} = c(F_1(x_1), \ldots, F_d(x_d)) \tag{6.3}$$

with c denoting the copula density which is linked to the distribution C by

$$c(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \ldots du_d} C(\mathbf{u}). \tag{6.4}$$

Therefore the copula density is the ratio of the joint density $P(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ and the density under independence which is equivalent to the product of the marginal density function $P_1(x_1) \cdots P_d(x_d)$. It is therefore possible to interpret the copula as the adjustment that we need to make to convert the independence pdf into the joint pdf.

However this adjustment can also have another interpretation: a Radon-Nikodym (R-N) derivative. In our specific case it will be a R-N derivative for the discrete measure $P^{(n)}$ constructed through the local Gaussian CTMC approximation.

If in eq. (6.3) we set $P(\mathbf{u}) = P(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))$ and $P^{\perp}(\mathbf{u}) = P_1(F_1^{-1}(u_1)) \cdots P_d F_d^{-1}(u_d))$ for convenience of notation then

$$P(\mathbf{u}) = c(\mathbf{u}) P^{\perp}(\mathbf{u}). \tag{6.5}$$

Let us give some more insights about the interpretation of eq. (6.5) in a more general mathematical setting and the way the copula $c^{(n)}(\mathbf{u})$ is constructed in our proposed model.

For this purpose $X = (X, \mathcal{F})$ is a measurable space, and we shall refer to the elements of $\mathcal{F}$ simply as measurable sets omitting the σ-algebra $\mathcal{F}$. All subsets of X and all functions on X appearing below are measurable unless otherwise indicated.

Definition 39 (Radon-Nikodym Theorem)**.** *Let μ and ν be two continuous measures on X, with ν is σ-finite and μ absolutely continuous with respect to ν. Then there exists a positive function such that*

$$\mu(B) = \int_B f d\nu \tag{6.6}$$

for all subsets B of X. The function f is unique in the sense that if $\mu(B) = \int_B g d\nu$ for all B then $f = g$. The function f is called the Radon-Nikodym derivative or density of μ with respect to ν and is denoted by $\frac{d\mu}{d\nu}$.

Note that a necessary and sufficient condition that the Radon-Nikodym derivative f be integrable is that μ is σ-finite. If we restrict the Radon-Nikodym derivative f to the space $(X, \mathcal{G})$ with $\mathcal{G} \in \mathcal{F}$ a σ-subalgebra, the Radon-Nikodym derivative is

$$\frac{d\mu}{d\nu}|_{\mathcal{G}} = \frac{1}{P(\mathcal{G})} \int_{\mathcal{G}} f d\mu = \mathbb{E}[f|\mathcal{G}]. \tag{6.7}$$

We also remind the law of total expectation which plays an import role in the way we construct the approximated copula density.

Definition 40. *If X is a random variable whose expected value is $\mathbb{E}[X]$, and if $\{B_j\}$, $i = 1, 2, \ldots$ is a finite or countable partition of the sample space, then*

$$\mathbb{E}[X] = \sum_j \mathbb{E}[X|B_j]P(B_j) \tag{6.8}$$

We are now in a position to explain what is the relationship between a R-N derivative λ and all its restrictions $g_j = \mathbb{E}[\lambda|B_j]$ over disjoint sets $\{B_j\}$, $i = 1, 2, \ldots$, and we do this in the discrete settings of the copula $c^{(n)}$.

Proposition 15 (Construction of the Approximated Copula c^n)**.** *The copula $c^{(n)} : \mathbf{U} \to [0, \infty)$ in eq. (6.5) is the approximating density or R-N derivative of the approximated measure $P^{(n)}$ with respect to $P^{(n)\perp}$. Furthermore $c(n)$ it is the results of the averaging of R-N $g_j = \mathbb{E}[\lambda|B_j]$ which are the restriction of c(n) on disjoint set B_j spanning $\mathcal{X} = \bigotimes_i \mathcal{X}^i$.*

Proof. From eq. (6.5) we have that the copula density $c^{(n)}$ is interpretable as R-N derivative of the approximated correlated measure $P^{(n)}$ with respect to the absolutely continuous measure $P^{(n)\perp}$ for all sets in $\mathbf{U}$. Equivalently for the approximated copula $c^{(n)}$ we have that $P^{(n)}(A) = \mathbb{E}[1_A c^{(n)}]$ for $A \in \mathcal{X}$. We then observe that in our framework the density support $\mathcal{X} = \bigcup_{j=1}^n B_j$ can be represented as the union of countable disjoint sets B_j. Therefore we can apply the law of total expectation to the discrete R-N derivative and obtain:

$$\mathbb{E}[1_A c^{(n)}] = \mathbb{E}\Big[\sum_j \frac{\mathbb{E}[c^{(n)} 1_{B_j}]}{P(B_j)} 1_{B_j}\Big] = \sum_j \mathbb{E}[c^{(n)} 1_{B_j}] = g_j \mathbb{E}[1_{B_j}] \tag{6.9}$$

being the variable g_j is measurable with respect to B_j. □

In our proposed framework we approximate the first two moments of each conditional variable $g_j = \mathbb{E}[\lambda|B_j]$ with the first two moments of a conditional normal variable. Details of the local construction of each approximated conditional variable are in Appendix 5.1.3.

Using eq. (6.9) we can calculate a local conditional value, which is Gaussian in our case, matching the local target copula value, and this is done by means of the local minimizer of eq. (5.31). The existence of the local minimizer is explained by the above properties of the R-N derivative. The fact that in this setting always exists a bounded R-N derivative linking the two densities, and the approximation of the restricted R-N derivative to a conditional Gaussian density, guarantees the existence of a local Gaussian copula value and its corresponding correlation $\rho(\mathbf{u})$.

6.2.1 Direct measure of directional dependence

In this section we show how our framework gives immediate measure of directional dependence. At this purpose we recall some basic definitions.

Definition 41. *[Directional Dependence] The pair (U_1, U_2) is directionally dependent in joint behaviour if*

$$\mathbb{E}[U_1|U_2 = w] \neq \mathbb{E}[U_2|U_1 = w] \tag{6.10}$$

Directional dependence can be expressed in terms of regression using a copula distribution function, see Sungur [2005].

Definition 42 (Directional Dependence through regression, see Sungur [2005])**.** *Let (U_1, U_2) denote a pair of random variables whose marginal distributions have uniform distribution on $[0, 1]$ and the joint distribution is a copula function $C(u_1, u_2)$. Let $C_{u_2}(u_1)$ denote the conditional distribution of U_1 given $U_2 = \bar{u}$*

$$C_{u_2}(u_1) = P(U_1 \leq u_1|U_2 = \bar{u}) = \frac{\partial C(u_1, u_2)}{\partial u_2}. \tag{6.11}$$

The copula regression function of U_2 on U_1 is the conditional expectation of U_1 given $U_2 = \bar{u}$, which can be expressed by the copula as

$$r_{U_1|U_2}(\bar{u}) := \mathbb{E}[U_1|U_2 = \bar{u}] = 1 - \int_0^1 C_{u_2}(u_1)du_1. \tag{6.12}$$

The directional dependence from U_2 to U_1 is defined by using the copula regression function on U_1 as

$$\rho^2_{U_2 \to U_1} = \frac{Var(r_{U_1|U_2}(\bar{u}))}{Var(U_1)} = \frac{\mathbb{E}[(r_{U_1|U_2}(\bar{u}) - 0.5)^2]}{1/12} = 12\mathbb{E}[(r_{U_1|U_2}(\bar{u}))^2] - 3 \tag{6.13}$$

which can be interpreted as the proportion of total variance of U_1 that has been explained by the copula regression function $r_{U_1|U_2}(\bar{u})$.

We can immediately calculate eq. (6.12) in our framework which can be rewritten in terms of approximated local Gaussian copula as

$$r^{(n)}_{U_1|U_2}(\bar{u}) = 1 - \sum_i C^{(n)}_{u^{(2)}}(u^1_i)\Delta u^{(1)}. \tag{6.14}$$

where

$$C^{(n)}_{u^{(2)}_j}(u^{(1)}_i) = \frac{1}{\Delta u^{(2)}_j}\Big(C^{(n)}(u^{(1)}_i, u^{(2)}_j + \Delta u^{(2)}_j) - C^{(n)}(u^{(1)}_i, u^{(2)}_j)\Big) \tag{6.15}$$

with $\mathbf{U}^k = \{u^{(k)}_1, u^{(k)}_2, \ldots, u^{(k)}_{n_k}\}$ the discretized support and $\Delta u^{(k)}_i = u^{(k)}_{i+1} - u^{(k)}_i$. Note that each support grid U^k is not uniformly spaced.

6.3 Understanding Relationships between Copula Parameter(s) and Concordance Measures

In this section we discuss the relationship between concordance measures of dependence and the copula representations presented in the previous section. In particular we illustrate how to utilise such copula representations to develop a clearer understanding of the relationship between a copula parameter and different measures of concordance which can be generally applied to any form of dependence model captured by a copula. The choice of dependence measure or concordance measure is influenced by the type of dependence one seeks to study, such as: multivariate upper negative (positive) dependence, lower negative (positive) dependence and negative (positive) dependence; multivariate negative and positive quadrant dependence; multivariate association, co-monotonicity and stochastic ordering; positive and negative regression dependence; and extreme dependence, tail dependence and intermediate tail dependence. We start by discussing recent relationships between general notions of concordance measures between random variables and their characterization through copula parametric models. We recall first a general definition of a concordance measure for a random vector.

Definition 43 (Concordance Measures for Random Vectors)**.** *A pair of random variables are concordant if 'large' values of one tend to be associated with 'large' values of the other and 'small' values of one with 'small' values of the other. Analogous definitions of discordance are available in reverse directions.*

There are numerous ways of mathematically trying to quantify this statement, so consequently, many measures of concordance are available. Such a definition, offers the intuition behind the need for measuring and quantifying concordance. However, from a statistical perspective a more formal understanding of concordance measures was mathematically stated with regard to copula models in the recent works of Taylor [2007]. This selection of axioms that are desirable for a concordance measure to satisfy are based on the copula characterization of the earlier framework originally developed by Scarsini [1984] and presented in Proposition 14. In Taylor [2007] the following axioms for general concordance measures κ specified generically via copula C are detailed below in Proposition 16.

Proposition 16 (Multivariate Concordance Measures via Copula)**.** *Consider a sequence of maps $\kappa_d : Cop(d) \mapsto \mathbb{R}$ and a sequence of numbers $\{r_d\}$, such that if A, B, C and C_m are d-copulas and $n \geq 2$ then:*

- *Normalization:*$\kappa\left(M^d\right) = 1$ *and* $\kappa\left(\Pi^d\right) = 0$.
- *Monotonicity: If* $A <_{st} B$ *and* $\overline{A} \leq_{st} \overline{B}$ *then* $\kappa_d(A) \leq \kappa_d(B)$
- *Continuity: If* $C_m \to C$, *then* $\kappa_d(C_m) \to \kappa_d(C)$ *as* $m \to \infty$.
- *Permutation Invariance: If* $(i_1, \ldots, i_d)$ *is a permutation of* $(1, \ldots, d)$ *then* $\kappa(c(u_{i_1}, \ldots, u_{i_d})) = \kappa(c(u_1, \ldots, u_d)$.
- *Duality:* $\kappa_d(c(1-u_1, \ldots, 1-u_d) = \kappa_d(c(u_1, \ldots, u_d))$
- *Reflection Symmetry:* $\sum_{\Psi \in \mathcal{R}_d} \kappa_d\left(C^{\Psi}\right) = 0$, *where* Ψ *is a reflection,* $\Psi \in \mathcal{R}_d$ *is an element of the subgroup of reflections in the group of symmetries under composition* $\mathcal{S}([0,1]^d)$.

- *Transition:*

$$r_n \kappa_d\left(C\right) = \kappa_{n+1}\left(E\right) + \kappa_{n+1}\left(E(1-u_1, u_2, \ldots, u_d)\right)$$

whenever E *is an* $(d+1)$*-copula s.t.* $C(u_1, \ldots, u_d) = E(1, u_1, \ldots, u_d)$.

Although we know these general relationships between the axioms that describe concordance and dependence structures in general and we now know them in general terms as specified by a statistical model in the form of a copula, it still remains a real challenge to evaluate and understand these relationships with regard to model parameters for a given copula model. We aim to address this challenge using the representations of copula developed previously in the chapter.

We note that one can also state the following theorem regarding the properties of concordance measures that satisfy these axioms, see details in Taylor [2007].

Theorem 16 (Properties of Concordance Measures Satisfying Proposition 16)**.** *Consider the d-copula that is permutation symmetric ie.* $C^{\zeta} = C$ *for all permuations* ζ *of* $[0,1]^d$*. Then for all measures of concordance* κ *and for all symmetries* Ψ *and* ζ *of* $[0,1]^d$ *one has*

$$\kappa_d(C^{\Psi}) = \kappa_d(C^{\zeta}) \tag{6.16}$$

whenever $|\Psi| = |\zeta|$ *or* $|\Psi| + |\zeta| = d$

Recall: symmetry length $|\cdot|$ corresponds to the number elementary reflections required to obtain it.

Corollary 1. *For all* $d \geq 2$ *and for all symmetries* Ψ *and* ζ *of* $[0,1]^d$ *such that* $|\Psi| = |\zeta|$ *or* $|\Psi| + |\zeta| = d$ *one has*

$$\kappa_d(M^{\Psi}) = \kappa_d(M^{\zeta}). \tag{6.17}$$

where M *is the d-Frechet-Hoffding Upper Bound copula under permutation.*

With regard to copula models there are numerous measures one can consider that satisfy all or some subset of such axioms. Typically, the evaluation of these concordance measures and the understanding of the relationship between these concordance measures and the copula parameter(s) is non-trivial and changes as a function of both dimension and copula parameter(s) value highly non-linearly. In this section we first discuss some important examples of concordance measures and then we detail how the framework developed previously may be utilise to provide a natural framework for evaluation of such measures, thus providing the necessary understanding of the relationships between the copula family, copula parameter(s) and the behaviour of the concordance measure.

6.3.1 Functional Copula Mapping for Approximation of General Multivariate Concordance Measures

Arguably the most widely known and utilised measure of dependence, Pearson's Product Moment Correlation Coefficient, was developed by Karl Pearson, see Pearson [1896], building on Sir Francis Galton's approach using the median and semi-interquartile range, see Galton [1894]. Pearson's correlation coefficient is a measure of how well the two random variables can be described by a linear function. The other popular measures often used in practice when considering notions of pairwise dependence are Spearman's ρ (Spearman

[1904]), Kendall's τ (Kendall [1938]) and Blomqvist's β (Blomqvist [1950]). In the case of Spearman's ρ it is a measure that assesses how well the dependence between two random variables can be described by a monotonic function. As such it is equivalent to the Pearson's correlation coefficient between the ranked variables. In addition, one can consider Spearman's rho and Kendall's τ each as a simple scalar measures of dependence that depend on the copula of two random variables but not on their marginal distributions. As noted in Fredricks & Nelsen [2007], one can consider Spearman's ρ as forming a measure of average quadrant dependence, while Kendall's τ is a measure of average likelihood ratio dependence. In the case of Blomqvist's β the generalized form is discussed in any dimension in Joe [1990], Dolati & Úbeda-Flores [2006], and Nelsen [2002] where it is shown to be expressed directly in terms of a copula distribution function. However, unlike Kendalls τ and Spearmans ρ which are functionals integrated against the copula distribution, the Blomqvists β is simple to evaluate given the copula distribution function. It simply requires the evaluation of the copula distribution at a point $\bar{u} = [1/2, \ldots, 1/2]$. We therefore focus below on the evaluation of functionals with regard to the constructed copula and demonstrate how to evaluate these efficiently.

First we observe the following results for Spearman's and Kendall's Rank correlations, both can be specified directly in terms of a copula distribution in both the bivariate and multivariate settings. We begin with the specification of multivariate Spearman rank.

Multivariate Spearman Rank Correlation Approximations

Definition 44 (Spearman's Rank Correlation via Copula)**.** *The bivariate Spearman's Rank Correlation can be expressed explicitly via the bivariate copula C according to*

$$\rho = 12 \int_{[0,1]} \int_{[0,1]} u_1 u_2 dC\left(u_1, u_2\right) - 3. \tag{6.18}$$

In addition, a general multivariate extension of Spearman's Rank Correlation is developed for d-dimensional loss random vectors and given below, see details in Nelsen [2002].

Definition 45 (Multivariate Spearman's Rho via Copula)**.** *Consider the d-copula given by C, then Spearman's Rho concordance measure of dependence is given by*

$$\rho(C) = h(d)\Big[2^d \int_{[0,1]^d} C(u)du - 1\Big] = h(d)\Big[2^d \int_{[0,1]^d} \Pi(u)dC(u) - 1\Big] \tag{6.19}$$

where $h(d) = \frac{d+1}{2^d-(d+1)}$ is the normalizing factor derived such that the maximum correlation is equal to 1.

Definition 46 (Multivariate Generalized Spearman's Rho via Copula)**.** *Consider the d-copula given by C and the permuted copula C^σ, then the generalized Spearman's Rho concordance measure of dependence is given according to*

$$\rho_d(C) = \alpha_d \left(\int_{[0,1]^d} \left(C + C^\sigma\right) d\Pi^d - \frac{1}{2^{d-1}} \right) \tag{6.20}$$

where one has $\alpha_d = \frac{(d+1)2^{d-1}}{2^d-(d+1)}$ and Π^d is the d-Independence Copula.

From this definition of the generalized Spearman's Rho rank correlation, we can also define the local as well as the constrained versions of the state-space rank correlation.

Definition 47 (Local and Constrained Multivariate Generalized Spearman's Rho via Copula)**.** *Consider the d-copula given by C and the permuted copula C^{σ}, then the local generalized Spearman's Rho concordance measure of dependence for vectors $\mathbf{u}$ localized in some sub-space $\mathbf{u} \in \Omega \subset [0,1]^d$ is given according to*

$$\rho_d(C;\Omega) = \alpha_d \left(\int_{\Omega} \left(C + C^{\sigma}\right) d\Pi^d - \frac{1}{2^{d-1}} \right) \tag{6.21}$$

where one has $\alpha_d = \frac{(d+1)2^{d-1}}{2^d-(d+1)}$ and Π^d is the d-Independence Copula. The constrained version of the Spearman's Rho is then given for vectors $\mathbf{u}$ constrained in some sub-space $\mathbf{u} \in \Omega \subset [0,1]^d$ according to

$$\widetilde{\rho}_d(C;\Omega) = \alpha_d \left(\int_{\Omega} \left(\widetilde{C} + \widetilde{C}^{\sigma}\right) d\widetilde{\Pi}^d - \frac{1}{2^{d-1}} \right) \tag{6.22}$$

where one has $\alpha_d = \frac{(d+1)2^{d-1}}{2^d-(d+1)}$ and $\widetilde{\Pi}^d$ is the d-Independence Copula. Where the $\widetilde{C}$ notation refers to the copula distribution given after renormalization by the truncation according to

$$\widetilde{C}(B;\Omega) = \int_B \frac{c(u)}{C(\Omega)} du \tag{6.23}$$

for all sets $B \in \Omega$.

An example where such a correlation measure is of interest arises when one considers intermediate and extremal dependence measures, such as the following illustrative bivariate examples where given a pair of uniform variable such that $(U_1, U_2) \sim C(u_1, u_2)$ and one wishes to evaluate the truncated correlation or restricted correlations given for instance by

$$\begin{aligned} \mathbb{E}\left[U_1U_2|U_1 > u_1, U_2 > u_2\right] &= \int_{[0,1]}\int_{[0,1]} u_1u_2 dC(u_1u_2|U_1 > u_1, U_2 > u_2) \\ \mathbb{E}[U_1U_2|U_1 \leq u_1, U_2 > u_2] &= \int_{[0,1]}\int_{[0,1]} u_1u_2 dC(u_1u_2|U_1 \leq u_1, U_2 > u_2). \end{aligned} \tag{6.24}$$

The important observation is that in general it can be very difficult to evaluate such quantities even using Monte Carlo procedures, as sampling the conditional or constrained copulas is a great challenge in many families of copula. In the the following we specify an efficient approximation based on the general functional copula mapping developed previously.

Proposition 17 (Approximating Local and Constrained Multivariate Generalized Spearman's Rho via Functional Copula Mappings)**.** *Consider the d-copula given by C which is discretely approximated by the functional copula mapping given by $C^{(n)}(\mathbf{u})$ for all vectors $\mathbf{u} \in \mathbf{U}^k$, $k = 1, \ldots, d$, and $C^{(n)\sigma}$ represents the functional copula mapping of the permuted copula C obtained after permuting the base mimicking local Gaussian copula mapping $C^{(n)}$. Then one has the functional approximations to the localized generalized Spearman's Rho* *concordance measure of dependence in some sub-space* $\Omega \subseteq \mathbf{U} \subset [0,1]^d$ *is given according to*

$$\rho_d^{(n)}(C^{(n)};\Omega) = \alpha_d \sum_{\mathbf{u}\in\Omega} \left[C^{(n)}(\mathbf{u}) + C^{(n)\sigma}(\mathbf{u})\right] \prod_{i=1}^{d}(\mathbf{u}-\mathbf{u}') - \frac{\alpha_d}{2^{d-1}} \tag{6.25}$$

where one has $\alpha_d = \frac{(d+1)2^{d-1}}{2^d-(d+1)}$ *and* $\mathbf{u}' \in \Omega$ *is the neighbour of* $\mathbf{u}$.
The constrained version of the Spearman's Rho is then given for vectors $\mathbf{u} \in \bar{\Omega} \subseteq \mathbf{U}$, *therefore for vectors in a constrained some sub-space* $\bar{\Omega} \subset [0,1]^d$ *according to*

$$\widetilde{\rho}_d^{(n)}(C;\bar{\Omega}) = \alpha_d \sum_{\mathbf{u}\in\bar{\Omega}} \left[\widetilde{C}^{(n)}(\mathbf{u}) + \widetilde{C}^{(n)\sigma}(\mathbf{u})\right] \frac{1}{\mathcal{C}\text{ard}\left\{\boldsymbol{U}\cap\bar{\Omega}\right\}} \prod_{i=1}^{d}(\mathbf{u}-\mathbf{u}') - \frac{\alpha_d}{2^{d-1}} \tag{6.26}$$

where one has $\alpha_d = \frac{(d+1)2^{d-1}}{2^d-(d+1)}$ *and where the* $\widetilde{C}^{(n)}$ *notation refers to the functional copula mapping distribution given after renormalization by the truncation according to*

$$\widetilde{C}^{(n)}(B;\bar{\Omega}) = \sum_{\mathbf{u}\in\bar{\Omega}\cap B} \frac{c^{(n)}(\mathbf{u})}{\sum_{\mathbf{u}\in\bar{\Omega}} c^{(n)}(\mathbf{u})} \tag{6.27}$$

for all sets $B \in \bar{\Omega}$.

Understanding Multivariate Spearman's Rank Correlations in Copula Parameter Space

We illustrate the approximation of the multivariate form of Spearman's rank correlation for a range of different copula models: skewed Student's t-copula; Archimax; and the Clayton copula member of the Archimedean class. Note, we adopt identical parameterizations for these copula models as detailed in the previous Chapter 5. Furthermore, the accuracy of the reconstruction via the mimicking local Gaussian copula mapping for a general class of copula is detailed in Section 5.3.

The first study we illustrate is to demonstrate the accuracy of our approximation method for evaluation of multivariate Spearman's rank correlation. To achieve this we consider one of the few cases which involves a non-trivial copula model that admits an explicit closed form relationship between the copula parameter and Spearman's rank correlation. This is the 2-dimensional Archimedean copula with the Clayton generator function. We illustrate below in Figure 6.1 the exact Spearman's Rho rank correlation as a function of θ the Clayton copula parameter and we plot our approximation for a range of copula parameter values as crosses. The local Gaussian copula distribution $C^{(n)}$ mimicking the parametric Clayton copula is obtained through the mapping of Proposition 11 on a unit square $\mathbf{U} = \mathbf{U}^1 \otimes \mathbf{U}^2$, with $n_1 = n_2 = 200$ discretization points. The approximated Spearman's Rho $\rho_d^{(n)}(C^{(n)};\mathbf{U})$ is then computed following proposition 17, where the distribution $C^{(n)}$ is integrated over the set $\mathbf{U}$. The resulting approximated correlation measure is compared with the closed form and it proves to be very accurate.

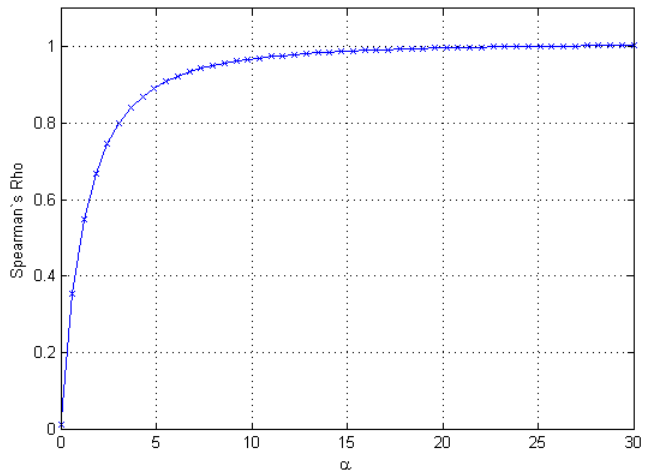

Figure 6.1: Spearman's Rho for a Clayton copula with parameter α. Solid line is the exact rank correlation as a function of the copula parameter α ranging from $\alpha = 0$ to $\alpha = 30$, whilst the crosses represent the approximated Spearman's rank correlation $\rho_d^{(n)}(C^{(n)};\mathbf{U})$ obtained using the mapped local Gaussian copula distributions $C^{(n)}$. Several Clayton copula mappings $C_{Clayton}(\alpha) \mapsto C^{(n)}$ have been calculated each one corresponding to a specific value of the Clayton copula parameter $\alpha \in [0, 30]$. The values of the approximated rank correlation $\rho_d^{(n)}(C^{(n)};\mathbf{U})$ coefficients are in line with the closed form ones and they exhibit also the desired asymptotic behaviour $\rho_d^{(n)} \to 1$ as $\alpha \gg 1$.

Next we illustrate how this decomposition can be used to study the behaviour of the copula parameter(s) and the effect varying such parameters will have on the resulting induced concordance measure of interest which in this case will first correspond to the multivariate rank correlation.

In a 2-dimensional illustration, we first consider the skewed Student's t-copula model. We present the case of two different values of skew parameter $\gamma \in \{1, 2\}$ and we demonstrate the range of the multivariate Spearman's rho rank correlation as a function of the degree of freedom parameter ν and the correlation parameter ρ. Below in Figure 6.2 we display the results of the mapping

$$C_{\text{skewed t-copula}}(\mathbf{u}; \{\rho, \gamma, \nu\}) \to C^{(n)} \to \rho_d^{(n)}(C^{(n)};\mathbf{U}) \text{ for } \gamma \in \{1, 2\}, \ \rho \in [-1, 1], \ \nu \in [10, 30]. \quad (6.28)$$

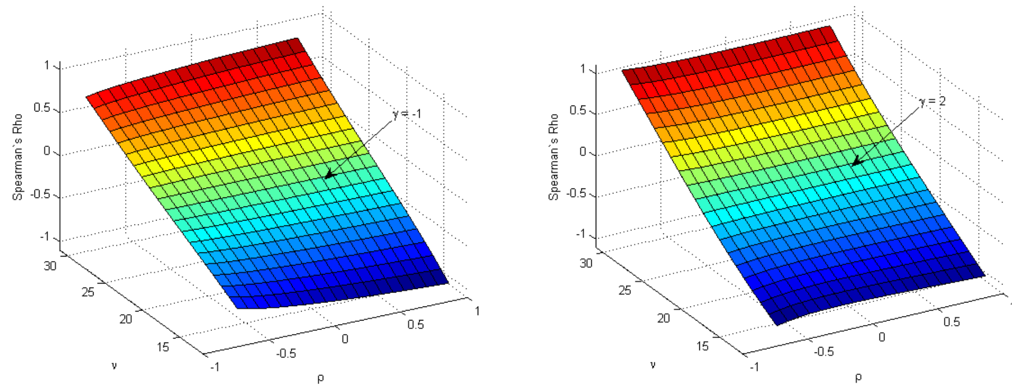

Figure 6.2: Approximated Spearman's Rho $\rho_d^{(n)}(C^{(n)}; \mathbf{U})$ for a skewed t-copula with parameters ν, γ and correlation ρ.

Firstly, we learn from this analysis that as the skewness parameter γ increases, the Spearman's rank correlation surface as a function of correlation ρ and degree of freedom ν goes from a concave to a convex relationship. Furthermore, this inversion of the correlation surface in the parameter space of ρ and ν is more pronounced at extreme correlations for small degrees of freedom ν, compared to the Gaussian like tail cases, which begin to arise as ν increases. In addition, we see from this analysis that as the skewness parameter γ increases, there is a tendency for the model to more readily attain the maximal Spearman's rank correlation of 1 uniformly in the correlation parameter ρ as the degree of freedom parameter ν increases. We conjecture from this analysis that as $\gamma \downarrow 0$ the maximal attainable Spearman's rank correlation is strictly less than 1 and ordered in the maximal correlation attained in the parameter ρ for any given ν.

In the next illustration, see Figure 6.3, we demonstrate how the multivariate Spearman rank correlation behaves for the Archimax copula which we constructed using a Pickands function parameterised by α and β parameters and a Clayton generator with parameter θ. Even for the 2-dimensional illustration example we present, very little is known about the Spearman's rank correlation for such models as a function of the parameters in this increasingly popular Archimax model. Again, we illustrate the behavior of the Spearman's rank correlation for different fixed values of the Clayton copula parameter, as we vary the Pickands distortion function parameters.

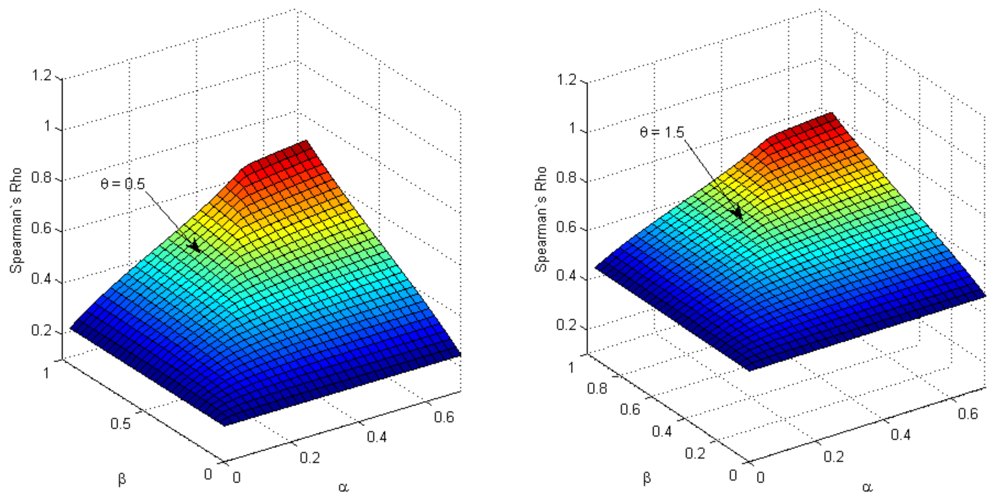

Figure 6.3: Approximated Spearman's Rho $\rho_d^{(n)}(C^{(n)}; \mathbf{U})$ for a Archimax copula with Pickands parmaters (α, β) and Clayton parameter θ.

We observe from this study several interesting features. Firstly, there is a lower bound for the Spearman's rank correlation for these models which is uniform across the parameter space for the Pickands distortion function parameters α and β. Secondly, this lower bound increases in magnitude towards the maximum of 1 as the Clayton generator parameter θ increases. Thirdly, we observe from this analysis that the maximal values of Spearman's rank correlation for a given combination of (α, β) lie on the line $\alpha = \beta$ for any give Clayton generator parameter θ.

Multivariate Kendall's Tau Rank Correlation Approximations

In terms of the multivariate extension of Kendall's tau rank correlation, one can obtain the copula representation of the concept of this rank correlation as follows. Consider the concordance function κ quantifying the difference in probabilities of concordance and discordance for bi-variate random vectors (X_1, Y_1) and (X_2, Y_2) which are specified as follows:

- Assume X_1 and X_2 have common continuous marginal F_X;
- Assume Y_1 and Y_2 have common continuous marginal F_Y;
- Assume (X_1, Y_1) and (X_2, Y_2) have different copula C_1 and C_2 respectively.

Then Nelsen [2002] proposed to consider an alternative copula specified concordance function κ for the equivalent Kendall's τ in a copula form as measuring the probability of concordance and discordance given by

$$\begin{aligned} \kappa &= \mathbb{P}\mathrm{r}\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - \mathbb{P}\mathrm{r}\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right] \\ &= 4\int_0^1 \int_0^1 C_2(u,v)dC_1(u,v) - 1. \end{aligned} \tag{6.29}$$

Therefore, with this copula based specification, one can then define analogously the functional copula mapping approximation for a given copula for this concordance-discordance copula based Kendall's τ measure, as well as the local and constrained variants.

Proposition 18 (Approximating the Generalized, Local and Constrained Multivariate Kendall's Tau via Functional Copula Mappings)**.** *Consider the d-copula given by C which is discretely approximated by the functional copula mapping given by $C^{(n)}(\mathbf{u})$ for $\mathbf{u} \in \boldsymbol{U}$. Hence, one obtains the functional copula approximation for the copula based Kendall's τ according to*

$$\kappa^{(n)}(C^{(n)}; \mathbf{U}) = 4\left|(u_1 - u_1')(u_2, u_2')\right| \sum_{\mathbf{u} \in \boldsymbol{U}} C_2^{(n_1,n_2)}(u_1, u_2) c_1^{(n_1,n_2)}(u_1, u_2) - 1. \tag{6.30}$$

In addition, the functional approximations to the localized Kendall's τ for some sub-space $\Omega \subseteq \mathbf{U} \subset [0,1]^d$ is given according to

$$\kappa^{(n)}(C^{(n)}; \Omega) = 4\left|(u_1 - u_1')(u_2, u_2')\right| \sum_{\mathbf{u} \in \boldsymbol{U} \cap \Omega} C_2^{(n_1,n_2)}(u_1, u_2) c_1^{(n_1,n_2)}(u_1, u_2) - 1. \tag{6.31}$$

The constrained version of the Kendall's τ is then given for vectors $\boldsymbol{u}$ constrained in some sub-space $\boldsymbol{u} \in \Omega \subset [0,1]^d$ according to

$$\widetilde{\kappa}^{(n)}(C^{(n)}; \Omega) = 4\left|(u_1 - u_1')(u_2, u_2')\right| \sum_{\mathbf{u} \in \boldsymbol{U} \cap \Omega} \widetilde{C}_2^{(n_1,n_2)}(u_1, u_2) c_1^{(n_1,n_2)}(u_1, u_2) - 1. \tag{6.32}$$

where $\widetilde{C}^{(n)}$ notation refers to the functional copula mapping distribution given after normalization by the truncation, as in eq. (6.27).

Understanding Multivariate Kendall's Tau Rank Correlations in Copula Parameter Space

In this section we illustrate the approximation of the multivariate form of Kendall's tau rank correlation for a range of different copula models: skewed Student's t-copula; Archimax; and the Clayton copula member of the Archimedean class. Again, the first study we illustrate is to demonstrate the accuracy of our approximation method for evaluation of multivariate Kendall's tau rank correlation. To achieve this we consider one of the few cases which involves a non-trivial copula model that admits an explicit closed form relationship between the copula parameter and Kendall's tau rank correlation. This is again the model given by the 2-dimensional Archimedean copula with the Clayton generator function. We illustrate below in Figure 6.4 the exact Kendall's Tau rank correlation as a function of θ the Clayton copula parameter and we plot our approximation for a range of copula parameter values as crosses.

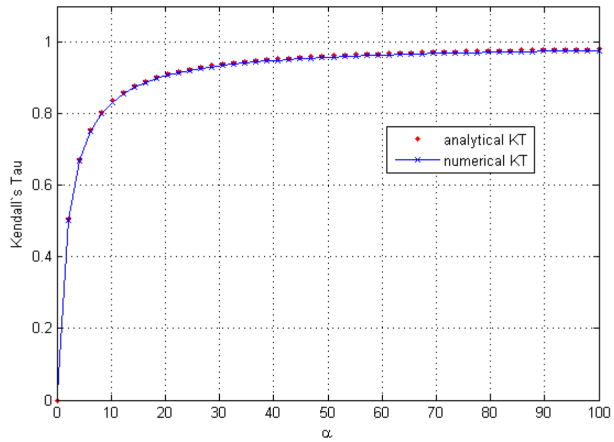

Figure 6.4: Kendall's tau for a Clayton copula with parameter α. Solid line is the exact rank correlation as a function of copula parameter θ whilst the crosses represent the approximated Kendall's Tau rank correlation $\kappa^{(n)}(C^{(n)}; \mathbf{U})$ of eq. (6.30) we obtained via the use of the functional copula mapping summing over the unit square $\mathbf{U} = \mathbf{U}^1 \otimes \mathbf{U}^2$, with $n_1 = n_2 = 200$ discretization points.

To further illustrate the accuracy of the approximation, see Figure 6.5, we construct via the functional copula mapping approach we also include the approximation for the calculation of Kendall's tau coefficient for a second Archimedean copula model given by the generator for the Frank copula, see Nelsen [1999] . The copula generator is $\varphi(t) = -\ln \frac{e^{-\theta t}-1}{e^{\theta}-1}$, where $\theta \in \mathbb{R}\backslash\{0\}$. This gives the Frank family

$$C_\theta(u,v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right) \tag{6.33}$$

It can be shown that Kendall's tau is

$$\tau_\theta = 1 - 4\frac{1 - D_1(\theta)}{\theta} \tag{6.34}$$

where $D_k(x)$ is the Debye function given by $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$, for $k > 0$.

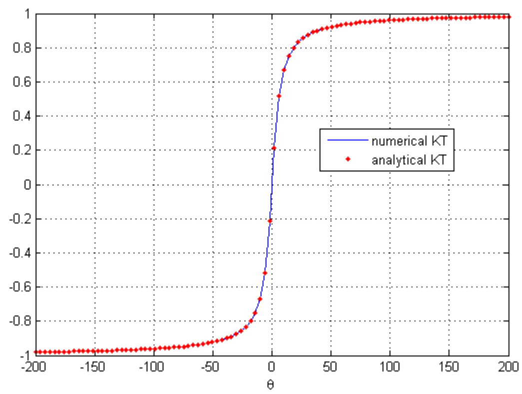

Figure 6.5: Approximated Kendall's tau rank correlation $\kappa^{(n)}(C^{(n)};\mathbf{U})$ of eq. (6.30) obtained via the use of the functional copula mapping summing over the unit square $\mathbf{U} = \mathbf{U}^1 \otimes \mathbf{U}^2$, with $n_1 = n_2 = 200$ discretization points for a Frank copula with parameters $\theta \in [-200, 200]$.

Constructing the Generalized Gaussian Copula Functional Copula for Target Frank Copula

The Frank copula has generator inverse equal to

$$-\frac{1}{\theta}\log\Big(1+\exp(-t)(\exp(-\theta)-1)\Big). \tag{6.35}$$

The density in the $d = 3$ case therefore takes the form,

$$C(u_1,u_2,u_3;\theta) = -\frac{1}{\theta}\log\Big(1+\frac{(\exp(-\theta u_1)-1)(\exp(-\theta u_2)-1)(\exp(-\theta u_3)-1)}{(\exp(-t)-1)^2}\Big),\ \ \theta \in \mathbb{R}\backslash\{0\}. \tag{6.36}$$

We demonstrate how one can utilise the copula mapping applied to eq. (6.36) to study the local effect of dependence induced in different regions of the state space, arising as one varies the parameter θ of the Frank copula model. Specifically we calculate

$$\mathbb{E}[U_1-\mathbb{E}[U_1], U_1-\mathbb{E}[U_1], U_2-\mathbb{E}[U_2]], \tag{6.37}$$

and

$$\mathbb{E}[U_1-\mathbb{E}[U_1], U_2-\mathbb{E}[U_2], U_3-\mathbb{E}[U_3]], \tag{6.38}$$

which can be computed using

$$\sum_{\mathbf{u}} c^{(n)}(\mathbf{u})(\mathbf{u}-\hat{\mathbf{u}}), \quad \text{for all } \mathbf{u} \in \mathbf{U} = \bigotimes_{k=1}^{3} \mathbf{U}^k, \tag{6.39}$$

or equivalently

$$\sum_{u^{(1)}}\sum_{u^{(2)}}\sum_{u^{(3)}} c^{(n)}(u^{(1)},u^{(2)},u^{(3)})(u^{(1)}-\hat{u}, u^{(2)}-\hat{u}, u^{(3)}-\hat{u}),\ \text{ for all } \mathbf{u} = (u^{(1)},u^{(2)},u^{(3)}) \in \mathbf{U}, \hat{u} = \frac{1}{2}. \tag{6.40}$$

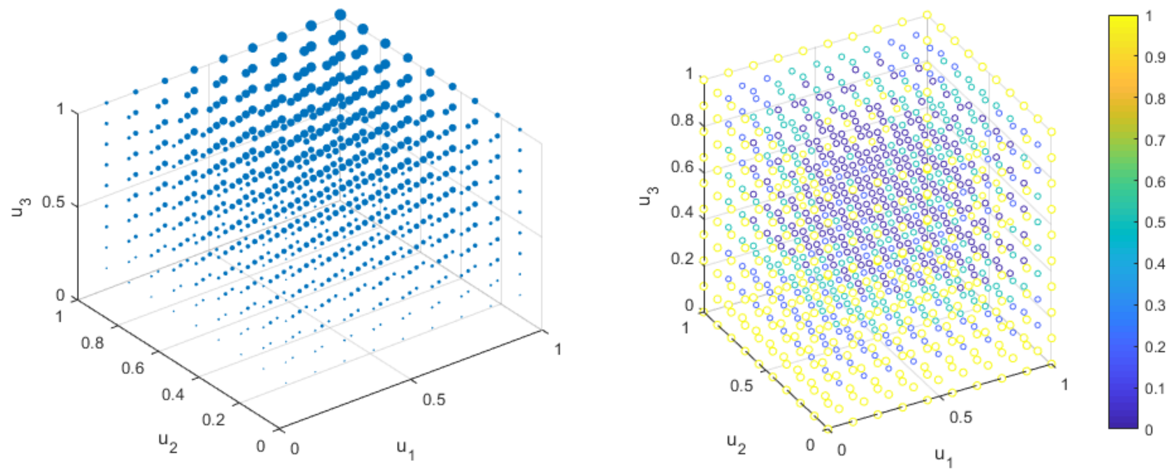

Figure 6.6: The left plot reports the approximated 3-dimensions Frank copula distribution $C : \mathbf{U} = \bigotimes_{k=1}^{3} \mathbf{U}^k \rightarrow [0,1]$ as per eq. (6.36), with parameter $\theta = 6$ and $n_1 = n_2 = n_3 = 20$. The scatter plot on the right report the mapped local Gaussian correlation $\rho(\mathbf{u}) \in [-1,1]$, for all $\mathbf{u} \in \mathbf{U}$. Note that for a Frank copula parameter $\theta = 6$, we have only positive intermediate tail dependence, which is evident after performing the local Gaussian mapping.

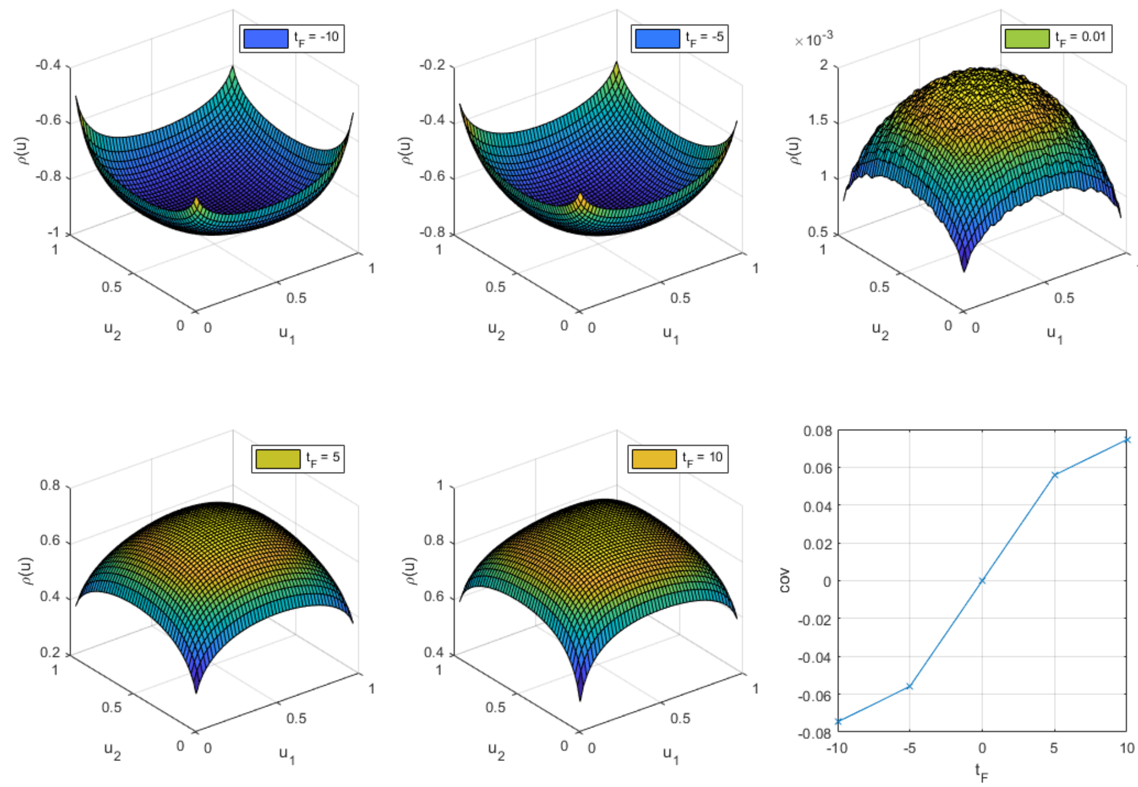

Figure 6.7: The plots report the mapped local Gaussian correlation $\rho(\mathbf{u}) : \bigotimes_{k=1}^{2} \mathbf{U}^k \to [-1, 1]$ for two dimensional Frank copulas with parameters $\theta = \{-10, -5, 0.01, 5, 10\}$. It is evident the symmetry of the local Gaussian correlation when the Frank copula parameter θ has opposite sign and this is confirmed in the last bottom right plot where we report the Frank copula covariance $cov = \mathbb{E}[(U_1 - \mathbb{E}[U_1])(U_2 - \mathbb{E}[U_2])]$. The number of discretization points used in this calculation is $n_1 = n_2 = 200$.

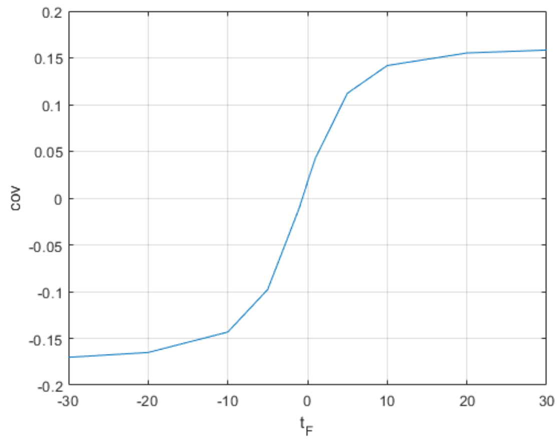

Figure 6.8: The plot reports the behaviour of the covariance for a mapped 3-dimensions Frank copula with parameters $\theta = \{-10, -5, 0.01, 5, 10\}$. The covariance $cov = \mathbb{E}[(U_1 - \mathbb{E}[U_1])(U_2 - \mathbb{E}[U_2])(U_3 - \mathbb{E}[U_3])]$ is computed using eq. (6.40). The covariance reflects the symmetry of the Frank copula. The number of discretization points used in this calculation is $n_1 = n_2 = n_3 = 200$.

Next, as with the case of the Spearman's Rho rank correlation, we now repeat the studies of the Kendall's tau approximation to the multivariate rank correlation for the skewed Student's t-copula and the Archimax copula comprised of Clayton generator and Pickands' distortion function, the results are presented in Figure 6.9 and Figure 6.10.

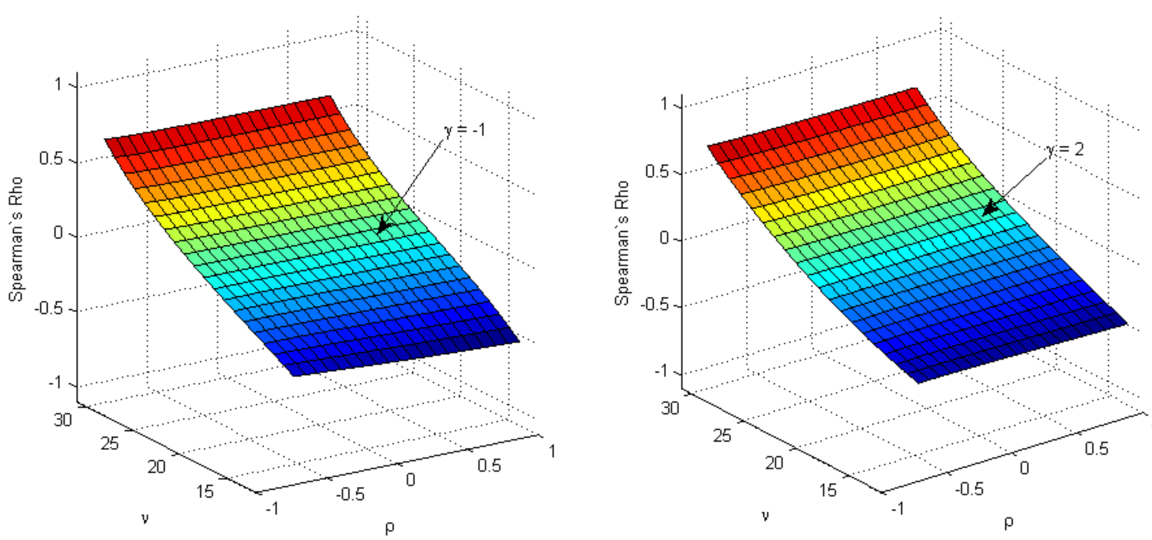

Figure 6.9: Approximated Kendall's tau $\kappa^{(n)}(C^{(n)}; \mathbf{U})$ for a skewed t-copula with parameters ν, γ and correlation ρ. The computation of eq. (6.30) has been done over the unit square $\mathbf{U} = \mathbf{U}^1 \otimes \mathbf{U}^2$, with $n_1 = n_2 = 200$ discretization points.

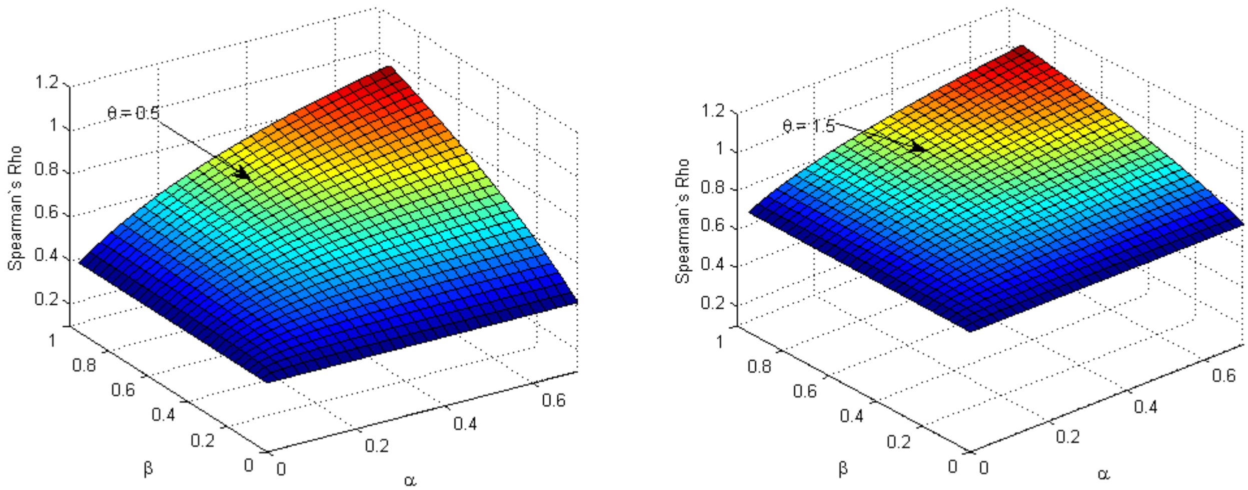

Figure 6.10: Approximated Kendall's tau $\kappa^{(n)}(C^{(n)}; \mathbf{U})$ for an Archimax copula with Pickands parameters (α, β) and Clayton parameter θ. The computation of eq. (6.30) has been done over the unit square $\mathbf{U} = \mathbf{U}^1 \otimes \mathbf{U}^2$, with $n_1 = n_2 = 200$ discretization points.

6.3.2 Multivariate Directional Dependence Approximations

After considering these multivariate copula specifications of rank correlations, one can also consider copula specifications of notions of directional dependence as detailed in for instance the notions of intermediate directional dependence in the 3-dimensional context, see details in Nelsen [2002]. We will denote such dependence measures as rho-directional dependence, say in for instance the 3-dimensional setting are then easily obtained using these functional copula approximations of the rank correlations as follows.

Proposition 19 (Functional Copula Approximations for 3-Copula ρ-Directional Dependence)**.** *Let $C(\mathbf{u})$, $\mathbf{u} = (u_1, u_2, u_3)$ a 3-dimensional copula distribution function with continuous marginals $\boldsymbol{X} = (X_1, X_2, X_3)$ uniform on $[0,1]^3$. Then for any direction $(\alpha_1, \alpha_2, \alpha_3)$ characterised by the vector components $\alpha_i \in \{-1, 1\}$ for $i \in \{1, 2, 3\}$, one has the ρ-directional dependence, expressed by a discretized copula function $C^{(n)}$, is given by*

$$\begin{aligned}\rho_{X_1,X_2,X_3}^{(n_1,n_2,n_3)(\alpha_1,\alpha_2,\alpha_3)} &= \frac{\alpha_1\alpha_2\rho_{X_1,X_2}^{(n_1,n_2)} + \alpha_2\alpha_3\rho_{X_2,X_3}^{(n_2,n_3)} + \alpha_3\alpha_1\rho_{X_3,X_1}^{(n_3,n_1)}}{3} \\ &+ \alpha_1\alpha_2\alpha_3\frac{\left[\rho_{X_1,X_2,X_3}^{(n_1,n_2,n_3)}\right]^+ - \left[\rho_{X_1,X_2,X_3}^{(n_1,n_2,n_3)}\right]^-}{2}\end{aligned} \tag{6.41}$$

with pairwise Spearman's rho obtained via the functional copula approximations given above *in Proposition 17 and*

$$
\begin{aligned}
\left[\rho_{X_1,X_2,X_3}^{(n_1,n_2,n_3)}\right]^{+}\left(C^{(n_1,n_2,n_3)}\right) &= 8\left|\prod_{i=1}^{3}(u_i - u_i')\right| \sum_{\boldsymbol{u}\in \boldsymbol{U}\cap\Omega} \overline{C}^{(n_1,n_2,n_3)}(\boldsymbol{u}) - 1, \\
\left[\rho_{X_1,X_2,X_3}^{(n_1,n_2,n_3)}\right]^{-}\left(C^{(n_1,n_2,n_3)}\right) &= 8\left|\prod_{i=1}^{3}(u_i - u_i')\right| \sum_{\boldsymbol{u}\in \boldsymbol{U}\cap\Omega} C^{(n_1,n_2,n_3)}(\boldsymbol{u}) - 1.
\end{aligned}
\tag{6.42}
$$

Remark 10. *The eight vectors which characterize directions* $(\alpha_1, \alpha_2, \alpha_3)$ *where* $\alpha_i \in \{-1, 1\}$ *for* $i \in \{1, 2, 3\}$ *in* $[0, 1]^3$ *allow one to utilise the* ρ*-directional dependence to measure directional dependence in different quadrants.*

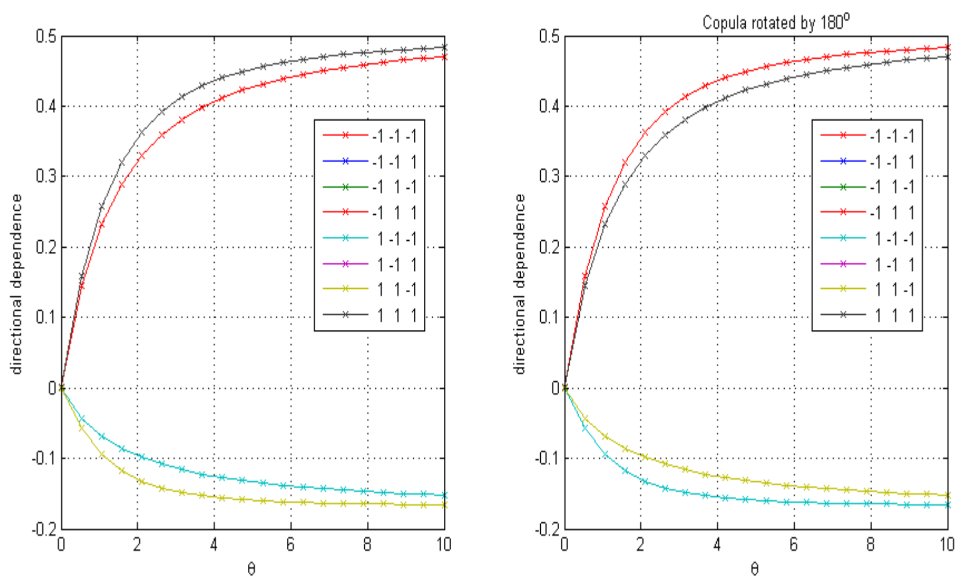

Figure 6.11: Approximated Spearman's Rho directional dependence for a 3-D Clayton copula with parameter θ computed using eq. (6.42), over the unit cube $\mathbf{U} = \bigotimes_{i=1}^{3} \mathbf{U}^i$, with $n_1 = n_2 = n_3 = 60$. The plot on the left displays the set of directional dependence curves for a standard 3-D Clayton copula, each of them is a different permutation of three distinct direction coefficients $(\alpha_1, \alpha_2, \alpha_3)$. The plot on the right reports directional dependence curves for a 3-D Clayton copula rotated by 180^o. The rotation of the copula imposes a change in tail dependence that is accurately reflected by the directional dependence curves. This behaviour is shown in the above plots and specifically by the red and black curves, which interchange values.

6.3.3 General Copula Approximations for Intermediate and Asymptotic Tail Dependence

The joint behaviour of extremes for a multivariate copula distribution often are directly influenced through a non-trivial and often non-linear relationship between the copula parameter(s), the dimension and the choice of copula and direction in which extremes are considered. For instance, if we consider extremes in the direction of some angle θ, special cases often get singled out, such as $\theta \in \{\pi/4, 5\pi/4\}$ which correspond to particular types of quadrant dependence and are typically known as upper and lower intermediate and asymptotic tail dependencies.

Tail dependence provides one approach to quantification of the dependence in extremes of a multivariate distribution. Traditionally this notion of dependence was considered from a pairwise construction due to tractability of expressions for the pairwise construction when applied to copula models. However there is no reason to restrict this notion to just pairwise analysis and below we consider first the pairwise definition and then the generalized definition for d-variate random vectors.

In addition, there is also interest in studying related functions known as tail dependence functions and tail orders. In some cases the tail dependence has been specified, for instance by De Luca & Rivieccio [2012] and Li [2009], according to the formulation where for an arbitrary d-variate cases, such that $d > 2$ one may quantify the tail dependence present between sub-vector partitions of the multivariate random vector with regard to joint tail dependence behaviours. Such a specification can be made as shown in Definition 48.

Definition 48 (Multivariate Tail Dependence)**.** *Let $\mathbf{X} = (X_1, ..., X_d)$ be a d-dimensional random vector with marginal distribution functions $F_1, ..., F_d$ and copula C.*

1. *One may define the coefficient of multivariate upper tail dependence (upper orthant dependence) by:*

$$\begin{aligned}\lambda_u^{1,\dots,h|h+1,\dots,d} &= \lim_{\nu\to 1-} P\left(X_1 > F^{-1}(\nu),\dots,X_h > F^{-1}(\nu)|X_{h+1} > F^{-1}(\nu),\dots,X_d > F^{-1}(\nu)\right)\\ &= \lim_{\nu\to 1-}\frac{\bar{C}_n(1-\nu,\dots,1-\nu)}{\bar{C}_{n-h}(1-\nu,\dots,1-\nu)}\end{aligned} \tag{6.43}$$

   where $\bar{C}$ is the survival copula of C.

2. *One may define the coefficient of multivariate lower tail dependence (lower orthant dependence) by:*

$$\begin{aligned}\lambda_l^{1,\dots,h|h+1,\dots,d} &= \lim_{\nu\to 0+} P\left(X_1 < F^{-1}(\nu),\dots,X_h < F^{-1}(\nu)|X_{h+1} < F^{-1}(\nu),\dots,X_d < F^{-1}(\nu)\right)\\ &= \lim_{\nu\to 0+}\frac{C_n(\nu,\dots,\nu)}{C_{n-h}(\nu,\dots,\nu)}\end{aligned} \tag{6.44}$$

   where, h is the number of variables conditioned on (from the d considered).

Alternative definitions of multivariate tail dependence are also popular in the literature, for instance the specification given in Klüppelberg *et al.* [2008] where they define the tail dependence function according to Definition 49.

Definition 49 (Tail Dependence Function of a Multivariate Distribution)**.** *Consider a d-dimensional random vector $\mathbf{X} = (X_1, ..., X_d) \in \mathbb{R}^d$, $d \geq 2$ with marginal distribution functions $F_1, ..., F_d$, then the tail dependence function is given by*

$$\lambda(\mathbf{u}) := \lambda\left(x_1, x_2, \dots, x_d\right) = \lim_{t\to 0}\frac{1}{t}\mathbb{P}r\left[\overline{F}_1(X_1) \leq tx_1, \dots, \overline{F}_2(X_d) \leq tx_d\right], \quad \mathbf{x} \in \mathbb{R}^d. \tag{6.45}$$

In Joe *et al.* [2010] they studied properties of tail dependence functions and conditional tail dependence functions, which they defined via the joint distribution on the unit d-dimensional hyper-cube known as a copula, for a multi-variate distribution. The definition adopted in Joe *et al.* [2010] for the upper and lower tail dependence functions differs to that provided in Definition 49 via the fact that each marginal can go to the limit at different rates according to the functions:

- Lower Tail Dependence Function: the tail dependence function for the copula distribution $C(u_1, \ldots, u_d)$ is given by

$$\lambda_l(t; C(\mathbf{u})) = \lim_{t \downarrow 0} \frac{C(tu_1, \ldots, tu_d)}{t}, \quad \forall\, \boldsymbol{u} = (u_1, \ldots, u_d) \in \mathbb{R}_+^d \tag{6.46}$$

- Upper Tail Dependence Function: the tail dependence function for the copula distribution $C(u_1, \ldots, u_d)$ is given by

$$\lambda_u(t; C(\mathbf{u})) = \lim_{t \uparrow 1} \frac{\overline{C}(tu_1, \ldots, tu_d)}{t}, \quad \forall\, \boldsymbol{u} = (u_1, \ldots, u_d) \in \mathbb{R}_+^d \tag{6.47}$$

  where we denote by $\overline{C}(u_1, \ldots, u_d) = C(1 - u_1, \ldots, 1 - u_d)$ the survival copula distribution.

The existence of such limits in the definition of the tail dependence functions can be linked to existence of multivariate regular variation on the copula distribution tails.

Given these definitions, one may now define the following intermediate and extremal tail dependence measures

Definition 50 (Intermediate and asymptotic Directional Tail Dependence Functions)**.** *Consider the d-dimensional random vector* $\mathbf{X} = (X_1, X_2, \ldots, X_d) \in \mathbb{R}^d$ *with joint distribution characterized by marginals* $F_i(x_i)$ *and joint copula distributions* $C(F_1(x_1), \ldots, F_d(x_d))$*. The intermediate tail dependence function for a k-dimensional sub-vector* $\boldsymbol{y}$*,* $k \leq d$ *for some angular region* $\Omega = \prod_{i=1}^{d-1} [\theta_i, \theta_i'] \subseteq \prod_{i=1}^{d-2} [0, \pi] \times [0, 2\pi]$ *and a radius r by*

$$\widetilde{\lambda}(\boldsymbol{x}, k, d, \Omega) = \mathbb{P}r\left[X_1 \leq F_1^{-1}(u(r)), \ldots, X_k \leq F_k^{-1}(u(r)) \middle| \mathcal{H}(\boldsymbol{x}, k, d, \Omega)\right] \tag{6.48}$$

where $\mathcal{H}(\boldsymbol{x}, k, d, \Omega) := \left\{X_{k+1} \leq F_{k+1}^{-1}(u(r)), \ldots, X_d \leq F_d^{-1}(u(r))\right\} \cap \{\boldsymbol{x} \in \mathcal{C}_{d-1}(\boldsymbol{\theta} \in \Omega)\}$*. The asymptotic tail dependence coefficient is then given by the limiting behaviour of this intermediate tail dependence function according to*

$$\lambda(k, d, \Omega) = \lim_{r \to \infty} \mathbb{P}r\left[X_1 \leq F_1^{-1}(u(r)), \ldots, X_k \leq F_k^{-1}(u(r)) \middle| \mathcal{H}(\boldsymbol{x}, k, d, \Omega)\right] \tag{6.49}$$

As with the other concordance measures, we may also approximate these measures using the developed functional copula mapping approach. In the following example we illustrate the behaviour of the tail dependence function specified directly above, in the 2-dimensional case of the Clayton copula.

If we consider the two dimensional case with random variables X_1 and X_2 with distributions $F_i, i = 1, 2$ and a density describing their dependence on the unit cube known as a copula C. Then we may define for $\Omega = [0, \pi/2]$ the upper tail dependence coefficient by:

$$\lambda_u := \lim_{u \uparrow 1} \Pr\left[X_2 > F_2^{-1}(u) \,|X_1 > F_1^{-1}(u)\right] = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \tag{6.50}$$

and similarly we define for $\Omega = [-\pi, -\pi/2]$ the coefficient of lower tail dependence given by:

$$\lambda_l := \lim_{u \downarrow 0} \mathbb{P}\mathrm{r}\left[X_2 \leq F_2^{-1}(u) \,|X_1 \leq F_1^{-1}(u)\right] = \lim_{u \downarrow 0} \frac{C(u,u)}{u} \tag{6.51}$$

Note that $\widetilde{C}(1-u, 1-u) = 1 - 2u - C(u,u)$ is known the survival copula of C. The above relationships show that the upper tail dependence coefficients of copula C is also equal to the lower tail dependence coefficient of the survival copula of C. Analogously, the lower tail dependence coefficient of copula C is equivalent to the upper tail dependence coefficient of the survival copula.

Remark 11. *Similar to rank correlations, the tail dependence coefficient is a simple scalar measure of dependence that depends on the copula of two random variables but not on their marginal distributions.*

Following we show how it is possible to compute coefficients of intermediate tail dependence using the copula approximation framework proposed in the previous section. After selecting few representative target copula functions we use the mapping of eq. (5.31) to obtain the approximated copula distributions $C^{(n)}$ which are function of the local Gaussian correlations. In the examples we used $n = 200$ as a number of discretization point across the dimensions of the unit hypercube, which is the support of the mapped and discretized copula distribution functions. Specifically in Figure 6.12 we calculate the approximated intermediate and asymptotic lower tail directional dependence

$$\lambda_l^{(n)}(C^{(n)}) = \frac{C^{(n)}(\mathbf{u},\mathbf{u})}{\mathbf{u}}, \quad \mathbf{u} = \mathbf{0}, \dots, \mathbf{1}, \tag{6.52}$$

for a mapped Clayton copula $C^{(n)}$ with parameters $\theta = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. In Figure 6.13 we calculate the intermediate and asymptotic upper tail for a skewed Student's t-copula with parameters $\rho \in [-1, 1]$, $\nu = 11$ and $g = -1$. In Figure 6.14 we calculate the approximated intermediate and asymptotic upper tail (UT) coefficient

$$\lambda_u^{(n)}(C^{(n)}) = \frac{1 - 2\mathbf{u} - C^{(n)}(\mathbf{u},\mathbf{u})}{1 - \mathbf{u}}, \quad \mathbf{u} = \mathbf{0}, \dots, \mathbf{1}, \tag{6.53}$$

for an approximated two dimensional mapped Archimax Copula $C^{(n)}$ with Clayton generating function $\varphi(x, \theta)$ and Pickands function $\mathcal{A}(t)$ with parameters $\theta = 3$, $a = \{0, 0.3, 0.6, 0.9\}$ and $b \in [0, 2]$. The examples highlight how the proposed framework can deal with complex dependence structure like the case of the Clayton copula with Archimax distortion function.

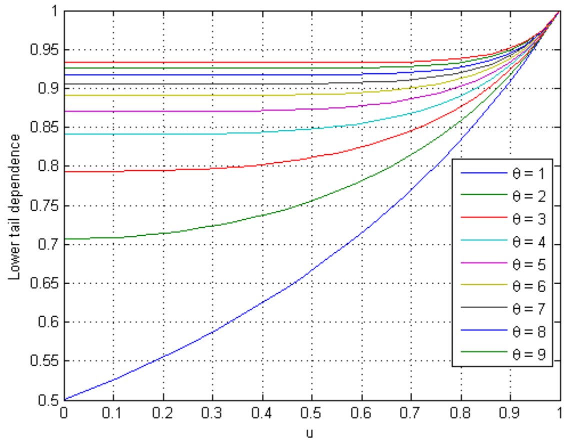

Figure 6.12: Intermediate and asymptotic approximated lower tail (LT) coefficient $\lambda_l^{(n)}(C^{(n)})$ for a Clayton copula with parameter θ computed using eq. (6.52). For $u \to 0$ the value of the LT coefficient correspond to the asymptotic one.

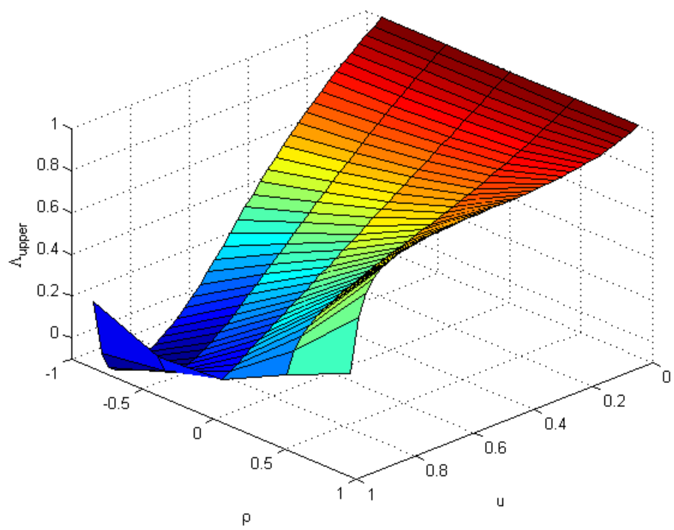

Figure 6.13: Intermediate and asymptotic approximated upper tail (UT) coefficient $\lambda_u^{(n)}(C^{(n)})$ for a skewed Student's t-copula with parameters $g = -1$, $\nu = 11$ and correlation parameter $\rho \in [-1, 1]$, computed using eq. (6.53). For $u \to 1$ the value of the UT coefficient corresponds to the asymptotic one.

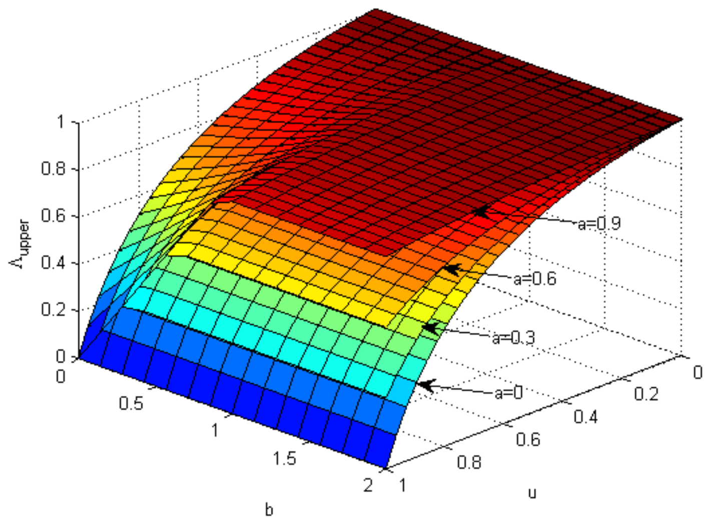

Figure 6.14: Intermediate and asymptotic upper tail (UT) coefficient for a 2-D Archimax Copula with Clayton generating function $\varphi(x, \theta)$ and Pickands function $\mathcal{A}(t)$. The parameters are $\theta = 3$, $a = [0, 0.3, 0.6, 0.9]$ and $b \in [0, 2]$. Note that when $(a, b) = (0, 0)$ then the Archimax copula is equivalent to the Clayton one with the same θ parameter that is equal to 3 in this specific example. For $u \to 1$ the value of the UT coefficient corresponds to the asymptotic one.

6.4 Conclusions

In this chapter we proposed a class of new methods that involves a highly accurate and computationally efficient procedure to evaluate concordance measures for a given copula, applicable when sampling from the copula is not easily achieved or the curse of dimensionality makes computation using the copula function impossible. Our methodology is based on the numerical procedure introduced in Chapter 5 which allows us to map any copula function to a generalized Gaussian copula function and to visualize a given copulas' characteristics and better quantify the local correlation generated by the copula function. We showed how to reconstruct maps of concordance measures locally in all regions of the state space for any range of copula parameters. We believe this technique will be a valuable tool for practitioners to understand better the behaviour of copula models and associated concordance measures expressed in terms of these copula models and also to capture, measure and quantify different notions of dependence in stochastic processes. We demonstrated the simplicity and intuition behind the concordance measures computation by copula mapping through illustrative examples.

Chapter 7

Conclusion

We have presented numerical methods and algorithms based on CTMC which allow us to compute the approximate solution of a multidimensional SDE and characterize its correlation structure.

In Chapter 3 we introduced concepts and elementary notions about continuous time Markov chains (CTMC) and the way can be used to both construct an approximate solution to an SDE. The proposed approximation method is based on linear and tensor algebra. The basic theory on Markov chains was accompanied by illustrative examples and Matlab code listings.

In Chapter 4 we gave a full characterization of multidimensional approximation of correlated diffusion by CTMCs, with the main results being the specification of the conditional infinitesimal generator matrix in a tensor space given by Proposition 5. We also showed that the mimicking process generates an approximated solution which converges in distribution to a multidimensional diffusion process and analysed its weak convergence in Section 4.3.4.

In Chapter 5 we first proposed and then investigated the novel concept of local copula operator derived from a CTMC approximation of diffusions and presented a numerical procedure to map any copula function to a generalized Gaussian copula function. The proposed copula mapping is entirely based on tensor algebra and is part of a new modelling framework we also introduced consisting of the tensor approximation of the infinitesimal generator associated to multidimensional correlated diffusions. The mapping procedure has been described in Section 5.2.3 and we proved that there is always existence of a solution to the mapping problem in Proposition 11.

We demonstrated the simplicity and intuition behind the copula mapping through illustrative examples and all the copula functions mapping results proved to be exact up to the tensor space local discretization error and the error due to approximating the generator matrix exponential.

In Chapter 6 we also showed how it is straightforward to study and quantify within the proposed framework concordance measures of dependence. This allowed us to visualize copula's characteristics and better quantify the local correlation generated by the copula function. In fact copulas' parametric expressions are neither intuitive nor reflecting the dependence structure that create. The proposed methodology proved to be an invaluable tool to understand dependence of a given copula.

There are several open questions and applications related to this thesis. First, it would be instructive to further investigate the scalability properties of the introduced mimicking approximation, based on linear and tensor algebra, when the number of correlated marginals to approximate by CTMCs is very large ($\gg 1000$), being a natural application a search engine exhibiting local correlations or a large scale stochastic control problem. Second it would be important to study ways to improve the convergence rate, which can be achieved either by constructing the mimicking CTMC using higher moments or by transformation of the state space, such as unary transformations or interpolation by mean of transcendental maps.

The third development can be related to a financial application and specifically to the classification of stochastic volatility models. It would be beneficial, especially for the financial practitioners' community, to have a deeper insight into our proposed methodology and formulate the approximate solution of existing and new stochastic volatility models by means of CTMCs along the lines of this thesis. This would constitute an important and useful numerical exercise, leading to a better insights of the computational advantages associated to our methodology in comparison to existing numerical schemes for stochastic volatility frameworks.

Appendix A

Matlab® Code

This appendix reports the code listings for the examples of Chapter 3.

A.1 Examples

```
% Copyright Antonio Dalessandro

% 1-D example of infinitesimal generator matrix
% for any generalized diffusion
% Specifically this is the Black-Scholes operator

function [w, L] = markovGenerator(S, S0, vol, rate, beta, T)

%  Generator for the chain that approximates
%  diffusion: dS/S = rate  dt + vol S^beta dW

N = length(S);       % number of points in the state-space
L(1,:)=0; L(N,:)=0;  % absorbing boundaries

for x = 2:N-1
    dSp = S(x+1)-S(x);
    dSm  = S(x-1)-S(x);

    muS = rate;
    volS = vol^beta;

    A = [dSp dSm; dSp^2 dSm^2];
    b = [muS; volS^2];
    lambda = A\b;
    L(x,x+1) = lambda(1);
    L(x,x-1) = lambda(2);
    L(x,x)   = -(L(x,x-1)+L(x,x+1));
end
v = zeros(N, 1);
v(S0)= 1;

W = expm((L)*T);
w = W'*v;
end
```

Listing A.1: ©MATLAB code for the infinitesimal generator matrix of the 1D Black-Scholes operator.

```
% Copyright Antonio Dalessandro

% Standard Black-Scholes option formula
function [C, P]= option_bs(S,X,r,sigma,T)

    % C: call price
    % P: put price
    % S: stock price at time 0
    % X:  strike price
    % r: risk-free interest rate
    % sigma: volatility of the stock price measured as annual standard ...
        deviation
    % Black-Scholes formula:
    %  C = S N(d1) - Xe-rt N(d2)
    %  P = Xe-rt N(-d2) - S N(-d1)

    % for call
    d1 = (log(S/X) + (r + 0.5*sigma^2)*T)/sigma/sqrt(T);
    d2 = d1 - sigma*sqrt(T);
    N1 = normcdf(d1);
    N2 = normcdf(d2);

    C = S*N1 - X*exp(-r*T)*N2;

    % for put
    N1 = 0.5*(1+erf(-d1/sqrt(2)));
    N2 = 0.5*(1+erf(-d2/sqrt(2)));

    P = X*exp(-r*T)*N2 - S*N1;
end
```

Listing A.2: ©MATLAB code for the closed form Black-Scholes option pricing formula.

```
% Copyright Antonio Dalessandro

% 1-D example of call option pricing using the Black-Scholes infinitesimal
% operator matrix

function main

n = 100;

% grid can be generated to be non-homogeneous and also
% with specific points of interest
S = linspace(0,20,n);

% SPOT AND STRIKE
K = 5; % strike
SpotIdx = max(find(S≤K));
Spot = S(SpotIdx);
K = Spot; % ATM

vol = .2;
T = 2;
rate = 0.05;
beta = 1;
disc = exp(-rate * T);

% here we calculate the transition probability using the infinitesimal
% generator matrix L
[Pr, L] = markovGenerator(S, SpotIdx, vol, rate, beta,T);

```

```
payoff = max(S - K,0)';
plot(S,payoff,'r');
hold on
plot(S,Pr)
grid on
ylim([0,.07])

CallPrice = disc*(Pr'*payoff);

% B-S CLOSED FORM (this is just a standard BS closed form option pricing ...
    function)
[C]= option_bs(Spot,K,rate,vol,T);

fprintf('\n')
fprintf('---------------- BS Call Options -----------------\n')
fprintf('time to maturity (Years):')
disp(T);
format long
fprintf('strike:')
disp(K);
fprintf('Call Prices: \n')
fprintf('Closed Form :    ')
disp(C)
fprintf('Operator Price:  ')
disp(CallPrice)

end
```

Listing A.3: ©MATLAB code for the Black-Scholes option pricing through the approximated infinitesimal generator matrix.

```
% Copyright Antonio Dalessandro
% state space normalization

close all
clear all

n = 200;
S = linspace(0.001,10,n);
logS = log(S);

K    = 4;
vol  = .15;
T    = 1;
rate = 0.0;
beta = 1;


SpotIdx = sum(S≤K);
[Pr, L] = markovGenerator(logS, SpotIdx, vol, rate - .5*vol^2, beta,T);

shift = rate - .5*vol^2;
SpotIdx_r = sum((logS ≤ log(K)+ shift*T));

[Pr_2, L] = markovGenerator((logS - shift)/vol , SpotIdx_r , 1, 0, beta,T);

plot(S,Pr);
hold on
plot(S,Pr_2,'rx'); grid on
legend('N(\mu - 0.5* \sigma^2, \sigma)', 'N(0,1)')
```

Listing A.4: ©MATLAB code that illustrates how it is possible to build equivalent infinitesimal generators of a diffusion. In this case the process dynamics is log-normal and we apply state space transformations.

```
close all; clear all
% INPUT parameters
T = 7;
n1 = 80; n2 = 200;
X1 = linspace(0.001,25,n1); % discretization state space 1
X2 = linspace(0.001,7,n2); % discretization state space 2
S1 = log(X1); S2 = log(X2);

F = 3;
v0 = 1;

X1_0 = sum(X1 <= F);
X2_0 = sum(X2 <= v0);
beta = 1;
RR = [0, -.3];

% volatility process
G2 = zeros(n2);
v2 = .15; r = .05;
mu2 = r - .5*v2^2;

for i = 2:n2-1
h2 = S2(i) - S2(i-1);
G2(i,i+1) = (v2^2)/(2*h2^2) + mu2/(2*h2);
G2(i,i-1) = (v2^2)/(2*h2^2) - mu2/(2*h2);
G2(i,i) = -(G2(i,i+1) + G2(i,i-1));
end
P2 = expm(G2*T);
B = P2(X2_0,:);
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
for ir = 1:length(RR)
corrI = RR(ir);
corrMtx = ones(n1,n2)*corrI; % correlation
%%%%%%%%%%%%%%%%%%%%%%%%%%%
v1 = .2;
mu1 = r - .5*v1^2;

G1 = zeros(n1);
P1cP2 = zeros(n1,n1,n2);
for j = 1:n2
for i = 2:n1-1
h1 = S1(i+1) - S1(i);

mcc = mu1 + corrMtx(i,j)*(v1/v2)*((S2(j)- S2(X2_0))/T - mu2 );
cv = sqrt((1-corrMtx(i,j)*corrMtx(i,j))*v1*v1);

G1(i,i+1) = (cv^2)/(2*h1^2) + mcc/(2*h1);
G1(i,i-1) = (cv^2)/(2*h1^2) - mcc/(2*h1);
G1(i,i) = -(G1(i,i+1) + G1(i,i-1));
end
P1cP2(:,:,j) = expm(sparse(G1)*T);
end
A = P1cP2(X1_0,:,:);
P = repmat(B,n1,1).*reshape(A,n1,n2);
m2(:,ir) = sum(P,1);
m1(:,ir) = sum(P,2);
```

```
end

%%
sum(sum(P))
plot(X1,m1(:,1)); hold on
plot(X1,m1(:,2),'r-X')
grid on
```

Listing A.5: ©MATLAB code for the operator matrix associated to paired lognormal processes with instantaneous conditional moments as per eq. (3.29). In this case the joint processes are log-normal and we apply state space transformations as per section 3.3 .

```

% INPUT parameters

clear all
close all
T = 1;
mu1 = .02; v1 = .4; % marginal 1 drift and vol
mu2 = 0.03; v2 = .4; % marginal 2 drift and vol
n1 = 40; S1 = linspace(-2,2,n1); % discretization state space 1
n2 = 40; S2 = linspace(-2,2,n2); % discretization state space 2


corrI = -.6;

corrMtx = ones(n1,n2)*corrI; % correlation
h1 = S1(3) - S1(2);
h2 = S2(3) - S2(2);
G1 = zeros(n1);


% identity matrices
I1 = eye(n1);
I2 = eye(n2);

I2_G1 =  zeros(n1*n2);

%P1cP2 = zeros(n1,n1,n2);
for j = 1:n2
for i = 2:n1-1
mcc = mu1 + corrMtx(i,j)*(v1/v2)*(S2(j) - mu2);
cv = sqrt((1-corrMtx(i,j)*corrMtx(i,j))*v1*v1);
G1(i,i+1) = (cv^2)/(2*h1^2) + mcc/(2*h1);
G1(i,i-1) = (cv^2)/(2*h1^2) - mcc/(2*h1);
G1(i,i) = -(G1(i,i+1) + G1(i,i-1));
end
I2_G1((j-1)*n1 + 1:(j-1)*n1 + n1,(j-1)*n1 + 1:(j-1)*n1 + n1) = G1;
%P1cP2(:,:,j) = expm(sparse(G1)*T);
end
G2 = zeros(n2);
for i = 2:n2-1
G2(i,i+1) = (v2^2)/(2*h2^2) + mu2/(2*h2);
G2(i,i-1) = (v2^2)/(2*h2^2) - mu2/(2*h2);
G2(i,i) = -(G2(i,i+1) + G2(i,i-1));
end
%P2 = expm(G2*T);
G2_I1 = kron(G2, I1);

G = G2_I1 + I2_G1;
P = expm(sparse(G));
```

```
%P_split = expm(sparse(G2_I1))*expm(sparse(I2_G1));

%%
% spot calculation
X1_0 = n1/2;
X2_0 = n2/2;
xx = n1*(X2_0 - 1) + X1_0 ;

% full generator
Pm = reshape(P(xx,:),n1,n2);

% rank-1 approximation
%Ps2 = reshape(P_split(xx,:),n1,n2);
% A = P1cP2(X1_0,:,:);
% B = P2(X2_0,:);
% Ps = repmat(B,n1,1).*reshape(A,n1,n2);

% plot join densities
[X1, X2] = meshgrid(S1,S2);
% surf(X1,X2,Ps');
% figure
surf(X1,X2,Pm');
```

Listing A.6: ©MATLAB code for the Computation of the approximated transition density for two coupled diffusions by tensor approximation .

Bibliography

Aas, Kjersti, Czado, Claudia, Frigessi, Arnoldo, & Bakken, Henrik. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, **44**(2), 182–198.

Asmussen, Soren. 2003. *Applied Probability and Queues*. Springer Science & Business Media.

Asmussen, Soren, & Glynn, Peter W. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag New York.

Bacigál, T, & Mesiar, R. 2012. *3-dimensional Archimax copulas and their fitting to real data*.

Baldi, Paolo. 2017. *Stochastic Calculus*. Springer International Publishing.

Balkema, G., Embrechts, P., & Nolde, N. 2012. Sensitivity of the limit shape of sample clouds from meta densities. *Bernoulli*, 13861404.

Balkema, G., Embrechts, P., & Nolde, N. 2013. The Shape of Asymptotic Dependence. *In: Shiryaev A., Varadhan S., Presman E. (eds) Prokhorov and Contemporary Probability Theory*.

Billingsley, Patrick. 2013. *Convergence of probability measures*. John Wiley & Sons.

Black, Fischer, & Scholes, Myron. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81**(3), 637–54.

Blomqvist, Nils. 1950. On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, 593–600.

Brigo, Damiano, & Mercurio, Fabio. 2008. *Interest Rate Models, Theory and Practice*. Springer.

Brigo, Damiano, Pallavicini, Andrea, & Torresetti, Roberto. 2010a. *Credit models and the crisis: A journey into CDOs, copulas, correlations and dynamic models*. John Wiley & Sons.

Brigo, Damiano, Pallavicini, Andrea, & Torresetti, Roberto. 2010b. *Credit models and the crisis: A journey into CDOs, copulas, correlations and dynamic models*. John Wiley & Sons.

Brunick, G. 2013. Uniqueness in law for a class of degenerate diffusions with continuous covariance. *Probab. Theory Related Fields*, 265302265302.

Brunick, G., & Shreve, S. 2013. Mimicking an Ito process by a solution of a stochastic differential equation. *Ann. Appl. Probab*, **23**(4), 1584–1628.

Brzeniak, Z., & Zastawniak, T. 1999. *Basic Stochastic Processes*. Springer-Verlag London.

Cambanis, Stamatis, Simons, Gordon, & Stout, William. 1976. Inequalities for E k (x, y) when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **36**(4), 285–294.

Capéraà, Philippe, Fougères, Anne-Laure, & Genest, Christian. 2000. Bivariate Distributions with Given Extreme Value Attractor. *Journal of Multivariate Analysis*, **72**, 30–49.

Chang, Meng-Shiuh, & Wu, Ximing. 2015. Transformation-based nonparametric estimation of multivariate densities. *Journal of Multivariate Analysis*, **135**, 71–88.

Charpentier, Arthur, Fougeres, A-L, Genest, Christian, & Nešlehová, JG. 2014. Multivariate archimax copulas. *Journal of Multivariate Analysis*, **126**, 118–136.

Cherubini, Umberto, Luciano, Elisa, & Vecchiato, Walter. 2004. *Copula methods in finance*. John Wiley & Sons.

Cruz, Marcelo G, Peters, Gareth W, & Shevchenko, Pavel V. 2015a. Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk. *Wiley, Handbook of Financial Engineering and Econometrics*.

Cruz, Marcelo G, Peters, Gareth W, & Shevchenko, Pavel V. 2015b. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. John Wiley & Sons.

Dalessandro, Antonio, & Peters, Gareth W. 2016a. Evaluating Concordance Measures via a Tensor Approximation of Generalized Correlated Diffusions. *Preprint available at SSRN*.

Dalessandro, Antonio, & Peters, Gareth W. 2016b. Functional Tensor Characterizations of Copulas and Generalized Copula Mappings. *Preprint available at SSRN*.

Dalessandro, Antonio, & Peters, Gareth W. 2017. Tensor Approximation of Generalized Correlated Diffusions and Functional Copula Operators. *Methodology and Computing in Applied Probability*, 1–35.

Daruich, Diego. 2014. *Notes on Measure Theory and Markov Processes*.

Daul, Stéphane, De Giorgi, Enrico G, Lindskog, Filip, & McNeil, Alexander. 2003. The grouped t-copula with an application to credit risk. *Available at SSRN 1358956*.

De Luca, Giovanni, & Rivieccio, Giorgia. 2012. Multivariate tail dependence coefficients for Archimedean copulae. *Pages 287–296 of: Advanced Statistical Methods for the Analysis of Large Data-Sets*. Springer.

Demarta, Stefano, & McNeil, Alexander J. 2005. The t copula and related copulas. *International statistical review*, **73**(1), 111–129.

Di Graziano, Giuseppe, & Rogers, L.G.C. 2009. A Dynamic approach to the modeling of correlation credit derivatives using Markov chains. *Int. J. Theor. Appl. Finan.*, **12**(45).

Dolati, Ali, & Úbeda-Flores, Manuel. 2006. On measures of multivariate concordance. *Journal of Probability and Statistical Science*, **4**(2), 147–164.

E. Ekström, D. Hobson, S. Janson, & Tysk, J. 2013. Can time-homogeneous diffusions produce any distribution? *Probab. Theory Related Fields*, 493520.

Eban, E., Rothschild, R., Mizrahi, A., Elidan, G., Carvalho, C., & Ravikumar, P. 2013. Dynamic Copula Networks for Modeling Real-valued Time Series. *Journal of Machine Learning Research*, **31**.

Embrechts, Paul, Lindskog, Filip, & McNeil, Alexander. 2001. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.

Ethier, Stewart N, & Kurtz, Thomas G. 2009. *Markov processes: characterization and convergence*. Vol. 282. John Wiley & Sons.

Fredricks, Gregory A, & Nelsen, Roger B. 2007. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, **137**(7), 2143–2150.

Frees, Edward W, & Valdez, Emiliano A. 1998. Understanding relationships using copulas. *North American actuarial journal*, **2**(1), 1–25.

Galton, Francis. 1894. *Natural inheritance*. Macmillan.

Gardiner, Crispin W. 2004. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer Series in Synergetics.

Genest, Christian, & Favre, Anne-Catherine. 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, **12**(4), 347–368.

Genest, Christian, & Rivest, Louis-Paul. 2001. On the multivariate probability integral transformation. *Statistics & Probability Letters*, **53**(4), 391–399.

Genest, Christian, Kojadinovic, Ivan, Nešlehová, Johanna, Yan, Jun, *et al.* 2011. A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, **17**(1), 253–275.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer.

Grimmett, G.R., & Stirzaker, D.R. 2001. *Probability and random processes*. Vol. 80. Oxford university press.

Gyöngy, I. 1986. Mimicking the one-dimensional marginal distributions of processes having an ito differential. *Probability Theory and Related Fields*, **71**, 501–516.

Hackbusch, Wolfgang. 2012. *Tensor spaces and numerical tensor calculus*. Vol. 42. Springer Science & Business Media.

Haff, Ingrid Hobæk, *et al.* 2013. Parameter estimation for pair-copula constructions. *Bernoulli*, **19**(2), 462–491.

Jacod, J., & Protter, P. 2004. *Probability Essentials*. Universitext. Springer Berlin Heidelberg.

Joe, Harry. 1990. Multivariate concordance. *Journal of multivariate analysis*, **35**(1), 12–30.

Joe, Harry. 1997. *Multivariate models and multivariate dependence concepts*. CRC Press.

Joe, Harry, Li, Haijun, & Nikoloulopoulos, Aristidis K. 2010. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, **101**(1), 252–270.

Karatzas, I., & Shreve, S. 2000. *Brownian Motion and Stochastic Calculus - Second Edition*. Springer.

Kendall, Maurice G. 1938. A new measure of rank correlation. *Biometrika*, 81–93.

Kloeden, Peter E., & Platen, Eckhard. 2011. *Numerical Solution of Stochastic Differential Equations*. Springer.

Klüppelberg, Claudia, Kuhn, Gabriel, & Peng, Liang. 2008. Semi-Parametric Models for the Multivariate Tail Dependence Function–the Asymptotically Dependent Case. *Scandinavian Journal of Statistics*, **35**(4), 701–718.

Kuang, L., Hao, F., Yang, L. T., Lin, M., Luo, C., & Min, G. 2014. A Tensor-Based Approach for Big Data Representation and Dimensionality Reduction. *IEEE Transactions on Emerging Topics in Computing*, **2**(3), 280–291.

Kuo, H.H. 2006. *Introduction to Stochastic Integration*. Universitext. Springer New York.

Kushner, H. J., & Dupuis, P. 2001. *Numerical Methods for Stochastic Control Problems in Continuous Time - Second Edition*. Springer Science & Business Media.

Li, Haijun. 2009. Orthant tail dependence of multivariate extreme value distributions. *Journal of Multivariate Analysis*, **100**, 243–256.

Li, Haijun, & Wu, Peiling. 2013. Extremal dependence of copulas: A tail density approach. *Journal of multivariate analysis*, **114**(1), 99–111.

Lukacs, E. 1958. *Characteristic Functions*. Charles Griffin & Company Limited London.

Luo, Xiaolin, & Shevchenko, Pavel V. 2010. The t copula with multiple parameters of degrees of freedom: bivariate characteristics and application to risk management. *Quantitative Finance*, **10**(9), 1039–1054.

Marchuk, G.I. 2011. *Methods of Numerical Mathematics*. Vol. 510. Springer-Verlag New York.

Mari, Dominique Drouet, & Kotz, Samuel. 2001. *Correlation and dependence*. Vol. 518. World Scientific.

McNeil, Alexander J., & Nešlehová, Johanna. 2009. Multivariate Archimedean Copulas, d-Monotone Functions and l_1-Norm Symmetric Distributions. *The Annals of Statistics*, **37**, 3059–3097.

McNeil, Alexander J, Frey, Rüdiger, & Embrechts, Paul. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools: Concepts, Techniques and Tools*. Princeton university press.

Mesiar, Radko, & Jágr, Vladimír. 2013. d-Dimensional dependence functions and Archimax copulas. *Fuzzy Sets and Systems*, **228**, 78–87.

Meucci, Attilio. 2011. A new breed of copulas for risk and portfolio management. *Risk*, **24**(9), 122–126.

Milstein, Grigori N., & Tretyakov, Michael V. 2004. *Stochastic Numerics for Mathematical Physics*. Vol. 596. Springer Berlin Heidelberg.

Morillas, Patricia Mariela. 2005. A method to obtain new copulas from a given one. *Metrika*, **61**(2), 169–184.

Nelsen, Roger B. 1999. *An introduction to copulas*. Vol. 139. Springer Science & Business Media.

Nelsen, Roger B. 2002. Concordance and copulas: A survey. *Pages 169–177 of: Distributions with given marginals and statistical modelling*. Springer.

Øksendal, Bernt. 2003. *Stochastic Differential Equations*. Springer-Verlag Berlin Heidelberg.

Patton, Andrew J. 2009. Copula–based models for financial time series. *Pages 767–785 of: Handbook of financial time series*. Springer.

Pearson, Karl. 1896. Mathematical Contributions to the Theory of Evolution.–On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs. *Proceedings of the royal society of london*, **60**(359-367), 489–498.

Rogers, L Chris G, & Williams, David. 2000. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*. Vol. 2. Cambridge university press.

Ruan, S., Swaminathan, N., & O.Darbyshire. 2014. Modelling of turbulent lifted jet flames using flamelets: a priori assessment and a posteriori validation. *Combustion Theory and Modelling*, **15**(2), 201–218.

Scarsini, Marco. 1984. On measures of concordance. *Stochastica: revista de matemática pura y aplicada*, **8**(3), 201–218.

Sklar, Abe. 1959. Fonction de répartition dont les marges sont données. *Inst. Stat. Univ. Paris*, **8**, 229–231.

Sklar, Abe. 1996. Random Variables, Distribution Functions, and Copulas: a personal look backward and forward. *Lecture notes-monograph series, JSTOR.*

Spearman, Charles. 1904. " General Intelligence," objectively determined and measured. *The American Journal of Psychology*, **15**(2), 201–292.

Stroock, Daniel W, & Varadhan, SR Srinivasa. 2007. *Multidimensional diffusion processes.* Springer.

Sungur, Engin A. 2005. A Note on Directional Dependence in Regression Setting. *Communications in Statistics - Theory and Methods*, **34**(9-10), 1957–1965.

Tavella, Domingo, & Randall, Curt. 2000. *Pricing financial instruments: The finite difference method*. John Wiley & Sons.

Taylor, MD. 2007. Multivariate measures of concordance. *Annals of the Institute of Statistical Mathematics*, **59**(4), 789–806.

Tchen, André H. 1980. Inequalities for distributions with given marginals. *The Annals of Probability*, 814–827.

Varadhan, S. R. S. 2007. *Probability Theory.* American Mathematical Society (September 1, 2001).

Venter, Gary, Barnett, Jack, & Rodney Kreps, John Major. 2007. Multivariate copulas for financial modeling. *Variance*, **1**, 103–119.

Wilcox, R. M. 1967. Exponential Operators and Parameter Differentiation in Quantum Physics. *Journal of Mathematical Physics*, **8**(4), 962–982.

Wu, Shaomin. 2014. Construction of asymmetric copulas and its application in two-dimensional reliability modelling. *European Journal of Operational Research*, **238**(2), 476–485.

Yanagimoto, Takemi, & Okamoto, Masashi. 1969. Partial orderings of permutations and monotonicity of a rank correlation statistic. *Annals of the Institute of Statistical Mathematics*, **21**(1), 489–506.

Zhang, H., & Ding, F. 2013. On the Kronecker Products and Their Applications. *Journal of Applied Mathematics*, **1**(Article ID 296185), 1–8.