## UvA-DARE (Digital Academic Repository)

## Essays on mutually exciting point processes

Karim, R.S.

## Publication date

2023
Document Version
Final published version

Link to publication

## Citation for published version (APA):

Karim, R. S. (2023). Essays on mutually exciting point processes. [Thesis, fully internal, Universiteit van Amsterdam].

## General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

## Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Essays on

## Mutually Exciting Point Processes



## Essays on Mutually Exciting Point Processes

Cover design: Esther Scheide - proefschriftomslag.nl
Print: Ridderprint - ridderprint.nl

# Essays on Mutually Exciting Point Processes 

## ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam op gezag van de Rector Magnificus prof. dr. ir. P.P.C.C. Verbeek ten overstaan van een door het College voor Promoties ingestelde commissie, in het openbaar te verdedigen in de Aula der Universiteit op maandag 3 juli 2023, te 11.00 uur<br>door<br>\section*{Raviar Shirwam Karim}

geboren te Sulaymaniyah

## Promotiecommissie

Promotores: prof. dr. R.J.A. Laeven prof. dr. M.R.H. Mandjes

Overige leden: prof. dr. H.P. Boswijk
prof. dr. P.J.C. Spreij
dr. S.U. Can
prof. dr. ir. O.J. Boxma
prof. dr. W.T.F. den Hollander

Universiteit van Amsterdam Universiteit van Amsterdam

Universiteit van Amsterdam
Universiteit van Amsterdam
Universiteit van Amsterdam
Technische Universiteit Eindhoven
Universiteit Leiden

Faculteit Economie en Bedrijfskunde

## Preface

This PhD thesis is the result of three and a half years of academic research on mutually exciting point processes, a specific discipline of science. The contents of the thesis are research papers that I worked on during my PhD at the University of Amsterdam, each strongly related to the underlying topic and yet quite different in methods and results. Before diving into the subject matter, I want to thank a number of people that were on this journey with me.

First, I would like to express my deep gratitude to my doctoral advisors Roger Laeven and Michel Mandjes, who have given me the opportunity to embark on this research with them. Under their guidance, I learned how to ask and answer the right questions, and how to become a good researcher. Roger, thank you for your clear and sharp feedback, which has made every aspect of our research better. Your view on the broader implications of our research continuously steered me in the right direction, while your attention to detail made me become ever more precise. Michel, thank you for your constructive feedback and your support throughout the years. Your critical questions taught me how to clearly communicate my thoughts and ideas. Aside from research, you always showed an interest in my personal life and I look back on our bike rides with joy. It has been a great pleasure to work with both of you.

Second, I thank my colleagues at the Faculty of Economics and Business of the University of Amsterdam that I got to know over the past couple of years. There are too many to name, so I hope that you, dear colleague reader, recognize the subset you belong to and know that I am thankful for you and our encounters. First, a thank you to all the full-time academic staff for the interesting seminars and lunches. I am particularly grateful to Katrien and Umut, for being exemplary teachers and for the very pleasant cooperation in teaching. Next, a big thank you to the secretariat, for creating and facilitating a nice environment to work in. A special thanks goes to Lisa, who has always supported me and actively helped with many endeavors. Finally, a major thank you to all PhD candidates. During the lunches, seminars, and social events, I had a great time getting to know you all. Special thanks are due to those who helped me with my research, taking their time to attempt to solve yet another mathematical problem that I encountered. A very special thanks goes to Evgenii, who has been my academic brother throughout the years.

## Preface

I want to thank my fellow PhD candidates who co-created the faculty PhD council with me and helped strengthen the PhD community, most notably after the pandemic. It has been wonderful collaborating with you to advocate PhD interests in faculty policy and in organizing social and research-related events. I believe we have made real impact on the well-being of current PhD candidates and changed the policy for prospective PhD candidates for the better. I was inspired to put my efforts into this by Eva and Emilie, whom I want to thank in particular. I also thank the Dean and research directors of our faculty for their cooperation in this regard. I am glad to see the willingness of fellow PhD candidates to continue this work.

I also thank my fellow PhD candidates in the Central PhD Council. I had the joy of representing the Faculty of Economics and Business for two years in this universitywide council and getting to know PhDs from other faculties. I enjoyed our regular meetings to discuss and represent PhD interests, as well as writing and analyzing the bi-annual survey.

I would not have finished this PhD without close friends and family. I am grateful to my groups of friends from my high schools and university. I am deeply thankful for my family, their unquestionable support and warmth have made all the difference.

Finally, I am most grateful to Charlotte. You have lovingly and patiently supported me throughout this journey, celebrating every advancement of my research, however tiny. At the start of this PhD, you were my girlfriend. Now, at the end, I am proud to call you my wife. Asking you to marry me was the greatest question I will ever ask.

I dedicate this work to my beloved parents, who have done everything to create a life of possibilities for me.
R.S.K., Amsterdam, March 2023

## Contents

Preface ..... v
1 Introduction ..... 3
1.1 Modeling Stochastic Phenomena Using Point Processes ..... 3
1.2 Modeling Stochastic Phenomena Using Mutually Exciting Point Processes ..... 4
1.3 Main Contributions and Outline ..... 5
2 Exact and Asymptotic Analysis of Multivariate Hawkes Population Pro- cesses ..... 11
2.1 Introduction ..... 11
2.2 Model and Underlying Branching Structure ..... 14
2.2.1 Definition and properties ..... 15
2.2.2 Cluster representation and branching structure ..... 19
2.3 Joint Transforms ..... 21
2.3.1 Analyzing distributional equalities ..... 22
2.3.2 Transform characterization ..... 25
2.4 Fixed-Point Theorem ..... 29
2.4.1 Spaces of joint transforms ..... 29
2.4.2 Fixed point and convergence results ..... 31
2.5 Tail Probabilities ..... 36
2.5.1 Power-law tails and the Hawkes graph ..... 37
2.5.2 Tails of the marginal distributions ..... 38
2.5.3 The tail index is a class property ..... 44
2.5.4 The asymptotic behavior of linear combinations ..... 45
2.6 Numerical Examples ..... 46
2.6.1 Exact analysis ..... 47
2.6.2 Asymptotic analysis ..... 49
2.7 Conclusions ..... 52
Appendix 2.A Additional Proofs ..... 53

## Contents

3 Multivariate Compound Hawkes Processes: Large Deviations and Rare Event Simulation ..... 61
3.1 Introduction ..... 61
3.2 Multivariate Compound Hawkes Processes ..... 65
3.3 Transform Analysis ..... 69
3.4 Large Deviations ..... 76
3.4.1 LDP of multivariate compound Hawkes processes ..... 77
3.4.2 LDP of a single component ..... 81
3.5 Ruin Probabilities ..... 81
3.6 Rare Event Simulation ..... 86
3.6.1 Identification of the alternative distribution ..... 86
3.6.2 Ruin probabilities ..... 89
3.6.3 Exceedance probabilities ..... 93
3.7 Examples and Numerical Illustrations ..... 97
3.7.1 Ruin probability ..... 97
3.7.2 Exceedance probability ..... 101
3.8 Concluding Remarks ..... 103
Appendix 3.A Steepness in the Univariate Case ..... 104
4 Markovian Multivariate Hawkes Population Processes: Efficient Evalu- ation of Moments ..... 111
4.1 Introduction ..... 111
4.2 Multivariate Hawkes Populations ..... 114
4.3 Transform and Joint Moments ..... 116
4.3.1 Transform characterization ..... 117
4.3.2 Joint moments ..... 119
4.4 Recursive Procedure ..... 125
4.4.1 Transient moments ..... 126
4.4.2 Stationary moments ..... 130
4.5 Nested Block Matrices: Bivariate Setting ..... 131
4.6 The Nearly Unstable Behavior ..... 134
4.7 Numerical Experiments ..... 138
4.7.1 Multivariate ..... 139
4.7.2 Bivariate ..... 141
4.8 Concluding Remarks ..... 145
Appendix 4.A Proofs ..... 146
Appendix 4.B Computations ..... 152
4.B. 1 Transform computations ..... 152
4.B. 2 Joint moments: computations ..... 156
Appendix 4.C Explicit Examples ..... 161
4.C. 1 Recursive procedure ..... 161
4.C. 2 Transient moments ..... 163
4.C. 3 Stationary moments ..... 170
4.C. 4 Higher order moments ..... 171
Bibliography ..... 175
Summary ..... 183
Nederlandse Samenvatting (Summary in Dutch) ..... 185
List of Co-Authors ..... 189

# Chapter 1 

Introduction

## Chapter 1

## Introduction

This chapter serves as a motivation and a brief introduction to the model studied in this thesis. Each subsequent chapter has its own introduction pertaining to the specific problems studied in that chapter.

### 1.1 Modeling Stochastic Phenomena Using Point Processes

Point processes are a class of stochastic models that are often applied to model events that occur over time and space. They have proven to be a powerful tool to analyze sequences of events in many fields of science, such as economics, epidemiology, seismology, forestry, neuroscience, and many more. Typical examples of events include earthquakes, financial crashes, and outbreaks of diseases. These models have the capacity and the flexibility to accurately describe how event sequences are generated in all these different settings. Such a model allows one to calculate the likelihood of a new event occurring, or the expected number and the associated variance of events to occur within a certain time frame.

In a nutshell, point processes provide a probabilistic description of how (random) sequences of events are generated. One key aspect of all point processes is that their behavior can be completely characterized by their intensity. The intensity is strongly related to the occurrence of new events, in that it describes in each time frame, what the proportional likelihood is of a new event occurring in that specific time frame. Intuitively, a large intensity will lead to the occurrence of more events on average, while a small intensity will have fewer events on average. The intensity should be higher when events occur in a somewhat regular fashion, such as the arrival of new customers in a store, compared to a situation where events occur very rarely, such as the occurrence of a large earthquake, where a very small intensity is appropriate. This intensity can be a number, but it can also be a function, and even a random function.

## Chapter 1 Introduction

The prototypical example is the Poisson point process, whose intensity is a fixed positive number. The rate at which new events occur is proportional to the intensity, and because the intensity is in this case a fixed number, the sequences of events it generates are similar for each time frame of equal size. If the sequences of events are different in distinct time frames, but there is a structure to it, e.g. known busy periods with many events, one can add time dependence in the intensity to accommodate for that. This yields the slightly more complicated case of the Inhomogeneous Poisson point process, whose intensity is a function of time. When the intensity changes over time, it allows one to incorporate periods where events are more, or less, likely to occur than others. If there is more complicated behavior inherent to the sequence of events, one has to resort to more advanced methods, such as a random intensity function.

Depending on the type of behavior of events one is interested in modeling or analyzing, one has to carefully attend to the choice of parameters of the point process, in the intensity as well as the dimensionality. Here, the dimensionality of the point process simply means how many distinct sequences of events one considers. For instance, when one is interested in the number of people infected with a virus in two different cities, one could use a two-dimensional point process to model this. It need not be a physical location such as a city however, it could be a more abstract location. One could consider 10 stock exchanges around the world and use a 10 -dimensional point process to model event sequences in each stock exchange. This added dimensionality makes point processes flexible and useful in modeling the occurrence of events that have a spatial aspect.

### 1.2 Modeling Stochastic Phenomena Using Mutually Exciting Point Processes

The general theory of point processes has been well-developed and the before-mentioned considerations have, among other things, naturally led to the study of many different sub-classes of point processes, each with its own characteristics and uses. One such sub-class is that of mutually exciting point processes, which is studied in this thesis. Mutually exciting point processes are also known as multivariate Hawkes processes, named after the mathematician Alan Hawkes ([45, 46]), who first studied them in the 70s. More recently, there has been a resurgence of interest in these Hawkes processes, exploring further theoretical properties, such as implications of parameter choices and estimation methods, as well as finding practical applications in various contexts.

What makes mutually exciting point processes interesting, is the built-in feedback system that makes events cluster in both time and space. This entails that events at
a specific time and location can influence the occurrence of events in the future and in other locations. The mechanism that describes the feedback is defined through the intensity, which in this setting depends on the history of past events in a specific way. Namely, the occurrence of an initial event in one location will increase, or excite, the likelihood that new events occur in the future, both in the same location and in other locations. The occurrence of a second event will in turn excite the likelihood of new events occurring. This is where the term mutually exciting comes from, as events can have mutual effects on each other.

This mutually exciting feedback system makes the sequences of events exhibit contagious behavior, through the described temporal and spatial dependence. As a consequence, the sequences of events are grouped in clusters that are highly complex and intertwined, making them accordingly difficult to analyze. To analyze these clusters, one can resort to an alternative and useful description of mutually exciting point processes $([47])$, the so-called branching representation. This is a tree-like structure that describes how initial events are generated and how these create clusters of events over time and across space.

### 1.3 Main Contributions and Outline

This thesis makes contributions to the class of mutually exciting point processes.
Chapter 2 of this thesis characterizes the distribution of sequences of events in a general setting. In particular, it extends the mutually exciting point process to a population process, which does not simply count the number of events over time as the point process does, but in fact, allows for the number of events to decrease as well. A practical example is e.g. in the field of epidemiology, with the population process modeling the number of infected people.

By exploiting the branching representation in combination with the mutually exciting feedback system, we obtain a full characterization of the distribution of the sequences of events. This characterization is given in terms of a fixed-point representation of a suitably defined transform of both the mutually exciting population process and the underlying intensity process. The generality of this result allows for a flexible choice of parameters of the model, such as random marks, i.e. random jump sizes in the intensity, as well as any decay function, i.e. the function that prescribes how the intensity decays over time.

We also derive two asymptotic results, namely a convergence result of the transforms, and the asymptotic tail behavior in the setting of heavy-tailed intensity jumps.

## Chapter 1 Introduction

The convergence result relies on an iterative procedure based on the fixed-point representation, which is computationally attractive as it is easy to implement. The asymptotic tail behavior displays how a property such as heavy-tailed-ness propagates throughout the locations. With the established results, one can compute mathematical objects of interest using numerical techniques, and we illustrate this with some examples.

Chapter 3 considers the compound version of the mutually exciting point process. This compound model consists of the sum of non-negative random variables, drawn for each event of the mutually exciting point process. In the context of insurance, events can trigger insurance claims, which are modeled as non-negative random variables, that the insurance company has to pay out. With a compound model, an insurance company can estimate the expected number of claims as well as their sizes at a certain time in the future, which gives them an estimate of how much initial capital they need to set aside to cover the costs. This is intimately related to the probability of ruin, which is the probability that the insurance company goes insolvent by not having enough initial capital to cover costs.

To investigate this, we derive a Large Deviations Principle, for the mutually exciting point process as well as the compound model. This result requires us to prove a property of a certain function called steepness, for which we partially exploit the results of Chapter 2, namely the fixed-point representation. It turns out particularly tricky and involved to prove this, due to the required multi-dimensional analysis. The Large Deviations Principle gives an important tool to establish two further results. First, we characterize the asymptotic properties of the probability of ruin, yielding the asymptotic logarithmic decay rate. Second, we obtain a method to calculate exceedance probabilities, which is the probability that the compound model exceeds a given threshold.

Typically, the above-mentioned probabilities are very small, to such a degree that it is infeasible to use standard Monte Carlo simulation methods to approximate them. The problem is that extreme events, such as insolvency or exceeding a high threshold, happen very irregularly (as they should!) according to the compound model. To overcome this, we exponentially twist the compound model such that extreme events become regular events, which allows us to apply simulation methods. We show that the exponentially twisted compound model is again the compound version of another mutually exciting point process, with different parameters. With this at hand, we use importance sampling estimators to efficiently approximate the probabilities of ruin and the exceedance probabilities. We prove that these estimators are in a sense optimal.

Chapter 4 focuses on the more narrow class of Markovian mutually exciting point processes. By choosing the intensity decay functions to be of exponential type, the process obtains the Markov property, yielding a powerful tool to analyze the process. In practical settings where one has quick and sudden outbursts of sequences of events, which also fade out quickly, choosing exponential type decay is particularly suitable.

As in Chapter 2, we consider the population process. Using the Markov property, we are able to characterize more explicitly the joint behavior of the population process and the underlying intensity, given in terms of a system of ordinary differential equations. Moreover, we show that this characterization holds when one conditions on different initial values and when one considers multiple points in time. This opens the door to numerically computing any combination of moments of the model.

Moreover, from this more explicit characterization, we reveal a recursive and nested structure hidden within the moments. This method yields explicit expressions of moments up to arbitrary order and brings important computational advantages. We show the superiority of our method in terms of accuracy and speed compared to conventional methods of obtaining moments, such as Monte Carlo simulation and finite difference schemes.

## Chapter 2

## Exact and Asymptotic Analysis of Multivariate Hawkes Population Processes



Abstract
This paper studies multivariate population processes in which multivariate Hawkes processes dictate the stochastic arrivals. We establish results to determine their time-dependent joint probability distributions allowing for general arrival intensity decay functions possibly inducing non-Markovianity, general intensity jump distributions, and general sojourn time distributions. We obtain an exact, full characterization of the time-dependent joint transform of the population process and its underlying intensity process in terms of a fixed-point representation and corresponding convergence results. We also derive the asymptotic tail behavior of the population process and its underlying intensity process in the setting of heavy-tailed intensity jumps. By exploiting the results we establish, arbitrary joint spatial-temporal moments and other distributional properties can now be readily evaluated using standard transform differentiation and inversion techniques, and we illustrate this in a few examples.

## Chapter 2

## Exact and Asymptotic Analysis of Multivariate Hawkes Population Processes

### 2.1 Introduction

As the world grows more interconnected, event shocks tend to spread and cluster more easily and more forcefully. Prototypical examples include the contagious spread of diseases across populations, financial contagion across equity or credit markets, and cyber infections across technological networks. The amplification of event shocks over time and in space, that is, across populations, markets and networks, arguably constitutes one of the core challenges to modern risk measurement and risk management.

In principle, multivariate point processes can provide a probabilistic description of the occurrence of events, and their stochastic dependencies, in time and space. Among multivariate point processes, the class of Hawkes processes ([45]), or mutually exciting processes, provides a natural contender for modeling contagious phenomena. Different from multivariate Poisson, or more generally Lévy, processes, they allow for clustering to occur not just in the spatial (i.e., cross-sectional) dimension, but also in the temporal (i.e., time-series) dimension. Originally introduced to stochastically describe epidemics and earthquake occurrences, Hawkes processes have seen increased interest over the past few years, in finance ( $[2,3,7,33,48]$ ), social interaction ( $[17,66]$ ), neuroscience and genome analysis ( $[13,65]$ ), and so on.

Analyzing distributional properties of multivariate Hawkes, and related, processes is, however, challenging. The existing literature often focuses on the special case of exponential excitation functions under which the Hawkes process (more precisely, the vector consisting of the counting process and its intensity process jointly) can be shown to be Markovian ([2, 3, 24, 26, 28, 33, 45, 62]), or on other specific cases or asymptotic
regimes (discussed below in more detail). The importance of non-Markovian models in describing contagious phenomena was emphasized recently in [35, 59].

In this paper, we establish exact and asymptotic results on distributional properties of general - in our context, not necessarily Markovian - multivariate Hawkes processes and induced population processes, with widespread use across various applications. In a series of exact results, we obtain a full characterization of the joint transform of the multivariate population process and its underlying arrival intensity process, in terms of a fixed-point representation and corresponding convergence results. These results allow for general intensity decay functions, general intensity jump distributions, and general sojourn time distributions. From these results, arbitrary joint moments, including auto- and cross-covariances, and other distributional properties such as joint event probabilities, can be readily obtained using standard transform differentiation and inversion techniques. This paves the way to important applications such as momentbased statistical inference and multivariate risk measurement.

Our exact results exploit a cluster process representation of the Hawkes process, for the univariate self-exciting process first described in [47]; see also [23]. The crossexcitation phenomenon that is present in multivariate Hawkes processes significantly complicates the situation compared to the univariate setting where only self-excitation is present. More specifically, cross-excitation leads to clusters with branches (i.e., offspring) in the time-series as well as in the cross-sectional dimensions, generating complex intertwined clusters. Our analysis of a d-dimensional population process whose arrivals are described by a general multivariate Hawkes process and whose departures admit general sojourn time distributions, yields a fixed-point theorem that characterizes the joint transform of the respective processes at, potentially multiple, future time points, involving suitably defined $d \times d$-dimensional objects to represent the full range of cross-sectional and time-series dependencies. We establish convergence of successive iterations of the fixed-point mapping, thus obtaining the joint transform uniquely.

In a series of asymptotic results, we characterize the tail behavior of our population process and the underlying intensity process in a setting of heavy-tailed intensity jumps. These results pertain directly to the respective probability distributions and enable us to derive associated tail probabilities. We also establish several class properties, and irreducibility results, using the nomenclature of Markov chains.

When analyzing the asymptotic tail behavior in our general multivariate setting, one might naively conjecture that the heaviest tail among the tails of the distributions of intensity jumps that excite component $i$ 'dominates', and therefore dictates the tail behavior of component $i$. Our asymptotic results reveal that this is not necessarily true as, due to the cross-excitation phenomenon, heavy-tailed intensity jumps

### 2.1 Introduction

originating in different components may propagate to component $i$ indirectly, through other components in the system with lighter-tailed intensity jumps. From the full representation of the mutually exciting behavior within the system as provided by our fixed-point theorem, along with suitable Tauberian theorems ([10]), we derive a system of vector-valued renewal equations that jointly characterize the asymptotic tail behavior of our population process and the underlying intensity process. Both our exact and asymptotic results are directly amenable to numerical evaluation and we illustrate our results in a collection of numerical examples.

This paper relates to several branches in the existing literature, which we categorize along three dimensions. First, a starting point for our analysis is provided by the generalized branching structure that underlies a Poissonian cluster process representation of multivariate Hawkes processes. In the univariate self-exciting setting, this branching structure was first discussed in [47]; see also e.g., the formal and extensive treatment in [23] and [60, Chapter 4]. We use a cluster process representation for the complex intertwined spatial-temporal structure of general multivariate Hawkes processes and induced population processes to obtain suitable general and micro-level distributional equalities for our population process and its underlying intensity process, which we next exploit to characterize the joint transform.

Second, transform analysis and the derivation of transient and stationary moments for Hawkes processes has received considerable attention in the literature, especially under Markovian assumptions. In a univariate Markov setting, [24, 21] characterize the joint transform of the point process and the underlying intensity process by relying on the infinitesimal Markov generator, yielding systems of ODEs for the moments of both processes; see also $[26,28,57]$ and the recent closed-form expressions in [29]. In addition, [57] and [38] consider the case of non-exponential decay for the probability generating function of the univariate point process and its joint Laplace transform, respectively; see also the distinct elementary approach in [18]. In a multivariate Markov setting, and more generally in the context of (Markovian) affine point processes, [30] provide semi-analytic expressions of conditional characteristic functions, involving solutions to systems of ODEs; see also [3, 33]. Closed-form expressions of stationary moments as Taylor series approximations over short time intervals in a multivariate Markov setting are derived in [2] by exploiting operator methods. Furthermore, [31] derive an integral equation for the characteristic function of the multivariate point process allowing for non-exponential decay, as an intermediate step to a fractional stochastic volatility model. We characterize the joint transform for, possibly non-Markovian, multivariate Hawkes processes and induced population processes, at possibly multiple time points, through a fixed-point representation and corresponding convergence results, allowing for general decay functions, general distributions of the intensity jump

## Chapter 2 Multivariate Hawkes Population Processes

sizes, and general distributions of the sojourn times. We are not aware of other work on exact transform and moment characterizations for Hawkes processes that allows for a comparable degree of generality along all these dimensions.

Third, asymptotic results such as LLNs and FCLTs for multivariate Hawkes processes have been established in e.g., [6, 39]. Furthermore, the nearly unstable situation is analyzed in [52] and the setting of a large initial intensity is considered in [41]. Large and precise deviation results are obtained in e.g., $[11,43]$ for large times and general univariate Hawkes processes and in [40] for a large initial intensity in the Markov case. These large and precise deviation results are obtained in a setting of light-tailed counting and intensity processes. By contrast, we study the asymptotic tail behavior of the general population process and its underlying intensity process in the setting of heavy-tailed intensity jumps.

We finally note that the population process analyzed in this paper may be naturally interpreted and applied in the - now highly relevant - context of epidemiological modeling. It provides an appealing probabilistic description of the contagious amplification of viruses among populations across the globe. Recent work that uses Hawkes processes, potentially in conjunction with SIR-models, to stochastically describe pandemics such as COVID-19 includes $[14,16]$.

The remainder of this paper is organized as follows. In Section 2.2, we define the multivariate Hawkes process and induced population process, discuss some of their properties, and describe the cluster representation. In Section 2.3, we exploit the branching structure to obtain distributional equalities, which we use to characterize the joint transform of the processes under consideration. In Section 2.4, we represent the joint transform as the fixed point of a certain mapping and establish corresponding convergence results. In Section 2.5, we derive the asymptotic tail behavior. Section 2.6 contains our numerical illustrations. Conclusions are in Section 2.7. The proofs of some auxiliary results are relegated to the Appendix.

### 2.2 Model and Underlying Branching Structure

In this section, we provide a formal definition and a cluster process representation of multivariate Hawkes processes and induced population processes, and discuss their properties and branching structure. Throughout this paper, we adopt the notation $\boldsymbol{x}=\left(x_{1}, \ldots, x_{d}\right)^{\top}, d \in \mathbb{N}$ and $\mathbb{R}_{+}=\{t \in \mathbb{R}: t \geqslant 0\}$.

### 2.2.1 Definition and properties

Multivariate Hawkes processes constitute a class of multivariate point processes. We denote a $d$-dimensional càdlàg point process by $\boldsymbol{N}(\cdot) \equiv(\boldsymbol{N}(t))_{t \in \mathbb{R}_{+}}$, where each $N_{j}(t)$ records the number of points in component $j \in[d]:=\{1, \ldots, d\}$ in the time interval $(0, t]$.

For each $j \in[d]$, consider an a.s. increasing sequence of positive random variables $\boldsymbol{T}_{j}=\left\{T_{j, r}\right\}_{r \in \mathbb{N}}=\left\{T_{j, 1}, T_{j, 2}, \ldots\right\}$. To this sequence we associate $N_{j}(\cdot)$ by setting $N_{j}(t):=N_{\boldsymbol{T}_{j}}(0, t]=\sum_{r=1}^{\infty} \mathbf{1}_{\left\{T_{j, r} \leq t\right\}}$. That is, the process $\boldsymbol{N}(\cdot)=\left(N_{1}(\cdot), \ldots, N_{d}(\cdot)\right)^{\top}$ is the $d$-dimensional counting process associated to the sequences $\boldsymbol{T}_{1}, \ldots, \boldsymbol{T}_{d}$, compactly denoted by $\boldsymbol{N}(t)=\boldsymbol{N}_{\boldsymbol{T}}(0, t]$. The points will be referred to as events and the $T_{j, r}$ 's will be interpreted as the event times of the $r$-th event in component $j$. We assume $\boldsymbol{N}(0)=\mathbf{0}$ throughout this paper. We use the terms point and counting process interchangeably for $\boldsymbol{N}(\cdot)$.

As is well-known, a point process can be characterized by its conditional intensity $\boldsymbol{\lambda}(\cdot)$. The $i$-th component of $\boldsymbol{\lambda}(t), i \in[d]$ and $t>0$, is given by

$$
\begin{equation*}
\lambda_{i}(t)=\lim _{h \downarrow 0} \frac{\mathbb{E}\left[N_{i}(t+h)-N_{i}(t) \mid \mathcal{F}_{t-}\right]}{h} \tag{2.1}
\end{equation*}
$$

where $\mathcal{F}_{t-}=\sigma(\boldsymbol{N}(s): s<t)$ is the sigma algebra of events up to, but not including, time $t$. We refer to $\boldsymbol{\lambda}(\cdot)$ simply as the intensity, where it is noted that it may itself be a stochastic process. Clearly, $\boldsymbol{\lambda}(\cdot)$ is predictable; see e.g., [23, Chapter 7 ] for further details.

Definition 2.1. A multivariate Hawkes process ([45]) is a point process $\boldsymbol{N}(\cdot)$ whose components $N_{i}(\cdot)$, for $i \in[d]$, satisfy

$$
\left\{\begin{array}{l}
\mathbb{P}\left(N_{i}(t+\Delta)-N_{i}(t)=0 \mid \mathcal{F}_{t}\right)=1-\lambda_{i}(t) \Delta+o(\Delta),  \tag{2.2}\\
\mathbb{P}\left(N_{i}(t+\Delta)-N_{i}(t)=1 \mid \mathcal{F}_{t}\right)=\lambda_{i}(t) \Delta+o(\Delta) \\
\mathbb{P}\left(N_{i}(t+\Delta)-N_{i}(t)>1 \mid \mathcal{F}_{t}\right)=o(\Delta)
\end{array}\right.
$$

as $\Delta \downarrow 0$, with $\boldsymbol{N}(0)=\mathbf{0}$. Here, $\mathcal{F}_{t}=\sigma(\boldsymbol{N}(s): s \leqslant t)$ is the natural filtration generated by $\boldsymbol{N}(\cdot)$. With $\bar{\lambda}_{i}>0$ and $g_{i j}(\cdot)$ non-negative integrable functions, the intensity $\lambda_{i}(\cdot)$ takes the form

$$
\begin{equation*}
\lambda_{i}(t)=\bar{\lambda}_{i}+\sum_{j=1}^{d} \int_{0}^{t} B_{i j}(s) g_{i j}(t-s) \mathrm{d} N_{j}(s) \tag{2.3}
\end{equation*}
$$

where, for each $i, j \in[d]$, the $\left(B_{i j}(s)\right)_{0<s<t}$ constitutes a sequence of cross-sectionally and serially independently distributed random variables that are distributed as the

## Chapter 2 Multivariate Hawkes Population Processes

generic non-negative random variable $B_{i j}$. The integral in (2.3) is understood as $\int_{(0, t)}$, excluding $t$.

Informally, the definition above is understood as follows. The constant $\bar{\lambda}_{i}$ represents the base rate corresponding to component $i$. An event generated by $N_{j}(\cdot)$ in component $j$ leads to a jump in the intensity $\lambda_{i}(\cdot)$ of component $i$, with $i, j \in[d]$; its size is distributed as the random variable $B_{i j}$. After the occurrence of this event, the decay functions $g_{i j}(\cdot)$ govern the path of the intensity $\lambda_{i}(\cdot)$ back to the base rate $\bar{\lambda}_{i}$. The base rates can, in principle, be allowed to be time-varying, and take the value zero for a strict subset of $[d]$, at the expense of additional notation in the results that follow. The multivariate Hawkes process $\boldsymbol{N}(\cdot)$ is also known as a mutually exciting point process. When an event in component $i$ impacts the intensity of component $i$, we speak of self-excitation-a purely temporal effect. When an event in component $j$ impacts the intensity of component $i$, with $i \neq j$, we speak of cross-excitation-a temporal as well as spatial effect. Mutually exciting point processes accommodate both effects.

One may introduce functions $h_{i j}(\cdot)=B_{i j} g_{i j}(\cdot)$, where the $B_{i j}$ is understood to be sampled at every event in $N_{j}(\cdot)$ in the manner described in Definition 2.1. These $h_{i j}(\cdot)$, frequently called excitation or impact functions (see e.g., [45, 60]), couple the jump size and the decay function in a multiplicative manner. We can thus compactly rewrite (2.3) in vector notation by setting

$$
\begin{equation*}
\boldsymbol{\lambda}(t)=\overline{\boldsymbol{\lambda}}+\int_{0}^{t} \boldsymbol{H}(t-s) \mathrm{d} \boldsymbol{N}(s) \tag{2.4}
\end{equation*}
$$

where $\boldsymbol{H}(\cdot)=\left(h_{i j}(\cdot)\right)_{i, j \in[d]}$. The matrix form $\boldsymbol{H}(\cdot)$ in (2.4), where for each $j \in[d]$ we define the random vector $\boldsymbol{H}_{j}(\cdot)=\left(h_{1 j}(\cdot), \ldots, h_{d j}(\cdot)\right)^{\top}$, justifies the indexing convention of $h_{i j}(\cdot)$ : $h_{i j}(\cdot)$ describes the impact of events in source component $j$ on the intensity $\lambda_{i}(t)$ of target component $i$.

Whereas $\boldsymbol{N}(\cdot)$ is right-continuous, the Hawkes intensity $\boldsymbol{\lambda}(\cdot)$ is left-continuous. Existence, uniqueness and positivity of $\lambda_{i}(t)$ is guaranteed if $g_{i j}(\cdot)$ satisfies the conditions in Definition 2.1 and $B_{i j}<\infty$ with probability one; see e.g., [23, Example 7.2(b)]. To guarantee stability of the Hawkes process, [45] shows that an additional condition must be imposed; in the current setting, this condition takes the following form:

Assumption 2.1 (Stability condition). Assume that the matrix $\|\boldsymbol{H}\|=\left(\left\|h_{i j}\right\|\right)_{i, j \in[d]}$ with

$$
\left\|h_{i j}\right\|=\mathbb{E}\left[B_{i j}\right] \cdot\left\|g_{i j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}=\mathbb{E}\left[B_{i j}\right] \int_{0}^{\infty} g_{i j}(t) \mathrm{d} t
$$

satisfies $\rho(\|\boldsymbol{H}\|)<1$, where $\rho(\cdot)$ denotes the spectral radius.

In the sequel, we assume the stability condition applies. We emphasize that the Hawkes process in our definition has empty history on the interval $(-\infty, 0]$ by setting $\boldsymbol{N}(0)=\mathbf{0}$, which implies that there are no events prior to $t=0$. In [23, Example $12.5(\mathrm{c})]$ it is shown that, under Assumption 2.1, $\boldsymbol{N}(\cdot)$ converges weakly to the stationary version $\boldsymbol{N}^{\dagger}(\cdot)$ defined on $\mathbb{R}$ with complete intensity $\boldsymbol{\lambda}^{\dagger}(\cdot)$, which implies that $\boldsymbol{N}(\cdot)$ satisfies weak asymptotic stationarity. Furthermore, as shown in [45], under Assumption 2.1, the entries of the expected stationary intensity vector are the constants $\lambda_{i}:=\mathbb{E}\left[\lambda_{i}^{\dagger}(t)\right]=\mathbb{E}\left[\mathrm{d} N_{i}^{\dagger}(t)\right] / \mathrm{d} t$ and, due to the weak asymptotic stationarity, we have $\mathbb{E}\left[\lambda_{i}(t)\right] \rightarrow \lambda_{i}$ as $t \rightarrow \infty$. In vector form these intensities can be expressed as $\boldsymbol{\lambda}=(\boldsymbol{I}-\|\boldsymbol{H}\|)^{-1} \overline{\boldsymbol{\lambda}}$.

We note that the random vectors $\boldsymbol{B}_{j}=\left(B_{1 j}, \ldots, B_{d j}\right)^{\top}$ can be considered as marks associated to an event in component $j \in[d]$; see [23, Chapter 6.4]. More precisely, for each $j \in[d],\left\{\left(T_{j, r}, \boldsymbol{B}_{j, r}\right)\right\}_{r \in \mathbb{N}}$ with $\boldsymbol{B}_{j, r} \sim \boldsymbol{B}_{j}$ constitutes the $j$-th component of a multivariate marked Hawkes process. In other words, Eqn. (2.3) can be rewritten as

$$
\begin{equation*}
\lambda_{i}(t)=\bar{\lambda}_{i}+\sum_{j=1}^{d} \sum_{T_{j, r}<t} B_{i j, r} g_{i j}\left(t-T_{j, r}\right) . \tag{2.5}
\end{equation*}
$$

Note that in the special case that $B_{i j} \equiv 0$ for all $i, j \in[d], \lambda_{i}(\cdot) \equiv \bar{\lambda}_{i}$ such that $\boldsymbol{N}(\cdot)$ reduces to a $d$-dimensional homogeneous Poisson process with rate $\overline{\boldsymbol{\lambda}}$.

While the methodology developed in this work applies to general non-negative integrable functions $g_{i j}(\cdot)$, a parameterization of special interest is that of exponential decay.

Example 2.1 (Exponential). Let the decay functions $g_{i j}(\cdot)$ be of exponential form $g_{i j}(t)=e^{-\alpha_{i} t}$, for some $\alpha_{i}>0$ known as the decay rate. Eqn. (2.3) is in this case given by

$$
\begin{equation*}
\lambda_{i}(t)=\bar{\lambda}_{i}+\sum_{j=1}^{d} \int_{0}^{t} B_{i j}(s) e^{-\alpha_{i}(t-s)} \mathrm{d} N_{j}(s), \tag{2.6}
\end{equation*}
$$

and can, by Itô's Lemma, alternatively be expressed in SDE notation as

$$
\begin{equation*}
\mathrm{d} \lambda_{i}(t)=\alpha_{i}\left(\bar{\lambda}_{i}-\lambda_{i}(t)\right) \mathrm{d} t+\sum_{j=1}^{d} B_{i j}(t) \mathrm{d} N_{j}(t) . \tag{2.7}
\end{equation*}
$$

A distinctive property of exponential decay is that the joint process $(\boldsymbol{N}(\cdot), \boldsymbol{\lambda}(\cdot))$ constitutes a Markov process; see [60, 62].

The Markov property yields a number of useful tools to analyze distributional properties of the Hawkes process in the case that $g_{i j}(\cdot)$ is of exponential form. The explicit

## Chapter 2 Multivariate Hawkes Population Processes

treatment in [2] relies on operator methods applied to the Markov infinitesimal generator. In $[21,24]$, the Markov infinitesimal generator is used to characterize the conditional joint transform of $(N(\cdot), \lambda(\cdot))$ as the solution to a system of ODEs; see also [30]. A similar characterization is given in [57], which exploits the Markov property directly. Departing from exponential decay renders the Hawkes process to be non-Markov in general. An important example, extensively used to model e.g., the temporal clustering of earthquake occurrences, is the power-law parameterization proposed in [64]; see also the recent [49] and the references therein.

Example 2.2 (PowEr). Consider the decay functions $g_{i j}(\cdot)$ to be of power-law type by setting $g_{i j}(t)=1 /\left(c_{i j}+t\right)^{p_{i j}}$ for some $c_{i j} \in \mathbb{R}_{+}$and $p_{i j}>1$. Note that $p_{i j}>1$ ensures integrability. In this case, by Eqn. (2.3),

$$
\begin{equation*}
\lambda_{i}(t)=\bar{\lambda}_{i}+\sum_{j=1}^{d} \int_{0}^{t} \frac{B_{i j}(s)}{\left(c_{i j}+t-s\right)^{p_{i j}}} \mathrm{~d} N_{j}(s) . \tag{2.8}
\end{equation*}
$$

Analyzing distributional properties of the Hawkes process with power-law decay is considerably more complex than under exponential decay, due to the system being non-Markovian. From an applications point of view, the power-law type and related parameterizations are of significant interest, as they enable to model multivariate dynamic behavior that exhibits long-memory properties across time and space.

The Hawkes process can be used in conjunction with other models, such as (affine) jump-diffusion models [2, 33] or epidemiological models [14, 16]. In this paper, we consider a model that also allows for departures. Considering the events generated by $\boldsymbol{N}(\cdot)$ as arrivals, we introduce the Hawkes population process $\boldsymbol{Q}(\cdot)$ by setting for $t>0$,

$$
\begin{equation*}
Q_{i}(t):=\int_{0}^{t} \mathbf{1}_{\left\{E_{i}(s)>t-s\right\}} \mathrm{d} N_{i}(s), \tag{2.9}
\end{equation*}
$$

where $\left(E_{i}(s)\right)_{0<s \leq t}$ is a sequence of i.i.d. random variables, also independent of $\boldsymbol{N}(\cdot)$, representing sojourn times and distributed as the generic non-negative random variable $E_{i}$, and where the integral is understood as $\int_{(0, t]}$, including $t$. Hence,

$$
\begin{equation*}
D_{i}(t):=N_{i}(t)-Q_{i}(t) \tag{2.10}
\end{equation*}
$$

can be interpreted as the process that counts the number of departures in component $i \in[d]$ up to, and including, time $t$. Of course, if $E_{i} \equiv \infty$ for all $i \in[d]$, then $\boldsymbol{Q}(\cdot)=\boldsymbol{N}(\cdot)$, and hence $\boldsymbol{D}(\cdot) \equiv 0$.

Two canonical examples of population processes occur in demography and epidemiology. First, suppose $Q_{i}(\cdot)$ represents the number of (living or active) people in
population $i \in[d]$ and $D_{i}(\cdot)$ the number of deaths. Assuming the individual lifetimes $E_{i}$ to be exponentially distributed with mean $\mu_{i}^{-1}$ for some $\mu_{i} \geqslant 0$, the process $D_{i}(\cdot)$ is an inhomogeneous Poisson process with rate $\mu_{i} Q_{i}(t)$; cf. [4]. Second, $Q_{i}(\cdot)$ may represent the number of infected people in geographic location $i \in[d]$ and $D_{i}(\cdot)$ the recovery process.

### 2.2.2 Cluster representation and branching structure

The Hawkes process admits an alternative definition as a Poisson cluster process, which will play a pivotal role in our analysis. This so-called cluster process representation is first described in [47] in the setting of the conventional single-dimensional Hawkes process; see also [23, Example 6.3(c)] and [60, Ch. IV]. It essentially consists of immigrants generated from the base rates and clusters of offspring generated from rates that account for self- and cross-excitation. In our multivariate setting, the cluster process representation can be described as follows.

Definition 2.2 (Cluster process representation). Define a d-dimensional counting process $\boldsymbol{N}(\cdot)$ componentwise by $N_{j}(t)=N_{\boldsymbol{T}_{j}}(0, t]$ for $j \in[d]$ and $t>0$, with the sequences of event times $\boldsymbol{T}_{1}, \ldots, \boldsymbol{T}_{d}$ generated according to the following procedure:
(1) For each $j \in[d]$, consider a sequence of immigrant event times $\left\{T_{j, r}^{(0)}\right\}_{r \in \mathbb{N}}$ taking values in $\mathbb{R}_{+}$, generated by a homogeneous Poisson process $I_{j}(\cdot)$ with rate $\bar{\lambda}_{j}$.
(2) Next, each immigrant event independently generates a d-dimensional cluster $\boldsymbol{C}_{j} \equiv \boldsymbol{C}_{T_{j, r}^{(0)}} \subseteq \mathbb{R}_{+}^{d}$, which consists of event times associated with generations of events:
(a) The immigrant with event time $T_{j, r}^{(0)}$, considered to be of generation 0 , generates into each component $m \in[d]$ a sequence of first-generation event times $\left\{T_{m, r}^{(1)}\right\}_{r \in \mathbb{N}}$ taking values in $\left(T_{j, r}^{(0)}, \infty\right)$, according to $K_{m j}\left(\cdot-T_{j, r}^{(0)}\right)$, with $K_{m j}(\cdot)$ a non-homogeneous Poisson process with rate $B_{m j, r} g_{m j}(\cdot)$, where $B_{m j, r}$ is the jump size associated to $T_{j, r}^{(0)}$.
(b) Upon iterating (a) above, with $T_{m, r}^{(n-1)}$ denoting the $r$-th event time of generation $n-1$ in component $m \in[d]$, we obtain generation $n$ event times $\left\{T_{l, r}^{(n)}\right\}_{r \in \mathbb{N}}$ in component $l \in[d]$ on the interval $\left(T_{m, r}^{(n-1)}, \infty\right)$, drawn according to $K_{l m}\left(\cdot-T_{m, r}^{(n-1)}\right)$.

By taking the union over all generations, we obtain for each component $j \in[d]$,

$$
\boldsymbol{T}_{j}=\left\{T_{j, r}\right\}_{r \in \mathbb{N}}=\bigcup_{n=0}^{\infty}\left\{T_{j, r}^{(n)}\right\}_{r \in \mathbb{N}} .
$$

## Chapter 2 Multivariate Hawkes Population Processes

The process $\boldsymbol{N}(\cdot)$ thus defined for $t>0$ and with $\boldsymbol{N}(0)=\mathbf{0}$, constitutes a multivariate Hawkes process equivalent to that in Definition 2.1.

The cluster process representation in Definition 2.2 agrees with the intensity-based Definition 2.1 (provided the stability condition is satisfied), as shown in [47, 23]. (The stability condition corresponds to the sub-criticality condition in the context of branching processes.) The proof in [47] amounts to comparing the cluster process representation to an age-dependent birth-death process allowing for immigration, and where the death process is set to zero. Hence, we can naturally extend the cluster process representation to our population process $\boldsymbol{Q}(\cdot)$, by including departures seen as a death process.

The richness, and complexity, of the cluster representation is apparent from the cluster processes that each immigrant event generates. Consider immigrant event time $T_{j, r}^{(0)}$ and, henceforth, set $u=t-T_{j, r}^{(0)}$ as the remaining time after the arrival of the immigrant event. We associate to the $d$-dimensional cluster $\boldsymbol{C}_{j} \equiv \boldsymbol{C}_{T_{j, r}^{(0)}}$ the cluster point process $\boldsymbol{S}_{j}^{\boldsymbol{N}}(\cdot)$, by setting

$$
\begin{equation*}
\boldsymbol{S}_{j}^{\boldsymbol{N}}(u) \equiv \boldsymbol{S}_{C_{j}}^{\boldsymbol{N}}(0, u] \subseteq \mathbb{N}^{d} \tag{2.11}
\end{equation*}
$$

which counts the number of events in $\boldsymbol{C}_{j}$ on the interval ( $0, u$ ]. More explicitly, we have

$$
\boldsymbol{S}_{j}^{\boldsymbol{N}}(u):=\left[\begin{array}{c}
S_{1 \leftarrow j}^{N}(u)  \tag{2.12}\\
\vdots \\
S_{d \leftarrow j}^{N}(u)
\end{array}\right],
$$

where each entry $S_{i \leftarrow j}^{N}(u)$ records the number of events in the cluster generated into component $i \in[d]$ with as oldest ancestor the immigrant with event time $T_{j, r}^{(0)}$. The immigrant with event time $T_{j, r}^{(0)}$ is itself included in the cluster point process (only) when $i=j$, to avoid double counts. Note that due to our multivariate setting, events recorded in $S_{i \leftarrow j}^{N}(\cdot)$ may have propagated through other dimensions $m \in[d]$ before arriving in $i$. Jointly, $\boldsymbol{S}_{j}^{\boldsymbol{N}}(u)$ meticulously keeps track of the components of descendant events and whether their event times are in $\left(T_{j, r}^{(0)}, t\right]$.

Three basic distributional properties, which we will later exploit to operationalize the cluster representation, are noteworthy. First, each immigrant event from source component $j \in[d]$ generates a cluster according to the same distribution modulo a time shift to account for the arrival time; see also [23, Section 6.3]. Hence, we index the cluster process $\boldsymbol{S}_{j}^{\boldsymbol{N}}(\cdot)$ by the source component it corresponds to, not by individual immigrant. Second, each event from the same source component generates offspring
according to the same iterative procedure and, as such, there is a branching structure, and self-similarity, underlying each cluster. Third, we emphasize that the clusters are generated independently of each other.

Our analysis will exploit the branching structure that underlies the cluster representation for all three processes $\boldsymbol{N}(\cdot), \boldsymbol{Q}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$ to characterize distributional properties of the processes jointly. To explicitly describe their dynamics, we introduce the following two cluster processes, resembling the cluster process $\boldsymbol{S}_{j}^{\boldsymbol{N}}(\cdot)$ in (2.11). Denote the $\mathbb{N}^{d}$-valued cluster process $\boldsymbol{S}_{j}^{\boldsymbol{Q}}(\cdot)$ and the $\mathbb{R}_{+}^{d}$-valued cluster process $\boldsymbol{S}_{j}^{\boldsymbol{\lambda}}(\cdot)$, corresponding to $\boldsymbol{Q}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$, at time $u$ by

$$
\boldsymbol{S}_{j}^{\boldsymbol{Q}}(u):=\left[\begin{array}{c}
S_{1 \leftarrow j}^{\boldsymbol{Q}}(u)  \tag{2.13}\\
\vdots \\
S_{d \leftarrow j}^{\boldsymbol{Q}}(u)
\end{array}\right], \quad \boldsymbol{S}_{j}^{\boldsymbol{\lambda}}(u):=\left[\begin{array}{c}
S_{1 \leftarrow j}^{\boldsymbol{\lambda}}(u) \\
\vdots \\
S_{d \leftarrow j}^{\boldsymbol{\lambda}}(u)
\end{array}\right] .
$$

These cluster processes are defined for each entry $i \in[d]$ by

$$
\begin{equation*}
S_{i \leftarrow j}^{Q}(u)=\int_{0}^{u} \mathbf{1}_{\left\{E_{i}(s)>u-s\right\}} \mathrm{d} S_{i \leftarrow j}^{\boldsymbol{N}}(s), \tag{2.14}
\end{equation*}
$$

and

$$
\begin{equation*}
S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u)=\sum_{m=1}^{d} \int_{0}^{u} B_{i m}(s) g_{i m}(u-s) \mathrm{d} S_{m \leftarrow j}^{\boldsymbol{N}}(s) . \tag{2.15}
\end{equation*}
$$

That is, $S_{i \leftarrow j}^{\boldsymbol{Q}}(\cdot)$ and $S_{i \leftarrow j}^{\boldsymbol{\lambda}}(\cdot)$ are defined as the cluster-level analogues of the processes $Q_{i}(\cdot)$ and $\lambda_{i}(\cdot)$ in Eqns. (2.9) and (2.3), respectively. The integrals in (2.14) and (2.15) are understood as in their counterparts (2.9) and (2.3).

Remark 2.1. Most of the results in this paper pertain to the joint process $(\boldsymbol{Q}(\cdot), \boldsymbol{\lambda}(\cdot))$, which includes $(\boldsymbol{N}(\cdot), \boldsymbol{\lambda}(\cdot))$ as a special case when $E_{i} \equiv \infty$ for all $i \in[d]$. We remark that it is possible to extend our results to cover the joint process $(\boldsymbol{N}(\cdot), \boldsymbol{Q}(\cdot), \boldsymbol{\lambda}(\cdot))$ at the cost of heavier notation.

### 2.3 Joint Transforms

In this section, by exploiting the branching structure underlying the cluster process representation, we first derive a collection of distributional equalities that play a key role in next characterizing a general joint transform of the random object $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ in terms of a semi-closed-form expression. The generality of this characterization also allows us to obtain, as corollaries in specific cases, several additional new transform results that are of independent interest.

## Chapter 2 Multivariate Hawkes Population Processes

### 2.3.1 Analyzing distributional equalities

First, to illustrate how the cluster processes behave, Figure 2.1 displays a realization of the paths of $\boldsymbol{N}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$ with corresponding cluster processes $\boldsymbol{S}_{j}^{\boldsymbol{N}}(\cdot)$ and $\boldsymbol{S}_{j}^{\boldsymbol{\lambda}}(\cdot)$ in the bivariate case $d=2$. In the first two subplots, the dotted arrows between events (crosses and diamonds) in $\boldsymbol{N}(\cdot)$ indicate how events are generated across time and components, revealing the branching structure of the cluster processes. The last four subplots display the cluster processes $S_{i \leftarrow j}^{N}(\cdot)$ and $S_{i \leftarrow j}^{\lambda}(\cdot)$, for $i, j=1,2$, so as to make visible how they relate to the processes $N_{i}(\cdot)$ and $\lambda_{i}(\cdot)$.

As the visualization in Figure 2.1 suggests, the cluster processes in (2.12) and (2.13) are formally connected to the processes $N_{i}(\cdot), Q_{i}(\cdot)$ and $\lambda_{i}(\cdot)$ via a set of distributional equalities. More precisely, the equivalence between the cluster process representation in Definition 2.2 and the intensity-based Definition 2.1 allows us to probabilistically describe events, and their impact on $N_{i}(\cdot), Q_{i}(\cdot)$ and $\lambda_{i}(\cdot)$, in terms of the cluster processes. Indeed, for given $t \in \mathbb{R}+$,

$$
\begin{align*}
& N_{i}(t) \stackrel{\mathrm{d}}{=} \sum_{j=1}^{d} \sum_{r=1}^{I_{j}(t)} S_{i \leftarrow j}^{N}\left(t-T_{j, r}^{(0)}\right), \\
& Q_{i}(t) \stackrel{\mathrm{d}}{=} \sum_{j=1}^{d} \sum_{r=1}^{I_{j}(t)} S_{i \leftarrow j}^{\boldsymbol{Q}}\left(t-T_{j, r}^{(0)}\right),  \tag{2.16}\\
& \lambda_{i}(t) \stackrel{\mathrm{d}}{=} \bar{\lambda}_{i}+\sum_{j=1}^{d} \sum_{r=1}^{I_{j}(t)} S_{i \leftarrow j}^{\boldsymbol{\lambda}}\left(t-T_{j, r}^{(0)}\right),
\end{align*}
$$

where $T_{j, r}^{(0)}$ are the immigrant events times and $I_{j}(\cdot)$ is as in Definition 2.2. The distributional equality concerning $\lambda_{i}(t)$ may be compared to Eqn. (2.5), where we expressed $\lambda_{i}(t)$ as the pathwise aggregated change in intensity due to all events strictly prior to time $t$.

We now fix an immigrant with event time $T_{j, r}^{(0)}$ in source component $j \in[d]$ and analyze the generated clusters $\boldsymbol{S}_{j}^{\star}(\cdot)$, where $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$. To exploit the underlying branching structure, and the self-similarity, the idea consists in recognizing that this immigrant event generates first-generation events into all $m$ components, and these in turn generate their own sub-clusters $\boldsymbol{S}_{m}^{\star}(\cdot)$. In order to formally capture this mechanism, we define the matrix processes $\boldsymbol{S}^{\boldsymbol{N}}(\cdot), \boldsymbol{S}^{\boldsymbol{Q}}(\cdot)$ and $\boldsymbol{S}^{\boldsymbol{\lambda}}(\cdot)$ by

$$
\boldsymbol{S}^{\star}(\cdot):=\left[\boldsymbol{S}_{1}^{\star}(\cdot)|\cdots| \boldsymbol{S}_{d}^{\star}(\cdot)\right]=\left[\begin{array}{ccc}
S_{1 \leftarrow 1}^{\star}(\cdot) & \cdots & S_{1 \leftarrow d}^{\star}(\cdot)  \tag{2.17}\\
\vdots & \ddots & \vdots \\
S_{d \leftarrow 1}^{\star}(\cdot) & \cdots & S_{d \leftarrow d}^{\star}(\cdot)
\end{array}\right],
$$



Figure 2.1. Sample paths of: $\boldsymbol{N}(\cdot)=\left(N_{1}(\cdot), N_{2}(\cdot)\right) ; \boldsymbol{\lambda}(\cdot)=\left(\lambda_{1}(\cdot), \lambda_{2}(\cdot)\right)$; the cluster processes originating in component 1 (diamonds), $\boldsymbol{S}_{1}^{\boldsymbol{N}}(\cdot)=\left(S_{1 \leftarrow 1}^{N}(\cdot), S_{2 \leftarrow 1}^{N}(\cdot)\right)$ and $\boldsymbol{S}_{1}^{\boldsymbol{\lambda}}(\cdot)=$ $\left(S_{1 \leftarrow 1}^{\boldsymbol{\lambda}}(\cdot), S_{2 \leftarrow 1}^{\boldsymbol{\lambda}}(\cdot)\right)$; and the cluster processes originating in component $2($ crosses $), \boldsymbol{S}_{2}^{N}(\cdot)=$ $\left(S_{1 \leftarrow 2}^{N}(\cdot), S_{2 \leftarrow 2}^{N}(\cdot)\right)$ and $\boldsymbol{S}_{2}^{\boldsymbol{\lambda}}(\cdot)=\left(S_{1 \leftarrow 2}^{\boldsymbol{\lambda}}(\cdot), S_{2 \leftarrow 2}^{\boldsymbol{\lambda}}(\cdot)\right)$, under exponential decay. Parameters: $\bar{\lambda}_{1}=\bar{\lambda}_{2}=0.5, \alpha_{11}=\alpha_{12}=2.3, \alpha_{21}=\alpha_{22}=2, B_{11} \equiv 1.3, B_{12} \equiv 0.6, B_{21} \equiv 0.8, B_{22} \equiv 0.5$.

## Chapter 2 Multivariate Hawkes Population Processes

for $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$. The columns $\boldsymbol{S}_{j}^{\star}(\cdot)$ correspond to the cluster processes defined in (2.12) and (2.13) and keep track of offspring events that originate from component $j$, while the rows, in the sequel denoted by $\boldsymbol{S}_{(i)}^{\star}(\cdot)$, record offspring events that arrive into component $i$. Observe that the right-hand side expressions in (2.16) contain precisely the entries of the rows.

From Section 2.2 we know that the underlying branching structure is similar for the clusters corresponding to $\boldsymbol{N}(\cdot), \boldsymbol{Q}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$. This fact leads us to introduce unifying notation. To this end, we define the functional $\mathcal{A}_{j}, j \in[d]$, that acts on $\boldsymbol{X}(\cdot)=\left(X_{1}(\cdot), \ldots, X_{d}(\cdot)\right)$, a (row-)vector-valued process taking values in $\mathbb{R}_{+}^{d}$, and $P \geqslant 0$, for each time $u$ by

$$
\begin{equation*}
\mathcal{A}_{j}\{P, \boldsymbol{X}(\cdot)\}(u)=P+\sum_{m=1}^{d} \sum_{r=1}^{K_{m j}(u)} X_{m}\left(u-T_{m, r}^{(1)}\right), \tag{2.18}
\end{equation*}
$$

where $P$ accounts for the impact of the immigrant event and the terms in the summations account for the impact of offspring events, with $K_{m j}(\cdot)$ as in Definition 2.2. Note the time shift to account for the arrival time of the first generation event, $T_{m, r}^{(1)}$.

The functional $\mathcal{A}_{j}$ allows us to compactly and coherently formulate distributional equalities for the respective cluster processes. Indeed, zooming in on specific components $S_{i \leftarrow j}^{\star}(\cdot)$, with $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$, yields the micro-level distributional equalities at time $u$ given by

$$
\begin{align*}
& S_{i \leftarrow j}^{\boldsymbol{N}}(u) \stackrel{\mathrm{d}}{=} \mathcal{A}_{j}\left\{\mathbf{1}_{\{i=j\}}, \boldsymbol{S}_{(i)}^{\boldsymbol{N}}(\cdot)\right\}(u), \\
& S_{i \leftarrow j}^{\boldsymbol{Q}}(u) \stackrel{\mathrm{d}}{=} \mathcal{A}_{j}\left\{\mathbf{1}_{\{i=j\}} \mathbf{1}_{\left\{E_{i}>u\right\}}, \boldsymbol{S}_{(i)}^{\boldsymbol{Q}}(\cdot)\right\}(u),  \tag{2.19}\\
& S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u) \stackrel{\mathrm{d}}{=} \mathcal{A}_{j}\left\{B_{i j} g_{i j}(u), \boldsymbol{S}_{(i)}^{\boldsymbol{\lambda}}(\cdot)\right\}(u),
\end{align*}
$$

which will prove to play a crucial role in the analysis of the cluster processes that follows (in Section 2.4). Note the difference in the first arguments for the different processes, and note that the second argument $\boldsymbol{S}_{(i)}^{\star}(\cdot)=\left(S_{i \leftarrow 1}^{\star}(\cdot), \ldots, S_{i \leftarrow d}^{\star}(\cdot)\right)$ is the $i$-th row of the matrix $\boldsymbol{S}^{\star}(\cdot)$, which accounts for the offspring events. The $B_{i j}$ 's in the expression for $S_{i \leftarrow j}^{\lambda}(\cdot)$ are understood to be sampled for each event in the cluster. Intuitively, Eqn. (2.19) says that the total impact of a cluster process from source component $j$ on target component $i$ is equal in distribution to the superposition of first-generation events and the impact of their offspring. The equality for $S_{i \leftarrow j}^{Q}(\cdot)$ in (2.19) is the multivariate counterpart of [57, Eqn. (4.20)], in the sense that they coincide when setting $d=1$ in our setup.

Remark 2.2. The distributional equalities in (2.19) can be extended to the vectors
$\boldsymbol{S}_{j}^{\star}(\cdot)$ for $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$ using the mapping $\boldsymbol{\mathcal { A }}_{j}$, defined at time u by

$$
\begin{equation*}
\mathcal{A}_{j}\{\boldsymbol{P}, \boldsymbol{X}(\cdot)\}(u)=\boldsymbol{P}+\sum_{m=1}^{d} \sum_{k=1}^{K_{m j}(u)} \boldsymbol{X}_{m}\left(u-T_{k}\right) \tag{2.20}
\end{equation*}
$$

Here, $\boldsymbol{P} \in \mathbb{R}_{+}^{d}$ accounts for the immigrant event, $\boldsymbol{X}(\cdot)$ is an $\mathbb{R}_{+}^{d \times d}$-valued matrix process, and $\boldsymbol{X}_{m}(\cdot)$ is its $m$-th column vector. For $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$, one can substitute appropriate values $\boldsymbol{P}^{\star}(u)$ for $\boldsymbol{P}$ and use the matrix $\boldsymbol{S}^{\star}(\cdot)$ defined in (2.17) to account for the offspring events, to obtain the vector-valued versions of Eqn. (2.19). Note that (2.20) describes the underlying branching structure of entire clusters $\boldsymbol{S}_{j}^{\star}(\cdot)$, and that the entries of $\mathcal{A}_{j}$ correspond to $\mathcal{A}_{1}, \ldots, \mathcal{A}_{d}$.

### 2.3.2 Transform characterization

The distributional equalities are key to characterize a joint transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$. We first make precise what we mean by joint transform.

Definition 2.3. Let $(\boldsymbol{X}(\cdot), \boldsymbol{Y}(\cdot))$ be a stochastic process taking values in $\mathbb{N}_{+}^{d} \times \mathbb{R}_{+}^{d}$. For any $t \in \mathbb{R}_{+}$, the joint transform of $(\boldsymbol{X}(t), \boldsymbol{Y}(t))$ is given by

$$
\begin{equation*}
\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}(t) \equiv \mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}(t, \boldsymbol{s}, \boldsymbol{z}):=\mathbb{E}\left[\boldsymbol{z}^{\boldsymbol{X}(t)} e^{-\boldsymbol{s}^{\top} \boldsymbol{Y}(t)}\right] \equiv \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{X_{i}(t)} e^{-s_{i} Y_{i}(t)}\right] \tag{2.21}
\end{equation*}
$$

where $\boldsymbol{s} \in \mathbb{R}_{+}^{d}$ and $\boldsymbol{z} \in[-1,1]^{d}$, and we denote the space of such transforms by $\mathbb{J}$, such that $\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}(\cdot) \in \mathbb{J}$. Here, $\mathbb{E}[\cdot]$ is understood as $\mathbb{E}_{0}[\cdot]$, i.e., expectation conditional upon the respective filtration at $t=0$, i.e., the value of $(\boldsymbol{X}(0), \boldsymbol{Y}(0))$.

Note that Eqn. (2.21) is well-defined, i.e., exists for any $t \in \mathbb{R}_{+}, \boldsymbol{s} \in \mathbb{R}_{+}^{d}$, and $\boldsymbol{z} \in$ $[-1,1]^{d}$. Throughout the paper, $\boldsymbol{s} \in \mathbb{R}_{+}^{d}$ and $\boldsymbol{z} \in[-1,1]^{d}$ remain fixed, unless stated otherwise, and are therefore sometimes suppressed in the notation for readability. The results that follow can, in principle, be extended to allow for complex-valued $s$ and $\boldsymbol{z}$. In our setting, we consider the joint transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$, with $\boldsymbol{Q}(0)=\mathbf{0}$ and $\boldsymbol{\lambda}(0)=\overline{\boldsymbol{\lambda}}$, given by

$$
\begin{equation*}
\mathcal{J}_{Q, \boldsymbol{\lambda}}(t)=\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t)} e^{-s_{i} \lambda_{i}(t)}\right] . \tag{2.22}
\end{equation*}
$$

We proceed to show that we can obtain a semi-closed-form expression for $\mathcal{J}_{Q, \boldsymbol{\lambda}}(t)$ using the distributional properties derived in Section 2.3.1. Specifically, we use Eqn. (2.16) to describe the entries $Q_{i}(\cdot)$ and $\lambda_{i}(\cdot)$ in terms of the respective cluster processes $\boldsymbol{S}_{j}^{Q}(\cdot)$

## Chapter 2 Multivariate Hawkes Population Processes

and $\boldsymbol{S}_{j}^{\lambda}(\cdot)$. To that end, we also need to consider the joint transform of $\left(\boldsymbol{S}_{j}^{\boldsymbol{Q}}(u), \boldsymbol{S}_{j}^{\boldsymbol{\lambda}}(u)\right)$, with $\boldsymbol{S}_{j}^{\boldsymbol{Q}}(0)=\boldsymbol{e}_{j}$, the unit vector with $j$-th entry equal to 1 , and $\boldsymbol{S}_{j}^{\boldsymbol{\lambda}}(0)=\boldsymbol{B}_{j}$, given by

$$
\begin{equation*}
\mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\boldsymbol{\lambda}}}(u)=\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i<j}^{Q}(u)} e^{-s_{i} S_{i \nless j}^{\lambda}(u)}\right] . \tag{2.23}
\end{equation*}
$$

We can now state the first main result regarding the joint transform $\mathcal{J}_{Q, \boldsymbol{\lambda}}(t)$, expressed in terms of $\mathcal{J}_{S_{j}^{Q}, S_{j}^{\lambda}}(u)$ for $j \in[d]$ and $u \in[0, t]$; later in this paper (in Section 2.4) it is shown how the $\mathcal{J}_{S_{j}^{Q}, S_{j}^{\lambda}}(u)$ can be identified. By using the cluster representation, the independence of the cluster processes across components, and exploiting the derived distributional equalities, we establish the following identity.

Theorem 2.1. The joint transform $\mathcal{J}_{Q, \boldsymbol{\lambda}}(t)$ satisfies

$$
\begin{equation*}
\mathcal{J}_{Q, \boldsymbol{\lambda}}(t, \boldsymbol{s}, \boldsymbol{z})=\prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j}\left(t+s_{j}\right)+\bar{\lambda}_{j} \int_{0}^{t} \mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\boldsymbol{\lambda}}}(u, \boldsymbol{s}, \boldsymbol{z}) \mathrm{d} u\right) . \tag{2.24}
\end{equation*}
$$

Proof. We start by conditioning on the number of immigrants (generation 0 events) in each component, and use the fact that these arrive independently. For brevity, we introduce the vectors $\boldsymbol{I}(t)=\left(I_{1}(t), \ldots, I_{d}(t)\right)^{\top}$ of immigrant processes and $\boldsymbol{n}=$ $\left(n_{1}, \ldots, n_{d}\right)^{\top} \in \mathbb{N}_{0}^{d}\left(\right.$ with $\left.\mathbb{N}_{0}:=\{0,1,2, \ldots\}\right)$. From the distributional equalities (2.16), we obtain

$$
\begin{aligned}
\mathcal{J}_{\boldsymbol{Q}, \boldsymbol{\lambda}}(t, \boldsymbol{s}, \boldsymbol{z}) & =\sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \boldsymbol{z}^{\boldsymbol{Q}(t)} \mid \boldsymbol{I}(t)=\boldsymbol{n}\right] \mathbb{P}(\boldsymbol{I}(t)=\boldsymbol{n}) \\
& =\sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \mathbb{E}\left[\prod_{j=1}^{d} e^{-s_{j} \bar{\lambda}_{j}} \prod_{i=1}^{d} e^{-s_{i} \sum_{r=1}^{n_{j}} S_{i \leftarrow j}^{\boldsymbol{\lambda}}\left(t-T_{j, r}^{(0)}\right)} z_{i}^{\sum_{r=1}^{n_{j}} S_{i \leftarrow j}^{Q}\left(t-T_{j, r}^{(0)}\right)}\right] \mathbb{P}(\boldsymbol{I}(t)=\boldsymbol{n}) \\
& =\sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \prod_{j=1}^{d} e^{-s_{j} \bar{\lambda}_{j}} \mathbb{E}\left[\prod_{i=1}^{d} e^{-s_{i} \sum_{r=1}^{n_{j}} S_{i \leftarrow j}^{\lambda}\left(t-T_{j, r}^{(0)}\right)} z_{i}^{\sum_{r=1}^{n_{j}} S_{i \leftarrow j}^{\boldsymbol{Q}}\left(t-T_{j, r}^{(0)}\right)}\right] \mathbb{P}(\boldsymbol{I}(t)=\boldsymbol{n}),
\end{aligned}
$$

where we have used the independence between the clusters and immigrant processes in the second equality, the independence among clusters in the last equality, and write

$$
\mathbb{P}(\boldsymbol{I}(t)=\boldsymbol{n})=\prod_{j=1}^{d} \mathbb{P}\left(I_{j}(t)=n_{j}\right)
$$

for brevity. Recalling that each $I_{j}(t)$ is a Poisson process, we can use the property that conditional on the number of events at time $t$, the event arrival times are i.i.d.
according to a uniformly distributed random variable on the interval $[0, t]$. With $T^{(j)}$ being uniformly distributed on $[0, t]$ (independent of anything else, that is), we thus have that, for each $j \in[d]$, the sequence $\left(T_{j, r}^{(0)}\right)_{r \in\left[n_{j}\right]}$ are i.i.d. copies of $T^{(j)}$. This allows us to write

$$
\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{d} e^{-s_{i} \sum_{r=1}^{n_{j}} S_{i \leftarrow j}^{\boldsymbol{\lambda}}\left(t-T_{j, r}^{(0)}\right)} z_{i}^{\sum_{r=1}^{n_{j}} S_{i \leftarrow j}^{Q}\left(t-T_{j, r}^{(0)}\right)}\right] & =\left(\mathbb{E}\left[\prod_{i=1}^{d} e^{-s_{i} S_{i \leftarrow j}^{\boldsymbol{\lambda}}\left(t-T^{(j)}\right)} z_{i}^{S_{i \leftarrow j}^{Q}\left(t-T^{(j)}\right)}\right]\right)^{n_{j}} \\
& =\left(\mathcal{J}_{S_{j}^{Q}, S_{j}^{\boldsymbol{\lambda}}}\left(u-T^{(j)}\right)\right)^{n_{j}},
\end{aligned}
$$

by the definition of $\mathcal{J}_{S_{j}^{Q}, S_{j}^{\lambda}}$. Now using that $T^{(j)}$ is uniformly distributed on $[0, t]$ and that $I_{j}(\cdot)$ are Poisson processes with rate $\bar{\lambda}_{j}$, we obtain that

$$
\begin{aligned}
\mathcal{J}_{Q, \boldsymbol{\lambda}}(t) & =\sum_{n \in \mathbb{N}_{0}^{d}} \prod_{j=1}^{d} e^{-s_{j} \bar{\lambda}_{j}}\left(\frac{1}{t} \int_{0}^{t} \mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\boldsymbol{\lambda}}}(t-u) \mathrm{d} u\right)^{n_{j}} \frac{\left(\bar{\lambda}_{j} t\right)^{n_{j}}}{n_{j}!} e^{-\bar{\lambda}_{j} t} \\
& =\sum_{n \in \mathbb{N}_{0}^{d}} \prod_{j=1}^{d} e^{-\bar{\lambda}_{j}\left(t+s_{j}\right)} \frac{\left(\bar{\lambda}_{j} \int_{0}^{t} \mathcal{J}_{S_{j}^{Q}, S_{j}^{\boldsymbol{\lambda}}}(u) \mathrm{d} u\right)^{n_{j}}}{n_{j}!} \\
& \left.=\prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j}\left(t+s_{j}\right)+\bar{\lambda}_{j} \int_{0}^{t} \mathcal{J}_{S_{j}^{Q},,_{j}^{\lambda}}(u)\right) \mathrm{d} u\right),
\end{aligned}
$$

where the second equality holds by an elementary change of variables.
Remark 2.3. Due to the generality of the joint transform, one can easily obtain a number of special cases. Taking $d=1$, we obtain the univariate joint transform of the processes $(Q(t), \lambda(t))$. Choosing $s_{i}=0$ for all $i \in[d]$ yields the probability generating function of $\boldsymbol{Q}(t)$. Choosing $z_{i}=1$ for all $i \in[d]$ yields the Laplace-Stieltjes transform of $\boldsymbol{\lambda}(t)$. Finally, taking $\mu_{i} \equiv 0$ for all $i \in[d]$, we can obtain all above transforms for the counting process $\boldsymbol{N}(t)$.

In multivariate time-series analysis, where multivariate point processes play an important role, one often wants to compute auto- and cross-covariances, involving expressions of the form $\mathbb{E}\left[Q_{i}(t) Q_{j}(t+\tau)\right]$ for a combination of $i, j \in[d]$ and where $\tau>0$, which in turn enable one to compute the respective auto- and cross-correlation functions. Indeed, these functions are central objects in the identification and statistical inference of multivariate Hawkes processes; see e.g., [2] in the Markovian case. In the next proposition, we provide a characterization of the probability generating function of $\boldsymbol{Q}(\cdot)$ associated with different time points, by extending Theorem 2.1. We note that this characterization may be extended further to include $\boldsymbol{\lambda}(\cdot)$ as well as to cover any finite number of time points.

## Chapter 2 Multivariate Hawkes Population Processes

Proposition 2.1. For $\boldsymbol{y}, \boldsymbol{z} \in[-1,1]^{d}$ and $\tau>0$, we have that

$$
\begin{align*}
\mathbb{E}\left[\prod_{i=1}^{d} y_{i}^{Q_{i}(t)} z_{i}^{Q_{i}(t+\tau)}\right]= & \prod_{j=1}^{d} \exp \left(\bar{\lambda}_{j} \int_{0}^{t}\left(\mathbb{E}\left[\prod_{i=1}^{d}\left(y_{i} z_{i}\right)^{S_{i \leftarrow j}^{Q}(u)}\right]-1\right) \mathrm{d} u\right) \\
& \times \exp \left(\bar{\lambda}_{j} \int_{t}^{t+\tau}\left(\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i \leftarrow j}^{\boldsymbol{Q}}(u)}\right]-1\right) \mathrm{d} u\right) . \tag{2.25}
\end{align*}
$$

Proof. The proof is similar to the proof of Theorem 2.1. We therefore omit the details and only explain the general structure of the proof. Compared to Theorem 2.1, one now has to condition twice: first one conditions on the number of immigrant events up to time $t$, and next, given the information up to time $t$, one re-conditions on the number of immigrant events up to time $t+\tau$. By the independent increments property of the (immigrant) Poisson processes, the immigrant event arrival times for the respective conditioning events are uniformly distributed among the intervals $[0, t]$ and $[t, t+$ $\tau]$. Properly using the independence among the clusters similar to Theorem 2.1 and collecting terms, then yields the stated result.

Remark 2.4. A related process, with ample applications in e.g., insurance and risk, is the multivariate compound Hawkes process, constructed as follows. For each $i \in[d]$, let $\left(U_{i}^{(n)}\right)_{n \in \mathbb{N}}$ be a sequence of non-negative i.i.d. random variables independent of $\boldsymbol{N}(\cdot)$. Define the multivariate compound Hawkes process $\boldsymbol{Z}(\cdot):=\left(Z_{1}(\cdot), \ldots, Z_{d}(\cdot)\right)^{\top}$ entrywise for $t \in \mathbb{R}$ by

$$
\begin{equation*}
Z_{i}(t):=\sum_{n=1}^{N_{i}(t)} U_{i}^{(n)} \tag{2.26}
\end{equation*}
$$

For fixed $t \in \mathbb{R}$, the Laplace-Stieltjes transform $\mathcal{T}\{\boldsymbol{Z}(t)\}(\boldsymbol{s})$ of $\boldsymbol{Z}(t)$ satisfies

$$
\begin{equation*}
\mathcal{T}\{\boldsymbol{Z}(t)\}(\boldsymbol{s})=\mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{Z}(t)}\right]=\mathbb{E}\left[\prod_{i=1}^{d}\left(\mathcal{T}\left\{U_{i}\right\}\left(s_{i}\right)\right)^{N_{i}(t)}\right], \tag{2.27}
\end{equation*}
$$

where $\mathcal{T}\left\{U_{i}\right\}(s)=\mathbb{E}\left[e^{-s U_{i}}\right]$ is the Laplace-Stieltjes transform of $U_{i}$ evaluated in s. Also, observe that $\mathcal{T}\{\boldsymbol{Z}(t)\}(\boldsymbol{s})$ can be expressed in terms of quantities discussed earlier in this section, as the right-hand side of (2.27) can be interpreted as the probability generating function of $\boldsymbol{N}(t)$ evaluated in $\boldsymbol{z}=\mathcal{T}\{\boldsymbol{U}\}(\boldsymbol{s}) \equiv\left(\mathcal{T}\left\{U_{1}\right\}\left(s_{1}\right), \ldots, \mathcal{T}\left\{U_{d}\right\}\left(s_{d}\right)\right)$. In other words,

$$
\begin{equation*}
\mathcal{T}\{\boldsymbol{Z}(t)\}(\boldsymbol{s})=\mathcal{J}_{\boldsymbol{N}, \boldsymbol{\lambda}}(t, \mathbf{0}, \mathcal{T}\{\boldsymbol{U}\}(\boldsymbol{s})) . \tag{2.28}
\end{equation*}
$$

### 2.4 Fixed-Point Theorem

In the previous section, we expressed the joint transform $\mathcal{J}_{Q, \boldsymbol{\lambda}}(t)$ of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ in terms of

$$
\begin{equation*}
\mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\lambda}}(u) \equiv \mathcal{J}_{S_{j}^{Q}, S_{j}^{\lambda}}(u, \boldsymbol{s}, \boldsymbol{z}) ; \tag{2.29}
\end{equation*}
$$

see in particular Eqn. (2.24) in the characterization of $\mathcal{J}_{\boldsymbol{Q}, \boldsymbol{\lambda}}(t)$ that is given in Theorem 2.1. In this section, we focus our analysis on (2.29). More specifically, by employing the previously derived distributional equalities, we first characterize the joint transform (2.29) in terms of the fixed point of a certain mapping. Next, we provide an iteration procedure involving this mapping that, as we formally prove, converges to the fixed point, yielding the joint transform uniquely.

### 2.4.1 Spaces of joint transforms

Recall from Definition 2.3 the space $\mathbb{J}$ consisting of time-dependent joint transforms of $d$-dimensional vector-valued processes $(\boldsymbol{X}(\cdot), \boldsymbol{Y}(\cdot))$. To handle the matrices $\boldsymbol{S}^{\star}(\cdot)$ for $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$, we extend the space $\mathbb{J}$ to include matrices, as follows.

Definition 2.4. Set $\mathbb{J}^{d}$ to be the d-dimensional analogue of $\mathbb{J}$, in the sense that an element $\mathcal{J}_{X, Y}(\cdot) \in \mathbb{J}^{d}$ is given at time $u$ by

$$
\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}(u):=\left[\begin{array}{c}
\mathcal{J}_{\boldsymbol{X}_{1}, \boldsymbol{Y}_{1}}(u)  \tag{2.30}\\
\vdots \\
\mathcal{J}_{\boldsymbol{X}_{d}, \boldsymbol{Y}_{d}}(u)
\end{array}\right],
$$

where for each $j \in[d]$, the entry $\mathcal{J}_{\boldsymbol{X}_{j}, \boldsymbol{Y}_{j}}(\cdot) \in \mathbb{J}$ is the joint transform corresponding to $\left(\boldsymbol{X}_{j}(u), \boldsymbol{Y}_{j}(u)\right)=\left(\left(X_{1 j}(u), Y_{1 j}(u)\right), \ldots,\left(X_{d j}(u), Y_{d j}(u)\right)\right)^{\top}$ as defined in Eqn. (2.21).

Note that an entry on the right-hand side of Eqn. (2.30) can be viewed as the joint transform of the columns of the matrix-valued random object $(\boldsymbol{X}(u), \boldsymbol{Y}(u))=$ $\left(X_{i j}(u), Y_{i j}(u)\right)_{i, j \in[d]}$.

When considering the processes $\left(\boldsymbol{S}^{\boldsymbol{Q}}(\cdot), \boldsymbol{S}^{\boldsymbol{\lambda}}(\cdot)\right)$, recall that for each $j \in[d]$ we have that the transform $\mathcal{J}_{S_{j}^{Q}, S_{j}^{\lambda}}(\cdot)$ is an element of $\mathbb{J}$ by Eqn. (2.23). The space $\mathbb{J}^{d}$ plays an important role in the exploitation of the distributional equalities given in Eqn. (2.19). Since in our general multivariate Hawkes model a cluster originating in source component $j$ can in principle generate events in any of the components, characterizing $\mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\lambda}}(\cdot)$ requires us to simultaneously consider $\mathcal{J}_{S_{m}^{Q}, \boldsymbol{S}_{m}^{\lambda}}(\cdot)$ for all $m \in[d]$. This

## Chapter 2 Multivariate Hawkes Population Processes

explains why we work with the following vector of time-dependent joint transforms:

$$
\mathcal{J}_{\boldsymbol{S}^{Q}, \boldsymbol{S}^{\boldsymbol{\lambda}}}(u)=\left[\begin{array}{c}
\mathcal{J}_{S_{1}^{Q}, \boldsymbol{S}_{1}^{\lambda}}(u)  \tag{2.31}\\
\vdots \\
\mathcal{J}_{S_{d}^{Q}, \boldsymbol{S}_{d}^{\lambda}}(u)
\end{array}\right] .
$$

Here, as for any $j \in[d]$ the entry $\mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\lambda}}(\cdot)$ is in $\mathbb{J}$ by Eqn. (2.23), we have that $\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\boldsymbol{\lambda}}}(\cdot) \in \mathbb{J}^{d}$. In words, $\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\boldsymbol{\lambda}}}(u)$ is the vector containing the time-dependent joint transforms corresponding to all pairs of cluster processes $\left(\boldsymbol{S}_{j}^{\boldsymbol{Q}}(u), \boldsymbol{S}_{j}^{\boldsymbol{\lambda}}(u)\right)$ for $j \in[d]$ at time $u$, i.e., the columns of $\left(\boldsymbol{S}^{\boldsymbol{Q}}(u), \boldsymbol{S}^{\boldsymbol{\lambda}}(u)\right)$. This informally entails that the object $\mathcal{J}_{S^{Q}, S^{\lambda}}(u)$ contains a full probabilistic description of all underlying components.

Next, we state the following definition.
Definition 2.5. Consider the mapping $\phi: \mathbb{J}^{d} \rightarrow \mathbb{J}^{d}$ which maps an element $\mathcal{J} \equiv$ $\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}(\cdot) \in \mathbb{J}^{d}$ to:

$$
\mathcal{J}(\cdot)=\left[\begin{array}{c}
\mathcal{J}_{1}(\cdot)  \tag{2.32}\\
\vdots \\
\mathcal{J}_{d}(\cdot)
\end{array}\right] \mapsto\left[\begin{array}{c}
\phi_{1}\left(\mathcal{J}_{1}, \ldots, \mathcal{J}_{d}\right)(\cdot) \\
\vdots \\
\phi_{d}\left(\mathcal{J}_{1}, \ldots, \mathcal{J}_{d}\right)(\cdot)
\end{array}\right]=\left[\begin{array}{c}
\phi_{1}(\mathcal{J})(\cdot) \\
\vdots \\
\phi_{d}(\mathcal{J})(\cdot)
\end{array}\right]=\phi(\mathcal{J})(\cdot),
$$

where each entry $\phi_{j}(\mathcal{J})(\cdot) \in \mathbb{J}$ is defined at time $u$ by

$$
\begin{align*}
\phi_{j}(\mathcal{J})(u) \equiv & \phi_{j}\left(\mathcal{J}_{1}, \ldots, \mathcal{J}_{d}\right)(u, \boldsymbol{s}, \boldsymbol{z}) \\
:= & \mathbb{E}\left[z_{j}^{\mathbf{1}_{\left\{E_{j}>u\right\}}}\right] \prod_{i=1}^{d} \mathbb{E}\left[\exp \left(-s_{i} B_{i j} g_{i j}(u)\right)\right]  \tag{2.33}\\
& \times \prod_{m=1}^{d} \mathbb{E}\left[\exp \left(-B_{m j} \int_{0}^{u} g_{m j}(v)\left(1-\mathcal{J}_{m}(u-v)\right) \mathrm{d} v\right)\right]
\end{align*}
$$

Note that, in Definition 2.5, we suppressed the function notation $\mathcal{J} \equiv \mathcal{J}(\cdot)$ in the argument of $\phi$ for ease of readability, allowing us to denote the image of $\mathcal{J}(\cdot)$ as $\phi(\mathcal{J})(\cdot)$. It is not immediately clear that the mapping $\phi$ is well-defined, i.e., that for any $\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}(\cdot) \in \mathbb{J}^{d}$, we have $\phi\left(\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}\right)(\cdot) \in \mathbb{J}^{d}$ as well. We can show that $\phi\left(\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}\right)(\cdot) \in$ $\mathbb{J}^{d}$, as desired, by suitably modifying the arguments used e.g., in the proof of $[1$, Theorem 1] to this more complex setting.

Lemma 2.1. The mapping $\phi$ in Eqn. (2.32) is well-defined.
Proof. See Section 2.A in the Appendix.

### 2.4.2 Fixed point and convergence results

In this subsection, we characterize $\mathcal{J}_{S^{Q}, S^{\boldsymbol{\lambda}}}(u)$ in terms of a fixed point involving the mapping $\phi$. We then show that iterating the mapping $\phi$ leads us to a unique limit (i.e., the value of $\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\lambda}}(u)$ that we are after). To facilitate the analysis, we need to define an appropriate notion of distance for the space $\mathbb{J}^{d}$. We endow the space $\mathbb{J}^{d}$ with the topology induced by the norm $\|\cdot\|_{\mathbb{J}^{d}}$, defined by

$$
\|\mathcal{J}\|_{\mathbb{J}^{d}}:=\sup _{\substack{u \in[0, t] \\ s \in \mathbb{R}_{+}^{d} \\ \boldsymbol{z} \in[-1,1]^{d}}}\|\mathcal{J}(u, \boldsymbol{s}, \boldsymbol{z})\|_{\mathbb{R}^{d}} \equiv \sup _{u, \boldsymbol{s}, \boldsymbol{z}}\|\mathcal{J}(u, \boldsymbol{s}, \boldsymbol{z})\|_{\mathbb{R}^{d}}
$$

The following result can be proven by suitably applying standard topological methods.
Lemma 2.2. The mapping $\phi$ in Eqn. (2.32) is continuous with respect to the norm $\|\cdot\|_{J^{d}}$.

Proof. See Section 2.A in the Appendix.
Before we can state the main results of this section, we need an intermediate result. In Definition 2.2, we saw that every event in source component $j$ generates events into target component $m$ according to an inhomogeneous Poisson process $K_{m j}(\cdot)$, with intensity $B_{m j} g_{m j}(\cdot)$, with $B_{m j}$ understood to be sampled at every event in $N_{j}(\cdot)$. We need to specify when the offspring events arrive exactly, since these can generate further offspring only after they arrive. Given $u$ as the remaining time after the arrival of the source event, let $v \leqslant u$ and denote by $H_{i j}(v \mid u)$ the probability that an offspring event was already generated before $v$, conditional on it being generated before $u$. Also recall that each first-generation event generates a sub-cluster, as part of the original cluster.

Lemma 2.3. Consider the cluster process $\boldsymbol{S}_{j}^{\star}(\cdot)$ for $\star \in\{\boldsymbol{N}, \boldsymbol{Q}, \boldsymbol{\lambda}\}$ generated by an immigrant event $T_{j, r}^{(0)}$ in component $j \in[d]$ and let $u=t-T_{j, r}^{(0)}$ be the time after this arrival. Then the following statements hold:
(i) Sub-clusters are i.i.d.; more precisely, for each $m \in[d]$, the sequence

$$
\left(\boldsymbol{S}_{m}^{\star}\left(u-T_{m, r}^{(1)}\right)\right)_{r \in[n]},
$$

is an i.i.d. sequence, conditional on $\left\{K_{m j}(u)=n\right\}$ for some $n \in \mathbb{N}$ and with $\left\{T_{m, r}^{(1)}\right\}_{r \in[n]}$ the arrival times of the first-generation events.
(ii) For $v \leqslant u$, the density $h_{i j}(v \mid u)=\frac{\mathrm{d}}{\mathrm{d} v} H_{i j}(v \mid u)$ is given by

$$
\begin{equation*}
h_{i j}(v \mid u)=\frac{g_{i j}(v)}{\int_{0}^{u} g_{i j}(w) \mathrm{d} w} . \tag{2.34}
\end{equation*}
$$

## Chapter 2 Multivariate Hawkes Population Processes

Proof. To prove part (i), fix $m \in[d]$. Then, conditional on $\left\{K_{m j}(u)=n\right\}$, the number of first-generation events, the sub-clusters $\boldsymbol{S}_{m}^{\star}\left(u-T_{m, r}^{(1)}\right)$ can be considered clusters generated by an immigrant in component $m \in[d]$, which are i.i.d. due to the construction in part (2) of Definition 2.2, modulo the time shift.

To prove (ii), we note that $K_{i j}(t)$ is distributed as a Poisson random variable with parameter $\int_{0}^{t} B_{i j} g_{i j}(s) \mathrm{d} s$, conditional on the realization of $B_{i j}$; see Definition 2.2. Using Bayes rule, we compute

$$
\begin{aligned}
H_{i j}(v \mid u) & =\mathbb{P}\left(K_{i j}(v)=1, K_{i j}(u)-K_{i j}(v)=0 \mid K_{i j}(u)=1\right) \\
& =\frac{\exp \left(-\int_{0}^{v} B_{i j} g_{i j}(w) \mathrm{d} w\right) \int_{0}^{v} B_{i j} g_{i j}(w) \mathrm{d} w \exp \left(-\int_{v}^{u} B_{i j} g_{i j}(w) \mathrm{d} w\right)}{\exp \left(-\int_{0}^{u} B_{i j} g_{i j}(w) \mathrm{d} w\right) \int_{0}^{u} B_{i j} g_{i j}(w) \mathrm{d} w} \\
& =\frac{\int_{0}^{v} g_{i j}(w) \mathrm{d} w}{\int_{0}^{u} g_{i j}(w) \mathrm{d} w},
\end{aligned}
$$

which yields the stated result.
We proceed to our characterization of the transform $\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\lambda}}(u)$, for given $\boldsymbol{s}, \boldsymbol{z}$, in terms of a fixed point of the mapping $\phi$. Here, by 'fixed point' we mean that there exists an element in $\mathbb{J}^{d}$ such that applying the mapping $\phi$ to it returns the same element. (Uniqueness considerations will be dealt with later.)

Theorem 2.2. The vector of time-dependent joint transforms $\mathcal{J}_{S^{Q}, S^{\boldsymbol{\lambda}}}(u)$ defined in (2.31) satisfies the fixed-point equation

$$
\begin{equation*}
\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\lambda}}(u)=\phi\left(\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\lambda}}\right)(u) . \tag{2.35}
\end{equation*}
$$

Proof. The structure of the proof is as follows. We start with the law of total expectation, then use the i.i.d. nature of (sub)-clusters, write out the distribution of the first-generation events, and then collect terms.

We fix $j \in[d]$, and show that Eqn. (2.35) holds for the entry $\phi_{j}\left(\mathcal{J}_{S^{Q}, S^{\lambda}}\right)(u)$. We throughout keep $\boldsymbol{s}, \boldsymbol{z}$ fixed. We start the computation of $\mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\lambda}}(u)$ by applying the tower property and conditioning on the number of first-generation events $\left\{T_{m, r}^{(1)}\right\}_{r \in \mathbb{N}}$. For brevity, introduce the vector $\boldsymbol{K}_{j}(u)=\left(K_{1 j}, \ldots, K_{d j}(u)\right)^{\top}$, and note that $\mathbb{P}\left(\boldsymbol{K}_{j}(u)=\boldsymbol{n}\right)=\prod_{m=1}^{d} \mathbb{P}\left(K_{m j}(u)=n_{m}\right)$. Using the distributional equalities (2.19), in combination with the conditional independence between $S_{i \leftarrow j}^{\boldsymbol{Q}}(u), S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u)$ and $K_{m j}(u)$, we have

$$
\begin{equation*}
\mathcal{J}_{\boldsymbol{S}_{j}^{Q}, \boldsymbol{S}_{j}^{\lambda}}(u)=\sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i \leftarrow j}^{\boldsymbol{Q}}(u)} e^{-s_{i} S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u)} \mid \boldsymbol{K}_{j}(u)=\boldsymbol{n}\right] \mathbb{P}\left(\boldsymbol{K}_{j}(u)=\boldsymbol{n}\right) \tag{2.36}
\end{equation*}
$$

$$
=c(u) \sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{\sum_{i=1}^{d} \sum_{r=1}^{n_{m}} S_{i \leftarrow m}^{\boldsymbol{Q}}\left(u-T_{m, r}^{(1)}\right)} e^{-s_{i}} \sum_{m=1}^{d} \sum_{r=1}^{n_{m}} S_{i \leftarrow m}^{\lambda}\left(u-T_{m, r}^{(1)}\right)\right] \mathbb{P}\left(\boldsymbol{K}_{j}(u)=\boldsymbol{n}\right),
$$

where, for brevity, we introduced the constant

$$
c(u):=\mathbb{E}\left[z_{j}^{\mathbf{1}_{\left\{E_{j}>u\right\}}}\right] \prod_{i=1}^{d} \mathbb{E}\left[e^{-s_{i} B_{i j} g_{i j}(u)}\right] .
$$

We now use the i.i.d. nature of the sub-clusters, as described by Lemma 2.3, to write the inner expectation in (2.36) as a product over the source components of the firstgeneration events. To that end, let $T^{(m j)}$ be a random variable with probability density function $h_{m j}(\cdot)$ as given in (ii) of Lemma 2.3, such that $T^{(m j)}$ is distributed as $T_{m, r}^{(1)}$ if it was generated by $K_{m j}(\cdot)$. With these observations and the definition of $\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\lambda}}(\cdot)$, we can write

$$
\begin{aligned}
& \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{\sum_{i=1}^{d} \sum_{r=1}^{n_{m}} S_{i \leftarrow m}^{Q}\left(u-T_{m, r}^{(1)}\right)} e^{-s_{i} \sum_{m=1}^{d} \sum_{r=1}^{n_{m}} S_{i \leftarrow m}^{\lambda}\left(u-T_{m, r}^{(1)}\right)}\right] \\
&=\prod_{m=1}^{d} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i \leftarrow m}^{Q}\left(u-T^{(m j)}\right)} e^{-s_{i} S_{i \leftarrow m}^{\lambda}\left(u-T^{(m j)}\right)}\right]^{n_{m}}=\prod_{m=1}^{d}\left(\mathcal{J}_{S_{m}^{Q}, S_{m}^{\lambda}}\left(u-T^{(m j)}\right)\right)^{n_{m}} .
\end{aligned}
$$

Using that the $K_{m j}(\cdot)$ are Poisson processes with intensity $B_{m j} g_{m j}(\cdot)$, and writing out the density $h_{m j}(\cdot \mid u)$, we thus find

$$
\begin{aligned}
& \mathcal{J}_{S_{j}^{Q}, \boldsymbol{S}_{j}^{\lambda}}(u) \\
&=c(u) \mathbb{E} {\left[\sum_{n \in \mathbb{N}_{0}^{d}} \prod_{m=1}^{d}\left(\int_{0}^{u} h_{m j}(v \mid u) \mathcal{J}_{S_{m}^{Q}, \boldsymbol{S}_{m}^{\lambda}}(u-v) \mathrm{d} v\right)^{n_{m}}\right.} \\
&\left.\times \frac{\left(B_{m j} \int_{0}^{u} g_{m j}(v) \mathrm{d} v\right)^{n_{m}}}{n_{m}!} \exp \left(-B_{m j} \int_{0}^{u} g_{m j}(v) \mathrm{d} v\right)\right] \\
&=c(u) \mathbb{E}\left[\sum_{n \in \mathbb{N}_{0}^{d}} \prod_{m=1}^{d} \frac{1}{n_{m}!}\left(B_{m j} \int_{0}^{u} g_{m j}(v) \mathcal{J}_{S_{m}^{Q}, \boldsymbol{S}_{m}^{\lambda}}(u-v) \mathrm{d} v\right)^{n_{m}} \exp \left(-B_{m j} \int_{0}^{u} g_{m j}(v) \mathrm{d} v\right)\right] \\
&=c(u) \prod_{m=1}^{d} \mathbb{E}\left[\exp \left(B_{m j} \int_{0}^{u} g_{m j}(v)\left(\mathcal{J}_{S_{m}^{Q}, \boldsymbol{S}_{m}^{\mathbf{\lambda}}}(u-v)-1\right) \mathrm{d} v\right)\right],
\end{aligned}
$$

where the last equality holds due to independence between the random variables $B_{m j}$. Note that this last expression equals $\phi_{j}\left(\mathcal{J}_{\boldsymbol{S}^{Q}, \boldsymbol{S}^{\boldsymbol{\lambda}}}\right)(u, \boldsymbol{s}, \boldsymbol{z})$ as was introduced in (2.33), which finishes the proof.

## Chapter 2 Multivariate Hawkes Population Processes

The functional equation described in Eqn. (2.35) can be exploited to numerically approximate the joint transforms of the cluster processes. The convergence result of Theorem 2.3 below entails that iterating the map $\phi$ leads to the desired fixed point, thus having found $\mathcal{J}_{S^{Q}, S^{\lambda}}(u)$ uniquely. Once $\mathcal{J}_{S^{Q}, S^{\lambda}}(u)$ has been obtained, numerical inversion can be applied to obtain the corresponding joint densities and distribution functions; likewise, arbitrary joint spatial-temporal moments can be evaluated by differentiation.

We now proceed to establish the convergence result. Consider a joint transform $\mathcal{J}^{(0)}(\cdot) \in \mathbb{J}^{d}$. Define the sequence $\left(\mathcal{J}^{(n)}(\cdot)\right)_{n \in \mathbb{N}_{0}}$ by $\mathcal{J}^{(n)}(\cdot):=\phi^{n}(\mathcal{J})(\cdot)$ for $n \in \mathbb{N}$, where $\phi$ is the mapping in Eqn. (2.33) and $\phi^{n}$ denotes the $n$-fold iteration. Note that $\mathcal{J}^{(n)}(\cdot) \in \mathbb{J}^{d}$ for all $n \in \mathbb{N}_{0}$ by Lemma 2.1 and induction.

Theorem 2.3. For any $\mathcal{J}^{(0)}(\cdot) \in \mathbb{J}^{d}$, the sequence $\left(\mathcal{J}^{(n)}\right)_{n \in \mathbb{N}_{0}}(u)$ converges pointwise to the fixed point $\mathcal{J}_{S^{Q}, S^{\boldsymbol{\lambda}}}(u)$, i.e., as $n \rightarrow \infty$, for any $u \leqslant t$,

$$
\begin{equation*}
\mathcal{J}^{(n)}(u) \equiv \mathcal{J}^{(n)}(u, \boldsymbol{s}, \boldsymbol{z}) \rightarrow \mathcal{J}_{\boldsymbol{S}^{Q}, \boldsymbol{S}^{\lambda}}(u, \boldsymbol{s}, \boldsymbol{z}) \equiv \mathcal{J}_{\boldsymbol{S}^{Q}, \boldsymbol{S}^{\boldsymbol{\lambda}}}(u) \tag{2.37}
\end{equation*}
$$

Proof. Consider, for $\mathcal{J}_{\mathrm{A}}^{(0)}(\cdot), \mathcal{J}_{\mathrm{B}}^{(0)}(\cdot) \in \mathbb{J}^{d}$, the sequences $\mathcal{J}_{\mathrm{A}}^{(n)}(\cdot), \mathcal{J}_{\mathrm{B}}^{(n)}(\cdot) \in \mathbb{J}^{d}$ by

$$
\mathcal{J}_{\mathrm{A}}^{(n)}(\cdot):=\phi\left(\mathcal{J}_{\mathrm{A}}^{(n-1)}(\cdot)\right), \quad \mathcal{J}_{\mathrm{B}}^{(n)}(\cdot):=\phi\left(\mathcal{J}_{\mathrm{B}}^{(n-1)}(\cdot)\right),
$$

where $n \in \mathbb{N}$. We show that the sequences have a unique limit by first proving that there exists a constant $M>0$ such that, uniformly in $n \in \mathbb{N}_{0}$ and $u \leqslant t$,

$$
\begin{equation*}
\left|\left(\mathcal{J}_{\mathrm{A}}^{(n)}\right)_{j}(u)-\left(\mathcal{J}_{\mathrm{B}}^{(n)}\right)_{j}(u)\right| \leqslant \frac{1}{n!}(M u)^{n} \tag{2.38}
\end{equation*}
$$

where $\left(\mathcal{J}_{i}^{(n)}\right)_{j}(u)$ it the $j$-th entry of $\mathcal{J}_{i}^{(n)}(u)$, for $i \in\{\mathrm{~A}, \mathrm{~B}\}$. We prove (2.38) inductively. For $n=1$, using the bound in Eqn. (2.70) in the proof of Lemma 2.2, we have

$$
\begin{aligned}
\left|\left(\mathcal{J}_{\mathrm{A}}^{(1)}\right)_{j}(u)-\left(\mathcal{J}_{\mathrm{B}}^{(1)}\right)_{j}(u)\right|^{2} & =\left|\phi_{j}\left(\mathcal{J}_{\mathrm{A}}^{(0)}\right)(u)-\phi_{j}\left(\mathcal{J}_{\mathrm{B}}^{(0)}\right)(u)\right|^{2} \\
& \left.\leqslant d \max _{m \in[d]} \mathbb{E}\left[B_{m j}\right]^{2}\left\|g_{m j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}^{2} \int_{0}^{u} \mid \mathcal{J}_{\mathrm{A}}^{(0)}\right)_{j}(v)-\left.\left(\mathcal{J}_{\mathrm{B}}^{(0)}\right)_{j}(v)\right|^{2} \mathrm{~d} v \\
& \leqslant 4 \max _{m \in[d]} \mathbb{E}\left[B_{m j}\right]^{2}\left\|g_{m j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}^{u} \\
& \leqslant M u,
\end{aligned}
$$

where we choose $M=\max _{j \in[d]} M_{j}$, with $M_{j}:=4 d \max _{i \in[d]} \mathbb{E}\left[B_{i j}\right]^{2}\left\|g_{i j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}^{2}$, which is finite by assumption. Note that the second inequality holds since $\mid\left(\mathcal{J}_{\mathrm{A}}^{(0)}\right)_{j}(v)-$ $\left(\mathcal{J}_{\mathrm{B}}^{(0)}\right)_{j}(v) \mid \leqslant 2$ and that the base case is thus proven. For the induction step, assume
that (2.38) holds for some $n \in \mathbb{N}$ for every component $j \in[d]$. Again applying the bounds in Eqn. (2.70), we have

$$
\begin{aligned}
\mid\left(\mathcal{J}_{\mathrm{A}}^{(n+1)}\right)_{j}(u) & -\left.\left(\mathcal{J}_{\mathrm{B}}^{(n+1)}\right)_{j}(u)\right|^{2} \\
& \leqslant M \int_{0}^{u}\left|\left(\mathcal{J}_{\mathrm{A}}^{(n)}\right)_{m}(v)-\left(\mathcal{J}_{\mathrm{B}}^{(n)}\right)_{m}(v)\right| \mathrm{d} v \\
& \leqslant M \frac{1}{n!} \int_{0}^{u}(M v)^{n} \mathrm{~d} v=\frac{1}{(n+1)!}(M u)^{n+1} .
\end{aligned}
$$

Since this holds for all $j \in[d]$, it is clear that (2.38) implies that

$$
\begin{equation*}
\left\|\mathcal{J}_{\mathrm{A}}^{(n)}(u)-\mathcal{J}_{\mathrm{B}}^{(n)}(u)\right\|_{\mathbb{R}^{d}}^{2} \leqslant \sum_{j=1}^{d} \frac{1}{n!}(M u)^{n} \leqslant \frac{d}{n!}(M u)^{n}, \tag{2.39}
\end{equation*}
$$

which implies

$$
\begin{equation*}
\left\|\mathcal{J}_{\mathrm{A}}^{(n)}(u)-\mathcal{J}_{\mathrm{B}}^{(n)}(u)\right\|_{\mathbb{R}^{d}} \rightarrow 0 \tag{2.40}
\end{equation*}
$$

Next, we prove that the sequence $\left(\mathcal{J}_{\mathrm{A}}^{(n)}(u)\right)_{n \in \mathbb{N}_{0}}$ converges for any $u \leqslant t$, by showing that it is a Cauchy sequence. Note that the bound in Eqn. (2.39) holds for any initial $\mathcal{J}_{\mathrm{B}}^{(0)}(\cdot) \in \mathbb{J}^{d}$. In particular, for any $k \in \mathbb{N}$, set $\mathcal{J}_{\mathrm{B}}^{(0)}(\cdot)=\mathcal{J}_{\mathrm{A}}^{(k)}(\cdot)$, such that $\mathcal{J}_{\mathrm{B}}^{(n)}(\cdot)=\phi^{n}\left(\mathcal{J}_{\mathrm{A}}^{(k)}(\cdot)=\mathcal{J}_{\mathrm{A}}^{(n+k)}(\cdot)\right.$, which implies

$$
\begin{equation*}
\left\|\mathcal{J}_{\mathrm{A}}^{(n)}(u)-\mathcal{J}_{\mathrm{A}}^{(n+k)}(u)\right\|_{\mathbb{R}^{d}}^{2}=\left\|\mathcal{J}_{\mathrm{A}}^{(n)}(u)-\mathcal{J}_{\mathrm{B}}^{(n)}(u)\right\|_{\mathbb{R}^{d}}^{2} \leqslant \frac{d}{n!}(M u)^{n} \tag{2.41}
\end{equation*}
$$

Observe that the bound is uniform in $k \in \mathbb{N}$. Let $\epsilon>0$ and choose $N \in \mathbb{N}$ such that $(d / N!)(M u)^{N}<\epsilon / 2$. Then we have $\forall m, n \geqslant N$ that

$$
\begin{aligned}
\left\|\mathcal{J}_{\mathrm{A}}^{(n)}(u)-\mathcal{J}_{\mathrm{A}}^{(m)}(u)\right\|_{\mathbb{R}^{d}}^{2} & =\left\|\mathcal{J}_{\mathrm{A}}^{(N+(n-N))}(u)-\mathcal{J}_{\mathrm{A}}^{(N+(m-N))}(u)\right\|_{\mathbb{R}^{d}}^{2} \\
& \leqslant\left\|\mathcal{J}_{\mathrm{A}}^{(N+(n-N))}(u)-\mathcal{J}_{\mathrm{A}}^{(N)}(u)\right\|_{\mathbb{R}^{d}}^{2}+\left\|\mathcal{J}_{\mathrm{A}}^{(N)}(u)-\mathcal{J}_{\mathrm{A}}^{(N+(m-N))}(u)\right\|_{\mathbb{R}^{d}}^{2} \\
& <\epsilon / 2+\epsilon / 2=\epsilon,
\end{aligned}
$$

by the triangle inequality and using the bound in Eqn. (2.41). Hence, the sequence $\left(\mathcal{J}_{\mathrm{A}}^{(n)}(u)\right)_{n \in \mathbb{N}_{0}}$ is Cauchy and it converges. Mirroring this argument shows that the sequence $\left(\mathcal{J}_{\mathrm{B}}^{(n)}(u)\right)_{n \in \mathbb{N}_{0}}$ converges. Moreover, their difference vanishes by Eqn. (2.40) so both sequences have the same limit, i.e.

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \mathcal{J}_{\mathrm{A}}^{(n)}(u)=\lim _{n \rightarrow \infty} \mathcal{J}_{\mathrm{B}}^{(n)}(u)=\mathcal{J}(u), \tag{2.42}
\end{equation*}
$$

for some $\mathcal{J}(u)$. This implies that for any initial transform $\mathcal{J}^{(0)}(\cdot) \in \mathbb{J}^{d}$, we have that $\mathcal{J}^{(n)}(u) \rightarrow \mathcal{J}(u)$ as $n \rightarrow \infty$.

## Chapter 2 Multivariate Hawkes Population Processes

We now show that $\mathcal{J}(\cdot) \in \mathbb{J}^{d}$, which we argue to hold by applying Lévy's continuity theorem for Laplace-Stieltjes transforms. (It suffices to consider $\boldsymbol{z} \in[0,1]^{d}$; the case $\boldsymbol{z} \in[-1,0)^{d}$ can be dealt with analogously.) Note that $\left(\mathcal{J}^{(n)}\right)_{j}(u)$ can be considered to be a Laplace-Stieltjes transform; indeed, as a consequence of $\mathcal{J}^{(n)}(\cdot) \in \mathbb{J}^{d}$, we can rewrite, for certain $Y_{m}(u) \equiv Y_{m}^{(j, n)}(u), X_{m}(u) \equiv X_{m}^{(j, n)}(u)$, and with $\boldsymbol{Z}^{(j, n)}(u):=$ $\left(\boldsymbol{Y}^{(j, n)}(u), \boldsymbol{X}^{(j, n)}(u)\right)^{\top}$ a random vector of length $2 d$,

$$
\begin{aligned}
\left(\mathcal{J}^{(n)}\right)_{j}(u) & =\mathbb{E}\left[\prod_{m=1}^{d} z_{m}^{Y_{m}(u)} e^{-s_{m} X_{m}(u)}\right]=\mathbb{E}\left[\prod_{m=1}^{d} z_{m}^{Y_{m}(u)} \prod_{m=d+1}^{2 d} e^{-s_{m-d} X_{m-d}(u)}\right] \\
& =\mathbb{E}\left[\prod_{m=1}^{d} e^{-r_{m} Y_{m}(u)} \prod_{m=d+1}^{2 d} e^{-r_{m} X_{m-d}(u)}\right]=\mathbb{E}\left[\exp \left(-\boldsymbol{r}^{\top} \boldsymbol{Z}^{(j, n)}(u)\right)\right],
\end{aligned}
$$

where we set, for $m=1, \ldots, d, e^{-r_{m}}=z_{m}$ and, for $m=d+1, \ldots, 2 d, r_{m}=s_{m-d}$, with $r_{m} \geqslant 0$ for all $m \in[2 d]$. As a result, $\left(\mathcal{J}^{(n)}\right)_{j}(u)$ is the Laplace-Stieltjes transform of the random vector $\boldsymbol{Z}^{(j, n)}(u)$. Since we established above that the limit of $\mathcal{J}^{(n)}(u)$, as $n \rightarrow \infty$, exists, Lévy's continuity theorem, see e.g., [34, XIII: Theorem 2], now implies that there is a random vector $\boldsymbol{Z}^{(j)}$ such that $\boldsymbol{Z}^{(j, n)}(u)$ converges weakly to $\boldsymbol{Z}^{(j)}$ as $n \rightarrow \infty$, with the Laplace-Stieltjes transform of $\boldsymbol{Z}^{(j)}$ being $(\mathcal{J})_{j}(u)$. This implies $(\mathcal{J})_{j}(\cdot) \in \mathbb{J}$, and hence also $\mathcal{J}(\cdot) \in \mathbb{J}^{d}$.

Finally, by Lemma 2.2 we know that $\phi$ is continuous, so

$$
\mathcal{J}(u)=\lim _{n \rightarrow \infty} \mathcal{J}^{(n+1)}(u)=\lim _{n \rightarrow \infty} \phi\left(\mathcal{J}^{(n)}\right)(u)=\phi\left(\lim _{n \rightarrow \infty} \mathcal{J}^{(n)}\right)(u)=\phi(\mathcal{J})(u),
$$

implying that the limit point is a fixed point of $\phi$. By Theorem 2.2 we know that $\mathcal{J}_{S^{Q}, \boldsymbol{S}^{\lambda}}(\cdot)$ is a fixed point of $\phi$ and combined with (2.42), where we derived that every sequence has the same (unique) limit, we have that $\mathcal{J}^{(n)}(u) \rightarrow \mathcal{J}_{S^{Q}, S^{\lambda}}(u)$ as $n \rightarrow \infty$, irrespective of $\mathcal{J}^{(0)}(\cdot) \in \mathbb{J}^{d}$.

### 2.5 Tail Probabilities

In the previous sections, we have provided an exact analysis of the probabilistic behavior of (primarily) the random object $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ for $t>0$. We have characterized in particular the associated joint transform and established a corresponding fixed-point representation, allowing for general decay functions $g_{i j}(\cdot)$, general distributions of the jump sizes $B_{i j}$, and exponential distributions of the sojourn times $E_{i}$. In the present section, we provide an asymptotic analysis pertaining directly to the probability distribution functions of $\boldsymbol{Q}(t)$ and $\boldsymbol{\lambda}(t)$, in the setting where the jump sizes' tail distributions are essentially of a power-law nature.

### 2.5.1 Power-law tails and the Hawkes graph

We start our exposition by introducing the concept of asymptotically power-law tails.
Definition 2.6 (Asymptotically Power-Law Tail (APT)). We say that a scalar-valued non-negative random variable $X$ has an asymptotically power-law tail if there exist positive constants $C$ and $\gamma$ such that

$$
\mathbb{P}(X>x) x^{\gamma} \rightarrow C,
$$

as $x \rightarrow \infty$. We write: $X \in A P T(C, \gamma)$ and refer to $\gamma$ as the tail index.
In the sequel, we assume that, for each $i, j \in[d]$, either $B_{i j} \in \operatorname{APT}\left(C_{i j}, \gamma_{i j}\right)$ for constants $C_{i j}>0$ and $\gamma_{i j}>1$, or $B_{i j} \equiv 0$. Later in this section we discuss a few generalizations.

In the literature, a substantial amount of attention has been devoted to probabilistic systems in which some of the underlying random variables have a distribution function with a power-law tail. One often works with a class of distributions that is closely related to, but slightly wider than, APT, namely the class of regularly varying distributions: then, $\mathbb{P}(X>x)$ behaves as $\ell(x) x^{-\gamma}$ when $x \rightarrow \infty$, for some slowly-varying function $\ell(\cdot)$ (i.e., for all $a>0$ we have that $\ell(a x) / \ell(x) \rightarrow 1$ ). A detailed exposition of regular variation in the context of insurance and finance can, for instance, be found in the monograph [32], and for examples of its use in queueing theory we refer to [36, 78]; a general treatment is in [10]. A powerful concept in this branch of the literature is the so-called 'principle of a single big jump': in many systems, rare events happen with 'overwhelming' probability due to a single extreme outcome of a random quantity that features in the model. In a simple single-dimensional counterpart of our model, it is shown in [57] that in the spirit of this principle $Q(t)$ essentially inherits the tail behavior of the jump size $B$.

An important result that we exploit in the general context of this section, is a closure property related to the sum of independent random variables in APT. If $X_{i}$ is in APT with tail index $\gamma_{i}$, for $i=1,2$, and $X_{1}$ and $X_{2}$ are independent, then $X_{1}+X_{2}$ is also in APT with tail index equal to $\min \left\{\gamma_{1}, \gamma_{2}\right\}$, i.e., the heaviest tail dominates; cf. [32]. Based on this property, one could naïvely guess that in the present multivariate setting, the tail of $Q_{i}(t)$ will resemble the tail of the heaviest among $B_{i 1}, \ldots, B_{i d}$. This is however not necessarily the case: due to potential cross-excitation, heavy tails originating in another component may indirectly propagate to component $i$. This concept of propagation can be conveniently reasoned about relying on socalled Hawkes graphs, defined in the present setting as follows (see also, e.g., [7, 56] for related, different definitions).

## Chapter 2 Multivariate Hawkes Population Processes

Definition 2.7 (Hawkes Graph). Let $V=\{1, \ldots, d\}$ be a set of vertices. Let an edge $e_{i j}$ from $j$ to $i$ exist if $B_{i j} \in \operatorname{APT}\left(C_{i j}, \gamma_{i j}\right)$ for $C_{i j}>0, \gamma_{i j}>1$ (i.e., $B_{i j}$ is not identical to 0 ), and call the resulting set of directed edges $E$. Then, the directed graph $(V, E)$ is called the Hawkes graph.

Note that the vertices $V$ of the Hawkes graph are associated to the components of the Hawkes process $(\boldsymbol{Q}(\cdot), \boldsymbol{\lambda}(\cdot), \boldsymbol{N}(\cdot))$, i.e., $i \in V$ corresponds to $Q_{i}(t)$ (or $\lambda_{i}(t), N_{i}(t)$ ). The states in the Hawkes graph $(V, E)$ can be classified similarly to how this is conducted for Markov chains. To this end, we set $P_{i \leftarrow j}=1$ if a path from $j$ to $i$ in $(V, E)$ exists, and $P_{i \leftarrow j}=0$ otherwise. (The order of the indices in $P_{i \leftarrow j}$ might look unnatural at first sight, but it should be borne in mind that $B_{i j} \in \operatorname{APT}\left(C_{i j}, \gamma_{i j}\right)$ implies that there is an edge from $j$ to $i$ in the Hawkes graph, with corresponding jump sizes generated according to the random variable $B_{i j}$.) We say that vertices $i$ and $j$ belong to the same class if $P_{i \leftarrow j}=P_{j \leftarrow i}=1$. We shall call a class recurrent if there is no path to vertices outside the class, otherwise it is transient.

The following lemma relates the cluster processes to the Hawkes graph path indicators. The proof follows using standard techniques and is for completeness given in Section 2.A in the Appendix.

Lemma 2.4. Consider $u>0$. The following statements are equivalent:
(i) $\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]>0$;
(ii) $\mathbb{E}\left[S_{i \leftarrow j}^{\lambda}(u)\right]>0$; and
(iii) $P_{i \leftarrow j}=1$.

Proof. See Section 2.A in the Appendix.

### 2.5.2 Tails of the marginal distributions

In this subsection, we establish the asymptotic behavior of $\mathbb{P}\left(Q_{i}(t)>x\right)$ and $\mathbb{P}\left(\lambda_{i}(t)>\right.$ $x)$ as $x \rightarrow \infty$, for any $i \in[d]$. We start by introducing a few objects that play a pivotal role in our analysis:

$$
\begin{aligned}
\delta_{i j}:=\min _{m \in[d]}\left\{\gamma_{m j}: P_{i \leftarrow m}=1\right\}, & \bar{\gamma}_{i}:=\min _{j \in[d]} \delta_{i j}, \\
I_{i}:=\underset{j \in[d]}{\arg \min } \delta_{i j}, & I_{i j}:=\underset{m \in[d]}{\arg \min }\left\{\gamma_{m j}: j \in I_{i}\right\} .
\end{aligned}
$$

Given the existence of a path from $m$ to $i, \delta_{i j}$ determines the smallest $\gamma_{m j}$ associated to $B_{m j}$ over all such $m$ for a given $j$. Here, it is noted that we do not assume that $\min _{j \in[d]} \delta_{i j}$ is attained at a unique argument, i.e., $I_{i}$ is a set that potentially consists of more than one element; the same applies to $I_{i j}$.

In addition, we introduce two functionals, in the same spirit as Eqn. (2.18). First, with $\boldsymbol{x}(\cdot)=\left(x_{1}(\cdot), \ldots, x_{d}(\cdot)\right)$ a (row-)vector-valued function and $P \geqslant 0$, for $j \in[d]$,

$$
\begin{equation*}
\mathcal{B}_{j}\{P, \boldsymbol{x}(\cdot)\}(u)=P+\sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v) x_{m}(u-v) \mathrm{d} v \tag{2.43}
\end{equation*}
$$

Second, with $\boldsymbol{y}(\cdot)=\left(y_{1}(\cdot), \ldots, y_{d}(\cdot)\right)$ denoting another (row-)vector-valued function and $\delta \in(1,2)$, for $i, j \in[d]$,

$$
\begin{array}{rl}
\mathcal{B}_{i j}^{\delta}\{P, \boldsymbol{x}(\cdot) \mid \boldsymbol{y}(\cdot)\}(u)=\omega_{i} P+\sum_{m \in I_{i}} & \mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v) x_{m}(u-v) \mathrm{d} v \\
& +\omega_{i} \sum_{m \in I_{i j}} C_{m j}\left(\int_{0}^{u} g_{m j}(v) y_{m}(u-v) \mathrm{d} v\right)^{\delta} \tag{2.44}
\end{array}
$$

where $\omega_{i}=\Gamma(1-\delta)$ with $\Gamma(\cdot)$ the Gamma function.
In the sequel, we will intensively work with functions $R_{i j}^{\star}(\cdot)$ and $R_{i j}^{\star, \delta}(\cdot), \star \in$ $\{\boldsymbol{Q}, \boldsymbol{\lambda}\}$, defined using (2.43)-(2.44), and their row-vector-valued counterparts $\boldsymbol{R}_{(i)}^{\star}(\cdot)$ and $\boldsymbol{R}_{(i)}^{\star, \delta}(\cdot)$, defined analogously to $\boldsymbol{S}_{(i)}^{\star}(\cdot)$. The functions $R_{i j}^{\boldsymbol{Q}}(\cdot)$ and $R_{i j}^{\boldsymbol{\lambda}}(\cdot)$ for $i, j \in[d]$ satisfy the system of coupled functional equations

$$
\begin{equation*}
R_{i j}^{\boldsymbol{Q}}(u)=\mathcal{B}_{j}\left\{\mathbf{1}_{\{i=j\}} \mathbb{P}\left(E_{i}>u\right), \boldsymbol{R}_{(i)}^{Q}(\cdot)\right\}(u), \quad R_{i j}^{\lambda}(u)=\mathcal{B}_{j}\left\{\mathbb{E}\left[B_{i j}\right] g_{i j}(u), \boldsymbol{R}_{(i)}^{\lambda}(\cdot)\right\}(u), \tag{2.45}
\end{equation*}
$$

whereas the functions $R_{i j}^{\boldsymbol{Q}, \delta}(\cdot)$ and $R_{i j}^{\boldsymbol{\lambda}, \delta}(\cdot)$, with $j \in I_{i}$ and $\delta \in(1,2)$, satisfy the system of coupled functional equations

$$
\begin{equation*}
R_{i j}^{\boldsymbol{Q}, \delta}(u)=\mathcal{B}_{i j}^{\delta}\left\{0, \boldsymbol{R}_{(i)}^{\boldsymbol{Q}, \delta}(\cdot) \mid \boldsymbol{R}_{(i)}^{\boldsymbol{Q}}(\cdot)\right\}(u), \quad R_{i j}^{\boldsymbol{\lambda}, \delta}(u)=\mathcal{B}_{i j}^{\delta}\left\{\omega_{i} C_{i j}, \boldsymbol{R}_{(i)}^{\boldsymbol{\lambda}, \delta}(\cdot) \mid \boldsymbol{R}_{(i)}^{\boldsymbol{\lambda}}(\cdot)\right\}(u) . \tag{2.46}
\end{equation*}
$$

The following theorem reveals how APT-behavior is inherited by $Q_{i}(t)$ and $\lambda_{i}(t)$, $i \in[d]$. Henceforth, for positive functions $f(\cdot)$ and $g(\cdot)$, we write $f(x) \sim g(x)$ as $x \rightarrow x_{0}$ to mean $\lim _{x \rightarrow x_{0}} f(x) / g(x)=1$.

Theorem 2.4. Fix $i \in[d]$ and $t \in \mathbb{R}_{+}$. Assume that $\bar{\gamma}_{i} \in(1,2)$. Then, $Q_{i}(t) \in$ $\operatorname{APT}\left(\bar{C}_{i}^{Q}, \bar{\gamma}_{i}\right)$ and $\lambda_{i}(t) \in A P T\left(\bar{C}_{i}^{\lambda}, \bar{\gamma}_{i}\right)$ for some $\bar{C}_{i}^{Q}, \bar{C}_{i}^{\lambda}>0$. More precisely,

$$
\begin{aligned}
& \mathbb{E}\left[z^{Q_{i}(t)}\right]-1+(1-z) \mathbb{E}\left[Q_{i}(t)\right] \sim-(1-z)^{\bar{\gamma}_{i}} \sum_{j \in I_{i}} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\boldsymbol{Q}, \bar{\gamma}_{i}}(u) \mathrm{d} u, \\
& \mathbb{E}\left[e^{-s \lambda_{i}(t)}\right]-1+s \mathbb{E}\left[\lambda_{i}(t)\right] \sim-s^{\bar{\gamma}_{i}} \sum_{j \in I_{i}} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\boldsymbol{\lambda}, \bar{\gamma}_{i}}(u) \mathrm{d} u,
\end{aligned}
$$

## Chapter 2 Multivariate Hawkes Population Processes

as $z \uparrow 1$ and $s \downarrow 0$, respectively, and the first moments equal

$$
\mathbb{E}\left[Q_{i}(t)\right]=\sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{Q}(u) \mathrm{d} u, \quad \mathbb{E}\left[\lambda_{i}(t)\right]=\sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\lambda}(u) \mathrm{d} u
$$

Before we give the proof of Theorem 2.4, we discuss the systems of Eqns. (2.45)(2.46). First, observe that they can be considered as vector-valued renewal equations, owing to the structure of $(2.43)-(2.44)$. We point out how to solve these for (2.45); (2.46) can be dealt with analogously, hence we only provide its final result. In the sequel, we denote the Laplace transform of $f(\cdot)$ by

$$
\mathcal{L}\{f(\cdot)\}(r)=\int_{0}^{\infty} e^{-r u} f(u) \mathrm{d} u
$$

for $r \geqslant 0$. Taking the Laplace transform of (2.45), recognizing the convolution structure, we readily obtain, with $\bar{F}_{i}(\cdot)=\mathbb{P}\left(E_{i}>\cdot\right)$,

$$
\begin{equation*}
\mathcal{L}\left\{R_{i j}^{\boldsymbol{Q}}(\cdot)\right\}(r)=\mathbf{1}_{\{i=j\}} \mathcal{L}\left\{\bar{F}_{i}(\cdot)\right\}(r)+\sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right] \cdot \mathcal{L}\left\{g_{m j}(\cdot)\right\}(r) \cdot \mathcal{L}\left\{R_{i m}^{\boldsymbol{Q}}(\cdot)\right\}(r) \tag{2.47}
\end{equation*}
$$

For a given argument $r \geqslant 0$ and component $i \in[d]$, Eqn. (2.47) is a linear system from which the unknowns $z_{i j}(r)=\mathcal{L}\left\{R_{i j}^{Q}(\cdot)\right\}(r)$ can be solved. This is done by first solving for any recurrent class the corresponding linear subsystem, and then iteratively for any transient class leading to these recurrent classes. Having computed $z_{i j}(\cdot)$, standard Laplace inversion can be invoked to identify the functions $R_{i j}^{\boldsymbol{Q}}(\cdot)$; see also Section 2.6.

A similar argumentation applies when taking the Laplace transform of (2.46), leading to,

$$
\begin{array}{r}
\mathcal{L}\left\{R_{i j}^{\boldsymbol{Q}, \delta}(\cdot)\right\}(r)=\sum_{m \in I_{i}} \mathbb{E}\left[B_{m j}\right] \cdot \mathcal{L}\left\{g_{m j}(\cdot)\right\}(r) \cdot \mathcal{L}\left\{R_{i m}^{\boldsymbol{Q}, \delta}(\cdot)\right\}(r) \\
+\omega_{i} \sum_{m \in I_{i j}} C_{m j} \mathcal{L}\left\{\left(g_{m j} * R_{i m}^{\boldsymbol{Q}}\right)^{\delta}(\cdot)\right\}(r) \tag{2.48}
\end{array}
$$

where, as usual, $*$ denotes the convolution operator, i.e.,

$$
\left(g_{m j} * R_{i m}^{\boldsymbol{Q}}\right)^{\delta}(t)=\left(\int_{0}^{t} g_{m j}(v) R_{i m}^{Q}(t-v) \mathrm{d} v\right)^{\delta}
$$

This is again a linear system, from which the $\bar{z}_{i j}(r)=\mathcal{L}\left\{R_{i j}^{\boldsymbol{Q}, \delta}(\cdot)\right\}(r)$ can be solved for any given $r \geqslant 0$. Recall that at this stage the functions $R_{i m}^{Q}(\cdot)$ are available, thus allowing us to evaluate the last term in (2.48). We can obtain similar equations for the transforms pertaining to $\lambda_{i}(t)$; as these are fully analogous, we leave them out.

Proof of Theorem 2.4. We fix $i \in[d]$ and prove the result for $Q_{i}(t)$; the arguments for $\lambda_{i}(t)$ are similar, mutatis mutandis. To that end, we set $s_{m}=0$ for all $m \in[d]$ and $z_{k}=1$ for $k \neq i$ and $z_{i} \equiv z \in[-1,1]$ in the joint transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t)), \mathcal{J}_{Q, \boldsymbol{\lambda}}(t)$. We thus obtain the $z$-transform of $Q_{i}(t)$ in terms of the cluster process entries $S_{i \leftarrow j}^{\boldsymbol{Q}}(u)$,

$$
\begin{align*}
\mathbb{E}\left[z^{Q_{i}(t)}\right] & =\prod_{j=1}^{d} \exp \left(\bar{\lambda}_{j} \int_{0}^{t}\left(\mathbb{E}\left[z^{S_{i \leftarrow j}^{Q}(u)}\right]-1\right) \mathrm{d} u\right), \\
\mathbb{E}\left[z^{S_{i \leftarrow j}^{Q}(u)}\right] & =\mathbb{E}\left[z^{\mathbf{1}_{\{i=j\}} \mathbf{1}_{\left\{E_{i}>u\right\}}}\right] \prod_{m=1}^{d} \mathcal{T}\left\{B_{m j}\right\}\left(\int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v\right), \tag{2.49}
\end{align*}
$$

where the latter equation holds by our fixed-point theorem, i.e., Theorem 2.2, and $\mathcal{T}\left\{B_{m j}\right\}(s)=\mathbb{E}\left[e^{-s B_{m j}}\right]$ denotes the Laplace-Stieltjes transform of $B_{m j}$. The idea is to analyze expansions of the $z$-transform appearing in (2.49) by using Tauberian theorems, so as to establish the tail behavior of $Q_{i}(t)$.

Under the assumption that $B_{m j} \in \operatorname{APT}\left(C_{m j}, \gamma_{m j}\right)$, the $B_{m j}$ that have index $\gamma_{m j} \in$ $(1,2)$ satisfy

$$
\begin{equation*}
\mathcal{T}\left\{B_{m j}\right\}(r) \sim 1-r \mathbb{E}\left[B_{m j}\right]-C_{m j} \Gamma\left(1-\gamma_{m j}\right) r^{\gamma_{m j}} \tag{2.50}
\end{equation*}
$$

as $r \downarrow 0$, where $\Gamma(\cdot)$ is the Gamma function, by virtue of a Tauberian theorem; see e.g., [10, Theorem 8.1.6]. Observe that

$$
\int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v \downarrow 0
$$

as $z \uparrow 1$. Using this in (2.50), we obtain

$$
\begin{align*}
& \mathcal{T}\left\{B_{m j}\right\}\left(\int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v\right) \\
& \sim 1-\mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v  \tag{2.51}\\
& \quad-C_{m j} \Gamma\left(1-\gamma_{m j}\right)\left(\int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v\right)^{\gamma_{m j}}
\end{align*}
$$

We substitute (2.51) in the expression for $\mathbb{E}\left[z^{S_{i \leftarrow j}^{Q}(u)}\right]$ in (2.49) in combination with the calculation $\mathbb{E}\left[z^{\left.\mathbf{1}_{\{i=j\}} \mathbf{1}_{\left\{E_{i}>u\right\}}\right]}=\mathbb{P}\left(E_{i} \leqslant u\right)+z^{\mathbf{1}_{\{i=j\}}} \mathbb{P}\left(E_{i}>u\right)=1-\left(1-z^{\mathbf{1}_{\{i=j\}}}\right) \mathbb{P}\left(E_{j}>u\right)\right.$. This allows us to write, up to $O\left((1-z)^{2}\right)$ terms,
$1-\mathbb{E}\left[z^{S_{i \leftarrow j}^{Q}(u)}\right] \sim 1-\left(1-\left(1-z^{\mathbf{1}_{\{i=j\}}}\right) \mathbb{P}\left(E_{i}>u\right)\right)$

## Chapter 2 Multivariate Hawkes Population Processes

$$
\begin{align*}
\times & \prod_{m=1}^{d}\left\{1-\mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v\right. \\
& \left.\quad-C_{m j} \Gamma\left(1-\gamma_{m j}\right)\left(\int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v\right)^{\gamma_{m j}}\right\} \\
= & \left(1-z^{\mathbf{1}_{\{i=j\}}}\right) \mathbb{P}\left(E_{i}>u\right)+\sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v \\
+ & \sum_{m=1}^{d} C_{m j} \Gamma\left(1-\gamma_{m j}\right)\left(\int_{0}^{u} g_{m j}(v)\left(1-\mathbb{E}\left[z^{S_{i \leftarrow m}^{Q}(u-v)}\right]\right) \mathrm{d} v\right)^{\gamma_{m j}} . \tag{2.52}
\end{align*}
$$

We now focus on the linear terms, and the next term of order strictly between 1 and 2 , yet to be determined. As for the linear terms, consider the expansion

$$
\begin{aligned}
1-\mathbb{E}\left[z^{S_{\ll j}^{Q}(u)}\right] & =1-\left(1-(1-z) \mathbb{E}\left[S_{i \leftarrow j}^{\boldsymbol{Q}}(u)\right]\right)+O\left((1-z)^{2}\right) \\
& =(1-z) \mathbb{E}\left[S_{i \leftarrow j}^{\boldsymbol{Q}}(u)\right]+O\left((1-z)^{2}\right)=(1-z) R_{i j}^{\boldsymbol{Q}}(u)+O\left((1-z)^{2}\right),
\end{aligned}
$$

where the last equality follows from the observation $R_{i j}^{Q}(u)=\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]$; applying the distributional equality (2.19) to $\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]$ yields Eqn. (2.45).

For the next term (beyond the linear terms, that is), we expand $1-\mathbb{E}\left[z^{S_{i \leftarrow j}^{Q}(u)}\right]$ again, now including a term of fractional order $\vartheta_{i j} \in(1,2)$ (whose value will be determined below), yielding

$$
1-\mathbb{E}\left[z^{S_{i<j}^{Q}(u)}\right]=(1-z) R_{i j}^{\boldsymbol{Q}}(u)+(1-z)^{\vartheta_{i j}} R_{i j}^{\boldsymbol{Q}, \delta}(u)+O\left((1-z)^{2}\right),
$$

where $R_{i j}^{\boldsymbol{Q}, \delta}(u)$ is the solution to (2.46), which we argue next. By substituting this expansion in (2.52), equating the terms of order between 1 and 2 , and ignoring higherorder terms, we obtain

$$
\begin{aligned}
(1-z)^{\vartheta_{i j}} R_{i j}^{\boldsymbol{Q}, \delta}(u)= & \sum_{m=1}^{d}(1-z)^{\vartheta_{i m}} \mathbb{E}\left[B_{m j}\right] \int_{0}^{t} g_{m j}(v) R_{i m}^{\boldsymbol{Q}, \delta}(u-v) \mathrm{d} v \\
& +\sum_{m=1}^{d} C_{m j}(1-z)^{\gamma_{m j}} \Gamma\left(1-\gamma_{m j}\right)\left(\int_{0}^{u} g_{m j}(v) R_{i m}^{\boldsymbol{Q}}(u-v) \mathrm{d} v\right)^{\gamma_{m j}}
\end{aligned}
$$

The problem of solving this system of equations comes down to determining $\vartheta_{i 1}, \ldots, \vartheta_{i d}$ and finding the term(s) of lowest order among the $\gamma_{m j}$. In other words, we need to equate $\vartheta_{i j}$ on the LHS with the minimal $\gamma_{m j}$ on the RHS among the non-zero terms. As $R_{i m}^{Q}(u)=\mathbb{E}\left[S_{i \leftarrow m}^{Q}(u)\right]$, we have by Lemma 2.4 that $R_{i m}^{Q}(u)$ is non-zero if there is a path from $m$ to $i$ in the Hawkes graph. We therefore obtain the condition $\vartheta_{i j}=\delta_{i j}$, with $\delta_{i j}$
introduced at the beginning of Section 2.5.2. Note that $\vartheta_{i j}$ does not necessarily equal $\gamma_{i j}$, but rather corresponds to the heaviest tail originating from $j$ with a potential path of propagation to $i$. Taking into account the source and target component of this heaviest tail, we obtain Eqn. (2.46). Finally, we substitute this back into the $z$-transform of $Q_{i}(t)$ in (2.49), yielding, up to $O\left((1-z)^{2}\right)$ terms, as $z \uparrow 1$,

$$
\begin{aligned}
\mathbb{E}\left[z^{Q_{i}(t)}\right] & \sim \prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \int_{0}^{t}\left((1-z) R_{i j}^{\boldsymbol{Q}}(u)+(1-z)^{\vartheta_{i j}} R_{i j}^{\boldsymbol{Q}, \delta}(u)\right) \mathrm{d} u\right) \\
& \sim 1-(1-z) \sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\boldsymbol{Q}}(u) \mathrm{d} u-\sum_{j=1}^{d}(1-z)^{\vartheta_{i j}} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\boldsymbol{Q}, \delta}(u) \mathrm{d} u
\end{aligned}
$$

where in the last asymptotic equality we used the Taylor expansion of the exponential expression. To obtain the lowest order term among the $\vartheta_{i j}$, we set $\bar{\gamma}_{i}:=\min _{j \in[d]}\left\{\vartheta_{i j}\right\}$, and the components from where the heaviest tails originate are given by the set $I_{i}:=$ $\arg \min _{j \in[d]}\left\{\vartheta_{i j}\right\} \subseteq[d]$. After rewriting this last expression and observing that the linear term equals $\mathbb{E}\left[Q_{i}(t)\right]$, we obtain

$$
\begin{equation*}
\mathbb{E}\left[z^{Q_{i}(t)}\right]-1+(1-z) \mathbb{E}\left[Q_{i}(t)\right] \sim-(1-z)^{\bar{\gamma}_{i}} \sum_{j \in I_{i}} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\boldsymbol{Q}, \bar{\gamma}_{i}}(u) \mathrm{d} u \tag{2.53}
\end{equation*}
$$

Note that (2.53) is now in the general form stated in [10, Theorem 8.1.6], as an expansion of the $z$-transform of $Q_{i}(t)$, which yields that $Q_{i}(t) \in \operatorname{APT}\left(\bar{C}_{i}^{\boldsymbol{Q}}, \bar{\gamma}_{i}\right)$ for some $\bar{C}_{i}^{Q}>0$.

The result of Theorem 2.4 in combination with suitable Tauberian theorems allows us to describe the asymptotic tail behavior of $Q_{i}(t)$ and $\lambda_{i}(t)$. In order to properly take care of the constants that appear in the application of this theorem, we introduce, for $\star \in\{\boldsymbol{Q}, \boldsymbol{\lambda}\}$, the function $\bar{R}_{i j}^{\star, \delta}(\cdot)$ through

$$
\begin{equation*}
R_{i j}^{\star, \delta}(u)=\Gamma(1-\delta) \bar{R}_{i j}^{\star, \delta}(u) . \tag{2.54}
\end{equation*}
$$

Note that $\bar{R}_{i j}^{\star, \delta}(u)$ satisfies Eqn. (2.46) without the $\omega_{i}$ term. We can then rewrite Eqn. (2.53) as

$$
\mathbb{E}\left[z^{Q_{i}(t)}\right]-1+(1-z) \mathbb{E}\left[Q_{i}(t)\right] \sim-(1-z)^{\bar{\gamma}_{i}} \Gamma\left(1-\bar{\gamma}_{i}\right) \sum_{j \in I_{i}} \bar{\lambda}_{j} \int_{0}^{t} \bar{R}_{i j}^{Q, \bar{\gamma}_{i}}(u) \mathrm{d} u
$$

Corollary 2.1. Fix $i \in[d]$ and $t \in \mathbb{R}_{+}$. Assume that $\bar{\gamma}_{i} \in(1,2)$. Then, as $x \rightarrow \infty$,

$$
\begin{align*}
& \mathbb{P}\left(Q_{i}(t)>x\right) \sim\left(\sum_{j \in I_{i}} \bar{\lambda}_{j} \int_{0}^{t} \bar{R}_{i j}^{\boldsymbol{Q}, \bar{\gamma}_{i}}(u) \mathrm{d} u\right) x^{-\bar{\gamma}_{i}}, \\
& \mathbb{P}\left(\lambda_{i}(t)>x\right) \sim\left(\sum_{j \in I_{i}} \bar{\lambda}_{j} \int_{0}^{t} \bar{R}_{i j}^{\lambda, \bar{\gamma}_{i}}(u) \mathrm{d} u\right) x^{-\bar{\gamma}_{i}} \tag{2.55}
\end{align*}
$$

An interesting special case of Theorem 2.4, and Corollary 2.1, concerns the situation in which the Hawkes graph is irreducible, i.e., the case where there exists a directed path between every pair of vertices. Theorem 2.4 then implies that all components $Q_{i}(t)$ and $\lambda_{i}(t), i \in[d]$, inherit the minimum $\gamma=\min _{i, j \in[d]}\left\{\gamma_{i j}\right\}$, i.e., all components $Q_{i}(t)$ and $\lambda_{i}(t)$ are APT with tail index $\gamma$.

Remark 2.5. In the setting of Theorem 2.4, we have assumed $\bar{\gamma}_{i} \in(1,2)$. In case $\bar{\gamma}_{i} \in$ $(m, m+1)$ for some $m \in\{2,3, \ldots\}$, a similar analysis can be performed, but additional terms are to be included in the expansions of (2.49). The statement that $Q_{i}(t)$ and $\lambda_{i}(t)$ are APT with tail index $\bar{\gamma}_{i}$ carries over. While the proof is conceptually analogous to the case of $\bar{\gamma}_{i} \in(1,2)$, various expressions that will appear in the corresponding proof are substantially more involved.

Remark 2.6. We conclude this subsection with a remark on extensions to cases in which some of the jump size distributions are not of APT type. Following Definition 2.6, we have assumed that the $B_{i j}$ are either of APT type or identical to 0, for a transparent exposition. At the expense of additional notation and administration in the proof, also the cases that some of the $B_{i j}$ are identical to a positive constant $b_{i j}$ can be covered. Likewise, using a similar but more cumbersome treatment, one can also handle the situation in which, besides the dominant jump sizes $B_{i j}$ of APT type, there are light-tailed jump sizes as well.

### 2.5.3 The tail index is a class property

To fully appreciate how the chain structure of the Hawkes graph is reflected in the tail indices of $Q_{i}(t)$ and $\lambda_{i}(t), i \in[d]$, this subsection systematically studies this feature, starting by revisiting the two examples discussed earlier. We focus primarily on $Q_{i}(t)$; the analysis for $\lambda_{i}(t)$ is analogous.

Proposition 2.2. For $i \in[d]$ in a given class of the Hawkes graph, all $\mathbb{P}\left(Q_{i}(t)>x\right)$ have the same tail index, i.e., the tail index is a class property.

Proof. The proof follows from Theorem 2.4, in particular, from the role played by $\delta_{i j}$ in the proof of that theorem.

The following somewhat more elaborate example demonstrates how in general to iteratively compute the tail indices corresponding to the individual classes.

Example 2.3 (Hawkes graph). In this example, we consider a system of 6 states such that the vertex set $V=\left\{Q_{1}, \ldots, Q_{6}\right\}$ and the directed edges $E$ are drawn below.


The different colors (drawing styles) represent the class each vertex belongs to. Observe that there are three transient classes, viz. green (G, solid), cyan ( $C$, dashed-dotted) and blue ( $B$, dotted), and one recurrent class, viz. red ( $R$, dashed). We start by determining the tail index of the transient class that is the 'furthest away' from the recurrent class, viz. C. This class has distance 2 to the recurrent class, as one has to go through another transient class to reach the recurrent class. In evident notation,

$$
\gamma_{C}=\bar{\gamma}_{4}=\bar{\gamma}_{5}=\min \left\{\gamma_{45}, \gamma_{54}\right\} .
$$

We proceed with the transient classes with distance 1 to the recurrent class, i.e., $G$ and $B$, and find

$$
\begin{aligned}
& \gamma_{G}=\bar{\gamma}_{1}=\min \left\{\gamma_{C}, \gamma_{14}, \gamma_{11}\right\}=\min \left\{\gamma_{54}, \gamma_{45}, \gamma_{14}, \gamma_{11}\right\}, \\
& \gamma_{B}=\bar{\gamma}_{6}=\gamma_{66} .
\end{aligned}
$$

We finally determine the tail indices of the recurrent classes. In this case, there is just one recurrent class, viz. $R$ :

$$
\begin{aligned}
\gamma_{R} & =\bar{\gamma}_{2}=\bar{\gamma}_{3}=\min \left\{\gamma_{G}, \gamma_{B}, \gamma_{21}, \gamma_{36}, \gamma_{32}, \gamma_{23}, \gamma_{33}\right\} \\
& =\min \left\{\gamma_{45}, \gamma_{54}, \gamma_{14}, \gamma_{11}, \gamma_{66}, \gamma_{21}, \gamma_{36}, \gamma_{32}, \gamma_{23}, \gamma_{33}\right\} .
\end{aligned}
$$

### 2.5.4 The asymptotic behavior of linear combinations

In the above, we have focused on the asymptotic (marginal) tail behavior of the $i$-th components $Q_{i}(t)$ and $\lambda_{i}(t)$. To gain insight into the corresponding joint asymptotic behavior, we now consider the tail behavior of $\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle$ for some $\boldsymbol{c} \in \mathbb{R}_{+}^{d}$, where $\langle\boldsymbol{x}, \boldsymbol{y}\rangle$ denotes the inner product. As before, the case $\langle\boldsymbol{c}, \boldsymbol{\lambda}(t)\rangle$ can be dealt with analogously. For convenience, we do so for the case that the Hawkes graph is irreducible; the case that the Hawkes graph is not irreducible can be addressed as well, but its analysis is somewhat cumbersome and mechanical, as it requires distinguishing various cases.

As demonstrated above, in the irreducible case, $\bar{\gamma}_{i}=\gamma$ for all $i \in[d]$. Likewise, also the set $I_{i}$ does not depend on $i$, and is therefore denoted simply by $I$.

Proposition 2.3. Fix $t \in \mathbb{R}_{+}$and $\boldsymbol{c} \in \mathbb{R}_{+}^{d}$. Let the Hawkes graph be irreducible. Assume $\gamma \in(1,2)$. Then, $\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle \in A P T(\widehat{C}, \gamma)$ for some $\widehat{C}>0$. More precisely,

$$
\begin{equation*}
\mathbb{E}\left[z^{\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle}\right]-1+(1-z) \mathbb{E}[\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle] \sim-(1-z)^{\gamma} \sum_{i=1}^{d} c_{i}^{\gamma} \sum_{j \in I} \bar{\lambda}_{j} \int_{0}^{t} R_{i j}^{\boldsymbol{Q}, \gamma}(u) \mathrm{d} u, \tag{2.56}
\end{equation*}
$$

as $z \uparrow 1$. The first moment equals

$$
\mathbb{E}[\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle]=\sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{t} \sum_{i=1}^{d} c_{i} R_{i j}^{\boldsymbol{Q}}(u) \mathrm{d} u
$$

Proof. The proof follows the lines of the proof of Theorem 2.4; we therefore omit the details. One has to properly take care of the multivariate setting when taking expansions and equating linear and fractional order terms. Due to irreducibility, the term of order $\gamma$ effectively propagates throughout the system.

Proposition 2.3 can be used to determine the asymptotic tail behavior of $\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle$ analogous to Corollary 2.1: under the assumptions of Proposition 2.3, as $x \rightarrow \infty$, we have that

$$
\begin{equation*}
\mathbb{P}(\langle\boldsymbol{c}, \boldsymbol{Q}(t)\rangle>x) \sim\left(\sum_{i=1}^{d} c_{i}^{\gamma} \sum_{j \in I} \bar{\lambda}_{j} \int_{0}^{t} \bar{R}_{i j}^{\boldsymbol{Q}, \gamma}(u) \mathrm{d} u\right) x^{-\gamma} . \tag{2.57}
\end{equation*}
$$

### 2.6 Numerical Examples

This section provides a collection of numerical examples to illustrate the results derived in the previous sections. All the numerical computations in this section are conducted in Python and the computer code is available from the authors upon request. The first part of this section focuses on the exact analysis. The characterization of the joint transform in Theorem 2.1 and the fixed-point representation and convergence results in Theorems 2.2-2.3 enable us to numerically compute arbitrary joint moments of $N_{i}(t)$, $Q_{i}(t)$ and $\lambda_{i}(t)$, for any $i \in[d]$ and $t \in \mathbb{R}_{+}$, using standard numerical techniques. The second part of this section considers the asymptotic analysis. The characterization of the heavy-tailed asymptotic behavior of $Q_{i}(t)$ and $\lambda_{i}(t)$ in Theorem 2.4 allows us to numerically evaluate their tail probabilities. For both the exact and asymptotic analyses, we compare our numerically evaluated results to Monte Carlo simulated counterparts. Our simulation procedure is based on Ogata's [63] thinning algorithm (see also [60, Algorithm 1.21] for details), which in our general multivariate setting essentially relies on the cluster representation in Definition 2.2. Once a sample path
of $\boldsymbol{N}(t)$ has been simulated for some $t \in \mathbb{R}_{+}$, one can easily obtain the corresponding $\boldsymbol{\lambda}(t)$ and $\boldsymbol{Q}(t)$. The simulation procedure is, in principle, elementary, but can be highly time consuming when $t \in \mathbb{R}_{+}$is large, in the presence of heavy tails, or when a high level of precision is pursued. By contrast, the numerical evaluation of our exact and asymptotic results is nearly instantaneous.

Because our results allow for general decay functions $g_{i j}(\cdot)$, we illustrate two examples of parameterizations, namely exponential and power-law decay. Throughout this section, we consider the bivariate case $d=2$. In the instances considered, we let the stability condition be satisfied, i.e., the spectral radius of the matrix $\|\boldsymbol{H}\|=\left(\left\|h_{i j}\right\|\right)_{i, j \in[2]}$ is smaller than 1 , which results in the condition

$$
\begin{equation*}
\left(1-\left\|h_{11}\right\|\right)\left(1-\left\|h_{22}\right\|\right)>\left\|h_{12}\right\|\left\|h_{21}\right\| . \tag{2.58}
\end{equation*}
$$

The stability condition (2.58) implies that the processes $N_{i}(t), Q_{i}(t)$ and $\lambda_{i}(t)$ converge to a steady state as $t$ grows.

### 2.6.1 Exact analysis

We focus attention on the processes $\boldsymbol{Q}(t)=\left(Q_{1}(t), Q_{2}(t)\right)$ and $\boldsymbol{\lambda}(t)=\left(\lambda_{1}(t), \lambda_{2}(t)\right)$ and illustrate the joint transform characterization of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ under both exponential and power-law decay. More specifically, by means of a standard finite difference method, we approximate the derivatives of the joint transform $\mathcal{J}_{Q, \boldsymbol{\lambda}}(t)$ obtained by relying on the fixed-point characterization of Theorems 2.1-2.3, so as to numerically evaluate arbitrary joint moments. We focus on moments corresponding to the first components, i.e.,

$$
\mathbb{E}\left[Q_{1}(t)\right]=\mathbb{E}\left[Q_{1}(t) \mid Q_{1}(0)=0\right], \quad \mathbb{E}\left[\lambda_{1}(t)\right]=\mathbb{E}\left[\lambda_{1}(t) \mid \lambda_{1}(0)=\bar{\lambda}_{1}\right],
$$

and associated variances, as well as the joint moments $\mathbb{E}\left[Q_{1}(t) Q_{2}(t)\right]$ and $\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right]$. Recall that the departures of events in $Q_{i}(t)$ are governed by the non-negative random variable $E_{i}$, distributed as an exponential random variable with parameter $\mu_{i}>0$.

## Exponential

We let the decay functions $g_{i j}(\cdot)$, with $i, j \in[2]$, be of exponential form. In our bivariate setting, we assume

$$
\begin{equation*}
g_{11}(t)=g_{12}(t)=e^{-\alpha_{1} t}, \quad g_{21}(t)=g_{22}(t)=e^{-\alpha_{2} t} \tag{2.59}
\end{equation*}
$$

for $\alpha_{1}, \alpha_{2}>0$. Further, in this example we take the random variables $B_{i j}$ to be positive constants, i.e., $B_{i j} \equiv b_{i j} \in \mathbb{R}_{+}$. This yields, for $i, j \in[2]$, that $\left\|h_{i j}\right\|=b_{i j} / \alpha_{i}$, and


Figure 2.2. Plots of the expectations and variances of $Q_{1}(\cdot)$ and $\lambda_{1}(\cdot)$ and the joint moments $\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right]$ and $\mathbb{E}\left[Q_{1}(t) Q_{2}(t)\right]$ (solid lines), compared to Monte Carlo simulated averages (dashed lines), in the bivariate model $(d=2)$ under exponential decay, with $t \in[0,10]$. Parameters: $\bar{\lambda}_{1}=\bar{\lambda}_{2}=0.5, \mu_{1}=\mu_{2}=2, \alpha_{11}=\alpha_{12}=2.3, \alpha_{21}=\alpha_{22}=2, B_{11} \equiv 1.3$, $B_{12} \equiv 0.6, B_{21} \equiv 0.8, B_{22} \equiv 0.5$.
we will select parameters such that the bivariate stability condition in Eqn. (2.58) is satisfied.

In Figure 2.2, we plot the expectations $\mathbb{E}\left[Q_{1}(t)\right]$ and $\mathbb{E}\left[\lambda_{1}(t)\right]$, the variances $\operatorname{Var}\left[Q_{1}(t)\right]$ and $\operatorname{Var}\left[\lambda_{1}(t)\right]$, and the joint moments $\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right]$ and $\mathbb{E}\left[Q_{1}(t) Q_{2}(t)\right]$, obtained from the fixed-point characterization. These are plotted against their Monte Carlo simulated counterparts based on $M=10^{5}$ runs, for $t \in[0,10]$. In all cases, the numerical evaluation of our exact results closely matches the Monte Carlo simulated counterparts. Around $t=6$, the effect of the initial state has vanished, in that the processes enter the stationary regime. We note that the CPU time associated with the exact results is negligible compared to the CPU time required for the simulation results. Furthermore, for large $t$, the exact results remain smooth while the simulation results become a little rougher.

## Power

We next let the decay functions $g_{i j}(\cdot)$, with $i, j \in[2]$, be of power-law type. In particular, we take

$$
\begin{equation*}
g_{11}(t)=g_{12}(t)=\frac{1}{\left(c_{1}+t\right)^{p_{1}}}, \quad g_{21}(t)=g_{22}(t)=\frac{1}{\left(c_{2}+t\right)^{p_{2}}}, \tag{2.60}
\end{equation*}
$$

where $c_{1}, c_{2}>0$ and $p_{1}, p_{2}>1$ to ensure integrability. Again taking $B_{i j} \equiv b_{i j} \in \mathbb{R}_{+}$, we now have $\left\|h_{i j}\right\|=b_{i j} c_{i}^{1-p_{i}}\left(p_{i}-1\right)^{-1}$. As before, we choose the parameters such that the bivariate stability condition in Eqn. (2.58) is satisfied.

Figure 2.3 displays the expectations $\mathbb{E}\left[Q_{1}(t)\right]$ and $\mathbb{E}\left[\lambda_{1}(t)\right]$, the variances $\operatorname{Var}\left[Q_{1}(t)\right]$ and $\operatorname{Var}\left[\lambda_{1}(t)\right]$, as well as the joint moments $\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right]$ and $\mathbb{E}\left[Q_{1}(t) Q_{2}(t)\right]$, plotted against their Monte Carlo simulated counterparts. We observe again a highly accurate match. We note that, compared to the case of exponential decay, it now takes longer for the processes to reach the stationary regime. This is due to the fatter tails of power-law decay: heuristically, the processes now 'have more memory'.

### 2.6.2 Asymptotic analysis

We proceed by numerically illustrating our asymptotic results on the tail probabilities of $N_{i}(t)$, as established in Corollary 2.1. In particular, we compute the $\bar{R}_{i j}^{N, \delta}(\cdot)$,function appearing in Corollary 2.1 and the corresponding tail probability approximations, and compare these to tail probabilities estimated using Monte Carlo simulation.

In our specific bivariate example, we consider only one direction of cross-excitation and one heavy-tailed random variable. More precisely, we set $B_{11} \equiv 0, B_{21} \equiv 0$, $B_{12}=1$ and assume $B_{22} \in \operatorname{APT}(1, \gamma)$ with $\gamma=1.8$, such that this system can be represented by the following Hawkes graph:


Figure 2.3. Plots of the expectations and variances of $Q_{1}(\cdot)$ and $\lambda_{1}(\cdot)$ and the joint moments $\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right]$ and $\mathbb{E}\left[Q_{1}(t) Q_{2}(t)\right]$ (solid lines), compared to Monte Carlo simulated averages (dashed lines), in the bivariate model $(d=2)$ under power-law decay, with $t \in[0,20]$. Parameters: $\bar{\lambda}_{1}=\bar{\lambda}_{2}=1, \mu_{1}=\mu_{2}=1.5, c_{1}=1.5, c_{2}=2, p_{1}=2.5, p_{2}=3, B_{11} \equiv 1.5$, $B_{12} \equiv 0.5, B_{21} \equiv 1, B_{22} \equiv 0.5$.


The index $\gamma$ of the heavy-tailed random variable $B_{22}$ is inherited by both $N_{1}(\cdot)$ and $N_{2}(\cdot)$. In this setting, we compute the functions $R_{i j}^{N}(\cdot)$ and $\bar{R}_{i j}^{N, \gamma}(\cdot)$, with $i, j \in[2]$, using the system of equations of their Laplace transforms given in Eqns. (2.47), (2.48) and (2.54). By solving this linear system and inverting back, we obtain $R_{i j}^{N}(\cdot)$ at time $u$ by

$$
\begin{align*}
& R_{12}^{N}(u)=\mathbb{E}\left[B_{12}\right] \mathcal{L}^{-1}\left\{\frac{\mathcal{L}\left\{G_{12}\right\}(\cdot)}{1-\mathbb{E}\left[B_{22}\right] \mathcal{L}\left\{g_{22}\right\}(\cdot)}\right\}(u),  \tag{2.61}\\
& R_{22}^{N}(u)=\mathcal{L}^{-1}\left\{\frac{\mathcal{L}\{1\}(\cdot)}{1-\mathbb{E}\left[B_{22}\right] \mathcal{L}\left\{g_{22}\right\}(\cdot)}\right\}(u),
\end{align*}
$$

where $G_{12}(u)=\int_{0}^{u} g_{12}(v) \mathrm{d} v$. Note that $R_{11}^{N}(\cdot) \equiv 1$ by Eqn. (2.45) since $B_{11} \equiv 0$, and $R_{21}^{N}(\cdot) \equiv 0$ by Lemma 2.4 as $P_{2 \leftarrow 1}=0$. In a similar manner, we obtain $\bar{R}_{i j}^{N, \gamma}(\cdot)$ at time $u$ by

$$
\begin{align*}
& \bar{R}_{12}^{N, \gamma}(u)=\mathcal{L}^{-1}\left\{\frac{\mathcal{L}\left\{\left(g_{22} * \bar{R}_{12}^{N}\right)^{\gamma}\right\}(\cdot)}{1-\mathbb{E}\left[B_{22}\right] \mathcal{L}\left\{g_{22}\right\}(\cdot)}\right\}(u), \\
& \bar{R}_{22}^{N, \gamma}(u)=\mathcal{L}^{-1}\left\{\frac{\mathcal{L}\left\{\left(g_{22} * \bar{R}_{22}^{N}\right)^{\gamma}\right\}(\cdot)}{1-\mathbb{E}\left[B_{22}\right] \mathcal{L}\left\{g_{22}\right\}(\cdot)}\right\}(u) . \tag{2.62}
\end{align*}
$$

This allows us to compute the analytical expressions that appear in Corollary 2.1, which in our bivariate setting are $\int_{0}^{t} \bar{R}_{12}^{\boldsymbol{N}, \gamma}(u) \mathrm{d} u$ and $\int_{0}^{t} \bar{R}_{22}^{\boldsymbol{N}, \gamma}(u) \mathrm{d} u$ for components $N_{1}(t)$ and $N_{2}(t)$, respectively. We next discuss two parameterizations of the decay functions to compute these terms explicitly.

## Exponential

We choose our decay functions as in Eqn. (2.59), and select the remaining parameters such that the stability condition (2.58) is satisfied. In Figure 2.4, we plot the analytical expressions from Corollary 2.1 against the Monte Carlo simulation-based approximations of the tail probabilities $\mathbb{P}\left(N_{1}(t)>x\right)$ and $\mathbb{P}\left(N_{2}(t)>x\right)$ as $x$ grows, for fixed $t=1$. The simulations are performed by sampling $M=2 \cdot 10^{6}$ runs of $\boldsymbol{N}(1)=\left(N_{1}(1), N_{2}(1)\right)$ and counting the proportion of these runs that lead to values larger than any given threshold $x$. The number of simulation runs is chosen sufficiently large to obtain reasonable estimates of small tail probabilities. As expected, we see that as $x$ grows, the simulation approximations of the tail probabilities converge toward the analytical asymptotic expressions.


Figure 2.4. Plots of the tail probabilities $\mathbb{P}\left(N_{1}(t)>x\right)$ and $\mathbb{P}\left(N_{2}(t)>x\right)$ for fixed $t=1$ in the bivariate model $(d=2)$ under exponential decay, using the analytical expressions appearing in Corollary 2.1 (solid lines), compared to Monte Carlo simulated approximations (dashed lines). Parameters: $\bar{\lambda}_{1}=0.5, \bar{\lambda}_{2}=1.5, \alpha_{11}=\alpha_{12}=\alpha_{21}=\alpha_{22}=1.5, B_{11} \equiv 0$, $B_{12} \equiv 1, B_{21} \equiv 0, B_{22} \in \operatorname{APT}(1,1.8)$.

## Power

We now choose our decay functions as in Eqn. (2.60), and the remaining parameters such that the stability condition (2.58) is satisfied. We substitute the decay functions into (2.61) and (2.62), and compute the analytical expressions appearing in Corollary 2.1. Figure 2.5 is the counterpart of Figure 2.4, but now for power-law decay. We observe again that both Monte Carlo simulated approximations converge to the analytical asymptotic expressions.

### 2.7 Conclusions

This paper has focused on multivariate Hawkes processes and associated population processes, establishing both exact and asymptotic results. Importantly, we allow our model to be non-Markovian, in that the processes' decay functions can be chosen generally. We have characterized their joint transform via a fixed-point theorem, by exploiting suitable extensions of the existing results on cluster representations for Hawkes processes and associated distributional equalities induced by the underlying branching structure. For the case that the intensity jumps are heavy-tailed, we have also succeeded in determining the corresponding asymptotic tail behavior.

A (partially) more general setting than that analyzed in our paper is provided by the


Figure 2.5. Plots of the tail probabilities $\mathbb{P}\left(N_{1}(t)>x\right)$ and $\mathbb{P}\left(N_{2}(t)>x\right)$ for fixed $t=1$ in the bivariate model $(d=2)$ under power-law decay, using the analytical expressions appearing in Corollary 2.1 (solid lines), compared to Monte Carlo simulated approximations (dashed lines). Parameters: $\bar{\lambda}_{1}=0.5, \bar{\lambda}_{2}=1.5, c_{1}=c_{2}=1, p_{1}=2.5, p_{2}=3.5, B_{11} \equiv 0, B_{12} \equiv 1$, $B_{21} \equiv 0, B_{22} \in \operatorname{APT}(1,1.8)$.
extensions of Hawkes processes introduced in [12], referred to as non-linear Hawkes processes, and versatilely analyzed further in [70, 74, 77]. Then, the form of the intensity process is governed by a general auxiliary function, rather than the additive structure (2.3). Importantly, the cluster representation applies to the additive case only. As a consequence, our results cannot be readily generalized to the non-linear case, which will require an intrinsically different approach.

## 2.A Additional Proofs

Proof of Lemma 2.1. We start by considering a pair of matrix-valued processes

$$
(\boldsymbol{X}(\cdot), \boldsymbol{Y}(\cdot))=\left(X_{i j}(\cdot), Y_{i j}(\cdot)\right)_{i, j \in[d]}
$$

and denote by $\boldsymbol{X}_{(i)}(\cdot)$ and $\boldsymbol{Y}_{(i)}(\cdot)$ the $i$-th rows of the respective matrices. We define a mapping $\Phi$ acting on the space of such pairs of processes at time $u$ by

$$
\begin{align*}
\left(X_{i j}(u), Y_{i j}(u)\right)_{i, j \in[d]} & \mapsto \Phi\left(\left(X_{i j}(\cdot), Y_{i j}(\cdot)\right)_{i, j \in[d]}\right)(u) \equiv \Phi(\boldsymbol{X}, \boldsymbol{Y})(u), \\
& =:\left(\mathcal{A}_{j}\left\{\mathbf{1}_{\{i=j\}} \mathbf{1}_{\left\{E_{i}>u\right\}}, \boldsymbol{X}_{(i)}(\cdot)\right\}(u), \mathcal{A}_{j}\left\{B_{i j} g_{i j}(u), \boldsymbol{Y}_{(i)}(\cdot)\right\}(u)\right)_{i, j \in[d]}, \tag{2.63}
\end{align*}
$$

where $\mathcal{A}_{j}$ is as defined in Eqn. (2.18). We note that the above mapping could be equivalently expressed in terms of distribution functions of $\boldsymbol{X}(\cdot)$ and $\boldsymbol{Y}(\cdot)$. Now observe

## Chapter 2 Multivariate Hawkes Population Processes

that the mapping in Eqn. (2.63) is the probabilistic analogue of the mapping $\phi$ given in Eqn. (2.33). We can verify this, as the time-dependent joint transform $\mathcal{J}_{\Phi(\boldsymbol{X}, \boldsymbol{Y})}(\cdot)$ is, at time $u$, due to Definition 2.4, explicitly given by

$$
\mathcal{J}_{\Phi(\boldsymbol{X}, \boldsymbol{Y})}(u)=\left[\begin{array}{c}
\mathcal{J}_{\Phi_{1}(\boldsymbol{X}, \boldsymbol{Y})}(u)  \tag{2.64}\\
\vdots \\
\mathcal{J}_{\Phi_{d}(\boldsymbol{X}, \boldsymbol{Y})}(u)
\end{array}\right],
$$

where each entry, for $j \in[d]$ corresponding to the $j$-th column $\Phi_{j}(\boldsymbol{X}, \boldsymbol{Y})(\cdot)$, is given by

$$
\begin{align*}
\mathcal{J}_{\Phi_{j}(\boldsymbol{X}, \boldsymbol{Y})}(u) & \equiv \mathcal{J}_{\Phi_{j}(\boldsymbol{X}, \boldsymbol{Y})}(u, \boldsymbol{s}, \boldsymbol{z}) \\
& =\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{\mathcal{A}_{j}\left\{\mathbf{1}_{\{i=j\}} \mathbf{1}_{\left\{E_{i}>u\right\}}, \boldsymbol{X}_{(i)} \cdot(\cdot)\right\}(u)} e^{-s_{i} \mathcal{A}_{j}\left\{B_{i j} g_{i j}(u), \boldsymbol{Y}_{(i)} \cdot(\cdot)\right\}(u)}\right] . \tag{2.65}
\end{align*}
$$

Mimicking the steps of the proof of Theorem 2.2, we derive that the right-hand side of Eqn. (2.65) equals $\phi_{j}\left(\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}\right)(u)$. Therefore, the joint transform of $\Phi_{j}(\boldsymbol{X}, \boldsymbol{Y})(\cdot)$ satisfies $\mathcal{J}_{\Phi_{j}(\boldsymbol{X}, \boldsymbol{Y})}(\cdot)=\phi_{j}\left(\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}\right)(\cdot) \in \mathbb{J}$. Since this holds for every entry $j \in[d]$, we have that $\phi\left(\mathcal{J}_{\boldsymbol{X}, \boldsymbol{Y}}\right)(\cdot) \in \mathbb{J}^{d}$.

Proof of Lemma 2.2. Let $\mathcal{J}(\cdot), \tilde{\mathcal{J}}(\cdot) \in \mathbb{J}^{d}$ and $\epsilon>0$. We will show that for a certain choice of $\delta>0$, we have

$$
\|\mathcal{J}-\tilde{\mathcal{J}}\|_{\mathbb{J}^{d}}<\delta \Longrightarrow\|\phi(\mathcal{J})-\phi(\widetilde{\mathcal{J}})\|_{\mathbb{J}^{d}}<\epsilon
$$

It suffices to prove continuity in each entry separately. Considering the $j$-th entry of $\phi$, which is the mapping $\phi_{j}$ defined in (2.33), observe that it can be rewritten as

$$
\begin{aligned}
& \phi_{j}\left(\mathcal{J}_{1}, \ldots, \mathcal{J}_{d}\right)(u, \boldsymbol{s}, \boldsymbol{z})=\mathbb{E}\left[z_{j}^{\mathbf{1}_{\left\{E_{j}>u\right\}}}\right] \\
& \quad \times \prod_{m=1}^{d} \mathbb{E}\left[\exp \left(-B_{m j}\left(s_{m} g_{m j}(u)+\int_{0}^{u} g_{m j}(v)\left(1-\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right)\right)\right] .
\end{aligned}
$$

We then have

$$
\begin{align*}
& \left\|\phi_{j}(\mathcal{J})-\phi_{j}(\widetilde{\mathcal{J}})\right\|_{\mathbb{J}}^{2}  \tag{2.66}\\
& =\sup _{u, \boldsymbol{s}, \boldsymbol{z}}\left|\phi_{j}(\mathcal{J})(u, \boldsymbol{s}, \boldsymbol{z})-\phi_{j}(\widetilde{\mathcal{J}})(u, \boldsymbol{s}, \boldsymbol{z})\right|^{2} \\
& \leqslant \sup _{u, \boldsymbol{s}, \boldsymbol{z}} \mid \mathbb{E}\left[\exp \left(-\sum_{m=1}^{d} B_{m j}\left(s_{m} g_{m j}(u)+\int_{0}^{u} g_{m j}(v)\left(\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})-1\right) \mathrm{d} v\right)\right)\right.  \tag{2.67}\\
& \left.\quad-\exp \left(-\sum_{m=1}^{d} B_{m j}\left(s_{m} g_{m j}(u)+\int_{0}^{u} g_{m j}(v)\left(\widetilde{\mathcal{J}}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})-1\right) \mathrm{d} v\right)\right)\right]\left.\right|^{2}
\end{align*}
$$

where we have used that $\left|z_{j}\right| \leqslant 1$. We then apply the mean value theorem to the difference of the exponential terms; for an exponential function, it states that $e^{a}-e^{b}=$ $(a-b) e^{c}$ for some $c \in[a, b]$. We have that

$$
-\sum_{m=1}^{d} B_{m j}\left(s_{m} g_{m j}(u)+\int_{0}^{u} g_{m j}(v)\left(1-\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right) \leqslant 0
$$

due to $\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z}) \leqslant 1, B_{m j}$ being non negative, $s \in \mathbb{R}_{+}^{d}$ and $g_{m j}(v) \geqslant 0$ by assumption. The same holds for the terms involving $\widetilde{\mathcal{J}}_{m}$. Hence, we can apply the mean value theorem with some $c \leqslant 0$. We thus obtain the following upper bound on (2.67):

$$
\begin{equation*}
\sup _{u, \boldsymbol{s} \boldsymbol{z}}\left|\sum_{m=1}^{d} \mathbb{E}\left[B_{m j} \int_{0}^{u} g_{m j}(v)\left(\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})-\widetilde{\mathcal{J}}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right]\right|^{2}, \tag{2.68}
\end{equation*}
$$

since the $B_{m j} s_{m} g_{m j}(u)$ terms and the constants cancel. By an application of the triangle inequality we can bound this expression further. Indeed, (2.68) is dominated by

$$
\begin{align*}
& \sum_{m=1}^{d} \sup _{u, \boldsymbol{s}, \boldsymbol{z}} \mathbb{E}\left|B_{m j} \int_{0}^{u} g_{m j}(v)\left(\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})-\widetilde{\mathcal{J}}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right|^{2} \\
& \leqslant \sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right]^{2} \sup _{u, \boldsymbol{s}, \boldsymbol{z}}\left|\int_{0}^{u} g_{m j}(v)\left(\mathcal{J}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})-\widetilde{\mathcal{J}}_{m}(u-v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right|^{2}, \tag{2.69}
\end{align*}
$$

where $\mathbb{E}\left[B_{m j}\right]^{2}<\infty$ by assumption. Finally, this term can be bounded by applying Young's inequality for convolutions, which yields that (2.69) is dominated by

$$
\begin{align*}
& \sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right]^{2} \sup _{u, \boldsymbol{s}, \boldsymbol{z}}\left|\int_{0}^{u} g_{m j}(v) \mathrm{d} s \int_{0}^{u}\left(\mathcal{J}_{m}(v, \boldsymbol{s}, \boldsymbol{z})-\widetilde{\mathcal{J}}_{m}(v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right|^{2} \\
& \leqslant d \max _{m, j \in[d]} \mathbb{E}\left[B_{m j}\right]^{2}\left|\int_{0}^{t} g_{m j}(v) \mathrm{d} v\right|^{2} \sup _{u, \boldsymbol{s}, \boldsymbol{z}}\left|\int_{0}^{u}\left(\mathcal{J}_{m}(v, \boldsymbol{s}, \boldsymbol{z})-\widetilde{\mathcal{J}}_{m}(v, \boldsymbol{s}, \boldsymbol{z})\right) \mathrm{d} v\right|^{2}  \tag{2.70}\\
& \leqslant d \max _{m, j \in[d]} \mathbb{E}\left[B_{m j}\right]^{2}\left\|g_{m j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}^{2} t \delta^{2}
\end{align*}
$$

where $\left.\left\|g_{m j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}^{2}=\left(\int_{0}^{\infty} g_{m j}(v)\right] \mathrm{d} v\right)^{2}<\infty$ and $\left|\mathcal{J}_{m}(v, \boldsymbol{s}, \boldsymbol{z})-\widetilde{\mathcal{J}}_{m}(v, \boldsymbol{s}, \boldsymbol{z})\right|<\delta$ by assumption, in combination with the standard inequality $\sum_{m=1}^{d} a_{i} \leqslant d \max _{i \in[d]}\left\{a_{i}\right\}$ for real numbers $a_{i}$. Hence, choosing $\delta$ as

$$
\delta^{2}=\frac{\epsilon}{t d^{2} \max _{m, j \in[d]} \mathbb{E}\left[B_{m j}\right]^{2}\left\|g_{m j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}^{2}},
$$

## Chapter 2 Multivariate Hawkes Population Processes

implies that (2.66) becomes $\left\|\phi_{j}(\mathcal{J})-\phi_{j}(\tilde{\mathcal{J}})\right\|_{\mathcal{J}}^{2} \leqslant \epsilon^{2}$. Doing this for all $j \in[d]$ yields the result.

Of course, for the proof of Lemma 2.2 to work, we need $t>0, \mathbb{E}\left[B_{m j}\right]>0$ and $g_{m j} \neq 0$ for at least some combination of $m, j \in[d]$. However, it is clear that choosing all of these to be 0 yields an irrelevant model.

Proof of Lemma 2.4. (i) $\Rightarrow$ (iii). We have that $\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]>0$ implies that $\mathbb{P}\left(S_{i \leftarrow j}^{Q}(u)>\right.$ 0 ) is a positive probability, implying that a path must exist from component $j$ to $i$, which proves the result immediately.
(iii) $\Rightarrow(i)$. Using the distributional equality of $S_{i \leftarrow j}^{Q}(u)$, as given in (2.19), we have

$$
\begin{equation*}
\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]=\mathbf{1}_{\{i=j\}} \mathbb{P}\left(E_{i}>u\right)+\sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v) \mathbb{E}\left[S_{i \leftarrow m}^{Q}(u-v)\right] \mathrm{d} v . \tag{2.71}
\end{equation*}
$$

If $P_{i \leftarrow j}=1$, then the path must start with an edge $e_{k j}$, for some $k \in[d]$ and so $\mathbb{E}\left[B_{k j}\right]>0$ by definition. If $k=i$, then $\mathbb{E}\left[B_{i j}\right]>0$ and so this direct link implies $\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]>0$. If $k \neq i$, then we apply Eqn. (2.71) to $\mathbb{E}\left[S_{i \leftarrow k}^{Q}(u-v)\right]$ on the RHS to obtain the next edge along the path. Iterating this procedure for the non-zero terms on the RHS, we obtain a path from component $j$ to $i$, proving that $\mathbb{E}\left[S_{i \leftarrow j}^{Q}(u)\right]>0$. $(i) \Leftrightarrow(i i)$. We can use the same argument, now for $\boldsymbol{\lambda}$. We have that $\mathbb{E}\left[S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u)\right]$ satisfies

$$
\begin{equation*}
\mathbb{E}\left[S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u)\right]=\mathbb{E}\left[B_{i j}\right] g_{i j}(u)+\sum_{m=1}^{d} \mathbb{E}\left[B_{m j}\right] \int_{0}^{u} g_{m j}(v) \mathbb{E}\left[S_{i \leftarrow m}^{\boldsymbol{\lambda}}(u-v)\right] \mathrm{d} v, \tag{2.72}
\end{equation*}
$$

due to the distributional equality for $S_{i \leftarrow j}^{\boldsymbol{\lambda}}(u)$ that is stated in (2.19). A similar reasoning as above yields the result.

## Chapter 3

## Multivariate Compound Hawkes Processes: <br> Large Deviations and Rare Event Simulation


#### Abstract

In this paper, we establish a large deviations principle for a multivariate compound process induced by a multivariate Hawkes process with random marks. Our proof hinges on showing essential smoothness of the limiting cumulant of the multivariate compound process, resolving the inherent complication that this cumulant is implicitly characterized through a fixed-point representation. We employ the large deviations principle to derive logarithmic asymptotic results on the marginal ruin probabilities of the associated multivariate risk process. We also show how to conduct rare event simulation in this multivariate setting using importance sampling, and prove the asymptotic efficiency of our importance sampling based estimators. The paper is concluded with a systematic assessment of the performance of our rare event simulation procedure.


## Chapter 3

## Multivariate Compound Hawkes Processes: Large Deviations and Rare Event Simulation

### 3.1 Introduction

Mutually exciting processes, or multivariate Hawkes processes ([45, 47]), constitute an important class of point processes, particularly suitable to describe stochastic dependences among occurrences of events across time and space. Due to their built-in feedback mechanism, they are a natural contender to model contagious phenomena where clusters of events occur in both the temporal and spatial dimensions. Over the past decades, Hawkes processes have been increasingly applied across a broad variety of fields, such as finance ([2, 3, 6, 48]), social interaction ([17]), neuroscience ([65]), seismology $([64,49])$, and many others.

The key property of a Hawkes process is that it exhibits 'self-exciting' behavior: informally, any event instantaneously increases the likelihood of, hence potentially triggers, additional future events. A crucial element in its definition is the so-called decay function that quantifies how quickly the effect of an initial event on future events vanishes. Choosing this function to be exponential renders the model Markovian, which facilitates the explicit evaluation of various relevant risk and performance metrics (e.g., transient and stationary moments). In practical applications, however, it can be more natural to allow for other, i.e., not necessarily exponential, decay functions admitting non-Markovian, and long-memory, properties but making the analysis substantially more challenging; see, e.g., $[35,59]$ who indicate the relevance of nonMarkovian models in describing contagious phenomena.

In applied probability and mathematical risk theory, Hawkes processes have been used to model the claim arrival process, and, likewise, compound Hawkes processes

## Chapter 3 Multivariate Compound Hawkes Processes

to model the associated cumulative claim process that an insurance firm is facing; see the related literature discussed in detail below. In collective risk theory, multivariate Hawkes processes provide an appealing candidate for modeling, for example, the claim arrival process associated with technological risks; see, e.g., [8] who apply the contagion model of [2] to model cyber attacks and also [9]. Bearing in mind that ruin and exceedance probabilities ought to be kept small, a primary research goal concerns their analysis in the asymptotic regime in which the initial reserve level of the insurer or the time horizon of aggregation grows large, such that the event of interest becomes increasingly rare. This explains the interest in deriving large deviations principles for (compound) Hawkes processes, providing a formal tool to assess their rare event behavior, and facilitating in particular the identification of the asymptotics of ruin and exceedance probabilities. At the same time, it is noted, however, that large deviations results usually yield rough, logarithmic asymptotics only, in that they focus on identifying the associated decay rate. To remedy this, one could attempt to develop 'large deviations informed' simulation techniques by which rare events can be evaluated fast and accurately. This is particularly useful when the probabilities of rare events are too small to be estimated with reasonable accuracy using regular Monte Carlo simulations.

Whereas large deviations for univariate Hawkes processes are well understood, their multivariate counterpart is to a large extent unexplored. In this context, we mention [73], which considers a broad class of multivariate affine processes of Markovian type, covering the special case of the multivariate Hawkes process with an exponential decay function; see also [42] for refinements. In addition, a moderate deviations result has been derived in [72]. To the best of our knowledge, however, large deviations principles for multivariate compound Hawkes processes allowing for general decay functions have not been established, and in addition, no rare event simulation techniques have been developed in this setting. These are the main subjects of this paper.

In the univariate case, large deviations results for compound Hawkes processes with general decay function have been derived in [69] building on [11]. The underlying argumentation relies on the cluster representation of the driving Hawkes process, as developed in the seminal work [47], from which it is concluded that the cluster size follows a so-called Borel distribution. A crucial element in proving the large deviations principle lies in showing that the limiting cumulant of the random object under study is steep, entailing that its derivative grows to infinity when approaching the boundary of its domain, such that the Gärtner-Ellis Theorem can be invoked. In the univariate (compound) Hawkes case considered in [69], steepness could be established by using the explicit expression for the cluster size distribution. When studying large deviations for multivariate (compound) Hawkes processes, however, a main technical difficulty that
arises is that the cluster size distribution of the process is not known in closed form. In the multivariate setting, all one has is a vector-valued fixed-point representation for the limiting cumulant of the (multivariate) cluster size, as was derived in [55]. Importantly, this relation does not allow the closed-form identification of the limiting cumulant (let alone that one can find the distribution itself), entailing that one cannot explicitly characterize the boundary of its domain.

In light of the gaps in the literature described above, the contributions of this paper are the following.

- First, we establish a large deviations principle for multivariate compound Hawkes processes allowing for general decay functions. We also allow the Hawkes process to be marked, such that the intensity process experiences jumps of random size. We have succeeded in establishing steepness based on an implicit fixedpoint representation for the limiting cumulant of the joint cluster size distribution. Specifically, without having an explicit expression for the limiting cumulant, and without having an explicit characterization of its domain, we prove that the derivative of the limiting cumulant grows to infinity when approaching the boundary of its domain. This steepness property facilitates the use of the Gärtner-Ellis Theorem, so as to establish the desired large deviations principle. The mathematical details of the required multivariate analysis are involved.
- Second, we characterize the asymptotic behavior of the ruin probability for the marginal ruin processes in the regime that the initial reserve level grows large. We prove that this ruin probability decays essentially exponentially, with the corresponding decay rate being equal to the unique zero of the limiting cumulant pertaining to that marginal. The proof of the lower bound on the decay rate is a direct application of our large deviations principle for multivariate compound Hawkes processes. The corresponding upper bound is established by a timediscretization argument, a union bound, and by showing that one can neglect the contributions to the resulting sum due to small and large time scales, in combination with frequent use of the well-known Chernoff bound.
- Third, we develop an importance sampling algorithm for estimating rare event probabilities in our multivariate setting. More precisely, we first derive the parameters of the exponentially twisted multivariate Hawkes process. The twisted marginal ruin process has positive drift, yielding ruin with probability one under the new measure, but with the likelihood ratio being bounded by a function that decays exponentially in the initial reserve, thus leading to a considerable speedup in importance sampling relative to regular Monte Carlo simulation. We
prove that this estimator is, in fact, asymptotically efficient in the sense of Siegmund ([68]), in passing also establishing a Lundberg-type upper bound on the ruin probability. In addition, we devise an asymptotically efficient importance sampling algorithm for estimating the probability of the multivariate compound Hawkes process (at a given point in time, that is) attaining a rare large value. The attainable speedup, relative to regular Monte Carlo simulation, is quantified through a series of simulation experiments.

Without attempting to provide an exhaustive overview, we now review a few relevant related papers, all of which focus on the univariate setting. We already mentioned [69], which analyzes the asymptotic behavior of ruin probabilities, under the assumption of light-tailed claims, drawing upon earlier large deviations results derived in [11] for more general Poisson cluster processes. Furthermore, in [69], an importance sampling based algorithm is proposed that is capable of efficiently generating estimates of the rare event probabilities of interest. In [54], the limiting cumulant of the cluster size distribution is implicitly characterized for the setting with random marks using a fixed-point argument, while proving a large deviations principle using the Gärtner-Ellis theorem for the upper bound and an exponential tilting method for the lower bound. Where the contributions above focus primarily on the case of light-tailed claims, subexponentially distributed claims are studied in [74], in the context of a non-stationary version of the Hawkes process. For (non-compound) Hawkes processes (i.e., not involving claims), 'precise' large deviations results, providing asymptotics beyond the leading order term, are obtained in [43]. The setting of a large initial intensity is studied in [40] and [41]. For the more general class of non-linear Hawkes processes, [76] proves the process-level large deviations, and [77] derives large deviations in the Markovian setting.

The rest of this paper is organized as follows. In Section 3.2, we introduce the relevant processes and discuss some basic properties that are used throughout the paper. Section 3.3 derives results on the transform of the joint cluster size distribution, and provides an implicit characterization of the domain of the limiting cumulant. Section 3.4 then establishes the large deviations principle for the multivariate compound Hawkes process with general decay function and random marks. In Section 3.5, we consider the associated multivariate risk process, with the objective to characterize the decay rate of the marginal ruin probability. Then, in Section 3.6, we exploit the large deviations principle to develop an importance sampling algorithm to efficiently estimate rare event probabilities. Section 3.7 numerically demonstrates the performance of our importance sampling based estimators. Concluding remarks are in Section 3.8. Some auxiliary proofs are relegated to the Appendix.

### 3.2 Multivariate Compound Hawkes Processes

In this section, we first provide the definitions of multivariate Hawkes and compound Hawkes processes, and next introduce some objects and discuss some properties that are relevant in the context of this paper. Throughout, we use the boldface notation $\boldsymbol{x}=\left(x_{1}, \ldots, x_{d}\right)^{\top}$ to denote a $d$-dimensional vector, for a given dimension $d \in \mathbb{N}$. Inequalities between vectors are understood componentwise, e.g., $\boldsymbol{x}>\boldsymbol{y}$ means $x_{i}>y_{i}$ for all $i=1, \ldots, d$.

Consider a $d$-dimensional càdlàg counting process $\boldsymbol{N}(\cdot) \equiv(\boldsymbol{N}(t))_{t \in \mathbb{R}_{+}}$, where each increment $N_{i}(t)-N_{i}(s)$ records the number of points in component $i \in[d]:=\{1, \ldots, d\}$ in the time interval $(s, t]$, with $s<t$. We label the points by considering, for each $j \in[d]$, a sequence of a.s. increasing positive random variables $\boldsymbol{T}_{j}=\left\{T_{j, r}\right\}_{r \in \mathbb{N}}=$ $\left\{T_{j, 1}, T_{j, 2}, \ldots\right\}$ representing the event times. We associate to this sequence the onedimensional counting process $N_{j}(\cdot)$ by setting

$$
N_{j}(t):=N_{\boldsymbol{T}_{j}}(0, t]=\sum_{r=1}^{\infty} \mathbf{1}_{\left\{T_{j, r} \leq t\right\}} .
$$

The process $\boldsymbol{N}(\cdot)=\left(N_{1}(\cdot), \ldots, N_{d}(\cdot)\right)^{\top}$ is then the $d$-dimensional counting process associated with the sequences of event times in all components, $\boldsymbol{T}_{1}, \ldots, \boldsymbol{T}_{d}$, compactly denoted by $\boldsymbol{N}(t)=\boldsymbol{N}_{\boldsymbol{T}}(0, t]$. Throughout, the points will be referred to as events and the terms point process and counting process are used interchangeably for $\boldsymbol{N}(\cdot)$. We assume that the point process starts empty, i.e., $\boldsymbol{N}(0)=\mathbf{0}=(0, \ldots, 0)^{\top}$.

In the original work [45], the Hawkes process is defined by relying on the concept of the conditional intensity function. An alternative, equivalent definition, known as the cluster process representation, can be given by representing the Hawkes process as a Poisson cluster process; it was first described in [47] in the setting of the conventional univariate Hawkes process, see also [23, Example 6.3(c)] and [60, Ch. IV]. The cluster process representation distinguishes between two types of events: first, there are immigrant events generated according to a homogeneous Poisson process with a given rate; and second, there are offspring events generated by an inhomogeneous Poisson process with rates that account for self-excitation and, in the multivariate context also, cross-excitation. In the following, we introduce the relevant terminology and provide a formal definition of the process.

For $j \in[d]$, we consider base rates $\bar{\lambda}_{j} \geqslant 0$, with at least one of the base rates being strictly positive. For each combination $i, j \in[d]$, we let the decay function $g_{i j}(\cdot)$ : $\mathbb{R}_{+} \rightarrow \mathbb{R}_{+}$be non-negative, non-increasing, and integrable. Also, for $j \in[d]$, we define the random marks through the generic non-negative, non-degenerate random vector $\boldsymbol{B}_{j}=\left(B_{1 j}, \ldots, B_{d j}\right)$, asserting that the sequence of random marks $\left\{\boldsymbol{B}_{j, r}\right\}_{r \in \mathbb{N}}$ consists

## Chapter 3 Multivariate Compound Hawkes Processes

of i.i.d. random vectors that are distributed as $\boldsymbol{B}_{j}$. We allow the random variables $B_{i j, r}$ to be dependent for fixed $j$ and $r$. Finally, we let $K_{i j}(\cdot)$ be an inhomogeneous Poisson process with intensity $B_{i j, r} g_{i j}(\cdot)$, given the value of $B_{i j, r}$. With these elements in place, the following definition describes the cluster process representation for a multivariate Hawkes process.
Definition 3.1 (multivariate Hawkes process). Define a d-dimensional point process $\boldsymbol{N}(\cdot)$ componentwise by $N_{j}(t)=N_{\boldsymbol{T}_{j}}(0, t]$ for $j \in[d]$ and $t>0$, where the sequences of event times $\boldsymbol{T}_{1}, \ldots, \boldsymbol{T}_{d}$ are generated as follows:
(i) First, for each $j \in[d]$, let there be a sequence of immigrant event times $\left\{T_{j, r}^{(0)}\right\}_{r \in \mathbb{N}}$ on the interval $(0, \infty)$ generated by a homogeneous Poisson process $I_{j}(\cdot)$ with rate $\bar{\lambda}_{j}$.
(ii) Second, let each immigrant event independently generate a d-dimensional cluster $\boldsymbol{C}_{j} \equiv \boldsymbol{C}_{T_{j, r}^{(0)}}$, consisting of event times associated with generations of events:
(a) The immigrant with event time $T_{j, r}^{(0)}$ is labeled to be of generation 0 and into each component $m \in[d]$, it generates a sequence of first-generation event times $\left\{T_{m, r}^{(1)}\right\}_{r \in \mathbb{N}}$ on the interval $\left(T_{j, r}^{(0)}, \infty\right)$, according to $K_{m j}\left(\cdot-T_{j, r}^{(0)}\right)$ with $B_{m j, r}$ the random mark associated to $T_{j, r}^{(0)}$.
(b) Iterating (a) above, with $T_{m, r}^{(n-1)}$ designating the $r$-th event time of generation $n-1$ in component $m \in[d]$, yields generation $n$ event times $\left\{T_{l, r}^{(n)}\right\}_{r \in \mathbb{N}}$ in component $l \in[d]$ on the interval $\left(T_{m, r}^{(n-1)}, \infty\right)$, generated according to $K_{l m}\left(\cdot-T_{m, r}^{(n-1)}\right)$.

Upon taking the union over all generations, we obtain, for each component $j \in[d]$,

$$
\boldsymbol{T}_{j}=\left\{T_{j, r}\right\}_{r \in \mathbb{N}}=\bigcup_{n=0}^{\infty}\left\{T_{j, r}^{(n)}\right\}_{r \in \mathbb{N}} .
$$

The process $\boldsymbol{N}(\cdot)$ defined above for $t>0$ and with $\boldsymbol{N}(0)=\mathbf{0}$ constitutes a multivariate Hawkes process.

To ensure that the clusters described in part (ii) of Definition 3.1 are a.s. finite, we assume that a stability condition applies throughout this paper. It is shown in [45] that this stability condition guarantees non-explosiveness of $\boldsymbol{N}(\cdot)$, see also [23, Example 8.3(c)].

Assumption 3.1. Assume that the matrix $\boldsymbol{H}:=\left(h_{m j}\right)_{m, j \in[d]}$ with elements

$$
\begin{equation*}
h_{m j}:=\mathbb{E}\left[B_{m j}\right] c_{m j}, \tag{3.1}
\end{equation*}
$$

with $c_{m j}=\int_{0}^{\infty} g_{m j}(v) \mathrm{d} v$, has spectral radius strictly smaller than 1 .

We next define the multivariate compound Hawkes process as follows. Let $d^{\star} \in \mathbb{N}$ be fixed and note that we allow $d \neq d^{\star}$. Let $\left\{\boldsymbol{U}_{j, r}\right\}_{r \in \mathbb{N}}=\left\{\left(U_{1 j, r}, \ldots, U_{d^{\star} j, r}\right)^{\top}\right\}_{r \in \mathbb{N}}$ be a sequence of non-negative, non-degenerate i.i.d. random vectors of length $d^{\star}$, for each $j \in[d]$. We allow the random variables $U_{i j, r}$ to be dependent for fixed $j$ and $r$.

Definition 3.2 (multivariate compound Hawkes process). Define the process $\boldsymbol{Z}(\cdot)$ := $\left(Z_{1}(\cdot), \ldots, Z_{d^{\star}}(\cdot)\right)^{\top}$ for each component $Z_{i}(\cdot)$ with $i \in\left[d^{\star}\right]$ by

$$
\begin{equation*}
Z_{i}(t):=\sum_{j=1}^{d} \sum_{r=1}^{N_{j}(t)} U_{i j, r}, \quad t>0, \tag{3.2}
\end{equation*}
$$

where $\boldsymbol{U}_{j, r}=\left(U_{1 j, r}, \ldots, U_{d^{\star} j, r}\right)^{\top}$ is drawn independently for every event in $N_{j}(t)$, with $j \in[d]$. The process $\boldsymbol{Z}(\cdot)$ defined above for any $t>0$ constitutes a multivariate compound Hawkes process.

If we define the random matrix $\boldsymbol{U} \in \mathbb{R}_{+}^{d^{\star} \times d}$ as

$$
\boldsymbol{U} \equiv\left[\boldsymbol{U}_{1}, \ldots, \boldsymbol{U}_{d}\right]:=\left[\begin{array}{cccc}
U_{11} & U_{12} & \ldots & U_{1 d}  \tag{3.3}\\
U_{21} & U_{22} & \ldots & U_{2 d} \\
\vdots & \vdots & \ddots & \vdots \\
U_{d^{\star} 1} & U_{d^{\star} 2} & \cdots & U_{d^{\star} d}
\end{array}\right]
$$

we can represent Eqn. (3.2) in vector-matrix form by

$$
\begin{equation*}
\boldsymbol{Z}(t)=\boldsymbol{N}(t) \star \boldsymbol{U}:=\sum_{j=1}^{d} \sum_{r=1}^{N_{j}(t)} \boldsymbol{U}_{j, r}, \tag{3.4}
\end{equation*}
$$

where $\star$ denotes the compound sum operation.
We proceed with a brief discussion of the dimensionality of the objects appearing in Definition 3.2. The Hawkes process $\boldsymbol{N}(\cdot)$ is of dimension $d$ and the random vectors $\boldsymbol{U}_{j}$ are of dimension $d^{\star}$, which results in the compound Hawkes process $\boldsymbol{Z}(\cdot)$ being also of dimension $d^{\star}$. This reflects that the random variables of the type $U_{i j}$, with $i \in\left[d^{\star}\right]$ and $j \in[d]$, can be interpreted as the effect that an event in the $j$-th component of the Hawkes process $\boldsymbol{N}(\cdot)$ has on the $i$-the component of the compound Hawkes process $\boldsymbol{Z}(\cdot)$. As stated before, we allow $d \neq d^{\star}$. Intuitively, for instance, in the context of insurance, this means that the number of risk drivers may be larger $\left(d>d^{\star}\right)$ or smaller $\left(d<d^{\star}\right)$ than the number of insurance product categories.

We now introduce some objects related to the cluster process representation that are relevant for later analysis. Recall that for each immigrant in component $j \in[d]$, the $d$-dimensional cluster $\boldsymbol{C}_{j}$ from Definition 3.1 contains the sequences of event times

## Chapter 3 Multivariate Compound Hawkes Processes

in each component that have the immigrant with event time $T_{j, r}^{(0)}$ as oldest ancestor. We associate to $\boldsymbol{C}_{j}$ the $d$-dimensional cluster point process $\boldsymbol{S}_{j}(\cdot)$, by setting

$$
\begin{equation*}
\boldsymbol{S}_{j}(u):=\boldsymbol{S}_{\boldsymbol{C}_{j}}(0, u], \tag{3.5}
\end{equation*}
$$

such that it counts the number of events of $\boldsymbol{C}_{j}$ on the interval $(0, u$ ], where $u=$ $t-T_{j, r}^{(0)}>0$ is the remaining time after the arrival of the immigrant event. Concretely, we have

$$
\boldsymbol{S}_{j}(u):=\left[\begin{array}{c}
S_{1 \leftarrow j}(u)  \tag{3.6}\\
\vdots \\
S_{d \leftarrow j}(u)
\end{array}\right],
$$

where each entry $S_{i \leftarrow j}(u)$ records the number of events generated into component $i \in[d]$ in the cluster $\boldsymbol{C}_{j}$, with as oldest ancestor the immigrant event in component $j$ that generated the cluster. To avoid double counts, we let the immigrant itself be included in the cluster (only) when $i=j$.

If we let $u$ tend to infinity, the entries of the random vector $\boldsymbol{S}_{j}(u)$ ultimately count the total number of events within the cluster $\boldsymbol{C}_{j}$ generated into each component $i \in[d]$. Observe that $u \mapsto \boldsymbol{S}_{j}(u)$ is increasing componentwise and $\sup _{u \in \mathbb{R}_{+}}\left\|\boldsymbol{S}_{j}(u)\right\|_{\mathbb{R}^{d}}<\infty$ with probability 1 due to Assumption 3.1. Hence, we can define a random vector that counts the total number of events in all components, or simply cluster size, by setting $\boldsymbol{S}_{j}:=\lim _{u \rightarrow \infty} \boldsymbol{S}_{j}(u)$, where convergence is understood in the a.s. sense.

One can interpret these clusters in terms of $d$-type Galton-Watson processes, where the total progeny equals the sum of all generations of offspring that descend from one individual ([51]). Suppose the Galton-Watson process starts with an individual of type $j \in[d]$, and let $\boldsymbol{S}_{j}^{(k)}$ denote the $k$-th generation of descendants. Then one can write

$$
\begin{equation*}
\boldsymbol{S}_{j}=\sum_{k=0}^{\infty} \boldsymbol{S}_{j}^{(k)} \tag{3.7}
\end{equation*}
$$

where $\boldsymbol{S}_{j}^{(0)}=\boldsymbol{e}_{j}$, the unit vector (i.e., with $j$-th entry equal to 1 , and other entries equal to 0). In [51], the total progeny, i.e. cluster size, of such a process is analyzed in the one-dimensional setting and shown to have a so-called Borel distribution. For higher-dimensional Hawkes processes, by using [15, Theorem 1.2], it is, in principle, also possible to derive a representation of the multivariate cluster size distribution. However, the resulting expression is neither explicit nor workable for the goal at hand due to (highly) convoluted sums that arise in the derivation. More specifically, the multiplicity of the different possible sample paths to generate a certain number of events in each component yields a complex combinatorial problem.

We conclude this section by stating two convergence results that will be needed later in this paper. Under the stability condition, we have that the Hawkes process $\boldsymbol{N}(\cdot)$ satisfies the following strong law of large numbers, as shown in [6]: as $t \rightarrow \infty$, we have

$$
\begin{equation*}
\frac{\boldsymbol{N}(t)}{t} \rightarrow(\boldsymbol{I}-\boldsymbol{H})^{-1} \overline{\boldsymbol{\lambda}} \tag{3.8}
\end{equation*}
$$

a.s., where $\overline{\boldsymbol{\lambda}}=\left(\bar{\lambda}_{1}, \ldots, \bar{\lambda}_{d}\right)^{\top}$. This result naturally extends to the corresponding compound Hawkes process $\boldsymbol{Z}(\cdot)$ (see [44, Theorem 1]): as $t \rightarrow \infty$, a.s.,

$$
\begin{equation*}
\frac{\boldsymbol{Z}(t)}{t} \rightarrow \mathbb{E}[\boldsymbol{U}](\boldsymbol{I}-\boldsymbol{H})^{-1} \overline{\boldsymbol{\lambda}} \tag{3.9}
\end{equation*}
$$

### 3.3 Transform Analysis

In this section, we discuss probability generating functions and moment generating functions pertaining to the processes introduced in the previous section, viz. the multivariate Hawkes and compound Hawkes processes. These functions will play a pivotal role when deriving large deviations results later in this paper.

It is directly seen from the definition of $\boldsymbol{Z}(t)$ that, for fixed $t>0$, its moment generating function satisfies the following composite expression in terms of the probability generating function of $\boldsymbol{N}(t)$ :

$$
\begin{equation*}
m_{\boldsymbol{Z}(t)}(\boldsymbol{\theta}) \equiv \mathbb{E}\left[e^{\boldsymbol{\theta}^{\top} \boldsymbol{Z}(t)}\right]=\mathbb{E}\left[\prod_{l=1}^{d}\left(m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})\right)^{N_{l}(t)}\right] \tag{3.10}
\end{equation*}
$$

where

$$
m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta}) \equiv \mathbb{E}\left[e^{\boldsymbol{\theta}^{\top} \boldsymbol{U}_{l}}\right]=\mathbb{E}\left[\prod_{i=1}^{d^{\star}} e^{\theta_{i} U_{i l}}\right] .
$$

For now, we assume $\boldsymbol{\theta} \in \mathbb{R}^{d^{\star}}$ is chosen such that (3.10) exists-we will further discuss the domain of convergence below. We are interested in the limiting cumulant of $\boldsymbol{Z}(t)$ as $t \rightarrow \infty$, that is, we wish to analyze the limiting cumulant

$$
\begin{equation*}
\lim _{t \rightarrow \infty} \frac{1}{t} \log m_{\boldsymbol{Z}(t)}(\boldsymbol{\theta}) \tag{3.11}
\end{equation*}
$$

To derive an expression for (3.11), we use a characterization of the probability generating function of $\boldsymbol{N}(t)$ in terms of the cluster point processes $\boldsymbol{S}_{j}(u)$, obtained in [55, Theorem 1]:

$$
\begin{equation*}
\mathbb{E}\left[\prod_{l=1}^{d} z_{l}^{N_{l}(t)}\right]=\prod_{j=1}^{d} \exp \left(\bar{\lambda}_{j} \int_{0}^{t}\left(\mathbb{E}\left[\prod_{l=1}^{d} z_{l}^{S_{l \leftarrow j}(u)}\right]-1\right) \mathrm{d} u\right) \tag{3.12}
\end{equation*}
$$

## Chapter 3 Multivariate Compound Hawkes Processes

where, for each $j \in[d]$, the probability generating function of $\boldsymbol{S}_{j}(u)$ appearing on the right-hand side of (3.12) satisfies the fixed-point representation

$$
\begin{equation*}
f_{j}(\boldsymbol{z}, u):=\mathbb{E}\left[\prod_{l=1}^{d} z_{l}^{S_{l \leftarrow j}(u)}\right]=z_{j} \mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} \int_{0}^{u} g_{m j}(v)\left(f_{m}(\boldsymbol{z}, u-v)-1\right) \mathrm{d} v\right)\right] ; \tag{3.13}
\end{equation*}
$$

see [55, Theorem 2]. In order to exploit this characterization to establish our large deviations result in the multivariate compound Hawkes setting, we need to analyze the domain of $\boldsymbol{z}=\left(z_{1}, \ldots, z_{d}\right)^{\top}$ for which Eqns. (3.12) and (3.13) hold, i.e., where the probability generating functions of $\boldsymbol{S}_{j}(u)$ exist. More precisely, since we focus on the regime $t \rightarrow \infty$ in (3.11), we actually need to consider the probability generating function of the total cluster size $\boldsymbol{S}_{j}$ instead of $\boldsymbol{S}_{j}(u)$, which in the sequel is denoted by

$$
\begin{equation*}
f_{j}(\boldsymbol{z}):=\lim _{u \rightarrow \infty} f_{j}(\boldsymbol{z}, u)=\mathbb{E}\left[\prod_{l=1}^{d} z_{l}^{S_{l \leftarrow j}}\right] \tag{3.14}
\end{equation*}
$$

and analyze its domain. Let $\boldsymbol{f}: \mathbb{R}_{+}^{d} \rightarrow \overline{\mathbb{R}}^{d}$ be given by $\boldsymbol{f}(\boldsymbol{z})=\left(f_{1}(\boldsymbol{z}), \ldots, f_{d}(\boldsymbol{z})\right)^{\top}$ and denote its effective domain by $\mathscr{D}_{f}:=\left\{\boldsymbol{z} \in \mathbb{R}_{+}^{d}:\|\boldsymbol{f}(\boldsymbol{z})\|_{\mathbb{R}^{d}}<\infty\right\}$. For the set $\mathscr{D}_{f}$, denote the interior by $\mathscr{D}_{f}^{\circ}$ and the boundary by $\partial \mathscr{D}_{f}$.

Observe that the right-hand side of Eqn. (3.13) is expressed in terms of the moment generating function of the random vector $\boldsymbol{B}_{j}$. We assume the following to hold throughout the paper.

Assumption 3.2. For some $\boldsymbol{\vartheta} \in \mathbb{R}_{+}^{d}$, assume that for all $j \in[d]$,

$$
\begin{equation*}
m_{\boldsymbol{B}_{j}}(\boldsymbol{\vartheta})=\mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} \vartheta_{m}\right)\right]<\infty . \tag{3.15}
\end{equation*}
$$

The following result gives an implicit characterization of $\boldsymbol{f}(\cdot)$ and its domain $\mathscr{D}_{f}$ in terms of a fixed-point representation.

Proposition 3.1. The vector-valued function $\boldsymbol{f}(\boldsymbol{z})$ is the unique increasing function that satisfies

$$
\begin{equation*}
f_{j}(\boldsymbol{z})=z_{j} \mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(f_{m}(\boldsymbol{z})-1\right)\right)\right] \tag{3.16}
\end{equation*}
$$

for $\boldsymbol{z} \in \mathbb{R}_{+}^{d}$ such that $\boldsymbol{z} \leqslant \widehat{\boldsymbol{z}}_{\boldsymbol{r}} \equiv \widehat{\boldsymbol{z}}$, where, for an arbitrarily given positive vector $\boldsymbol{r} \in \mathbb{R}_{+}^{d}, \widehat{\boldsymbol{z}}=\left(\widehat{z}_{1}, \ldots, \widehat{z}_{d}\right)^{\top}$ is given for each $j \in[d]$ by

$$
\begin{equation*}
\widehat{z}_{j}=r_{j}\left(\sum_{k=1}^{d} r_{k} \mathbb{E}\left[B_{k j} c_{k j} \exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(\widehat{x}_{m}-1\right)\right)\right]\right)^{-1} \tag{3.17}
\end{equation*}
$$

and $\widehat{\boldsymbol{x}}=\left(\widehat{x}_{1}, \ldots, \widehat{x}_{d}\right)^{\top}$ solves

$$
\begin{align*}
& x_{j}\left(\sum_{k=1}^{d} r_{k} \mathbb{E}\left[B_{k j} c_{k j} \exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(x_{m}-1\right)\right)\right]\right) \\
& =r_{j} \mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(x_{m}-1\right)\right)\right] \tag{3.18}
\end{align*}
$$

Proof. The proof consists of three parts: (i) identifying the limit of $\boldsymbol{f}(\boldsymbol{z}, u)$ as $u \rightarrow \infty$; (ii) implicit characterization of the domain $\mathscr{D}_{f}$; (iii) proving uniqueness of $\boldsymbol{f}(\cdot)$ that satisfies (3.16).

- Proof of (i). We show that for $\boldsymbol{z} \in \mathscr{D}_{f}$, we have that $f_{j}(\boldsymbol{z}, u) \rightarrow f_{j}(\boldsymbol{z})$ for all $j \in[d]$. At this point, we do not yet know the precise domain $\mathscr{D}_{f}$, but we do know it is a convex subset of $\mathbb{R}_{+}^{d}$ and we implicitly derive it later in the proof.

When $\boldsymbol{z}=\mathbf{1}$, we have $\boldsymbol{f}(\mathbf{1}, u) \equiv \mathbf{1} \equiv \boldsymbol{f}(\mathbf{1})$ and convergence follows trivially. Observe that when $\mathbf{0} \leqslant \boldsymbol{z}<\mathbf{1}, f_{j}(\boldsymbol{z}, u)$ is decreasing in $u$ and $0 \leqslant f_{j}(\boldsymbol{z}, u)<1$. Hence, $f_{j}(\boldsymbol{z}, u)$ converges by the monotone convergence theorem to a finite limit as $u \rightarrow \infty$, satisfying the limit of (3.13). When $\boldsymbol{z}>\mathbf{1}, f_{j}(\boldsymbol{z}, u)$ is increasing in $u$ and either diverges to $\infty$ or converges to a finite limit, satisfying the limit of (3.13). In the intermediate case, where for some $k, m \in[d]$ one has $z_{k} \leqslant 1$ and $z_{m}>1$, we proceed as follows. Recall that for each $j \in[d]$, the map $u \mapsto \boldsymbol{S}_{j}(u)$ is a.s. increasing in all components. We obtain the following upper bound

$$
\begin{aligned}
\limsup _{u \rightarrow \infty} f_{j}(\boldsymbol{z}, u) & =\limsup _{u \rightarrow \infty} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i \leftarrow j}(u)}\right] \\
& =\limsup _{u \rightarrow \infty} \mathbb{E}\left[\prod_{k: z_{k} \leqslant 1} z_{k}^{S_{k \leftarrow j}(u)} \prod_{m: z_{m}>1} z_{m}^{S_{m \leftarrow j}(u)}\right] \\
& \leqslant \limsup _{u \rightarrow \infty} \mathbb{E}\left[\prod_{k: z_{k} \leqslant 1} z_{k}^{S_{k \leftarrow j}(u)} \prod_{m: z_{m}>1} z_{m}^{S_{m \leftarrow j}}\right] \\
& \stackrel{(*)}{=} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i \leftarrow j}}\right]=f_{j}(\boldsymbol{z}),
\end{aligned}
$$

if the limit is finite, where in $(*)$ we have used the monotone convergence on the product over $k$, as this product is decreasing. Similarly, we obtain the lower bound

$$
\begin{aligned}
\liminf _{u \rightarrow \infty} f_{j}(\boldsymbol{z}, u) & \geqslant \liminf _{u \rightarrow \infty} \mathbb{E}\left[\prod_{k: z_{k} \leqslant 1} z_{k}^{S_{k \leftarrow j}} \prod_{m: z_{m}>1} z_{m}^{S_{m \leftarrow j}(u)}\right] \\
& \stackrel{(*)}{=} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{S_{i \leftarrow j}}\right]=f_{j}(\boldsymbol{z}),
\end{aligned}
$$

which implies that the liminf and $\lim \sup$ coincide, and so $f_{j}(\boldsymbol{z}, u) \rightarrow f_{j}(\boldsymbol{z})$ for all $j \in$ [d]. Provided all components converge, we have convergence of the vector $\boldsymbol{f}(\boldsymbol{z}, u) \rightarrow$ $\boldsymbol{f}(\boldsymbol{z})$. Hence, from (3.13) we obtain for each $j \in[d]$ and with $\boldsymbol{z} \in \mathscr{D}_{f}$ that

$$
\begin{equation*}
f_{j}(\boldsymbol{z})=z_{j} \mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(f_{m}(\boldsymbol{z})-1\right)\right)\right] \tag{3.19}
\end{equation*}
$$

yielding a vector-valued fixed-point representation for $\boldsymbol{f}(\boldsymbol{z})$.

- Proof of (ii). We now implicitly characterize the domain $\mathscr{D}_{f}$. To that end, consider the function $\boldsymbol{G}: \mathbb{R}^{d} \times \mathbb{R}^{d} \rightarrow \mathbb{R}^{d}$, where each entry $G_{j}: \mathbb{R}^{d} \times \mathbb{R}^{d} \rightarrow \mathbb{R}$, with $j \in[d]$, is given by

$$
\begin{equation*}
G_{j}(\boldsymbol{z}, \boldsymbol{x})=z_{j} \mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(x_{m}-1\right)\right)\right]-x_{j} . \tag{3.20}
\end{equation*}
$$

Note that $\boldsymbol{G}(\cdot)$ is continuously differentiable for all $\boldsymbol{z} \in \mathbb{R}^{d}$ and $\boldsymbol{x} \in \mathbb{R}^{d}$ for which the respective moment generating functions of $\boldsymbol{B}_{j}$ are defined. To obtain a characterization of $\mathscr{D}_{f}$, we need to find the set of $\boldsymbol{z}$ on the boundary of $\mathscr{D}_{f}$, for which $\boldsymbol{f}(\boldsymbol{z})$ exists and satisfies (3.16). Since Eqn. (3.16) is analogous to solving $\boldsymbol{G}(\boldsymbol{z}, \boldsymbol{f}(\boldsymbol{z}))=\mathbf{0}$, we can find the domain of $\boldsymbol{f}(\cdot)$ by investigating the set $\boldsymbol{G}^{-1}(\mathbf{0})=\{(\boldsymbol{z}, \boldsymbol{x}): \boldsymbol{G}(\boldsymbol{z}, \boldsymbol{x})=$ $\mathbf{0}\}$. Since the preimage $\boldsymbol{G}^{-1}(\mathbf{0})$ can be a complicated set, we resort to the preimage theorem, a variation of the implicit function theorem also known as the regular level set theorem, see e.g., [71, Theorem 9.9], which states two results. First, the preimage has codimension equal to the dimension of the image, and second, the tangent space at a point of the preimage coincides with the kernel of the Jacobian at that point, provided that the Jacobian is of full-rank.

We proceed by providing a specification of the preimage theorem in our setting. The first part of the preimage theorem states that $\boldsymbol{G}^{-1}(\mathbf{0})$ is a $d$-dimensional space. Note that in the univariate $(d=1)$ setting, $\boldsymbol{G}^{-1}(0)$ would be a curve embedded in $\mathbb{R} \times \mathbb{R}$, and the tangent space would be a line. In our multivariate $(d>1)$ setting, $\boldsymbol{G}^{-1}(\mathbf{0})$ is a $d$-dimensional manifold embedded in $\mathbb{R}^{d} \times \mathbb{R}^{d}$, and the tangent space is again $d$-dimensional. The second part concerns tangent spaces, defined as follows: for any $(\boldsymbol{z}, \boldsymbol{x}) \in \boldsymbol{G}^{-1}(\mathbf{0})$, the tangent space $T_{\boldsymbol{z}, \boldsymbol{x}}\left(\boldsymbol{G}^{-1}(\mathbf{0})\right)$ consists of the set of vectors $(\boldsymbol{q}, \boldsymbol{r}) \in \mathbb{R}_{+}^{d} \times \mathbb{R}_{+}^{d}$ for which there exists a curve $\gamma \subseteq \boldsymbol{G}^{-1}(\mathbf{0})$ with $\gamma(0)=(\boldsymbol{z}, \boldsymbol{x})$ and $\gamma^{\prime}(0)=(\boldsymbol{q}, \boldsymbol{r})$. The second part of the preimage theorem then states

$$
\begin{equation*}
\operatorname{Ker}\left(\boldsymbol{J}_{\boldsymbol{G}}(\boldsymbol{z}, \boldsymbol{x})\right)=T_{\boldsymbol{z}, \boldsymbol{x}}\left(\boldsymbol{G}^{-1}(\mathbf{0})\right) \tag{3.21}
\end{equation*}
$$

with $\boldsymbol{J}_{\boldsymbol{G}}(\boldsymbol{z}, \boldsymbol{x}) \in \mathbb{R}^{d \times 2 d}$ denoting the full Jacobian of $\boldsymbol{G}$ evaluated at $(\boldsymbol{z}, \boldsymbol{x})$. We compute the $d \times d$-dimensional Jacobian matrices of partial derivatives of $\boldsymbol{G}$ w.r.t. $\boldsymbol{z}$
and $\boldsymbol{x}$ separately by

$$
\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{z}}:=\left[\frac{\partial G_{j}}{\partial z_{k}}(\boldsymbol{z}, \boldsymbol{x})\right]_{j, k \in[d]}=\left[\mathbf{1}_{\{j=k\}} \mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(x_{m}-1\right)\right)\right]\right]_{j, k \in[d]}
$$

and
$\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}:=\left[\frac{\partial G_{j}}{\partial x_{k}}(\boldsymbol{z}, \boldsymbol{x})\right]_{j, k \in[d]}=\left[z_{j} \mathbb{E}\left[B_{k j} c_{k j} \exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(x_{m}-1\right)\right)\right]-\mathbf{1}_{\{j=k\}}\right]_{j, k \in[d]}$,
such that $\boldsymbol{J}_{\boldsymbol{G}}=\left(\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{z}} \mid \boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}\right)$. To utilize Eqn. (3.21), we need to look for vectors $(\boldsymbol{q}, \boldsymbol{r}) \in \mathbb{R}_{+}^{d} \times \mathbb{R}_{+}^{d}$ such that $\boldsymbol{J}_{\boldsymbol{G}} \cdot(\boldsymbol{q}, \boldsymbol{r})=\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{z}} \cdot \boldsymbol{q}+\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}} \cdot \boldsymbol{r}=\mathbf{0}$, where we denote $\boldsymbol{J}_{\boldsymbol{G}} \equiv \boldsymbol{J}_{\boldsymbol{G}}(\boldsymbol{z}, \boldsymbol{x})$ for brevity. Note that $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{z}}$ is a diagonal matrix with positive entries, such that $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{z}} \cdot \boldsymbol{q}=\mathbf{0}$ only if $\boldsymbol{q}=\mathbf{0}$; so we can focus on $\boldsymbol{r}$. Observe that the set of points where the determinant of the Jacobian $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}$ has the same sign is a connected set, due to strict convexity in each entry of the functions $G_{j}(\boldsymbol{z}, \boldsymbol{x})$, with $j \in[d]$, and by continuity of the determinant and partial derivatives. For a fixed vector $\boldsymbol{r} \in \mathbb{R}_{+}^{d}$, we can establish systems of equations for $\boldsymbol{z}$ and $\boldsymbol{x}$ such that $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}} \cdot \boldsymbol{r}=\mathbf{0}$. As will become clear later in the proof, convexity will play a crucial role in determining uniqueness of these solutions.

With the objective of substantiating the claim in (3.21), we compute the solution to the systems of equations $\boldsymbol{G}(\boldsymbol{z}, \boldsymbol{x})=\mathbf{0}$ and $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}} \cdot \boldsymbol{r}=\mathbf{0}$ by using the expression for $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}$ and solving for $\boldsymbol{z}$ and $\boldsymbol{x}$. This yields the solutions $\widehat{\boldsymbol{z}}=\left(\widehat{z}_{1}, \ldots, \widehat{z}_{d}\right)^{\top}$ and $\widehat{\boldsymbol{x}}=\left(\widehat{x}_{1}, \ldots \widehat{x}_{d}\right)^{\top}$ given in Eqns. (3.17) and (3.18), with $(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{x}}) \equiv\left(\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}\right)$ parameterized by vectors $\boldsymbol{r} \in \mathbb{R}_{+}^{d}$, such that $\boldsymbol{G}(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{x}})=\mathbf{0}$ and $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{x}}) \cdot \boldsymbol{r}=\mathbf{0}$. Moreover, for any given $\boldsymbol{r} \in \mathbb{R}_{+}^{d}$, we show that the associated pair $\left(\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}\right)$ is unique. The condition $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}\left(\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}\right) \cdot \boldsymbol{r}=\mathbf{0}$ stated for each row yields $\nabla_{\boldsymbol{x}} G_{j}\left(\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}\right) \cdot \boldsymbol{r}=0$ for all $j \in[d]$, with $\nabla_{x} G_{j}(\cdot)$ the $j$-th row of the Jacobian. Note that each $G_{j}(\cdot)$ only depends on $z_{j}$ and $\boldsymbol{x}$. Due to strict convexity of $G_{j}(\boldsymbol{z}, \boldsymbol{x})$ in each entry, we have that the sub-level set $G_{j}^{-1}(\leqslant 0):=\left\{(\boldsymbol{z}, \boldsymbol{x}): G_{j}(\boldsymbol{z}, \boldsymbol{x}) \leqslant 0\right\}$ is a strictly convex set, and the level set $G_{j}^{-1}(0)$ is the boundary of $G_{j}^{-1}(\leqslant 0)$. This implies that

$$
\boldsymbol{G}^{-1}(\mathbf{0})=\left\{(\boldsymbol{z}, \boldsymbol{x}): G_{j}(\boldsymbol{z}, \boldsymbol{x})=0, \forall j \in[d]\right\}=\bigcap_{j=1}^{d} G_{j}^{-1}(0),
$$

is the boundary of a strictly convex set, namely $\boldsymbol{G}^{-1}(\leqslant \mathbf{0})$, as the latter is the intersection of strictly convex sets. Since $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}\left(\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}\right) \cdot \boldsymbol{r}=\mathbf{0}$ means $\boldsymbol{r} \in T_{\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}}\left(\boldsymbol{G}^{-1}(\mathbf{0})\right)$ by Eqn. (3.21), and since $\boldsymbol{G}^{-1}(\mathbf{0})$ is the boundary of a strictly convex set, we have that $\boldsymbol{r}$ uniquely determines the point $\left(\widehat{\boldsymbol{z}}_{\boldsymbol{r}}, \widehat{\boldsymbol{x}}_{\boldsymbol{r}}\right)$.

## Chapter 3 Multivariate Compound Hawkes Processes

The next step amounts to relating what we found so far to the domain $\mathscr{D}_{f}$. A given value of $\boldsymbol{z} \in \mathbb{R}_{+}^{d}$ determines whether one can find $\boldsymbol{x} \in \mathbb{R}_{+}^{d}$ for which $\boldsymbol{G}(\boldsymbol{z}, \boldsymbol{x})=\mathbf{0}$, such that $(\boldsymbol{z}, \boldsymbol{x}) \in \boldsymbol{G}^{-1}(\mathbf{0})$. Observe that the set $R_{z}:=\left\{\boldsymbol{z} \in \mathbb{R}_{+}^{d}: \boldsymbol{z}=\widehat{\boldsymbol{z}}_{r}, \boldsymbol{r} \in \mathbb{R}_{+}^{d}\right\}$ divides the positive quadrant $\mathbb{R}_{+}^{d}$ into two disjoint sets. The first set is the inner (convex) region, defined as the set of $\boldsymbol{z} \in \mathbb{R}_{+}^{d}$ enclosed by the origin, the $\boldsymbol{z}$ axes and the set $R_{z}$, with $R_{z}$ included; denote this set by $\mathcal{Z}$. The second set is the outer region, denoted by $\mathcal{Z}^{c}$, and it is the complement of $\mathcal{Z}$, such that $\mathcal{Z} \cup \mathcal{Z}^{c}=\mathbb{R}_{+}^{d}$. Note that when $\boldsymbol{z} \in \mathcal{Z}^{c}$, then $\boldsymbol{G}(\boldsymbol{z}, \boldsymbol{x}) \neq \mathbf{0}$ for any $\boldsymbol{x} \in \mathbb{R}_{+}^{d}$, since $\mathcal{Z}^{c} \times \mathbb{R}_{+}^{d}$ does not intersect $\boldsymbol{G}^{-1}(\mathbf{0})$. This yields $\boldsymbol{G}^{-1}(\mathbf{0})=\{(\boldsymbol{z}, \boldsymbol{x}): \boldsymbol{z} \in \mathcal{Z}, \boldsymbol{G}(\boldsymbol{z}, \boldsymbol{x})=\mathbf{0}\}$, and using that $\boldsymbol{G}(\boldsymbol{z}, \boldsymbol{f}(\boldsymbol{z}))=\mathbf{0}$ for all $\boldsymbol{z} \in \mathscr{D}_{f}$, we find $\mathscr{D}_{f} \subseteq \mathcal{Z}$, which proves (ii). However, for $\boldsymbol{z} \in \mathcal{Z}$, we may have multiple $\boldsymbol{x} \in \mathbb{R}_{+}^{d}$ such that $\boldsymbol{G}(\boldsymbol{z}, \boldsymbol{x})=\mathbf{0}$, so we investigate this further.

- Proof of (iii). We are left with proving uniqueness of $\boldsymbol{f}(\cdot)$. We prove this by considering points $(\boldsymbol{z}, \boldsymbol{x}) \in \boldsymbol{G}^{-1}(\mathbf{0})$ and relating them to $\boldsymbol{f}(\cdot)$. From the preimage theorem, we know that $\boldsymbol{G}^{-1}(\mathbf{0})$ is $d$-dimensional, so we need only $d$ parameters to describe this set. We can use the implicit function theorem to describe the $\boldsymbol{x}$ coordinate of $(\boldsymbol{z}, \boldsymbol{x}) \in \boldsymbol{G}^{-1}(\mathbf{0})$ in terms of an implicit function of $\boldsymbol{z}$. We consider a particular point in this set and then show how the argument extends to other points.

Consider the point $(\boldsymbol{z}, \boldsymbol{x})=(\mathbf{1}, \mathbf{1}) \in \boldsymbol{G}^{-1}(\mathbf{0})$ since it satisfies $\boldsymbol{G}(\mathbf{1}, \mathbf{1})=\mathbf{0}$, where we use Assumption 3.2 to ensure existence of the moment generating functions of $\boldsymbol{B}_{j}$ around this point. Evaluated at the point $(\mathbf{1}, \mathbf{1})$, the Jacobian of $\boldsymbol{G}$ with respect to $\boldsymbol{x}$ is given by

$$
\begin{equation*}
J_{G, x}(\mathbf{1}, \mathbf{1})=H^{\top}-I, \tag{3.22}
\end{equation*}
$$

which is invertible due to Assumption 3.1. Then by the implicit function theorem, there exist open sets $V, W \subseteq \mathbb{R}_{+}^{d}$ both containing $\mathbf{1}$, and a unique continuously differentiable function $\tilde{\boldsymbol{f}}: V \rightarrow W$ such that $\tilde{\boldsymbol{f}}(\mathbf{1})=\mathbf{1}$ and $\boldsymbol{G}(\boldsymbol{z}, \tilde{\boldsymbol{f}}(\boldsymbol{z}))=\mathbf{0}$ for all $\boldsymbol{z} \in V$. Note that this implies $V \subseteq \mathcal{Z}$ and that $\tilde{\boldsymbol{f}}(\cdot)$ satisfies the fixed equation in (3.16). Moreover, since $\tilde{\boldsymbol{f}}(\cdot)$ is unique, we have $\tilde{\boldsymbol{f}}(\cdot)=\boldsymbol{f}(\cdot)$ on $V$ and $V \subseteq \mathscr{D}_{f}^{\circ}$, provided that $\tilde{\boldsymbol{f}}(\cdot)$ is increasing in all entries, since we know by definition that $\boldsymbol{f}(\cdot)$ is increasing in all entries.

The point $(\mathbf{1}, \mathbf{1})$ is not particularly special; if we take another point $\left(\boldsymbol{z}_{0}, \boldsymbol{x}_{0}\right) \in$ $\boldsymbol{G}^{-1}(\mathbf{0})$, we find that the Jacobian $\boldsymbol{J}_{\boldsymbol{G}, \boldsymbol{x}}\left(\boldsymbol{z}_{0}, \boldsymbol{x}_{0}\right)$ is invertible provided $\left(\boldsymbol{z}_{0}, \boldsymbol{x}_{0}\right) \neq(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{x}})$. We can then apply the implicit function theorem to obtain open sets $\boldsymbol{z}_{0} \in V_{0} \subseteq \mathbb{R}_{+}^{d}$, $\boldsymbol{x}_{0} \in W_{0} \subseteq \mathbb{R}_{+}^{d}$ and a unique map $\tilde{\boldsymbol{f}}_{0}: V_{0} \rightarrow W_{0}$ that satisfies $\boldsymbol{G}\left(\boldsymbol{z}, \tilde{\boldsymbol{f}}_{0}(\boldsymbol{z})\right)=\mathbf{0}$ for all $\boldsymbol{z} \in V_{0}$, again with $V_{0} \subseteq \mathcal{Z}$. As before, we obtain $\tilde{\boldsymbol{f}}(\cdot)=\boldsymbol{f}(\cdot)$ on $V_{0}$ and $V_{0} \subseteq \mathscr{D}_{f}^{\circ}$, due to uniqueness of $\tilde{\boldsymbol{f}}_{0}$, provided $\tilde{\boldsymbol{f}}_{0}(\cdot)$ is increasing in all entries. Since we can do this for arbitrary points, we obtain uniqueness of $\boldsymbol{f}(\cdot)$ on all of $\mathcal{Z}$, such that $\mathcal{Z} \subseteq \mathscr{D}_{f}^{\circ}$. Finally, for any pair of solutions ( $\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{x}})$ to Eqns. (3.17) and (3.18), we have by monotonicity of $\boldsymbol{f}(\cdot)$ that $\lim _{\boldsymbol{z} \nearrow \bar{z}} \boldsymbol{f}(\boldsymbol{z})=\widehat{\boldsymbol{x}}$, which yields the characterization $\mathcal{Z}=\mathscr{D}_{f}$.

We remark that taking $d=1$ in Proposition 3.1 yields agreement with the results obtained in [54], where the condition that we impose that the implicit function is increasing, is equivalent to the condition in [54] where they take the minimal solution of the equation $G(z, x)=0$ for fixed $z<\widehat{z}$. Next, we focus on the limiting cumulant of $\boldsymbol{Z}(\cdot)$ given in (3.11). Note that the moment generating function of $\boldsymbol{Z}(t)$ in (3.10), and hence also in (3.11), is expressed in terms of the moment generating functions of the random vectors $\boldsymbol{U}_{1}, \ldots, \boldsymbol{U}_{d}$. Denote $\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta})=\left(m_{\boldsymbol{U}_{1}}(\boldsymbol{\theta}), \ldots, m_{\boldsymbol{U}_{d}}(\boldsymbol{\theta})\right)^{\top}$ as the vector of moment generating functions of $\boldsymbol{U}_{1}, \ldots, \boldsymbol{U}_{d}$. We impose the following condition, assumed to hold throughout the paper.
Assumption 3.3. Assume that for any $\widehat{\boldsymbol{z}}$ in (3.17), there exists a vector $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{d^{\star}}$ such that

$$
\begin{equation*}
\boldsymbol{m}_{U}(\widehat{\boldsymbol{\theta}})=\widehat{\boldsymbol{z}} \tag{3.23}
\end{equation*}
$$

Define the function $\Lambda: \mathbb{R}^{d^{\star}} \rightarrow \mathbb{R}$ by

$$
\begin{equation*}
\Lambda(\boldsymbol{\theta})=\sum_{j=1}^{d} \bar{\lambda}_{j}\left(f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta})\right)-1\right) \tag{3.24}
\end{equation*}
$$

Also define the domain of convergence $\mathscr{D}_{\Lambda}:=\left\{\boldsymbol{\theta} \in \mathbb{R}^{d^{\star}}: \Lambda(\boldsymbol{\theta})<\infty\right\}$, denote its interior by $\mathscr{D}_{\Lambda}^{\circ}$ and denote by $\partial \mathscr{D}_{\Lambda}$ its boundary. We now characterize the limiting cumulant of $\boldsymbol{Z}(\cdot)$ in (3.11).

Lemma 3.1. We have $\mathbf{0} \in \mathscr{D}_{\Lambda}^{\circ}$ and for $\boldsymbol{\theta} \in \mathbb{R}^{d^{\star}}$ such that $\boldsymbol{\theta} \leqslant \widehat{\boldsymbol{\theta}}$, where $\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})=\widehat{\boldsymbol{z}}$ and with $\widehat{\boldsymbol{z}}$ the solution to (3.17), we have

$$
\begin{equation*}
\lim _{t \rightarrow \infty} \frac{1}{t} \log m_{\boldsymbol{Z}(t)}(\boldsymbol{\theta})=\Lambda(\boldsymbol{\theta}) . \tag{3.25}
\end{equation*}
$$

Proof. We showed that $\mathbf{1} \in \mathscr{D}_{f}^{\circ}$ in the proof of Proposition 3.1. We then immediately have by Assumption 3.3 that the vector of moment generating functions $\boldsymbol{m}_{\boldsymbol{U}}(\cdot)$ is defined in a neighborhood of the origin. Hence, by taking $\boldsymbol{\theta}=\mathbf{0}$, we have $\boldsymbol{m}_{\boldsymbol{U}}(\mathbf{0})=\mathbf{1}$, which implies $\mathbf{0} \in \mathscr{D}_{\Lambda}^{\circ}$.

We now prove that Eqn. (3.25) holds. Combining Eqns. (3.10) and (3.12), we obtain

$$
m_{\boldsymbol{Z}(t)}(\boldsymbol{\theta})=\prod_{j=1}^{d} \exp \left(\bar{\lambda}_{j} \int_{0}^{t}\left(\mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}(u)}\right]-1\right) \mathrm{d} u\right),
$$

or, equivalently,

$$
\begin{equation*}
\frac{1}{t} \log m_{\boldsymbol{Z}(t)}(\boldsymbol{\theta})=\sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{1}\left(\mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}(v t)}\right]-1\right) \mathrm{d} v, \tag{3.26}
\end{equation*}
$$

performing an elementary change of variables. We want to take limits $t \rightarrow \infty$ in Eqn. (3.26); to that end we first focus on the expectation in the integrand. By mimicking the proof of Proposition 3.1 in combination with the convexity of $m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})$ in each entry, we can apply the monotone convergence theorem on the integrand to find

$$
\lim _{t \rightarrow \infty} \mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}(v t)}\right]=\mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}}\right]
$$

provided $\boldsymbol{\theta} \in \mathscr{D}_{\Lambda}$. Now since we took $\boldsymbol{\theta} \in \mathscr{D}_{\Lambda}$, we have by Assumption 3.3 that there exists $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{d^{\star}}$ such that $\boldsymbol{\theta} \leqslant \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})=\widehat{\boldsymbol{z}}$. Using a similar argument as in the proof of $(i)$ in Proposition 3.1, distinguishing between indices $k, n \in[d]$ for which $\widehat{z}_{k}=m_{\boldsymbol{U}_{k}}(\widehat{\boldsymbol{\theta}}) \leqslant 1$ and $\widehat{z}_{n}=m_{\boldsymbol{U}_{n}}(\widehat{\boldsymbol{\theta}})>1$, we can apply the dominated convergence theorem to obtain

$$
\begin{aligned}
& \lim _{t \rightarrow \infty} \sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{1}\left(\mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}(v t)}\right]-1\right) \mathrm{d} v \\
& =\sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{1} \lim _{t \rightarrow \infty}\left(\mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}(v t)}\right]-1\right) \mathrm{d} v \\
& =\sum_{j=1}^{d} \bar{\lambda}_{j}\left(f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta})\right)-1\right),
\end{aligned}
$$

which proves Eqn. (3.25).
This result can be seen as the multivariate generalization of [54, Theorem 3.1.1], which considers random marks, combined with the approach of [69], which has constant marks but considers the compound process.

### 3.4 Large Deviations

In this section, we show that the multivariate compound Hawkes process satisfies a large deviations principle (LDP). The proof proceeds by establishing the required conditions on the limiting cumulant $\Lambda(\boldsymbol{\theta})$-essential smoothness, most notably - such that the Gärtner-Ellis theorem (see e.g., [19, Theorem 2.3.6]) can be invoked. This section also briefly covers the special case of the LDP of a single component of a multivariate compound Hawkes process.

First recall the definition of an LDP for $\mathbb{R}^{d}$-valued random vectors; see [19, Section 1.2] for more background. Let $\mathcal{B}\left(\mathbb{R}^{d}\right)$ be the Borel $\sigma$-field on $\mathbb{R}^{d}$. Consider a family of random vectors $\left\{\boldsymbol{X}_{\epsilon}\right\}_{\epsilon \in \mathbb{R}_{+}}$taking values in $\left(\mathbb{R}^{d}, \mathcal{B}\left(\mathbb{R}^{d}\right)\right)$. We say that $\left\{\boldsymbol{X}_{\epsilon}\right\}_{\epsilon \in \mathbb{R}_{+}}$satisfies
the LDP with rate function $I(\cdot)$ if $I: \mathbb{R}^{d} \rightarrow[0, \infty]$ is a lower semicontinuous mapping, and if for every Borel set $A \in \mathcal{B}\left(\mathbb{R}^{d}\right)$,

$$
\begin{equation*}
-\inf _{\boldsymbol{x} \in A^{\circ}} I(\boldsymbol{x}) \leqslant \liminf _{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(\boldsymbol{X}_{\epsilon} \in A\right) \leqslant \limsup _{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(\boldsymbol{X}_{\epsilon} \in A\right) \leqslant-\inf _{\boldsymbol{x} \in \bar{A}} I(\boldsymbol{x}), \tag{3.27}
\end{equation*}
$$

where $A^{\circ}$ and $\bar{A}$ denote the interior and closure of $A$. Also, recall that $I(\cdot)$ is lower semicontinuous if, for all $\alpha \geqslant 0$, the level sets $\left\{\boldsymbol{x} \in \mathbb{R}^{d}: I(\boldsymbol{x}) \leqslant \alpha\right\}$ are closed; we call $I(\cdot)$ a good rate function if the level sets are compact.

### 3.4.1 LDP of multivariate compound Hawkes processes

In this subsection, we establish that the process $(\boldsymbol{Z}(t) / t)_{t \in \mathbb{R}_{+}}$satisfies an LDP on $\left(\mathbb{R}^{d^{\star}}, \mathcal{B}\left(\mathbb{R}^{d^{\star}}\right)\right)$, as stated in the following theorem.

A distinguishing feature of our proof is that, due to the fact that the distribution of $\boldsymbol{S}_{j}$ is not explicitly known, we prove steepness-a key step in proving essential smoothness-implicitly, i.e., through the fixed-point representation (3.16) that the probability generating functions $f_{j}(\cdot)$ satisfy. In particular, we cannot mimic the proof that was developed in [69] for the univariate case, as that proof heavily rests on explicit expressions for the univariate cluster size distribution.

Our steepness proof, as given below, may be somewhat obscured by the involved notation and complex objects needed due to the fact that we work in a multivariate setting. To remedy this, we have also included in Appendix 3.A a separate proof for the univariate setting that is based on the precise same reasoning as the one below, but is considerably more transparent.

Theorem 3.1. The process $(\boldsymbol{Z}(t) / t)_{t \in \mathbb{R}_{+}}$satisfies on $\left(\mathbb{R}^{d^{\star}}, \mathcal{B}\left(\mathbb{R}^{d^{\star}}\right)\right)$ the LDP with good rate function

$$
\begin{equation*}
\Lambda^{*}(\boldsymbol{x})=\sup _{\boldsymbol{\theta} \in \mathbb{R}^{d^{\star}}}\left(\boldsymbol{\theta}^{\top} \boldsymbol{x}-\Lambda(\boldsymbol{\theta})\right) . \tag{3.28}
\end{equation*}
$$

Proof. The proof relies on an application of the Gärtner-Ellis theorem, for which we need to show that the limiting cumulant $\Lambda(\boldsymbol{\theta})$ is an essentially smooth, lower semicontinuous function. For essential smoothness, we need to show that $\mathscr{D}_{\Lambda}^{\circ}$ is nonempty and that $\mathbf{0} \in \mathscr{D}_{\Lambda}^{\circ}$, that $\Lambda(\cdot)$ is differentiable on $\mathscr{D}_{\Lambda}^{\circ}$, and finally that $\Lambda(\cdot)$ is steep; see [19, Section 2.3] for further details.

Lemma 3.1 shows that $\mathscr{D}_{\Lambda}^{\circ}$ is non-empty and $\mathbf{0} \in \mathscr{D}_{\Lambda}^{\circ}$. To show that $\Lambda(\cdot)$ is differentiable on $\mathscr{D}_{\Lambda}^{\circ}$, recall from the proof of Proposition 3.1 that $\boldsymbol{f}(\cdot)$ is continuously differentiable on $\mathscr{D}_{f}^{\circ}$, exploiting Assumptions 3.1 and 3.2. Using this property, in combination with the fact that the moment generating functions $m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})$ are differentiable for $\boldsymbol{\theta} \in \mathscr{D}_{\Lambda}^{\circ}$ by invoking Assumption 3.3, we conclude differentiability of $\Lambda(\cdot)$ on $\mathscr{D}_{\Lambda}^{\circ}$.

Next, we prove that $\Lambda(\cdot)$ is steep, i.e., for any $\overline{\boldsymbol{\theta}} \in \partial \mathscr{D}_{\Lambda}^{\circ}$ and a sequence $\boldsymbol{\theta}_{n} \nearrow \overline{\boldsymbol{\theta}}$ as $n \rightarrow \infty$, we have that $\lim _{n \rightarrow \infty}\left\|\nabla \Lambda\left(\boldsymbol{\theta}_{n}\right)\right\|_{\mathbb{R}^{d^{\star}}}=\infty$. For any $i \in\left[d^{\star}\right]$, we first observe that

$$
\begin{align*}
\frac{\partial}{\partial \theta_{i}} \Lambda(\boldsymbol{\theta}) & =\sum_{j=1}^{d} \bar{\lambda}_{j} \frac{\partial}{\partial \theta_{i}} f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta})\right) \\
& =\sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d}\left(\frac{\partial}{\partial \theta_{i}} m_{\boldsymbol{U}_{k}}(\boldsymbol{\theta})\right) \mathbb{E}\left[S_{k \leftarrow j} \prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}(\boldsymbol{\theta})^{S_{l \leftarrow j}-\mathbf{1}_{\{k=l\}}}\right] . \tag{3.29}
\end{align*}
$$

This identity entails that entries of $\nabla \Lambda(\cdot)$ are given in terms of the partial derivatives of the probability generating function of $\boldsymbol{S}_{j}$, for all $j \in[d]$.

To establish steepness of $\Lambda(\cdot)$, it suffices to show that the partial derivatives of $f_{j}(\cdot)$ diverge on the boundary of $\mathscr{D}_{\Lambda}^{\circ}$. Recall that the input for the probability generating function $\boldsymbol{f}(\cdot)$ in Eqn. (3.24) is the vector $\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta}) \in \mathbb{R}_{+}^{d}$. In the remainder of the proof, we first derive steepness of $\boldsymbol{f}(\cdot)$ at a specific $\boldsymbol{z} \in \mathbb{R}_{+}^{d}$, after which we consider the setting in which $\boldsymbol{f}(\cdot)$ is evaluated in the vector $\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta})$.

Define the matrix $\widehat{\boldsymbol{B}}(\boldsymbol{z})=\left(\widehat{B}_{m j}(\boldsymbol{z})\right)_{m, j \in[d]}$ by

$$
\begin{equation*}
\widehat{B}_{m j}(\boldsymbol{z}):=z_{j} \mathbb{E}\left[B_{m j} c_{m j} \exp \left(\sum_{i=1}^{d} B_{i j} c_{i j}\left(f_{i}(\boldsymbol{z})-1\right)\right)\right] \tag{3.30}
\end{equation*}
$$

Taking the partial derivative of the fixed-point representation (3.16) with respect to $z_{k}$, for $k \in[d]$, yields

$$
\begin{align*}
\frac{\partial f_{j}(\boldsymbol{z})}{\partial z_{k}} & =\mathbb{E}\left[\exp \left(\sum_{m=1}^{d} B_{m j} c_{m j}\left(f_{m}(\boldsymbol{z})-1\right)\right)\right] \mathbf{1}_{\{k=j\}}+\sum_{m=1}^{d} \frac{\partial f_{m}(\boldsymbol{z})}{\partial z_{k}} \widehat{B}_{m j}(\boldsymbol{z})  \tag{3.31}\\
& =\frac{f_{j}(\boldsymbol{z})}{z_{j}} \boldsymbol{1}_{\{k=j\}}+\sum_{m=1}^{d} \frac{\partial f_{m}(\boldsymbol{z})}{\partial z_{k}} \widehat{B}_{m j}(\boldsymbol{z})
\end{align*}
$$

where the second equality is due to the fixed-point representation (3.16) itself. We can write (3.31) compactly in matrix-vector form by considering the Jacobian $\boldsymbol{J}_{\boldsymbol{f}}$ of $\boldsymbol{f}(\cdot)$, which yields

$$
\begin{equation*}
\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{z})=\left(\boldsymbol{I}-\widehat{\boldsymbol{B}}(\boldsymbol{z})^{\top}\right)^{-1} \operatorname{diag}(\boldsymbol{f}(\boldsymbol{z}) / \boldsymbol{z}), \tag{3.32}
\end{equation*}
$$

where the division $\boldsymbol{f}(\boldsymbol{z}) / \boldsymbol{z}$ is to be understood componentwise, provided the inverse exists. We now explore for which values of $\boldsymbol{z}$ the inverse appearing in (3.32) fails to exist, i.e., when the associated determinant equals 0 . Consider an element $\widehat{\boldsymbol{z}}$ on the boundary of $\mathscr{D}_{f}$. Recall from Eqn. (3.17) that this $\widehat{\boldsymbol{z}}$ is parametrized by some
positive vector $\boldsymbol{r} \in \mathbb{R}_{+}^{d}$. Moreover, in this point we have $\boldsymbol{f}(\widehat{\boldsymbol{z}})=\widehat{\boldsymbol{x}}$, with $\widehat{\boldsymbol{x}}$ the solution to (3.18), and hence

$$
\begin{equation*}
\widehat{B}_{m j}(\widehat{\boldsymbol{z}})=\widehat{z}_{j} \mathbb{E}\left[B_{m j} c_{m j} \exp \left(\sum_{i=1}^{d} B_{i j} c_{i j}\left(\widehat{x}_{i}-1\right)\right)\right] \tag{3.33}
\end{equation*}
$$

Combining this with Eqns. (3.17) and (3.18), we obtain

$$
\begin{equation*}
\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{z}}) \cdot r=r \Longleftrightarrow(\boldsymbol{I}-\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{z}})) \cdot r=0 \tag{3.34}
\end{equation*}
$$

implying that $\boldsymbol{r}$ is in the kernel of $\boldsymbol{I}-\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{z}})$. Since $\boldsymbol{r}$ is a positive (non-zero) vector, we obtain that $\boldsymbol{I}-\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{z}})$ is not invertible and so

$$
\begin{equation*}
\operatorname{det}\left(\boldsymbol{I}-\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{z}})^{\top}\right)=0 \tag{3.35}
\end{equation*}
$$

Then by Eqn. (3.32), we find that the directional derivative into the positive quadrant diverges, i.e., for any $\boldsymbol{q} \in \mathbb{R}_{+}^{d}$ and sequence $\left\{\boldsymbol{z}_{n}\right\} \subseteq \mathscr{D}_{f}^{o}$ such that $\boldsymbol{z}_{n} \nearrow \widehat{\boldsymbol{z}}$, we have

$$
\lim _{\boldsymbol{z}_{n} \nearrow \tilde{z}}\left\|\boldsymbol{J}_{\boldsymbol{f}}\left(\boldsymbol{z}_{n}\right) \cdot \boldsymbol{q}\right\|_{\mathbb{R}^{d}}=\infty
$$

since each element of $\operatorname{diag}(\boldsymbol{f}(\widehat{\boldsymbol{z}}) / \widehat{\boldsymbol{z}})=\operatorname{diag}(\widehat{\boldsymbol{x}} / \widehat{\boldsymbol{z}})$ is positive and bounded. This proves that $\boldsymbol{f}(\cdot)$ is steep in each argument.

We now use the above observations to prove steepness of $\Lambda(\cdot)$. By Assumption 3.3, there exists $\widehat{\boldsymbol{\theta}}$ on the boundary of $\mathscr{D}_{\Lambda}$ such that $\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})=\widehat{\boldsymbol{z}}$. With the same argument as above, we find $\operatorname{det}\left(\boldsymbol{I}-\widehat{\boldsymbol{B}}\left(\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})\right)^{\top}\right)=0$, such that $\boldsymbol{I}-\widehat{\boldsymbol{B}}\left(\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})\right)^{\top}$ is not invertible at the boundary of $\mathscr{D}_{\Lambda}$. Hence, for any positive vector $\boldsymbol{q} \in \mathbb{R}^{d}$ and a sequence $\left\{\boldsymbol{\theta}_{n}\right\} \subseteq$ $\mathscr{D}_{\Lambda}^{\circ}$ such that $\boldsymbol{\theta}_{n} \nearrow \widehat{\boldsymbol{\theta}}$, we have

$$
\begin{equation*}
\liminf _{\boldsymbol{\theta}_{n} \nearrow \widehat{\boldsymbol{\theta}}}\left\|\boldsymbol{J}_{\boldsymbol{f}}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}_{n}\right)\right) \cdot \boldsymbol{q}\right\|_{\mathbb{R}^{d}}=\infty \tag{3.36}
\end{equation*}
$$

If we denote the entries of $\boldsymbol{J}_{\boldsymbol{f}}$ by $\boldsymbol{J}_{\boldsymbol{f}}^{(j k)}=\partial f_{j} / \partial z_{k}$, then from Eqn. (3.29), we have

$$
\begin{align*}
& \liminf _{\boldsymbol{\theta}^{\prime} \nearrow \widehat{\boldsymbol{\theta}}}\left\|\nabla \Lambda\left(\boldsymbol{\theta}_{n}\right)\right\|_{\mathbb{R}^{d^{\star}}}=\liminf _{\boldsymbol{\theta}_{n} \nearrow \widehat{\boldsymbol{\theta}}}\left\|\left(\frac{\partial}{\partial \theta_{1}} \Lambda\left(\boldsymbol{\theta}_{n}\right), \ldots, \frac{\partial}{\partial \theta_{d^{\star}}} \Lambda\left(\boldsymbol{\theta}_{n}\right)\right)\right\|_{\mathbb{R}^{d^{\star}}} \\
&=\| \|\left(\sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \lim _{\boldsymbol{\theta}_{n} \nearrow \widehat{\boldsymbol{\theta}}} \inf \frac{\partial}{\partial \theta_{1}} m_{\boldsymbol{U}_{k}}\left(\boldsymbol{\theta}_{n}\right) \mathbb{E}\left[S_{k \leftarrow j} \prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}-\mathbf{1}_{\{k=l\}}}\right]\right. \\
&\left.\quad \ldots, \sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \liminf _{\boldsymbol{\theta}_{n} \nearrow \widehat{\boldsymbol{\theta}}} \frac{\partial}{\partial \theta_{d^{\star}}} m_{\boldsymbol{U}_{k}}\left(\boldsymbol{\theta}_{n}\right) \mathbb{E}\left[S_{k \leftarrow j} \prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}-\mathbf{1}_{\{k=l\}}}\right]\right) \|_{\mathbb{R}^{d^{\star}}} \\
& \geqslant \|\left(\sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \frac{\partial}{\partial \theta_{1}} m_{\boldsymbol{U}_{k}}(\widehat{\boldsymbol{\theta}}) \mathbb{E}\left[\liminf _{\boldsymbol{\theta}_{n} \nearrow \widehat{\boldsymbol{\theta}}} S_{k \leftarrow j} \prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}-\mathbf{1}_{\{k=l\}}}\right]\right. \tag{3.37}
\end{align*}
$$

$$
\begin{aligned}
& \left.\ldots, \sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \frac{\partial}{\partial \theta_{d^{\star}}} m_{\boldsymbol{U}_{k}}(\widehat{\boldsymbol{\theta}}) \mathbb{E}\left[\lim _{\boldsymbol{\theta}_{n} \backslash \widehat{\boldsymbol{\theta}}} \inf _{S_{k \leftarrow j}} \prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}-\mathbf{1}_{\{k=l\}}}\right]\right) \|_{\mathbb{R}^{d^{\star}}} \\
= & \|\left(\sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \frac{\partial}{\partial \theta_{1}} m_{\boldsymbol{U}_{k}}(\widehat{\boldsymbol{\theta}}) \boldsymbol{J}_{\boldsymbol{f}}^{j k}\left(\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})\right), \ldots, \sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \frac{\partial}{\partial \theta_{d^{\star}}} m_{\boldsymbol{U}_{k}}(\widehat{\boldsymbol{\theta}}) \boldsymbol{J}_{\boldsymbol{f}}^{j k}\left(\boldsymbol{m}_{\boldsymbol{U}}(\widehat{\boldsymbol{\theta}})\right) \|_{\mathbb{R}^{d^{\star}}}\right. \\
= & \infty,
\end{aligned}
$$

where the inequality is an application of Fatou's lemma and the last equality is a consequence of (3.36).

We finally prove lower semicontinuity of $\Lambda(\cdot)$. Since we consider a metric space $\mathbb{R}^{d^{\star}}$, it suffices to show lower semicontinuity through sequences. Consider $\boldsymbol{\theta}_{n} \nearrow \boldsymbol{\theta} \in \mathscr{D}_{\Lambda}^{\circ}$ and observe that by Fatou's lemma, we have

$$
\begin{equation*}
\lim _{\boldsymbol{\theta}_{n} \nmid \boldsymbol{\theta}} \inf _{\mathbb{\theta}} \mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}}\right] \geqslant \mathbb{E}\left[\prod_{l=1}^{d} \lim _{\boldsymbol{\theta}_{n} \nmid \boldsymbol{\theta}} \inf _{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l<j}}\right] \tag{3.39}
\end{equation*}
$$

for any $j \in[d]$. Furthermore, it is easily shown that, for any integer $k \in \mathbb{N}$, another application of Fatou's lemma yields

$$
\begin{equation*}
\liminf _{\boldsymbol{\theta}_{n} \nmid \boldsymbol{\theta}} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{k}=\liminf _{\boldsymbol{\theta}_{n} \nmid \boldsymbol{\theta}} \mathbb{E}\left[\exp \left(\boldsymbol{\theta}_{n}^{\top} \boldsymbol{U}_{l}\right)\right]^{k} \geqslant \mathbb{E}\left[\exp \left(\boldsymbol{\theta}^{\top} \boldsymbol{U}_{l}\right)\right]^{k} \tag{3.40}
\end{equation*}
$$

Since the random variables $S_{l \leftarrow j}$ are non-negative, we obtain

$$
\begin{align*}
\liminf _{\boldsymbol{\theta}_{n} \nearrow \boldsymbol{\theta}} \Lambda\left(\boldsymbol{\theta}_{n}\right) & =\liminf _{\boldsymbol{\theta}_{n} \nearrow \boldsymbol{\theta}} \sum_{j=1}^{d} \bar{\lambda}_{j}\left(\mathbb{E}\left[\prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}}\right]-1\right) \\
& \geqslant \sum_{j=1}^{d} \bar{\lambda}_{j}\left(\mathbb{E}\left[\lim _{\boldsymbol{\theta}_{n} \nearrow \boldsymbol{\theta}} \prod_{l=1}^{d} m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}_{n}\right)^{S_{l \leftarrow j}}\right]-1\right) \geqslant \Lambda(\boldsymbol{\theta}) . \tag{3.41}
\end{align*}
$$

We have now verified that the limiting cumulant $\Lambda(\cdot)$ satisfies all conditions for the Gärtner-Ellis theorem [19, Theorem 2.3.6] to apply. This concludes the proof of the LDP.

The consequence of this LDP is that, for any Borel set $A \in \mathcal{B}\left(\mathbb{R}^{d^{\star}}\right)$, we have that the measure $\nu_{t}: \mathcal{B}\left(\mathbb{R}^{d^{\star}}\right) \rightarrow[0,1]$ defined by $\nu_{t}(A):=\mathbb{P}(\boldsymbol{Z}(t) / t \in A)$ satisfies

$$
\begin{equation*}
-\inf _{\boldsymbol{x} \in A^{\circ}} \Lambda^{*}(\boldsymbol{x}) \leqslant \liminf _{t \rightarrow \infty} \frac{1}{t} \log \nu_{t}(A) \leqslant \limsup _{t \rightarrow \infty} \frac{1}{t} \log \nu_{t}(A) \leqslant-\inf _{\boldsymbol{x} \in \bar{A}} \Lambda^{*}(\boldsymbol{x}) \tag{3.42}
\end{equation*}
$$

### 3.4.2 LDP of a single component

As a special case of interest, explicitly required later in this paper, we now provide the LDP of a single component. Hence, throughout this subsection, fix $i \in\left[d^{\star}\right]$ and consider the $\mathbb{R}$-valued component $Z_{i}(\cdot)$ of the multivariate compound Hawkes process, as defined in Eqn. (3.2). The associated limiting cumulant is

$$
\begin{equation*}
\Lambda_{i}(\theta):=\lim _{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta Z_{i}(t)}\right]=\Lambda(0, \ldots, \theta, \ldots, 0) \tag{3.43}
\end{equation*}
$$

where the input vector of $\Lambda(\cdot)$ is non-zero on the $i$-th position. It is noted that whereas $Z_{i}(t)$ is a one-dimensional object, it is still driven by the multivariate Hawkes process $N(\cdot)$.

Compared to the multivariate setting in Section 3.4.1, we now have to work with the domain $\mathscr{D}_{\Lambda_{i}}:=\left\{\theta \in \mathbb{R}: \Lambda_{i}(\theta)<\infty\right\}$, where the argument of the probability generating function of $\boldsymbol{S}_{j}$ is given by

$$
\boldsymbol{m}_{U_{(i)}}(\theta):=\left(m_{U_{i 1}}(\theta), \ldots, m_{U_{i d}}(\theta)\right)^{\top}
$$

with $m_{U_{i j}}(\theta)=\mathbb{E}\left[e^{\theta U_{i j}}\right]$. Here, the vector $\boldsymbol{U}_{(i)}=\left(U_{i 1}, \ldots, U_{i d}\right)^{\top}$ is the $i$-th row of the matrix $\boldsymbol{U}$, where $i \in\left[d^{\star}\right]$. From Assumption 3.3, we find that for any $\widehat{\boldsymbol{z}}$ as the solution to (3.17), there exists $\widehat{\theta}>0$ such that $\boldsymbol{m}_{U_{(i)}}(\widehat{\theta})=\widehat{\boldsymbol{z}}$, and for $\theta \leqslant \widehat{\theta}$, we have by Lemma 3.1,

$$
\begin{equation*}
\Lambda_{i}(\theta)=\sum_{j=1}^{d} \bar{\lambda}_{j}\left(\mathbb{E}\left[\prod_{l=1}^{d} m_{U_{i l}}(\theta)^{S_{l<j}}\right]-1\right) . \tag{3.44}
\end{equation*}
$$

Corollary 3.1. The process $\left(Z_{i}(t) / t\right)_{t \in \mathbb{R}_{+}}$satisfies on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the LDP with good rate function

$$
\begin{equation*}
\Lambda_{i}^{*}(x)=\sup _{\theta \in \mathbb{R}}\left(\theta x-\Lambda_{i}(\theta)\right) . \tag{3.45}
\end{equation*}
$$

### 3.5 Ruin Probabilities

In this section, we analyze a risk process in which the claims are generated by a multivariate compound Hawkes process. In particular, using the LDP results for $Z_{i}(t) / t$ established in the previous section, we characterize the asymptotic behavior of the ruin probabilities of the corresponding risk process.

We assume a constant premium rate $r>0$ per unit of time and consider, for a given $i \in\left[d^{\star}\right]$, the net cumulative claim process (or: risk process)

$$
\begin{equation*}
Y_{i}(t):=Z_{i}(t)-r t . \tag{3.46}
\end{equation*}
$$

## Chapter 3 Multivariate Compound Hawkes Processes

Our objective is to find the asymptotics of the associated ruin probability, i.e., the probability that this net cumulative process ever exceeds level $u$, for some $u>0$.

From the LLN result for $\boldsymbol{Z}(t) / t$ given in Eqn. (3.9), we have that

$$
\begin{equation*}
\frac{Z_{i}(t)}{t} \rightarrow \mathbb{E}\left[\boldsymbol{U}_{(i)}\right](\boldsymbol{I}-\boldsymbol{H})^{-1} \overline{\boldsymbol{\lambda}}, \tag{3.47}
\end{equation*}
$$

a.s., with $\mathbb{E}\left[\boldsymbol{U}_{(i)}\right]=\left(\mathbb{E}\left[U_{i 1}\right], \ldots, \mathbb{E}\left[U_{i d}\right]\right)$. To make sure that ruin is rare, we impose throughout the net profit condition:

$$
\begin{equation*}
r>\mathbb{E}\left[\boldsymbol{U}_{(i)}\right](\boldsymbol{I}-\boldsymbol{H})^{-1} \overline{\boldsymbol{\lambda}}, \tag{3.48}
\end{equation*}
$$

such that the process $Y_{i}(t)$ drifts towards $-\infty$. For some initial capital $u>0$, the time of ruin is defined as

$$
\tau_{u}:=\inf \left\{t>0: u+r t-Z_{i}(t)<0\right\}=\inf \left\{t>0: Y_{i}(t)>u\right\}
$$

and the associated infinite horizon ruin probability is defined as

$$
\begin{equation*}
p(u):=\mathbb{P}\left(\tau_{u}<\infty\right) \tag{3.49}
\end{equation*}
$$

We study the behavior of $p(u)$ for $u$ large.
From Lemma 3.1, it immediately follows that the limiting cumulant function of $Y_{i}(\cdot)$ satisfies

$$
\begin{equation*}
\Psi_{i}(\theta):=\lim _{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta Y_{i}(t)}\right]=\Lambda_{i}(\theta)-r \theta \tag{3.50}
\end{equation*}
$$

with $\Lambda_{i}(\cdot)$ given in Eqn. (3.44). By [19, Lemma 2.3.9], we know that $\Lambda_{i}(\cdot)$ is a convex function, which implies that $\Psi_{i}(\cdot)$ is also convex. We assume throughout the paper that we are in the light-tailed regime, in the sense that there exists $\theta^{\star}>0$ such that

$$
\begin{equation*}
\Psi_{i}\left(\theta^{\star}\right)=0 . \tag{3.51}
\end{equation*}
$$

In the rest of this section, we prove that $p(u)$ decays essentially exponentially as $u$ increases, as made precise in the following theorem.

Theorem 3.2. For fixed $i \in\left[d^{\star}\right]$, the ruin probability $p(u)$ associated to the risk process $Y_{i}(\cdot)$ has logarithmic decay rate $-\theta^{\star}$, i.e.,

$$
\begin{equation*}
\lim _{u \rightarrow \infty} \frac{1}{u} \log p(u)=-\theta^{\star} \tag{3.52}
\end{equation*}
$$

where $\theta^{\star}$ is the unique positive solution of (3.51).

Proof. First note that $\theta^{\star}>0$ is unique by combining the observations $\Psi_{i}(0)=0$, $\Psi_{i}^{\prime}(0)<0$ as a consequence of (3.48), and the convexity of $\Psi_{i}(\cdot)$. The rest of the proof consists of two main steps, which yield a lower bound and an upper bound on the left-hand side of Eqn. (3.52).

- Lower bound. The objective is to prove that

$$
\begin{equation*}
\liminf _{u \rightarrow \infty} \frac{1}{u} \log p(u) \geqslant-\theta^{\star} \tag{3.53}
\end{equation*}
$$

We first observe that, for any $t>0$, we evidently have $p(u) \geqslant p(u, t):=\mathbb{P}\left(Y_{i}(t)>u\right)$. This directly implies that, for any $t>0$,

$$
\begin{aligned}
\liminf _{u \rightarrow \infty} \frac{1}{u} \log p(u) & \geqslant \lim _{u \rightarrow \infty} \frac{1}{u} \log p(u, u t) \\
& =\liminf _{u \rightarrow \infty} \frac{1}{u} \log \mathbb{P}\left(Z_{i}(u t)-r u t>u\right) \\
& =t \liminf _{u \rightarrow \infty} \frac{1}{u t} \log \mathbb{P}\left(Z_{i}(u t) / u t>r+1 / t\right)
\end{aligned}
$$

Then we can use that, by Corollary 3.1, $\left\{Z_{i}(u t) / u t\right\}_{t \in \mathbb{R}_{+}}$satisfies the LDP with rate function $\Lambda_{i}^{*}(\cdot)$. We thus obtain

$$
\liminf _{u \rightarrow \infty} \frac{1}{u} \log p(u) \geqslant-t \inf _{x>r+1 / t} \Lambda_{i}^{*}(x)
$$

Recall the assumption given in Eqn. (3.48) that $r>\mu_{i}:=\mathbb{E}\left[\boldsymbol{U}_{(i)}\right](\boldsymbol{I}-\boldsymbol{H})^{-1} \overline{\boldsymbol{\lambda}}$. Further, note that $\Lambda_{i}^{*}(\cdot)$ is continuous and non-decreasing on $\left(\mu_{i}, \infty\right)$ by [37, Lemma 2.7], and in addition $\Lambda_{i}^{*}\left(\mu_{i}\right)=0$. Upon combining all these elements, we have that

$$
\liminf _{u \rightarrow \infty} \frac{1}{u} \log p(u) \geqslant-t \Lambda_{i}^{*}(r+1 / t)
$$

To compute this final expression, observe that since $\Psi_{i}(0)=0, \Psi_{i}\left(\theta^{\star}\right)=0$ and $\Psi_{i}(\cdot)$ is convex, we have that $\Psi_{i}^{\prime}\left(\theta^{\star}\right)=\Lambda_{i}^{\prime}\left(\theta^{\star}\right)-r>0$. Since $t>0$ was arbitrary, we can in particular take $t=t^{\star}:=\left(\Lambda_{i}^{\prime}\left(\theta^{\star}\right)-r\right)^{-1}$. This implies that

$$
\begin{aligned}
\Lambda_{i}^{*}\left(r+1 / t^{\star}\right) & =\Lambda_{i}^{*}\left(\Lambda_{i}^{\prime}\left(\theta^{\star}\right)\right)=\sup _{\theta \in \mathbb{R}}\left(\theta \Lambda_{i}^{\prime}\left(\theta^{\star}\right)-\Lambda_{i}(\theta)\right) \\
& =\theta^{\star} \Lambda_{i}^{\prime}\left(\theta^{\star}\right)-\Lambda_{i}\left(\theta^{\star}\right)=\theta^{\star}\left(\Lambda_{i}^{\prime}\left(\theta^{\star}\right)-r\right)=\frac{\theta^{\star}}{t^{\star}} ;
\end{aligned}
$$

the third equality follows by noting that the supremum is attained at $\theta^{\star}$ since $g(\theta):=$ $\theta \Lambda_{i}^{\prime}\left(\theta^{\star}\right)-\Lambda_{i}(\theta)$ is maximal if $g^{\prime}(\theta)=0$, which is when $\theta=\theta^{\star}$; the fourth equality follows from $\Psi_{i}\left(\theta^{\star}\right)=0$, which is equivalent to $\Lambda_{i}\left(\theta^{\star}\right)=r \theta^{\star}$; and the last equality follows from the definition of $t^{\star}$. Hence, we conclude that

$$
\liminf _{u \rightarrow \infty} \frac{1}{u} \log p(u) \geqslant-t^{\star} \Lambda_{i}^{*}\left(r+1 / t^{\star}\right)=-\theta^{\star}
$$

## Chapter 3 Multivariate Compound Hawkes Processes

which proves Eqn. (3.53).

- Upper bound. We next prove the upper bound, i.e.,

$$
\begin{equation*}
\limsup _{u \rightarrow \infty} \frac{1}{u} \log p(u) \leqslant-\theta^{\star} . \tag{3.54}
\end{equation*}
$$

To do so, we majorize the ruin probability $p(u)$ by a finite sum and a sum with countably many terms, as follows. First, observe that for $u>r$, we have

$$
p(u) \leqslant q(u):=\mathbb{P}\left(\inf _{n \in \mathbb{N}} Y_{i}(n) \geqslant u-r\right)
$$

where we have used that the net cumulative claim process can go down with maximally an amount $r$ per time unit. Define, as before, $t^{\star}=\left(\Lambda_{i}^{\prime}\left(\theta^{\star}\right)-r\right)^{-1}$. For any $L>0$, the union bound gives

$$
\begin{aligned}
q(u) & \leqslant \sum_{n \leqslant(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant u-r\right)+\sum_{n>(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant u-r\right) \\
& \leqslant \sum_{n \leqslant(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant u-r\right)+\sum_{n>(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant 0\right),
\end{aligned}
$$

where $n$ takes integer values in both sums. The intuition is that the first sum contains the contribution of paths corresponding to the most likely timescale $t^{\star} u$ of exceeding $u$. This means that the first sum is expected to dominate the second sum as $u \rightarrow \infty$, which is proven to be correct in the remainder of the proof.

Consider the finite sum, and note that the Chernoff bound yields for any $\theta>0$ that

$$
\begin{align*}
\sum_{n \leqslant(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant u-r\right) & \leqslant \sum_{n \leqslant(1+L) t^{\star} u} e^{-\theta(u-r)} \mathbb{E}\left[e^{\theta Y_{i}(n)}\right]  \tag{3.55}\\
& \leqslant(1+L) t^{\star} u e^{-\theta(u-r)} \max _{n \in\left[(1+L) t^{\star} u\right]} \mathbb{E}\left[e^{\theta Y_{i}(n)}\right] .
\end{align*}
$$

In particular, this inequality holds for $\theta=\theta^{\star}$. Inserting the familiar expression for the cumulant of $Z_{i}(t)$, we have for any $t>0$ that

$$
\begin{aligned}
\log \mathbb{E}\left[e^{\theta^{\star} Y_{i}(t)}\right] & =\log \mathbb{E}\left[e^{\theta^{\star} Z_{i}(t)}\right]-r \theta^{\star} t \\
& =\sum_{j=1}^{d} \bar{\lambda}_{j} \int_{0}^{t} \mathbb{E}\left[\prod_{l=1}^{d} m_{U_{i l}}\left(\theta^{\star}\right)^{S_{l \leftarrow j}(v)}\right] \mathrm{d} v-t \sum_{j=1}^{d} \bar{\lambda}_{j}-r \theta^{\star} t \\
& \leqslant t \sum_{j=1}^{d} \bar{\lambda}_{j} \sup _{v \leqslant t} \mathbb{E}\left[\prod_{l=1}^{d} m_{U_{i l}}\left(\theta^{\star}\right)^{S_{l \leftarrow j}(v)}\right]-t \sum_{j=1}^{d} \bar{\lambda}_{j}-r \theta^{\star} t \\
& \leqslant t\left(\sum_{j=1}^{d} \bar{\lambda}_{j} \mathbb{E}\left[\prod_{l=1}^{d} m_{U_{i l}}\left(\theta^{\star}\right)^{S_{l \leftarrow j}}\right]-\sum_{j=1}^{d} \bar{\lambda}_{j}-r \theta^{\star}\right)=t \Psi_{i}\left(\theta^{\star}\right)=0,
\end{aligned}
$$

where in the second inequality we used that $S_{l \leftarrow j}(v)$ increases in $v$, and in the final equality we used the definition of $\theta^{\star}$, as given in Eqn. (3.51). The next step is to combine this with the upper bound that was found in Eqn. (3.55). This yields

$$
\begin{aligned}
& \limsup _{u \rightarrow \infty} \frac{1}{u} \log \sum_{n \leqslant(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant u-r\right) \\
& \leqslant \limsup _{u \rightarrow \infty} \frac{1}{u} \log \left((1+L) t^{\star} u \max _{n \in\left[(1+L) t^{\star} u\right]} e^{-\theta^{\star}(u-r)} \mathbb{E}\left[e^{\theta^{\star} Y_{i}(n)}\right]\right) \\
& \leqslant \limsup _{u \rightarrow \infty} \frac{1}{u} \log \left((1+L) t^{\star} u\right)+\lim _{u \rightarrow \infty} \frac{-\theta^{\star} u+\theta^{\star} r}{u}+0=-\theta^{\star} .
\end{aligned}
$$

For the sum with countably many terms, we again apply the Chernoff bound. For any $\theta>0$, we thus obtain

$$
\sum_{n>(1+L) t^{\star} u} \mathbb{P}\left(Y_{i}(n) \geqslant 0\right) \leqslant \sum_{n>(1+L) t^{\star} u} \mathbb{E}\left[e^{\theta Y_{i}(n)}\right] .
$$

To proceed, we observe that the assumption $\Psi_{i}\left(\theta^{\star}\right)=0$ together with $\Psi_{i}(0)=0$ implies, by the mean value theorem, that there exists $\theta^{\circ}>0$ such that $\Psi_{i}^{\prime}\left(\theta^{\circ}\right)=0$. It requires elementary calculus to verify that

$$
\Psi_{i}^{\prime}(0)=\Lambda_{i}^{\prime}(0)-r=\sum_{j=1}^{d} \bar{\lambda}_{j} \sum_{k=1}^{d} \mathbb{E}\left[U_{i k}\right] \mathbb{E}\left[S_{k \leftarrow j}\right]-r=\mathbb{E}\left[\boldsymbol{U}_{(i)}\right] \mathbb{E}[\boldsymbol{S}] \overline{\boldsymbol{\lambda}}-r,
$$

cf. Eqn. (3.9). We conclude that $\Psi_{i}^{\prime}(0)<0$ from the net profit condition (3.48). Combining the above observations with the fact that $\Psi_{i}(\cdot)$ is convex, it follows that $\Psi_{i}\left(\theta^{\circ}\right)<0$. Hence, we can choose $n_{0} \in \mathbb{N}$ and $\delta \in\left(0,\left|\Psi_{i}\left(\theta^{\circ}\right)\right|\right)$ such that for all $n$ larger than $n_{0}$, we have that

$$
\frac{1}{n} \log \mathbb{E}\left[e^{\theta^{\circ} Y_{i}(n)}\right]<\Psi_{i}\left(\theta^{\circ}\right)+\delta .
$$

Finally, since $\Psi_{i}\left(\theta^{\circ}\right)+\delta<0$, we have that $z:=\exp \left(\Psi_{i}\left(\theta^{\circ}\right)+\delta\right)<1$, so that we can apply the geometric series to bound

$$
\sum_{n>(1+L) t^{\star} u} \mathbb{E}\left[e^{\theta^{\circ} Y_{i}(n)}\right] \leqslant \sum_{n>(1+L) t^{\star} u} \exp \left(n\left(\Psi_{i}\left(\theta^{\circ}\right)+\delta\right)\right) \leqslant \frac{z^{(1+L) t^{\star} u}}{1-z}
$$

We conclude the proof by combining the two upper bounds. We take $L$ large enough such that $z^{(1+L) t^{\star}}<e^{-\theta^{\star}}$, for which it is sufficient that

$$
L>\frac{\theta^{\star}}{\left(\Phi\left(\theta^{\circ}\right)+\delta\right) t^{\star}} .
$$

## Chapter 3 Multivariate Compound Hawkes Processes

As a consequence,

$$
\begin{equation*}
\limsup _{u \rightarrow \infty} \frac{1}{u} \log p(u) \leqslant \limsup _{u \rightarrow \infty} \frac{1}{u} \log \left((1+L) t^{\star} u e^{-\theta^{\star}(u-r)}+\frac{z^{(1+L) t^{\star} u}}{1-z}\right)=-\theta^{\star} . \tag{3.56}
\end{equation*}
$$

We have thus established (3.54).

### 3.6 Rare Event Simulation

In this section, we use importance sampling to efficiently estimate rare event probabilities. This is accomplished by first exponentially twisting the underlying probability measure $\mathbb{P}$. In Section 3.6.1, we describe how to identify the model primitives under this new measure, which we throughout refer to as $\mathbb{Q}$. In the two subsequent subsections, we specifically consider the probability of ruin in component $i \in[d]$ (of which the logarithmic asymptotics have been derived in Section 3.5), and the probability of the multivariate compound Hawkes process attaining rare values (of which the logarithmic asymptotics have been established in Theorem 3.1).

### 3.6.1 Identification of the alternative distribution

In this subsection, we describe how to construct the exponentially twisted version of the multivariate compound Hawkes process, which we associate with the probability measure $\mathbb{Q}$. More specifically, we identify a stochastic process of which the limiting cumulant equals, for a vector $\boldsymbol{\theta}^{\star} \in \mathbb{R}^{d^{\star}}$,

$$
\Psi^{\mathbb{Q}}(\boldsymbol{\theta}):=\Psi\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)-\Psi\left(\boldsymbol{\theta}^{\star}\right),
$$

with $\Psi(\boldsymbol{\theta})=\Lambda(\boldsymbol{\theta})-\boldsymbol{r}^{\top} \boldsymbol{\theta}$, and, by virtue of Lemma 3.1,

$$
\begin{equation*}
\Lambda(\boldsymbol{\theta})=\sum_{j=1}^{d} \bar{\lambda}_{j}\left(f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}(\boldsymbol{\theta})\right)-1\right) \tag{3.57}
\end{equation*}
$$

To this end, it is first verified that

$$
\Lambda\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)-\Lambda\left(\boldsymbol{\theta}^{\star}\right)=\sum_{j=1}^{d} \bar{\lambda}_{j} f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\left(\frac{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)\right)}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)}-1\right) .
$$

Next, for $j \in[d]$,

$$
\frac{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)\right)}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)}=\frac{1}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)} \sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \mathbb{P}\left(\boldsymbol{S}_{j}=\boldsymbol{n}\right) \prod_{l=1}^{d}\left(m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)\right)^{n_{l}}
$$

$$
=\sum_{\boldsymbol{n} \in \mathbb{N}_{0}^{d}} \frac{\mathbb{P}\left(\boldsymbol{S}_{j}=\boldsymbol{n}\right)}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)} \prod_{l=1}^{d}\left(m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}^{\star}\right)\right)^{n_{l}}\left(\frac{m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)}{m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}^{\star}\right)}\right)^{n_{l}}
$$

Now define, for $j \in[d]$,

$$
\mathbb{Q}\left(\boldsymbol{S}_{j}=\boldsymbol{n}\right):=\frac{\mathbb{P}\left(\boldsymbol{S}_{j}=\boldsymbol{n}\right)}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)} \prod_{l=1}^{d}\left(m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}^{\star}\right)\right)^{n_{l}}
$$

which induces a probability distribution (i.e., non-negative and summing to 1 ) by its very construction. Define by $f_{j}^{\mathbb{Q}}(\boldsymbol{z})$ the corresponding probability generating function, which is the counterpart of $f_{j}(\boldsymbol{z})$ under $\mathbb{Q}$ : for $j \in[d]$,

$$
\begin{equation*}
f_{j}^{\mathbb{Q}}(\boldsymbol{z})=\sum_{n \in \mathbb{N}_{0}^{d}} \mathbb{Q}\left(\boldsymbol{S}_{j}=\boldsymbol{n}\right) \prod_{l=1}^{d} z_{l}^{n_{l}}=\frac{f_{j}\left(m_{\boldsymbol{U}_{1}}\left(\boldsymbol{\theta}^{\star}\right) z_{1}, \ldots, m_{\boldsymbol{U}_{d}}\left(\boldsymbol{\theta}^{\star}\right) z_{d}\right)}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)} \tag{3.58}
\end{equation*}
$$

In addition, define, for $l \in[d]$,

$$
\begin{equation*}
\mathbb{Q}\left(U_{1 l} \in \mathrm{~d} x_{1}, \ldots, U_{d^{\star} l} \in \mathrm{~d} x_{d^{\star}}\right):=\frac{\mathbb{P}\left(U_{1 l} \in \mathrm{~d} x_{1}, \ldots, U_{d^{\star} l} \in \mathrm{~d} x_{d^{\star}}\right)}{m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}^{\star}\right)} \prod_{k=1}^{d^{\star}} e^{\theta_{k}^{\star} x_{k}} \tag{3.59}
\end{equation*}
$$

which generates a probability distribution (i.e., non-negative and integrating to 1 ); let $m_{\boldsymbol{U}_{l}}^{\mathbb{Q}}(\boldsymbol{\theta})$ be the associated moment generating function, given by

$$
m_{\boldsymbol{U}_{l}}^{\mathbb{Q}}(\boldsymbol{\theta})=\frac{m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)}{m_{\boldsymbol{U}_{l}}\left(\boldsymbol{\theta}^{\star}\right)} .
$$

We finally define the base rates under $\mathbb{Q}$ via

$$
\begin{equation*}
\bar{\lambda}_{j}^{\mathbb{Q}}:=\bar{\lambda}_{j} f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right), \tag{3.60}
\end{equation*}
$$

for $j \in[d]$.
Upon combining the objects defined above, it now requires an elementary verification to conclude that

$$
\Lambda\left(\boldsymbol{\theta}+\boldsymbol{\theta}^{\star}\right)-\Lambda\left(\boldsymbol{\theta}^{\star}\right)=\sum_{j=1}^{d} \bar{\lambda}_{j}^{\mathbb{Q}}\left(f_{j}^{\mathbb{Q}}\left(\boldsymbol{m}_{U}^{\mathbb{Q}}(\boldsymbol{\theta})\right)-1\right),
$$

as desired; cf. (3.57). This means that we have uniquely characterized the joint distribution of the cluster sizes $\boldsymbol{S}_{j}$ (for $j \in[d]$ ), the joint distribution of the claim sizes $\boldsymbol{U}_{l}$ (for $l \in[d]$ ), and the base rates under the alternative measure $\mathbb{Q}$.

The only question left is: How does one sample a cluster size $\boldsymbol{S}_{j}$ under $\mathbb{Q}$ ? More concretely: What is the distribution of the marks $B_{l j}$ under the alternative measure

## Chapter 3 Multivariate Compound Hawkes Processes

$\mathbb{Q}$, and how should the corresponding decay functions $g_{l j}(\cdot)$ be adapted? To this end, we revisit (3.16), which we rewrite to

$$
\begin{equation*}
f_{j}(\boldsymbol{z})=z_{j} m_{\boldsymbol{B}_{j}}\left(c_{1 j}\left(f_{1}(\boldsymbol{z})-1\right), \ldots, c_{d j}\left(f_{d}(\boldsymbol{z})-1\right)\right), \tag{3.61}
\end{equation*}
$$

using the self-evident notation

$$
m_{\boldsymbol{B}_{j}}(\boldsymbol{\theta}):=\mathbb{E} \exp \left(\sum_{m=1}^{d} \theta_{m} B_{m j}\right) .
$$

Introduce the compact notation $\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z}):=\left(m_{\boldsymbol{U}_{1}}\left(\boldsymbol{\theta}^{\star}\right) z_{1}, \ldots, m_{\boldsymbol{U}_{d}}\left(\boldsymbol{\theta}^{\star}\right) z_{d}\right)^{\top}$. Hence, as an immediate consequence of (3.58), we obtain

$$
f_{j}^{\mathbb{Q}}(\boldsymbol{z})=\frac{f_{j}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)}{f_{j}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)} .
$$

Upon combining the two previous displays, we conclude that we can rewrite $f_{j}^{\mathbb{Q}}(\boldsymbol{z})$ as

$$
\begin{aligned}
f_{j}^{\mathbb{Q}}(\boldsymbol{z}) & =\frac{m_{\boldsymbol{U}_{j}}\left(\boldsymbol{\theta}^{\star}\right) z_{j} \cdot m_{\boldsymbol{B}_{j}}\left(c_{1 j}\left(f_{1}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)-1\right), \ldots, c_{d j}\left(f_{d}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)-1\right)\right)}{m_{\boldsymbol{U}_{j}}\left(\boldsymbol{\theta}^{\star}\right) \cdot m_{\boldsymbol{B}_{j}}\left(c_{1 j}\left(f_{1}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)-1\right), \ldots, c_{d j}\left(f_{d}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)-1\right)\right)} \\
& =z_{j} \frac{m_{\boldsymbol{B}_{j}}\left(c_{1 j}\left(f_{1}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)-1\right), \ldots, c_{d j}\left(f_{d}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)-1\right)\right)}{m_{\boldsymbol{B}_{j}}\left(c_{1 j}\left(f_{1}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)-1\right), \ldots, c_{d j}\left(f_{d}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)-1\right)\right)} .
\end{aligned}
$$

To simplify this further, we write

$$
\begin{equation*}
c_{l j}^{\mathbb{Q}}=c_{l j} f_{l}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right), \quad \bar{c}_{l j}^{\mathbb{Q}}:=c_{l j}^{\mathbb{Q}}-c_{l j}, \tag{3.62}
\end{equation*}
$$

such that

$$
f_{j}^{\mathbb{Q}}(\boldsymbol{z})=z_{j} \frac{m_{\boldsymbol{B}_{j}}\left(c_{1 j}^{\mathbb{Q}}\left(\frac{f_{1}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)}{f_{1}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)}-1\right)+\bar{c}_{1 j}^{\mathbb{Q}}, \ldots, c_{d j}^{\mathbb{Q}}\left(\frac{f_{d}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\boldsymbol{z})\right)}{f_{d}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)}-1\right)+\bar{c}_{d j}^{\mathbb{Q}}\right)}{m_{\boldsymbol{B}_{j}}\left(c_{1 j}^{\mathbb{Q}}, \ldots, \bar{c}_{d j}^{\mathbb{Q}}\right)} .
$$

We now focus on the distribution of the marks under the alternative measure $\mathbb{Q}$. Denoting $\overline{\boldsymbol{c}}_{j}^{\mathbb{Q}}=\left(\bar{c}_{1 j}^{\mathbb{Q}}, \ldots, \bar{c}_{d j}^{\mathbb{Q}}\right)^{\top}$, we define

$$
\begin{equation*}
\mathbb{Q}\left(B_{1 j} \in \mathrm{~d} x_{1}, \ldots, B_{d j} \in \mathrm{~d} x_{d}\right):=\frac{\mathbb{P}\left(B_{1 j} \in \mathrm{~d} x_{1}, \ldots, B_{d j} \in \mathrm{~d} x_{d}\right)}{m_{B_{j}}\left(\boldsymbol{c}_{j}^{\mathbb{Q}}\right)} \prod_{l=1}^{d} e^{\bar{c}_{l j}^{\mathrm{Q}} x_{l}} \tag{3.63}
\end{equation*}
$$

so that

$$
m_{\boldsymbol{B}_{j}}^{\mathbb{Q}}(\boldsymbol{\theta})=\frac{m_{\boldsymbol{B}_{j}}\left(\boldsymbol{\theta}+\overline{\boldsymbol{c}}_{j}^{\mathbb{Q}}\right)}{m_{\boldsymbol{B}_{j}}\left(\boldsymbol{c}_{j}^{\mathbb{Q}}\right)} .
$$

Combining the above relations, we thus conclude that

$$
f_{j}^{\mathbb{Q}}(\boldsymbol{z})=z_{j} m_{\boldsymbol{B}_{j}}^{\mathbb{Q}}\left(c_{1 j}^{\mathbb{Q}}\left(f_{1}^{\mathbb{Q}}(\boldsymbol{z})-1\right), \ldots, c_{d j}^{\mathbb{Q}}\left(f_{d}^{\mathbb{Q}}(\boldsymbol{z})-1\right)\right),
$$

which has, appealing to Eqn. (3.61), the right structure. This means that we have identified the distribution of the marks and the decay functions under $\mathbb{Q}$.

The following summarizes the above findings. Most importantly, the exponentially twisted version of the multivariate compound Hawkes process is again a multivariate compound Hawkes process, but (evidently) with different model primitives. Specifically, the $\boldsymbol{\theta}^{\star}$-twisted version of the multivariate compound Hawkes process can be constructed as follows:

- the base rate $\bar{\lambda}_{j}$ is replaced by $\bar{\lambda}_{j}^{\mathbb{Q}}=\bar{\lambda}_{j} f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)$; cf. (3.60).
- the density of $\boldsymbol{U}_{l}$ is replaced by $\mathbb{Q}\left(U_{1 l} \in \mathrm{~d} x_{1}, \ldots, U_{d^{\star} l} \in \mathrm{~d} x_{d^{\star}}\right)$, as given by (3.59);
- the density of $\boldsymbol{B}_{j}$ is replaced by $\mathbb{Q}\left(B_{1 j} \in \mathrm{~d} x_{1}, \ldots, B_{d j} \in \mathrm{~d} x_{d}\right)$, as given by (3.63);
- the decay function $g_{l j}(\cdot)$ is replaced by $g_{l j}^{\mathbb{Q}}(\cdot):=g_{l j}(\cdot) f_{l}\left(\boldsymbol{y}_{\boldsymbol{\theta}^{\star}}(\mathbf{1})\right)=g_{l j}(\cdot) f_{l}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)$; cf. (3.62).

This exponentially twisting mechanism generalizes the one identified for the univariate compound Hawkes process with unit marks, featuring in the statement of [69, Theorem 2.2].

### 3.6.2 Ruin probabilities

In this subsection, we return to the problem of assessing the ruin probability pertaining to the net cumulative claim process $Y_{i}(t)$, as defined by (3.46). We show that twisting $\boldsymbol{Y}(t)$ by $\boldsymbol{\theta}^{\star}=\left(0, \ldots, 0, \theta^{\star}, 0, \ldots, 0\right)^{\top}$, with the $\theta^{\star}$ corresponding to the $i$-th entry and solving $\Psi_{i}\left(\theta^{\star}\right)=0$, leads to an estimator that is asymptotically efficient (also sometimes referred to as logarithmically efficient, or asymptotically optimal); in the remainder of this subsection, we refer to this specific alternative measure by $\mathbb{Q}$. For more background on optimality notions of importance sampling procedures, such as asymptotic efficiency, we refer to [5, Section VI.1]. Our proof, in principle, follows the same structure as the one given in [69, Section 4]; as various steps directly mimic their counterparts in [69], we leave out some evident details and focus on the main innovations in this general, multivariate setting with random marks.

Recall that $p(u)=\mathbb{P}\left(\tau_{u}<\infty\right)$, with $\tau_{u}$ the first time that $Y_{i}(\cdot)$ exceeds level $u$. First note that $p(u)=\mathbb{E}_{\mathbb{Q}}\left[L_{\tau_{u}} I\right]$, where $I$ is the indicator function of the event $\left\{\tau_{u}<\infty\right\}$ and $L_{\tau_{u}}$ is the appropriate likelihood ratio, which quantifies the likelihood of the sampled

## Chapter 3 Multivariate Compound Hawkes Processes

path under $\mathbb{P}$ relative to $\mathbb{Q}$. More precisely, $L_{\tau_{u}}$ is the Radon-Nikodym derivative of the sampled path under the measure $\mathbb{P}$ relative to the measure $\mathbb{Q}$, evaluated at the ruin time $\tau_{u}$. As in [69, Lemma 4.3], it can be concluded that - essentially due to the fact that we changed the drift of the risk process from a negative value (under $\mathbb{P}$ ) into a positive value (under $\mathbb{Q}$ )—under $\mathbb{Q}$ eventually any positive value is reached by the process $Y_{i}(\cdot)$. Thus, $I \equiv 1$ with $\mathbb{Q}$-probability 1 , and hence $p(u)=\mathbb{E}_{\mathbb{Q}} L_{\tau_{u}}$.

Following the reasoning in [69] (i.e., effectively relying on a general result in [50]), we can express the likelihood ratio in terms of the various quantities pertaining to the original measure $\mathbb{P}$ and their counterparts under $\mathbb{Q}$. Indeed, the likelihood ratio at time $t$ equals

$$
\begin{align*}
\left.\frac{\mathrm{d} \mathbb{P}}{\mathrm{~d} \mathbb{Q}}\right|_{\mathcal{F}_{t}}=L_{t}= & \exp \left(-\sum_{j=1}^{d} \int_{0}^{t}\left(\lambda_{j}(s)-\lambda_{j}^{\mathbb{Q}}(s)\right) \mathrm{d} s\right) \exp \left(\sum_{j=1}^{d} \int_{0}^{t} \log \frac{\lambda_{j}(s)}{\lambda_{j}^{\mathbb{Q}}(s)} \mathrm{d} N_{j}(s)\right) \\
& \times \exp \left(\sum_{j=1}^{d} \sum_{r=1}^{N_{j}(t)} \log \ell_{j}\left(\boldsymbol{B}_{j, r}\right)\right) \prod_{j=1}^{d} \frac{m_{\boldsymbol{U}_{j}}\left(\boldsymbol{\theta}^{\star}\right)^{N_{j}(t)}}{e^{\theta_{j}^{*} Z_{j}(t)}}, \tag{3.64}
\end{align*}
$$

where

$$
\begin{equation*}
\lambda_{j}(s)=\bar{\lambda}_{j}+\sum_{l=1}^{d} \sum_{r=1}^{N_{l}(s)} B_{j l, r} g_{j l}\left(s-T_{l, r}\right), \quad \lambda_{j}^{\mathbb{Q}}(s)=\bar{\lambda}_{j}^{\mathbb{Q}}+\sum_{l=1}^{d} \sum_{r=1}^{N_{l}(s)} B_{j l, r} g_{j l}^{\mathbb{Q}}\left(s-T_{l, r}\right), \tag{3.65}
\end{equation*}
$$

with $\bar{\lambda}_{j}^{\mathbb{Q}}$ and $g_{j l}^{\mathbb{Q}}(\cdot)$ as defined in Section 3.6.1 and all random objects sampled under $\mathbb{Q}$, and with $\ell_{j}(\boldsymbol{x})$ denoting the ratio of the density of the random marks $\boldsymbol{B}_{j}$ under $\mathbb{P}$ and its counterpart under $\mathbb{Q}$ evaluated in the argument $\boldsymbol{x}$. It is directly seen, from the construction of the measure $\mathbb{Q}$, that

$$
\lambda_{j}^{\mathbb{Q}}(s)=\lambda_{j}(s) f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)>\lambda_{j}(s) .
$$

Note that the relation between $\lambda_{j}^{\mathbb{Q}}(s)$ and $\lambda_{j}(s)$, and the fact that $\boldsymbol{\theta}^{\star}$ is non-zero in the $i$-th entry, allows us to express the likelihood ratio as

$$
\begin{align*}
L_{t}= & \exp \left(-\sum_{j=1}^{d}\left(1-f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\right) \int_{0}^{t} \lambda_{j}(s) \mathrm{d} s\right) e^{-\theta^{\star} Z_{i}(t)} \\
& \times \exp \left(\sum_{j=1}^{d} \sum_{r=1}^{N_{j}(t)} \log \ell_{j}\left(\boldsymbol{B}_{j, r}\right)\right) \prod_{j=1}^{d}\left(\frac{m_{\boldsymbol{U}_{j}}\left(\boldsymbol{\theta}^{\star}\right)}{f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)}\right)^{N_{j}(t)} . \tag{3.66}
\end{align*}
$$

We now introduce the importance sampling estimator and prove its efficiency. With $n \in \mathbb{N}$, we define the importance sampling estimator of $p(u)$ by

$$
\begin{equation*}
p_{n}(u):=\frac{1}{n} \sum_{m=1}^{n} L_{\tau_{u}}^{(m)}, \tag{3.67}
\end{equation*}
$$

where $L_{\tau_{u}}^{(m)}$ (for $m=1, \ldots, n$ ) are independent replications of $L_{\tau_{u}}$, sampled under $\mathbb{Q}$. In our context, asymptotic efficiency is to be understood as

$$
\lim _{u \rightarrow \infty} \frac{1}{u} \log \sqrt{\operatorname{Var}_{\mathbb{Q}} L_{\tau_{u}}} \leqslant \lim _{u \rightarrow \infty} \frac{1}{u} \log p(u)
$$

that is, the measure $\mathbb{Q}$ is asymptotically efficient for simulations; see Siegmund's criterion [68].

Theorem 3.3. The importance sampling estimator $p_{n}(u)$ of the ruin probability $p(u)$, using the alternative measure $\mathbb{Q}$ that corresponds to the exponential twist

$$
\boldsymbol{\theta}^{\star}=\left(0, \ldots, 0, \theta^{\star}, 0, \ldots, 0\right),
$$

is asymptotically efficient.
Proof. In order to eventually prove that our estimator is asymptotically efficient, the main idea is to find an upper bound on $L_{\tau_{u}}$. By definition of the conditional intensities $\lambda_{j}(s)$ and $\lambda_{j}^{\mathbb{Q}}(s)$ in (3.65), we have

$$
\begin{array}{r}
\exp \left(-\int_{0}^{\tau_{u}} \sum_{j=1}^{d}\left(1-f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\right) \lambda_{j}(s) \mathrm{d} s\right)=\exp \left(-\int_{0}^{\tau_{u}} \sum_{j=1}^{d} \bar{\lambda}_{j}\left(1-f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\right) \mathrm{d} s\right) \\
\times \exp \left(-\int_{0}^{\tau_{u}} \sum_{l=1}^{d} \sum_{j=1}^{d} \sum_{r=1}^{N_{j}(s)} B_{l j, r} g_{l j}\left(s-T_{j, r}\right)\left(1-f_{l}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\right) \mathrm{d} s\right)
\end{array}
$$

where we switched the order of the summations over $j$ and $l$. Then observe that, recalling that $\theta^{\star}$ solves the equation $\Psi_{i}\left(\theta^{\star}\right)=0$,

$$
\exp \left(-\int_{0}^{\tau_{u}} \sum_{j=1}^{d} \bar{\lambda}_{j}\left(1-f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\right) \mathrm{d} s\right)=\exp \left(-\tau_{u} \sum_{j=1}^{d} \bar{\lambda}_{j}\left(1-f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)\right)\right)=e^{r \theta^{\star} \tau_{u}} .
$$

In addition, note that $Z_{i}\left(\tau_{u}\right)-r \tau_{u}>u$ due to the definition of $\tau_{u}$, which implies that

$$
e^{r \theta^{\star} \tau_{u}} e^{-\theta^{\star} Z_{i}\left(\tau_{u}\right)} \leqslant e^{-\theta^{\star} u} .
$$

## Chapter 3 Multivariate Compound Hawkes Processes

Also observe that since $c_{l j}=\left\|g_{l j}\right\|_{L^{1}\left(\mathbb{R}_{+}\right)}$, we have the bound

$$
\begin{align*}
& \exp \left(\int_{0}^{\tau_{u}} \sum_{j=1}^{d} \sum_{l=1}^{d} \sum_{r=1}^{N_{j}(s)} B_{l j, r} g_{l j}\left(s-T_{j, r}\right)\left(f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)-1\right) \mathrm{d} s\right) \\
& \leqslant \exp \left(\sum_{j=1}^{d} \sum_{l=1}^{d} \sum_{r=1}^{N_{j}\left(\tau_{u}\right)} B_{l j, r} c_{l j}\left(f_{l}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)-1\right)\right) \tag{3.68}
\end{align*}
$$

Finally, the contribution to the likelihood ratio $L_{\tau_{u}}$ due to the random marks is given by

$$
\begin{equation*}
\ell_{j}(\boldsymbol{v})=\exp \left(-\boldsymbol{v}^{\top} \overline{\boldsymbol{c}}^{\mathbb{Q}}\right) m_{\boldsymbol{B}_{j}}\left(\overline{\boldsymbol{c}}^{\mathbb{Q}}\right), \tag{3.69}
\end{equation*}
$$

where $\overline{\boldsymbol{c}}^{\mathbb{Q}}$ is the twist parameter for $\boldsymbol{B}_{j}$ as defined in Eqn. (3.62). As a consequence,

$$
\begin{equation*}
\exp \left(\sum_{j=1}^{d} \sum_{r=1}^{N_{j}\left(\tau_{u}\right)} \log \ell_{j}\left(\boldsymbol{B}_{j, r}\right)\right)=\exp \left(-\sum_{j=1}^{d} \sum_{r=1}^{N_{j}\left(\tau_{u}\right)} \boldsymbol{B}_{j, r}^{\top} \bar{c}^{\mathbb{Q}}\right) \prod_{j=1}^{d} m_{\boldsymbol{B}_{j}}\left(\overline{\boldsymbol{c}}^{\mathbb{Q}}\right)^{N_{j}\left(\tau_{u}\right)}, \tag{3.70}
\end{equation*}
$$

which implies that the expression given in (3.68) cancels against the exponential term in (3.70), noting that $\overline{\boldsymbol{c}}^{\mathbb{Q}}=\left(c_{1 j}\left(f_{1}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)-1\right), \ldots, c_{d j}\left(f_{d}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)-1\right)\right)^{\top}$.

Upon combining the above, and after some rewriting, we obtain that

$$
p(u) \leqslant e^{-\theta^{\star} u} \mathbb{E}_{\mathbb{Q}}\left[\prod_{j=1}^{d} \exp \left(-\log f_{j}\left(\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)\right)+\log m_{\boldsymbol{U}_{j}}\left(\boldsymbol{\theta}^{\star}\right)+\log m_{\boldsymbol{B}_{j}}\left(\overline{\boldsymbol{c}}^{\mathbb{Q}}\right)\right)^{N_{j}\left(\tau_{u}\right)}\right]
$$

As it turns out, the expression in the previous display simplifies considerably, as can be seen as follows. By (3.16), for any $\boldsymbol{z}$,

$$
\log f_{j}(\boldsymbol{z})=\log z_{j}+m_{\boldsymbol{B}_{j}}\left(c_{1 j}\left(f_{1}(\boldsymbol{z})-1\right), \ldots, c_{d j}\left(f_{d}(\boldsymbol{z})-1\right)\right) .
$$

Now plugging in $\boldsymbol{z}=\boldsymbol{m}_{\boldsymbol{U}}\left(\boldsymbol{\theta}^{\star}\right)$, we conclude that the expectation under the new measure $\mathbb{Q}$ fully reduces to unity, again by definition of $\overline{\boldsymbol{c}}^{\mathbb{Q}}$. This means that we have arrived at the upper bound $p(u)=\mathbb{E}_{\mathbb{Q}} L_{\tau_{u}} \leqslant e^{-\theta^{\star} u}$ and we are now in a position to conclude the statement. It follows directly from the observation

$$
\operatorname{Var}_{\mathbb{Q}} L_{\tau_{u}}=\mathbb{E}_{\mathbb{Q}} L_{\tau_{u}}^{2}-\left(\mathbb{E}_{\mathbb{Q}} L_{\tau_{u}}\right)^{2} \leqslant \mathbb{E}_{\mathbb{Q}} L_{\tau_{u}}^{2} \leqslant e^{-2 \theta^{\star} u}
$$

in combination with Theorem 3.2.
Remark 3.1. In the final part of the proof of Theorem 3.3, we have in passing derived a Lundberg-type inequality for this non-standard ruin model. Indeed, we have that
the ruin probability $p(u)$, corresponding to the net cumulative claim process $Y_{i}(\cdot)$, satisfies the upper bound

$$
\begin{equation*}
p(u) \leqslant e^{-\theta^{\star} u} \tag{3.71}
\end{equation*}
$$

uniformly in $u>0$.
The immediate consequence of the above theorem, which substantially generalizes [69, Theorem 4.5], is the following. Suppose that we wish to obtain an estimate with a certain precision, defined as the ratio of the confidence interval's half-width (which is proportional to the standard deviation of the estimate) and the estimate itself. Using simulation under the actual measure $\mathbb{P}$, the number of runs required to obtain a given precision is inversely proportional to the probability to be estimated. In our specific case, this means that, due to Theorem 3.2, under $\mathbb{P}$ this number grows exponentially in $u$ (roughly like $e^{\theta^{\star} u}$, that is). Under the alternative measure $\mathbb{Q}$, however, Theorem 3.3 entails that the number of runs to achieve this precision grows subexponentially, thus yielding a substantial variance reduction. This means that, despite the fact that the ruin probability decays very rapidly as $u$ grows, the simulation effort required to estimate it grows at a relatively modest pace.

### 3.6.3 Exceedance probabilities

In this subsection, we consider the estimation of probabilities of the type

$$
q_{t}(\boldsymbol{a}):=\mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1}, \ldots, \frac{Z_{d^{\star}}(t)}{t} \geqslant a_{d^{\star}}\right)
$$

where the set $A:=\left[a_{1}, \infty\right) \times \cdots \times\left[a_{d^{\star}}, \infty\right)$ does not contain the vector $\boldsymbol{\mu}$, with, as before

$$
\mu_{i}=\mathbb{E}\left[\boldsymbol{U}_{(i)}\right](\boldsymbol{I}-\boldsymbol{H})^{-1} \overline{\boldsymbol{\lambda}},
$$

the asymptotic value of the process $Z_{i}(t) / t$. We consider the regime that $t$ grows large, in which the event of interest becomes increasingly rare by Theorem 3.1. We show that the associated importance sampling estimator is asymptotically efficient. In our exposition, we start by detailing the case $d^{\star}=1$ and then provide a proof by example for the case $d^{\star}=2$, where it turns out that multiple cases need to be distinguished. We finish by briefly discussing how the case of general $d^{\star}$ can be dealt with, relying on the principles developed for the $d^{\star}=2$ case. As will become clear in our analysis, the underlying computations strongly resemble those in Section 3.6.2.

We now introduce the importance sampling estimator for the exceedance probability and then prove its efficiency. Let $I \equiv I_{\boldsymbol{a}}$ be the indicator for the rare event, i.e., set
$I_{a}:=\left\{Z_{1}(t) \geqslant a_{1} t, \ldots, Z_{d^{\star}}(t) \geqslant a_{d^{\star}} t\right\}$ for any given $t>0$. We define the importance sampling estimator for the probability of this event by

$$
\begin{equation*}
q_{t, n}(\boldsymbol{a}):=\frac{1}{n} \sum_{m=1}^{n} L_{t}^{(m)} I_{\boldsymbol{a}}^{(m)}, \tag{3.72}
\end{equation*}
$$

where $L_{t}^{(m)}$ are independent replications of $L_{t}$, sampled under $\mathbb{Q}$ with a twist parameter depending on $\boldsymbol{a}$. Here, $I_{a}^{(m)}$ are the associated indicators.

Theorem 3.4. The importance sampling estimator $q_{t, n}(\boldsymbol{a})$ of the exceedance probability $q_{t}(\boldsymbol{a})$, using the alternative measure $\mathbb{Q}$ that corresponds to the exponential twist

$$
\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)=\arg \sup _{\boldsymbol{\theta}}\left(\boldsymbol{\theta}^{\top} \boldsymbol{a}^{\star}-\Lambda(\boldsymbol{\theta})\right),
$$

with $\boldsymbol{a}^{\star}:=\arg \inf _{\boldsymbol{x} \in A} \Lambda^{\star}(\boldsymbol{x})$, is asymptotically efficient.
Proof. We provide the proof for $d^{\star}=1$, followed by a proof by example for $d^{\star}=2$, which is easily extended for $d^{\star}>2$. For $d^{\star}=1$, we first observe that Theorem 3.1 yields that, using that we assumed $a_{1}>\mu_{1}$,

$$
\lim _{t \rightarrow \infty} \frac{1}{t} \log q_{t}\left(a_{1}\right)=-\inf _{x \geqslant a_{1}} \Lambda^{\star}(x)=-\Lambda^{\star}\left(a_{1}\right) .
$$

Define

$$
\theta\left(a_{1}\right):=\arg \sup _{\theta}\left(\theta a_{1}-\Lambda(\theta)\right) ;
$$

it is straightforward to argue that $\theta\left(a_{1}\right)$ is positive for $a_{1}>\mu_{1}$. Letting $I$ be the indicator function of the event $\left\{Z_{1}(t) \geqslant a_{1} t\right\}, \mathbb{Q}$ the probability measure corresponding to exponentially twisting the original measure by $\theta\left(a_{1}\right)$, and $L_{t}$ the appropriate likelihood, we now have that

$$
q_{t}(\boldsymbol{a})=\mathbb{E}_{\mathbb{Q}}\left[L_{t} I\right] .
$$

As an aside, observe that in this setting, unlike the one discussed in Section 3.6.2, we do not have that $L_{t}=1$ almost surely under $\mathbb{Q}$. This is an immediate consequence of the fact that we constructed $\mathbb{Q}$ such that $\lim _{t \rightarrow \infty} \mathbb{E}_{\mathbb{Q}} Z_{1}(t) / t=a$, so that the central limit theorem implies that in roughly half of the runs, we have that $I=1$. The likelihood ratio can be evaluated by mimicking the calculations in Section 3.6.2. We thus obtain, leaving out indices in this single-dimensional case,

$$
\begin{aligned}
L_{t}= & \exp \left(-\left(1-f\left(m_{U}\left(\theta\left(a_{1}\right)\right)\right)\right) \int_{0}^{t} \lambda(s) \mathrm{d} s\right) e^{-\theta\left(a_{1}\right) Z_{1}(t)} \\
& \times \exp \left(\sum_{r=1}^{N(t)} \log \ell\left(B_{r}\right)\right)\left(\frac{m_{U}\left(\theta\left(a_{1}\right)\right)}{f\left(m_{U}\left(\theta\left(a_{1}\right)\right)\right)}\right)^{N(t)} .
\end{aligned}
$$

Using that $\Lambda(\theta)=\bar{\lambda}\left(f\left(m_{U}(\theta)\right)-1\right)$ and $m_{B}(c(f(z)-1))-\log f(z)+\log z=0$ (the latter identity being a consequence of (3.16)), and applying essentially the same majorizations as the ones used in Section 3.6.2, we readily obtain that

$$
q_{t}(a)=\mathbb{E}_{\mathbb{Q}}\left[L_{t} I\right] \leqslant e^{\Lambda\left(\theta\left(a_{1}\right)\right) t} e^{-\theta\left(a_{1}\right) Z_{1}(t)} I .
$$

The event $\{I=1\}$ is equivalent to $\left\{Z_{1}(t) \geqslant a t\right\}$, so that

$$
q_{t}(a) \leqslant e^{\Lambda\left(\theta\left(a_{1}\right)\right) t} e^{-\theta\left(a_{1}\right) a t}=e^{-\Lambda^{\star}\left(a_{1}\right) t} .
$$

We have thus obtained that

$$
q_{t}\left(a_{1}\right) \leqslant e^{-\Lambda^{\star}\left(a_{1}\right) t}
$$

which can be seen as a variant of the classical Chernoff bound. The asymptotic efficiency for $d^{\star}=1$ follows directly now. To this end, first note that using the very same reasoning we also find that

$$
\mathbb{E}_{\mathbb{Q}}\left[L_{t}^{2} I\right] \leqslant e^{-2 \Lambda^{\star}\left(a_{1}\right) t}
$$

so that also $\operatorname{Var}_{\mathbb{Q}}\left[L_{t} I\right] \leqslant e^{-2 \Lambda^{\star}\left(a_{1}\right) t}$. Combining this with Theorem 3.1, we conclude that in this single-dimensional case, we have asymptotic efficiency under the measure $\mathbb{Q}$ defined above.

We now move to the case $d^{\star}=2$, for which Theorem 3.1 gives

$$
\lim _{t \rightarrow \infty} \frac{1}{t} \log q_{t}(\boldsymbol{a})=-\inf _{\left(x_{1}, x_{2}\right) \in A} \Lambda^{\star}(\boldsymbol{x})
$$

Due to the convexity of the contour lines of $\Lambda^{\star}(\boldsymbol{a})$, with $\boldsymbol{a}^{\star}$ the optimizing $\boldsymbol{a} \in A$, three situations can occur: (i) $\boldsymbol{a}^{\star}=\boldsymbol{a}$, (ii) $a_{1}^{\star}=a_{1}$ and $a_{2}^{\star}>a_{2}$, and (iii) $a_{1}^{\star}>a_{1}$ and $a_{2}^{\star}=a_{2}$. As, by symmetry, cases (ii) and (iii) are conceptually the same and can therefore be treated identically, we restrict ourselves to discussing cases (i) and (ii) only. Let, as before, $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$ be the optimizing argument in the definition of $\Lambda^{\star}(\boldsymbol{a})$.

In case (i), using standard properties of the Legendre transform, we have that

$$
\theta_{1}\left(\boldsymbol{a}^{\star}\right)=\frac{\partial}{\partial a_{1}} \Lambda^{\star}\left(\boldsymbol{a}^{\star}\right)>0, \quad \theta_{2}\left(\boldsymbol{a}^{\star}\right)=\frac{\partial}{\partial a_{2}} \Lambda^{\star}\left(\boldsymbol{a}^{\star}\right)>0 .
$$

We let $\mathbb{Q}$ correspond to the $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$-twisted version of the original probability measure. Going through the same steps as in the case $d^{\star}=1$, we obtain that

$$
q_{t}(\boldsymbol{a})=\mathbb{E}_{\mathbb{Q}}\left[L_{t} I\right] \leqslant e^{\Lambda\left(\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)\right) t} e^{-\theta_{1}\left(\boldsymbol{a}^{\star}\right) Z_{1}(t)-\theta_{2}\left(\boldsymbol{a}^{\star}\right) Z_{2}(t)}
$$

Then note that the right-hand side of the expression in the previous display is, on the set $\{I=1\}=\left\{Z_{1}(t) \geqslant a_{1}^{\star} t, Z_{2}(t) \geqslant a_{2}^{\star} t\right\}$, bounded from above by

$$
e^{\Lambda\left(\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)\right) t} e^{-\theta_{1}\left(\boldsymbol{a}^{\star}\right) a_{1}^{\star} t-\theta_{2}\left(\boldsymbol{a}^{\star}\right) a_{2}^{\star} t}=e^{-\Lambda^{\star}\left(\boldsymbol{a}^{\star}\right) t} .
$$

## Chapter 3 Multivariate Compound Hawkes Processes

This implies $q_{t}(\boldsymbol{a}) \leqslant e^{-\Lambda^{\star}\left(a^{\star}\right) t}$, but in addition that $\operatorname{Var}_{\mathbb{Q}}\left[L_{t} I\right] \leqslant e^{-2 \Lambda^{\star}\left(a^{\star}\right) t}$. We conclude, using the same reasoning as before, that in this case twisting by $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$ yields asymptotic efficiency.

Case (ii) works similarly. Observe that now (using that the line $x=a_{1}$ is a tangent of the contour lines of the Legendre transform)

$$
\theta_{1}\left(\boldsymbol{a}^{\star}\right)=\frac{\partial}{\partial a_{1}} \Lambda^{\star}\left(\boldsymbol{a}^{\star}\right)>0, \quad \theta_{2}\left(\boldsymbol{a}^{\star}\right)=\frac{\partial}{\partial a_{2}} \Lambda^{\star}\left(\boldsymbol{a}^{\star}\right)=0 .
$$

The intuition is that in this case, if $Z_{1}(t) \geqslant a_{1} t$, then with high probability also $Z_{2}(t) \geqslant a_{2} t$, as reflected by the fact that

$$
\lim _{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1}, \frac{Z_{2}(t)}{t} \geqslant a_{2}\right)=\lim _{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1}\right) .
$$

Let $\mathbb{Q}$ be the $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$-twisted version of the original probability measure. After straightforward algebra, we now obtain that

$$
q_{t}(\boldsymbol{a})=\mathbb{E}_{\mathbb{Q}} L_{t} I \leqslant e^{\Lambda\left(\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)\right) t} e^{-\theta_{1}\left(\boldsymbol{a}^{\star}\right) Z_{1}(t)}=e^{-\Lambda^{\star}\left(\boldsymbol{a}^{\star}\right) t}
$$

and $\operatorname{Var}_{\mathbb{Q}}\left[L_{t} I\right] \leqslant e^{-2 \Lambda^{\star}\left(\boldsymbol{a}^{\star}\right) t}$. Hence, also in this case twisting by $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$ yields asymptotic efficiency.

It can be seen in a direct manner that the same procedure (i.e., working with a twist $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$ with non-negative entries) can be followed for any $d^{\star}$ larger than 2 . We have thus established the stated result.

We finish this subsection by briefly discussing a related, intrinsically more complicated, rare-event probability. Observe that $q_{t}(\boldsymbol{a})$ corresponds to the intersection of the events $\left\{Z_{1}(t) \geqslant a_{1} t\right\}, \ldots,\left\{Z_{d^{\star}}(t) \geqslant a_{d^{\star}} t\right\}$. Consider now, rather than the intersection of events, their union, e.g., in the case $d^{\star}=2$,

$$
\mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1} \text { or } \frac{Z_{2}(t)}{t} \geqslant a_{2}\right) .
$$

Assume that $\boldsymbol{\mu}<\boldsymbol{a}$, to make sure that we are dealing with a rare event probability and let $I$ denote the indicator of the union event.

Let $A^{c}$ be the complement of $\left(-\infty, a_{1}\right) \times\left(-\infty, a_{2}\right)$. Theorem 3.1 asserts that, as $t \rightarrow \infty$,

$$
\frac{1}{t} \log \mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1} \text { or } \frac{Z_{2}(t)}{t} \geqslant a_{2}\right) \rightarrow-\inf _{\boldsymbol{x} \in A^{c}} \Lambda^{\star}(\boldsymbol{x}) .
$$

With $\boldsymbol{a}^{\star}:=\arg \inf _{\boldsymbol{x} \in A^{c}} \Lambda^{\star}(\boldsymbol{x})$, suppose that we twist by $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)$. If it happens that $a_{1}^{\star}<a_{1}$ and $a_{2}^{\star}=a_{2}$, then, following the reasoning applied above, $\theta_{1}(\boldsymbol{a})=0$ and
$\theta_{2}(\boldsymbol{a})>0$. When simulating under this measure, one could however have that $I=1$ due to $Z_{1}(t) \geqslant a_{1} t$ (despite the fact that it is more likely to see $Z_{2}(t) \geqslant a_{2} t$ ). A similar effect occurs when $a_{1}^{\star}=a_{1}$ and $a_{2}^{\star}<a_{2}$. The implication is that in those cases we do not have a bound on the likelihood ratio $L_{t}$ (as opposed to the case of an intersection of events; see the above analysis for $q_{t}(\boldsymbol{a})$ ). There are various ways to deal with this inherent complication; see e.g., the discussions on this issue in [20, 61]. The most straightforward solution is to use the asymptotically efficient algorithm featured in Theorem 3.4 to separately estimate the three probabilities

$$
\mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1}, \frac{Z_{2}(t)}{t} \geqslant a_{2}\right), \mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1}\right), \mathbb{P}\left(\frac{Z_{2}(t)}{t} \geqslant a_{2}\right)
$$

and to add up the resulting estimates.

### 3.7 Examples and Numerical Illustrations

In this section, we provide a set of simulation experiments that illustrate the proposed rare event simulation algorithms and assess the achievable efficiency gains relative to conventional simulation methods. All simulations have been conducted in Python; the code is available from the authors upon request. Throughout this section, we consider the bivariate setting for both the Hawkes process $\boldsymbol{N}(\cdot)=\left(N_{1}(\cdot), N_{2}(\cdot)\right)^{\top}$ and the compound process $\boldsymbol{Z}(\cdot)=\left(Z_{1}(\cdot), Z_{2}(\cdot)\right)^{\top}$, i.e., we set $d=d^{\star}=2$.

### 3.7.1 Ruin probability

Without loss of generality, we focus in this subsection on the net cumulative claim process corresponding to the first component, i.e., $Y_{1}(t)=Z_{1}(t)-r t$. Our objective is to compute the ruin probability

$$
p(u)=\mathbb{P}\left(\exists t>0: Y_{1}(t)>0\right)=\mathbb{P}\left(\tau_{u}<\infty\right)=\mathbb{E}_{\mathbb{Q}}\left[L_{\tau_{u}}\right],
$$

where in the previous display we have used the notations that have been introduced in Sections 3.5 and 3.6. As before, it is assumed that the net profit condition (3.48) is in place. With $n$ denoting the number of runs, we let $p_{n}(u)$ denote our importance sampling estimator, see Eqn. (3.67).

We distinguish between the cases where the marks are deterministic and random. It is anticipated that, due to the increased variability of the driving Hawkes process, the ruin probabilities will be higher under random marks than under deterministic marks (obviously assuming that the mean mark sizes in the random mark model equal their

## Chapter 3 Multivariate Compound Hawkes Processes



Figure 3.1. Convergence of $u^{-1} \log p_{n}(u)$ to the logarithmic decay rate $-\theta^{\star}$ for the marginal process $Y_{1}(\cdot)$ in the bivariate model with deterministic marks. Chosen parameters are $\bar{\lambda}_{1}=$ $\bar{\lambda}_{2}=0.5, \alpha_{1}=2, \alpha_{2}=1.5, \beta_{11}=0.5, \beta_{12}=0.25, \beta_{21}=0.3, \beta_{22}=0.4, \mathbb{E} U_{11}=2$, $\mathbb{E} U_{12}=\mathbb{E} U_{21}=2.5, \mathbb{E} U_{22}=3$ and $r=8$. In this setting, solving $\Psi_{1}\left(\theta^{\star}\right)=0$ yields $\theta^{\star}=0.097$.
counterparts in the deterministic mark model). By applying our efficient simulation approach we can quantify this effect.

In case of deterministic marks, the model primitives are assumed to take the following form. For $i, j=1,2$, we take

$$
g_{i j}(t) \equiv g_{i}(t)=e^{-\alpha_{i} t}, \quad B_{i j}=\beta_{i j}, \quad U_{i j} \sim \operatorname{Exp}\left(u_{i j}\right),
$$

where $\alpha_{i}, \beta_{i j}, u_{i j}>0$. Note that due to the deterministic marks, the likelihood ratio $L_{\tau_{u}}$ given in Eqn. (3.66), simplifies considerably. Also note that in this case $m_{U_{i j}}(\theta)=$ $u_{i j}\left(u_{i j}-\theta\right)^{-1}$ for any $\theta<u_{i j}$ and $i, j=1,2$. The specific parameters used in the simulation experiments are provided in the captions of the figures.

In order to be able to evaluate the likelihood ratio $L_{\tau_{u}}$, we first calculate the 'twist vector' $\left(\theta^{\star}, 0\right)$, where $\theta^{\star}$ is found by solving $\Psi_{1}\left(\theta^{\star}\right)=0$. We then exponentially twist $\boldsymbol{Y}(\cdot)$ by $\left(\theta^{\star}, 0\right)$, using the change of measure introduced in Section 3.6.1, enabling us to sample $n$ times the likelihood ratio $L_{\tau_{u}}$ under the measure $\mathbb{Q}$, after which we can compute the importance sampling estimator $p_{n}(u)$. Recall that under the measure $\mathbb{Q}$, we have that the event $\left\{\tau_{u}<\infty\right\}$ happens with probability one since the twisted process has positive drift and will hit any level $u>0$. Figure 3.1 illustrates the validity of Theorem 3.2 by showing the convergence of $u^{-1} \log p_{n}(u)$ to the logarithmic decay rate $-\theta^{\star}$ as $u$ grows large; it, in addition, provides insight into the speed of convergence for this specific instance.


Figure 3.2. Convergence of $u^{-1} \log p_{n}(u)$ to the logarithmic decay rate $-\theta^{\star}$ for the marginal process $Y_{1}(\cdot)$ in the bivariate model with random marks. Chosen parameters are $\bar{\lambda}_{1}=\bar{\lambda}_{2}=$ $0.5, \alpha_{1}=2, \alpha_{2}=1.5, \mathbb{E} B_{11}=0.5, \mathbb{E} B_{12}=0.25, \mathbb{E} B_{21}=0.3, \mathbb{E} B_{22}=0.4, \mathbb{E} U_{11}=2$, $\mathbb{E} U_{12}=\mathbb{E} U_{21}=2.5, \mathbb{E} U_{22}=3$ and $r=8$. In this setting, solving $\Psi_{1}\left(\theta^{\star}\right)=0$ yields $\theta^{\star}=0.082$.

In the case of random marks, we consider the following instance:

$$
g_{i j}(t) \equiv g_{i}(t)=e^{-\alpha_{i} t}, \quad B_{i j} \sim \operatorname{Exp}\left(\gamma_{i j}\right), \quad U_{i j} \sim \operatorname{Exp}\left(u_{i j}\right)
$$

where $\alpha_{i}, \gamma_{i j}, u_{i j}>0$. To study the effect of the random marks, in our experiment we take $\mathbb{E} B_{i j}=1 / \gamma_{i j}=\beta_{i j}$ (recalling that the $\beta_{i j}$ were the deterministic marks that we have used in the first experiment). As before, we need to evaluate the likelihood ratio $L_{\tau_{u}}$, for which we first solve the equation $\Psi_{1}\left(\theta^{\star}\right)=0$. This requires us to compute the fixed point Eqn. (3.16) with random marks, where the i.i.d. assumption that was imposed on the marks $B_{i j}$ implies that, for $j=1,2$, we have

$$
\begin{equation*}
f_{j}(\boldsymbol{z})=z_{j} \frac{\gamma_{1 j}}{\gamma_{1 j}-c_{1}\left(f_{1}(\boldsymbol{z})-1\right)} \frac{\gamma_{2 j}}{\gamma_{2 j}-c_{2}\left(f_{2}(\boldsymbol{z})-1\right)} . \tag{3.73}
\end{equation*}
$$

Then we again exponentially twist $\boldsymbol{Y}(\cdot)$ by $\left(\theta^{\star}, 0\right)$ to sample under the measure $\mathbb{Q}$ and compute $L_{\tau_{u}}$. Figure 3.2 confirms convergence of $u^{-1} \log p_{n}(u)$ to $-\theta^{\star}$. Note that the decay rate $-\theta^{\star}$ is higher than in the case with deterministic marks, reflecting that the increased variability due to the random marks leads to a higher ruin probability.

In the next experiment we study the efficiency of the proposed estimator in terms of the number of runs needed to reach a predefined level of precision. As before, we consider ruin in the first component in the bivariate model. We compare the importance sampling estimators in the setting with deterministic and random marks. We continue generating runs until the relative standard error of the importance sampling

## Chapter 3 Multivariate Compound Hawkes Processes

estimator becomes less than the precision parameter $\epsilon>0$. More precisely, we denote the relative standard error (after $n$ runs, that is) and the number of required runs by

$$
\begin{equation*}
\epsilon_{n}:=\frac{\sqrt{v_{n, \mathrm{IS}}(u)}}{p_{n}(u) \sqrt{n}}, \quad \widehat{n}:=\inf \left\{n \in \mathbb{N}: \epsilon_{n}<\epsilon\right\}, \tag{3.74}
\end{equation*}
$$

respectively, where

$$
v_{n, \mathrm{IS}}(u):=\frac{1}{n} \sum_{m=1}^{n}\left(L_{\tau_{u}}^{(m)}-p_{n}(u)\right)^{2} .
$$

We also compare the number of runs needed for the deterministic marks $\widehat{n}_{d}$ and the random marks $\widehat{n}_{r}$, with the associated Lundberg bounds, see Eqn. (3.71). Note that $\widehat{n}_{d}$ and $\widehat{n}_{r}$ vary per experiment; we remedy this by performing the entire procedure multiple times and taking the average of the obtained numbers. The twist parameter corresponding to deterministic marks is denoted by $\theta_{d}^{\star}$, its counterpart for random marks by $\theta_{r}^{\star}$. The numbers in Table 3.1 confirm that random marks consistently lead to higher ruin probabilities. In addition, as expected, the number of runs needed grows at a very modest pace (despite the fact that $p(u)$ decays essentially exponentially in $u$ ).

| $u$ | $e^{-\theta \theta_{d}^{*} u}$ | $p_{\widehat{n}_{d}}(u)$ | $\widehat{n}_{d}$ | $e^{-\theta_{r}^{*} u}$ | $p_{\widehat{n}_{r}}(u)$ | $\widehat{n}_{r}$ |
| :---: | :---: | :--- | :---: | :---: | :---: | :---: |
| 1 | $9.07 \cdot 10^{-1}$ | $3.15 \cdot 10^{-1}$ | 109 | $9.21 \cdot 10^{-1}$ | $3.32 \cdot 10^{-1}$ | 150 |
| 2 | $8.23 \cdot 10^{-1}$ | $2.65 \cdot 10^{-1}$ | 122 | $8.48 \cdot 10^{-1}$ | $2.69 \cdot 10^{-1}$ | 173 |
| 5 | $6.15 \cdot 10^{-1}$ | $1.52 \cdot 10^{-2}$ | 128 | $6.62 \cdot 10^{-1}$ | $1.75 \cdot 10^{-1}$ | 199 |
| 10 | $3.79 \cdot 10^{-1}$ | $7.89 \cdot 10^{-2}$ | 136 | $4.39 \cdot 10^{-1}$ | $8.45 \cdot 10^{-2}$ | 232 |
| 20 | $1.43 \cdot 10^{-1}$ | $2.28 \cdot 10^{-2}$ | 142 | $1.92 \cdot 10^{-1}$ | $2.68 \cdot 10^{-2}$ | 267 |
| 50 | $7.78 \cdot 10^{-3}$ | $8.89 \cdot 10^{-4}$ | 204 | $1.62 \cdot 10^{-2}$ | $1.64 \cdot 10^{-3}$ | 386 |
| 100 | $6.05 \cdot 10^{-5}$ | $5.83 \cdot 10^{-6}$ | 245 | $2.64 \cdot 10^{-4}$ | $2.18 \cdot 10^{-5}$ | 533 |
| 200 | $3.66 \cdot 10^{-9}$ | $3.49 \cdot 10^{-10}$ | 302 | $6.95 \cdot 10^{-8}$ | $4.70 \cdot 10^{-9}$ | 595 |

Table 3.1. Ruin probabilities, Lundberg bounds, and number of runs needed to reach precision of $\epsilon=0.05$, for deterministic and random marks. Chosen parameters are: $\bar{\lambda}_{1}=\bar{\lambda}_{2}=0.5, \alpha_{1}=2, \alpha_{2}=1.5, \mathbb{E} B_{11}=0.5, \mathbb{E} B_{12}=0.25, \mathbb{E} B_{21}=0.3, \mathbb{E} B_{22}=0.4$, $\mathbb{E} U_{11}=2, \mathbb{E} U_{12}=\mathbb{E} U_{21}=2.5, \mathbb{E} U_{22}=3$ and $r=8$. The twist parameters are $\theta_{d}^{\star}=0.097$ and $\theta_{r}^{\star}=0.082$.

In the remainder of this subsection, we consider the setting of random marks. In the next experiment we assess the computational advantage of the importance sampling estimator (using the measure $\mathbb{Q}$; indicated by subscript IS) over the conventional Monte Carlo estimator (using the measure $\mathbb{P}$; indicated by subscript MC). Our goal is to compare the time it takes for both estimators to get a sufficiently precise estimate
of $p(u)$. For the IS estimator, we use Eqn. (3.74). For the conventional MC estimator based on $n$ runs, denoted by $p_{n, \mathrm{MC}}(u)$, we have that

$$
\begin{equation*}
\epsilon_{n}=\frac{\sqrt{v_{n, \mathrm{MC}}(u)}}{p_{n, \mathrm{MC}}(u) \sqrt{n}} \approx \frac{1}{\sqrt{p_{n, \mathrm{MC}}(u) n}}, \tag{3.75}
\end{equation*}
$$

since the variance $v_{n, \mathrm{MC}}(u):=p_{n, \mathrm{MC}}(u)\left(1-p_{n, \mathrm{MC}}(u)\right) \approx p_{n, \mathrm{MC}}(u)$ for small probabilities.

In Table 3.2, we show the estimates of the ruin probabilities using MC and IS, including the average number of runs needed (in the table denoted by $\widehat{n}_{\text {MC }}$ and $\widehat{n}_{\text {IS }}$ ). As the absolute duration of each run highly depends on the specific hardware used, the programming language, the number of cores, etc., we decided to work with the speedup ratio, denoted by $\kappa$, which is the ratio of the simulation time needed under MC (to obtain the desired precision, that is) and its counterpart under IS.

While MC is actually more efficient for very low values of $u$, IS takes over already for moderate $u$. For $u$ larger than 60, it turned out even infeasible to obtain an MC estimate within a reasonable amount of time, whereas IS estimates can still be efficiently produced. For instance, by extrapolating the results we found for smaller values of $u$, MC would take approximately 18 hours for $u=70$ in our simulation environment; it would take even 1200 hours for $u=100$. In these cases, we estimated $\kappa$ by extrapolation of the running times under MC (growing effectively exponentially in $u$ ) and those under IS (growing effectively linearly in $u$ ); in the table these estimated values are given in italics. We conclude from the table that the speedup achieved by applying IS can be huge, particularly in the domain that ruin is rare; for $u=300$ the speedup is expected to be as high as $9.00 \cdot 10^{15}$. We also remark that our approach yields the same quantitative results as the approach taken in [73], since our choice of $\epsilon=0.05$ corresponds to their relative precision of $10 \%$ with respect to a $95 \%$ confidence interval (when approximating the quantile value 1.96 by 2 ).

### 3.7.2 Exceedance probability

In this subsection, we numerically illustrate the results of the simulation procedure proposed in Section 3.6.3. We consider the simulation-based computation of the bivariate exceedance probability

$$
q_{t}\left(a_{1}, a_{2}\right)=\mathbb{P}\left(\frac{Z_{1}(t)}{t} \geqslant a_{1}, \frac{Z_{2}(t)}{t} \geqslant a_{2}\right)
$$

where we assume $\left(a_{1}, a_{2}\right) \nless\left(\mu_{1}, \mu_{2}\right):=\lim _{t \rightarrow \infty}\left(Z_{1}(t) / t, Z_{2}(t) / t\right)$ to ensure that we are dealing with an event that becomes increasingly rare as $t \rightarrow \infty$. By Theorem 3.1,

$$
\begin{equation*}
\lim _{t \rightarrow \infty} \frac{1}{t} \log q_{t}\left(a_{1}, a_{2}\right)=-\inf _{\left(x_{1}, x_{2}\right) \in A} \Lambda^{\star}\left(x_{1}, x_{2}\right) \tag{3.76}
\end{equation*}
$$

| $u$ | $p_{n, \mathrm{MC}}(u)$ | $\widehat{n}_{\mathrm{MC}}$ | $p_{n, \mathrm{IS}}(u)$ | $v_{n, \mathrm{IS}}(u)$ | $\widehat{n}_{\mathrm{IS}}$ | $\kappa$ |
| :---: | :---: | ---: | :---: | ---: | :---: | :---: |
| 1 | $3.31 \cdot 10^{-1}$ | 807 | $3.34 \cdot 10^{-1}$ | $4.92 \cdot 10^{-2}$ | 150 | $4.62 \cdot 10^{-1}$ |
| 2 | $2.71 \cdot 10^{-1}$ | 1095 | $2.69 \cdot 10^{-1}$ | $3.47 \cdot 10^{-2}$ | 173 | $3.60 \cdot 10^{-1}$ |
| 3 | $2.39 \cdot 10^{-1}$ | 1357 | $2.37 \cdot 10^{-1}$ | $2.41 \cdot 10^{-2}$ | 187 | $3.22 \cdot 10^{-1}$ |
| 5 | $1.65 \cdot 10^{-1}$ | 2013 | $1.61 \cdot 10^{-1}$ | $1.61 \cdot 10^{-2}$ | 199 | $3.85 \cdot 10^{-1}$ |
| 10 | $8.40 \cdot 10^{-2}$ | 4368 | $8.45 \cdot 10^{-2}$ | $4.19 \cdot 10^{-3}$ | 232 | $5.89 \cdot 10^{-1}$ |
| 20 | $2.71 \cdot 10^{-2}$ | 14012 | $2.68 \cdot 10^{-2}$ | $4.92 \cdot 10^{-4}$ | 267 | $1.38 \cdot 10^{0}$ |
| 30 | $9.71 \cdot 10^{-3}$ | 38639 | $1.02 \cdot 10^{-2}$ | $7.29 \cdot 10^{-5}$ | 306 | $3.84 \cdot 10^{0}$ |
| 40 | $4.08 \cdot 10^{-3}$ | 96074 | $4.01 \cdot 10^{-3}$ | $1.16 \cdot 10^{-5}$ | 355 | $1.12 \cdot 10^{1}$ |
| 50 | $1.62 \cdot 10^{-3}$ | 263106 | $1.64 \cdot 10^{-3}$ | $2.44 \cdot 10^{-6}$ | 386 | $3.96 \cdot 10^{1}$ |
| 60 | $6.55 \cdot 10^{-4}$ | 747083 | $6.46 \cdot 10^{-4}$ | $4.36 \cdot 10^{-7}$ | 401 | $1.43 \cdot 10^{2}$ |
| 70 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $2.77 \cdot 10^{-4}$ | $8.45 \cdot 10^{-8}$ | 410 | $4.98 \cdot 10^{2}$ |
| 80 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $1.12 \cdot 10^{-4}$ | $1.64 \cdot 10^{-8}$ | 460 | $1.67 \cdot 10^{3}$ |
| 100 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $2.18 \cdot 10^{-5}$ | $5.98 \cdot 10^{-10}$ | 533 | $2.04 \cdot 10^{4}$ |
| 200 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $4.70 \cdot 10^{-9}$ | $3.51 \cdot 10^{-17}$ | 595 | $1.18 \cdot 10^{10}$ |
| 300 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $1.25 \cdot 10^{-12}$ | $2.71 \cdot 10^{-24}$ | 683 | $9.00 \cdot 10^{15}$ |

Table 3.2. Estimation of the ruin probability $p(u)$ for the marginal ruin process $Y_{1}(\cdot)$. Chosen parameters are as in the caption of Figure 3.1.
where $A=\left[a_{1}, \infty\right) \times\left[a_{2}, \infty\right)$. Denote the minimizer of the infimum on the RHS of (3.76) by $\boldsymbol{a}^{\star}=\left(a_{1}^{\star}, a_{2}^{\star}\right)$, which can be obtained using standard optimization techniques since $\Lambda^{*}(\cdot)$ is a convex function.

Consider estimating $q_{t}\left(a_{1}, a_{2}\right)$ by applying importance sampling. More precisely, with $I_{\left(a_{1}, a_{2}\right)}$ denoting the indicator of the event we are interested in, and $L_{t}$ the likelihood ratio given in Eqn. (3.66), we have

$$
\begin{equation*}
q_{t}\left(a_{1}, a_{2}\right)=\mathbb{E}_{\mathbb{Q}}\left[L_{t} I_{\left(a_{1}, a_{2}\right)}\right] . \tag{3.77}
\end{equation*}
$$

With $n$ denoting the number of runs, let $q_{t, n}\left(a_{1}, a_{2}\right)$ be the importance sampling estimator for $q_{t}\left(a_{1}, a_{2}\right)$ as defined in Eqn. (3.72). To find the twist parameter, we solve the optimization problem

$$
\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)=\left(\theta_{1}\left(a_{1}^{\star}, a_{2}^{\star}\right), \theta_{2}\left(a_{1}^{\star}, a_{2}^{\star}\right)\right)=\operatorname{argsup}_{\theta \in \mathscr{D}_{\Lambda}}\left(\boldsymbol{\theta}^{\top} \boldsymbol{a}^{\star}-\Lambda(\boldsymbol{\theta})\right) .
$$

Figure 3.3 illustrates the behavior of $q_{t}\left(a_{1}, a_{2}\right)$ as $t$ grows, converging to $-\inf _{\boldsymbol{x} \in A} \Lambda^{\star}(\boldsymbol{x})=$ $-\Lambda^{\star}(\boldsymbol{a})$, as was stated in Eqn. (3.76). We chose $a_{1}=10>3.90=\mu_{1}$ and $a_{2}=12>$ $4.76=\mu_{2}$, such that $\left\{Z_{1}(t) \geqslant a_{1} t, Z_{2}(t) \geqslant a_{2} t\right\}$ is an increasingly rare event as $t$ grows. For our specific parameters $\boldsymbol{\theta}\left(\boldsymbol{a}^{\star}\right)>\mathbf{0}$ (componentwise, that is), corresponding to $\boldsymbol{a}^{\star}=\boldsymbol{a}=\left(a_{1}, a_{2}\right)$. Intuitively, this means that the most probable way that the process $\left(Z_{1}(t) / t, Z_{2}(t) / t\right)_{t \in \mathbb{R}_{+}}$reaches the region $A=\left[a_{1}, \infty\right) \times\left[a_{2}, \infty\right)$ is a straight line from the origin to ( $a_{1}, a_{2}$ ).


Figure 3.3. Convergence of $\widehat{q}_{t}\left(a_{1}, a_{2}\right)$ to the logarithmic decay rate $-\Lambda^{\star}\left(\boldsymbol{a}^{\star}\right)$ for the bivariate process $\left(Z_{1}(t) / t, Z_{2}(t) / t\right)_{t \in \mathbb{R}_{+}}$. Chosen parameters are: $a_{1}=10$, $a_{2}=12, \bar{\lambda}_{1}=\bar{\lambda}_{2}=0.5$, $\alpha_{1}=2, \alpha_{2}=1.5, \mathbb{E} B_{11}=0.5, \mathbb{E} B_{12}=0.25, \mathbb{E} B_{21}=0.3, \mathbb{E} B_{22}=0.4, \mathbb{E} U_{11}=2$, $\mathbb{E} U_{12}=\mathbb{E} U_{21}=2.5, \mathbb{E} U_{22}=3$. Since $\boldsymbol{a}^{\star}=\boldsymbol{a}$, the decay rate is $-\Lambda^{\star}(\boldsymbol{a})=-0.276$ and the twist parameter is $\boldsymbol{\theta}^{\star}(\boldsymbol{a})=(0.0376,0.0256)$.

Next, we quantify the computational advantage of the IS estimator over the conventional MC estimator, using the same approach as the one underlying Table 3.2. As before, we run the simulation for both methods until the relative standard error $\epsilon_{n}$ is under our desired precision $\epsilon$. Table 3.3 shows the comparison between MC and IS, for different values of $t$. When increasing $t$, MC becomes infeasible due to the very steeply increasing number of runs needed as well as the simulation time needed per run. Already at $t=15$, it would take approximately 69 hours in our simulation environment. IS, however, remains feasible, even in the domain of extremely small probabilities. Note that the number of runs needed for IS does not increase in a monotone fashion, which is due to the fact that for small $t$ the process is not yet in the regime where the exceedance event is rare. We also note the speedup ratio $\kappa$ of the exceedance probabilities increases more steeply (in $t$ ) than that of the ruin probabilities (in $u$ ).

### 3.8 Concluding Remarks

This paper has studied large deviations of multivariate compound Hawkes processes, with the underlying Hawkes process having a general decay function and random marks. In order to prove the LDP, the main technical hurdle concerned proving that the limiting cumulant is steep. Our steepness proof is methodologically novel, in that we manage to show that the derivative of the cumulant grows to infinity when ap-

| $t$ | $q_{t, \mathrm{MC}}(\boldsymbol{a})$ | $\widehat{n}_{\mathrm{MC}}$ | $q_{t, \mathrm{IS}}(\boldsymbol{a})$ | $v_{t, \mathrm{IS}}(\boldsymbol{a})$ | $\widehat{n}_{\mathrm{IS}}$ | $\kappa$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | $2.58 \cdot 10^{-2}$ | 16620 | $2.61 \cdot 10^{-2}$ | $8.27 \cdot 10^{-3}$ | 4879 | $6.56 \cdot 10^{-1}$ |
| 2 | $1.83 \cdot 10^{-2}$ | 21983 | $1.78 \cdot 10^{-2}$ | $3.47 \cdot 10^{-3}$ | 4378 | $9.98 \cdot 10^{-1}$ |
| 3 | $1.45 \cdot 10^{-2}$ | 29076 | $1.42 \cdot 10^{-2}$ | $1.92 \cdot 10^{-3}$ | 3800 | $1.63 \cdot 10^{0}$ |
| 5 | $7.88 \cdot 10^{-3}$ | 50869 | $7.95 \cdot 10^{-3}$ | $5.67 \cdot 10^{-4}$ | 3589 | $3.66 \cdot 10^{0}$ |
| 10 | $1.75 \cdot 10^{-3}$ | 205934 | $1.69 \cdot 10^{-3}$ | $2.73 \cdot 10^{-5}$ | 3818 | $3.04 \cdot 10^{1}$ |
| 15 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $3.64 \cdot 10^{-4}$ | $1.40 \cdot 10^{-6}$ | 4223 | $3.24 \cdot 10^{2}$ |
| 20 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $7.83 \cdot 10^{-5}$ | $7.10 \cdot 10^{-8}$ | 4631 | $3.92 \cdot 10^{3}$ |
| 25 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $1.87 \cdot 10^{-5}$ | $4.48 \cdot 10^{-9}$ | 5123 | $5.04 \cdot 10^{4}$ |
| 30 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $3.84 \cdot 10^{-6}$ | $2.23 \cdot 10^{-10}$ | 6043 | $6.25 \cdot 10^{5}$ |
| 40 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $2.18 \cdot 10^{-7}$ | $8.05 \cdot 10^{-13}$ | 6749 | $1.28 \cdot 10^{8}$ |
| 50 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $1.15 \cdot 10^{-8}$ | $2.69 \cdot 10^{-15}$ | 8209 | $3.43 \cdot 10^{10}$ |
| 75 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $9.25 \cdot 10^{-12}$ | $2.16 \cdot 10^{-24}$ | 10106 | $3.93 \cdot 10^{16}$ |
| 100 | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $6.31 \cdot 10^{-15}$ | $1.44 \cdot 10^{-27}$ | 11926 | $5.56 \cdot 10^{22}$ |

Table 3.3. Estimation of the exceedance probability $q_{t}(\boldsymbol{a})$. Chosen parameters are as in the caption of Figure 3.3
proaching the boundary of its domain, but, remarkably, without having an explicit characterization of this domain. Then the logarithmic asymptotics of the corresponding ruin probability are identified, the proof using (in the lower bound) the LDP. In addition, the logarithmic asymptotics of the multivariate net cumulative claim attaining a rare value (at a given point in time, that is) are established. The final contribution of the paper concerns importance sampling based rare event simulation procedures, shown to be asymptotically efficient.

An interesting topic for future research is to consider other types of deviations for multivariate Hawkes processes, such as precise or process-level large deviations. Another interesting problem is to derive large deviations results for non-linear multivariate Hawkes processes. Since the non-linear Hawkes process is not a branching process, this setting requires a genuinely different type of analysis.

## 3.A Steepness in the Univariate Case

In this appendix, we prove steepness of the limiting cumulant in the univariate case. The appendix has two objectives. First, in this single-dimensional setting, elements that look intricate in the proof of Theorem 3.1 now simplify and become significantly more transparent; indeed, this univariate proof helps the reader navigating the proof that we gave for the multivariate case. Second, our proof is of a generic nature, in that
it does not rely on the fact that in this univariate setting the distribution of the cluster size is explicitly known. This distinguishes our approach from the one followed in [69], which explicitly uses that the cluster size has a Borel distribution. In the multivariate case, the explicit distribution of the joint cluster size is unknown, thus prohibiting the approach of [69].

Consider the case of $d=d^{\star}=1$; we leave out all indices. Let $S$ denote the total number of events in a cluster. We know that

$$
\begin{equation*}
\Lambda(\theta):=\lim _{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta Z(t)}\right]=\bar{\lambda}\left(\mathbb{E}\left[m_{U}(\theta)^{S}\right]-1\right) \tag{3.78}
\end{equation*}
$$

where $m_{U}(\theta)=\mathbb{E}\left[e^{\theta U}\right]$ is the moment generating function of $U$ and $\bar{\lambda}$ is the base rate. To guarantee stability of the Hawkes process, we assume the univariate counterpart of Assumption 3.1, that is,

$$
\begin{equation*}
\mu=\mathbb{E}[B] c=\mathbb{E}[B] \int_{0}^{\infty} g(v) \mathrm{d} v \in(0,1) . \tag{3.79}
\end{equation*}
$$

In this univariate case, the probability generating function $f(\cdot)$ of the cluster size $S$ satisfies the fixed-point representation

$$
\begin{equation*}
f(z)=z \mathbb{E}\left[e^{B c(f(z)-1)}\right] \tag{3.80}
\end{equation*}
$$

First, we take derivatives on both sides of Eqn. (3.80) to obtain

$$
f^{\prime}(z)=\mathbb{E}\left[e^{B c(f(z)-1)}\right]+z f^{\prime}(z) \mathbb{E}\left[B c e^{B c(f(z)-1)}\right]
$$

which, due to (3.80), can be rewritten as

$$
f^{\prime}(z)=\frac{f(z) / z}{1-b(z)}
$$

where $b(z):=z \mathbb{E}\left[B c e^{B c(f(z)-1)}\right]$. The next step is to show that there exists $\widehat{z} \geqslant 1$ such that $b(\widehat{z})=1$. The domain $\mathscr{D}_{f}$ is obtained from Proposition 3.1 when $d=1$ as $\mathscr{D}_{f}=[0, \widehat{z}]$, where $\widehat{z}>1$ is given by

$$
\begin{equation*}
\widehat{z}=\mathbb{E}\left[B c e^{B c(\widehat{x}-1)}\right]^{-1}, \tag{3.81}
\end{equation*}
$$

and where $\widehat{x}>1$ solves the equation

$$
\begin{equation*}
x \mathbb{E}\left[B c e^{B c(x-1)}\right]=\mathbb{E}\left[e^{B c(x-1)}\right], \tag{3.82}
\end{equation*}
$$

see also [54, Theorem 3.1.1]. Since $f^{\prime}(z)$ is well-defined for $0<z<\widehat{z}$, we consider what happens when we approach the boundary $\widehat{z}$. We compute, using Eqns. (3.81) and (3.82) and the fact $f(\widehat{z})=\widehat{x}$, that

$$
\begin{equation*}
b(\widehat{z})=\widehat{z} \mathbb{E}\left[B c e^{B c(\widehat{x}-1)}\right]=1 \tag{3.83}
\end{equation*}
$$

## Chapter 3 Multivariate Compound Hawkes Processes

Hence, we obtain

$$
\begin{equation*}
\lim _{z \uparrow \widehat{z}} f^{\prime}(z)=\lim _{z \uparrow \widehat{z}} \frac{f(z) / z}{1-b(z)}=\infty \tag{3.84}
\end{equation*}
$$

also noting that for the numerator $f(\widehat{z}) / \widehat{z}=\widehat{x} / \widehat{z}>0$.
Concerning the steepness of $\Lambda(\cdot)$, let there be a $\widehat{\theta}>0$ such that $m_{U}(\widehat{\theta})=\widehat{z}$, which is the univariate counterpart of Assumption 3.3. We then have that

$$
\begin{equation*}
\underset{\theta \uparrow \widehat{\theta}}{\liminf } \Lambda^{\prime}(\theta) \geqslant \bar{\lambda} \mathbb{E}[U] \mathbb{E}\left[S m_{U}(\widehat{\theta})^{S-1}\right] \geqslant \bar{\lambda} \mathbb{E}[U] f^{\prime}(\widehat{z})=\infty \tag{3.85}
\end{equation*}
$$

where the first inequality is due to Fatou's lemma, and the second inequality is due to $m_{U}^{\prime}(\theta) \geqslant \mathbb{E}[U]$ for any $\theta \geqslant 0$.

## Chapter 4

## Markovian Multivariate Hawkes Population Processes:

Efficient Evaluation of Moments


#### Abstract

We provide probabilistic and computational results on Markovian multivariate Hawkes processes and induced population processes. By applying the Markov property, we characterize in closed form a joint transform, bijective to the probability distribution, of the population process and its underlying intensity process. We demonstrate a method that exploits the transform to obtain analytic expressions for transient and stationary multivariate moments up to arbitrary order, as well as auto- and crosscovariances. We reveal a nested sequence of block matrices that yields the moments in explicit form and brings important computational advantages. We also establish the asymptotic behavior of the intensity of the multivariate Hawkes process in its nearly unstable regime, under a specific parameterization. In extensive numerical experiments, we analyze the computational complexity, accuracy and efficiency of the established results.


## Chapter 4

## Markovian Multivariate Hawkes Population Processes: Efficient Evaluation of Moments

### 4.1 Introduction

Since the onset of globalization, the mechanisms according to which events spread have become increasingly complex. Events involving disease contagion, financial panic, news that goes viral, are all subject to forms of propagation that occur through time and space, be it across human populations, equity markets, or news outlets. A multitude of mathematical models have been proposed to describe the corresponding underlying dynamics.

Multivariate point processes constitute one such class of models that describe the random nature of the arrival, and subsequent spread, of events, in the time as well as space domain. In particular, the subclass of multivariate Hawkes processes ([45, 46]) provides a rich structure that is capable of capturing contagious dynamics. These Hawkes processes allow for flexible dependencies of events, due to the inherent feedback mechanisms known as self- and cross-excitation. Recently, the Hawkes process has been increasingly used as the arrival process of infinite-server queues ([26, 57]), in which arrived events at some point leave the system. Such infinite-server queues can be seen as population processes, where individuals are born (i.e., arrive) according to a Hawkes process and die (i.e., leave) after a random time. Throughout this paper, we use the terminology of infinite-server queues and population processes interchangeably.

A clear and relevant motivation to study a Hawkes-fed infinite-server queue is to account for infected and recovered subjects in a population. Hawkes processes have been widely applied in epidemiology, in conjunction with SIR models ([67]) and dynamic contagion models, e.g., in the context of the COVID-19 pandemic ([16]). An impor-

## Chapter 4 Markovian Multivariate Hawkes Population Processes

tant strand of research considers a multivariate framework to allow for cross-exciting dynamics between subpopulations (for instance, residing at different geographic locations). Also in various other domains this type of model offers a natural framework, for example when considering the number of simultaneous online visitors of a specific website or to describe social interaction. Multivariate Hawkes processes have recently gained interest in the context of modeling customer support contact centers ([25]).

When analyzing infinite-server queues with multivariate Hawkes input, the main challenge lies in unraveling, and tractably representing, its inherently complex probabilistic structure. In this paper our main objective is to devise techniques to efficiently and accurately evaluate moments. We consider the Markovian case, i.e., the case in which the excitation functions are exponentially decaying; see ([45, 62]). The Markov property is used to obtain a characterization of the transform of the joint process, that is, the Hawkes intensity process and the infinite-server queue, in terms of a system of differential equations.

Contributions. This paper makes five contributions. We start by deriving in closed form a joint transform - bijective to the joint probability distribution - of the multivariate population process and its underlying intensity process, exploiting their Markovian nature. We allow for general distributions of the intensity jump sizes and allow the processes to be evaluated at potentially multiple future time instances, thus characterizing all cross-sectional and temporal probabilistic features. Second, we employ this joint transform to obtain explicit, recursive expressions for both the transient and stationary multivariate cross-moments. We thus extend earlier results pertaining to the case that the underlying Hawkes process is single-dimensional ([24, 26]). Third, we show that the higher-order transient and stationary moments can be obtained in closed form from a nested sequence of block matrices, having important computational advantages. Fourth, for a specifically chosen set of parameters, we establish the limiting behavior of the underlying intensity process of the multivariate population process in the practically relevant, nearly unstable situation, where the stability condition is close to being violated; cf. the heavy-traffic regime in queueing theory.

The final contribution concerns the use of our analytic findings when devising efficient computational techniques for evaluating moments. To this end, we analyze in numerical experiments the methods developed in this paper and assess their accuracy and efficiency. The methods we develop show superior performance in terms of speed and accuracy when compared to the computational alternative of applying finite difference schemes of the joint transform. Our computational approach is fast and accurate: it provides near-instant response that is exact up to machine precision. When using the nested block matrices, it has the attractive feature that the computation speed does
not depend on the value of the considered time horizon. Moreover, when compared to the simulation alternative, using a large number of Monte Carlo simulation runs to approximate the object of interest, the performance gain is even orders of magnitude larger. The computer codes that implement the methods developed in this paper are available at https://github.com/RaviarKarim/HawkesMarkov.

In the computation of the transient and stationary moments, we focus on two settings, which are in a sense each others dual. Namely the bivariate setting with moments of arbitrary order and the multivariate model with moments up to second order. The generic recursive structure for the transient and stationary moments is similar. However, in the transient setting, we obtain recursive systems of non-homogeneous linear differential equations, whereas in the stationary setting, we obtain recursive systems of linear algebraic equations.

An important application that our results make possible, is moment-based estimation of multivariate Hawkes processes. Such an estimation approach requires evaluating a collection of moments, auto-covariances and cross-covariances a (very) large number of times, for the parameter vector proposed at each new iteration of the optimization routine. Existing approaches to evaluating moments and other distributional characteristics are computationally prohibitively expensive for this purpose. One might employ the explicit approximate moments, obtained from the infinitesimal Markov generator using operator methods and Taylor expansions applied to short time intervals, as in [2]. This approach, however, requires data sampled at least at the daily frequency. Our approach is based on exact expressions of the moments, autocovariances and cross-covariances over time intervals of arbitrary length, and their numerical evaluation remains fast and accurate. Another advantage of our approach is that it makes comparative statics possible, i.e., it provides an efficient, tractable link between the objects of interest and the parameters of the process.

Organization. This paper is organized as follows. In Section 4.2, we introduce the Markovian multivariate Hawkes process and the induced population process. In Section 4.3, we derive a characterization of the joint transform using the Markov property of the process. Subsequently, we use this result to obtain relations for the transient and stationary moments. In Section 4.4, we show the underlying recursive structure in the computation of moments of certain order and dimension. In Section 4.5, we focus on the bivariate setting, revealing a nested block structure of matrices that characterize the moments, enabling fast computation. In Section 4.6, we analyze the nearly unstable case for the intensity of the process. Section 4.7 provides our numerical experiments. Conclusions are in Section 4.8 and some proofs are relegated to the Appendix.

### 4.2 Multivariate Hawkes Populations

In this section we define the multivariate Hawkes process, by means of the associated conditional intensity function, as well as its induced population process. Hawkes processes, first introduced in a series of papers [45, 46, 47], are a class of point processes that exhibit self-exciting behavior, in the sense that the current value of the associated intensity function depends on the history of the point process.

In this paper, we focus on multivariate Hawkes processes of the Markovian type. Consider a $d$-dimensional point process $\boldsymbol{N}(\cdot) \equiv(\boldsymbol{N}(t))_{t \in \mathbb{R}_{+}}=\left(\left(N_{1}(t), \ldots, N_{d}(t)\right)\right)_{t \in \mathbb{R}_{+}}^{\top}$, which records the number of points in each component $N_{i}(t)$, with $i \in[d]:=\{1, \ldots, d\}$, up to and including time $t$. It is well known that a point process is characterized by its conditional intensity function $\boldsymbol{\lambda}(t)=\left(\lambda_{1}(t), \ldots, \lambda_{d}(t)\right)^{\top}$, see [23, Chapter 7].

Definition 4.1. A Markovian multivariate Hawkes process is a point process $\boldsymbol{N}(\cdot)$, with $\boldsymbol{N}(0)=\mathbf{0}$, whose components $N_{i}(\cdot)$ satisfy

$$
\begin{align*}
& \mathbb{P}\left(N_{i}(t+\Delta)-N_{i}(t)=0 \mid \mathcal{F}_{t}\right)=1-\lambda_{i}(t) \Delta+o(\Delta), \\
& \mathbb{P}\left(N_{i}(t+\Delta)-N_{i}(t)=1 \mid \mathcal{F}_{t}\right)=\lambda_{i}(t) \Delta+o(\Delta),  \tag{4.1}\\
& \mathbb{P}\left(N_{i}(t+\Delta)-N_{i}(t)>1 \mid \mathcal{F}_{t}\right)=o(\Delta),
\end{align*}
$$

as $\Delta \downarrow 0$, with $\mathcal{F}_{t}=\sigma(\boldsymbol{N}(s): s \leqslant t)$ denoting the natural filtration generated by $\boldsymbol{N}(\cdot)$. Here, each component $\lambda_{i}(t)$ of the intensity function satisfies the mean-reverting dynamics

$$
\begin{equation*}
\mathrm{d} \lambda_{i}(t)=\alpha_{i}\left(\bar{\lambda}_{i}-\lambda_{i}(t)\right) \mathrm{d} t+\sum_{j=1}^{d} B_{i j}(t) \mathrm{d} N_{j}(t) \tag{4.2}
\end{equation*}
$$

where $\lambda_{i}(0)=\bar{\lambda}_{i} \geqslant 0, \alpha_{i} \geqslant 0$, and for each combination of $i, j \in[d],\left(B_{i j}(t)\right)_{t}$ is a sequence of independent random variables, distributed as the generic non-negative random variable $B_{i j}$.

The intuitive explanation behind Definition 4.1 goes as follows. The constants $\bar{\lambda}_{i}$ are referred to as the base rates. When a point is generated in component $j \in[d], N_{j}(t)$ increases by one and makes intensity $\lambda_{i}(t)$, for all $i \in[d]$, jump by a value $B_{i j}(t)$ that is distributed as the random variable $B_{i j}$. This jump in the intensity $\lambda_{i}(t)$ caused by $B_{i j}(t)$ increases the probability of new points being generated in component $i$, thereby more likely to increase $N_{i}(t)$. After the jump has occurred, the intensity $\lambda_{i}(t)$ decays exponentially with rate $\alpha_{i}$ back to the base rate $\bar{\lambda}_{i}$. These jumps in intensities and the subsequent decay is what makes points cluster across time and space. That is, for any $j \in[d]$, when points in $N_{j}(t)$ cause $\lambda_{j}(t)$ to jump, there is a pure temporal effect
and we speak of self-excitation, while an effect on $\lambda_{i}(t)$, with $i \neq j$, has an additional spatial effect and we speak of cross-excitation.

It is well-known that with exponential decay the joint process $(\boldsymbol{N}(\cdot), \boldsymbol{\lambda}(\cdot))$ is a Markov process, see e.g. [58, 60, 62]. Furthermore, by applying Itô's Lemma to $f\left(t, \lambda_{i}(t)\right)=e^{\alpha_{i} t} \lambda_{i}(t)$, Eqn. (4.2) can alternatively be expressed as

$$
\begin{equation*}
\lambda_{i}(t)=\bar{\lambda}_{i}+\sum_{j=1}^{d} \int_{0}^{t} B_{i j}(s) e^{-\alpha_{i}(t-s)} \mathrm{d} N_{j}(s), \tag{4.3}
\end{equation*}
$$

where we used that $\boldsymbol{N}(0)=\mathbf{0}$, implying $\lambda_{i}(0)=\bar{\lambda}_{i}$. The exponential term in Eqn. (4.3) can be defined in a function $g_{i}(t):=e^{-\alpha_{i} t}$, known in the literature as the decay function. We emphasize that only exponentially decaying $g_{i}(\cdot)$ render the joint process $(\boldsymbol{N}(\cdot), \boldsymbol{\lambda}(\cdot))$ a Markov process.

To ensure stability of the multivariate Hawkes process, $[46,12]$ show that we must impose a stability condition. In what follows, we assume that the stability condition applies.

Assumption 4.1. With $\rho(\cdot)$ denoting the spectral radius of a matrix, assume that $\rho(\boldsymbol{H})<1$, where the matrix $\boldsymbol{H}=\left(h_{i j}\right)_{i, j \in[d]}$ has elements

$$
\begin{equation*}
h_{i j} \equiv \mathbb{E}\left[B_{i j}\right] \int_{0}^{\infty} e^{-\alpha_{i} t} \mathrm{~d} t=\mathbb{E}\left[B_{i j}\right] / \alpha_{i} . \tag{4.4}
\end{equation*}
$$

In the model discussed in this paper, the Markovian multivariate Hawkes process will serve as the arrival process of a multivariate population process. Our overarching goal is to compute various quantities pertaining to the (joint) distribution of the arrival process and the population process. A similar setup has been considered in [26, 57], where the univariate Hawkes process serves as the arrival process for an infinite-server queue in which arriving customers reside for independent and identically distributed (i.i.d.) amounts of time (often assumed exponentially distributed). The following definition generalizes that framework to the multivariate setting.

Definition 4.2. Let $\boldsymbol{N}(\cdot)$ be a Markovian multivariate Hawkes process as given in Definition 4.1. Define the associated Hawkes population process $\boldsymbol{Q}(\cdot)$, with $\boldsymbol{Q}(0)=\mathbf{0}$, by setting for each component $Q_{i}(t)$

$$
\begin{equation*}
Q_{i}(t):=\int_{0}^{t} \mathbf{1}_{\left\{E_{i}(s)>t-s\right\}} \mathrm{d} N_{i}(s) \tag{4.5}
\end{equation*}
$$

for any $t \geqslant 0$ and where $\left(E_{i}(s)\right)_{s}$ is a sequence of independent random variables, exponentially distributed with parameter $\mu_{i}$, also independent of the multivariate arrival process $\boldsymbol{N}(t)$.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

The above definition entails that each arrival in component $i \in[d]$ remains in the system for an exponentially distributed amount of time. In demography, one can think of $Q_{i}(\cdot)$ as the number of individuals in a subpopulation $i \in[d]$, where each $E_{i}(s)$ models the lifetime of an individual in subpopulation $i$. In epidemiology, $Q_{i}(\cdot)$ would represent the number of individuals infected by a disease at location $i \in[d]$, where $E_{i}(s)$ would model the duration from infection to recovery (or death). In the queuing literature, this can be interpreted as a special case of an infinite-server queue. We note that if $\mu_{i} \equiv 0$, then no points ever depart from component $i$, and hence $Q_{i}(\cdot) \equiv N_{i}(\cdot)$. However, in some of the expressions we will encounter in this paper one must take proper care of taking the limit $\mu_{i} \downarrow 0$, so as to avoid dividing by zero; note that if $\mu_{i}=0$, then $Q_{i}(\cdot)$ eventually grows unbounded.

In this paper, we consider Markovian multivariate Hawkes processes and associated population processes. We conclude this section by briefly discussing various more general variants. In the first place, one can introduce decay functions $g_{i j}(\cdot)$ that depend on both the origin component $j$ and the destination component $i$. Moreover, where in this work we exclusively focus on exponential decay functions, one can work with more general non-increasing and integrable decay functions, thus also leaving the Markovian setting. Our analysis fully depends on 'Markovian reasoning', and can therefore not be used in this non-Markovian case. Instead, the process may be analyzed through the cluster representation, first described in [47], where the Hawkes process is described as a Poisson cluster process. This approach has been followed in [55] to study distributional properties in the multivariate setting. Another extension concerns the nonlinear case, where the intensity depends on an auxiliary function $h_{i}: \mathbb{R} \rightarrow \mathbb{R}$; see e.g., $[12,74]$. In this nonlinear setting, the Hawkes process is not even a Poisson cluster process and requires different methods of analysis entirely.

The application of our methodology to the class of multivariate Hawkes population processes may be viewed as a "proof of principle". Our methodology can, in principle, be applied to general multivariate Markov processes.

### 4.3 Transform and Joint Moments

The objective of this section is to analyze the joint transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ for any fixed $t \in \mathbb{R}_{+}$. To this end, define, for given initial values $\boldsymbol{Q}\left(t_{0}\right)=\boldsymbol{Q}_{0} \in \mathbb{N}_{0}^{d}$ and $\boldsymbol{\lambda}\left(t_{0}\right)=\boldsymbol{\lambda}_{0} \in \mathbb{R}_{+}^{d}$ for some $0 \leqslant t_{0}<t$, the conditional joint transform

$$
\begin{equation*}
\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})=\mathbb{E}_{t_{0}}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t)} e^{-s_{i} \lambda_{i}(t)}\right]:=\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t)} e^{-s_{i} \lambda_{i}(t)} \mid \boldsymbol{Q}\left(t_{0}\right)=\boldsymbol{Q}_{0}, \boldsymbol{\lambda}\left(t_{0}\right)=\boldsymbol{\lambda}_{0}\right], \tag{4.6}
\end{equation*}
$$

where $t \geqslant 0, s \in \mathbb{R}_{+}^{d}$ and $\boldsymbol{z} \in[-1,1]^{d}$. In the specific case that $t_{0}=0$, with the assumed initial conditions $\boldsymbol{Q}(0)=\mathbf{0}$ and $\boldsymbol{\lambda}(0)=\overline{\boldsymbol{\lambda}}$, we set

$$
\begin{align*}
\zeta(t, \boldsymbol{s}, \boldsymbol{z}) \equiv \zeta_{0}(t, \boldsymbol{s}, \boldsymbol{z}): & =\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t)} e^{-s_{i} \lambda_{i}(t)}\right] \\
& =\mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t)} e^{-s_{i} \lambda_{i}(t)} \mid \boldsymbol{Q}(0)=\mathbf{0}, \boldsymbol{\lambda}(0)=\overline{\boldsymbol{\lambda}}\right] \tag{4.7}
\end{align*}
$$

where the expectation operator $\mathbb{E}[\cdot]$ is understood as the conditional $\mathbb{E}_{0}[\cdot]$.

### 4.3.1 Transform characterization

The following theorem identifies $\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})$.
Theorem 4.1. Fix $t \in \mathbb{R}_{+}$, and assume $\boldsymbol{Q}\left(t_{0}\right)=\boldsymbol{Q}_{0} \in \mathbb{N}_{0}^{d}$ and $\boldsymbol{\lambda}\left(t_{0}\right)=\boldsymbol{\lambda}_{0} \in \mathbb{R}_{+}^{d}$ for some $0 \leqslant t_{0}<t$. Then, for any $\boldsymbol{z} \in[-1,1]^{d}, \boldsymbol{s} \in \mathbb{R}_{+}^{d}$,

$$
\begin{equation*}
\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})=\prod_{j=1}^{d} \widehat{z}_{j}\left(t_{0}\right)^{Q_{j, 0}} \exp \left(-\tilde{s}_{j}(t) \lambda_{j, 0}-\bar{\lambda}_{j} \alpha_{j} \int_{t_{0}}^{t} \tilde{s}_{j}(u) \mathrm{d} u\right), \tag{4.8}
\end{equation*}
$$

where, for $t_{0} \leqslant u \leqslant t$ and $j \in[d]$, the functions $\widehat{z}_{j}(\cdot)$ and $\tilde{s}_{j}(\cdot)$ satisfy

$$
\begin{align*}
& \widehat{z}_{j}(u)=1+\left(z_{j}-1\right) e^{-\mu_{j}(t-u)}, \\
& \frac{\mathrm{d} \tilde{s}_{j}(u)}{\mathrm{d} u}+\alpha_{j} \tilde{s}_{j}(u)+\left(1+\left(z_{j}-1\right) e^{-\mu_{j}\left(u-t_{0}\right)}\right) \beta_{j}(\tilde{\boldsymbol{s}}(u))-1=0, \tag{4.9}
\end{align*}
$$

with boundary condition $\tilde{s}_{j}\left(t_{0}\right)=s_{j}$, and $\beta_{j}(\boldsymbol{s}):=\mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right]=\mathbb{E}\left[\exp \left(-\sum_{i=1}^{d} s_{i} B_{i j}\right)\right]$. Proof. See Appendix 4.A.

Corollary 4.1. Fix $t \in \mathbb{R}_{+}$, and assume $\boldsymbol{Q}(0)=\mathbf{0}$ and $\boldsymbol{\lambda}(0)=\overline{\boldsymbol{\lambda}}$. Then, for any $z \in[-1,1]^{d}, s \in \mathbb{R}_{+}^{d}$,

$$
\begin{equation*}
\zeta(t, \boldsymbol{s}, \boldsymbol{z})=\prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \tilde{s}_{j}(t)-\bar{\lambda}_{j} \alpha_{j} \int_{0}^{t} \tilde{s}_{j}(v) \mathrm{d} v\right) \tag{4.10}
\end{equation*}
$$

where, for $j \in[d]$ and $0 \leqslant v \leqslant t$, the functions $\tilde{s}_{j}(\cdot)$ satisfy

$$
\begin{equation*}
\frac{\mathrm{d} \tilde{s}_{j}(v)}{\mathrm{d} v}+\alpha_{j} \tilde{s}_{j}(v)+\left(1+\left(z_{j}-1\right) e^{-\mu_{j} v}\right) \beta_{j}(\tilde{\boldsymbol{s}}(v))-1=0 \tag{4.11}
\end{equation*}
$$

with boundary condition $\tilde{s}_{j}(0)=s_{j}$.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

Using the expression presented in Theorem 4.1, in principle any conditional joint moment can be obtained. Indeed, for any $n_{\lambda_{i}}, n_{Q_{i}} \in \mathbb{N}_{0}$, we have

$$
\begin{equation*}
\left.\frac{\mathrm{d}^{n_{\lambda_{1}}}}{\mathrm{~d} s_{1}^{n_{1}}} \cdots \frac{\mathrm{~d}^{n_{\lambda_{d}}}}{\mathrm{~d} s_{d}^{n_{d}}} \frac{\mathrm{~d}^{n_{Q_{1}}}}{\mathrm{~d} z_{1}^{n_{Q_{1}}}} \cdots \frac{\mathrm{~d}^{n_{Q_{d}}}}{\mathrm{~d} z_{d}^{n_{Q_{d}}}} \zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})\right|_{\substack{s=0 \\ \boldsymbol{z}=1}}=\mathbb{E}_{t_{0}}\left[\prod_{i=1}^{d}(-1)^{n_{\lambda_{i}}} \lambda_{i}(t)^{n_{\lambda_{i}}} Q_{i}(t)^{\left[n_{Q_{i}}\right]}\right], \tag{4.12}
\end{equation*}
$$

i.e., an object composed from (standard) moments of $\lambda_{i}(t)$ and reduced moments of $Q_{i}(t)$. Note that in (4.12) we have used the standard Pochhammer notation: for integers $m$ and $n$ we denote $m^{[n]}:=m(m-1) \cdots(m-n+1)$, by convention setting $m^{[0]}:=1$.

In the above, we focused on identifying transforms pertaining to a single point in time. Using similar methods, however, it is possible to derive the transform of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ and $(\boldsymbol{Q}(t+\tau), \boldsymbol{\lambda}(t+\tau))$ jointly, for some $\tau>0$. More precisely, for $\boldsymbol{y}, \boldsymbol{z} \in[-1,1]^{d}$ and $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}_{+}^{d}$, we consider the object

$$
\begin{equation*}
\zeta_{\tau}(t, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{z}):=\mathbb{E}\left[\prod_{i=1}^{d} y_{i}^{Q_{i}(t)} e^{-r_{i} \lambda_{i}(t)} z_{i}^{Q_{i}(t+\tau)} e^{-s_{i} \lambda_{i}(t+\tau)}\right] \tag{4.13}
\end{equation*}
$$

where as before, $\mathbb{E}[\cdot]$ is understood as $\mathbb{E}_{0}[\cdot]$, with $\boldsymbol{Q}(0)=\mathbf{0}$ and $\boldsymbol{\lambda}(0)=\overline{\boldsymbol{\lambda}}$. In addition, $\boldsymbol{y} \odot \boldsymbol{z}$ is the component-wise product of the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$.

Theorem 4.2. Fix $t, \tau \in \mathbb{R}_{+}$, and assume $\boldsymbol{Q}(0)=\mathbf{0}$ and $\boldsymbol{\lambda}(0)=\overline{\boldsymbol{\lambda}}$. Then, for any $\boldsymbol{y}, \boldsymbol{z} \in[-1,1]^{d}, \boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}_{+}^{d}$,

$$
\begin{align*}
\zeta_{\tau}(t, \boldsymbol{r}, \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{z}) & =\zeta(t, \boldsymbol{y} \odot \widehat{\boldsymbol{z}}(t), \boldsymbol{r}+\tilde{\boldsymbol{s}}(t+\tau)) \prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \alpha_{j} \int_{t}^{t+\tau} \tilde{s}_{j}(u) \mathrm{d} u\right) \\
& =\prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \tilde{r}_{j}(t)-\bar{\lambda}_{j} \alpha_{j} \int_{0}^{t} \tilde{r}_{j}(v) \mathrm{d} v-\bar{\lambda}_{j} \alpha_{j} \int_{t}^{t+\tau} \tilde{s}_{j}(u) \mathrm{d} u\right) \tag{4.14}
\end{align*}
$$

where $\zeta(\cdot)$ is given by Eqn. (4.10), and where for $j \in[d]$, the functions $\widehat{z}_{j}(\cdot)$, $\tilde{s}_{j}(\cdot)$ and $\tilde{r}_{j}(\cdot)$ satisfy

$$
\begin{align*}
& \widehat{z}_{j}(u)=1+\left(z_{j}-1\right) e^{-\mu_{j}(t+\tau-u)} \\
& \frac{\mathrm{d} \tilde{s}_{j}(u)}{\mathrm{d} u}+\alpha_{j} \tilde{s}_{j}(u)+\left(1+\left(z_{j}-1\right) e^{-\mu_{j}(u-t)}\right) \beta_{j}(\tilde{\boldsymbol{s}}(u))-1=0  \tag{4.15}\\
& \frac{\mathrm{~d} \tilde{r}_{j}(v)}{\mathrm{d} v}+\alpha_{j} \tilde{r}_{j}(v)+\left(1+\left(y_{j}-1\right) e^{-\mu_{j} v}+y_{j}\left(z_{j}-1\right) e^{-\mu_{j}(v+\tau)}\right) \beta(\tilde{\boldsymbol{r}}(v))-1=0,
\end{align*}
$$

with boundary condition $\tilde{s}_{j}(t)=s_{j}$ and $\tilde{r}_{j}(0)=r_{j}+\tilde{s}_{j}(t+\tau)$, and where $0 \leqslant v \leqslant t$ and $t \leqslant u \leqslant t+\tau$.

Proof. See Appendix 4.A.
Observe that the result of Theorem 4.2 can be extended to include arbitrarily many time points $t<t_{1}<t_{2}<\cdots<t_{k}, k \in \mathbb{N}$, by repeated conditioning and applying Eqn. (4.8). Further, as in Eqn. (4.12), we can obtain corresponding joint moments by differentiation. This in particular allows us to compute the auto-correlation and autocovariance functions of the multivariate Hawkes process and its associated population process. More precisely, for any $t \geqslant 0$ and $\tau>0$, we can compute the auto-correlation function by

$$
\begin{aligned}
R_{\boldsymbol{Q}}(t, \tau) & =\mathbb{E}\left[\boldsymbol{Q}(t) \boldsymbol{Q}(t+\tau)^{\top}\right], \\
R_{\boldsymbol{\lambda}}(t, \tau) & =\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t+\tau)^{\top}\right],
\end{aligned}
$$

and the auto-covariance function by

$$
\begin{aligned}
C_{\boldsymbol{Q}}(t, \tau) & =\mathbb{E}\left[\boldsymbol{Q}(t) \boldsymbol{Q}(t+\tau)^{\top}\right]-\mathbb{E}[\boldsymbol{Q}(t)] \mathbb{E}[\boldsymbol{Q}(t+\tau)]^{\top}, \\
C_{\boldsymbol{\lambda}}(t, \tau) & =\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t+\tau)^{\top}\right]-\mathbb{E}[\boldsymbol{\lambda}(t)] \mathbb{E}[\boldsymbol{\lambda}(t+\tau)]^{\top} .
\end{aligned}
$$

### 4.3.2 Joint moments

We proceed by exploiting the characterization of the joint transform $\zeta(t, s, \boldsymbol{z})$, as given in Theorem 4.1, to derive a system of linear differential equations for the joint transient moments pertaining to $(\boldsymbol{\lambda}(t), \boldsymbol{Q}(t))$ for a given $t \in \mathbb{R}_{+}$, as well as a system of linear (algebraic) equations for the corresponding stationary moments pertaining to ( $\boldsymbol{\lambda}, \boldsymbol{Q}$ ), where

$$
\begin{align*}
\boldsymbol{\lambda} & =\left(\lambda_{1}, \ldots, \lambda_{d}\right):=\lim _{t \rightarrow \infty}\left(\lambda_{1}(t), \ldots, \lambda_{d}(t)\right), \\
\boldsymbol{Q} & =\left(Q_{1}, \ldots, Q_{d}\right):=\lim _{t \rightarrow \infty}\left(Q_{1}(t), \ldots, Q_{d}(t)\right), \tag{4.16}
\end{align*}
$$

where the limits exist if the stability condition is in place, and in addition all $\mu_{i}$ are positive.

We start with the joint transient moments and derive a system of linear differential equations. Let $n_{Q_{i}}$ and $n_{\lambda_{i}} \in \mathbb{N}$ be such that $\sum_{i=1}^{d} n_{Q_{i}}=n_{Q}$ and $\sum_{i=1}^{d} n_{\lambda_{i}}=n_{\lambda}$. Consider, for given $n_{Q}, n_{\lambda} \in \mathbb{N}$ and $t \in \mathbb{R}_{+}$, the object

$$
\begin{equation*}
\varphi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right):=\mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} Q_{i}(t)^{n_{Q_{i}}}\right] \tag{4.17}
\end{equation*}
$$

we call these the joint moments of total order $n_{\lambda}$ and $n_{Q}$. To make sure the objects that we consider are well-defined, we throughout assume that for any $j \in[d]$,

$$
\begin{equation*}
\mathbb{E}\left[\prod_{i=1}^{d} B_{i j}^{n \lambda_{i}}\right]<\infty \tag{4.18}
\end{equation*}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

By significantly generalizing the approach of [57] to the multivariate setting, we obtain a vector-valued ODE to derive the joint transient moments (4.17).

In Appendix 4.A we give the proof of Theorem 4.1. In an intermediate step, summarized in Eqn. (4.79), the following PDE has been derived:

$$
\begin{align*}
& \frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]-\sum_{j=1}^{d}\left(\alpha_{j} s_{j}+z_{j} \beta_{j}(\boldsymbol{s})-1\right) \mathbb{E}\left[\lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& +\sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) e^{-s^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)-\mathbf{1}_{\{n=j\}}}\right]  \tag{4.19}\\
& =-\sum_{j=1}^{d} \alpha_{j} s_{j} \bar{\lambda}_{j} \mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]
\end{align*}
$$

where we wrote out the terms using the definition of $\zeta(t, \boldsymbol{s}, \boldsymbol{z})$. The next step is to repeatedly differentiate the PDE in Eqn. (4.19). This means that, for a given $n_{\lambda} \in \mathbb{N}$, we start with the $\operatorname{PDE}$ (4.19) and differentiate $n_{\lambda_{1}}, \ldots, n_{\lambda_{d}}$ times with respect to $s_{1}, \ldots, s_{d}$, respectively, and then substitute $s_{1}=s_{2}=\cdots=s_{d}=0$. Subsequently, we differentiate $n_{Q_{1}}, \ldots, n_{Q_{d}}$ times with respect to $z_{1}, \ldots, z_{d}$ respectively, and then substitute $z_{1}=\cdots=z_{d}=1$. As will become clear below, this procedure yields a relation between the joint moments of total order $n_{\lambda}$ and $n_{Q}$. To make the notation concise, we introduce the transient joint moments of total order $n_{\lambda}$ and $n_{Q}$ by

$$
\begin{equation*}
\psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right):=\mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} Q_{i}(t)^{\left[n_{Q_{i}}\right]}\right], \tag{4.20}
\end{equation*}
$$

where $\boldsymbol{n}_{\boldsymbol{Q}}=\left(n_{Q_{1}}, \ldots, n_{Q_{d}}\right)$ and $\boldsymbol{n}_{\boldsymbol{\lambda}}=\left(n_{\lambda_{1}}, \ldots, n_{\lambda_{d}}\right)$; these should be interpreted as reduced moments as far as the $Q_{i}(t)$ are concerned. Note that once we obtain an expression for $\psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)$, we can compute the joint moments $\varphi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)$ as given in Eqn. (4.17). Doing the repeated differentiation on Eqn. (4.19) and after elementary algebraic computations, we obtain

$$
\begin{align*}
& \frac{\mathrm{d}}{\mathrm{~d} t} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)+\sum_{j=1}^{d}\left(n_{\lambda_{j}}\left(\alpha_{j}-\mathbb{E}\left[B_{j j}\right]\right)+n_{Q_{j}} \mu_{j}\right) \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right) \\
& =\sum_{\substack{ \\
j=1}}^{d} \sum_{\substack{i=1 \\
i \neq j}}^{d} n_{\lambda_{i}} \mathbb{E}\left[B_{i j}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right)+\sum_{j=1}^{d} n_{Q_{j}} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{n}_{\boldsymbol{\lambda}}+\boldsymbol{e}_{\boldsymbol{j}}\right)  \tag{4.21}\\
& \quad+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} n_{\lambda_{j}} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{j}\right)+\sum_{i=1}^{d} \sum_{j=1}^{d} n_{\lambda_{i}} n_{Q_{j}} \mathbb{E}\left[B_{i j}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right)
\end{align*}
$$

$$
\begin{aligned}
& +\sum_{j=1}^{d} \sum_{m_{1}=0}^{n_{\lambda_{1}}} \cdots \sum_{m_{d}=0}^{n_{\lambda_{d}}} \mathbf{1}_{\left\{m \leqslant n_{\lambda}-2\right\}} \prod_{k=1}^{d}\binom{n_{\lambda_{k}}}{m_{k}}\left\{n_{Q_{j}} \prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right. \\
& \left.\quad+\prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right\}
\end{aligned}
$$

see Appendix 4.B. 2 for details.
Observe that Eqn. (4.21) gives a relation involving mixed moments of $\lambda_{1}(t), \ldots, \lambda_{d}(t)$ and of $Q_{1}(t), \ldots, Q_{d}(t)$ of total order $n_{\lambda}$ and $n_{Q}$, expressed in terms of mixed (factorial) moments of lower total order. Note that this identity generalizes Eqn. (3.9) in [57], with extra terms containing (products of) $\mathbb{E}\left[B_{i j}\right]$ arising from cross-excitation. We have thus obtained a linear vector-valued ODE that enables the computation of the transient joint moments of total order $n_{\lambda}$ and $n_{Q}$.

The joint stationary moments of total order $n_{\lambda}$ and $n_{Q}$, to be interpreted as reduced moments as far as the $Q_{i}$ are concerned,

$$
\begin{equation*}
\psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right):=\lim _{t \rightarrow \infty} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)=\mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}^{n_{\lambda_{i}}} Q_{i}^{\left[n_{Q_{i}}\right]}\right] \tag{4.22}
\end{equation*}
$$

can be identified analogously. Instead of a system of ODEs, we find for the stationary moments the following system of equations:

$$
\begin{align*}
& \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right) \sum_{j=1}^{d}\left(n_{\lambda_{j}}\left(\alpha_{j}-\mathbb{E}\left[B_{j j}\right]\right)+n_{Q_{j}} \mu_{j}\right) \\
& =\sum_{j=1}^{d} \sum_{\substack{i=1 \\
i \neq j}}^{d} n_{\lambda_{i}} \mathbb{E}\left[B_{i j}\right] \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right)+\sum_{j=1}^{d} n_{Q_{j}} \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{n}_{\boldsymbol{\lambda}}+\boldsymbol{e}_{\boldsymbol{j}}\right)  \tag{4.23}\\
& \quad+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} n_{\lambda_{j}} \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{j}\right)+\sum_{i=1}^{d} \sum_{j=1}^{d} n_{\lambda_{i}} n_{Q_{j}} \mathbb{E}\left[B_{i j}\right] \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right) \\
& \quad+\sum_{j=1}^{d} \sum_{m_{1}=0}^{n_{\lambda_{1}}} \cdots \sum_{m_{d}=0}^{n_{\lambda_{d}}} \mathbf{1}_{\left\{m \leqslant n_{\lambda}-2\right\}} \prod_{k=1}^{d}\binom{n_{\lambda_{k}}}{m_{k}}\left\{n_{Q_{j}} \prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right. \\
& \left.\quad+\prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right\} .
\end{align*}
$$

Note that Eqn. (4.23) can be interpreted as the stationary version of Eqn. (4.21). Where in transient case we had to solve a system of linear differential equations, in the stationary case this has turned into a system of linear algebraic equations.

We illustrate how this can be used to obtain moments of order $n \in\{1,2\}$, for the processes $\boldsymbol{Q}(\cdot)=\left(Q_{1}(\cdot), \ldots, Q_{d}(\cdot)\right)$ and $\boldsymbol{\lambda}(\cdot)=\left(\lambda_{1}(\cdot), \ldots, \lambda_{d}(\cdot)\right)$. We first focus

## Chapter 4 Markovian Multivariate Hawkes Population Processes

on transient moments and then proceed with the corresponding stationary moments. We introduce the relevant objects along the way, starting with the matrix $\mathbb{E}[\boldsymbol{B}]=$ $\left(\mathbb{E}\left[B_{i j}\right]\right)_{i, j \in[d]}$ and the diagonal matrices

$$
\boldsymbol{D}_{\alpha}:=\operatorname{diag}\left(\alpha_{1}, \alpha_{2}, \ldots, \alpha_{d}\right), \quad \boldsymbol{D}_{\mu}:=\operatorname{diag}\left(\mu_{1}, \mu_{2}, \ldots, \mu_{d}\right) .
$$

## Transient moments

We focus on the transient moments $\psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)$, where now $\boldsymbol{n}_{\boldsymbol{Q}}=\left(n_{Q_{1}}, \ldots, n_{Q_{d}}\right)$ and $\boldsymbol{n}_{\boldsymbol{\lambda}}=\left(n_{\lambda_{1}}, \ldots, n_{\lambda_{d}}\right)$. We separately consider the cases $n=1$ and $n=2$. To describe the joint moments of equal order in vector/matrix-form, it turns out that for $n=1$ we need a stacked vector, and for $n=2$ a stacked matrix, given in detail below.

For $n=1$ we define the stacked vector

$$
\begin{equation*}
\boldsymbol{\Sigma}_{t}^{(1)}:=(\mathbb{E}[\boldsymbol{\lambda}(t)], \mathbb{E}[\boldsymbol{Q}(t)])^{\top} \tag{4.24}
\end{equation*}
$$

For each entry of the vector, we use Eqn. (4.21) to obtain the vector-valued ODEs

$$
\begin{align*}
\frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}[\boldsymbol{\lambda}(t)] & =\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right) \mathbb{E}[\boldsymbol{\lambda}(t)]+\boldsymbol{L}^{(0,1)}, \\
\frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}[\boldsymbol{Q}(t)] & =-\boldsymbol{D}_{\mu} \mathbb{E}[\boldsymbol{Q}(t)]+\mathbb{E}[\boldsymbol{\lambda}(t)], \tag{4.25}
\end{align*}
$$

where $\boldsymbol{L}^{(0,1)}=\left(\alpha_{1} \bar{\lambda}_{1}, \alpha_{2} \bar{\lambda}_{2}, \ldots, \alpha_{d} \bar{\lambda}_{d}\right)^{\top}$. It is immediately verified that the solution is given by

$$
\begin{align*}
\mathbb{E}[\boldsymbol{\lambda}(t)] & =e^{t\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)} \overline{\boldsymbol{\lambda}}+\int_{0}^{t} e^{(t-s)\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)} \mathrm{d} s(\boldsymbol{\alpha} \odot \overline{\boldsymbol{\lambda}}) \\
& =e^{t\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)} \overline{\boldsymbol{\lambda}}+\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{-1}\left(e^{\left.t \mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)}-\boldsymbol{I}\right)(\boldsymbol{\alpha} \odot \overline{\boldsymbol{\lambda}}),  \tag{4.26}\\
\mathbb{E}[\boldsymbol{Q}(t)] & =\int_{0}^{t} e^{-(t-s) \boldsymbol{D}_{\mu}} \mathbb{E}[\boldsymbol{\lambda}(s)] \mathrm{d} s
\end{align*}
$$

We proceed with $n=2$. As indicated at the start of this subsection, in this case we should work with a stacked matrix. To this end, define

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{Q}(t)^{[2]}\right] & :=\mathbb{E}\left[\boldsymbol{Q}(t) \boldsymbol{Q}(t)^{\top}\right]-\operatorname{diag}(\mathbb{E}[\boldsymbol{Q}(t)]) \\
& =\mathbb{E}\left[\begin{array}{cccc}
Q_{1}(t)^{[2]} & Q_{1}(t) Q_{2}(t) & \cdots & Q_{1}(t) Q_{d}(t) \\
Q_{2}(t) Q_{1}(t) & Q_{2}(t)^{[2]} & \cdots & Q_{2}(t) Q_{d}(t) \\
\vdots & \vdots & \ddots & \vdots \\
Q_{d}(t) Q_{1}(t) & Q_{d}(t) Q_{2}(t) & \cdots & Q_{d}(t)^{[2]}
\end{array}\right],
\end{aligned}
$$

and we also consider $\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right]$ and $\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right]$. Note that these are all $d \times d$ matrices and we define the stacked matrix $\boldsymbol{\Sigma}_{t}^{(2)}$ given by

$$
\begin{equation*}
\boldsymbol{\Sigma}_{t}^{(2)}:=\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right] \oplus \mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right] \oplus \mathbb{E}\left[\boldsymbol{Q}(t)^{[2]}\right] \tag{4.27}
\end{equation*}
$$

where $\oplus$ indicates the direct sum, such that $\boldsymbol{\Sigma}_{t}^{(2)}$ is a $3 d \times 3 d$-matrix. We now describe the three steps required to compute each submatrix of the stacked matrix. For each entry of a submatrix, we derive its associated ODE from Eqn. (4.21) which we combine to obtain matrix-valued ODEs. We have that the matrices $\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right]$, $\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right]$ and $\mathbb{E}\left[\boldsymbol{Q}(t)^{[2]}\right]$, satisfy the matrix-valued ODEs

$$
\begin{align*}
\frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right]= & \left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right) \mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right]+\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right]\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{\top} \\
& +\mathbb{E}\left[\boldsymbol{B} \operatorname{diag}(\mathbb{E}[\boldsymbol{\lambda}(t)]) \boldsymbol{B}^{\top}\right]+\boldsymbol{D}_{\alpha}\left(\overline{\boldsymbol{\lambda}} \mathbb{E}[\boldsymbol{\lambda}(t)]^{\top}\right)+\left(\mathbb{E}[\boldsymbol{\lambda}(t)] \overline{\boldsymbol{\lambda}}^{\top}\right) \boldsymbol{D}_{\alpha}, \\
\frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right]= & \left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right) \mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right]-\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right] \boldsymbol{D}_{\mu}+\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{\lambda}(t)^{\top}\right]  \tag{4.28}\\
& +(\boldsymbol{\alpha} \odot \overline{\boldsymbol{\lambda}}) \mathbb{E}[\boldsymbol{Q}(t)]^{\top}+\mathbb{E}[\boldsymbol{B}] \operatorname{diag}(\mathbb{E}[\boldsymbol{\lambda}(t)]), \\
\frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}\left[\boldsymbol{Q}(t)^{[2]}\right]= & -\boldsymbol{D}_{\mu} \mathbb{E}\left[\boldsymbol{Q}(t)^{[2]}\right]-\mathbb{E}\left[\boldsymbol{Q}(t)^{[2]}\right] \boldsymbol{D}_{\mu}+\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right]+\left(\mathbb{E}\left[\boldsymbol{\lambda}(t) \boldsymbol{Q}(t)^{\top}\right]\right)^{\top} .
\end{align*}
$$

We end our account of the transient moments with a series of brief remarks. The ODEs for $\boldsymbol{\Sigma}_{t}^{(1)}$ and $\boldsymbol{\Sigma}_{t}^{(2)}$ are related to those derived in Lemmas 1 and 3 in [22]. Concretely, the solution for the first moment $\mathbb{E}[\boldsymbol{\lambda}(t)]$ agrees with Eqn. (8) in [22]. Furthermore, by taking the limit $\boldsymbol{\mu} \downarrow \mathbf{0}$ in our expression for $\boldsymbol{Q}(t)$ we obtain

$$
\begin{align*}
\mathbb{E}[\boldsymbol{N}(t)]=\lim _{\boldsymbol{\mu} \downarrow \mathbf{0}} \mathbb{E}[\boldsymbol{Q}(t)]= & \left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{-1}\left(e^{t\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)}-I\right) \overline{\boldsymbol{\lambda}} \\
& +\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{-2}\left(e^{t\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)}-I\right)(\boldsymbol{\alpha} \odot \overline{\boldsymbol{\lambda}})  \tag{4.29}\\
& +t\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{-1}(\boldsymbol{\alpha} \odot \overline{\boldsymbol{\lambda}}),
\end{align*}
$$

which agrees with the result in Eqn. (10) in [22]. Considering the second order moments, upon taking $\boldsymbol{\mu} \downarrow 0$ in Eqn. (4.28), these agree with those in Lemma 3 in [22]. Note that the solution of $\mathbb{E}\left[\boldsymbol{N}(t)^{[2]}\right]$ is slightly different since we study the reduced moment instead of $\mathbb{E}\left[\boldsymbol{N}(t) \boldsymbol{N}(t)^{\top}\right]$. We finally remark that for orders $n=3,4, \ldots$, it becomes more cumbersome to obtain a clean and transparent way in which one can encode the ODEs that describe the transient moments.

## Stationary moments

We continue by considering the joint stationary moments of order at most 2 for $\boldsymbol{Q}$ and $\boldsymbol{\lambda}$ of general dimension $d \in \mathbb{N}$, see Eqn. (4.16). By exploiting the relation between

## Chapter 4 Markovian Multivariate Hawkes Population Processes

the joint stationary moments $\psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)$ as described Eqn. (4.23), we derive the linear systems of equations that the joint moments satisfy. In the sequel we use the notation from Subsection 4.3.2.

For order $n=1$, define the stacked vector

$$
\begin{equation*}
\boldsymbol{\Sigma}^{(1)}:=\lim _{t \rightarrow \infty} \boldsymbol{\Sigma}_{t}^{(1)}=(\mathbb{E}[\boldsymbol{\lambda}], \mathbb{E}[\boldsymbol{Q}])^{\top} \tag{4.30}
\end{equation*}
$$

which is the stationary version of Eqn. (4.24). We derive from Eqn. (4.23) that the elements of this stacked vector satisfy

$$
\begin{align*}
\mathbb{E}[\boldsymbol{\lambda}] & =-\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{-1} \boldsymbol{L}^{(0,1)}, \\
\mathbb{E}[\boldsymbol{Q}] & =\boldsymbol{D}_{\mu}^{-1} \mathbb{E}[\boldsymbol{\lambda}] \tag{4.31}
\end{align*}
$$

For order $n=2$, we define the stacked matrix

$$
\begin{equation*}
\boldsymbol{\Sigma}^{(2)}:=\lim _{t \rightarrow \infty} \boldsymbol{\Sigma}_{t}^{(2)}=\mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{\lambda}^{\top}\right] \oplus \mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{Q}^{\top}\right] \oplus \mathbb{E}\left[\boldsymbol{Q}^{[2]}\right] . \tag{4.32}
\end{equation*}
$$

From the procedure followed for the transient moment stacked matrix $\boldsymbol{\Sigma}_{t}^{(2)}$, in combination with Eqn. (4.23), we conclude that the stationary submatrices satisfy

$$
\begin{align*}
0= & \left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right) \mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{\lambda}^{\top}\right]+\mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{\lambda}^{\top}\right]\left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right)^{\top}+\mathbb{E}\left[\boldsymbol{B} \operatorname{diag}(\mathbb{E}[\boldsymbol{\lambda}]) \boldsymbol{B}^{\top}\right] \\
& +\boldsymbol{D}_{\alpha}\left(\overline{\boldsymbol{\lambda}} \mathbb{E}[\boldsymbol{\lambda}]^{\top}\right)+\left(\mathbb{E}[\boldsymbol{\lambda}] \overline{\boldsymbol{\lambda}}^{\top}\right) \boldsymbol{D}_{\alpha}, \\
0= & \left(\mathbb{E}[\boldsymbol{B}]-\boldsymbol{D}_{\alpha}\right) \mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{Q}^{\top}\right]-\mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{Q}^{\top}\right] \boldsymbol{D}_{\mu}+\mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{\lambda}^{\top}\right]+(\boldsymbol{\alpha} \odot \overline{\boldsymbol{\lambda}}) \mathbb{E}[\boldsymbol{Q}]^{\top}+\mathbb{E}[\boldsymbol{B}] \operatorname{diag}(\mathbb{E}[\boldsymbol{\lambda}])^{\top}, \\
0= & -\boldsymbol{D}_{\mu} \mathbb{E}\left[\boldsymbol{Q}^{[2]}\right]-\mathbb{E}\left[\boldsymbol{Q}^{[2]}\right] \boldsymbol{D}_{\mu}+\mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{Q}^{\top}\right]+\left(\mathbb{E}\left[\boldsymbol{\lambda} \boldsymbol{Q}^{\top}\right]\right)^{\top} . \tag{4.33}
\end{align*}
$$

The matrix-valued equations in (4.33) are all Sylvester equations, i.e., equations of the form

$$
\begin{equation*}
A \boldsymbol{X}+\boldsymbol{X} \boldsymbol{B}=\boldsymbol{C} \tag{4.34}
\end{equation*}
$$

for known matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$, with the matrix $\boldsymbol{X}$ being unknown. It is a known result that a unique solution for $\boldsymbol{X}$ exists if and only $\boldsymbol{A}$ and $-\boldsymbol{B}$ do not share any eigenvalue.

We conclude this subsection with two results on higher order stationary moments. By applying Eqn. (4.23), we can obtain expressions for the moments

$$
\begin{equation*}
\psi\left(\mathbf{0}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)=\mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}^{n_{\lambda_{i}}}\right], \quad \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \mathbf{0}\right)=\mathbb{E}\left[\prod_{i=1}^{d} Q_{i}^{\left[n_{Q_{i}}\right]}\right] \tag{4.35}
\end{equation*}
$$

by straightforward substitution. Indeed, for fixed $n_{\lambda} \in \mathbb{N}$ with $n_{\lambda}=n_{\lambda_{1}}+\cdots+n_{\lambda_{d}}$, we substitute $n_{Q_{j}} \equiv 0$ for all $j \in[d]$ in Eqn. (4.23) and rearrange terms to obtain

$$
\begin{align*}
& \psi\left(\mathbf{0}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)=\left(\sum_{j=1}^{d} n_{\lambda_{j}}\left(\alpha_{j}-\mathbb{E}\left[B_{j j}\right]\right)\right)^{-1} \sum_{j=1}^{d}\left\{\sum_{\substack{i=1 \\
i \neq j}}^{d} n_{\lambda_{i}} \mathbb{E}\left[B_{i j}\right] \psi\left(\mathbf{0}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right)\right.  \tag{4.36}\\
& \left.+\alpha_{j} \bar{\lambda}_{j} n_{\lambda_{j}} \psi\left(\mathbf{0}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{j}\right)+\sum_{m_{1}=0}^{n_{\lambda_{1}}} \cdots \sum_{m_{d}=0}^{n_{\lambda_{d}}} \mathbf{1}_{\left\{m \leqslant n_{\lambda}-2\right\}} \prod_{k=1}^{d}\binom{n_{\lambda_{k}}}{m_{k}} \prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi\left(\mathbf{0}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right\},
\end{align*}
$$

under the assumption that $\mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}}\right]<\infty$ for all combinations $i, j \in[d]$. Observe that in order to obtain a final closed-form expression for $\psi\left(\mathbf{0}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)$, we need to solve a linear system of equations of equal order moments, i.e., the $\psi\left(\mathbf{0}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right)$ terms.

A similar result holds for the joint moments of $\boldsymbol{Q}$, given below. For fixed $n_{Q} \in \mathbb{N}$ with $n_{Q}=n_{Q_{1}}+\cdots+n_{Q_{d}}$, we substitute $n_{\lambda_{j}} \equiv 0$ for all $j \in[d]$ in Eqn. (4.23), yielding

$$
\begin{equation*}
\psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \mathbf{0}\right)=\left(\sum_{j=1}^{d} n_{Q_{j}} \mu_{j}\right)^{-1} \sum_{j=1}^{d} n_{Q_{j}} \psi\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{e}_{\boldsymbol{j}}\right) \tag{4.37}
\end{equation*}
$$

We note that one can use the factorial moments $\psi\left(\boldsymbol{n}_{\boldsymbol{Q}}, \mathbf{0}\right)$ to determine the nonfactorial moments of $\boldsymbol{Q}$. More precisely, for a fixed $i \in[d]$ and $n_{Q_{i}} \in \mathbb{N}$, we use the well-known relationship

$$
\begin{equation*}
\mathbb{E}\left[Q_{i}^{n_{Q_{i}}}\right]=\sum_{k=1}^{n_{Q_{i}}} \Delta_{k, n_{Q_{i}}} \mathbb{E}\left[Q_{i}^{[k]}\right]=\sum_{k=1}^{n_{Q_{i}}} \Delta_{k, n_{Q_{i}}} \psi\left(k \boldsymbol{e}_{i}, \mathbf{0}\right) \tag{4.38}
\end{equation*}
$$

where $\Delta_{k, n_{Q_{i}}}$ are the Stirling numbers of the second kind, i.e., $\Delta_{k, n_{Q_{i}}}$ denotes the number of ways one can partition a set of $n_{Q_{i}}$ elements into $k$ non-empty subsets.

### 4.4 Recursive Procedure

In this section, we consider the bivariate setting $d=2$ and focus on the structure behind the joint moments of arbitrary order $n \in \mathbb{N}$. While we focus on the bivariate $d=2$ setting here, this method can be extended to higher dimensions $d \in \mathbb{N}$, at the cost of heavier notation and more intricate objects. As such, this section serves as a proof of principle on how to identify the recursive structure. This novel approach leads to a transparent and exhaustive method to compute all possible joint moments of the bivariate Hawkes process. Moreover, it provides the necessary insight into the dependence structure between different moments, which can be exploited for fast computation in Section 4.5 below.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

We construct, based on the results obtained in the previous sections, a recursive procedure to compute the joint transient moments $\psi_{t}\left(\boldsymbol{n}_{\boldsymbol{\lambda}}, \boldsymbol{n}_{Q}\right)$ as well as the joint stationary moments $\psi\left(\boldsymbol{n}_{\boldsymbol{\lambda}}, \boldsymbol{n}_{\boldsymbol{Q}}\right)$. To make the analysis as transparent as possible, we will express the main objects in vector/matrix-form. As it turns out, there is a strong similarity between the structure of the algorithm to evaluate the transient moments on one hand and its counterpart for the stationary moments on the other hand.

In the sequel we let $n$ be the total order of the joint moments, i.e.,

$$
n=n_{Q}+n_{\lambda}=n_{Q_{1}}+n_{Q_{2}}+n_{\lambda_{1}}+n_{\lambda_{2}} .
$$

As a first step we rewrite the system of linear differential equations, as given by Eqns. (4.21), and the system of linear equations, as given by (4.23), in vector-matrix form. We then use these to set up a procedure to compute the transient and stationary joint moments of any order. We first focus on deriving a procedure for the transient moments $\psi_{t}\left(\boldsymbol{n}_{\boldsymbol{\lambda}}, \boldsymbol{n}_{\boldsymbol{Q}}\right)$, as the stationary moments $\psi\left(\boldsymbol{n}_{\boldsymbol{\lambda}}, \boldsymbol{n}_{\boldsymbol{Q}}\right)$ will follow from that.

### 4.4.1 Transient moments

We construct a recursive procedure to compute transient joint moments $\psi_{t}\left(\boldsymbol{n}_{\boldsymbol{\lambda}}, \boldsymbol{n}_{\boldsymbol{Q}}\right)$ by introducing properly defined vector- and matrix-valued objects, such that we can exploit the ODE in Eqn. (4.21). Since $d=2$, we aim to compute

$$
\psi_{t}\left(\left(n_{Q_{1}}, n_{Q_{2}}\right),\left(n_{\lambda_{1}}, n_{\lambda_{2}}\right)\right)=\mathbb{E}\left[\lambda_{1}(t)^{n_{\lambda_{1}}} \lambda_{2}(t)^{n_{\lambda_{2}}} Q_{1}(t)^{\left[n_{Q_{1}}\right]} Q_{2}(t)^{\left[n_{Q_{2}}\right]}\right] .
$$

The multivariate setting has the intrinsic complication that the number of combinations of possible joint moments increases rapidly in $n$ and $d$; already in this bivariate setting, there are many possible combinations of joint moments of order $n$. To collect all joint moments $\psi_{t}\left(\left(n_{Q_{1}}, n_{Q_{2}}\right),\left(n_{\lambda_{1}}, n_{\lambda_{2}}\right)\right)$ in a single vector, we need to specify an ordering of the different moments. To that end, we introduce the stacked vector

$$
\begin{equation*}
\mathbf{\Psi}_{t}^{(n)}:=\left(\mathbf{\Psi}_{t}^{(0, n)}, \mathbf{\Psi}_{t}^{(1, n-1)}, \ldots, \mathbf{\Psi}_{t}^{(n, 0)}\right)^{\top} \tag{4.39}
\end{equation*}
$$

where, for each $k \in\{0,1, \ldots, n\}$, the vector $\mathbf{\Psi}_{t}^{(k, n-k)}$ exhaustively contains all combinations of joint moments such that $n_{Q_{1}}+n_{Q_{2}}=k$ and $n_{\lambda_{1}}+n_{\lambda_{2}}=n-k$. We illustrate this concept by detailing a few special cases. In the first place, for $k=0$

$$
\begin{aligned}
\boldsymbol{\Psi}_{t}^{(0, n)} & =\left(\psi_{t}(\mathbf{0},(n, 0)), \psi_{t}(\mathbf{0},(n-1,1)), \ldots, \psi_{t}(\mathbf{0},(0, n))\right)^{\top} \\
& =\left(\mathbb{E}\left[\lambda_{1}(t)^{n}\right], \mathbb{E}\left[\lambda_{1}(t)^{n-1} \lambda_{2}(t)\right], \ldots, \mathbb{E}\left[\lambda_{2}(t)^{n}\right]\right)^{\top},
\end{aligned}
$$

whereas for $k=n$

$$
\begin{aligned}
\boldsymbol{\Psi}_{t}^{(n, 0)} & =\left(\psi_{t}((n, 0), \mathbf{0}), \psi_{t}((n-1,1), \mathbf{0}), \ldots, \psi_{t}((0, n), \mathbf{0})\right)^{\top} \\
& =\left(\mathbb{E}\left[Q_{1}(t)^{[n]}\right], \mathbb{E}\left[Q_{1}(t)^{[n-1]} Q_{2}(t)\right], \ldots, \mathbb{E}\left[Q_{2}(t)^{[n]}\right]\right)^{\top} .
\end{aligned}
$$

The cases corresponding with $k \in\{1, \ldots, n-1\}$ are notationally considerably more burdensome since one has to include all possible combinations of order $k$ as well as $n-k$.

For concrete examples of the stacked vector in (4.39) for orders $n=1,2,3$, see Appendix 4.C.1. It is readily verified that for general $d$ the dimension of $\mathbf{\Psi}_{t}^{(n)}$ equals

$$
\begin{equation*}
\mathfrak{D}(d, n):=\sum_{k=1}^{n}\binom{2 d}{k}\binom{n-1}{k-1} \tag{4.40}
\end{equation*}
$$

where the $2 d$ is due to the fact that we include moments of both $\boldsymbol{Q}(t)$ and $\boldsymbol{\lambda}(t)$. For $d=2$, we thus have that the size of $\boldsymbol{\Psi}_{t}^{(n)}$ is given by $\mathfrak{D}(2, n)$.

The stacked vector $\boldsymbol{\Psi}_{t}^{(n)}$ satisfies a vector-valued ODE, which we describe now. By Eqn. (4.21), it is immediate that we can write

$$
\begin{equation*}
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(n)}=\boldsymbol{M} \boldsymbol{\Psi}_{t}^{(n)}+\boldsymbol{L}\left(\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}\right)^{\top} \tag{4.41}
\end{equation*}
$$

for certain matrices $\boldsymbol{M}$ and $\boldsymbol{L}$ of appropriate dimension. Here, the matrix $\boldsymbol{M}$ is of dimension $\mathfrak{D}(2, n) \times \mathfrak{D}(2, n)$, and $\boldsymbol{L}$ of dimension $\mathfrak{D}(2, n) \times \overline{\mathfrak{D}}(2, n)$, where

$$
\begin{equation*}
\overline{\mathfrak{D}}(2, n)=\sum_{m=1}^{n-1} \mathfrak{D}(2, m) . \tag{4.42}
\end{equation*}
$$

As a next step, we identify blocks of $\boldsymbol{M}$ that correspond to subvectors $\boldsymbol{\Psi}_{t}^{(k, n-k)}$ of the stacked vector $\boldsymbol{\Psi}_{t}^{(n)}$. Upon inspecting (4.21) we observe that, when considering in (4.41) the differential equations that correspond to $\mathrm{d} \boldsymbol{\Psi}_{t}^{(k, n-k)} / \mathrm{d} t$, in the right hand side only $\boldsymbol{\Psi}_{t}^{(k, n-k)}$ and $\boldsymbol{\Psi}_{t}^{(k-1, n-k+1)}$ appear, besides a linear combination of the lower-order objects $\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}$. It means that we can write

$$
\begin{align*}
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(k, n-k)}= & \boldsymbol{M}^{(k, n-k)} \boldsymbol{\Psi}_{t}^{(k, n-k)}+\boldsymbol{K}^{(k, n-k)} \boldsymbol{\Psi}_{t}^{(k-1, n-k+1)}  \tag{4.43}\\
& +\boldsymbol{L}^{(k, n-k)}\left(\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}\right)^{\top}
\end{align*}
$$

for matrices $\boldsymbol{M}^{(k, n-k)}, \boldsymbol{K}^{(k, n-k)}$, and $\boldsymbol{L}^{(k, n-k)}$, where we set $\boldsymbol{K}^{(k, n-k)} \equiv 0$ when $m=0$. Eqn. (4.43) reveals a recursive procedure to compute $\boldsymbol{\Psi}_{t}^{(n)}$, where in the $n$-th iteration

## Chapter 4 Markovian Multivariate Hawkes Population Processes

a non-homogeneous linear system of ODEs has to be solved, with $\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}$, as derived in the previous steps, appearing in the non-homogeneous part.

We proceed by introducing the notation needed to set up the recursive procedure. A tridiagonal matrix in $\mathbb{R}^{n \times n}$ is a matrix with elements on the main diagonal, the first diagonal above and below the main diagonal only, for which we use the notation, with $\boldsymbol{a}, \boldsymbol{c} \in \mathbb{R}^{n-1}$ and $\boldsymbol{d} \in \mathbb{R}^{n}$,

$$
\operatorname{tridiag}(\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{c}):=\left[\begin{array}{ccccc}
d_{1} & c_{1} & 0 & \cdots & 0 \\
a_{1} & d_{2} & c_{2} & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & a_{n-2} & d_{n-1} & c_{n-1} \\
0 & 0 & 0 & a_{n-1} & d_{n}
\end{array}\right] .
$$

Given the vectors $\boldsymbol{n}_{\boldsymbol{Q}}=\left(n_{Q_{1}}, n_{Q_{2}}\right)$ and $\boldsymbol{n}_{\boldsymbol{\lambda}}=\left(n_{\lambda_{1}}, n_{\lambda_{2}}\right)$, we set

$$
\begin{aligned}
v\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right) & :=-n_{\lambda_{1}} \bar{\alpha}_{1}-n_{Q_{1}} \mu_{1}-n_{\lambda_{2}} \bar{\alpha}_{2}-n_{Q_{2}} \mu_{2}, \\
\boldsymbol{w}\left(\boldsymbol{n}_{\boldsymbol{Q}}, n\right) & :=\left(v\left(\boldsymbol{n}_{\boldsymbol{Q}},(n, 0)\right), v\left(\boldsymbol{n}_{\boldsymbol{Q}},(n-1,1)\right), \ldots, v\left(\boldsymbol{n}_{\boldsymbol{Q}},(1, n-1)\right), v\left(\boldsymbol{n}_{\boldsymbol{Q}},(0, n)\right)\right)^{\top},
\end{aligned}
$$

where $\bar{\alpha}_{i}:=\alpha_{i}-\mathbb{E}\left[B_{i i}\right]$ for $i=1,2$, corresponding to the LHS of Eqn. (4.21). With the vectors

$$
\mathbf{1}_{(n)}:=(1,2, \ldots, n), \quad \mathbf{1}^{(n)}:=(n, n-1, \ldots, 1),
$$

and with $e_{j}$ the unit vector with 1 on the $j$-th component, we finally define the matrix, for each $k \in\{0,1, \ldots, n-1, n\}$

$$
\begin{equation*}
\boldsymbol{M}^{(k, n-k)}:=\bigoplus_{m=0}^{k} \operatorname{tridiag}\left(\mathbf{1}_{(n-k)} \mathbb{E}\left[B_{21}\right], \boldsymbol{w}\left((n-m) \boldsymbol{e}_{1}+m \boldsymbol{e}_{2}, n-k\right), \mathbf{1}^{(n-k)} \mathbb{E}\left[B_{12}\right]\right), \tag{4.44}
\end{equation*}
$$

where $\bigoplus$ denotes the direct sum notation for matrices. For concrete examples of the matrix $\boldsymbol{M}^{(k, n-k)}$ and how it appears in the ODE in (4.43), see Appendix 4.C. We can now present our algorithm to compute the transient moments after which we provide a method to solve the underlying ODEs.

Algorithm 4.1. Fix $n \in \mathbb{N}$, and suppose we know the sequence of lower-order stacked vectors $\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}$. The vector-valued ODE in Eqn. (4.41) for the stacked vector $\mathbf{\Psi}_{t}^{(n)}$ can be solved as follows.

Step 0: The vector $\mathbf{\Psi}_{t}^{(0, n)}$ satisfies the vector-valued ODE

$$
\begin{equation*}
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(0, n)}=\boldsymbol{M}^{(0, n)} \boldsymbol{\Psi}_{t}^{(0, n)}+\boldsymbol{L}^{(0, n)}\left(\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}\right) \tag{4.45}
\end{equation*}
$$

where the matrix $\boldsymbol{L}^{(0, n)}$ follows from Eqn. (4.21), with initial condition $\mathbf{\Psi}_{0}^{(0, n)}$ determined by $\boldsymbol{Q}(0)$ and $\boldsymbol{\lambda}(0)$.

Step m: For any $k=1,2, \ldots, n-1, n$, the vector $\boldsymbol{\Psi}_{t}^{(k, n-k)}$ satisfies the vector-valued ODE

$$
\begin{gather*}
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(k, n-k)}=\boldsymbol{M}^{(k, n-k)} \boldsymbol{\Psi}_{t}^{(k, n-k)}+\boldsymbol{K}^{(k, n-k)} \boldsymbol{\Psi}_{t}^{(k-1, n-k+1)}  \tag{4.46}\\
+\boldsymbol{L}^{(k, n-k)}\left(\boldsymbol{\Psi}_{t}^{(1)}, \ldots, \boldsymbol{\Psi}_{t}^{(n-1)}\right)
\end{gather*}
$$

where the matrices $\boldsymbol{K}^{(k, n-k)}$ and $\boldsymbol{L}^{(k, n-k)}$ are derived in a similar manner from Eqn. (4.21), with initial condition $\mathbf{\Psi}_{0}^{(k, n-k)}$.

It is clear that Eqn. (4.21) uniquely defines the matrices $\boldsymbol{K}^{(k, n-k)}$ and $\boldsymbol{L}^{(k, n-k)}$ needed in the above algorithm. However, their explicit definition would require objects that are even more notationally involved. For moments of orders $n=1$ and $n=2$, we can still explicitly write down the matrix for the stacked vector ODE, we can combine the blocks of matrices into $4 \times 4$ and $10 \times 10$ matrices $\boldsymbol{M}$ respectively and also construct the corresponding matrix $\boldsymbol{L}$, see Appendix 4.C. However, for order $n=3,4, \ldots$, we would need very large matrices which are cumbersome to write down explicitly.

Due to the direct sum structure of $\boldsymbol{M}^{(k, n-k)}$, it consists of blocks. This allows us to decompose the $k$-th step in the algorithm into smaller steps, by considering the parts of the vector $\boldsymbol{\Psi}_{t}^{(k, n-k)}$ associated with the individual blocks of the matrix. The solution to the ODEs in Algorithm 4.1 can be given in terms of a matrix exponential.

Proposition 4.1. For fixed $t \in \mathbb{R}_{+}, n \in \mathbb{N}$ and $k=0,1, \ldots, n-1, n$, the solution for the vector-valued ODE for $\boldsymbol{\Psi}_{t}^{(k, n-k)}$ in Eqn. (4.46) is given by

$$
\begin{align*}
\boldsymbol{\Psi}_{t}^{(k, n-k)}= & e^{t \boldsymbol{M}^{(k, n-k)}} \boldsymbol{\Psi}_{0}^{(k, n-k)}  \tag{4.47}\\
& +\int_{0}^{t} e^{(t-s) \boldsymbol{M}^{(k, n-k)}}\left(\boldsymbol{K}^{(k, n-k)} \boldsymbol{\Psi}_{s}^{(k-1, n-k+1)}+\boldsymbol{L}^{(k, n-k)}\left(\boldsymbol{\Psi}_{s}^{(1)}, \ldots, \boldsymbol{\Psi}_{s}^{(n-1)}\right)\right) \mathrm{d} s .
\end{align*}
$$

Proof. Computing $\frac{\mathrm{d}}{\mathrm{d} t} \mathbf{\Psi}_{t}^{(k, n-k)}$ in Eqn. (4.47) immediately yields (4.46) by inspection and Leibniz' integral rule.

In case the objective is to obtain closed-form expressions, this becomes prohibitive for orders $n=3,4, \ldots$. In Appendix 4.C, we give explicit results by working out the details for transient moments of order $n=1$ and $n=2$.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

### 4.4.2 Stationary moments

In this subsection, we focus on computing the joint stationary order $n \in \mathbb{N}$ moments: with $n=n_{Q}+n_{\lambda}$ the total order, we consider

$$
\begin{equation*}
\psi\left(\left(n_{Q_{1}}, n_{Q_{2}}\right),\left(n_{\lambda_{1}}, n_{\lambda_{2}}\right)\right)=\mathbb{E}\left[\lambda_{1}^{n_{\lambda_{1}}} \lambda_{2}^{n \lambda_{\lambda_{2}}} Q_{1}^{\left[n_{Q_{1}}\right]} Q_{2}^{\left[n Q_{2}\right]}\right] . \tag{4.48}
\end{equation*}
$$

By exploiting Eqn. (4.23), we develop a recursive procedure similar to the one for the transient moments. The central object of study is

$$
\begin{equation*}
\boldsymbol{\Psi}^{(n)}:=\lim _{t \rightarrow \infty} \boldsymbol{\Psi}_{t}^{(n)}=\left(\boldsymbol{\Psi}^{(0, n)}, \boldsymbol{\Psi}^{(1, n-1)}, \ldots, \boldsymbol{\Psi}^{(n, 0)}\right)^{\top} \tag{4.49}
\end{equation*}
$$

For the following result, we use the notation from Section 4.4.1; recall in particular the matrices defined in Eqn. (4.44). We have the following recursive procedure to compute joint stationary order $n$ moments, which is the stationary analog of Algorithm 4.1.
Algorithm 4.2. For any $n \in \mathbb{N}$, the elements of the stacked vector $\boldsymbol{\Psi}^{(n)}$ can be computed using the following $(n+1)$-step recursive procedure, given the lower-order stacked vectors $\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}$.

Step 0: The vector $\boldsymbol{\Psi}^{(0, n)}$ satisfies the linear equation

$$
\begin{equation*}
0=\boldsymbol{M}^{(0, n)} \boldsymbol{\Psi}^{(0, n)}+\boldsymbol{L}^{(0, n)}\left(\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}\right)^{\top} \tag{4.50}
\end{equation*}
$$

where the matrix $\boldsymbol{L}^{(0, n)}$ follows from Eqn. (4.23).
Step $k$ : For any $k=1,2, \ldots, n-1, n$, the vector $\boldsymbol{\Psi}^{(k, n-k)}$ satisfies the linear equations

$$
\begin{align*}
0= & \boldsymbol{M}^{(k, n-k)} \boldsymbol{\Psi}^{(k, n-k)}+\boldsymbol{K}^{(k, n-k)} \boldsymbol{\Psi}^{(k-1, n-k+1)} \\
& +\boldsymbol{L}^{(k, n-k)}\left(\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}\right)^{\top}, \tag{4.51}
\end{align*}
$$

where the matrices $\boldsymbol{K}^{(k, n-k)}$ and $\boldsymbol{L}^{(k, n-k)}$ are derived in a similar manner from Eqn. (4.23).

The solution to the linear equations in Algorithm 4.2 is the following result.
Proposition 4.2. For $n \in \mathbb{N}$ and $k=0,1, \ldots, n-1, n$, the solution for the linear equation for $\boldsymbol{\Psi}^{(k, n-k)}$ in Eqn. (4.51) is given by

$$
\boldsymbol{\Psi}^{(k, n-k)}=-\left(\boldsymbol{M}^{(k, n-k)}\right)^{-1}\left\{\boldsymbol{K}^{(k, n-k)} \boldsymbol{\Psi}^{(k-1, n-k+1)}+\boldsymbol{L}^{(k, n-k)}\left(\boldsymbol{\Psi}^{(1)}, \ldots, \boldsymbol{\Psi}^{(n-1)}\right)^{\top}\right\}
$$

Proof. Follows immediately from solving Eqn. (4.51).
Observe the strong similarity between the Algorithms 4.1 and 4.2, in the sense that the underlying recursive structures fully match. Where for the transient moments in the algorithm linear differential equations need to be solved, for their stationary counterparts corresponding linear algebraic equations need to be solved.

### 4.5 Nested Block Matrices: Bivariate Setting

In this section, we investigate the nested structure of the matrices associated with the ODEs of the moments more thoroughly, again in the bivariate setting $d=2$. The motivation behind this is to identify a fast computational method for the transient and stationary moments by exploiting the recursive procedure since the recursive method requires nested integration or inversion of complicated matrices. It turns out that one can find a nested sequence of well-behaved matrices that describe the relations between the moments and enhance the computational effort significantly. We will see this section is a bivariate version of Section 3.2 in [27], where they derive the structure of ODEs associated with the transient moments using lower triangular matrices with scalar entries. The difference is that in our case, they are replaced by block lower triangular matrices, containing matrix entries. As was the case in Section 4.4, the method can be extended to higher dimensions $d \in \mathbb{N}$.

We first revisit the transient moments of $\boldsymbol{\lambda}(t)$ for a fixed $t \in \mathbb{R}_{+}$so as to illustrate the nested structure of the matrices. After that, we consider the joint transient moments of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$, which has a similar but more complex nested structure. To motivate our analysis, consider the ODEs associated with the vectors $\boldsymbol{\Psi}_{t}^{(0,1)}$ and $\boldsymbol{\Psi}_{t}^{(0,2)}$ containing the first and second order moments of $\boldsymbol{\lambda}(t)$ (see Eqns. (4.85) and (4.87) in Appendix 4.C.1). Observe that in stacked form, they can be represented in a block lower triangular matrix structure, as follows:

$$
\frac{\mathrm{d}}{\mathrm{~d} t}\left[\begin{array}{l}
\boldsymbol{\Psi}_{t}^{(0,1)}  \tag{4.52}\\
\boldsymbol{\Psi}_{t}^{(0,2)}
\end{array}\right]=\left[\begin{array}{ll}
\boldsymbol{A}_{1}^{2 \times 2} & \mathbf{0}_{2}^{2 \times 3} \\
\boldsymbol{D}_{2}^{3 \times 2} & \boldsymbol{C}_{2}^{3 \times 3}
\end{array}\right]\left[\begin{array}{l}
\boldsymbol{\Psi}_{t}^{(0,1)} \\
\boldsymbol{\Psi}_{t}^{(0,2)}
\end{array}\right]+\left[\begin{array}{l}
\boldsymbol{b}^{2 \times 1} \\
\mathbf{0}^{3 \times 1}
\end{array}\right]
$$

with

$$
\boldsymbol{A}_{1}^{2 \times 2}=\boldsymbol{M}^{(0,1)}, \boldsymbol{C}_{2}^{3 \times 3}=\boldsymbol{M}^{(0,2)}, \boldsymbol{D}_{2}^{3 \times 2}=\boldsymbol{L}^{(0,2)}
$$

where the superscripts denote the dimensionality of the matrices. The matrices $\boldsymbol{M}^{(0,1)}$ and $\boldsymbol{M}^{(0,2)}$ are defined in Eqn. (4.44) and $\boldsymbol{L}^{(0,2)}$ is given in Eqn. (4.87). Further, the notation $\mathbf{0}^{k \times l} \in \mathbb{R}^{k \times l}$ represents the all-zeros matrix and $\boldsymbol{b}^{2 \times 1}=\left(\alpha_{1} \bar{\lambda}_{1}, \alpha_{2} \bar{\lambda}_{2}\right)^{\top}$. Observe that this stacked form contains the previously defined matrices as blocks. A similar form occurs for higher orders $n$ of the stacked vector $\left(\Psi_{t}^{(0,1)}, \ldots, \boldsymbol{\Psi}_{t}^{(0, n)}\right)^{\top}$, containing mixed moments of $\boldsymbol{\lambda}(t)$ up to order $n$.

Careful inspection of previous results reveals a nested sequence of block lower triangular matrices. Let $m_{n}:=n+1$ and define $\mathfrak{m}_{n}:=m_{1}+\cdots+m_{n}$. Then, consider a nested sequence of block lower triangular matrices $\left\{\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}\right\}_{n \in \mathbb{N}}$ given by

$$
\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}=\left[\begin{array}{cc}
\boldsymbol{A}_{n-1}^{\mathfrak{m}_{n-1} \times \mathfrak{m}_{n-1}} & \mathbf{0}_{n}^{\mathfrak{m}_{n-1} \times m_{n}}  \tag{4.53}\\
\boldsymbol{D}_{n}^{m_{n} \times \mathfrak{m}_{n-1}} & \boldsymbol{C}_{n}^{m_{n} \times m_{n}}
\end{array}\right],
$$

where $\boldsymbol{A}_{1}^{2 \times 2} \equiv \boldsymbol{C}_{1}^{2 \times 2}=\boldsymbol{M}^{(0,1)}, \boldsymbol{C}_{n}^{m_{n} \times m_{n}}=\boldsymbol{M}^{(0, n)}$ and $\boldsymbol{D}_{n}^{m_{n} \times \mathfrak{m}_{n-1}}=\boldsymbol{L}^{(0, n)}$. Clearly, the first term on the RHS of (4.52) occurs as a special case of (4.53) when $n=2$. Recall that we know the structure of the matrices $\boldsymbol{M}^{(0, n)}$ as given in Eqn. (4.44). In Appendix 4.C.4, we give some further details to the structure of $\boldsymbol{L}^{(0, n)}$ for the case $n=3$. Going back to the sequence of matrices $\left\{\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}\right\}_{n \in \mathbb{N}}$ as defined in (4.53), the stacked vector $\left(\mathbf{\Psi}_{t}^{(0,1)}, \mathbf{\Psi}_{t}^{(0,2)}, \ldots, \mathbf{\Psi}_{t}^{(0, n)}\right)^{\top}$ satisfies the following ODE

$$
\frac{\mathrm{d}}{\mathrm{~d} t}\left[\begin{array}{c}
\boldsymbol{\Psi}_{t}^{(0,1)}  \tag{4.54}\\
\vdots \\
\boldsymbol{\Psi}_{t}^{(0, n)}
\end{array}\right]=\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}\left[\begin{array}{c}
\boldsymbol{\Psi}_{t}^{(0,1)} \\
\vdots \\
\boldsymbol{\Psi}_{t}^{(0, n)}
\end{array}\right]+\left[\begin{array}{c}
\boldsymbol{b}^{2 \times 1} \\
\mathbf{0}^{\left(\mathfrak{m}_{n}-m_{1}\right) \times 1}
\end{array}\right]
$$

with initial condition $\left(\Psi_{0}^{(0,1)}, \ldots, \boldsymbol{\Psi}_{0}^{(0, n)}\right)^{\top}$.
Proposition 4.3. If $C_{i}^{m_{i} \times m_{i}}$ is invertible for all $i \in\{1, \ldots, n\}$, then

$$
\left[\begin{array}{c}
\boldsymbol{\Psi}_{t}^{(0,1)}  \tag{4.55}\\
\vdots \\
\boldsymbol{\Psi}_{t}^{(0, n)}
\end{array}\right]=e^{\boldsymbol{A}_{n}^{\boldsymbol{m}_{n} \times \mathfrak{m}_{n}} t}\left[\begin{array}{c}
\boldsymbol{\Psi}_{0}^{(0,1)} \\
\vdots \\
\boldsymbol{\Psi}_{0}^{(0, n)}
\end{array}\right]-\left(\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}\right)^{-1}\left(\boldsymbol{I}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}-e^{\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}} t}\right)\left[\begin{array}{c}
\boldsymbol{b}^{2 \times 1} \\
\mathbf{0}^{\left(\mathfrak{m}_{n}-m_{1}\right) \times 1}
\end{array}\right]
$$

Proof. Taking the time derivative of Eqn. (4.55) immediately yields Eqn. (4.54).
Proposition 4.3 allows for the simultaneous computation of the first $n$ transient moments of $\boldsymbol{\lambda}(t)=\left(\lambda_{1}(t), \lambda_{2}(t)\right)^{\top}$. However, we do need to compute the matrix exponential and the inverse of $\boldsymbol{A}_{n}^{\mathfrak{m}_{n} \times \mathfrak{m}_{n}}$.

This idea can be extended to consider the joint transient moments of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$ in the bivariate setting $d=2$. As we did before, we show the details of the first and second order, and show how this extends to higher order moments. For the first order moments, we consider the stacked vector $\boldsymbol{\Psi}_{t}^{(1)}$ stated here for completeness sake:

$$
\boldsymbol{\Psi}_{t}^{(1)}=\left(\mathbf{\Psi}_{t}^{(0,1)}, \mathbf{\Psi}_{t}^{(1,0)}\right)^{\top}=\left(\mathbb{E}\left[\lambda_{1}(t)\right], \mathbb{E}\left[\lambda_{2}(t)\right], \mathbb{E}\left[Q_{1}(t)\right], \mathbb{E}\left[Q_{2}(t)\right]\right)^{\top}
$$

Close inspection of the associated ODEs (given in Eqns. (4.85) and (4.86) in Appendix 4.C.1) yields that

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(1)}=\boldsymbol{F}_{1}^{4 \times 4} \boldsymbol{\Psi}_{t}^{(1)}+\left[\begin{array}{l}
\boldsymbol{b}^{2 \times 1}  \tag{4.56}\\
\mathbf{0}^{2 \times 1}
\end{array}\right]
$$

with

$$
\boldsymbol{F}_{1}^{4 \times 4}=\left[\begin{array}{cc}
\boldsymbol{M}^{(0,1)} & \boldsymbol{0}^{2 \times 2} \\
\boldsymbol{I}^{2 \times 2} & \boldsymbol{M}^{(1,0)}
\end{array}\right], \quad \boldsymbol{b}_{1}^{2 \times 1}=\left[\begin{array}{c}
\alpha_{1} \bar{\lambda}_{1} \\
\alpha_{2} \bar{\lambda}_{2}
\end{array}\right] .
$$

As before, the superscripts denote the dimensionality of the matrices, with $\boldsymbol{I}^{k \times l} \in \mathbb{R}^{k \times l}$ the identity matrix. Note the block lower triangular shape of $\boldsymbol{F}_{1}^{4 \times 4}$.

In fact, if we include the second order moments contained in the vector by considering $\left(\boldsymbol{\Psi}_{t}^{(1)}, \boldsymbol{\Psi}_{t}^{(2)}\right)^{\top}$, we can infer from the associated ODEs (given in Eqns.(4.87), (4.88) and (4.89) in Appendix 4.C.1), in combination with Eqn. (4.56) that

$$
\frac{\mathrm{d}}{\mathrm{~d} t}\left[\begin{array}{l}
\boldsymbol{\Psi}_{t}^{(1)}  \tag{4.57}\\
\boldsymbol{\Psi}_{t}^{(2)}
\end{array}\right]=\boldsymbol{F}_{2}^{14 \times 14}\left[\begin{array}{l}
\boldsymbol{\Psi}_{t}^{(1)} \\
\boldsymbol{\Psi}_{t}^{(2)}
\end{array}\right]+\left[\begin{array}{l}
\boldsymbol{b}_{1}^{2 \times 1} \\
0^{12 \times 1}
\end{array}\right],
$$

with $\boldsymbol{F}_{2}^{14 \times 14}$ again a lower triangular matrix, given by

$$
\boldsymbol{F}_{2}^{14 \times 14}=\left[\begin{array}{cc}
\boldsymbol{F}_{1}^{4 \times 4} & \mathbf{0}^{4 \times 10} \\
\boldsymbol{G}_{2}^{10 \times 4} & \boldsymbol{H}_{2}^{10 \times 10}
\end{array}\right],
$$

with the matrices contained in it defined by

$$
\begin{aligned}
\boldsymbol{G}_{2}^{10 \times 4} & =\left[\begin{array}{cc}
\boldsymbol{L}^{(0,2)} & \mathbf{0}^{3 \times 2} \\
\boldsymbol{L}_{\lambda}^{(1,1)} & \boldsymbol{L}_{Q}^{(1,1)} \\
\mathbf{0}^{3 \times 2} & \mathbf{0}^{3 \times 2}
\end{array}\right], \quad \boldsymbol{H}_{2}^{10 \times 10}=\left[\begin{array}{cc}
\boldsymbol{M}^{(0,2)} & \mathbf{0}^{3 \times 4} \\
\boldsymbol{K}^{(1,1)} & \boldsymbol{M}^{3 \times 3} \\
\mathbf{0}^{3 \times 3} & \boldsymbol{K}^{(1,1)} \\
\mathbf{0}^{3 \times 3} & \boldsymbol{M}^{(2,0)}
\end{array}\right], \\
\boldsymbol{L}_{\lambda}^{(1,1)} & =\left[\begin{array}{cc}
\mathbb{E}\left[B_{11}\right] & 0 \\
\mathbb{E}\left[B_{21}\right] & 0 \\
0 & \mathbb{E}\left[B_{12}\right] \\
0 & \mathbb{E}\left[B_{22}\right]
\end{array}\right], \quad \boldsymbol{L}_{Q}^{(1,1)}=\left[\begin{array}{cc}
\alpha_{1} \bar{\lambda}_{1} & 0 \\
\alpha_{2} \bar{\lambda}_{2} & 0 \\
0 & \alpha_{1} \bar{\lambda}_{1} \\
0 & \alpha_{2} \bar{\lambda}_{2}
\end{array}\right],
\end{aligned}
$$

where $\boldsymbol{L}^{(0,2)}$ and the matrices $\boldsymbol{K}^{(1,1)}$ and $\boldsymbol{K}^{(2,0)}$ are known, see Appendix 4.C.1.
Continuing in this fashion we can consider vectors of arbitrary length $n \in \mathbb{N}$, namely $\left(\boldsymbol{\Psi}_{t}^{(1)}, \boldsymbol{\Psi}_{t}^{(2)}, \ldots, \boldsymbol{\Psi}_{t}^{(n)}\right)^{\top}$. Recall that we know the size of the vector $p_{n} \equiv$ $\mathfrak{D}(2, n)=\left|\Psi_{t}^{(n)}\right|$ from Eqn. (4.40), which yields the size of the stacked vector $\mathfrak{p}_{n} \equiv$ $\left|\left(\boldsymbol{\Psi}_{t}^{(1)}, \boldsymbol{\Psi}_{t}^{(2)}, \ldots, \boldsymbol{\Psi}_{t}^{(n)}\right)^{\top}\right|$, by setting $\mathfrak{p}_{n}=p_{1}+\cdots+p_{n}$. Consider the sequence of matrices $\left\{\boldsymbol{F}_{n}^{\boldsymbol{p}_{n} \times \mathfrak{p}_{n}}\right\}_{n \in \mathbb{N}}$ given by

$$
\boldsymbol{F}_{n}^{\mathfrak{p}_{n} \times \mathfrak{p}_{n}}=\left[\begin{array}{cc}
\boldsymbol{F}_{n-1}^{\mathfrak{p}_{n-1} \times \mathfrak{p}_{n-1}} & \boldsymbol{0}^{\mathfrak{p}_{n-1} \times p_{n}} \\
\boldsymbol{G}_{n}^{p_{n} \times \mathfrak{p}_{n-1}} & \boldsymbol{H}_{n}^{p_{n} \times p_{n}}
\end{array}\right],
$$

with

$$
\boldsymbol{H}_{n}^{p_{n} \times p_{n}}=\left[\begin{array}{cccc}
\boldsymbol{M}^{(0, n)} & \mathbf{0} & \cdots & \mathbf{0} \\
\boldsymbol{K}^{(1, n-1)} & \boldsymbol{M}^{(1, n-1)} & \cdots & \mathbf{0} \\
\vdots & \ddots & \ddots & \vdots \\
\mathbf{0} & \cdots & \boldsymbol{K}^{(n, 0)} & \boldsymbol{M}^{(n, 0)}
\end{array}\right] .
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

The matrices $\boldsymbol{G}_{n}^{p_{n} \times \boldsymbol{p}_{n-1}}$ are not as elegantly expressed for general $n \in \mathbb{N}$, but can be explicitly obtained through Eqn. 4.21. We have used these matrices for $n=1,2$ to compute explicit moments in Appendix 4.C.2, and in Appendix 4.C.4 we give further details to these matrices for order $n=3$. Observe that Eqns. (4.56) and (4.57) are special cases. The stacked vector $\left(\boldsymbol{\Psi}_{t}^{(1)}, \boldsymbol{\Psi}_{t}^{(2)}, \ldots, \boldsymbol{\Psi}_{t}^{(n)}\right)^{\top}$ satisfies the ODE

$$
\frac{\mathrm{d}}{\mathrm{~d} t}\left[\begin{array}{c}
\boldsymbol{\Psi}_{t}^{(1)}  \tag{4.58}\\
\vdots \\
\boldsymbol{\Psi}_{t}^{(n)}
\end{array}\right]=\boldsymbol{F}_{n}^{\mathfrak{p}_{n} \times \mathfrak{p}_{n}}\left[\begin{array}{c}
\boldsymbol{\Psi}_{t}^{(1)} \\
\vdots \\
\boldsymbol{\Psi}_{t}^{(n)}
\end{array}\right]+\left[\begin{array}{c}
\boldsymbol{b}^{2 \times 1} \\
\mathbf{0}^{\left(\mathfrak{p}_{n}-p_{1}\right) \times 1}
\end{array}\right],
$$

with initial condition given by $\left(\mathbf{\Psi}_{0}^{(1)}, \mathbf{\Psi}_{0}^{(2)}, \ldots, \boldsymbol{\Psi}_{0}^{(n)}\right)^{\top}$. The solution is given in the following proposition.

Proposition 4.4. If $\boldsymbol{H}_{i}^{p_{i} \times p_{i}}$ is invertible for all $i \in\{1, \ldots, n\}$, then

$$
\left[\begin{array}{c}
\boldsymbol{\Psi}_{t}^{(1)}  \tag{4.59}\\
\vdots \\
\boldsymbol{\Psi}_{t}^{(n)}
\end{array}\right]=e^{\boldsymbol{F}_{n}^{\boldsymbol{p}_{n} \times \boldsymbol{p}_{n}} t}\left[\begin{array}{c}
\boldsymbol{\Psi}_{0}^{(1)} \\
\vdots \\
\boldsymbol{\Psi}_{0}^{(n)}
\end{array}\right]-\left(\boldsymbol{F}_{n}^{\mathfrak{p}_{n} \times \mathfrak{p}_{n}}\right)^{-1}\left(\boldsymbol{I}^{\mathfrak{p}_{n} \times \mathfrak{p}_{n}}-e^{\boldsymbol{F}_{n}^{\boldsymbol{p}_{n} \times \boldsymbol{p}_{n}} t}\right)\left[\begin{array}{c}
\boldsymbol{b}^{2 \times 1} \\
\mathbf{0}^{\left(\mathfrak{p}_{n}-p_{1}\right) \times 1}
\end{array}\right] .
$$

Proof. Taking the time derivative of Eqn. (4.59) immediately yields Eqn. (4.58).
The advantage of Proposition 4.4 is that the computation of joint moments does not require integration of matrix exponentials as in Proposition 4.1. By considering the joint moments, the matrix in the associated ODE is more complex, but we reduced the non-homogeneous part of the ODE in Eqn. (4.58) to a constant, which allows for an analytic expression of the solution. In particular, the run time of computing these transient moments does not increase in $t$, as opposed to the other computational methods. We will see in the numerical computations done in Section 4.7 what the advantage in terms of computation time is, compared to previously derived results.

### 4.6 The Nearly Unstable Behavior

In this section, we revisit the stability condition in Assumption 4.1 for the stationary distributions of $\boldsymbol{\lambda}$. More precisely, we analyze the behavior of $\boldsymbol{\lambda}$ when the spectral radius of $\boldsymbol{H}$ approaches 1, see Eqn. (4.4), i.e., when the system becomes nearly unstable. The situation of near instability is commonly encountered in real (financial) data, and deriving the corresponding limiting distributions provides an insightful analysis of the probabilistic behavior in such cases. The nearly unstable situation is also referred to as heavy traffic in OR and queueing theory.

We consider the general $d \in \mathbb{N}$ dimensional setting and focus first on results for $\boldsymbol{\lambda}=\left(\lambda_{1}, \ldots, \lambda_{d}\right)^{\top}$. For certain choices of symmetric parameters, we calculate the Laplace transform of $\boldsymbol{\lambda}$ and identify its limit. To that end, we introduce the notation $\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})$ for the Laplace transform of $\boldsymbol{\lambda}$ evaluated in $\boldsymbol{s}=\left(s_{1}, \ldots, s_{d}\right)^{\top} \in \mathbb{R}_{+}^{d}$, given explicitly by

$$
\begin{equation*}
\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=\mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}}\right]=\mathbb{E}\left[\prod_{i=1}^{d} e^{-s_{i} \lambda_{i}}\right] . \tag{4.60}
\end{equation*}
$$

We assume a specific parameterization for sake of tractability, namely that for each $i \in[d]$, we have

$$
\begin{equation*}
\alpha_{i}=\alpha \geqslant 0, \quad B_{1 i} \stackrel{d}{\xlongequal{d}} \ldots \stackrel{d}{=} B_{d i} \stackrel{d}{=} B_{i}, \quad \bar{\lambda}_{i}=\bar{\lambda}>0, \tag{4.61}
\end{equation*}
$$

where $B_{i}$ are independent non-negative random variables with $\mathbb{E}\left[B_{i}^{2}\right]<\infty$. This choice of parameters induces symmetry, since it implies that each component $\lambda_{i}$ has the same base rate $\bar{\lambda}$, the same decay rate $\alpha$, and that it is self- or cross-excited by all $B_{1}, \ldots, B_{d}$. Hence, this choice of parameters implies $\lambda_{i} \stackrel{d}{=} \lambda_{j}$ for any combination $i, j \in[d]$. The following result yields an explicit solution for $\mathcal{T}\{\boldsymbol{\lambda}\}(s)$.

Lemma 4.1. Assume Eqn. (4.61) and let $\beta_{i}(u)=\mathbb{E}\left[e^{-u B_{i}}\right]$ for any $u \geqslant 0$. Then, with $\bar{s}=s_{1}+\cdots+s_{d}$, we have

$$
\begin{equation*}
\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=\exp \left(-\alpha \bar{\lambda} \int_{0}^{\bar{s}} \frac{u}{\alpha u+\sum_{i=1}^{d} \beta_{i}(u)-d} \mathrm{~d} u\right) . \tag{4.62}
\end{equation*}
$$

Proof. From the PDE in Eqn. (4.19) and our assumptions on the parameters, we derive that $\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})$ satisfies the PDE

$$
\sum_{i=1}^{d}\left(\alpha s_{i}+\beta_{i}(\bar{s})-1\right) \frac{\mathrm{d}}{\mathrm{~d} s_{i}} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=-\alpha \bar{\lambda} \sum_{i=1}^{d} s_{i} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}),
$$

upon substituting $z_{1}=\cdots=z_{d}=1$. Further observe that for any $i, j \in[d]$, we have

$$
\frac{\mathrm{d}}{\mathrm{~d} s_{i}} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=\mathbb{E}\left[-\lambda_{i} e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}}\right]=\mathbb{E}\left[-\lambda_{j} e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}}\right]=\frac{\mathrm{d}}{\mathrm{~d} s_{j}} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}),
$$

since $\lambda_{i} \stackrel{d}{=} \lambda_{j}$, again by our assumptions on the parameters. Hence, we obtain the ODE

$$
\begin{equation*}
\left.\frac{\mathrm{d}}{\mathrm{~d} s_{1}} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=\frac{-\alpha \bar{\lambda} \bar{s}}{\alpha \bar{s}+\sum_{i=1}^{d} \beta_{i}(\bar{s})-d} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})=:-f\left(s_{1}, \ldots, s_{d}\right)\right) \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}), \tag{4.63}
\end{equation*}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

where we note that the choice of index on the LHS is arbitrary. The solution to this ODE is derived in a number of steps. First, we have that the solution may be expressed as

$$
\begin{equation*}
\log (\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}))=-\int_{0}^{s_{1}} f\left(u, s_{2}, \ldots, s_{d}\right) \mathrm{d} u+K, \quad K=\log \left(\mathcal{T}\{\boldsymbol{\lambda}\}\left(0, s_{2}, \ldots, s_{d}\right)\right) \tag{4.64}
\end{equation*}
$$

Second, observe that since $\lambda_{i} \stackrel{d}{=} \lambda_{j}$, the Laplace transforms of the marginals satisfy, for any $s \in \mathbb{R}_{+}$,

$$
\mathcal{T}\{\boldsymbol{\lambda}\}(s, 0, \ldots, 0)=\mathcal{T}\{\boldsymbol{\lambda}\}(0, s, 0, \ldots, 0)=\cdots=\mathcal{T}\{\boldsymbol{\lambda}\}(0,0, \ldots, s)
$$

We are then able to derive the joint Laplace transform of, say, $\left(\lambda_{1}, \lambda_{2}\right)^{\top}$ from Eqn. (4.64), since

$$
\begin{align*}
& \mathcal{T}\{\boldsymbol{\lambda}\}\left(s_{1}, s_{2}, \ldots, 0\right) \\
& =\exp \left(-\int_{0}^{s_{1}} f\left(u, s_{2}, \ldots, 0\right) \mathrm{d} u\right) \mathcal{T}\{\boldsymbol{\lambda}\}\left(0, s_{2}, \ldots, 0\right) \\
& =\exp \left(-\alpha \bar{\lambda} \int_{0}^{s_{1}} \frac{u+s_{2}}{\alpha\left(u+s_{2}\right)+\sum_{i=1}^{d} \beta_{i}\left(u+s_{2}\right)-d} \mathrm{~d} u\right) \\
& \quad \times \exp \left(-\alpha \bar{\lambda} \int_{0}^{s_{2}} \frac{u}{\alpha u+\sum_{i=1}^{d} \beta_{i}(u)-d} \mathrm{~d} u\right) \\
& =\exp \left(-\alpha \bar{\lambda}\left(\int_{s_{2}}^{s_{1}+s_{2}} \frac{v}{\alpha v+\sum_{i=1}^{d} \beta_{i}(v)-d} \mathrm{~d} v+\int_{0}^{s_{2}} \frac{u}{\alpha u+\sum_{i=1}^{d} \beta_{i}(u)-d} \mathrm{~d} u\right)\right) \\
& =\exp \left(-\alpha \bar{\lambda} \int_{0}^{s_{1}+s_{2}} \frac{v}{\alpha v+\sum_{i=1}^{d} \beta_{i}(v)-d} \mathrm{~d} v\right), \tag{4.65}
\end{align*}
$$

where we used a change of variable $v=u+s_{2}$. Note that by symmetry, we can do this for any pair $\left(\lambda_{i}, \lambda_{j}\right)^{\top}$, with $i, j \in[d]$. Iterating the derivation in Eqn. (4.65), we obtain the full solution

$$
\begin{aligned}
\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}) & =\exp \left(-\int_{0}^{s_{1}} f\left(u, s_{2}, \ldots, s_{d}\right) \mathrm{d} u\right) \mathcal{T}\{\boldsymbol{\lambda}\}\left(0, s_{2}, \ldots, s_{d}\right) \\
& =\exp \left(-\alpha \bar{\lambda} \int_{0}^{s_{1}+\cdots+s_{d}} \frac{u}{\alpha u+\sum_{i=1}^{d} \beta_{i}(u)-d} \mathrm{~d} u\right) .
\end{aligned}
$$

We now use this result to derive the desired limit result in the nearly unstable case. Observe that the stability condition of the matrix $\boldsymbol{C}$, see Eqn. (4.4), having a
maximum eigenvalue smaller than 1 , is in our setting explicitly given by

$$
\begin{equation*}
\theta:=\frac{1}{\alpha} \sum_{i=1}^{d} \mathbb{E}\left[B_{i}\right]<1 \tag{4.66}
\end{equation*}
$$

Further let $\sigma=2 \alpha\left(\sum_{i=1}^{d} \mathbb{E}\left[B_{i}^{2}\right]\right)^{-1}$.
Theorem 4.3. Assuming Eqn. (4.61), the heavy-traffic Laplace transform of $\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})$ is given by

$$
\begin{equation*}
\lim _{\theta \uparrow 1} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}(1-\theta))=\left(\frac{\sigma}{\sigma+\bar{s}}\right)^{\sigma \bar{\lambda}} \tag{4.67}
\end{equation*}
$$

Proof. The proof follows from the expression for $\mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s})$ in Eqn. (4.62), Taylor expanding the $\beta_{i}(\cdot)$, and computing the limit. Since the second moments of $B_{i}$ exist, we have $\beta_{i}(u)=1-u \mathbb{E}\left[B_{i}\right]+\frac{1}{2} u^{2} \mathbb{E}\left[B_{i}^{2}\right]+o\left(u^{2}\right)$ as $u \downarrow 0$. Substituting $\boldsymbol{s}(1-\theta)$ as the argument in Eqn. (4.62) and the Taylor expansion of $\beta_{i}(\cdot)$, we obtain as $\theta \uparrow 1$, that

$$
\begin{aligned}
& \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}(1-\theta)) \\
& =\exp \left(-\alpha \bar{\lambda} \int_{0}^{\bar{s}} \frac{u(1-\theta)}{\alpha u(1-\theta)+\sum_{i=1}^{d} \beta_{i}(u(1-\theta))-d}(1-\theta) \mathrm{d} u\right) \\
& =\exp \left(-\bar{\lambda} \int_{0}^{\bar{s}} \frac{\alpha(1-\theta)}{\alpha-\sum_{i=1}^{d} \mathbb{E}\left[B_{i}\right]+\frac{u}{2}(1-\theta) \sum_{i=1}^{d} \mathbb{E}\left[B_{i}^{2}\right]+o(1-\theta)} \mathrm{d} u\right) \\
& =\exp \left(-\bar{\lambda} \int_{0}^{\bar{s}} \frac{1}{1+\frac{u}{2 \alpha} \sum_{i=1}^{d} \mathbb{E}\left[B_{i}^{2}\right]+o(1)} \mathrm{d} u\right) .
\end{aligned}
$$

Finally, by definition of $\sigma$ and an elementary computation, we have

$$
\lim _{\theta \uparrow 1} \mathcal{T}\{\boldsymbol{\lambda}\}(\boldsymbol{s}(1-\theta))=\lim _{\theta \uparrow 1} \exp \left(-\bar{\lambda} \int_{0}^{\bar{s}} \frac{1}{1+u \sigma^{-1}+o(1)} \mathrm{d} u\right)=\left(\frac{\sigma}{\sigma+\bar{s}}\right)^{\sigma \bar{\lambda}}
$$

One can recognize the familiar shape of the Laplace transform of a (multivariate) Gamma distribution. Since there are multiple ways of defining a multivariate Gamma distribution, we describe a straightforward way that takes into account the covariance between components. More precisely, we have for any $i, j \in[d]$ that

$$
\begin{equation*}
\lim _{\theta \uparrow 1} \operatorname{Cov}\left((1-\theta) \lambda_{i},(1-\theta) \lambda_{j}\right)=\bar{\lambda} / \sigma, \tag{4.68}
\end{equation*}
$$

by virtue of Eqn. (4.67). This yields the following result.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

Corollary 4.2. Assume Eqn. (4.61) holds. Then, for a random vector $\boldsymbol{X}$, we have as $\theta \uparrow 1$

$$
(1-\theta) \boldsymbol{\lambda} \xrightarrow{d} \boldsymbol{X},
$$

where each marginal $X_{i} \sim \Gamma(\sigma \bar{\lambda}, \sigma)$ and $\operatorname{Cov}\left(X_{i}, X_{j}\right)=\bar{\lambda} / \sigma$ for any $i, j \in[d]$.
Proof. Follows immediately from Theorem 4.3 combined with Lévy's continuity theorem. The covariance expression follows from Eqn. (4.68).

### 4.7 Numerical Experiments

The primary goal of the findings of the previous sections is to numerically evaluate moments. In this section we compare the resulting output, in terms of efficiency and accuracy, to alternatives such as finite difference schemes and Monte Carlo simulations. All coding has been performed in Python. The underlying code can be obtained upon request.

The first subsection focuses on computing moments of order $n=1$ and $n=2$ in the $d$-dimensional setting of Section 4.3.2. In the trivariate setting $d=3$ we compute first and second order moments by evaluating the solutions of the ODEs derived in Section 4.3.2, which are used as a benchmark to compare the alternative methods to. In the second subsection, the bivariate setting $d=2$ is considered, where we apply the results of the nested block-matrix developed in Section 4.5, allowing us to compute moments up to any order $n \in \mathbb{N}$. By this method, relying on analytical closed-form expression, we compute moments of order up to three, which are then used as a benchmark.

Throughout this section, two alternative numerical evaluation techniques are used. The first is based on finite differences (FD). We perform numerical differentiation of the relevant transform, using the characterization in Theorem 4.1 and Corollary 4.1. The moments of interest can be obtained by appropriately differentiating the joint transform with respect to $s$ and $\boldsymbol{z}$ and then setting $\boldsymbol{s}=\mathbf{0}$ and $\boldsymbol{z}=\mathbf{1}$. In this approach these derivatives are approximated by the corresponding (central) finite differences. The precision of these approximations depends on the 'width parameter' $h>0$. The second alternative technique is based on Monte Carlo simulation (MC). To simulate the Hawkes process we use an algorithm based on Ogata's thinning algorithm; see [63] and [60, Algorithm 1.21] for details. The sampling mechanism is based on the cluster representation of [55, Definition 2]. The performance of this method depends on the number of runs $m \in \mathbb{N}$.

The performance of the various approaches is quantified in terms of run time and error (relative to the benchmark, abbreviated by BM). Two types of errors are distinguished, namely the Mean Absolute Error (MAE) and Mean Relative Error (MRE):

$$
\begin{equation*}
\mathrm{MAE}=\sum_{j=1}^{k}\left|m_{j}^{(\mathrm{BM})}-m_{j}^{(\mathrm{FD} / \mathrm{MC})}\right|, \quad \mathrm{MRE}=\sum_{j=1}^{k} \frac{\left|m_{j}^{(\mathrm{BM})}-m_{j}^{(\mathrm{FD} / \mathrm{MC})}\right|}{m_{j}^{(\mathrm{BM})}}, \tag{4.69}
\end{equation*}
$$

where $m_{j}$ denotes our BM value of the $j$-th moment, $k$ is the number of moments computed, and the superscript indicates the underlying computational method.

### 4.7.1 Multivariate

In the trivariate $d=3$ setting we numerically evaluate the first and second order moments. While the methods of Section 4.3 .2 work for any $d \in \mathbb{N}$, we focus on $d=3$. We compute (joint) moments of $\boldsymbol{Q}(t)=\left(Q_{1}(t), Q_{2}(t), Q_{3}(t)\right)$ and $\boldsymbol{\lambda}(t)=$ $\left(\lambda_{1}(t), \lambda_{2}(t), \lambda_{3}(t)\right)$. We start by taking $t=5$, later we study the impact of $t$.

We let the random marks be exponentially distributed: for any combination $i, j \in$ $\{1,2,3\}$, we set $B_{i j} \sim \operatorname{Exp}\left(b_{i j}\right)$ for some $b_{i j}>0$. For simplicity we assume independence among the $B_{i j}$. In the experiments we take
$\overline{\boldsymbol{\lambda}}=\left[\begin{array}{c}0.3 \\ 1 \\ 0.5\end{array}\right], \quad \mathbb{E} \boldsymbol{B}=\left[\begin{array}{ccc}0.5 & 0.3 & 0.4 \\ 0.7 & 0.5 & 0.5 \\ 0.4 & 0.2 & 0.5\end{array}\right], \quad \boldsymbol{D}_{\alpha}=\left[\begin{array}{ccc}2 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 2.5\end{array}\right], \quad \boldsymbol{D}_{\mu}=\left[\begin{array}{ccc}1.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1\end{array}\right]$.
One readily verifies that for these parameters the stability condition of Assumption 4.1 is met.

Recall that the stacked vector $\boldsymbol{\Sigma}_{t}^{(1)}$ and the stacked matrix $\boldsymbol{\Sigma}_{t}^{(2)}$ contain all the first order moments and combinations of second order moments, respectively; see Eqns. (4.24) and (4.27). In this subsection, the benchmark BM corresponds to the solution of the vector- and matrix-valued ODEs as given in Eqns. (4.25) and (4.28), obtained using the SciPy package in Python. We used the default precision of the SciPy ODE solver; the output presented in the next subsection indicates that this provides sufficiently precise results.

Table 4.1, displaying the resulting run times and errors, quantifies the superior performance of our approach. The table shows that the ODE method is faster than the FD method, and orders of magnitude faster than the Monte Carlo simulation, where the latter method in addition typically yields substantial errors. Run times are reliably estimated by taking the average of sufficiently many experiments. We see that in the FD method smaller values of $h$ lead to lower run times: in this method, we vary the arguments $\boldsymbol{s}$ and $\boldsymbol{z}$ with $h$ when evaluating the joint transform, which is faster for

## Chapter 4 Markovian Multivariate Hawkes Population Processes

smaller values of $h$. Further, observe that the MAE and MRE are not monotone in $h$ : for larger $h$ the derivative is poorly approximated by the finite difference, while for smaller $h$ numerical stability issues have a detrimental effect. There is an optimal width where the error is smallest, which in our instance happens to be around $h=10^{-3}$.

| $n$ | BM | FD |  |  |  | MC |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | RT | $h$ | RT | MAE | MRE | $m$ | RT | MAE | MRE |
| 1 | $7.17 \cdot 10^{-3}$ | $10^{-2}$ | $5.26 \cdot 10^{-2}$ | $1.27 \cdot 10^{-3}$ | $3.42 \cdot 10^{-4}$ | $10^{2}$ | 9 | $8.92 \cdot 10^{-1}$ | $4.98 \cdot 10^{-1}$ |
|  | . | $10^{-3}$ | $3.94 \cdot 10^{-2}$ | $1.81 \cdot 10^{-4}$ | $9.84 \cdot 10^{-5}$ | $10^{3}$ | 96 | $2.41 \cdot 10^{-1}$ | $1.43 \cdot 10^{-1}$ |
|  | . | $10^{-4}$ | $3.35 \cdot 10^{-2}$ | $6.60 \cdot 10^{-4}$ | $3.69 \cdot 10^{-4}$ | $10^{4}$ | 956 | $6.87 \cdot 10^{-2}$ | $3.68 \cdot 10^{-2}$ |
| 2 | $9.65 \cdot 10^{-2}$ | $10^{-2}$ | $3.29 \cdot 10^{-1}$ | $9.96 \cdot 10^{-2}$ | $1.17 \cdot 10^{-2}$ | $10^{2}$ | 13 | $2.48 \cdot 10^{1}$ | $4.56 \cdot 10^{0}$ |
|  |  | $10^{-3}$ | $2.62 \cdot 10^{-1}$ | $2.03 \cdot 10^{-3}$ | $4.18 \cdot 10^{-4}$ | $10^{3}$ | 115 | $3.89 \cdot 10^{0}$ | $7.69 \cdot 10^{-1}$ |
|  |  | $10^{-4}$ | $2.20 \cdot 10^{-1}$ | $1.13 \cdot 10^{-2}$ | $1.18 \cdot 10^{-3}$ | $10^{4}$ | 1025 | $9.07 \cdot 10^{-1}$ | $2.42 \cdot 10^{-1}$ |

Table 4.1. Run times (RT) in seconds and errors (MAE, MRE) for first ( $n=1$ ) and second ( $n=2$ ) order moments in the trivariate setting: performance of the benchmark ODE-based method relative to FD and MC.

To assess whether the effects observed in the previous experiment hold in general, we have performed experiments with a set of intrinsically different parameter settings. In these experiments, we study run times and errors, while we fix the 'width parameter' $h \equiv 10^{-3}$ and number of simulation runs $m \equiv 10^{3}$. Since varying each entry in each of the vectors and matrices would lead to a large set of instances, we decided to focus on altering only the parameters directly pertaining to $\lambda_{1}(\cdot)$ and $Q_{1}(\cdot)$, while respecting Assumption 4.1. Note that the effect will propagate to other components due to crossexcitation. Table 4.2 shows the resulting run times and errors. The main conclusion is that the experiments reveal that, uniformly across all instances, the benchmark ODE-based method remains the fastest, with a run time that is hardly affected by the parameters chosen. We note that increasing the value of $\mathbb{E}\left[B_{11}\right]$ or decreasing the value of $\alpha_{1}$ results in the system approaching the boundary of the stability condition in Assumption 4.1, thus leading to larger relative errors.

We also studied the effect of varying the time parameter $t$ on the run times. Recall that the FD method uses the (conditional) joint transform, where the latter requires solving systems of ODEs. Figure 4.1 shows that the run times of the ODE-based method and the FD method scale effectively linearly with $t$, with the ODE method having the smallest slope. The run time for MC increases superlinearly; we took its logarithm to be able to show it in the same plot. This superlinear behavior is an inherent consequence of the branching structure underlying the Hawkes process. We note that this qualitative behavior is observed for all choices of parameters that we considered.

| Parameter | BM | FD |  |  | MC |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | RT | RT | MAE | MRE | RT | MAE | MRE |
| $\bar{\lambda}_{1}=3$ | $1.13 \cdot 10^{-1}$ | $3.04 \cdot 10^{-1}$ | $2.48 \cdot 10^{-2}$ | $7.44 \cdot 10^{-4}$ | 286 | $7.19 \cdot 10^{0}$ | $3.34 \cdot 10^{-1}$ |
| $\bar{\lambda}_{1}=5$ | $1.20 \cdot 10^{-1}$ | $3.10 \cdot 10^{-1}$ | $7.68 \cdot 10^{-2}$ | $1.02 \cdot 10^{-3}$ | 432 | $1.47 \cdot 10^{1}$ | $3.98 \cdot 10^{-1}$ |
| $\bar{\lambda}_{1}=10$ | $1.36 \cdot 10^{-1}$ | $3.32 \cdot 10^{-1}$ | $5.37 \cdot 10^{-1}$ | $2.09 \cdot 10^{-3}$ | 812 | $1.18 \cdot 10^{2}$ | $6.19 \cdot 10^{-1}$ |
| $\mathbb{E} B_{11}=1$ | $1.04 \cdot 10^{-1}$ | $3.05 \cdot 10^{-1}$ | $1.02 \cdot 10^{-2}$ | $1.14 \cdot 10^{-3}$ | 118 | $5.48 \cdot 10^{0}$ | $7.55 \cdot 10^{-1}$ |
| $\mathbb{E} B_{11}=1.3$ | $1.06 \cdot 10^{-1}$ | $3.07 \cdot 10^{-1}$ | $1.52 \cdot 10^{-2}$ | $1.07 \cdot 10^{-3}$ | 131 | $1.12 \cdot 10^{1}$ | $9.38 \cdot 10^{-1}$ |
| $\mathbb{E} B_{11}=1.6$ | $1.02 \cdot 10^{-1}$ | $3.08 \cdot 10^{-1}$ | $1.19 \cdot 10^{-1}$ | $2.62 \cdot 10^{-3}$ | 152 | $8.72 \cdot 10^{1}$ | $2.38 \cdot 10^{0}$ |
| $\alpha_{1}=1$ | $1.03 \cdot 10^{-1}$ | $2.97 \cdot 10^{-1}$ | $7.91 \cdot 10^{-3}$ | $5.11 \cdot 10^{-3}$ | 130 | $1.29 \cdot 10^{1}$ | $1.09 \cdot 10^{0}$ |
| $\alpha_{1}=3$ | $1.11 \cdot 10^{-1}$ | $3.27 \cdot 10^{-1}$ | $2.89 \cdot 10^{-3}$ | $1.92 \cdot 10^{-3}$ | 93 | $5.85 \cdot 10^{0}$ | $2.04 \cdot 10^{0}$ |
| $\alpha_{1}=10$ | $1.73 \cdot 10^{-1}$ | $5.45 \cdot 10^{-1}$ | $1.14 \cdot 10^{-2}$ | $8.89 \cdot 10^{-3}$ | 80 | $2.29 \cdot 10^{0}$ | $9.83 \cdot 10^{-1}$ |
| $\mu_{1}=0.5$ | $1.01 \cdot 10^{-1}$ | $3.01 \cdot 10^{-1}$ | $2.49 \cdot 10^{-3}$ | $5.05 \cdot 10^{-4}$ | 98 | $3.84 \cdot 10^{0}$ | $6.30 \cdot 10^{-1}$ |
| $\mu_{1}=2$ | $1.06 \cdot 10^{-1}$ | $3.04 \cdot 10^{-1}$ | $2.23 \cdot 10^{-3}$ | $5.82 \cdot 10^{-4}$ | 97 | $4.98 \cdot 10^{0}$ | $1.25 \cdot 10^{0}$ |
| $\mu_{1}=5$ | $1.26 \cdot 10^{-1}$ | $3.18 \cdot 10^{-1}$ | $2.42 \cdot 10^{-3}$ | $1.12 \cdot 10^{-3}$ | 99 | $5.22 \cdot 10^{0}$ | $1.45 \cdot 10^{0}$ |

Table 4.2. Run times ( $R T$ ) in seconds and errors (MAE, MRE) for combined first and second order moments in the trivariate setting: effect of parameter changes of the benchmark ODE-based method relative to FD and MC.

### 4.7.2 Bivariate

In this subsection, we compute for the bivariate setting $(d=2)$ the transient moments of $\boldsymbol{Q}(t)=\left(Q_{1}(t), Q_{2}(t)\right)$ and $\boldsymbol{\lambda}(t)=\left(\lambda_{1}(t), \lambda_{2}(t)\right)$, of orders $n=1,2,3$. Again we take $t=5$, but later assess the effect of the choice of $t$. As before the random marks are exponentially distributed, i.e., for $i, j \in\{1,2\}$, we set $B_{i j} \sim \operatorname{Exp}\left(b_{i j}\right)$ for some $b_{i j}>0$, with independence between the $B_{i j}$. The parameters are

$$
\overline{\boldsymbol{\lambda}}=\left[\begin{array}{l}
0.5 \\
0.5
\end{array}\right], \quad \mathbb{E} \boldsymbol{B}=\left[\begin{array}{cc}
1.5 & 0.5 \\
0.75 & 1.25
\end{array}\right], \quad \boldsymbol{D}_{\alpha}=\left[\begin{array}{ll}
3 & 0 \\
0 & 2
\end{array}\right], \quad \boldsymbol{D}_{\mu}=\left[\begin{array}{ll}
1 & 0 \\
0 & 2
\end{array}\right] .
$$

The stability condition of Eqn. (4.4) is met, which in the bivariate setting reads

$$
\begin{equation*}
\left(\alpha_{1}-\mathbb{E}\left[B_{11}\right]\right)\left(\alpha_{2}-\mathbb{E}\left[B_{22}\right]\right)>\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right] . \tag{4.70}
\end{equation*}
$$

We compute all the first, second and third order moments, i.e., all entries of the stacked vectors $\boldsymbol{\Psi}_{t}^{(1)}, \boldsymbol{\Psi}_{t}^{(2)}$, and $\boldsymbol{\Psi}_{t}^{(3)}$. As our benchmark serves the main result of Section 4.5, namely Proposition 4.4. This result exploits the block-matrix structure, by which we can simultaneously compute moments of multiple orders, thus greatly increasing the computational performance. As it turns out, the value difference with the ODE-based approach of the previous subsection is negligible (i.e., in the order of $10^{-8}$ ). The difference in computational effort, however, is substantial: for this instance the block-matrix method is about 200 times faster.

Table 4.3 shows that the BM method is much faster than FD and MC, especially for second and third order moments. We also see that the absolute and relative errors of


Figure 4.1. Run times of the BM, FD, and MC method for moments up to second order in the trivariate setting.

FD and MC significantly grow as the order of moments increase. Particularly for the third order moments, the poor stability of the FD method significantly degrades the performance, as can be seen by the variability of the error when changing the precision parameter $h$.

| $n$ | BM | FD |  |  |  | MC |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | RT | $h$ | RT | MAE | MRE | $m$ | RT | MAE | MRE |
| 1 | $4.77 \cdot 10^{-4}$ | $10^{-2}$ | $4.42 \cdot 10^{-2}$ | $2.55 \cdot 10^{-3}$ | $1.12 \cdot 10^{-3}$ | $10^{2}$ | 5 | $7.16 \cdot 10^{-1}$ | $3.32 \cdot 10^{-1}$ |
|  |  | $10^{-3}$ | $3.31 \cdot 10^{-2}$ | $8.55 \cdot 10^{-5}$ | $4.61 \cdot 10^{-5}$ | $10^{3}$ | 62 | $3.19 \cdot 10^{-1}$ | $1.49 \cdot 10^{-1}$ |
|  |  | $10^{-4}$ | $2.97 \cdot 10^{-2}$ | $4.38 \cdot 10^{-4}$ | $2.64 \cdot 10^{-4}$ | $10^{4}$ | 589 | $1.69 \cdot 10^{-2}$ | $7.30 \cdot 10^{-3}$ |
| 2 | $5.61 \cdot 10^{-4}$ | $10^{-2}$ | $8.12 \cdot 10^{-2}$ | $2.70 \cdot 10^{-1}$ | $2.45 \cdot 10^{-2}$ | $10^{2}$ | 6 | $2.98 \cdot 10^{1}$ | $2.76 \cdot 10^{0}$ |
|  |  | $10^{-3}$ | $6.42 \cdot 10^{-2}$ | $2.03 \cdot 10^{-3}$ | $4.18 \cdot 10^{-4}$ | $10^{3}$ | 67 | $8.01 \cdot 10^{0}$ | $9.71 \cdot 10^{-1}$ |
|  | . | $10^{-4}$ | $5.73 \cdot 10^{-2}$ | $1.14 \cdot 10^{-2}$ | $1.38 \cdot 10^{-3}$ | $10^{4}$ | 631 | $5.36 \cdot 10^{0}$ | $6.92 \cdot 10^{-1}$ |
| 3 | $9.26 \cdot 10^{-4}$ | $10^{-2}$ | $2.49 \cdot 10^{-1}$ | $2.23 \cdot 10^{2}$ | $1.77 \cdot 10^{0}$ | $10^{2}$ | 7 | $9.57 \cdot 10^{2}$ | $1.09 \cdot 10^{1}$ |
|  |  | $10^{-3}$ | $2.07 \cdot 10^{-1}$ | $2.99 \cdot 10^{2}$ | $5.61 \cdot 10^{1}$ | $10^{3}$ | 69 | $2.21 \cdot 10^{2}$ | $2.87 \cdot 10^{0}$ |
|  |  | $10^{-4}$ | $1.72 \cdot 10^{-1}$ | $1.34 \cdot 10^{6}$ | $3.34 \cdot 10^{4}$ | $10^{4}$ | 639 | $3.98 \cdot 10^{1}$ | $4.47 \cdot 10^{-1}$ |

Table 4.3. Run times ( $R T$ ) in seconds and errors (MAE, MRE) for first ( $n=1$ ), second $(n=2)$, and third $(n=3)$ order moments in the bivariate setting: comparison of the benchmark block-matrix method relative to FD and MC.

Figure 4.2 shows the effect of varying the time parameter $t$. As before, the run time of FD scales linearly with $t$, and that of MC superlinearly. This should be contrasted with the fact that the block-matrix method has the attractive feature that its run time does not depend on $t$.

We now consider the block-matrix stationary moments. The first and second order stationary moments can be immediately obtained from the results of Section 4.3.2,


Figure 4.2. Run times of the BM, FD, and MC method for moments up to third order in the bivariate setting.
by solving the associated Sylvester matrix equations. We now study the obvious alternatives when not knowing these stationary moments, based on picking a 'large' value of $t$ in the FD and MC methods. These methods have two intrinsic drawbacks: (1) run times increase in $t$, and (2) we do not know a priori what value of $t$ guarantees that the error made is sufficiently small. As we already saw that MC is typically outperformed by FD, we focus on FD only. To select a sufficiently large value of $t$, we compute the FD-based approximation of first and second order transient moments for successive integer values of $t$, until the difference of the respective MREs is smaller than some given threshold $\epsilon$. We compare the resulting approximation to our benchmark, i.e., the values obtained by solving the Sylvester matrix equations, so as to quantify the error made.

Table 4.4 presents the results for the FD method with precision level $\epsilon=0.01$. We have performed the above procedure for different choices of parameters, where the parameter that is altered is given in the table. Note that the benchmark method is exact and provides near-instant response. Observe that for specific sets of parameters there is a substantial effect on the value of $t$ (the time at which the procedure terminates, that is), the run time, and the MRE, in particular when the parameters are close to the boundary of the stability condition in Assumption 4.1, e.g., $\mathbb{E}\left[B_{11}\right]=2.25$ or $\alpha_{1}=2.1$.

We proceed by studying the numerical evaluation of the objects featured in Section 4.3.1. From the joint transform characterization in Theorem 4.2, we compute the mixed moments, for any $t \geqslant 0, \tau>0$ and any combination $i, j=1,2$,

$$
\begin{equation*}
\mathbb{E}\left[Q_{i}(t) Q_{j}(t+\tau)\right], \quad \mathbb{E}\left[\lambda_{i}(t) \lambda_{j}(t+\tau)\right], \tag{4.71}
\end{equation*}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

|  | FD |  |  |
| :---: | :---: | :---: | :---: |
| Parameter | $t$ | RT | MRE |
| $\bar{\lambda}_{1}=3$ | 18 | $3.25 \cdot 10^{0}$ | $2.17 \cdot 10^{-2}$ |
| $\bar{\lambda}_{1}=10$ | 18 | $3.38 \cdot 10^{0}$ | $1.94 \cdot 10^{-2}$ |
| $\mathbb{E} B_{11}=0.5$ | 14 | $2.75 \cdot 10^{0}$ | $1.51 \cdot 10^{-2}$ |
| $\mathbb{E} B_{11}=2.25$ | 43 | $1.08 \cdot 10^{1}$ | $6.87 \cdot 10^{-2}$ |
| $\alpha_{1}=2.1$ | 86 | $2.89 \cdot 10^{1}$ | $1.58 \cdot 10^{-1}$ |
| $\alpha_{1}=5$ | 15 | $3.77 \cdot 10^{0}$ | $1.55 \cdot 10^{-2}$ |
| $\mu_{1}=0.5$ | 19 | $3.97 \cdot 10^{0}$ | $2.11 \cdot 10^{-2}$ |
| $\mu_{1}=5$ | 18 | $3.38 \cdot 10^{0}$ | $2.84 \cdot 10^{-2}$ |

Table 4.4. Run times ( $R T$ ) in seconds until first and second order transient moments approximate stationary moments in the bivariate setting: $F D$ method with precision $\epsilon=0.01$.
as before using finite differences. It is noted that along the same lines objects of the type $\mathbb{E}\left[Q_{i}(t) \lambda_{j}(t+\tau)\right]$ can be evaluated, and in addition various types of autocovariances and auto-correlations (cf. [22] for the auto-covariance in the market microstructure setting). The joint transform characterization allows for efficient and fast computation of these cross-moments, also for large $t>0$, which makes it practical in these settings.

To assess the performance, we have conducted a numerical experiment with the same parameters as earlier in this subsection. To analyze the effect of the $\tau$ parameter in (4.71), in Figure 4.3 we fix $t=1.5$ and plot the quantities of interest as functions of $\tau$. The solid lines are the moments computed by applying FD to the joint transform, and the dotted lines represent the results from the MC method (based on $10^{4}$ runs). We see that MC performs increasingly poorly as $\tau$ increases, in particular for the population processes $Q_{i}(\cdot)$, which is due to the fact that there are more events (i.e., arrivals and departures) for larger $\tau$, and hence more variation. Further, the different shapes in the plots indicate that the effect of $\tau$ on the specific cross-moment depends on the chosen parameters.

We now turn our attention to the effect of the initial values. In Theorem 4.1, we characterized the joint transform with the processes being initialized at $\boldsymbol{Q}\left(t_{0}\right)=$ $\left(Q_{1}\left(t_{0}\right), Q_{2}\left(t_{0}\right)\right)=\left(q_{1,0}, q_{2,0}\right) \in \mathbb{N}^{2}$ and $\boldsymbol{\lambda}\left(t_{0}\right)=\left(\lambda_{1}\left(t_{0}\right), \lambda_{2}\left(t_{0}\right)\right)=\left(\lambda_{1,0}, \lambda_{2,0}\right) \in \mathbb{R}_{+}^{2}$ for some $t_{0}>0$. By applying FD, we can compute the moments of our interest for any initial value, for instance

$$
\begin{equation*}
\mathbb{E}\left[Q_{i}(t) \mid Q_{i}\left(t_{0}\right)=q_{i, 0}\right], \quad \mathbb{E}\left[\lambda_{i}(t) \mid \lambda_{i}\left(t_{0}\right)=\lambda_{i, 0}\right] \tag{4.72}
\end{equation*}
$$

with $i=1,2$, where $q_{i, 0} \in \mathbb{N}$ and $\lambda_{i, 0} \in \mathbb{R}_{+}$. In our experiment we focus on the moments of $Q_{1}(\cdot)$ and $\lambda_{1}(\cdot)$, studying the effect of three different choices of $q_{i, 0}$ and


Figure 4.3. Computation of cross-moments for $t=1.5$ and $\tau \in[0,10]$ using the joint transform characterization (solid lines) compared to Monte Carlo simulated averages (dashed lines)
$\lambda_{i, 0}$ on the first moments and variances as a function of $t$. Note that a different value of $Q_{i}\left(t_{0}\right)=q_{i, 0}$ will not influence $\lambda_{j}(\cdot)$, since the population processes do not directly affect the intensity processes, but due to mutual excitation, the values $\lambda_{1}\left(t_{0}\right)=$ $\lambda_{1,0}$ and $\lambda_{2}\left(t_{0}\right)=\lambda_{2,0}$ do matter. When computing $\mathbb{E}\left[Q_{1}(t) \mid Q_{1}\left(t_{0}\right)=q_{1,0}\right]$, we leave $\lambda_{i}\left(t_{0}\right)=\lambda_{i}(0)=\bar{\lambda}_{i}$ for $i=1,2$, and we only change $q_{1,0}$. Similarly, when computing $\mathbb{E}\left[\lambda_{1}(t) \mid \lambda_{1}\left(t_{0}\right)=\lambda_{1,0}\right]$, we leave $\lambda_{2}\left(t_{0}\right)=\lambda_{2}(0)=\bar{\lambda}_{2}$ and only change $\lambda_{1,0}$.

Figure 4.4 shows the expectations and variances, where we introduced compact notation $\mathbb{E}\left[Q_{1}(t) \mid q_{1,0}\right]=\mathbb{E}\left[Q_{1}(t) \mid Q_{1}\left(t_{0}\right)=q_{1,0}\right]$ and $\mathbb{E}\left[\lambda_{1}(t) \mid \lambda_{1,0}\right]=\mathbb{E}\left[\lambda_{1}(t) \mid \lambda_{1}\left(t_{0}\right)=\right.$ $\left.\lambda_{1,0}\right]$; similarly for the variances. For the moments of $Q_{1}(t)$, we observe a vertical translation of the plots, which is expected since the arrived individuals depart independently and according to the same distribution. For the moments of $\lambda_{1}(t)$, we see that the effect of the $\lambda_{1,0}$-value is substantial. For both the mean and the variance there is convergence to their respective steady-state values.

### 4.8 Concluding Remarks

This paper has studied multivariate Hawkes-fed Markovian infinite-server queues, which can be alternatively interpreted as population processes. Our objective was to devise accurate and efficient algorithms to compute transient and stationary moments.


Figure 4.4. Computation of expected values and variances of $Q_{1}(t)$ and $\lambda_{1}(t)$, with different initial values at $t_{0}=2$, with $t \in\left[t_{0}, 10\right]$, using the joint transform characterization.

We succeeded in doing so, heavily relying on having access to the joint transform of the Hawkes intensity process and the population process. When the multivariate Hawkes process is of general dimension $d$, this transform is expressed in terms of systems of ODEs, allowing for the computation of joint moments. This includes joint moments where the components pertain to the same as well as to different points in time, thus covering also the evaluation of the processes' autocovariance functions. We then proceeded with finding expressions for the first and second order, transient and stationary moments for the $d$-dimensional processes. Next, in the 2-dimensional setting we derived a recursive procedure, revealing a block-matrix structure for the computation of moments of any order. Our numerical experiments show that our approach outperforms its alternatives: it provides highly accurate, near instant response.

## 4.A Proofs

Proof of Theorem 4.1. The proof is comprised of a number of steps. First, we use the Markov property on the distribution function of the joint process $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$, next we take partial derivatives to obtain an expression for the density. Then, we consecutively apply the Laplace and $\boldsymbol{z}$-transform to obtain a PDE. Finally, we use the method of characteristics to obtain a system of ODEs. We describe below the main steps; some technical details are relegated to the Appendix.

We note that the probabilities considered in this proof are conditional on the value of the processes at time $t_{0}$, i.e. $\boldsymbol{Q}\left(t_{0}\right)=\boldsymbol{Q}_{0}$ and $\boldsymbol{\lambda}\left(t_{0}\right)=\boldsymbol{\lambda}_{0}$. To start off, for $t \in \mathbb{R}_{+}$, $\boldsymbol{k} \in \mathbb{N}_{+}^{d}$ and $\boldsymbol{\nu} \in \mathbb{R}_{+}^{d}$, set

$$
F(t, \boldsymbol{\nu}, \boldsymbol{k})=\mathbb{P}(\boldsymbol{\lambda}(t) \leqslant \boldsymbol{\nu}, \boldsymbol{Q}(t)=\boldsymbol{k}), \quad \frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \boldsymbol{\nu}}=\left[\begin{array}{c}
\frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{1}}  \tag{4.73}\\
\vdots \\
\frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{d}}
\end{array}\right]
$$

Also define

$$
\begin{equation*}
f(t, \boldsymbol{\nu}, \boldsymbol{k})=\frac{\partial^{d} F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{1} \cdots \partial \nu_{d}}, \tag{4.74}
\end{equation*}
$$

as the joint density of $(\boldsymbol{Q}(t), \boldsymbol{\lambda}(t))$. For some $\delta>0$, consider the probability

$$
\begin{equation*}
F(t+\delta, \boldsymbol{\nu}-\boldsymbol{\alpha} \odot(\boldsymbol{\nu}-\overline{\boldsymbol{\lambda}}) \delta, \boldsymbol{k}) \tag{4.75}
\end{equation*}
$$

where the interpretation of the term $\boldsymbol{\nu}-\boldsymbol{\alpha} \odot(\boldsymbol{\nu}-\overline{\boldsymbol{\lambda}})$ is a decay factor, in the sense that for small $\delta$, no new arrival of a point in $(t, t+\delta$ ] makes the intensity $\boldsymbol{\lambda}(\cdot)$ decay with rate $\boldsymbol{\alpha}$ back to the mean reversion level $\overline{\boldsymbol{\lambda}}$.

To compute this probability, we apply the Markov property to Eqn. (4.75), which leaves us to consider the possibilities to get in the state of exactly $\boldsymbol{k}$ active points and intensity equal to $\boldsymbol{\nu}-\boldsymbol{\alpha} \odot(\boldsymbol{\nu}-\overline{\boldsymbol{\lambda}})$. There are three distinct ways to get to this state at time $t+\delta$ from time $t$ : we have exactly $\boldsymbol{k}$ active points with no arrivals or departures; we have $\boldsymbol{k}-\boldsymbol{e}_{j}$ active points and exactly one arrival in component $j$; or we have $\boldsymbol{k}+\boldsymbol{e}_{j}$ active points and one departure in component $j$. This yields up to $o(\delta)$ terms that

$$
\begin{aligned}
F & (t+\delta, \boldsymbol{\nu}-\boldsymbol{\alpha} \odot(\boldsymbol{\nu}-\overline{\boldsymbol{\lambda}}) \delta, \boldsymbol{k}) \\
= & \sum_{j=1}^{d} \int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}} \delta y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathbb{P}\left(\boldsymbol{B}_{j} \leqslant \boldsymbol{\nu}-\boldsymbol{y}\right) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d} \\
& +\sum_{j=1}^{d}\left(k_{j}+1\right) \delta \mu_{j} F\left(t, \boldsymbol{\nu}, \boldsymbol{k}+\boldsymbol{e}_{j}\right) \\
& +\int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}}\left(1-\sum_{j=1}^{d} \delta \mu_{j} k_{j}-\sum_{j=1}^{d} \delta y_{j}\right) f(t, \boldsymbol{y}, \boldsymbol{k}) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d}+o(\delta) .
\end{aligned}
$$

Subtracting $F(t, \boldsymbol{k}, \boldsymbol{\lambda})$ on both sides, dividing by $\delta$ and taking $\delta \downarrow 0$ yields

$$
\frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial t}-(\boldsymbol{\alpha} \odot(\boldsymbol{\nu}-\overline{\boldsymbol{\lambda}}))^{\top} \frac{\partial F(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \boldsymbol{\nu}}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

$$
\begin{aligned}
= & \sum_{j=1}^{d} \int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}} y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathbb{P}\left(\boldsymbol{B}_{j} \leqslant \boldsymbol{\nu}-\boldsymbol{y}\right) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d} \\
& +\sum_{j=1}^{d}\left(k_{j}+1\right) \mu_{j} F\left(t, \boldsymbol{\nu}, \boldsymbol{k}+\boldsymbol{e}_{j}\right) \\
& -\sum_{j=1}^{d} \int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}}\left(\mu_{j} k_{j}+y_{j}\right) f(t, \boldsymbol{y}, \boldsymbol{k}) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d}
\end{aligned}
$$

where the LHS follows from the definition of the directional derivative. Next, we take the partial derivatives with respect to $\nu_{1}, \ldots, \nu_{d}$, so as to rewrite above equation in terms of the probability density function $f(t, \boldsymbol{\nu}, \boldsymbol{k})$. By the definitions of $F$ and $f$ given in Eqns. (4.73) and (4.74), and using Leibniz' integral rule on the integral terms, we obtain

$$
\begin{align*}
& \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial t}-\sum_{j=1}^{d} \alpha_{j} \frac{\partial}{\partial \nu_{j}} \nu_{j} f(t, \boldsymbol{\nu}, \boldsymbol{k})+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}} \\
& =\sum_{j=1}^{d} \int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}} y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \frac{\partial^{d}}{\partial \nu_{1} \cdots \partial \nu_{d}} \mathbb{P}\left(\boldsymbol{B}_{j} \leqslant \boldsymbol{\nu}-\boldsymbol{y}\right) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d}  \tag{4.76}\\
& \quad+\sum_{j=1}^{d}\left(f\left(t, \boldsymbol{\nu}, \boldsymbol{k}+\boldsymbol{e}_{j}\right)\left(k_{j}+1\right) \mu_{j}-f(t, \boldsymbol{\nu}, \boldsymbol{k})\left(k_{j} \mu_{j}+\nu_{j}\right)\right) .
\end{align*}
$$

Denote the $d$-dimensional Laplace transform with respect to $\boldsymbol{\nu}$ by

$$
\xi(t, \boldsymbol{s}, \boldsymbol{k}):=\mathcal{L}(f(t, \boldsymbol{\nu}, \boldsymbol{k}))(\boldsymbol{s})=\int_{0}^{\infty} \cdots \int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \nu_{1} \cdots \mathrm{~d} \nu_{d} .
$$

Taking the Laplace transform of Eqn. (4.76) yields

$$
\begin{align*}
& \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial t}+\sum_{j=1}^{d} \alpha_{j} s_{j} \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}}+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} s_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})  \tag{4.77}\\
& =\sum_{j=1}^{d}\left(-\frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \beta_{j}(\boldsymbol{s})+\left(k_{j}+1\right) \mu_{j} \xi\left(t, \boldsymbol{s}, \boldsymbol{k}+\boldsymbol{e}_{j}\right)-k_{j} \mu_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})+\frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}}\right)
\end{align*}
$$

see Appendix 4.B. 1 for the term-by-term derivation. The computations boil down to applying integration by parts, convolution arguments and the properties of $F$ and $f$.

Rewriting the expression to have all derivatives on one side yields

$$
\begin{align*}
& \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial t}+\sum_{j=1}^{d}\left(\alpha_{j} s_{j}-1\right) \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}}+\sum_{j=1}^{d} \frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \beta_{j}(\boldsymbol{s})  \tag{4.78}\\
& =\sum_{j=1}^{d}\left(\left(k_{j}+1\right) \mu_{j} \xi\left(t, \boldsymbol{s}, \boldsymbol{k}+\boldsymbol{e}_{j}\right)-k_{j} \mu_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})-\alpha_{j} \bar{\lambda}_{j} s_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})\right)
\end{align*}
$$

Next, we rewrite equation (4.78) by taking the $\boldsymbol{z}$-transform, which gives us the joint transform introduced in Eqn. (4.6), i.e.,
$\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z}):=\mathcal{Z}(\xi(t, \boldsymbol{s}, \boldsymbol{k}))(\boldsymbol{z})=\sum_{k_{1}=0}^{\infty} \cdots \sum_{k_{d}=0}^{\infty} z_{1}^{k_{1}} \cdots z_{d}^{k_{d}} \xi(t, \boldsymbol{s}, \boldsymbol{k})=\mathbb{E}_{t_{0}}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{i=1}^{d} z_{i}^{Q_{i}(t)}\right]$,
yielding

$$
\begin{align*}
\frac{\partial \zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})}{\partial t} & +\sum_{j=1}^{d}\left(\alpha_{j} s_{j}+z_{j} \beta_{j}(\boldsymbol{s})-1\right) \frac{\partial \zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})}{\partial s_{j}}+\sum_{j=1}^{d}\left(\mu_{j}\left(z_{j}-1\right)\right) \frac{\partial \zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})}{\partial z_{j}} \\
& =-\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z}) \sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} s_{j} \tag{4.79}
\end{align*}
$$

where we added the subscript $t_{0}$ to emphasize the dependence on this initial time value. We refer to Appendix 4.B. 1 for the term-by-term derivation.

By employing the method of characteristics, we can rewrite the PDE in Eqn. (4.79) into a system of ODEs. To that end, consider a curve in $\mathbb{R}^{2 d}$ parameterized by $(\widehat{\boldsymbol{s}}(u), \widehat{\boldsymbol{z}}(u))$ as a function of $u$, where $t_{0} \leqslant u \leqslant t$, which terminates at the set of parameters $(s, \boldsymbol{z})$, i.e. $(\widehat{\boldsymbol{s}}(t), \widehat{\boldsymbol{z}}(t))=(\boldsymbol{s}, \boldsymbol{z})$. Since we have a first-order PDE, we easily obtain the characteristic system of ODEs by

$$
\begin{align*}
& \frac{\mathrm{d} \widehat{s}_{j}(u)}{\mathrm{d} u}=\alpha_{j} \widehat{s}_{j}(u)+\widehat{z}_{j}(u) \beta_{j}(\widehat{\boldsymbol{s}}(u))-1, \\
& \frac{\mathrm{~d} \widehat{z}_{j}(u)}{\mathrm{d} u}=\mu_{j}\left(\widehat{z}_{j}(u)-1\right), \tag{4.80}
\end{align*}
$$

for each $j \in[d]$. For $\widehat{z}_{j}(\cdot)$, the solution can be directly computed as

$$
\widehat{z}_{j}(u)=1+C_{j} e^{u \mu_{j}},
$$

where $C_{j}$ is derived by the boundary condition $\widehat{z}_{j}(t)=z_{j}$, yielding $C_{j}=\left(z_{j}-1\right) e^{-t \mu_{j}}$, and thus $\widehat{z}_{j}(u)=1+\left(z_{j}-1\right) e^{-\mu_{j}(t-u)}$. Upon substituting the solution for $\widehat{z}_{j}(u)$ in the equation of $\widehat{s}_{j}(u)$ in (4.80), we obtain the ODE

$$
\begin{equation*}
-\frac{\mathrm{d} \widehat{s}_{j}(u)}{\mathrm{d} u}+\alpha_{j} \widehat{s}_{j}(u)+\left(1+\left(z_{j}-1\right) e^{-\mu_{j}(t-u)}\right) \beta_{j}(\widehat{\boldsymbol{s}}(u))-1=0 \tag{4.81}
\end{equation*}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

with terminal condition $\widehat{s}_{j}(t)=s_{j}$. For later purposes, we rephrase the ODE in Eqn. (4.81) into an ODE subject to an initial condition. To that end, let $v=t_{0}+t-u$ such that $t_{0} \leqslant v \leqslant t$ and the ODE for $\widehat{s}_{j}(\cdot)$ becomes

$$
\frac{\mathrm{d} \widehat{s}_{j}\left(t_{0}+t-v\right)}{\mathrm{d} v}+\alpha_{j} \widehat{s}_{j}\left(t_{0}+t-v\right)+\left(1+\left(z_{j}-1\right) e^{-\mu_{j}\left(v-t_{0}\right)} \beta_{j}\left(\widehat{\boldsymbol{s}}\left(t_{0}+t-v\right)\right)-1=0\right.
$$

Upon defining $\tilde{s}_{j}(v)=\widehat{s}_{j}\left(t_{0}+t-v\right)$, we have that $\tilde{s}_{j}(\cdot)$ satisfies Eqn. (4.9), with initial condition $\tilde{s}_{j}\left(t_{0}\right)=\widehat{s}_{j}(t)=s_{j}$.

We can now solve the characteristic equation of $\zeta_{t_{0}}(\cdot)$. Since the original PDE in Eqn. (4.79) is inhomogeneous, we know the solution $\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})$ is not constant along characteristics, but evolves according to the RHS of (4.79). Therefore, if we set $\widehat{\zeta}_{t_{0}}(u):=\zeta_{t_{0}}(u, \widehat{\boldsymbol{s}}(u), \widehat{\boldsymbol{z}}(u))$ to be the solution restricted to the characteristics, then $\widehat{\zeta}_{t_{0}}(\cdot)$ satisfies

$$
\frac{\partial \widehat{\zeta}_{t_{0}}(u)}{\partial u}=-\widehat{\zeta}_{t_{0}}(u) \sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} \widehat{s}_{j}(u)
$$

subject to initial condition

$$
\widehat{\zeta}_{t_{0}}\left(t_{0}\right)=\zeta_{t_{0}}\left(t_{0}, \widehat{\boldsymbol{z}}\left(t_{0}\right), \widehat{\boldsymbol{s}}\left(t_{0}\right)\right)=\prod_{j=1}^{d} \widehat{z}_{j}\left(t_{0}\right)^{Q_{j, 0}} \exp \left(-\widehat{s}_{j}\left(t_{0}\right) \lambda_{j, 0}\right)
$$

Solving this yields

$$
\widehat{\zeta}_{t_{0}}(u)=\prod_{j=1}^{d} \widehat{z}_{j}\left(t_{0}\right)^{Q_{j, 0}} \exp \left(-\widehat{s}_{j}\left(t_{0}\right) \lambda_{j, 0}-\alpha_{j} \bar{\lambda}_{j} \int_{t_{0}}^{u} \widehat{s}_{j}(v) \mathrm{d} v\right) .
$$

Finally, the solution of the PDE is given at the endpoint of the characteristic $(t, s, \boldsymbol{z})$, implying that $\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z})=\widehat{\zeta}_{t_{0}}(t)$. Using the relation $\widehat{s}_{j}\left(t_{0}\right)=\tilde{s}_{j}(t)$, we obtain

$$
\begin{aligned}
\zeta_{t_{0}}(t, \boldsymbol{s}, \boldsymbol{z}) & =\prod_{j=1}^{d} \widehat{z}_{j}\left(t_{0}\right)^{Q_{j, 0}} \exp \left(-\widehat{s}_{j}\left(t_{0}\right) \lambda_{j, 0}-\alpha_{j} \bar{\lambda}_{j} \int_{0}^{t} \widehat{s}_{j}\left(t_{0}+t-u\right) \mathrm{d} u\right) \\
& =\prod_{j=1}^{d} \widehat{z}_{j}\left(t_{0}\right)^{Q_{j, 0}} \exp \left(-\tilde{s}_{j}(t) \lambda_{j, 0}-\alpha_{j} \bar{\lambda}_{j} \int_{t_{0}}^{t} \tilde{s}_{j}(u) \mathrm{d} u\right),
\end{aligned}
$$

which finishes the proof.
Proof of Theorem 4.2. The proof follows by conditioning on $\boldsymbol{Q}(t)$ and $\boldsymbol{\lambda}(t)$, using the tower property, and then applying Theorem 4.1 and techniques from the proof. To
start, we have

$$
\begin{align*}
& \mathbb{E}\left[\prod_{i=1}^{d} y_{i}^{Q_{i}(t)} e^{-r_{i} \lambda_{i}(t)} z_{i}^{Q_{i}(t+\tau)} e^{-s_{i} \lambda_{i}(t+\tau)}\right] \\
& =\mathbb{E}\left[\prod_{i=1}^{d} y_{i}^{Q_{i}(t)} e^{-r_{i} \lambda_{i}(t)} \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t+\tau)} e^{-s_{i} \lambda_{i}(t+\tau)} \mid \boldsymbol{Q}(t), \boldsymbol{\lambda}(t)\right]\right], \tag{4.82}
\end{align*}
$$

by the tower property. The inner expectation can be derived from Theorem 4.1 and is given by

$$
\begin{aligned}
& \mathbb{E}\left[\prod_{i=1}^{d} z_{i}^{Q_{i}(t+\tau)} e^{-s_{i} \lambda_{i}(t+\tau)} \mid \boldsymbol{Q}(t), \boldsymbol{\lambda}(t)\right] \\
& =\prod_{j=1}^{d} \widehat{z}_{j}(t)^{Q_{j}(t)} e^{-\tilde{s}_{j}(t+\tau) \lambda_{j}(t)} \exp \left(-\bar{\lambda}_{j} \alpha_{j} \int_{t}^{t+\tau} \tilde{s}_{j}(u) \mathrm{d} u\right),
\end{aligned}
$$

where $\widehat{z}_{j}(\cdot)$ and $\tilde{s}_{j}(\cdot)$ satisfy Eqn. (4.15). Substituting this back in Eqn. (4.82) yields

$$
\begin{aligned}
& \mathbb{E}\left[\prod_{i=1}^{d} y_{i}^{Q_{i}(t)} e^{-r_{i} \lambda_{i}(t)} \prod_{j=1}^{d} \widehat{z}_{j}(t)^{Q_{j}(t)} e^{-\tilde{s}_{j}(t+\tau) \lambda_{j}(t)} \exp \left(-\bar{\lambda}_{j} \alpha_{j} \int_{t}^{t+\tau} \tilde{s}_{j}(u) \mathrm{d} u\right)\right] \\
& =\mathbb{E}\left[\prod_{j=1}^{d}\left(y_{j} \widehat{z}_{j}(t)\right)^{Q_{j}(t)} e^{-\left(r_{j}+\tilde{s}_{j}(t+\tau)\right) \lambda_{j}(t)}\right] \prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \alpha_{j} \int_{t}^{t+\tau} \tilde{s}_{j}(u) \mathrm{d} u\right) \\
& =\zeta(t, \boldsymbol{y} \odot \widehat{\boldsymbol{z}}(t), \boldsymbol{r}+\tilde{\boldsymbol{s}}(t+\tau)) \prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \alpha_{j} \int_{t}^{t+\tau} \tilde{s}_{j}(u) \mathrm{d} u\right) .
\end{aligned}
$$

Applying Corollary 4.1 to the $\zeta(\cdot)$ term on the RHS, specifically Eqn. (4.10), we obtain

$$
\zeta(t, \boldsymbol{y} \odot \widehat{\boldsymbol{z}}(t), \boldsymbol{r}+\tilde{\boldsymbol{s}}(t+\tau))=\prod_{j=1}^{d} \exp \left(-\bar{\lambda}_{j} \tilde{r}_{j}(t)-\bar{\lambda}_{j} \alpha_{j} \int_{0}^{t} \tilde{r}_{j}(v) \mathrm{d} v\right),
$$

where $\tilde{r}_{j}(\cdot)$ satisfies, for each $j \in[d]$, the ODE

$$
\frac{\mathrm{d} \tilde{r}_{j}(v)}{\mathrm{d} v}+\alpha_{j} \tilde{r}_{j}(v)+\left(1+\left(y_{j} \widehat{z}_{j}(t)-1\right) e^{-\mu_{j} v}\right) \beta(\tilde{\boldsymbol{r}}(v))-1=0 .
$$

Since $\widehat{z}_{j}(t)=1+\left(z_{j}-1\right) e^{-\mu_{j} \tau}$, substituting this into the ODE for $\tilde{r}_{j}(\cdot)$ and rearranging terms finishes the proof.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

## 4.B Computations

## 4.B.1 Transform computations

In this section, we provide the details behind taking the Laplace and $\boldsymbol{z}$-transform of Eqns. (4.76) and (4.77) respectively. First we show the Laplace transform, denoted by $\mathcal{L}(\cdot)$, of (4.76), which we restate here for convenience

$$
\begin{aligned}
& \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial t}-\sum_{j=1}^{d} \alpha_{j} \frac{\partial}{\partial \nu_{j}} \nu_{j} f(t, \boldsymbol{\nu}, \boldsymbol{k})+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}} \\
& =\sum_{j=1}^{d} \int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}} y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \frac{\partial^{d}}{\partial \nu_{1} \cdots \partial \nu_{d}} \mathbb{P}\left(\boldsymbol{B}_{j} \leqslant \boldsymbol{\nu}-\boldsymbol{y}\right) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d} \\
& \quad+\sum_{j=1}^{d}\left(f\left(t, \boldsymbol{\nu}, \boldsymbol{k}+\boldsymbol{e}_{j}\right)\left(k_{j}+1\right) \mu_{j}-f(t, \boldsymbol{\nu}, \boldsymbol{k})\left(k_{j} \mu_{j}+\nu_{j}\right)\right)
\end{aligned}
$$

and we introduce the shorthand notation
$\xi(t, \boldsymbol{s}, \boldsymbol{k}):=\mathcal{L}(f(t, \boldsymbol{\nu}, \boldsymbol{k}))(\boldsymbol{s})=\int_{0}^{\infty} \cdots \int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \nu_{1} \cdots \mathrm{~d} \nu_{d} \equiv \int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu}$.
We consider the term-by-term derivation in the equation that yields the transformed version as given in Eqn. (4.77). For the first term, it is clear that

$$
\mathcal{L}\left(\frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial t}\right)(s)=\frac{\partial \xi(t, \boldsymbol{k}, \boldsymbol{s})}{\partial t} .
$$

For the second term, we need to show

$$
-\mathcal{L}\left(\sum_{j=1}^{d} \alpha_{j} \frac{\partial}{\partial \nu_{j}} \nu_{j} f(t, \boldsymbol{\nu}, \boldsymbol{k})\right)(\boldsymbol{s})=\sum_{j=1}^{d} \alpha_{j} s_{j} \frac{\partial}{\partial s_{j}} \xi(t, \boldsymbol{k}, \boldsymbol{s}) .
$$

The argument of the Laplace transform is

$$
\sum_{j=1}^{d} \alpha_{j} \frac{\partial}{\partial \nu_{j}} \nu_{j} f(t, \boldsymbol{\nu}, \boldsymbol{k})=\sum_{j=1}^{d} \alpha_{j} f(t, \boldsymbol{\nu}, \boldsymbol{k})+\sum_{j=1}^{d} \alpha_{j} \nu_{j} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}} .
$$

We then use the linearity of the Laplace transform and apply integration by parts to obtain

$$
\sum_{j=1}^{d} \alpha_{j} \mathcal{L}(f(t, \boldsymbol{\nu}, \boldsymbol{k}))(\boldsymbol{s})+\sum_{j=1}^{d} \alpha_{j} \mathcal{L}\left(\nu_{j} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}}\right)(\boldsymbol{s})
$$

$$
\begin{aligned}
& =\sum_{j=1}^{d} \alpha_{j} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu}+\sum_{j=1}^{d} \alpha_{j} \int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}_{j}} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}} \mathrm{~d} \boldsymbol{\nu} \\
& =\sum_{j=1}^{d} \alpha_{j} \int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu}+\sum_{j=1}^{d} \alpha_{j}\left[\nu_{j} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k})\right]_{\mathbf{0}}^{\infty} \\
& -\sum_{j=1}^{d} \alpha_{j} \int_{\mathbf{0}}^{\infty}\left(1-\nu_{j} s_{j}\right) e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu} \\
& =0+\sum_{j=1}^{d} \alpha_{j} s_{j} \int_{\mathbf{0}}^{\infty} \nu_{j} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu} \\
& =\sum_{j=1}^{d} \alpha_{j} s_{j} \int_{\mathbf{0}}^{\infty}-\frac{\partial}{\partial s_{j}} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu} \\
& =-\sum_{j=1}^{d} \alpha_{j} s_{j} \frac{\partial}{\partial s_{j}} \xi(t, \boldsymbol{k}, \boldsymbol{s}) .
\end{aligned}
$$

For the third term, we need to show

$$
\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} \mathcal{L}\left(\frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}}\right)(\boldsymbol{s})=\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} s_{j} \xi(t, \boldsymbol{k}, \boldsymbol{s})
$$

Using integration by parts and that $f(t, \boldsymbol{k}, \mathbf{0})=0$, we have

$$
\begin{aligned}
\mathcal{L}\left(\frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}}\right)(\boldsymbol{s}) & =\int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} \frac{\partial f(t, \boldsymbol{\nu}, \boldsymbol{k})}{\partial \nu_{j}} \mathrm{~d} \boldsymbol{\nu} \\
& =\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k})\right]_{0}^{\infty}+s_{j} s_{j} \int_{0}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu} \\
& =0+s_{j} \xi(t, \boldsymbol{k}, \boldsymbol{s}) .
\end{aligned}
$$

For the fourth term, let us denote the probability density function of $\boldsymbol{B}_{j}=\left(B_{1 j}, \ldots, B_{d j}\right)^{\top}$ by $h_{j} \equiv h_{\boldsymbol{B}_{j}}$. We need to show

$$
\begin{aligned}
& \sum_{j=1}^{d} \mathcal{L}\left(\int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}} y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \frac{\partial^{d}}{\partial \nu_{1} \cdots \partial \nu_{d}} \mathbb{P}\left(\boldsymbol{B}_{j} \leqslant \boldsymbol{\nu}-\boldsymbol{y}\right) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d}\right) \\
& =\sum_{j=1}^{d} \mathcal{L}\left(\int_{0}^{\nu} y_{j} h_{j}(\boldsymbol{\nu}-\boldsymbol{y}) f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathrm{d} \boldsymbol{y}\right) \\
& =-\sum_{j=1}^{d} \beta_{j}(\boldsymbol{s}) \frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}},
\end{aligned}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

with $\beta_{j}(\boldsymbol{s})=\mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right]$. To show that this holds, we need the property that relates convolutions with integration, which states that

$$
\int_{\mathbb{R}^{d}}(f * g)(\boldsymbol{x}) \mathrm{d} \boldsymbol{x}=\left(\int_{\mathbb{R}^{d}} f(\boldsymbol{x}) \mathrm{d} \boldsymbol{x}\right)\left(\int_{\mathbb{R}^{d}} g(\boldsymbol{x}) \mathrm{d} \boldsymbol{x}\right),
$$

for given integrable functions $f$ and $g$. Using this, we have

$$
\begin{aligned}
& -\sum_{j=1}^{d} \beta_{j}(\boldsymbol{s}) \frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \\
& =-\sum_{j=1}^{d} \beta_{j}(\boldsymbol{s}) \int_{\mathbf{0}}^{\infty} \frac{\partial}{\partial s_{j}} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f\left(t, \boldsymbol{\nu}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathrm{d} \boldsymbol{\nu} \\
& =\sum_{j=1}^{d} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}^{2}} h_{j}(\boldsymbol{\nu}) \mathrm{d} \boldsymbol{\nu} \int_{\mathbf{0}}^{\infty} \nu_{j} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f\left(t, \boldsymbol{\nu}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathrm{d} \boldsymbol{\nu} \\
& =\sum_{j=1}^{d} \int_{\mathbf{0}}^{\infty} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^{\top}(\boldsymbol{\nu}-\boldsymbol{y})} h_{j}(\boldsymbol{\nu}-\boldsymbol{y}) y_{j} e^{-\boldsymbol{s}^{\top} \boldsymbol{y}} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathrm{d} \boldsymbol{y} \mathrm{~d} \boldsymbol{\nu} \\
& \stackrel{(\star)}{=} \sum_{j=1}^{d} \int_{\mathbf{0}}^{\infty} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} \int_{\mathbf{0}}^{\nu} h_{j}(\boldsymbol{\nu}-\boldsymbol{y}) y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \mathrm{d} \boldsymbol{y} \mathrm{~d} \boldsymbol{\nu} \\
& =\sum_{j=1}^{d} \mathcal{L}\left(\int_{0}^{\nu_{d}} \cdots \int_{0}^{\nu_{1}} y_{j} f\left(t, \boldsymbol{y}, \boldsymbol{k}-\boldsymbol{e}_{j}\right) \frac{\partial^{d}}{\partial \nu_{1} \cdots \partial \nu_{d}} \mathbb{P}\left(\boldsymbol{B}_{j} \leqslant \boldsymbol{\nu}-\boldsymbol{y}\right) \mathrm{d} y_{1} \cdots \mathrm{~d} y_{d}\right),
\end{aligned}
$$

where $(\star)$ holds because the non negativity $\mathbb{P}\left(\boldsymbol{B}_{j} \geqslant \mathbf{0}\right)=1$ implies $h_{j}(\boldsymbol{\nu}-\boldsymbol{y})=0$ if $\boldsymbol{\nu} \leqslant \boldsymbol{y}$.

The fifth term follows immediately by linearity since

$$
\sum_{j=1}^{d} \mathcal{L}\left(f\left(t, \boldsymbol{\nu}, \boldsymbol{k}+\boldsymbol{e}_{j}\right)\left(k_{j}+1\right) \mu_{j}\right)(\boldsymbol{s})=\sum_{j=1}^{d}\left(k_{j}+1\right) \mu_{j} \xi\left(t, \boldsymbol{s}, \boldsymbol{k}+\boldsymbol{e}_{j}\right) .
$$

Finally, the sixth term follows from the elementary computation

$$
\begin{aligned}
-\sum_{j=1}^{d} \mathcal{L}\left(f(t, \boldsymbol{\nu}, \boldsymbol{k})\left(k_{j} \mu_{j}+\nu_{j}\right)\right)(\boldsymbol{s}) & =-\sum_{j=1}^{d} k_{j} \mu_{j} \mathcal{L}\left(f(t, \boldsymbol{\nu}, \boldsymbol{k})(\boldsymbol{s})-\sum_{j=1}^{d} k_{j} \nu_{j} \mathcal{L}(f(t, \boldsymbol{\nu}, \boldsymbol{k}))(\boldsymbol{s})\right. \\
& =-\sum_{j=1}^{d} k_{j} \mu_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})+\sum_{j=1}^{d} \int_{0}^{\infty} \frac{\partial}{\partial s_{j}} e^{-\boldsymbol{s}^{\top} \boldsymbol{\nu}} f(t, \boldsymbol{\nu}, \boldsymbol{k}) \mathrm{d} \boldsymbol{\nu} \\
& =-\sum_{j=1}^{d} k_{j} \mu_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})+\sum_{j=1}^{d} \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}}
\end{aligned}
$$

We now show the Zeta transform, denoted by $\mathcal{Z}(\cdot)$ of Eqn. (4.78), restated here for convenience

$$
\begin{aligned}
& \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial t}+\sum_{j=1}^{d}\left(\alpha_{j} s_{j}-1\right) \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}}+\sum_{j=1}^{d} \frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \beta_{j}(\boldsymbol{s}) \\
& =\sum_{j=1}^{d}\left(\left(k_{j}+1\right) \mu_{j} \xi\left(t, \boldsymbol{s}, \boldsymbol{k}+\boldsymbol{e}_{j}\right)-k_{j} \mu_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})-\alpha_{j} \bar{\lambda}_{j} s_{j} \xi(t, \boldsymbol{s}, \boldsymbol{k})\right) .
\end{aligned}
$$

We introduce the shorthand notation

$$
\zeta(t, \boldsymbol{s}, \boldsymbol{z})=\mathcal{Z}(\xi(t, \boldsymbol{s}, \boldsymbol{k}))(\boldsymbol{z})=\sum_{k_{1}=0}^{\infty} \cdots \sum_{k_{d}=0}^{\infty} z_{1}^{k_{1}} \cdots z_{d}^{k_{d}} \xi(t, \boldsymbol{s}, \boldsymbol{k}) \equiv \sum_{\boldsymbol{k} \in \mathbb{N}_{0}^{d}} \boldsymbol{z}^{\boldsymbol{k}} \xi(t, \boldsymbol{s}, \boldsymbol{k})
$$

with $\mathbb{N}_{0}^{d}=\{0,1,2, \ldots\}^{d}$. As before, we take the term-by-term zeta transformation and show that we obtain Eqn. (4.79). The first and second terms are immediate by construction and linearity, since we have

$$
\begin{aligned}
\mathcal{Z}\left(\frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial t}\right)(\boldsymbol{z}) & =\frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial t} \\
\sum_{j=1}^{d} \mathcal{Z}\left(\alpha_{j} s_{j} \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}}\right)(\boldsymbol{z}) & =\sum_{j=1}^{d}\left(\alpha_{j} s_{j}-1\right) \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial s_{j}} .
\end{aligned}
$$

For the third term, we need to show

$$
\sum_{j=1}^{d} \mathcal{Z}\left(\frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \beta_{j}(\boldsymbol{s})\right)(\boldsymbol{z})=\sum_{j=1}^{d} z_{j} \beta_{j}(\boldsymbol{s}) \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial s_{j}}
$$

With the notation $\mathbb{N}_{j}^{d}=\left\{\boldsymbol{n} \in \mathbb{N}^{d}: n_{j} \geqslant 1\right\}$ for $j \in[d]$, we write out the left hand side to obtain

$$
\begin{aligned}
\sum_{j=1}^{d} \mathcal{Z}\left(\frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \beta_{j}(\boldsymbol{s})\right)(\boldsymbol{z}) & =\sum_{j=1}^{d} \sum_{\boldsymbol{k} \in \mathbb{N}_{j}^{d}} \boldsymbol{z}^{\boldsymbol{k}} \beta_{j}(\boldsymbol{s}) \frac{\partial \xi\left(t, \boldsymbol{s}, \boldsymbol{k}-\boldsymbol{e}_{j}\right)}{\partial s_{j}} \\
& =\sum_{j=1}^{d} \beta_{j}(\boldsymbol{s}) z_{j} \sum_{k \in \mathbb{N}_{0}^{d}} \boldsymbol{z}^{k} \frac{\partial \xi(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}} \\
& =\sum_{j=1}^{d} \beta_{j}(\boldsymbol{s}) z_{j} \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{k})}{\partial s_{j}} .
\end{aligned}
$$

For the fourth and fifth term, we need to show

$$
\sum_{j=1}^{d} \mathcal{Z}\left(\mu_{j}\left(k_{j}+1\right) \xi\left(t, \boldsymbol{k}+\boldsymbol{e}_{j}, \boldsymbol{s}\right)-\mu_{j} k_{j} \xi(t, \boldsymbol{k}, \boldsymbol{s})\right)(\boldsymbol{z})=-\sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial z_{j}}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

Using elementary computations, we have

$$
\begin{aligned}
& \sum_{j=1}^{d} \mathcal{Z}\left(\mu_{j}\left(k_{j}+1\right) \xi\left(t, \boldsymbol{k}+\boldsymbol{e}_{j}, \boldsymbol{s}\right)-\mu_{j} k_{j} \xi(t, \boldsymbol{k}, \boldsymbol{s})\right)(\boldsymbol{z}) \\
& =\sum_{j=1}^{d} \mu_{j} \sum_{\boldsymbol{k} \in \mathbb{N}_{0}^{d}}\left(k_{j}+1\right) \boldsymbol{z}^{\boldsymbol{k}} \xi\left(t, \boldsymbol{k}+\boldsymbol{e}_{j}, \boldsymbol{s}\right)-k_{j} \boldsymbol{z}^{\boldsymbol{k}} \xi(t, \boldsymbol{k}, \boldsymbol{s}) \\
& =\sum_{j=1}^{d} \mu_{j} \sum_{\boldsymbol{k} \in \mathbb{N}_{j}^{d}} k_{j} \boldsymbol{z}^{\boldsymbol{k}-\boldsymbol{e}_{j}} \xi(t, \boldsymbol{k}, \boldsymbol{s})-k_{j} z_{j} \boldsymbol{z}^{\boldsymbol{k}-\boldsymbol{e}_{j}} \xi(t, \boldsymbol{k}, \boldsymbol{s}) \\
& =\sum_{j=1}^{d} \mu_{j}\left(1-z_{j}\right) \frac{\partial}{\partial z_{j}} \sum_{\boldsymbol{k} \in \mathbb{N}_{0}^{d}} \boldsymbol{z}^{\boldsymbol{k}} \xi(t, \boldsymbol{k}, \boldsymbol{s}) \\
& =-\sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \frac{\partial \zeta(t, \boldsymbol{s}, \boldsymbol{z})}{\partial z_{j}} .
\end{aligned}
$$

Finally, the sixth term follows immediately from the definition since

$$
\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} s_{j} \mathcal{Z}(\xi(t, \boldsymbol{s}, \boldsymbol{k}))(\boldsymbol{z})=\zeta(t, \boldsymbol{s}, \boldsymbol{z}) \sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} s_{j}
$$

## 4.B. 2 Joint moments: computations

In this section, we provide the details behind the derivation of the PDE to ODE as given in Eqns. (4.83) and (4.21). Since we are taking partial derivatives with respect to multiple variables, it is not immediate to see where all terms come from. Some terms are straightforward to compute, so we focus on the ones that require careful attention. We first show the result and then provide details about the complicated terms.

Differentiating Eqn. (4.19) $n_{\lambda_{1}}, \ldots, n_{\lambda_{d}}$ times with respect to $s_{1}, \ldots, s_{d}$, respectively,
and then substituting $s_{1}=s_{2}=\cdots=s_{d}=0$, yields

$$
\begin{align*}
& \frac{\mathrm{d}}{\mathrm{~d} t} \mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)}\right]+\sum_{j=1}^{d} n_{\lambda_{j}} \alpha_{j} \mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)}\right] \\
& \quad-\sum_{l=1}^{d} n_{\lambda_{l}} \sum_{j=1}^{d} \mathbb{E}\left[B_{l j}\right] \mathbb{E}\left[z_{j} \lambda_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}-\mathbf{1}_{\{i=l\}}} z_{i}^{Q_{i}(t)}\right] \\
& \quad+\sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=j\}}}\right] \\
& =\sum_{j=1}^{d}\left(z_{j}-1\right) \mathbb{E}\left[\lambda_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)}\right]+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} n_{\lambda_{j}} \mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}-\mathbf{1}_{\{i=j\}}} z_{i}^{Q_{i}(t)}\right] \\
& \quad+\sum_{j=1}^{d} \sum_{m_{1}=0}^{n_{\lambda_{1}}} \cdots \sum_{m_{d}=0}^{n_{\lambda_{d}}} \mathbf{1}_{\left\{m \leqslant n_{\lambda}-2\right\}} \prod_{k=1}^{d}\binom{n_{\lambda_{k}}}{m_{k}} \mathbb{E}\left[z_{j} \prod_{i=1}^{d} B_{i j}^{n_{\lambda_{i}}-m_{i}} \lambda_{i}(t)^{m_{i}+\mathbf{1}_{\{i=j\}}} z_{i}^{Q_{i}(t)}\right] \tag{4.83}
\end{align*}
$$

where $m=m_{1}+\cdots+m_{d}$ and we collected the $\mathbb{E}\left[B_{i j}\right]$ combinations of first order on the LHS and higher orders on the RHS. Next, we take Eqn. (4.83) and differentiate $n_{Q_{1}}, \ldots, n_{Q_{d}}$ times with respect to $z_{1}, \ldots, z_{d}$, respectively, and then substitute $z_{1}=$ $\cdots=z_{d}=1$. After elementary calculus we obtain

$$
\begin{align*}
& \frac{\mathrm{d}}{\mathrm{~d} t} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right)+\sum_{j=1}^{d}\left(n_{\lambda_{j}}\left(\alpha_{j}-\mathbb{E}\left[B_{j j}\right]\right)+n_{Q_{j}} \mu_{j}\right) \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right) \\
& =\sum_{\substack{ \\
j=1}}^{d} \sum_{\substack{i=1 \\
i \neq j}}^{d} n_{\lambda_{i}} \mathbb{E}\left[B_{i j}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right)+\sum_{j=1}^{d} n_{Q_{j}} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{n}_{\boldsymbol{\lambda}}+\boldsymbol{e}_{\boldsymbol{j}}\right)  \tag{4.84}\\
& \quad+\sum_{j=1}^{d} \alpha_{j} \bar{\lambda}_{j} n_{\lambda_{j}} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{j}\right)+\sum_{i=1}^{d} \sum_{j=1}^{d} n_{\lambda_{i}} n_{Q_{j}} \mathbb{E}\left[B_{i j}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{n}_{\boldsymbol{\lambda}}-\boldsymbol{e}_{i}+\boldsymbol{e}_{j}\right) \\
& \quad+\sum_{j=1}^{d} \sum_{m_{1}=0}^{n_{\lambda_{1}}} \cdots \sum_{m_{d}=0}^{n_{\lambda_{d}}} \mathbf{1}_{\left\{m \leqslant n_{\lambda}-2\right\}} \prod_{k=1}^{d}\binom{n_{\lambda_{k}}}{m_{k}}\left\{n_{Q_{j}} \prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}-\boldsymbol{e}_{j}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right. \\
& \left.\quad+\prod_{i=1}^{d} \mathbb{E}\left[B_{i j}^{n_{\lambda_{i}}-m_{i}}\right] \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{m}+\boldsymbol{e}_{j}\right)\right\} .
\end{align*}
$$

To obtain the ODE in Eqn. (4.83), the starting point is Eqn. (4.19) and we differentiate $n_{\lambda_{1}}, \ldots, n_{\lambda_{d}}$ times with respect to $s_{1}, \ldots, s_{d}$ respectively, and then substitute $s_{1}=\cdots=s_{d}=0$. The terms that are not immediate to compute are those where we

## Chapter 4 Markovian Multivariate Hawkes Population Processes

need to apply the product rule repeatedly. Consider the computation of

$$
\frac{\partial^{n_{\lambda_{1}}} \cdots \partial^{n_{\lambda_{d}}}}{\partial s_{1}^{n_{\lambda_{1}}} \cdots \partial s_{d}^{n_{\lambda_{d}}}} \sum_{j=1}^{d} \alpha_{j} s_{j} \mathbb{E}\left[\lambda_{j}(t) e^{-s^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] .
$$

We first focus on differentiation with respect to the first component, yielding

$$
\begin{aligned}
& \frac{\partial^{n_{\lambda_{1}}}}{\partial s_{1}^{n_{\lambda_{1}}}} \sum_{j=1}^{d} \alpha_{j} s_{j} \mathbb{E}\left[\lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& =\frac{\partial^{n_{\lambda_{1}}-1}}{\partial s_{1}^{n_{\lambda_{1}-1}}} \alpha_{1} \mathbb{E}\left[\lambda_{1}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]-\frac{\partial^{n_{\lambda_{1}}-1}}{\partial s_{1}^{n_{\lambda_{1}}-1}} \sum_{j=1}^{d} \alpha_{j} s_{j} \mathbb{E}\left[\lambda_{1}(t) \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& =-2 \frac{\partial^{n_{\lambda_{1}-2}}}{\partial s_{1}^{n_{\lambda_{1}-2}}} \alpha_{1} \mathbb{E}\left[\lambda_{1}(t)^{2} e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]+\frac{\partial^{n_{\lambda_{1}-2}}}{\partial s_{1}^{n_{\lambda_{1}-2}}} \sum_{j=1}^{d} \alpha_{j} s_{j} \mathbb{E}\left[\lambda_{1}(t)^{2} \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& \vdots \\
& =n_{\lambda_{1}}(-1)^{n_{\lambda_{1}}-1} \alpha_{1} \mathbb{E}\left[\lambda_{1}(t)^{n_{\lambda_{1}}} e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]+\sum_{j=1}^{d} \alpha_{j} s_{j} \mathbb{E}\left[\lambda_{1}(t)^{n_{\lambda_{1}}} \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] .
\end{aligned}
$$

Note that all the terms in the latter sum vanish when we substitute $s_{1}=\cdots=s_{d}=0$. An analogous expression holds for the other components. If we now combine the differentiation with respect to all components and substitute $s_{1}=\cdots=s_{d}=0$, we have

$$
\frac{\partial^{n_{\lambda_{1}}} \cdots \partial^{n_{\lambda_{d}}}}{\partial s_{1}^{n_{\lambda_{1}}} \cdots \partial s_{d}^{n_{\lambda_{d}}}} \sum_{j=1}^{d} \alpha_{j} s_{j} \mathbb{E}\left[\lambda_{j}(t) e^{-s^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]=\sum_{j=1}^{d} n_{\lambda_{j}} \alpha_{j} \mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)}\right]
$$

Another, more complicated, term we need to compute is

$$
\frac{\partial^{n_{\lambda_{1}}} \cdots \partial^{n_{\lambda_{d}}}}{\partial s_{1}^{n_{\lambda_{1}}} \cdots \partial s_{d}^{n_{d}}} \sum_{j=1}^{d} z_{j} \beta_{j}(s) \mathbb{E}\left[\lambda_{j}(t) e^{-s^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]
$$

with $\beta_{j}(\boldsymbol{s})=\mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right]$. It is clear that taking higher order derivatives means that we have to successively apply the product rule. Moreover, since we are taking partial derivatives with respect to multiple components, we will get many cross terms. Let us focus on the first component, which yields

$$
\frac{\partial^{n_{\lambda_{1}}}}{\partial s_{1}^{n_{1}}} \sum_{j=1}^{d} z_{j} \beta_{j}(s) \mathbb{E}\left[\lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right]
$$

$$
\begin{aligned}
= & (-1)^{1} \sum_{j=1}^{d} z_{j} \frac{\partial^{n_{\lambda_{1}}-1}}{\partial s_{1}^{n_{\lambda_{1}-1}}} \mathbb{E}\left[B_{1 j} e^{-s^{\top} \boldsymbol{B}_{j}}\right] \mathbb{E}\left[\lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& +(-1)^{1} \sum_{j=1}^{d} z_{j} \frac{\partial^{n_{\lambda_{1}-1}}}{\partial s_{1}^{n_{\lambda_{1}}-1}} \mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right] \mathbb{E}\left[\lambda_{1}(t) \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
= & (-1)^{2} \sum_{j=1}^{d} z_{j} \frac{\partial^{n_{\lambda_{1}}-2}}{\partial s_{1}^{n_{\lambda_{1}-2}}} \mathbb{E}\left[B_{1 j}^{2} e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right] \mathbb{E}\left[\lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& +2(-1)^{2} \sum_{j=1}^{d} z_{j} \frac{\partial^{n_{\lambda_{1}-2}-2}}{\partial s_{1}^{n_{\lambda_{1}-2}}} \mathbb{E}\left[B_{1_{j}} e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right] \mathbb{E}\left[\lambda_{1}(t) \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& +(-1)^{2} \sum_{j=1}^{d} z_{j} \frac{\partial^{n_{\lambda_{1}}-2}}{\partial s_{1}^{n_{1}-2}} \mathbb{E}\left[e^{-\boldsymbol{s}^{\top} \boldsymbol{B}_{j}}\right] \mathbb{E}\left[\lambda_{1}(t)^{2} \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
= & \\
= & n_{\lambda_{1}}(-1)^{n_{\lambda_{1}}} \sum_{j=1}^{d} z_{j} \mathbb{E}\left[B_{1 j} e^{-s^{\top} \boldsymbol{B}_{j}}\right] \mathbb{E}\left[\lambda_{1}(t)^{n_{\lambda_{1}-1}} \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& +(-1)^{n_{\lambda_{1}}} \sum_{j=1}^{d} z_{j} \mathbf{1}_{\left\{n_{\lambda_{1}} \geqslant 2\right\}} \sum_{m_{1}=0}^{n_{\lambda_{1}-2}}\binom{n_{\lambda_{1}}}{m_{1}} \mathbb{E}\left[B_{1 j}^{n_{\lambda_{1}-m_{1}}}\right] \mathbb{E}\left[\lambda_{1}(t)^{m} \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& +(-1)^{n_{\lambda_{1}}} \sum_{j=1}^{d} z_{j} \mathbb{E}\left[\lambda_{1}(t)^{n_{\lambda_{1}}} \lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right],
\end{aligned}
$$

since the number of terms is doubled in every step of the derivation. The computation for the other components is entirely analogous. Upon taking the joint derivative and substituting $s_{1}=s_{2}=\cdots=s_{d}=0$, we obtain

$$
\begin{aligned}
& \frac{\partial^{n_{\lambda_{1}}} \cdots \partial^{n_{\lambda_{d}}}}{\partial s_{1}^{\lambda_{1}} \cdots \partial s_{d}^{n_{d}}} \sum_{j=1}^{d} z_{j} \beta_{j}(s) \mathbb{E}\left[\lambda_{j}(t) e^{-\boldsymbol{s}^{\top} \boldsymbol{\lambda}(t)} \prod_{n=1}^{d} z_{n}^{Q_{n}(t)}\right] \\
& =\sum_{l=1}^{d} n_{\lambda_{l}} \sum_{j=1}^{d} \mathbb{E}\left[B_{l j}\right] \mathbb{E}\left[z_{j} \lambda_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}-\mathbf{1}_{\{i=l\}}} z_{i}^{Q_{i}(t)}\right]+\sum_{j=1}^{d} z_{j} \mathbb{E}\left[\lambda_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)}\right] \\
& \quad+\sum_{j=1}^{d} \sum_{m_{1}=0}^{n_{\lambda_{1}}} \cdots \sum_{m_{d}=0}^{n_{\lambda_{d}}} \mathbf{1}_{\left\{m \leqslant n_{\lambda}-2\right\}} \prod_{k=1}^{d}\binom{n_{\lambda_{k}}}{m_{k}} \mathbb{E}\left[z_{j} \prod_{i=1}^{d} B_{i j}^{n_{\lambda_{i}}-m_{i}} \lambda_{i}(t)^{m_{i}+\mathbf{1}_{\{i=j\}}} z_{i}^{Q_{i}(t)}\right],
\end{aligned}
$$

where $m=m_{1}+\cdots m_{d}$.
We now focus on the terms to obtain the ODE in Eqn. (4.21). The starting point is Eqn. (4.83), which we differentiate $n_{Q_{1}}, \ldots, n_{Q_{d}}$ times with respect to $z_{1}, \ldots, z_{d}$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

respectively and substitute $z_{1}=\cdots=z_{d}=1$. There are multiple terms in (4.83) that require the product rule when differentiating.

We consider one such term and take the appropriate derivative, i.e.,

$$
\frac{\partial^{n_{Q_{1}}} \cdots \partial^{n_{Q_{d}}}}{\partial z_{1}^{n_{Q_{1}}} \cdots \partial z_{d}^{n_{Q_{d}}}} \sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=j\}}}\right] .
$$

Again, we focus on differentiation with respect to the first component, which yields

$$
\begin{aligned}
& \frac{\partial^{n_{Q_{1}}}}{\partial z_{1}^{n_{Q_{1}}}} \sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=j\}}}\right] \\
& =\frac{\partial^{n_{Q_{1}}-1}}{\partial z_{1}^{n_{Q_{1}-1}}} \mu_{1} \mathbb{E}\left[Q_{1}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=1\}}}\right] \\
& +\frac{\partial^{n_{Q_{1}}-1}}{\partial z_{1}^{n_{Q_{1}}-1}} \mu_{1}\left(z_{1}-1\right) \mathbb{E}\left[Q_{1}(t)\left(Q_{1}(t)-1\right) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-2 \mathbf{1}_{\{i=1\}}}\right] \\
& +\sum_{j=2}^{d} \frac{\partial^{n_{Q_{1}}-1}}{\partial z_{1}^{n_{Q_{1}}-1}} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) Q_{1}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=j\}}-\mathbf{1}_{\{i=1\}}}\right] \\
& =2 \frac{\partial^{n_{Q_{1}}-2}}{\partial z_{1}^{n_{Q_{1}-2}}} \mu_{1} \mathbb{E}\left[Q_{1}(t)\left(Q_{1}(t)-1\right) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-2 \mathbf{1}_{\{i=1\}}}\right] \\
& +\frac{\partial^{n Q_{1}-2}}{\partial z_{1}^{n_{Q_{1}}-2}} \mu_{1}\left(z_{1}-1\right) \mathbb{E}\left[Q_{1}(t)\left(Q_{1}(t)-1\right)\left(Q_{1}(t)-2\right) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-3 \mathbf{1}_{\{i=1\}}}\right] \\
& \sum_{j=2}^{d} \frac{\partial^{n_{Q_{1}}-2}}{\partial z_{1}^{n_{Q_{1}-2}}} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) Q_{1}(t)\left(Q_{1}(t)-1\right) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=j\}}-2 \mathbf{1}_{\{i=1\}}}\right] \\
& = \\
& =n_{Q_{1}} \mu_{1} \mathbb{E}\left[Q_{1}(t)^{\left[n_{Q_{1}}\right]} \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-n_{Q_{1}} \mathbf{1}_{\{i=1\}}}\right],
\end{aligned}
$$

where we substituted $z_{1}=\cdots=z_{d}=1$ in the last step, canceling out all the terms that contain the factor $\left(z_{j}-1\right)$. We can compute the derivatives with respect to other components in a similar manner, which results in

$$
\frac{\partial^{n_{Q_{1}}} \cdots \partial^{n Q_{d}}}{\partial z_{1}^{n_{Q_{1}}} \cdots \partial z_{d}^{n_{Q_{d}}}} \sum_{j=1}^{d} \mu_{j}\left(z_{j}-1\right) \mathbb{E}\left[Q_{j}(t) \prod_{i=1}^{d} \lambda_{i}(t)^{n_{\lambda_{i}}} z_{i}^{Q_{i}(t)-\mathbf{1}_{\{i=j\}}}\right]
$$

$$
=\sum_{j=1}^{d} n_{Q_{j}} \mathbb{E}\left[\prod_{i=1}^{d} \lambda_{i}^{n_{\lambda_{i}}} Q_{i}(t)^{\left[n_{Q_{i}}\right]}\right]=\sum_{j=1}^{d} n_{Q_{j}} \psi_{t}\left(\boldsymbol{n}_{\boldsymbol{Q}}, \boldsymbol{n}_{\boldsymbol{\lambda}}\right) .
$$

Note that all other terms in Eqn. (4.83) follow the same structure of differentiation.

## 4.C Explicit Examples

In this section, we provide more explicit details for the moments in the bivariate setting $d=2$. We provide examples by writing out the recursive procedure outlined in Section 4.4. The main objective is to derive near-explicit results for both the transient moments $\psi_{t}\left(\left(n_{Q_{1}}, n_{Q_{2}}\right),\left(n_{\lambda_{1}}, n_{\lambda_{2}}\right)\right)$ and stationary moments $\psi\left(\left(n_{Q_{1}}, n_{Q_{2}}\right),\left(n_{\lambda_{1}}, n_{\lambda_{2}}\right)\right)$, where the focus is on moments of order 1 and 2 . In both cases, we apply the recursive procedures described in Section 4.4.

## 4.C. 1 Recursive procedure

We illustrate the stacked vector $\mathbf{\Psi}_{t}^{(n)}$ for orders $n=1$ and $n=2$, and derive the ODEs associated with the recursive procedure.

Example 4.1 (first order, bivariate). For $n=1$, we have $\mathfrak{D}(2,1)=4$, and

$$
\boldsymbol{\Psi}_{t}^{(1)}=\left(\mathbf{\Psi}_{t}^{(0,1)}, \boldsymbol{\Psi}_{t}^{(1,0)}\right)^{\top},
$$

where

$$
\boldsymbol{\Psi}_{t}^{(0,1)}=\left(\mathbb{E}\left[\lambda_{1}(t)\right], \mathbb{E}\left[\lambda_{2}(t)\right]\right)^{\top}, \quad \mathbf{\Psi}_{t}^{(1,0)}=\left(\mathbb{E}\left[Q_{1}(t)\right], \mathbb{E}\left[Q_{2}(t)\right]\right)^{\top} .
$$

By Step 0 of Algorithm 4.1, we obtain the ODE

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(0,1)}=\left[\begin{array}{cc}
-\bar{\alpha}_{1} & \mathbb{E}\left[B_{12}\right]  \tag{4.85}\\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}
\end{array}\right] \boldsymbol{\Psi}_{t}^{(0,1)}+\left[\begin{array}{c}
\alpha_{1} \bar{\lambda}_{1} \\
\alpha_{2} \bar{\lambda}_{2}
\end{array}\right]
$$

whose solution gives us an expression for $\mathbf{\Psi}_{t}^{(0,1)}=\left(\mathbb{E}\left[\lambda_{1}(t)\right], \mathbb{E}\left[\lambda_{2}(t)\right]\right)^{\top}$. We need this $\mathbf{\Psi}_{t}^{(0,1)}$ in Step 1, which states

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(1,0)}=\left[\begin{array}{cc}
-\mu_{1} & 0  \tag{4.86}\\
0 & -\mu_{2}
\end{array}\right] \boldsymbol{\Psi}_{t}^{(1,0)}+\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right] \boldsymbol{\Psi}_{t}^{(0,1)}
$$

whose solution yields an expression for $\mathbf{\Psi}_{t}^{(1,0)}=\left(\mathbb{E}\left[Q_{1}(t)\right], \mathbb{E}\left[Q_{2}(t)\right]\right)^{\top}$.

Example 4.2 (second order, bivariate). For $n=2$, we have $\mathfrak{D}(2,2)=10$, and

$$
\boldsymbol{\Psi}_{t}^{(2)}=\left(\mathbf{\Psi}_{t}^{(0,2)}, \mathbf{\Psi}_{t}^{(1,1)}, \mathbf{\Psi}_{t}^{(2,0)}\right)^{\top},
$$

where

$$
\begin{aligned}
& \mathbf{\Psi}_{t}^{(0,2)}=\left(\mathbb{E}\left[\lambda_{1}(t)^{2}\right], \mathbb{E}\left[\lambda_{1}(t) \lambda_{2}(t)\right], \mathbb{E}\left[\lambda_{2}(t)^{2}\right]\right)^{\top}, \\
& \mathbf{\Psi}_{t}^{(1,1)}=\left(\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right], \mathbb{E}\left[Q_{1}(t) \lambda_{2}(t)\right], \mathbb{E}\left[Q_{2}(t) \lambda_{1}(t)\right], \mathbb{E}\left[Q_{2}(t) \lambda_{2}(t)\right]\right)^{\top}, \\
& \mathbf{\Psi}_{t}^{(2,0)}=\left(\mathbb{E}\left[Q_{1}(t)^{[2]}\right], \mathbb{E}\left[Q_{1}(t) Q_{2}(t)\right], \mathbb{E}\left[Q_{2}(t)^{[2]}\right]\right)^{\top} .
\end{aligned}
$$

For order $n=2$, our objective is to compute $\boldsymbol{\Psi}_{t}^{(0,2)}, \boldsymbol{\Psi}_{t}^{(1,1)}$, and $\boldsymbol{\Psi}_{t}^{(2,0)}$. Step 0 of Algorithm 4.1 yields

$$
\begin{align*}
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(0,2)}= & {\left[\begin{array}{ccc}
-2 \bar{\alpha}_{1} & 2 \mathbb{E}\left[B_{12}\right] & 0 \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{1}-\bar{\alpha}_{2} & \mathbb{E}\left[B_{12}\right] \\
0 & 2 \mathbb{E}\left[B_{21}\right] & -2 \bar{\alpha}_{2}
\end{array}\right] \boldsymbol{\Psi}_{t}^{(0,2)} } \\
& +\left[\begin{array}{cc}
2 \alpha_{1} \bar{\lambda}_{1}+\mathbb{E}\left[B_{11}^{2}\right] & \mathbb{E}\left[B_{12}^{2}\right] \\
\mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}\right]+\alpha_{2} \bar{\lambda}_{2} & \mathbb{E}\left[B_{22}\right] \mathbb{E}\left[B_{12}\right]+\alpha_{1} \bar{\lambda}_{1} \\
\mathbb{E}\left[B_{21}^{2}\right] & 2 \alpha_{2} \bar{\lambda}_{2}+\mathbb{E}\left[B_{22}^{2}\right]
\end{array}\right] \boldsymbol{\Psi}_{t}^{(0,1)}, \tag{4.87}
\end{align*}
$$

which depends on the lower-order vector $\boldsymbol{\Psi}_{t}^{(0,1)}$ (which was found in Example 1). For Step 1, note that $\mathbf{\Psi}_{t}^{(1,1)}$ is a 4-dimensional vector, which satisfies

$$
\begin{align*}
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(1,1)}= & {\left[\begin{array}{cc}
-\bar{\alpha}_{1}-\mu_{1} & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}-\mu_{1}
\end{array}\right] \oplus\left[\begin{array}{ccc}
-\bar{\alpha}_{1}-\mu_{2} & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}-\mu_{2}
\end{array}\right] \boldsymbol{\Psi}_{t}^{(1,1)} }  \tag{4.88}\\
& +\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{array}\right] \boldsymbol{\Psi}_{t}^{(0,2)}+\left[\begin{array}{cccc}
\alpha_{1} \bar{\lambda}_{1} & 0 & \mathbb{E}\left[B_{11}\right] & 0 \\
\alpha_{2} \bar{\lambda}_{2} & 0 & \mathbb{E}\left[B_{21}\right] & 0 \\
0 & \alpha_{1} \bar{\lambda}_{1} & 0 & \mathbb{E}\left[B_{12}\right] \\
0 & \alpha_{2} \bar{\lambda}_{2} & 0 & \mathbb{E}\left[B_{22}\right]
\end{array}\right] \boldsymbol{\Psi}_{t}^{(1)},
\end{align*}
$$

where we see the dependence on the lower-order stacked vector $\boldsymbol{\Psi}_{t}^{(1)}$ (which was found in Example 1). Regarding the final step, i.e., Step 2,

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t}^{(2,0)}=\left[\begin{array}{ccc}
-2 \mu_{1} & 0 & 0  \tag{4.89}\\
0 & -\mu_{1}-\mu_{2} & 0 \\
0 & 0 & -2 \mu_{2}
\end{array}\right] \boldsymbol{\Psi}_{t}^{(2,0)}+\left[\begin{array}{cccc}
2 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 2
\end{array}\right] \boldsymbol{\Psi}_{t}^{(1,1)} .
$$

Example 4.3 (third order, bivariate). For $n=3$, we have $\mathfrak{D}(2,3)=20$, and

$$
\boldsymbol{\Psi}_{t}^{(3)}=\left(\boldsymbol{\Psi}_{t}^{(0,3)}, \boldsymbol{\Psi}_{t}^{(1,2)}, \boldsymbol{\Psi}_{t}^{(2,1)}, \boldsymbol{\Psi}_{t}^{(3,0)}\right)^{\top}
$$

where

$$
\begin{aligned}
\Psi_{t}^{(0,3)}= & \left(\mathbb{E}\left[\lambda_{1}(t)^{3}\right], \mathbb{E}\left[\lambda_{1}(t)^{2} \lambda_{2}(t)\right], \mathbb{E}\left[\lambda_{1}(t) \lambda_{2}(t)^{2}\right], \mathbb{E}\left[\lambda_{2}(t)^{3}\right]\right)^{\top}, \\
\Psi_{t}^{(1,2)}= & \left(\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)^{2}\right], \mathbb{E}\left[Q_{1}(t) \lambda_{1} \lambda_{2}(t)\right], \mathbb{E}\left[Q_{1}(t) \lambda_{2}(t)^{2}\right],\right. \\
& \left.\mathbb{E}\left[Q_{2}(t) \lambda_{1}(t)^{2}\right], \mathbb{E}\left[Q_{2}(t) \lambda_{1}(t) \lambda_{2}\right], \mathbb{E}\left[Q_{2}(t) \lambda_{2}(t)^{2}\right]\right)^{\top}, \\
\Psi_{t}^{(2,1)}= & \left(\mathbb{E}\left[Q_{1}(t)^{[2]} \lambda_{1}(t)\right], \mathbb{E}\left[Q_{1}(t)^{[2]} \lambda_{2}(t)\right], \mathbb{E}\left[Q_{1}(t) Q_{2}(t) \lambda_{1}(t)\right], \mathbb{E}\left[Q_{1}(t) Q_{2}(t) \lambda_{2}(t)\right],\right. \\
& \left.\mathbb{E}\left[Q_{2}(t)^{[2]} \lambda_{1}(t)\right], \mathbb{E}\left[Q_{2}(t)^{[2]} \lambda_{2}(t)\right]\right)^{\top}, \\
\boldsymbol{\Psi}_{t}^{(3,0)}= & \left(\mathbb{E}\left[Q_{1}(t)^{[3]}\right], \mathbb{E}\left[Q_{1}(t)^{[2]} Q_{2}(t)\right], \mathbb{E}\left[Q_{1}(t) Q_{2}(t)^{[2]}\right], \mathbb{E}\left[Q_{2}(t)^{[3]}\right]\right)^{\top} .
\end{aligned}
$$

We could explicitly write down the ODEs of these vectors using Algorithm 4.1, but the exposition would be rather tedious with large matrices.

## 4.C. 2 Transient moments

The goal of this subsection is to find near-explicit expressions for $\boldsymbol{\Psi}_{t}^{(1)}$ and $\boldsymbol{\Psi}_{t}^{(2)}$ by further solving the associated ODEs. It is clear that we can obtain the solution in terms of a matrix exponential, which can be made more explicit in terms of its eigenvalues, namely

$$
\begin{equation*}
e^{t \boldsymbol{M}^{(k, n-k)}}=\sum_{\ell=1}^{\bar{k}} e^{t \eta_{\ell}^{(k)}} \prod_{\substack{m=1 \\ m \neq \ell}}^{\bar{k}} \frac{\boldsymbol{M}^{(k, n-k)}-\eta_{m}^{(k)} \boldsymbol{I}}{\eta_{\ell}^{(k)}-\eta_{m}^{(k)}} \tag{4.90}
\end{equation*}
$$

with $\bar{k}$ denoting the dimension of $\boldsymbol{\Psi}_{t}^{(k, n-k)}$ and $\eta_{1}^{(k)}, \ldots, \eta_{\bar{k}}^{(k)}$ the eigenvalues of $\boldsymbol{M}^{(k, n-k)}$, and $\boldsymbol{I}$ the identity matrix.

We first consider the transient moments of order $n=1$, working out the entries of the stacked vector $\mathbf{\Psi}_{t}^{(1)}$, in particular solving the ODE of $\boldsymbol{\Psi}_{t}^{(0,1)}$ as given in Eqn. (4.85). By Proposition 4.1, the solution requires us to find the eigenvalues of the matrix

$$
\boldsymbol{M}^{(0,1)}=\left[\begin{array}{cc}
-\bar{\alpha}_{1} & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}
\end{array}\right],
$$

so as to compute the matrix exponential $e^{t \boldsymbol{M}^{(0,1)}}$. With $\eta \equiv \eta_{1}, \eta_{2}$ denoting the two eigenvalues, we straightforwardly obtain

$$
\begin{equation*}
\eta=\frac{1}{2}\left(-\bar{\alpha}_{1}-\bar{\alpha}_{2} \pm \sqrt{\bar{\alpha}_{1}^{2}-2 \bar{\alpha}_{1} \bar{\alpha}_{2}+\bar{\alpha}_{2}^{2}+4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}\right) \equiv \frac{1}{2}\left(-\bar{\alpha}_{1}-\bar{\alpha}_{2} \pm \sqrt{D_{1}}\right) \tag{4.91}
\end{equation*}
$$

where $\bar{\alpha}_{i}=\alpha_{i}-\mathbb{E}\left[B_{i i}\right]$ for $i=1,2$. We let $\eta_{1}$ and $\eta_{2}$ denote the plus- and minus-variant of $\eta$ respectively. Note that $D_{1} \geqslant 0$ since it involves a square and $B_{i j}$ are non-negative

## Chapter 4 Markovian Multivariate Hawkes Population Processes

random variables. Using Eqn. (4.90) and doing some elementary computations, the matrix exponential can be expressed as

$$
\begin{aligned}
e^{t \boldsymbol{M}^{(0,1)}} & =\frac{1}{\eta_{1}-\eta_{2}}\left(e^{t \eta_{1}}\left(\boldsymbol{M}^{(0,1)}-\eta_{2} \boldsymbol{I}\right)-e^{t \eta_{2}}\left(\boldsymbol{M}^{(0,1)}-\eta_{1} \boldsymbol{I}\right)\right) \\
& =\frac{1}{\sqrt{D_{1}}}\left(\left\{e^{t \eta_{1}}-e^{t \eta_{2}}\right\}\left[\begin{array}{cc}
\frac{1}{2}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right) & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & \frac{1}{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)
\end{array}\right]+\left\{e^{t \eta_{1}}+e^{t \eta_{2}}\right\}\left[\begin{array}{cc}
\frac{1}{2} \sqrt{D_{1}} & 0 \\
0 & \frac{1}{2} \sqrt{D_{1}}
\end{array}\right]\right) \\
& =\frac{1}{2}\left\{e^{t \eta_{1}}+e^{t \eta_{2}}\right\} \boldsymbol{I}+\frac{1}{\sqrt{D_{1}}\left\{e^{t \eta_{1}}-e^{t \eta_{2}}\right\}\left[\begin{array}{cc}
\frac{1}{2}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right) & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & \frac{1}{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)
\end{array}\right] ;}
\end{aligned}
$$

we use curly brackets to distinguish scalar terms from the vectors and matrices.
Next, we consider $\boldsymbol{\Psi}_{t}^{(1,0)}$ for which we need to find the eigenvalues of the matrix $\boldsymbol{M}^{(1,0)}=\operatorname{diag}\left(-\mu_{1},-\mu_{2}\right)$; cf. the ODE in Eqn. (4.86). Since this is a diagonal matrix, these are simply $-\mu_{1}$ and $-\mu_{2}$, such that the matrix exponential is just

$$
e^{t \boldsymbol{M}^{(1,0)}}=\left[\begin{array}{cc}
e^{-t \mu_{1}} & 0 \\
0 & e^{-t \mu_{2}}
\end{array}\right] .
$$

Before we get to the solution, we define a number of functions needed for the solution of $\boldsymbol{\Psi}_{t}^{(1,0)}$, namely

$$
\begin{aligned}
& \boldsymbol{u}_{1}(t):=\left[\begin{array}{l}
\left(\mu_{1}+\eta_{1}\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{1}}\right)+\left(\mu_{1}+\eta_{2}\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{1}}\right) \\
\left(\mu_{2}+\eta_{1}\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{2}}\right)+\left(\mu_{2}+\eta_{2}\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{2}}\right)
\end{array}\right], \\
& \boldsymbol{u}_{2}(t):=\left[\begin{array}{l}
\left(\mu_{1}+\eta_{1}\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{1}}\right)-\left(\mu_{1}+\eta_{2}\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{1}}\right) \\
\left(\mu_{2}+\eta_{1}\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{2}}\right)-\left(\mu_{2}+\eta_{2}\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{2}}\right)
\end{array}\right], \\
& \boldsymbol{u}_{3}(t):=\left[\begin{array}{l}
\left(\eta_{1}\left(\mu_{1}+\eta_{1}\right)\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{1}}\right)+\left(\eta_{2}\left(\mu_{1}+\eta_{2}\right)\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{1}}\right) \\
\left(\eta_{1}\left(\mu_{2}+\eta_{1}\right)\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{2}}\right)+\left(\eta_{2}\left(\mu_{2}+\eta_{2}\right)\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{2}}\right)
\end{array}\right], \\
& \boldsymbol{u}_{4}(t):=\left[\begin{array}{l}
\left(\eta_{1}\left(\mu_{1}+\eta_{1}\right)\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{1}}\right)-\left(\eta_{2}\left(\mu_{1}+\eta_{2}\right)\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{1}}\right) \\
\left(\eta_{1}\left(\mu_{2}+\eta_{1}\right)\right)^{-1}\left(e^{t \eta_{1}}-e^{-t \mu_{2}}\right)-\left(\eta_{2}\left(\mu_{2}+\eta_{2}\right)\right)^{-1}\left(e^{t \eta_{2}}-e^{-t \mu_{2}}\right)
\end{array}\right] .
\end{aligned}
$$

We can now give the first moments explicitly through the application of Proposition 4.1. Given the initial conditions $\boldsymbol{\Psi}_{0}^{(0,1)}=\left(\bar{\lambda}_{1}, \bar{\lambda}_{2}\right)^{\top}$ and $\boldsymbol{\Psi}_{0}^{(1,0)}=(0,0)^{\top}$, the solutions
to the ODEs in (4.85) and (4.86) are given by

$$
\begin{align*}
\boldsymbol{\Psi}_{t}^{(0,1)}= & e^{t \boldsymbol{M}^{(0,1)}}\left[\begin{array}{l}
\bar{\lambda}_{1} \\
\bar{\lambda}_{2}
\end{array}\right]+\int_{0}^{t} e^{(t-s) \boldsymbol{M}^{(0,1)}}\left[\begin{array}{c}
\alpha_{1} \bar{\lambda}_{1} \\
\alpha_{2} \bar{\lambda}_{2}
\end{array}\right] \mathrm{d} s \\
= & \frac{1}{\bar{\alpha}_{1} \bar{\alpha}_{2}-\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}\left[\begin{array}{l}
\alpha_{1} \bar{\lambda}_{1} \bar{\alpha}_{2}+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right] \\
\alpha_{2} \bar{\lambda}_{2} \bar{\alpha}_{1}+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right] \\
& +\frac{1}{2}\left\{e^{t \eta_{1}}+e^{t \eta_{2}}\right\}\left[\begin{array}{l}
\bar{\lambda}_{1} \\
\bar{\lambda}_{2}
\end{array}\right]+\frac{1}{\sqrt{D_{1}}\left\{e^{t \eta_{1}}-e^{t \eta_{2}}\right\}\left[\begin{array}{l}
\frac{1}{2} \bar{\lambda}_{1}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right)+\bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right] \\
\frac{1}{2} \bar{\lambda}_{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)+\bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right]} \\
& +\frac{1}{2}\left\{\eta_{1}^{-1} e^{t \eta_{1}}+\eta_{2}^{-1} e^{t \eta_{2}}\right\}\left[\begin{array}{l}
\alpha_{1} \bar{\lambda}_{1} \\
\alpha_{2} \bar{\lambda}_{2}
\end{array}\right] \\
& +\frac{1}{\sqrt{D_{1}}}\left\{\eta_{1}^{-1} e^{t \eta_{1}}-\eta_{2}^{-1} e^{t \eta_{2}}\right\}\left[\begin{array}{l}
\frac{1}{2} \alpha_{1} \bar{\lambda}_{1}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right)+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right] \\
\frac{1}{2} \alpha_{2} \bar{\lambda}_{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right] \tag{4.92}
\end{align*}
$$

and

$$
\begin{align*}
\boldsymbol{\Psi}_{t}^{(1,0)}= & \int_{0}^{t} e^{(t-s) \boldsymbol{M}^{(1,0)}} \boldsymbol{\Psi}_{s}^{(0,1)} \mathrm{d} s \\
= & \frac{1}{\bar{\alpha}_{1} \bar{\alpha}_{2}-\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}\left[\begin{array}{l}
\mu_{1}^{-1}\left(1-e^{-t \mu_{1}}\right) \\
\mu_{1}^{-2}\left(1-e^{-t \mu_{2}}\right)
\end{array}\right] \odot\left[\begin{array}{l}
\alpha_{1} \bar{\lambda}_{1} \bar{\alpha}_{2}+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right] \\
\alpha_{2} \bar{\lambda}_{2} \bar{\alpha}_{1}+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right] \\
& +\frac{1}{2} \boldsymbol{u}_{1}(t) \odot\left[\begin{array}{l}
\bar{\lambda}_{1} \\
\bar{\lambda}_{2}
\end{array}\right]+\frac{1}{\sqrt{D_{1}}} \boldsymbol{u}_{2}(t) \odot\left[\begin{array}{l}
\frac{1}{2} \bar{\lambda}_{1}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right)+\bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right] \\
\frac{1}{2} \bar{\lambda}_{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)+\bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right]  \tag{4.93}\\
& +\frac{1}{2} \boldsymbol{u}_{3}(t) \odot\left[\begin{array}{l}
\alpha_{1} \bar{\lambda}_{1} \\
\alpha_{2} \bar{\lambda}_{2}
\end{array}\right]+\frac{1}{\sqrt{D_{1}}} \boldsymbol{u}_{4}(t) \odot\left[\begin{array}{l}
\frac{1}{2} \alpha_{1} \bar{\lambda}_{1}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right)+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right] \\
\frac{1}{2} \alpha_{2} \bar{\lambda}_{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right] .
\end{align*}
$$

Observe that in order for the solution in (4.92) to remain stable and to obtain finite moments, we need that both eigenvalues are strictly smaller than 0 . By some elementary algebra, it is seen that we should have that

$$
\begin{equation*}
\bar{\alpha}_{1} \bar{\alpha}_{2}>\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right] . \tag{4.94}
\end{equation*}
$$

Note that this is the explicit version of the stability condition $\rho(\boldsymbol{H})<1$ for the bivariate setting; see Assumption 4.1. Also note that if one is interested in the Hawkes process $\boldsymbol{N}(t)=\left(N_{1}(t), N_{2}(t)\right)^{\top}$ rather than the population process $\boldsymbol{Q}(t)=$ $\left(Q_{1}(t), Q_{2}(t)\right)^{\top}$, one needs to take $\mu_{1}=\mu_{2} \equiv 0$. The corresponding moments $\mathbb{E}[\boldsymbol{N}(t)]=$ $\left(\mathbb{E}\left[N_{1}(t)\right], \mathbb{E}\left[N_{2}(t)\right]\right)^{\top}$ can be derived from Eqn. (4.93) by taking the limit $\left(\mu_{1}, \mu_{2}\right) \downarrow$ $(0,0)$ and the use of L'Hopital's rule.

## Chapter 4 Markovian Multivariate Hawkes Population Processes

We now turn to order 2 and compute elements of the stacked vector $\boldsymbol{\Psi}_{t}^{(2)}$. As before, we start by considering $\boldsymbol{\Psi}_{t}^{(0,2)}$, the vector containing the (mixed) moments corresponding to $\boldsymbol{\lambda}(t)$. By Proposition 4.1, and recalling that $\boldsymbol{\Psi}_{t}^{(0,2)}$ satisfies the ODE in Eqn. (4.87), we need to find the eigenvalues of

$$
\boldsymbol{M}^{(0,2)}=\left[\begin{array}{ccc}
-2 \bar{\alpha}_{1} & 2 \mathbb{E}\left[B_{12}\right] & 0 \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{1}-\bar{\alpha}_{2} & \mathbb{E}\left[B_{12}\right] \\
0 & 2 \mathbb{E}\left[B_{21}\right] & -2 \bar{\alpha}_{2}
\end{array}\right] .
$$

Let $\kappa \equiv \kappa_{1}, \kappa_{2}, \kappa_{3}$ denote the eigenvalues of $\boldsymbol{M}^{(0,2)}$. We compute

$$
\begin{aligned}
& \quad\left|\begin{array}{ccc}
-2 \bar{\alpha}_{1}-\kappa & 2 \mathbb{E}\left[B_{12}\right] & 0 \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{1}-\bar{\alpha}_{2}-\kappa & \mathbb{E}\left[B_{12}\right] \\
0 & 2 \mathbb{E}\left[B_{21}\right] & -2 \bar{\alpha}_{2}-\kappa
\end{array}\right|=0 \\
& \Longleftrightarrow\left(-2 \bar{\alpha}_{1}-\kappa\right)\left|\begin{array}{cc}
-\bar{\alpha}_{1}-\bar{\alpha}_{2}-\kappa & \mathbb{E}\left[B_{12}\right] \\
2 \mathbb{E}\left[B_{21}\right] & -2 \bar{\alpha}_{2}-\kappa
\end{array}\right|-2 \mathbb{E}\left[B_{12}\right]\left|\begin{array}{cc}
\mathbb{E}\left[B_{21}\right] & \mathbb{E}\left[B_{12}\right] \\
0 & -2 \bar{\alpha}_{2}-\kappa
\end{array}\right|=0 \\
& \Longleftrightarrow \\
& \\
& \quad\left(-2 \bar{\alpha}_{1}-\kappa\right)\left(\kappa^{2}+3 \bar{\alpha}_{2} \kappa+2 \bar{\alpha}_{2}^{2}+2 \bar{\alpha}_{1} \bar{\alpha}_{2}+\bar{\alpha}_{1} \kappa-2 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]\right) \\
& \quad+2 \mathbb{E}\left[B_{12}\right]\left(\mathbb{E}\left[B_{21}\right] \kappa+2 \bar{\alpha}_{2} \mathbb{E}\left[B_{21}\right]=0\right. \\
& \Longleftrightarrow \kappa^{3}+\kappa^{2} 3\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)+\kappa\left(2\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}-4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]\right)-4\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right) \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]=0 \\
& \Longleftrightarrow \kappa^{3}+\kappa^{2} b+\kappa c+d=0,
\end{aligned}
$$

with $b, c, d$ defined as the constants of the square, linear and constant term respectively. To apply the formula for the solutions to this cubic equation, we compute $p$ and $q$, given by

$$
\begin{aligned}
p & =\frac{1}{3}\left(3 c-b^{2}\right) \\
& =2\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}-4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]-3\left(\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}\right. \\
& =-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}-4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
q= & \frac{1}{27}\left\{2 b^{3}-9 b c+27 d\right\} \\
= & \frac{1}{27}\left\{2\left(3\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)\right)^{3}-27\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)\left(2\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}-4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]\right)\right. \\
& \left.\quad-4 \cdot 27\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right) \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]\right\} \\
= & 2\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{3}-2\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{3}+4\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right) \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]-4\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right) \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right] \\
= & 0 .
\end{aligned}
$$

It is well-known that the cubic equation has three real roots if $4 p^{3}+27 q^{2}<0$. Since $q=0$, the condition becomes $4 p^{3}<0$, which holds since $p<0$ because of the square term and $\mathbb{E}\left[B_{i j}\right] \geqslant 0$. Hence, the eigenvalues $\kappa_{m}$, with $m=1,2,3$, are given by the trigonometric solution

$$
\begin{aligned}
& \kappa_{k}=-\frac{b}{3}+2 \sqrt{\frac{-p}{3}} \cos \left(\theta_{m}\right)=-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)+2 \sqrt{\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}+4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]} \cos \left(\theta_{m}\right), \\
& \theta_{m}=\frac{1}{3} \arccos \left(\frac{3 q}{2 p} \sqrt{\frac{3}{-p}}\right)-\frac{2 \pi}{3}(m-1)=\frac{\pi}{6}-\frac{2 \pi}{3}(m-1)
\end{aligned}
$$

This yields the eigenvalues

$$
\begin{aligned}
& \kappa_{1}=-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right) \\
& \kappa_{2}=-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)+\sqrt{3} \sqrt{\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}+4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}=-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)+\sqrt{D_{2}} \\
& \kappa_{3}=-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)-\sqrt{3} \sqrt{\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}+4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}=-\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)-\sqrt{D_{2}},
\end{aligned}
$$

with $D_{2}=3\left(\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}+4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]\right)$. We apply these eigenvalues in the computation of the matrix exponential, as described in Eqn. (4.90), to obtain

$$
\begin{aligned}
& e^{t \boldsymbol{M}^{(0,2)}} \\
& =e^{\kappa_{1} t} \frac{1}{\kappa_{1}-\kappa_{2}} \frac{1}{\kappa_{1}-\kappa_{3}}\left(\boldsymbol{M}^{(0,2)}-\kappa_{2} \boldsymbol{I}\right)\left(\boldsymbol{M}^{(0,2)}-\kappa_{3} \boldsymbol{I}\right) \\
& +e^{\kappa_{2} t} \frac{1}{\kappa_{2}-\kappa_{1}} \frac{1}{\kappa_{2}-\kappa_{3}}\left(\boldsymbol{M}^{(0,2)}-\kappa_{1} \boldsymbol{I}\right)\left(\boldsymbol{M}^{(0,2)}-\kappa_{3} \boldsymbol{I}\right) \\
& +e^{\kappa_{3} t} \frac{1}{\kappa_{3}-\kappa_{1}} \frac{1}{\kappa_{3}-\kappa_{2}}\left(\boldsymbol{M}^{(0,2)}-\kappa_{1} \boldsymbol{I}\right)\left(\boldsymbol{M}^{(0,2)}-\kappa_{2} \boldsymbol{I}\right) \\
& =\frac{e^{\kappa_{1} t}}{D_{2}}\left[\begin{array}{ccc}
c_{\kappa_{1}} & -2 \mathbb{E}\left[B_{12}\right]\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right) & -2 \mathbb{E}\left[B_{12}\right]^{2} \\
-\mathbb{E}\left[B_{21}\right]\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right) & 3\left(\bar{\alpha}_{1}+\bar{\alpha}_{2}\right)^{2}+8 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right] & -\mathbb{E}\left[B_{12}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right) \\
-2 \mathbb{E}\left[B_{21}\right]^{2} & -2 \mathbb{E}\left[B_{21}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right) & c_{\kappa_{1}}
\end{array}\right] \\
& +\frac{e^{\kappa_{2} t}}{2 D_{2}}\left[\begin{array}{ccc}
c_{-} & -2 \mathbb{E}\left[B_{12}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{2}}\right) & 2 \mathbb{E}\left[B_{12}\right]^{2} \\
-\mathbb{E}\left[B_{21}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{2}}\right) & 4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right] & \mathbb{E}\left[B_{12}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}+\sqrt{D_{2}}\right) \\
2 \mathbb{E}\left[B_{21}\right]^{2} & 2 \mathbb{E}\left[B_{21}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}+\sqrt{D_{2}}\right) & c_{+}
\end{array}\right] \\
& +\frac{e^{\kappa_{3} t}}{2 D_{2}}\left[\begin{array}{ccc}
c_{+} & -2 \mathbb{E}\left[B_{12}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}+\sqrt{D_{2}}\right) & 2 \mathbb{E}\left[B_{12}\right]^{2} \\
-\mathbb{E}\left[B_{21}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{2}}\right) & 4 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right] & \mathbb{E}\left[B_{12}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{2}}\right) \\
2 \mathbb{E}\left[B_{21}\right]^{2} & 2 \mathbb{E}\left[B_{21}\right]\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{2}}\right) & c_{-}
\end{array}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
c_{\kappa_{1}} & =2 \bar{\alpha}_{1}^{2}+8 \bar{\alpha}_{1} \bar{\alpha}_{2}+2 \bar{\alpha}_{2}^{2}+10 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right], \\
c_{-} & =2 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]+\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{2}}\right),
\end{aligned}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

$$
c_{+}=2 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]+\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}+\sqrt{D_{2}}\right) .
$$

We now derive the matrix exponential corresponding to $\boldsymbol{\Psi}_{t}^{(1,1)}$, appearing in the ODE given in Eqn. (4.88). Observe that (4.88) reveals that to compute $\boldsymbol{\Psi}_{t}^{(1,1)}$, we need to know the non-homogeneous part of the equation, i.e. $\boldsymbol{\Psi}_{t}^{(0,2)}$, as well as the lower order $(n=1)$ stacked vector $\boldsymbol{\Psi}_{t}^{(1)}$. Further notice that the $4 \times 4$ matrix $\boldsymbol{M}^{(1,1)}$ is the direct sum of two $2 \times 2$ matrices, which implies that we can split the 4 -dimensional ODE into two 2-dimensional ODEs. We introduce the relevant objects by setting

$$
\boldsymbol{\Psi}_{t}^{(1,1)}=\left(\mathbf{\Psi}_{t, Q_{1}}^{(1,1)}, \boldsymbol{\Psi}_{t, Q_{2}}^{(1,1)}\right)^{\top},
$$

where

$$
\begin{aligned}
\boldsymbol{\Psi}_{t, Q_{1}}^{(1,1)} & =\left(\mathbb{E}\left[Q_{1}(t) \lambda_{1}(t)\right], \mathbb{E}\left[Q_{1}(t) \lambda_{2}(t)\right]\right)^{\top}, \\
\mathbf{\Psi}_{t, Q_{2}}^{(1,1)} & =\left(\mathbb{E}\left[Q_{2}(t) \lambda_{1}(t)\right], \mathbb{E}\left[Q_{2}(t) \lambda_{2}(t)\right]\right)^{\top} .
\end{aligned}
$$

Focus on the solution of $\boldsymbol{\Psi}_{t, Q_{1}}^{(1,1)}$, where we note that the solution for $\boldsymbol{\Psi}_{t, Q_{2}}^{(1,1)}$ can be obtained in an analogous manner. Observe that one can derive from Eqn. (4.88) that $\Psi_{t, Q_{1}}^{(1,1)}$ satisfies the ODE

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \boldsymbol{\Psi}_{t, Q_{1}}^{(1,1)}=\left[\begin{array}{cc}
-\bar{\alpha}_{1}-\mu_{1} & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}-\mu_{1}
\end{array}\right] \boldsymbol{\Psi}_{t, Q_{1}}^{(1,1)}+\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0
\end{array}\right] \boldsymbol{\Psi}_{t}^{(0,2)}+\left[\begin{array}{ccc}
\alpha_{1} \bar{\lambda}_{1} & 0 & \mathbb{E}\left[B_{11}\right] \\
\alpha_{2} \bar{\lambda}_{2} & 0 & \mathbb{E}\left[B_{21}\right]
\end{array} 0\right] \boldsymbol{\Psi}_{t}^{(1)} .
$$

This means that we need the matrix exponential of

$$
\boldsymbol{M}_{Q_{1}}^{(1,1)}=\left[\begin{array}{cc}
-\bar{\alpha}_{1}-\mu_{1} & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}-\mu_{1}
\end{array}\right],
$$

which requires us to find the corresponding two eigenvalues, denoted by $\gamma_{1}^{\left(Q_{1}\right)}$ and $\gamma_{2}^{\left(Q_{1}\right)}$. These eigenvalues are similarly derived as in Eqn. (4.91), in this case given by

$$
\gamma_{1}^{\left(Q_{1}\right)}=-\mu_{1}+\eta_{1}, \quad \gamma_{2}^{\left(Q_{1}\right)}=-\mu_{1}+\eta_{2},
$$

with $\eta_{1}=\frac{1}{2}\left(-\bar{\alpha}_{1}-\bar{\alpha}_{2}+\sqrt{D_{1}}\right)$ and $\eta_{2}=\frac{1}{2}\left(-\bar{\alpha}_{1}-\bar{\alpha}_{2}-\sqrt{D_{1}}\right)$. Substituting this in the Lagrange interpolation formula of Eqn. (4.90), we obtain

$$
\begin{aligned}
e^{t \boldsymbol{M}_{Q_{1}}^{(1,1)}} & =\frac{1}{\gamma_{1}^{\left(Q_{1}\right)}-\gamma_{2}^{\left(Q_{1}\right)}}\left\{e^{t \gamma_{1}^{\left(Q_{1}\right)}}\left(\boldsymbol{M}_{Q_{1}}^{(1,1)}-\gamma_{2}^{\left(Q_{1}\right)} \boldsymbol{I}_{2}\right)-e^{t \gamma_{2}^{\left(Q_{1}\right)}}\left(\boldsymbol{M}_{Q_{1}}^{(1,1)}-\gamma_{1}^{\left(Q_{1}\right)} \boldsymbol{I}_{2}\right)\right\} \\
& =\frac{1}{2}\left\{e^{t \gamma_{1}^{\left(Q_{1}\right)}}+e^{t \gamma_{2}^{\left(Q_{1}\right)}}\right\} \boldsymbol{I}+\frac{1}{\sqrt{D_{1}}}\left\{e^{t \gamma_{1}^{\left(Q_{1}\right)}}-e^{t \gamma_{2}^{\left(Q_{1}\right)}}\right\}\left[\begin{array}{cc}
\frac{1}{2}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right) & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & \frac{1}{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)
\end{array}\right] .
\end{aligned}
$$

In a very similar manner, the matrix exponential needed to evaluate $\boldsymbol{\Psi}_{t, Q_{2}}^{(1,1)}$ can be obtained from the ODE in Eqn. (4.88), and it is given by

$$
\begin{aligned}
e^{t \boldsymbol{M}_{Q_{2}}^{(1,1)}} & =\frac{1}{\gamma_{1}^{\left(Q_{2}\right)}-\gamma_{2}^{\left(Q_{2}\right)}}\left\{e^{t \gamma_{1}^{\left(Q_{2}\right)}}\left(\boldsymbol{M}_{Q_{2}}^{(1,1)}-\gamma_{2}^{\left(Q_{2}\right)} \boldsymbol{I}_{2}\right)-e^{t \gamma_{2}^{\left(Q_{2}\right)}}\left(\boldsymbol{M}_{Q_{2}}^{(1,1)}-\gamma_{1}^{\left(Q_{2}\right)} \boldsymbol{I}_{2}\right)\right\} \\
& =\frac{1}{2}\left\{e^{t \gamma_{1}^{\left(Q_{2}\right)}}+e^{t \gamma_{2}^{\left(Q_{2}\right)}}\right\} \boldsymbol{I}+\frac{1}{\sqrt{D_{1}}}\left\{e^{t \gamma_{1}^{\left(Q_{2}\right)}}-e^{t \gamma_{2}^{\left(Q_{2}\right)}}\right\}\left[\begin{array}{cc}
\frac{1}{2}\left(\bar{\alpha}_{2}-\bar{\alpha}_{1}\right) & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & \frac{1}{2}\left(\bar{\alpha}_{1}-\bar{\alpha}_{2}\right)
\end{array}\right],
\end{aligned}
$$

with $\gamma_{1}^{\left(Q_{2}\right)}=-\mu_{2}+\eta_{1}$ and $\gamma_{2}^{\left(Q_{2}\right)}=-\mu_{2}+\eta_{2}$.
Finally, for the solution of $\Psi_{t}^{(2,0)}$, the vector containing the mixed factorial moments of $\boldsymbol{Q}(t)$, we derive the matrix exponential of $\boldsymbol{M}^{(2,0)}=\operatorname{diag}\left(-2 \mu_{1},-\mu_{1}-\mu_{2},-2 \mu_{2}\right)$, which is simply

$$
e^{t \boldsymbol{M}^{(2,0)}}=\left[\begin{array}{ccc}
e^{-2 t \mu_{1}} & 0 & 0 \\
0 & e^{-t\left(\mu_{1}+\mu_{2}\right)} & 0 \\
0 & 0 & e^{-2 t \mu_{2}}
\end{array}\right]
$$

Applying Proposition 4.1 to $\boldsymbol{\Psi}_{t}^{(0,2)}, \boldsymbol{\Psi}_{t}^{(1,1)}$ and $\boldsymbol{\Psi}_{t}^{(2,0)}$, we obtain the following result. Given the initial conditions $\boldsymbol{\Psi}_{0}^{(0,2)}=\left(\bar{\lambda}_{1}^{2}, \bar{\lambda}_{1} \bar{\lambda}_{2}, \bar{\lambda}_{2}^{2}\right)^{\top}, \boldsymbol{\Psi}_{0, Q_{1}}^{(1,1)}=(0,0)^{\top}, \boldsymbol{\Psi}_{0, Q_{2}}^{(1,1)}=(0,0)^{\top}$ and $\boldsymbol{\Psi}_{0}^{(2,0)}=(0,0,0)^{\top}$, the solutions to the ODEs in Eqns. (4.87), (4.88) and (4.89), are given by, respectively,

$$
\begin{align*}
\boldsymbol{\Psi}_{t}^{(0,2)}= & e^{t \boldsymbol{M}^{(0,2)}} \boldsymbol{\Psi}_{0}^{(0,2)} \\
& +\int_{0}^{t} e^{(t-s) \boldsymbol{M}^{(0,2)}}\left[\begin{array}{cc}
2 \alpha_{1} \bar{\lambda}_{1}+\mathbb{E}\left[B_{11}^{2}\right] & \mathbb{E}\left[B_{12}^{2}\right] \\
\mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}\right]+\alpha_{2} \bar{\lambda}_{2} & \mathbb{E}\left[B_{22}\right] \mathbb{E}\left[B_{12}\right]+\alpha_{1} \bar{\lambda}_{1} \\
\mathbb{E}\left[B_{21}^{2}\right] & 2 \alpha_{2} \bar{\lambda}_{2}+\mathbb{E}\left[B_{22}^{2}\right]
\end{array}\right] \boldsymbol{\Psi}_{s}^{(0,1)} \mathrm{d} s, \tag{4.95}
\end{align*}
$$

and $\boldsymbol{\Psi}_{t}^{(1,1)}=\left(\Psi_{t, Q_{1}}^{(1,1)}, \Psi_{t, Q_{2}}^{(1,1)}\right)^{\top}$, with

$$
\begin{align*}
& \boldsymbol{\Psi}_{t, Q_{1}}^{(1,1)}=\int_{0}^{t} e^{(t-s) \boldsymbol{M}_{Q_{1}}^{(1,1)}}\left\{\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0
\end{array}\right] \boldsymbol{\Psi}_{s}^{(0,2)}+\left[\begin{array}{lll}
\alpha_{1} \bar{\lambda}_{1} & 0 & \mathbb{E}\left[B_{11}\right] \\
\alpha_{2} \bar{\lambda}_{2} & 0 & \mathbb{E}\left[B_{21}\right]
\end{array}\right] \boldsymbol{\Psi}_{s}^{(1)}\right\} \mathrm{d} s  \tag{4.96}\\
& \boldsymbol{\Psi}_{t, Q_{2}}^{(1,1)}=\int_{0}^{t} e^{(t-s) \boldsymbol{M}_{Q_{2}}^{(1,1)}}\left\{\left[\begin{array}{lll}
0 & 1 & 0 \\
0 & 0 & 1
\end{array}\right] \boldsymbol{\Psi}_{s}^{(0,2)}+\left[\begin{array}{llll}
0 & \alpha_{1} \bar{\lambda}_{1} & 0 & \mathbb{E}\left[B_{12}\right] \\
0 & \alpha_{2} \bar{\lambda}_{2} & 0 & \mathbb{E}\left[B_{22}\right]
\end{array}\right] \boldsymbol{\Psi}_{s}^{(1)}\right\} \mathrm{d} s
\end{align*}
$$

## Chapter 4 Markovian Multivariate Hawkes Population Processes

and

$$
\begin{align*}
\boldsymbol{\Psi}_{t}^{(2,0)} & =\int_{0}^{t} e^{(t-s) \boldsymbol{M}^{(2,0)}}\left[\begin{array}{llll}
2 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 2
\end{array}\right] \boldsymbol{\Psi}_{s}^{(1,1)} \mathrm{d} s  \tag{4.97}\\
& =\int_{0}^{t}\left[\begin{array}{c}
2 e^{-2(t-s) \mu_{1}} \mathbb{E}\left[Q_{1}(s) \lambda_{1}(s)\right] \\
e^{-(t-s)\left(\mu_{1}+\mu_{2}\right)}\left(\mathbb{E}\left[Q_{1}(s) \lambda_{2}(s)\right]+\mathbb{E}\left[Q_{2}(s) \lambda_{1}(s)\right]\right) \\
2 e^{-2(t-s) \mu_{2}} \mathbb{E}\left[Q_{2}(s) \lambda_{2}(s)\right]
\end{array}\right] \mathrm{d} s .
\end{align*}
$$

We remark that explicit evaluation of the expressions for the second-order moments is tedious, but in principle possible. For instance, it is evident that substituting the matrix exponential of $\boldsymbol{M}^{(0,2)}$ and the lower-order solution $\boldsymbol{\Psi}_{s}^{(0,1)}$ will yield rather involved expressions. Evidently, for higher order moments, these expressions become even more complex. However, the near-explicit solution in terms of matrix exponentials and lower order terms is useful for practical purposes. Due to the availability of fast and robust algorithms for the matrix exponential and the solution of ODEs, it is relatively straightforward to numerically compute higher-order moments.

## 4.C. 3 Stationary moments

This subsection deals with the application of Algorithm 4.2 to evaluate the stationary moments $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Psi}^{(2)}$.

We start by computing the stationary moments of order 1. By Step 1 of Algorithm 4.2, $\boldsymbol{\Psi}^{(0,1)}$ satisfies the linear equation

$$
0=\left[\begin{array}{cc}
-\bar{\alpha}_{1} & \mathbb{E}\left[B_{12}\right] \\
\mathbb{E}\left[B_{21}\right] & -\bar{\alpha}_{2}
\end{array}\right] \boldsymbol{\Psi}^{(0,1)}+\left[\begin{array}{c}
\alpha_{1} \bar{\lambda}_{1} \\
\alpha_{2} \bar{\lambda}_{2}
\end{array}\right],
$$

yielding

$$
\Psi^{(0,1)}=\frac{1}{\bar{\alpha}_{1} \bar{\alpha}_{2}-\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}\left[\begin{array}{l}
\alpha_{1} \bar{\lambda}_{1} \bar{\alpha}_{2}+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right]  \tag{4.98}\\
\alpha_{2} \bar{\lambda}_{2} \bar{\alpha}_{1}+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]
\end{array}\right] .
$$

Then, regarding Step 2, we find after some calculus

$$
\boldsymbol{\Psi}^{(1,0)}=\frac{1}{\bar{\alpha}_{1} \bar{\alpha}_{2}-\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{21}\right]}\left[\begin{array}{l}
\mu_{1}^{-1}\left(\alpha_{1} \bar{\lambda}_{1} \bar{\alpha}_{2}+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[B_{12}\right]\right)  \tag{4.99}\\
\mu_{2}^{-1}\left(\alpha_{2} \bar{\lambda}_{2} \bar{\alpha}_{1}+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[B_{21}\right]\right)
\end{array}\right] .
$$

We have thus found the stacked vector $\boldsymbol{\Psi}^{(1)}$. These expressions could also have been derived by sending $t \rightarrow \infty$ in the expressions of the transient moments $\boldsymbol{\Psi}_{t}^{(0,1)}$ and $\boldsymbol{\Psi}_{t}^{(1,0)}$, respectively.

For the second order stationary moments, we again go over the steps of Algorithm 4.2. Step 0 yields

$$
\begin{align*}
\boldsymbol{\Psi}^{(0,2)}= & {\left[\begin{array}{ccc}
2 \bar{\alpha}_{1} & -2 \mathbb{E}\left[B_{12}\right] & 0 \\
-\mathbb{E}\left[B_{21}\right] & \bar{\alpha}_{1}+\bar{\alpha}_{2} & -\mathbb{E}\left[B_{12}\right] \\
0 & -2 \mathbb{E}\left[B_{21}\right] & 2 \bar{\alpha}_{2}
\end{array}\right]^{-1} }  \tag{4.100}\\
& \times\left[\begin{array}{c}
\mathbb{E}\left[\lambda_{1}\right]\left(2 \alpha_{1} \bar{\lambda}_{1}+\mathbb{E}\left[B_{11}^{2}\right]\right)+\mathbb{E}\left[\lambda_{2}\right] \mathbb{E}\left[B_{12}^{2}\right] \\
\mathbb{E}\left[\lambda_{1}\right]\left(\mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}\right]+\alpha_{2} \bar{\lambda}_{2}\right)+\mathbb{E}\left[\lambda_{2}\right]\left(\mathbb{E}\left[B_{22}\right] \mathbb{E}\left[B_{12}\right]+\alpha_{1} \bar{\lambda}_{1}\right) \\
\mathbb{E}\left[\lambda_{1}\right] \mathbb{E}\left[B_{21}^{2}\right]+\mathbb{E}\left[\lambda_{2}\right]\left(2 \alpha_{2} \bar{\lambda}_{2}+\mathbb{E}\left[B_{22}^{2}\right]\right)
\end{array}\right]
\end{align*}
$$

where the inverse may be explicitly computed in specific cases. For Step 1, we have, after some elementary matrix computations, that

$$
\boldsymbol{\Psi}^{(1,1)}=\left[\begin{array}{cccc}
\bar{\alpha}_{1}+\mu_{1} & -\mathbb{E}\left[B_{12}\right] & 0 & 0  \tag{4.101}\\
-\mathbb{E}\left[B_{21}\right] & \bar{\alpha}_{2}+\mu_{1} & 0 & 0 \\
0 & 0 & \bar{\alpha}_{1}+\mu_{2} & -\mathbb{E}\left[B_{12}\right] \\
0 & 0 & -\mathbb{E}\left[B_{21}\right] & \bar{\alpha}_{2}+\mu_{2}
\end{array}\right]^{-1}\left[\begin{array}{c}
\mathbb{E}\left[\lambda_{1}^{2}\right]+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[Q_{1}\right]+\mathbb{E}\left[B_{11}\right] \mathbb{E}\left[\lambda_{1}\right] \\
\mathbb{E}\left[\lambda_{1} \lambda_{2}\right]+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[Q_{1}\right]+\mathbb{E}\left[B_{21}\right] \mathbb{E}\left[\lambda_{1}\right] \\
\mathbb{E}\left[\lambda_{1} \lambda_{2}\right]+\alpha_{1} \bar{\lambda}_{1} \mathbb{E}\left[Q_{2}\right]+\mathbb{E}\left[B_{12}\right] \mathbb{E}\left[\lambda_{2}\right] \\
\mathbb{E}\left[\lambda_{2}^{2}\right]+\alpha_{2} \bar{\lambda}_{2} \mathbb{E}\left[Q_{2}\right]+\mathbb{E}\left[B_{22}\right] \mathbb{E}\left[\lambda_{2}\right]
\end{array}\right] .
$$

Finally, for Step 2 we have

$$
\boldsymbol{\Psi}^{(2,0)}=\left[\begin{array}{ccc}
1 /\left(2 \mu_{1}\right) & 0 & 0  \tag{4.102}\\
0 & 1 /\left(\mu_{1}+\mu_{2}\right) & 0 \\
0 & 0 & 1 /\left(2 \mu_{2}\right)
\end{array}\right]\left[\begin{array}{llll}
2 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 2
\end{array}\right] \boldsymbol{\Psi}^{(1,1)} .
$$

In line with earlier observations, the quasi-explicit results for the transient and stationary moments become involved for larger values of the order $n$. We have checked all above results using SymPy, a symbolic programming package.

## 4.C. 4 Higher order moments

In this section, we provide some more detailed explicit matrices discussed in Section 4.5, for the bivariate $d=2$ setting for moments of order $n=3$. For $n=3$ we have $\mathfrak{D}(3,2)=20$, and we consider the stacked vector (size 34)

$$
\left(\mathbf{\Psi}_{t}^{(3)}, \mathbf{\Psi}_{t}^{(2)}, \mathbf{\Psi}_{t}^{(1)}\right)^{\top},
$$

which satisfies the ODE

$$
\frac{\mathrm{d}}{\mathrm{~d} t}\left[\begin{array}{l}
\boldsymbol{\Psi}_{t}^{(3)}  \tag{4.103}\\
\boldsymbol{\Psi}_{t}^{(2)} \\
\boldsymbol{\Psi}_{t}^{(1)}
\end{array}\right]=\boldsymbol{A}_{3}^{35 \times 35}\left[\begin{array}{l}
\mathbf{\Psi}_{t}^{(3)} \\
\mathbf{\Psi}_{t}^{(2)} \\
\boldsymbol{\Psi}_{t}^{(1)}
\end{array}\right]+\left[\begin{array}{c}
\boldsymbol{b}^{2 \times 1} \\
\mathbf{0}^{32 \times 1}
\end{array}\right] .
$$

The matrix $\boldsymbol{F}_{3}^{34 \times 34}$ is given by

$$
\boldsymbol{F}_{3}^{34 \times 34}=\left[\begin{array}{cc}
\boldsymbol{F}_{2}^{14 \times 14} & \mathbf{0}^{14 \times 20}  \tag{4.104}\\
\boldsymbol{G}_{3}^{20 \times 14} & \boldsymbol{H}_{3}^{14 \times 14}
\end{array}\right]
$$

where

$$
\boldsymbol{H}_{3}^{14 \times 14}=\left[\begin{array}{cccc}
\boldsymbol{M}^{(3,0)} & \boldsymbol{0}^{4 \times 6} & \mathbf{0}^{4 \times 6} & \boldsymbol{0}^{4 \times 4} \\
\boldsymbol{K}^{(2,1)} & \boldsymbol{M}^{(2,1)} & \mathbf{0}^{6 \times 6} & \mathbf{0}^{6 \times 4} \\
\mathbf{0}^{6 \times 4} & \boldsymbol{K}^{(1,2)} & \boldsymbol{M}^{(1,2)} & \mathbf{0}^{6 \times 4} \\
\mathbf{0}^{4 \times 4} & \mathbf{0}^{4 \times 6} & \boldsymbol{K}^{(0,3)} & \boldsymbol{M}^{(0,3)}
\end{array}\right], \quad \boldsymbol{G}_{3}^{20 \times 14}=\left[\begin{array}{c}
\boldsymbol{L}^{(0,3)} \\
\boldsymbol{L}^{(1,2)} \\
\boldsymbol{L}^{(2,1)} \\
\mathbf{0}^{4 \times 20}
\end{array}\right] .
$$

The elements of $\boldsymbol{G}_{3}^{20 \times 20}$ require some more notation to describe, by introducing a number of sub-matrices. First, we have $\boldsymbol{L}^{(0,3)}=\left[\begin{array}{llll}\boldsymbol{L}_{\lambda^{1}}^{(0,3)} & \mathbf{0}^{4 \times 2} & \boldsymbol{L}_{\lambda^{2}}^{(0,3)} & \mathbf{0}^{4 \times 7}\end{array}\right]$, with
$\boldsymbol{L}_{\lambda^{1}}^{(0,3)}=\left[\begin{array}{cc}\mathbb{E}\left[B_{11}^{3}\right] & \mathbb{E}\left[B_{12}^{3}\right] \\ \mathbb{E}\left[B_{11}^{2}\right] \mathbb{E}\left[B_{21}\right] & \mathbb{E}\left[B_{12}^{2}\right] \mathbb{E}\left[B_{22}\right] \\ \mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}^{2}\right] & \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{22}^{2}\right] \\ \mathbb{E}\left[B_{21}^{3}\right] & \mathbb{E}\left[B_{22}^{3}\right]\end{array}\right]$,
$\boldsymbol{L}_{\lambda^{2}}^{(0,3)}=\left[\begin{array}{ccc}3 \mathbb{E}\left[B_{11}^{2}\right]+3 \alpha_{1} \bar{\lambda}_{1} & 3 \mathbb{E}\left[B_{12}^{2}\right] & 0 \\ 2 \mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}\right]+\alpha_{2} \bar{\lambda}_{2} & 2 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{22}\right]+\mathbb{E}\left[B_{11}^{2}\right]+2 \alpha_{1} \bar{\lambda}_{1} & \mathbb{E}\left[B_{12}^{2}\right] \\ \mathbb{E}\left[B_{21}^{2}\right] & 2 \mathbb{E}\left[B_{21}\right] \mathbb{E}\left[B_{11}\right]+\mathbb{E}\left[B_{22}^{2}\right]+2 \alpha_{2} \bar{\lambda}_{2} & 2 \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{22}\right]+\alpha_{1} \bar{\lambda}_{1} \\ 0 & 3 \mathbb{E}\left[B_{21}^{2}\right] & 3 \mathbb{E}\left[B_{22}^{2}\right]+3 \alpha_{2} \bar{\lambda}_{2}\end{array}\right]$.
Second, we have $\boldsymbol{L}^{(1,2)}=\left[\begin{array}{lllll}\boldsymbol{L}_{\lambda^{1}}^{(1,2)} & \mathbf{0}^{6 \times 2} & \boldsymbol{L}_{\lambda^{2}}^{(1,2)} & \boldsymbol{L}_{Q^{1} \lambda^{1}}^{(1,2)} & \mathbf{0}^{6 \times 3}\end{array}\right]$, where
$\boldsymbol{L}_{\lambda^{1}}^{(1,2)}=\left[\begin{array}{cc}\mathbb{E}\left[B_{11}^{2}\right] & 0 \\ \mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}\right] & 0 \\ \mathbb{E}\left[B_{21}^{2}\right] & 0 \\ 0 & \mathbb{E}\left[B_{12}^{2}\right] \\ 0 & \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{22}\right] \\ 0 & \mathbb{E}\left[B_{22}^{2}\right]\end{array}\right], \quad \boldsymbol{L}_{\lambda^{2}}^{(1,2)}=\left[\begin{array}{ccc}2 \mathbb{E}\left[B_{11}\right] & 0 & 0 \\ \mathbb{E}\left[B_{21}\right] & \mathbb{E}\left[B_{11}\right] & 0 \\ 0 & 2 \mathbb{E}\left[B_{21}\right. & 0 \\ 0 & 2 \mathbb{E}\left[B_{12}\right] & 0 \\ 0 & \mathbb{E}\left[B_{22}\right] & \mathbb{E}\left[B_{12}\right] \\ 0 & 0 & 2 \mathbb{E}\left[B_{22}\right]\end{array}\right]$,
$\boldsymbol{L}_{Q^{1} L^{1}}^{(1,2)}$

$$
=\left[\begin{array}{cccc}
\mathbb{E}\left[B_{11}^{2}\right]+2 \alpha_{1} \bar{\lambda}_{1} & \mathbb{E}\left[B_{12}^{2}\right] & 0 & 0 \\
\mathbb{E}\left[B_{11} 1\right] \mathbb{E}\left[B_{21}\right]+\alpha_{2} \bar{\lambda}_{2} & \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{22}\right]+\alpha_{1} \bar{\lambda}_{1} & 0 & 0 \\
\mathbb{E}\left[B_{21}^{2}\right] & \mathbb{E}\left[B_{22}^{2}\right]+2 \alpha_{2} \bar{\lambda}_{2} & 0 & 0 \\
0 & 0 & \mathbb{E}\left[B_{11}^{2}\right]+2 \alpha_{1} \bar{\lambda}_{1} & \mathbb{E}\left[B_{12}^{2}\right] \\
0 & 0 & \mathbb{E}\left[B_{11}\right] \mathbb{E}\left[B_{21}\right]+\alpha_{2} \bar{\lambda}_{2} & \mathbb{E}\left[B_{12}\right] \mathbb{E}\left[B_{22}\right]+\alpha_{1} \bar{\lambda}_{1} \\
0 & 0 & \mathbb{E}\left[B_{21}^{2}\right] & \mathbb{E}\left[B_{22}^{2}\right]+2 \alpha_{2} \bar{\lambda}_{2}
\end{array}\right] .
$$

Finally, $\boldsymbol{L}^{(2,1)}=\left[\begin{array}{lll}\mathbf{0}^{6 \times 7} & \boldsymbol{L}_{Q^{1} L^{1}}^{(2,1)} & \boldsymbol{L}_{Q^{2}}^{(2,1)}\end{array}\right]$, where

$$
\boldsymbol{L}_{Q^{1} L^{1}}^{(2,1)}=\left[\begin{array}{cccc}
2 \mathbb{E}\left[B_{11}\right] & 0 & 0 & 0 \\
2 \mathbb{E}\left[B_{21}\right] & 0 & 0 & 0 \\
0 & \mathbb{E}\left[B_{12}\right] & \mathbb{E}\left[B_{11}\right] & 0 \\
0 & \mathbb{E}\left[B_{22}\right] & \mathbb{E}\left[B_{21}\right] & 0 \\
0 & 0 & 0 & 2 \mathbb{E}\left[B_{12}\right] \\
0 & 0 & 0 & 2 \mathbb{E}\left[B_{22}\right]
\end{array}\right], \quad \boldsymbol{L}_{Q^{2}}^{(2,1)}=\left[\begin{array}{ccc}
\alpha_{1} \bar{\lambda}_{1} & 0 & 0 \\
\alpha_{2} \bar{\lambda}_{2} & 0 & 0 \\
0 & \alpha_{1} \bar{\lambda}_{1} & 0 \\
0 & \alpha_{2} \bar{\lambda}_{2} & 0 \\
0 & 0 & \alpha_{1} \bar{\lambda}_{1} \\
0 & 0 & \alpha_{2} \bar{\lambda}_{2}
\end{array}\right] .
$$

## Bibliography

[1] Abate, J. \& Whitt, W. (1992). Solving probability transform functional equations for numerical inversion. Operations Research Letters 12, 275-281.
[2] Aїt-Sahalia, Y., Cacho-Diaz, J. A. \& Laeven, R. J. A. (2015). Modeling financial contagion using mutually exciting jump processes. Journal of Financial Economics 117, 585-606.
[3] Aїт-Sahalia, Y., Laeven, R. J. A. \& Pelizzon, L. (2014). Mutual excitation in Eurozone sovereign CDS. Journal of Econometrics 183, 151-167.
[4] Andersen, P. K., Borgan, Ø., Gill, R. D. \& Keiding, N. (1993). Statistical Models Based on Counting Processes. Springer Series in Statistics. SpringerVerlag, New York.
[5] Asmussen, S \& Glynn, P. (2007). Stochastic Simulation: Algorithms and Analysis. Springer, New York.
[6] Bacry, E., Delattre, S., Hoffmann, M. \& Muzy, J. F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. Stochastic Processes and their Applications 123, 2475-2499.
[7] Bacry, E., Mastromatteo, I. \& Muzy, J. (2015). Hawkes processes in finance. Market Microstructure and Liquidity 01.
[8] Baldwin, A., Gheyas, I., Ioannidis, Ch., Pym, D. \& Williams, J. (2017). Contagion in cyber security attacks. Journal of the Operational Research Society 68, 780-791.
[9] Bessy-Roland Y., Boumezoued, A. \& Hillairet, C. (2021). Multivariate Hawkes process for cyber risk insurance. Annals of Actuarial Science 15, 14-39.
[10] Bingham, N. H., Goldie, C. M. \& Teugels, L. (1987). Regular Variation. Cambridge University Press, Cambridge.
[11] Bordenave, C., Torrisi, G. L. (2007). Large deviations of Poisson cluster processes. Stochastic Models 23, 593-625.
[12] Brémaud, P. \& Massoulié, L. (1996). Stability of nonlinear Hawkes processes. Annals of Probability 24, 1562-1588.
[13] Brillinger, D. R. (1988). Some statistical methods for random process data from seismology and neurophysiology. Annals of Statistics 16, 1-54.
[14] Champin, P.-A., Gandon, F. \& Médini, L. (2018). SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. Preprint. Available at https://arxiv.org/abs/1711.01679.
[15] Chaumont, L. \& Liu, R. (2014). Coding multitype forests: Application to the law of the total population of branching forests. Transactions of the American Mathematical Society 368, 2723-2747.
[16] Chen, Z., Dassios, A., Kuan, V., Lim, J. W., Qu, Y., Surya, B. \& Zhao, H. (2020). A two-phase dynamic contagion model for COVID19. Preprint. Available at https://www.lse.ac.uk/Statistics/Assets/ Documents/Paper-Covid-19-DCP.pdf.
[17] Crane, R. \& Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. Proceedings of the National Academy of Sciences 105, 15649-15653.
[18] Cui, L., Hawkes, A. \& Yi, H. (2020). An elementary derivation of moments of Hawkes processes. Advances in Applied Probability 52, 102-137.
[19] Dembo, A. \& Zeitouni, O. (1998). Large Deviations Techniques and Applications. Second Edition. Springer, New York.
[20] Dieker, T. \& Mandjes, M. (2006). Fast simulation of overflow probabilities in a queue with Gaussian input. ACM Transactions on Modeling and Computer Simulation 16, 119-151.
[21] Da Fonseca, J. \& Zaatour, R. (2014). Hawkes process: Fast calibration, application to trade clustering and diffusive limit. Journal of Futures Markets 34, 548-579.
[22] Da Fonseca, J. \& Zaatour, R. (2015). Clustering and mean reversion in a Hawkes microstructure model. Journal of Futures Markets 35, 813-838.
[23] Daley, D. J. \& Vere-Jones, D. (2003). An Introduction to the Theory of Point Processes. Volume 1: Elementary Theory and Methods. Volume 2: General Theory and Structure. Second Edition. Springer Science and Business Media, New York.
[24] Dassios, A. \& Zhou, H. (2011). A dynamic contagion process. Advances in Applied Probability 43, 814-846.
[25] Daw, A., Castellanos, A., Yom-Tov, G., Pender, J. \& GruEndlinger, L. (2020). The co-production of service: Modeling service times in contact centers using Hawkes processes. Preprint. Available at https://arxiv. org/pdf/2004.07861.pdf.
[26] Daw, A. \& Pender, J. (2018). Queues driven by Hawkes processes. Stochastic Systems 8, 192-229.
[27] Daw, A. \& Pender, J. (2019). Matrix calculations for moments of Markov processes. Preprint. Available at https://arxiv.org/abs/1909.03320.
[28] Daw, A. \& Pender, J. (2022a). An ephemerally self-exciting point process. Advances in Applied Probability 54, 340-403.
[29] Daw, A. \& Pender, J. (2022b). Matrix calculations for moments of Markov processes. Advances in Applied Probability, forthcoming.
[30] Duffie, D., Pan, J. \& Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343-1376.
[31] El Euch, O. \& Rosenbaum, M. (2019). The characteristic function of rough Heston models. Mathematical Finance 29, 3-38.
[32] Embrechts, P., Klüppelberg, C. \& Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer, New York.
[33] Errais, E., Giesecke, K. \& Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. Journal of Financial Mathematics 1, 642-665.
[34] Feller, W. (1970). An Introduction to Probability Theory and Its Applications: Volume II. Second Edition. Wiley Series in Probability and Statistics. Wiley, New York.
[35] Feng, M., Cai, S.-M., Tang, M., \& Lai, Y.-C. (2019). Equivalence and its invalidation between non-Markovian and Markovian spreading dynamics on complex networks. Nature Communications 10, 3748.
[36] Foss, S., Konstantopoulos, T., \& Zachary, S. (2007). Discrete and continuous time modulated random walks with heavy-tailed increments. Journal of Theoretical Probability 20, 581-612.
[37] Ganesh, A., O’Connell, N. \& Wischik, D. (2004). Big Queues. Springer, Berlin.
[38] Gao, X., Zhou, X. \& Zhu, L. (2018). Transform analysis for Hawkes processes with applications in dark pool trading. Quantitative Finance 18, 265-282.
[39] Gao, X. \& Zhu, L. (2018). Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. Queueing Systems: Theory and Applications 90, 161-206.
[40] Gao, X. \& Zhu, L. (2018). Large deviations and applications for Markovian Hawkes processes with a large initial intensity. Bernoulli 24, 2875-2905.
[41] Gao, X. \& Zhu, L. (2018). Limit theorems for Markovian Hawkes processes with a large initial intensity. Stochastic Processes and their Applications 128, 3807-3839.
[42] Gao, X. \& Zhu, L. (2019). Affine point processes: Refinements to large-time asymptotics. Preprint. Available at https://arxiv.org/pdf/1903.06371.pdf.
[43] Gao, X. \& Zhu, L. (2021). Precise deviations for Hawkes processes. Bernoulli 27, 221-248.
[44] Guo, Q., Remillard, B. \& Swishchuk, A. (2020) Multivariate general compound point processes in limit order books. Risks 8, 98.
[45] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. Biometrika 58, 83-90.
[46] Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. Journal of the Royal Statistical Society. Series B (Methodological), 438-443.
[47] Hawkes, A. G. \& Oakes, D. (1974). A cluster representation of a self-exciting process. Journal of Applied Probability 11, 493-503.
[48] Hawkes, A. G. (2018). Hawkes processes and their applications to finance: A review. Quantitative Finance 18, 193-198.
[49] Ikefuji, M., R. J. A. Laeven, J. R. Magnus \& Y. Yue (2022). Earthquake risk embedded in property prices: Evidence from five Japanese cities. Journal of the American Statistical Association 117, 82-93.
[50] Jacod, J. \& Mémin, J (1976). Caractéristiques locales et conditions de continuité absolue pour les semi-martingales. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 35, 1-37.
[51] Jagers, P. (1975). Branching Processes with Biological Applications. First edition. Wiley, London.
[52] Jaisson, T. \& Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. Annals of Applied Probability 25, 600-631.
[53] Jaisson, T. \& Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. Annals of Applied Probability 26, 2860-2882.
[54] Karabash, D. \& Zhu, L. (2015). Limit theorems for marked Hawkes processes with application to a risk model. Stochastic Models 31, 433-451.
[55] Karim, R. S., Laeven, R. J. A. \& Mandjes, M. (2021). Exact and asymptotic analysis of general multivariate Hawkes processes and induced population processes. Preprint. Available at https://arxiv.org/pdf/2106.03560.pdf.
[56] Kirchner, M. (2017). Perspectives on Hawkes Processes. PhD Thesis, ETH Zürich.
[57] Koops, D. T., Saxena, M., Boxma, O. J. \& Mandjes, M. (2018). Infiniteserver queues with Hawkes input. Journal of Applied Probability 55, 920-943.
[58] Laub, P. J., Taimre, T. \& Pollett, P. K. (2015). Hawkes Processes. Preprint. Available at https://arxiv.org/abs/1507.02822.
[59] Lin, Z.-H., Feng, M., Tang, M., Liu, Z., Xu, C., Hui, P. M. \& Lai, Y.-C. (2020). Non-Markovian recovery makes complex networks more resilient against large-scale failures. Nature Communications 11, 2490.
[60] Liniger, T. J. (2009). Multivariate Hawkes Processes. PhD Thesis, ETH Zürich.

## Bibliography

[61] Mandjes, M. \& Blanchet, J. (2009). Rare-event simulation for queues. In: G. Rubino and B. Tuffin (eds.): Rare Event Simulation Using Monte Carlo Methods, Wiley, Chichester, pp. 87-124.
[62] Oakes, D. (1975). The Markovian self-exciting process. Journal of Applied Probability 12, 69-77.
[63] Ogata, Y. (1981). On Lewis' simulation method for point processes. IEEE Transactions on Information Theory 27, 23-31.
[64] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical Association 83, 9-27.
[65] Reynaud-Bouret, P. \& Schbath, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. Annals of Statistics 38, 27812822.
[66] Rizoiu, M., Lee, Y., Swapnil, M. \& Xie, L. (2017). A tutorial on Hawkes processes for events in social media. Preprint. Available at https://arxiv.org/ abs/1708. 06401.
[67] Rizoiu, M.-A., Mishra, S. Kong, Q., Carman, M. \& Xie, L. (2018). SiRHawkes: linking epidemic models and Hawkes processes to model diffusions in finite populations. Preprint. Available at https://arxiv.org/abs/1711.01679
[68] Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. Annals of Statistics 4, 673-684.
[69] Stabile, G. \& Torrisi, G. L. (2010). Risk processes with non-stationary Hawkes claims arrivals. Methodology and Computing in Applied Probability 12, 415-429.
[70] Torrisi, G. L. (2016). Gaussian approximation of nonlinear Hawkes processes. Annals of Applied Probability 26, 2106-2140.
[71] Tu, L. W. (2011). An Introduction to Manifolds. Second edition. Springer, New York.
[72] Yao, N. (2018). Moderate deviations for multivariate Hawkes processes. Statistics and Probability Letters 140, 71-76.
[73] Zhang, X., Blanchet, J., Giesecke, K. \& Glynn, P. (2015). Affine point processes: approximation and efficient simulation. Mathematics of Operations Research 4, 797-819.
[74] Zhu, L. (2013). Central limit theorem for nonlinear Hawkes processes. Journal of Applied Probability 50, 760-771.
[75] ZHU, L. (2013). Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. Insurance: Mathematics and Economics 53, 544-550.
[76] ZHu, L. (2014). Process-level large deviations for nonlinear Hawkes point processes. Annales de l'Institut Henri Poincaré - Probabilités et Statistiques 50, 845-871.
[77] Zhu, L. (2015). Large deviations for Markovian nonlinear Hawkes processes. Annals of Applied Probability 25, 548-581.
[78] Zwart, B., Borst, S. \& Mandjes, M. (2004). Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off sources. Annals of Applied Probability 14, 903-957.

## Summary

Our world is becoming increasingly interconnected, in an economic, social, political, and environmental sense. An event occurring at one time and location can trigger new events, in the future and other locations. As a result, such sequences of events have a contagious character: an earthquake that causes aftershocks, a virus that spreads in different countries, or financial panic that ripples across the international markets. The consequence of our interconnected world is that the behavior of such sequences of events has become increasingly complex, spreading over time and space, which warrants the search for mathematical models to describe this behavior.

It is a challenging task to build a mathematical contagion model that is general and flexible enough to accurately describe how different types of sequences of events spread. One candidate class of models is the class of mutually exciting point processes, which is studied in this thesis. A point process counts the points, which we call events, that occur over time. What makes these processes interesting is the mutually exciting part, which is a built-in feedback system that captures the contagious character of how events spread. When an initial event occurs at a certain location, it increases, or excites, the likelihood that events will occur in the future, at the same location as well as other locations. After a second event has occurred, it can in turn increase the likelihood of even more events in the future, both in the original location and in other locations. This is where the term mutually exciting comes from since events at different times and in different locations can have mutual effects on each other.

This thesis studies certain mathematical properties of mutually exciting point processes. In a general setting, it characterizes what the distribution is of the sequences of events. This is then used to obtain results on specific mathematical objects of interest: certain probabilities, expected values, variances, and more. Another aspect was to have flexible parameters by imposing limited restrictions on the model, allowing for greater applicability. The derived results can be used to efficiently calculate certain objects of interest and allow us to answer related questions. In what follows, I give an impression of the results in each chapter.

In Chapter 2, we introduce the mathematical objects that are needed to fully describe the dynamics of events that spread across time and locations. We heavily use the so-called branching representation, a tree-like structure that is an alternative de-

## Summary

scription of the mutually exciting point process. This representation, in combination with the exploitation of the feedback mechanism, allowed us to fully characterize the distribution of the sequences of events, by using a fixed-point argument. The obtained characterization is in a general setting in terms of parameter choice and the number of locations, opening the door to calculating probabilities, expected values, variances, and more. We also extended the point process to an induced population process, which means that we do not only count the number of events, but we allow for the number of events to decrease as well. One can think of practical applications in the context of epidemiology, where the population process would represent the number of infected individuals since this can fluctuate over time.

In Chapter 3, we focus on the compound version of the mutually exciting point process, motivated by the insurance industry. From the point of view of insurance companies, events such as an earthquake can trigger insurance claims that they have to pay out. The compound model consists of the sum of these claims and gives the insurance company a grip on the expected number of claims as well as their sizes, which is important in determining how much initial capital they need to allocate for these claims. For the mutually exciting point process as well as for the compound version, we derive the so-called Large Deviations Principle, which allows us to establish two important results. First, we characterize the probability of ruin for the compound model, which is the probability that the insurance company goes bankrupt because of too many or too large claims, and not having enough initial capital to cover it. Second, we derive a method to calculate exceedance probabilities, the probability that the compound model exceeds a certain threshold. Both these probabilities are usually very small, difficult to calculate, and computationally heavy to simulate. To mediate these problems, we develop and apply an importance sampling method to approximate these probabilities, which we then prove to be optimal.

In Chapter 4, we focus the setting of Chapter 2 on a more narrow class of mutually exciting point processes. By choosing the parameters in a certain way, the process obtains the Markov property, giving us a powerful tool to analyze the process. We obtain a more explicit characterization of the distribution of the spread of events, given in terms of systems of ordinary differential equations. We then reveal a recursive and nested structure, allowing us to calculate moments up to a certain order. Our method yields explicit expressions of the moments and brings important computational advantages.

## Nederlandse Samenvatting (Summary in Dutch)

Onze wereld is in de loop der tijd steeds meer met elkaar verbonden geraakt, in economische, sociale, politieke en milieutechnische zin. Een gebeurtenis die zich op een bepaalde tijd en plaats voordoet, kan nieuwe gebeurtenissen in de toekomst en op andere plaatsen veroorzaken. Hierdoor hebben dergelijke reeksen gebeurtenissen een besmettelijk karakter: een aardbeving die naschokken veroorzaakt, een virus dat zich in verschillende landen verspreidt, of financiële paniek die zich over internationale markten verspreidt. Het gevolg van onze onderling verbonden wereld is dat de dynamiek van dergelijke reeksen gebeurtenissen steeds complexer is geworden, wat de zoektocht naar wiskundige modellen rechtvaardigt om deze te kunnen analyseren.

Het is een uitdagende taak om een wiskundig besmettingsmodel te bouwen dat algemeen en flexibel genoeg is om nauwkeurig te beschrijven hoe verschillende soorten gebeurtenissen zich verspreiden over tijd en ruimte. Een potentiële klasse van modellen zijn mutually exciting punt processen, die in dit proefschrift worden bestudeerd. Een punt proces telt punten, wat wij gebeurtenissen noemen, die in de loop van de tijd plaatsvinden. Wat deze processen interessant maakt, is het mutally exciting deel, wat in essentie een ingebouwd feedbacksysteem is dat het besmettelijke karakter vastlegt van hoe gebeurtenissen zich verspreiden. Wanneer een eerste gebeurtenis op een bepaalde tijd en locatie plaatsvindt, vergroot, of prikkelt (excite), het de kans dat gebeurtenissen in de toekomst zullen plaatsvinden, op dezelfde locatie en op andere locaties. Nadat een tweede gebeurtenis heeft plaatsgevonden, kan het op zijn beurt de kans op nog meer gebeurtenissen in de toekomst vergroten, zowel op de oorspronkelijke locatie als op andere locaties. Hier komt de term mutually exciting vandaan, omdat gebeurtenissen op verschillende tijden en locaties wederzijds effect op elkaar kunnen hebben.

Dit proefschrift bestudeert bepaalde wiskundige eigenschappen van mutually exciting punt processen om vervolgens resultaten te verkrijgen over specifieke wiskundige objecten van belang: bepaalde kansen, verwachtingswaarden, varianties en meer. In deze algemene setting met flexibele parameters door weinig beperkingen op het model op te leggen, zijn de verkregen resulaten breed toepasbaar en omvattend. Deze theo-

## Nederlandse Samenvatting (Summary in Dutch)

retische resultaten kunnen worden gebruikt om bepaalde objecten van belang efficiënt te berekenen en stellen ons in staat gerelateerde vragen te beantwoorden. Hieronder geef ik een indruk van de resultaten in elk hoofdstuk.

In Hoofdstuk 2 introduceren we de wiskundige objecten die nodig zijn om de dynamiek van gebeurtenissen die zich verspreiden over tijd en locaties volledig te beschrijven. We maken veel gebruik van de zogenaamde branching representation, een boomachtige structuur die een alternatieve beschrijving is van het mutually exciting punt proces. Deze representatie, in combinatie met het slim benutten van het feedbackmechanisme, stelt ons in staat om de verdeling van de reeks gebeurtenissen volledig te karakteriseren. Dit verkrijgen we door een vast-punt relatie uit de boomachtige structuur af te leiden. Deze karakterisering is behoorlijk algemeen, in de zin dat we een flexibele parameterkeuze hebben en een vrije keuze in aantal locaties. We hebben ook het punt proces uitgebreid tot een populatie proces, wat betekent dat we niet alleen het aantal gebeurtenissen tellen, maar ook toestaan dat het aantal gebeurtenissen afneemt. Praktische toepassingen in dit opzicht zijn te vinden in de context van epidemiologie, waarbij het populatie proces het aantal geïnfecteerde personen zou vertegenwoordigen.

In Hoofdstuk 3 richten we ons op de compound versie van het mutually exciting punt proces, gemotiveerd door de verzekeringsindustrie. Voor verzekeringsmaatschappijen kunnen gebeurtenissen zoals een aardbeving aanzet geven tot verzekeringsclaims, die ze vervolgens moeten uitbetalen. Het compound model bestaat uit de som van deze claims en geeft de verzekeringsmaatschappij grip op het verwachte aantal claims en hun omvang, wat belangrijk is om te bepalen hoeveel solvabilitetitskapitaal ze moeten toewijzen om de claims uit te kunnen betalen. Zowel voor het mutually exciting punt proces als voor de compound model, leiden we, gebruikmakend van de resultaten van Hoofdstuk 2, het zogenaamde Large Deviations Principle af, wat ons in staat stelt twee belangrijke resultaten af te leiden. Ten eerste karakteriseren we de kans op faillissement voor het compound model, d.w.z. de kans dat de verzekeringsmaatschappij failliet gaat vanwege te veel of te grote claims, en niet genoeg solvabilitetitskapitaal heeft om het te dekken. Ten tweede leiden we een methode af om overschrijdingskansen te berekenen, d.w.z. de kans dat het compound model een bepaalde grens overschrijdt. Beide kansen zijn meestal erg klein, moeilijk te berekenen en computationeel zwaar om te simuleren. Om dit te verhelpen, ontwikkelen en passen we een importance sampling methode toe om deze kansen te benaderen, en vervolgens bewijzen we dat deze methode optimaal is.

In Hoofdstuk 4 richten beschouwen we een iets specifieker model van Hoofdstuk 2, een subklasse van mutually exciting punt processen. Door de parameters op een bepaalde manier te kiezen, verkrijgt het proces de Markov eigenschap, wat ons een krachtig hulpmiddel geeft om het proces te analyseren. We verkrijgen een meer expli-
ciete karakterisering van de verdeling van de reeks gebeurtenissen, gegeven in termen van systemen van gewone differentiaalvergelijkingen. Vervolgens ontsluieren we een recursieve en ingebedde structuur, waardoor we momenten tot een bepaalde orde kunnen berekenen. Onze methode levert expliciete uitdrukkingen van de momenten op en brengt belangrijke computationele voordelen met zich mee.

## List of Co-Authors

While the theoretical results in Chapters 2, 3 and 4 have been developed in collaboration with the advisors (Prof. Roger J. A. Laeven and Prof. Michel R. H. Mandjes), Raviar S. Karim has made very significant independent contributions to all three chapters.

More specifically, the three authors have contributed equally to the development of all the theoretical results, where Laeven has focused on the theoretical developments in Chapter 2, the first half of Chapter 3, and on the nested matrix structure in Chapter 4, Mandjes on those in Chapter 2, the second half of Chapter 3, and on the nearly unstable behavior in Chapter 4, and Karim on those in all three chapters.

Furthermore, Karim has conducted all the elements of the numerical experiments in all three chapters.

