EINDHOVEN UNIVERSITY OF TECHNOLOGY

Queues and risk models

Citation for published version (APA):
Badila, E. S. (2015). *Queues and risk models*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Technische Universiteit Eindhoven.

Document status and date:
Published: 01/01/2015

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

Queues and Risk Models

This work is part of the research project *Queues and Risk Models* (QUARM 613.001.017) funded by

Netherlands Organisation for Scientific Research

© Emil Şerban Bădilă, 2015.

Queues and Risk Models / by E.Ş. Bădilă. –

Mathematics Subject Classification (2010):
Primary 60K25 (Queueing theory), 91B30 (Risk theory, Insurance),
Secondary 60G51 (Processes with independent increments; Lévy processes), 62P05 (Applications to actuarial sciences and financial mathematics),
90B22 (Queues and service).

A catalogue record is available from the Eindhoven University of Technology Library.
ISBN: 978-90-386-3877-5

Printed by: Gildeprint

Queues and Risk Models

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op maandag 22 juni 2015 om 14.00 uur

door

Emil Şerban Bădilă

geboren te Târgu-Mureş, Roemenië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. E.H.L. Aarts
1e promotor:	prof.dr.ir. O.J. Boxma
co-promotor:	dr. J.A.C. Resing
leden:	prof.dr.ir. I.J.B.F. Adan
	prof.dr. H. Albrecher (Université de Lausanne)
	prof.dr. J.S.H. van Leeuwaarden
	prof.dr. M.R.H. Mandjes (University of Amsterdam)
	prof.dr. Z. Palmowski (University of Wrocław)

Acknowledgments

The contents of this book are the result of the four-year work on the topic presented in the title. Intensive as it was, it wouldn't have been possible without the contribution and support of many people, to whom I am most grateful.

Firstly, I am indebted to my supervisors. Onno was always kind, careful and eager to help and inspire me with a fantastic mathematical common sense and intuition; And on top of this with a great sense of humor. I am a lucky fellow to have you as a supervisor. I want to thank Jacques for all the fruitful discussions we had. Throughout all our projects, you always managed to make the right remarks and raise just objections, especially when I was being sloppy. When I had a question, the doors of your offices were always open.

I would also like to thank Hansjoerg Albrecher, Zbigniew Palmowski, Michel Mandjes, Ivo Adan and Johan van Leeuwaarden for agreeing to be part of the committee. Special thanks go to Hansjoerg for inviting me to visit his group in Lausanne and to Jevgenij for taking me around the city. I also want to thank Zbyszek for the visit I paid to Wrocław and for the good work we have done together; This visit has also been a great opportunity to have some nice discussions with Thomasz Rolski.

EURANDOM felt like a big family. Without trying to be precise, I witnessed here two successive generations of PhD students. I have to mention the pleasant evenings at Botond, Robert and Elena's house - the EURANDOM house as we used to call it. The barbecues, the games and the pub quizzing have always been a good distraction. It was great to have around the "old" bunch of jolly fellows (in random order): Tim, Stella, Paulo, Julia, Peter, Jan-Pieter, Eleni, Enrico, Carlo, Martin, Florence, Sander. And a special mention for Jaron whom I've never seen performing live yet and Alessandro, with whom I had the trip to the other end of the world. Then there is the new generation; I had many nice evenings, pizzas and board games with Britt, Jori, Alessandro, Maria Luisa, Fabio, Thomas, Gianmarco, Murtaza, Szilard, Clara, Rick, Bart, Fiona and Mirko. Besides, there were also the squash and the basketball games with Fabio. Thank you all for the great time.

I also wish to thank Răzvan, Ada, Andrei, Andreea, Dragoş, Ionuţ, Delia, Radu; I could not write this without mentioning you here. I wish I would see you more often.

Finally, I am grateful to Maria, Daniela and to my parents for their unconditional love and support.

Şerban Bădilă
Eindhoven, May 2015

Contents

Chapter 1

Introduction

There are many daily life examples related to traffic and the problem of congestion, which occur whenever a given resource cannot keep up with the rate of arrival of service requests. Examples of offered resources include Internet bandwidth, the width of a highway as part of a traffic network, the speed of the central processing unit (CPU) in a computer, or the working speed of a server at a counter in a supermarket. A natural way to mitigate the unavoidable congestion issues is to consider a buffer in which arriving customers/jobs can queue up waiting for service. The study of phenomena related to congestion relies significantly on the mathematics of Queueing Theory, and this includes related problems such as the design and optimization of queueing systems. There is a dual perspective which makes queueing processes appear as mirror images of so-called risk reserve processes (or better called surplus processes). These latter describe the dynamics of the surplus process of an insurance portfolio, or the evolution of assets on the stock market.

The amount of processing time demanded from a CPU is not constant among the various requested jobs. Even more so, one cannot expect fixed deterministic time points at which cars arrive at an intersection. Similarly, arrival epochs of accident claims incurred by an insurance portfolio are subject to hazard and so is their severity. Thus one is naturally led to consider the input processes in these systems as being stochastic in nature.

Among the associated key performance measures one can think of are the waiting time of a typical customer before receiving service, the workload demanded from the server at any point in time, the queue length etc., and in the set-up of surplus processes one is typically interested in the ruin probability, the time to ruin and the deficit of the surplus process at ruin. In view of the above, these performance measures are to be regarded as random variables or closely related entities, and this leads to considering the input in the form of a stochastic process.

The mathematical problems that appear in applications coming from queueing theory and insurance can typically be reduced to the study of the supremum and infimum functionals of a stochastic process, together with their associated excursion processes (as excursions away from the minimum). Part of the study of fluctuations is related to ergodic theorems that specify conditions on the traffic load under which the above extreme functionals converge as time goes to infinity, and respectively conditions under which the distribution of an excursion length is a probability distribution (which may be defective). By the theory of fluctuations we mean the study of these problems and related ones, like the rate of convergence of the above functionals in the ergodic case. These appear in many related areas of Probability and Statistics, however, from an applications point of view, the focus in this thesis will be on queueing and risk reserve processes.

There is a convenient way to talk about queues and risk reserve processes at the same time without explicitly mentioning any of them. Let $(t_n)_n$ be a countable sequence of random points on $\mathbb{R}$ (thought of as the time axis), and make correspond to each point t_n a random element B_n of some probability space. We will call the sequence of tuples $\{(t_n, B_n)\}_n$, a marked point process (allow $n \in \mathbb{Z}$). The first chapters are dedicated to the study of marked point processes with the special feature that the inter-arrival process is renewal and in the positive flow of time, the current mark size is allowed to depend on the time until the next arrival epoch. The positive direction for the time flow corresponds to queueing applications, so that if we denote the inter-arrivals by $A_n := t_{n+1} - t_n$, the sequence $\{(A_n, B_n)\}_n$ is assumed independent and identically distributed (i.i.d.). For the setting of risk reserve processes, the direction in which time flows is reversed so that the current mark size may depend on the time elapsed since the previous mark epoch. The reason for changing the direction of time is that, in some cases, queueing and surplus processes can be backwards coupled one with another in such a way that their key performance measures can be related. And then the processes are said to be in duality. A special focus will be on the case when the point process is Poisson with the Lebesgue measure as intensity. Also of special interest is when the marks are random vectors. It may happen that the distributions of the vectors of marks may be supported on proper linear subspaces of the embedding space. In the case of Poisson point processes, this can happen if we merge different arrival processes that mark only proper subspaces. For example, in queueing terms, for two servers working in parallel we can have arriving customers that demand service only from one of the two available servers, but there can also be customers that demand work from both servers simultaneously. Similarly, an insurer can choose to reinsure only part of its portfolio; so that some types of incoming risks will not be shared.

About the motivation behind this thesis, there are two main problems investigated that are overlapping:

1. The study of multidimensional queueing processes and insurance/risk processes. This is a difficult topic in itself and a significant part of the thesis is dedicated to its study. The focus will be on calculating the stationary distribution of the maximum (under various orderings on $\mathbb{R}^n$) of the vector-valued input process, under suitable stability conditions that ensure these maxima have proper distributions in the limit. Similarly as in one dimension, from the queueing perspective these are related to the vector-valued amount of work in the whole system, but the maxima also give various types of ruin probabilities in the related surplus process (there are several ways in which ruin can occur, for example it can ultimately happen in all the components of the surplus process or in just one of them, etc.).

A special class of risk reserve processes is related to the so-called proportional reinsurance contracts for which a generic claim (which corresponds to the marks of the arrival process) is shared by two insurance companies in fixed proportions. A relevant question in general and for this problem in particular is to understand the asymptotic behaviour of the ruin probability as a function of the initial capital, for large values of this capital.

Another special feature in several dimensions is that the marginal processes may interact with each other. This is the case for the so-called coupled processor model; for two processors, this is an interacting queueing system in which as soon as a processor becomes idle (the marginal amount of work hits level 0) it switches to assist the other processor, if there is any work in the other buffer at this instant.

2. To try to remove, as much as possible, the requirement that the inter-arrival times and the mark sizes form independent families of random variables. In many applications arrival patterns are sometimes periodic or they can become bursty from time to time. Going to the heart of the matter, the question is whether one can study fluctuations of sums of variables that are neither independent nor stationary and even more, if, when dealing with such problems, one can forge weapons that are as effective as the Fourier transform methods that are traditionally used in the study of sums of independent variables.

These two topics are intertwined: in several dimensions, there are various correlations between the components of the vector-valued processes that are considered as input. Without having to complicate the model by allowing the marginal systems to interact, the possibility to have dedicated arrivals into a subset of the components causes correlations in the dynamics of the whole system and this is already a source for many difficulties. This is the case for instance when studying networks of queues or several insurance companies that share risks.

The second problem is relevant on its own. The assumption of renewal arrivals, which was used systematically by Pollaczek, Lindley, Kendall and many others, is appealing from a mathematical perspective because, while it generalizes the assumption of Poisson arrivals, it is still tractable if one uses the analytic theory of Fourier(-Stieltjes) transforms and the related Laplace-Stieltjes transforms. However, the renewal structure is broken as soon as one allows the rate of the Poisson process to be time dependent or even stochastic by itself (as in a Cox process). Even more so, known measurements of the internet traffic show that the packet sizes (these stand for the successive marks of the arrival process) are correlated with each other even at large lags (long range dependence). Both these extensions and their associated fluctuation problems are intensively studied in the literature; they are broad research topics so we will end up barely scratching the surface and leaving many contributions unmentioned in the following sections.

The rest of this chapter is organized as follows: Section 1.1 gives an overview of the single server queue G/G/1 with various correlation devices in the input process and the related work in the insurance literature on the Sparre-Andersen insurance model. Section 1.2 reviews the literature on multivariate queueing systems and contributions from the related insurance literature. Section 1.3 is an informal introduction to the topic of duality and prepares the ground for Chapter 2. Finally, Section 1.4 gives an overview of the thesis and briefly presents its main contributions.

1.1 Single server queues and Sparre-Andersen risk reserve processes, with correlations

We begin by introducing the basic queueing system, the single server queue, and the related risk reserve process. The key performance measures that will be studied throughout the thesis are also introduced.

We will use Kendall's notation for queues, so the G/G/1 queueing system takes the marked point process with renewal arrivals as input which has both the inter-arrivals A_n and the marks B_n generally distributed (hence the acronym G). We will also use the notation GI/G/1 to mean that the sequence of inter-arrival times is renewal and that the service times are correlated either among themselves or with the inter-arrival sequence in a way that will always be specified when describing the model. Much of the literature concerned with obtaining exact results (in terms of transforms or by direct calculations) assumes the model to have both the service time sequence and the inter-arrival sequence i.i.d. and independent one from the other. GI/GI/1 is then usually used as a notation instead, although it is clear at this point that these are simple (and imprecise) conventions and we will use them like this. There is a single server that handles the service request B_n of customer n; the server works at constant rate $c > 0$. Customers that find the server busy upon arrival queue up in a buffer. Typically, we are interested in the amount of work waiting in the buffer at any point in time. We assume that the buffer is initially empty and we denote by V_t the workload at time t. We take by convention V_t to be left-continuous with right limits, so that V_{t_n} is the amount of work in the system right before customer n arrives.

It follows from the definition that the embedded workload at arrival epochs V_{t_n} must be the solution of Lindley's recursion

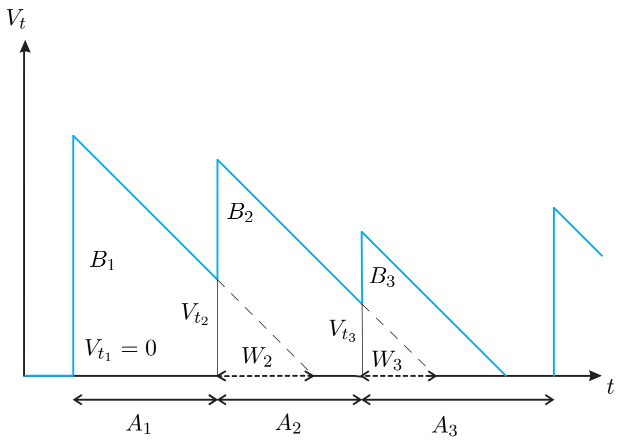


FIGURE 1.1: A path of the single server queue.

$$\begin{aligned} V_{t_{n+1}} &= \max(V_{t_n} + B_n - cA_n, 0), \quad n = 1, 2, \ldots, \\ V_{t_1} &= 0. \end{aligned} \tag{1.1}$$

The workload process has positive jumps of size B_n at the nth arrival instant and decreases linearly in between arrival epochs such that level 0 is an impenetrable barrier: after it hits level 0 (which can only happen by drifting towards it) the process stays at 0 until the next arrival epoch. Because of this, V_t is usually called the reflected version of the input process (V_{t_n} is the reflected version of the random walk).

We will also denote by W_n the waiting time of customer n in the buffer, $n \geq 1$. Under the "first come first served" (FCFS) server policy, $W_n = V_{t_n}/c$ because the server is working at constant rate (V_{t_n} is sometimes called the virtual waiting time because of this identity). The above assumptions on the input process imply that $(V_{t_n})_n$ is a Markov chain on $\mathbb{R}_+$ conditioned to start from 0. It can be shown that the unique solution of Recursion (1.1) is given by

$$V_{t_{n+1}} = S_n - \inf_{k \leq n} S_k, \tag{1.2}$$

where we denote by $(S_n)_n$ the random walk with increments $X_k := B_k - cA_k$, $k \geq 1$, and $S_0 = 0$.

The workload is partitioned into regeneration cycles w.r.t. state 0. The length of the excursion of V_t above level 0 is the busy period of the server. Other measures of performance are the idle periods of the server, and the number of customers that are served during a busy period. The stability issue is whether state 0 is positive recurrent for the chain $(V_{t_n})_n$. The idle period between epochs t_n and t_{n+1}, $n \geq 1$, is complementary to $W_{n+1} = V_{t_{n+1}}/c$:

$$I_{n+1} := -\min(V_{t_n} + B_n - cA_n, 0)/c, \quad n \geq 1. \tag{1.3}$$

The related surplus process $(R_t)_{t\geq 0}$ evolves in the following way: starting with a level $u > 0$ (initial capital) R_t drifts linearly upwards at a positive rate (premium income rate) and at an arrival epoch t_n a claim B_n is requested which immediately decreases the surplus level:

$$R_t(u) = u + ct - \sum_{i=1}^{n(t)} B_i,$$

with $n(t)$ the number of arrivals up to time t. We take $(R_t)_t$ to have right-continuous paths with left limits (càdlàg). If we focus on arrival epochs (since ruin can only occur at these points in time)

$$R_{t_{n+1}}(u) = u - \sum_{i=1}^{n} (B_i - cA_i) = u - \sum_{i=1}^{n} X_i, \quad n \geq 1.$$

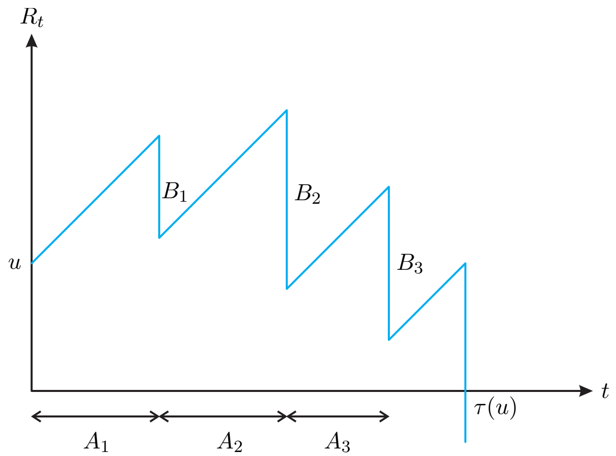


FIGURE 1.2: A path of the surplus process.

We define the time to ruin $\tau \equiv \tau(u)$ of the process $R_t(u)$ as the hitting time of the negative half-axis

$$\tau(u) = \inf\{t > 0\ ;\ R_t(u) < 0\}. \tag{1.4}$$

From our perspective, it will turn out to be useful to let the process continue if it happens to hit state 0, or if started from 0. The ruin probabilities for the finite and infinite time horizon are

$$\mathbb{P}(\tau(u) < t_n),\ n \geq 1, \quad \text{and} \quad \mathbb{P}(\tau(u) < \infty),$$

with extra conditions needed for the latter to be strictly less than 1. The above ruin time is closely related to the so-called maximum aggregate loss of the risk reserve process. For finite horizon this is just

$$M_n := \max(S_0, S_1, \dots, S_n), \tag{1.5}$$

for $S_n := \sum_{i=1}^n X_i$ and $S_0 = 0$. Under suitable conditions which ensure convergence, the all time maximum aggregate loss is $M := \lim_n M_n$, so that we have directly from their definitions:

$$\tau(u) \leq t_n \Leftrightarrow M_n > u; \qquad \tau(u) < \infty \Leftrightarrow M > u.$$

The second equivalence holds because the sequence $(M_n)_n$ is non-decreasing. Finally, another quantity of interest is the deficit at ruin; this is defined as $-R_{\tau(u)}(u)$.

Queueing models: When studying the single server queue G/G/1, it is usually assumed that all inter-arrival times and service requirements are independent.

A general approach to allowing correlations between successive inter-arrivals/service requirements is the class of Markov Arrival processes; see Neuts [90] and Lucantoni [85].

Phase-type inter-arrivals and Markov-Modulated Poisson arrival processes are contained in this class.

Following Lucantoni [85], a natural way of correlating inter-arrivals with mark sizes is the Batch Markov Arrival Process (BMAP) and its associated queueing system, the BMAP/G/1 queue, where the arrival of a batch of size $k \geq 1$ causes the request to process k i.i.d. service components, each having some generally distributed size. The BMAP/G/1 queue provides a versatile framework to model dependence between successive inter-arrival times but also dependence between inter-arrival times and service requirements. In Combé and Boxma [45] the BMAP is also used to study an M/G/1 queue in which service requirements depend on the previous inter-arrival times; see Borst et al. [33] for a different approach to the latter form of dependence, which does not use the MAP machinery.

An important paper regarding dependence between inter-arrival and service requirements is the one by Adan and Kulkarni [2]. They consider a single server queue with Markov-dependent inter-arrival and service requirements: a service requirement and subsequent inter-arrival time have a bivariate distribution that depends on an underlying Markov chain which jumps at customer arrival epochs. The inter-arrival times in [2] are exponentially distributed, with rate λ_j when the Markov chain jumps to state j. See also Constantinescu et al. [48], where a different approach is used to study similar models for the risk reserve process. The methods used therein are based on operator theory (Heaviside operational calculus).

It should be observed that the analysis of a GI/G/1 queue with some dependence structure between a service requirement B_i and the *subsequent* inter-arrival time A_i is intrinsically easier than that of a GI/G/1 queue with some dependence structure between A_i and the *next* B_{i+1}. The reason is that B_i and A_i only appear as a difference in the Lindley recursion (1.1) for the waiting time W_i of the ith arriving customer. In Chapter 3 we will study the problem of fluctuations of a queue which has matrix exponential service requirements (see Bladt and Nielsen [29]) that are also correlated with subsequent inter-arrival times. The working assumption is given in terms of the Laplace-Stieltjes transform of the vector composed of the inter-arrival time together with the corresponding service requirement, namely, it is assumed that this is a rational function in both arguments. The classes of phase-type random vectors known in the literature as Assaf's [18] or Kulkarni's [79] are special instances. This class allows us to obtain detailed, explicit, results for the steady-state waiting time, workload and idle period, and even to stochastically compare the waiting time distributions for various types of correlation.

By focusing on the arrival instants of marks, the study of the waiting time and idle time distribution in the GI/G/1 queue reduces to the study of a random walk with increments that have a rational characteristic function and can thus be continued analytically in the whole plane except for a finite number of poles.

Insurance risk: Having discussed the queueing literature with dependence between inter-arrival time and service requirement, let us now turn to the insurance risk literature with dependence between inter-claim time and claim size. In recent years, this has been a hot topic in risk theory. Albrecher and Boxma [4] derive exact formulas for the ruin probability in a Cramér-Lundberg model with a threshold-type dependence between a claim size and the next inter-claim time. In Albrecher and Boxma [5] a much more general semi-Markovian risk model is being considered, which bears some resemblance to the queueing model in Adan and Kulkarni [2]. Kwan and Yang [80] consider a specific threshold-type dependence of claim size on previous inter-claim time; in Albrecher et al. [6] this is put in the larger framework of Markov Additive Processes. Another specific dependence structure between claim size and previous inter-claim time is treated in Boudreault et al. [34], where it is assumed that the conditional density of a claim, given the previous inter-arrival is a mixture of two arbitrary probability density functions. Asymptotic results were obtained in Albrecher and Kantor [7], where the relation between the dependence structure and the Lundberg exponent is studied. Also Albrecher and Teugels [8] give asymptotic results for the finite and infinite horizon ruin probabilities when the current claim size and the previous inter-claim time are dependent according to an arbitrary copula structure.

1.2 Queueing systems and risk reserve processes with multiple components

An important part of the thesis is dedicated to the study of multidimensional queueing systems and the related risk reserve processes. In particular, we look into the possibility of extending the duality of Siegmund [97] to processes on higher dimensional state spaces. First, we will focus on queueing models which have a compound Poisson input process with a negative drift and their dual insurance risk counterparts which take a similar compound Poisson input with a positive drift. The negative/positive drift is in terms of the componentwise ordering of $\mathbb{R}^n$.

Queueing models with several queues in parallel and with an arrival process in which there may be simultaneous arrivals - often called fork-join queues - have many applications in computer-, communication- and production systems. These are models where jobs are split among a number of different processors, communication channels or machines. The servers work in parallel and each has its own dedicated buffer where jobs wait to be processed. Clearly, the queues in these models are dependent due to the simultaneous arrivals. In general this makes an exact analysis of the model very difficult, and by exact analysis is meant the derivation of (Laplace-Stieltjes transforms for) the distributions of the relevant performance measures. Only in the case of two queues, exact results are available; see, e.g., Flatto and Hahn [59], Wright [105], Baccelli [21], De Klein [76] and Cohen [43]. An early paper, classical for the use of complex functions and singular integrals in queueing theory, is Pollaczek's [92]. The above mentioned works rely heavily on advanced complex function theory, including the theory of boundary value problems. This not only makes the methods involved, it is often difficult to recognize the stochastic nature of the initial problem in the manipulations. It comes as no surprise that approximations and asymptotic studies are more popular than exact methods for these multidimensional models. Given that it is still imperative to understand fluctuation theory in higher dimensions, a starting point would be to understand how one can leverage the stochastic nature of the problem to actually guide the analysis (complex as it may be).

For the model with more than two servers no exact analytical results are available in the literature. In this case, bounds and approximations for several performance measures have been developed, see e.g. Baccelli et al. [22] and Nelson and Tantawi [88, 89].

Studies of multidimensional risk reserve processes are scarce in the insurance literature, although results about risk measures related to such models are highly relevant both from a theoretical and a practitioner's perspective. Multivariate ruin problems are relevant because they give insight into the behaviour of risk measures under various types of correlations between the insurance lines. Like in queueing theory, it is natural to study insurance risk processes with simultaneous arrivals of claims in several insurance lines. One example is presented by multiple insurance lines within the same company which are interacting with each other as they evolve in time, via, say, coupled income rates. Another typical example is an umbrella type of insurance model, where a claim occurrence event generates multiple types of claims which may be correlated, and each type of claim is paid from its corresponding component, such as car insurance together with health insurance or insurance against earthquakes. Yet another class of models is related to reinsurance problems, where a claim is shared between the insurer and one or more reinsurers.

Avram et al. [19, 20] have studied the joint ruin problem for the special case of proportional reinsurance. In particular, they derive the double Laplace transform with respect to the two initial reserves of the survival probabilities of the two companies. One of the key observations in [19, 20] is that, due to the fact that companies divide claims in some specific proportions, the two-dimensional ruin problem may be viewed as a one-dimensional crossing problem over a piecewise linear barrier. Badescu et al. [23] have extended the two-dimensional model of [19, 20] by allowing, next to the arrivals of claims for which the two insurers divide the claim in some specific proportions, also extra arrivals of claims which are fully paid by one of the insurers (e.g., insurer 1). They show that under some conditions also in this model the previously mentioned reduction to a one dimensional problem still holds. However, in [23] the authors do not consider the double Laplace transform with respect to the two initial reserves of the survival probabilities of the two companies (their main focus is on the Laplace transform of the time until ruin of at least one insurer).

An early study of multivariate risk measures can be found in the paper of Sundt [101] about developing multivariate Panjer recursions which are then used to compute the distribution of the aggregate claim process, assuming simultaneous claim events and discrete claim sizes. Other approaches are deriving integro-differential equations for the various measures of risk and then iterating these equations to find numerical approximations as in Chan et al. [37] and Gong et al. [67], or computing bounds for the different types of ruin probabilities that can occur in a setting where more than one insurance line is considered (see Cai and Li [36] who consider multivariate phase-type claims). It is worth mentioning that very few papers (like Avram et al. [19], Badescu et al. [23]), analytically determine, e.g., the ruin probability for insurance models with more than one company (see also [17], Ch. XIII.9 for an overview of this topic).

In an attempt to solve the integro-differential equations that arise from such models, Chan et al. [37] derive a Riemann-Hilbert boundary value problem for the bivariate Laplace transform of the joint survival function (see Section 5.5 for details about such problems arising in the context of risk and queueing theory and Cohen and Boxma [41] for an extended analysis of similar models in queueing). However, Chan et al. [37] do not solve this functional equation. The law of the bivariate risk reserve process usually considered in the above mentioned works is that of a compound Poisson process with vector-valued jumps supported on the negative quadrant in $\mathbb{R}^2$, conditioned to start at some positive level, and linearly drifting along a direction vector that belongs to the positive quadrant. In Chapter 5 a similar functional equation is taken as a departure point, and it is explained how one can find transforms of ruin related performance measures via solutions of the above mentioned boundary value problems. It is also shown that the boundary value problem has an *explicit* solution in terms of transforms, if the claim sizes are ordered.

Another line of research in queueing theory regarding queues in parallel is the study of queueing systems in which the servers interact with each other. A natural example is the so-called coupled processor model where, given two processors in parallel, as soon as one of them becomes idle, it switches to process work from the other buffer, thus strictly improving the performance of the system. A pioneering paper in this area is Fayolle and Iasnogorodski [55], who consider two parallel M/M/1 queues with independent Poisson arrival processes, and such that the service rate in one of the queues changes as soon as the other queue becomes idle. This system is solved for the steady state number of customers in both queues by reducing the problem to a boundary value problem of a Riemann-Hilbert type. In Cohen and Boxma [41], this model is generalized by dropping the assumption that the service requirements are exponentially distributed. It is shown that the problem of determining the joint workload distribution reduces again to a Riemann-Hilbert boundary value problem. In Cohen [43], the analysis is further extended for the case when, with some probability, arriving customers may also request service simultaneously from both queues. Moreover, the service requirement of a customer is allowed to depend on whether he finds one of the queues to be empty, the so-called semi-homogeneous workload process. Both in [41] and [43] the focus is on the transient problem, that is the study of the time dependent amount of work/queue lengths. A related paper is Ivanovs and Boxma [71], about a two-dimensional insurance model where capital is being transferred - if available - from one of the two components, if one has negative surplus level.

1.3 Duality

There are several connections between single server queueing systems and risk reserve processes. Such connections were already revealed by Sparre-Andersen [10], [12], [13] (see also [11] for the introduction of the ruin model related to the GI/GI/1 queueing system) and then by Lindley [82] and Feller [57]. These relations have sometimes been called duality.

A famous duality is the one between the classical M/G/1 queue (M stands for memoryless inter-arrivals; i.e. the arrival process is compound Poisson) and the classical Cramér-Lundberg model, where the inter-arrival times in both models have the same exponential distribution while also the service times and claim sizes have the same general distribution. To be more precise, the distribution of the amount of work V_{t_n} at arrival epochs in the queue is related to the ruin probability in the corresponding surplus process, as defined in Section 1.1 by

$$\mathbb{P}(V_{t_n} > u | V_0 = 0) = \mathbb{P}(\tau(u) \le t_n), \tag{1.6}$$

where we took $t_1 = 0$ as the reference time and $\tau(u)$ is the exit time of $(R_t)_t$ from the non-negative half-axis. In this way, the reflected version V_t of the input process has 0 as a reflecting barrier, whereas the dual surplus process has the negative half-axis as an absorbing set.

The input processes which are among the most tractable and that emphasize these relations in a clear way, are the random walks. By a random walk we will mean in this thesis the process in discrete time which consists of partial sums of independent and identically distributed random variables. A more general class is that for which the sequence of increments is assumed to be stationary only, and then allowing the increments $(X_n)_n$ for $n \in \mathbb{Z}$, a stationary version of the workload can be constructed as in Loynes [83]:

$$V_{t_{n+1}} \stackrel{d}{=} \max(0, X_n, X_n + X_{n-1}, \ldots, X_n + \ldots + X_0, \ldots). \tag{1.7}$$

Since the sequence $(X_n)_n$ is stationary, the same holds for this construction of the embedded workload sequence: $V_{t_n} \stackrel{d}{=} V_{t_1}$, for all $n \ge 1$. In words, time is reverted starting from the nth arrival epoch (w.r.t. the queueing time, say), and the horizon is infinite. If we introduce

$$M_k := \max(0, X_n, X_n + X_{n-1}, \ldots, X_n + \ldots + X_{n-k+1}),$$

then the right-hand side of (1.7) can be seen as the infinite horizon maximum aggregate loss (1.5) in a risk reserve process driven by the increments $X_k^* := X_{n-k+1}$, $k \ge 1$. This means that the kth inter-arrival time and claim size pair is (A_{n-k+1}, B_{n-k+1}) (we will come back to this coupling in Chapter 2). If we denote with $M = \lim_{k\to\infty} M_k$, the all time maximum aggregate loss which appears on the right-hand side of (1.7), then M is directly related to the infinite horizon ruin probability by $\mathbb{P}(\tau(u) < \infty) = \mathbb{P}(M > u)$ (see (1.4) for the definition of $\tau(u)$). Thus using Loynes' construction, we obtain a relation between the stationary version of the workload and the infinite horizon ruin probability. In this case, (1.6) becomes

$$\mathbb{P}(V_{t_\infty} > u) = \mathbb{P}(\tau < \infty \,|R_{t_0} = u),$$

where in this case, V_{t_∞} stands for the stationary version of the workload embedded at arrival epochs.

A more general instance of the above duality is studied in Siegmund [97], where it is shown that for any stochastically monotone Markov process on $\mathbb{R}$, the law of the process with a reflecting barrier can be put into a relation with the law of the version of the process which has the same barrier as an absorbing one instead. The motivation comes from the remark that it is numerically more effective to simulate hitting probabilities than distributions of reflected processes in equilibrium. These results are all for the real line and it would be interesting to know if these can be generalized to higher dimensions. In this generality however, one has to construct in a case by case fashion the transition probabilities of the dual process, and this can become complicated in some cases. See Asmussen [15], [16], Ch. XI.2, for duality extended to Markov modulated processes as well, and [15] for several open problems.

Another type of duality is obtained by simply changing the sign of the increments X_n. In terms of the input processes, the inter-arrival times in the queueing system will correspond to claim sizes in the risk reserve process, and the service requirements of customers become inter-arrival times for the surplus process. For example, the standard Cramér-Lundberg model that consists of a Poisson arrival process and i.i.d. generally distributed claims is obtained from the G/M/1 queueing model (whereas the previously discussed duality linked it to the M/G/1 queue).

This duality with the G/M/1 queue is useful because it relates performance measures other than the workload/maximum aggregate loss; for example, the length of a busy period (the excursion away from the infimum) jointly with the number of customers served during this period in the single server queue corresponds under this type of duality to the time to ruin in the risk reserve jointly with the number of claims paid up to this time. In addition, the length of the idle period during a busy cycle relates to the deficit at ruin of the risk reserve process: $(-R_\tau)$. This relation was pointed out already in Prabhu [93] and the references therein; see also Frostig [64] and Löpker and Perry [84].

1.4 Outline and contributions

The thesis is structured in the following way: In Chapter 2, we extend the duality relation in the sense of Siegmund [97], as introduced in the first part of the previous section, to several multidimensional queueing systems with parallel servers with and without interactions, which will enable us to relate them to risk reserve processes with multiple branches. The existence of interactions between the servers (the coupled processor model) is changing the geometry of the absorbing sets of the dual processes. Key for this type of duality is the possibility to realize the state space as an ordered vector space.

In Chapter 3 we study the GI/G/1 queue which has the current service time correlated with the time until the next arrival epoch. At the same time we will consider the dual Sparre-Andersen insurance model. The focus here will be on calculating the waiting time distribution and the idle period, which is done under the assumption that the distribution of the inter-arrivals jointly with the service requirements is bivariate matrix exponential – see Bladt and Nielsen [29]. By duality, the waiting time corresponds to the ruin probability of the risk reserve process which has the current claim size correlated with the time elapsed since the previous arrival epoch. We also explore the relation between the stationary workload and the stationary waiting time. It is shown that this relation is analogous to the one that connects the ruin probability for the delayed risk reserve process and the ruin probability for the ordinary risk reserve process. By definition, the delayed risk reserve process has the first arrival epoch distributed as the forward recurrence time of a typical inter-arrival (the renewal arrival process is started in stationarity). The ordinary risk reserve process has the same distribution of the first inter-arrival time as the subsequent inter-arrival times (this is a version of the risk reserve process conditional on an arrival happening at time 0). The results obtained give insight into the effect of the correlation between inter-arrivals and service requirements/claim sizes. It is shown that a negative correlation increases the waiting time distribution/ruin probability and a positive correlation decreases these performance measures when compared to independent input. The increase/decrease are both in the sense of convex ordering (see Stoyan [100], Ch. 1).

In Chapter 4, we continue with the set-up of the random walk with scalar increments but this time we study it in greater generality. The structure of the increment is still given as the difference of marginals of a sample from (A, B). Without making any assumptions on the distribution of the pair (A, B), we obtain integral representations for the busy period, idle period and workload in the underlying queueing model. These are obtained by generalising a well known relation that represents functionals of probability distributions in terms of their characteristic functions (Hewitt's inversion formula [69]). Obtaining the above mentioned representations is equivalent to solving a special kind of Riemann boundary value problem for the imaginary axis (this is related to what is usually called the Wiener-Hopf factorization in the probability literature). By virtue of the duality relation described at the end of Section 1.3, the integral representations are also valid in the context of ruin, because the busy period corresponds to the time to ruin and the idle period to the deficit at ruin in the dual risk reserve process. If the two-dimensional Laplace-Stieltjes transform of the pair (A, B) is a rational function in at least one of its arguments, then the transforms of these performance measures can be evaluated explicitly, by contour integration.

In Chapter 5, we study a two dimensional queueing system composed of two parallel processors which receive input according to a compound Poisson arrival process, with simultaneous arrivals. We show that under ordered service times, the steady state workload has an explicit form, and moreover a stochastic decomposition holds in steady state, which can be interpreted probabilistically in terms of the busy periods of one of the processors (the excursion lengths of the compound Poisson process above its successive minima). The results are further extended to k processors in parallel. We also explain how the more general problem, without the ordering assumption, can be related to the theory of boundary value problems and singular integrals. By virtue of results from the multivariate duality which is discussed in Section 2.2, the distribution of the steady state waiting time vector is related to the ruin probability as a function of the initial vector of starting capital levels in the risk reserve process.

The Poisson arrival assumption is generalized in Chapter 6, allowing for a two-dimensional renewal arrival process with general inter-arrivals which may also be coupled with the corresponding two-dimensional mark size. In this way, we extend the BMAP set-up of Chapter 3 to two dimensions. Using the duality relations from Chapter 2, the results obtained also give the ruin probability for two-dimensional risk reserve processes, as a function of two arguments which represent the starting capital in both marginal risk reserve processes. As a particularly important example, the ruin probability for proportional reinsurance contracts is obtained.

The asymptotic behaviour of the ruin function for the proportional reinsurance process is studied in Chapter 7, assuming that the common claim distribution that is being partitioned belongs to the class of subexponential distributions (long-tailed), see Foss et al. [60]. This is carried out using only probabilistic methods.

Chapter 8 is dedicated to the study of two coupled processors. The focus here is on the stationary workload in the system. It is shown that if the service time vector is ordered, the amount of work in this model can be related to the amount of work for two parallel processors without coupling, and thus the stability condition and the steady-state waiting time follow from the results of Chapter 5. Then the results are extended to several coupled processors.

Remark. *The results of Chapter 3 are published in [24]; Chapters 5, 6, and 8 are published in [25], [26] and [27], respectively. Chapter 7 is part of an ongoing project together with Sergey Foss, Zbigniew Palmowski and Tomasz Rolski. Finally, Chapter 4 has not yet been submitted for publication.*

Chapter 2

Duality

In Section 1.3 we pointed out that there is a duality relation between single server queues and risk reserve processes that involves time reversion. In the present chapter we will build on this (Section 2.1) and we shall explore this type of duality between queueing systems and risk reserve processes in more depth, with a special focus on multivariate models. In Section 2.2 we present the standard model for d queues in parallel with correlated service requirements and show that this model has a dual risk reserve process that consists of d insurance companies which receive correlated claims. In Section 2.3 it is proven that there exists a duality relation which connects the two coupled processors model (an interacting system of queues) to an absorbing process which has a certain open convex subset of the plane as the absorbing domain. This is then used to derive stability conditions for the queueing system.

For a random walk $(S_n)_n$, the duality frequently used in this thesis relies on the fact that the reflected version of S_n (the solution of (1.1)) is the same as the supremum of the time-reversed walk:

$$S_n - \inf_{0 \le k \le n} S_k = \sup_{0 \le k \le n} S_k^*,$$

where $S_k^* := S_n - S_{n-k}$, for a fixed epoch $n > 0$ and $0 \le k \le n$. This will be used to relate the amount of work in a queue to the probability of ruin in a corresponding risk reserve process, similarly as in (1.6). We will prove this in the more general case when $(S_n)_{n \ge 0}$ is vector-valued, in which case one can still construct a dual risk reserve process by using geometric arguments.

This type of duality can be argued with the help of a coupling, as in Asmussen [15]. As in Chapter 1, let $X_j := B_j - cA_j$ be the increments that determine $(V_{t_j})_j$ via (1.1). For fixed $n \ge 1$ and $(X_1, \ldots, X_n)$ the sample vector of the increments up to time n, take the input in the risk reserve process as $(X_1^*, \ldots, X_n^*)$ where $X_j^* = X_{n-j+1}$, so that the backwards coupled risk reserve process is given by definition as

$$R_{t_{k+1}^*}(u) = u - \sum_{j=1}^{k} X_j^*, \quad 0 \le k \le n, \tag{2.1}$$

with the time reversed epochs $t_k^* := t_n - t_{n-k+1}$ and $t_1 = 0$. This determines the entire risk reserve process $(R_t)_t$, because in between the arrival epochs t_k^* it drifts linearly upwards at constant rate c. We will investigate this coupling in more detail and derive some first consequences.

2.1 The embedded workload as a potential loss for the risk reserve process

The workload process was defined below (1.1) to be right-continuous with left limits. The dual reserve process $(R_t)_t$ is defined by time reversion starting from a fixed arrival epoch t_n of V_t, so that $(R_t)_t$ has cádlág paths. The following relation is a first consequence of the above coupling. It expresses the reserve level in terms of (differences of) the embedded workload process and the idle periods I_n, which are defined in (1.3).

Proposition 2.1.1. *For a fixed epoch t_n it holds that*

$$R_{t_{k+1}^*}(u) = u + \sum_{j=n-k+1}^{n} cI_{j+1} - \sum_{j=n-k+1}^{n} (V_{t_{j+1}} - V_{t_j}), \quad 0 \le k \le n, \tag{2.2}$$

Proof. The differences of V_{t_j} are related to the backwards accumulated idle period:

$$V_{t_{n+1}} - V_{t_{n-k+1}} = \sum_{j=n-k+1}^{n} X_j + \sum_{j=n-k+2}^{n+1} cI_j, \quad 1 \le k \le n. \tag{2.3}$$

To show this, notice that it follows at once from (1.1) and (1.3) that

$$V_{t_{j+1}} - V_{t_j} = X_j + cI_{j+1}, \quad j \ge 1.$$

Then (2.3) is obtained by summing these differences for $n-k+1 \le j \le n$. Replacing (2.3) in (2.1) via the coupling $X_j^* = X_{n-j+1}$ obtains (2.2). The proof is complete.

See Figure 2.1 for a coupled sample path of V_t and R_t with $n = 3$. For this trajectory, $I_2 = 0$ because $V_{t_2} > 0$.

Consider t_{k_0} to be the last arrival epoch before t_n such that $V_{t_{k_0}} = 0$ (think in terms of regeneration cycles for the workload). If nowhere sooner, then $k_0 = 1$ and with this choice, the backwards accumulated idle period appearing on the right-hand side of (2.2) is null, hence

$$R_{t_{k_0}^*}(u) = u - V_{t_{n+1}}. \tag{2.4}$$

From this follows immediately that for fixed u and a time horizon $t_n (= t_n^*)$,

$$\{V_{t_n} > u\} = \{\tau(u) \le t_n\}, \tag{2.5}$$

with $\tau(u)$ the time to ruin as defined in Section 1.4. In terms of probabilities, we have

$$\mathbb{P}(V_{t_n} > u \,|\, V_0 = 0) = \mathbb{P}(\exists \tau(u) \le t_n : R_{\tau(u)} < 0 \,|\, R_0 = u). \tag{2.6}$$

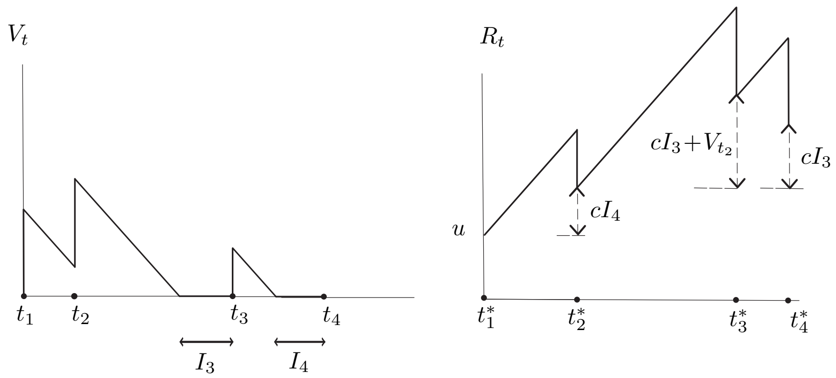


FIGURE 2.1: The workload process (left) and the dual risk reserve process (right).

The coupling presented above gives a direct interpretation of the embedded workload process in relation to the risk reserve process. Imagine that R_t gives the evolving price of a risky asset. There is just one investment made into this asset at time t_0 and this is equal to its initial value. Then the identity (2.4) interprets V_{t_n} as the amount lost through the evolution of the asset's value up to and including the nth epoch; that is, by taking $u = V_{t_n}$, we have

$$R_{t^*_{k_0}}(V_{t_n}) = 0 \quad \text{a.s.} \tag{2.7}$$

From epoch $t^*_{k_0}$ onwards, the asset's value evolves as starting from 0 but in any case, the initial capital has been lost by time t^*_n. Of course, everything is put in a stochastic perspective: V_{t_n} is to be seen as a sample from the workload distribution at epoch t_n, which is backwards coupled to a sample path of the risk reserve process $(R_t)_t$ up to the fixed time horizon t_n.

One of the performance measures traditionally associated to $(R_t)_{t\geq 0}$ is the so-called finite horizon *survival function*

$$F^s_n(u) := \mathbb{P}(\tau(u) > t_n) = \mathbb{P}(V_{t_n} \leq u | V_0 = 0).$$

The last identity follows from (2.6). In the insurance literature, the survival function is traditionally defined and handled as a function in the classical sense, with u as the argument. In light of the last term above, it is more convenient to regard this as a function in the generalized sense, more precisely as the (cumulative) distribution function of V_{t_n}, as it follows from (2.6). This function is non-decreasing and right-continuous, hence we can define Lebesgue-Stieltjes integrals w.r.t. it. If we consider the example of the investment in an asset, then directly from (2.7), the expected loss incurred by the asset up to time t^*_n is given by

$$\int_0^\infty u \, \mathrm{d}\mathbb{P}_0(V_{t_n} \leq u) = \mathbb{E}_0 V_{t_n},$$

where by $\mathbb{P}_0$ ($\mathbb{E}_0$) we denoted the probability law of $(V_t)_t$ (expectation operator) conditional on $V_0 = 0$. Similarly, the higher moments of the finite horizon loss distribution are equal to the respective moments of V_{t_n}.

The maximum aggregate loss M_n of the risk reserve process is simply defined as the running maximum of the partial sums $S^*_k = \sum_{i=1}^{k} X^*_i$. We have the following identity, still using the coupling:

$$V_{t_{n+1}} = S_n - \min(S_0, S_1, ..., S_n) = \max(S^*_0, S^*_1, ..., S^*_n) = M_n. \tag{2.8}$$

This identity is in line with (2.5) because it holds that $M_n > u \Leftrightarrow \tau(u) \leq t_n$.

Before we move on, it should be emphasized that the above duality is valid only for fixed, but arbitrary time horizons (the coupling depends on the time horizon). The paths of the maximum aggregate loss are non-decreasing, whereas this is not true for the paths of the embedded workload process, hence these two processes cannot coincide in law.

2.2 Multivariate duality

For random walks with vector-valued increments, it turns out that the ordered vector space structure of the state space is the essential ingredient for multivariate duality. However, in higher dimensions the orderings are *not total*. Those that make $\mathbb{R}^d$ into a partially ordered vector space are in one to one correspondence with the family of positive cones. These cones will appear in the following sections.

Several queues in parallel with simultaneous arrivals. The results of this section are valid for $\mathbb{R}^d$, but in order to keep formulae short we will work with $d = 2$.

There is a single arrival stream of customers that requests service from two servers that work in parallel, at rates c_i, and they are not interacting with each other.
Let the service request vector of customer n be $B_n := (B_n^{(1)}, B_n^{(2)})$. We will extend the notations from the previous subsections, so that the increments of the random walk S_n are $X_k := (B_k^{(1)} - c_1 A_k, B_k^{(2)} - c_2 A_k)$, with A_n still denoting the time between the nth and $(n+1)$th arrival epochs.

The key observation is that, for this queueing system, one can extend the previous results (2.5)–(2.8) if the canonical ordering of $\mathbb{R}^2$ is considered instead. For $x := (x^{(1)}, x^{(2)}), y := (y^{(1)}, y^{(2)}) \in \mathbb{R}^2$, we can abuse notation and still denote this ordering using $\leq$; then we have by definition

$$x \leq y \Leftrightarrow x^{(1)} \leq y^{(1)},\ x^{(2)} \leq y^{(2)}.$$

Also set

$$x \vee y := (\max(x^{(1)}, y^{(1)}),\ \max(x^{(2)}, y^{(2)})),$$

$$x \wedge y := (\min(x^{(1)}, y^{(1)}),\ \min(x^{(2)}, y^{(2)})).$$

Let $S_n := \sum_{k=1}^n X_k$, $n \geq 1$ and $S_0 = 0$, the origin of $\mathbb{R}^2$. Similarly as for the one dimensional single server queue, the workload process embedded at arrival epochs is the reflected version of the random walk given in terms of the Lindley recursion

$$V_{t_{n+1}} = (V_{t_n} + X_n) \vee 0, \tag{2.9}$$

and initial condition $V_{t_1} = 0$. This process evolves as a random walk which has the boundary of the non-negative orthant (which is also the positive cone that defines the canonical ordering) as an impenetrable barrier. The process $(V_t)_t$ is then defined as continuous from the right with left limits. Between arrival epochs the process drifts linearly along the direction vector $c := (c_1, c_2)$.

Proposition 2.2.1. *The sequence $(V_{t_n})_{n\geq 1}$ satisfies the following identity:*

$$V_{t_{n+1}} = S_n - \bigwedge_{k=0}^{n} S_k.$$

Roughly speaking, as soon as one of the components of S_n reaches a new minimum, the running infimum is updated accordingly and therefore the corresponding component of V_{t_n} is set to zero.

Proof. Proceed by induction. Assume the identity is valid for V_{t_n} (it trivially holds for $n = 1$). The proof follows by exploring all four possibilities, depending on the position of $V_{t_n} + X_n = S_n - \bigwedge_{i=0}^{n-1} S_i$ relative to the origin. For example, if $V_{t_n} + X_n$ is in the second quadrant, that is, if $S_n^{(1)} \leq \min_{i\leq n-1} S_i^{(1)}$ and $S_n^{(2)} \geq \min_{i\leq n-1} S_i^{(2)}$, then

$$S_n^{(1)} = \min_{i\leq n} S_i^{(1)}, \text{ and } S_n^{(2)} \geq \min_{i\leq n} S_i^{(2)}.$$

On the other hand, $S_n - \bigwedge_{i=0}^{n} S_i = (0,\, S_n^{(2)} - \min_{i\leq n} S_i^{(2)})$, and remark that this is the same as $(V_{t_n} + X_n) \vee 0$. The other cases follow by analogous considerations, which completes the proof.

Using the same coupling as in the one-dimensional case, $X_k^* = X_{n-k+1}$ for $1 \leq k \leq n$ and $S_n^* := \sum_{k=1}^n X_k^*$, consider the running maximum $M_n := \bigvee_{i=0}^n S_i^*$.

Lemma 2.2.1. *For all $n \geq 0$,*

$$S_n - \bigwedge_{i=0}^{n} S_i = M_n.$$

Proof. We can write

$$S_n - \bigwedge_{i=0}^{n} S_i = S_n + \bigvee_{i=0}^{n} (-S_i) = \bigvee_{i=0}^{n} (S_n - S_i) = \bigvee_{i=0}^{n} S_i^*.$$

Here we used the ordered vector space structure. The above are all sample-path identities because of the coupling. The proof is complete.

To this queueing system there corresponds an insurance risk model that consists of two insurance lines (or companies) which receive income at fixed rates. Claims are being simultaneously requested at random epochs from each marginal risk reserve.

The risk reserve process $R_t = (R_t^{(1)}, R_t^{(2)})$ evolves as

$$R_t = u + ct - \sum_{k=1}^{n(t)} B_k = u - \sum_{k=1}^{n(t)} X_k^*, \tag{2.10}$$

where $c = (c_1, c_2)$ and $u := (u^{(1)}, u^{(2)})$ is the initial capital, with $n(t)$ the number of arrivals before time t.

We will be concerned with measuring the event that both risk reserve processes survive indefinitely, i.e., we aim to determine the *survival function*

$$F^s(u^{(1)}, u^{(2)}) := \mathbb{P}(R_t^{(1)} \geq 0,\ \forall t > 0 \text{ and } R_t^{(2)} \geq 0,\ \forall t > 0 \mid R_0 = (u^{(1)}, u^{(2)})),$$

under some ergodicity conditions that ensure F^s is not null. In terms of times to ruin, F^s is related to the first time at least one of the two insurance lines is ruined,

$$\tau_\wedge(u^{(1)}, u^{(2)}) = \inf\{t;\ \min(R_t^{(1)}, R_t^{(2)}) < 0\} = \tau^{(1)}(u^{(1)}) \wedge \tau^{(2)}(u^{(2)}), \tag{2.11}$$

where $\tau^{(i)}(u^{(i)})$ are the marginal times to ruin, $i = 1, 2$. In particular,

$$F^s(u^{(1)}, u^{(2)}) = 1 - \mathbb{P}(\tau_\wedge(u^{(1)}, u^{(2)}) < \infty).$$

We can also define the first time at which both insurance lines are ruined:

$$\tau_\vee(u^{(1)}, u^{(2)}) = \tau^{(1)}(u^{(1)}) \vee \tau^{(2)}(u^{(2)}).$$

It is similarly related to the probability that *at least* one of the two branches survives indefinitely,

$$F^s_{OR}(u^{(1)}, u^{(2)}) := \mathbb{P}(R_t^{(1)} \geq 0,\ \forall t > 0 \text{ or } R_t^{(2)} \geq 0,\ \forall t > 0 \mid R_0 = (u^{(1)}, u^{(2)})),$$

by

$$F^s_{OR}(u^{(1)}, u^{(2)}) = 1 - \mathbb{P}(\tau_\vee(u^{(1)}, u^{(2)}) < \infty).$$

Notice that $\tau_\vee$ is not the same as $\inf\{t;\ \max(R_t^{(1)}, R_t^{(2)}) < 0\}$, that is, joint ruin does not have to happen simultaneously.

The above survival functions are then related by

$$F^s_{OR}(u^{(1)}, u^{(2)}) = F^s(u^{(1)}, \infty) + F^s(\infty, u^{(2)}) - F^s(u^{(1)}, u^{(2)}),$$

where $F^s(u^{(1)}, \infty)$ and $F^s(\infty, u^{(2)})$ are the marginal survival functions. Moreover, we also have $F^{i,j}(u^{(1)}, u^{(2)})$, the probability that component i survives indefinitely, while component j ruins, for $i, j = 1, 2$, $i \neq j$.

$$F^{1,2}(u^{(1)}, u^{(2)}) = F^s(u^{(1)}, \infty) - F^s(u^{(1)}, u^{(2)}),$$

and similarly for $F^{2,1}(u^{(1)}, u^{(2)})$.

In view of the above, it suffices to determine $F^s(u^{(1)}, u^{(2)})$ and the marginal survival functions in order to obtain all the other survival/ruin functions.

Ruin can only occur at arrival epochs, and since arrivals are simultaneous, we have the following relation for $\tau_\wedge$, the exit time defined in (2.11):

$$\{\tau_\wedge(u^{(1)}, u^{(2)}) > t_n\} = \{M_{n-1} \leq (u^{(1)}, u^{(2)})\}. \tag{2.12}$$

Notice also that $\tau_\wedge$ can now be rewritten in terms of the order relation '$\geq$':

$$\tau_\wedge(u^{(1)}, u^{(2)}) = \inf\{t_n \,; R_{t_n} \ngeq 0 \,|\, R_0 = (u^{(1)}, u^{(2)})\}.$$

We can regard the finite horizon survival function

$$F_n^s(u^{(1)}, u^{(2)}) := \mathbb{P}(\tau_\wedge(u^{(1)}, u^{(2)}) > t_n)$$

as the c.d.f. of a *survival measure.* Relation (2.12), Lemma 2.2.1 and Proposition 2.2.1 imply that this survival measure is nothing else but the distribution of the *reflected random walk* V_{t_n} inside the non-negative quadrant of $\mathbb{R}^2$.

Theorem 2.2.1 (Duality)**.** *The following identity relates the finite horizon survival functions of the risk reserve process to the distribution of the embedded workload process in the associated parallel queueing system:*

$$\mathbb{P}(R_{t_i} \geq 0,\, i = 1, ..., n \,|\, R_0 = (u^{(1)}, u^{(2)})) = \mathbb{P}(V_{t_n} \leq (u^{(1)}, u^{(2)}) \,|\, V_0 = 0). \tag{2.13}$$

Proof. That V_{t_n} is the reflected version of the random walk follows directly from the fact that it is the solution of the recursive equation in Proposition 2.2.1. In view of Lemma 2.2.1 and (2.12), the duality relation (2.13) is also obvious, so this concludes the proof.

2.3 Siegmund duality for coupled processors models

The purpose of this section is to show that the workload process embedded at arrival epochs which appears in the study of two interacting queues can be represented as a special kind of reflected process and this can further be related to a class of random walks which have certain absorbing sets in the plane. The geometry of these sets is tied to the special way in which the reflection works for the buffer content of the queueing system. This relation -which is a direct extension of the Duality Theorem 2.2.1- is the topic of Theorem 2.3.2, which is also the main result of this subsection. The dual absorbing processes are killed upon exiting domains which contain the non-negative quadrant as a subset, so they are allowed to have a negative component before the exit time (it is not clear if this has a ruin interpretation, see Theorem 2.3.2).

As introduced in Section 1.2, the coupled processor model is a queueing system consisting of two parallel processors that are switching to process work from the other buffer instead of entering an idle state (this happens if the other buffer is not empty already).

There are operators $\wedge_\alpha^\beta$ defined in (2.15) that are acting on the (unrestricted) inventory level and they give a representation for the embedded workload process (Theorem 2.3.1). It will be shown that one can associate two kinds of absorbing processes using the related operators defined in (2.15-2.16) and then the embedded workload distribution can be squeezed in between the survival probabilities of these dual processes. As in the previous subsection, the ordered vector space structure is key for duality.

The Lindley recursion for the amount of work at the arrival epoch t_{n+1} in a system with coupled processors reads as follows:

$$\begin{aligned} V_{t_{n+1}}^{(1)} &= \left[V_{t_n}^{(1)} + B_n^{(1)} - c_1 A_n + \frac{c_2^*}{c_2}(V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \wedge 0 \right] \vee 0, \\ V_{t_{n+1}}^{(2)} &= \left[V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n + \frac{c_1^*}{c_1}(V_{t_n}^{(1)} + B_n^{(1)} - c_1 A_n) \wedge 0 \right] \vee 0. \end{aligned} \tag{2.14}$$

Here c_i is server i speed and c_i^* is the working speed when processing from the other buffer. The system is initially empty. An interpretation for the above recursion is that

$$\frac{1}{c_2}(V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \wedge 0$$

is the length of the idle period of the second component (if non-zero and assuming non-coupled processors), so when multiplied with c_2^*, it becomes the total capacity that server 2 can process from buffer 1 before it receives an arrival of its own.

There is a lot of geometry behind the dynamics of this queueing system. Define the reflection angles $\alpha := \arctan c_1^*/c_1$, $\beta := \arctan c_2^*/c_2$, and let the (non-negative) cones $\mathcal{C}_\alpha$ and $\mathcal{C}_\beta$ be as in Figure 2.2. By a cone is meant any subset of the vector space $\mathbb{R}^2$ which is closed under linear combinations with non-negative scalars. We will take these cones to be closed in the usual topology of $\mathbb{R}^2$.

Each cone $\mathcal{C}_\alpha$ and $\mathcal{C}_\beta$ defines an order relation well behaved with the linear structure of $\mathbb{R}^2$ by setting

$$x \geq_\alpha y \Leftrightarrow x - y \in \mathcal{C}_\alpha, \quad x \geq_\beta y \Leftrightarrow x - y \in \mathcal{C}_\beta.$$

Denote the suprema of two vectors $x \vee_\alpha y$ ($x \vee^\beta y$) to be the least upper bound of x and y w.r.t. the above order relations. Geometrically, $x \vee_\alpha y$ is obtained by taking the intersection of the shifted positive cones $(x + \mathcal{C}_\alpha) \cap (y + \mathcal{C}_\alpha)$. This is still a cone and its vertex lies precisely at $x \vee_\alpha y$. The same construction gives $x \vee^\beta y$, when using the cone $\mathcal{C}_\beta$. Similarly, $x \wedge_\alpha y$ ($x \wedge^\beta y$) is defined using the negative cone $-\mathcal{C}_\alpha$ ($-\mathcal{C}_\beta$) shifted at x and y.

For the duality we need the following operator $\wedge_\alpha^\beta : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$,

$$x \wedge_\alpha^\beta y := \begin{cases} x \wedge_\alpha y, & \text{if } x_1 \geq y_1,\ x_2 \leq y_2 \\ x \wedge^\beta y, & \text{if } x_1 \leq y_1,\ x_2 \geq y_2 \\ x \wedge_\alpha y = x \wedge^\beta y, & \text{if } x_i \leq y_i, \text{ or } x_i \geq y_i,\ i = 1, 2. \end{cases} \tag{2.15}$$

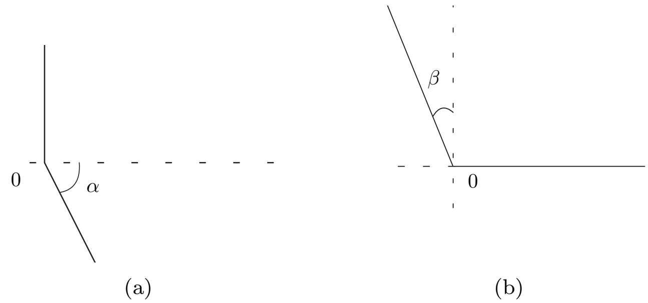


FIGURE 2.2: The cones $\mathcal{C}_\alpha$ and $\mathcal{C}_\beta$.

Define also the operator $\vee_\alpha^\beta : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$,

$$x \vee_\alpha^\beta y := \begin{cases} x \vee_\alpha y, & \text{if } x_1 \geq y_1,\ x_2 \leq y_2 \\ x \vee^\beta y, & \text{if } x_1 \leq y_1,\ x_2 \geq y_2 \\ x \vee_\alpha y = x \vee^\beta y, & \text{if } x_i \leq y_i, \text{ or } x_i \geq y_i,\ i = 1, 2. \end{cases} \tag{2.16}$$

$\wedge_\alpha^\beta$ will be used to define the compensator for the workload process (a running 'infimum' of the input process, see Figure 2.3) below, while $\vee_\beta^\alpha$ will appear in the running maximum associated to the workload process in Theorem 2.3.1 (Figure 2.4 below); which is the first step towards the duality Theorem 2.3.2. At this point, let us remark that these operators are not commutative (they are asymmetric). They are however related one to the other by

$$-\left(x \wedge_\alpha^\beta y\right) = (-y) \vee_\alpha^\beta (-x), \tag{2.17}$$

and the order in which the vectors appear matters. This is a key relation that follows at once because of the identity $-(x \wedge_\alpha y) = (-x) \vee_\alpha (-y)$, together with the similar one for $\wedge^\beta$. Besides the above relation, we have the following properties.

Lemma 2.3.1. *The following compatibility relations between the operator $\wedge_\alpha^\beta$ and the linear structure on $\mathbb{R}^2$ hold*

i) $z + x \wedge_\alpha^\beta y = (x + z) \wedge_\alpha^\beta (y + z), \quad z \in \mathbb{R}^2,$

ii) $\omega(x \wedge_\alpha^\beta y) = (\omega x) \wedge_\alpha^\beta (\omega y), \quad \omega \in \mathbb{R}_+.$

The same is valid for $\wedge_\alpha^\beta$ replaced by $\vee_\alpha^\beta$.

Proof. Notice that the regions of the plane used to define the operators $\wedge_\alpha^\beta$, $\vee_\alpha^\beta$ in (2.15), (2.16) are preserved by vector addition and positive scalar multiplication, hence the proof follows from the fact that the lattice operations $\wedge_\alpha(\vee_\alpha)$, $\wedge^\beta(\vee^\beta)$ are generated by cones.

Before we give the next result, let us make some conventions related to the lack of commutativity: when taking the successive infima of the process $(S_n)_n$, we will always "minimize" to the right, that is by convention set

$$x_1 \wedge_\alpha^\beta x_2 \wedge_\alpha^\beta x_3 := (x_1 \wedge_\alpha^\beta x_2) \wedge_\alpha^\beta x_3,$$

and we will "maximize" to the left:

$$x_1 \vee_\alpha^\beta x_2 \vee_\alpha^\beta x_3 := x_1 \vee_\alpha^\beta (x_2 \vee_\alpha^\beta x_3).$$

These are successively defined for n vectors by iteration, and these conventions are consistent with (2.17). This convention extends to n vectors when taking their successive infima and suprema, as in (2.18) below.

As usual, let $S_n = \sum_{k=1}^n X_k$, $X_k^{(i)} := B_k^{(i)} - c_i A_k$, $i = 1, 2$, together with the backwards coupled walk S_k^*. The reason for introducing the above operators is the following

Theorem 2.3.1. *The solution to the Lindley recursion (2.14) can be represented as*

$$V_{t_{n+1}} = S_n - \bigwedge_{i=0}^{n}{}_\alpha^\beta\, S_i.$$

Moreover, the compatibility between order and vector space structure implies the following relation:

$$S_n - \bigwedge_{i=0}^{n}{}_\alpha^\beta\, S_i \;=\; \bigvee_{i=0}^{n}{}_\alpha^\beta\, S_i^*. \tag{2.18}$$

Proof. The first part of the proof is similar to the proof of Proposition 2.2.1. Let us denote the quadrants of the plane with $qI - qIV$. The positive quadrant is qI and the rest are defined based on the trigonometric order. Proceeding by induction, (2.3.1) holds trivially for V_{t_1}. Assume the identity to be valid for V_{t_n}. If we denote by

$$C_{n-1} := \bigwedge_{i=0}^{n-1}{}_\alpha^\beta\, S_i,$$

then the following cases will be considered, depending on the position of S_n relative to C_{n-1}.

Case I: $S_n - C_{n-1} \in qI$, and then, by definition, $\wedge_\alpha^\beta \equiv \wedge_\alpha \equiv \wedge^\beta$, so that $C_n = C_{n-1} \wedge_\alpha^\beta S_n = C_{n-1}$. Using the induction hypothesis, we can write $S_n - C_n = V_{t_n} + X_n$, this vector belonging to the first quadrant, so that it is equal to $V_{t_{n+1}}$ by (2.14).

Case II: $S_n - C_{n-1} \in qII$. Then $C_{n-1} \wedge_\alpha^\beta S_n = C_{n-1} \wedge_\alpha S_n$. There are two subcases to be considered:

Firstly, if $S_n - C_{n-1} \in qII \cap (-\mathcal{C}_\alpha)$, then $S_n \leq_\alpha C_{n-1} \Rightarrow C_{n-1} \wedge_\alpha S_n = S_n$, so that $S_n - C_n = 0$ (see Figure 2.3(a)). On the other hand, by the induction hypothesis, $S_n - C_{n-1} = V_{t_n} + X_n$ and the case we are in dictates

$$\begin{cases} V_{t_n}^{(1)} + X_n^{(1)} \leq 0, \\ V_{t_n}^{(2)} + X_n^{(2)} + \tan\alpha \left[V_{t_n}^{(1)} + X_n^{(1)} \right] \leq 0, \end{cases}$$

which means that $V_{t_{n+1}} = S_n - C_n = 0$, from (2.14). In words, at epoch t_{n+1} server one has been idle long enough to drain the inventory level that buffer two would have had.

Secondly, if $S_n - C_{n-1} \in qII\backslash(-\mathcal{C}_\alpha)$, then

$$S_n^{(1)} - C_n^{(1)} = 0, \qquad S_n^{(2)} - C_n^{(2)} = S_n^{(2)} - C_{n-1}^{(2)} + \tan\alpha\left[S_n^{(1)} - C_{n-1}^{(1)}\right]. \qquad (2.19)$$

The geometric argument for these is in Figure 2.3(b). In the triangle $\triangle S_n PQ$, the length of $\overrightarrow{PQ}$ is equal to $\tan\alpha\,|\overrightarrow{PS_n}| = \tan\alpha\,[S_n^{(1)} - C_{n-1}^{(1)}]$. The origin is at C_{n-1} and the square is placed on top of C_n. Thus we have the identity $S_n^{(2)} - C_n^{(2)} = |\overrightarrow{PC_{n-1}}| - |\overrightarrow{PQ}|$. On the other hand, by hypothesis, $S_{n-1} - C_{n-1} = V_{t_n}$, so the second identity in (2.19) becomes

$$S_n^{(2)} - C_n^{(2)} = V_{t_n}^{(2)} + X_n^{(2)} + \tan\alpha\left[V_{t_n}^{(1)} + X_n^{(1)}\right] \geq 0.$$

Thus (2.19) is identical with the right-hand side of the recursion (2.14), which means that also in this case $V_{t_{n+1}} = S_n - C_n$.

The case $S_n - C_{n-1} \in qIII$ is similar to the first case, whereas the proof for the case $S_n - C_{n-1} \in qIV$ is analogous to that for Case II, with the difference that $\wedge_\alpha^\beta$ becomes $\wedge^\beta$. This completes the argument for the first part of the theorem.

Using Lemma 2.3.1 together with relation (2.17), gives

$$S_n - \bigwedge_{i=0}^{n}{}_{\alpha}^{\beta} S_i = S_n + \bigvee_{i=n}^{0}{}_{\alpha}^{\beta}(-S_i) = \bigvee_{i=n}^{0}{}_{\alpha}^{\beta}(S_n - S_i) = \bigvee_{i=0}^{n}{}_{\alpha}^{\beta} S_i^*.$$

In the intermediate terms, the supremum runs from $i = n$ down to $i = 0$, because, as noted above (2.17), these operators do not commute, so the order in which they are applied makes a difference (see also the conventions made before the theorem). The proof is now complete.

The right-hand side of (2.18) defines a "maximum" process, call it $(\tilde{M}_n)_n$ (here maximum should be taken in a loose sense because $\vee_\alpha^\beta$ is not defined by an order relation). As in the classical set-up, we will next associate to the process $(\tilde{M}_n)_n$ a hitting time that will describe the version of the random walk which has an absorbing barrier. However, in order to have a Siegmund type of duality for the process V_t, we need to impose a condition on the reflection angles, see Remark 2.3.1 below.

Denote for brevity $\mathcal{C}_\alpha(x)$, $\mathcal{C}_\beta(x)$ the shifted cones at x. Set also $\mathcal{C}_{\alpha,\beta}(x) := \mathcal{C}_\alpha(x) \cup \mathcal{C}_\beta(x)$, and set $\mathcal{C}_{\alpha,\beta} := \mathcal{C}_{\alpha,\beta}(0)$, so that we have

$$y \geq_\alpha x \text{ or } y \geq_\beta x \quad \Leftrightarrow \quad y \in C_{\alpha,\beta}(x).$$

Let us call the left-hand side of (2.18) the $\alpha\beta$-reflected version of the random walk $(S_n)_n$.

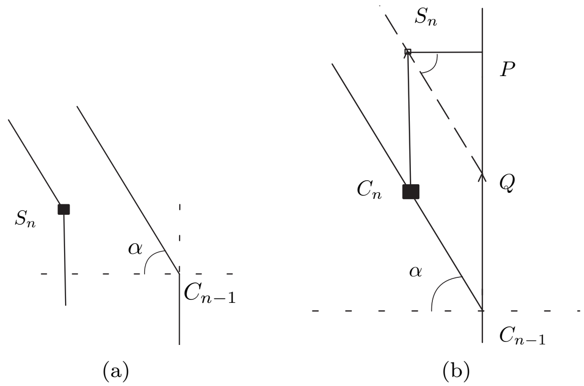


FIGURE 2.3: Positions of S_n relative to C_{n-1} and the construction of infima.

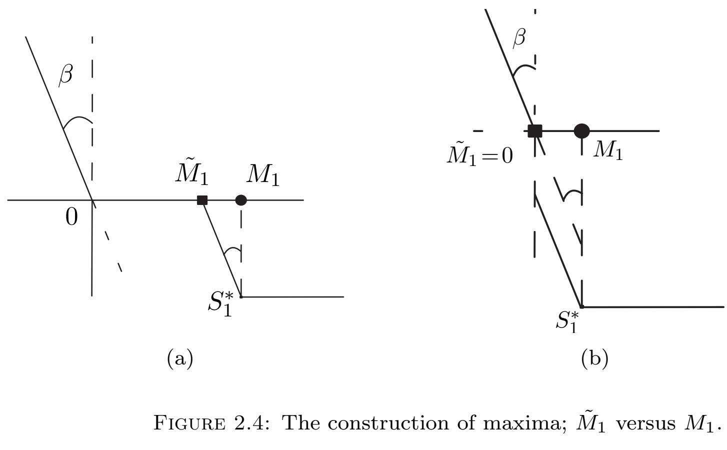


FIGURE 2.4: The construction of maxima; $\tilde{M}_1$ versus M_1.

Remark 2.3.1. *The possibility to have an identity between the law of the $\alpha\beta$-reflected process $(V_{t_n})_n$ and a process with an absorbing domain relies on the validity of the following equivalence (see also (2.18))*

$$x \vee_{\alpha}^{\beta} y \in -\mathcal{C}_{\alpha,\beta} \Leftrightarrow x,\, y \in -\mathcal{C}_{\alpha,\beta},$$

for arbitrary vectors $x, y \in \mathbb{R}^2$. In words, knowing that the running $\vee_{\alpha,\beta}$-supremum belongs to some subset of $\mathbb{R}^2$, we must be able to infer on the position of each component that appears in the supremum. In contrast to the set-up of Section 2.2, where we can find such an equivalence for the negative cones, see for example (2.12), in the present set-up this is not always possible anymore. It is easy to see that this equivalence holds only when $\alpha + \beta = \pi/2$ ($\alpha > 0$ or $\beta > 0$, otherwise, if $\alpha = \beta = 0$ the equivalence holds, but we are back in the instance from Subsection 2.2).
In more generality, let $\overline{\mathcal{C}^c}_{\alpha,\beta}(u)$ be the topological closure of the complementary of $\mathcal{C}_{\alpha,\beta}(u)$.

If $\alpha + \beta \le \pi/2$, then $\overline{\mathcal{C}^c}_{\alpha,\beta}(u) \supseteq -\mathcal{C}_{\alpha,\beta}(u)$, and the following chain of implications is valid

$$x \vee_{\alpha}^{\beta} y \in -\mathcal{C}_{\alpha,\beta}(u) \Rightarrow x,\, y \in -\mathcal{C}_{\alpha,\beta}(u) \Rightarrow x \vee_{\alpha}^{\beta} y \in \overline{\mathcal{C}^c}_{\alpha,\beta}(u) \Rightarrow x,\, y \in \overline{\mathcal{C}^c}_{\alpha,\beta}(u).$$

If $\alpha + \beta \ge \pi/2$, then $\overline{\mathcal{C}^c}_{\alpha,\beta}(u) \subseteq -\mathcal{C}_{\alpha,\beta}(u)$ and the converse chain holds. These can be argued directly from definition (2.16), by considering all the cases, based on the position of y relative to x, just like in the proof of Theorem 2.3.1.

Consider $(\overline{S}^a_n)_n$, the version of the random walk $u - S^*_n =: R_{t^*_{n+1}}$, killed upon exiting $-\overline{\mathcal{C}^c}_{\alpha,\beta}$, and $(\underline{S}^a_n)_n$, the version killed upon exiting $\mathcal{C}_{\alpha,\beta}$.

Theorem 2.3.2 (Siegmund Duality)**.** *There exists a duality relation between the $\alpha\beta$-reflected version $(V_{t_n})_n$ of $(S_n)_n$, conditional on $V_{t_1} = x$ and the laws of the absorbed random walks $(\underline{S}^a_n)_n$ and $(\overline{S}^a_n)_n$. More precisely, under the condition $\alpha + \beta \le \pi/2$, it holds for $x \ge 0, n \ge 1$ that*

$$\begin{aligned} &\mathbb{P}(V_{t_n} \in -\mathcal{C}_{\alpha,\beta}(u) |\, V_{t_1} = x) \le \mathbb{P}(\underline{S}^a_n \in \mathcal{C}_{\alpha,\beta}(x) |\, S^a_0 = u) \le \\ &\mathbb{P}(V_{t_n} \in \overline{\mathcal{C}^c}_{\alpha,\beta}(u) |\, V_{t_1} = x) \le \mathbb{P}(\overline{S}^a_n \in -\overline{\mathcal{C}^c}_{\alpha,\beta}(x) |\, \overline{S}^a_1 = u),\ u \in \mathcal{C}_{\alpha,\beta}, \end{aligned} \tag{2.20}$$

and if $\alpha + \beta \ge \pi/2$, the chain of inequalities is reversed and valid whenever $u \in -\overline{\mathcal{C}^c}_{\alpha,\beta}$ (see Remark 2.3.1).

In particular, if $\alpha + \beta = \pi/2$, the $\alpha\beta$-reflected version of the random walk has the same law as the version $(\underline{S}^a_n)_n \equiv (\overline{S}^a_n)_n$, absorbed upon exiting $\mathcal{C}_{\alpha,\beta} \equiv -\overline{\mathcal{C}^c}_{\alpha,\beta}$, (cf. (2.13)):

$$\mathbb{P}(V_{t_n} \in -\mathcal{C}_{\alpha,\beta}(u) |\, V_{t_1} = x) = \mathbb{P}(\underline{S}^a_n \in \mathcal{C}_{\alpha,\beta}(x) |\, \underline{S}^a_0 = u),\ u \in \mathcal{C}_{\alpha,\beta}. \tag{2.21}$$

Notice that the event on the right-hand side of (2.21) implies $\underline{S}^a_n$ has not been killed up to epoch t^*_n, that is we have

$$\mathbb{P}(\underline{S}_n^a \in \mathcal{C}_{\alpha,\beta}(x) |\, \underline{S}_0^a = u) = \mathbb{P}(\underline{S}_k^a \in \mathcal{C}_{\alpha,\beta},\ k \le n-1,\ \underline{S}_n^a \in \mathcal{C}_{\alpha,\beta}(x) |\, \underline{S}_0^a = u). \quad (2.22)$$

Proof of Theorem 2.3.2. The proof is a variation of Theorem 2.3.1. The extra feature is that the solution of the Lindley recursion was given under the condition $V_{t_1} = 0$, and now we adapt it to the case $V_{t_1} = x$. Making use of the operators introduced, we begin with the remark that recursion (2.14) can be represented as (mind the order in which the supremum is taken)

$$V_{t_{n+1}} = 0 \vee_\alpha^\beta (V_{t_n} + X_{n+1}).$$

Iterating this recursion via Lemma 2.3.1 with initial condition $V_{t_1} = x$, gives along the same lines as in the proof of Theorem 2.3.1:

$$V_{t_{n+1}} = 0 \vee_\alpha^\beta S_1^* \vee_\alpha^\beta S_2^* \vee_\alpha^\beta \dots \vee_\alpha^\beta (S_n^* + x), \quad (2.23)$$

where $V_{t_1} = x$ appears in the rightmost term only.

Under the condition $\alpha + \beta \le \pi/2$, the chain of inequalities from Remark 2.3.1 is in force, and thus we can write using (2.23), for an arbitrary $u \in \mathcal{C}_{\alpha,\beta}$:

$$V_{t_{n+1}} \in -\mathcal{C}_{\alpha,\beta}(u) \Rightarrow S_k^* \in -\mathcal{C}_{\alpha,\beta}(u),\ k \le n-1,\ S_n^* + x \in -\mathcal{C}_{\alpha,\beta}(u),$$

which is the same as $\underline{S}_n^a \in \mathcal{C}_{\alpha,\beta}(x)$, via (2.22). Finally, (2.20) follows similarly from the rest of the chain of inequalities from Remark 2.3.1.

The case $\alpha + \beta \ge \pi/2$ follows from the second part of Remark 2.3.1, because in this case we have the converse implications. This completes the proof.

Concluding remarks

Ergodicity. Focus first on the case $\alpha + \beta \le \pi/2$. For the embedded random walk $(u - S_n^*)_n$ starting at $u \in \mathcal{C}_{\alpha,\beta}$, define its first exit time from $\mathcal{C}_{\alpha,\beta}$:

$$\tau_{\alpha,\beta}(u) = \min\{t_n^*;\ S_n^* \not\succ_\alpha u, \text{ and } S_n^* \not\succ_\beta u\}.$$

The second inequality from (2.20) becomes via (2.22), for $x = 0$, $u \in \mathcal{C}_{\alpha,\beta}$:

$$\mathbb{P}(V_{t_n} \in \overline{\mathcal{C}^c}_{\alpha,\beta}(u) |\, V_{t_1} = 0) \ge \mathbb{P}(\underline{S}_k^a \in \mathcal{C}_{\alpha,\beta},\ k \le n |\, \underline{S}_0^a = u) = \mathbb{P}(\tau_{\alpha,\beta}(u) > t_n^*), \quad (2.24)$$

(cf. (2.13)). Since $(V_{t_n})_n$ is a regenerative process with 0 as regeneration point, V_{t_n} converges weakly as $n \to \infty$ to a proper probability distribution V_{t_∞}, whenever it has 0 as a positive recurrent state, and then by (2.24), the hitting time $\tau_{\alpha,\beta}(0)$ must have positive probability of never occurring, i.e., (2.24) with $u = 0$ must read in the limit

$$\mathbb{P}(V_{t_\infty} = 0) \ge \mathbb{P}(\tau_{\alpha,\beta}(0) = \infty) > 0.$$

Thus, a sufficient condition that ensures ergodicity for $(V_{t_n})_n$ is that the mean increment $\mathbb{E}(-X_1)$ of $\underline{S}^a_n$ must lie in the topological interior $\mathcal{C}^\circ_{\alpha,\beta}$. In words, $(\underline{S}^a_n)_n$ must drift away from its absorbing domain.

That the necessary condition is $\mathbb{E}(-X_1) \in \mathcal{C}_{\alpha,\beta}$, follows along the same lines by using the first inequality from (2.20) (which gives the upper bound on the distribution function of V_{t_∞}).

For the case $\alpha + \beta \geq \pi/2$, the discussion is analogous. One has to redefine $\tau_{\alpha,\beta}$ to be the exit time of $\overline{S}^a_n$ from the (closed, convex) set $-\overline{\mathcal{C}^c}_{\alpha,\beta}$. The chain of inequalities in (2.20) is reversed so one should use the last inequality that reads in the current case:

$$\mathbb{P}(V_{t_n} \in \overline{\mathcal{C}^c}_{\alpha,\beta}(u) | \, V_{t_1} = x) \geq \mathbb{P}(\overline{S}^a_n \in -\overline{\mathcal{C}^c}_{\alpha,\beta}(x) | \, \overline{S}^a_1 = u),$$

for $u \in \overline{\mathcal{C}^c}_{\alpha,\beta}$. The same arguments as in the previous case give that the mean drift $\mathbb{E}(-X_1)$ of $\overline{S}^a_n$ must belong to $-\mathcal{C}^c_{\alpha,\beta}$, and in the current case, this is sufficient to ensure the queueing system is stable. Remark that in both cases, the domain inside which the random walk is allowed to evolve before absorbtion is closed and convex.

In Cohen and Boxma [41], the study of the amount of work in the coupled processor model is reduced to a functional equation in terms of Laplace-Stieltjes transforms, which is then further related to a boundary value problem of Riemann type. They assume independent Poisson arrivals with intensities λ_i and independent service requirements between the queues. In our case, the arrival streams are merged with common arrival intensity, say, λ ($\lambda = \lambda_1 + \lambda_2$, hence service requirements are correlated among the queues, and since a type 1 arrival will demand no work from queue 2, the traffic loads remain the same). Synchronizing notations - the model description is different, but equivalent, see [41, p. 288, pp. 294-296] - they have

$$\rho_1 := 1 + \frac{c_2^*}{c_1} \geq 1, \quad \rho_2 := 1 + \frac{c_1^*}{c_2} \geq 1,$$

and further, the natural server rates are normalized to $c_2 = c_1 = 1$, so that the reflection angles in our notation become $\tan\alpha = \rho_2 - 1$, $\tan\beta = \rho_1 - 1$ (the ρ_i's in [41] have a different meaning than the usual load coefficients). It is easy to see that the ergodicity condition extracted in [41] from the study of the functional equation reads in our notation

$$b_1 := \lambda \mathbb{E}B^{(1)} \left(1 - \frac{1}{1 + c_1^*}\right) + \frac{\lambda \mathbb{E}B^{(2)}}{1 + c_1^*} < 1, \quad b_2 := \frac{\lambda \mathbb{E}B^{(1)}}{1 + c_2^*} + \lambda \mathbb{E}B^{(2)} \left(1 - \frac{1}{1 + c_2^*}\right) < 1,$$

and is equivalent to the conditions

$$\mathbb{E}X_1^{(2)} + \tan\alpha \, \mathbb{E}X_1^{(1)} < 0, \quad \mathbb{E}X_1^{(1)} + \tan\beta \, \mathbb{E}X_1^{(2)} < 0, \tag{2.25}$$

which in case $\alpha + \beta \leq \pi/2$ is the same as $\mathbb{E}(-X_1) \in \mathcal{C}^\circ_{\alpha,\beta}$, and it is equivalent to $\mathbb{E}(-X_1) \in \mathcal{C}^c_{\alpha,\beta}$ for $\alpha + \beta \geq \pi/2$. In any case, (2.25) means the mean drift vector must point into the interior of the intersection of the lower hyperplanes at 0 with direction vectors $(\tan\alpha, 1)$ and respectively $(1, \tan\beta)$.

In conclusion, the stability condition for the coupled processor model studied in Cohen and Boxma [41] is the same as the ergodicity condition for the dual absorbed process S_n^a, as predicted by the duality relation (2.20). The result of Theorem 2.3.2 is valid under very general assumptions on the arrival process; for example, a semi-Markov structure for the arrival process preserves this ergodicity condition.

Large deviations. The possibility to infer on the positions of the walks $\underline{S}_n^a$ and $\overline{S}_n^a$ based on the position of the all-time maximum $\tilde{M}_\infty$ is the usual method for analyzing overflow probabilities, either when increments are light-tailed or heavy-tailed. Assume for simplicity $\alpha+\beta=\pi/2$, then under the stability condition (2.25), (2.20) becomes in the limit

$$\mathbb{P}(V_{t_\infty} \notin -\mathcal{C}_{\alpha,\beta}(u)) = \mathbb{P}(\exists n_0;\ \underline{S}_{n_0}^a \notin \mathcal{C}_{\alpha,\beta} \,|\, \underline{S}_0^a = u),$$

and the right-hand side is typically the starting point for the study of large deviations.

Simulation. One of the reasons why Siegmund [97] studied such duality relations is that estimating the hitting probabilities for the absorbed process is more efficient numerically than the problem of estimating the equilibrium distribution for the dual reflecting process (this is the usual Markov-Chain-Monte-Carlo scheme, the efficiency of which depends severely on the mixing properties of the chain). For $\alpha+\beta=\pi/2$, Theorem 2.3.2 gives a numerical scheme for simulating exactly the equilibrium distribution of the embedded workload in the system, and it provides bounds otherwise.

Chapter 3

Single server queues and Sparre-Andersen risk reserve processes with correlations

This chapter deals with the study of performance measures related to single server queues and their dual Sparre-Andersen risk reserve processes which have a form of dependence between inter-arrival times and service requirements/claim sizes. These models were introduced and described in Section 1.1. The *main contributions* of the current chapter are the following: (i) We provide an exact analysis of the waiting time distribution in a GI/G/1 queue with correlation between a service requirement B and the subsequent inter-arrival time A, B and A having a multivariate matrix-exponential distribution. Via (2.6), the distribution of the waiting time is then immediately related to the ruin probability of the dual risk reserve process. (ii) We prove that the simple relation which holds between steady-state workload and waiting time distributions in the ordinary GI/GI/1 queue remains valid in the case of correlated B and A. (iii) We consider the dual Sparre-Andersen insurance risk model with correlation between inter-claim time and subsequent claim size, and in particular we show that the Takács relation (cf. [63], Corollary 4.5.4) between the ordinary ruin probability and the delayed ruin probability remains valid; and this is analogous to the relation between the steady-state waiting time and the workload. (iv) Finally, we show that, in comparison with the classical set-up with mutually independent sequences $(A_k)_{k\geq 1}$ and $(B_k)_{k\geq 1}$, positive and negative correlation respectively decreases and increases the waiting times in the sense of convex ordering. We also illustrate with numerical results the influence of dependence on the expected values of the waiting times but also on the 95%-percentiles of the ruin functions (values at risk) - the quantiles of the ruin function correspond to the values of the waiting time tail, again via (2.6).

The chapter is organized as follows. Section 3.1 contains a detailed model description, which in particular includes a description of the class of bivariate distributions under consideration. It also presents the waiting time analysis. The relation between the steady-state waiting time and workload distributions is exposed in Section 3.2.

Section 3.3 is devoted to the dual insurance risk model. In Section 3.4 we consider several examples of bivariate distributions for (A_k, B_k). For these examples, we present numerical results on the mean and tail of the waiting time distribution (and, by duality, on the ruin probability), which exhibit the effect of (positive or negative) correlation on waiting time and ruin probability, together with stochastic ordering results and by consequence, ordering between the waiting times.

3.1 Model description and analysis of waiting times

We study a generalization of the classical GI/GI/1 model, where we allow for an arbitrary correlation between the service requirement of the nth customer and the inter-arrival time between the nth and $(n+1)$th customer. As a key performance measure in this model, we first consider the waiting time process in an initially empty system. It follows from Section 1.1 that the waiting time W_n is the same as V_{t_n}, the embedded workload at epoch t_n, up to the factor c^{-1}. In Section 3.2, we show a relation between the steady-state waiting time (embedded workload) and the steady-state workload, which is similar to the independent case (Cohen [39]).

Let B_i be the service requirement of the ith customer, A_i the inter-arrival time between the ith and the $(i+1)$th customer, and c the server's speed. We assume that (A_i, B_i) are i.i.d. sequences of random vectors. This implies that the arrival process of customers is renewal and that the quantities $(B_i - cA_i)$ are i.i.d. However, within a pair, A_i and B_i are dependent, hence the ith service requirement and the subsequent inter-arrival time are correlated. We denote by (A, B) a generic pair made up of a service requirement and the subsequent inter-arrival time. In Figure 1.1, the workload process $\{V_t, t \geq 0\}$ and the waiting time process $\{W_n, n = 1, 2, \dots\}$ are displayed; here V_t denotes the work in the system at time t, and W_n denotes the waiting time of the nth arriving customer. These are defined in Section 1.1. It follows from the relation $W_{n+1} = c^{-1}V_{t_{n+1}}$ and Recursion (1.1) that the waiting time process satisfies the Lindley recursion:

$$W_{n+1} = \max(W_n + c^{-1}B_n - A_n, 0).$$

We assume that both B and A have finite first moments, and then under the stability condition $\mathbb{E}(c^{-1}B - A) < 0$, W_n converges in distribution to a proper random variable W and we can write:

$$W \stackrel{d}{=} \max\left(W + c^{-1}B - A, 0\right). \tag{3.1}$$

The dependence structure: We model the dependence structure using the class of multivariate matrix-exponential distributions (MVME), which was introduced by Bladt and Nielsen [29]. This class contains other known classes of distributions with interesting probabilistic interpretations, like the multivariate phase-type distributions studied in Assaf et al. [18] and further in Kulkarni [79]. We will further discuss this class in Section 3.4 where we also give examples which admit a probabilistic interpretation. Below we cite Definition 4.1 of Bladt and Nielsen [29]:

Definition 3.1.1. *A non-negative random vector (A, B) is said to have a bivariate matrix-exponential distribution if the joint Laplace-Stieltjes transform (LST) $\mathbb{E}e^{-s_1 A - s_2 B}$ is a rational function in (s_1, s_2), i.e. it can be written as $\frac{G(s_1,s_2)}{D(s_1,s_2)}$, where $G(s_1, s_2)$ and $D(s_1, s_2)$ are polynomial functions in s_1 and s_2.*

As a consequence of this defining property, the transform of the difference $X = c^{-1}B - A$ is also a rational function. The distribution of X is called a bilateral matrix exponential (see Bladt et al. [31], Thm. 3.1). For simplicity, let us denote $\mathbb{E}e^{-sX} := \frac{f(s)}{g(s)}$. We rewrite identity (3.1) in terms of Laplace-Stieltjes transforms. After some straightforward computations, one obtains:

$$\mathbb{E}e^{-sW}\left[1 - \mathbb{E}e^{-sX}\right] = \mathbb{P}\left(W + X \leq 0\right) - \mathbb{E}e^{-s(W+X)}1_{\{W+X\leq 0\}}. \tag{3.2}$$

Using the rationality of the transform of X, we can rewrite (3.2):

$$\mathbb{E}e^{-sW}\frac{g(s) - f(s)}{g(s)} = R_-(s),$$

where $R_-(s)$ is the function on the right-hand side of (3.2), which is analytic in $\mathcal{R}e\, s < 0$ and continuous in $\mathcal{R}e\, s \leq 0$. Also, since $W \geq 0$ by definition, $\mathbb{E}e^{-sW}$ is analytic in $\mathcal{R}e\, s > 0$ and continuous in $\mathcal{R}e\, s \geq 0$.

In the next theorem we calculate the Wiener-Hopf factors of the associated random walk. These factors are the solution to the following *boundary value problem*:

Given the rational function $1 - \mathbb{E}e^{-sX}$, find two functions $K^+(s)$ and $K^-(s)$ with the following properties:

1. *$K^+(s)$ is analytic in $\mathcal{R}e\, s > 0$, $K^-(s)$ is analytic in $\mathcal{R}e\, s < 0$, and both are continuous up to the imaginary axis.*
2. *On the imaginary axis, $K^+(s)$ and $K^-(s)$ satisfy the identity*

$$(1 - K^+(s))(1 - K^-(s)) = 1 - \mathbb{E}e^{-sX}.$$

The above factorization is unique and the Wiener-Hopf factors $1-K^+(s)$, $1-K^-(s)$ can be represented using Spitzer's identity (Prabhu [94], Ch. 1). For the random walk with increments X_n, $n \geq 1$, $K^+(s)$ and $K^-(s)$ are the transforms of the first ascending ladder height and the first descending ladder height, respectively (for a probabilistic introduction, see Cohen [38]).

Using the Wiener-Hopf factorization, we now obtain the LST of the steady-state waiting time distribution:

Theorem 3.1.1. *For (A, B) having a bivariate matrix exponential distribution, the LST of the steady state waiting time is given by*

$$Ee^{-sW} = \frac{\prod_{\tilde{s}_j^-}(1 - \frac{s}{\tilde{s}_j^-})}{\prod_{s_k^-}(1 - \frac{s}{s_k^-})}, \tag{3.3}$$

where s_k^- are the zeros of $1 - \mathbb{E}e^{-sX}$ in $\mathcal{R}e\, s < 0$ and $\tilde{s}_j^-$ are its poles in $\mathcal{R}e\, s < 0$.

Proof. Let m_+ be the number of zeros of $g(s)$ in $\mathcal{R}e\, s \geq 0$. We move these to the right-hand side of the identity above:

$$\mathbb{E}e^{-sW}\frac{g(s)-f(s)}{g_-(s)} = g_+(s)R_-(s), \tag{3.4}$$

where $g_+(s) = \prod_{k=1}^{m_+}(s-\tilde{s}_k^+)$, the product being over the zeros of g with $\mathcal{R}e\, \tilde{s}_k^+ \geq 0$; and $g_-(s) = g(s)/g_+(s)$. Now the left-hand side of (3.4) is analytic in $\mathcal{R}e\, s \geq 0$, the right-hand side remains analytic in $\mathcal{R}e\, s < 0$; therefore (3.4) represents a function that is analytic everywhere (entire).

We use a version of Liouville's theorem 3.5.2 (see Appendix), which states that an entire function with asymptotic behavior $O(|s|^{m_+})$ must be a polynomial of degree at most m_+.

Liouville's theorem implies that the left-hand side of (3.4) is a polynomial $P(s)$ of degree $deg(P) \leq deg(g_+) = m_+$. Therefore we can write

$$\mathbb{E}e^{-sW} = \frac{g_-(s)}{g(s)-f(s)}P(s), \quad \mathcal{R}e\, s \geq 0. \tag{3.5}$$

Since $g_-(s)$ has zeros only in $\mathcal{R}e\, s < 0$, $P(s)$ must have all the zeros of $g-f$ from $\mathcal{R}e\, s \geq 0$ because otherwise $\mathbb{E}e^{-sW}$ would have a pole in $\mathcal{R}e\, s \geq 0$ which is not possible.

Now all boils down to showing that $g(s)-f(s)$ and $g(s)$ have the same number of zeros (i.e. m_+) in $\mathcal{R}e\, s \geq 0$. Rouché's theorem 3.5.1 in the Appendix is the right tool for this, and in Lemma 3.5.1 in the Appendix we show that indeed $|g(s)| > |f(s)|$ in $\mathcal{R}e\, s \geq 0$.

Since $P(s)$ must have these m_+ zeros of $g(s)-f(s)$ as its own, and at the same time $deg(P) \leq m_+$ from above, this determines $P(s)$ up to a constant: $P(s) = \alpha(g-f)_+(s)$, where $(g-f)_+(s) := \prod_{s_k^+}(s-s_k^+)$, s_k^+ being the zeros of $(g-f)(s)$ with $\mathcal{R}e\, s \geq 0$ (this also includes the zero at $s_0 = 0$). After replacing $P(s)$ and reducing the factors in Formula (3.5), we obtain the following formula for $\mathbb{E}e^{-sW}$:

$$\mathbb{E}e^{-sW} = \alpha\frac{\prod_{\tilde{s}_j^-}(s-\tilde{s}_j^-)}{\prod_{s_k^-}(s-s_k^-)}. \tag{3.6}$$

Setting $s=0$ determines the constant: $\alpha = \prod_{s_k^-}(-s_k^-)/\prod_{\tilde{s}_j^-}(-\tilde{s}_j^-)$, hence (3.3) follows and the proof is complete.

Remark 3.1.1. *The PASTA property (see Wolff [104]) does not hold, and hence the distribution of the steady-state workload differs in principle from cW, the steady-state workload as seen by an arriving customer. In particular, we have $\mathbb{P}(V=0) \neq \mathbb{P}(cW=0)$. Actually, we find the atom at zero of cW if we take $s \to \infty$ in (3.3), with the additional remark that the numerator has the same number of factors as the denominator, which follows from Rouché's theorem:*

$$\mathbb{P}(cW=0) = \alpha = \prod_{s_k^-} s_k^- \Big/ \prod_{\tilde{s}_j^-} \tilde{s}_j^-. \tag{3.7}$$

On the other hand, from first principles we have, with $\rho := \frac{\mathbb{E}B}{c\mathbb{E}A}$*, for the steady-state probability of an empty system:*

$$\mathbb{P}(V = 0) = 1 - \rho.$$

The factorization used in the proof of identity (3.3) can be also used to obtain the transform $\mathbb{E}e^{sI}$ of I, the steady state idle period of the system.

Corollary 3.1.1. *The transform of the idle period is given by*

$$\mathbb{E}e^{sI} = 1 - \prod_{s_k^+}(s - s_k^+)/\prod_{\tilde{s}_j^+}(s - \tilde{s}_j^+), \quad \mathcal{R}e\, s \leq 0,$$

with s_k^+ *being the zeroes of* $\frac{g(s)-f(s)}{g(s)}$ *in* $\mathcal{R}e\, s \geq 0$ *and* $\tilde{s}_j^+$ *its poles in* $\mathcal{R}e\, s > 0$.

Proof. Conditional on $W + X \leq 0$, $I = -(W + X)$, so we may write

$$\mathbb{E}e^{sI} = \frac{1}{\mathbb{P}(W + X \leq 0)} \mathbb{E}e^{s(-W-X)} 1_{\{W+X\leq 0\}}, \quad \mathcal{R}e\, s \leq 0.$$

The transform $\mathbb{E}e^{s(-W-X)} 1_{\{W+X\leq 0\}}$ already appears on the right-hand side of (3.2), hence the transform of the idle period can be rewritten as

$$\mathbb{E}e^{sI} = 1 - \frac{1}{\mathbb{P}(W + X \leq 0)} \cdot \mathbb{E}e^{-sW} \cdot \frac{g(s) - f(s)}{g(s)}. \tag{3.8}$$

As in the proof of Theorem 3.1.1, we make use of the factorizations $g(s) = g_+(s) \cdot g_-(s)$ and $(g - f)(s) = (g - f)_+(s) \cdot (g - f)_-(s)$ which were obtained via Rouché's theorem. Therefore, using (3.5), (3.6) and (3.7) we may write

$$\mathbb{E}e^{sI} = 1 - \frac{\mathbb{P}(W = 0)}{\mathbb{P}(W + X \leq 0)} \cdot \frac{g_-(s)}{(g - f)_-(s)} \cdot \frac{g(s) - f(s)}{g(s)}.$$

Note that the identity in law (3.1) implies $\mathbb{P}(W = 0) = \mathbb{P}(W + X \leq 0)$. After cancelling the factors above, $\mathbb{E}e^{sI}$ reduces to

$$\mathbb{E}e^{sI} = 1 - \prod_{s_k^+}(s - s_k^+)/\prod_{\tilde{s}_j^+}(s - \tilde{s}_j^+).$$

□

Remark 3.1.2. *Alternatively we can use Formula (6.20) in Cohen [39], p. 21, which makes use of the regenerative structure of the workload process w.r.t. the busy cycles of the queue. It can be shown that the formula remains valid even in the dependent case. The connection with (3.8) is then* $\frac{1}{\mathbb{P}(W+X\leq 0)} = \mathbb{E}N$*, the mean number of customers served during a busy cycle.*

In the next section we show that similar arguments involving regeneration as the ones employed in [39], can be extended in our setting to give the relation between the steady-state workload and waiting time distributions.

3.2 The steady-state workload

In this section we consider the steady-state workload in the queueing model with correlation between service requirement B and subsequent inter-arrival time A. We shall prove that the known relation between the steady-state workload and waiting time for the single server queue with *independent* service requirement and inter-arrival time (Asmussen [16], p. 274, Cohen [39], p. 19-20, or [40], p. 296, 297) remains valid. For this purpose we adapt the proof in [39], which is based on the fact that the workload process regenerates at the beginning of each busy cycle. The LST of the workload and waiting time distributions can then be written as stochastic mean values of the LST over one full busy cycle.

Theorem 3.2.1. *The steady-state workload V and the waiting time W are related in the following way:*

$$\mathbb{P}(V \leq v) = 1 - \rho + \rho\,\mathbb{P}(cW + B^{res} \leq v), \tag{3.9}$$

with $\rho = \frac{\mathbb{E}B}{c\mathbb{E}A}$ and B^{res} the marginal distribution of a residual service requirement, viz.,

$$\mathbb{P}(B^{res} \leq v) = \tfrac{1}{\mathbb{E}B} \int\limits_0^v \mathbb{P}(B > u)\,\mathrm{d}u.$$

Remark that only the marginal distribution of the residual service requirement appears in the above, not the joint distribution of A and B.

Proof. Let 0 be the beginning of a busy period and P be its length. Following Cohen [39], within this busy period, we may write (cf. Figure 1.1):

$$V_t = cW_{n(t)} + B_{n(t)} - c(t - t_{n(t)}),$$

where V_t is the workload at time t, $n(t)$ is the number of arrivals in $[0,t]$ and $t_{n(t)}$ is the last arrival epoch before t. The following identities hold path-wise:

$$\int_0^P e^{-sV_t}\mathrm{d}t = \int_0^P e^{-s\left[cW_{n(t)}+B_{n(t)}-c(t-t_{n(t)})\right]}\,\mathrm{d}t$$

$$= \sum_{i=1}^{N-1} \int_0^{A_i} e^{-s(cW_i+B_i-ct)}\mathrm{d}t + \int_0^{A_N-I} e^{-s(cW_N+B_N-ct)}\,\mathrm{d}t. \tag{3.10}$$

Here N is the number of customers served during a busy period. The key observation is that the following relation holds even when A_i and B_i are dependent:

$$\int_0^{A_i} e^{-s(cW_i+B_i-ct)}\mathrm{d}t = e^{-s(cW_i+B_i)}\frac{1}{cs}(e^{csA_i} - 1).$$

There is no expectation taken so integration is carried out as usual, all these being path-wise identities. Formula (3.10) now becomes

$$\int_0^P e^{-sV_t}\,\mathrm{d}t = \frac{1}{cs}\sum_{i=1}^{N-1} e^{-s(cW_i+B_i)}\left[e^{csA_i}-1\right] + \frac{1}{cs}e^{-s(cW_N+B_N)}\left[e^{cs(A_N-I)}-1\right]$$
$$= \frac{1}{cs}\sum_{i=1}^{N-1}\left[e^{-s(cW_i+B_i-cA_i)} - e^{-s(cW_i+B_i)}\right] + \frac{1}{cs}e^{-s[cW_N+B_N-c(A_N-I)]}$$
$$- \frac{1}{cs}e^{-s(cW_N+B_N)}.$$

We make use of the following identities for the waiting time during a busy period: For $i \le N-1$, $cW_i + B_i - cA_i = cW_{i+1}$; and $cW_N + B_N - cA_N = -cI$, hence

$$\int_0^P e^{-sV_t}\,\mathrm{d}t = \frac{1}{cs}\sum_{i=1}^{N-1}\left(e^{-scW_{i+1}} - e^{-scW_i-sB_i}\right) + \frac{1}{cs}\left[1-e^{-scW_N-sB_N}\right]$$

$$= \frac{1}{cs}\sum_{i=1}^{N} e^{-scW_i}(1-e^{-sB_i}). \tag{3.11}$$

All derivations up to this point are path-wise manipulations, hence insensitive to correlations between A_i and B_i. Remark that B_n is independent of W_n but also of the r.v. $1_{\{N\ge n\}}$. So if we take expectations in (3.11)

$$\mathbb{E}\int_0^P e^{-sV_t}\,\mathrm{d}t = \frac{1}{cs}\mathbb{E}\sum_{n=1}^{\infty}\left[e^{-scW_n}1_{\{N\ge n\}}(1-e^{-sB_n})\right]$$
$$= \mathbb{E}\left(\sum_{n=1}^{\infty} e^{-scW_n}1_{\{N\ge n\}}\right)\frac{1-\mathbb{E}e^{-sB_1}}{cs},$$

so that

$$\mathbb{E}\int_0^P e^{-sV_t}\,\mathrm{d}t = \mathbb{E}\left(\sum_{i=1}^{N} e^{-scW_i}\right)\frac{1-\mathbb{E}e^{-sB_1}}{cs}. \tag{3.12}$$

A key remark is that the workload process is still regenerative with respect to the renewal sequence given by the epochs at which busy periods begin. Under the stability condition, the mean cycle length $\mathbb{E}C$ (C is the length of a regeneration cycle) of the workload process is finite, hence the stochastic mean value results still hold in this case (cf. Cohen [39], Thm. 4.1) and we have the identities:

$$\mathbb{E}e^{-sV} = \frac{1}{\mathbb{E}C}\mathbb{E}\int_0^C e^{-sV_t}\,\mathrm{d}t,$$

and

$$\mathbb{E}e^{-sW} = \frac{1}{\mathbb{E}N}\mathbb{E}\sum_{1}^{N} e^{-sW_i}.$$

We can now use these identities together with (3.12) and $\mathbb{E}C = \mathbb{E}P + \mathbb{E}I$, so we may write

$$\mathbb{E}e^{-sV} = \frac{\mathbb{E}\int_0^P e^{-sV_t}\,\mathrm{d}t + \mathbb{E}I}{\mathbb{E}P + \mathbb{E}I} = \frac{\mathbb{E}N\,\mathbb{E}B}{\mathbb{E}C}\mathbb{E}e^{-scW}\frac{1-\mathbb{E}e^{-sB}}{cs\mathbb{E}B} + \frac{\mathbb{E}I}{\mathbb{E}C}.$$

Note that by definition, $P = \sum_{i=1}^{N} c^{-1}B_i$, with $(B_i)_i$ the i.i.d. sequence such that B_i is the service requirement of the ith customer in a busy cycle. Hence Wald's identity gives $c\mathbb{E}P = \mathbb{E}N\,\mathbb{E}B$, and using in addition $\frac{\mathbb{E}P}{\mathbb{E}C} = \rho$, $\frac{\mathbb{E}I}{\mathbb{E}C} = 1-\rho$, we can rewrite the above as

$$\mathbb{E}e^{-sV} = \rho\,\frac{1-\mathbb{E}e^{-sB}}{s\mathbb{E}B}\mathbb{E}e^{-scW} + (1-\rho),\ \mathcal{R}e\,s \geq 0.$$

This can immediately be inverted to give the desired relation (3.9), which concludes the proof.

3.3 Duality between the insurance and queueing processes

It is well known that there are duality relations between the classical GI/GI/1 queue and the corresponding classical Sparre-Andersen insurance risk model, with independence between service requirements (respectively claim sizes) and inter-arrival times. In this case 'corresponding' means: the same inter-arrival distributions, and that the service requirement distribution equals the claim size distribution, the service rate c is the same as the premium rate. There are two versions of the duality result (cf. Asmussen and Albrecher [17], p. 45, 161):

$$(i)\quad \Psi_0(u) = \mathbb{P}(cW > u), \tag{3.13}$$

$$(ii)\ \Psi(u) = \mathbb{P}(V > u). \tag{3.14}$$

Here $\mathbb{P}(cW > u)$ is the tail of the amount of work as seen by an arriving customer in equilibrium, and $\mathbb{P}(V > u)$ is the tail of the steady-state workload in the $G/G/1$ queue. $\Psi_0(u)$ is the ruin probability in the Sparre-Andersen model, when at time $t = 0$ the capital is u and a new inter-arrival time begins, i.e., $t = 0$ is an arrival epoch. $\Psi(u)$ is the ruin probability when the risk process is started *in stationarity*, i.e., $t = 0$ is independent of the process itself. In this case the time elapsed until the first claim arrives has a residual distribution. We will call $\Psi_0(u)$ the ordinary ruin probability and $\Psi(u)$ the delayed ruin probability. It is also worth to notice that the relations obtained in this section between Ψ_0 and Ψ do not use the bilateral matrix exponential structure of the increments X_n.

We pose and answer three questions in this section, for the dependencies under consideration (between service requirement and subsequent inter-arrival time, respectively between inter-claim time and subsequent claim size):
(1) Does the duality relation (3.13) still hold?
(2) Does the duality relation (3.14) still hold?
(3) Does the relation between steady-state workload and waiting time from Theorem 3.2.1 translate to a relation between delayed ruin probability and ordinary ruin probability, just as it does in the independent case (cf. p. 69 of Grandell [68])?

The answer to question (1) is immediately seen to be positive, as shown in Asmussen and Albrecher [17] p. 45, because this relation uses only the random walk structure of the risk/queueing process embedded at arrival epochs, which is preserved in the model we study (B_i and A_i only appear in the random walk via the difference $B_i - cA_i$). The Laplace transform of the ruin probability now immediately follows from the waiting time LST in Theorem 3.1.1, by observing that the relation $\Psi_0(u) = \mathbb{P}(cW > u)$ becomes in terms of transforms: $\Psi_0^*(s) = \frac{1}{s}\left(1 - \mathbb{E}e^{-scW}\right)$. Hence we have:

Corollary 3.3.1. *The Laplace transform of* $\Psi_0(u)$, $\Psi_0^*(s) := \int_0^\infty e^{-su}\Psi_0(u)\,\mathrm{d}u$ *equals*

$$\Psi_0^*(s) = \frac{1}{s}\left[1 - \frac{\prod_{\tilde{s}_j^-}(1 - \frac{cs}{\tilde{s}_j^-})}{\prod_{s_k^-}(1 - \frac{cs}{s_k^-})}\right].$$

This result was also obtained in Constantinescu et al. [48], using operator theory.

We shall prove that the answer to question (3) is also affirmative. In combination with the duality relation (3.13), this implies that the answer to question (2) is also affirmative: the duality relation (3.14) still holds in the dependent case.

For the purpose of studying the relation between the ordinary and the delayed ruin functions below, we assume that the pair (A, B) has a joint density, $f_{A,B}(r, z)$. Let $\phi_0(u) := 1 - \Psi_0(u)$ and $\phi(u) := 1 - \Psi(u)$ be the survival functions for the ordinary risk process and for its stationary version, respectively. In addition, denote by u the initial capital, and let $\alpha := \frac{1}{\mathbb{E}A}$ be the arrival rate of claims.

Theorem 3.3.1. *The relation between the survival functions for the two versions of the ruin process is*

$$\phi(u) = \phi(0) + \frac{\alpha}{c}\int_{v=0}^{\infty}\int_{w=0}^{u}\phi_0(u-w)\int_{z=w}^{\infty} f_{A,B}(v,z)\,\mathrm{d}z\,\mathrm{d}w\,\mathrm{d}v.$$

Let us make some remarks about this formula before proving it.

Remark 3.3.1. *In the stationary version of the ruin process, the first claim arrival happens after a time distributed as the residual inter-arrival time. Because of the correlation between claim sizes and their inter-arrival times, the claim size that corresponds to the residual arrival time will have a distinguished distribution; therefore let us denote the first pair by* (A^{res}, B^*). *Regarding the density function, it can be shown that (see Lemma 3.5.3)*

$$f_{A^{res},B^*}(r,z) = \alpha \int\limits_{v=r}^{\infty} f_{A,B}(v,z)\,\mathrm{d}v. \tag{3.15}$$

Remark 3.3.2. *The double integral that appears in the last term from Theorem 3.3.1 above:*

$$\int\limits_{v=0}^{\infty}\int\limits_{z=w}^{\infty} f_{A,B}(v,z)\,\mathrm{d}z\,\mathrm{d}v$$

equals the marginal tail of a claim size, $\mathbb{P}(B > w)$. If we replace this in the relation from Theorem 3.3.1, we obtain the same formula as in Grandell [68] p. 69:

$$\phi(u) = \phi(0) + \frac{\alpha\mathbb{E}B}{c}\int\limits_{w=0}^{u} \phi_0(u-w)\frac{\mathbb{P}(B>w)}{\mathbb{E}B}\,\mathrm{d}w. \tag{3.16}$$

This is also known as Takács' formula (see [63], Corollary 4.5.4). (3.16) shows that only the marginal residual service requirement appears in this relation between $\phi(\cdot)$ and $\phi_0(\cdot)$, even if we have the correlation between a pair (A,B).

By using the fact that $\phi(u)$, $\phi_0(u) \to 1$ as $u \to \infty$, together with dominated convergence, to argue that it is allowed to interchange limit and integration, one can easily show that $\phi(0) = 1 - \frac{\alpha\mathbb{E}B}{c}$. Now observe that Relation (3.16) between delayed and ordinary survival function is the precise counterpart/equivalent of relation (3.9) between the workload and waiting time distributions.

Proof of Theorem 3.3.1. We follow the derivation that Grandell [68] (p. 69, see also p. 5) has given for the case when A and B are independent. Starting with the stationary risk process, we condition on the arrival time of the first claim, together with its size:

$$\phi(w) = \int\limits_{r=0}^{\infty}\int\limits_{z=0}^{w+cr} \phi_0(w+cr-z) f_{A^{res},B^*}(r,z)\,\mathrm{d}z\,\mathrm{d}r.$$

Using (3.15) we obtain:

$$\phi(w) = \alpha\int\limits_{r=0}^{\infty}\int\limits_{v=r}^{\infty}\int\limits_{z=0}^{w+cr} \phi_0(w+cr-z) f_{A,B}(v,z)\,\mathrm{d}z\,\mathrm{d}v\,\mathrm{d}r.$$

By changing the order of integration between variables v and r, we have:

$$\phi(w) = \alpha\int\limits_{v=0}^{\infty}\int\limits_{r=0}^{v}\int\limits_{z=0}^{w+cr} f_{A,B}(v,z)\,\phi_0(w+cr-z)\,\mathrm{d}z\,\mathrm{d}r\,\mathrm{d}v.$$

We use the change of variable $x := w + cr$:

$$\phi(w) = \frac{\alpha}{c}\int\limits_{v=0}^{\infty}\int\limits_{x=w}^{w+cv}\int\limits_{z=0}^{x} f_{A,B}(v,z)\,\phi_0(x-z)\,\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}v. \tag{3.17}$$

Let us take the derivative of $\phi(w)$. In Lemma 3.5.2 in the Appendix we argue that this is allowed.

$$\phi'(w) = \frac{\alpha}{c}\int\limits_{v=0}^{\infty}\left[\int\limits_{z=0}^{w+cv} f_{A,B}(v,z)\,\phi_0(w+cv-z)\mathrm{d}z - \int\limits_{z=0}^{w} f_{A,B}(v,z)\,\phi_0(w-z)\mathrm{d}z\right]\mathrm{d}v$$

$$= \frac{\alpha}{c}\phi_0(w) - \frac{\alpha}{c}\int\limits_{v=0}^{\infty}\int\limits_{z=0}^{w} f_{A,B}(v,z)\,\phi_0(w-z)\,\mathrm{d}z\,\mathrm{d}v.$$

Here we replaced the first term in the right-hand side by virtue of the renewal equation for the *ordinary* survival probability. We can now integrate w between 0 and u:

$$\phi(u) - \phi(0) = \frac{\alpha}{c}\int\limits_{w=0}^{u}\phi_0(w)\,\mathrm{d}w - \frac{\alpha}{c}\int\limits_{w=0}^{u}\int\limits_{v=0}^{\infty}\int\limits_{z=0}^{w}\phi_0(w-z)\,f_{A,B}(v,z)\,\mathrm{d}z\,\mathrm{d}v\,\mathrm{d}w. \quad (3.18)$$

Let us focus on the last term from (3.18), to be called L. Integration over v yields, with $f_B(\cdot)$ the density of the service requirement B:

$$L = \frac{\alpha}{c}\int\limits_{w=0}^{u}\int\limits_{z=0}^{w}\phi_0(w-z)f_B(z)\mathrm{d}z\,\mathrm{d}w. \quad (3.19)$$

Partial integration gives:

$$\begin{aligned}
L &= \frac{\alpha}{c}\int\limits_{w=0}^{u}\phi_0(0)\mathbb{P}(B\le w)\,\mathrm{d}w + \frac{\alpha}{c}\int_{w=0}^{u}\int_{z=0}^{w}\mathbb{P}(B\le z)\phi_0'(w-z)\,\mathrm{d}z\,\mathrm{d}w \\
&= \frac{\alpha}{c}\int\limits_{w=0}^{u}\phi_0(0)\mathbb{P}(B\le w)\,\mathrm{d}w + \frac{\alpha}{c}\int_{z=0}^{u}\mathbb{P}(B\le z)\int_{w=z}^{u}\phi_0'(w-z)\,\mathrm{d}w\,\mathrm{d}z \\
&= \frac{\alpha}{c}\int_{w=0}^{u}\phi_0(0)\mathbb{P}(B\le w)\,\mathrm{d}w + \frac{\alpha}{c}\int_{z=0}^{u}\mathbb{P}(B\le z)[\phi_0(u-z)-\phi_0(0)]\,\mathrm{d}z \\
&= \frac{\alpha}{c}\int_{z=0}^{u}\mathbb{P}(B\le z)\phi_0(u-z)\,\mathrm{d}z. \qquad (3.20)
\end{aligned}$$

Substitution of (3.20) in (3.18) gives (3.16) and thus the result of the theorem.

3.4 Examples and numerical results

In this section we present examples of dependence structures which are tractable and have a probabilistic interpretation. We also numerically illustrate the effect of correlations on the waiting time distribution/ruin probability. Throughout the section we take for simplicity $c = 1$.

A comprehensive survey of multivariate matrix-exponential distributions (MVME) can be found in Bladt and Nielsen [29]. As a special subclass of these, Kulkarni [79] introduced multivariate phase-type (MPH) distributions (see also Assaf et al. [18]). In the bivariate case, these are defined as follows: Consider a continuous-time Markov chain $J(t)$ with finite state space $\mathcal{S}$, with an absorbing state Δ, and generator matrix

$$\mathbf{Q} = \begin{pmatrix} Q & -Q\mathbf{1} \\ 0 & 0 \end{pmatrix}$$

together with a reward matrix $(r_x^{(k)})_{x,k}$, $r_x^{(k)} \geq 0$ for $x \in \mathcal{S}\backslash\{\Delta\}$, $k = 1, 2$. Assume that as long as the chain is in state x, we earn at rate vector $\mathbf{r}_x = (r_x^{(1)}, r_x^{(2)})$. We look at the bivariate distribution of the random vector (Z_1, Z_2), where the marginals of this vector are defined to be the total accumulated rewards until absorption:

$$Z_k = \int_0^{\zeta} r_{J(t)}^{(k)} \mathrm{d}t,$$

with ζ the time to absorption. Remark that Z_k can be rewritten as

$$Z_k = \sum_{i=1}^{\kappa} r_{J_i}^{(k)} H_i, \quad k = 1, 2, \tag{3.21}$$

κ being the number of jumps until absorption of the embedded discrete-time Markov chain J_i and H_i the holding time in state J_i. The H_i's are independent exponentials with rates $-Q_{J_i J_i}$. The dependence structure between Z_1 and Z_2 is thus given by the underlying continuous-time Markov chain $J(t)$. That this is indeed a subclass of MVME, follows from [29], Theorem 4.1.

As a special case of Kulkarni's bivariate-phase type distributions, one can obtain a fairly large class of distributions by a partial decoupling of the bivariate phase-type: For the discrete-time Markov chain J_i, and for a fixed i, let $H_i^{(1)}, H_i^{(2)}$ be independent, having exponential distributions with rates λ_{J_i} and μ_{J_i}, respectively. Without loss of generality we can consider $r_{J_i}^{(k)} = 1$, $k = 1, 2$ and set

$$A = \sum_{i=1}^{\kappa} H_i^{(1)}, \quad B = \sum_{i=1}^{\kappa} H_i^{(2)}.$$

The difference with Formula (3.21) is that now the dependence structure is given only by the common underlying discrete-time Markov chain J_i. Furthermore, if we assume the jump rates to be the same in each state, i.e. $H_i^{(1)} \sim \exp(\lambda)$, $H_i^{(2)} \sim \exp(\mu)$, then the number of jumps κ before absorption is a sufficient statistic for the joint distribution of (A, B). More precisely, conditional on κ, A and B are independent Erlang(κ, λ), Erlang(κ, μ) respectively.

Remark 3.4.1. *This dependence structure can be realized as in the description of Kulkarni's class. More precisely, we obtain the partial decoupling by doubling all states of the underlying Markov Process: replace each transient state x with x_1, x_2 and allow only the corresponding component of (A, B) to increase while in state x_i (formally, put*

$r_{x_1}^{(1)} = r_x^{(1)}$, $r_{x_2}^{(1)} = 0$ and similarly $r_{x_1}^{(2)} = 0$, $r_{x_2}^{(2)} = r_x^{(2)}$). Extend the transition matrix of the Markov Chain such that after visiting state x_1, it always jumps to state x_2 and thereafter jumps according to the original transition matrix.

If we denote by $\boldsymbol{\alpha}$ the initial distribution of $(J_n)_n$, by T the transient component of its transition matrix, and by $\boldsymbol{t}$ the vector of exit probabilities, then by conditioning on κ we obtain the following result as a probabilistic alternative to Theorem 3.2 in Bladt and Nielsen [29]:

Lemma 3.4.1. *The following are valid for the random vector (A, B):*

a) The Laplace-Stieltjes transform of (A, B) is:

$$\mathbb{E}e^{-s_1 A - s_2 B} = \boldsymbol{\alpha}' \left[\frac{(\lambda + s_1)(\mu + s_2)}{\lambda\mu} I - T \right]^{-1} \boldsymbol{t}.$$

b) The transform $\mathbb{E}e^{-sX}$ of the difference $(B - A)$, is a rational function of the form $\frac{f(s)}{g(s)}$, with f and g polynomial functions such that $\deg(f) < \deg(g)$.

Proof. see Appendix A.

Examples: **1**. Kibble and Moran's bivariate Gamma distribution (Kotz et al. [78]) can be realized as above. Consider the state space $\{1, ..., m, \Delta\}$. Assume the Markov Chain $(J_n)_n$ starts in 1 and jumps from i to $i+1$ w.p. p or stays in state i w.p. $1-p$. Furthermore, assume the same rates for the holding times in every state: $H_n^{(1)} \sim$exp(λ), $H_n^{(2)} \sim$exp(μ), for λ, $\mu > 0$. Hence this distribution is the $m-$fold convolution of Kibble and Moran's bivariate exponential with itself (cf. [78]), where this bivariate exponential distribution can be represented as

$$(Erlang(\kappa, \lambda), Erlang(\kappa, \mu)),$$

with κ having a geometric distribution. In the insurance risk setting, the analysis for this example has been done in Ambagaspitiya [9] and in Constantinescu et al. [48] using operator theory (see also Albrecher and Teugels [8]). The Laplace transform of the ordinary ruin probability $\Psi_0(u)$ is given by

$$\Psi_0^*(s) = \frac{1}{s} \left[1 - \frac{(1 - \frac{s}{b})^m}{\prod_{s_k} (1 - \frac{s}{s_k})} \right],$$

with b the pole of order m of $1 - \mathbb{E}e^{-sX}$ such that $\mathcal{R}e\, b < 0$.

2. Cheriyan and Ramabhadran's bivariate Gamma is another example of Kulkarni's bivariate phase-type. This was also analyzed in Ambagaspitiya [9] in the insurance risk setting.

For non-negative integers m_0, m_1, m_2, consider the state space $\mathcal{S} = \{1, ..., m_0 + m_1 + m_2, \Delta\}$, with the set of transient states partitioned as: $\mathcal{S} \backslash \{\Delta\} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \mathcal{S}_2$ with $\mathcal{S}_0 = \{1, ..., m_0\}$, $\mathcal{S}_1 = \{m_0 + 1, ..., m_0 + m_1\}$, $\mathcal{S}_2 = \{m_0 + m_1 + 1, ..., m_0 + m_1 + m_2\}$. The chain starts in state 1 and jumps from state i to $i+1$. The jump rates are β_k

while in state $x \in \mathcal{S}_k$, $k \in \{0,1,2\}$. The reward rates in state x are $r_x^{(1)} = r_x^{(2)} = 1$ for $x \in \mathcal{S}_0$; $r_x^{(1)} = 1, r_x^{(2)} = 0$ for $x \in \mathcal{S}_1$, and $r_x^{(1)} = 0$, $r_x^{(2)} = 1$ for $x \in \mathcal{S}_2$. Then the bivariate total accumulated reward has a distribution of the form

$$(A, B) \stackrel{d}{=} (Z_0 + Z_1, Z_0 + Z_2),$$

where Z_k are mutually independent $\sim$Erlang(m_k, β_k), $k \in \{0,1,2\}$.

3. In the class of MVME, it is possible to achieve negative correlation as well. Consider κ to be a discrete random variable with finite support: $\kappa \in \{1, ..., k\}$, for k some fixed positive integer. Negative correlation can be achieved if we consider the following mixture of Erlang distributions:

$$(A, B)_- \stackrel{d}{=} (Erlang(\kappa, \lambda), Erlang(k - \kappa + 1, \mu)).$$

For more examples of negatively correlated phase-type distributions, we refer to [30].

Stochastic ordering results. We compare the tails of the waiting times for the mixed Erlang distributions in the following scenarios: the negatively correlated one from Example 3 versus the positively correlated case

$$(A, B)_+ \stackrel{d}{=} (Erlang(\kappa, \lambda), Erlang(\kappa, \mu)),$$

and the corresponding independent pair obtained by sampling twice from the distribution of κ; i.e. for κ_1 and κ_2 i.i.d. copies of κ.

$$(A, B)_0 \stackrel{d}{=} (Erlang(\kappa_1, \lambda), Erlang(\kappa_2, \mu)).$$

Here κ is taken to have finite support, as in Example 3 above.

Denote respectively by D_-, D_+ and D_0, the differences $A - B$ in the three scenarios above. In Theorem 3.4.1 below we show that under a mild assumption on the distribution of κ, there exists *convex ordering* between the random variables D_+, D_0 and D_-. For two r.v.'s X and Y, $X \preceq_{cx} Y$ means, by definition, that for an arbitrary convex function $\varphi(x)$,

$$\mathbb{E}\varphi(X) \leq \mathbb{E}\varphi(Y). \tag{3.22}$$

For more about the notion of convex order and other related stochastic orderings, we refer the reader to [100], Ch. 1. Before we give the result, let us recall a useful criterion (cf. [100], Prop. 1.5.1):

Proposition 3.4.1 (Karlin and Novikoff's cut criterion)**.** *For X, Y r.v.'s with c.d.f.'s F_X and F_Y respectively, and finite first moments, assume that $\mathbb{E}X = \mathbb{E}Y$, and that there exists an x_0 such that $F_X(x) \leq F_Y(x)$, for $x \leq x_0$ and $F_X(x) \geq F_Y(x)$ for $x \geq x_0$. Then $X \preceq_{cx} Y$.*

Theorem 3.4.1. *With the above definitions and notations, it holds that*

$$D_+ \preceq_{cx} D_0. \tag{3.23}$$

Moreover, if κ has a symmetric distribution, $\kappa \overset{d}{=} k+1-\kappa$, then we also have

$$D_0 \preceq_{cx} D_-. \tag{3.24}$$

Proof. Let C_i^λ and C_j^μ respectively be Erlang(i,λ) and Erlang(j,μ) distributed random variables independent of each other, for $i = 1,\dots,k$; also denote $\pi_i := \mathbb{P}(\kappa = i)$.

We will first prove *icx* ordering, i.e., the functional inequality (3.22) is restricted to *increasing* convex functions φ. This together with the fact that the expected values of D_-, D_+ and D_0 are the same implies *cx* ordering (see [100], Thm. 1.3.1, p. 9).

Take φ to be any convex and increasing function. Firstly, we prove (3.23), that is, we must show that $\mathbb{E}\varphi(D_+) \le \mathbb{E}\varphi(D_0)$, or equivalently,

$$\sum_{i=1}^k \pi_i \mathbb{E}\varphi(C_i^\lambda - C_i^\mu) \le \sum_{i=1}^k \sum_{j=1}^k \pi_i\pi_j \mathbb{E}\varphi(C_i^\lambda - C_j^\mu).$$

Let us put for simplicity $\varphi(i,j) := \mathbb{E}\varphi(C_i^\lambda - C_j^\mu)$, so we can rewrite the above as

$$\sum_i \pi_i\varphi(i,i) \le \sum_i \sum_j \pi_i\pi_j\varphi(i,j). \tag{3.25}$$

Note that (3.25) is an association type of inequality, similar to Cebishev's inequality (see [14], Lemma 2.3 and the references therein). Using that the $\pi_j's$ form a probability distribution, we can further rewrite (3.25)

$$\sum_i \sum_j \pi_i\pi_j\varphi(i,i) \le \sum_i \sum_j \pi_i\pi_j\varphi(i,j)$$

$$\Leftrightarrow \sum_i \sum_{j>i} \pi_i\pi_j\left[\varphi(i,i) - \varphi(i,j)\right] \le \sum_m \sum_{l<m} \pi_m\pi_l[\varphi(m,l) - \varphi(m,m)]. \tag{3.26}$$

Remark that there is an equal number of terms on the two sides of (3.26) because we sum over indices that lie respectively above and below the main diagonal of the tableaux $(\varphi(i,j))_{i,j}$. We are done as soon as we show that the inequality holds for a one-to-one correspondence between these indices; more precisely, for the correspondence $(i,j) \leftrightarrow (j,i)$, $j > i$, we will prove that

$$\varphi(i,i) - \varphi(i,j) \le \varphi(j,i) - \varphi(j,j), \tag{3.27}$$

that is, (3.26) holds term by term, and remark that the coefficients $\pi_i\pi_j$ and $\pi_j\pi_i$ cancel against each other. Put $u := j - i$ and denote

$$\gamma(x) := \mathbb{E}\varphi(x + C_i^\lambda - C_j^\mu).$$

Obviously, $\gamma(x)$ is increasing and convex, because φ is. Consider the decomposition of C_j^μ and C_j^λ as sums of independent r.v.'s $C_j^\mu := C_i^\mu + C_u^\mu$, and $C_j^\lambda := C_i^\lambda + C_u^\lambda$ with C_u^μ, C_u^λ Erlang distributed of order u and rates μ and λ, respectively. By conditioning on C_u^λ and C_u^μ, we can write

$$\varphi(j,i) = \mathbb{E}\{\mathbb{E}[\varphi(y + C_i^\lambda - C_i^\mu - x + x)|C_u^\lambda = y, C_u^\mu = x]\} \Leftrightarrow$$

$$\varphi(j,i) = \mathbb{E}\{\mathbb{E}[\gamma(y+x)|C_u^\lambda = y, C_u^\mu = x]\} = \mathbb{E}\gamma(C_u^\lambda + C_u^\mu).$$

Similarly, we obtain $\varphi(i,i) = \mathbb{E}\gamma(C_u^\mu)$ and $\varphi(j,j) = \mathbb{E}\gamma(C_u^\lambda)$, so that (3.27) becomes

$$\mathbb{E}\gamma(C_u^\mu) + \mathbb{E}\gamma(C_u^\lambda) \leq \mathbb{E}\gamma(C_u^\lambda + C_u^\mu) + \gamma(0). \tag{3.28}$$

All boils down to proving (3.28). In order to achieve this, let U be a r.v. with a Bernoulli(1/2) distribution and let $c_\mu \neq c_\lambda$ be two arbitrary positive constants. Consider the following r.v.'s

$$Z_1 := (c_\lambda + c_\mu)U, \quad Z_2 := c_\lambda U + c_\mu(1-U).$$

We have the following identities in distribution

$$Z_1 \stackrel{d}{=} \frac{1}{2}[\delta_0 + \delta_{c_\lambda + c_\mu}], \quad Z_2 \stackrel{d}{=} \frac{1}{2}[\delta_{c_\lambda} + \delta_{c_\mu}],$$

with δ_x being the Dirac measure at x. Now it follows easily from the cut criterion in Proposition 3.4.1 above that $Z_2 \preceq_{cx} Z_1$. Hence, in particular, we can choose $\gamma(x)$ as a test function to obtain

$$\mathbb{E}\gamma(c_\lambda U + c_\mu(1-U)) \leq \mathbb{E}\gamma(c_\lambda U + c_\mu U).$$

Because U is a Bernoulli(1/2), the inequality above becomes

$$\gamma(c_\lambda) + \gamma(c_\mu) \leq \gamma(c_\lambda + c_\mu) + \gamma(0).$$

Finally, taking the double mixture over c_λ and c_μ according to the distributions of C_u^λ and C_u^μ respectively, shows that (3.28) is true, and this proves (3.23).

Now, for inequality (3.24) we have to prove that $\mathbb{E}\varphi(D_0) \leq \mathbb{E}\varphi(D_-)$, that is, keeping the same notation as in (3.25),

$$\sum_i \sum_j \pi_i \pi_j \varphi(i,j) \leq \sum_i \sum_j \pi_i \pi_j \varphi(i, k+1-i),$$

and upon regrouping terms it becomes

$$\sum_i \sum_{j:\, j<k+1-i} \pi_i \pi_j [\varphi(i,j) - \varphi(i,k+1-i)] \leq \sum_m \sum_{l:\, l>k+1-m} \pi_m \pi_l [\varphi(m,k+1-m) - \varphi(m,l)].$$

This is the analogue of (3.26). Again, it suffices to prove the term by term inequalities similar to (3.27). The symmetry axis in this case is the second diagonal of the tableaux. This means that the correspondence is $(i,j) \leftrightarrow (k+1-j, k+1-i)$, so the analogue of (3.27) that we prove is, for i,j fixed, $j < k+1-i$,

$$\varphi(i,j) - \varphi(i,k+1-i) \leq \varphi(k+1-j,j) - \varphi(k+1-j,k+1-i). \tag{3.29}$$

In (3.29) we dropped the coefficients $\pi_i \pi_j$ and $\pi_{k+1-i}\pi_{k+1-j}$ because these are equal since κ is assumed to have a symmetric distribution. If we set $u = (k+1-i) - j =$

$(k+1-j)-i$, from this point on the analysis is essentially the same. Consider the analogue of γ,

$$\eta(x) := \mathbb{E}\varphi\left(x + C_i^\lambda - C_{k+1-i}^\mu\right),$$

then (3.29) becomes

$$\mathbb{E}\eta(C_u^\mu) - \eta(0) \leq \mathbb{E}\eta(C_u^\lambda + C_u^\mu) - \mathbb{E}\eta(C_u^\lambda).$$

This is precisely (3.28) with $\gamma(x)$ replaced by $\eta(x)$, and since φ was taken to be an arbitrary increasing convex function, the proof is complete.

Remark 3.4.2. *The requirement for κ to have a symmetric distribution may be too strong in general. Some assumption on the distribution of κ is necessary but only for the ordering $D_0 \preceq_{cx} D_-$. For example, if we let $k = 2$ and $\kappa \stackrel{d}{=} \delta_1$ (Dirac mass in 1) then D_0 is the difference of two independent Erlang-1, whereas D_- is an Erlang-1 minus an Erlang-2 so D_- is cx-dominated in this case.*

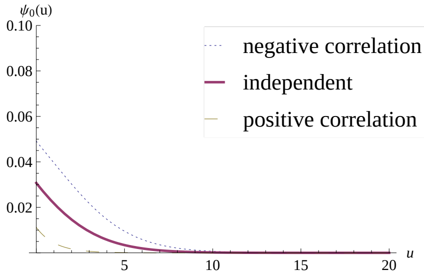
The above proof of the inequality between D_+ and D_0 does not require the finiteness of the support of κ; κ discrete phase type is also a possible case in which the sums that appear in the proof become series. There are no convergence problems and we are allowed to change summation order as well, due to probabilistic interpretations. There are restrictions if we look for negative correlation when κ has infinite support. More about the possibility for realizing such correlations can be found in Bladt and Nielsen [30] on negatively correlated exponentials.

Proposition 3.4.2. *Let W_-, W_0, and W_+, be the steady-state waiting times, that correspond to the increments of the random walk distributed as $-D_-$, $-D_0$, and $-D_+$, respectively. Then we have convex ordering between the waiting times in the three scenarios*

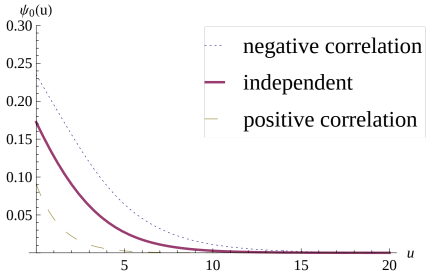
$$W_+ \preceq_{cx} W_0 \preceq_{cx} W_-.$$

Proof. From the definition of convex ordering, $D_+ \preceq_{cx} D_0$ is the same as $-D_+ \preceq_{cx} -D_0$, and similarly $D_0 \preceq_{cx} D_-$ is the same as $-D_0 \preceq_{cx} -D_-$. Therefore the external monotonicity result from Daley and Stoyan [100] (Thm. 5.2.1, p. 80) implies that the steady state workloads are convex ordered in the three scenarios, according to the increments of the random walk. This can also be seen in the numerical tables and the plots below. □

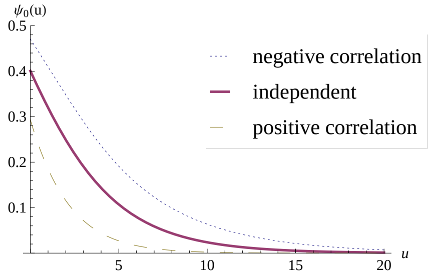
In Table 3.1 we vary the load coefficient ρ and we keep the mixing distribution κ uniform on $\{1, ..., 5\}$ (i.e., $k = 5$). In Table 3.2 below, we keep ρ fixed, say $\rho = .5$, and we vary k. The tables contain the mean waiting times, their atoms at zero and q, the 95% quantile of the survival function/waiting time (i.e., q is the value of the initial capital for which $\mathbb{P}(W \leq q) = \phi_0(q) = .95$). The plots of the tails of the ruin functions are in Figure 3.1 and Figure 3.2 below.



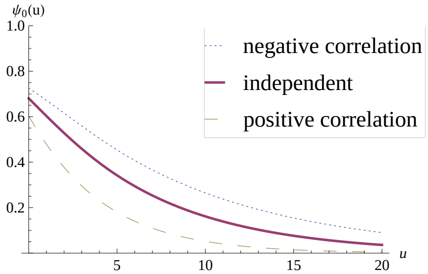
(a) $\rho = .05,\ k = 5$



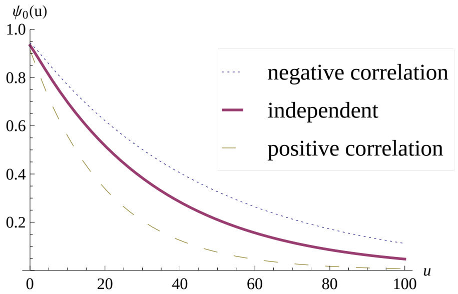
(b) $\rho = .25,\ k = 5$



(c) $\rho = .5,\ k = 5$



(d) $\rho = .75,\ k = 5$



(e) $\rho = .95,\ k = 5$

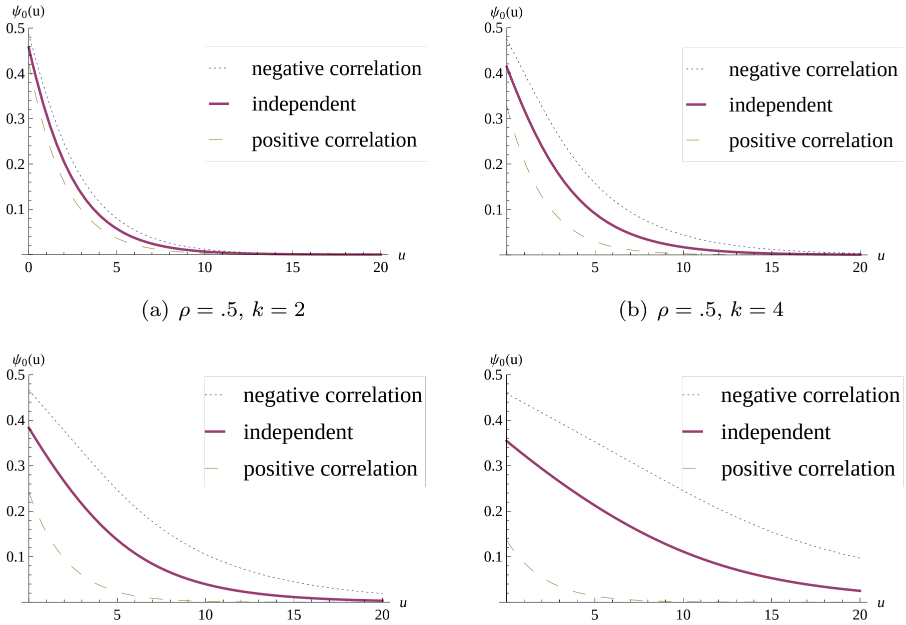
FIGURE 3.1: $\mathbb{P}(W > u) = \Psi_0(u)$.

ρ	$\mathbb{E}W_+$	$\mathbb{E}W_0$	$\mathbb{E}W_-$	$\mathbb{P}(W_+ = 0)$	$\mathbb{P}(W_0 = 0)$	$\mathbb{P}(W_- = 0)$	q_+	q_0	q_-
.05	0.01	0.07	0.15	0.988	0.96	0.95	0	0	0
.25	0.12	0.47	0.88	0.90	0.82	0.76	0.85	3.54	5.72
.5	0.62	1.50	2.48	0.70	0.59	0.52	3.74	7.54	11.1
.75	2.48	4.77	7.15	0.39	0.32	0.27	10.14	17.81	25.5
.95	18.4	31.4	44.48	0.08	0.066	0.056	58.26	97.89	137.5

TABLE 3.1: Mean waiting times, atoms at 0 and 95% percentiles for $k = 5$ and various values of ρ.

k	$\mathbb{E}W_+$	$\mathbb{E}W_0$	$\mathbb{E}W_-$	$\mathbb{P}(W_+ = 0)$	$\mathbb{P}(W_0 = 0)$	$\mathbb{P}(W_- = 0)$	q_+	q_0	q_-
2	0.86	1.11	1.36	0.57	0.54	0.51	4.36	5.31	6.25
4	0.68	1.37	2.11	0.67	0.58	0.52	3.93	6.78	9.48
7	0.51	1.78	3.22	0.75	0.61	0.53	3.39	9.09	14.35
14	0.31	2.79	5.82	0.85	0.64	0.540	2.33	14.58	25.74

TABLE 3.2: Mean waiting times, atoms at 0 and 95% percentiles for $\rho = .5$ and various values of k.



(a) $\rho = .5,\ k = 2$ (b) $\rho = .5,\ k = 4$

(c) $\rho = .5,\ k = 7$ (d) $\rho = .5,\ k = 14$

FIGURE 3.2: $\mathbb{P}(W > u) = \Psi_0(u)$.

3.5 Appendix A

Theorem 3.5.1 (Rouché, [102], p. 116)**.** *If two functions $g(s)$ and $f(s)$ are analytic inside and on a closed contour C, and $|g(s)| > |f(s)|$ on C, then $g(s)$ and $g(s) - f(s)$ have the same number of zeros inside C.*

Theorem 3.5.2 (Liouville, [102], p. 85)**.** *If $f(s)$ is analytic for all finite values of s, and as $|s| \to \infty$,*

$$f(s) = O(|s|^m),$$

then $f(s)$ is a polynomial of order $\leq m$.

We can now formulate and prove the following lemma.

Lemma 3.5.1. *Let $f(s)$ and $g(s)$ be the numerator and the denominator of $\mathbb{E}e^{-s(c^{-1}B-A)}$. Then $g(s) - f(s)$ and $g(s)$ have the same number of zeros in $\mathcal{R}e\, s \geq 0$.*

Proof. Via Rouché's theorem, we first prove that $|g(s)| > |f(s)|$ on a suitably chosen contour in the complex plane. The fact that $f(0) = g(0)$ and that the transform is rational (so it is also analytic on a strip in $\mathcal{R}e\ s < 0$) suggests that we consider the following contour made up from the extended semi-circle

$$\mathcal{C}_\epsilon := \{R(\cos\varphi + i\sin\varphi);\ \varphi \in [-\pi/2 - \arccos\epsilon, \pi/2 + \arccos\epsilon]\},$$

together with the vertical line segment $S := \left\{-\epsilon + i\omega;\ |\omega| \in \left[0, R\sqrt{1-\epsilon^2}\right]\right\}$.

We show that $|g(s)| > |f(s)|$ on this contour, for ϵ sufficiently small.

First on $\mathcal{C}_\epsilon$: $\left|\frac{f(Re^{i\varphi})}{g(Re^{i\varphi})}\right| \leq \mathbb{E}e^{-R\cos\varphi(c^{-1}B-A)} \to \mathbb{P}(c^{-1}B - A = 0)$ as $R \to \infty$.

We can assume $\mathbb{P}(A = c^{-1}B) < 1$, else there is nothing to prove. This means $|\frac{f(Re^{i\varphi})}{g(Re^{i\varphi})}| < 1$ for R sufficiently large.

In order to prove the inequality on the line segment S, we use the stability condition: $\mathbb{E}(A - c^{-1}B) = \frac{d}{ds}\frac{f(s)}{g(s)}|_{s=0} > 0$. So for ϵ sufficiently small, $\frac{f(-\epsilon)}{g(-\epsilon)} < \frac{f(0)}{g(0)} = 1$. Then on S we have:

$$\left|\frac{f(-\epsilon + i\omega)}{g(-\epsilon + i\omega)}\right| = |\mathbb{E}e^{-(-\epsilon+i\omega)(c^{-1}B-A)}| \leq \mathbb{E}e^{\epsilon(c^{-1}B-A)}|e^{-i\omega(c^{-1}B-A)}| = \frac{f(-\epsilon)}{g(-\epsilon)} < 1.$$

Hence $|f(s)| < |g(s)|$ on the whole contour. These being polynomials, Rouché's theorem 3.5.1 ensures that g and $g - f$ have the same number of zeros inside $\mathcal{C}_\epsilon$, and since ϵ was arbitrarily small, this also holds on $\cap_{\epsilon>0}\mathcal{C}_\epsilon^\circ = \{s;\ \mathcal{R}e\, s \geq 0\} \cap \{s;\ |s| \leq R\}$, where $\mathcal{C}_\epsilon^\circ$ is the interior of C_ϵ. Finally, letting $R \to \infty$, proves the assertion. □

Proof of Lemma 3.4.1. a) We can write the joint Laplace-Stieltjes transform by conditioning on κ:

$$\mathbb{E}e^{-s_1A-s_2B} = \sum_{n=1}^{\infty}\mathbb{P}(\kappa = n)\left(\frac{\lambda}{\lambda + s_1}\right)^n\left(\frac{\mu}{\mu + s_2}\right)^n.$$

If we set $z = \frac{\lambda}{\lambda+s_1}\frac{\mu}{\mu+s_2}$, we can recognize the probability generating function of κ at z, call it $P_\kappa(z)$.

κ has a discrete phase-type distribution with representation $(\boldsymbol{\alpha}, T)$ (Neuts [90]), such that $I - T$ is non-singular (here I is the identity matrix), and the probability vector $\boldsymbol{\alpha}$ is supported on the transient states. Thus

$$\mathbb{P}(\kappa = n) = \boldsymbol{\alpha}'T^{n-1}\boldsymbol{t}$$

for $n \geq 1$, $\boldsymbol{t} = (I - T)\mathbf{1}$, $P(\kappa = 0) = 0$. If we now focus on this generating function, we have the following (Asmussen [16] Prop. 4.1, p. 83):

$$P_\kappa(z) = \boldsymbol{\alpha}'(z^{-1}I - T)^{-1}\boldsymbol{t},$$

and we have proved part a).

b) To see why $\mathbb{E}e^{-sX} = P_\kappa(\frac{\lambda}{\lambda-s}\frac{\mu}{\mu+s})$ is a rational function, rewrite the inverse :

$$(z^{-1}I - T)^{-1} = \frac{1}{\det(z^{-1}I - T)}(z^{-1}I - T)^*.$$

Remark that the denominator $\det(z^{-1}I - T)$ is a polynomial of order $|\mathcal{S}| - 1$ (the number of transient states) in z^{-1}, because z^{-1} appears only on the diagonal of the matrix $(z^{-1}I-T)$. $(z^{-1}I-T)^*$ is the algebraic complement of $(z^{-1}I-T)$ (also known as matrix of cofactors). Its entries are of the form $(-1)^{i+j}\det(M_{ij})$, where M_{ij} is the matrix obtained by deleting row i and column j of $(z^{-1}I-T)$. These are polynomials in z^{-1} of order $< |\mathcal{S}|-1$ (because of the deleted rows and columns in the entries, the degree of the determinants of these sub-blocks as polynomials in z^{-1} is always smaller than the dimension of the matrix T) and hence so is the bilinear form $\boldsymbol{\alpha}'(z^{-1}I-T)^*\boldsymbol{t}$, which is the numerator of $P_\kappa(z)$. □

Lemma 3.5.2. *$\phi(w)$ in (3.17) is differentiable.*

Proof. Let $h_w(v) := \int\limits_{x=w}^{w+cv}\int\limits_{z=0}^{x} f_{A,B}(v,z)\,\phi_0(x-z)\,\mathrm{d}z\,\mathrm{d}x$. Using the triangle inequality, we have the following upper bound

$$|h_{w+\epsilon}(v)-h_w(v)| \le \int\limits_{w}^{w+\epsilon}\int\limits_{0}^{x} f_{A,B}(v,z)\phi_0(x-z)\,\mathrm{d}z\,\mathrm{d}x + \int\limits_{w+cv}^{w+cv+\epsilon}\int\limits_{0}^{x} f_{A,B}(v,z)\phi_0(x-z)\,\mathrm{d}z\,\mathrm{d}x.$$

Let us denote by I and II the first and the second term that appear above, respectively. If we use the fact that $\phi_0(x)\le 1$, we find the upper bounds on I and II:

$$I \le \int\limits_{x=w}^{w+\epsilon}\int\limits_{z=0}^{w+\epsilon} f_{A,B}(v,z)\,\mathrm{d}z\,\mathrm{d}x = \epsilon\int\limits_{z=0}^{w+\epsilon} f_{A,B}(v,z)\,\mathrm{d}z,$$

and similarly,

$$II \le \epsilon \int\limits_{z=0}^{w+cv+\epsilon} f_{A,B}(v,z)\,\mathrm{d}z.$$

So if we denote $D_\epsilon(v) := \frac{h_{w+\epsilon}(v)-h_w(v)}{\epsilon}$,

$$|D_\epsilon(v)| \le \int\limits_{z=0}^{w+\epsilon} f_{A,B}(v,z)\,\mathrm{d}z + \int\limits_{z=0}^{w+cv+\epsilon} f_{A,B}(v,z)\,\mathrm{d}z \le 2f_A(v),$$

and clearly the upper bound is integrable as a function of v. By virtue of dominated convergence

$$\phi'(w) = \lim_{\epsilon\to 0}\int\limits_{v=0}^{\infty} D_\epsilon(v)\,\mathrm{d}v = \int\limits_{v=0}^{\infty}\lim_{\epsilon\to 0} D_\epsilon(v)\,\mathrm{d}v = \int\limits_{v=0}^{\infty}\frac{\partial}{\partial w}h_w(v)\,\mathrm{d}v.$$

□

Lemma 3.5.3. *Under the conditions from Remark 3.3.1, the density of the pair (A^{res}, B^*) is*

$$f_{(A^{res},B^*)}(r,z) = \alpha\int\limits_{v=r}^{\infty} f_{(A,B)}(v,z)\,\mathrm{d}v.$$

Proof. Consider the augmented pair $(\tilde{A}, B^*)$ which by definition has density

$$f_{(\tilde{A},B^*)}(v,z) := \alpha v f_{A,B}(v,z),$$

where α acts as the normalizing factor: $\frac{1}{\alpha} = \mathbb{E}A = \int_z \int_v v f_{A,B}(v,z)\,\mathrm{d}v\,\mathrm{d}z$. Let U be a standard uniform r.v., independent of both A and B. Then $(A^{res}, B^*) \stackrel{d}{=} ((1-U)\tilde{A}, B^*)$, therefore conditional on $\tilde{A}$, A^{res} is uniformly distributed over the interval $[0, \tilde{A}]$, so we may write in terms of density functions

$$f_{(A^{res},B^*)}(v,z) = \int_{r=v}^{\infty} \frac{1}{r} f_{(\tilde{A},B^*)}(r,z)\,\mathrm{d}r = \alpha \int_{r=v}^{\infty} f_{(A,B)}(r,z)\,\mathrm{d}r.$$

The proof is complete.

Chapter 4

Integral representations for one-dimensional random walks

A usual assumption in the theory of fluctuations of random walks as they appear in, e.g., queueing or insurance applications, is that the increment of the random walk can be represented as the difference of two *independent* random variables. In the context of queueing theory, random walks with increments which *do not have* this property appear as embedded at arrival epochs of customers in a GI/G/1 queue in which the service requirement of the current customer is correlated with the time until the next arrival. In risk reserve processes that appear in insurance, the independence assumption is violated when the current claim size depends on the time elapsed since the previous arrival, and hence on the premium gained meanwhile. A useful relation between the queueing system and the risk reserve model is given via the duality relation described at the end of Section 1.3.

The purpose of the present chapter is to show how much can be done for random walks which do not satisfy this independence assumption, regarding their maxima (as in the waiting time/maximum aggregate loss), their minima (idle periods/deficit at ruin) and their excursions (related to busy periods/time to ruin).

From a queueing perspective, it turns out that the busy period is a more sensitive issue to study than the idle period or the waiting time, as it appears from the proof of Theorem 4.2.1 below. For this purpose, we present a generalization of an inversion formula for characteristic functions/Fourier-Stieltjes transforms due to Hewitt [69], which in turn is an extension of P. Lévy's inversion formula. All these results are essentially variations on the Dirichlet integral for complex-valued functions of bounded variation.

From a stochastic point of view, the information contained by the increments of the random walk is sufficient to infer about the extreme statistics; and since successive increments are independent, the usual form of Hewitt's inversion formula is sufficient to obtain the integral representations; this is how it was used originally by Spitzer

[99] to derive the Laplace-Stieltjes transform of the maxima of partial sums. He also related these transforms to the Wiener-Hopf problem (see in addition Cohen [40], Ch. II.5, for the relation with the Wiener-Hopf equation as it appears in Probability Theory). For the derivation of the length of an excursion above the starting level, there is more information needed, namely, that given by the partial sums $\sum_{i=1}^{n} B_i$ *together* with the partial sums of the embedded random walk $S_n = \sum_i^n B_i - \sum_i^n A_i$. It is possible to derive the excursion lengths still using Hewitt's formula when the A_i's are independent of the B_i's, and this was carried out in Kingman [74]. It is shown in the present chapter that if one extends Hewitt's inversion formula, a similar derivation is possible for the case when there is arbitrary correlation inside the vectors (A_i, B_i), which means the random walk S_n can have generally distributed increments.

Hewitt's approach was to find a most general inversion identity for Laplace-Stieltjes transforms with a view on Harmonic Analysis; this is much more than we need for our purposes. Instead of trying to find a most general instance of inversion, the focus in the present chapter is to obtain a sufficiently broad result to apply to the kind of random walks that appear in the study of workload/insurance related problems. One can only hope then that the result itself will find applications in other related areas of probability and statistics.

In the queueing literature, Conolly obtained (the transform of) the busy period together with the number of customers served in an Erlang queue with independent exponential inter-arrivals [46], and he obtained continued fraction expansions for the Erlang queue with state dependent parameters of both the exponential inter-arrivals and the service times in [47]. Conolly's results from [46] were then extended to general independent inter-arrivals and service times in Finch [58] and in Kingman [74].

The time to ruin has been studied in the insurance literature by deriving recursion formulae, typically obtained by discretising the claim sizes. For example, Dickson and Waters [50] present various approximation methods for its numerical computation. Studying the time to ruin is analytically the same as inferring on the busy period, by virtue of the alternative form of duality, see the end of Section 1.3. Another good reference is Prabhu [93] §3, who obtained an integral equation for the time to ruin starting from a positive capital u, in the Cramer-Lundberg risk reserve model; Borovkov and Dickson [32] obtain series representations for the distribution of the time to ruin in the Sparre-Andersen risk reserve model with exponentially distributed claims and general renewal inter-arrivals. Besides these results, there exists a significant amount of literature on the Gerber-Shiu functions which contain the time to ruin as a special case. We will come back to the problem of analyzing the time to ruin in Section 4.4.

The chapter is organized in the following way: In Section 4.1, we extend Hewitt's inversion formula to allow for probability distribution functions on $\mathbb{R}^2$ which do not have a product form (see Remark 4.1.1). One of the ingredients of the proof consists of having a precise meaning for the conditional distribution of B given A; this is settled as a preliminary. The approach used for studying fluctuations of random walks involves obtaining integral representations for the above-mentioned quantities. The busy period, idle period, transient workload, and the related insurance functionals can still be determined in the form of a Cauchy integral, once Theorem 4.1.1 is combined with a version of Spitzer-Baxter's identity (Proposition 4.2.1), and this is carried out in detail in Section 4.2 for the correlated GI/G/1 queue and the Sparre-Andersen risk reserve process. Roughly speaking, all the transforms of the relevant performance measures are obtained by reading the inversion formula in Theorem 4.1.1 from right to left. The combination of Spitzer-Baxter's identity with Hewitt's inversion has been used by Kingman [74] to determine busy periods when inter-arrivals are independent from service requirements.

Having obtained a Cauchy integral representation for the busy period, one can then evaluate it when the transform of the generic pair (B, A) is a rational function in the argument that corresponds to the service requirement B (Section 4.3). Finally, we point out in Section 4.4 how one can obtain the distribution of the time to ruin in the dual risk setting, when starting with a non-negative initial capital.

4.1 On Hewitt's inversion formula

The starting point is P. Lévy's inversion formula which gives a precise form to the well known assertion that a characteristic function uniquely determines a probability measure ϕ on the real line:

$$\lim_{T\to\infty} \frac{1}{2\pi i} \int_{-iT}^{iT} \left\{ \int_{-\infty}^{\infty} e^{\xi x} \phi(\mathrm{d}x) \right\} \frac{e^{-\xi a} - e^{-\xi b}}{\xi} \,\mathrm{d}\xi = \frac{1}{2}\phi((a,b)) + \frac{1}{2}\phi((a,b]).$$

Thus the value $\phi((a,b]) := \int \chi_{(a,b]}(u)\,\phi(\mathrm{d}u)$ can be recovered, $\chi_{(a,b]}$ being the indicator function of the interval $(a, b]$. Hewitt's formula extends this result to recover directly functionals of the form

$$\phi(f) := \frac{1}{2} \int [f(u+) + f(u-)]\phi(\mathrm{d}u),$$

for functions f of bounded variation; if f is also continuous, then the integral above becomes $\int f(u)\,\phi(\mathrm{d}u)$.

The inversion formula in the form given by Lévy is further generalized to higher dimensions and some other topological groups in Hewitt [69]. It is not however the purpose of the current chapter to explore the possibility of a most general form for it. Such an attempt may not even yield satisfactory results, as was already pointed out in [69].

We will generalize the above in Theorem 4.1.1 to probability measures on $\mathbb{R}^2$ related to the random vector (B, A). The probabilistic structure is never really lost, the conditional distribution of B given A appears throughout the proof of Theorem 4.1.1 disguised as the conditional kernel $q(u, y)$, which is defined below as a "partial" Radon-Nikodým derivative.

4.1.1 Preliminaries

We say that $f : \mathbb{R} \to \mathbb{C}$ is of bounded variation if both its real and imaginary parts are of bounded variation; this is the same as $|f|$ being of bounded variation because all norms are equivalent on $\mathbb{C}$. We will also work with (complex-valued) measures on the Borel subsets of $\mathbb{R}^2$, which are of finite total variation. For such a measure ϕ, we will denote by $|\phi|$ the total variation measure of ϕ, and with $\|\phi\|$ its total variation. This will suffice for our purposes, but this set-up is fully detailed and generalized in [69], and the references therein.

Let (B, A) be a random vector on some probability space, having an arbitrary probability distribution. Denote by $\mathbb{P}$ the probability measure and by H the joint c.d.f. of (B, A):

$$H(x, y) = \mathbb{P}(B \leq x,\ A \leq y).$$

The correlation device for the increment of the random walk is given by this distribution, not necessarily having a product form, and its Fourier-Stieltjes transform

$$h(s_1, s_2) := \mathbb{E}e^{-s_1 B - s_2 A} = \int e^{-s_1 x - s_2 y}\, H(\mathrm{d}x, \mathrm{d}y).$$

In general this is convergent only for $\mathcal{R}e\, s_1 = 0$, $\mathcal{R}e\, s_2 = 0$, but if (B, A) is supported on the non-negative quadrant in $\mathbb{R}^2$, then h can be continued analytically to $\mathcal{R}e\, s_1 \geq 0$, $\mathcal{R}e\, s_2 \geq 0$. The characteristic function of the increment $A - B$ will also be relevant:

$$h(\xi, -\xi) = \int e^{\xi x}\, \mathrm{d}\mathbb{P}(A - B \leq x),\ \ \mathcal{R}e\, \xi = 0.$$

Let λ be the probability measure associated with the random vector (B, A), and ν be the marginal measure associated with B,

$$\lambda(\mathcal{U} \times \mathcal{V}) := \mathbb{P}(B \in \mathcal{U}, A \in \mathcal{V}), \quad \nu(\mathcal{U}) := \lambda(\mathcal{U} \times \mathbb{R}), \quad \mathcal{U}, \mathcal{V} \in \mathcal{B}(\mathbb{R}),$$

with $\mathcal{B}(\mathbb{R})$ the family of Borel sets on the line. The notation $H(\mathrm{d}u, y)$ will be used to suggest that we are integrating w.r.t. the measure $\lambda_y(\mathcal{U}) := \lambda(\mathcal{U} \times (-\infty, y])$. We will work with a version of the *conditional* cumulative distribution function (c.d.f.) of A given B and this is made precise below.

It clearly holds that $\lambda_y(\mathcal{U}) \leq \nu(\mathcal{U})$, in particular $\lambda_y \ll \nu$, so let

$$q(u, y) := \frac{\mathrm{d}\lambda_y}{\mathrm{d}\nu}(u)$$

be its Radon-Nikodým derivative. Heuristically, $q(u, y)$ is to be regarded as $q(u, y) = \mathbb{P}(A \leq y \,|\, B \in \mathrm{d}u)$, and we have the disintegration identity

$$\int_{\mathcal{U}} H(\mathrm{d}u, y) \equiv H(\mathcal{U}, y) \equiv \int_{\mathcal{U}} q(u, y)\, \nu(\mathrm{d}u),$$

with any of the terms above meaning $\mathbb{P}(A \leq y,\ B \in \mathcal{U})$. We will be working with a *regular version* of q, which exists by virtue of the separability of $\mathbb{R}$, see for instance Kallenberg [72] Thm. 5.3, p. 84. Further, we have more than just regularity for this kernel, the same result gives that $q(u, y)$ is *regularly monotone* as a function in the argument y, i.e., $q(u, y)$ is non-decreasing in y outside a set of $\nu-$measure zero which does not depend on y.

These considerations are quite intuitive because of the probabilistic nature of the measure associated with H. It turns out, however, that we will have to consider instances of the inversion theorem for the slightly more general case of complex valued functions H which are also of bounded variation, and for this purpose we will show below that $H(\mathrm{d}u, x)$ can be given a meaning in an analogous way.

Confusing H with its associated complex-valued measure, see Hewitt [69], we can reduce it to a monotonically increasing function, by splitting into real and imaginary parts and using the Jordan decomposition

$$\mathcal{R}e\, H \equiv (\mathcal{R}e\, H)^+ - (\mathcal{R}e\, H)^-,$$

for the signed measure $\mathcal{R}e\, H$, and similarly for $\mathcal{I}m\, H$. Setting $\nu^\pm(\mathcal{U}) := (\mathcal{R}e\, H)^\pm(\mathcal{U} \times \mathbb{R})$, it holds with the similar notation as in the probabilistic case that $(\mathcal{R}e\, H)_y^\pm \ll \nu^\pm$, so we can define again the Radon-Nikodým derivatives

$$q_1^+(u, y) := \frac{\mathrm{d}(\mathcal{R}e\, H)_y^+}{\mathrm{d}\nu^+}(u), \;\; q_1^-(u, y) := \frac{\mathrm{d}(\mathcal{R}e\, H)_y^-}{\mathrm{d}\nu^-}(u),$$

and similarly for $\mathcal{I}m\, H$. Now we can reconstruct $H(\mathrm{d}u, y)$ in an obvious way, using the linearity of the Radon-Nikodým derivative.

Alternatively, we could have used the total variation measures:

$$|\mathcal{R}e\, H_y|(\mathcal{U}) \leq |\mathcal{R}e\, H|(\mathcal{U} \times \mathbb{R})$$

and thus $\mathcal{R}e\, H_y \ll \nu$, but a simple argument relying on the Hahn-Jordan decomposition shows that this construction yields the same result for $H(\mathrm{d}u, x)$.

Moreover, the monotone regularity property from the probabilistic instance extends to $q(u, y)$ being of bounded variation in y outside a set of ν-measure zero which does not depend on y. This property will be useful in the proof of the next result.

Theorem 4.1.1 (Generalized Hewitt inversion). *Let H be a totally bounded (complex-valued) measure on $\mathbb{R}^2$, and let $f : \mathbb{R} \to \mathbb{C}$ be a function of bounded variation which is also absolutely integrable (w.r.t. the Lebesgue measure). Then the following Cauchy principal value can be represented as a Lebesgue-Stieltjes integral:*

$$\lim_{T \to \infty} \frac{1}{2\pi i} \int_{-iT}^{iT} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\xi(u-y)} H(\mathrm{d}u, y) f(y) \mathrm{d}y \right\} \mathrm{d}\xi$$
$$= \frac{1}{2} \int_{-\infty}^{\infty} \{ f(u+) H(du, u+) + f(u-) H(du, u-) \} .$$

f need not be integrable w.r.t. $H(\mathrm{d}u, u)$. If one of the sides above converges, so does the other one.

Remark 4.1.1. *If H is of the form $H_1 H_2$, the double integral inside the Cauchy principal value factorizes into the Fourier-Stieltjes transform of H_1 and the Fourier transform of $f(y) H_2(y)$, this function being again absolutely integrable and of bounded* *variation. Thus the above reduces to the inversion formula in Hewitt [69], Thm. (3.1.1):*

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{-iT}^{iT}\left\{\int_{-\infty}^{\infty} e^{\xi u}H_1(\mathrm{d}u)\right\}\left\{\int_{-\infty}^{\infty} e^{-\xi y}g(y)\,\mathrm{d}y\right\}\mathrm{d}\xi=\frac{1}{2}\int_{-\infty}^{\infty}[g(u+)+g(u-)]\,H_1(du),$$

with $g(y)=f(y)H_2(y)$.

Proof of Theorem 4.1.1. For fixed u, change the variable $x:=u-y$, so that the double integral inside the Cauchy principal value becomes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\xi(u-y)}H(\mathrm{d}u,y)f(y)\,\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\xi x}H(\mathrm{d}u,u-x)f(u-x)\,\mathrm{d}x.$$

We can bound

$$\int \mathrm{d}x\left|\int H(\mathrm{d}u,u-x)f(u-x)\right|\le\int\mathrm{d}x\int|f(u-x)|\,|H|(\mathrm{d}u,\infty)=\|H\|\int|f(v)|\,\mathrm{d}v,$$

hence $\int H(\mathrm{d}u,u-x)f(u-x)$ is absolutely integrable in x because f is. We can now change the order of integration in the Cauchy principal value, which becomes after integrating over ξ:

$$\lim_{T\to\infty}\iint\frac{1}{2\pi i}\int_{-iT}^{iT}e^{\xi x}\,\mathrm{d}\xi\,H(\mathrm{d}u,u-x)f(x)\mathrm{d}x=\lim_{T\to\infty}\int\frac{\sin Tx}{\pi x}\mathrm{d}x\int H(\mathrm{d}u,u-x)f(u-x).\tag{4.1}$$

At this point, disintegrate the kernel $H(\mathrm{d}u,u-x)=q(u,u-x)\,\nu(\mathrm{d}u)$, so that, by the Radon-Nikodým Theorem, we can rewrite the right-hand side in (4.1) as

$$\lim_{T\to\infty}\int\nu(\mathrm{d}u)\left\{\int\frac{\sin Tx}{\pi x}q(u,u-x)f(u-x)\,\mathrm{d}x\right\}.\tag{4.2}$$

The main remark is that $q(u,u-x)f(u-x)$ is of bounded variation as a function in x for ν-almost all u, but regularity is the key, as can be seen from the following:

Let $\{x_i\}_{i\in I}$ be some ordered sequence determined by the edges of an interval partition of $\mathbb{R}$. Since f is assumed of bounded variation, it can be reduced to a real valued and monotonically increasing function, again by splitting it into its real and imaginary parts and making use of Jordan's decomposition. Similarly, $q(u,\cdot)$ can be reduced to a regularly non-decreasing function. Then we can write by exploiting the monotonicity of both $q(u,\cdot)$ and f:

$$\sum_{i=1}^{m}[q(u,u-x_i)f(u-x_i)-q(u,u-x_{i+1})f(u-x_{i+1})]$$

$$\leq \sum_{i=1}^{\infty} q(u, u - x_i)[f(u - x_i) - f(u - x_{i+1})]$$
$$\leq q(u, \infty) \sum_{i=1}^{m} [f(u - x_i) - f(u - x_{i+1})], \quad u \notin \mathcal{Q},$$

where, by virtue of the regularity of q, $\mathcal{Q}$ is a $\nu-$negligible set, outside which $q(u, \cdot)$ is monotone.

Notice that in the upper bound the arguments of q do not depend anymore on the partition $\{x_i\}_i$. Even more so, $q(u, \infty) = 1$ $\nu-$a.e. by definition. Since by regularity, $\mathcal{Q}$ does not depend on the sequence $(x_i)_i$, we can take the supremum over all such sequences and using that f is of bounded variation, gives that also $q(u, u - x)f(u - x)$ is of bounded variation for $\nu-$almost all u.

We have arrived at the following limit:

$$\lim_{T \to \infty} \int_{-\infty}^{\infty} \frac{\sin Tx}{\pi x} \varphi(u, x) \, \mathrm{d}x = \frac{1}{2}[\varphi(u, 0+) + \varphi(u, 0-)],$$

for fixed u and $\varphi(u, x) := q(u, u - x)f(u - x)$. This identity is known as Dirichlet's integral. The integrability condition for the left-hand side that assures the limit exists is that $\varphi(u, \cdot)$ be of bounded variation. As seen from the above, this assumption is only slightly more general than Dirichlet's original condition of monotonicity for $\varphi(u, \cdot)$. See Doetsch [51] Ch. 24, or Titchmarsh [102] §13.2 (the condition is also known as Jordan's test). The limit equals what is usually called the normalized function in the origin.

The proof will be complete as soon as we show that the limit in T can be taken inside the $\nu(\mathrm{d}u)$ integral (4.2). Reduce f again to a non-negative and monotonically increasing function. Using the second mean value theorem for $\varphi(u, x)$ which is now decreasing in x, for some $\beta > 0$, we can find $\alpha > 0$ such that

$$\left| \int_0^{\beta} \varphi(u, x) \frac{\sin Tx}{x} \mathrm{d}x \right| \leq \left| \varphi(u, 0) \int_0^{\alpha} \frac{\sin Tx}{x} \, \mathrm{d}x \right| < \varphi(u, 0) \int_0^{\pi} \frac{\sin x}{x} \, \mathrm{d}x.$$

The final upper bound is obtained by a change of variable $x \to x/T$. This is the same argument as given in Hewitt [69], and it is clearly sufficient to allow the interchanging of limit and integration in (4.2); this together with the fact that Dirichlet's integral identity holds for ν-almost all u, completes the proof.

4.2 The GI/G/1 queue with correlations

Having laid down the inversion result in the previous section, let us start with the study of the special queueing system GI/G/1 briefly described in the introduction.

It is assumed that the law of the random walk $\{S_n\}_{n \geq 0}$

$$S_n = S_0 + \sum_{i=1}^{n} (B_i - A_i)$$

is the law $\mathbb{P}$ conditional on $S_0 = 0$. Also set $b_n = \sum_{i=1}^{n} B_i$ and $a_n = \sum_{i=1}^{n} A_i$.

Let N be a random variable which is distributed as the number of customers served during a busy cycle of the server. Then (assuming unit server speed) b_N, a_N, $-S_N$ stand for the length of the busy period, that of the busy cycle and respectively the idle time of the server.

In queueing terms, one can think of the pair (B, A) as the service time of a generic customer together with the time until the arrival of the next customer in a GI/G/1 queueing system - which can be always normalized to unit server speed without losing generality. In terms of insurance and risk theory, this pair can be interpreted as the time elapsed (and hence the premium B gained) since the last claim incurred together with the amount A claimed through an insurance policy. Then, conditional on starting with 0 initial capital, b_N is the time to ruin (after normalizing the risk reserve process in order to have unit income rate), a_N is the total amount claimed until ruin (including the claim that causes it) and $-S_N$ is the deficit at ruin.

The representation given below was obtained in Wendel [103] §4 as an algebraic identity, slightly more general than Spitzer's identity, who originally obtained in [99] the representation for the LST of the successive maxima of the partial sums $(S_n)_n$ (see also Baxter and Donsker [28] for a similar derivation that holds for Lévy processes). Theorem 1 and Identity (9) in Kingman [74] are much closer to our purposes. For the sake of completeness, we cite the relevant result in the following proposition

Proposition 4.2.1 (Spitzer, Wendel, Baxter). *With the above notations, it holds that*

$$\mathbb{E}\{z^N e^{-s_1 b_N - s_2 S_N}\} = 1 - \exp\left\{ -\sum_{n=1}^{\infty} \frac{z^n}{n} \mathbb{E}[e^{-s_1 b_n - s_2 S_n} 1_{\{S_n < 0\}}] \right\}, \tag{4.3}$$

which holds for $\mathcal{R}e\, s_1 \geq 0$, $\mathcal{R}e\, s_2 \leq 0$, $|z| \leq 1$.

In Kingman [74], this identity is obtained by 'killing' inside the Spitzer-Wendel identity, i.e. in the three-dimensional space where $(a_n, b_n, S_n)_n$ is evolving, replace the reflecting hyperplane at $x_3 = 0$ with an absorbing hyperplane. This is formally carried out in [74] by replacing the projection operator used by Wendel [103] with an absorption operator.

We will use Theorem 4.1.1 in conjunction with the version of Spitzer's identity from Proposition 4.2.1 to obtain integral representations for the transforms of the busy period, idle period and the number of customers served during a busy period.

Throughout the rest of this chapter, we will use the dashed integral sign as a replacement for the cumbersome limit

$$\lim_{T\to\infty} \int_{-iT}^{iT} \equiv \fint_{-i\infty}^{i\infty} .$$

Theorem 4.2.1. *We have the following integral representations for* $P := b_N$, $I := -S_N$, *valid whenever* $\mathcal{R}e\, s > 0$, $|z| < 1$*:*

$$\mathbb{E}\{z^N e^{-sP}\} = 1 - \exp\left\{ \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s-\xi} \{\log[1 - zh(\xi, 0)] - \log[1 - zh(\xi, s-\xi)]\} \right\}, \tag{4.4}$$

$$\mathbb{E}\{z^N e^{-sI}\} = 1 - \exp\left\{ \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s+\xi} \log[1 - zh(\xi, -\xi)] \right\}. \tag{4.5}$$

Here log *is the principal branch, that has the cut taken along the negative real axis between* 0 *and* ∞, *so that it admits the power series representation*

$$\log \frac{1}{1-r} = \sum_{n=1}^{\infty} \frac{r^n}{n}, \quad |r| < 1.$$

Proof. Below we will use an integration by parts argument and for this reason it will be convenient to introduce the function $G(u, x) = \mathbb{P}(B \le u, A > x)$. We can write by integrating over the possible values of b_n:

$$\mathbb{P}(b_n < a_n) = \int_0^{\infty} G^{*n}(\mathrm{d}x, x) \quad \text{and} \quad \mathbb{P}(b_n \le a_n) = \int_0^{\infty} G^{*n}(\mathrm{d}x, x-)$$

where $G^{*n}(x, y) := \mathbb{P}(b_n \le x,\, a_n > y)$, is the n-fold convolution of G with itself and $G^{*n}(\mathrm{d}x, y)$ is the associated integral kernel, as described in Section 4.1,

$$G^{*n}(\mathrm{d}x, y) \equiv \mathbb{P}(b_n \in \mathrm{d}x,\, a_n > y), \quad G^{*n}(\mathrm{d}x, y-) \equiv \mathbb{P}(b_n \in \mathrm{d}x,\, a_n \ge y).$$

Let us begin with (4.4), which means we start with (4.3) for $s_2 = 0$ and $s_1 = s$. The first step is to represent the expected value inside the series in (4.3):

$$\frac{1}{2}\mathbb{E}[e^{-sb_n}(1_{\{S_n \le 0\}} + 1_{\{S_n < 0\}})] = \int_0^{\infty} e^{-sx}[\frac{1}{2}G^{*n}(\mathrm{d}x, x-) + \frac{1}{2}G^{*n}(\mathrm{d}x, x+)]. \tag{4.6}$$

Define the following z-harmonic measure associated to $G(\mathrm{d}x, \mathrm{d}y) = -H(\mathrm{d}x, \mathrm{d}y)$:

$$H_z^*(\mathrm{d}x, \mathrm{d}y) = \sum_{n=1}^{\infty} \frac{z^n}{n} G^{*n}(\mathrm{d}x, \mathrm{d}y),$$

so that, in particular,

$$H_z^*(\mathrm{d}x, y) = \sum_{n=1}^{\infty} \frac{z^n}{n} G^{*n}(\mathrm{d}x, y). \tag{4.7}$$

H_z^* is a complex valued measure proper (i.e. it has finite total variation) for $|z|<1$, since we have

$$||H_z^*|| \le \sum_{n=1} \frac{|z|^n}{n} = \log \frac{1}{1-|z|};$$

moreover, its LST equals

$$\int e^{-s_1 x - s_2 y} H_z^*(\mathrm{d}x, \mathrm{d}y) = \sum_{n=1}^{\infty} \frac{z^n}{n} h^n(s_1, s_2), \tag{4.8}$$

where the interchanging of the integral with the series is allowed because of absolute integrability:

$$\iint |e^{-s_1 x - s_2 y}|\, |H_z^*|(\mathrm{d}x, \mathrm{d}y) \le \sum_{n=1}^{\infty} \frac{|z|^n}{n} \iint G^{*n}(\mathrm{d}x, \mathrm{d}y) = \log \frac{1}{1-|z|}.$$

Now let us use Theorem 4.1.1 for $f(y) = e^{-sy}\chi_{[0,\infty)}(y)$, $\mathcal{R}e\, s > 0$, which means we can write via (4.6) and (4.7):

$$\sum_{n=1}^{\infty} \frac{z^n}{n} \mathbb{E}[e^{-sb_n}(1_{\{S_n \le 0\}} + 1_{\{S_n < 0\}})] = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \left\{ \int_0^{\infty} \int_0^{\infty} e^{\xi x - (s+\xi) y} H_z^*(\mathrm{d}x, y)\, \mathrm{d}y \right\} \mathrm{d}\xi. \tag{4.9}$$

Assume for simplicity that $\mathbb{P}(A = B) = 0$ (i.e. $\mathbb{P}(S_n = 0)$ is null for all n). This assumption is not essential, see for example the discussion in Cohen [40], p. 284. Then the normalized indicator function $\frac{1}{2}(1_{\{S_n \le 0\}} + 1_{\{S_n < 0\}})$ that appears on the left-hand side of (4.6) simplifies to $1_{\{S_n < 0\}}$. Since $H_z^*(\mathrm{d}x, y)$ is of bounded variation in y, we can use the integration by parts formula for Lebesgue-Stieltjes integrals, so that (4.9) becomes

$$\begin{aligned} &\frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s+\xi} \left\{ \int_0^{\infty} e^{\xi x} \left[H_z^*(\mathrm{d}x, 0-) + \int_{y=0}^{\infty} e^{-(s+\xi)y} H_z^*(\mathrm{d}x, \mathrm{d}y) \right] \right\} \\ &= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s+\xi} \left[\sum_{n=1}^{\infty} \frac{z^n}{n} h^n(-\xi, 0) - \sum_{n=1}^{\infty} \frac{z^n}{n} h^n(-\xi, s+\xi) \right], \end{aligned} \tag{4.10}$$

where we used (4.8) and the identity $G(x, 0-) = \mathbb{P}(B \le x)$. Changing the variable $\xi \to -\xi$, the exponent in (4.3) can be rewritten via (4.10), for $\mathcal{R}e\, s > 0$, $|z| < 1$:

$$-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathbb{E}[e^{-sb_n} 1_{\{S_n < 0\}}] = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s-\xi} \left\{ \log[1 - zh(\xi, 0)] - \log[1 - zh(\xi, s-\xi)] \right\}. \tag{4.11}$$

Thus (4.4) follows from these considerations and Spitzer's identity (4.2.1).

For the integral representation (4.5), the extension of Hewitt's formula is not needed. The starting point is (4.3) with $s_1 = 0$, $s_2 = -s$, $\mathcal{R}e\, s > 0$, together with the identity

$$\mathbb{E}e^{sS_n} 1_{\{S_n<0\}} = \frac{1}{2}\int_{-\infty}^{0+} e^{sx}\mathrm{d}\mathbb{P}(S_n \le x) + \frac{1}{2}\int_{-\infty}^{0-} e^{sx}\mathrm{d}\mathbb{P}(S_n \le x)$$

valid because $\mathbb{P}(S_1 = 0) = 0$. Similarly as above, set $F(x) = \mathbb{P}(S_1 \le x)$ and introduce

$$F_z^*(\mathrm{d}x) := \sum_{n=1}^{\infty} \frac{z^n}{n} F^{*n}(\mathrm{d}x). \tag{4.12}$$

Use the inversion formula in Remark 4.1.1 with $g(y) = e^{sy}\chi_{(-\infty,0]}(y)$, $\mathcal{R}e\, s > 0$, so that we can write similarly as for (4.9)-(4.10):

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{z^n}{n}\mathbb{E}e^{sS_n} 1_{\{S_n<0\}} &= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \left\{ \int_{-\infty}^{\infty} e^{\xi x} F_z^*(\mathrm{d}x)\right\}\left\{\int_{-\infty}^{0} e^{(s-\xi)y}\,\mathrm{d}y\right\} \mathrm{d}\xi \\
&= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s-\xi} \sum_{n=1}^{\infty} \frac{z^n}{n} h^n(-\xi,\xi) \\
&= -\frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s+\xi} \log[1 - zh(\xi,-\xi)], \quad \mathcal{R}e\, s > 0,
\end{aligned} \tag{4.13}$$

after the change of variable $\xi \to -\xi$.

Once (4.13) is replaced into (4.3), it immediately yields (4.5). The proof is complete.

Remark 4.2.1. *If B is independent of A, then $h(s_1, s_2)$ is the product of the marginal transforms and the integral representation (4.4) reduces to that from Kingman [74], Thm. 4 (see also Cohen [40], p. 304 for (4.5)).*

Remark 4.2.2. *The integral representations (4.4) and (4.5) hold under very general conditions (there are no regularity assumptions of the distribution of (B, A), these can even be discrete random variables, in which case the LSTs become generating functions). The reason is that these representations are given for the interior of their convergence domains ($\mathcal{R}e\, s>0$, $|z|<1$). If we want to take any of the arguments to their respective boundary, we have to require extra conditions to ensure convergence. For example, when letting s converge towards the imaginary axis, there is a singularity appearing, because the factor $1/(\xi - s)$ gains a simple pole located at $\xi = s$.*

It turns out one can give a definite meaning to these integrals, for $\mathcal{R}e\, s = 0$, if they are regarded as singular integrals w.r.t. the Cauchy kernel $1/(\xi - s)$. Then one considers the Cauchy principal value obtained by removing a circle of arbitrarily small radius around the singularity s and then taking its radius to 0. Now we are dealing *with a double principal value: first one coming from the pole at s of the Riemann integral along the segment $L := [-iT, iT]$ (T large enough, so that $s \in L$) and the second one obtained by letting $T \to \infty$. A standard condition (see Gakhov [65], Muskhelishvili [87]) that ensures the first principal value converges is that the density functions $\varphi_1(\xi) := \log[1-h(\xi,-\xi)]$, $\varphi_2(s,\xi) := \log[1-h(\xi, s-\xi)]$ are Hölder continuous along the imaginary axis, with some positive indices. The Hölder continuity of $\varphi_1(\xi)$ is fairly close to Spitzer's [98] integrability condition, which requires (upon taking $s \to 0$) that $(1-h(\xi,-\xi))/\xi$ be integrable in a neighbourhood of $s=0$ on the imaginary axis.*

4.2.1 The number of arrivals during an excursion

Further along the lines of Remark 4.2.2, we will use the doubly dashed integral sign to denote the double Cauchy principal value. The choice for the branch of log is essential for the definiteness of the first principal value. It turns out to be convenient to work with a branch of log which has the cut between 0 and ∞ taken inside the negative half-plane (the same principal branch as the one used in Theorem 4.2.1), so we have by definition,

$$\frac{1}{2\pi i}\ \ {\int\!\!\!\!\!\!-}_{L} \frac{\varphi(\xi)}{\xi - s}\mathrm{d}\xi := -\frac{1}{2}\varphi(s) + \frac{\varphi(s)}{2\pi i}[\log(iT-s) - \log(-iT-s)] + \frac{1}{2\pi i}\int_L \frac{\varphi(\xi)-\varphi(s)}{\xi - s}\mathrm{d}\xi. \tag{4.14}$$

The integral on the right is well defined as a Riemann integral as soon as $\varphi(s)$ is Hölder continuous along the line L. By choice of logarithm, the argument of $\log(iT-s) - \log(-iT-s)$ equals πi for all T. This means that the first two terms cancel in the limit $T \to \infty$, and the above definition becomes

$$\frac{1}{2\pi i}\ \ {\int\!\!\!\!\!\!=}_{-i\infty}^{i\infty} \frac{\varphi(\xi)}{\xi - s}\mathrm{d}\xi = \frac{1}{2\pi i}\ \ {\int\!\!\!\!\!\!-}_{-i\infty}^{i\infty} \frac{\varphi(\xi)-\varphi(s)}{\xi - s}\mathrm{d}\xi. \tag{4.15}$$

Formula (4.14) differs slightly from the definition given in Mushkelishvili [87], p. 27 or Gakhov [65], p. 16, because therein the cut of the logarithm is taken in the opposite half-plane. To be more precise, in our case, Cauchy's integral representation reads:

$$\ \ {\int\!\!\!\!\!\!=}_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{\xi - s} = \begin{cases} 2\pi i, \ s \in L^+, \\ 0, \ s \in L^-, \\ 0, \ s \in L. \end{cases}$$

The first two values are the well known Cauchy integral identities; the third one is the Cauchy principal value for this specific choice of log (use (4.15) with $\varphi \equiv 1$).

It helps to think about the Riemann sphere as the one point compactification of the complex plane so that the imaginary axis is closed into a large circle on the sphere. Then all of the conventions above are specifying a means of integrating on the large circle of the sphere, the integrands being extended by continuity at infinity; the interior of the imaginary axis is by definition the left hemisphere and the exterior is the right hemisphere. Define

$$\Phi(s) = \int_{-i\infty}^{i\infty} \frac{\varphi(\xi)}{\xi - s} \mathrm{d}\xi, \quad s \in \mathbb{C},$$

and the integral is defined in the sense of (4.15), when $\mathcal{R}e\, s = 0$. If we denote by $\Phi^-(s)$ the limit as s approaches the imaginary axis from its exterior, and by $\Phi^+(s)$, the limit taken from the interior, then the Plemelj–Sokhotski formulae (cf. Gakhov [65], p. 25) become with the above conventions:

$$\Phi^-(s) = \Phi(s), \quad \Phi^+(s) = \Phi(s) + \varphi(s), \quad \mathcal{R}e\, s = 0. \tag{4.16}$$

Now we can calculate the limit as s approaches the imaginary axis from its exterior in (4.4), and from its interior in (4.5), using (4.16). Still denoting the interior limit with $\Phi^+(s)$, (4.5) becomes for $|z| < 1$ and as s tends to the imaginary axis:

$$\Phi^+(s) = 1 - [1 - zh(-s, s)] \exp\left\{ \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{\xi + s} \log[1 - zh(\xi, -\xi)] \right\}. \tag{4.17}$$

Before we give the limiting values of (4.4), we point out how one can simplify it. Consider the first integral in the exponent of (4.4):

$$\Psi(\eta) := -\frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{\xi - \eta} \log[1 - zh(\xi, 0)], \quad \mathcal{R}e\, \eta > 0.$$

The key remark is that $h(\xi, 0) = \mathbb{E}e^{-\xi B}$ is analytic for $\mathcal{R}e\, \xi > 0$ and this implies that the density $\log[1 - zh(\xi, 0)]$ is again analytic for $\mathcal{R}e\, \xi > 0$. Since the integrand has a simple pole in the positive half-plane located at $\eta = \xi$, it follows at once from Cauchy's Theorem applied to the imaginary axis that $\Psi(\eta)$ equals

$$\Psi(\eta) = \log[1 - zh(\eta, 0)],$$

for any $\mathcal{R}e\, \eta > 0$. This can be seen by closing the segment $[-iT, iT]$ with the half circle spanning between its endpoints inside the positive half-plane. The contour integral thus obtained equals $\log[1 - zh(s, 0)]$ for T large enough, so that the pole $\eta = \xi$ lies inside the contour (remark that because of the conventions on the interior of the imaginary axis, this contour is traversed in the clockwise direction). Finally, the contribution along the half-circle tends to 0 as $T \to \infty$ because the integrand behaves as $o(|\xi|^{-1})$ along the half-circle.

Having settled the first term in the exponent of (4.4), we can use the Plemelj–Sokhotski's formula (the continuity of the exterior limit from (4.16)) for the other term, to obtain the identity

$$\Phi^-(s) = 1 - \exp\left\{\log[1 - zh(s,0)] + \frac{1}{2\pi i} \fint_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{\xi - s} \log[1 - zh(\xi, s-\xi)]\right\}, \; \mathcal{R}e\, s = 0, \tag{4.18}$$

hence we can rewrite (4.18) as

$$\Phi^-(s) = 1 - [1 - zh(s,0)] \exp\left\{\frac{1}{2\pi i} \fint_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{\xi - s} \log[1 - zh(\xi, s-\xi)]\right\}, \quad \mathcal{R}e\, s = 0. \tag{4.19}$$

In particular, for $s = 0$, the two limits (4.17) and (4.19) agree and these must then coincide with the generating function of N. The following has been proven

Proposition 4.2.2. *With the above notations and conventions, it holds for $|z| < 1$:*

$$\mathbb{E}z^N = 1 - (1-z) \exp\left\{\frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{\xi} \log \frac{1 - zh(\xi, -\xi)}{1-z}\right\}.$$

4.3 Examples

In this section we point out that the integral representation (4.4) can be explicitly evaluated under the assumption that the transform of the generic pair (B, A) is a rational function in the argument that corresponds to the service requirement B.

Assume that for all s_2, the joint LST $h(s_1, s_2)$ is a rational function in the argument s_1, which can be represented as

$$h(s_1, s_2) = \frac{h_1(s_1, s_2)}{h_2(s_1, s_2)}, \tag{4.20}$$

where $h_i(\cdot, s_2)$ are polynomial functions. Moreover we will assume that for $\mathcal{R}e\, s \geq 0$, $h(\xi, s-\xi)$ has a finite number of poles in the negative half-plane as a function in the argument ξ ($h(\xi, s-\xi)$ is already meromorphic in this region, because of the previous assumption). This is an algorithmically friendly assumption, which will also give a representation for the busy period transform in terms of a finite number of factors.

We may still assume without losing generality that $\mathbb{P}(B - A = 0) = 0$, which implies $h(s_1, s_2) \to 0$, as $s_1 \to \infty$, $\mathcal{R}e\, s_1 > 0$, and the convergence is uniform in s_2, for $\mathcal{R}e\, s_2 \geq 0$. In particular, we have for any $\mathcal{R}e\, s_2 \geq 0$, $\deg h_1(\cdot, s_2) < \deg h_2(\cdot, s_2)$.

Before we proceed with the analysis, let us point out some ways of creating correlation between the inter-arrivals and the corresponding service times.

Example 1 (Threshold dependence) This is one of the simplest ways of making B depend on the size of A: for a fixed threshold $l > 0$, $B \sim B_1$ on the event $A \leq l$ and $B \sim B_2$ otherwise; with B_i independent of A and having rational transforms $f_i(s_1)$, $i = 1, 2$; thus

$$h(s_1, s_2) = f_1(s_1)a_1(s_2) + f_2(s_1)a_2(s_2),$$

$$a_1(s_2) = \int_{y=0}^{l} e^{-s_2 y}\, \mathrm{d}\mathbb{P}(A \le y), \qquad a_2(s_2) = \int_{y=l+}^{\infty} e^{-s_2 y}\, \mathrm{d}\mathbb{P}(A \le y),$$

so that $a_1(s_1)$ is an entire function and $a_2(s_2)$ is analytic for $\mathcal{R}e\, s_2 > 0$. This construction can be naturally extended to k thresholds, giving

$$h(s_1, s_2) = \sum_{i=1}^{k} f_i(s_1)a_i(s_2),$$

with $a_i(s_2)$ entire functions, $i < k$, and $a_k(s_2)$ analytic for $\mathcal{R}e\, s_2 > 0$.

Example 2 (Markov Modulation) This is similar to the class of examples given in Lemma 3.4.1,

$$(A, B) = \sum_{i=1}^{\kappa} (A_i, B_i),$$

with κ the number of jumps until absorbtion of a finite state Markov chain. The difference is that the component A_1 is now allowed to be generally distributed with $g_0(s_2) = \mathbb{E}e^{-s_2 A_1}$, and B_1 has a rational transform of the form $f_1(s_1)/f_2(s_1)$. With the same notations for the transition structure of the absorbing Markov chain as in Section 3.4, the transform of (A, B) is

$$h(s_1, s_2) = \boldsymbol{\alpha}' \left[\frac{f_2(s_1)}{f_1(s_1)g_0(s_2)} I - T \right]^{-1} \boldsymbol{t}.$$

Both these examples are of the form assumed by (4.20) (see the proof of Lemma 3.4.1, for the second example).

Remark that $h_2(\cdot, s_2)$ can only have zeroes with negative real part, because of the regularity domain of $h(\cdot, s_2)$. With these assumptions, the exponent in (4.4) becomes

$$\frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s - \xi} \log \frac{h_2(\xi, 0) - z h_1(\xi, 0)}{h_2(\xi, 0)} - \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\mathrm{d}\xi}{s - \xi} \log[1 - z h(\xi, s - \xi)]. \quad (4.21)$$

For the principal branch of the logarithm which has the cut taken along the negative real axis, the single valued functions $\log[1 - zh(\xi, 0)]$ and $\log[1 - zh(\xi, s - \xi)]$ are *holomorphic*, for ξ lying in a neighbourhood of infinity, $\mathcal{R}e\, \xi < 0$. The reason is that for such values of ξ, $|zh(\xi, 0)| < 1$, $|zh(\xi, s - \xi)| < 1$, and then a simple geometric argument shows that both $1 - zh(\xi, 0)$ and $1 - zh(\xi, s - \xi)$ lie in the positive half-plane. With this choice of the cut, the evaluation of the integrals (4.21) becomes an application of the theorem of residues. Before we can evaluate (4.21), the zeroes and poles of the arguments of the logarithm must be localized. The following lemma will also be useful later on.

Lemma 4.3.1. *The functions $h_2(\xi, s-\xi)$ and $h_2(\xi, s-\xi) - zh_1(\xi, s-\xi)$ have the same number $n \equiv n(s)$ of zeroes in the negative half of the complex ξ-plane, either when $|z| < 1$, $\mathcal{R}e\, s \geq 0$, or $|z| \leq 1$, $\mathcal{R}e\, s > 0$.*

Assuming $\mathcal{R}e\, s \geq 0$ and $\mathbb{E}B$, $\mathbb{E}A < \infty$, then under the extra ergodicity condition $\mathbb{E}B < \mathbb{E}A$, the functions $h_2(\xi, s-\xi)$ and $h_2(\xi, s-\xi) - h_1(\xi, s-\xi)$ have the same number $m \equiv m(s)$ of zeroes with negative real part.

Proof. The statements will follow from Rouché's theorem (cf. Titchmarsh [102], p. 116) as soon as we show that these functions are analytic in the interior of some suitably chosen contours and on their boundary it holds that

$$|h_2(\xi, s-\xi)| > |zh_1(\xi, s-\xi)|, \quad |h_2(\xi, -\xi)| > |h_1(\xi, -\xi)|. \tag{4.22}$$

Fix $R > 0$ and consider the contour $\mathcal{C}$ consisting of the segment of the imaginary axis between $-iR$ and iR together with the semicircle with radius R that spans in the negative half-plane. For the segment of the imaginary axis, we have the following bounds

$$|zh(\xi, s-\xi)| = |z||\mathbb{E}e^{-\xi B-(s-\xi)A}| \leq |z|\,\mathbb{E}|e^{-\xi B-(s-\xi)A}| \leq |z|\,\mathbb{E}e^{-(\mathcal{R}e\, s)A}, \quad \mathcal{R}e\, \xi = 0. \tag{4.23}$$

For the bound on the half-circle, consider the following representation for $h(s_1, s_2)$:

$$h(s_1, s_2) = \frac{a_1(s_2)s_1^{n-1} + a_2(s_2)s_1^{n-2} + \ldots + a_n(s_2)}{b_1(s_2)s_1^n + b_2(s_2)s_1^{n-1} \ldots + b_{n+1}(s_2)},$$

where $n \equiv n(s_2)$ (remark that for the examples presented above, the denominator does not depend on s_2, hence neither does the degree n). The functions $a_i(s_2)$ can be taken to be bounded for $\mathcal{R}e\, s_2 \geq 0$, because it holds that $|h(1, s_2)| \to 0$ as $s_2 \to \infty$. Moreover, we can assume $b_1(s_2) \equiv 1$, for $\mathcal{R}e\, s_2 \geq 0$, after normalizing the fraction. For fixed s_2, let $\xi_i(s_2)$ be the zeroes (all with negative real part) of $h_2(s_1, s_2)$; when bounding the above representation of $h(s_1, s_2)$, use the triangle inequality for the numerator and use the inequality $|z_1 - z_2| \geq ||z_1| - |z_2||$ for the denominator:

$$|h_2(s_1, s_2)| = \left|\prod_{i=1}^{n}(s_1 - \xi_i(s_2))\right| \geq \left|\prod_{i=1}^{n}(|s_1| - |\xi_i(s_2)|)\right|,$$

the right-hand side being a polynomial function in $|s_1|$ of the same degree as $h_2(s_1, s_2)$. Thus we have the upper bound

$$|h(\xi, s-\xi)| \leq \frac{|a_1(s-\xi)|\,|\xi|^{n-1} + |a_2(s-\xi)|\,|\xi|^{n-2} + \ldots}{|\prod_{i=1}^{n}(|\xi| - |\xi_i(s-\xi)|)|}.$$

Then it follows from the facts that $\deg h_1(\xi, s-\xi) < \deg h_2(\xi, s-\xi)$ and that the a_i are bounded, that

$$|h(\xi, s-\xi)| = o(|\xi|^{-1}), \;\; |\xi| = R \to \infty,\, \mathcal{R}e\, s \geq 0, \tag{4.24}$$

for ξ running along the half-circle that closes the contour $\mathcal{C}$.

From the bounds (4.23) and (4.24), it follows that $|zh_1(\xi, s-\xi)| < |h_2(\xi, s-\xi)|$ when ξ is on $\mathcal{C}$, for R large enough, either if $|z| < 1$, $\mathcal{R}e\, s \geq 0$, or if $|z| \leq 1$, $\mathcal{R}e\, s > 0$. This yields the first part of the lemma, via Rouché's theorem.

For the second part, consider the contour $\mathcal{C}_\epsilon$ made up from the segment that runs in parallel to the imaginary axis and lying to its left at distance ϵ, together with the arc of the circle with radius R spanning in the negative half-plane between the edges of this segment.

It is essential that $h(\xi, -\xi)$ is meromorphic in $\mathcal{R}e\,\xi < 0$ (see the discussion below (4.20)). Since it has isolated poles, we can find $\epsilon > 0$, such that $h(\xi, -\xi)$ is holomorphic in the thin strip $-2\epsilon < \mathcal{R}e\,\xi < 0$. Then the left derivative of the function $h(\xi, -\xi)$ exists at 0 and we have by hypothesis,

$$\lim_{\substack{\xi \to 0 \\ \mathcal{R}e\,\xi < 0}} \frac{\mathrm{d}}{\mathrm{d}\xi} h(\xi, -\xi) = \mathbb{E}A - \mathbb{E}B > 0,$$

in particular, $h(\mathcal{R}e\,\xi, -\mathcal{R}e\,\xi) < h(0,0) = 1$. We can now bound for $\mathcal{R}e\, s \geq 0$ and ξ lying on the segment of $\mathcal{C}_\epsilon$:

$$|h(\xi, s-\xi)| \leq \mathbb{E}|e^{-\xi B - sA + \xi A}| \leq h(\mathcal{R}e\,\xi, -\mathcal{R}e\,\xi) < 1.$$

The bound for ξ lying on the arc component of $\mathcal{C}_\epsilon$ follows in the same way as (4.24). By virtue of Rouché's theorem, the proof is complete.

Remark 4.3.1. *It follows in a similar way as for the first part of Lemma 4.3.1 that the polynomials $h_2(\xi, 0)$ and $h_2(\xi, 0) - zh_1(\xi, 0)$ have the same number of zeroes with negative real part, $|z| < 1$. But since $h_2(\xi, 0)$ has only such zeroes and $\deg h_2(\cdot, 0) > \deg h_1(\cdot, 0)$, the same holds for $h_2(\xi, 0) - zh_1(\xi, 0)$.*

The idea for evaluating (4.21) is to use the theorem of residues for the contour integrals along $\mathcal{C}_\epsilon$ while arguing that the contributions from the integrals along the half-circle vanish as the radius $R \to \infty$. Focus on the contour integral of the second term in (4.21) taken along the semi-circle component, say $\mathcal{S}_\epsilon$, of $\mathcal{C}_\epsilon$. For R large enough $\mathcal{S}_\epsilon$ will be contained in the interior of a domain where $h(\xi, s-\xi)$ is holomorphic, and in addition, $|zh(\xi, s-\xi)| \leq 1$, hence the position vector $1 - zh(\xi, s-\xi)$ has positive real part when the argument ξ runs along $\mathcal{S}_\epsilon$, which means $\log[1 - zh(\xi, s-\xi)]$ is holomorphic in a neighbourhood around the arc $\mathcal{S}_\epsilon$. In conclusion, we can integrate by parts:

$$\begin{aligned}\int_{\mathcal{S}_\epsilon} \frac{\mathrm{d}\xi}{s-\xi} \log[1 - zh(\xi, s-\xi)] &= -\log(s-\xi)\log[1 - zh(\xi, s-\xi)]\Big|_{-\epsilon - iR}^{-\epsilon + iR} \\ &\quad + \int_{\mathcal{S}_\epsilon} \log(s-\xi) \frac{\mathrm{d}}{\mathrm{d}\xi} \log[1 - zh(\xi, s-\xi)]\mathrm{d}\xi.\end{aligned}$$

For large R, $|h(\xi, s-\xi)| \to 0$, which means $|\log[1 - zh(\xi, s-\xi)]| \sim |zh(\xi, s-\xi)|$, so the first term on the right behaves in absolute value as

$$\sim |\log(\xi - s)|\,|zh(\xi, s-\xi)| \sim |z\log(\xi - s)|\,|\xi|^{-1},$$

and the integrand on the left-hand side behaves as $|\xi|^{-2}$. Thus we have

$$\int_{\mathcal{S}_\epsilon} \log(\xi - s)\frac{\mathrm{d}}{\mathrm{d}\xi}\log[1 - zh(\xi, s-\xi)]\mathrm{d}\xi \to 0, \quad |\xi| \to \infty,\ \mathcal{R}e\,\xi < 0. \tag{4.25}$$

The same arguments that led to (4.25) apply to the function $\log[1 - zh(\xi, 0)]$, and similarly we have

$$\int_{\mathcal{S}_\epsilon} \log(\xi - s)\frac{\mathrm{d}}{\mathrm{d}\xi}\log[1 - zh(\xi, 0)]\mathrm{d}\xi \to 0, \quad |\xi| \to \infty,\ \mathcal{R}e\,\xi < 0. \tag{4.26}$$

Now we are ready to calculate the contour integrals (4.21). Fix s, $\mathcal{R}e\,s > 0$ and consider the integrals (4.21) taken along the contour $\mathcal{C}_\epsilon$ described in the proof of Lemma 4.3.1. ϵ is taken sufficiently small such that no poles of the integrands are lying between the segment and the imaginary axis, irrespective of R (this can be found since there are finitely many poles in the negative half-plane). The integrals in (4.21) can be approximated from the interior of the negative half-plane using the contours $\mathcal{C}_\epsilon$, for arbitrarily large R and small ϵ. Splitting the integrals along $\mathcal{C}_\epsilon$ based on the factors inside the logarithm and integrating by parts in each term, the expression in (4.21) becomes

$$\frac{1}{2\pi i}\int_{\mathcal{C}_\epsilon} \log(s-\xi)\,\frac{\frac{\mathrm{d}}{\mathrm{d}\xi}[h_2(\xi,0) - zh_1(\xi,0)]}{h_2(\xi,0) - zh_1(\xi,0)}\mathrm{d}\xi - \frac{1}{2\pi i}\int_{\mathcal{C}_\epsilon} \log(s-\xi)\,\frac{\frac{\mathrm{d}}{\mathrm{d}\xi}\,h_2(\xi,0)}{h_2(\xi,0)}\mathrm{d}\xi$$

$$+\frac{1}{2\pi i}\int_{\mathcal{C}_\epsilon} \log(s-\xi)\frac{\frac{\mathrm{d}}{\mathrm{d}\xi}\,h_2(\xi,s-\xi)}{h_2(\xi,s-\xi)}\mathrm{d}\xi - \frac{1}{2\pi i}\int_{\mathcal{C}_\epsilon} \log(s-\xi)\,\frac{\frac{\mathrm{d}}{\mathrm{d}\xi}[h_2(\xi,s-\xi) - zh_1(\xi,s-\xi)]}{h_2(\xi,s-\xi) - zh_1(\xi,s-\xi)}\mathrm{d}\xi.$$

By (4.25) and (4.26), the total contribution from the integrals along the semi-circle $\mathcal{S}_\epsilon$ vanishes as $R \to \infty$. Moreover, the branch of log was chosen such that the factors $\log(s - \xi)$ are analytic for $\mathcal{R}e\,\xi < 0$. Then the integrands have simple poles located at the zeroes of their denominators in the negative half of the complex plane. So if we denote by $\xi_i(s)$, $\xi_i(z, s)$, $i = 1, ..., n$, the zeroes with negative real part of $h_2(\xi, s - \xi)$, respectively $h_2(\xi, s - \xi) - zh_1(\xi, s - \xi)$ (see Lemma 4.3.1) and with η_j, $\eta_j(z)$, $j = 1, ..., m$ the zeroes (all having negative real part) of $h_2(\xi, 0)$, respectively $h_2(\xi, 0) - zh_1(\xi, 0)$, the integral in (4.21) is equal to

$$\log \frac{\prod_{j=1}^{m}[s - \eta_j(z)]\prod_{i=1}^{n}[s - \xi_i(s)]}{\prod_{j=1}^{m}[s - \eta_j]\prod_{i=1}^{n}[s - \xi_i(z,s)]} = \log \frac{[h_2(s,0) - zh_1(s,0)]\prod_{i=1}^{n}[s - \xi_i(s)]}{h_2(s,0)\prod_{i=1}^{n}[s - \xi_i(z,s)]},$$

after letting $R \to \infty$ and $\epsilon \to 0$. Remark also that the degree n is a (piecewise constant) function of the argument s (Lemma 4.3.1). In conclusion, we have

$$\mathbb{E}\{z^N e^{-sP}\} = 1 - \frac{[h_2(s,0) - zh_1(s,0)] \prod_{i=1}^{n} [s - \xi_i(s)]}{h_2(s,0) \prod_{i=1}^{n} [s - \xi_i(z,s)]}. \tag{4.27}$$

The calculations that led to (4.27) can be repeated for $z = 1$ and $\mathcal{R}e\, s \geq 0$. The contour of integration is the same as $\mathcal{C}_\epsilon$, and the second part of Lemma 4.3.1 must be used to conclude about the number of zeroes under the condition $\mathbb{E}A > \mathbb{E}B$. This condition is necessary for stability (P has a proper probability distribution, which must be verified when taking $z = 1$, $s = 0$). The conclusion is that we are allowed to formally replace $z = 1$ in (4.27), which becomes

$$\mathbb{E}e^{-sP} = 1 - \frac{[h_2(s,0) - h_1(s,0)] \prod_{i=1}^{n} [s - \xi_i(s)]}{h_2(s,0) \prod_{i=1}^{n} [s - \xi_i(1,s)]}, \tag{4.28}$$

with the remark that P has indeed a proper probability distribution.

Finally, it is possible to derive the transform of the idle period using a similar contour integral; and the methods of this section equally apply to the symmetric case of the joint transform $h(s_1, s_2)$ being a rational function in the argument s_2, for each fixed s_1, $\mathcal{R}e\, s_1 \geq 0$.

4.4 The time to ruin when starting at a positive level

We conclude this chapter with a derivation of the length of an excursion above level 0, when the random walk is started at a positive level u. From an insurance perspective, this is the time to ruin $\tau(u)$ when starting with initial capital u. $\tau(u)$ is a more insightful performance measure than the ruin probability. From the perspective of a risk reserve process, B stands for a generic inter-claim time and A for a generic claim size (see the description at the end of Section 1.3). As in the rest of this chapter, the elements B and A in a pair (B, A) can also be correlated with each other.

Using renewal arguments, we point out that it is possible to represent $\tau(u)$ in the form of a double transform:

$$G(s_1, s_2) := \iint e^{-s_1 u - s_2 v} \mathrm{d}\mathbb{P}(\tau(u) \leq v).$$

We will argue that the right-hand side is actually a double Stieltjes integral (the integrator $\mathbb{P}(\tau(u) \leq v)$ turns out to be a function of bounded variation in both arguments, as it is non-decreasing in v, for fixed u, and non-increasing in u, for fixed v).

From a probabilistic perspective, $G(s_1, s_2)$ represents the LST of the time to ruin given that it starts with a random initial capital having an exponential distribution (with positive rate s_1), independently of everything else.

We show in (4.30) that $G(s_1, s_2)$ can be represented in terms of the double transform of the time to ruin and the deficit at ruin when starting without any capital.

The key idea for calculating G is to exploit the regenerative structure of the descending ladder process: starting at level u, there is an excursion above u (which is distributed as P), followed by a descending ladder height (undershoot) of size distributed as I, which brings the reserve closer to ruin. From this point on, the process regenerates so we are interested in the number of jumps $N^-(u)$ required to bring the surplus in the negative for the first time. This discrete random variable then defines the compound sum of $N^-(u)$ i.i.d. copies P_k of P, and this is the actual time to ruin (with the remark that $N^-(u)$ is in general *not independent* of the components P_k):

$$\tau(u) \stackrel{d}{=} \sum_{k=1}^{N^-(u)} P_k.$$

$N^-(u)$ is the first index such that $\sum_{k=1}^{N^-(u)} I_k > u$. And since the increments are non-negative a.s., we have the identity of events

$$\{N^-(u) > n\} = \{I_1 + \ldots + I_n \leq u\}.$$

We can write

$$\begin{aligned}
\mathbb{P}(N^-(u) = n,\ \sum_{k=1}^{n} P_k \leq v) &= \mathbb{P}(\sum_{k=1}^{n-1} I_k \leq u,\ \sum_{k=1}^{n} I_k > u,\ \sum_{k=1}^{n} P_k \leq v) \\
&= \mathbb{P}(\sum_{k=1}^{n-1} I_k \leq u,\ \sum_{k=1}^{n} P_k \leq v) - \mathbb{P}(\sum_{k=1}^{n} I_k \leq u,\ \sum_{k=1}^{n} P_k \leq v);
\end{aligned} \tag{4.29}$$

the final decomposition holds because $\{\sum_{k=1}^{n} I_k \leq u\} \subset \{\sum_{k=1}^{n-1} I_k \leq u\}$.

In relation to the first term on the RHS of (4.29), remark that P_n is independent of both P_k and I_k, $k < n$, because of the regenerative structure of the excursion process. Moreover, the left-hand side is a function of bounded variation in both arguments u and v, because of the right-hand side, which is the difference of two functions that are non-decreasing both in u and v. In conclusion, if we denote the LST of the left-hand side with

$$G_n(s_1, s_2) = \iint e^{-s_1 u - s_2 v} \mathrm{d}\mathbb{P}(N^-(u) = n, \tau(u) \leq v),$$

then, by virtue of the regenerative structure of the excursion/ladder height process, (4.29) gives

$$G_n(s_1, s_2) = [\psi(0, s_2) - \psi(s_1, s_2)]\psi(s_1, s_2)^{n-1}, \quad n \geq 1,$$

where $\psi(s_1, s_2) := \mathbb{E}e^{-s_1 I - s_2 P}$, with I and P defined in Theorem 4.2.1. I stands for the deficit at ruin and P for the time to ruin, both for the risk reserve process without any initial capital ($P = \tau(0)$). Summing over $n \geq 1$, gives

$$G(s_1, s_2) = \iint e^{-s_1 u - s_2 v} \mathrm{d}\mathbb{P}(\tau(u) \le v) = \frac{\psi(0, s_2) - \psi(s_1, s_2)}{1 - \psi(s_1, s_2)}. \tag{4.30}$$

A simple instance for which $G(s_1, s_2)$ can be calculated explicitly, is if we assume the generic claim size A to follow an exponential distribution, independently from the inter-claim time; then under the condition $\mathbb{E}A > \mathbb{E}B$, the undershoot I will be again exponential with the same parameter, and independent of the excursion length, all this because of the memoryless property. In this case, (4.30) becomes:

$$G(s_1, s_2) = \mathbb{E}e^{-s_2 P} \frac{s_1}{s_1 + \mu(1 - \mathbb{E}e^{-s_2 P})},$$

with A and thus $I \sim \exp(\mu)$. The transform of the excursion length P is known, see Cohen [40], p. 250 (the corresponding queueing model is the M/G/1 queue). Finally, it is possible to numerically invert $G(s_1, s_2)$, using a similar inversion scheme as in Chapter 6, to obtain for fixed u the c.d.f. of $\tau(u)$.

Some concluding remarks.

1. The problem with identity (4.30) is that the quantity $\psi(s_1, s_2)$ is in general quite difficult to obtain. First of all, even though Spitzer's identity (4.3) is general enough to represent $\psi(s_1, s_2)$ in terms of the LST $h(s_1, s_2)$ of (B, A), the inversion theorem 4.1.1 is not sufficient to obtain a Cauchy principal value out of $h(s_1, s_2)$ (we would need a two-dimensional version of it which should apply to functions $f(u, v)$ of bounded variation, more general than the indicator functions for which higher dimensional versions are available in Hewitt [69]). These results are needed if one tries to replicate the proof of Theorem 4.2.1.

But even if we would be able to obtain integral representations for this joint transform, the real problem would still lie ahead of us. The reason is that lifting the theorem of residues we used in Section 4.3 to higher dimensions is a serious complex-analytic problem. To give the reader a basic idea, the poles and zeroes of meromorphic functions of a single variable are isolated points in $\mathbb{C}$, but already for meromorphic functions of two complex variables, the poles and zeroes become curves of a complex argument, related to the so-called *residue currents*, see for instance Coleff and Herrera [44]. These currents can be quite complicated to handle in general, and are still an active area of research in Analysis.

2. We conclude this chapter with a final remark about the stability condition $\mathbb{E}A > \mathbb{E}B$, used for the derivation of P in Section 4.3. In the dual ruin set-up, this condition implies negative safety loading, which means that P and implicitly the time to ruin $\tau(u)$ have a proper probability distribution: $\mathbb{P}(\tau(u) < \infty) = 1$. Thus, it would be useful for applications related to ruin, to also obtain the formula analogous to (4.28) under the opposite stability assumption $\mathbb{E}A < \mathbb{E}B$ (under this condition, the excursion length P above a fixed level has a defective distribution and the distribution of the deficit at ruin has an atom at 0, the size of this atom being precisely the defect of P: $\mathbb{P}(I = 0) = 1 - \mathbb{P}(P < \infty)$. For this purpose, notice that the integral representations obtained in Theorem 4.2.1 do not assume any stability condition (a close inspection of the proof of the first part in Lemma 4.3.1 and the contour integral used thereafter shows that (4.27) is always valid when the discount factor z lies in the interior of the

unit disc: $|z| < 1$). The technical problem thus reduces to either being able to use an extended version of Rouché's theorem for the case when the inequalities (4.22) are relaxed to weak inequalities along the contour of integration, or to argue that the limit $z \to 1$ is allowed from within the unit disk in (4.27) (and then the precise limit must depend on the relation between $\mathbb{E}A$ and $\mathbb{E}B$). For the study of problems related to queue lengths and their associated generating functions, extensions of Rouché's theorem have been obtained in Adan et al. [3], and Klimenok [77], the latter being based on the generalized principle of the argument from Gakhov et al. [66].

Chapter 5

Queues and risk processes with multivariate Poisson input

This chapter is devoted to the study of a two-dimensional queueing system composed of two parallel processors which receive input according to a compound Poisson arrival process with simultaneous arrivals. Moreover, the duality relation from Section 2.2 allows us to relate the workload process of the queueing system to the joint surplus process of two insurance companies that share risks using a proportional reinsurance contract. The chapter is organized as follows: In Section 5.1 we describe the model in detail. Section 5.2 is dedicated to the analysis of the 2-dimensional queueing model with ordered service times. After introducing the assumptions, we derive the Laplace-Stieltjes transform of the joint stationary workloads in the two queues and present a decomposition theorem for the stationary workload in the two queues. In Section 5.4 we extend the results of Section 5.2 to the k-dimensional queueing model. Section 5.3 is dedicated to relations to other models. We present connections with tandem and priority queues, but also with a reinsurance problem with proportional claim sizes. In Section 5.5 we discuss the case of a general two-dimensional service time (or claim size) distribution. We indicate that the two-dimensional workload problem has been solved in the queueing literature. The solution is very complicated; our ordered service times case is a degenerate case, but a case which has the advantage of a much more explicit solution which offers more probabilistic insight – and a case that can be generalized to higher dimensions.

Among the main contributions of this chapter, we mention an explicit result for the transform of the joint workload (respectively, of the joint survival probability) and its extension to the k-dimensional model. In addition, we mention the workload decomposition result. It seems to be new in this setting, although similar results – under the assumption of independent inputs – were obtained for parallel queues (cf. Kella [73]).

5.1 Model description

We consider a k-dimensional risk process in which claims arrive simultaneously in the k branches, according to a Poisson process with rate λ. The claim sizes in the k insurance lines are independent, identically distributed random vectors $(B_n^{(1)}, ..., B_n^{(k)})$, $n \geq 1$. In the sequel we denote with $(B^{(1)}, ..., B^{(k)})$ a random vector with the same distribution as $(B_1^{(1)}, ..., B_1^{(k)})$.

For the nth arriving claim vector, denote by A_n the time elapsed since the arrival of the previous claim vector, so that the A_n are independent and have an identical exponential distribution with parameter λ.

Using similar notations as in Chapter 1, let $R_t^{(i)}$, $i = 1, ..., k$ be k risk reserve processes with initial capital levels $u^{(i)}$, premium rates c_i and the *same* arrival instants t_n^*, $n \geq 1$. We have $A_n = t_{n+1}^* - t_n^*$ and $t_1^* = 0$ (no delay). Then

$$R_t^{(i)} = u^{(i)} + \sum_{j=1}^{n(t)} (c_i A_j - B_j^{(i)}) + c_i \left(t - t_{n(t)}^* \right), \tag{5.1}$$

where $n(t)$ is the number of arrivals before t (the first claim occurs at epoch t_2^*). Let $\tau^{(i)}(u^{(i)}) = \inf\{t > 0 : R_t^{(i)} < 0\}$ be the marginal times to ruin.

In connection with the ruin process, we consider k parallel M/G/1 queues with simultaneous (coupled) arrivals and correlated service requirements. As in the ruin setting, A_n are the inter arrival times of customers in the k queues and the vector $(B^{(1)}, ..., B^{(k)})$ denotes the generic service requirements. The speed of server i is denoted by c_i, meaning that server i handles c_i units of work per time unit, $i = 1, \ldots, k$.

Furthermore we denote by $\rho_i := \lambda \mathbb{E}(B^{(i)})$ the load of queue i, $i = 1, ..., k$ and we assume that $\rho_i < c_i$, to ensure that all queues can handle the offered traffic. These conditions imply positive safety loading in the ruin setting.

From the queueing perspective, the input process in this queueing system is again multivariate compound Poisson; this is because the Poisson process of arrival epochs is reversible. The vector of marks has the same distribution as the vector of claims and marks corresponding to different arrival epochs are i.i.d. vectors.

Let $(V_t^{(1)}, ..., V_t^{(k)})$ be the workload vector at time t in the system or, if we consider the nth arrival epoch, this is the workload vector $(V_{t_n}^{(1)}, \ldots, V_{t_n}^{(k)})$ seen by the customers of the nth batch arrival. Remark that, as in Section 1.1, $V_{t_n}^{(i)} = c_i W_n^{(i)}$, with $W_n^{(i)}$ the waiting time of the nth arrival in queue i. Under the stability conditions above, the vectors $(V_t^{(1)}, ..., V_t^{(k)})$ and $(V_{t_n}^{(1)}, ..., V_{t_n}^{(k)})$ converge in distribution to the steady-state joint workload at arbitrary epochs and at arrival epochs, respectively. Due to the PASTA property (Poisson Arrivals See Time Averages), these vectors are equal in distribution. Similarly, the vector $(W_n^{(1)}, \ldots, W_n^{(k)})$ converges in distribution to the steady state waiting time. We denote the Laplace-Stieltjes transform (LST) of the steady-state workload vector:

$$\psi(s_1, s_2, ..., s_k) := \mathbb{E}(e^{-s_1 V^{(1)} - s_2 V^{(2)} - \ldots - s_k V^{(k)}}).$$

The multivariate duality result of Section 2.2 is still in force, however, since we are dealing here with a Lévy process as input, we have a lot of extra structure at hand,

and we can obtain more than if we would exploit only the embedded random walk structure. The embedded random walk approach will be fully exploited in the next chapter, where we study a more general model driven by a semi-Markov process.

5.2 The analysis of the two-dimensional problem

In this section we derive the transform of the joint steady state workload process of the two-dimensional queueing model with simultaneous arrivals, as introduced in Section 5.1. We also present a probabilistic interpretation of the quantities involved in the formula of the joint workload. The results are of immediate relevance for the corresponding insurance problem, via the duality outlined in Section 2.2.

Before we start with the analysis, we make the following simplifying assumption.
Assumption 1. All premium rates, respectively all service speeds, are 1, viz., $c_1 = \cdots = c_k = 1$.
The following observation shows that this assumption is not restrictive. If we divide all terms in the righthand side of (5.1) by c_i, we arrive at a new risk model with initial capital $u^{(i)}/c_i$ and claim size $B^{(i)}/c_i$ and *unit* premium rates.

Remark 5.2.1. *For the study of survival functions, we can normalize the risk reserve processes by their respective income rates. The survival function is preserved, with the starting capital scaled accordingly. To be more precise, let*

$$\tilde{R}_t^{(i)} = u^{(i)}/c_i + t - \sum_{k=1}^{n(t)} B_k^{(i)}/c_i, \ i = 1, 2.$$

Since $R_t^{(i)} \geq 0 \Leftrightarrow \tilde{R}_t^{(i)} \geq 0$, for $\tilde{\tau}_\wedge$, $\tau_\wedge$ the exit times (2.11) of $\tilde{R}_t$ respectively R_t from the non-negative quadrant, we have the relation

$$\tilde{\tau}_\wedge(u^{(1)}, u^{(2)}) = \tau_\wedge(c_1 u^{(1)}, c_2 u^{(2)})$$

and then also $\tilde{F}^s(u^{(1)}, u^{(2)}) = F^s(c_1 u^{(1)}, c_2 u^{(2)})$, with $\tilde{F}^s$ the survival function of the scaled process.

This means that for our purposes it suffices to study the process R_t and the associated survival functions $F^s(u^{(1)}, u^{(2)})$ assuming unit income rates and properly scaled claim size vectors; and this is what we will do from now on.

Similarly, in the corresponding queueing model the service times at queue i are also divided by c_i and the service speeds are equal to 1. This will not change the nth waiting time $W_n^{(i)}$ at queue i, but the workload $V_n^{(i)}$ at the nth arrival epoch is divided by c_i. Hence the multivariate duality relation from Theorem 2.2.1 is preserved.

The LST of the joint service time/claim size vector is denoted by

$$\phi(s_1, s_2) := \mathbb{E}(e^{-s_1 B^{(1)} - s_2 B^{(2)}}).$$

Our key assumption is the following:
Assumption 2. $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$. In view of the above discussion, in case the speeds are c_i, our assumption would be $\mathbb{P}(B^{(1)}/c_1 \geq B^{(2)}/c_2) = 1$.

Remark 5.2.2. *This model allows for a dedicated Poisson arrival stream into queue 1. Merging this separate arrival process with the simultaneous arrival process at queue 1, the distribution of $B^{(2)}$ will have an atom in 0, which is the probability that a dedicated Poisson arrival happens instead of a simultaneous one (see Badescu et al. [23] for a reinsurance model with both dedicated and simultaneous arrivals). This modification preserves the ordering assumption.*

We are interested in the joint stationary distribution of the amount of work in the two queues

$$\psi(s_1, s_2) := \mathbb{E}(e^{-s_1 V^{(1)} - s_2 V^{(2)}}).$$

This can be obtained by considering the Lindley recursion for the embedded workload process (1.1)

$$(V^{(1)}_{t_{n+1}}, V^{(2)}_{t_{n+1}}) = (\max(V^{(1)}_{t_n} + B^{(1)}_n - A_n, 0), \max(V^{(2)}_{t_n} + B^{(2)}_n - A_n, 0)).$$

Or, for the LST

$$\psi_n(s_1, s_2) = \mathbb{E}\left(e^{-s_1 V^{(1)}_{t_n} - s_2 V^{(2)}_{t_n}}\right), \quad n = 1, 2, \ldots,$$

this gives after straightforward calculations

$$\begin{aligned} \psi_{n+1}(s_1, s_2) = {} & \tfrac{\lambda}{\lambda - s_1 - s_2} \left(\phi(s_1, s_2)\psi_n(s_1, s_2) - \phi(s_1, \lambda - s_1)\psi_n(s_1, \lambda - s_1)\right) \\ & + \tfrac{\lambda}{\lambda - s_1} \left(\phi(s_1, \lambda - s_1)\psi_n(s_1, \lambda - s_1) - \phi(\lambda, 0)\psi_n(\lambda, 0)\right) \\ & + \phi(\lambda, 0)\psi_n(\lambda, 0). \end{aligned} \tag{5.2}$$

Under the stability condition $\rho_1 < 1$, $\psi(s_1, s_2) := \lim_{n\to\infty} \psi_n(s_1, s_2)$ exists and

$$\begin{aligned} \left(1 - \tfrac{\lambda\phi(s_1,s_2)}{\lambda - s_1 - s_2}\right) \psi(s_1, s_2) = {} & \left(\tfrac{\lambda}{\lambda - s_1} - \tfrac{\lambda}{\lambda - s_1 - s_2}\right) \phi(s_1, \lambda - s_1)\psi(s_1, \lambda - s_1) \\ & + \left(1 - \tfrac{\lambda}{\lambda - s_1}\right) \phi(\lambda, 0)\psi(\lambda, 0). \end{aligned} \tag{5.3}$$

If we let A denote a generic inter-arrival time, then due to the PASTA property,

$$\phi(\lambda, 0)\psi(\lambda, 0) = \mathbb{P}(V^{(1)} + B^{(1)} \leq A) = \mathbb{P}(V^{(1)} = 0) = 1 - \rho_1. \tag{5.4}$$

This is the probability that queue 1 is empty at an arbitrary time instant.

On the regularity domains of $\psi(s_1, s_2)$ and $\phi(s_1, s_2)$: We remark that, because of the ordering $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$, we can rewrite the transform of the joint service times as:

$$\phi(s_1, s_2) = \mathbb{E}e^{-s_1(B^{(1)} - B^{(2)}) - (s_1 + s_2)B^{(2)}} =: \tilde{\phi}(s_1, s_1 + s_2),$$

and this function is always regular in $\mathcal{R}e\, s_1 > 0$, $\mathcal{R}e\,(s_1 + s_2) > 0$. If we consider $(B^{(1)}, B^{(2)})$ subject to $B^{(1)} \geq B^{(2)}$ a.s., $\phi(s_1, s_2)$ may not be regular beyond this domain. More precisely, if $B^{(2)}$ has a heavy-tailed distribution, this implies that $B^{(1)}$

is also heavy tailed because of the dependence structure. In this case $\phi(s_1, s_2)$ cannot be extended beyond $\mathcal{R}e\, s_1 \geq 0$, $\mathcal{R}e\,(s_1 + s_2) \geq 0$. Similar considerations hold for $\psi(s_1, s_2)$ because we must also have $\mathbb{P}(V^{(1)} \geq V^{(2)}) = 1$.

It can be shown using Rouché's Theorem that for every s_1 with $\mathcal{R}e\, s_1 > 0$ there exists a unique $s_2 = s_2(s_1)$ with $\mathcal{R}e\, s_2(s_1) > \mathcal{R}e\,(-s_1)$, that satisfies the identity $\lambda\phi(s_1, s_2) = \lambda - (s_1 + s_2)$. Moreover the function: $s_1 \to s_2(s_1)$ (which is in this case well defined) is analytic in $\mathcal{R}e\, s_1 > 0$. For the proof of this, see Lemma 5.6.1 in the Appendix B of this chapter.

Hence the pair $(s_1, s_2(s_1))$ is a zero of $\left(1 - \frac{\lambda\phi(s_1,s_2)}{\lambda - s_1 - s_2}\right)$ in (5.3), which is in the regularity domain of $\psi(s_1, s_2)$. Then the righthand side of (5.3) is also zero, i.e.

$$\lambda s_2(s_1)\phi(s_1, \lambda - s_1)\psi(s_1, \lambda - s_1) = -s_1(\lambda - s_2(s_1) - s_1)\phi(\lambda, 0)\psi(\lambda, 0). \tag{5.5}$$

If we substitute this in (5.3) and use (5.4), we obtain

$$\psi(s_1, s_2) = (1 - \rho_1)\frac{s_1}{s_1 + s_2 - \lambda(1 - \phi(s_1, s_2))} \cdot \frac{s_2(s_1) - s_2}{s_2(s_1)}. \tag{5.6}$$

The interpretation of the Rouché zero $s_2(s_1)$. Assume that a customer that starts a busy period $BP^{(2)}$ in queue 2 demands work x in queue 2 and work $x + y$ in queue 1. During the service time of this customer in the second queue, there are Poisson(λx) arriving customers, all of these generating independent busy sub-periods with the same distribution as $BP^{(2)}$ in queue 2. So if we denote with U the extra work in the first queue, at the end of a busy period in the second queue, and with $U^*(s_1)$ its Laplace-Stieltjes transform, we have the identity:

$$U^*(s_1) = \int_{x=0}^{\infty}\int_{y=0}^{\infty} e^{-s_1 y}\sum_{k=0}^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}[U^*(s_1)]^k \,\mathrm{d}\mathbb{P}(B^{(1)} - B^{(2)} \leq y, B^{(2)} \leq x).$$

The powers of $U^*(s_1)$ correspond to the extra work contributions at the end of the busy sub-periods started during the service time of the first customer in the busy period $BP^{(2)}$. We can rewrite the above identity as:

$$U^*(s_1) = \tilde{\phi}(s_1, \lambda[1 - U^*(s_1)]) = \phi(s_1, \lambda[1 - U^*(s_1)] - s_1). \tag{5.7}$$

Comparing this with the equation satisfied by $s_2(s_1)$, in terms of $\tilde{\phi}(s_1, s_1 + s_2)$, we have:

$$\begin{cases} \lambda\tilde{\phi}(s_1, s_1 + s_2(s_1)) = \lambda - (s_1 + s_2(s_1)), \\ \lambda\tilde{\phi}(s_1, \lambda[1 - U^*(s_1)]) = \lambda U^*(s_1). \end{cases}$$

We may assume w.l.o.g. that $\mathbb{P}(B^{(1)} > B^{(2)}) > 0$, otherwise the two queues are a.s. identical, which is not interesting. Then it follows that the real part of $\lambda(1 - U^*(s_1))$ is positive, and we must have $s_1 + s_2(s_1) = \lambda(1 - U^*(s_1))$ because $s_2(s_1)$ is unique in the region $\mathcal{R}e\,(s_1 + s_2) > 0$. We have thus proved:

Proposition 5.2.1. *The relation between $s_2(s_1)$ and the transform of the extra workload in queue 1 at the end of a busy period in the shortest queue is*

$$\lambda U^*(s_1) = \lambda - (s_1 + s_2(s_1)). \tag{5.8}$$

The transform of the joint workload in the two systems becomes

$$\psi(s_1, s_2) = (1-\rho_1)\frac{s_1 + s_2 - \lambda(1 - U^*(s_1))}{s_1 + s_2 - \lambda(1 - \phi(s_1, s_2))} \cdot \frac{s_1}{s_1 - \lambda(1 - U^*(s_1))}.$$

The workload decomposition. Based on Proposition 5.2.1, we show that the steady-state workload decomposes into an independent sum of a modified workload and an additional term, which represents the steady-state workload in a classical M/G/1 queue.

We start the joint workload process and let it run until the end of each busy period in the queue with the smallest workload. At this random time instant, we remove the extra content in queue 1, which has the largest workload of the two. Let us denote this modified joint workload process as $(\tilde{V}^{(1)}, V^{(2)})$. Then at the arrival instants of customers in the two queues, the following recurrence relation holds:

$$(\tilde{V}^{(1)}_{t_{n+1}}, V^{(2)}_{t_{n+1}}) = \begin{cases} (\tilde{V}^{(1)}_{t_n} + B^{(1)}_n - A_n, V^{(2)}_{t_n} + B^{(2)}_n - A_n), & \text{if } A_n < V^{(2)}_{t_n} + B^{(2)}_n, \\ (0, 0), & \text{if } A_n \geq V^{(2)}_{t_n} + B^{(2)}_n. \end{cases}$$

Remark that marginally, the shortest queue evolves unchanged. If we have ergodicity, then in steady state the above recurrence becomes:

$$(\tilde{V}^{(1)}, V^{(2)}) \stackrel{d}{=} \begin{cases} (\tilde{V}^{(1)} + B^{(1)} - A, V^{(2)} + B^{(2)} - A), & \text{if } A < V^{(2)} + B^{(2)}, \\ (0, 0), & \text{if } A \geq V^{(2)} + B^{(2)}. \end{cases}$$

Here and in the following, $\stackrel{d}{=}$ denotes equality in distribution. In terms of LST's, we obtain the following functional equation for $\tilde{\psi}(s_1, s_2) := \mathbb{E}e^{-s_1\tilde{V}^{(1)} - s_2 V^{(2)}}$:

$$(1 - \frac{\lambda\phi(s_1, s_2)}{\lambda - s_1 - s_2})\tilde{\psi}(s_1, s_2) = (1 - \rho_2) - \frac{\lambda}{\lambda - s_1 - s_2}\tilde{\psi}(s_1, \lambda - s_1)\phi(s_1, \lambda - s_1),$$

where $1 - \rho_2 = \mathbb{P}(V^{(2)} = 0)$.

Now follows a similar analysis as for $\psi(s_1, s_2)$. We already know from the Rouché problem that $s_2(s_1)$ is a zero of $(1 - \frac{\lambda\phi(s_1,s_2)}{\lambda - s_1 - s_2})$. We also have $\tilde{V}^{(1)} \geq V^{(2)}$ a.s. (even if we take out the extra workload at the largest queue at the end of each busy period, $\tilde{V}^{(1)}$ is still at least as large as $V^{(2)}$), therefore $(s_1, s_2(s_1))$ is in the regularity domain of $\tilde{\psi}(s_1, s_2)$ and therefore, at the point $(s_1, s_2(s_1))$, the right-hand side of the above identity is equal to zero:

$$\tilde{\psi}(s_1, \lambda - s_1)\phi(s_1, \lambda - s_1) = (1 - \rho_2)\frac{\lambda - s_1 - s_2(s_1)}{\lambda}.$$

Substituting back in the original identity, yields:

$$\tilde{\psi}(s_1, s_2) = (1 - \rho_2)\frac{s_1 + s_2 - \lambda(1 - \phi(s_1, s_2(s_1)))}{s_1 + s_2 - \lambda(1 - \phi(s_1, s_2))}. \tag{5.9}$$

This is a 2-dimensional Pollaczek-Khinchine type of representation. From an analytic point of view, the role of the numerator is to cancel the unique pole of the denominator in the region $\mathcal{R}e\,(s_1 + s_2) > 0$.

Substitute $s_2(s_1)$ from Proposition 5.2.1 and $\tilde{\psi}$ from (5.9) into (5.6):

$$\psi(s_1, s_2) = \frac{1 - \rho_1}{1 - \rho_2}\frac{s_1}{s_1 - \lambda[1 - U^*(s_1)]}\tilde{\psi}(s_1, s_2). \tag{5.10}$$

We can now state the main result:

Theorem 5.2.1 (Work decomposition)**.** *In steady state, we have the following representation of the joint workload at the two queues as an independent sum:*

$$(V^{(1)}, V^{(2)}) \stackrel{d}{=} (\tilde{V}^{(1)}, V^{(2)}) + (V^{(1),1}, 0),$$

where $V^{(1),1}$ is the workload in an independent, virtual M/G/1 queue with arrival rate λ and service requirements distributed as U, the extra workload at the end of a busy period $BP^{(2)}$ in the shortest queue.

Proof. It suffices to remark that the factor

$$\frac{1 - \rho_1}{1 - \rho_2}\frac{s_1}{s_1 - \lambda[1 - U^*(s_1)]} = \mathbb{E}e^{-s_1 V^{(1),1}}$$

in (5.10) is the Pollaczek-Khinchine formula for the transform of the workload in the virtual M/G/1 queue with service time that has the same distribution as U. This virtual queue is obtained by contracting the busy periods in the initial shortest queue, so that an arrival in the virtual queue happens at the end of this busy period and the inter arrival time is then the idle period in the initial queue, and so is exponentially distributed. To see that indeed $\frac{1-\rho_1}{1-\rho_2}$ is the atom of $V^{(1),1}$ at 0, differentiate the identity for $U^*(s_1)$ in (5.7):

$$\mathbb{E}(U) = -\frac{\mathrm{d}}{\mathrm{d}s_1}\phi(s_1, \lambda(1 - U^*(s_1)) - s_1)_{|s_1=0} = \mathbb{E}(B^{(1)} - B^{(2)}) + \lambda\mathbb{E}B^{(2)}\mathbb{E}(U),$$

so that $1 - \lambda\mathbb{E}(U) = \frac{1-\rho_1}{1-\rho_2}$.

5.3 Relation with other models

In this section we point out how the results of the previous section are related to results for a risk model with proportional reinsurance, a particular tandem fluid model and to a particular priority queue. We start by showing that (5.6) generalizes a result obtained in Avram et al. [19], for the risk setting.

The case of proportional reinsurance. In [19] the joint risk reserve process $(R^{(1)}, R^{(2)})$ is of the form: $R^{(i)}(t) = u^{(i)} + c_i t/\delta_i - S(t)$. Here $S(t)$ is a common Compound Poisson input process with generic claim sizes B and c_i are the premium rates. The claims are being divided in fixed proportions δ_i, respectively.

To bring this closer to our setting in Section 5.2, normalize the income rates: i.e. we consider $(\frac{1}{p_1}R^{(1)}, \frac{1}{p_2}R^{(2)})$ with $p_i = \frac{c_i}{\delta_i}$. The assumption in [19] is that $p_1 > p_2$, which means that, in our notation, the claim sizes are $B^{(1)} := \frac{1}{p_1}B < \frac{1}{p_2}B =: B^{(2)}$. Remark that the inequality between the $B^{(i)}$'s is reversed here (which means the role of the arguments in our transforms is interchanged, especially the Rouché zero).

Let us recall the main formula in Avram et al. [19] (Formula (23)):

$$\psi_{*R^{(1)},R^{(2)}}(p,q) = \frac{\kappa_2(0+)'}{p(\kappa_1(p+q) - q(p_1 - p_2))} \frac{q + p - q^+(q(p_1 - p_2))}{q - q^+(q(p_1 - p_2))}, \tag{5.11}$$

where $\psi_{*R^{(1)},R^{(2)}}(p,q)$ denotes the *Laplace transform* of the infinite horizon survival function:

$$\psi_{*R^{(1)},R^{(2)}}(p,q) = \int_{0-}^{\infty} e^{-pu^{(1)} - qu^{(2)}} F^s(u^{(1)}, u^{(2)})\, \mathrm{d}u^{(1)}\, \mathrm{d}u^{(2)}.$$

The relation between the ruin times of $(R^{(1)}, R^{(2)})$ and $(\frac{1}{p_1}R^{(1)}, \frac{1}{p_2}R^{(2)})$ is

$$\tau_{\frac{1}{p_1}R^{(1)}, \frac{1}{p_2}R^{(2)}}(u_1, u_2) = \tau_{R^{(1)},R^{(2)}}(p_1 u_1, p_2 u_2).$$

Hence the relation to our coordinates is $s_1 = p_1 p$, $s_2 = p_2 q$. From this, the relation between the LT of the survival functions become after a change of variables:

$$\psi_{*\frac{1}{p_1}R^{(1)}, \frac{1}{p_2}R^{(2)}}(s_1, s_2) = \frac{1}{p_1 p_2} \psi_{*R^{(1)},R^{(2)}}(p,q). \tag{5.12}$$

- $\kappa_i(\alpha)$ is the Laplace exponent of the Compound Poisson process with drift p_i per unit time. This means

  $$\kappa_i(\alpha) = p_i \alpha - \lambda(1 - \mathbb{E}e^{-\alpha B}).$$

  Because of the linear dependence between the $B^{(i)}$'s, their LST has the form $\mathbb{E}e^{-s_1 B^{(1)} - s_2 B^{(2)}} = \phi(s_1, s_2) =: \phi_{B^{(1)}}(s_1 + \frac{p_1}{p_2}s_2)$.

- $q^+(q)$ is the largest root of the equation $\kappa_1(\alpha) = q$. Then $q^+(q(p_1 - p_2))$ solves:

  $$p_1\alpha - \lambda(1 - \mathbb{E}e^{-\alpha p_1 B^{(1)}}) = q(p_1 - p_2).$$

  Remark that if we set $\alpha = p + q$, the above becomes:

  $$p_1 p + p_2 q - \lambda(1 - \phi_{B^{(1)}}(p_1 p + p_1 q)) = 0,$$

  or, written in the (s_1, s_2)-coordinates, this becomes the equation satisfied by $s_1(s_2)$ (s_1 and s_2 are now interchanged). Hence the relation between the zeroes in the two notations is: $s_1(s_2) = p_1(\alpha - q) = p_1[q^+(q(p_1 - p_2)) - q]$.

The constant $\kappa(0+)' = p_2 - \lambda \mathbb{E}B^{(2)} = p_2(1-\rho_2)$ is the probability that the queueing system is empty in steady state (now the second queue has a higher workload).

In conclusion, using the relation between the Laplace integral transform and the Laplace-Stieltjes transform:

$$\psi_{*R^{(1)},R^{(2)}}(s_1, s_2) = \frac{1}{s_1 s_2}\psi(s_1, s_2), \tag{5.13}$$

(5.11) written via (5.12) and (5.13) in the (s_1, s_2) coordinates becomes Formula (5.6):

$$\psi(s_2, s_1) = \frac{s_1(1-\rho_2)}{s_1 + s_2 - \lambda(1 - \phi_{B^{(1)}}(s_1 + \frac{p_1}{p_2}s_2))} \cdot \frac{s_1 - s_1(s_2)}{-s_1(s_2)},$$

with the arguments s_1 and s_2 interchanged.

Relation with work on tandem fluid queues. We now show that the workload model with ordered service times is equivalent with a particular tandem fluid queue. That is a model of two queues in series, in which the outflow from the first queue is a fluid, i.e., there is continuous outflow when the server is working (instead of customers leaving one by one). Such tandem fluid queues have been studied by various authors, see in particular Kella [73]. Consider the following two-station tandem fluid network with *independent* compound Poisson input at the two stations (with arrival rate λ_i and Laplace-Stieltjes transform of the service times $B_i^*(\cdot), i = 1, 2$). Then Theorem 4.1 of Kella [73] gives the Laplace-Stieltjes transform of the steady-state fluid levels W_1 and W_2 in the two nodes:

$$\psi_W(\alpha_1, \alpha_2) = \mathbb{E}\left(e^{-\alpha_1 W_1 - \alpha_2 W_2}\right) = \frac{(1-\rho_1-\rho_2)\alpha_2}{\phi_1(\alpha_1) - \phi_1(\hat{\eta}_2(\alpha_2))} \cdot \frac{\alpha_1 - \hat{\eta}_2(\alpha_2)}{\alpha_2 - \hat{\eta}_2(\alpha_2)}, \tag{5.14}$$

with

- $\rho_i = \lambda_i \mathbb{E}(B_i)$,
- $\phi_1(\alpha_1) = \alpha_1 - \eta_1(\alpha_1)$,
- $\eta_i(\alpha_i) = \lambda_i(1 - B_i^*(\alpha_i))$,
- $\hat{\eta}_2(\alpha_2)$ the solution of $\phi_1(\hat{\eta}_2(\alpha_2)) = \eta_2(\alpha_2)$.

Alternatively, the last relation can also be formulated as: $\hat{\eta}_2(\alpha_2)$ is the solution of

$$\lambda_1 B_1^*(\hat{\eta}_2(\alpha_2)) + \lambda_2 B_2^*(\alpha_2) = \lambda_1 + \lambda_2 - \hat{\eta}_2(\alpha_2).$$

This system is related to our model with arrival rate $\lambda = \lambda_1 + \lambda_2$ and Laplace-Stieltjes transform of service requirements

$$\phi(s_1, s_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} B_1^*(s_1 + s_2) + \frac{\lambda_2}{\lambda_1 + \lambda_2} B_2^*(s_1).$$

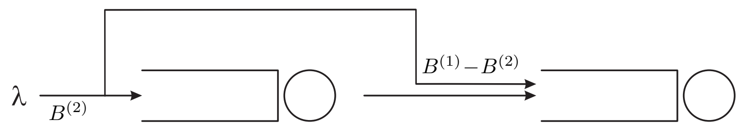


FIGURE 5.1: Tandem fluid queue

The corresponding notation is: $B_1 \stackrel{d}{=} B^{(2)}$ and $B_2 \stackrel{d}{=} B^{(1)} - B^{(2)}$. Here W_1 in the tandem model corresponds to the workload in the smallest queue in our model and $W_1 + W_2$ in the tandem model corresponds to the workload in the largest queue in our model. So we have

$$\begin{aligned}\psi(s_1, s_2) &= \mathbb{E}\left(e^{-s_1 V_1 - s_2 V_2}\right) = \mathbb{E}\left(e^{-s_1(W_1+W_2) - s_2 W_1}\right) = \psi_W(s_1 + s_2, s_1)\\ &= \frac{(1-\rho_1-\rho_2)s_1}{s_1+s_2-\lambda_1(1-B_1^*(s_1+s_2))-\lambda_2(1-B_2^*(s_1))} \cdot \frac{s_1+s_2-\hat{\eta}_2(s_1)}{s_1-\hat{\eta}_2(s_1)}.\end{aligned}$$

Now remark that

- The total traffic offered to the largest queue is $\rho_1 + \rho_2$, so indeed the factor $1 - \rho_1 - \rho_2$ in [73] corresponds to the factor $1 - \rho_1$ in (5.6);
- $\lambda(1 - \phi(s_1, s_2)) = \lambda_1(1 - B_1^*(s_1 + s_2)) + \lambda_2(1 - B_2^*(s_1))$;
- $\lambda\phi(s_1, s_2(s_1)) = \lambda_1 B_1^*(s_1 + s_2(s_1)) + \lambda_2 B_2^*(s_1) = \lambda_1 + \lambda_2 - (s_1 + s_2(s_1))$, so indeed $\hat{\eta}_2(s_1)$ corresponds to our $s_1 + s_2(s_1)$.

We conclude that (5.6) coincides with Theorem 4.1 of Kella [73] in the case of independent compound Poisson input. Kella's result is more general in the sense that he has Lévy input instead of compound Poisson input. Our result is more general in the sense that we have *dependent* compound Poisson input.

Relation with work on priority queues. As was already noticed in Kella [73], but also in several other places in the literature, the tandem fluid network described above is also related to a priority queue with preemptive resume priorities. Hence the same holds for our workload model. Consider the following model with two types of customers where customers of type-i arrive according to a Poisson process with rate λ_i having service times with Laplace-Stieltjes transform $B_i^*(\cdot), i = 1, 2$. Assume furthermore that customers of type-1 have preemptive resume priority over customers of type-2. If we denote by Y_1 and Y_2 the steady-state workloads in the two queues, then Y_1 and Y_2 are related to W_1 and W_2 in the tandem fluid network. The Laplace-Stieltjes transform of the steady-state workloads in the two queues satisfies

$$\begin{aligned}\psi_Y(s_1, s_2) &= \mathbb{E}\left(e^{-s_1 Y_1 - s_2 Y_2}\right) = \mathbb{E}\left(e^{-s_1 W_1 - s_2 W_2}\right)\\ &= \mathbb{E}\left(e^{-s_1 V_2 - s_2(V_1 - V_2)}\right) = \psi_V(s_2, s_1 - s_2)\end{aligned}$$

where again in our model we have to take arrival rate $\lambda = \lambda_1 + \lambda_2$ and Laplace-Stieltjes transform of service requirements

$$\phi(s_1, s_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} B_1^*(s_1 + s_2) + \frac{\lambda_2}{\lambda_1 + \lambda_2} B_2^*(s_1).$$

We conclude that (5.6) also gives the Laplace-Stieltjes transform of a priority queue. Again our result is more general in the sense that we have *dependent* compound Poisson input (i.e., we can have arrivals of customers who have both low and high priority work).

5.4 The k-dimensional problem

In this section we consider the k-queue system with simultaneous arrivals. We give the transform for the steady-state joint workload and we show that the decomposition in Theorem 5.2.1 extends to this case if we preserve the ordering between the service requirements/claim sizes. We use an iterative argument and for this purpose, the decomposition in Section 5.2 will be the starting point; the iteration step is essentially done with the help of Lemma 5.4.1 below as a work conservation identity.

We thus consider k parallel M/G/1 queues, numbered 1 to k, respectively, with simultaneous (coupled) arrivals and correlated service requirements. We use the same notations as in Section 5.1. The LST of the service time/claim size vector is denoted by

$$\phi(s_1, \dots, s_k) := \mathbb{E}(e^{-s_1 B^{(1)} - \dots - s_k B^{(k)}}).$$

The essential assumption in the model extends Assumption 2 for the 2-dimensional problem:

$$\mathbb{P}(B^{(1)} \geq B^{(2)} \geq \dots \geq B^{(k)}) = 1.$$

Furthermore we denote by $\rho_i := \lambda \mathbb{E} B^{(i)}$, $i = 1, \dots, k$, the load of queue i and we assume that $\rho_1 < 1$ (hence $\rho_i < 1$, $\forall i$), to assure that all queues can handle the offered work.

Remark 5.4.1. *Like in the two-dimensional case (cf. Remark 5.2.2), this model allows for a separate Poisson arrival stream into queue 1. Merging this separate arrival process with the simultaneous arrival process, the distribution of $(B^{(2)}, \dots, B^{(k)})$ will have an atom in (0,...,0), which is the probability that a dedicated Poisson arrival happens instead of a simultaneous one.*

The Laplace-Stieltjes transform of $(V^{(1)}, \dots, V^{(k)})$. The k-dimensional Lindley recursion holds for the random variables $(V_{t_n}^{(1)}, \dots, V_{t_n}^{(k)})$:

$$(V_{t_{n+1}}^{(1)}, \dots, V_{t_{n+1}}^{(k)}) = (\max(V_{t_n}^{(1)} + B_n^{(1)} - A_n, 0), \dots, \max(V_{t_n}^{(k)} + B_n^{(k)} - A_n, 0)).$$

For $\psi_n(s_1,\dots,s_k) := \mathbb{E}\left(e^{-s_1 V_{t_n}^{(1)} - \dots - s_k V_{t_n}^{(k)}}\right)$, $n \geq 1$, the Lindley recursion gives after straightforward calculations:

$$\begin{aligned}\psi_{n+1}(s_1,\dots,s_k) = \sum_{j=1}^{k} \tfrac{\lambda}{\lambda - \sum_{i=1}^{j} s_i} \Big[&\phi^{(j)}(s_1,\dots,s_j)\psi_n^{(j)}(s_1,\dots,s_j) \\ &- \phi^{(j-1)}(s_1,\dots,s_{j-1})\psi_n^{(j-1)}(s_1,\dots,s_{j-1})\Big] + \phi^{(0)}\psi_n^{(0)} \qquad (5.15)\end{aligned}$$

where we used the following notation for simplicity: $\psi_n^{(k)}(s_1,\dots,s_k) := \psi_n(s_1,\dots,s_k)$ with $\psi_n^{(0)} := \psi_n(\lambda,0,\dots,0)$, and

$$\psi_n^{(j)}(s_1,\dots,s_j) := \psi_n(s_1,\dots,s_j,\lambda - \sum_{i=1}^{j} s_i, \underbrace{0,\dots,0}_{k-j-1\ arguments}), \text{ for } 1 \leq j \leq k-1.$$

$\phi^{(j)}(s_1,\dots,s_j)$ is analogously defined for $j = 0,\dots,k$. By taking $n \to \infty$ in (5.15), we obtain for $\psi(s_1,\dots,s_k) := \lim_{n\to\infty} \psi_n(s_1,\dots,s_k)$,

$$\begin{aligned}\left(1 - \tfrac{\lambda\phi(s_1,\dots,s_k)}{\lambda - \sum_{i=1}^{k} s_i}\right)\psi(s_1,\dots,s_k) = \sum_{j=0}^{k-1} &\left(\tfrac{\lambda}{\lambda - \sum_{i=1}^{j} s_i} - \tfrac{\lambda}{\lambda - \sum_{i=1}^{j+1} s_i}\right) \\ &\cdot \phi^{(j)}(s_1,\dots,s_j)\psi^{(j)}(s_1,\dots,s_j), \qquad (5.16)\end{aligned}$$

with $\psi^{(j)} := \lim\limits_{n\to\infty} \psi_n^{(j)}$; and $\phi^{(0)}\psi^{(0)} = \mathbb{P}(V^{(1)} + B^{(1)} \leq A) = 1 - \rho_1$.

Formula (5.16) has a simple recursive structure, and we can rewrite it as:

$$\begin{aligned}\left(1 - \tfrac{\lambda\phi(s_1,\dots,s_k)}{\lambda - \sum_{i=1}^{k} s_i}\right)\psi(s_1,\dots,s_k) = \left(\tfrac{\lambda}{\lambda - \sum_{i=1}^{k-1} s_i} - \tfrac{\lambda}{\lambda - \sum_{i=1}^{k} s_i}\right)\phi^{(k-1)}(s_1,\dots,s_{k-1})\cdot \\ \psi^{(k-1)}(s_1,\dots,s_{k-1}) + \left(1 - \tfrac{\lambda\phi(s_1,\dots,s_{k-1},0)}{\lambda - \sum_{i=1}^{k-1} s_i}\right)\psi(s_1,\dots,s_{k-1},0). \\ (5.17)\end{aligned}$$

Denote by $C_j := \left(1 - \frac{\lambda\phi(s_1,\dots,s_j,0,\dots,0)}{\lambda - \sum_{i=1}^{j} s_i}\right)\psi(s_1,\dots,s_j,0,\dots,0)$, and remark that $\psi(s_1,\dots,s_j,0,\dots,0)$ is the transform of the workload in the j-dimensional system obtained by ignoring the last $(k-j)$ queues, $j = 1,\dots,k$.

Proposition 5.4.1. *The LST of the steady-state workload in the $k \geq 3$ systems is given by:*

$$\psi(s_1,\dots,s_k) = \frac{(1-\rho_k)(S_k - s_k)}{\sum_{i=1}^{k} s_i - \lambda(1 - \phi(s_1,\dots,s_k))} \prod_{j=2}^{k-1} \frac{1-\rho_j}{1-\rho_{j+1}} \frac{S_j - s_j}{S_{j+1}} \cdot \frac{1-\rho_1}{1-\rho_2}\frac{s_1}{S_2}, \qquad (5.18)$$

with $S_j = S_j(s_1,...,s_{j-1})$ the unique solution of the equation

$$\lambda\phi(s_1,\dots,s_j,0,\dots,0) = \lambda - \sum_{i=1}^{j} s_i,$$

with $\mathcal{R}e\,(s_1 + \cdots + s_{j-1} + S_j(s_1, \ldots, s_{j-1})) > 0$, *for all* $j = 2, \ldots, k$.

Proof. The key remark is that s_k is not among the arguments of the functions $\psi^{(j)}$ that appear in the righthand side of (5.16) or (5.17). Similarly as in Section 5.2, Rouché's theorem (Lemma 5.6.1) applied to s_1 replaced by $s_1 + \cdots + s_{k-1}$ and s_2 replaced by s_k, yields the existence of a unique solution $S_k = S_k(s_1, \ldots, s_{k-1})$ of the equation

$$\lambda\phi(s_1, \ldots, s_k) = \lambda - \sum_{i=1}^{k} s_i,$$

such that $S_k(s_1, \ldots, s_{k-1}) + \sum_{i=1}^{k-1} s_i$ has positive real part. Hence the hyper-surface given by $S_k = S_k(s_1, \ldots, s_{k-1})$ is contained in the regularity domain of $\psi(s_1, \ldots, s_k)$, and then the righthand side of (5.17) must be zero. This gives the following relation for $\psi^{(k-1)}(s_1, \ldots, s_{k-1})$:

$$(\phi^{(k-1)} \cdot \psi^{(k-1)})(s_1, \ldots, s_{k-1}) = \frac{(\lambda - \sum_{i=1}^{k-1} s_i)\phi(s_1, \ldots, s_{k-1}, S_k)}{S_k} C_{k-1}.$$

By substituting back into Equation (5.17), we obtain the recursion

$$C_k = \frac{\lambda - \sum_{i=1}^{k-1} s_i}{\lambda - \sum_{i=1}^{k} s_i} \cdot \frac{S_k - s_k}{S_k} C_{k-1}$$

with initial condition $C_2 = -(1 - \rho_1)\frac{s_1}{\lambda - s_1 - s_2}\frac{S_2 - s_2}{S_2}$, which follows from (5.6). From this, we obtain (5.18), after rearranging the factors. The proof is complete.

Interpretation of the Rouché zero. It is worthwhile to change the coordinates: $(s_1, s_2, \ldots, s_k) \to (s_1, s_2, \ldots, s_{k-1}, \sum_{i=1}^{k} s_i)$. We can rewrite

$$\phi(s_1, \ldots, s_k) = \mathbb{E}e^{-s_1(B^{(1)} - B^{(k)}) - \ldots - s_{k-1}(B^{(k-1)} - B^{(k)}) - (\sum_{i=1}^{k} s_i)B^{(k)}}$$

Let us denote it by $\tilde{\phi}(s_1, \ldots, s_{k-1}, \sum_{i=1}^{k} s_i)$. This is the transform of the extra service time (relative to the shortest queue) in the first $k-1$ queues, together with the shortest one. It turns out there is a connection between $S_k(s_1, \ldots, s_{k-1})$ and the joint extra work in systems 1 to $k-1$ at the end of a busy period in system k. Let us denote this extra work by $(U_1, U_2, \ldots, U_{k-1})$, with LST $U_k^*(s_1, \ldots, s_{k-1})$, and let $F(x_1, x_2, \ldots, x_k)$ be the c.d.f. of $(B^{(1)} - B^{(k)}, \ldots, B^{(k-1)} - B^{(k)}, B^{(k)})$. Then by a similar argument as the one leading to (5.7), $U_k^*(s_1, \ldots, s_{k-1})$ satisfies the identity

$$\begin{aligned} U_k^*(s_1, \ldots, s_{k-1}) &= \int e^{-\sum_{i=1}^{k-1} s_i x_i} \sum_{n=0}^{\infty} \frac{(\lambda x_k)^n}{n!} e^{-\lambda x_k} [U_k^*(s_1, \ldots, s_{k-1})]^n F(\mathrm{d}x_1 \ldots \mathrm{d}x_k) \\ &= \tilde{\phi}(s_1, \ldots, s_{k-1}, \lambda[1 - U_k^*(s_1, \ldots, s_{k-1})]). \end{aligned} \tag{5.19}$$

Comparing this with the identity for the Rouché zero,

$$\lambda - (s_1 + \cdots + s_{k-1} + S_k) = \lambda\tilde{\phi}(s_1, \ldots, s_{k-1}, s_1 + \cdots + s_{k-1} + S_k),$$

gives the relation analogous to (5.8):

$$\lambda U_k^*(s_1,\ldots,s_{k-1}) = \lambda - (s_1 + \cdots + s_{k-1} + S_k), \tag{5.20}$$

which follows because the Rouché zero is unique.

Let us fix our attention on the case $k = 3$ for the moment. Then identity (5.18) becomes

$$\psi(s_1,s_2,s_3) = \frac{(1-\rho_3)(S_3-s_3)}{s_1+s_2+s_3-\lambda[1-\phi(s_1,s_2,s_3)]}\cdot\frac{1-\rho_2}{1-\rho_3}\frac{S_2-s_2}{S_3}\cdot\frac{1-\rho_1}{1-\rho_2}\frac{s_1}{S_2}. \tag{5.21}$$

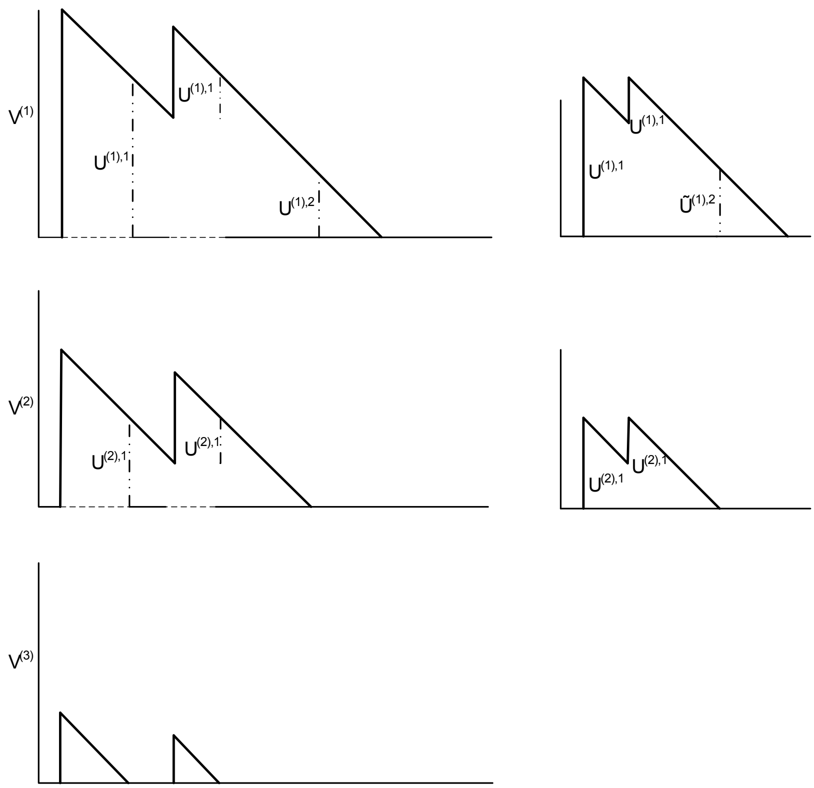


FIGURE 5.2: Work in the original system (left) and in the virtual system (right)

Work conservation. We would like to give a probabilistic interpretation of (5.21). In order to achieve this, we start by considering the joint extra work in queues 1 and 2 at the end of a busy period in queue 3. This has LST $U_3^*(s_1,s_2)$ as input in a 2-dimensional system with simultaneous Poisson arrivals, which is obtained by contracting the busy cycles in queue 3. We call this the 2-dimensional virtual system.

Remark that the inter-arrival times in the virtual system are precisely the idle periods in queue 3.

For this construction, the key observation is that the steady-state extra work in the virtual queue 1 at the end of the busy period in the virtual queue 2 is the same as the extra work in the initial queue 1 at the end of the busy period in the original queue 2. In analytic form, let $\tilde{U}_2^*(s_1)$ be the LST of the extra work in the virtual system and $U_2^*(s_1)$ be the LST of the extra work in the original system, see Figure 5.2.

Lemma 5.4.1.

$$\tilde{U}_2^*(s_1) = U_2^*(s_1).$$

Proof. We begin by remarking that the extra work $(U^{(1),1}, U^{(2),1})$ in the first 2 queues at the end of a busy period in queue 3 satisfies the a.s. inequality $U^{(1),1} \geq U^{(2),1}$. Since this is the input in the virtual system, from Proposition 5.2.1, $\tilde{U}_2^*(s_1)$ satisfies the identity (5.7) with $U_3^*(s_1, s_2)$ instead of $\phi(s_1, s_2)$:

$$U_3^*(s_1, \lambda[1 - \tilde{U}_2^*(s_1)] - s_1) = \tilde{U}_2^*(s_1). \tag{5.22}$$

At the same time, via (5.19), $U_3^*(s_1, s_2)$ satisfies

$$\phi(s_1, s_2, \lambda(1 - U_3^*(s_1, s_2)) - s_1 - s_2) = U_3^*(s_1, s_2).$$

If we substitute this fixed point identity in (5.22) above, we have

$$\phi(s_1, \lambda(1 - \tilde{U}_2^*(s_1)) - s_1, 0) = \tilde{U}_2^*(s_1).$$

On the other hand, this is also the identity (5.7) satisfied by $U_2^*(s_1)$, in the 2-dimensional system obtained by ignoring the last queue. Hence, from the uniqueness of Rouché's zero, $\tilde{U}_2^*(s_1) = U_2^*(s_1)$ (See Figure 5.2). This completes the proof.
We can rewrite (19) using (21):

$$\begin{aligned}\psi(s_1, s_2, s_3) = (1 - \rho_3) &\frac{s_1 + s_2 + s_3 - \lambda(1 - U_3^*(s_1, s_2))}{s_1 + s_2 + s_3 - \lambda(1 - \phi(s_1, s_2, s_3))} \\ &\cdot \frac{1 - \rho_2}{1 - \rho_3} \frac{s_1 + s_2 - \lambda(1 - U_2^*(s_1))}{s_1 + s_2 - \lambda(1 - U_3^*(s_1, s_2))} \cdot \frac{1 - \rho_1}{1 - \rho_2} \frac{s_1}{s_1 - \lambda(1 - \tilde{U}_2^*(s_1))}.\end{aligned} \tag{5.23}$$

Remark that the atom $\frac{1-\rho_1}{1-\rho_2}$ above is the conditional probability that queue 1 is empty, given that queue 2 is empty; and similarly for $\frac{1-\rho_2}{1-\rho_3}$. In addition, the last factor in (5.23) is the Pollaczek-Khinchine representation for an M/G/1 queue with service times having LST $\tilde{U}_2^*(s_1)$. Now we are ready to give the main result of this section.

Theorem 5.4.1. *In steady state, the joint workload distribution decomposes as an independent sum:*

$$(V^{(1)}, V^{(2)}, V^{(3)}) \stackrel{d}{=} (\tilde{V}^{(1),1}, \tilde{V}^{(2),1}, V^{(3)}) + (\tilde{V}^{(1),2}, V^{(2),2}, 0) + (V^{(1),3}, 0, 0).$$

The first term in the sum represents the steady-state distribution of the modified joint workload process obtained by removing the extra work in the first two queues at the *end of a busy period in the third queue. The second term is the workload in the first two queues obtained by removing the extra work in the first queue at the end of a busy cycle in the second queue. Finally the third term represents the workload in the virtual M/G/1 queue with input distributed as the extra work in queue 1, at the end of a busy period in queue 2.*

Proof. Consider the modified work process that evolves in steady state as

$$(\tilde{V}^{(1),1}, \tilde{V}^{(2),1}, V^{(3)}) \stackrel{d}{=} \left(\tilde{V}^{(1),1} + B^{(1)} - A, \tilde{V}^{(2),1} + B^{(2)} - A, V^{(3)} + B^{(3)} - A\right),$$

if $A < V^{(3)} + B^{(3)}$; and $(\tilde{V}^{(1),1}, \tilde{V}^{(2),1}, V^{(3)}) = (0,0,0)$, else.

By similar computations as the ones leading to Formula (5.9), we obtain

$$\tilde{\psi}(s_1, s_2, s_3) = (1-\rho_3)\frac{s_1 + s_2 + s_3 - \lambda(1 - U_3^*(s_1, s_2))}{s_1 + s_2 + s_3 - \lambda(1 - \phi(s_1, s_2, s_3))}.$$

This is the first factor in (5.23). For the second one, consider the following modified virtual workload process that evolves in steady state as

$$(\tilde{V}^{(1),2}, V^{(2),2}, 0) \stackrel{d}{=} \begin{cases} \left(\tilde{V}^{(1),2} + U^{(1),1} - A, V^{(2),2} + U^{(2),1} - A, 0\right), \\ \text{if } A < V^{(2),2} + U^{(2),1}, \\ (0,0,0), \text{ if } A \geq V^{(2),2} + U^{(2),1}, \end{cases}$$

with $(U^{(1),1}, U^{(2),1})$ the extra work vector in the first two queues at the end of a busy period in queue 3. Here we remove the excess workload in the virtual queue 1 at the end of the busy period in the virtual queue 2, which by Lemma 5.4.1 is the same as in the original system. In terms of LST's , this becomes

$$\tilde{\psi}_1(s_1, s_2) = \frac{1-\rho_1}{1-\rho_2}\frac{s_1 + s_2 - \lambda(1 - U_2^*(s_1))}{s_1 + s_2 - \lambda(1 - U_3^*(s_1, s_2))}.$$

Finally, the third factor in (5.23) is the Pollaczek-Khinchine representation of the steady-state workload in the M/G/1 queue with service time distributed as the extra work in queue 1 at the end of a busy period in queue 2. This completes the proof.

These considerations can be iterated now for the general k-dimensional system.

Corollary 5.4.1. *The steady-state joint workload in the k systems decomposes into the independent sum*

$$(V^{(1)}, \ldots, V^{(k)}) \stackrel{d}{=} (\tilde{V}^{(1),1}, \ldots, \tilde{V}^{(k-1),1}, V^{(k)}) + (\tilde{V}^{(1),2}, \ldots, \tilde{V}^{(k-2),2}, V^{(k-1),2}, 0) + \cdots + (\tilde{V}^{(1),k-1}, V^{(2),k-1}, 0, \ldots, 0) + (V^{(1),k}, 0, \ldots, 0),$$

where the jth term in the sum satisfies the identity in distribution (j=2,...,k):

$$(\tilde{V}^{(1),j}, \tilde{V}^{(2),j}, \ldots, \tilde{V}^{(k-j),j}, V^{(k-j+1),j}, 0, \ldots, 0) \stackrel{d}{=} \Big(\tilde{V}^{(1),j} + U^{(1),j-1} - A,$$

$$\tilde{V}^{(2),j} + U^{(2),j-1} - A, \ldots, V^{(k-j+1),j} - B^{(k-j+1)} - A, 0, \ldots, 0\Big),$$
$$\textit{if} \;\; A \leq V^{(k-j+1),j} - B^{(k-j+1)},$$

and $(0, \ldots, 0)$ *else.* $U^{(i),j}$ *is the extra workload in queue* i *at the end of a busy period in queue* $(k-j+1)$*, for* $i > k-j+1$*.*

5.5 The general two-dimensional workload/reinsurance problem

In this section we consider the general two-dimensional workload problem: pairs of customers arrive simultaneously at two parallel queues Q_1 and Q_2 according to a Poisson(λ) process, the nth pair requiring service times $(B_n^{(1)}, B_n^{(2)})$ with LST $\phi(s_1, s_2)$. We are interested in the steady-state workload vector $(V^{(1)}, V^{(2)})$ with LST $\psi(s_1, s_2)$. By the duality that is exposed in Section 2.2, $\psi(s_1, s_2)$ is also the Laplace transform (w.r.t. $u^{(1)}$ and $u^{(2)}$) of the probability that both portfolios of an insurance company with simultaneous claims $(B_n^{(1)}, B_n^{(2)})$, with initial capital $u^{(1)}$ and $u^{(2)}$, will survive.

In Section 5.2 we have determined $\psi(s_1, s_2)$ for the special case that $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$. We now show how the general case – $B_n^{(1)}$ and $B_n^{(2)}$ having an arbitrary joint distribution – has been solved in the literature (with the solution of that special case emerging as a degenerate solution). We shall successively discuss the contributions of Baccelli [21], De Klein [76] and Cohen [43], who have treated the two-dimensional workload problem with simultaneous arrivals in increasing generality. Starting point in all those three studies is the following functional equation for $\psi(s_1, s_2)$, which is derived by studying the 2-dimensional Markovian workload process during an infinitesimal amount of time Δt:

$$K(s_1, s_2)\psi(s_1, s_2) = s_2\psi_1(s_1) + s_1\psi_2(s_2), \quad \mathcal{R}e\ s_1, s_2 \geq 0. \tag{5.24}$$

Here the so-called *kernel* $K(s_1, s_2)$ is given by:

$$K(s_1, s_2) := s_1 + s_2 - \lambda(1 - \phi(s_1, s_2)), \tag{5.25}$$

and

$$\psi_1(s_1) := \mathbb{E}[e^{-s_1 V^{(1)}}(V^{(2)} = 0)], \quad \psi_2(s_2) := \mathbb{E}[e^{-s_2 V^{(2)}}(V^{(1)} = 0)], \tag{5.26}$$

with $(\cdot)$ denoting an indicator function.

Remark 5.5.1. *In the special case of Section 5.2, with* $\mathbb{P}(B^{(1)} \geq B^{(2)}) = 1$*, one has* $\psi_2(s_2) \equiv \mathbb{P}(V^{(1)} = 0)$*, because* $V^{(2)}$ *cannot be positive when* $V^{(1)} = 0$*. It then remains to find* $\psi_1(s_1)$*. This is done by observing (cf. Appendix B) that, for all* s_1 *with* $\mathcal{R}e\, s_1 > 0$*, there is a unique zero* $s_2(s_1)$ *of the kernel, with* $\mathcal{R}e\, s_2(s_1) > \mathcal{R}e\,(-s_1)$*. This immediately yields that* $\psi_1(s_1) = -\frac{s_1}{s_2(s_1)}\mathbb{P}(V^{(1)} = 0)$*, which is readily seen to be in agreement with (5.6).*

Equation (5.3), which was obtained by studying the workloads at arrival epochs (i.e., the waiting times; by PASTA they have the same distribution as the steady-state *workloads), looks slightly different from (5.24), but using (5.5) it is readily seen that they are equivalent.*

Globally speaking, the essential steps in [21, 76, 43] are the following.
Step 1: find a suitable set of zeroes $(\hat{s}_1, \hat{s}_2)$, *with* $\mathcal{R}e\, \hat{s}_1 \geq 0$, $\mathcal{R}e\, \hat{s}_2 \geq 0$, *of the kernel* $K(s_1, s_2)$, i.e., $K(\hat{s}_1, \hat{s}_2) = 0$. Because $\psi(s_1, s_2)$ is regular for all (s_1, s_2) with $\mathcal{R}e\ s_1, s_2 \geq 0$, one must have for all these zeroes:

$$\hat{s}_2 \psi_1(\hat{s}_1) = -\hat{s}_1 \psi_2(\hat{s}_2). \tag{5.27}$$

It is further observed that $\psi_1(s_1)$ is regular for $\mathcal{R}e\ s_1 > 0$, continuous for $\mathcal{R}e\ s_1 \geq 0$, and that $\psi_2(s_2)$ is regular for $\mathcal{R}e\ s_2 > 0$, continuous for $\mathcal{R}e\ s_2 \geq 0$.
Step 2: formulate a boundary value problem for $\psi_1(s_1)$ *and* $\psi_2(s_2)$. There are various types of boundary value problems, like the Riemann and the Wiener-Hopf boundary value problems. Typically, they ask to determine two functions $P_1(\cdot)$ and $P_2(\cdot)$, which satisfy a relation on a particular boundary $\hat{B}$, while $P_1(\cdot)$ is regular in the interior $\hat{B}^+$ and $P_2(\cdot)$ is regular in the exterior $\hat{B}^-$. $\hat{B}$ could be the unit circle (Riemann boundary value problem), or the imaginary axis (Wiener-Hopf boundary value problem; $\hat{B}^+$ now is the left-half plane). We refer to Gakhov [65] and Mushkelishvili [87] for excellent expositions of such boundary value problems and their variants, like the boundary value problem with a shift. The latter occurs in the approach of De Klein [76], see below.
Step 3: solve the boundary value problem for $\psi_1(\cdot)$ *and* $\psi_2(\cdot)$ *with boundary* $\hat{B}$. If $\hat{B}$ is a smooth closed contour that is not a circle, the use of a conformal mapping from $\hat{B}$ to the unit circle C is required to arrive at a Riemann boundary value problem for the unit circle, the solution of which can be found in [65, 87]. Thus one obtains $\psi_1(s_1)$ and $\psi_2(s_2)$ inside certain regions; subsequently, one may use analytic continuation to find them in $\mathcal{R}e\ s_1, s_2 \geq 0$. Finally, $\psi(s_1, s_2)$ follows from (5.24).

Remark 5.5.2. *Application of the boundary value method in queueing theory was pioneered by Fayolle and Iasnogorodski in [55]. They used this method to analyze the joint queue length process in two coupled processors, viz., two M/M/1 queues which operate at unit speeds when the other queue is not empty, but at different speeds when the other queue is empty. The method was subsequently developed in [41] for a large class of two-dimensional random walks; various queueing applications were also discussed in [41]. See [42] for a survey of the method in queueing theory, and see [56, 43] for two monographs which have further developed the theory of two-dimensional random walks. Part IV of [43] explores the analysis of* k*-dimensional random walks with* $k > 2$. *Results for* $k > 2$ *are very limited, and it seems fair to conclude that the boundary value method is, apart from a few special cases, restricted to two-dimensional random walks.*

Remark 5.5.3. *We strongly believe that the boundary value method also has a large potential in the analysis of two-dimensional risk models. Due to the duality between the reinsurance model and the* 2*-queue model with simultaneous arrivals, the publications of Baccelli [21], DeKlein [76], Cohen [43] are of immediate relevance to the reinsurance problem. These publications seem unknown in the insurance community (see, e.g., Chan et al. [37], who pose the two-dimensional risk problem and stop at Equation*

(5.24) (where [21, 76, 43] begin). They have remained largely unnoticed even in the queueing community, perhaps because of their complexity and because [21] and [76] did not appear in the open literature. For these reasons, we now successively expose the approaches in [21], [76] and [43] at some length.

The approach of Baccelli [21]
Baccelli [21] restricts himself to the case of exchangeable $(B^{(1)}, B^{(2)})$, i.e., $\mathbb{P}(B^{(1)} < x, B^{(2)} < y) = \mathbb{P}(B^{(1)} < y, B^{(2)} < x)$, or equivalently, $\phi(s_1, s_2) = \phi(s_2, s_1)$. We briefly review the three steps mentioned above.
Step 1 in [21] is as follows. Consider zero pairs $(\hat{s}_1, \hat{s}_2) = (g + iu, g - iu)$ of kernel $K(s_1, s_2)$, with $u \in \mathcal{R}$ and with $g = g(u)$ the unique zero in $\mathcal{R}e\ g \geq 0$ of

$$2g = \lambda(1 - \phi(g + iu, g - iu)).$$

Using the exchangeability, it can be shown that this unique zero is real and non-negative, while $g(-u) = g(u),\ u \in \mathcal{R}$.
Step 2. Consider the arc $\hat{A} = \{s_1 : s_1 = g(u) + iu,\ u \in R\}$, with $g(u)$ the zero defined above. This is a smooth arc, located in the right half-plane. Baccelli finds a conformal mapping $p(\cdot)$ of the interior C^+ of the unit circle C onto $\hat{A}^+$, the 'interior' of $\hat{A}$ located on the right of $\hat{A}$, and a conformal mapping $q(\cdot)$ of C^-, the exterior of the unit circle, onto $\hat{A}^+$; their limits on C are denoted by $p^+(z)$ and $q^-(z)$, which are each other's complex conjugates because of the exchangeability. Noticing that $p^+(-1) = q^-(-1) = 0$, he multiplies both sides of (5.27) with $1+z$. This yields (divide both sides of (5.27) by $\hat{s}_1\hat{s}_2$):

$$(1+z)\frac{\psi_1(p^+(z))}{p^+(z)} = -(1+z)\frac{\psi_2(q^-(z))}{q^-(z)}, \qquad |z| = 1. \tag{5.28}$$

Because of the regularity properties of the conformal mappings and of $\psi_1(s_1)$ and $\psi_2(s_2)$, $\mathcal{R}e\ s_1, s_2 > 0$, one now arrives at a simple boundary value problem: we have (5.28) for $|z| = 1$, while the left-hand side of (5.28) is regular for $|z| < 1$, and the right-hand side is regular for $|z| > 1$.
Step 3. The solution of this problem immediately follows from Liouville's theorem, cf. [102] p. 85:

$$\psi_1(p(z)) = \frac{\gamma + \delta z}{1+z} p(z), \ \ |z| < 1, \quad \psi_2(q(z)) = \frac{\gamma + \delta z}{1+z} q(z), \qquad |z| > 1.$$

Baccelli [21] shows that $\gamma = -\delta$, and determines the remaining unknown constant δ by normalization. Having thus determined $\psi_1(s_1)$ for $s_1 \in \hat{A}^+$, he uses analytic continuation to obtain $\psi_1(s_1)$ in the whole right half-plane; similarly for $\psi_2(s_2)$. Finally, substitution in (5.24) determines $\psi(s_1, s_2)$.

The approach of De Klein [76]
De Klein [76], pp. 119-168, studies the general case of an arbitrary joint distribution of $B^{(1)}$ and $B^{(2)}$.
Step 1 in [76] is as follows. He considers the same zero pairs as Baccelli (also suggesting another set of zero pairs on p. 132). $g(u)$ is no longer necessarily real, but for all real u there still is a unique zero $g(u)$.

Step 2. De Klein subsequently considers the simple, smooth arcs $\hat{A}_1 = \{s_1 : s_1 = g(u) + iu,\ u \in R\}$ and $\hat{A}_2 = \{s_2 : s_2 = g(u) - iu,\ u \in R\}$ in the right half-plane. Notice that $\hat{A}_1$ and $\hat{A}_2$ are each other's complex conjugates in the exchangeable case of Baccelli, but not in De Klein's more general case. De Klein now uses the (unique) one-to-one mapping $s_2 = \omega_2(s_1)$ from $\hat{A}_1$ onto $\hat{A}_2$ (with inverse $\omega_1(s_2)$) determined by the fact that, $\forall s_1 \in \hat{A}_1$, $(s_1, \omega_2(s_1))$ is a zero pair of the kernel. Similarly, $\forall s_2 \in \hat{A}_2$, $(\omega_1(s_2), s_2)$ is a zero pair. Hence the following must hold:

$$\psi_1(\omega_1(s_2)) = -\frac{\omega_1(s_2)}{s_2}\psi_2(s_2), \quad s_2 \in \hat{A}_2. \tag{5.29}$$

In addition, one has the regularity properties of the functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ which were listed below (5.27). Determination of functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ with these regularity properties and satisfying (5.29) is a so-called *shift problem*, a boundary value problem with a shift (cf. Sections 17 and 18 of [65]).

Step 3. Gakhov [65] mentions two methods to solve such problems: (i) reduce the problem to a Fredholm integral equation of the second kind, and (ii) reduce the problem to an ordinary Riemann boundary value problem, by means of conformal mappings. De Klein [76] explores the first method in Section II.4.2 and the second in Section II.4.3. We concentrate on the first method. De Klein first translates the shift problem to one on a finite smooth *closed* contour, via the conformal mapping $\zeta(z) = \frac{1-z}{1+z}$ (with inverse $z(\zeta) = \frac{1-\zeta}{1+\zeta}$) that maps $\hat{A}_i$ onto smooth closed contours T_i, $i = 1, 2$; he then applies Gakhov's first method. He obtains the following Fredholm integral equation of the second kind for an unknown function $G_1(\cdot)$ – which up to a constant equals $\log\{\psi_1(z(\cdot))\}$:

$$G_1(p_1) = \frac{1}{2\pi i}\int_{T_1} G_1(v_1)\Big[\frac{1}{v_1 - p_1} - \frac{\nu_2'(v_1)}{\nu_2(v_1) - \nu_2(p_1)} - \frac{1}{v_1 - c_1}\Big]\mathrm{d}v_1 + H_1(p_1), \quad p_1 \in T_1, \tag{5.30}$$

with $H_1(\cdot)$ some known function, c_1 some point in the interior of T_1, and for $\nu_2(v_1) = \zeta(\omega_2(z(v_1)))$, $v_1 \in T_1$. After having solved the integral equation (which can be done numerically in an efficient way, as shown by De Klein), one obtains $\psi_1(s_1)$ for $s_1 \in T_1$, and then $\psi_2(s_2)$ for $s_2 \in T_2$ via (5.27). The regularity of $\psi_1(s_1)$ in the interior T_1^+ subsequently allows one to obtain $\psi_1(s_1)$, $s_1 \in T_1^+$, as a Cauchy integral; similarly for $\psi_2(s_2)$, $s_2 \in T_2^+$. By analytic continuation, $\psi_1(s_1)$ and $\psi_2(s_2)$ are then also uniquely determined in $\mathcal{R}e\ s_1 \geq 0$ and $\mathcal{R}e\ s_2 \geq 0$, respectively. Finally, $\psi(s_1, s_2)$ again follows from (5.24).

De Klein also explores Gakhov's second method to treat the shift problem. However, this reduction to a Riemann boundary value problem requires a conformal mapping that itself must be determined by solving another Fredholm integral equation of the second kind. In Chapter II.6 he extensively investigates the numerical solution of both integral equations by means of the Nystrom or quadrature method. He obtains, a.o., accurate results for the mean sojourn time of a customer pair, viz., the time until both customers of a pair have left the system.

The approach of Cohen [43]

Cohen [43], Part III, considers a very general class of two-dimensional workload processes. Basically, he combines the model with simultaneous arrivals and the coupled processors model. The two servers have speeds r_1 and r_2 if they are both non-idle, and speeds $r^{(1)}$ and $r^{(2)}$ when the other server is idle. Furthermore, he also allows the possibility of different joint service requirement distributions if a customer pair arrives when at least one of the servers is idle. Finally, he explicitly allows single arrivals next to simultaneous arrivals (cf. also [23]). Much of Part III of [43] is devoted to a detailed study of the ergodicity conditions and of the so-called hitting point process and hitting point identity of the workload process, hitting point referring to the first entrance point of one of the axes.

In Chapter III.4 he determines the steady-state joint workload distribution for a variety of cases. For us, the most relevant cases are treated in Sections III.4.9 and III.4.10. Section III.4.9 treats the model of De Klein [76]. The same zero pairs are used (*Step 1*), and the same smooth closed contours T_1 and T_2; Cohen subsequently uses Gakhov's second method to arrive at a Riemann boundary value problem (*Step 2*). That boundary value problem actually is so simple that it can be solved straightforwardly by applying Liouville's theorem, cf. Baccelli's method above (*Step 3*); however, a conformal mapping is required, which is obtained as the solution of another Fredholm integral equation of the second kind. A nice feature in Section III.4.9 is that $\psi_1(s)$ and $\psi_2(t)$, after normalization, are expressed as LST's of waiting time or workload distributions of special M/G/1 queues (which are related to hitting points).

Section III.4.10 treats the model of De Klein with the additional feature that there is coupling of the servers, of a rather special form: $\frac{r_1}{r^{(1)}} + \frac{r_2}{r^{(2)}} = 1$. This does not change the kernel $K(s,t)$ (which only refers to the interior of the state space, with both servers active), so the same zero pairs and contours can still be used. However, it does change the right-hand side of (5.24), and hence a slightly different Riemann boundary value problem must be solved.

Remark 5.5.4. *It should be observed that Baccelli [21], De Klein [76] and Cohen [43] all also solve the more complicated* transient *problem, of determining the joint time-dependent distribution of the two workloads.*

5.6 Appendix B

Lemma 5.6.1 (Rouché zero)**.** *For every s_1 with $\mathcal{R}e\, s_1 > 0$ there exists a unique $s_2 = s_2(s_1)$ with $\mathcal{R}e\, s_2(s_1) > \mathcal{R}e\,(-s_1)$, that satisfies the identity*

$$\lambda\phi(s_1, s_2) = \lambda - (s_1 + s_2).$$

Moreover the function: $s_1 \to s_2(s_1)$ is analytic in $\mathcal{R}e\, s_1 > 0$.

Proof. For fixed s_1 with $\mathcal{R}e\, s_1 > 0$, let $f(s_1 + s_2) := \lambda - (s_1 + s_2)$. Consider in the right half-plane the contour $\mathcal{C}$ made up from the semicircle with center at $-s_1$ and radius $R > 2\lambda$ together with the line segment $I := \{-s_1 + iw | w \in [-R, R]\}$. We show that on this contour $|\lambda\phi(s_1, s_2)| < |f(s_1 + s_2)|$. We can bound $|\phi(s_1, s_2)|$ by

$$\lambda\,|\phi(s_1, s_2)| = \lambda|\tilde{\phi}(s_1, s_1 + s_2)| \le \lambda \mathbb{E} e^{-Re\ s_1(B^{(1)} - B^{(2)}) - Re\ (s_2 + s_1)B^{(2)}} < \lambda.$$

This holds everywhere in the domain of $\phi(s_1, s_2)$ if $B^{(1)} - B^{(2)}$ has positive mass on $(0, \infty)$.

Now we bound $|f(s_1+s_2)|$. When (s_1+s_2) is on the semicircle (i.e $|s_1+s_2| = R > 2\lambda$), apply the triangle inequality to the triangle with vertices at $0, \lambda, s_1+s_2$, to find $|\lambda - s_1 - s_2| > \lambda$. When $(s_1+s_2) \in I$, by a similar argument we obtain $|\lambda - s_1 - s_2| \geq \lambda$, with equality only when $s_1 + s_2 = 0$. Hence on the contour $\mathcal{C}$, $|f(s_1+s_2)| \geq \lambda$. We can now use Rouché's theorem to conclude that the equation $\lambda\phi(s_1, s_2) = \lambda - (s_1+s_2)$ has a unique solution $s_2(s_1)$ inside $\mathcal{C}$, because the polynomial $f(s_1+s_2)$ has only one zero inside $\mathcal{C}$, at λ. Letting $R \to \infty$, proves the assertion.

Chapter 6

Two parallel insurance lines with simultaneous arrivals

As mentioned in the introductory Section 1.2, in the existing risk and insurance literature, there are not many approaches towards analyzing multidimensional models.

The approach we take in this chapter combines ideas from Chapters 3 and 5. In Section 6.2 we derive a functional equation for the survival function related to a 2-dimensional risk reserve process, but unlike in Chapter 5, we do not assume that the claim intervals are exponentially distributed. A semi-Markov structure is assumed for the arrival process which is in many respects a two-dimensional (vector-valued) version of the BMAP structure of Chapter 3. To be more precise, we assume the claim size vector to be correlated with the time elapsed since the previous arrival. Such a correlation is quite natural; e.g., a claim event that generates very large claims could be subjected to additional administrative/regulatory delays. The type of correlation between the inter-arrival time and the vector of claim sizes is an extension to two dimensions of the dependence structure studied in Chapter 3 for the generalized Sparre-Andersen model. It involves making a rationality assumption regarding the trivariate LST of inter-arrival time and claim size vector (Assumption 6.1.1), which extends the case where the vector with the aforementioned components has a multivariate phase type distribution (MPH). In addition, we also make the assumption that the claim sizes are a.s. ordered (Assumption 6.1.2). Under these assumptions, we obtain our main result: An explicit expression for the (LST of the) two-dimensional survival function, for a large class of vectors of interclaim times and claim amounts of both risk reserves. For example, this is the case if we consider proportional reinsurance.

The chapter is organized in the following way. In Section 6.1 we describe the model in detail and present the main assumptions we will be working with.

For a start, an essential ingredient is the multivariate duality relation from Theorem 2.2.1. This relation makes it clear that determining the survival function is equivalent to determining the two-dimensional waiting time distribution in a dual two-queue two-server queueing model with simultaneous arrivals of customers at both queues. With the help of the embedded random walk/queueing process we derive, in Section 6.2, a stochastic recursion for the LST of the finite horizon survival function. In Section

6.3 we resolve the stationary version of this stochastic recursion, (6.8). The key tool used is a one-parameter Wiener-Hopf factorization of the bivariate kernel appearing in (6.8). More precisely, the Wiener-Hopf factors will depend on one parameter, which is the first argument of that bivariate kernel; see Proposition 6.3.1 in Section 6.3. The main result, Theorem 6.3.1, gives the LST of the survival function, or equivalently the stationary distribution of the waiting time/reflected random walk inside the non-negative quadrant (see also Theorem 2.2.1). In Section 6.4, we give an asymptotic decomposition result for the joint LST of the waiting time, which can be written as a product, with one of the factors being the transform of the waiting time in the first queue conditional that the other queue is empty. This is an extension of the decomposition result obtained in (5.10), Chapter 5.

In Section 6.5 we explain how to calculate the transform obtained in Theorem 6.3.1 for some examples, and we numerically calculate the ruin functions/waiting time distributions using an efficient inversion algorithm of den Iseger [70] - which is about using the Gauss quadrature method to discretize the LST and then apply the inversion scheme of Abate and Whitt [1]. Finally, we also point out that the numerical results suggest that the ruin functions appear to be stochastically ordered for various types of correlations between inter-arrival times and claim sizes, a positive correlation apparently leading to smaller ruin probabilities.

6.1 Model description

Let us begin with the general assumptions on the two risk reserves. These are started with non-negative initial capital $(u^{(1)}, u^{(2)})$; as long as there are no arrivals, the surplus levels increase linearly with positive rates (c_1, c_2). At the nth claim arrival epoch, claim sizes $B_n^{(1)}$ and $B_n^{(2)}$ are respectively requested from each component. The time between the $(n-1)$th and nth claim arrival is denoted by A_n. The sequence $\{(A_n, B_n^{(1)}, B_n^{(2)})\}_{n\geq 1}$, is assumed to be an i.i.d. sequence, but within a triple, $(A_n, B_n^{(1)}, B_n^{(2)})$ are allowed to be correlated. We will use A, $B^{(1)}$, $B^{(2)}$ respectively for the generic inter-arrival time and claim sizes. In the above-described very general set-up, the following assumption will allow us to explicitly determine the ruin/survival probabilities by using Wiener-Hopf factorization:

Assumption 6.1.1 (On the joint transform of A, $\mathbf{B}^{(1)}$, $\mathbf{B}^{(2)}$). *The triple transform*

$$H(q_0, q_1, q_2) := \mathbb{E}e^{-q_0 A - q_1 B^{(1)} - q_2 B^{(2)}} \tag{6.1}$$

is a rational function in q_i, i.e. it has representation $\frac{G(q_0,q_1,q_2)}{L(q_0,q_1,q_2)}$ such that $G(q_0, q_1, q_2)$ and $L(q_0, q_1, q_2)$ are polynomials in the variables q_i, $i = 0, 1, 2$.

$G(q_0, q_1, q_2)$ and $L(q_0, q_1, q_2)$ must satisfy some conditions, because their ratio is a transform, such as,

$$\lim_{|q_0|\to\infty,\, \mathcal{R}e\, q_0>0} H(q_0, q_1, q_2) = \mathbb{E}[e^{-q_1 B^{(1)} - q_2 B^{(2)}} 1_{\{A=0\}}].$$

We can assume without loss of generality that $A > 0$ *a.s.* Because of the above limit, this means the degree of G as a polynomial in q_0, $G_{q_1,q_2}(q_0)$, is strictly less than the degree of L as a polynomial in q_0: $L_{q_1,q_2}(q_0)$, for all q_1 and q_2.

Remark 6.1.1. *The class of rational multivariate Laplace-Stieltjes transforms contains the class of LSTs of multivariate Phase-Type distributions (MPH); see Bladt and Nielsen [29], where the rational transform class is called multivariate matrix-exponential (MME). All of the well-known classes of multivariate Phase-Type distributions are MME. Actually, all the examples we will present are MPH distributions with a specific correlation structure which are a special case of Kulkarni's MPH* class [79]. There is no point in restricting ourselves to any of these subclasses. We will fully exploit the algebraic representation of rational functions in order to derive explicitly the transforms of the two-dimensional survival/ruin functions.*

The risk reserve process $R_t = (R_t^{(1)}, R_t^{(2)})$ is defined in (2.10). The associated types of survival measures and the related ruin functions are defined in Section 2.2. This chapter is mainly concerned with measuring the event that both risk reserve processes survive indefinitely, i.e., we aim to determine the *survival function* $F^s(u^{(1)}, u^{(2)})$, or equivalently, the ruin function $\tau_\wedge(u^{(1)}, u^{(2)})$, as defined by (2.11).

In this model, the two reserves are correlated due to simultaneous claim arrivals *and* due to correlations that may exist in the claim size vector $(B^{(1)}, B^{(2)})$.

From this point on we scale the risk reserve process by the income rates in each component (see Remark 5.2.1). The main idea is to exploit the fact that the embedded process at arrival epochs of claims is a random walk in the plane with increments $(A_n - B_n^{(1)}/c_1, A_n - B_n^{(2)}/c_2)$, conditioned on starting at $(u^{(1)}, u^{(2)})$. The different ways in which the risk reserve process can be ruined correspond to the possible positions of the random walk

$$R_{t_{n+1}} := u - \sum_{i=1}^{n} X_i, \qquad X_n = (B_n^{(1)}/c_1 - A_n, B_n^{(2)}/c_2 - A_n), \tag{6.2}$$

at the time of exit from the non-negative quadrant.

This is a difficult model to analyze in full generality; in particular, it is more general than the two-dimensional ruin process described in Chan et al. [37] and in Chapter 5. However, in the case that the claims in the first insurance line are larger than those in the second, we are able to determine the two-dimensional survival function.

Assumption 6.1.2 (**Ordering assumption**)**.** *For a generic claim event* $(B^{(1)}, B^{(2)})$*:*

$$B^{(1)}/c_1 \geq B^{(2)}/c_2 \;\; a.s. \tag{6.3}$$

A special, important example for this ordering assumption is the case when there is a single arrival process such that the common claim is partitioned into fixed proportions $(\alpha, 1-\alpha)$, and we can always take w.l.o.g. $\alpha \in [1/2, 1]$ (α may even be a random variable with this interval as support). This special case then significantly generalizes the setting in Avram et al. [19], where it is assumed that the common arrival process is compound Poisson (and in particular the inter-arrival times are independent of the claim sizes).

One can go a step further and assume there is a dedicated renewal-arrival stream of claims into the line which pays the greater share, say α. This is in line with the assumption in Badescu et al. [23] of a dedicated Poissonian stream of claims, and extends it, once we combine it with Assumption 6.1.1. Clearly this is not a proportional reinsurance problem anymore. From a mathematical perspective, the analysis is more insightful than if one would just assume proportionality.

We take the following approach: using the duality arguments (Section 2.2), we derive a recursive equation for the survival function (Section 6.2). Using complex function theory, and under Assumptions 6.1.1 and 6.1.2, we solve the functional equation that corresponds to the stochastic recursion in terms of survival function LSTs (Section 6.3).

6.2 A functional equation

In this section we will consider the Laplace-Stieltjes transform of the survival function $F_n^s(u^{(1)},u^{(2)})$. Theorem 2.2.1 and the usual relation between the waiting time vector and the embedded workload vector $W_n = V_{t_n}$ (we assumed unit income/service rates, see Remark 5.2.1) implies that this equals the transform of the bivariate waiting time for the nth customer:

$$\mathbb{E}e^{-s_1 W_n^{(1)} - s_2 W_n^{(2)}} = \int e^{-s_1 u^{(1)} - s_2 u^{(2)}} \mathrm{d}F_n^s(u^{(1)}, u^{(2)}), \quad \mathcal{R}e\, s_i \geq 0, \quad i = 1, 2. \tag{6.4}$$

Our main goal in this section is to obtain a recursion between the LSTs of $(W_{n+1}^{(1)}, W_{n+1}^{(2)})$ and $(W_n^{(1)}, W_n^{(2)})$ using Lindley's recursion (2.9). But first we point out a sample-path relation between the risk reserve process and the reflected random walk which will be useful in the next section (the relation in Theorem 2.2.1 holds in distribution only). The event that any of the risk reserves is running at the maximum is the same as the event that the corresponding component of the reflected random walk is at 0.

Lemma 6.2.1. *For $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, it holds*

$$\{W_{n+1}^{(i)} = 0\} = \{R_{t_{n+1}}^{(i)} = u^{(i)} + \max(0, -S_1^{(i)}, ..., -S_n^{(i)})\}, \quad i = 1, 2.$$

Proof. $R_{t_{n+1}}^{(i)} = u^{(i)} - S_n^{(i)}$ and notice the following equivalence:

$$S_n^{(i)} - \min(0, S_1^{(i)}, \ldots, S_n^{(i)}) = 0 \Leftrightarrow -S_n^{(i)} = \max(0, -S_1^{(i)}, \ldots, -S_n^{(i)}).$$

The statement now follows from Proposition 2.2.1.

Remark 6.2.1. *The event $\{W_n^{(i)} = 0\}$ does not depend on the initial capital. The event on the RHS in Lemma 6.2.1 above does not restrict the risk reserve process to staying above 0; equivalently $W_n^{(i)}$ on the LHS is not restricted to staying below level $u^{(i)}$. This is in line with (2.13) in Theorem 2.2.1.*

Below we point out how one can obtain a recursive equation for the LST of the survival function $F_n^s(u^{(1)}, u^{(2)})$ in the general case without the ordering assumption. In Section 6.3 we solve this equation for the case when the risks are ordered. The Lindley recursion (2.9) becomes in terms of LSTs:

$$\mathbb{E}e^{-s_1 W_{n+1}^{(1)} - s_2 W_{n+1}^{(2)}} = \mathbb{E}e^{-s_1(W_n^{(1)}+X_n^{(1)})^+ - s_2(W_n^{(2)}+X_n^{(2)})^+},$$

with $(x)^+ = \max(x, 0)$. Hence, with $1_{\{E\}}$ the indicator function of event E,

$$\begin{aligned}\mathbb{E}e^{-s_1 W_{n+1}^{(1)} - s_2 W_{n+1}^{(2)}} =& \mathbb{E}\left[e^{-s_1(W_n^{(1)}+X_n^{(1)}) - s_2(W_n^{(2)}+X_n^{(2)})} 1_{\{-X_n^{(1)}<W_n^{(1)},\, -X_n^{(2)}<W_n^{(2)}\}}\right] \\ &+ \mathbb{E}\left[e^{-s_1(W_n^{(1)}+X_n^{(1)})} 1_{\{-X_n^{(1)}<W_n^{(1)},\, -X_n^{(2)}\geq W_n^{(2)}\}}\right] \\ &+ \mathbb{E}\left[e^{-s_2(W_n^{(2)}+X_n^{(2)})} 1_{\{-X_n^{(1)}\geq W_n^{(1)},\, -X_n^{(2)}<W_n^{(2)}\}}\right] \\ &+ \mathbb{P}(-X_n^{(1)} \geq W_n^{(1)}, -X_n^{(2)} \geq W_n^{(2)}). \end{aligned} \tag{6.5}$$

In view of (6.4), the left-hand side of (6.5) represents the LST of the survival measure F_{n+1}^s. Below we also interpret each of the four terms in the righthand side in terms of transforms of survival measures.

Term 1: In terms of excursions away from the maximum of the surplus process (Lemma 6.2.1), the first term on the RHS represents the transform of the survival measure in the event that both risk reserves are during an excursion below the running maximum at time t_{n+1},

$$\begin{aligned}&\mathbb{E}e^{-s_1 W_{n+1}^{(1)} - s_2 W_{n+1}^{(2)}} 1_{\{W_{n+1}^{(1)}>0,\, W_{n+1}^{(2)}>0\}} = \\ &\qquad \int e^{-s_1 u^{(1)} - s_2 u^{(2)}} 1_{\{R_{t_{n+1}} < \bigvee_{k=0,\ldots,n} R_{t_k}\}} \mathrm{d}F_{n+1}^s(u^{(1)}, u^{(2)}).\end{aligned}$$

Above we used that on the event $\{W_{n+1}^{(i)} > 0\}$, it holds that $W_n^{(i)} + X_n^{(i)} = W_{n+1}^{(i)}$.

Terms 2 and 3: These can be translated in terms of survival functions using (6.4) again:

$$\begin{aligned}&\mathbb{E}e^{-s_1 W_{n+1}^{(1)}} 1_{\{W_{n+1}^{(1)}>0,\, W_{n+1}^{(2)}=0\}} = \\ &\qquad \int_{0+}^{\infty} e^{-s_1 u^{(1)}} 1_{\{R_{t_{n+1}}^{(1)} < \max_{k=0,\ldots,n} R_{t_k}^{(1)};\, R_{t_{n+1}}^{(2)} \geq \max_{k=0,\ldots,n} R_{t_k}^{(2)}\}} \mathrm{d}F_{n+1}^s(u^{(1)}, 0),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}e^{-s_2 W_{n+1}^{(2)}} 1_{\{W_{n+1}^{(1)}=0,\, W_{n+1}^{(2)}>0\}} = \\ &\qquad \int_{0+}^{\infty} e^{-s_2 u^{(2)}} 1_{\{R_{t_{n+1}}^{(1)} \geq \max_{k=0,\ldots,n} R_{t_k}^{(1)};\, R_{t_{n+1}}^{(2)} < \max_{k=0,\ldots,n} R_{t_k}^{(2)}\}} \mathrm{d}F_{n+1}^s(0, u^{(2)}),\end{aligned}$$

are 'boundary' transforms. The survival measure that corresponds to F_n^s can have positive mass on the axes of the non-negative quadrant, if nowhere else, at least $F_n^s(0,0) = \mathbb{P}(W_n = 0 \,|\, W_0 = 0)$ is positive, i.e., it has an atom in the origin.
Term 4: This is the probability that both risk reserves are surviving and running at a maximum, which by Lemma 6.2.1 and Theorem 2.2.1 is $F_{n+1}^s(0,0)$.

From an analytic point of view it is more convenient to rewrite (6.5) as a recursion. After adding and subtracting appropriate terms, one obtains,

$$
\begin{aligned}
\mathbb{E}e^{-s_1 W_{n+1}^{(1)} - s_2 W_{n+1}^{(2)}} = \\
&\mathbb{E}e^{-s_1 X_n^{(1)} - s_2 X_n^{(2)}}\, \mathbb{E}e^{-s_1 W_n^{(1)} - s_2 W_n^{(2)}} \\
&+ \mathbb{E}\left\{ e^{-s_1(W_n^{(1)}+X_n^{(1)})}[1 - e^{-s_2(W_n^{(2)}+X_n^{(2)})}]1_{\{W_{n+1}^{(1)}>0,\, W_{n+1}^{(2)}=0\}} \right\} \\
&+ \mathbb{E}\left\{ e^{-s_2(W_n^{(2)}+X_n^{(2)})}[1 - e^{-s_1(W_n^{(1)}+X_n^{(1)})}]1_{\{W_{n+1}^{(1)}=0,\, W_{n+1}^{(2)}>0\}} \right\} \\
&+ \mathbb{E}\left\{ [1 - e^{-s_1(W_n^{(1)}+X_n^{(1)}) - s_2(W_n^{(2)}+X_n^{(2)})}]1_{\{W_{n+1}^{(1)}=0,\, W_{n+1}^{(2)}=0\}} \right\}.
\end{aligned}
\tag{6.6}
$$

Above we used that the increment X_n is independent of W_n. Under the assumption that the vector W_n has a limit in distribution, W, as $n \to \infty$, (6.6) becomes

$$
\begin{aligned}
K(s_1, s_2)\mathbb{E}e^{-s_1 W^{(1)} - s_2 W^{(2)}} = \\
&\mathbb{E}\left\{ e^{-s_1(W^{(1)}+X^{(1)})}[1 - e^{-s_2(W^{(2)}+X^{(2)})}]1_{\{W^{(1)}>-X^{(1)},\, W^{(2)}\le -X^{(2)}\}} \right\} \\
&+ \mathbb{E}\left\{ e^{-s_2(W^{(2)}+X^{(2)})}[1 - e^{-s_1(W^{(1)}+X^{(1)})}]1_{\{W^{(1)}\le -X^{(1)},\, W^{(2)}>-X^{(2)}\}} \right\} \\
&+ \mathbb{E}\left\{ [1 - e^{-s_1(W^{(1)}+X^{(1)}) - s_2(W^{(2)}+X^{(2)})}]1_{\{W^{(1)}\le -X^{(1)},\, W^{(2)}\le -X^{(2)}\}} \right\},
\end{aligned}
\tag{6.7}
$$

with "kernel" $K(s_1, s_2) := 1 - \mathbb{E}e^{-s_1 X^{(1)} - s_2 X^{(2)}}$ and $\mathcal{R}e\, s_i = 0$, $i = 1, 2$.

Remark 6.2.2. *From Lemma 2.2.1, W has the same distribution as the all-time supremum $M := \lim_{n\to\infty} M_n$; and M_n being a sequence of non-decreasing random vectors w.r.t. the componentwise order '$\le$', the limit always exists a.s., although it may have a defective distribution. In the next section we give a sufficient condition for M to have a proper distribution under the assumption that risks are ordered.*

6.3 Wiener-Hopf analysis of the stochastic recursion

In this section we resolve the functional equation (6.7) under the Assumptions 6.1.1 and 6.1.2, i.e., we find the LST of the infinite horizon survival function:

$$F^s = \lim_{n\to\infty} F_n^s,$$

the limit being considered in distribution. Theorem 2.2.1 together with a limit argument shows that this weak limit is the same as the c.d.f. of the stationary version of the waiting time process $(W_n)_{n\geq 0}$.

The section is divided into three subsections. Subsection 6.3.1 prepares the ground, by making a key observation about the functional equation (6.7), introducing some notation and discussing the stability condition. Subsection 6.3.2 expresses the two-dimensional LST $\psi(s_1, s_2)$ of F^s in a one-dimensional unknown function $C(s_1)$ (Proposition 6.3.1). That function is determined in Subsection 6.3.3, yielding our main result: Theorem 6.3.1.

6.3.1 Preparations

Introduce the extra claim amount $\delta_n := B_n^{(1)}/c_1 - B_n^{(2)}/c_2 = X_n^{(1)} - X_n^{(2)}$, so that the increments of the random walk S_n can be represented as $X_n = (X_n^{(2)} + \delta_n, X_n^{(2)})$.

We first make the following *key observation*: The ordering assumption (6.3) implies that when $R_{t_n}^{(1)}$ is at a maximum, necessarily $R_{t_n}^{(2)}$ is at a maximum. Via Lemma 6.2.1, this corresponds to the fact that the events

$$\{W_n^{(1)} \leq -X_n^{(1)}, W_n^{(2)} > -X_n^{(2)}\} = \{W_{n+1}^{(1)} = 0, W_{n+1}^{(2)} > 0\}$$

are null for all $n \geq 0$. This means that the third term on the RHS of (6.6) is null and hence the second term on the RHS of (6.7) vanishes as well, so that after regrouping terms, (6.7) can be rewritten as

$$K(s_1, s_2)\psi(s_1, s_2) = -\psi_1(s_1, s_2) + \psi_2(s_1) + \mathbb{P}(W^{(1)} + X^{(1)} \leq 0), \tag{6.8}$$

where

$$\psi_1(s_1, s_2) = \mathbb{E}e^{-s_1(W^{(1)}+X^{(2)}+\delta)-s_2(W^{(2)}+X^{(2)})}1_{\{-X^{(2)}\geq W^{(2)}\}},$$
$$\psi_2(s_1) = \mathbb{E}e^{-s_1(W^{(1)}+X^{(2)}+\delta)}1_{\{W^{(1)}+\delta>-X^{(2)}\geq W^{(2)}\}}.$$

Consider the following function:

$$\tilde{K}(s_1, z) := 1 - \mathbb{E}e^{-s_1\delta - zX^{(2)}}, \quad \mathcal{R}e\, s_1 \geq 0,\ \mathcal{R}e\, z = 0.$$

This is related to $K(s_1, s_2)$ that appears in (6.7) through the change of coordinates: $\tilde{K}(s_1, z) = K(s_1, z - s_1)$. In addition, remark that for fixed z, $\tilde{K}(s_1, z)$ is indeed analytic in $\mathcal{R}e\, s_1 > 0$ because $\delta \geq 0$ a.s. Now let us change the coordinates: $(s_1, s_2) \to (s_1, s_1 + s_2) =: (s_1, z)$. Denote $\tilde{\psi}(s_1, z) := \psi(s_1, z - s_1)$, and similarly for $\tilde{\psi}_1(s_1, z) := \psi_1(s_1, z - s_1)$. Then $\psi_1(s_1, s_2)$ becomes

$$\psi_1(s_1, s_2) = \mathbb{E}\left[e^{-s_1(W^{(1)}-W^{(2)}+\delta)-z(W^{(2)}+X^{(2)})}1_{\{-X^{(2)}\geq W^{(2)}\}}\right] =: \tilde{\psi}_1(s_1, z),$$

and therefore (6.8) can be rewritten as

$$\tilde{K}(s_1, z)\tilde{\psi}(s_1, z) = -\tilde{\psi}_1(s_1, z) + \psi_2(s_1) + \mathbb{P}(-X^{(1)} \geq W^{(1)}). \tag{6.9}$$

Running example: One of the simplest examples (cf. Example 3 of Section 3.4) that we will use throughout is obtained when taking the joint distribution of $(A, \delta, B^{(2)})$ to be such that conditional on a random variable κ, these A, δ and $B^{(2)}$ are independent and have Erlang distributions of order κ and rates respectively λ, μ_δ and μ. To keep things as simple as possible, we choose $\mathbb{P}(\kappa = 1) = \mathbb{P}(\kappa = 2) = \frac{1}{2}$ and rates $\lambda = 1$, $\mu_\delta = 3$, $\mu = 2$; we also choose the income rates c_1 and c_2 to be equal to 1.

The kernel $\tilde{K}(s_1, z)$ has the following simple form:

$$\tilde{K}(s_1, z) = 1 - \frac{3(3 + s_1)(1 - z)(2 + z) + 18}{(3 + s_1)^2(1 - z)^2(2 + z)^2}. \tag{6.10}$$

From a specific example like this it becomes clear that the coefficients of the numerator, for example, as a polynomial in z are themselves polynomials in s_1 and vice versa.

We are now ready to formulate a Wiener-Hopf boundary value problem in variable z. For fixed s_1, $\mathcal{R}e\, s_1 > 0$, $\tilde{\psi}_1(s_1, z)$ is analytic in $\mathcal{R}e\, z < 0$ (by analytic continuation), while $\tilde{\psi}(s_1, z)$ is analytic (by analytic continuation) in $\mathcal{R}e\, z > 0$. These statements follow easily from the probabilistic nature of these functions. In particular, notice that $\tilde{\psi}(s_1, z) = \mathbb{E}\mathrm{e}^{-s_1(W^{(1)} - W^{(2)}) - zW^{(2)}}$. On the event $\{W^{(2)} < -X^{(2)}\}$, the random variable $e^{-z(W^{(2)} + X^{(2)})}$ is uniformly bounded in $\mathcal{R}e\, z \le 0$, hence the analyticity of $\tilde{\psi}_1(s_1, z)$ follows by an application of Lebesgue's dominated convergence theorem.

The approach we take in order to solve (6.9) uses a Wiener-Hopf factorization with a parameter. More precisely, for each fixed s_1, $\mathcal{R}e\, s_1 > 0$, we will factorize the bivariate kernel $\tilde{K}(s_1, z) = \tilde{K}_{s_1}(z)$ that appears in (6.9) into $\tilde{K}_{s_1}(z) = \tilde{K}^{+}_{s_1}(z)\tilde{K}^{-}_{s_1}(z)$, such that $\tilde{K}^{+}_{s_1}(z)$ can be analytically continued in $\mathcal{R}e\, z > 0$ and $\tilde{K}^{-}_{s_1}(z)$ can be analytically continued in $\mathcal{R}e\, z < 0$. The Wiener-Hopf factorization that solves (6.9) is discussed in the next subsection. Finally, once we resolve (6.9), the solution to (6.8) follows by reverting to the original coordinate system (s_1, s_2).

Remark 6.3.1. *A reason why we prefer the notation using the argument s_1 as a subscript $\tilde{K}^{\pm}_{s_1}(z)$ for these factors is that they are in general obtained by pasting together different branches of multi-valued complex functions in s_1 using analytic continuation. More precisely, since $\tilde{K}(s_1, z)$ is a rational function, the 1-parameter Wiener-Hopf factors may have branch cuts in $\mathcal{R}e\, s_1 > 0$ (discontinuities) as functions of the argument s_1; then as it follows from Proposition 6.3.1 below, we must choose the values of the zeroes of the kernel that have positive real part for $\mathcal{R}e\, s_1 > 0$ and glue them together (using analytic continuation). Because of this, the 1-parameter Wiener-Hopf factors $\tilde{K}^{+}_{s_1}(z)$ and $\tilde{K}^{-}_{s_1}(z)$ are not functions of s_1 in the real sense. This will also be the case with the zeroes of the kernel from Example 2 in Section 6.5.*

Finally, a word about conditions under which the limiting distribution of the two-dimensional waiting time process $\{W^{(1)}_n, W^{(2)}_n\}_{n=1,2,\ldots}$ exists, or equivalently, under which indefinite survival of both risk reserves has a positive probability. It will turn out from the analysis below that a necessary condition for the existence of a proper limit in distribution W is $\rho_1 := \mathbb{E}B^{(1)}/(c_1\mathbb{E}A) < 1$. This is easy to interpret in our case, because it is sufficient to ensure positive safety loading for line 1 which receives always larger claims - we then automatically have positive safety loading for the second insurance line. The safety loading condition for the second risk reserve process $\rho_2 := \mathbb{E}B^{(2)}/(c_2\mathbb{E}A) < 1$ will be necessary for the Wiener-Hopf factorization to hold. The two Wiener-Hopf factors will be initially determined up to a certain unknown 'boundary' function $C(s_1)$ that appears in Equation (6.11) below; we further determine this boundary function by noting that the marginal risk reserve process $R_t^{(1)}$ behaves as a (one-dimensional) generalized Sparre-Andersen risk reserve process with dependence between inter-arrival times and subsequent claim sizes, for which the analysis of the survival function is available in Chapter 3. At this point the safety loading condition $\rho_1 < 1$ becomes necessary.

6.3.2 A Wiener-Hopf factorization

In this subsection we determine the double transform $\tilde{\psi}(s_1, z)$ up to a – yet – unknown single argument function $C(s_1)$. In the next subsection we will determine this function, which turns out to be related to the first insurance line only.

Proposition 6.3.1 (Wiener-Hopf factorization with a parameter). *Under Assumption 6.1.1 the double LST $\tilde{\psi}(s_1, z)$ is of the form*

$$\tilde{\psi}(s_1, z) = \mathbb{E}e^{-s_1(W^{(1)}-W^{(2)})-zW^{(2)}} = C(s_1)\, \tilde{K}^+_{s_1}(z)^{-1}. \tag{6.11}$$

$C(s_1)$ is a yet to be determined analytic function, $\mathcal{R}e\, s_1 > 0$. For fixed $\mathcal{R}e\, s_1 > 0$, $\tilde{K}^+_{s_1}(z)$ is analytic for $\mathcal{R}e\, z > 0$, continuous up to the boundary and it factorizes $\tilde{K}(s_1, z)$ into

$$\tilde{K}(s_1, z) = \tilde{K}^+_{s_1}(z)\, \tilde{K}^-_{s_1}(z),$$

such that $\tilde{K}^-_{s_1}(z)$ is analytic for $\mathcal{R}e\, z < 0$ and continuous up to the boundary.

Proof. Under the above Assumption 6.1.1, the triple transform of A, $B^{(1)}/c_1$, $B^{(2)}/c_2$ is
$H(q_0, q_1/c_1, q_2/c_2)$ and the kernel is $\tilde{K}(s_1, z) = 1 - H(-z, s_1/c_1, (z - s_1)/c_2)$ which is also a rational function with representation $\tilde{K}(s_1, z) := 1 - \frac{f(s_1,z)}{g(s_1,z)}$. The assumption that $A > 0$ a.s. implies that the degree of $f_{s_1}(z)$ is strictly less than the degree of $g_{s_1}(z)$ as polynomial functions in the argument z (see the discussion in Remark 6.3.1). Now the functional equation (6.9) becomes

$$\frac{g(s_1, z) - f(s_1, z)}{g(s_1, z)} \tilde{\psi}(s_1, z) = -\tilde{\psi}_1(s_1, z) + \psi_2(s_1) + \mathbb{P}(W^{(1)} + X^{(1)} \leq 0). \tag{6.12}$$

The first step is to factorize the kernel into two factors with respect to the z variable and regroup (6.12) into an analytic function in $\mathcal{R}e\, z > 0$ on the LHS and an analytic function in $\mathcal{R}e\, z < 0$ on the RHS. Once we have this representation, we can use Liouville's Theorem to determine both sides of (6.12) up to the function $C(s_1)$.

Remove all the poles with non-negative real part from the LHS of (6.12). We keep for now s_1 fixed with $\mathcal{R}e\, s_1 \geq 0$, and denote by

$$g_{s_1}^-(z) := \prod_{i:Re\ z_i(s_1)<0} (z - z_i(s_1)),$$

where $z_i(s_1)$ are zeroes of $g(s_1, z) = g_{s_1}(z)$; also put $g_{s_1}^+(z) := \frac{g_{s_1}(z)}{g_{s_1}^-(z)}$, so that we have the factorization $g_{s_1}(z) = g_{s_1}^+(z)\, g_{s_1}^-(z)$. Upon multiplying both sides of (6.12) by $g_{s_1}^+(z)$, the LHS becomes analytic for all $\mathcal{R}e\, z > 0$ and continuous up to the imaginary axis. Similarly, the RHS is analytic for all $\mathcal{R}e\, z < 0$ and continuous for $\mathcal{R}e\, z \leq 0$. Since these two coincide for $\mathcal{R}e\, z = 0$, they are analytic continuations of each other, in particular $\frac{g_{s_1}(z) - f_{s_1}(z)}{g_{s_1}^-(z)} \tilde{\psi}_{s_1}(z)$ is an entire function in z. Because $\deg f_{s_1}(z) \leq \deg g_{s_1}(z)$, asymptotically

$$\frac{g_{s_1}(z) - f_{s_1}(z)}{g_{s_1}^-(z)} \tilde{\psi}_{s_1}(z) = O(z^{m_+(s_1)}),$$

where $m_+(s_1) := \deg g_{s_1}^+(z)$. By virtue of Liouville's theorem ([102], p. 85),

$$\tilde{\psi}_{s_1}(z) = \frac{g_{s_1}^-(z)}{g_{s_1}(z) - f_{s_1}(z)} P_{s_1}(z), \tag{6.13}$$

where (for fixed $s_1 \geq 0$), $P_{s_1}(z)$ is a polynomial in z with $\deg P_{s_1}(z) \leq m_+(s_1)$. From (6.13) it follows immediately that $P_{s_1}(z)$ must have all the zeroes with non-negative real part of the denominator $g_{s_1}(z) - f_{s_1}(z)$. Now a key part in the argument is the fact that the denominator $g_{s_1}(z) - f_{s_1}(z)$ has the same number of zeroes in $\mathcal{R}e\, z \geq 0$ as $g_{s_1}^+(z)$. The proof of this statement is deferred to the Appendix in Proposition 6.6.1. Thus we have $\deg P_{s_1}(z) \geq m_+$. Together with the upper bound on the degree of $P_{s_1}(z)$, this implies $\deg P_{s_1}(z) = m_+$; moreover, this determines $P_{s_1}(z)$ up to a constant factor (the constant being relative to z !)

$$P_{s_1}(z) = C(s_1) \prod_{i:\mathcal{R}e\, v_i(s_1) \geq 0} (z - v_i(s_1)),$$

where $v_i(s_1)$ are zeroes of $g_{s_1}(z) - f_{s_1}(z)$. Upon replacing the above in (6.13), we have found the one-parameter positive Wiener-Hopf factor

$$\tilde{K}_{s_1}^+(z) = \frac{\prod_{i:\mathcal{R}e\, v_i(s_1)<0} (z - v_i(s_1))}{g_{s_1}^-(z)}. \tag{6.14}$$

And in particular the above also determines $\tilde{K}_{s_1}^-(z)$:

$$\tilde{K}_{s_1}^-(z) = \frac{\tilde{K}(s_1, z)}{\tilde{K}_{s_1}^+(z)},$$

and the proof is complete.

Running example: For fixed s_1, we can carry out the factorization for the kernel (6.10) in the running example. Below we give the zeroes of the numerator

$$v_1(s_1) = \frac{-s_1-3-\sqrt{3}\sqrt{(s_1+3)(1+3s_1)}}{2(3+s_1)},\ v_2(s_1) = \frac{-s_1-3+\sqrt{3}\sqrt{(s_1+3)(3s_1+1)}}{2(3+s_1)},$$

$$v_3(s_1) = \frac{-s_1-3-\sqrt{(s_1+3)(3s_1+13)}}{2(3+s_1)},\ v_4(s_1) = \frac{-s_1-3+\sqrt{(s_1+3)(3s_1+13)}}{2(3+s_1)}.$$

The radicals above are defined when the cut in the complex plane is taken along the negative half of the real axis and the complex arguments are measured from $-\pi$ to π. This convention determines the so-called principal value of the square root function. The negative real half-axis will be a discontinuity line for the square root function $\sqrt{z}$ as a function of a complex variable because as the argument z approaches the negative real half-axis,

$$(z^- - z_0)^{1/2} = e^{i\pi}(z^+ - z_0)^{1/2},$$

where z_0 lies on the negative real half-axis, that is, $\mathcal{R}e\, z_0 < 0$ and $\mathcal{I}m\, z_0 = 0$; $z^\pm$ denotes the limit of z towards z_0 from respectively above and below the real axis. We will call such lines of discontinuity branch cuts. The branch points of v_1 (and v_2) are -3 and $-1/3$. The branch cuts are then the curves generated by the equations

$$\arg(s_1+3) + \arg(1+3s_1) = \pm\pi.$$

It is a problem of plane geometry to see that these branch cuts are constituted by the line segment joining the two branch points, together with the perpendicular line on this segment that passes through its mid-point. The situation is similar for $v_3(s_1)$ and $v_4(s_1)$.

By inspecting the zeroes of the numerator for $\mathcal{R}e\, s_1 > 0$, exactly $v_1(s_1)$ and $v_3(s_1)$ are negative, where by positive/negative values of complex numbers we will always refer to their real parts. There are as many negative zeroes in the denominator, which is already confirmed by Proposition 6.6.1. Moreover, the branch cuts of neither v_1 nor v_3 are located in the right half-plane, which means these zeroes are regular functions for positive values of s_1.

Having isolated the negative zeroes, the one-parameter positive Wiener-Hopf factor from (6.14) is

$$\tilde{K}^+_{s_1}(z) = \frac{\left(z - \frac{-s_1-3-\sqrt{3}\sqrt{(s_1+3)(1+3s_1)}}{2(s_1+3)}\right)\left(z - \frac{-s_1-3-\sqrt{(s_1+3)(3s_1+13)}}{2(s_1+3)}\right)}{(z+2)^2}. \tag{6.15}$$

Remark 6.3.2. *Interestingly, $\tilde{K}^+_{s_1}$ is not a rational function anymore in the argument s_1 (however, in this example it is meromorphic in the argument s_1, for $\mathcal{R}e\, s_1 > 0$). A queueing theoretic explanation of this remark can be found by comparing the double transform (6.16) obtained below with the decomposition results for a particular case of the present model in Chapter 5. For the process with Markov arrivals studied therein,* *the stationary waiting time stochastically decomposes into two components, one of which is related to the extra busy period length in the longest queue and it is well known that busy periods in general do not have rational transforms (already in the case of an M/M/1 system, the busy period has a non-rational transform).*

Moreover, it is easy to see that the marginal factor $\tilde{K}^+_{s_1}(s_2)|_{s_1=0}$ is a rational function. From this and (6.16) below follows that the marginal transform $\psi(0, s_2)$ is a rational function. Also, $\psi(s_1, 0)$ is a rational function because for $s_2 = 0$ the factor $\tilde{K}^+_{s_1}$ cancels against itself in (6.16). The rationality of the marginal transforms is clear from their queueing interpretation because these are the transforms of the univariate survival functions/waiting times for the two insurance lines/queueing systems in isolation (see the discussion in Cohen [40] p. 325 and the references therein).

6.3.3 The main result

We are now ready to formulate and prove the main result.

Theorem 6.3.1. *Under the safety loading condition for the riskier line 1, $\rho_1 < 1$, the infinite horizon survival function $F^s(u^{(1)}, u^{(2)})$ is a (proper) probability distribution function with support the non-negative quadrant in $\mathbb{R}^2$, and its LST is given by*

$$\psi(s_1, s_2) = \int e^{-s_1 u^{(1)} - s_2 u^{(2)}} \mathrm{d}F^s(u^{(1)}, u^{(2)}) = \frac{K^+_{pr}(0)}{K^+_{pr}(s_1)} \frac{\tilde{K}^+_{s_1}(s_1)}{\tilde{K}^+_{s_1}(s_1 + s_2)}, \tag{6.16}$$

for $\mathcal{R}e\, s_i \geq 0$, $i = 1, 2$. $\tilde{K}^+_{s_1}(z)$ is given in (6.14) and here is evaluated at $z = s_1$ and at $z = s_1 + s_2$. $K^+_{pr}(s_1)$ is the positive Wiener-Hopf factor of the projected one-dimensional kernel $\tilde{K}(s_1, s_1) = K(s_1, 0)$, i.e., the unique function analytic in $\mathcal{R}e\, s_1 > 0$, continuous in $\mathcal{R}e\, s_1 \geq 0$ that factorizes $K(s_1, 0)$ into

$$K(s_1, 0) = K^+_{pr}(s_1)\, K^-_{pr}(s_1),$$

with $K^-_{pr}(s_1)$ analytic in $\mathcal{R}e\, s_1 < 0$ and continuous in $\mathcal{R}e\, s_1 \leq 0$ (see for instance Prabhu [94], Thm. 7 p. 55). Under Assumption 6.1.1 it is of the form (see also Theorem 3.1.1)

$$K^+_{pr}(s_1) = \frac{\prod_j (s_1 - \tilde{v}^-_j)}{\prod_j (s_1 - v^-_j)}, \tag{6.17}$$

with $\tilde{v}^-_j$ the negative zeroes of $K(s_1, 0)$ and v^-_j its negative poles.

Proof. Our starting-point is (6.11), and our goal is to determine the one yet unknown function $C(s_1)$ in that formula. The idea is that, since $C(s_1)$ stays the same irrespective of the value of z, we are free to choose any z. Since by definition, $\tilde{K}^+_{s_1}(z)$ is analytic for all $\mathcal{R}e\, z > 0$ and continuous in $\mathcal{R}e\, z \geq 0$, take $z = s_1$:

$$\tilde{\psi}(s_1, s_1) = \mathbb{E}e^{-s_1 W^{(1)}} = C(s_1)[\tilde{K}^+_{s_1}(s_1)]^{-1}. \tag{6.18}$$

We can determine $C(s_1)$ from (6.18), because $\tilde{\psi}(s_1, s_1) = \psi(s_1, 0) = \mathbb{E}e^{-s_1 W^{(1)}}$ is the steady-state waiting time transform in the marginal $G/G/1$ queue with dependence between inter-arrival times and service requirements, which has been determined in Chapter 3. This chapter was devoted to an analysis of a one-dimensional risk/queueing model that amounts to the present model with $B^{(2)} \equiv 0$. The kernel of the functional identity for this marginal queue with generic service requirement $B^{(1)}$ and correlated inter-arrival time A is $\tilde{K}(s_1, s_1)$; the corresponding Rouché problem is to prove that $g(s_1, s_1)$ and $g(s_1, s_1) - f(s_1, s_1)$ have the same number of non-negative zeroes. This has been carried out in Lemma 3.5.1. Formula (3.3) reads in the current notation

$$\tilde{\psi}(s_1, s_1) = \mathbb{E}e^{-s_1 W^{(1)}} = K_{pr}^{+}(0)[K_{pr}^{+}(s_1)]^{-1}, \tag{6.19}$$

with $K_{pr}^{+}(s_1)$ the positive Wiener-Hopf factor of the projected kernel $K(s_1, 0)$:

$$K_{pr}^{+}(s_1) = \frac{\prod_j (s_1 - \tilde{v}_j^-)}{\prod_j (s_1 - v_j^-)},$$

such that $\tilde{v}_j^-$ are the negative zeroes of $K(s_1, 0)$ and v_j^- its negative poles, and the normalizing constant $K_{pr}^{+}(0)$ is equal to the atom at 0 of $W^{(1)}$:

$$\mathbb{P}(W^{(1)} = 0) = \prod_k (-v_k^-) / \prod_j (-\tilde{v}_j^-).$$

Formula (6.19) together with (6.18) now determine $C(s_1)$:

$$C(s_1) = K_{pr}^{+}(0)\, [K_{pr}^{+}(s_1)]^{-1}\, \tilde{K}_{s_1}^{+}(s_1), \tag{6.20}$$

and with this we obtain the transform of the joint waiting time distribution, or equivalently of the survival function $F^s(u^{(1)}, u^{(2)})$, from (6.11), upon switching back to the original coordinates. The proof is complete.

Remark 6.3.3. *An important remark is that the factors in (6.16), $\tilde{K}_{s_1}^{+}(s_1)$ and $K_{pr}^{+}(s_1)$, as defined in Proposition 6.3.1 and in Theorem 6.3.1, are not the same. One can already compare (6.15) with (6.27) for Example 1 in the following section.*
More precisely, the operations of taking the projection and carrying out the factorization do not commute with each other, in contrast with the one-dimensional Fluctuation Theory of random walks. See also Section 13 in Kingman [75].
$\tilde{K}_{s_1}^{+}(s_1)$ is defined by first carrying out the one-parameter Wiener-Hopf factorization for $\tilde{K}(s_1, z)$ and then projecting the positive factor onto the main diagonal of the 2-dimensional complex space: $z = s_1$.
On the other hand, $K_{pr}^{+}(s_1)$ is obtained by first projecting $\tilde{K}(s_1, z)$ onto the main diagonal $z = s_1$ and then carrying out the Wiener-Hopf factorization for the projected kernel $\tilde{K}(s_1, s_1) = K(s_1, 0)$.

Remark 6.3.4. *The LST in (6.16) has a product form. It can be shown that the bivariate LST of the reflected random walk decomposes in a similar way as in Theorem 5.2.1, where one of the factors is related to a modified workload process.*

6.4 A probabilistic decomposition

By the duality arguments of Section 2.2, Formula (6.16) also provides an expression for the LST of the two-dimensional waiting time distribution in the dual queueing model with simultaneous arrivals and with generic input vector $(A, B^{(1)}, B^{(2)})$. In the queueing setting, Formula (6.16) allows for an interesting decomposition/interpretation. In particular, the constant $C(s_1)$ has a probabilistic interpretation that makes it the key to finding the other two unknown functionals $\tilde{\psi}_1(s_1, z)$ and $\psi_2(s_1)$. If we let $z \to \infty$ through the positive half-plane in (6.11) we see that, by dominated convergence,

$$C(s_1) = \mathbb{E}[e^{-s_1 W^{(1)}} 1_{\{W^{(2)}=0\}}] = \int_{0-}^{\infty} e^{-s_1 u^{(1)}} \mathrm{d}F^s(u^{(1)}, 0).$$

This holds because by Proposition 6.6.1, the limit of (6.14) as $|z| \to \infty$ is equal to 1. On the other hand, from the definition of $\psi_2(s_1)$, we see that the last two terms in the RHS of (6.12) add up to

$$\psi_2(s_1) + \mathbb{P}(W^{(1)} + X^{(1)} \leq 0) = \mathbb{E}[e^{-s_1 (W^{(1)}+X^{(1)})^+} 1_{\{W^{(2)}=0\}}].$$

Taking into account that $W^{(1)} \stackrel{d}{=} (W^{(1)} + X^{(1)})^+$, we have

$$C(s_1) = \psi_2(s_1) + \mathbb{P}(W^{(1)} + X^{(1)} \leq 0). \tag{6.21}$$

Motivated by the decomposition result for the case with exponential inter-arrivals in Chapter 5, let us consider the following modified waiting time process in equilibrium:

$$(W_*^{(1)}, W^{(2)}) \stackrel{d}{=} \begin{cases} (W_*^{(1)} + X^{(1)}, W^{(2)} + X^{(2)}), & \text{if } W^{(2)} + X^{(2)} > 0 \\ (0, 0), & \text{if } W^{(2)} + X^{(2)} \leq 0. \end{cases}$$

Marginally $W^{(2)}$ evolves unchanged. Denote by $\eta(s_1, s_2) = \mathbb{E}e^{-s_1 W_*^{(1)} - s_2 W^{(2)}}$, the joint transform of the modified waiting time. Roughly speaking, as soon as the second component of the reflected version of the random walk attempts to exit the positive half-line (and is thus set to 0), the other component is also set to 0.

To the above distributional identity corresponds the functional equation in terms of LST's:

$$\tilde{K}(s_1, s_1 + s_2)\, \eta(s_1, s_2) = -\eta_1(s_1, s_2) + \mathbb{P}(W^{(2)} + X^{(2)} \leq 0), \tag{6.22}$$

with $\eta_1(s_1, s_2) := \mathbb{E}[e^{-s_1 (W_*^{(1)}+X^{(1)}) - s_2 (W^{(2)}+X^{(2)})} 1_{\{W^{(2)}+X^{(2)} \leq 0\}}]$.

Since the kernel is the same as in (6.8), we can use a similar approach as in Proposition 6.3.1 in order to analyze (6.22). We already have the kernel factorization in the z argument, for fixed s_1. Therefore upon passing to the usual coordinates (s_1, z), and applying Liouville's theorem, we get for $\tilde{\eta}(s_1, z) := \eta(s_1, z - s_1)$:

$$\tilde{\eta}(s_1, z) = \frac{g_{s_1}^-(z)}{g_{s_1}(z) - f_{s_1}(z)} P_{s_1}^*(z), \tag{6.23}$$

using the same notation as in Proposition 6.3.1 and with $P^*_{s_1}(z)$ another polynomial to be determined. Again, Proposition 6.6.1 determines $P^*_{s_1}(z)$ up to a constant $C^*(s_1)$:

$$P^*_{s_1}(z) = C^*(s_1) \prod_{\mathcal{R}e\, v_i(s_1)\geq 0} (z - v_i(s_1)),$$

where $v_i(s_1)$ are the same zeroes of $g_{s_1}(z) - f_{s_1}(z)$. In relation to $P_{s_1}(z)$, we can thus write

$$P^*_{s_1}(z) = \frac{C^*(s_1)}{C(s_1)} P_{s_1}(z).$$

$C^*(s_1)$ is a priori a function of s_1. However, it turns out that it is constant: $C^*(s_1) = \mathbb{P}(W^{(2)} + X^{(2)} \leq 0)$. To see this, replace $P^*_{s_1}(z)$ in (6.23) to obtain the analogue of (6.11):

$$\tilde{\eta}(s_1, z) = C^*(s_1) \frac{g^-_{s_1}(z)}{\prod\limits_{\mathcal{R}e\, v_i(s_1)<0} (z - v_i(s_1))}. \tag{6.24}$$

Taking $z \to \infty$, $\mathcal{R}e\, z > 0$ in (6.24),

$$C^*(s_1) = \mathbb{E}\left[e^{-s_1 W^{(1)}_*} 1_{\{W^{(2)}=0\}}\right] = \mathbb{E}1_{\{W^{(1)}_*=0, W^{(2)}=0\}} = \mathbb{P}(W^{(2)} + X^{(2)} \leq 0).$$

The last two identities follow because in this modified process $W^{(1)}_*$ is 0 as soon as $W^{(2)}$ becomes 0. To conclude so far, we obtain the relation between $\tilde{\psi}(s_1, z)$ and $\tilde{\eta}(s_1, z)$, by comparing (6.24) to (6.16):

$$\tilde{\psi}(s_1, z) = \frac{C(s_1)}{\mathbb{P}(W^{(2)} = 0)} \tilde{\eta}(s_1, z),$$

which holds because $W^{(2)}$ is the stationary version of the waiting time, thus we have the stationary Lindley identity $[W^{(2)} + X^{(2)}]^+ \stackrel{d}{=} W^{(2)}$. It follows from (6.21) that

$$C(s_1)/\mathbb{P}(W^{(2)} = 0) = \mathbb{E}[e^{-s_1 W^{(1)}} \,|\, W^{(2)} = 0] = \frac{1}{F^s(\infty, 0)} \int_{0-}^{\infty} e^{-s_1 u^{(1)}} \mathrm{d}F^s(u^{(1)}, 0),$$

and we find the following LST decomposition:

$$\tilde{\psi}(s_1, z) = \mathbb{E}[e^{-s_1 W^{(1)}} \,|\, W^{(2)} = 0]\; \tilde{\eta}(s_1, z). \tag{6.25}$$

Compare this to the decomposition (5.10) obtained for the two-dimensional compound Poisson input process in Chapter 5.

6.5 Examples and numerical inversion

In the previous sections we dealt with some theoretical aspects related to obtaining the transform of the survival function. It turns out that additional insight can be obtained by applying the previous results to some specific examples. Our aims in this section are:

(i) to provide examples for which the Laplace-Stieltjes transforms of the survival measures can be calculated, based on the general results obtained in the previous section.

(ii) to explain the various analytic challenges that appear when one tries to determine the Laplace-Stieltjes transform of the survival measure for some specific classes of distributions for the input $(A, \delta, B^{(2)})$.

(iii) numerical inversion of the bivariate Laplace-Stieltjes transform in (6.16) and the comparison between the risks for various possible correlations between the claim sizes $(B_n^{(1)}, B_n^{(2)})$ and inter-arrival times A_n.

We begin by explaining the inversion algorithm and how we applied it. However, in order to obtain the input for the algorithm, we need to follow the steps in Section 6.3 and construct the Wiener-Hopf factors. It turns out that this presents a challenge because of the branch cuts (discontinuities) that the zeroes of the kernel might have in the right half-plane of the complex s_1-plane. The running example from Section 6.3.1 can thus be considered a simple instance of the inversion algorithm.

Numerical inversion For the purpose of inverting (6.16), consider the Laplace transform of the bivariate tail probability of the waiting time. By a straight-forward integration by parts, this can be related to the Laplace-Stieltjes transform of the waiting time/survival function:

$$\iint e^{-s_1u_1-s_2u_2}\,\mathbb{P}(W^{(1)}>u_1, W^{(2)}>u_2)\,\mathrm{d}u_1\mathrm{d}u_2 = \frac{1-\psi(s_1,0)-\psi(0,s_2)+\psi(s_1,s_2)}{s_1s_2}. \tag{6.26}$$

The key remark is that under mild conditions, this transform is continuous up to the boundary of the non-negative quadrant, as opposed to the Laplace transform of the survival function:

$$\iint e^{-s_1u_1-s_2u_2}\,\mathbb{P}(W^{(1)}\le u_1, W^{(2)}\le u_2)\,\mathrm{d}u_1\mathrm{d}u_2$$

which has by definition a singularity at $(s_1,s_2)=(0,0)$. It is easy to see that for example

$$\lim_{\substack{s_1\to 0\\ \mathcal{R}e\, s_1>0}} \frac{1}{s_1s_2}[1-\psi(s_1,0)-\psi(0,s_2)+\psi(s_1,s_2)] = \frac{1}{s_2}\left[\frac{\partial\psi}{\partial s_1}(0,s_2)-\frac{\partial\psi}{\partial s_1}(0,0)\right],$$

and even further

$$\lim_{\substack{s_1,s_2\to 0\\ \mathcal{R}e\, s_1,s_2>0}} \frac{1}{s_1 s_2}[1-\psi(s_1,0)-\psi(0,s_2)+\psi(s_1,s_2)] = \frac{\partial^2\psi}{\partial s_1\partial s_2}(0,0),$$

and it is clear that this mixed derivative is equal to $\mathbb{E}[W^{(1)}W^{(2)}]$, which is the same as the left-hand side of (6.26) evaluated at $s_1=s_2=0$. The partial derivatives on the right-hand side above must be considered as limits from the interior of the positive quadrant. The Laplace transform of the ruin function is continuous up to the boundary given that the above partial derivatives exist.

The main point of the above discussion is that we may now use the standard form of the multidimensional inversion algorithm developed in [70], for which it is essential that the Laplace transform is regular and continuous up to the boundary of the positive quadrant. The above trick of passing to tail probabilities thus frees one from considering modifications of the inversion algorithm for non-smooth functions (see [70]). Once the tail probability/ruin function has been obtained, the survival function follows from an identity similar to (6.26):

$$\mathbb{P}(W^{(1)}>x_1, W^{(2)}>x_2) = 1-\mathbb{P}(W^{(1)}\le x_1)-\mathbb{P}(W^{(2)}\le x_2)+\mathbb{P}(W^{(1)}\le x_1,\, W^{(2)}\le x_2),$$

for any $x_1,\, x_2 \ge 0$.

There are no regularity problems with the transforms we will be working with throughout this section because they are all meromorphic in both arguments for positive real values (in some cases they are constructed from branches of various locally meromorphic functions via analytic continuation – see Example 2 below).

Above we discussed how to consider the input for the inversion algorithm; some remarks are also needed about the output. This is an $m_1 \times m_2$ matrix, that represents the values of the ruin function $\mathbb{P}(\tau_\vee(\cdot,\cdot)<\infty)$ on a grid: the entry (k,l), $k\le m_1-1$, $l \le m_2-1$ stands for

$$\mathbb{P}(\tau_\vee((k-1)\Delta_1,(l-1)\Delta_2)<\infty) = \mathbb{P}(W^{(1)}>(k-1)\Delta_1, W^{(2)}>(l-1)\Delta_2),$$

where Δ_i are division sizes. The values of the inverted transform are plotted in the figures below for various examples.

Example 1. This is the running example that started at (6.10). We have calculated the one-parameter Wiener-Hopf factor for this example in (6.15). For the LST of the survival function we need also the positive Wiener-Hopf factor of the projected kernel:

$$\tilde{K}(s_1,s_1) = K(s_1,0) = \frac{-9s_1-35s_1^2-s_1^3+18s_1^4+8s_1^5+s_1^6}{(-1+s_1)^2(2+s_1)^2(3+s_1)^2}.$$

The numerator can be factorized as $s_1(s_1^2+4s_1+1)(s_1^3+4s_1^2+s_1-9)$ where the order 2 polynomial further factorizes as

$$s_1^2+4s_1+1 = (s_1+2-\sqrt{3})(s_1+2+\sqrt{3})$$

and the zeroes of the factor $s_1^3+4s_1^2+s_1-9$ are

$$v_0 := -\frac{4}{3} + \frac{1}{3\sqrt[3]{2}}\left(151 - 9\sqrt{173}\right)^{1/3} + \frac{1}{3\sqrt[3]{2}}\left(151 + 9\sqrt{173}\right)^{1/3},$$

$$v_1 := -\frac{4}{3} - \frac{1}{6\sqrt[3]{2}}(1+i\sqrt{3})\left(151 - 9\sqrt{173}\right)^{1/3} - \frac{1}{6\sqrt[3]{2}}(1-i\sqrt{3})\left(151 + 9\sqrt{173}\right)^{1/3},$$

$$v_2 = -\frac{4}{3} - \frac{1}{6\sqrt[3]{2}}(1-i\sqrt{3})\left(151 - 9\sqrt{173}\right)^{1/3} - \frac{1}{6\sqrt[3]{2}}(1+i\sqrt{3})\left(151 + 9\sqrt{173}\right)^{1/3}, \text{ with } v_2 = \bar{v}_1.$$

The positive Wiener-Hopf factor of the projected kernel becomes, cf. (6.17),

$$K^+_{pr}(s_1) = \frac{(s_1^2 + 4s_1 + 1)(s_1 - v_1)(s_1 - \bar{v}_1)}{(s_1+2)^2(s_1+3)^2}. \tag{6.27}$$

(6.15) and (6.27) are the necessary components for constructing the LST of the survival function in this example, as given by (6.16). The atom at $(0,0)$ is

$$\mathbb{P}(W^{(1)} = 0, W^{(2)} = 0) = \mathbb{P}(W^{(1)} = 0) = K^+_{pr}(0) = \frac{|v_1|^2}{36} \approx 0.204.$$

Here we used the ordering between $W^{(1)}$ and $W^{(2)}$. Below is given the final formula for the transform of the survival function in the original coordinates:

$$\psi(s_1, s_2) = \frac{[(s_1+3)(2s_1+1) + \sqrt{(3s_1+9)(1+3s_1)}]}{[(s_1+3)(2s_1+2s_2+1) + \sqrt{(3s_1+9)(1+3s_1)}]}.$$

$$\frac{[(s_1+3)(2s_1+1) + \sqrt{(s_1+3)(3s_1+13)}]}{[(s_1+3)(2s_1+2s_2+1) + \sqrt{(s_1+3)(3s_1+13)}]}\,\frac{|v_1|^2(s_1+s_2+2)^2(s_1+3)^2}{36(s_1^2+4s_1+1)(s_1-v_1)(s_1-\bar{v}_1)}. \tag{6.28}$$

Remark 6.5.1. $\sqrt{s_1+3}$ *cannot be simplified above. When choosing a branch for the square root as a function of a complex variable, one has in general for* $a \neq b$*:*

$$\sqrt{(z-a)(z-b)} \neq \sqrt{z-a}\sqrt{z-b}.$$

The marginal transforms are

$$\psi(s_1, 0) = \frac{|v_1|^2}{36}\,\frac{(2+s_1)^2(3+s_1)^2}{(s_1^2+4s_1+1)(s_1-v_1)(s_1-\bar{v}_1)},$$

$$\psi(0, s_2) = \frac{(3+\sqrt{39})(s_2+2)^2}{4(6s_2+3+\sqrt{39})(s_2+1)}.$$

The atom at zero of $W^{(2)}$ is approximately 0.575. We have carried out the numerical inversion for the transform in Example 1: the division size is chosen $\Delta_1 = \Delta_2 = .1$ and the grid size is $m_1 = m_2 = 2^6$. An important performance measure is the 5% quantile curve of the tail probability – the two-dimensional version of the 5% quantile. This is the (not necessarily continuous in general) curve that contains all (u_1, u_2), such that $\mathbb{P}(W^{(1)} > u_1, W^{(2)} > u_2) \geq .05$ and $\mathbb{P}(W^{(1)} > u_1+, W^{(2)} > u_2+) \leq .05$. To put it simply, the ruin function $\mathbb{P}(\tau_\vee(\cdot,\cdot) < \infty)$ (see Formula (2.11) and Remark 5.2.1) is less than 5% whenever it is evaluated at a point which lies outside the region bounded by this curve in the non-negative quadrant. This, together with several other quantile curves, is displayed in Figure 6.1 below.

Finally, as a verification, we estimated the ruin function using simulation. Upon choosing suitable bin sizes that account for the atom at $(0,0)$ of $(W^{(1)}, W^{(2)})$, the uniform distance between the output of the inversion algorithm and the simulated ruin function is of the order of 10^{-3}. If we denote with $R(x_1,x_2) = \mathbb{P}(\tau_\vee(x_1,x_2) < \infty)$, the joint ruin function, the simulation comparison is in Table 6.1.

(x_1,x_2)	$(0,0)$	(2,0)	(2,2)	$(4,0)$	$(4,2)$	$(4,4)$	$(6,0)$	$(6,2)$	$(6,4)$	(6,6)
$\hat{R}(x_1,x_2)$	.423	.297	.060	.180	.050	.008	.107	.034	.007	.001
$R(x_1,x_2)$	.424	.301	.060	.184	.050	.008	.110	.035	.007	.001

TABLE 6.1: Comparison between the simulated ruin function ($\hat{R}$) and the inverted function (R) for various values of the initial capital (x_1,x_2).

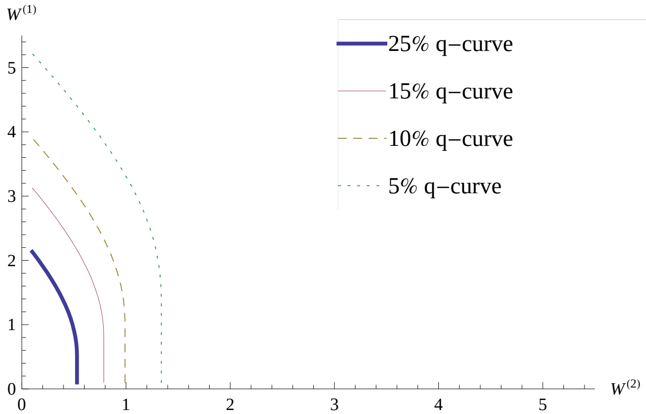


FIGURE 6.1: 25%, 15%, 10%, and respectively 5%-quantile curves for the ruin function in Example 1. The abscissa corresponds to the values at risk in the second insurance line/the marginal tail of $W^{(2)}$.

Remark 6.5.2. *The quantile curve plots in Figure 6.1 contain lines which below the main diagonal are straight. This is a consequence of the ordering assumption on the claims (and implicitly on the waiting times) since we can write for $x_1 \le x_2$*

$$\mathbb{P}(W^{(1)} > x_1, W^{(2)} > x_2) = \mathbb{P}(W^{(2)} > x_2),$$

because $W^{(2)} > x_2$ implies $W^{(1)} > x_1$ for all $x_1 \le x_2$.

Example 2. The parameters are the same as in Example 1, except now the order is $n = 3$. The kernel is

$$\tilde{K}(s_1,z)=1-\frac{72}{(3+s_1)^3(1-z)^3(2+z)^3}-\frac{12}{(3+s_1)^2(1-z)^2(2+z)^2}$$
$$-\frac{2}{(3+s_1)(1-z)(2+z)}.$$

Below we list the zeroes of the numerator. The radicals are again defined when the cut is taken along the negative half of the real axis.

$$v_1(s_1)=\frac{-(s_1+3)^2-\sqrt{(s_1+3)^4-4(s_1+3)^2(-24-14s_1-2s_1^2+2\sqrt{2}\sqrt{-(3+s_1)^2})}}{2(s_1+3)^2},$$

$$v_2(s_1)=\frac{-(s_1+3)^2+\sqrt{(s_1+3)^4-4(s_1+3)^2(-24-14s_1-2s_1^2+2\sqrt{2}\sqrt{-(3+s_1)^2})}}{2(s_1+3)^2},$$

$$v_3(s_1)=\frac{-(s_1+3)^2-\sqrt{(s_1+3)^4-4(s_1+3)^2(-24-14s_1-2s_1^2-2\sqrt{2}\sqrt{-(3+s_1)^2})}}{2(s_1+3)^2},$$

$$v_4(s_1)=\frac{-(s_1+3)^2+\sqrt{(s_1+3)^4-4(s_1+3)^2(-24-14s_1-2s_1^2-2\sqrt{2}\sqrt{-(3+s_1)^2})}}{2(s_1+3)^2},$$

$$v_5(s_1)=\frac{-(s_1+3)-\sqrt{3}\sqrt{(s_1+3)(1+3s_1)}}{2(s_1+3)},\ \ v_6(s_1)=\frac{-(s_1+3)+\sqrt{3}\sqrt{(s_1+3)(1+3s_1)}}{2(s_1+3)}.$$

The above formulae cannot be simplified (Remark 6.5.1). In addition, $\sqrt{-(3+s_1)^2}$ is discontinuous (its discontinuity line is $\mathcal{I}m\, s_1=0$) and it contributes towards the discontinuities of the zeroes $v_i(s_1)$, $i=\overline{1,4}$.

It is not clear a priori which one of the four zeroes to choose when constructing the one-parameter factor $\tilde{K}_{s_1}(z)$ from (6.14) because, in contrast to Example 1, the branch cuts of $v_1(s_1)$ up to $v_4(s_1)$ cross inside the right half-plane, and the zeroes $v_i(s_1)$ jump from positive to negative real values when the argument passes between the regions bounded by the cuts in $\mathcal{R}e\, s_1>0$.

The key observation is that $v_1(s_1)$ is an analytic continuation of $v_2(s_1)$, and $v_3(s_1)$ is an analytic continuation for $v_4(s_1)$. In order to obtain $\tilde{K}_{s_1}(z)$, one has to glue together (using analytic continuation) the negative branches of $v_1(s_1)$ and $v_2(s_1)$ on the one hand, and of $v_3(s_1)$ and $v_4(s_1)$ on the other, for $\mathcal{R}e\, s_1>0$. $v_5(s_1)$ is negative for any $\mathcal{R}e\, s_1>0$ so it will always enter the formula for $K^+_{s_1}(\cdot)$ as opposed to $v_6(s_1)$ which is positive and doesn't play any role. Moreover the branch cuts of $v_1(s_1)$ up to $v_4(s_1)$ partition the complex half-plane in 4 regions symmetric around the real axis and the cuts are pairwise parallel lines at angles $\pm\pi/4$. To be more precise, we have to use the 3 different branches of $\tilde{K}^+_{s_1}(z)$:

$$\tilde{K}^{1,3}_{s_1}(z)=\frac{(z-v_1(s_1))(z-v_3(s_1))(z-v_5(s_1))}{(z+1)^3},$$

$$\tilde{K}^{2,3}_{s_1}(z)=\frac{(z-v_2(s_1))(z-v_3(s_1))(z-v_5(s_1))}{(z+1)^3},$$

$$\tilde{K}^{2,4}_{s_1}(v) = \frac{(z - v_2(s_1))(z - v_4(s_1))(z - v_5(s_1))}{(z+1)^3}. \tag{6.29}$$

The branches $C^{1,3}(s_1)$, $C^{2,3}(s_1)$ and $C^{2,4}(s_1)$ are obtained similarly because from (6.20), these are related to the corresponding branches of $\tilde{K}^{+}_{s_1}(z)$ by setting $z = s_1$; the branch cuts and the partition of the complex plane are therefore the same. Since both $C(\cdot)$ and $K^{+}_{s_1}(\cdot)$ enter Formula (6.16), the expression for the LST of the survival function/joint waiting time (see Theorem 2.2.1) is obtained by patching together (via analytic continuation) the positive branches (in the s_1-plane) of the generalized Wiener-Hopf factors from Proposition 6.3.1 and Theorem 6.3.1.

In Figure 6.2 below we plot a section in the three branches of the real part of the LST of the survival measure in Example 2. More precisely consider the section

$$\zeta(y) := \mathcal{R}e\, \psi(iy, 14 + iy) \tag{6.30}$$

that runs along the imaginary axis in the first argument s_1 (the argument that generates the discontinuities). From this figure it becomes apparent how the three different branches of $\zeta(y)$ are continuations of each other: the central branch belongs to $\mathcal{R}e\, \psi^{1,3}(iy, 14+iy)$ (the blue curve). This is continued by the branch $\mathcal{R}e\, \psi^{2,3}(iy, 14+iy)$ (the dashed red curve) which in turn is continued by $\mathcal{R}e\, \psi^{2,4}(iy, 14 + iy)$ (the orange curve segment) towards the ends of the plot.

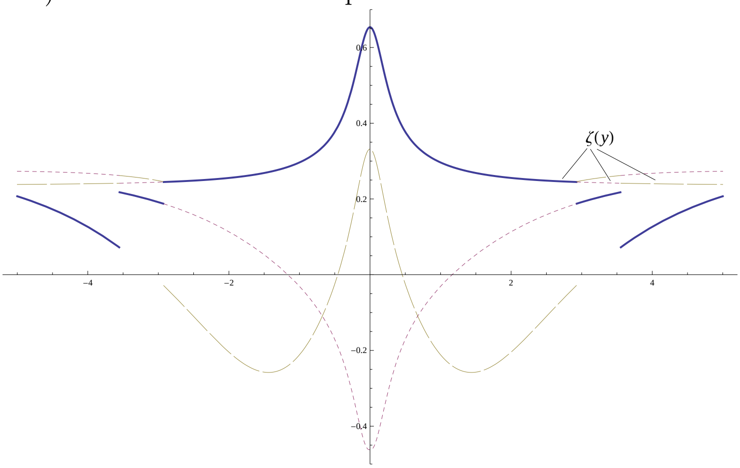


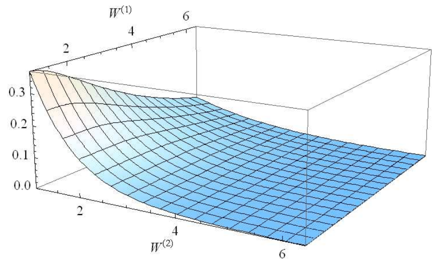
FIGURE 6.2: The plot of the three branches of the section $\zeta(y)$ from (6.30).

We numerically inverted the expression obtained for the transform $\psi(s_1, s_2)$ using den Iseger's algorithm [70]. Again, the division size is $\Delta = .1$ and the grid size is 2^6 in both directions. In Figure 6.3(a) the plot of the ruin function $\mathbb{P}(\tau_{\wedge}(\cdot,\cdot) < \infty)$ is presented (see Formula (2.11) and Remark 5.2.1) or equivalently the tail of the equilibrium distribution for the bivariate waiting time $(W^{(1)}, W^{(2)})$ (Theorem 2.2.1) in Example 2. We can write in general

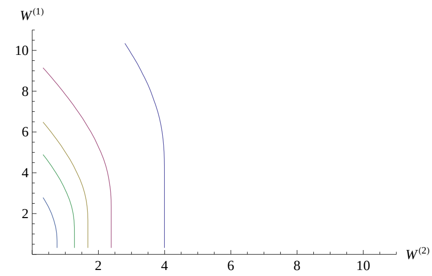
$$\mathbb{P}(W^{(1)} > 0, W^{(2)} > 0) = 1 - \mathbb{P}(W^{(1)} = 0) - \mathbb{P}(W^{(2)} = 0) + \mathbb{P}(W^{(1)} = 0, W^{(2)} = 0),$$

and because of the ordering, $\mathbb{P}(W^{(1)} = 0, W^{(2)} = 0) = \mathbb{P}(W^{(1)} = 0)$. And then the value at $(0, 0)$ of the joint ruin function is $\mathbb{P}(W^{(2)} > 0)$, which in this example approximately equates 0.37.

Finally, in Figure 6.3(b) we present various quantile curves for the ruin function/stationary tail of the waiting time.

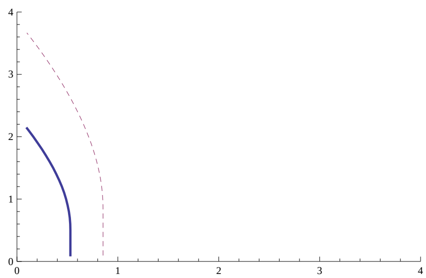


(a) The joint ruin function /the bivariate tail of the waiting time from Example 2.

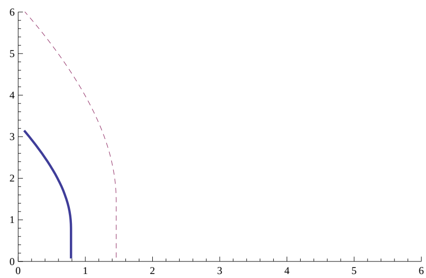


(b) 25%, 15%, 10%, 5%, and 1%-quantile curves for the ruin function in Example 2.

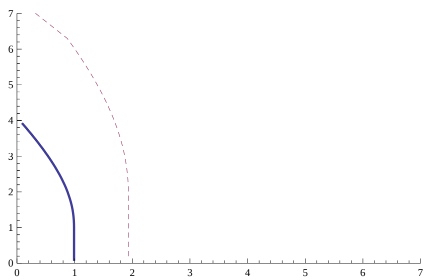
FIGURE 6.3: Numerical results for the risk reserve process in Example 2: (a) ruin function and (b) quantile curves.



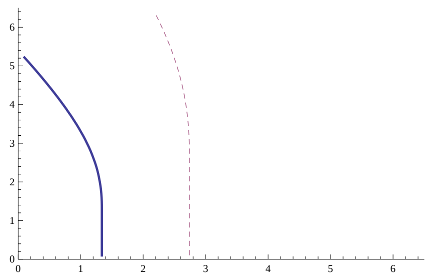
(a) 25% quantile curves



(b) 15% quantile curves



(c) 10% quantile curves



(d) 5% quantile curves

FIGURE 6.4: Comparison of risks. Dashed curves correspond to decoupled input.

Comparison of risks. In Figure 6.4 we compare the results for quantile curves of the ruin function of Example 1 with the quantile curves of the ruin function for the case where the input is decoupled. By this we mean we take three samples κ_1, κ_2 and κ_3 from a uniform distribution on $\{1,2\}$ and define the random variables $A_{dec} = \sum_{i=1}^{\kappa_1} A_i$, $\delta_{dec} = \sum_{i=1}^{\kappa_2} \delta_i$, $B_{dec}^{(2)} = \sum_{i=1}^{\kappa_3} B_i^{(2)}$, where $(A_i)_{i\le n}$, $(B_i^{(2)})_{i\le n}$, and $(\delta_i)_{i\le n}$ are mutually independent sequences of exponential random variables with rates λ, μ and μ_δ respectively. In this case the inter-arrival time becomes independent of the claim size vector, while marginally A_{dec}, $B_{dec}^{(2)}$ and $B_{dec}^{(1)}$ have the same distribution as in Example 1. The kernel for this instance is

$$\tilde{K}(s_1,z) = \left(\frac{9}{2(3+s_1)^2} + \frac{3}{2(3+s_1)}\right)\left(\frac{1}{2(1-z)^2} + \frac{1}{2(1-z)}\right)\left(\frac{2}{(2+z)^2} + \frac{1}{2+z}\right).$$

The zeroes of the numerator as a polynomial in z are already too complicated to present here. This instance is similar to Example 2 in terms of the analytic behaviour of these zeroes.

The main point is that, similarly to the ordering result obtained in Section 3.4, numerical data suggests that the ruin functions corresponding to positively correlated input on the one hand, and the ruin functions for decoupled input on the other are stochastically ordered (Figure 6.4).

Example 3 (proportional reinsurance) This is the case with proportional claims. There is a common arrival process such that the inter-arrival time A_n is correlated with the claim size B_n, and αB_n is deducted from the first insurance line and $(1-\alpha)B_n$ from the second.

We take $\kappa \sim \text{Uniform}\{1,2,3\}$, $\lambda = \mu = 1$, $\alpha = 3/4$ and unit income rates. For the purpose of comparing risks, we will consider three instances for the random vector (A,B):

positive correlation: $(A,B)_{pos} \sim (\text{ Erlang}(\kappa,\lambda),\ \text{Erlang}(\kappa,\mu))$,
independence: $(A,B)_0 \sim (\text{Erlang}(\kappa_1,\lambda),\ \text{Erlang}(\kappa_2,\mu))$, with κ_1, κ_2 two copies of κ.
negative correlation: $(A,B)_{neg} \sim\ (\text{Erlang}(\kappa,\lambda),\ \text{Erlang}(4-\kappa,\mu))$.

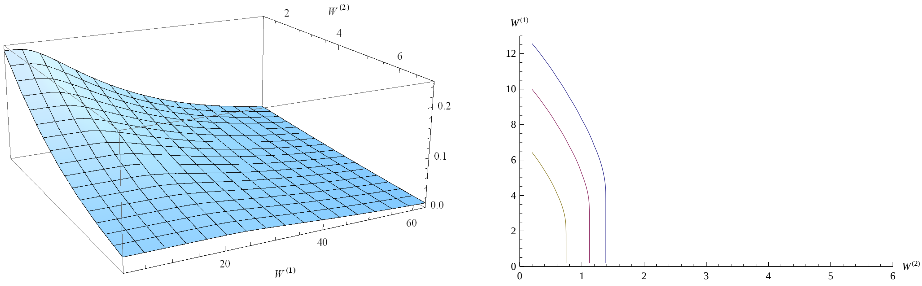
The kernels corresponding to these instances are

$$\begin{aligned}\tilde{K}_{pos}(s_1,z) = &\frac{1}{3(1-z)(1+s_1/2+z/4)} + \frac{1}{3(1-z)^2(1+s_1/2+z/4)^2}\\ &+ \frac{1}{3(1-z)^3(1+s_1/2+z/4)^3},\end{aligned}$$

$$\begin{aligned}\tilde{K}_0(s_1,z) = &\left(\frac{1}{3(1-z)} + \frac{1}{3(1-z)^2} + \frac{1}{3(1-z)^3}\right)\left(\frac{1}{3(1+s_1/2+z/4)} + \frac{1}{3(1+s_1/2+z/4)^2}\right.\\ &\left.+ \frac{1}{3(1+s_1/2+z/4)^3}\right),\end{aligned}$$

$$\tilde{K}_{neg}(s_1, z) = \frac{1}{3(1-z)(1+s_1/2+z/4)^3} + \frac{1}{3(1-z)^2(1+s_1/2+z/4)^2} + \frac{1}{3(1-z)^3(1+s_1/2+z/4)}.$$

These functions stand for $\mathbb{E}e^{-[s_1(2\alpha-1)+z(1-\alpha)]B+zA}$ under the three couplings. The correlations between the variables A and B can also be read directly from the shapes of these transforms. Numerical illustrations are in Figure 6.5.



(a) (b)

FIGURE 6.5: Bivariate tail (left) and respectively 10%, 5%, and 3% quantile curves (right) for proportional reinsurance with negative correlation.

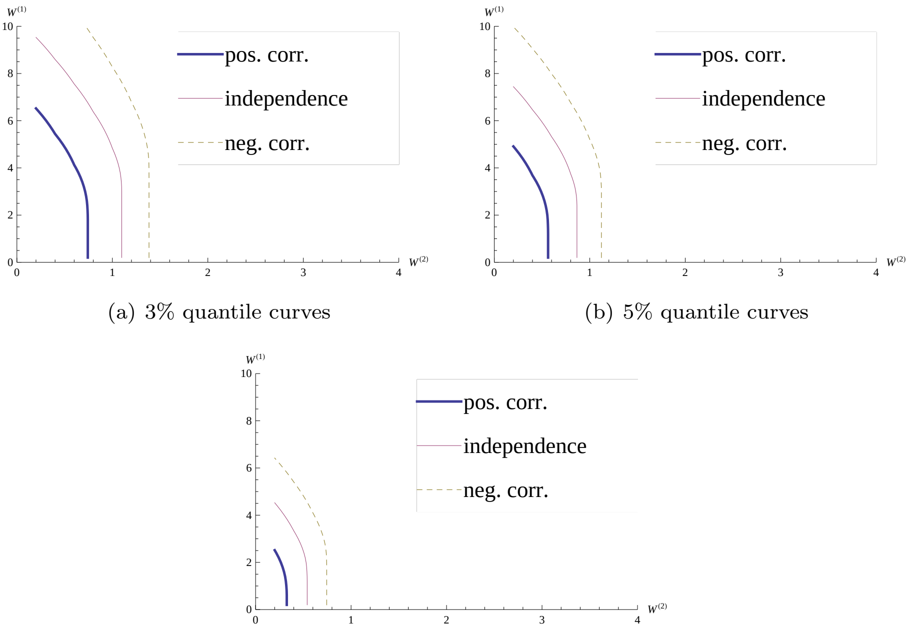
Comparison of risks In the table below we present various points at which a specific ruin probability is achieved. Given a fixed value for the ruin probability, any starting capital (x_1, x_2) lying on the respective quantile curve will achieve it. Interestingly, the risks are ordered between the various types of correlations (see also Figure 6.6). Let us denote by $R(x_1, x_2) = \mathbb{P}(\tau_\vee(x_1, x_2) < \infty)$, the probability that eventually, both lines are ruined. Then for the three types of correlation we give the values of the function R in Table 6.2. Positive correlation gives the lowest ruin probabilities for any starting capital considered.

(x_1, x_2)	$(0,0)$	(2.4,0)	(4.8,0)	(4.8, .4)	(6.4, .4)	(6.4, .8)	(9,.4)	(9, .8)
$R_{neg}(x_1, x_2)$	.2388	.1995	.1309	.0862	.0648	.0402	.0397	.0253
$R_0(x_1, x_2)$	.1922	.1516	.0896	.0536	.0375	.0214	.0203	.0120
$R_{pos}(x_1, x_2)$	.1381	.0979	.0486	.0237	.0148	.0070	.0065	.0033

TABLE 6.2: Comparison between the joint ruin functions R_{neg}, R_0 and R_{pos} respectively, for the various types of correlation.

Conclusions

The working assumptions of rationality and ordering for the trivariate input made up from the generic claim vector together with the preceding inter-arrival time, allow



(a) 3% quantile curves

(b) 5% quantile curves

(c) 10% quantile curves

FIGURE 6.6: Comparison of risks for proportional reinsurance.

one to obtain detailed numerical results for the joint ruin probability as a function of the initial risk reserves. Our numerical results suggest that when comparing ruin functions that correspond to various correlation structures between claim intervals and claim sizes, positive correlation among these cause lower values of the ruin probability compared to zero correlation, and even more so, when compared to negative correlation.

By using the relations described in Section 2.2, one can recover the other types of ruin or survival functions, where the numerical inversion of the marginal transforms, when needed, can be carried out using the one-dimensional inversion algorithm.

6.6 Appendix C

Proposition 6.6.1 (On the zeroes of the kernel $1-\tilde{K}(s_1,z)$)**.** *For each s_1, $\mathcal{R}e\, s_1 \geq 0$, $g(s_1,z) - f(s_1,z)$ and $g(s_1,z)$ have the same number of zeroes in $\mathcal{R}e\, z \geq 0$.*

Proof. We show that $g(s_1,z)$ dominates $f(s_1,z)$ on a suitably chosen contour in the complex z-plane. From this the claim in the proposition will follow via Rouché's theorem [102], p. 116. Consider the contour which is made up from the extended arc:

$$\mathcal{C}_\epsilon := \left\{re^{i\varphi};\ \varphi \in [-\pi/2 - \arccos\epsilon, \pi/2 + \arccos\epsilon]\right\},$$

together with the line segment

$$I := \left\{-\epsilon + i\omega;\ |\omega| \in \left[0, r\sqrt{1-\epsilon^2}\right]\right\}.$$

The rationality of the transform ensures that $\tilde{K}(s_1,z)$ can be analytically continued on a thin strip: $\mathcal{R}e\, z < 0$, $|\mathcal{R}e\, z| < \epsilon$. We first consider the contour $\mathcal{C}_\epsilon$. On the one hand we have the triangle inequality for f: $|f(s_1,z)| \leq \bar{f}(|s_1|,|z|)$, where $\bar{f}$ is a polynomial with the same degree as f. On the other hand, if we represent $g_{s_1}(z)$ for fixed s_1 as $g_{s_1}(z) = a_m(s_1)\prod_i (z - \xi_i(s_1))$, with $\xi_i(s_1)$ its zeroes and $a_m(s_1)$ the coefficient of z^m, $m = m(s_1) = \deg g_{s_1}(z)$, then the same triangle inequality gives a lower bound for $|g(s_1,z)|$:

$$|g(s_1,z)| = |a_m(s_1)|\prod_i |z - \xi_i(s_1)| \geq |a_m(s_1)|\prod_i (|z| - |\xi_i(s_1)|) =: \bar{g}(s_1,|z|),$$

with the remark that $\bar{g}_{s_1}(|z|)$ has the same degree as $g_{s_1}(z)$.

Now we can bound for r sufficiently large, such that the interior of $\mathcal{C}_\epsilon \cup I$ contains all the zeroes of $\bar{g}_{s_1}(|z|)$ and $z \in \mathcal{C}_\epsilon$:

$$\left|\frac{f(s_1,z)}{g(s_1,z)}\right| \leq \frac{\bar{f}(|s_1|,|z|)}{\bar{g}(s_1,|z|)} \to 0, \quad \text{as } r \to \infty.$$

Convergence holds because the degree of the numerator is strictly less than that of the denominator, by Assumption 6.1.1. This establishes the bound $|g(s_1,z)| > |f(s_1,z)|$ on $\mathcal{C}_\epsilon$, for r large enough.

For the segment I, we use the safety loading condition for the second(!) line: $c_2\mathbb{E}A - \mathbb{E}B^{(2)} > 0$. That is, we start with the fact $\frac{d}{dz}\frac{f(0,z)}{g(0,z)}|_{z=0} = c_2\mathbb{E}A - \mathbb{E}B^{(2)} > 0$. So for $\epsilon > 0$ sufficiently small, $\frac{f(0,-\epsilon)}{g(0,-\epsilon)} < \frac{f(0,0)}{g(0,0)} = 1$. Then we can write for $z \in I$:

$$\begin{aligned}\left|\frac{f(s_1,\epsilon+i\omega)}{g(s_1,\epsilon+i\omega)}\right| &\leq \mathbb{E}\left(|e^{-s_1\delta}|\cdot|e^{-\epsilon(A-B^{(2)})}|\cdot|e^{-i\omega(A-B^{(2)})}|\right)\\ &\leq \mathbb{E}e^{-\epsilon(A-B^{(2)})} = \frac{f(0,-\epsilon)}{g(0,-\epsilon)} < 1.\end{aligned}$$

Above we used the rough bound $|e^{-s_1\delta}| \leq 1$. This completes the proof.

Notice that the key role in the proof is played by $\rho_2 < 1$ and not $\rho_1 < 1$ ($\rho_2 < \rho_1$).

Chapter 7

Proportional reinsurance with subexponential claims

In this chapter, we will study the asymptotic behaviour of the ruin probability for a large initial capital for a proportional reinsurance contract. This problem can be formulated as the first crossing probability of the process of aggregate claims above a non-linear barrier. A detailed description of the problem is given in Section 7.1; the main result and its probabilistic interpretation is given in Section 7.2. The 'one large jump that causes ruin' heuristic is shown to be asymptotically valid as the initial capital grows arbitrarily large along a ray in the plane. The proofs and the technical results are found in Section 7.3.

This is part of an ongoing project with Sergey Foss, Zbigniew Palmowski and Tomasz Rolski, and the contents of this chapter are slightly related to an unpublished manuscript of S. Foss, T. Rolski and S. Zachary [62].

7.1 Introduction

Palmowski and Pistorius [91] studied boundary crossing probabilities of a stochastic process with regularly varying increments. This study was motivated by ruin probabilities of two insurance companies with proportional claims (see Avram et al. [20]) and the steady state distribution of a tandem queue with two servers (see Lieshout and Mandjes [86]). In this paper we generalize this result to the case of a strongly subexponential distribution of increments. We define

$$S_t = \sum_{k=1}^{n(t)} B_k \tag{7.1}$$

for $n(t)$ a renewal process with i.i.d. inter-arrival times A_k, and the claims B_k are i.i.d. random variables independent of $n(t)$, with the distribution function $F(x)$. Let the barriers b_1, b_2 be given by

$$b_1(t) = b_1(t; x_1) = x_1 + p_1 t, \qquad b_2(t) = b_2(t; x_2) = x_2 + p_2 t,$$

where $x_i = u^{(i)}/\delta_i$, $p_i = c_i/\delta_i$, with δ_i being the specific proportion used to split a claim B, $\delta_1 + \delta_2 = 1$, so that risk reserve i receives claim sizes distributed as $\delta_i B$, $i = 1, 2$. We assume that

$$p_1 > p_2, \qquad p_2 > \mathbb{E}B/\mathbb{E}A, \tag{7.2}$$

for generic A and B. We will consider the following boundary crossing probabilities:

$$\begin{aligned}\psi_\wedge(x_1, x_2) &= \mathbb{P}(\exists t \geq 0 : S_t > (x_1 + p_1 t) \wedge (x_2 + p_2 t)),\\ \psi_\vee(x_1, x_2) &= \mathbb{P}(\exists t \geq 0 : S_t > (x_1 + p_1 t) \vee (x_2 + p_2 t)),\end{aligned}$$

where $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$.

The $\psi_\wedge(x_1, x_2)$ describes the ruin probability of at least one insurance company, and $\psi_\vee(x_1, x_2)$ corresponds to the eventual ruin of both insurance companies. The first assumption in (7.2) means that the second company receives less premium per amount paid out and the second one is the stability condition under which the surplus levels of both insurance companies tend to infinity. The solutions of the "degenerate two-dimensional" ruin problems strongly depend on the position of the vector of premium rates $p = (p_1, p_2)$. Namely, if the initial capital levels satisfy $x_2 \leq x_1$, the two lines do not intersect, hence the "$\wedge-$" and the "$\vee-$"ruin always happen for the second and first company respectively. In this case the asymptotics follow from one-dimensional ruin theory – see e.g. Rolski et al. [96]. Therefore we focus here on the opposite case, when $x_1 < x_2$.

In this chapter, we derive the exact first order asymptotics of these ruin probabilities if x_1, x_2 tend to infinity along a ray in the positive quadrant (i.e. $x_1/x_2 = a < 1$ is constant) if the claims follow a subexponential distribution. We model the claims by subexponential distributions since many catastrophic events like earthquakes, storms, terrorist attacks etc. are modelled by them. Insurance companies use e.g. the Lognormal distribution (which is subexponential) to model car claims – see Foss et al. [60], Rolski et al. [96] or Embrechts et al. [53] for further background.

The chapter is organized as follows. In the next section we present the main results which will be proved in Section 7.3.

7.2 Main results

In order to state our results we start with recalling some notions. We assume that inter-arrival times $(A_k)_k$ are light-tailed, that is $\mathbb{E}\exp\{\theta A\} < \infty$ for some $\theta > 0$ and generic A. Moreover, we assume that the distribution F of a generic claim $B \geq 0$ belongs to the class $\mathcal{S}$ of *subexponential* distribution functions, where a distribution function $G \in \mathcal{S}$ if and only if $\overline{G}(x) > 0$ for all x and

$$\lim_{x \to \infty} \overline{G^{*2}}(x)/\overline{G}(x) = 2 \tag{7.3}$$

(where G^{*2} is the convolution of G with itself). Here $\overline{G}$ denotes the tail distribution given by $\overline{G}(x) = 1 - G(x)$. The above means that asymptotically, the only way in which a sum of two independent copies from G can be large is that precisely one of them is large. Since these are independent copies, any one of them is equally likely to be the large one, hence the limit above, having a combinatorial flavor. This limit can be extended by induction to the n-fold convolution, and then the limit becomes n, see Foss et al. [60], Cor. 3.20.

We further assume throughout that $F \in \mathcal{S}^*$, the class of *strong subexponential* distributions. A distribution function G on $\mathbb{R}$ belongs to the class $\mathcal{S}^*$ if $\overline{G}(x) > 0$ for all x, and

$$\int_0^x \overline{G}(x-y)\overline{G}(y)\,dy \sim 2m_G\overline{G}(x), \quad \text{as } x \to \infty, \tag{7.4}$$

where we denote $f(x) \sim g(x)$ to mean $\lim_{x\to\infty} f(x)/g(x) = 1$. The constant m_G represents

$$m_G = \int_0^\infty \overline{G}(x)\,dx.$$

It is known that the property $G \in \mathcal{S}^*$ depends only on the tail of G. Further, if $G \in \mathcal{S}^*$ then $G \in \mathcal{S}$ and also $G^s \in \mathcal{S}^*$, where

$$\overline{G^s}(x) = \min\left(1, \int_x^\infty \overline{G}(t)\,dt\right)$$

is the integrated, or *second-tail*, distribution function determined by G; see Foss et al. [60] §3.4 for details.

Theorem 7.2.1. *We have, for $a < 1$:*

$$\psi_\wedge(aK, K) \sim H(K), \tag{7.5}$$
$$\psi_\vee(aK, K) \sim U(K), \tag{7.6}$$

as $K \to \infty$, where

$$H(K) = \int_0^\infty \overline{F}(\min\{aK + t(p_1\mathbb{E}A - \mathbb{E}B), K + t(p_2\mathbb{E}A - \mathbb{E}B)\})\mathrm{d}t,$$

$$U(K) = \int_0^\infty \overline{F}(\max\{aK + t(p_1\mathbb{E}A - \mathbb{E}B), K + t(p_2\mathbb{E}A - \mathbb{E}B)\})\mathrm{d}t.$$

7.3 Proof of the main result

Before we give the formal proof, we owe some explanations. Identities (7.5) and (7.6) can be explained by the 'single large jump' principle, that is, for large values of the starting capital, the most likely way in which the corresponding type of ruin can happen is through a single large claim that makes the aggregate claims process S_n jump directly above the corresponding barrier. Up to the time of ruin, the risk reserve process behaves in a typical way - looked at on a large scale, component i drifts

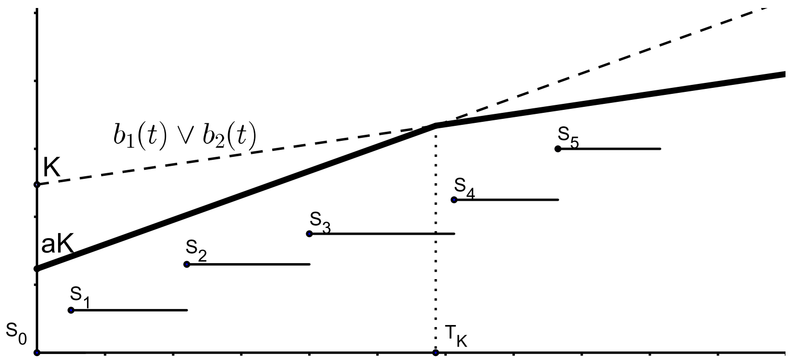


FIGURE 7.1: A path of the aggregate claims process. The upper (dashed) barrier is $b_1(t) \vee b_2(t)$.

upwards at rate $p_i\mathbb{E}A - \mathbb{E}B > 0$, $i = 1, 2$ - that is, the risk reserve processes receive proportions of claims of a size that is relatively close to their mean, compared to the claim that causes ruin.

Up to this point, the heuristic is the same as in the one-dimensional exit problem of a random walk with heavy-tailed increments above a linear barrier. But there is an additional key insight that appears also in a formal way in the proof: asymptotically in K, the large jump will always happen after the process of inter-arrivals $\sum_{k=1}^{n} A_k$ had stabilized around its mean drift (by the law of large numbers). Formally, with the help of Lemma 7.3.1, we show that there exists a rate function $\epsilon(K) \to 0$ sufficiently slowly, as $K \to \infty$, such that the residual terms $I_1(K, \epsilon(K))$ and $I_0(K, \epsilon(K))$ that appear in (7.9) and (7.12) are asymptotically negligible w.r.t. $H(K)$ as $K \to \infty$. Below we review some results about heavy-tailed distributions, which will be used in the proof.

A known property of a long-tailed distribution F with $\int_0^\infty \overline{F}(y)\mathrm{d}y < \infty$, is that F^s is long-tailed again and it holds that (see Foss et al. [60], Lemma 2.26)

$$\overline{F}(x) = o(\overline{F}^s(x)), \text{ as } x \to \infty. \tag{7.7}$$

By a long-tailed distribution function F, it is meant one such that $\overline{F}(x) > 0$, for all x, and, for any fixed $y > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1.$$

All of the heavy-tailed distributions that typically appear in practice are also long tailed. Since F^s is long-tailed, it admits an insensitivity function h which is by definition such that

$$\lim_{x\to\infty} \frac{\overline{F}^s(x+h(x))}{\overline{F}^s(x)} = 1, \tag{7.8}$$

and $h(x) \to \infty$ as $x \to \infty$ sufficiently slowly. The rate at which the insensitivity function h tends to infinity depends on how heavy the tail of F^s is. For such an h we can strengthen the asymptotic behaviour (7.7) to

$$h(x)\overline{F}(x) = o(\overline{F}^s(x)),$$

and this will be a useful consequence for our purposes, see (7.14) and (7.16).

Proof of Theorem 7.2.1. Let $S_n = \sum_{k=1}^n B_k$ and $\tilde{S}_n := S_n - n\mathbb{E}B$ be the centered random walk. Denote by

$$N_\epsilon^* = \inf\{n : \forall k \geq n, \quad t_k \in [(\mathbb{E}A - \epsilon)k, (\mathbb{E}A + \epsilon)k]\}$$

for some $\epsilon > 0$, to be the first time the random walk $t_n = \sum_{k=1}^n A_k$ will stabilize around its mean. By the strong law of large numbers (SLLN), we have $N_\epsilon^* < \infty$ a.s. and we can bound

$$I_2 \leq \psi_\wedge(aK, K) \leq I_1 + I_2, \tag{7.9}$$

where

$$\begin{aligned} I_1 &:= \mathbb{P}(\exists n \leq N_\epsilon^* : \tilde{S}_n > \min\{aK + p_1 t_n - n\mathbb{E}B, K + p_2 t_n - n\mathbb{E}B\}) \\ I_2 &:= \mathbb{P}(\exists n > N_\epsilon^* : \tilde{S}_n > \min\{aK + p_1 t_n - n\mathbb{E}B, K + p_2 t_n - n\mathbb{E}B\}) \end{aligned}$$

We will show below that $I_1 = o(H(K))$ and $I_2 \sim H(K)$, as $K \to \infty$. Observe that, by the definition of N_ϵ^*,

$$I_2 \leq \mathbb{P}(\exists n \geq 0 : \tilde{S}_n > \min\{aK + p_1(\mathbb{E}A - \epsilon)n - n\mathbb{E}B, K + p_2(\mathbb{E}A - \epsilon)n - n\mathbb{E}B\}). \tag{7.10}$$

If we denote with $g(n) := \min\{aK + p_1(\mathbb{E}A - \epsilon)n - n\mathbb{E}B, K + p_2(\mathbb{E}A - \epsilon)n - n\mathbb{E}B\}$ that appears inside the upper bound in (7.10), then we will have for some positive c

$$g(n+1) \geq g(n) + c,$$

and if we take $\epsilon > 0$ sufficiently small, we can choose $0 < c < p_2\mathbb{E}A - \mathbb{E}B - p_2\epsilon$, and such that it does not depend on ϵ. From the main result of Foss et al. [61], the upper bound in (7.10) is asymptotically equivalent to:

$$(1 + \delta_c(K)) \int_0^\infty \overline{F}(\min\{aK + (p_1\mathbb{E}A - \mathbb{E}B - p_1\epsilon)t, K + (p_2\mathbb{E}A - \mathbb{E}B - p_2\epsilon)t\})\mathrm{d}t, \tag{7.11}$$

with $\delta_c(K) \to 0$ as $K \to \infty$. This establishes the asymptotic behaviour of the upper bound of I_2. Now we turn to the lower bound on I_2.

$$\begin{aligned} I_2 \geq\ & \mathbb{P}(\exists n \geq 0 : \tilde{S}_n > \min\{aK + p_1(\mathbb{E}A + \epsilon)n - n\mathbb{E}B, K + p_2(\mathbb{E}A + \epsilon)n - n\mathbb{E}B\}) \\ & - \mathbb{P}(\exists n \leq N_\epsilon^* : \tilde{S}_n > \min\{aK + p_1(\mathbb{E}A + \epsilon)n - n\mathbb{E}B, K + p_2(\mathbb{E}A + \epsilon)n - n\mathbb{E}B\}). \end{aligned} \tag{7.12}$$

We denote the second increment on the RHS of the above inequality by I_0. We will show below that this together with I_1 is asymptotically negligible $o(H(K))$ as $K \to \infty$.

Using again the main result in Foss et al. [61], gives the asymptotic behaviour of the first term on the RHS of (7.12):

$$(1-\delta_c(aK))\int_0^\infty \overline{F}(\min\{aK+(p_1\mathbb{E}A-\mathbb{E}B+p_1\epsilon)t, K+(p_2\mathbb{E}A-\mathbb{E}B+p_2\epsilon)t\})\mathrm{d}t. \quad (7.13)$$

We will prove now that for sufficiently small $\epsilon > 0$,

$$I_1 = \mathbb{P}(\exists n \le N_\epsilon^* : \tilde{S}_n > \min\{aK + p_1 t_n - n\mathbb{E}B, K + p_2 t_n - n\mathbb{E}B\})$$

and

$$I_0 = \mathbb{P}(\exists n \le N_\epsilon^* : \tilde{S}_n > \min\{aK + p_1(\mathbb{E}A+\epsilon)n - n\mathbb{E}B, K + p_2(\mathbb{E}A+\epsilon)n - n\mathbb{E}B\})$$

are asymptotically negligible.

Since $F^s \in \mathcal{S}^*$, a useful characterization of long-tailed distribution functions yields the existence of a non decreasing function $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that $h(x) \to \infty$ as $x \to \infty$ and (7.8) holds. An immediate consequence of this insensitivity property is that (see Foss et al. [60], Lemmae 2.19, 2.25),

$$h(x)\overline{F}(x) = o(F^s(x)). \quad (7.14)$$

Since N_ϵ^* is independent of S_n, we have the following upper bound for both I_0 and I_1 (see [60], Thm. 3.37, Cor. 3.20):

$$I_0,\ I_1 \le \mathbb{P}(\exists n \le N_\epsilon^* : S_n > aK) \sim \mathbb{E}N_\epsilon^*\overline{F}(aK), \quad (7.15)$$

once we show that $\mathbb{E}N_\epsilon^* < \infty$ for any $\epsilon > 0$. The latter statement follows from Cramér's Large Deviations Principle (LDP) for the light-tailed sequence $\{A_k\}_{k\ge1}$, which has the good rate function $\Lambda^*(x)$:

$$\mathbb{P}(|\tilde{t}_l| \ge \epsilon l) \le 2e^{-l\inf_{x\ge\epsilon}\Lambda^*(x)},$$

with $\tilde{t}_l := \sum_{i=1}^l A_i - l\mathbb{E}A$, so that we can write

$$\mathbb{P}(N_\epsilon^* > k) \le \sum_{l\ge k}\mathbb{P}(|\tilde{t}_l| \ge \epsilon l) \le \sum_{l\ge k} 2e^{-I(\epsilon)l} \le C_\epsilon e^{-I(\epsilon)k},$$

with $I(\epsilon) := \inf_{x\ge\epsilon}\Lambda^*(x)$, and C_ϵ is some positive constant that depends on ϵ. See Cramér's Theorem 2.2.3 and the subsequent Remark c) in Dembo and Zeitouni [49], p. 27. For the arguments below, it is essential that $\mathbb{E}N_\epsilon^*$ explodes as ϵ tends to 0.

Define $f(x) := \mathbb{E}N_{1/x}^*$; by construction, f is non-decreasing and $f(x) \to \infty$ as $x \to \infty$. We show in Lemma 7.3.1 that there exists a non-negative function $f^*(y)$, such that $f^*(y) \to \infty$ as $y \to \infty$ and

$$f \circ f^*(y) \le y, \quad \text{for all } y > 0.$$

With this construction, take $\epsilon = 1/f^*(h(K))$, so that we have by the definitions of f^* and of h that

$$\mathbb{E}N_\epsilon^* = f \circ f^* \circ h(K) \le h(K),$$

and from (7.15), with this choice for ϵ, the asymptotic behaviour of the upper bound on $I_1(K)$ and $I_0(K)$ is

$$\mathbb{E}N_\epsilon^* \overline{F}(aK) \le h(K)\overline{F}(aK) = \mathrm{o}(F^s(aK)) = \mathrm{o}(H(K)), \tag{7.16}$$

by virtue of (7.14). This completes the argument for the asymptotic behaviour of $I_1(K)$ and $I_0(K)$.

All that is left is a limit argument for taking $\epsilon \to 0$. Let us denote the integrated tails that appear in (7.11) and (7.13)

$$\begin{aligned} H_{-\epsilon}(K) &:= \int_0^\infty \overline{F}(\min\{aK + (p_1\mathbb{E}A - \mathbb{E}B - p_1\epsilon)t, K + (p_2\mathbb{E}A - \mathbb{E}B - p_2\epsilon)t\})\mathrm{d}t, \\ H_{\epsilon}(K) &:= \int_0^\infty \overline{F}(\min\{aK + (p_1\mathbb{E}A - \mathbb{E}B + p_1\epsilon)t, K + (p_2\mathbb{E}A - \mathbb{E}B + p_2\epsilon)t\})\mathrm{d}t. \end{aligned} \tag{7.17}$$

It follows from Lemma 7.3.2 below that both $H_\epsilon(K)/H(K)$ and $H_{-\epsilon}(K)/H(K)$ converge to 1 as $K \to \infty$. The proof of (7.5) is complete because we can now write for ϵ small, using (7.10), (7.11), (7.13) and (7.16):

$$(1 - \delta_c(aK))\frac{H_\epsilon(K)}{H(K)} - o(1) \le \frac{\psi_\wedge(K)}{H(K)} \le (1 + \delta_c(aK))\frac{H_{-\epsilon}(K)}{H(K)} + o(1), \quad K \to \infty,$$

and with the choice $\epsilon(K) = 1/f^*(h(K))$.

The proof of (7.6) follows along exactly the same lines. The only place where the shape of the barrier is used is Lemma 7.3.2. But also in this case, it can be shown that the corresponding functionals $U_{-\epsilon}$, U_ϵ are proportional to U, and the proportionality constants tend to 1 as $\epsilon \to 0$. This completes the proof.

Lemma 7.3.1. *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function such that $f(x) \to \infty$ as $x \to \infty$. Then there exists a non-decreasing function $f^* : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f^*(y) \to \infty$ as $y \to \infty$ and*

$$f \circ f^*(y) \le y, \quad \textit{for all } y > 0.$$

Proof of Lemma 7.3.1. There are several possible choices for f^*. A natural one is the following: let f^+ be the right-continuous version of f. Since f is non-decreasing, $f(x) \le f^+(x)$. For a fixed y, set

$$f^*(y) := \sup\{x;\ f^+(x) \le y\}.$$

We have by construction that

$$f^+ \circ f^*(y) \leq y.$$

Since f is non-decreasing, we can write

$$f \circ f^*(y) \leq f^+ \circ f^*(y) \leq y. \tag{7.18}$$

The monotonicity of f^* together with $f^*(y) \to \infty$ as $y \to \infty$ follow from its definition and the assumption $f(x) \to \infty$. Thus f^* is constructed as the right-inverse of f^+, and this is the mirror-image construction of the well known quantile function (the left inverse) of a non-decreasing and right-continuous function – a c.d.f. (see Embrechts and Hofert [52] for the analogue of (7.18)). The proof is complete.

Lemma 7.3.2. *With the notations and definitions from the proof of Theorem 7.2.1, the functions $H_\epsilon(K)$, $H(K)$ and $H_{-\epsilon}(K)$ are proportional for any $K > 0$ and $\epsilon > 0$:*

$$H_\epsilon(K) = \frac{m_1}{m_{\epsilon,1}} H(K).$$

$$H_{-\epsilon}(K) = \frac{m_1}{m_{-\epsilon,1}} H(K).$$

Here we denoted by $m_i := p_i \mathbb{E}A - \mathbb{E}B$ and $m_{\pm\epsilon,i} := p_i(\mathbb{E}A \pm \epsilon) - \mathbb{E}B$.

Proof of Lemma 7.3.2.

Define $T := \frac{(1-a)K}{m_1 - m_2}$ the (unique) crossing epoch of lines $aK + m_1 t$ and $K + m_2 t$ and let

$$T_{\pm\epsilon} := \frac{(1-a)K}{(m_{\pm\epsilon,1} - m_{\pm\epsilon,2})}$$

be the (unique) epochs at which the lines $aK + p_1(\mathbb{E}A \pm \epsilon)t - t\mathbb{E}B$ and $K + p_2(\mathbb{E}A \pm \epsilon)t - t\mathbb{E}B$ cross.

There are several key relations that these crossing times satisfy, and they are easy to check:

$$m_{\epsilon,1} T_\epsilon = m_1 T, \quad m_{\epsilon,2} T_\epsilon = m_2 T, \quad \frac{m_1}{m_{\epsilon,1}} = \frac{m_2}{m_{\epsilon,2}}. \tag{7.19}$$

Similar relations hold for $T_{-\epsilon}$ and $m_{-\epsilon,i}$. Using the above definitions, we can write

$$H(K) = \frac{1}{m_1} \int\limits_{aK}^{aK+m_1T} \overline{F}(u)\mathrm{d}u + \frac{1}{m_2} \int\limits_{K+m_2T}^{\infty} \overline{F}(u)\mathrm{d}u,$$

$$H_\epsilon(K) = \frac{1}{m_{\epsilon,1}} \int\limits_{aK}^{aK+m_{\epsilon,1}T_\epsilon} \overline{F}(u)\mathrm{d}u + \frac{1}{m_{\epsilon,2}} \int\limits_{K+m_{\epsilon,2}T_\epsilon}^{\infty} \overline{F}(u)\mathrm{d}u,$$

$$H_{-\epsilon}(K) = \frac{1}{m_{-\epsilon,1}} \int\limits_{aK}^{aK+m_{-\epsilon,1}T_{-\epsilon}} \overline{F}(u)\mathrm{d}u + \frac{1}{m_{-\epsilon,2}} \int\limits_{K+m_{-\epsilon,2}T_{-\epsilon}}^{\infty} \overline{F}(u)\mathrm{d}u.$$

Above we partitioned the integrated tails based on the crossing times of the barriers and we changed the variables of integration in each term. We did this in order to be better able to compare the three integrated tails.

From (7.19), the integration ends are termwise identical for $H(K)$ and $H_\epsilon(K)$. Moreover, using the last identity in (7.19), $H(K)$ and $H_\epsilon(K)$ are proportional:

$$H_\epsilon(K) = \frac{m_1}{m_{\epsilon,1}} H(K).$$

It follows in the same way as for H_ϵ that

$$H_{-\epsilon}(K) = \frac{m_1}{m_{-\epsilon,1}} H(K).$$

The proof is complete.

Chapter 8

A coupled processor model with simultaneous arrivals

In this chapter we study a coupled processor model which receives service requirements at both queues simultaneously. We assume that the service requirement at, say, station 1 is always greater than the service requirement from station 2. Moreover, when server 2 is idle, it switches to process work from the first queue, if there is any. As described in Section 1.2, for the models introduced by Fayolle and Iasnogorodski [55] and Cohen and Boxma [41], the server i rate is r_i when processing from buffer i and r_j^* when processing from buffer j, $i = 1, 2$, $j \neq i$, so that during an idle period of server i, the other server works at speed $r_j + r_j^*$ (the notation we use is different than the one used in [41], but equivalent). By studying the dynamics of this system during infinitesimal time intervals, it is possible to derive a functional equation for the transform of the joint amount of work. This analysis was carried out in Cohen [43]. The steady-state version of the functional equation for the workload vector $(V^{(1)}, V^{(2)})$ from [43, p. 186, (1.10)] reads, using our notation:

$$K(s_1, s_2)\psi(s_1, s_2) = (r_2 s_2 - r_1^* s_1)\psi_1(s_1) + (r_1 s_1 - r_2^* s_2)\psi_2(s_2) + (r_1^* s_1 + r_2^* s_2)\psi_0, \quad (8.1)$$

with $\psi(s_1, s_2)$ the Laplace-Stieltjes transform (LST) of the stationary amount of work in the system and with the unknown boundary functions

$$\psi_i(s_i) = \mathbb{E}[e^{-s_i V^{(i)}}(V^{(j)} = 0)],\ i \neq j \in \{1, 2\}, \quad \psi_0 = \mathbb{P}(V^{(1)} = V^{(2)} = 0).$$

The function $K(s_1, s_2)$ is the so-called Poisson kernel. For $\phi(s_1, s_2)$ the joint transform of a generic service time vector,

$$K(s_1, s_2) = r_1 s_1 + r_2 s_2 - \lambda[1 - \phi(s_1, s_2)].$$

Actually, in Cohen [43] the functional equation of the time dependent workload is given. The stationary version above is obtained by multiplying the functional equation

[43, p. 186, (1.10)] with the discount factor of the Laplace-Stieltjes transform over time, and then taking the discount factor to 0, while keeping it positive (this is by virtue of Abel's theorem).

The analysis in the above mentioned works relies heavily on the theory of complex functions which makes it highly non-trivial, and in addition, it is difficult to recognize the probabilistic nature of the initial problem.

We will show in Section 8.2 that under the additional ordering assumption between the claims, it is possible to relate the coupled processor model to a parallel queueing system without coupling. Then the transform of the amount of work in the coupled system follows from that obtained in the decoupled parallel system, by using Formula (5.6). This gives an explicit representation for the steady-state amount of work in the coupled system (Section 8.3), which can be extended to multiple coupled queues by making suitable assumptions on the coupling rates (Section 8.4).

8.1 Model description

We consider two parallel M/G/1 queues, with simultaneous arrivals and correlated service requirements. The arrival process is a Poisson process with rate λ. We will denote by A_n the time elapsed between arrival epochs n and $n+1$, $n \geq 1$. The service requirements at the two queues of successive customers are independent, identically distributed random vectors $(B_n^{(1)}, B_n^{(2)}), n \geq 1$. In the sequel we denote with $(B^{(1)}, B^{(2)})$ a random vector with the same distribution as the vectors $(B_n^{(1)}, B_n^{(2)}), n \geq 1$. The joint Laplace-Stieltjes transform of this vector is

$$\phi(s_1, s_2) := \mathbb{E}(e^{-s_1 B^{(1)} - s_2 B^{(2)}}).$$

We will work with the processing rates of server i, when working in buffer $j \neq i$ for a three dimensional system in Section 8.4. For this reason, let us change the rates notation and set c_1 and c_2 for the servers' processing rates from their respective buffer and c_i^j, $i \neq j$ for the processing rate of server i in buffer j. An essential assumption in the model is that, after normalizing the system with the server rates, with probability one, each customer has a bigger service requirement in queue 1 than in queue 2, i.e.,

$$\mathbb{P}(B^{(1)}/c_1 \geq B^{(2)}/c_2) = 1.$$

The processors are coupled in the sense that as soon as server 2 becomes idle, it switches its capacity to serving work from buffer 1 so that it works at rate c_2^1 when processing from buffer 1. Because of the ordering assumption, server 1 will never become idle while server 2 is busy, so there is no need for server 2 to help server 1. We are interested in the joint stationary distribution of the amount of work in the two queues. Let us still denote with $V^{(1)}$ the stationary amount of work in queue 1 and with $V^{(2)}$ the stationary amount of work in queue 2.

Our aim is to study such an interacting queueing system under the above assumption of ordered service times. In particular we want to find an expression for the Laplace-Stieltjes transform of the joint stationary amount of work.

8.2 Recursive equations for the amount of work in the coupled system

In this section we will derive stochastic recursive equations for the joint amount of work in the system.

With $V_{t_n}^{(i)}$ the amount of work in queue i as seen by customer n upon arrival, we assume that at time 0 the first customer arrives in an empty system. Then we have the following recursion for the random variables $(V_{t_n}^{(1)}, V_{t_n}^{(2)}), n \geq 1$, which is a special case of Lindley's recursion for the coupled processor model (2.14):

$$
\begin{aligned}
&(V_{t_1}^{(1)}, V_{t_1}^{(2)}) = (0,0), \\
&V_{t_{n+1}}^{(1)} = \left[(V_{t_n}^{(1)} + B_n^{(1)} - c_1 A_n) \vee 0 + \frac{c_2^1}{c_2} (V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \wedge 0 \right] \vee 0, \\
&V_{t_{n+1}}^{(2)} = (V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \vee 0.
\end{aligned} \tag{8.2}
$$

The second queue evolves without the assistance of the first queue because of the ordering assumption. Remark that

$$
-\frac{1}{c_2}[(V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \wedge 0]
$$

is the amount of time that server 2 has been idle between the arrival epochs t_n and t_{n+1} (cf. (1.3)). Then the second term in the recursion of $V_{t_{n+1}}^{(1)}$ is minus the amount of work server 2 processes at rate c_2^1 from buffer 1 during his idle period (if it has an idle period).

This extra term appears because the servers are coupled. It is useful to compare this with the recursion for a system without coupling between the servers. Consider two parallel queues simultaneously receiving service requirements distributed as the vector $(\tilde{B}^{(1)}, \tilde{B}^{(2)})$. The servers are not coupled anymore and server 1 always works at speed $\tilde{c}_1$, while server 2 always works at speed $\tilde{c}_2$. Let $(\tilde{V}_{t_n}^{(1)}, \tilde{V}_{t_n}^{(2)})$ be the amount of work at arrival epochs in such a system, then the following recursion holds

$$
\begin{aligned}
&(\tilde{V}_{t_1}^{(1)}, \tilde{V}_{t_1}^{(2)}) = (0,0), \\
&\tilde{V}_{t_{n+1}}^{(1)} = (\tilde{V}_{t_n}^{(1)} + \tilde{B}_n^{(1)} - \tilde{c}_1 \tilde{A}_n) \vee 0, \\
&\tilde{V}_{t_{n+1}}^{(2)} = (\tilde{V}_{t_n}^{(2)} + \tilde{B}_n^{(2)} - \tilde{c}_2 \tilde{A}_n) \vee 0,
\end{aligned} \tag{8.3}
$$

with $\tilde{A}_n$ the inter-arrival time between customer n and $n+1$. This is Lindley's recursion because both queues evolve in isolation as one-dimensional systems. Notice that, marginally, queue 2 evolves as if there was no coupling in (8.2).

We are ready to give the main result of this section which connects the amount of work in the coupled system to a workload process in a system without coupling between the servers. For further usage we will denote by system (C) the coupled system and by system (D), the system without coupling.

Proposition 8.2.1. *Let* $(V_{t_n}^{(1)}, V_{t_n}^{(2)})_{n\geq 1}$ *be the workload process at arrival epochs in system (C). Then the process* $(V_{t_n}^{(1)} + \frac{c_2^1}{c_2} V_{t_n}^{(2)}, V_{t_n}^{(2)})_{n\geq 1}$ *is the workload process in a system of type (D) with generic input of the form* $(B^{(1)} + \frac{c_2^1}{c_2} B^{(2)}, B^{(2)})$*, where the servers have speed* $(c_1 + c_2^1, c_2)$ *and do not interact with each other.*

We will show that $(V_{t_n}^{(1)} + \frac{c_2^1}{c_2} V_{t_n}^{(2)}, V_{t_n}^{(2)})$ has the same distribution as $(\tilde{V}_{t_n}^{(1)}, \tilde{V}_{t_n}^{(2)})$, the solution to the recursive system (8.3), by using a probabilistic coupling between systems (C) and (D), that is we will let the two systems evolve on the same probability space given by the sequences $(A_n)_{n\geq 1}$ and $(B_n^{(1)}, B_n^{(2)})_{n\geq 1}$. The choice for the input variables in system (D) is the following:

$$\tilde{A}_n := A_n,\ (\tilde{B}_n^{(1)}, \tilde{B}_n^{(2)}) := (B_n^{(1)} + \frac{c_2^1}{c_2} B_n^{(2)}, B_n^{(2)}),$$

$$\tilde{c}_1 := c_1 + c_2^1, \quad \tilde{c}_2 := c_2. \tag{8.4}$$

To be more precise, start both systems empty at time $t_1 = 0$. At the nth arrival epoch t_n, system (C) receives input $(B_n^{(1)}, B_n^{(2)})$ whereas system (D) receives input $(B_n^{(1)} + \frac{c_2^1}{c_2} B_n^{(2)}, B_n^{(2)})$. Let us focus on system (D). The key idea is to partition the amount of work at queue 1 in system (D) into $V_t^{(1)}$ and $\frac{c_2^1}{c_2} V_t^{(2)}$, then during the busy periods of server 2, distribute the total capacity per time unit $c_1 + c_2^1$ of server 1 in the following way: c_1 is dedicated to processing $V_t^{(1)}$ while c_2^1 is used to process the remaining $\frac{c_2^1}{c_2} V_t^{(2)}$. In this way, during the busy periods of server 2, the amount of work in queue one of system (C) and the work in the c_1-dedicated component of queue one in system (D) evolve in the same way.

As soon as queue 2 becomes empty (which now happens at the same moment in system (C) as in system (D) because the second queue evolves unchanged between the two systems), in both systems (C) and (D), server 1 will process work $V_t^{(1)}$ at speed $c_1 + c_2^1$. Another remark is that due to the ordering between the service requirements, queue 2 will always become idle before queue 1 in any of the systems (C) or (D) (see also Remark 8.2.1 below).

We give below the formal proof of Proposition 8.2.1. The idea of the proof is to verify that $V_{t_n}^{(1)} + \frac{c_2^1}{c_2} V_{t_n}^{(2)}$ satisfies (8.3) with the input variables from (8.4).

Proof of Proposition 8.2.1. First remark that we can drop the maximum w.r.t. 0 in the first term of recursion (8.2):

$$V_{t_{n+1}}^{(1)} = \left[V_{t_n}^{(1)} + B_n^{(1)} - c_1 A_n + \frac{c_2^1}{c_2} (V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \wedge 0 \right] \vee 0, \tag{8.5}$$

the reason being that the other term is either 0 or negative as pointed out below (8.2), so it can only decrease the term between the square brackets in the recursion of the coupled queue 1.

Adding the term $\frac{c_2^1}{c_2} V_{t_{n+1}}^{(2)} = \frac{c_2^1}{c_2} (V_{t_n}^{(2)} + B_n^{(2)} - c_2 A_n) \vee 0$ to both sides of (8.5) gives

$$V^{(1)}_{t_{n+1}} + \frac{c_2^1}{c_2} V^{(2)}_{t_{n+1}} = \left[V^{(1)}_{t_n} + B^{(1)}_n - c_1 A_n + \frac{c_2^1}{c_2} (V^{(2)}_{t_n} + B^{(2)}_n - c_2 A_n) \right] \\ \vee \frac{c_2^1}{c_2} (V^{(2)}_{t_n} + B^{(2)}_n - c_2 A_n) \vee 0.$$

We used the fact that the operator $\vee$ is distributive w.r.t. addition and the obvious decomposition $(x \wedge 0) + (x \vee 0) = x$.

There are two possible cases:

In the event that $V^{(1)}_{t_n} + B^{(1)}_n - c_1 A_n > 0$, the RHS above is of the form $a \vee b \vee 0$, with $a > b$, hence b can be removed. In the end we can rewrite the above as

$$V^{(1)}_{t_{n+1}} + \frac{c_2^1}{c_2} V^{(2)}_{t_{n+1}} = \left(V^{(1)}_{t_n} + \frac{c_2^1}{c_2} V^{(2)}_{t_n} + B^{(1)}_n + \frac{c_2^1}{c_2} B^{(2)}_n - (c_1 + c_2^1) A_n \right) \vee 0.$$

This is the desired Lindley recursion for $V^{(1)}_{t_n} + \frac{c_2^1}{c_2} V^{(2)}_{t_n}$.

In the event that $V^{(1)}_{t_n} + B^{(1)}_n - c_1 A_n < 0$, queue 1 would empty at epoch t_{n+1} without any additional help, so that $V^{(1)}_{t_{n+1}} = 0$. Then by the ordering assumption, $V^{(2)}_{t_{n+1}} = 0$ as well, and the above identity is trivially satisfied. The proof is complete and it shows that the ordering between the normalized claims is an essential assumption.

Remark 8.2.1. *Notice that the normalized input in the related system (D) remains ordered:*

$$\frac{B^{(1)}_n + \frac{c_2^1}{c_2} B^{(2)}_n}{c_1 + c_2^1} = \frac{c_1 \frac{B^{(1)}_n}{c_1} + c_2^1 \frac{B^{(2)}_n}{c_2}}{c_1 + c_2^1} \geq \frac{B^{(2)}_n}{c_2},$$

because by assumption, $B^{(1)}_n / c_1 \geq B^{(2)}_n / c_2$.

8.3 The transform of the equilibrium amount of work at arrival epochs

In Chapter 5 has been shown how to calculate the Laplace-Stieltjes transform of the joint, stationary amount of work in a system of type (D) under the same ordering assumption (Chapter 5, Formula (5.6)). Thus, by inverting the correspondence from Proposition 8.2.1, and using Remark 8.2.1, we can recover without any additional effort the joint transform of $(V^{(1)}, V^{(2)})$, the steady-state amounts of work in the coupled system. Moreover, because of Poisson arrivals, we have the PASTA property, which means that in equilibrium, the amount of work is the same as the workload seen by an arriving customer.

The inverse relation from Proposition 8.2.1 is

$$(V^{(1)}, V^{(2)}) = (\tilde{V}^{(1)} - \frac{c_2^1}{c_2}\tilde{V}^{(2)}, \tilde{V}^{(2)}). \tag{8.6}$$

If we denote by

$$\psi_{(C)}(s_1, s_2) := \mathbb{E}(e^{-s_1 V^{(1)} - s_2 V^{(2)}}), \quad \psi_{(D)}(s_1, s_2) := \mathbb{E}(e^{-s_1 \tilde{V}^{(1)} - s_2 \tilde{V}^{(2)}}),$$

the LST of the equilibrium amount of work in system (C) and respectively system (D), then via (8.6), the relation between the LSTs becomes:

$$\psi_{(C)}(s_1, s_2) = \psi_{(D)}(s_1, s_2 - s_1 c_2^1/c_2). \tag{8.7}$$

Remark 8.3.1. *Equation (8.1) can be adapted to describe the stationary workload $(\tilde{V}^{(1)}, \tilde{V}^{(2)})$ in the decoupled system by setting the coupling rates r^* equal to 0. The kernel $K(s_1, s_2)$ has to be modified as well to*

$$K_{(D)}(s_1, s_2) = (c_1 + c_2^1)s_1 + c_2 s_2 - \lambda[1 - \phi(s_1, s_2 + s_1 c_2^1/c_2)]$$

because server 1 receives the extra input $c_2^1/c_2 B^{(2)}$ and always works at speed $c_1 + c_2^1$. Then (8.1) becomes

$$K_{(D)}(s_1, s_2)\psi_{(D)}(s_1, s_2) = c_2 s_2 \psi_{(D),1}(s_1) + (c_1 + c_2^1)s_1 \psi_{(D),0}, \tag{8.8}$$

since by the ordering relation from Remark 8.2.1, $\psi_{(D),2}(s_2)$ is constant and equal to $\psi_{(D),0}$:

$$\psi_{(D),2}(s_2) = \mathbb{E}[e^{-s_2 \tilde{V}^{(2)}}(\tilde{V}^{(1)} = 0)] = \mathbb{P}(\tilde{V}^{(1)} = \tilde{V}^{(2)} = 0) =: \psi_{(D),0}.$$

On the other hand, the kernel for the coupled system (C) is

$$K_{(C)}(s_1, s_2) = c_1 s_1 + c_2 s_2 - \lambda[1 - \phi(s_1, s_2)].$$

The equation that has to be satisfied by $\psi_{(C)}(s_1, s_2)$ now reads (with $\psi_{(C),2}(s_2) \equiv \psi_{(C),0}$, again because of the ordering)

$$K_{(C)}(s_1, s_2)\psi_{(C)}(s_1, s_2) = (c_2 s_2 - c_2^1 s_1)\psi_{(C),1}(s_1) + (c_1 + c_2^1)s_1 \psi_{(C),0} \tag{8.9}$$

with the key remark that the two boundary functions $\psi_{(D),1}(s_1)$ and $\psi_{(C),1}(s_1)$ are identical (Proposition 8.2.1):

$$\mathbb{E}[e^{-s_1 \tilde{V}^{(1)}}(\tilde{V}^{(2)} = 0)] \equiv \mathbb{E}[e^{-s_1 V^{(1)}}(V^{(2)} = 0)]$$

and the same holds for $\psi_{(C),0}$ and $\psi_{(D),0}$.

Now it is easy to check that if $\psi_{(D)}(s_1, s_2)$ is the solution of (8.8), then $\psi_{(D)}(s_1, s_2 - s_1 c_2^1/c_2)$ as in (8.7) is the solution of (8.9), and conversely, if $\psi_{(C)}(s_1, s_2)$ is the solution *of (8.9) then $\psi_{(C)}(s_1, s_2 + c_2^1/c_2 s_1)$ is the solution of (8.8). In particular, it follows from Chapter 5 that the amount of work in system (C) is ergodic under the condition*

$$\lim_{\substack{s_1\to 0\\ s_1>0}} \frac{\partial}{\partial s_1} K_{(D)}(s_1, 0) > 0 \Leftrightarrow \mathbb{E}(B^{(1)}) + c_2^1/c_2\mathbb{E}(B^{(2)}) < (c_1 + c_2^1)\mathbb{E}(A). \tag{8.10}$$

This simply means that the first queue in system (D) is capable to handle the amount of input per time unit while working at speed $(c_1 + c_2^1)$ and this is sufficient to ensure that the entire system is stable, because of the ordering assumption. It may happen that $\mathbb{E}(B^{(1)}) > c_1\mathbb{E}(A)$, i.e. that queue 1 of system (C) would be supercritical if it were to work only on its own. Inequality (8.10) together with Remark 8.2.1, implies $\mathbb{E}(B^{(2)}) < c_2\mathbb{E}(A)$, thus queue 2 is ergodic and during its (non-degenerate) idle periods it is capable to maintain queue 1 stable because of the coupling.

Using the relation between $\psi_{(C)}(s_1, s_2)$ and $\psi_{(D)}(s_1, s_2)$ we obtain

Theorem 8.3.1. *Under the stability condition (8.10), with $\tilde{\rho}_1 := \lambda\mathbb{E}(\tilde{B}^{(1)})/(c_2^1 + c_1)$, the joint transform of the stationary amount of work in a system of type (C) is*

$$\psi_{(C)}(s_1, s_2) = (1-\tilde{\rho}_1)\frac{(c_1 + c_2^1)s_1}{c_1 s_1 + c_2 s_2 - \lambda[1-\phi(s_1, s_2)]} \cdot \frac{c_2 S_2(s_1) - c_2 s_2 + s_1 c_2^1}{c_2 S_2(s_1)}. \tag{8.11}$$

For each fixed s_1 with $\mathcal{R}e\, s_1 > 0$, $S_2(s_1)$ is the zero of the equation

$$K_{(C)}(s_1, S_2(s_1) + s_1 c_2^1/c_2) = 0,$$

that is unique in the positive half of the complex plane.

Proof. The derivation for the decoupled system is known (c.f. Chapter 5, Formula (5.6)). It is easy to adapt the analysis in Section 5.2 to give the transform of the workload when the servers' speeds are not normalized. The joint transform of the stationary amount of work in the decoupled system becomes

$$\psi_{(D)}(s_1, s_2) = (1-\tilde{\rho}_1)\frac{\tilde{c}_1 s_1}{\tilde{c}_1 s_1 + \tilde{c}_2 s_2 - \lambda[1-\phi_{(D)}(s_1, s_2)]} \cdot \frac{\tilde{c}_2 S_2(s_1) - \tilde{c}_2 s_2}{\tilde{c}_2 S_2(s_1)}, \tag{8.12}$$

with $\phi_{(D)}(s_1, s_2)$ the joint LST of the generic input

$$\phi_{(D)}(s_1, s_2) = \mathbb{E}(e^{-s_1\tilde{B}^{(1)} - s_2\tilde{B}^{(2)}}).$$

The stability condition for this system is $\tilde{\rho}_1 < 1$, and for each fixed s_1 with $\mathcal{R}e\, s_1 > 0$, $S_2(s_1)$ is the zero of the equation

$$K_{(D)}(s_1, s_2) = \tilde{c}_1 s_1 + \tilde{c}_2 s_2 - \lambda[1-\phi_{(D)}(s_1, s_2)] = 0, \tag{8.13}$$

that is unique in the positive half of the complex plane.

We have the analogous relation to (8.7):

$$\phi(s_1, s_2) = \phi_{(D)}(s_1, s_2 - s_1 c_2^1/c_2). \tag{8.14}$$

Combining (8.7), (8.12) and (8.14) we obtain

$$\psi_{(C)}(s_1, s_2) = (1 - \tilde{\rho}_1)\frac{(c_1 + c_2^1)s_1}{c_1 s_1 + c_2 s_2 - \lambda[1 - \phi(s_1, s_2)]} \cdot \frac{c_2 S_2(s_1) - c_2 s_2 + s_1 c_2^1}{c_2 S_2(s_1)}.$$

The kernel identity $K_{(D)}(s_1, s_2) = K_{(C)}(s_1, s_2 + s_1 c_2^1/c_2)$ (see Remark 8.3.1) together with (8.13) gives that $S_2(s_1)$ is then the unique zero with positive real part of

$$K_{(C)}(s_1, s_2 + s_1 c_2^1/c_2) = 0.$$

This yields the desired result and the proof is complete.

8.4 The *k*-dimensional model

In this section we consider multiple coupled servers in parallel which receive simultaneous requirements. It is shown that also in this case, the coupled system can be reduced to a decoupled system, upon modifying the input. However for three servers we have to specify in addition how to divide the extra service capacity of an idle server over the other queues. This was trivial for two servers since one can only assist the other during its idle periods. With such specifications in place, the formal idea of the proof for the k dimensional system is analogous to the case $k = 3$, and it relies on the result for two coupled queues. Thus, we can work with $k = 3$, to keep formulae still accessible, without losing generality.

We extend the ordering assumption between the service requirements to

$$\mathbb{P}(B^{(1)}/c_1 \geq B^{(2)}/c_2 \geq B^{(3)}/c_3) = 1.$$

In addition, while server 3 is idle and server 2 is busy, we denote by c_3^1 and c_3^2 the processing rate of server 3 into buffers 1 and 2 respectively. If also server 2 becomes idle, we denote by c_2^1 the processing rate of server 2 into buffer 1 during its idle time, and moreover, server 3 contributes an extra rate $\hat{c}_3^1$ into buffer 1, so that the total contribution from server 3 becomes $c_3^1 + \hat{c}_3^1$, while server 2 is idle.

In addition, we assume that $c_3^1/c_1 \leq c_3^2/c_2$ in order to ensure that the amount of work in queue 1 remains above the amount of work in queue 2, at all times. Because of this assumption, the amount of work in the system is again ordered:

$$\mathbb{P}(V^{(1)}/c_1 \geq V^{(2)}/c_2 \geq V^{(3)}/c_3) = 1.$$

The Lindley type recursion for queue 2 is similar to (8.5). In the sequel, we will derive the recursion for queue 1. Because of the coupling, the idle period of queue 2 plays a role in the dynamics of queue 1 so for this reason (and to keep notations short) we introduce

$$J_n^{(2)} := (c_3^2 + c_2)^{-1}[V_{n-1}^{(2)} + B_{n-1}^{(2)} - c_2 A_{n-1} + c_3^2/c_3\,(V_{n-1}^{(3)} + B_{n-1}^{(3)} - c_3 A_{n-1}) \wedge 0],$$

$$J_n^{(3)} := c_3^{-1}(V_{n-1}^{(3)} + B_{n-1}^{(3)} - c_3 A_{n-1}).$$

$-(J_n^{(2)} \wedge 0)$ is the idle period in queue 2 right before epoch n, and it follows from (8.2) that

$$(c_3^2 + c_2)J_n^{(2)} \vee 0 = V_n^{(2)}. \tag{8.15}$$

Also $-(J_n^{(3)} \wedge 0)$ is the idle period in queue 3. The fact that $-(J_n^{(2)} \wedge 0)$ is an idle period is again a consequence of the ordering, because if server 2 is idle then server 3 must also be idle and hence coupled to queue 2. We can combine the terms above into the identity

$$J_n^{(2)} = (c_3^2 + c_2)^{-1}(V_{n-1}^{(2)} + B_{n-1}^{(2)} - c_2 A_{n-1} + c_3^2 J_n^{(3)} \wedge 0). \tag{8.16}$$

Now we can write the stochastic recursion for the amount of work in queue 1 at arrival epoch $n+1$:

$$V_{n+1}^{(1)} = [V_n^{(1)} + B_n^{(1)} - c_1 A_n + (c_2^1 + \hat{c}_3^1)J_{n+1}^{(2)} \wedge 0 + c_3^1 J_{n+1}^{(3)} \wedge 0] \vee 0. \tag{8.17}$$

In addition, $-[(c_2^1 + \hat{c}_3^1)J_{n+1}^{(2)} \wedge 0]$ is the extra amount of work that server 2 and server 3 are capable of processing while working coupled to server 1 during the idle period of server 2. Similarly, $-(c_3^1 J_{n+1}^{(3)} \wedge 0)$ is the amount of work that can be processed by server 3 while coupled directly to server 1.

Proposition 8.4.1. *Let $(V_n^{(1)}, V_n^{(2)}, V_n^{(3)})$ be the amount of work at epoch n in the coupled system (C), then the following process defined for $n \geq 1$*

$$\begin{aligned}\tilde{V}_n^{(1)} &:= V_n^{(1)} + c_2^*/c_2 V_n^{(2)} + c_3^*/c_3 V_n^{(3)}, \\ \tilde{V}_n^{(2)} &:= V_n^{(2)} + c_3^2/c_3 V_n^{(3)}, \\ \tilde{V}_n^{(3)} &:= V_n^{(3)},\end{aligned} \tag{8.18}$$

with

$$c_2^* := \frac{c_2^1 + \hat{c}_3^1}{c_2 + c_3^2} c_2, \quad c_3^* := \frac{c_2^1 + \hat{c}_3^1}{c_2 + c_3^2} c_3^2 + c_3^1,$$

represents the amount of work at epoch n in a queueing system without coupling between the servers. The service rates are $\tilde{c}_1 := c_1 + c_2^* + c_3^* = c_1 + c_2^1 + c_3^1 + \hat{c}_3^1$ for server 1, $\tilde{c}_2 := c_2 + c_3^2$ for server 2 and $\tilde{c}_3 := c_3$ for server 3. The input in the three queues at epoch n is $\tilde{B}_n^{(1)} := B_n^{(1)} + c_2^*/c_2 B_n^{(2)} + c_3^*/c_3 B_n^{(3)}$, $\tilde{B}_n^{(2)} := B_n^{(2)} + c_3^2/c_3 B_n^{(3)}$ and $\tilde{B}_n^{(3)} := B_n^{(3)}$ respectively.

By a similar coupling argument as in the case $k = 2$, assume that all queues start empty and that the arrival epochs are the same as in the coupled system. Server 1 processes at rate $\tilde{c}_1 = c_1 + c_2^1 + c_3^1 + \hat{c}_3^1$, server 2 at rate $\tilde{c}_2 = c_2 + c_3^2$, and server 3 at the same rate c_3. Queue 3 evolves again unchanged. Using a bit of algebra, $\tilde{V}_n^{(1)}$ from (8.18) can be rewritten as

$$\tilde{V}_n^{(1)} = V_n^{(1)} + \frac{c_2^1 + \hat{c}_3^1}{c_2 + c_3^2}(V_n^{(2)} - c_2\frac{V_n^{(3)}}{c_3}) + (c_2^1 + c_3^1 + \hat{c}_3^1)\frac{V_n^{(3)}}{c_3}.$$

Focus on the ends of the successive busy periods in the three queues. The first one to empty is queue 3. Up to the moment the third queue is empty, partition the work in queue 2 as in Section 8.3. Actually queue 2 together with queue 3 make up precisely the two-dimensional system studied in Section 8.3.

The service rate in queue 1 can be partitioned in the following way: c_1 is dedicated to processing type $V_t^{(1)}$ work, and $(c_2^1 + c_3^1 + \hat{c}_3^1)$ is dedicated to processing the work $(c_2^1+c_3^1+\hat{c}_3^1)V_t^{(3)}/c_3$. Remark that $V_t^{(2)} - c_2V_t^{(3)}/c_3$ is a.s. non-negative due to ordering. The remainder of $\tilde{V}_n^{(1)}$ is waiting in the buffer up to the moment queue 3 empties. At this point in time, work $V_t^{(2)} - c_2V_t^{(3)}/c_3$ is still left in buffer 2, and it is processed at rate $c_2 + c_3^2$, whereas in buffer 1 the amount left equals

$$V_t^{(1)} - c_1\frac{V_t^{(3)}}{c_3} + \frac{c_2^1 + \hat{c}_3^1}{c_2 + c_3^2}(V_t^{(2)} - c_2\frac{V_t^{(3)}}{c_3}).$$

From this point on, partition server 1 capacity in the following way: dedicate rate $(c_1+c_3^1)$ to process work $V_t^{(1)}-c_1V_t^{(3)}/c_3$, and rate $(c_2^1+\hat{c}_3^1)$ to process work $\frac{c_2^1+\hat{c}_3^1}{c_2+c_3^2}(V_t^{(2)} - c_2V_t^{(3)}/c_3)$. In this way, at the moment queue 2 empties, there is still work left in queue 1, that is

$$V_t^{(1)} - c_1\frac{V_t^{(3)}}{c_3} - \frac{c_1 + c_3^1}{c_2 + c_3^2}(V_t^{(2)} - c_2\frac{V_t^{(3)}}{c_3}),$$

and this is processed at speed $c_1 + c_2^1 + \hat{c}_3^1 + c_3^1$ as long as there is no new arrival.

Proof of Proposition 8.4.1. The idea is to add successively the terms $(c_2^1 + c_3^2)J_{n+1}^{(2)} \vee 0$ and $c_3^* J_{n+1}^{(3)} \vee 0 = c_3^*/c_3 V_{n+1}^{(3)}$ to the recursion in (8.17) in order to compensate for the minima with 0 in the first bracket.

First add the term $(c_2^1 + \hat{c}_3^1)J_{n+1}^{(2)} \vee 0 = c_2^*/c_2 V_{n+1}^{(2)}$ to both sides of (8.17). Making use of (8.15) for the left-hand side and of (8.16) for the right-hand side, (8.17) becomes after rearranging terms

$$\begin{aligned} V_{n+1}^{(1)} + \frac{c_2^*}{c_2}V_{n+1}^{(2)} = \\ &\left[V_n^{(1)} + B_n^{(1)} - c_1A_n + \frac{c_2^*}{c_2}(V_n^{(2)} + B_n^{(2)} - c_2A_n) + \frac{c_3^*}{c_3}(V_n^{(3)} + B_n^{(3)} - c_3A_n) \wedge 0\right] \\ \vee &\left[\frac{c_2^*}{c_2}(V_n^{(2)} + B_n^{(2)} - c_2A_n) + \frac{c_3^* - c_3^1}{c_3}(V_n^{(3)} + B_n^{(3)} - c_3A_n) \vee 0\right] \vee 0. \end{aligned} \tag{8.19}$$

Now add the term $c_3^* J_{n+1}^{(3)} \vee 0$, which is the same as $c_3^*/c_3 V_{n+1}^{(3)}$, by Lindley's recursion. After regrouping terms, (8.19) becomes

$$
\begin{aligned}
V_{n+1}^{(1)}&+\frac{c_2^*}{c_2}V_{n+1}^{(2)}+\frac{c_3^*}{c_3}V_{n+1}^{(3)}= \\
&\left[V_n^{(1)}+B_n^{(1)}-c_1A_n+\frac{c_2^*}{c_2}(V_n^{(2)}+B_n^{(2)}-c_2A_n)+\frac{c_3^*}{c_3}(V_n^{(3)}+B_n^{(3)}-c_3A_n)\right] \\
&\vee\left[\frac{c_2^*}{c_2}(V_n^{(2)}+B_n^{(2)}-c_2A_n)+\frac{c_3^*-c_3^1}{c_3}(V_n^{(3)}+B_n^{(3)}-c_3A_n)\right. \\
&\left.+\frac{c_3^1}{c_3}(V_n^{(3)}+B_n^{(3)}-c_3A_n)\vee 0\right]\vee\frac{c_3^*}{c_3}(V_n^{(3)}+B_n^{(3)}-c_3A_n)\vee 0.
\end{aligned}
\tag{8.20}
$$

We show now that the two middle terms that appear in the maximum sequence above are always dominated by either one of the extremal terms. There are three cases to be considered. If

$$V_n^{(3)}+B_n^{(3)}-c_3A_n>0$$

then by the ordering assumption, $V_n^{(2)}+B_n^{(2)}-c_2A_n$ and $V_n^{(1)}+B_n^{(1)}-c_1A_n$ are also positive, and it is easy to see that the first term on the right-hand side is the largest one.

The alternative is

$$V_n^{(3)}+B_n^{(3)}-c_3A_n<0,$$

and there are two sub-cases to be considered: if

$$-\frac{c_3^1}{c_3}(V_n^{(3)}+B_n^{(3)}-c_3A_n)<V_n^{(1)}+B_n^{(1)}-c_1A_n,$$

then the first term dominates the second one on the right-hand side of (8.20) and the third term is negative by assumption, so one can ignore the intermediate terms again.

The other subcase is when

$$-\frac{c_3^1}{c_3}(V_n^{(3)}+B_n^{(3)}-c_3A_n)>V_n^{(1)}+B_n^{(1)}-c_1A_n,$$

thus the first term on the right-hand side of (8.20) is smaller than the second term. We show that this second term is now negative: the above inequality means that queue 3 is able to empty queue 1 at epoch $n+1$ without the help of queue 2, so as a consequence of the ordering assumptions, then queue 3 will also empty queue 2 at epoch $n+1$: $V_{n+1}^{(2)}=0$, and it is easy to see using the definitions of c_2^* and c_3^* and the recursion (8.5) for the two coupled queues, that this latter fact is equivalent to the second term being negative.

In conclusion the intermediate terms on the right-hand side of (8.20) can be ignored, and this gives after rearranging terms

$$\tilde{V}_{n+1}^{(1)}=\{\tilde{V}_n^{(1)}+\tilde{B}_n^{(1)}-\tilde{c}_1A_n\}\vee 0.$$

By ignoring queue 1, it follows at once from Proposition 8.2.1 that $\tilde{V}_n^{(2)}$ satisfies the corresponding Lindley recursion, and since queue 3 evolves unchanged, the proof is complete.

As in Section 8.2, if we denote by $\psi_{(C)}$ and $\psi_{(D)}$ the transforms of the amount of work in the coupled and respectively, in the related decoupled system, then the relation analogous to (8.7) is

$$\psi_{(C)}(s_1, s_2, s_3) = \psi_{(D)}(s_1, s_2 - \frac{c_2^*}{c_2} s_1, s_3 - \frac{c_3^2}{c_3} s_2 - \frac{c_3^*}{c_3} s_1),$$

and it is easy to see that the input in the decoupled system remains ordered, which means, in principle, one can determine $\psi_{(C)}$ from the relation above using the available expression for $\psi_{(D)}$ obtained in Chapter 5 (Section 5.2, Formulae (5.6) and (5.10)).

8.5 Conclusions and final remarks

We have pointed out a relation between a coupled processor model and two parallel queues without coupling, under the assumption of ordered input. This relation was further used to derive the joint Laplace-Stieltjes transform of the amount of work in equilibrium.

We have also derived the relation explicitly for the case of three coupled processors. This can be used in principle to determine the joint transform of the equilibrium amount of work by using the expression derived in Proposition 5.4.1 for the queueing system with the processors not coupled.

Relation with two coupled queues in tandem There is also a relation between the coupled processor model with two queues described above and two tandem queues which are coupled. The relation is similar to the one between the systems without coupling, which was pointed out for Lévy input in Kella [73] (see also Section 5.3).

Consider two queues working in tandem, and having a compound Poisson arrival process which at epoch t_n brings work $B_n^{(2)}$ in the first queue and work $B_n^{(1)} - B_n^{(2)}$ in the second queue. Both queues work at unit speed and the output from queue 1 flows into queue 2. The queues are also coupled, which means that as soon as queue 1 is empty it switches its capacity to help queue 2, so that the processing rate of queue 2 doubles during the idle periods of queue 1. The assumption of equal rates seems to be necessary to have this model related to the coupled processor model studied in the previous sections.

Since the amount of work does not depend on the server's policy, one can assume that the amount of fluid coming from server 1 is processed with priority over the accumulated exogenous input into buffer 2. Then the point of the assumption of equal rates is that server 2 finishes processing the fluid at the same instant that server 1 becomes idle. For this reason, the amount of work in queue 1 together with the total amount of work in the tandem system taken as a whole is the same as the workload vector in the coupled system (C).

As a final remark, we mention that the number of jobs waiting to be served in a tandem queueing system with coupled processors but without simultaneous arrivals has been studied in Resing and Örmeci [95] related to data transfer in cable networks. The focus is on the number of jobs at two stations which receive a Poisson input at only the first station (no exogenous arrivals) and the server distributes its capacity among the queues while both are non-empty and it switches full capacity to one queue when the other has no jobs to be served, hence the system behaves as a coupled processor model. The functional equation is solved by relating it to a Riemann-Hilbert boundary-value problem, and in van Leeuwaarden and Resing [81] it is also pointed out how to derive performance measures as the mean delay at one station, based on its solutions.

Bibliography

[1] J. Abate and W. Whitt (1992) Numerical inversion of probability generating functions. *O. R. Letters* **12**, 245–251.

[2] I.J.B.F. Adan and V. Kulkarni (2003) Single-server queue with Markov-dependent inter-arrival and service times. *Queueing Systems* **45**, 113–134.

[3] I.J.B.F. Adan, J.S.H. van Leeuwaarden and E.M.M. Winands (2006) On the application of Rouché's theorem in queueing theory. *O. R. Letters* **34**, 355–360.

[4] H. Albrecher and O.J. Boxma (2004) A ruin model with dependence between claim sizes and claim intervals. *Insurance: Mathematics and Economics* **35**, 245–254.

[5] H. Albrecher and O.J. Boxma (2005) On the discounted penalty function in a Markov-dependent risk model. *Insurance: Mathematics and Economics* **37**, 650–672.

[6] H. Albrecher, O.J. Boxma and J. Ivanovs (2014) On simple ruin expressions in dependent Sparre Andersen risk models. *J. Appl. Probab.* **51**, 293–296.

[7] H. Albrecher and J. Kantor (2002) Simulation of ruin probabilities for risk processes of Markovian type. *Monte Carlo Methods and Applications* **8**, 111–127.

[8] H. Albrecher and J.L. Teugels (2006) Exponential behavior in the presence of dependence in risk theory. *J. Appl. Probab.* **43**, 257–273.

[9] R.S. Ambagaspitiya (2009) Ultimate ruin probability in the Sparre Andersen model with dependent claim sizes and claim occurrence times. *Insurance: Mathematics and Economics* **44**, 464–472.

[10] E. Sparre Andersen (1953) On sums of symmetrically dependent random variables. *Skandinavisk Aktuarietidskrift* **36**, 123–138.

[11] E. Sparre Andersen (1957) On the collective theory of risk in case of contagion between the claims. *Transactions XVIth International Congress of Actuaries, New York, II* 219–229.

[12] E. Sparre Andersen (1953) On the fluctuations of sums of random variables. *Math. Scand.* **1**, 263–285.

[13] E. Sparre Andersen (1954) On the fluctuations of sums of random variables II. *Math. Scand.* **2**, 195–223.

[14] S. Asmussen, A. Frey, T. Rolski and V. Schmidt (1995) Does Markov-modulation increase the risk? *ASTIN Bull* **25**, 49–66.

[15] S. Asmussen (1995) Stationary distributions via first passage times. *Advances in Queueing: Models, Methods and Problems* (J. Dshalalow, ed.), CRC Press, 79–102.

[16] S. Asmussen (2003) *Applied Probability and Queues. Second edition,* Springer-Verlag, New York.

[17] S. Asmussen and H. Albrecher (2010) *Ruin Probabilities.* World Scientific Publ. Cy., Singapore.

[18] D. Assaf, N. A. Langberg, T. Savits and M. Shaked (1984) Multivariate phase-type distributions. *Operations Research* **32**, 688–701.

[19] F. Avram, Z. Palmowski and M. Pistorius (2008) A two-dimensional ruin problem on the positive quadrant. *Insurance: Mathematics and Economics* **42**, 227–234.

[20] F. Avram, Z. Palmowski and M. Pistorius (2008) Exit problem of a two-dimensional risk process from the quadrant: Exact and asymptotic results. *Annals of Applied Probability* **18**, 2421–2449.

[21] F. Baccelli (1985) Two parallel queues created by arrivals with two demands: The M/G/2 symmetrical case. *Technical Report, INRIA-Rocquencourt.*

[22] F. Baccelli, A.M. Makowski and A. Shwartz (1989) The fork-join queue and related systems with synchronization constraints: Stochastic ordering and computable bounds. *Adv. Appl. Probab.* **21**, 629–660.

[23] A.L. Badescu, E.C.K. Cheung and L. Rabehasaina (2011) A two-dimensional risk model with proportional reinsurance. *J. Appl. Probab.* **48**, 749–765.

[24] E.S. Badila, O.J. Boxma and J.A.C. Resing (2013) Queues and risk processes with dependencies. *Stochastic Models* **30**, 390–419.

[25] E.S. Badila, O.J. Boxma, J.A.C. Resing and E.M.M. Winands (2014) Queues and risk processes with simultaneous arrivals. *Adv. Appl. Probab.* **46**, 812–831.

[26] E.S. Badila, O.J. Boxma and J.A.C. Resing (2014) Two parallel insurance lines with simultaneous arrivals and risks correlated with inter-arrival times. *Insurance: Mathematics and Econonomics* **61**, 48–61.

[27] E.S. Badila and J.A.C. Resing (2015) A coupled processor model with simultaneous arrivals and ordered service requirements. *Accepted for publication in Queueing Systems.*

[28] G. Baxter and M.D. Donsker (1957) On the distribution of the supremum functional for processes with stationary independent increments. *Trans. Am. Math. Soc.* **85**, 73–87.

[29] M. Bladt and B.F. Nielsen (2010) Multivariate matrix-exponential distributions. *Stochastic Models* **26**, 1–26.

[30] M. Bladt and B.F. Nielsen (2010) On the construction of bivariate exponential distributions with an arbitrary correlation coefficient. *Stochastic Models* **26**, 295–308.

[31] M. Bladt, L.J.R. Esparza and B.F. Nielsen (2013) Bilateral matrix-exponential distributions. In: *Matrix-Analytic Methods in Stochastic Models*, Springer Proc. Math. Stat. **27**, 41–56, Springer, New York.

[32] K.A. Borovkov and D.C. Dickson (2008) On the ruin time distribution for a Sparre Andersen process with exponential claim sizes. *Insurance: Mathematics and Economics*, **42**, 1104–1108.

[33] S.C. Borst, O.J. Boxma and M.B. Combé (1993) An M/G/1 queue with dependence between interarrival and service times. *Stochastic Models* **9**, 341–371.

[34] M. Boudreault, H. Cossette, D. Landriault and E. Marceau (2006) On a risk model with dependence between interclaim arrivals and claim sizes. *Scand. Actuar. J.* **5**, 265–285.

[35] O.J. Boxma and D. Perry (2001) A queueing model with dependence between service and interarrival times. *European J. Oper. Res.* **128**, 611–624.

[36] J. Cai and H. Li (2005) Multivariate risk model of phase type. *Insurance: Mathematics and Economics* **36**, 137–152.

[37] W.-S. Chan, H. Yang and L. Zhang (2003) Some results on ruin probabilities in a two-dimensional risk model. *Insurance: Mathematics and Economics* **32**, 345–358.

[38] J.W. Cohen (1975) The Wiener-Hopf technique in applied probability. In: *Perspectives in Probability and Statistics, ed. J.Gani*, 145–156, Academic Press, London.

[39] J.W. Cohen (1976) *On Regenerative Processes in Queueing Theory.* Springer-Verlag, New York.

[40] J.W. Cohen (1982) *The Single Server Queue.* North-Holland, Amsterdam.

[41] J.W. Cohen and O. J. Boxma (1983) *Boundary Value Problems in Queueing System Analysis.* North-Holland, Amsterdam.

[42] J.W. Cohen (1988) Boundary value problems in queueing theory. *Queueing Systems* **3**, 97–128.

[43] J.W. Cohen (1992) *Analysis of Random Walks.* IOS Press, Amsterdam, the Netherlands.

[44] N.R. Coleff and M.E. Herrera (1978) *Les courants résiduels associés à une forme méromorphe.* Springer-Verlag, Berlin.

[45] M.B. Combé and O.J. Boxma (1998) BMAP modelling of a correlated queue. In: *Network Performance Modeling and Simulation,* eds. J. Walrand, K. Bagchi and G.W. Zobrist, 177–196.

[46] B.W. Conolly (1960) The busy period in relation to the single-server queueing system with general independent arrivals and erlangian service-time. *Journal of the Royal Statistical Society. Series B* **22**, 89–96.

[47] B.W. Conolly (1974) The generalised state-dependent queue: the busy period erlangian. *J. Appl. Probab.* **11**, 618–623.

[48] C. Constantinescu, D. Kortschak and V. Maume-Deschamps (2013) Ruin probabilities in models with a Markov chain dependence structure. *Scand. Actuar. J.* 453–476.

[49] A. Dembo and O. Zeitouni (1998) *Large Deviations Techniques and Applications.* Springer-Verlag, New York.

[50] D.C.M. Dickson and H.R. Waters (2002) The distribution of the time to ruin in the classical risk model. *ASTIN Bull.* **32**, 299–313.

[51] G. Doetsch (1974) *Introduction to the Theory and Application of the Laplace Transformation.* Springer-Verlag, Berlin.

[52] P. Embrechts and M. Hofert (2013) A note on generalized inverses. *Math. Meth. Oper. Res.* **77**, 423–432.

[53] P. Embrechts, C. Klüppelberg and T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

[54] P. Embrechts and N. Veraverbeke (1982) Estimates for the probability of ruin with special emphasis on the probability of large claims. *Insurance: Mathematics and Economics* **1**, 55–72.

[55] G. Fayolle and R. Iasnogorodski (1979) Two coupled processors: The reduction to a Riemann-Hilbert problem. *Z. Wahrsch. Verw. Gebiete* **47**, 325–351.

[56] G. Fayolle, R. Iasnogorodski and V. A. Malyshev (1999) *Random Walks in the Quarter Plane.* Springer, Berlin.

[57] W. Feller (1971) *An Introduction to Probability Theory and its Applications. Vol. 2* (2nd ed.) John Wiley & Sons Inc., New York.

[58] P.D. Finch (1961) On the busy period in the queueing system GI/G/1. *J. Austr. Math. Soc.* **2**, 217–228.

[59] L. Flatto and S. Hahn (1984) Two parallel queues created by arrivals with two demands. *SIAM Journal of Applied Mathematics* **44**, 1041–1053.

[60] S. Foss, D. Korshunov and S. Zachary (2011) *An Introduction to Heavy-tailed and Subexponential Distributions.* Springer, New York.

[61] S. Foss, Z. Palmowski and S. Zachary (2005) The probability of exceeding a high boundary on a random time interval for a heavy-tailed random walk. *Annals of Applied Probability* **15**, 1936–1957.

[62] S. Foss, T. Rolski and S. Zachary (2007) Reinsurance and ruin problem; asymptotics in case of heavy-tailed claims. *Unpublished manuscript.*

[63] P. Franken, D. König, U. Arndt and V. Schmidt (1983) *Queues and Point Processes.* John Wiley & Sons, Chichester.

[64] E. Frostig (2004). Upper bounds on the expected time to ruin and on the expected recovery time. *Adv. Appl. Probab.* **36**, 377–397.

[65] F.D. Gakhov (1990) *Boundary Value Problems.* Pergamon Press, Oxford.

[66] F.D. Gakhov, E.I. Zverovich and S.G. Samko (1973) Increment of the argument, logarithmic residue and a generalized principle of the argument. *Dokl. Akad. Nauk SSSR* **213**, 1233–1236 (in Russian) (translated in *Soviet Math. Dokl.* **14**, 1856–1860).

[67] L. Gong, A.L. Badescu and E.C.K. Cheung (2012) Recursive methods for a multi-dimensional risk process with common shocks. *Insurance: Mathematics and Economics* **50**, 109–120.

[68] J. Grandell (1992) *Aspects of Risk Theory.* Springer-Verlag, New York.

[69] E. Hewitt (1953) Remarks on the inversion of Fourier-Stieltjes transforms. *Annals of Mathematics* **57**, 458–474.

[70] P. den Iseger (2006) Numerical transform inversion using Gaussian quadrature. *Probab. Eng. Inform. Sc.* **20**, 1–44.

[71] J. Ivanovs and O.J. Boxma (2015) A bivariate risk model with mutual deficit coverage. *arXiv: http://arxiv.org/abs/1501.02927.*

[72] O. Kallenberg (2002) *Foundations of Modern Probability. Second edition.* Probability and its Applications. Springer-Verlag, New York.

[73] O. Kella (1993) Parallel and tandem fluid networks with dependent Lévy inputs. *Annals of Applied Probability* **3,** 682–695.

[74] J.F.C. Kingman (1962) The use of Spitzer's identity in the investigation of the busy period and other quantities in the queue GI/G/1. Austral. J. Math. **2**, 345–356.

[75] J.F.C. Kingman (1966) On the algebra of queues. *J. Appl. Probab.* **3**, 285–386.

[76] S. J. de Klein (1988) *Fredholm Integral Equations in Queueing Analysis.* PhD Thesis, University of Utrecht.

[77] V. Klimenok (2001) On the modification of Rouche's theorem for the queueing theory problems. *Queueing Systems* **38**, 431–434.

[78] S. Kotz, N. Balakrishnan and N.L. Johnson (2000) *Continuous Multivariate Distributions*, Vol. 1. John Wiley & Sons Inc, New York.

[79] V.G. Kulkarni (1989) A new class of multivariate phase type distributions. *Operations Research* **37**, 151–158.

[80] I.K.M. Kwan and H. Yang (2007) Ruin probability in a threshold insurance risk model. *Belg. Actuar. Bull.* **7**, 41–49.

[81] J.S.H. van Leeuwaarden and J.A.C. Resing (2005) A tandem queue with coupled processors: Computational issues. *Queueing Systems* **51**, 29–52.

[82] D. Lindley (1952) The theory of a queue with a single server. *Proc. Cambr. Philos. Soc.* **B21**, 22–23.

[83] R. M. Loynes (1962) The stability of a queue with non-independent inter-arrival and service times. *Proc. Cambr. Philos. Soc.* **58**, 497–520.

[84] A.H. Löpker and D. Perry (2010). The idle period of the finite G/M/1 queue with an interpretation in risk theory. *Queueing Systems* **64**, 395–407.

[85] D.M. Lucantoni (1991) New results on the single-server queue with a batch Markovian arrival process. *Stochastic Models* **7**, 1–46.

[86] P.M.D. Lieshout and M. Mandjes (2007) Tandem Brownian queues. *Math. Meth. Oper. Res.* **66**, 275–298.

[87] N. I. Mushkelishvili (1953) *Singular Integral Equations.* P. Noordhoff, Groningen–Holland.

[88] R. Nelson and A. N. Tantawi (1987) *Approximating task response times in fork/join queues.* Report IBM T.J. Watson Research Center.

[89] R. Nelson and A. N. Tantawi (1988) Approximate analysis of fork/join synchronization in parallel queues. *IEEE Transactions on Computers* **37,** 739–743.

[90] M.F. Neuts (1991) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* Dover Publications, New York.

[91] Z. Palmowski and M. Pistorius (2008) The probability of exceeding a piecewise deterministic barrier by the heavy-tailed renewal compound process. *http://arxiv.org/abs/0805.1631.*

[92] F. Pollaczek (1952) Fonctions caracteristiques de certaines répartitions définies au moyen de la notion a d'ordre. Application à la théorie des attentes. *C. R. Acad. Sci. Paris* **234**, 2334–2336.

[93] N. U. Prabhu (1961) On the ruin problem of collective risk theory. *Ann. Math. Stat.* **32**, 757–764.

[94] N.U. Prabhu (1980) *Stochastic Storage Processes. Queues, Insurance Risk, and Dams.* Springer Verlag, New York.

[95] J.A.C. Resing and L. Örmeci (2003) A tandem queueing model with coupled processors. *O.R. Letters* **31**, 383–389.

[96] T. Rolski, H. Schmidli, V. Schmidt and J.L. Teugels (1999) *Stochastic Processes for Insurance and Finance.* John Wiley and Sons, Inc., New York.

[97] D. Siegmund (1976) The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov Processes. *The Annals of Probability* **4**, 914–924.

[98] F. Spitzer (1956) A combinatorial lemma and its application to probability theory. *Trans. Am. Math. Soc.* **82**, 323–339.

[99] F. Spitzer (1957) The Wiener-Hopf equation whose kernel is a probability density. *Duke Math. J.* **24**, 327–343.

[100] D. Stoyan (1983) *Comparison Methods for Queues and Other Stochastic Models.* John Wiley & Sons, New York.

[101] B. Sundt (1999) On multivariate Panjer recursions. *ASTIN Bull.* **29**, 29–45.

[102] E.C. Titchmarsh (1939) *The Theory of Functions.* Oxford University Press, 2nd edition, Oxford.

[103] J.G. Wendel (1960) Order statistics of partial sums. *Ann. Math. Stat.* **31**, 1034–1044.

[104] R.W. Wolff (1982) Poisson arrivals see time averages. *Operations Research* **30**, 223–231.

[105] P. E. Wright (1992) Two parallel processors with coupled inputs. *Adv. Appl. Probab.* **24**, 986–1007.

Summary

Queues and Risk Models

The research in this thesis is at the boundary between Queueing Theory and Insurance Risk. Queueing Theory appears for example from the need to analyze the operations in an electronic device; it is also very much related to the mathematics of networks. Networks of queues appear in many real life applications such as traffic networks or the internet. On the other hand, Insurance and Risk theory applies to almost any financial reserve, be it the fluctuating capital of an insurance company or the evolving value of an asset. Related areas are also Credit Risk and derivatives. These topics are studied in the thesis in a unified way, using the probabilistic theory of random walks and the related fluctuation theory of stochastic processes, and combining various probabilistic and analytic methods.

A recurrent theme of the thesis are the duality relations between queues and risk reserve processes. The second chapter is dedicated to the question whether a so-called 'reflected' process (like the amount of work that has to be processed in a queueing system) can be related to a dual process with absorbtion (as defined by a reserve process stopped upon becoming ruined). This question is particularly interesting in several dimensions, for queues with many servers, respectively for a group of insurance companies that share risks. Such duality relations have quite some applications: First of all, it might help us to use results from Queueing Theory to obtain results in Insurance Risk, or vice versa. Furthermore, duality relations can be used to study the problem of stability for queueing systems, they are an essential ingredient in the study of large deviations and the related asymptotic problems, and they can also provide efficient simulation methods (Chapter 2).

From a different mathematical perspective, the range of problems studied can be grouped into two overlapping categories. Firstly, the fluctuation theory for one-dimensional systems is revisited, but this time by allowing the input processes that feed such systems to have various correlations (Chapters 3 and 4). For example, we allow correlations between inter-arrivals and service times in a single server queue, and correlations between inter-arrival times of claims and claim sizes in a reserve process consisting of a single insurance line. This can severely complicate any kind of exact analysis. Nevertheless, it is shown that key performance measures of queueing systems with particular correlations can be derived in terms of Laplace transforms. These correspond, by duality, to the performance measures of the related risk reserve. We also reveal the impact that the various kinds of correlations have on the amount of time a customer has to wait in the queue (and also on the survival function associated to the reserve process). This insight is given in terms of convex ordering and is relevant for the modelling, design and optimization of such systems.

Secondly, the analysis of fluctuations of multidimensional queueing systems and the related risk reserve processes is another contribution. The fact that we managed to give explicit formulae for a multidimensional system is a breakthrough; in particular, in the Insurance Risk literature there were hardly any known results. Key performance measures can be derived using probabilistic interpretations of the quantities that arise from the complex analysis, even more so, one can use the probabilistic insight to guide the analysis (Chapters 5, 6, and 8). In addition, multidimensional systems with interacting components are analyzed in Chapters 2 and 8. In the queueing literature, a special example is the so-called coupled processor model.

The asymptotic behaviour of the steady state waiting time of customers in a queueing network is important for design and optimization purposes. The dual problem is to understand the rate at which the ruin probability of an insurance company decreases when the initial capital grows indefinitely. This is particularly important for tuning the capital requirements of the company against the severity of requested claims. In Chapter 7, we give results that specify these asymptotic regimes for the so-called reinsurance contracts under very general assumptions on the distributions of the requested claims (subexponentiality). The rate depends on the frequency of arrivals, but especially on the distribution of the requested claims.

Curriculum Vitae

Emil Şerban Bădilă was born in Târgu-Mureş on April 28, 1985. In 2004, he finished secondary education at Al. Papiu Ilarian high-school in Târgu-Mureş. Afterwards, he went on to study Mathematics at the University of Bucharest. After graduation, he was accepted on a scholarship for a Master program at Utrecht University, where he graduated 'cum laude' in 2010.

Having obtained the MSc degree, in 2011 he started a research project at Eindhoven University of Technology, under the supervision of Onno Boxma and Jacques Resing. The topic was at the boundary between Queueing Theory and Insurance Risk and the current thesis is a compilation of most of the results obtained during the past four years. The research has led to the publication of several papers in international journals. In addition, he has paid a couple of visits to the group of Hansjoerg Albrecher at the University of Lausanne, and at the Mathematical Institute of the University of Wrocław, for a collaboration with Zbigniew Palmowski and Tomasz Rolski.

He was also one of the organizers of the 'Queues and Risk Workshop', in March 2013. The research project is concluded with a thesis entitled 'Queues and Risk Models' which is to be defended on June 22, 2015.