Graz University of Technology

Analysis & Number Theory TU Graz

Technische Universität Graz

Progress in Risk Theory and Dependence Modeling

with Applications in Finance and Insurance

Doctoral Thesis

by

Michael Preischl

to achieve the university degree

Doktor der technischen Wissenschaften

Supervisor: Prof. Dr. Robert F. Tichy
Second Supervisor: Prof. Dr. Stefan Thonhauser

Graz, May 2019

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date	Signature

Ein großes Dankeschön

Die Promotion mag nur eine Etappe im lebenslangen Weg des Lernens sein, für mich ist sie in jedem Fall Anlass, den Menschen zu danken, die diesen Weg erst möglich, leichter, spannender oder einfach nur angenehmer gemacht haben.
Zunächst möchte ich Markus Meiringer danken, dass er mir die Augen für die Schönheit der Mathematik geöffnet hat. Von den vielen engagierten Lehrenden, die meine Ausbildung an der TUM geprägt haben, bin ich besonders Nina Gantert dankbar für die festen Grundlagen und die tiefgehenden Einblicke in viele Bereiche der Wahrscheinlichkeitstheorie.
Seit ich mein Doktorat hier an der TU Graz begonnen habe, habe ich beinahe jeden Tag Grund, Hermine Panzenböck, Irene Pfeifer-Wilfinger und Jochen Resch zu danken, für die immer schnelle und kompetente Hilfe bei all den kleinen und großen Fragen zu administrativen und technischen Abläufen. Darüber hinaus möchte ich ihnen und allen anderen Kollegen des Instituts für Analysis und Zahlentheorie Danke sagen, für die herzliche Aufnahme, die gegenseitige Unterstützung, die netten Momente beim Kaffee oder den Schöcklwanderungen und ganz allgemein für ein Miteinander in dem man sich wohl fühlt. In diesem Sinne auch vielen Dank an den Institutsvorstand Peter Grabner, der dieses Arbeitsklima gut gedeihen lässt und bei der Verteilung von Lehraufgaben immer darauf geachtet hat, dass Zeit für die Forschung bleibt.
Danke auch an Hansjörg Albrecher, dessen genaue Begutachtung diese Dissertation an vielen Stellen verbessert hat. Meinem Doktorvater Robert Tichy bin ich sehr dankbar für seine fachliche Anleitung, seine wertvollen Anregungen und nicht zuletzt für seine pragmatische Unterstützung, die mir sehr vieles erleichtert hat. Das Vertrauen, mich mit allen Problemen an ihn wenden zu können war mir immer ein großer Rückhalt. Zu besonderem Dank bin ich Stefan Thonhauser verpflichtet, der mir von Anfang an Mentor und Freund war. Danke Stefan, für die unzähligen Stunden gemeinsamen Rechnens, die vielen wertvollen Tipps, die Begeisterung für unsere Projekte und die unendliche Geduld, die du immer mit mir hattest.
Schließlich gilt mein Dank auch meiner Familie und allen Freunden, die diesen Weg begleitet haben. Danke an meine Frau Michaela, dass du mit mir in ein fremdes Land gezogen bist und diese Zeit so viel schöner gemacht hast. Danke an meine Eltern und meinen Bruder, eure Unterstützung trägt mich schon mein Leben lang. Ihr seid nicht nur das Fundament, auf dem ich alles aufbauen konnte, ihr habt selbst mitgebaut, seid immer Hilfe und Antrieb.

Zusammenfassung

Die vorliegende Arbeit gliedert sich in zwei Teile. Im ersten Abschnitt werden von der Versicherungsmathematik motivierte Risikoprozesse behandelt und es werden neue Methoden und Ansätze für klassische Fragestellungen aus diesem Fachbereich präsentiert. Dabei liegt der Fokus auf speziellen, abdiskontierten Straffunktionen, in der Literatur auch als Gerber-Shiu Funktionen bekannt. So wird zunächst eine effiziente und allgemein anwendbare, numerische Approximation, basierend auf Quasi-Monte Carlo Integration, vorgestellt um Gerber-Shiu Funktionen zu berechnen. Im Anschluss wird das Problem optimaler, dynamischer Rückversicherung untersucht, wobei Gerber-Shiu Funktionen minimiert werden sollen. Dies wird mittels optimaler stochastischer Kontrolle auf sehr allgemeine Art und Weise erreicht, sodass dieses Verfahren für verschiedene Rückversicherungsarten, Prämienkalkulationsprinzipien und Schadenshöhenverteilungen gültig ist.
Im zweiten Abschnitt werden Abhängigkeitsstrukturen zwischen den einzelnen Komponenten höherdimensionaler Zufallsverktoren untersucht und über die Modellierung bestimmter Abhängigkeitskonzepte diskutiert. Als nützliches Werkzeug werden dabei insbesondere Copulas hervorgehoben und es wir kurz auf deren Bedeutung in Forschung und Praxis eingegangen. Motiviert durch zahlentheoretische Überlegungen wird dann ein Problem des optimalen Transports mit einem diskreten, linearen Optimierungsproblem identifiziert. Anschließend wird ein Lösungsverfahren für Probleme in mehr als zwei Dimensionen vorgestellt, diskutiert und im Vergleich zu vorhandenen Verfahren evaluiert. Zur Illustration werden bestehende, theoretische Resultate über Extremwerte von Abhängigkeitsmaßen numerisch bestätigt und Schranken angegeben für Fälle, in denen keine exakten Lösungen bekannt sind. Der letzte Teil liefert schließlich einen genaueren Blick auf strukturelle Aspekte von Abhängigkeit und verknüpft diese Thematik mit der Theorie von Majorisierungen zwischen Vektoren. Das führt zu einer Verallgemeinerung eines in der Praxis angewendeten Verfahrens um Extremfälle negativer Abhängigkeit in höheren Dimensionen zu berechnen.

Summary

This thesis is organized in two parts. The first section deals with risk theory, a common subject in insurance mathematics, and presents new methods and approaches to this topic. In this, the main focus will be on particular discounted penalty functions, also known as Gerber-Shiu functions in literature. At first, an efficient and broadly applicable numerical approximation based on Quasi-Monte Carlo Integration, is presented to calculate Gerber-Shiu functions. Then the problem of optimal dynamic reinsurance with the objective to minimize a Gerber-Shiu function is analyzed. The presented approach uses optimal stochastic control theory and is general enough to allow for a wide variety of models, including different reinsurance types, premium principles and claim-height distributions.
The second part focuses on dependence structures between the components of a multidimensional random vector and discusses the modeling of certain dependence concepts. As a useful tool in this area, copulas are mentioned and their relevance in research and industry is briefly outlined. Motivated by a number theoretical result, a problem from the theory of optimal transport is identified with a discrete, linear optimization problem. In the following, a numerical calculation method for more than two dimensions is presented, discussed and evaluated against existing methods. The proposed approach is then illustrated by numerically confirming analytic results on extremal bounds of dependence measures and also by stating bounds in cases, where no explicit solutions are known. Lastly, a more rigorous examination of dependence structures is attempted and the subject is linked to majorization theory. This leads to a generalization of an algorithm that is commonly used in practice.

Contents

Part I.

Risk Theory

1. Introduction to Risk Theory

This first part of the thesis covers various problems in the framework of risk theory and more specifically from the field of risk reserve processes and ruin probabilities. Simply speaking, this theory analyzes the probability and consequences of a ruin event. Often such an event is interpreted in the context of monetary reserves held by an insurance company. Given that this field of research naturally aims for applicability, both theoretical foundations and concrete computational procedures are called for in many problems. Catering to these needs, Chapter 2 applies numerical methods based on Quasi-Monte-Carlo (QMC) techniques to risk reserve processes, whereas Chapter 3 develops theory to find an optimal reinsurance strategy with respect to a certain target functional. The results in both of these chapters are novel in their respective form and are published as indicated in the corresponding chapters. In order to present these results in detail, it will be necessary to establish some background, notions and definitions.

1.1. A Short Note on History and Motivation

Although many models and concepts from risk theory have made their way into finance, this discipline of applied mathematics originated from and is usually considered as insurance mathematics. Insurance mathematics or actuarial science as it is also called has been practiced, at least in a basic sense, for several hundreds of years, however it was mainly in the first half of the 20th century that this field of science emerged as the rigorous mathematical discipline it is today. Much of this boom in the early 1900s was fueled by a very active group of Swedish mathematicians around Filip Lundberg and Harald Cramér, to name just the probably most notable pioneers in actuarial mathematics.

The basic idea of actuarial science is to provide the mathematical tools an insurance company needs to operate its business. This mainly includes quantifying and modeling the risks that are to be insured and, of course, ways to secure these risks, for example by calculating premiums or necessary reserve heights. In fact, the importance of doing this risk management in a proper way has by now been recognized by defining frameworks whose implementation is compulsory by law. In the European Union, this is given by the so called solvency directive which, among other things, propagates sound financial protection against ruin. For an introductory overview on solvency, see e.g. Sandstrom (2005).

1.2. Some Basics

As already mentioned, we will focus on risk reserve processes in this first part. A risk reserve process, or sometimes also a surplus process is a stochastic process that describes the evolution of the reserve heights of an insurance company over time. Its dynamics are governed by

- the premium inflow and
- claims of random size that occur at random times.

Throughout the first two chapters, we will denote the premium rate, that is the amount of money, the insurer gains from its customers per unit time by $c(t)$. To reach a higher flexibility, the premium rate can change over time. The claim height distribution and the distribution of the waiting times between claims however are assumed to stay unchanged. In more mathematical terms this means, we want the heights of the claims to be independent and identically distributed (i.i.d.) and we assume the same for the inter-claim times. Furthermore, inter-claim times and claim heights are independent of each other. We write F_Y for the distribution of the claim heights and F_T for the distribution of the inter-claim times. We will use the symbols F_Y and F_T also to denote the respective distribution functions and, whenever we assume they exist, we write f_Y and f_T for the corresponding density functions.
The reserve process itself will usually be denoted by X_t for $t \in \mathbb{R}_0^+$. For mathematical completeness, we assume that F_T, F_Y and X_t are defined on a probability space $\mathcal{P}(\Omega, \mathbb{P}, \mathcal{F})$, with probability measure $\mathbb{P}$ and σ-algebra $\mathcal{F}$, that is large enough to carry all the needed quantities. However, since the measure theoretical details are not directly relevant for the upcoming results, we will, at least for the moment, not specify this any further.

If we denote the initial reserve X_0 by x and write N_t for the number of claims up to time t, then a simple example of a risk reserve process might look like this

$$X_t = \underbrace{x}_{\text{initial capital}} + \underbrace{ct}_{\text{premiums}} - \underbrace{\sum_{k=1}^{N_t} Y_k}_{\text{aggregate claims}}.$$

The case above is particularly easy, since the premium rate $c(t) \equiv c$ is constant here, meaning that the premiums up to time t can simply be computed as ct. A possible path of this reserve process is shown in Figure 1.1

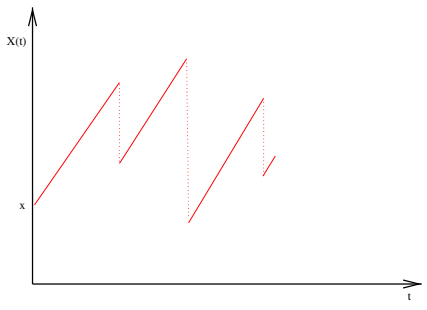

Figure 1.1.: Evolution of X_t.

What we introduced so far servers as the basis for most models built around risk processes. In order to describe more specialized problems or to cater for certain needs in applications, a large number of modifications of this basic setup has been studied. These include for example adding a diffusion term (Dufresne and Gerber (1991)), considering multiple processes at once (Dembo et al. (1994), Collamore et al. (2002)), introducing a reflection level (De Finetti (1957), Gerber and Shiu (2006), Albrecher and Thonhauser (2009)) or allowing the insurer to buy reinsurance (Waters (1983)). The latter two variations will be studied in the upcoming chapters.

As stated in the beginning, a crucial part of risk theory is the study of ruin and its consequences. To make this more formal, we define the time of ruin τ as

$$\tau(x) := \inf\{t \geq 0 : X_t < 0 | X_0 = x\}.$$

Obviously, τ depends on the starting value x and it is a stopping time with respect to $\mathcal{F}_t$, the σ-algebra generated by the process X_t. Of course, it might also happen that we never see the reserve dropping below zero. Therefore it makes sense to study the probability of ruin, usually denoted by $\psi(x)$ and defined as

$$\psi(x) := \mathbb{P}(\inf_{t \geq 0} X_t < 0 | X_0 = x).$$

These two quantities also have a straightforward relation to each other, as $\psi(x) = \mathbb{P}(\tau(x) < \infty)$. Indeed, the analysis of τ and ψ is so central to this field that some authors prefer to call it ruin theory instead of risk theory.

It should not be surprising, that the choice of the inter-claim time distribution F_T and the claim height distribution F_Y is determinant for much of the behaviour shown by the risk process. Concerning the inter-claim times, the most widely used assumption is that of exponentially distributed time intervals between single claims. This setting is known as the Cramér-Lundberg model and has been popular since the early days of modern actuarial science. The reason for this is certainly that the memoryless property of the exponential distribution makes many considerations a lot easier than they would otherwise be. In the Cramér-Lundberg model, the claim arrivals are given as a Poisson process which gives access to a rich toolbox of mathematical methods.
For the claim height distribution one usually distinguishes between light- and heavy-tailed distributions. Here, F_Y is called light-tailed, if the moment generating function $\hat{m}(s) = \mathbb{E}[e^{sY}]$ is finite for some $s > 0$ and heavy-tailed otherwise, however there are slightly different definitions used in the literature as well. Generally, light-tailed distributions are often easier to handle, whereas the heavy-tailed ones offer a better description of real claim data in many cases.

2. Quasi-Monte-Carlo Methods in Ruin Theory

To understand the motivation behind the research that was undertaken for this project, it is worth noting that finding concrete solutions is quite difficult in many problems in risk theory. Even in the most friendly case, the Cramér-Lundberg model with light-tailed claims, the set of quantities with a known closed form expression is surprisingly limited. Incorporating slight tweaks in the model assumptions will often mean that no analytic solution is available for the problem at hand. As a consequence, numerical procedures are then the only viable way to reach the assertions one is interested in.

Being a part of the special research program (SFB) for Quasi-Monte-Carlo methods, it may be seen as a natural research task to apply QMC techniques for risk processes, providing a tool that has not been very common in this field yet.

2.1. On Quasi-Monte Carlo Theory

Quasi-Monte Carlo techniques are mainly used for numerical integration. Just like with the more widespread "regular" Monte Carlo approach, the idea is to approximate an integral $\int_0^1 f(x)\,dx$ by evaluating f at N points that are evenly distributed in the unit interval.

The main difference is that while Monte Carlo simulation relies on pseudorandom numbers drawn from a uniform distribution for the evaluation points, the Quasi-Monte Carlo approach is to choose a deterministic sequence whose points are as evenly distributed across the unit interval as possible. To phrase this in more mathematical terms, a notion is introduced that measures how uniformly a sequence fills the unit interval.

Definition 2.1.1. *The discrepancy of N numbers $\{x_1, \ldots, x_N\} \in [0,1)^s$ is given as*

$$D_N := \sup_J \left| \frac{A(J, \{x_1, \ldots, x_N\})}{N} - \lambda(J) \right|,$$

where $J = \Pi_{i=1}^s [0, u_i)$ is a half-open subinterval of $[0,1)^s$, the expression $A(J, \{x_1, \ldots, x_N\})$ is the number of points from $\{x_1, \ldots, x_N\}$ that fall in the interval J and $\lambda(J)$ denotes the Lebesgue measure of J.

Looking at the definition, it becomes intuitively clear that D_N measures how much $\{x_1, \ldots, x_N\}$ deviates from a perfect uniform distribution. The above is however just one possible way of defining a discrepancy and depending on the situation, one might want (or need) to consider another variant, as it will also be done later in this chapter. A comprehensive treatment of the whole subject can be found in Kuipers and Niederreiter (1974)

A sequence $(x_n)_{n\in\mathbb{N}}$ is called a low discrepancy sequence, if for all $N \in \mathbb{N}$, $N \geq 2$ the discrepancy of $\{x_1, \ldots, x_N\}$ is low, where "low" generally means

$$N \cdot D_N \leq C_s(\log N)^s + \mathcal{O}\left((\log N)^{s-1}\right)$$

for a constant C_s which should be smallest possible. For details see e.g. Niederreiter (1988). As they are essential for Quasi-Monte Carlo integration and the possibility to consider uniformity in multiple dimensions or on different surfaces offers many ways of generalization, the construction of these low discrepancy sequences has been and still is a very active field of research. The most common low discrepancy sequences on the d-dimensional unit hypercube include the Halton sequence, the Van der Corput sequence and the Sobol sequence. The pictures below show a comparison between pseudorandom numbers and Sobol points on $[0, 1]^2$.

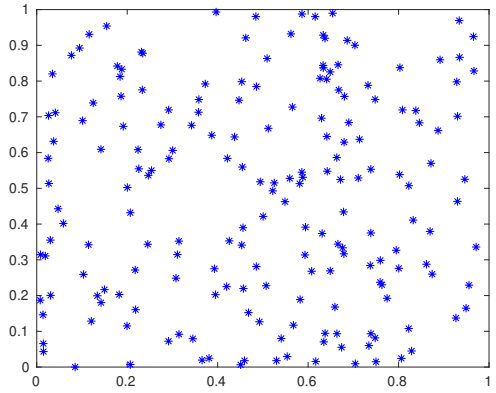

Figure 2.1.: 190 pseudorandom numbers.

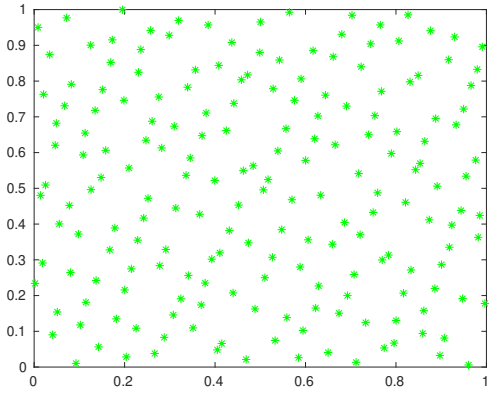

Figure 2.2.: 190 points of 2-dim. Sobol sequence.

Since among true (pseudo) random numbers there are naturally clusters but also less densely populated areas, it becomes reasonable that the same number of Quasi-Monte Carlo points achieves a better approximation of the uniform distribution. Therefore, Quasi-Monte Carlo integration features a convergence rate close to (but not actually reaching) $\mathcal{O}\left(\frac{1}{n}\right)$, in comparison to regular Monte Carlo integration, which has a probabilistic convergence rate of $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$. Although they show faster convergence behaviour

and are usually equally easy to implement, Quasi-Monte Carlo methods are still not nearly as popular as their regular counterpart. The reason for this might be that theoretical error bounds are eventually crude and in some cases the empirically shown good performance of Quasi-Monte Carlo integration lacks rigorous understanding and justification. Note however, that the overview given here is by no means comprehensive and that there exist many variations and also combinations of MC and QMC methods that play a role in research and application. Gaining more insights into unifomrly distributed sequences in general, as well as bridging the gap between theoretical and practical performance of certain methods are some of the goals of the mentioned SFB for Quasi-Monte-Carlo methods.

The remainder of this chapter, that is sections 2.2 through 2.5, was published as Preischl et al. (2018) and is therefore adopted verbatim, with only one reference, namely Tang and Wei (2010), added in section 2.5.

2.2. Introduction

During the last two decades quasi-Monte-Carlo methods (QMC-methods) have been applied to various problems in numerical analysis, statistical modeling and mathematical finance. In this paper we will give a brief survey on some of these developments and present new applications to more refined risk models involving discontinuous processes. Let us start with Fredholm integral equations of the second kind:

$$f(\mathbf{x}) = g(\mathbf{x}) + \int_{[0,1]^s} K(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}, \tag{2.1}$$

where the kernel is given by $K(\mathbf{x}, \mathbf{y}) = k(\mathbf{x} - \mathbf{y})$ with $k(\mathbf{x})$ having period 1 in each component of $\mathbf{x} = (x_1, \ldots, x_s)$. As it is quite common in applications of QMC-methods (see for example Dick et al. (2007), Sloan and Woźniakowski (2001), Kuo (2003)) it is assumed that g and k belong to a weighted Korobov space. Of course, there exists a vast literature concerning the numerical solution of Fredholm equations, see for instance Ikebe (1972), Atkinson (1967) or Twomey (1963). In particular, we want to mention the work of I. Sloan in the late 1980's where he explored various quadrature rules for solving integral equations and applications to engineering problems (Sloan and Lyness (1989), Sloan (1988) and Kumar and Sloan (1987)), which have also, after some modifications, been applied to Volterra type integral equations (see Brunner (1984) or Brunner (1992)). In Dick et al. (2007) the authors approximate f using the Nyström method based on

QMC rules.
For points $\mathbf{t}_1, \ldots, \mathbf{t}_N$ in $[0,1]^s$ the N-th approximation of f is given by

$$f_N(\mathbf{x}) := g(\mathbf{x}) + \frac{1}{N} \sum_{n=1}^{N} K(\mathbf{x}, \mathbf{t}_n) f_N(\mathbf{t}_n), \tag{2.2}$$

where the function values $f_N(\mathbf{t}_1), \ldots, f_N(\mathbf{t}_N)$ are obtained by solving the linear system

$$f_N(\mathbf{t}_j) = g(\mathbf{t}_j) + \frac{1}{N} \sum_{n=1}^{N} K(\mathbf{t}_j, \mathbf{t}_n) f_N(\mathbf{t}_n), \ j = 1, \ldots, N. \tag{2.3}$$

Under some mild conditions on K, N, and the integration points $\mathbf{t}_1, \ldots, \mathbf{t}_N$, it is shown in Dick et al. (2007) that there exists a unique solution of (2.3). Furthermore, the authors analyze the worst case error of this, so-called QMC-Nyström method. In addition, good lattice point sets $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are presented. Its convergence rate is best possible. A special focus of this important paper lies on the study of tractability and strong tractability of the QMC-Nyström method. For tractability theory in general we refer to the fundamental monograph of Novak and Woźniakowski (2010). Using ideas of Hlawka (1961) the third author of the present paper worked on iterative methods for solving Fredholm and Volterra equations, see also Hua and Wang (1981).

The idea is to approximate the solution of integral equations by means of iterated (i.e. multi-dimensional) integrals. The convergence of this procedure follows from Banach's fixed point theorem and error estimates can be established following the proof of the Picard-Lindelöf approximation for ordinary differential equations. To be more precise, let us consider integration points $\mathbf{t}_1, \ldots, \mathbf{t}_N \in [0,1]^s$ with star discrepancy D_N^* defined as usual by

$$D_N^* = \sup_{J \subset [0,1]^s} \left| \frac{1}{N} \sharp\{n \leq N : \mathbf{t}_n \in J\} - \lambda(J) \right|, \tag{2.4}$$

where the supremum is taken over all axis-aligned boxes J with one vertex in the origin and Lebesgue measure $\lambda(J)$. In Tichy (1984) the following system of r integral equations has been considered for given functions g_j on $[0,1]^{s+r}$ and h_j on $[0,1]^s$:

$$f_j(\mathbf{x}) = \int_0^{x_1} \ldots \int_0^{x_s} g_j(\xi_1, \ldots, \xi_s, f_1(\boldsymbol{\xi}), \ldots, f_r(\boldsymbol{\xi})) d\xi_s \ldots d\xi_1 + h_j(\mathbf{x}), \ j = 1, \ldots, r \tag{2.5}$$

where we have used the notations $\mathbf{x} = (x_1, \ldots, x_s) \in [0,1]^s$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_s)$. Furthermore, we assume that the partial derivatives up to order s of the functions g_j and h_j, $j = 1, \ldots, r$, are bounded by some constants G and H, respectively. Then, for a given point set $\mathbf{t}_1, \ldots, \mathbf{t}_N$ in $[0,1]^s$ with discrepancy D_N^*, the solution $\mathbf{f} = (f_1, \ldots, f_r)$ of the system (2.5) can be approximated by the quantities $\mathbf{f}^{(k)} = (f_1^{(k)}, \ldots, f_r^{(k)})$, given recursively by

$$f_j^{(k+1)}(\mathbf{x}) = \frac{x_1 \cdots x_s}{N} \sum_{n=1}^{N} g_j(x_1 t_{1,n}, \ldots, x_s t_{s,n}, f_1^{(k)}(\mathbf{x} \cdot \mathbf{t}_n), \ldots, f_r^{(k)}(\mathbf{x} \cdot \mathbf{t}_n)); \tag{2.6}$$

here $\mathbf{x} \cdot \mathbf{t}_n$ stands for the inner product $x_1 t_{1,n} + \ldots + x_s t_{s,n}$, where $\mathbf{t}_n = (t_{1,n}, \ldots, t_{s,n})$. In Tichy (1984) it is shown, that based on the classical Koksma-Hlawka inequality the worst case error, i.e., $\| \mathbf{f}^{(k)} - \mathbf{f} \|_\infty$ (sum of componentwise supremum norms) can be estimated in terms of the bounds G and H and the discrepancy D_N^* of the integration points. This method was also extended to integral equations with singularities, such as Abel's integral equation. The main focus of the present paper lies on applications in mathematical finance. In Albrecher and Kainhofer (2002) the above method was used for the numerical solution of certain Cramér-Lundberg models in risk theory. However, it turned out that in these models certain discontinuities occur. This means, that one cannot assume bounds for the involved partial derivatives and simply apply the classical Koksma-Hlawka inequality. Moreover, the involved functions are indicator functions of simplices, thus not of bounded variation in the sense of Hardy and Krause, see Drmota and Tichy (1997) and Kuipers and Niederreiter (1974).
Albrecher and Kainhofer (2002) considered a risk model with non-linear dividend barrier and made some assumptions to overcome the difficulties caused by discontinuities. For such applications it could help to use a different notion of variation for multivariate functions. Götz (2002) proved a version of the Koksma-Hlawka inequality for general measures, Aistleitner and Dick (2015) considered functions of bounded variation with respect to signed measures and Brandolini et al. (2013a,b) replaced the integration domain $[0,1]^s$ by an arbitrary bounded Borel subset of $\mathbb{R}^s$ and proved the inequality for piecewise smooth integrands. Based on fundamental work of Harman Harman (2010), a new concept of variation was developed for a wide class of functions, see Pausinger and Svane (2015) and Aistleitner et al. (2017).

In the following we give a brief overview on concepts of multivariate variation and how they can be applied for error estimates in numerical integration. Let $f(\mathbf{x})$ be a function on $[0,1]^s$ and $\mathbf{a} = (a_1, \ldots, a_s) \leq \mathbf{b} = (b_1, \ldots, b_s)$ points in $[0,1]^s$, where $\leq$ denotes the natural componentwise partial order. Following the notation of Owen (2005) and Aistleitner et al. (2017) for a subset $u \subseteq \{1, \ldots, s\}$ we denote by $\mathbf{a}^u : \mathbf{b}^{-u}$ the point with ith coordinate equal to a_i if $i \in u$ and equal to b_i otherwise. Then for the box $R = [\mathbf{a}, \mathbf{b}]$ we introduce the $s-$dimensional difference operator

$$\Delta^{(d)}(f; R) = \Delta(f; R) = \sum_u (-1)^{|u|} f(\mathbf{a}^u : \mathbf{b}^{-u}),$$

where the summation is extended over all subsets $u \in \{1, \ldots, s\}$ with cardinality $|u|$ and complement $-u$. Next we define partitions of $[0,1]^s$ as they are used in the theory of multivariate Riemann integrals, which we call here *ladder*. A ladder $\mathcal{Y}$ in $[0,1]^s$ is the cartesian product of one-dimensional partitions $0 = y_1^j < \ldots < y_{k_j}^j < 1$ (in any dimension $j = 1, \ldots, s$). Define the successor $(y_i^j)_+$ of y_i^j to be y_{i+1}^j if $i < k_j$ and $(y_{k_j}^j)_+ = 1$. For $\mathbf{y} = (y_{i_1}^1, \ldots, y_{i_s}^s) \in \mathcal{Y}$ we define the successor $\mathbf{y}_+ = ((y_{i_1}^1)_+, \ldots, (y_{i_s}^s)_+)$ and have

$$\Delta(f; [0,1]^s) = \sum_{\mathbf{y} \in \mathcal{Y}} \Delta(f; [\mathbf{y}, \mathbf{y}_+]).$$

Using the notation

$$V_{\mathcal{Y}}(f;[0,1]^s) = \sum_{\mathbf{y}\in\mathcal{Y}} \Delta(f;[\mathbf{y},\mathbf{y}_+])$$

the Vitali variation of f over $[0,1]^s$ is defined by

$$V(f;[0,1]^s) = \sup_{\mathcal{Y}} V_{\mathcal{Y}}(f;[0,1]^s). \tag{2.7}$$

Given a subset $u \subseteq \{1,\dots,s\}$, let

$$\Delta_u(f;[\mathbf{a},\mathbf{b}]) = \sum_{v\subseteq u} (-1)^{|v|} f(\mathbf{a}^v : \mathbf{b}^{-v})$$

and set $\mathbf{0} = (0,\dots,0), \mathbf{1} = (1,\dots,1) \in [0,1]^s$. For a ladder $\mathcal{Y}$ there is a corresponding ladder $\mathcal{Y}_u$ on the $|u|$-dimensional face of $[0,1]^s$ consisting of points of the form $\mathbf{x}^u : \mathbf{1}^{-u}$. Clearly,

$$\Delta_u(f;[0,1]^s) = \sum_{\mathbf{y}\in\mathcal{Y}_u} \Delta_u(f;[\mathbf{y},\mathbf{y}_+]).$$

Using the notation

$$V_{\mathcal{Y}_u}(f;[0,1]^s) = \sum_{\mathbf{y}\in\mathcal{Y}_u} \Delta_u(f;[\mathbf{y},\mathbf{y}_+])$$

for the variation over the ladder $\mathcal{Y}_u$ of the restriction of f to the face of $[0,1]^s$ specified by u, the Hardy-Krause variation is defined as

$$\mathcal{V}(f) = \mathcal{V}_{HK}(f;[0,1]^s) = \sum_{\emptyset\neq u\subseteq\{1,\dots,s\}} \sup_{\mathcal{Y}_u} V_{\mathcal{Y}_u}(f;[0,1]^s).$$

Assuming that f is of bounded Hardy-Krause variation, the classical Koksma-Hlawka inequality reads as follows:

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(\mathbf{x}_n) - \int_{[0,1]^s} f(\mathbf{x})d\mathbf{x}\right| \leq \mathcal{V}(f)D_N^*, \tag{2.8}$$

where $\mathbf{x}_1,\dots,\mathbf{x}_N$ is a finite point set in $[0,1]^s$ with star discrepancy D_N^*. In the case $f:[0,1]^s \to \mathbb{R}$ has continuous mixed partial derivatives up to order s the Vitali variation (2.7) is given by

$$\mathcal{V}(f;[0,1]^s) = \int_{[0,1]^s} \left|\frac{\partial^s f}{\partial x_1\cdots\partial x_s}(\mathbf{x})\right| d\mathbf{x}. \tag{2.9}$$

Summing over all non-empty subsets $u \subseteq [0,1]^s$ immediately yields an explicit formula for the Hardy-Krause variation in terms of integrals of partial derivatives, see (Leobacher and Pillichshammer, 2014, Ch.3, p. 59). In particular, the Hardy-Krause variation can be estimated from above by an absolute constant if we know global bounds on all partial

derivatives up to order s.

In the remaining part of the introduction we briefly sketch a more general concept of multidimensional variation which was recently developed in Pausinger and Svane (2015). Let $\mathcal{D}$ denote an arbitrary family of measurable subsets of $[0,1]^s$ which contains the empty set $\emptyset$ and $[0,1]^s$. Let $\mathcal{L}(\mathcal{D})$ denote the $\mathbb{R}-$vectorspace generated by the system of indicator functions $\mathbf{1}_A$ with $A \in \mathcal{D}$.

A set $A \subseteq [0,1]^s$ is called an algebraic sum of sets in $\mathcal{D}$ if there exist $A_1, \dots, A_m \in \mathcal{D}$ such that

$$\mathbf{1}_A = \sum_{i=1}^{n} \mathbf{1}_{A_i} - \sum_{i=n+1}^{m} \mathbf{1}_{A_i},$$

and $\mathcal{A}$ is defined to be the collection of algebraic sums of sets in $\mathcal{D}$. As in Pausinger and Svane (2015) we define the Harman complexity $h(A)$ of a non-empty set $A \in \mathcal{A}, A \neq [0,1]^s$ as the minimal number m such there exist $A_1, \dots, A_m$ with

$$\mathbf{1}_A = \sum_{i=1}^{n} \mathbf{1}_{A_i} - \sum_{i=n+1}^{m} \mathbf{1}_{A_i},$$

for some $1 \leq n \leq m$ and $A_i \in \mathcal{D}$ or $[0,1]^s \setminus A_i \in \mathcal{D}$. Moreover, set $h([0,1]^s) = h(\emptyset) = 0$ and for $f \in \mathcal{L}(\mathcal{D})$

$$V_{\mathcal{D}}^*(f) = \inf \left\{ \sum_{i=1}^{m} |\alpha_i| h_{\mathcal{D}(A_i)} : f = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i},\ \alpha_i \in \mathbb{R},\ A_i \in \mathcal{D} \right\}.$$

Furthermore, let $\mathcal{V}_\infty(\mathcal{D})$ denote the collection of all measurable, real-valued functions on $[0,1]^s$ which can be uniformly approximated by functions in $\mathcal{L}(\mathcal{D})$. Then the $\mathcal{D}-$variation of $f \in \mathcal{V}_\infty(\mathcal{D})$ is defined by

$$V_{\mathcal{D}}(f) = \inf\{ \liminf_{i\to\infty} V_{\mathcal{D}}^*(f_i) :\ f_i \in \mathcal{L}(\mathcal{D}),\ f = \lim_{i\to\infty} f_i \}, \tag{2.10}$$

and set $V_{\mathcal{D}}(f) = \infty$ if $f \notin \mathcal{V}_\infty(\mathcal{D})$. The space of functions of bounded $\mathcal{D}-$variation is denoted by $\mathcal{V}(\mathcal{D})$. Important classes of sets $\mathcal{D}$ are the class $\mathcal{K}$ of convex sets and the class $\mathcal{R}^*$ of axis aligned boxes containing $\mathbf{0}$ as a vertex. In Aistleitner et al. (2017) it is shown that the Hardy-Krause variation $\mathcal{V}(f)$ coincides with $\mathcal{V}_{\mathcal{R}^*}(f)$. For various applications the $\mathcal{D}-$variation seems to be a more natural and suitable concept. A convincing example concerning an application to computational geometry is due to Edelsbrunner and Pausinger (2016). Pausinger and Svane (2015) considered the variation $\mathcal{V}_{\mathcal{K}}(f)$ with respect to the class of convex sets. They proved the following version of the Koksma-Hlawka inequality:

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_n) - \int_{[0,1]^s} f(\mathbf{x}) d\mathbf{x} \right| \leq \mathcal{V}_{\mathcal{K}}(f) \tilde{D}_N,$$

where $\tilde{D}_N$ is the isotropic discrepancy of the point set $\mathbf{x}_1, \ldots, \mathbf{x}_N$, which is defined as follows

$$\tilde{D}_N = \sup_{C \in \mathcal{K}} \left| \frac{1}{N} \sharp\{n \leq N : \mathbf{x}_n \in C\} - \lambda(C) \right|.$$

Pausinger and Svane (2015) have shown that twice continuously differentiable functions f admit finite $\mathcal{V}_{\mathcal{K}}(f)$, and in addition they gave a bound which will be useful in our context.

Our paper is structured as follows. In Section 2 we introduce specific Markovian models in risk theory where in a natural way integral equations occur. These equations are based on arguments from renewal theory and only in particular cases they can be solved analytically. In Section 3 we develop a QMC method for such equations. We give an error estimate based on Koksma-Hlawka type inequalities for such models. In Section 4 we compare our numerical results to exact solutions in specific instances.

2.3. Discounted Penalties in the Renewal Risk Model

2.3.1. Stochastic Modeling of Risks

In the following we assume a stochastic basis $(\Omega, \mathcal{F}, P)$ which is large enough to carry all the subsequently defined random variables. In risk theory the surplus process of an insurance portfolio is modeled by a stochastic process $X = (X_t)_{t \geq 0}$. In the classical risk model, going back to Lundberg (1903), X takes the form

$$X_t = x + c\,t - \sum_{i=1}^{N_t} Y_i, \tag{2.11}$$

where the deterministic quantities $x \geq 0$ and $c \geq 0$ represent the initial capital and the premium rate. The stochastic ingredient $S_t = \sum_{i=1}^{N_t} Y_i$ is the cumulated claims process which is a compound Poisson process. The jump heights - or claim amounts - are $\{Y_i\}_{i \in \mathbb{N}}$ for which $Y_i \overset{iid}{\sim} F_Y$ with $F_Y(0) = 0$. The counting process $N = (N_t)_{t \geq 0}$ is a homogeneous Poisson process with intensity $\lambda > 0$. A crucial assumption in the classical model is the independence between $\{Y_i\}_{i \in \mathbb{N}}$ and N. A major topic in risk theory is the study of the ruin event. We introduce the time of ruin $\tau = \inf\{t \geq 0 \,|\, X_t < 0\}$, i.e., the first point in time at which the surplus becomes negative. In this setting τ is a stopping time with respect to the filtration generated by X, $\{\mathcal{F}_t^X\}_{t \geq 0}$ with $\mathcal{F}_t^X = \sigma\{X_s \,|\, 0 \leq s \leq t\}$. A first approach for quantifying the risk of X, is the study of the associated ruin probability

$$\psi(x) = P_x(X_t < 0 \ \text{ for some } \ t \geq 0) = P_x(\tau < \infty),$$

which is non-degenerate if $\mathbb{E}_x(X_1) > 0$, and satisfies the integral equation

$$\frac{c}{\lambda}\psi(x) = \int_x^\infty (1 - F_Y(y))dy + \int_0^x \psi(x-y)(1 - F_Y(y))dy.$$

In Gerber and Shiu (1998, 2005) so-called discounted penalty functions are introduced. This concept allows for an integral ruin evaluation and is based on a function $w : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ which links the deficit at ruin $|X_\tau|$ and the surplus prior to ruin $X_{\tau-} := \lim_{t \nearrow \tau} X_t$ via the function

$$V(x) = \mathbb{E}_x \left(e^{-\delta\tau} w(|X_\tau|, X_{\tau-}) \, \mathbf{1}_{\{\tau<\infty\}} \right).$$

The time of ruin τ is included by means of a discounting factor $\delta > 0$ which gives more weight to an early ruin event. In this setting specific choices of w allow for an unified treatment of ruin related quantities. In the literature, this kind of expected, discounted penalty function is often called a Gerber-Shiu function.

Remark 2.3.1. *When putting a focus on the study of $\psi(x)$, the condition $\mathbb{E}_x(X_1) > 0$ is crucial. It says that on average premiums exceed claim payments in one unit of time. Standard results, see Asmussen and Albrecher (2010), show that under this condition $\lim_{t\to\infty} X_t = +\infty$ P-a.s. From an economic perspective the accumulation of an infinite surplus is unrealistic and risk models including shareholder participation via dividends are introduced in the literature. We refer to Asmussen and Albrecher (2010) for model extensions in this direction.*

2.3.2. Markovian Risk Model

In the following we consider an insurance surplus process $X = (X_t)_{t\geq 0}$ of the form

$$X_t = x + \int_0^t c(X_{s-})ds - \sum_{i=1}^{N_t} Y_i.$$

The quantity $x \geq 0$ is called the initial capital, the cumulated claims are represented by $S_t = \sum_{i=1}^{N_t} Y_i$ and the state-dependent premium rate is $c(\cdot)$. The cumulated claims process $S = (S_t)_{t\geq 0}$ is given by a sequence $\{Y_i\}_{i\in\mathbb{N}}$ of positive, independently and identically distributed (iid) random variables and a counting process $N = (N_t)_{t\geq 0}$. For convenience we assume that the claims distribution admits a continuous density $f_Y : \mathbb{R}^+ \to \mathbb{R}^+$. In our setup we model the claim counting process $N = (N_t)_{t\geq 0}$ as a renewal counting process which is specified by the inter-jump times $\{W_i\}_{i\in\mathbb{N}}$ which are positive and iid random variables. Then, the time of the $i-$th jump is $T_i = W_1 + \ldots + W_i$ and if we assume that W_1 admits a density f_W, the jump intensity of the process X is $\lambda(t') = \frac{f_W(t')}{1-\int_0^{t'} f_W(s)ds}$. Here t' denotes the *time since the last jump*. A common assumption we are going to adopt, is the independence between $\{Y_i\}_{i\in\mathbb{N}}$ and $\{W_i\}_{i\in\mathbb{N}}$.
We choose, in contrast to classical models, a non-constant premium rate to model the effect of a so-called dividend barrier $a > 0$ in a smooth way. A barrier at level $a > 0$ has the purpose that every excess of surplus of this level is distributed as a dividend to shareholders which allows to include economic considerations in insurance modeling. Mathematically, this means that the process X is reflected at level a. Now instead of

directly reflecting the process we use the following construction. Fix $\varepsilon > 0$ and for some $\tilde{c} > 0$, define

$$c(x) = \begin{cases} \tilde{c}, & x \in [0, a-\varepsilon), \\ f(x), & x \in [a-\varepsilon, a], \\ 0, & x > a, \end{cases} \tag{2.12}$$

with some positive and twice continuously differentiable function f which fulfills $f(a-\varepsilon) = \tilde{c}$, $f(a) = 0$, $f'(a-\varepsilon) = f'(a) = f''(a-\varepsilon) = f''(a) = 0$. Altogether, we assume $c(\cdot) \in \mathcal{C}^2[0,a]$ with some Lipschitz constant $L > 0$ and $c'(a-\varepsilon) = c'(a) = 0$, $c''(a-\varepsilon) = c''(a) = 0$, $c' \leq 0$ and bounded derivatives c', c''. Then $\lim_{x \nearrow a} c(x) = 0$ and the process always stays below level a if started in $[0, a)$.
A concrete choice for f would be

$$\frac{c(a-x)^3 \left(15\varepsilon(x-a) + 6(a-x)^2 + 10\varepsilon^2\right)}{\varepsilon^5}. \tag{2.13}$$

In the following we do not specify f any further.
In this setting we add $X_0 = x$ into the definition of the time of ruin, i.e., $\tau_x = \inf\{t \geq 0 \,|\, X_t < 0,\, X_0 = x\}$.

Remark 2.3.2. *In this model setting ruin can only take place at some jump time T_k and since the process is bounded a.s. we have that $P_x(\tau_x < \infty) = 1$. If an approximation to classical reflection of the process at level a is implemented, then the process virtually started above a is forced to jump down to $a - \varepsilon$ and continue from this starting value. Consequently, we put the focus on starting values $x \in [0, a)$.*

In the remainder of this section we will study analytic properties of the discounted value function which in this framework takes the form

$$V(x) = \mathbb{E}_x \left(e^{-\delta \tau_x} w(|X_{\tau_x}|, X_{\tau_x -}) \right), \tag{2.14}$$

with $\delta > 0$ and a continuous penalty function $w : \mathbb{R}^+ \times [0, a) \to \mathbb{R}$.
To have a well defined function, typically the following integrability condition is used

$$\int_0^\infty \int_0^\infty |w(x,y)| f_Y(x+y) dy\, dx < \infty,$$

see Asmussen and Albrecher (2010). Since our process is kept below level a and w is supposed to be continuous in both arguments we can naturally replace the above condition by

$$\sup_{z \in [0,a)} \int_0^\infty |w(|z-y|, z)| f_Y(y) dy =: M < \infty, \tag{2.15}$$

which we will assume in the following. The condition from equation (2.15) holds true for example, if $|w(x,y)| \leq (1 + |x| + |y|)^p$ and F_Y admits a finite p-th moment for some $p \geq 1$. The condition (2.15) is motivated by the observation that $X_{\tau_x -} \in [0, a)$ and $|X_{\tau_x}| = |X_{\tau_x -} - Y_{N_{\tau_x}}|$ where $Y_{N_{\tau_x}} \overset{d}{\sim} f_Y$. Consequently, we get

$$V(x) \leq \mathbb{E}_x \left(|w(|X_{\tau_x}|, X_{\tau_x -})| \right) \leq \sup_{z \in [0,a)} \int_0^\infty |w(|z-y|, z)| f_Y(y) dy.$$

Remark 2.3.3. *From the construction of X we have that $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ with $\tilde{X}_t = (X_t, t'(t), t)$ is a piecewise-deterministic Markov process, see Davis (1993). Since the jump intensity depends on $t' = t - T_{N_t}$, one needs this additional component for the* Markovization *of X. But on the discrete time skeleton $\{T_i\}_{i\in\mathbb{N}}$ with $T_0 = 0$ the process $X = \{X_{T_k}\}_{k\in\mathbb{N}}$ has the Markov property.*

Remark 2.3.4. *In risk theory surplus models including a reflection at some level $a > 0$ with dynamics of the form*

$$dX_t = c\,\mathbf{1}_{\{X_t<a\}}dt - dS_t, \quad X_0 = x \geq 0,$$

arise when studying dividend strategies which pay out every excess over the level a immediately to shareholders. This is motivated by the following observation: when studying ruin probabilities it is crucial having $\mathbb{E}_x(X_1 - x) > 0$, which results in $P(X_t < 0$ for some $t > 0$ or $\lim_{t\to\infty} X_t = \infty) = 1$. This says that on the favourable set $\{\omega \in \Omega \,|\, \tau_x(\omega) = \infty\}$ the surplus becomes arbitrarily large. As a reaction to this unrealistic behaviour, a shareholder participation via dividend payments is introduced. An overview on the dividend problem in risk theory and related results can for instance be found in Albrecher and Thonhauser (2009). In the present setting, we introduce a smoothed reflection to make relevant computations accessible to an application of QMC methods, a feature which does not show up in the corresponding literature. Results on a classical QMC treatment in the situation of a non-linear dividend barrier can be found in Albrecher and Kainhofer (2002).

2.3.3. Analytic Properties and a Fixed Point Problem

We start with showing some elementary analytical properties of the function V defined in (2.14).

Theorem 2.3.5. *The function $V : [0, a) \to \mathbb{R}$ is bounded and continuous.*

Proof. The boundedness of V follows directly from the assumption made in (2.15).
For proving continuity we split off the expectation defining V into two parts which we separately deal with. Let $x > y$ and observe

$$\begin{aligned}
|\,V(x) - V(y)\,| =& \left|\,\mathbb{E}\left[e^{-\delta\tau_x} w(|X^x_{\tau_x}|, X^x_{\tau_x-}) - e^{-\delta\tau_y} w(|X^y_{\tau_y}|, X^y_{\tau_y-})\right]\right| \\
\leq& \mathbb{E}\left[e^{-\delta\tau_x}\left|w(|X^x_{\tau_x}|, X^x_{\tau_x-}) - w(|X^y_{\tau_x}|, X^y_{\tau_x-})\right| \mathbf{1}_{\{\tau_x=\tau_y\}}\right] \\
&+ \mathbb{E}\left[\left|e^{-\delta\tau_x} w(|X^x_{\tau_x}|, X^x_{\tau_x-}) - e^{-\delta\tau_y} w(|X^y_{\tau_y}|, X^y_{\tau_y-})\right| \mathbf{1}_{\{\tau_x>\tau_y\}}\right] \\
=& A + B.
\end{aligned}$$

For A we fix some $T > 0$ and notice the following bound

$$A \leq \mathbb{E}\left[e^{-\delta\tau_x}|w(|X^x_{\tau_x}|, X^x_{\tau_x-}) - w(|X^y_{\tau_x}|, X^y_{\tau_x-})|\mathbf{1}_{\{\tau_x=\tau_y\leq T\}}\right] + 2M\,P(\tau_x > T) \leq 2M. \tag{2.16}$$

Before going on we need some estimates on the difference of two paths, one starting in x and the other in y. For fixed $\omega \in \Omega$ we have that on $(0, T_1(\omega)$ the surplus fulfills $\frac{\partial X_t(\omega)}{\partial t} = c(X_t(\omega))$ with initial condition $X_0 = 0$, $T_1(\omega)$ is finite with probability one. Standard arguments on ordinary differential equations, see for instance (Stoer and Bulirsch, 2000, Th. 7.1.1 - 7.1.8), yield that an appropriate solution exists and is continuously differentiable in t and continuous in the initial value x. We even get the bound $|X_t^x - X_t^y| \leq e^{L\,t}\,|x - y|$ for fixed ω, where X_t^x denotes the path which starts in x and $L > 0$ the Lipschitz constant of $c(\cdot)$. From these results we directly obtain for a given path

$$|X_{T_1-}^x - X_{T_1-}^y| \leq e^{LT_1}|x - y|,$$

which by iteration results in

$$|X_{T_n-}^x - X_{T_n-}^y| = |X_{T_n}^x - X_{T_n}^y| \leq e^{LT_n}|x - y|,$$

because $|X_{T_n}^x - X_{T_n}^y| = |X_{T_n-}^x - Y_n - (X_{T_n-}^y - Y_n)| = |X_{T_n-}^x - X_{T_n-}^y|$. Since ruin takes place at some claim occurrence time T_k we get that on $\{\omega \in \Omega \,|\, \tau_x = \tau_y \leq T\}$ the quantities $|X_{\tau_x}^x|$ and $X_{\tau_x-}^x$ converge to the corresponding quantities started in y, all possible differences are bounded by $e^{LT}|x - y|$. Therefore, sending y to x in (2.16) and then sending T to infinity, we get that A converges to zero because $P(\tau_x < \infty) = 1$ and bounded convergence. We can repeat the argument for $x \to y$ when using $P(\tau_y > T)$ in (2.16).

Now consider part B. We first observe that $B \leq 2MP(\tau_x > \tau_y)$. Consequently, we need to show that $P(\tau_x > \tau_y)$ tends to zero if $y \to x$ or $x \to y$. Again, fix $\omega \in \Omega$ for which $\tau_x(\omega) > \tau_y(\omega)$, this implies that there is a claim amount Y_n, occurring at some point in time T_n, for which

$$X_{T_n-}^x(\omega) \geq Y_n(\omega) > X_{T_n-}^y(\omega),$$

i.e., causing ruin for the path started in y, (X_t^y), but not causing ruin for the one started in x, (X_t^x). From the construction of the drift $c(\cdot)$, it is decreasing to zero, we have that, surpressing the ω dependence,

$$0 < Y_n - X_{T_n-}^y \leq X_{T_n-}^x - X_{T_n-}^y \leq x - y.$$

Since $X_{T_n-}^y \in [0, a)$ we have

$$P(\tau_x > \tau_y) \leq \sup_{q \in [0,a)} P(0 < Y - q \leq x - y) = \sup_{q \in [0,a)} \{F_Y(x - y + q) - F_Y(q)\},$$

which approaches zero whenever x and y tend to each other since F_Y is continuous. □

Define for functions $f \in \mathcal{C}_b([0, a))$ the operator $\mathcal{A}$ by

$$\mathcal{A}f(x) := \mathbb{E}_x \left(e^{-\delta T_1} f(X_{T_1}) \mathbf{1}_{\{T_1 < \tau_x\}} + e^{-\delta \tau_x} w(|X_{T_1}|, X_{T_1-}) \mathbf{1}_{\{\tau_x = T_1\}} \right). \tag{2.17}$$

The Markov property of the sequence $\{X_{T_i}\}_{i\in\mathbb{N}}$ and the definition of V in (2.14) allow us to derive that $V = \mathcal{A}V$, or explicitly written

$$V(x) = \mathbb{E}_x\left[e^{-\delta T_1}V(X_{T_1})\mathbf{1}_{\{T_1<\tau_x\}} + e^{-\delta T_1}w(|X_{T_1}|, X_{T_1-})\mathbf{1}_{\{\tau_x=T_1\}}\right].$$

We can state the following lemma.

Lemma 2.3.6. *If $\delta > 0$, the operator $\mathcal{A} : \mathcal{C}_b([0,a)) \to \mathcal{C}_b([0,a))$ defined in (2.17) is a contraction with respect to $||\cdot||_\infty$.*

Proof. Let $f \in \mathcal{C}_b([0,a))$ be bounded by some constant M', then

$$\mathcal{A}f(x) = \mathbb{E}_x\left(e^{-\delta T_1}f(X_{T_1})\mathbf{1}_{\{T_1<\tau_x\}} + e^{-\delta\tau_x}w(|X_{T_1}|, X_{T_1-})\mathbf{1}_{\{\tau_x=T_1\}}\right),$$

is bounded by $\max\{M, M'\}$. From the integral representation of $\mathcal{A}f(x)$ we get continuity in x,

$$\mathcal{A}f(x) = \int_0^\infty e^{-\delta t_1}f_W(t_1)\Bigg[\int_0^{X_{t_1-}} f(X_{t_1-} - y_1)dF_Y(y_1) + \int_{X_{t_1-}}^\infty w(|X_{t_1-} - y_1|, X_{t_1-})dF_Y(y_1)\Bigg]dt_1,$$

where X_{t_1-} is the ODE's solution up to time t_1 with $X_0 = x$. From (Stoer and Bulirsch, 2000, Th. 7.1.4) we have that X_{t_1-} is continuous in its initial value which shows that $\mathcal{A}f(x)$ is continuous in x.
Let f, $g \in \mathcal{C}_b([0,a))$, then we have for all $x \in [0,a)$ that

$$\begin{aligned}|(\mathcal{A}f - \mathcal{A}g)(x)| &\le \int_0^\infty e^{-\delta t_1}f_W(t_1)\int_0^{X_{t_1}} |f(X_{t_1} - y_1) - g(X_{t_1} - y_1)|dF_Y(y_1)dt_1\\ &\le ||f-g||_\infty \int_0^\infty e^{-\delta t_1}f_W(t_1)dt_1 = ||f-g||_\infty \mathbb{E}[e^{-\delta T_1}].\end{aligned}$$

Since $\delta > 0$ and $T_1 > 0$ $P-$a.s., $\mathcal{A}$ is contractive with Lipschitz constant $\tilde{L} = \mathbb{E}[e^{-\delta T_1}] < 1$. □

For a possible application of quasi-Monte Carlo techniques we need to examine the structure of $\mathcal{A}$,

$$\begin{aligned}\mathcal{A}v(x) &= \int_0^\infty e^{-\delta t_1}f_W(t_1)\int_0^{X_{t_1-}} v(X_{t_1-} - y_1)dF_Y(y_1)dt_1 +\\ &\quad \int_0^\infty e^{-\delta t_1}f_W(t_1)\int_{X_{t_1-}}^\infty w(y_1 - X_{t_1-}, X_{t_1-})dF_Y(y_1)dt_1\\ &=: \mathcal{G}v(x) + \mathcal{H}(x).\end{aligned}$$

For $n \in \mathbb{N}$ the probabilistic interpretation of iterated applications of $\mathcal{A}$ is $\mathcal{A}^n v(x) = \mathbb{E}_x\left(e^{-\delta T_n}v(X_{T_n})\mathbf{1}_{\{T_n<\tau_x\}} + e^{-\delta\tau_x}w(|X_{\tau_x}|, X_{\tau_x-})\mathbf{1}_{\{\tau_x\le T_n\}}\right)$. Using $\mathcal{G}$ and $\mathcal{H}$ we can write

$$\mathcal{A}^n v(x) = \mathcal{G}^n v(x) + \sum_{k=0}^{n-1}\mathcal{G}^k\mathcal{H}(x),$$

where $\mathcal{G}^n v(x) = \mathbb{E}_x(e^{-\delta T_n} v(X_{T_n}) \mathbf{1}_{\{T_n<\tau\}})$ and

$$\mathcal{G}^{k-1}\mathcal{H}(x) = \int_0^\infty \cdots \int_0^\infty \int_{X_{\bar{t}_k-}}^\infty \int_0^{X_{\bar{t}_{k-1}-}} \cdots \int_0^{X_{\bar{t}_1-}} \left(\prod_{i=1}^k e^{-\delta t_i} f_W(t_i)\right) w(y_k - X_{\bar{t}_k-}, X_{\bar{t}_k-}) dF_Y(y_k) \cdots dF_Y(y_1) dt_k \cdots dt_1.$$

Here, $\bar{t} := \sum_{i=1}^k t_i$ and represents the time of the k-th jump. We see that via $X_{\bar{t}_k} = X_{\bar{t}_{k-1}} - y_k + \int_{\bar{t}_{k-1}}^{\bar{t}_k} c(X_s) ds$ the path of the process depends on all integration variables $(t_1, \ldots, t_k, y_1, \ldots, y_k)$.
For dealing with the situation $\delta = 0$, i.e., when the contraction argument fails, we can use a probabilistic argument. Since $\lim_{n\to\infty} T_n = \infty$ and $P(\tau_x < \infty) = 1$ we have that $\lim_{n\to\infty} \mathcal{G}^n v(x) = \lim_{n\to\infty} \mathbb{E}_x \left(e^{-\delta T_n} v(X_{T_n}) \mathbf{1}_{\{T_n<\tau\}}\right) = 0$ for $v \in \mathcal{C}_b([0,a))$. Using $|\mathcal{A}^n v(x) - V(x)| = |\,\mathcal{G}^n v(x) - \mathcal{G}^n V(x)\,|$ we get $\lim_{n\to\infty} \mathcal{A}^n v(x) = V(x)$ pointwise, even in the case if $\delta = 0$.
In what follows we put the focus on the determination of $\mathcal{G}^k\mathcal{H}(x)$.

2.4. Approximation Procedure

For the application of QMC methods we need to transform in a first step the integration domain in

$$\mathcal{G}^{k-1}\mathcal{H}(x) = \int_0^\infty \cdots \int_0^\infty \int_{X_{\bar{t}_k-}}^\infty \int_0^{X_{\bar{t}_{k-1}-}} \cdots \int_0^{X_{\bar{t}_1-}} \left(\prod_{i=1}^k e^{-\delta t_i} f_W(t_i)\right) w(y_k - X_{\bar{t}_k-}, X_{\bar{t}_k-}) dF_Y(y_k) \cdots dF_Y(y_1) dt_k \cdots dt_1$$

to $[0,1]^{2k}$. This is achieved by use of the following substitutions

$$\begin{aligned}
\alpha_i &:= e^{-t_i} \Rightarrow t_i = -\log \alpha_i \qquad \text{for } i \in \{1, \ldots, k\} \\
\beta_i &:= \frac{y_i}{X_{\bar{t}_i-}} \Rightarrow y_i = X_{\bar{t}_i-}\beta_i \qquad \text{for } i \in \{1, \ldots, k-1\} \\
\beta_k &:= e^{X_{\bar{t}_k-}} e^{-y_k} \Rightarrow y_k = X_{\bar{t}_k-} - \log \beta_k.
\end{aligned}$$

Here it has to be taken into account that the values of the reserve process X have to be calculated recursively, i.e., $X_{\bar{t}_i-}$ depends on $t_1, \ldots, t_i$ and $y_1, \ldots, y_{i-1}$. Since the Jacobian matrix of this transformation has a lower triangular form, the determinant can easily be found as $\frac{1}{\alpha_1 \ldots \alpha_k} X_{\bar{t}_1-} \cdots X_{\bar{t}_{k-1}-} \frac{1}{\beta_k}$. Altogether, we arrive at

$$\mathcal{G}^{k-1}\mathcal{H}(x) = \int_{[0,1]^{2k}} \prod_{i=1}^k \alpha_i^\delta f_W(t_i(\alpha_i)) \prod_{i=1}^k f_Y(y_i(\alpha_1, \ldots, \alpha_i, \beta_1, \ldots, \beta_i)) \frac{1}{\alpha_1 \ldots \alpha_k} X_{\bar{t}_1-} \cdots X_{\bar{t}_{k-1}-} \frac{1}{\beta_k} w(-\log \beta_k, X_{\bar{t}_k-})\, d\alpha_1 \ldots d\alpha_k d\beta_1 \ldots d\beta_k.$$

Consequently, for recovering the Koksma-Hlawka type error bound we need to examine the variation of the integrand:

$$F(\alpha_1,\dots,\alpha_k,\beta_1,\dots,\beta_k) = \left(\prod_{i=1}^{k-1}\alpha_i^{\delta-1} f_W(-\log(\alpha_i))\right)\left(\prod_{i=1}^{k-1} f_Y(\beta_i X_{\bar{t}_i-})X_{\bar{t}_i-}\right)\cdot \left(\alpha_k^{\delta-1} f_W(-\log(\alpha_k)) f_Y(X_{\bar{t}_k-}-\log(\beta_k))\frac{1}{\beta_k} w(-\log\beta_k, X_{\bar{t}_k-})\right). \quad (2.18)$$

Here we denote by $\phi(t,s)$ the solution to $\frac{\partial}{\partial t}x(t) = c(x(t))$ with $x(0) = s$. Consequently, we can write

$$X_{\bar{t}_i-} = X_{\bar{t}_{i-1}-} - y_{i-1} + \phi(t_i, X_{\bar{t}_{i-1}-} - y_{i-1}).$$

Or in terms of α_i, putting $\hat{x}_{i-1} = X_{\bar{t}_{i-1}-} - y_{i-1} = X_{\bar{t}_{i-1}-}(1-\beta_{i-1})$ and

$$X_{\bar{t}_i-} = \hat{x}_{i-1} + \phi(-\log(\alpha_i), \hat{x}_{i-1}). \quad (2.19)$$

In the following proposition we show that with a particular choice of model parameters it is possible to apply results from Pausinger and Svane (2015) to show that the integrand in (2.18) is in some sense of finite variation. Its proof shows that probabilistic and deterministic model ingredients are considerably interconnected.

Theorem 2.4.1. *Let $f_W(t) = \lambda e^{-\lambda t}\mathbf{1}_{\{t\geq 0\}}$ ($\lambda > 0$), $f_Y(y) = \mu e^{-\mu y}\mathbf{1}_{\{y\geq 0\}}$ ($\mu > 0$), $w \equiv 1$ and $c(\cdot)$ be specified by (2.13). Then, under the assumption $\lambda + \delta \geq 3$ and $\mu \geq 3$ the variation $\mathcal{V}_{\mathcal{K}}(F)$ (see (2.10) with $\mathcal{D} = \mathcal{K}$) of F, defined in (2.18), is finite.*

Proof. The main idea of the proof is the application of (Pausinger and Svane, 2015, Th. 3.12). For this purpose we need to show that $M(F) = \sup\{\|\mathrm{Hess}(F,x)\| \,|\, x \in [0,1]^{2k}\}$, $\sup F$ and $\inf F$ are finite, with the implication

$$\mathcal{V}_{\mathcal{K}}(F) \leq \sup F - \inf F + M(F).$$

Since in this theorem the operator (matrix) norm $\|\mathrm{Hess}(F,x)\|$ is arbitrary we use the 2-norm and exploit the relation

$$\|\mathrm{Hess}(F,x)\|_2 \leq \left(\sum_{i=1}^{2k}\sum_{j=1}^{2k}[\mathrm{Hess}(F,x)]_{ij}^2\right)^{\frac{1}{2}}.$$

We will show that $[\mathrm{Hess}(F,x)]_{ij}$ is finite for all $x \in [0,1]^{2k}$. At first we observe that when taking derivatives with respect to α_i and β_j, the structure of (2.19) implies the appearance of the following terms:

$$\frac{\partial}{\partial t}\phi(t,s) = c(\phi(t,s)),\ \frac{\partial^2}{\partial t^2}\phi(t,s) = c'(\phi(t,s))c(\phi(t,s)),$$
$$\frac{\partial}{\partial s}\phi(t,s) =: y(t,s) = e^{\int_0^t c'(\phi(u,s))du},\ \frac{\partial^2}{\partial t\partial s}\phi(t,s) = c'(\phi(t,s))y(t,s),$$
$$\frac{\partial^2}{\partial s^2}\phi(t,s) =: z(t,s) = y(t,s)\int_0^t c''(\phi(u,s))y(u,s)du.$$

The functions y, z correspond to the first and second derivative of the ODE's solution with respect to the initial value. They can be derived from the associated first and second order variational equations (see Grigorian (2009)). From our assumptions on $c(\cdot)$ we have that y is bounded by one ($c' \leq 0$) and all other derivatives including z are bounded as well. The boundedness of z can be derived from the boundedness of $c''(\cdot)$ and an analysis of the growth behaviour of y.
For the structure of $[\mathrm{Hess}(F, x)]_{ij}$ we can derive the following

$$\left(\prod_{l=1}^{k} \alpha_l^{\delta+\lambda-a_{ij}} \beta_k^{\mu-b_{ij}} e^{-\mu(y_1+\cdots+y_{k-1}+X_{\bar{t}_k-})}\right) \cdot Q_{ij}\left(\beta_1, \ldots, \beta_{k-1}, \phi, \frac{\partial}{\partial t}\phi, \frac{\partial^2}{\partial t^2}\phi, \frac{\partial}{\partial s}\phi, \frac{\partial^2}{\partial t \partial s}\phi, \frac{\partial^2}{\partial s^2}\phi\right),$$

for a_{ij}, $b_{ij} \in \{1, 2, 3\}$ and a function Q_{ij}. Q_{ij} is evaluated at the integration points and ϕ and its derivatives which themselves are evaluated in points of the form $(-\log(\alpha_l), \hat{x}_{l-1}) \in (0, \infty) \times [0, a)$ for $l \in \{1, \ldots, k\}$. If ϕ and its derivatives are considered to be variables, neglecting their dependence on the α_ls and β_ls, then Q_{ij} is a polynomial of degree k. The degree of the polynomial is produced by the recursive structure of the paths and its dependence on all previous jump times and sizes. From this inspection we get that under the conditions $\lambda + \delta \geq 3$ and $\mu \geq 3$ all entries of the Hessian matrix are bounded. Furthermore, the conditions on the parameters λ, δ, μ ensure that $\sup F$ is finite and $\inf F = 0$. □

Remark 2.4.2. *We can combine the above result with the convergence rate from Banach's fixed point theorem and obtain for our specific situation*

$$\begin{aligned}\left\|\sum_{k=0}^{n} \hat{\mathcal{G}}^k \mathcal{H} - V\right\|_\infty &\leq \left\|\sum_{k=0}^{n} (\hat{\mathcal{G}}^k \mathcal{H} - \mathcal{G}^k \mathcal{H})\right\|_\infty + \|\mathcal{A}^n - V\|_\infty + \|\mathcal{G}^n v\|_\infty \\ &\leq \sum_{k=0}^{n} \mathcal{V}_{\mathcal{K}}(F^k) \tilde{D}_{N_k} + \frac{\tilde{L}^n}{1-\tilde{L}} \|\mathcal{A}v - v\|_\infty + M' \left(\frac{\lambda}{\delta+\lambda}\right)^n.\end{aligned}$$

Here F^k denotes the integrand from (2.18) in dimension $2k$, $\tilde{D}_{N_k}$ the isotropic discrepancy of a pointset with N_k elements in $[0, 1]^{2k}$ and $\hat{\mathcal{G}}^k \mathcal{H}$ is the QMC approximation for $\mathcal{G}^k \mathcal{H}$. For the last term we used that v is bounded by some $M' > 0$ and the fact the T_n follows a Gamma distribution $\Gamma(n, \lambda)$.
From the type of arguments we used for the proof of Theorem 2.4.1, we expect that the result holds true for Γ-distributed inter-claim times and jump heights and $w(y, z) = y^k z^l$ with similar conditions on the parameters. Hence the method is also applicable for this more general situation. A detailed study of this claim is part of future research.

2.5. Numerical Results

In this section, we evaluate the integrals from Section 2.4 by applying Monte Carlo and quasi-Monte Carlo methods for different choices of the penalty function w.

Note that in the general case, determining the Gerber-Shiu function analytically is a profoundly hard problem, since only certain parameter constellations allow for explicit results. For constant parameter settings, in particular constant drift c, inter-claim times and jumps following *phase-type* distributions and special choices of w, the problem can be handled by matrix-analytic methods. For an overview on these techniques see Asmussen and Albrecher (2010), or for exemplary results one may consult Albrecher et al. (2010) and Lin et al. (2003). A main focus in the risk theoretic literature lies on asymptotic approximations as the initial value x becomes large, these results are referred to as Cramér-Lundberg type approximations, see Asmussen and Albrecher (2010) and Rolski et al. (1999). While there are examples for literature on the numerical treatment of Gerber-Shiu functions (e.g. Tang and Wei (2010)), these are relatively scarce in contrast to those on probabilistic approximations. For a survey on the use of collocation methods we refer to Makroglou (2004).

2.5.1. The Discounted Time of Ruin

Letting $w(y,z) := 1$, we arrive at $V(x) = \mathbb{E}_x(e^{-\delta\tau_x} w(|X_{\tau_x}|, X_{\tau_x-})) = \mathbb{E}_x(e^{-\delta\tau_x})$ which is the discounted time of ruin. Lin et al. Lin et al. (2003) found an analytic expression for this discounted time of ruin if both the inter-arrival times of the claims and the claim sizes are exponentially distributed. To have a reference value, we also adopt these assumptions and denote the parameters of the exponential distributions with λ for the parameter of the inter-arrival times and μ for the parameter of the claim sizes. The premium rate $c(\cdot)$ was chosen as in Section 2.3.2 with f from equation (2.13), with $\tilde{c} = 2$, $a = 3$ and ε was set to 0.001. Note that the results of Lin et al. (2003) were proved for a reflected process in the classical sense, which means $c(x) = \tilde{c}$ for $x \leq a$ and $c(x) = 0$ for $x > a$. Since Theorem 2.4.1 requires a premium rate satisfying certain smoothness conditions, we cannot use a discontinuous c and thus have a methodic error in our simulations. However, we will see that this error is, at least for small ε, very small.

We list the parameters together with the approximation values for increasing numbers of (Q)MC points and $k = 20$ iterations of the algorithm in Table 2.1, whereas Table 2.2 shows the approximation values for $k = 100$ iterations of the algorithm. Figure 2.3 and Figure 2.4 show the MC points (green) with 95% confidence intervals, together with QMC points from Sobol sequences (blue) and Halton sequences (orange).
The red line at height 0.7577 marks the analytically found value for the reflected process. We use it as a reference value here but, again, remark that it is not the exact value for our smoothed process. As can be seen in Figure 2.3, the algorithm has not yet converged for $k = 20$, whereas Figure 2.4 shows that $k = 100$ already yields a very good approximation. The computation time of the above example, for $k = 100$, was under 4 minutes for both choices of QMC sequences, whereas the MC method has a runtime of more than 30 minutes. The exact computation times are given in Table 2.3.

To illustrate the speed of convergence, we also plotted the absolute error, both for the

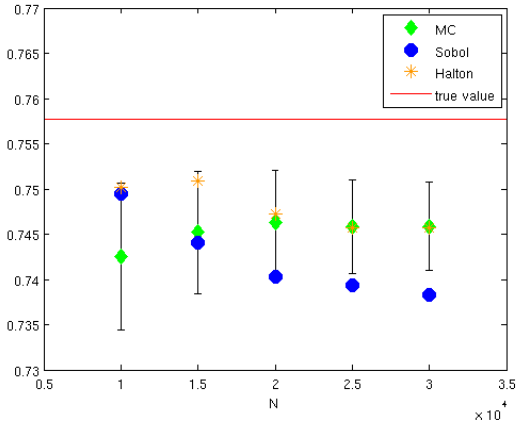

Figure 2.3.: $k = 20$ iterations of the algorithm.

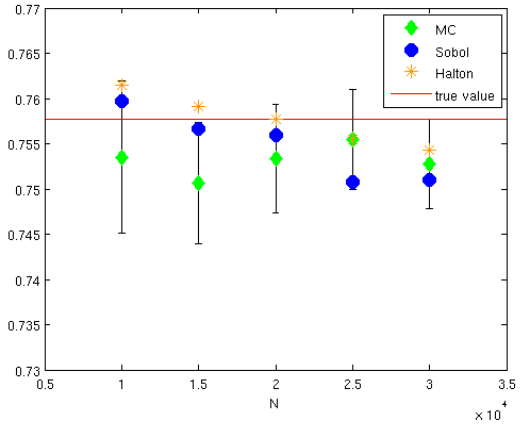

Figure 2.4.: $k = 100$ iterations of the algorithm.

x	λ	μ	$w(y,z)$	k	δ
1.2	1	0.8	1	20	0.05
N:	10000	15000	20000	25000	30000
MC:	0.7425	0.7452	0.7463	0.7458	0.7459
Sobol:	0.7494	0.7440	0.7403	0.7394	0.7383
Halton:	0.7502	0.7509	0.7473	0.7488	0.7457

Table 2.1.

x	λ	μ	$w(y,z)$	k	δ
1.2	1	0.8	1	100	0.05
N:	10000	15000	20000	25000	30000
MC:	0.7535	0.7507	0.7534	0.7555	0.7527
Sobol:	0.7597	0.7566	0.7560	0.7508	0.7510
Halton:	0.7615	0.7591	0.7577	0.7555	0.7543

Table 2.2.

k	MC points	Halton seq.	Sobol seq.
20	121	37	38
100	2436	198	200

Table 2.3.: Times to obtain the plots in Figures 2.1 and 2.2 in seconds.

MC approach as well as for QMC points (again taken from Sobol and Halton sequences) for varying numbers of points N. Figures 2.5 and 2.6 show the values obtained for $k = 40$ iterations of the algorithm. Obviously, $k = 40$ is also not yet enough to reach the actual value. But notice that the absolute error even for more iterations cannot converge to zero because of the smoothed reflection procedure. For both of the QMC methods, a scramble improved the results. In the Sobol case however, an "unlucky" choice in the scramble and the skip value (i.e. how many elements are dropped in the beginning) can lead to relatively high variation in the output, whereas the Halton set shows a more stable performance (compare Figures 2.5 and 2.6).

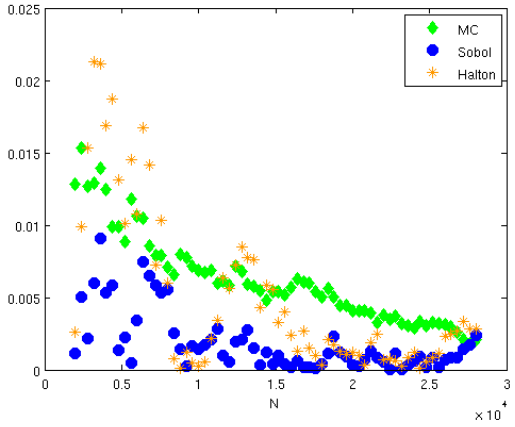

Figure 2.5.: "lucky" choice of QMC points.

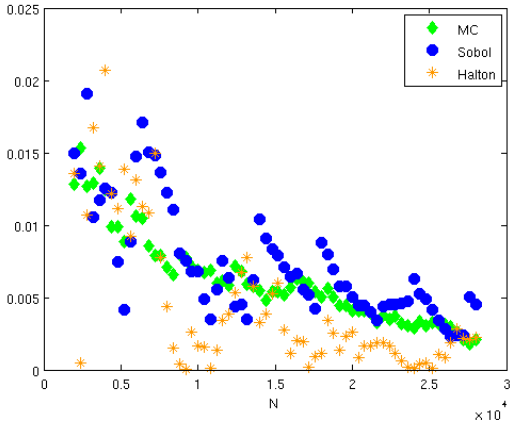

Figure 2.6.: "unlucky" choice of QMC points.

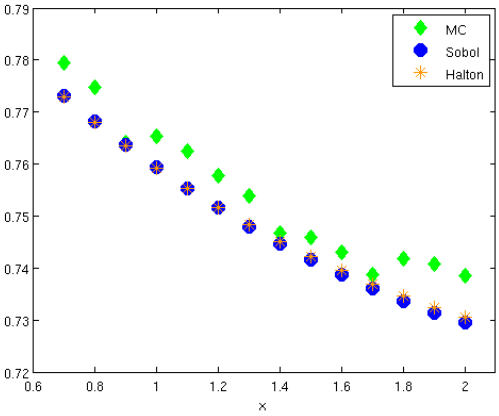

Figure 2.7.: Influence of the starting value

For Figure 2.7 we evaluated $k = 40$ iterations of the algorithm with $N = 30000$ (Q)MC points for different starting values x, ranging from 0.7 to 2. As expected, the discounted time of ruin decreases for increasing x.

2.5.2. The Deficit at Ruin

If we set $w(y, z) := y$, and $\delta = 0$, we have $V(x) = \mathbb{E}_x(|X_{\tau_x}|)$, the expected deficit at ruin. We use the same premium rate $c(\cdot)$ as before and again choose exponential distributions for the inter-arrival times and claim sizes with parameters λ and μ respectively, since also in this case the true value $\mathbb{E}_x(|X_{\tau_x}|) = \frac{1}{\mu}$ (for a classically reflected process) can be found in Lin et al. (2003). Figures 2.8 and 2.9 show the results for $k = 20$ and $k = 100$ iterations respectively. The reference value is again shown as a red line, in our case at 1.25. The MC points are drawn in green, the Sobol points blue and the Halton points in orange. Table 2.4 and Table 2.5 contain the precise values along with the corresponding parameters.

Note again the difference between Figure 2.8 and Figure 2.9, resulting from a different number of iterations k. The computation times for these plots deviate very little from those given in Figure 2.3.

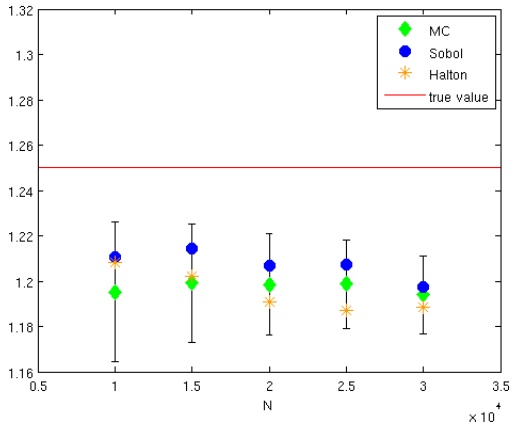

Figure 2.8.: $k = 20$ iterations of the algorithm.

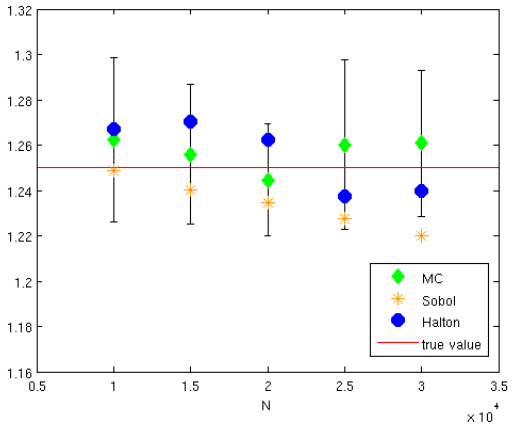

Figure 2.9.: $k = 100$ iterations of the algorithm.

x	λ	μ	$w(y,z)$	k	δ
1.2	1	0.8	y	20	0
N:	10000	15000	20000	25000	30000
MC:	1.1952	1.1991	1.1986	1.1987	1.1939
Sobol:	1.2105	1.2142	1.2070	1.2074	1.1975
Halton:	1.2084	1.2019	1.1906	1.1872	1.1885

Table 2.4.

x	λ	μ	$w(y,z)$	k	δ
1.2	1	0.8	y	100	0
N:	10000	15000	20000	25000	30000
MC:	1.2624	1.2558	1.2446	1.2602	1.2607
Sobol:	1.2669	1.2704	1.2624	1.2373	1.2398
Halton:	1.2487	1.2404	1.2344	1.2273	1.2199

Table 2.5.

Again, we plotted the absolute error for $k = 40$ iterations of the algorithm and a varying number of (Q)MC points N. Figure 2.10 shows the results using the same colorings as before.

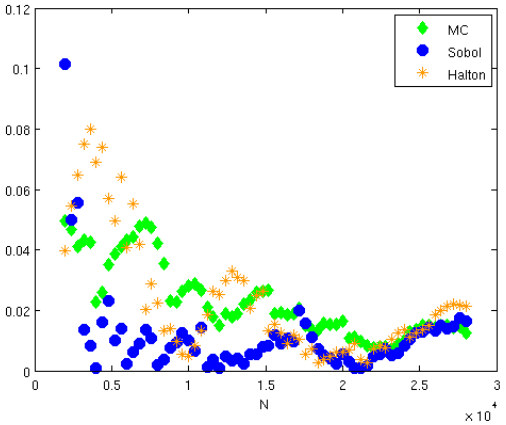

Figure 2.10.: The absolute error for the deficit at ruin

Remark 2.5.1. *We considered in our numerical examples two test cases for which explicit (approximate) reference values are available. Certainly our approach is not restricted to this particular choice of model ingredients - which are f_Y, f_W and the penalty function w.*
While there are several approximation techniques for discounted penalty functions, it is precisely this flexibility that makes the (Q)MC approach favourable in many situations.

3. Optimal Reinsurance for Gerber-Shiu Functions

If an insurance company decides to buy an insurance contract from another company, e.g. against very large claims, this is called reinsurance. From a practitioner's point of view, reinsurance is a central and highly necessary part of insurance business, be it for the company buying reinsurance, or the one offering it. It is therefore not surprising that actuaries have developed and improved ways to model reinsurance for several decades. However, it wasn't until the early 2000s that reinsurance was considered in a dynamic way, meaning the extent of the protection can be updated in response to business development. It may be questioned, whether this dynamic assumption is very realistic in applications, since contracts are generally renewed after fixed time periods (e.g. annually), but from a mathematical point of view, these dynamic contracts give rise to a very rich and interesting theory, which might eventually find its way into actuarial practice. It turned out that stochastic optimal control was a suitable approach to incorporate this dynamic setting and after some pioneering works this new branch in insurance mathematics spawned quite a lot of research. For an overview see e.g. Albrecher et al. (2017).

When something shall be *optimal* this instantly induces the question *with respect to what?* Indeed there are numerous different performance criteria for reinsurance and therefore also many concepts of optimality. So far, most literature focused on picking a specific feature that was to be optimized, usually ruin probabilities or dividend payouts, and then deriving a method tailored to solve this particular problem. The discounted penalty functions that were introduced in the last chapter, however offer a way to unify many relevant quantities connected to risk processes. So a method aiming to optimize these penalty functions is a useful generalization that will be presented in the following.

The remainder of this chapter, that is sections 3.1 through 3.4, was published as Preischl and Thonhauser (2019) and is therefore adopted verbatim, modulo some typographical errors that have been corrected in this thesis.

3.1. Introduction and Preliminaries

3.1.1. Motivation

The problem of choosing an optimal reinsurance contract has been a very active field inside actuarial mathematics for several years and numerous different frameworks have been considered in this context. The earlier works on this topic were inspired by Waters (1983) where the idea is to maximize the adjustment coefficient to achieve the fastest decay rate for the ruin probability with increasing initial capital. While this approach is focused on the asymptotic behaviour and therefore results in a static reinsurance strategy, Schmidli (2001, 2002), Hipp and Vogt (2003) and Hipp and Taksar (2010) considered dynamic control strategies, so the reinsurance policy can adapt to the evolution of the reserve process. A collection of results on optimal dynamic reinsurance can be found in Schmidli (2008). Like the papers cited above, most authors working on dynamic reinsurance take the perspective of optimal stochastic control. A comprehensive summary of these methods in insurance mathematics is provided by Azcue and Muler (2014).
Many different approaches can be made, depending on whether or not capital injections are considered, a diffusion term is added to the risk process and also which functional is to be optimized. For the latter question, the most popular choice is the ruin probability but other target values are thinkable and interesting. For example Azcue and Muler (2005) and Cani and Thonhauser (2017) ask for the strategy maximizing a dividend payoff and it is shown that results are qualitatively different from optimal strategies for minimizing the probability of ruin. In our manuscript, we will consider a quite general selection of functionals combined in the notion of discounted penalty functions, a concept that is widely used in many branches of insurance mathematics.

3.1.2. The Model

We consider a risk reserve process $(X_t)_{t\geq 0}$ in the classical Cramér-Lundberg model. That is, starting from some initial value x, the reserve process evolves over time subject to premium income and claim occurrence. The claim arrivals are given by a Poisson process with intensity λ, i.e. there are λ claims to be expected per unit time (equivalently, inter claim times follow an exponential distribution with mean $\frac{1}{\lambda}$). The claim heights are independent of this Poisson process and follow some continuous distribution F_Y on $(0,\infty)$. Although not strictly necessary, we will in general assume that F_Y has a density f_Y.
In our setting, reinsurance can be obtained in the form of a control function u in the following sense:
At each point in time t, a control parameter u is chosen from a compact set U (e.g. $U=[0,1]$). The map $u_t : \mathbb{R}^+ \to U$ is called the reinsurance strategy and by $\mathcal{U}$ we denote the set of processes on U that are previsible with respect to $\mathcal{F}_t^X$, the filtration generated by the process X_t. The functions in $\mathcal{U}$ are called *admissible control strategies*.
The effect of the reinsurance is modeled by the *retention function* $r : \mathbb{R}\times U \to \mathbb{R}_0^+$. If a claim of height y is encountered at time t, only the part $r(y,u_t)$ is to be paid by the first

insurer (in the following also called the cedent), the rest of the cost is transferred to the reinsurance company. Note that r is non-negative, meaning the cedent cannot reinsure more than the actual claim height. For the whole paper, we assume r to be monotone in y and continuous in u.
Of course, reinsurance is not for free and so the reinsurance strategy also influences the reinsurance premiums and thus ultimately the premium income of the cedent. Therefore, the premium rate at time t is calculated as

$$c(u_t) = c - p(u_t),$$

where c denotes the cedent's premiums without reinsurance and $p(u_t)$ is the reinsurer's premium. These premiums can be calculated in several ways, including the *expectation principle*, the *variance principle* and the *exponential principle* as some of the most popular ones. Throughout this article, we assume that U contains parameters corresponding to the two extremal cases of reinsurance, which are no reinsurance, denoted by $\mathring{u}$ and full reinsurance, denoted by u^*, i.e.

$$\begin{aligned} r(y, \mathring{u}) &= y, & c(\mathring{u}) &= c, \\ r(y, u^*) &= 0, & c(u^*) &=: \pi < 0. \end{aligned}$$

Note that we want to assume that the reinsurance premium is in relation higher than the cedent's premium. So buying full reinsurance will result in a negative premium rate π.

Combining these quantities, we define the process X_t^u controlled by the strategy $u \in \mathcal{U}$:

$$X_t^u = x + \int_0^t c(u_s)\, ds - \sum_{i=1}^{N_t} r(Y_i, u_{T_i}).$$

Here, and in the rest of the paper, N_t denotes the number of claims up to time t and T_i resp. Y_i denotes the time resp. the height of the i-th claim.

Let $\tau_x^u := \inf\{t \geq 0 : X_t^u < 0 | X_0^u = x\}$ denote the *time of ruin*, i.e. the first point in time at which X_t^u becomes negative. For convenience, we freeze the process after the ruin event, that is $X_t^u = X_{\tau_x^u}^u$ for all $t > \tau_x^u$. Following Gerber and Shiu (1998), we are interested in discounted penalty functions (or *Gerber-Shiu functions*) of the following form

$$\Phi^u(x) := \mathbb{E}_x \left[e^{-\delta \tau_x^u} w(X_{\tau_x^u -}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{\tau_x^u < \infty} \right].$$

Here, $X_{\tau_x^u -}^u$ is called *surplus prior to ruin*, $|X_{\tau_x^u}^u|$ is the *deficit at ruin* and $\delta > 0$ is a discounting factor. Throughout this article, we demand that $w : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \to \mathbb{R}^+$ is a continuous function. Furthermore, we want w to fulfill the two boundary conditions

$$\int_0^\infty \int_0^\infty w(x, y) f_Y(x + y)\, dx\, dy < \infty. \tag{3.1a}$$

and

$$\int_0^\infty w(x,y) f_Y(x+y)\,dy < M < \infty \qquad \forall x \in \mathbb{R}_0^+ \tag{3.1b}$$

Given that we want to minimize the penalty, we are left with finding

$$V(x) := \inf_{u \in \mathcal{U}} \Phi^u(x),$$

for $x > 0$. We will also call $V(x)$ the *value function*.

Remark 3.1.1. *Condition* (3.1a) *ensures that* $\Phi^{\mathring{u}}(x)$*, the Gerber-Shiu function with the strategy that does not buy any reinsurance, has a finite value. It is, for example, fulfilled if* w *is bounded. However, we didn't want to make boundedness an a priori assumption since many results only require the weaker statement* (3.1a)*. Both boundedness and* (3.1a) *imply that* $\Phi^{\mathring{u}}(x) \to 0$ *as* $x \to \infty$ *(see Asmussen and Albrecher (2010)). Note that conditions* (3.1a) *and* (3.1b) *do not imply each other.*

3.1.3. Properties of the Value Function

To conclude the preliminaries, we want to show two short but important lemmas, giving monotonicity, boundedness and, under mild conditions, Lipschitz continuity of V.

Lemma 3.1.2. $V(x)$ *is strictly monotonically decreasing. Furthermore, it holds that* $0 \le V(x) \le e^{\delta \frac{x}{\pi}} w(0,0)$*.*

Proof. Since $w(x,y) \ge 0$, it is obvious that $0 \le V(x)$. To achieve $V(x) \le e^{\delta \frac{x}{\pi}} w(0,0)$ just take full reinsurance, resulting in some negative drift π, until ruin occurs. We now show monotonicity. Let $x > y$ and, starting in x, buy continuously full reinsurance. Hence, deterministically, after time $\frac{y-x}{\pi}$ the process reaches level y. Taking an ε-optimal strategy from there means

$$V(x) \le e^{-\delta \frac{y-x}{\pi}} \left(V(y) + \varepsilon\right) < V(y)$$

for $\varepsilon < e^{\delta \frac{y-x}{\pi}} V(y) \left(1 - e^{-\delta \frac{y-x}{\pi}}\right)$. □

Remark 3.1.3. *Note that Lemma 3.1.2 is a statement about the discounted penalty function of the optimally controlled process. Monotonicity does not necessarily hold for an arbitrary control strategy. We also want to point out that boundedness of* V *does not require boundedness of* w *or any of the integrability conditions* (3.1a) *and* (3.1b) *but solely needs* $w(0,0) < \infty$*, as long as* U *contains full reinsurance. We will however need some integrability conditions on* w *respectively* Φ *to show the Lipschitz continuity of* V *in the next lemma.*

Lemma 3.1.4. *If* (3.1a) *and* (3.1b) *hold, then* $V(x)$ *is Lipschitz continuous.*

Proof. From (3.1a) it follows that the expected penalty function without reinsurance, $\mathbb{E}_x\left[w(X^{\mathring{u}}_{\tau^{\mathring{u}}_x-}, |X^{\mathring{u}}_{\tau^{\mathring{u}}_x}|)\right]$, is bounded by some constant M_0. Let $y < x$ and let $u_t \equiv \mathring{u}$ be the constant control strategy of no reinsurance. Under this strategy and without any claim, the process started at level y will reach level x at time $\frac{x-y}{c}$. For every $x > 0$ there is an ε-optimal strategy $\tilde{u}$ which fulfills

$$V(x) \geq \Phi^{\tilde{u}}(x) - \varepsilon.$$

We now define a new control strategy $\bar{u} = (\bar{u}^y_t)$ for the process starting in y by $\bar{u}_t = u_t$ for $0 \leq t \leq T_1 \wedge \frac{x-y}{c}$ and

$$\bar{u}_t = \mathbb{1}_{T_1 > \frac{x-y}{c}} \tilde{u}_{t-\frac{x-y}{c}} + \mathbb{1}_{T_1 \leq \frac{x-y}{c}} \mathring{u}$$

for $t > T_1 \wedge \frac{x-y}{c}$. So starting in y, the strategy $\bar{u}_t$ refuses to buy any reinsurance unless level x is reached before the first claim, in which case the strategy changes to the ε-optimal choice $\tilde{u}$.
We have

$$\begin{aligned}
V(y) \leq & \Phi^{\bar{u}}(y) \\
= & \mathbb{P}\left(T_1 > \frac{x-y}{c}\right) \mathbb{E}_y\left[e^{-\delta\tau^{\bar{u}}_y} w\left(X^{\bar{u}}_{\tau^{\bar{u}}_y-}, \left|X^{\bar{u}}_{\tau^{\bar{u}}_y}\right|\right) \middle| T_1 > \frac{x-y}{c}\right] \\
& + \mathbb{P}\left(T_1 \leq \frac{x-y}{c}\right) \mathbb{E}_y\left[e^{-\delta\tau^{\bar{u}}_y} w\left(X^{\bar{u}}_{\tau^{\bar{u}}_y-}, \left|X^{\bar{u}}_{\tau^{\bar{u}}_y}\right|\right) \mathbb{1}_{Y_1 \leq X^{\bar{u}}_{T_1-}} \middle| T_1 \leq \frac{x-y}{c}\right] \\
& + \mathbb{P}\left(T_1 \leq \frac{x-y}{c}\right) \mathbb{E}_y\left[e^{-\delta T_1} w\left(X^{\bar{u}}_{T1-}, \left|X^{\bar{u}}_{T_1}\right|\right) \mathbb{1}_{Y_1 > X^{\bar{u}}_{T_1-}} \middle| T_1 \leq \frac{x-y}{c}\right] \\
\leq & \mathbb{P}\left(T_1 > \frac{x-y}{c}\right) e^{-\delta\frac{x-y}{c}} \Phi^{\tilde{u}}(x) + \mathbb{P}\left(T_1 \leq \frac{x-y}{c}\right) \underbrace{(M_0 + M)}_{=:M_1},
\end{aligned}$$

where the last inequality follows from applying conditional expectation w.r.t $\mathcal{F}^X_{\frac{x-y}{c}}$ resp. $\mathcal{F}^X_{T_1}$, then using tower and markov property and finally conditions (3.1a) and (3.1b). So

$$\begin{aligned}
|V(x) - V(y)| &= V(y) - V(x) \\
&\leq V(x)\left(e^{-(\delta+\lambda)\frac{x-y}{c}} - 1\right) + \varepsilon e^{-(\delta+\lambda)\frac{x-y}{c}} + \left(1 - e^{-\lambda\frac{x-y}{c}}\right) M_1
\end{aligned}$$

Note that Lipschitz continuity implies absolute continuity of V. □

3.2. Main Results

Since we want to use the theory of stochastic optimal control, it is crucial to show that the value function is a solution to the problem's Hamilton-Jacobi-Bellman equation (HJB). For this, we need to show that the *dynamic programming principle* holds.

Lemma 3.2.1 (dynamic programming principle)**.** *For every bounded stopping time S, we have*

$$V(x) = \inf_{u\in\mathcal{U}} \mathbb{E}_x \left[e^{-\delta S} V(X_S^u) \mathbb{1}_{S<\tau_x^u} + e^{-\delta\tau_x^u} w(X_{\tau_x^u-}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{S\geq\tau_x^u} \right], \tag{3.2}$$

Proof. Let S be a bounded stopping time. We first show that for an arbitrary strategy u_t and any $\varepsilon > 0$, it is possible to find a strategy $\hat{u}_t$ in a measurable way such that $V(X_S^u) > \Phi^{\hat{u}}(X_S^u) - \varepsilon$, i.e. $\hat{u}$ is ε-optimal for the random starting point X_S^u. So let $\varepsilon > 0$. There exists $x_n > 0$ such that with constantly full reinsurance parameter u^* it holds $\Phi^{u^*}(x) = e^{\delta\frac{x}{\pi}} w(0,0) < \varepsilon$ for all $x \geq x_n$. Since V is continuous, we can choose a grid $x_0 < x_1 < \cdots < x_{n-1} < x_n$ such that $0 \leq V(x_{i-1}) - V(x_i) < \frac{\varepsilon}{2}$ for $i \in \{1, 2, \ldots, n\}$. Also, for every fixed x_i, there exists an $\frac{\varepsilon}{2}$-optimal strategy u^i with $\Phi^{u^i}(x_i) < V(x_i) + \frac{\varepsilon}{2}$. Note that for $X_S^u > x_n$, the strategy of full reinsurance is ε-optimal for X_S^u as it holds $\Phi^{u^*}(X_S^u) - V(X_S^u) < \varepsilon$, since $\Phi^{u^*}(X_S^u) < \varepsilon$ and $V(X_S^u) \geq 0$.
So assuming $X_S^u \in [x_{i-1}, x_i)$ we choose the strategy $\hat{u}^{i-1}$ which does the following: Buy continuously full reinsurance until the level x_{i-1} is reached, then apply the $\frac{\varepsilon}{2}$-optimal strategy u^{i-1}. We show that $\hat{u}^{i-1}$ is ε-optimal for X_S^u.

$$|\Phi^{\hat{u}^{i-1}}(X_S^u) - V(X_S^u)| \leq |\Phi^{\hat{u}^{i-1}}(X_S^u) - V(x_{i-1})| + \underbrace{|V(x_{i-1}) - V(X_S^u)|}_{<\frac{\varepsilon}{2}} < \varepsilon$$

because

$$|\Phi^{\hat{u}^{i-1}}(X_S^u) - V(x_{i-1})| = e^{-\delta\frac{x_{i-1}-X_S^u}{\pi}} \Phi^{u^{i-1}}(x_{i-1}) - V(x_{i-1}) < \frac{\varepsilon}{2},$$

which holds because $e^{-\delta\frac{x_{i-1}-X_S^u}{\pi}}$ can be made arbitrarily close to 1 by increasing the fineness of the grid.

We now define a composed strategy by choosing an arbitrary strategy u_t in the interval $[0, S]$ and, starting in S, we choose the strategy $\hat{u}_t := \sum_{i=1}^n u_{t-S}^{i-1} \mathbb{1}_{[x_{i-1},x_i)}(X_S^u)$. Above, we showed that $V(X_S^u) > \Phi^{\hat{u}}(X_S^u) - \varepsilon$. Call this composed strategy $\bar{u}_t$. We get

$$\begin{aligned} V(x) &\leq \Phi^{\bar{u}}(x) \\ &= \mathbb{E}_x \left[e^{-\delta S} \Phi^{\hat{u}}(X_S^u) \mathbb{1}_{S<\tau_x^u} + e^{-\delta\tau_x^u} w(X_{\tau_x^u-}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{S\geq\tau_x^u} \right] \\ &< \mathbb{E}_x \left[e^{-\delta S} V(X_S^u) \mathbb{1}_{S<\tau_x^u} + e^{-\delta\tau_x^u} w(X_{\tau_x^u-}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{S\geq\tau_x^u} \right] + \varepsilon. \end{aligned}$$

Since the inequality holds for arbitrary u and ε, we get that

$$V(x) \leq \inf_{u\in\mathcal{U}} \mathbb{E}_x \left[e^{-\delta S} V(X_S^u) \mathbb{1}_{S<\tau_x^u} + e^{-\delta\tau_x^u} w(X_{\tau_x^u-}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{S\geq\tau_x^u} \right].$$

To show the other direction, we take an arbitrary strategy u_t and get

$$\begin{aligned} \Phi^u(x) &= \mathbb{E} \left[e^{-\delta S} \Phi^u(X_S^u) \mathbb{1}_{S<\tau_x^u} + e^{-\delta\tau_x^u} w(X_{\tau_x^u-}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{S\geq\tau_x^u} \right] \\ &\geq \mathbb{E} \left[e^{-\delta S} V(X_S^u) \mathbb{1}_{S<\tau_x^u} + e^{-\delta\tau_x^u} w(X_{\tau_x^u-}^u, |X_{\tau_x^u}^u|) \mathbb{1}_{S\geq\tau_x^u} \right] \\ &\geq \inf_{\hat{u}\in\mathcal{U}} \mathbb{E} \left[e^{-\delta S} V(X_S^{\hat{u}}) \mathbb{1}_{S<\tau_x^{\hat{u}}} + e^{-\delta\tau_x^{\hat{u}}} w(X_{\tau_x^{\hat{u}}-}^{\hat{u}}, |X_{\tau_x^{\hat{u}}}^{\hat{u}}|) \mathbb{1}_{S\geq\tau_x^{\hat{u}}} \right] \end{aligned}$$

Since u was arbitrary, we can take the infimum on the left hand side and get the claim. □

Now we address the Hamilton-Jacobi-Bellman equation. The proof will follow similar arguments as the one of Lemma 3 in Cani and Thonhauser (2017).

Lemma 3.2.2. *The value function $V(x)$ is on $(0,\infty)$ a.e. a solution to*

$$0 = \inf_{u\in U}\left\{c(u)V'(x) - (\delta+\lambda)V(x) + \lambda\int_0^{\rho(x,u)} V(x-r(y,u))\,dF_Y(y) + \lambda\int_{\rho(x,u)}^{\infty} w(x,r(y,u)-x)\,dF_Y(y)\right\}. \tag{3.3}$$

Here, $\rho(a,u) := \inf\{y\in\mathbb{R}_0^+ : r(y,u)\geq a\}$ denotes the inverse of the retention function in the first component.

Proof. We first show the $\leq$ part.
Fix $x>0$ and $u\in U$ and take $h>0$ such that $x+h\pi>0$. Consider the strategy $\hat{u}_t\equiv u$ for $t\in[0,h]$ and $\hat{u}_t=\tilde{u}_{t-h}$ for $t>h$ for some $\tilde{u}\in\mathcal{U}$. With T_1 again being the time of the first claim, set $S:=\min\{h,T_1\}$. Obviously, S is a stopping time and the strategy $\hat{u}$ is constant in the time interval $[0,S]$. Setting $V(x)=0$ for $x<0$ and using (3.2), we have

$$0\leq \mathbb{E}_x\left[e^{-\delta S}V(X_s^{\hat{u}})\right] - V(x) + \mathbb{E}_x\left[e^{-\delta\tau_x^{\hat{u}}}w(X^{\hat{u}}_{\tau_x^{\hat{u}}-},|X^{\hat{u}}_{\tau_x^{\hat{u}}}|)\mathbb{1}_{S\geq\tau_x^{\hat{u}}}\right].$$

Applying Dynkin's formula and using $\hat{u}\equiv u$ on $[0,S]$ yields

$$\begin{aligned}0\leq\ & \mathbb{E}_x\left[V(x)+\int_0^S e^{-\delta t}\left(\mathcal{A}^uV(X^u_{t-}) - \delta V(X^u_{t-})\right)dt\right] - V(x)\\ &+\mathbb{E}_x\left[e^{-\delta\tau_x^u}w(X^u_{\tau_x^u-},|X^u_{\tau_x^u}|)\mathbb{1}_{S\geq\tau_x^u}\right],\end{aligned}$$

where $\mathcal{A}^u$ denotes the generator of the process X_t^u, which, according to Rolski et al. (1999), Theorem 11.2.2, is given by

$$\mathcal{A}^ug(x) = c(u)g'(x) - \lambda g(x) + \lambda\int_0^\infty g(x-r(y,u))\,dF_Y(y). \tag{3.4}$$

This leads to

$$\begin{aligned}0\leq&\mathbb{E}_x\left[\int_0^S e^{-\delta t}\Big(c(u)V'(X_t^u) - (\delta+\lambda)V(X_t^u) + \lambda\int_0^{\rho(X_t^u,u)} V(X_t^u - r(y,u))\,dF_Y(y)\Big)\,dt\right]\\ &+\mathbb{E}_x\left[e^{-\delta\tau_x^u}w(X^u_{\tau_x^u-},|X^u_{\tau_x^u}|)\mathbb{1}_{S\geq\tau_x^u}\right].\end{aligned}$$

Collecting the terms and dividing by h gives

$$\begin{aligned}0\leq&\frac{1}{h}\,\mathbb{E}_x\left[\int_0^S e^{-\delta t}c(u)V'(x+c(u)t)\,dt\right]\\ &+\frac{1}{h}\mathbb{E}_x\left[e^{-\delta T_1}w(x+c(u)T_1,|X^u_{T_1}|)\mathbb{1}_{S\geq\tau_x^u}\right]\\ &+\frac{1}{h}\mathbb{E}_x\left[\int_0^S e^{-\delta t}\left(-(\delta+\lambda)V(x+c(u)t) + \lambda\int_0^{\rho(x+c(u)t,u)} V(x+c(u)t-r(y,u))\,dF_Y(y)\right)dt\right].\end{aligned}$$

Having created an analogous situation as in the proof of Lemma 3 in Cani and Thonhauser (2017), we can use the same arguments to deduce

$$0 \leq \inf_{u\in\mathcal{U}} \Bigg\{ c(u)V'(x) - (\delta+\lambda)V(x) + \lambda \int_0^{\rho(x,u)} V(x - r(y,u))\, dF_Y(y) \\ + \lambda \int_{\rho(x,u)}^{\infty} w(x, r(y,u) - x)\, dF_Y(y) \Bigg\},$$

which is the first half of the proof.

For the other direction, we fix $x > 0$ and choose $h > 0$ such that $x + \pi h > 0$, where $\pi < 0$ is again the premium under full reinsurance. Let u^1 be an h^2-optimal strategy for (3.2) and take again $S := \min\{T_1, h\}$. Starting, as above, with (3.2), we get

$$0 > \mathbb{E}_x \left[e^{-\delta S} V(X_S^{u^1}) \mathbb{1}_{S<\tau_x^{u^1}} - V(x) \right] - h^2 - \varepsilon h \\ + \mathbb{E}_x \left[e^{-\delta S} w(X^{u^1}_{\tau_x^{u^1}-}, |X^{u^1}_{\tau_x^{u^1}}|) \mathbb{1}_{S\geq\tau_x^{u^1}} \right].$$

Conditioning on the time and height of the first claim and using the exponential distribution of the inter-claim times, this can be written as

$$\begin{aligned} 0 > {} & e^{-(\delta+\lambda)h} V\left(\tilde{x}_h\right) \\ & + \mathbb{E}_x \left[\int_0^h \lambda e^{-(\delta+\lambda)t} \int_0^{\rho(\tilde{x}_t, u_t^1)} V\left(\tilde{x}_t - r(y, u_t^1)\right) dF_Y(y)\, dt \right] \\ & + \mathbb{E}_x \left[\int_0^h \lambda e^{-(\delta+\lambda)t} \int_{\rho(\tilde{x}_t, u_t^1)}^{\infty} w\left(\tilde{x}_t, r(y, u_t^1) - \tilde{x}_t\right) dF_Y(y)\, dt \right] \\ & - V(x) - h^2 - h\varepsilon. \end{aligned}$$

Note that, to improve readability, we used the notational shortcuts $\tilde{x}_t := x + \int_0^t c(u_s^1) ds$ and $\tilde{x}_h := x + \int_0^h c(u_s^1) ds$.

At this point, we can again follow the proof of Lemma 3 in Cani and Thonhauser (2017) to deduce that

$$0 > c(u_0^1)V'(x) - (\delta+\lambda)V(x) + \lambda \int_0^{\rho(x,u_0^1)} V(x - r(y,u_0^1))\, dF_Y(y) \\ + \lambda \int_{\rho(x,u_0^1)}^{\infty} w(x, r(y,u_0^1) - x)\, dF_Y(y) - \varepsilon.$$

And letting $\varepsilon \to 0$ completes the proof. □

Having shown that the value function is a solution to the HJB-equation (3.3), we now need to show that it is the only one (at least with some given analytical properties).

3.3. Uniqueness of the Solution to the HJB-equation and Verification Statement

Ruin can either occur by a claim that is bigger than the current reserve (claim ruin) or by decreasing the reserve with a negative premium until the reserve becomes negative (smooth ruin). Under certain conditions it can actually be advantageous to deliberately induce smooth ruin and thus choose the penalty $e^{-\delta\tau_x^u}w(0,0)$. Later, we will see that the possibility of smooth ruin causes changes in the analytical framework of the model.

Write $\mathcal{C}^{+,b}[0,\infty)$ for the set of positive, continuous and bounded functions on $[0,\infty)$ and define the operator $\mathcal{G}$ on $\mathcal{C}^{+,b}[0,\infty)$ as

$$\begin{aligned}\mathcal{G}f(x) := \inf_{u\in\mathcal{U}}\Big\{&\mathbb{E}_x\left[e^{-\delta T_1}f(X_{T_1}^u)\mathbb{1}_{T_1<\tau_x^u}\right]\\&+\mathbb{E}_x\left[e^{-\delta T_1}w(X_{T_1-}^u,|X_{T_1}^u|)\mathbb{1}_{T_1=\tau_x^u}\right]\\&+\mathbb{E}_x\left[e^{-\delta\tau_x^u}w(0,0)\mathbb{1}_{T_1>\tau_x^u}\right]\Big\}.\end{aligned}$$

Lemma 3.3.1. *$\mathcal{G}f\in\mathcal{C}^{+,b}[0,\infty)$. Furthermore, $\mathcal{G}$ is a contraction on $\mathcal{C}^{+,b}[0,\infty)$.*

Proof. Positivity and boundedness follow immediately from (3.1a) and $w(x,y)\geq 0$. Now let $f\in\mathcal{C}^{+,b}[0,\infty)$ and $x,y\in[0,\infty)$ with $x>y$. With the same argumentation as in Lemma 3.1.2, we get that $\mathcal{G}f$ is monotonously decreasing as a function in x. Choose u_x as an ε-optimal strategy in $\mathcal{G}f(x)$ and write $\mathcal{G}^{u_x}f(x)$ for the right hand side of $\mathcal{G}f(x)$, with the control strategy u_x
In the following, we consider the reserve process pathwise. Write ${}_zX_t^u$ for the risk process at time t, started in z and controlled by the strategy u. Let $\mathring{u}\in U$ be the parameter corresponding to no reinsurance and define $\xi := \inf\{t : {}_yX_t^{\mathring{u}} = {}_xX_t^{u_x}\}$, so ξ is the time when the process started in y hits the path of the process started in x. Now set the strategy $(u_{y,t})_t\equiv\mathring{u}$ for $t\in[0,\xi]$ and $(u_{y,t})_t=(u_{x,t})_t$ for $t>\xi$. We have

$$|\mathcal{G}f(x)-\mathcal{G}f(y)| = \mathcal{G}f(y)-\mathcal{G}f(x)\leq\mathcal{G}^{u_y}f(y)-\mathcal{G}^{u_x}f(x)+\varepsilon.$$

Obviously, denoting the time of ruin of the process started in x and controlled by the strategy u_x by ${}_x\tau^{u_x}$, we have ${}_x\tau^{u_x}\geq{}_y\tau^{u_y}$. Expanding the above equation gives

$$\begin{aligned}\mathcal{G}^{u_y}f(y)-\mathcal{G}^{u_x}f(x)+\varepsilon =&\mathbb{E}\left[e^{-\delta T_1}f({}_yX_{T_1}^{u_y})\mathbb{1}_{T_1<{}_y\tau^{u_y}}\right]\\&+\mathbb{E}\left[e^{-\delta T_1}w({}_yX_{T_1-}^{u_y},|{}_yX_{T_1}^{u_y}|)\mathbb{1}_{T_1={}_y\tau^{u_y}}\right]\\&+\mathbb{E}\left[e^{-\delta\tau^{u_y}}w(0,0)\mathbb{1}_{T_1>{}_y\tau^{u_y}}\right]\\&-\mathbb{E}\left[e^{-\delta T_1}f({}_xX_{T_1}^{u_x})\mathbb{1}_{T_1<{}_x\tau^{u_x}}\right]\\&-\mathbb{E}\left[e^{-\delta T_1}w({}_xX_{T_1-}^{u_x},|{}_xX_{T_1}^{u_x}|)\mathbb{1}_{T_1={}_x\tau^{u_x}}\right]\\&-\mathbb{E}\left[e^{-\delta\tau^{u_x}}w(0,0)\mathbb{1}_{T_1>{}_x\tau^{u_x}}\right]+\varepsilon.\end{aligned}$$

After collecting terms, we see that

$$\begin{aligned}\mathcal{G}^{u_y} f(y) - \mathcal{G}^{u_x} f(x) = \mathbb{E}\left[e^{-\delta T_1}(f({}_yX_{T_1}^{u_y}) - f({}_xX_{T_1}^{u_x}))\mathbb{1}_{T_1<{}_y\tau^{u_y}}\right] \\ - \mathbb{E}\left[e^{-\delta T_1} f({}_xX_{T_1}^{u_x})\mathbb{1}_{{}_y\tau^{u_y}=T_1<{}_x\tau^{u_x}}\right] \\ + \mathbb{E}\left[e^{-\delta T_1} w({}_yX_{T_1-}^{u_y}, |{}_yX_{T_1}^{u_y}|)\mathbb{1}_{{}_y\tau^{u_y}=T_1={}_x\tau^{u_x}}\right] \\ - \mathbb{E}\left[e^{-\delta T_1} w({}_xX_{T_1-}^{u_x}, |{}_xX_{T_1}^{u_x}|)\mathbb{1}_{{}_y\tau^{u_y}=T_1={}_x\tau^{u_x}}\right].\end{aligned}$$

Note that the terms for smooth ruin before T_1 cancel out, since in this setting smooth ruin is only possible, after the processes started in x and y have merged. At this point it is helpful to distinguish the cases $\xi \leq T_1$ and $\xi > T_1$, so whether or not the merge has already happened before the first claim. Considering the summands separately yields

$$\begin{aligned}&\mathbb{E}\left[e^{-\delta T_1}(f({}_yX_{T_1}^{u_y}) - f({}_xX_{T_1}^{u_x}))\mathbb{1}_{T_1<{}_y\tau^{u_y}}\right] \\ &= \mathbb{E}\left[e^{-\delta T_1}(f({}_yX_{T_1}^{u_y}) - f({}_xX_{T_1}^{u_x}))\mathbb{1}_{T_1<{}_y\tau^{u_y}}\mathbb{1}_{\xi>T_1}\right] \\ &+ \underbrace{\mathbb{E}\left[e^{-\delta T_1}(f({}_yX_{T_1}^{u_y}) - f({}_xX_{T_1}^{u_x}))\mathbb{1}_{T_1<{}_y\tau^{u_y}}\mathbb{1}_{\xi\leq T_1}\right]}_{=0}.\end{aligned}$$

We see that for $\xi \leq T_1$ the terms cancel out. To analyze what happens for $\xi > T_1$, take $\varepsilon_c > 0$ and define

$$t^* = \inf\{t : \exists \varepsilon_t > 0 : c(u_x(\tilde{t})) < c - \varepsilon_c \text{ for } \tilde{t} \in [t, t+\varepsilon_t]\}.$$

In other words, at t^* starts the first open interval where the drift of the process started in x is by at least ε_c smaller than the drift of the process started in y. For $|x-y|$ small enough, this interval in time will, even for arbitrarily small ε_c, be enough for ${}_yX_t^{u_y}$ to reach the trajectory of ${}_xX_t^{u_x}$ so we know $\xi \in [t^*, t^*+\varepsilon_t]$ with $\varepsilon_t \to 0$ for $|x-y| \to 0$. Now let us consider the first claim occurrence T_1.

- For $T_1 < t^*$, the processes haven't merged yet, but their premium rates are at most ε_c apart and since the premium is a continuous, strictly monotone function, their control strategies are at most δ_c apart. Since the retention function is also continuous in u, and ε_c was arbitrary, we know that $|f({}_yX_{T_1}^{u_y}) - f({}_xX_{T_1}^{u_x})| \to 0$ as $|x-y| \to 0$.
- If $t^* < T_1 < \xi$, we cannot directly control the difference in the jump at T_1, but since we know that $\xi \in [t^*, t^*+\varepsilon_t]$ and because the distribution of T_1 is continuous, $\mathbb{P}(T_1 \in [t^*, t^*+\varepsilon_t])$ goes to zero for $\varepsilon_t \to 0$.

Similarly, for the second summand, we see that

$$\begin{aligned}&\mathbb{E}\left[e^{-\delta T_1} f({}_xX_{T_1}^{u_x})\mathbb{1}_{{}_y\tau^{u_y}=T_1<{}_x\tau^{u_x}}\right] \\ &= \mathbb{E}\left[e^{-\delta T_1} f({}_xX_{T_1}^{u_x})\mathbb{1}_{{}_y\tau^{u_y}=T_1<{}_x\tau^{u_x}}\mathbb{1}_{\xi<T_1}\right].\end{aligned}$$

Using the definition of t^* as before, we have again two cases to consider.

- For $T_1 < t^*$ we already argued that the two paths of the process are arbitrarily close for $|x-y|$ being sufficiently small. Since for a claim that ruins the process started in y but not the one started in x, we know that the claim height Y_1 must be in $[{}_yX_{T_1-}^{u_y}, {}_xX_{T_1-}^{u_x}]$ and since the claim height distribution is assumed to be continuous, we deduce $\mathbb{P}(Y_1 \in [{}_yX_{T_1-}^{u_y}, {}_xX_{T_1-}^{u_x}]) \to 0$ for $|x-y| \to 0$.

- In the case $t^* < T_1 < \xi$, we can use the same argumentation as above to reach the conclusion that $\mathbb{P}(T_1 \in [t^*, t^* + \varepsilon_t])$ goes to zero for $|x-y| \to 0$.

A combination of the arguments we used so far and exploiting the continuity of w will also send the remaining two summands to 0, showing continuity of $\mathcal{G}f$.

The analysis of $|\mathcal{G}f(x) - \mathcal{G}f(y)|$ showed that the operator $\mathcal{G}$ really acts on $\mathcal{C}^{+,b}[0,\infty)$, i.e. in particular that $\mathcal{G}f$ is again a continuous function. It now remains to prove that $\mathcal{G}$ is a contraction on $\mathcal{C}^{+,b}[0,\infty)$. Hence, given two functions f_1 and f_2, we want to bound the supremum norm $\|\mathcal{G}f_1 - \mathcal{G}f_2\|_\infty$. So let f_1, f_2 be positive, continuous and bounded and w.l.o.g. choose $x \in [0,\infty)$ such that $\mathcal{G}f_1(x) \geq \mathcal{G}f_2(x)$. Let u_2 be a ε-optimal strategy for f_2 in $\mathcal{G}$. We have

$$
\begin{aligned}
0 &< \mathcal{G}f_1(x) - \mathcal{G}f_2(x) \\
&= \inf_{u\in\mathcal{U}} \left\{ \mathbb{E}_x\left[e^{-\delta T_1} f_1(X_{T_1}^u)\mathbb{1}_{T_1<\tau^u}\right] + \mathbb{E}_x\left[e^{-\delta T_1} w(X_{T_1-}^u, |X_{T_1}^u|)\mathbb{1}_{T_1=\tau^u}\right] \right. \\
&\quad \left. + \mathbb{E}_x\left[e^{-\delta\tau^u} w(0,0)\mathbb{1}_{T_1>\tau^u}\right]\right\} \\
&\quad - \inf_{u\in\mathcal{U}} \left\{ \mathbb{E}_x\left[e^{-\delta T_1} f_2(X_{T_1}^u)\mathbb{1}_{T_1<\tau^u}\right] + \mathbb{E}_x\left[e^{-\delta T_1} w(X_{T_1-}^u, |X_{T_1}^u|)\mathbb{1}_{T_1=\tau^u}\right] \right. \\
&\quad \left. - \mathbb{E}_x\left[e^{-\delta\tau^u} w(0,0)\mathbb{1}_{T_1>\tau^u}\right]\right\} \\
&\leq \mathbb{E}_x\left[e^{-\delta T_1}(f_1(X_{T_1}^{u_2}) - f_2(X_{T_1}^{u_2}))\mathbb{1}_{T_1<\tau^{u_2}}\right] + \varepsilon \\
&= \int_0^\infty e^{-\delta t}\lambda e^{-\lambda t} \int_0^{\rho(X_{t-}^{u_2}, u_2)} f_1(X_{t-}^{u_2} - r(y,u_2)) - f_2(X_{t-}^{u_2} - r(y,u_2)) dF_Y(y) dt + \varepsilon \\
&\leq \underbrace{\mathbb{E}\left[e^{-\delta T_1}\right]}_{<1} \|f_1 - f_2\|_\infty + \varepsilon.
\end{aligned}
$$

Taking $\varepsilon < \left(1 - \mathbb{E}\left[e^{-\delta T_1}\right]\right)\|f_1 - f_2\|_\infty$ completes the proof. □

From the definition of $\mathcal{G}$, we see that $\mathcal{G}V = V$ holds by the dynamic programming principle. In the following, we want to establish the connection between $\mathcal{G}$ and the HJB-equation.

Lemma 3.3.2. *Let $f \in \mathcal{C}^{+,b}[0,\infty)$ be a solution to the HJB-equation (3.3) with $f(0) \leq$*

$w(0,0)$. *For* $x \in (0,\infty)$ *set*

$$u_f(x) = \arg\min_{u\in U} \left\{ c(u)f'(x) - (\delta+\lambda)f(x) + \lambda \int_0^{\rho(x,u)} f(x - r(y,u))\, dF_Y(y) \right. \tag{3.5}$$
$$\left. + \lambda \int_{\rho(x,u)}^{\infty} w(x, x - r(y,u))\, dF_Y(y) \right\}.$$

We complement the definition of u_f *by taking*

$$u_f(0) = \begin{cases} u_f(0+) & \text{if } f(0) < w(0,0) \\ u^* & \text{if } f(0) = w(0,0), \end{cases}$$

where u^* *denotes the strategy of full reinsurance. Then* f *is a fixed point of* $\mathcal{G}$ *and* u_f *is the minimizing strategy.*

Proof. We start with the HJB-equation

$$0 = \inf_{u\in U} \left\{ c(u)f'(x) - (\delta+\lambda)f(x) + \lambda \int_0^{\rho(x,u)} f(x - r(y,u))\, dF_Y(y) \right.$$
$$\left. + \lambda \int_{\rho(x,u)}^{\infty} w(x, x - r(y,u))\, dF_Y(y) \right\}.$$

This holds for arbitrary x and f is certainly defined at all X_t^u for $t \in [0, T_1 \wedge \tau_x^u]$. So we define the minimizing strategy $u_{f,t} = u_f\left(X_{t-}^{u_f}\right)$ (which exists by the continuity of all functions that are present in (3.5)). Using Dynkin's formula and writing u_f for $u_{f,t}$, we have

$$\begin{aligned}
0 &= \mathbb{E}_x \left[\int_0^{T_1\wedge\tau^{u_f}} e^{-\delta t} \Bigg(c(u_f) f'(X_{t-}^{u_f}) - (\delta+\lambda) f(X_{t-}^{u_f}) \right. \\
&\quad + \lambda \int_0^{\rho(X_{t-}^{u_f}, u_f)} f(X_{t-}^{u_f} - r(y,u_f))\, dF_Y(y) \\
&\quad \left. + \lambda \int_{\rho(X_{t-}^{u_f}, u_f)}^{\infty} w(X_{t-}^{u_f}, r(y,u_f) - X_{t-}^{u_f})\, dF_Y(y) \Bigg) dt \right] \\
&= \mathbb{E}_x \left[e^{-\delta(T_1\wedge\tau^{u_f})} f(X^{u_f}_{T_1\wedge\tau^{u_f}}) \mathbb{1}_{T_1\neq\tau^{u_f}} \right] - f(x) \\
&\quad + \mathbb{E}_x \left[\int_0^{T_1} e^{-\delta t} \lambda \int_{\rho(X_{t-}^{u_f}, u_f)}^{\infty} w(X_{t-}^{u_f}, r(y,u_f) - X_{t-}^{u_f})\, dF_Y(y) dt\, \mathbb{1}_{T_1=\tau^{u_f}} \right] \\
&= \mathbb{E}_x \left[e^{-\delta T_1} f(X_{T_1}^{u_f}) \mathbb{1}_{T_1<\tau^{u_f}} \right] + \mathbb{E}_x \left[e^{-\delta\tau^{u_f}} f(0) \mathbb{1}_{T_1>\tau^{u_f}} \right] - f(x) \\
&\quad + \mathbb{E}_x \left[\int_0^{T_1} e^{-\delta t} \lambda \int_{\rho(X_{t-}^{u_f}, u_f)}^{\infty} w(X_{t-}^{u_f}, r(y,u_f) - X_{t-}^{u_f})\, dF_Y(y) dt\, \mathbb{1}_{T_1=\tau^{u_f}} \right].
\end{aligned}$$

We now use the compensation theorem

$$\mathbb{E}_x \left[\int_0^{T_1} \lambda e^{-\delta t} H_t\, dt \right] = \mathbb{E}_x \left[\int_0^{T_1} e^{-\delta t} H_t\, dN_t \right] = \mathbb{E}_x \left[e^{-\delta T_1} H_{T_1} \right],$$

where λ is the intensity of the counting process N_t, for the previsible process

$$H_t := \int_{\rho(X_{t-}^{u_f}, u_f)}^{\infty} w(X_{t-}^{u_f}, r(y, u_f) - X_{t-}^{u_f})\, dF_Y(y).$$

Taking $u_f(0)$ as in the statement of the lemma yields

$$\begin{aligned} f(x) = \mathbb{E}_x \left[e^{-\delta T_1} f(X_{T_1}^{u_f}) \mathbb{1}_{T_1 < \tau^{u_f}} \right] + \mathbb{E}_x \left[e^{-\delta T_1} w(X_{T_1-}^{u_f}, |X_{T_1}^{u_f}|) \mathbb{1}_{T_1 = \tau^{u_f}} \right] \\ + \mathbb{E}_x \left[e^{-\delta \tau^{u_f}} w(0,0) \mathbb{1}_{T_1 > \tau^{u_f}} \right], \end{aligned}$$

because $\mathbb{1}_{T_1 > \tau^{u_f}} = 0$ if $c(u_f(0)) \geq 0$. So we showed $f \geq \mathcal{G}f$.

On the other hand

$$\begin{aligned} \mathcal{G}f(x) &= \inf_{u \in \mathcal{U}} \left\{ \mathbb{E}_x \left[e^{-\delta T_1} f(X_{T_1}^u) \mathbb{1}_{T_1 < \tau_x^u} \right] + \mathbb{E}_x \left[e^{-\delta T_1} w(X_{T_1-}^u, |X_{T_1}^u|) \mathbb{1}_{T_1 = \tau_x^u} \right] \right. \\ &\quad \left. + \mathbb{E}_x \left[e^{-\delta \tau_x^u} w(0,0) \mathbb{1}_{T_1 > \tau_x^u} \right] \right\} \\ &= \inf_{u \in \mathcal{U}} \left\{ f(x) + \mathbb{E}_x \left[\int_0^{T_1 \wedge \tau^u} e^{-\delta t} \left(c(u_t) f'(X_{t-}^u) - (\delta + \lambda) f(X_{t-}^u) \right. \right. \right. \\ &\quad \left. \left. \left. + \lambda \int_0^{\rho(X_{t-}^u, u_t)} f(X_{t-}^u - r(y, u_t))\, dF_Y(y) \right) dt \right] \right. \\ &\quad + \mathbb{E}_x \left[e^{-\delta \tau_x^u} w(0,0) \mathbb{1}_{T_1 > \tau_x^u} \right] \\ &\quad \left. + \mathbb{E}_x \left[\int_0^{T_1 \wedge \tau^u} e^{-\delta t} \lambda \int_{\rho(X_{t-}^u, u_t)}^{\infty} w(X_{t-}^u, r(y, u_t) - X_{t-}^u)\, dF_Y(y)\, dt \right] \right\} \\ &\geq f(x) \end{aligned}$$

where we again used the compensation theorem for the last expression and the last inequality follows from the HJB-equation.

□

Remark 3.3.3. *In the above Lemma, we write $u_f(x)$ to indicate that we are working with a Markov control, i.e. solely dependent on the current state. Furthermore, the described choice of $u_f(x)$ happens in a measurable way, as can be seen from arguments similar to those of Lemma 2.12 in Schmidli (2008).*

Remark 3.3.4. *The aim of this section is to show that the function f with the properties of Lemma 3.3.2 actually is the value function V. So demanding that $f(0) \leq w(0,0)$ is a natural condition since it is certainly fulfilled by V. The definition of u_f is also very intuitive as can be seen by the following consideration. Having $u_f(0) = u^*$ means a negative premium in zero and therefore the process can make the transition from "alive" to "ruined" without a jump. Interpreting the process $X_t^{u_f}$ as a piecewise deterministic Markov process (PDMP), this means the active boundary Γ is not empty here (see Chapter 2 Section 24 in Davis (1993)). In the theory of PDMPs, this goes along with*

the additional boundary condition $f(0) = w(0,0)$. Because smooth ruin is usually not considered in reinsurance scenarios where the ruin probability or dividend payments are to be optimized, it is an interesting feature of our model to (potentially) have $\Gamma \neq \emptyset$. For more details on this subject, we refer to Chapter 11 of Rolski et al. (1999).

The following Theorem is an immediate consequence of Lemmas 3.3.1 and 3.3.2 combined with Banach's fixed point theorem. It is also the central statement of this section as it establishes the HJB-equation as the crucial tool for finding the value function.

Theorem 3.3.5. *In the function space $\mathcal{C}^{+,b}[0,\infty)$, the value function V is the unique fixed point of $\mathcal{G}$ and hence it is also the unique solution to the HJB-equation.*

3.4. Numerical Examples

Following the results in the previous section, we can construct the value function by finding a solution to the Hamilton-Jacobi-Bellman equation. Our method of choice was *policy iteration* (for a detailed review of applicable methods see e.g. Kushner and Dupuis (2013)). In a first step, we discretized the interval $[x_0, x_N]$ where we want to find the solution. Then we started with an arbitrary strategy $u^{(0)}$ and used Monte-Carlo techniques to find the values for $\Phi^{u^{(0)}}(x_0)$ and $\Phi^{u^{(0)}}(x_N)$. For simplicity we chose $u^{(0)} = \mathring{u}$ i.e. the strategy of no reinsurance. Knowledge of these boundary values then enabled us to numerically solve the integro-differential equation that is given by the Feynman-Kac type equation

$$\begin{aligned} 0 = & c(u^{(0)}) \left(\Phi^{u^{(0)}}\right)'(x) - (\delta + \lambda)\Phi^{u^{(0)}}(x) \\ & + \lambda \int_0^{\rho(x,u^{(0)})} \Phi^{u^{(0)}}(x - r(y, u^{(0)}))\, dF_Y(y) \\ & + \lambda \int_{\rho(x,u^{(0)})}^{\infty} w(x, r(y, u^{(0)}) - x)\, dF_Y(y) \end{aligned}$$

as it is derived in Theorem 11.2.3 of Rolski et al. (1999). Here, we used a finite differences approach. Having calculated $\Phi^{u^{(0)}}(x)$ for all x on the grid corresponding to $[x_0, x_N]$ in this manner, we look for an improving strategy by taking

$$\begin{aligned} u^{(1)}(x) = \arg\min_{u \in \mathcal{U}} \Bigg\{ & c(u)(\Phi^{u^{(0)}})'(x) - (\delta + \lambda)\Phi^{u^{(0)}}(x) \\ & + \lambda \int_0^{\rho(x,u)} \Phi^{u^{(0)}}(x - r(y,u))\, dF_Y(y) \\ & + \lambda \int_{\rho(x,u)}^{\infty} w(x, r(y,u) - x)\, dF_Y(y) \Bigg\}. \end{aligned}$$

Now we repeat the procedure with $u^{(1)}$ in place of $u^{(0)}$ to construct $u^{(2)}, u^{(3)}, \ldots$ until no significant improvement can be achieved anymore.

For referencing, we chose similar parameters as in Schmidli (2008) Chapter 2 for the risk model. That is, we set the Poisson intensity λ to 1 and the interval under consideration to $[0, 14]$. The reinsurance shall be of proportional type, i.e. the retention function is given as $r(y, u) = u \cdot y$ for $u \in [0, 1]$. Furthermore, we calculated the premiums $c(u)$ following the expected value principle with the cedent's safety loading denoted by η and the reinsurer's safety loading θ. So

$$c(u) = \lambda\beta(\eta - \theta + u(1 + \theta)),$$

where β denotes the expected claim height. In all examples, we set $\eta = 0.5$ and $\theta = 0.7$.

3.4.1. Exponential Claims

First, we want to consider exponentially distributed claims. Setting the expected claim height to 1, this means $F_Y(y) = 1 - e^{-y}$. We start with the very simple penalty function $w_1(x, y) = 1$, so we want to minimize the discounted ruin probability. This exact setting was treated in Schmidli (2008) for $\delta = 0$. We undertook the calculation for the case $\delta = 0.05$ to see the effect of the discount factor on value function and strategy. The resulting strategy and the first 5 iterations of Φ are shown in Figures 3.1 and 3.2 respectively. While Figure 3.1 shows clear resemblance to the undiscounted case in Schmidli (2008), we see that in Figure 3.2, the difference between the first 3 Gerber Shiu functions (blue, red, yellow) is still significant, whereas there is almost no difference anymore between functions 3, 4 and 5 (depicted yellow, purple and green).

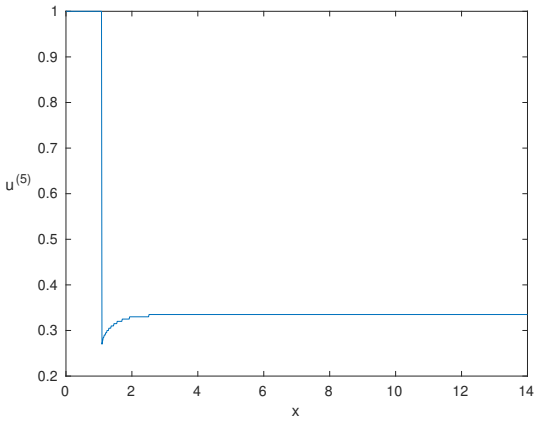

Figure 3.1.: Optimal strategy for exponential claims

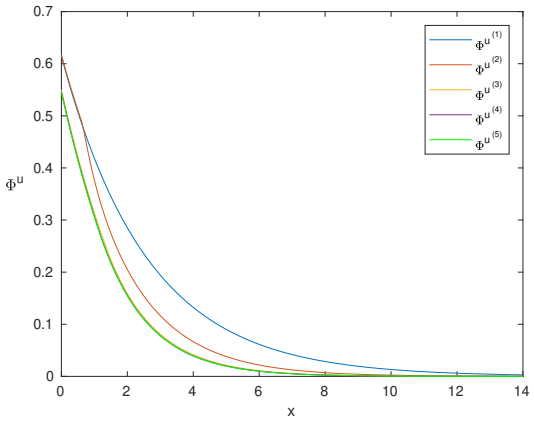

Figure 3.2.: functions $\Phi^{u^{(1)}}$ to $\Phi^{u^{(5)}}$.

To show the flexibility of our approach we want to consider a more general penalty function. So we will now use $w_2(x, y) = \min(10^{10}, (x + 0.5)(y + 1)^2)$ and also increase the discounting rate to $\delta = 0.1$. This choice of penalty function might seem arbitrary at first, but making the penalty actually depend on the surplus prior to- and deficit at ruin will trigger the incentive for smooth ruin in some situations. As before, we used

policy iteration and stopped when improvements fell under a predefined level. In Figure 3.4, we plotted the corresponding value of the HJB-equation. In the optimum this value is zero, values close to zero indicate a good approximation. The optimal strategy can be seen in Figure 3.3 where the red line is drawn at 0.1176, the zero of the premium function $c(u)$. So for $u < 0.1176$, the total premiums are negative.

The resulting strategy is particularly interesting since it leads to smooth ruin. That

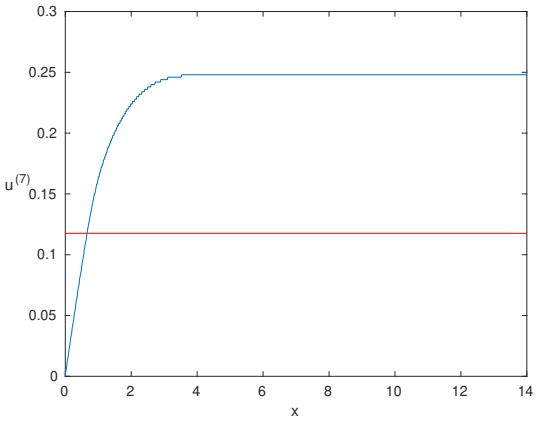

Figure 3.3.: Optimal strategy after 7 iterations, sign change at red line.

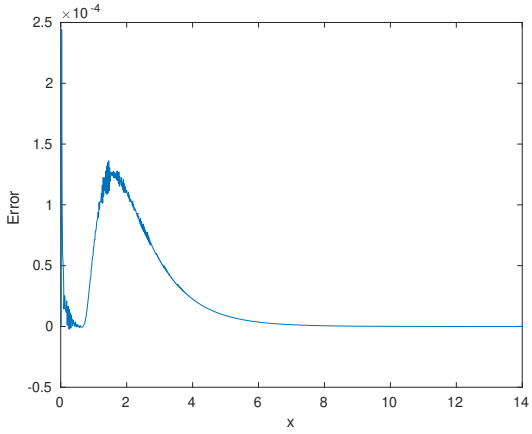

Figure 3.4.: Value of HJB-equation.

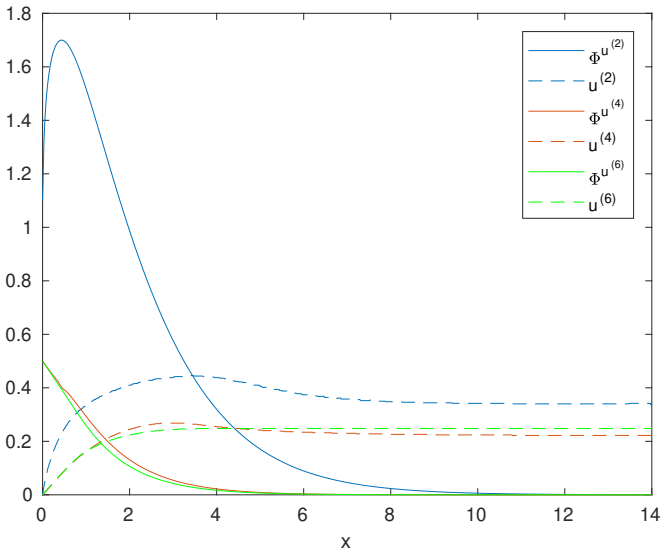

Figure 3.5.: functions $\Phi^{u^{(2)}}$, $\Phi^{u^{(4)}}$ and $\Phi^{u^{(6)}}$ along with strategies $u^{(2)}$, $u^{(4)}$ and $u^{(6)}$.

means, for low reserve values, the insurer prefers deliberately terminating the business and paying the comparably low penalty $w_2(0,0) = 0.5$ instead of taking the risk of

a much higher penalty. In Figure 3.5, we show the second (blue), fourth (red) and sixth (green) cost function with the respective minimizing strategies (dashed lines in the corresponding colors).

3.4.2. Pareto Claims

In insurance mathematics, a particular interest lies in the study of heavy-tailed distributions. To account for that, we also investigated the case of Pareto distributed claims. For w_1, that is the discounted ruin probability, we chose the claim distribution $F_Y(x) = 1 - (x+1)^{-2}$, resulting again in an expected claim height of 1. This claim distribution was also used in Schmidli (2008). The resulting strategy is shown in Figure 3.6, while Figure 3.7 gives again the first 5 cost functions in the order blue, red, yellow, purple and green.

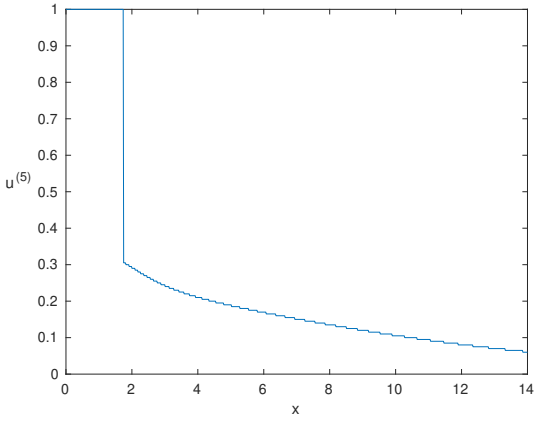

Figure 3.6.: Optimal strategy for Pareto claims

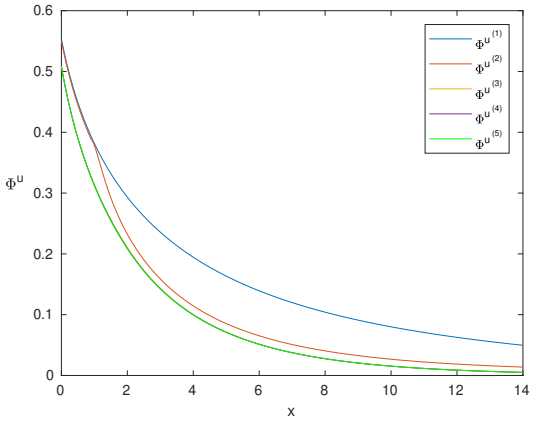

Figure 3.7.: functions $\Phi^{u^{(1)}}$ to $\Phi^{u^{(5)}}$.

As for the exponential case, we also want to find the optimal strategy for Pareto distributed claims and the penalty function w_2. Since the second moment for Pareto distributions exists only for shape parameters greater than 2, we chose the claim height distribution $F_Y(y) = 1 - (1+y)^{-3}$. In Figure 3.8, we again added the red line at the zero of $c(u)$. Note that on the whole interval the optimal strategy leads to negative premiums. This can be explained by the heavy tails of the Pareto distribution. At no level of the reserve does the chance to survive but under the risk of a potentially heavy ruin, outweigh the very moderate penalty of $w_2(0,0)e^{-\delta\tau} < 0.5$. In Figures 3.10 and 3.11, we also plotted the second to fifth iteration of the value function resp. the corresponding strategy. Note that in Figure 3.10, the earlier iterations are not monotone, whereas $\Phi^{u^{(4)}}$ and later iterations show the postulated monotonicity.

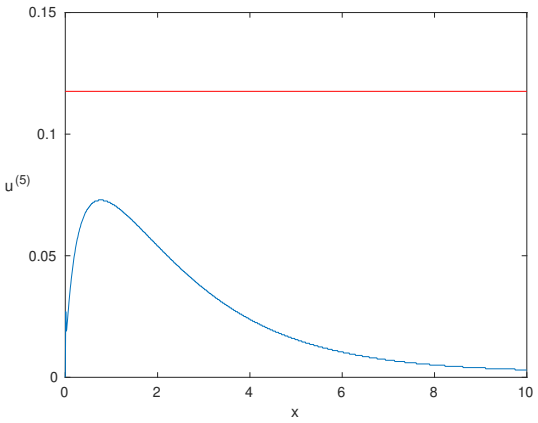

Figure 3.8.: Optimal strategy for Pareto claims

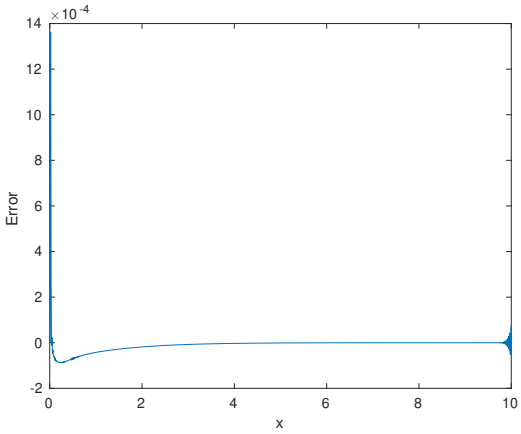

Figure 3.9.: HJB error.

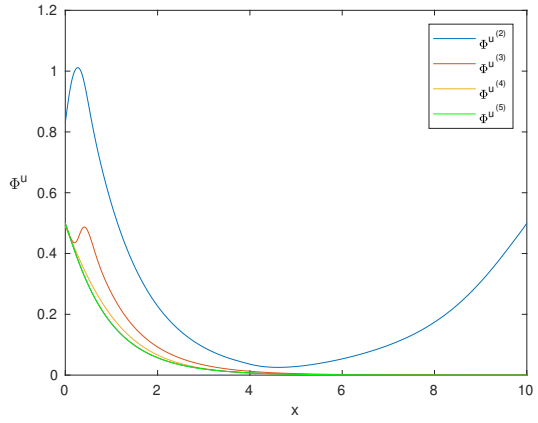

Figure 3.10.: functions $\Phi^{u^{(2)}}$ to $\Phi^{u^{(5)}}$.

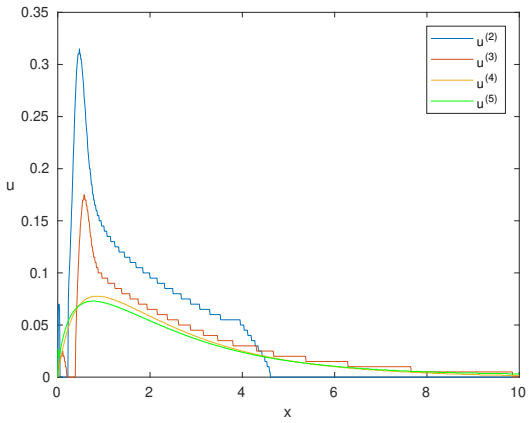

Figure 3.11.: strategies $u^{(2)}$ to $u^{(5)}$.

3.4.3. Excess of Loss Reinsurance

In the examples we have treated so far, the reinsurance was always of proportional type, i.e. we had the retention function $r_p(y,u) = u \cdot y$ with $u \in [0,1]$. Our method however is not restricted to this particular case. In the following we want to consider a prominent example of a non-proportional model, which is excess of loss reinsurance. Here the retention function is given as $r_{XL}(y,u) := \min\{y,u\}$ with $u \in [0,\infty]$. So the cedent always pays the part of the claim that is below u but any amount exceeding the threshold u is fully covered by the reinsurer. In this setting $u = 0$ means full reinsurance, while $u = \infty$ corresponds to not buying any reinsurance at all.

We used again the Pareto distribution $F_Y(y) = 1 - (1+y)^{-3}$ for the claim heights, the Poisson parameter $\lambda = 1$ for the claim arrivals and the discounting factor $\delta = 0.1$. As penalty function we chose $w_3(x,y) := x + y + 10$ which is the sum of surplus prior to ruin and deficit with a fixed penalty added for ruin. We calculated the premiums again by the expected value principle with the same safety loadings as before. The approximated optimal strategy after 8 iterations is shown in Figure 3.12 and the controlled Gerber-Shiu functions for iterations 1, 5, 7 and 8 are shown in Figure 3.13. The orange line in Figure 3.12 represents the value $u = \infty$, so here it is optimal to buy no reinsurance. Since the resulting premium $c(u)$ is positive for all values of the optimal strategy, we did not add a red line to Figure 3.12. The strategy has a similar shape as in Hipp and Vogt (2003), where the ruin probability is to be minimized. This is reasonable, since the constant 10 that is added in our penalty term is relatively high for the chosen parameters, hence the objective is essentially to prevent ruin.

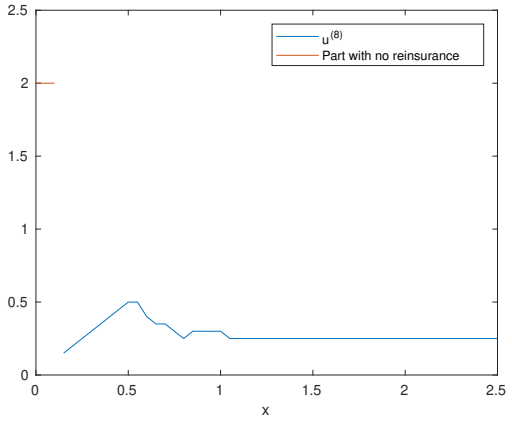

Figure 3.12.: Optimal strategy

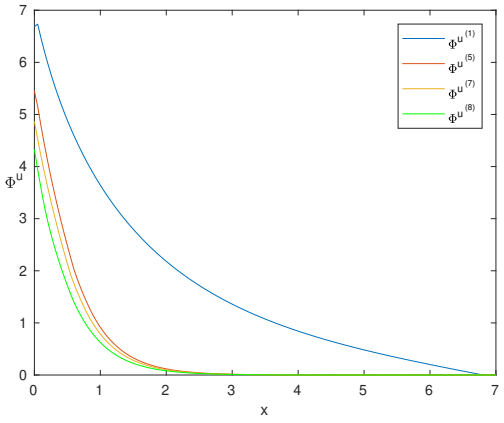

Figure 3.13.: Gerber-Shiu functions.

3.4.4. A Note on the Numerics

The calculations that were undertaken for this section turned out to be more laborious than expected. While some cases, like exponential claims without or with low discounting factor or Pareto claims without discounting factor didn't make much trouble, other

cases, namely the more general penalty function w_2 in combination with discount rates, proportional reinsurance and Pareto claims were quite demanding. The reason for this is that the finite differences approach in these cases was extremely sensitive to the right starting value, indeed to an extent where MC techniques could not provide the needed accuracy anymore. Relying on IDE solvers that treat the problem in a more continuous way is not immediately possible, since strategies crossing the zero of $c(u)$ result in singularities in the involved ODE terms.
The method that brought the best results was an individually chosen mix of central and backwards differences combined with a MC simulation for an initial guess, followed by a somewhat manual bisection technique to provide the correct initial values.

3.4.5. Asymptotic Behaviour

We also investigated the question of the asymptotically optimal strategy. In the case of exponentially distributed (that means light-tailed) claims, it is straightforward to proceed as in Hald and Schmidli (2004). One has to keep in mind though that the presence of a discount factor δ changes the associated Lundberg equation to

$$\lambda \hat{m}_Y(\alpha) = 1 + \frac{\delta}{\lambda} + \frac{\alpha c}{\lambda}, \tag{3.6}$$

where $\hat{m}_Y(\alpha) = \mathbb{E}\left[e^{\alpha Y}\right]$ is the moment-generating function of the claim height distribution F_Y. The positive solution γ for which (3.6) becomes zero (if such a solution exists) is usually called the (generalized) adjustment coefficient. Now consider the Cramér-Lundberg approximation for $\Psi(x)$, the ruin probability with initial value x, which reads

$$\lim_{x\to\infty} \Psi(x) e^{x\gamma} = C_\delta \tag{3.7}$$

for some constant C_δ. From (3.7), it becomes clear that maximizing the adjustment coefficient by means of the reinsurance parameter will lead to the maximally fast asymptotic decay rate for the (discounted) ruin probability. This approach goes back to Waters (1983). So if we now assume a constant reinsurance strategy u, proportional reinsurance and premiums calculated by the expected value principle as above, equation (3.6) becomes

$$\lambda(\hat{m}_Y(u\gamma) - 1) - \delta - \underbrace{(\lambda\beta(1+\eta) - (1-u)(1+\theta)\lambda\beta)}_{=c(u)}\gamma = 0.$$

Concavity arguments, differentiating and recollecting terms as in Schmidli (2008) now yield the following asymptotically optimal control strategy.

$$u^{asy} = \frac{\lambda(\theta-\eta)\left(1-\sqrt{\frac{1}{1+\theta}}\right)}{\delta + 2\lambda(1-\sqrt{1+\theta}) + \theta\lambda}. \tag{3.8}$$

It is, perhaps, a little surprising that for exponential claims, the optimal strategy does not depend on β, the expectation of F_Y. If we calculate u^{asy} for $\delta = 0.05$ and $\lambda = 1, \eta = 0.5$ and $\theta = 0.7$ as above, we get $u^{asy}_{0.05} = 0.3275$ which is also indicated by Figure 3.1.

Another very interesting fact is that the asymptotically optimal strategy does not depend on the actual penalty function w as well. This might seem counterintuitive at first, but using material from Asmussen and Albrecher (2010), we see that for a constant strategy u

$$\lim_{x\to\infty} \Phi^u(x)e^{\gamma(u)x} = C_{\delta,w}.$$

So only the constant $C_{\delta,w}$ depends on the penalty function w, while the asymptotic behaviour is governed by the adjustment coefficient just as in the case of the discounted ruin probability. The reason for this is of course the indicator function for ruin in the Gerber-Shiu function; for high starting values, ruin is just unlikely to occur. Evaluating (3.8) for $\delta = 0.1$ yields the asymptotically optimal strategy $u^{asy}_{0.1} = 0.2423$ which is confirmed by Figure 3.3. So for the same values of δ, Figures 3.1 and 3.3 converge to the same level.

Part II.

Dependence Modeling

4. Overview and Connection to QMC Theory

Dependence modeling as a mathematical discipline is a part of probability theory with the focus on analyzing and describing the structures that arise in the interplay of several quantities. This encompasses coming up with concepts to measure certain types of dependence and also finding constellations that are extreme in these metrics. The applications of this field can be of purely theoretical nature, however, many notions of dependence modeling are widely used in practice and have become very popular in financial mathematics in particular.

4.1. Quantitative Risk Management

With risk theory being mainly focused on insurance mathematics, quantitative risk management is the term commonly used when talking about a general financial institution, or a bank in particular. In that sense, the applied branch of dependence modeling can be seen as a part of quantitative risk management (QRM). Somewhat self explanatory, and similar to risk theory for insurance companies, the aim of QRM is to identify, quantify and manage the risks faced in the context of financial markets.

Since a portfolio, by the very meaning of the word, consists of various different positions, it is crucial to have an idea about the dependence between the single assets in order to make a statement about the portfolio's risk. It is therefore desirable to have a model for these dependencies. Of course, taking the single assets to be random variables $X_1, \dots, X_n$, a complete description of the dependence is given by the common distribution function $F_{X_1,\dots,X_n}$. However, as it turns out, the distribution function itself is composed of two parts: the marginal distributions, that is the distribution F_{X_i} for every $i \in \{1, \dots, n\}$ and the dependence structure that binds the assets together. This binding component is called a copula, a notion that we want to specify just a little further.

4.1.1. Copulas

Indeed, a copula is just a multidimensional distribution function, whose marginal distributions are uniform. This is formalized in the following definition

Definition 4.1.1. *A function* $C : [0,1]^d \to [0,1]$ *that fulfills*

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d) \quad u_1, \dots, u_d \in [0,1]$$

for random variables $U_1, \ldots, U_d \sim \mathcal{U}([0,1])$ *is called a d-copula.*

The importance of copulas in dependence modeling stems from the observation that they can be extracted as "building blocks" from any multidimensional distribution. This goes back to a famous result of Sklar (1959).

Theorem 4.1.2 (Sklar)**.** *A function* $F : \mathbb{R}^d \to [0,1]$ *is the distribution function of some random vector* $(X_1, \ldots, X_d)$ *if and only if there is a d-copula* C *and univariate distribution functions* $F_1, \ldots, F_d : \mathbb{R} \to [0,1]$ *such that*

$$C(F_1(x_1), \ldots, F_d(x_d)) = F(x_1, \ldots, x_d)$$

for any $x_1, \ldots, x_d \in \mathbb{R}$.

Additionally, if the marginal distributions $F_1, \ldots, F_d$ are continuous, the corresponding copula C is unique.

So this means that any multivariate distribution function can be seen as and decomposed into univariate margins and a copula. This allows to treat marginal distributions and dependence structures in a separate way. In statistics, this can be used to determine the influence of marginals resp. dependence on certain features of the data. When modeling random vectors, Sklar's theorem justifies restricting to copulas, because any desired marginal distributions can simply be applied afterwards.

Another very useful feature of copulas is that they provide easy representations for the extreme cases of dependency which are perfect positive dependence, usually called *comonotonicity*, perfect negative dependence or *countermonotonicity* and independence. In the case of two random variables X_1, X_2, perfect positive (resp. negative) dependence means X_2 is a monotonously increasing (resp. decreasing) function of X_1. The corresponding two dimensional copulas are given as

$$\Pi(x_1, x_2) = x_1 x_2 \qquad M(x_1, x_2) = \min(x_1, x_2) \qquad W(x_1, x_2) = \max(x_1 + x_2 - 1, 0).$$

for $x_1, x_2 \in [0,1]$. Here $(X_1, X_2) \sim \Pi$ is a vector of independent components, $(X_1, X_2) \sim M$ is a vector of comonotone components and $(X_1, X_2) \sim W$ is a vector of countermonotone components. Often, Π is called the independence copula and M and W are called the upper resp. lower Fréchet-Hoeffding bound. Note that while a generalization of independence and comonotonicity to higher dimensions is straightforward, it is not at all clear, what countermonotonicity should mean for three or more random variables. Indeed, finding extreme negative dependence concepts in higher dimensions is an active field of research and we will also get back to this question later in the thesis.
But also in between these extremal cases, copulas ease the description of dependence between two random variables. If we consider the probably most widespread measure of association, the linear correlation coefficient, or sometimes called Pearson's coefficient, defined as

$$\mathrm{Cor}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}}$$

we quickly realize it has some shortfalls. Most prominent among these is that it only measures linear correlation and is also not invariant under nonlinear transformations of X_1 or X_2, so it should not be used if another shape of connection is assumed. Furthermore, a correlation value of, say, 0.3 does not give much information about the strength of the correlation, since the absolute value of Pearson's coefficient is influenced by the marginal distributions of X_1 and X_2. In addition, although Pearson's coefficient is normed between -1 and 1, it is not a priori clear that these extreme values can be attained for any fixed marginal laws. The reason for most of these features is that linear correlation mixes aspects of dependence and marginal distributions, so they can be avoided when using measures of association that depend only on the copula. These measures are called rank correlations or concordance measures, we want to present two of them, namely Spearman's ρ and Kendall's τ. For two random variables X_1, X_2 with distribution functions F_1 and F_2, Spearman's ρ is simply defined as $\rho := \mathrm{Cor}(F_1(X_1), F_2(X_2))$, so, following Sklar, it is the Pearson coefficient of the underlying copula. To define Kendall's τ, for a random vector (X_1, X_2), we take another random vector (Y_1, Y_2) that is an independent copy of (X_1, X_2) and set $\tau := \mathbb{E}[\mathrm{sign}((X_1 - Y_1)(X_2 - Y_2))]$. Though ρ and τ are certainly the most widespread concordance measures, there are many more and defining good new measures for dependence tailored to certain situation is still an active field of interest. For a detailed account on these measures and a discussion of desirable properties, we refer to Nelsen (2007).
Moreover, because many copulas can be simulated easily, they are very well suited for procedures that use Monte-Carlo techniques and in applications it is usually viable to try out a variety of copula models to achieve a good fit on the data at hand.

4.1.2. Controversy About Copulas

Over the past decades, a broad variety of parametrized families of copulas has been defined that serve as blueprints for different kinds of dependence structures. From relatively simple Gaussian copulas that capture the dependence among components of Gaussian vectors to more sophisticated structures like the Gumbel copula, which is better suited to model tail dependencies there is a copula tailored to almost every situation.

Having such a rich toolbox of copulas that just require some parameter fitting to be readily applicable is, of course, very appealing to practitioners. As a result of this, usage of copulas boomed in many fields, especially financial mathematics and in particular credit risk modeling. Unfortunately, be it due to a lack of insight about the mathematical foundations or the desire to keep the underlying models as easy as possible, many analysts oversimplified the involved structures. In particular, the probabilities of joint defaults were often systematically underestimated which turned out to be fatal in the 2008 financial crisis. For this reason, copulas are sometimes seen controversial in financial literature and were even blamed in mainstream media (for a discussion on this topic, see e.g. Brigo et al. (2010)).
From a mathematical point of view however, using the concept of copulas offers many advantages, as long as sufficient care and understanding of the matter are applied.

4.2. Intersections with Uniform Distribution Theory

As it is crucial for (Q)MC techniques to be able to fill out the unit interval in a particularly uniform way, the theory of uniformly distributed points and sequences is a central aspect of Monte-Carlo and especially Quasi-Monte-Carlo simulation. Therefore, uniform distribution theory was also in the focus of the special research program (SFB) for Quasi-Monte-Carlo methods which the author of this dissertation is affiliated with.

The basic notion providing the bridge to copulas and dependence modeling is that of uniformly distributed sequences. A sequence $(x_n)_{n\in\mathbb{N}} \in [0,1)^{\mathbb{N}}$ is called uniformly distributed if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[a,b)}(x_n) = b - a$$

for all intervals $[a, b) \subset [0, 1)$.

Fialová and Strauch (2011) showed that for uniformly distributed sequences $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ and a continuous function $c : [0,1]^2 \to \mathbb{R}$, it holds

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} c(x_n, y_n) = \int_{[0,1)^2} c(x,y)\gamma(dx,dy), \tag{4.1}$$

with γ being a copula. Motivated by this result, Hofer and Iacò (2014) asked for maximizing and minimizing copulas, for (4.1). They were able to link this problem to optimal transport theory and its discrete analogue, the assignment problem. Doing this, they realized, they can find a solution to this optimization problem in a particular class of copulas, the *Shuffles of Min* and could find such a solution using convex optimization techniques. In a subsequent paper, Iacò et al. (2015) carried out a deeper structural analysis, including convergence results and approximations in number theory.
Aside from the structural analogies between copulas and QMC theory that were outlined above, some work has also been done on connecting these topics by using QMC methods in sampling algorithms for copulas. A reference would be Cambou et al. (2017).

5. Bounds on Integrals with Respect to Multivariate Copulas

From the perspective of number theory, the results of Fialová and Strauch (2011) are interesting, because they give information about the "average" value of a function applied to two infinite, uniform sequences and link the discrete setting to a finite integral. Looking through the eyes of probability theory, the integral on the right hand side of (4.1) is an expectation with respect to the 2 dimensional copula γ. So the method from Iacò et al. (2015) can be used to find maximal and minimal values for the expectation of a function of two random variables, when varying over all possible dependence structures.

Indeed, since information about marginal distributions of a random vector is much easier to acquire than about the dependence relation among the components, the assumption of known margins but unknown multidimensional distribution function is a common scheme in quantitative risk management. To assure stability, companies working in the finance sector, in particular banks and insurance companies, have to abide to certain regulations, commonly known as the Basel framework for banks or Solvency for insurers. In the aftermath of the 2008 financial crisis, many of these safety regulations have been enhanced. So a proper risk quantification is not only in the own interest of any financial institution, but also prescribed by law. A part of these frameworks is analyzing and hedging worst case scenarios. For risk measures that can be written as an expectation, this is exactly the maximization problem studied in Iacò et al. (2015).

Of course, having a method that is restricted to 2 dimensional random vectors might not be satisfying in practice. However, as it turns out, the transition from 2 to higher dimensions constitutes a structural break, caused by the higher complexity class of the multidimensional assignment problem. So the generalization to higher dimensions was left as an open problem in Iacò et al. (2015), that was eventually solved in Preischl (2016). We present these results in the following.

The remainder of this chapter, that is sections 5.1 through 5.4, was published as Preischl (2016) and is therefore adopted verbatim, with only one reference, namely Bernard et al. (2018), added to Remark 5.3.6.

5.1. Introduction

A multidimensional distribution function with uniform marginals is called a *copula*. In contrast to the simplified approach of quantifying risk and dependence by single numbers like Spearman's ρ or Kendall's τ, modeling with copulas makes it possible to describe and encapsulate the entire dependence structure between random variables. On the other hand, an obvious downside of copulas is that, unlike simple concordance measures, they are often hard to treat analytically, especially in dimensions higher than two. Hence, instead of modeling the actual dependence structure, it is naturally interesting to ask for a "worst case" respectively a "best case" behaviour. In this paper, we propose a flexible method to approximate those extremal cases.

Assume that we are given a d-dimensional random vector $(X_1, \dots, X_d)$ and a function $f : \mathbb{R}^d \to \mathbb{R}$ that describes the quantity associated with $(X_1, \dots, X_d)$ which we wish to optimize. We further assume that the marginal distributions of $X_1, \dots, X_d$ are known and the dependence structure (i.e. the common distribution function) is completely unknown. This assumption is called *dependence uncertainty* and it is widely used in applications, mainly because compared to finding the dependence structure, information about the marginal laws can be relatively easily obtained.
By Sklar's Theorem it is always possible to reduce this maximization respectively minimization to a similar problem involving uniformly distributed random variables $X_1, \dots, X_d$, see e.g. Nelsen (2007). Therefore, we are justified in restricting our focus to finding copulas $C_{\min}$ and $C_{\max}$ such that

$$\int_{[0,1]^d} f(x_1, \dots, x_d) dC_{\min} \leq \int_{[0,1]^d} f(x_1, \dots, x_d) dC \tag{5.1}$$

and

$$\int_{[0,1]^d} f(x_1, \dots, x_d) dC \leq \int_{[0,1]^d} f(x_1, \dots, x_d) dC_{\max} \tag{5.2}$$

for all d-dimensional copulas C.

Equations (5.1) and (5.2) are special cases of the *Monge-Kantorovich* problem. This problem is very well studied in the case $d = 2$, however, due to its complexity, most analytic approaches to the Monge-Kantorovich problem in higher dimensions are restricted to particular situations. For example, one of Rüschendorf's many contributions to this field considered the case where f is a Δ-monotone function (Rüschendorf, 1980).
A more flexible, numerical take on this optimization problem that had a significant impact in recent years is the *rearrangement algorithm*, introduced by Puccetti and Rüschendorf (2012). This algorithm is impressively efficient in approximating the desired bounds even in high dimensions and thus suffices for most real world applications. The price for this is that it only works when f is a supermodular function and that convergence

is not guaranteed. However, the cases where the algorithm does not converge are quite pathological and can be circumvented in practice. For a detailed description of how the rearrangement algorithm can be used to tackle optimization problems and also examples of non-convergence, see Puccetti and Rüschendorf (2015).
In two dimensions, Hofer and Iacò (2014) combined the spirit of optimization theory with rigorous structural considerations and developed an algorithm that converges to the minimal respectively maximal values of equations (5.1) and (5.2) for any continuous function f. Their method connects optimality with a particular class of copulas, the Shuffles of *Min*. We state their main results in Theorem 5.2.2 in section 2.

5.2. Mathematical Foundations

As stated previously, a d-copula is a d-dimensional distribution function on $[0,1]^d$ with uniform marginals. Every d-copula C defines a measure μ_C on $([0,1]^d, \mathcal{B}([0,1]^d))$ which is d-fold stochastic, i.e., it fulfills

$$\mu_C(\underbrace{[0,1]\times[0,1]\times\cdots\times[0,1]}_{i-1 \text{ times}} \times A \times \underbrace{[0,1]\times[0,1]\times\cdots\times[0,1]}_{d-i \text{ times}}) = \lambda(A), \tag{5.3}$$

for all $i = 1,\dots,d$ and any Borel set $A \subset [0,1]$. Conversely, every d-fold stochastic measure on $([0,1]^d, \mathcal{B}([0,1]^d))$ defines a copula. We write $\mathcal{C}^d$ for the set of all d-copulas.

We already mentioned that there is a class of copulas, which is closely related to assignment problems, the Shuffles of *Min* (or Shuffles of M). As the name suggests, these are obtained by a suitable rearrangement of the probability mass of the upper Fréchet-Hoeffding bound, or *Min* copula, $M(x_1,\dots,x_d) := \min(x_1,\dots,x_d)$. In two dimensions, C is a Shuffle of *Min* parametrized by $n \in \mathbb{N}$, a permutation π on $\{1,\dots,n\}$ and a function $\omega : \{1,\dots,n\} \to \{-1,1\}$ if C distributes the mass $\frac{1}{n}$ uniformly spread along the diagonal respectively antidiagonal of $[\frac{i-1}{n}, \frac{i}{n}] \times [\frac{\pi(i)-1}{n}, \frac{\pi(i)}{n}]$ whenever $\omega(i) = 1$ respectively $\omega(i) = -1$. The original, two dimensional definition is slightly more general and goes back to Mikusinski et al. (1992). In higher dimensions, there are several versions of Shuffles of *Min* (see e.g. Durante and Sánchez (2012) for a discussion). A basic property of these Shuffles is that they are dense in the set of all copulas with respect to weak convergence. For more details and a survey of different metrics also see Durante and Sánchez (2012).

A concept which has proven very useful when solving two dimensional Monge-Kantorovich problems is that of *c-cyclical monotonicity*. It is a famous result in optimal transport theory that, under mild assumptions on c, a probability measure μ is optimal for the two dimensional Monge-Kantorovich problem if and only if it is concentrated on a c-cyclically monotone set. This optimality result follows from a dual formulation of the problem, for which we refer to the book of Villani (2008).

Finding a similar statement for dimensions higher than two has been an open problem

for many years. Beiglböck and Griessler (2014) and Griessler (2016) recently generalized c-cyclical monotonicity to arbitrary dimensions:

Definition 5.2.1 (Beiglböck and Griessler (2014) and Griessler (2016))**.** *Let $X_1, \dots, X_d$ be Polish spaces and define $E := X_1 \times \cdots \times X_d$. Let $c : E \to \mathbb{R}$ be Borel measurable. A set $\Gamma \subset E$ is called c-cyclically monotone if it fulfills one of the following conditions:*

(i) For any N and any points $(x_1^{(1)}, \dots, x_d^{(1)}), \dots, (x_1^{(N)}, \dots, x_d^{(N)}) \in \Gamma$ and permutations $\sigma_2, \dots, \sigma_d : \{1, \dots, N\} \to \{1, \dots, N\}$, one has

$$\sum_{i=1}^{N} c(x_1^{(i)}, \dots, x_d^{(i)}) \leq \sum_{i=1}^{N} c(x_1^{(i)}, x_2^{(\sigma_2(i))}, \dots, x_d^{(\sigma_d(i))}).$$

(ii) Any finite measure α concentrated on finitely many points in Γ is a (with respect to c) cost-minimizing transport plan between its marginals; i.e. if α' has the same marginals as α, then

$$\int c d\alpha \leq \int c d\alpha'.$$

They were also able to show that for any measurable cost function c, a measure μ which is optimal for the multidimensional Monge-Kantorovich problem is always concentrated on some c-cyclically monotone set.
Griessler (2016) recently showed the converse statement under more assumptions on c: If the cost function c is continuous and bounded by a sum of integrable functions, any measure which is concentrated on a c-cyclically monotone set is an optimal solution to the multidimensional Monge-Kantorovich problem.

Next we give a short overview of assignment problems. The mathematical formulation of a Linear Sum Assignment Problem is the following:

$$\min_{x \in \mathbb{R}^{n \times n}} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ij} \tag{5.4}$$

subject to

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i \in \{1, \dots, n\}, \tag{5.5}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \forall j \in \{1, \dots, n\}, \tag{5.6}$$

$$x_{ij} \in \{0, 1\}. \tag{5.7}$$

The matrix (a_{ij}) is also called the *objective function* and the set of all $x \in \mathbb{R}^{n \times n}$ which fulfill all the constraints is called the *feasible region.*
It is not hard to see that a Linear Sum Assignment Problem can equivalently be written in the form

$$\min_{\pi \in S_n} \sum_{i=1}^{n} a_{i\pi(i)},$$

where S_n denotes the set of all permutations of $\{1, \dots, n\}$. Although the feasible region of this problem is actually n^2-dimensional, with n being the number of objects, we will refer to this version of the assignment problem as the "two dimensional assignment problem (2-AP)" since one can interpret this as matching two different kinds of objects. The assignment problem is, at least for the two dimensional case, very well-studied.

We are now ready to state the main result from Hofer and Iacò (2014) that connects Shuffles of *Min* and assignment problems to integrals with respect to copulas.

Theorem 5.2.2. *Let f be a continuous function on $[0,1]^2$ and let the partition I^n for any n be given as*

$$I_{ij}^n := \left[\frac{i-1}{n}, \frac{i}{n}\right) \times \left[\frac{j-1}{n}, \frac{j}{n}\right) \quad \text{for } i, j = 1, \dots, n.$$

Then define

$$f_n^{max}(x_1, x_2) = a_{ij} := \max_{(x_1,x_2) \in I_{ij}^n} f(x, y) \qquad \forall (x_1, x_2) \in I_{ij}^n.$$

Now a copula C_n^{max} which fulfills

$$\int_{[0,1]^2} f_n^{max}(x_1, x_2) dC_n^{max} = \max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f_n^{max}(x_1, x_2) dC(x_1, x_2) \tag{5.8}$$

is given as a shuffle of Min with parameters $(n, \pi^*, 1)$ where π^* is the permutation which solves the assignment problem

$$\max_{\pi \in S_n} \sum_{i=1}^{n} a_{i\pi(i)}.$$

Moreover, the maximal value of (5.8) is given as

$$\max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f_n^{max}(x_1, x_2) dC(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} a_{i\pi^*(i)}$$

and it holds

$$\lim_{n \to \infty} \max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f_n^{max}(x_1, x_2) dC(x_1, x_2) = \max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f(x_1, x_2) dC(x_1, x_2).$$

Furthermore, Iacò et al. (2015) showed that the sequence of maximizers C_n^{max} converges, at least along some subsequence, to a maximizer $C_{\max}$ of the problem

$$\max_{C \in \mathcal{C}^2} \int_{[0,1]^2} f(x_1, x_2) dC(x_1, x_2).$$

5.3. Main Results

As we will see, some structural analogies are destroyed in the d-dimensional case, which is why a direct application of the method from Hofer and Iacò (2014) is not possible. For our main result, an extension of Theorem 5.2.2 to arbitrary dimensions along with a similar convergence result as in Iacò et al. (2015), we start by introducing the concept of a multidimensional assignment problem.

Define the index sets $\mathcal{I} := \{1, \dots, n\}^d$ and $\mathcal{I}_m^k := \{(i_1, \dots, i_d) \in \mathcal{I} : i_k = m\}$. The (axial) d-dimensional assignment problem (d-AP) on n items with objective function (a_i) is given as follows:

$$\min_{x \in \mathbb{R}^{n^d}} \sum_{i \in \mathcal{I}} a_i x_i \tag{5.9}$$

subject to

$$\sum_{i \in \mathcal{I}_m^k} x_i = 1, \qquad \forall m \in \{1, \dots, n\}, \quad \forall k \in \{1, \dots, d\}, \tag{5.10}$$

$$x_i \in \{0, 1\}. \tag{5.11}$$

Again, "d-dimensional" is meant with respect to the number of different object types. The feasible region in this case would actually be n^d-dimensional.

Theorem 5.3.1. *Let n be a positive integer and $f : [0,1]^d \to \mathbb{R}$ be constant on the cubes I_i^n with*

$$I_i^n := \left[\frac{i_1 - 1}{n}, \frac{i_1}{n}\right) \times \cdots \times \left[\frac{i_d - 1}{n}, \frac{i_d}{n}\right),$$

for $i = (i_1, \dots, i_d) \in \mathcal{I}$. So $f(x) = a_i$ for all $x \in I_i^n$.
Then a copula $C_{\min}$ which fulfills

$$\int_{[0,1]^d} f(x)\, dC_{\min} = \min_{C \in \mathcal{C}^d} \int_{[0,1]^d} f(x)\, dC(x), \tag{5.12}$$

distributes uniformly on each square of type I_i^n the probability mass equal to x_i^*/n, where $(x_i^*)_{i \in \mathcal{I}}$ is an optimal solution to the relaxed d-AP with respect to the objective function $(a_i)_{i \in \mathcal{I}}$. By this construction, $C_{\min}$ is a so-called checkerboard copula (Durante et al. (2015), Mikusiński and Taylor (2010)).

Here "relaxed" means that we are considering the continuous relaxation of the axial d-AP, i.e., we replace the integer constraint (5.11) by $x_i \geq 0$ for all i. Even though a seemingly subtle difference, this change yields an entirely different problem from the perspective of optimization theory.
Theorem 5.3.1 holds for any dimension d and indeed for the case $d = 2$ we get precisely the method proposed in Hofer and Iacò (2014). This follows from Birkhoff's theorem, which states that the two dimensional assignment problem is identical to its continuous

relaxation. However, Birkhoff's theorem does not hold for any dimension greater than two, which is why it is not possible to restrict the solution space to Shuffles of *Min* in higher dimensions. As a result, the optimizing copula that comes from the assignment problem will, in general, not be given as a Shuffle of *Min*. We refer to Burkard et al. (2009) for further details about assignment problems.

That means, for dimensions greater than two, the maximizer we find via this procedure will not have the nicely parametrized form that made Shuffles of *Min* so appealing. On the other hand, working with the relaxed assignment problem instead of the integer problem brings great advantages concerning computability. While the classical integer assignment problems are $\mathcal{NP}$ hard, their continuous relaxations lie in $\mathcal{P}$. Here the computational complexity is with respect to the number of objects that should be assigned, or in the context of copulas, the coarseness of the partition of $[0,1]^d$.

Proof of Theorem 5.3.1. By definition, the value of (5.12) is given as

$$\int_{[0,1]^d} f(x)dC(x) = \sum_{i\in\mathcal{I}} a_i \mu_C(I_i^n).$$

Notice that the value of (5.12) does not depend on how the copula C distributes mass inside of each cube I_i^n, but only on how much mass is placed on each I_i^n. Hence, we can write $x_i := \mu_{C_{\min}}(I_i^n)$, with $i \in \mathcal{I}$ and are left with the following optimization problem:

$$\min \sum_{i\in\mathcal{I}} a_i x_i.$$

However, we still must encode constraints that ensure that the mass distribution x_i actually yields a copula. We recall that there is a one to one correspondence between d-copulas and d-fold stochastic measures, so it suffices to ensure that the measure $\mu_{C_{\min}}$ is d-fold stochastic. Since we already noted that the value of (5.12) is independent of the distribution inside the cubes I_i, we can assume that $\mu_{C_{\min}}$ distributes the mass inside of each cube I_i uniformly. The d-fold stochastic measures which distribute the mass x_i uniformly inside of the cube I_i for each $i \in \mathcal{I}$ are given by the equations

$$\sum_{i\in\mathcal{I}_m^k} x_i = \frac{1}{n}, \qquad \forall m \in \{1,\dots,n\}, \quad \forall k \in \{1,\dots,d\}. \tag{5.13}$$

This can be seen as follows: It is elementary that each d-fold stochastic measure satisfies the conditions (5.13). So let C fulfill (5.13) and let $0 \le a < b \le 1$. Now look for $1 \le i_- \le i^+ \le n$ such that $a \in [\frac{i_- -1}{n}, \frac{i_-}{n}]$ and $b \in [\frac{i^+-1}{n}, \frac{i^+}{n}]$. Without loss of generality

let us consider the first coordinate, it holds

$$\begin{aligned}\mu_C([a,b]\times[0,1]\times\cdots\times[0,1]) &= \mu_C\left(\left[a,\frac{i_-}{n}\right]\times[0,1]\times\cdots\times[0,1]\right)\\ &\quad+\sum_{i_1=i_-+1}^{i^+-1}\mu_C\left(\left[\frac{i_1-1}{n},\frac{i_1}{n}\right]\times[0,1]\times\cdots\times[0,1]\right)\\ &\quad+\mu_C\left(\left[\frac{i^+-1}{n},b\right]\times[0,1]\times\cdots\times[0,1]\right)\\ &= \frac{i_-}{n}-a+\left(\sum_{i_1=i_-+1}^{i^+-1}\sum_{j\in\mathcal{I}^1_{i_1}}x_j\right)+b-\frac{i^+-1}{n}\\ &= b-a=\lambda([a,b]).\end{aligned}$$

By standard arguments of measure theory, we can extend this result from intervals to arbitrary measurable sets A.
Hence the measure $C_{\min}$ is d-fold stochastic if and only if the constraints (5.13) are fulfilled. But those are exactly the constraints (5.10) from the d-AP with the right hand side $\frac{1}{n}$ instead of 1. Since scaling the right hand side of a linear optimization problem results in a similar scaling of the optimal solution, the optimal probability mass distribution (x_i) is given as $(x_i)=\frac{1}{n}(x_i^*)$ with (x_i^*) being the optimal solution to the general d-AP with objective function (a_i). □

With the necessary adjustments, Theorem 5.3.1 is equally valid for a maximization instead of a minimization.

Also in the multidimensional case, it is possible to approximate integrals over continuous functions by a sequence of integrals over functions that are piecewise constant.

Theorem 5.3.2. *Let f be continuous on $[0,1]^d$ and bounded by a sum of integrable functions and let the sets I_i^n for $i\in\mathcal{I}$ be given as before. Then, set*

$$\begin{aligned}f_n^{max}(x) &:= \max_{y\in I_i^n} f(y) \qquad \forall x\in I_i^n,\\ f_n^{min}(x) &:= \min_{y\in I_i^n} f(y) \qquad \forall x\in I_i^n.\end{aligned}$$

Now denote by C_n^{max} and C_n^{min} copulas which fulfill

$$\int_{[0,1]^d} f_n^{max}(x)\,dC_n^{max}(x) = \min_{C\in\mathcal{C}^d}\int_{[0,1]^d} f_n^{max}(x)\,dC(x) \tag{5.14}$$

and

$$\int_{[0,1]^d} f_n^{min}(x)\,dC_n^{min}(x) = \min_{C\in\mathcal{C}^d}\int_{[0,1]^d} f_n^{min}(x)\,dC(x).$$

Then

$$\lim_{n\to\infty}\int_{[0,1]^d} f_n^{min}(x)dC_n^{min}(x) = \lim_{n\to\infty}\int_{[0,1]^d} f_n^{max}(x)dC_n^{max}(x)$$
$$= \inf_{C\in\mathcal{C}^d}\int_{[0,1]^d} f(x)dC(x).$$

Furthermore, the sequences of minimizers C_n^{max} and C_n^{min} converge, at least along some subsequence, to a minimizer $C_{\min}$ of the problem

$$\min_{C\in\mathcal{C}_d}\int_{[0,1]^d} f(x)dC(x).$$

Proof. We show directly that

$$\lim_{n\to\infty}\int_{[0,1]^d} f_n^{min}(x)dC_n^{min}(x) = \lim_{n\to\infty}\int_{[0,1]^d} f_n^{max}(x)dC_n^{max}(x).$$

Let $\varepsilon > 0$. Since f is continuous, we may choose n large enough such that

$$|f_n^{min}(x) - f_n^{max}(x)| < \varepsilon \qquad \forall x \in [0,1]^d.$$

Furthermore, f_n^{min} is piecewise constant, so we have

$$\int_{[0,1]^d} f_n^{min}(x)dC_n^{min}(x) = \sum_{i\in\mathcal{I}} a_i \mu_{C_n^{min}}(I_i),$$

with $a_i := \min_{x\in I_i^n} f(x)$. Hence

$$\int_{[0,1]^d} f_n^{max}(x)\, dC_n^{max}(x) < \int_{[0,1]^d} f_n^{min}(x) + \varepsilon\, dC_n^{min}(x) = \sum_{i\in\mathcal{I}} \mu_{C_n^{min}}(I_i)(a_i + \varepsilon).$$

So we have

$$\left|\int_{[0,1]^d} f_n^{max}(x)\, dC_n^{max}(x) - \int_{[0,1]^d} f_n^{min}(x)\, dC_n^{min}(x)\right|$$
$$< \left|\sum_{i\in\mathcal{I}} \varepsilon\ \mu_{C_n^{min}}(I_i)\right| = \varepsilon.$$

For the proof that the sequence of optimizers converges to an optimizer for the continuous function, we start like in Iacò et al. (2015) by using Theorem 5.21 from Kallenberg (2002) to deduce that C_n^{max} converges weakly along some subsequence to a copula $C_{\min}$.

Now according to Theorem 2.4 from Beiglböck and Griessler (2014), any measure which is an optimal solution to a transportation problem is necessarily concentrated on a c-cyclical monotone set. So since C_n^{max} is optimal for the transportation problem (5.14) it

must be concentrated on a c-cyclical monotone set. Hence, for any $N \in \mathbb{N}$, the N-fold product measure $C_n^{max,\otimes N}$ is concentrated on the set $\mathcal{S}_n(N)$ of points $(x_1^{(1)}, \dots, x_d^{(1)}), \dots, (x_1^{(N)}, \dots, x_d^{(N)})$ for which for any permutations $\sigma_2, \dots, \sigma_d$

$$\sum_{i=1}^{N} f_n^{max}(x_1^{(i)}, \dots, x_d^{(i)}) \leq \sum_{i=1}^{N} f_n^{max}(x_1^{(i)}, x_2^{\sigma_2(i)}, \dots, x_d^{\sigma_d(i)}).$$

Now fix $\varepsilon > 0$. Since f is continuous, we can choose n large enough such that $C_n^{max,\otimes N}$ is concentrated on the set $\mathcal{S}_\varepsilon(N)$ of points with

$$\sum_{i=1}^{N} f(x_1^{(i)}, \dots, x_d^{(i)}) \leq \sum_{i=1}^{N} f(x_1^{(i)}, x_2^{\sigma_2(i)}, \dots, x_d^{\sigma_d(i)}) + \varepsilon.$$

Since f is continuous, the set $\mathcal{S}_\varepsilon(N)$ is closed. Therefore the limiting measure $C_{\min}^{\otimes N}$ is concentrated on $\mathcal{S}_\varepsilon(N)$ for all $\varepsilon > 0$. Now let $\varepsilon \to 0$ and we have that $C_{\min}^{\otimes N}$ is concentrated on a set of points with

$$\sum_{i=1}^{N} f(x_1^{(i)}, \dots, x_d^{(i)}) \leq \sum_{i=1}^{N} f(x_1^{(i)}, x_2^{\sigma_2(i)}, \dots, x_d^{\sigma_d(i)}),$$

which means that $C_{\min}$ is concentrated on a c-cyclically monotone set. Since f is continuous and bounded by a sum of integrable functions, we can apply Theorem 1.2 from Griessler (2016) to deduce that $C_{\min}$ is optimal. The proof for the sequence C_n^{min} is identical. □

In Hofer and Iacò (2014), the authors also give a convergence rate under the assumption of Lipschitz continuity. For completeness, we want to mention that this holds in very much the same way for the multidimensional setting.

Corollary 5.3.3. *Let the assumptions of Theorem 5.3.2 hold and, in addition, assume that f is Lipschitz continuous on $[0,1]^d$ with Lipschitz constant $L > 0$. Then*

$$\left| \int_{[0,1]^d} f_n^{max}(x) dC_n^{max}(x) - \int_{[0,1]^d} f_n^{min}(x) dC_n^{min}(x) \right| \leq L \frac{\sqrt{d}}{n}.$$

Proof. by the Lipschitz continuity of f and the construction of f_n^{min} and f_n^{max}, we have that

$$|f_n^{min}(x) - f_n^{max}(x)| \leq L \frac{\sqrt{d}}{n}.$$

Hence the proof follows by replacing ε with $L\frac{\sqrt{d}}{n}$ in the proof of Theorem 5.3.2. □

Remark 5.3.4. *The method presented generalizes the approach from Hofer and Iacò (2014) and furthermore facilitates the computation since the relaxed assignment problem is much easier to solve than the integer one. However, the solution vector is still n^d-dimensional. In practice this method can be applied on a standard laptop for $n^d \leq 10^7$ and*

since good approximations are already possible for $n \approx 25$, dimensions up to $d = 5$ can certainly be handled. So for most practical applications, the rearrangement algorithm by Puccetti and Rüschendorf (2015) will, whenever applicable, still be the method of choice. The merit of our approach lies in the generality of the statement. It is not limited to supermodular functions but requires merely continuity and a notion of boundedness. Both of these assumptions might possibly be relaxed as research in the field of optimal transport progresses.

Remark 5.3.5. *Considering only Shuffles of* Min *as minimizers respectively maximizers in equations* (5.1) *and* (5.2) *as was done in in Hofer and Iacò (2014) is in some sense also the notion behind the rearrangement algorithm. Puccetti and Wang (2015) consider* rearrangements*, show how they are linked to copulas and illustrate that Shuffles of* Min *can be seen as a construction of particular rearrangements. As the authors of Puccetti and Wang (2015) mention, the fact that Shuffles of* Min *are dense in the set of copulas could then be used to approximate solutions to copula optimization problems arbitrarily well with Shuffles of* Min*. This would be a direct extension of the method from Hofer and Iacò (2014) to general dimensions, however, due to the complexity of the integer assignment problem, this approach is of very limited practical relevance. The rearrangement algorithm, on the other hand, cuts the solution space down to* oppositely ordered rearrangements*, resulting in the restriction to supermodular functions but also in an enormous increase of efficiency. For more details, we refer to Puccetti and Rüschendorf (2015).*

Remark 5.3.6. *Following the spirit of Bernard et al. (2018), Bignozzi et al. (2015) or Lux et al. (2017), one might also consider including partial information about the distribution by simply adding suitable constraints to the linear program.*

5.4. Applications

5.4.1. Dependence Measures

A natural application for this technique is the approximation of upper and lower bounds on dependence measures. In the bivariate case, there are well-known and widely used measures such as *Spearman's* ρ, *Kendall's* τ, *Blomqvists* β and *Gini's* γ. See e.g. Mai and Scherer (2014) for definitions.

We now focus on a multidimensional version of Spearman's ρ. Define

$$\rho(C) := \frac{d+1}{2^d - (d+1)} \left(2^d \int_{[0,1]^d} \Pi(u) dC(u) - 1 \right). \tag{5.15}$$

Here, Π denotes the independence copula, i.e. $\Pi(x_1, \ldots, x_d) = \Pi_{i=1}^d x_i$. It is well-known that $\rho(C)$ is maximal when $C = M_d$, i.e. C is the *Min*-copula. It is also well-known that plugging in the lower Fréchet-Hoeffding bound (usually denoted by W_d) yields a lower

bound on $\rho(C)$:

$$\rho(W_d) = \frac{2^d - (d+1)!}{d!(2^d - (d+1))} \leq \rho(C) \leq \rho(M_d), \qquad C \in \mathcal{C}^d, \quad d \geq 2.$$

However, since W_d is not a copula for $d \geq 2$, it is not a priori clear whether this lower bound is attained or not. Indeed, this has been stated as an open problem in Schmid et al. (2010). In 2011, Wang and Wang (2011) found an analytical solution to this long unresolved question. They give a formula to explicitly compute

$$\Lambda_d := \min_{C \in \mathcal{C}^d} \int_{[0,1]^d} \Pi(u)\, dC(u)$$

for any $d \in \mathbb{N}$. Since the formula yields $\Lambda_3 = 5.4803 \times 10^{-2}$, it is straightforward that $\rho(W_3) = -\frac{2}{3}$ is not attained. We now want to give an alternative, numerical proof for this result. We chose this example because the fact that we know the exact analytical value will help us to validate the convergence of our method.

We use Theorem 5.3.2 to compute strict upper and lower bounds for Λ_3. In this case the monotonicity of Π facilitates finding the maximum respectively minimum functions in the algorithm as we simply have to evaluate Π at the vertices of the grid cubes that maximize respectively minimize the arguments. An approximated value is also obtained by evaluating Π in the center of each grid cube. Of course, this has to be adapted for other cost functions. Table 5.1 illustrates the results obtained using the method proposed in this paper with a grid of $n \in \{30, 40, 50, 60\}$ sections in each dimension[1] as well as the range the rearrangement algorithm (which uses the same discretization method as our algorithm) computes for a grid of 10^5 sections in each dimension Puccetti and Rüschendorf (2015).

Note that the range here is not a confidence interval but actually consists of deterministic upper and lower bounds on the true value. Already the lower bound for $n = 30$ suffices to prove that $\rho(C) > -\frac{2}{3} = \rho(W_3)$ for all copulas C. As can be seen, the rearrangement algorithm yields a more precise approximation for the same problem and that in considerably less time, even for higher dimensions d (more details can be found in Puccetti and Rüschendorf (2015)). However, since the lower bound computed by the rearrangement algorithm does not always have to be satisfied, we see the merit of our method in providing rigorous numerics for the fact that the lower bound $-\frac{2}{3}$ for $\rho(C)$ is not best-possible.

An interesting extension to the minimization of Λ_d is considering non-uniform marginal distributions. While the result of Wang and Wang (2011) is restricted to identical

[1]This result was obtained using the "lpSolve" package for the open source program *R*. This package is built on the free Mixed Integer Linear Program solver lp_solve, which utilizes the revised simplex method and the Branch-and-bound method. No presolve routines or any other kind of advanced techniques were used.

n	lower bd.	approx. value	upper bd.
30	0.044414	0.054971	0.066341
40	0.046906	0.054897	0.063349
50	0.048433	0.054863	0.061587
60	0.049466	0.054844	0.060427
10^5 (RA)	0.054800	-	0.054807
analytic value	-	0.054803	-

Table 5.1.: Approximation results for Λ_3. The values for $n = 30, \dots, 60$ were calculated by the method proposed in this paper, whereas the values for $n = 10^5$ were calculated using the rearrangement algorithm Puccetti and Rüschendorf (2015).

marginal distributions having an increasing density, Sklar's Theorem allows us to treat any marginal laws by inserting the quantile functions of the desired distributions. Table 5.2 contains the approximated value as well as upper and lower bounds of

$$\Theta(\mu_1, \mu_2, \mu_3) = \min_{\mu \in \mathcal{M}(\mu_1, \mu_2, \mu_3)} \int_{[0,1]^3} \Pi(x) d\mu(x)$$

for different choices of μ_1, μ_2, μ_3. Here $\mathcal{M}(\mu_1, \dots, \mu_d)$ denotes the set of probability measures on $[0,1]^d$ which have marginal distributions $\mu_1, \dots, \mu_d$. Distributions with unbounded support require some minor adjustments, however the method is still applicable. Again, the rearrangement algorithm as proposed by Puccetti and Rüschendorf (2015) will produce more accurate results in shorter calculation time and is thus probably the preferable choice in applications.

μ_1	μ_2	μ_3	lower bd.	approx. value	upper bd.
$\mathrm{Par}(2, 0.2)$	$\mathcal{U}([0,1])$	$\mathcal{U}([0,1])$	0.043123	0.046767	0.054226
$\mathcal{U}([0,1])$	$\mathcal{U}([0,1])$	$Beta(2,5)$	0.030489	0.033876	0.037232
$exp(1)$	$\mathcal{U}([0,1])$	$Beta(2,5)$	0.044647	0.051213	0.058365

Table 5.2.: Approximation results for inhomogenous marginals, obtained for a discretization of $n = 60$ sections in each dimension.

5.4.2. First-to-default Swaps

In the examples so far, we always minimized the expectation of the product of random variables. The product function is *supermodular* in the following sense.

Definition 5.4.1. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is called supermodular, if it fulfills*

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y) \qquad \forall x, y \in \mathbb{R}^d.$$

Here, $x \wedge y$ resp. $x \vee y$ means the componentwise minimum resp. maximum of x and y.

Since the rearrangement algorithm can only be applied to approximate the expectation of supermodular functions, it is interesting to look at an example involving a non-supermodular function. We want to consider so called *First-to-default* Swaps. They can be thought of as an insurance contract for portfolios of risky assets. The protection seller (PS) compensates the losses if one of the assets in the portfolio of the protection buyer (PB) defaults. In turn, the PB pays fixed premiums at fixed points in time (e.g. quarterly or annually) until the first default occurs or the maturity of the contract is reached. There are no payments for any event after the first default or after maturity. We consider a portfolio consisting of three risky assets and use the following assumptions and notations.

The default times τ_1, τ_2, and τ_3 of the assets follow an exponential distribution with parameters λ_1, λ_2 and λ_3 respectively. The notionals of all three assets are 1 and the recovery rates R_1, R_2 and R_3 describe the amount of money that can be liquidized if the corresponding asset defaults. So the total loss for a default of asset i is $(1 - R_i)$. The times of premium payments are denoted by $0 = t_0 < t_1 < \cdots < t_k = T$ with T denoting the time of maturity. Note in particular that the first payment is due at time $t_0 = 0$. Finally, we assume there is a constant, risk free interest rate $r \geq 0$. Now the premiums p are given as

$$\begin{aligned} p = \int_{[0,1)^3} & \frac{e^{-r \min(F_1^{-1}(x_1), F_2^{-1}(x_2), F_3^{-1}(x_3))}}{\sum_{i=0}^{k} e^{-rt_i} \mathbb{1}_{\{t_i < \min(F_1^{-1}(x_1), F_2^{-1}(x_2), F_3^{-1}(x_3))\}}} \\ & \cdot \Big(\mathbb{1}_{\{F_1^{-1}(x_1) \leq \min(F_2^{-1}(x_2), F_3^{-1}(x_3), T)\}} (1 - R_1) \\ & + \mathbb{1}_{\{F_2^{-1}(x_2) \leq \min(F_1^{-1}(x_1), F_3^{-1}(x_3), T)\}} (1 - R_2) \\ & + \mathbb{1}_{\{F_3^{-1}(x_3) \leq \min(F_1^{-1}(x_1), F_2^{-1}(x_2), T)\}} (1 - R_3) \Big) \, dC(x), \end{aligned}$$

with C denoting the copula of the distribution function of (τ_1, τ_2, τ_3) and F_i^{-1} being the quantile function corresponding to the distribution of τ_i. Since our assumptions and the valuation method we want to use are basically the same, we refer to Hofer and Iacò (2014) for the precise deduction of the premium heights. We calculate bounds for the minimal as well as for the maximal premium for three payment times $t_0 = 0$, $t_1 = 1$ and $t_2 = T = 2$. The results are listed in Table 5.3.

Remark 5.4.2. *Note that the integrand in the last example is not continuous. Our method is not restricted to continuous functions but can be applied as long as the integrand f can, with respect to the L^1 norm, be approximated by a (subsequence of a) sequence of functions $(f_n)_n$ that are constant on the cubes I_i^n with $i \in \mathcal{I}$ (as defined in Theorem 5.3.1). Hence, by a simple denseness argument, the algorithm actually works for any function f in $L^1([0,1]^d)$, which is why we still have valid bounds in our last example. However, the speed of convergence can be very slow for functions with many discontinuities. For example, it can happen that for $n_1 < n_2$, the bounds for a discretization of n_2 sections are worse than those for a discretization with n_1 sections. Also the convergence of the sequence of optimizers is not guaranteed if we choose an integrand f which is not continuous.*

$\lambda_1 = \frac{1}{3}, \quad R_1 = 0.5,$	$\lambda_2 = \frac{1}{2}, \quad R_2 = 0.7,$	$\lambda_3 = \frac{2}{5}, \quad R_3 = 0.6$
n	range min. premium	range max. premium
30	$0.14093 - 0.16104$	$0.37090 - 0.40287$
40	$0.14566 - 0.16072$	$0.37693 - 0.40086$
50	$0.14572 - 0.15775$	$0.37777 - 0.39690$
70	$0.14777 - 0.15424$	$0.37889 - 0.39102$

Table 5.3.: Upper and lower bounds for the minimal and the maximal premium. The interest rate was set to $r = 0.05$.

6. On Majorization in Dependence Modeling

In Chapter 4.1 we already discussed extremal dependence concepts in two dimensions and alluded to the problem of defining extremal negative dependence for more than two random variables. For two quantities, negative dependence is clear, high values of the first should come with low values of the second and vice versa. Because this intuition is not transferable to higher dimensions, other characteristics of negative dependence have to be found.

For this, consider d random variables $X_1, \dots, X_d$ with corresponding distributions $F_1, \dots, F_d$. It is not hard to see that the copula which maximizes

$$\operatorname{Var}\left(\sum_{i=1}^{d} X_i\right),$$

that is the variance of the sum of $X_1, \dots, X_d$, is given by the comonotonicity copula M. In fact, there is a whole class of functions that are maximized when choosing the copula M, namely the *supermodular functions*, that we have already seen in part 5.4. Interpreting countermonotonicity as the opposite of comonotonicity, it seems natural to call a copula that minimizes supermodular functions a negative dependence structure. Since for two random variables, the copula minimizing supermodular functions is indeed the lower Fréchet-Hoeffding bound W, this interpretation is consistent with the two dimensional case.
Over the past years, several concepts tying negative dependence to supermodular functions have been developed. Notable contributions to this field include Rüschendorf (1980) who considered Δ-monotone functions, Wang and Wang (2011) who introduce complete mixability (sometimes called joint mixability) and Puccetti and Wang (2015), where Σ-countermonotone distributions are defined. These concepts differ in whether or not they are attainable for any set of marginal distributions, whether they are unique distribution functions if they exist and whether or not they minimize any supermodular function. For questions concerning extremal dependence in applications see for example Wang et al. (2013) and Bernard et al. (2014). A good survey on extremal dependence concepts is provided in Puccetti and Wang (2015).

A very useful tool to find copulas minimizing supermodular functions in practice is the rearrangement algorithm (RA) that we already mentioned in Chapter 5. The RA is designed to return a Σ-countermonotone distribution for a given set of marginal laws

which will minimize a certain subclass of supermodular functions (for details see Puccetti and Rüschendorf (2015)). Since for any set of marginal distributions, there exists a Σ-countermonotone joint distribution, the RA will certainly terminate and return a solution. It can also be shown that the minimizing distribution can be found in the class of Σ-countermonotone distributions, unfortunately, though, Σ-countermonotonicity does not guarantee uniqueness so one cannot be sure that there isn't another Σ-countermonotone distribution that yields an even lower value for the supermodular function which is to be minimized. There are even theoretical examples that produce an arbitrarily large error because of this, however these examples exhibit a quite unrealistic structure and in contrast the RA works very well on real applications.

Still these shortcomings of the RA are somewhat unsatisfactory and in addition, the focus on finding Σ-countermonotone distributions further limits the use of this method. Following the motivation to broaden the set of functions the RA can be applied to and also to improve the understanding of the underlying structures, we analyzed the topic in more detail and found a strong connection to the field of majorization theory. In fact we were able to use majorization to find generalizations of statements from Puccetti and Rüschendorf (2015) also giving insights on the problem from a different angle. However, we learned that, independently of our own research, our results were already found by Jakobsons and Wang (2017), whose article had been under review while we were writing our paper. Since Jakobsons and Wang (2017) use the same foundations for their theoretical results, namely Marshall et al. (2011), and also provide more applications and connections to optimization theory, we felt that our paper couldn't add anything sufficiently new to the discussion. We therefore decided not to pursue publishing it but instead kept a preprint available online.

The remainder of this chapter, that is sections 6.1 through 6.3, is identical to an unpublished article that can be found on arXive as Preischl (2017).

6.1. Introduction and existing theory

When talking about dependence uncertainty, a major interest usually lies on finding a dependence structure that maximizes or minimizes a certain expectation. That is, given distributions $F^{(1)}, \dots, F^{(d)}$ and a function $f : \mathbb{R}^d \to \mathbb{R}$, what are

$$\inf \left\{ \mathbb{E}\left[f(X^{(1)}, \dots, X^{(d)}) \right] : X^{(i)} \sim F^{(i)}, 1 \leq i \leq d \right\} \quad \text{and}$$
$$\sup \left\{ \mathbb{E}\left[f(X^{(1)}, \dots, X^{(d)}) \right] : X^{(i)} \sim F^{(i)}, 1 \leq i \leq d \right\}?$$

For a start, assume that $X^{(1)}, \dots, X^{(d)}$ are discrete and take n not necessarily distinct values with an equal probability of $\frac{1}{n}$. In this case $X^{(i)}$ is said to have an *n-discrete* distribution for $i = 1, \dots, d$. Denote by $x^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})^T$ a vector containing the possible values for each $X^{(i)}$. Thus the $n \times d$ matrix $X = (x^{(1)}, \dots, x^{(d)})$ can be interpreted as the joint distribution of a random vector $X = (X^{(1)}, \dots, X^{(d)})$, where $X^{(i)} \sim F^{(i)}$ by giving each of its rows equal probability $\frac{1}{n}$. Obviously, for this distribution, we have

$$\mathbb{E}[f(X)] = \frac{1}{n} \sum_{k=1}^{n} f(x_k^{(1)}, \dots, x_k^{(d)}).$$

If we now rearrange the entries in one or multiple columns of X, we get a new distribution $\tilde{X}$ which has the same marginals as X. Write

$$\mathcal{P}(X) = \{(\tilde{x}^{(1)}, \dots, \tilde{x}^{(d)}) : \tilde{x}^{(i)} = \pi_i x^{(i)}, \, \pi_i \text{ is a permutation on } \{1, \dots, n\}\}$$

for the set of matrices that can be obtained from X by permuting elements within columns. In the following we will try to find

$$m_f(X) := \min_{\tilde{X} \in \mathcal{P}} \sum_{k=1}^{n} f(\tilde{x}_k^{(1)}, \dots, \tilde{x}_k^{(d)}).$$

Note that the factor $\frac{1}{n}$ was omitted since it does not affect optimality. This setting was considered in (Puccetti and Rüschendorf, 2015) where conditions are given under which the minimum is attained at a matrix which is in some sense oppositely ordered. More precisely, assume $h : \mathbb{R}^d \to \mathbb{R}$ can be decomposed into two functions $h^2 : \mathbb{R}^2 \to \mathbb{R}$ and $h^{d-1} : \mathbb{R}^{d-1} \to \mathbb{R}$ such that for all $x \in \mathbb{R}^d$

$$h(x_1, \dots, x_d) = h^2(x_i, h^{d-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)) \qquad \forall i = 1, \dots, d. \tag{6.1}$$

We write X_{-i} for the matrix $(x^{(1)}, \dots, x^{(i-1)}, x^{(i+1)}, \dots, x^{(d)})$, i.e. X with the ith column deleted and

$$h^{d-1}(X_{-i}) := \begin{pmatrix} h^{d-1}\left(x_1^{(1)}, \dots, x_1^{(i-1)}, x_1^{(i+1)}, \dots, x_1^{(d)}\right) \\ h^{d-1}\left(x_2^{(1)}, \dots, x_2^{(i-1)}, x_2^{(i+1)}, \dots, x_2^{(d)}\right) \\ \vdots \\ h^{d-1}\left(x_n^{(1)}, \dots, x_n^{(i-1)}, x_n^{(i+1)}, \dots, x_n^{(d)}\right) \end{pmatrix}$$

for the vector that is h^{d-1} applied to every row of X_{-i}.
Then define

$$\mathcal{O}_h(X) := \{\tilde{X} \in \mathcal{P}(X) : \tilde{x}^{(i)} \perp h^{d-1}(\tilde{X}_{-i}), \, i = 1, \dots, d\}.$$

Here $x \perp y$ denotes that the vectors x and y are oppositely ordered, meaning: for $x, y \in \mathbb{R}^n : x \perp y \iff (x_i - x_j)(y_i - y_j) \leq 0$ for all $i, j \in \{1, \dots, n\}$. A particular choice for h which satisfies (6.1) would be the sum operator, i.e. $h(x_1, \dots, x_d) = x_1 + \cdots + x_d$. Now Puccetti and Rüschendorf state two cases in which the minimum is attained at an element of $\mathcal{O}_h(X)$ (see Propositions 2.4 and 2.6 from (Puccetti and Rüschendorf, 2015)):

Theorem 6.1.1. *If* $f(x^{(1)}, \ldots, x^{(d)}) = g(h(x^{(1)}, \ldots, x^{(d)}))$ *where* $g : \mathbb{R} \to \mathbb{R}$ *is convex and* $h(x_1, \ldots, x_d) = x_1 + \cdots + x_d$ *is the sum operator, then*

$$m_f(X) = \min_{\tilde{X} \in \mathcal{O}_h(X)} \sum_{k=1}^{n} f(\tilde{x}_k^{(1)}, \ldots, \tilde{x}_k^{(d)}).$$

Theorem 6.1.2. *If* $h : \mathbb{R}^d \to \mathbb{R}$ *is supermodular, componentwise strictly monotonic and satisfies* (6.1) *with* h^2 *and* h^{d-1} *also being supermodular, then*

$$m_h(X) = \min_{\tilde{X} \in \mathcal{O}_h(X)} \sum_{k=1}^{n} h(\tilde{x}_k^{(1)}, \ldots, \tilde{x}_k^{(d)}).$$

Remember that a function $h : \mathbb{R}^d \to \mathbb{R}$ is called *supermodular* if
$h(x) + h(y) \leq h(x \wedge y) + h(x \vee y)$, where $\wedge$ and $\vee$ denote the componentwise minimum respectively maximum. Some authors prefer the term *L-superadditive* instead of supermodular.
The concept (and the implementation) of rearranging the matrix X until obtaining an element in $\mathcal{O}_h(X)$ is the rearrangement algorithm (RA).

6.2. Main results

As noted by Puccetti and Rüschendorf, Th. 6.1.1 and Th. 6.1.2 consider two distinct cases, though the difference might seem subtle. We will clarify this later, but for the moment, we want to mention that a minor extra assumption allows us to unify these cases.

Theorem 6.2.1. *If* $f(x^{(1)}, \ldots, x^{(d)}) = g(h(x^{(1)}, \ldots, x^{(d)}))$ *where* $g : \mathbb{R} \to \mathbb{R}$ *is increasing and convex and* $h : \mathbb{R}^d \to \mathbb{R}$ *is supermodular, componentwise strictly monotonic and satisfies* (6.1) *with* h^2 *also being supermodular, then*

$$m_f(X) = \min_{\tilde{X} \in \mathcal{O}_h(X)} \sum_{k=1}^{n} f(\tilde{x}_k^{(1)}, \ldots, \tilde{x}_k^{(d)}).$$

Note that Th. 6.1.2 is trivially included in Th. 6.2.1, whereas Th. 6.1.1 is not due to the extra monotonicity assumption on g. A proof of Th.6.2.1 will be given later along with the remark that it is actually possible to derive a stronger conclusion from the assumptions of 6.2.1 than the one actually stated. To see this, we need some notions from the theory of majorization, which are taken from the extremely rich summary (Marshall et al., 2011), from which also some of the arguments in (Puccetti and Rüschendorf, 2015) are taken.

The concept of majorization allows to find extreme values for certain functions $f : \mathbb{R}^n \to \mathbb{R}$ based on a (partial) ordering of the vectors $x \in \mathbb{R}^n$. Let $(x_{[1]}, \ldots, x_{[n]})$ denote

the decreasing rearrangement of x, i.e. $x_{[1]}$ is the largest component of x and $x_{[n]}$ is the smallest. Analogously, we write $(x_{(1)}, \ldots, x_{(n)})$ for the increasing rearrangement of x, i.e. $x_{[1]} = x_{(n)}$ and $x_{[n]} = x_{(1)}$. For reasons of convenience, we will also use the notation $x_\downarrow := (x_{[1]}, \ldots, x_{[n]})$ and $x_\uparrow := (x_{(1)}, \ldots, x_{(n)})$ to denote the decreasing and increasing rearrangements. Obviously it holds $x_\downarrow \perp x_\uparrow$.

Definition 6.2.2. *A vector $x = (x_1, \ldots, x_n)$ is said to be majorized by $y = (y_1, \ldots, y_n)$ (write $x \prec y$) if*

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \qquad k = 1, \ldots, n-1$$
$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}.$$

Furthermore, x is said to be weakly submajorized by y (write $x \prec_w y$) if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \qquad k = 1, \ldots, n.$$

Finally, x is said to be weakly supermajorized by y (write $x \prec^w y$) if

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)} \qquad k = 1, \ldots, n.$$

Obviously, we have $x \prec y$ if and only if $x \prec_w y$ and $x \prec^w y$.

When we observe $x \prec y$, we are naturally interested in the effects of this majorization. This leads to *Schur-convex* functions.

Definition 6.2.3. *A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be Schur-convex if $x \prec y$ implies $\phi(x) \leq \phi(y)$.*

Schur-convexity works with weak majorization as follows

Proposition 6.2.4 (see 3.A.8 in (Marshall et al., 2011))**.** *A real valued function ϕ, defined on $\mathbb{R}^n$ satisfies*

$$x \prec_w y \Rightarrow \phi(x) \leq \phi(y)$$

if and only if ϕ is Schur-convex and componentwise increasing. Analogously,

$$x \prec^w y \Rightarrow \phi(x) \leq \phi(y)$$

holds if and only if ϕ is Schur-convex and componentwise decreasing.

An important class of Schur-convex functions is given by the following proposition.

Proposition 6.2.5 (see 3.C.1 in (Marshall et al., 2011)). *If $\psi : \mathbb{R} \to \mathbb{R}$ is convex, then the function ϕ defined by*

$$\phi(x) = \sum_{i=1}^{n} \psi(x_i),$$

is Schur-convex. Obviously, if ψ is in addition increasing (decreasing) then ϕ is Schur-convex and componentwise increasing (decreasing).

Another natural question would be, which operations preserve majorization. It turns out that the ordering of vectors plays an important role here. A famous result for oppositely ordered vectors is the following.

Proposition 6.2.6 (see 6.A.2 in (Marshall et al., 2011)). *For any two vectors $x^{(1)}$ and $x^{(2)}$ on $\mathbb{R}^n$, it holds that*

$$x^{(1)}_{\downarrow} + x^{(2)}_{\uparrow} \prec x^{(1)} + x^{(2)}.$$

It is also possible to consider more general aggregation operators than the sum. This however requires more assumptions and yields only weak majorization (which is completely sufficient for our purpose).

Proposition 6.2.7 (see 6.C.4 in (Marshall et al., 2011)).

$$\left(h(x^{(1)}_{[1]}, x^{(2)}_{[n]}), \ldots, h(x^{(1)}_{[n]}, x^{(2)}_{[1]})\right) \prec_w \left(h(x^{(1)}_1, x^{(2)}_1), \ldots, h(x^{(1)}_n, x^{(2)}_n)\right),$$

holds for any two vectors $x^{(1)}$ and $x^{(2)}$ in $\mathbb{R}^n$, if and only if h is supermodular and either increasing in each component or decreasing in each component.

These statements are clearly restricted to two vectors since "oppositely ordered" does not make sense otherwise and this is also the reason why the decomposition property (6.1) is needed.
We are now ready to prove Theorem 6.2.1.

Proof of Theorem 6.2.1: We want to show that for every matrix $\tilde{X} \in \mathcal{P}(X) \setminus \mathcal{O}_h(X)$, there exists a matrix $\hat{X} \in \mathcal{O}_h(X)$ such that

$$\sum_{k=1}^{n} f(\hat{x}^{(1)}_k, \ldots, \hat{x}^{(d)}_k) \leq \sum_{k=1}^{n} f(\tilde{x}^{(1)}_k, \ldots, \tilde{x}^{(d)}_k). \tag{6.2}$$

If $\tilde{X} \in \mathcal{P}(X) \setminus \mathcal{O}_h(X)$, then there exists a column $\tilde{x}^{(i)}$ which is not oppositely ordered to $h^{d-1}(\tilde{X}_{-i})$. Denote by $\hat{X}$ the matrix obtained from $\tilde{X}$ by ordering $\tilde{x}^{(i)}$ oppositely to $h^{d-1}(\tilde{X}_{-i})$. By Proposition 6.2.7 it holds that

$$\begin{aligned} h(\hat{X}) &= \left(h^2\left(\hat{x}^{(i)}_1, h^{d-1}(\hat{X}_{-i})_1\right), \ldots, h^2\left(\hat{x}^{(i)}_n, h^{d-1}(\hat{X}_{-i})_n\right)\right)^T \\ &\prec_w \left(h^2\left(\tilde{x}^{(i)}_1, h^{d-1}(\tilde{X}_{-i})_1\right), \ldots, h^2\left(\tilde{x}^{(i)}_n, h^{d-1}(\tilde{X}_{-i})_n\right)\right)^T = h(\tilde{X}). \end{aligned}$$

Since g is increasing and convex, we get with the help of Proposition 6.2.5 that

$$\sum_{k=1}^{n} f(x_k^{(1)}, \dots, x_k^{(d)}) = \sum_{k=1}^{n} g(h(X)_k)$$

is Schur-convex and increasing, hence, (6.2) holds by Proposition 6.2.4. If $\hat{X} \in \mathcal{O}_h(X)$, we are done. It remains to show that this procedure eventually reaches an element in $\mathcal{O}_h(X)$. To see this, we look at the reordering of a certain column in more detail. Fix i and let $K \subset \{1, \dots, n\}$ be the set of all indices that appear in a pair which violates the opposite ordering of $\tilde{x}^{(i)}$ and $h^{d-1}(\tilde{X}_{-i})$. So $k_1, k_2 \in K \iff \tilde{x}_{k_1}^{(i)} < (>) \tilde{x}_{k_2}^{(i)}$ and $h^{d-1}(\tilde{X}_{-i})_{k_1} < (>) h^{d-1}(\tilde{X}_{-i})_{k_2}$. We construct $\hat{x}^{(i)}$ in the following way: let $x_{\underline{k}}^{(i)} := \min_{k \in K} x_k$ and let $h^{d-1}(\tilde{X}_{-i})_{\overline{k}} := \max_{k \in K} h^{d-1}(\tilde{X}_{-i})_k$. Since $\underline{k}, \overline{k} \in K$ and by construction, we know that

$$x_{\underline{k}}^{(i)} < x_{\overline{k}}^{(i)} \quad \text{and} \quad h^{d-1}(\tilde{X}_{-i})_{\underline{k}} < h^{d-1}(\tilde{X}_{-i})_{\overline{k}}. \tag{6.3}$$

Now exchange $x_{\underline{k}}^{(i)}$ and $x_{\overline{k}}^{(i)}$. It is not hard to see that this reduces the number of violating indices by at least one. We repeat this procedure until $K = \emptyset$ and have thus created a vector $\hat{x}^{(i)}$ which is ordered oppositely to $h^{d-1}(\tilde{X}_{-i})$. At this point it is important to see that due to (6.3), we have

$$h^2(\tilde{x}_{\overline{k}}^{(i)}, h^{d-1}(\tilde{X}_{-i})_{\overline{k}}) > h^2(\tilde{x}_{\underline{k}}^{(i)}, h^{d-1}(\tilde{X}_{-i})_{\overline{k}}) > h^2(\tilde{x}_{\underline{k}}^{(i)}, h^{d-1}(\tilde{X}_{-i})_{\underline{k}})$$

and

$$h^2(\tilde{x}_{\overline{k}}^{(i)}, h^{d-1}(\tilde{X}_{-i})_{\overline{k}}) > h^2(\tilde{x}_{\overline{k}}^{(i)}, h^{d-1}(\tilde{X}_{-i})_{\underline{k}}) > h^2(\tilde{x}_{\underline{k}}^{(i)}, h^{d-1}(\tilde{X}_{-i})_{\underline{k}}).$$

So after each exchanging step of the above type, the values of $h(\tilde{X})$ have changed in exactly two components and the new values are strictly between the old ones. Hence it is clear that $h(\hat{X}) \neq h(\tilde{X})$ and also that $h(\hat{X})$ is not a permutation of $h(\tilde{X})$. However, $x \prec_w y$ and $y \prec_w x$ at the same time implies that x is a permutation of y, so we know that $h(\tilde{X}) \prec_w h(\hat{X})$ cannot hold. This means that $h(\hat{X})$ is strictly below $h(\tilde{X})$ w.r.t $\prec_w$. Since $\mathcal{P}(X)$ is finite, it follows that after a finite number of steps, we arrive at a matrix $\hat{X} \in \mathcal{O}_h(X)$ which satisfies

$$\sum_{k=1}^{n} f(\hat{x}_k^{(1)}, \dots, \hat{x}_k^{(d)}) \leq \sum_{k=1}^{n} f(\tilde{x}_k^{(1)}, \dots, \tilde{x}_k^{(d)}).$$

□

Remark 6.2.8. *With the proof of Theorem 6.2.1 we actually showed that for any element $\tilde{X} \in \mathcal{P}(X)$, there is a chain of matrices $\hat{X}_{(1)}, \hat{X}_{(2)}, \dots, \hat{X}_{(n)}$ such that $h(\hat{X}_{(n)}) \prec_w h(\hat{X}_{(n-1)}) \prec_w \cdots \prec_w h(\hat{X}_{(1)}) \prec_w h(\tilde{X})$ and $\hat{X}_{(n)} \in \mathcal{O}_h(X)$. Indeed, weak majorization is much stronger than just the inequality* (6.2)*. This is also reflected by the fact that* (6.2) *can be installed without majorization, using a weaker statement by Lorentz, see (Puccetti and Rüschendorf, 2015) for details. There it is also shown that using Lorentz to obtain* (6.2) *does not require h to be monotone at all, whereas we showed, weak majorization needs (non strict) monotonicity. However, since strict monotonicity of h is, in general, required to reach an element in $\mathcal{O}_h(X)$ the weaker statement does not give any benefit.*

Remark 6.2.9. *The proof also shows, why Th. 6.1.1 does not require g to be monotone: According to Proposition 6.2.6, the sum as aggregation function yields strong majorization, which is why the inequality* (6.2) *holds for $f = g \circ h$ for arbitrary convex g. A complete description of all aggregating functions yielding strong monotonicity and hence* (6.2) *without monotonicity of g is given in the following statement.*

Proposition 6.2.10 (see 6.B.2 in (Marshall et al., 2011))**.**

$$\left(h(x_{[1]}^{(1)}, x_{[n]}^{(2)}), \ldots, h(x_{[n]}^{(1)}, x_{[1]}^{(2)})\right) \prec \left(h(x_1^{(1)}, x_1^{(2)}), \ldots, h(x_n^{(1)}, x_n^{(2)})\right)$$

holds for any two vectors $x^{(1)}$ and $x^{(2)}$ in $\mathbb{R}^n$, if and only if h is of the form $h(x_1, x_2) = \varphi_1(x_1) + \varphi_2(x_2)$, where φ_1 and φ_2 are monotone in the same direction.

At this point, we would like to stress the fact that the gain of Th. 6.2.1 over Th. 6.1.2 consists of dropping the strict monotonicity of the overall function f by showing that strict monotonicity is only needed for the aggregation function h. An example of a function that is included in Th. 6.2.1 but not in Th. 6.1.2 would be the stop-loss functional $f(x_1, \ldots, x_d) = \max(x_1 + \cdots + x_d - k, 0)$ for some $k \neq 0$. Except for the issue of strict monotonicity, the composition with an increasing convex function was already possible in Th. 6.1.2 as the next lemma shows.

Lemma 6.2.11. *If $h : \mathbb{R}^d \to \mathbb{R}$ has the decomposition property* (6.1) *then for any $g : \mathbb{R} \to \mathbb{R}$ the composition $f = g \circ h$ also satisfies* (6.1)*. If furthermore the decomposition of h is supermodular (in particular h^2 has to be supermodular) and g is increasing and convex, then the decomposition of f is also supermodular (in particular f^2 is supermodular).*

Proof. It holds

$$\begin{aligned} f(x_1, \ldots, x_d) &= (g \circ h)(x_1, \ldots, x_d) = g\left(h((x_1, \ldots, x_d)\right) \\ &= g\left(h^2(x_i, h^{d-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d))\right) \\ &= (g \circ h^2)(x_i, h^{d-1}((x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d))) \end{aligned}$$

for $i = 1, \ldots, d$. So take $f^2 = g \circ h^2$ and $f^{d-1} = h^{d-1}$ and we have (6.1). For the claim that f^2 is supermodular if h^2 is supermodular and g is increasing and convex, use Proposition 6.D.2 from (Marshall et al., 2011). □

Using majorization, we were able to identify new cases where the RA can be applied to compute bounds on expected values. In particular, this extension yields more flexibility when working with non-linear risk aggregation. While these models are relatively sparse in mathematical finance, they enjoy some popularity in the field of modeling medical risks (for example see (Brattin, 1994) or (Guo, 2000)). Another, even broader, extension is possible, when we generalize the decomposition property as follows.

Suppose that for every index $i = 1, \dots, d$, $h : \mathbb{R}^d \to \mathbb{R}$ can be decomposed into two functions ${}_ih^2 : \mathbb{R}^2 \to \mathbb{R}$ and ${}_{-i}h^{d-1} : \mathbb{R}^{d-1} \to \mathbb{R}$ such that for all $x \in \mathbb{R}^d$

$$h(x_1, \dots, x_d) = {}_ih^2(x_i, {}_{-i}h^{d-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)). \tag{6.4}$$

Note that in contrast to (6.1), here the decomposition may depend on the index. We want to use the notations $h(X)$ and ${}_{-i}h^{d-1}(X_{-i})$ just as before. So for all $i = 1, \dots, d$

$$h(X) = \begin{pmatrix} h(x_1^{(1)}, & x_1^{(2)}, & \dots, & x_1^{(d)}) \\ h(x_2^{(1)}, & x_2^{(2)}, & \dots, & x_2^{(d)}) \\ \vdots & \vdots & & \vdots \\ h(x_n^{(1)}, & x_n^{(2)}, & \dots, & x_n^{(d)}) \end{pmatrix} = \begin{pmatrix} {}_ih^2\Big(x_1^{(i)}, & {}_{-i}h^{d-1}(X_{-i})_1\Big) \\ {}_ih^2\Big(x_2^{(i)}, & {}_{-i}h^{d-1}(X_{-i})_2\Big) \\ \vdots & \\ {}_ih^2\Big(x_n^{(i)}, & {}_{-i}h^{d-1}(X_{-i})_n\Big) \end{pmatrix}.$$

Notice that after a column $x^{(i)}$ is fixed, the same decomposition with index i is applied in every row.
We now restate Theorem 6.2.1 in generalized form.

Theorem 6.2.12. *If $f(x^{(1)}, \dots, x^{(d)}) = g(h(x^{(1)}, \dots, x^{(d)}))$ where $g : \mathbb{R} \to \mathbb{R}$ is increasing and convex and $h : \mathbb{R}^d \to \mathbb{R}$ is supermodular, componentwise strictly monotonic and satisfies* (6.4) *with all ${}_ih^2$ also being supermodular, then*

$$m_f(X) = \min_{\tilde{X} \in \mathcal{O}_h(X)} \sum_{k=1}^{n} f(\tilde{x}_k^{(1)}, \dots, \tilde{x}_k^{(d)}).$$

Furthermore, the requirement of g being increasing can be dropped, whenever ${}_ih^2$ is of the form ${}_ih^2(x_1, x_2) = \varphi_1(x_1) + \varphi_2(x_2)$, where φ_1 and φ_2 are monotone in the same direction.

Proof. Since the majorization does not depend on the decomposition to be the same for each index, the proof of Theorem 6.2.1 carries over verbatim. The second statement follows from 6.2.10. □

Theorem 6.2.12 contains the statements 6.2.1 and 6.1.1 as special cases. Note that in contrast to (6.1), a function does not have to be symmetric to satisfy (6.4). An important case that is included in Theorem 6.2.12 but not in 6.2.1 or 6.1.1 is a weighted sum.

6.3. Applications

To illustrate the usefulness of the new cases the RA can be applied to, we want to compute some examples. That is, given some marginal distributions $F_1, \dots, F_d$ and a cost function $f : \mathbb{R}^d \to \mathbb{R}$ which fulfills our assumptions, we want to estimate

$$s_f := \inf\left\{\mathbb{E}\left[f(X^{(1)}, \dots, X^{(d)})\right] : X^{(i)} \sim F^{(i)}, 1 \le i \le d\right\}. \tag{6.5}$$

Here, we are only interested in the infimum since it is well known that for supermodular cost functions f, the supremum

$$S_f := \sup\left\{\mathbb{E}\left[f(X^{(1)},\dots,X^{(d)})\right] : X^{(i)} \sim F^{(i)}, 1 \le i \le d\right\}$$

is attained, when $X_1,\dots,X_d$ are comonotonic (i.e. their copula is the upper Fréchet-Hoeffding bound). Hence S_c is known in this case.

Considerig what we have seen in the preceding section, there are two immediate obstacles. The first one is that so far, we always talked about n-discrete distributions and now we want to work with general distributions F_i. This can be adressed by working with the two n-discrete distributions

$$\underline{F}_i = \frac{1}{n}\sum_{k=0}^{n-1} 1_{[q_k^i,\infty)}(x) \quad \text{and} \quad \overline{F}_i = \frac{1}{n}\sum_{k=1}^{n} 1_{[q_k^i,\infty)}(x)$$

where the q_k^i are defined in terms of the quantile functions F_i^{-1} by $q_k^i := F_i^{-1}(\frac{k}{n})$. Writing $\underline{s}_f$ resp. $\overline{s}_f$ for (6.5) with the F_i replaced by $\underline{F}_i$ resp. $\overline{F}_i$, it holds that if the cost function f is componentwise increasing we have

$$\underline{s}_f \le s_f \le \overline{s}_f.$$

This allows us to compute a range for s_f that will become small when n increases.

The second problem is that the rearrangement methodology aims to find m_f and not s_f. The difference here is that for m_f only distributions that give equal mass of $\frac{1}{n}$ to n out of n^d possible realizations of X are considered. However, for n large enough, we have

$$s_f \approx \frac{m_f}{n}.$$

Combining these two ideas leads to the approximation of s_f by the RA. For more details see section 3 in (Puccetti and Rüschendorf, 2015). Note that the requirement to have f componentwise increasing for this method eliminates the increased generality, 6.1.1 had over 6.1.2 and 6.2.1 in terms of monotonicity.

6.3.1. A weighted portfolio

Imagine a portfolio consisting of three different assets with value processes $X_1(t)$, $X_2(t)$ and $X_3(t)$. The value of the portfolio at time $t = 0$ is given by

$$L(0) = \alpha_1 X_1(0) + \alpha_2 X_2(0) + \alpha_3 X_3(0)$$

where α_i denotes the amount of positions held in asset X_i. Furthermore, the return of asset i is given by $R_i := \frac{X_i(1)-X_i(0)}{X_i(0)} \sim F_i$ with known distribution functions F_i. Then the return of the portfolio at time $t = 1$ is given by

$$h(R_1,R_2,R_3) = \frac{L(1)-L(0)}{L(0)} = \frac{\alpha_1 X_1(0)}{L(0)}R_1 + \frac{\alpha_2 X_2(0)}{L(0)}R_2 + \frac{\alpha_3 X_3(0)}{L(0)}R_3.$$

So interpreting $w_i := \frac{\alpha_i X_i(0)}{L(0)} \in [0,1]$ as the fraction of the original wealth that was invested in X_i, we get $h(R_1, R_2, R_3) = w_1 R_1 + w_2 R_2 + w_3 R_3$.
Assuming a guaranteed return of k on the portfolio leaves us to examine $\mathbb{E}\left[g(h(R_1, R_2, R_3))\right]$ with $g(x) = \max(x - k, 0)$. Obviously g and h satisfy the assumptions of Theorem 6.2.12, hence we can use the RA to compute s_f for $f = g \circ h$.

We considered varying sets of distributions F_1, F_2, F_3. Since the cost function f is unbounded from above, the discretizations $\overline{F}_i$ cannot be used in a meaningful way with distributions that are unbounded from above. Therefore, we truncated all distributions with infinite support at the 0.001% quantile and marked the modified distributions with an asterisk $(*)$. Note that in practice, the question where to truncate can be hard to decide, especially when dealing with heavy tails, see e.g. Clark (2013) for considerations concerning the Pareto distribution.
For consistency, also the lower bounds of (positively) unbounded distributions were calculated with the truncated distribution. To have a reference value, we also computed the range of $\mathbb{E}\left[g(h(R_1, R_2, R_3))\right]$ by solving a linear program. This procedure yields rigorous bounds but is computationally more costful and thus usually cannot provide the accuracy of the RA. For more details on the LP method and a comparison of the two approaches see Preischl (2016) and the references therein.

F_1	F_2	F_3	range LP	range RA
$\mathcal{U}([0, 0.4])$	$\mathcal{U}([0.1, 0.5])$	$\mathcal{U}([0, 1])$	0.0058-0.0148	0.0099-0.0100
$exp(1)^*$	$exp(2)^*$	$exp(4)^*$	0.3416-0.4711	0.3749-0.3750
$\mathcal{U}([0, 0.4])$	$exp(3)^*$	$\mathcal{U}([0, 1])$	0.0092-0.0303	0.0166-0.0167
$exp(1)^*$	$Pareto(2)^*$	$\mathcal{N}(0, 0.25)^*$	0.2876-1.4912	0.3990-0.4054

Table 6.1.: Approximation results for different marginals. LP values as a reference.

The RA was applied with a grid of 10^5 sections in each dimension, whereas the LP method used a grid of 60 sections in each dimension. The weight vector was set to $w := (0.5, 0.2, 0.3)$ throughout these computations.

For the LP approach, the computation time depends strongly on the marginal distributions and was on average around 10 minutes per value. The RA generally took less than one minute to compute both upper and lower bound. All results were obtained using the open source language **R**, where we used the implementation of the RA from the package *qrmtools* by (Hofert and Hornik, 2016).

Bibliography

Christoph Aistleitner and Josef Dick. Functions of bounded variation, signed measures, and a general Koksma-Hlawka inequality. *Acta Arith.*, 167(2):143–171, 2015.

Christoph Aistleitner, Florian Pausinger, Anne Marie Svane, and Robert F Tichy. On functions of bounded variation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 162, pages 405–418. Cambridge University Press, 2017.

Hansjörg Albrecher and Reinhold Kainhofer. Risk theory with a nonlinear dividend barrier. *Computing*, 68(4):289–311, 2002. ISSN 0010-485X.

Hansjörg Albrecher and Stefan Thonhauser. Optimality results for dividend problems in insurance. *RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 103(2):295–320, 2009.

Hansjörg Albrecher, Corina Constantinescu, Gottlieb Pirsic, Georg Regensburger, and Markus Rosenkranz. An algebraic operator approach to the analysis of Gerber–Shiu functions. *Insurance: Mathematics and Economics*, 46(1):42–51, 2010.

Hansjörg Albrecher, Jozef L Teugels, and Jan Beirlant. *Reinsurance: actuarial and statistical aspects*. John Wiley & Sons, 2017.

Søren Asmussen and Hansjörg Albrecher. *Ruin probabilities*. World Scientific Publishing Co. Pte. Ltd., 2nd edition, 2010.

Kendall E Atkinson. The numerical solution of Fredholm integral equations of the second kind. *SIAM J. Numer. Anal.*, 4(3):337–348, 1967.

Pablo Azcue and Nora Muler. Optimal reinsurance and dividend distribution policies in the Cramér-Lundberg model. *Math. Finance*, 15(2):261–308, 2005.

Pablo Azcue and Nora Muler. *Stochastic optimization in insurance: a dynamic programming approach*. Springer, 2014.

Mathias Beiglböck and Claus Griessler. An optimality principle with applications in optimal transport. *arXiv preprint arXiv:1404.7054*, 2014.

Carole Bernard, Xiao Jiang, and Ruodu Wang. Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, 54:93–108, 2014.

Carole Bernard, Michel Denuit, and Steven Vanduffel. Measuring portfolio risk under partial dependence information. *Journal of Risk and Insurance*, 85(3):843–863, 2018.

Valeria Bignozzi, Giovanni Puccetti, and Ludger Rüschendorf. Reducing model risk via positive and negative dependence assumptions. *Insurance Math. Econom.*, 61:17–26, 2015.

Luca Brandolini, Leonardo Colzani, Giacomo Gigante, and Giancarlo Travaglini. On the Koksma–Hlawka inequality. *J. Complexity*, 29(2):158 – 172, 2013a.

Luca Brandolini, Leonardo Colzani, Giacomo Gigante, and Giancarlo Travaglini. A Koksma–Hlawka inequality for simplices. In *Trends in Harmonic Analysis*, pages 33–46. Springer, 2013b.

W Brattin. Quantitative uncertainty analysis of radon risks attributable to drinking water. In *Proceedings of the Toxicology Forum, The Given Institute of Pathology, Aspen, Colorado, July*, volume 11, pages 338–350, 1994.

Damiano Brigo, Andrea Pallavicini, and Roberto Torresetti. *Credit models and the crisis: A journey into CDOs, copulas, correlations and dynamic models*. John Wiley & Sons, 2010.

Hermann Brunner. Iterated collocation methods and their discretizations for Volterra integral equations. *SIAM J. Numer. Anal.*, 21(6):1132–1145, 1984.

Hermann Brunner. Implicitly linear collocation methods for nonlinear Volterra equations. *Appl. Numer. Math.*, 9(3):235–247, 1992.

Rainer E Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment Problems, Revised Reprint*. Siam, 2009.

Mathieu Cambou, Marius Hofert, and Christiane Lemieux. Quasi-random numbers for copula models. *Statistics and Computing*, 27(5):1307–1329, 2017.

Arian Cani and Stefan Thonhauser. An optimal reinsurance problem in the Cramér–Lundberg model. *Mathematical methods of operations research*, 85(2):179–205, 2017.

David R Clark. A note on the upper-truncated Pareto distribution. In *Casualty Actuarial Society E-Forum, Winter*, pages 1–22, 2013.

Jeffrey F Collamore et al. Importance sampling techniques for the multidimensional ruin problem for general markov additive sequences of random vectors. *The Annals of Applied Probability*, 12(1):382–421, 2002.

M. H. A. Davis. *Markov Models and Optimization*. Chapman & Hall, London, 1993. ISBN 0-412-31410-X.

Bruno De Finetti. Su un'impostazione alternativa della teoria collettiva del rischio. In *Transactions of the XVth international congress of Actuaries*, volume 2, pages 433–443. New York, 1957.

Amir Dembo, Samuel Karlin, Ofer Zeitouni, et al. Large exceedances for multidimensional lévy processes. *The Annals of Applied Probability*, 4(2):432–447, 1994.

Josef Dick, Peter Kritzer, Frances Y Kuo, and Ian H Sloan. Lattice-Nyström method for Fredholm integral equations of the second kind with convolution type kernels. *J. Complexity*, 23(4):752–772, 2007.

Michael Drmota and Robert F. Tichy. *Sequences, discrepancies and applications*, volume 1651 of *Lecture Notes in Mathematics*. Springer, Berlin, 1997.

Francois Dufresne and Hans U Gerber. Risk theory for the compound poisson process that is perturbed by diffusion. *Insurance: mathematics and economics*, 10(1):51–59, 1991.

Fabrizio Durante and Juan Fernández Sánchez. On the approximation of copulas via shuffles of min. *Statist. Probab. Lett.*, 82(10):1761–1767, 2012.

Fabrizio Durante, Juan Fernández-Sánchez, José Juan Quesada-Molina, and Manuel Úbeda-Flores. Convergence results for patchwork copulas. *European J. Oper. Res.*, 247(2):525–531, 2015.

Herbert Edelsbrunner and Florian Pausinger. Approximation and convergence of the intrinsic volume. *Adv. Math.*, 287:674–703, 2016.

Jana Fialová and Oto Strauch. On two-dimensional sequences composed by one-dimensional uniformly distributed sequences. *Unif. Distrib. Theory*, 6(1):101–125, 2011.

Hans U. Gerber and Elias S. W. Shiu. On the time value of ruin. *N. Am. Actuar. J.*, 2 (1):48–78, 1998. ISSN 1092-0277.

Hans U. Gerber and Elias S. W. Shiu. The time value of ruin in a Sparre Andersen model. *N. Am. Actuar. J.*, 9(2):49–84, 2005. ISSN 1092-0277.

Hans U Gerber and Elias SW Shiu. On optimal dividends: from reflection to refraction. *Journal of Computational and Applied Mathematics*, 186(1):4–22, 2006.

M Götz. Discrepancy and the error in integration. *Monatsh. Math.*, 136(2):99–121, 2002.

Claus Griessler. c-cyclical monotonicity as a sufficient criterion for optimality in the multi-marginal Monge-Kantorovich problem. *arXiv preprint arXiv:1601.05608*, 2016.

Alexander Grigorian. Ordinary differential equation. Lecture notes, available at `https://www.math.uni-bielefeld.de/~grigor/odelec2009.pdf`, 2009.

Sun-Wei Guo. Familial aggregation of environmental risk factors and familial aggregation of disease. *American journal of epidemiology*, 151(11):1121–1131, 2000.

Morten Hald and Hanspeter Schmidli. On the maximisation of the adjustment coefficient under proportional reinsurance. *ASTIN Bulletin: The Journal of the IAA*, 34(1):75–83, 2004.

Glyn Harman. Variations on the Koksma-Hlawka inequality. *Unif. Distrib. Theory*, 5 (1):65–78, 2010. ISSN 1336-913X.

Christian Hipp and Michael Taksar. Optimal non-proportional reinsurance control. *Insurance: Mathematics and Economics*, 47(2):246–254, 2010.

Christian Hipp and Michael Vogt. Optimal dynamic XL- reinsurance. *ASTIN Bulletin: The Journal of the IAA*, 33(2):193–207, 2003.

Edmund Hlawka. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Ann. Mat. Pura Appl.*, 54(1):325–333, 1961.

Markus Hofer and Maria Rita Iacò. Optimal bounds for integrals with respect to copulas and applications. *J. Optim. Theory Appl.*, 161(3):999–1011, 2014.

Marius Hofert and Kurt Hornik. *qrmtools: Tools for Quantitative Risk Management*, 2016. URL `https://CRAN.R-project.org/package=qrmtools`. R package version 0.0-6.

Loo Keng Hua and Yuan Wang. *Applications of Number theory to Numerical Analysis*. Springer, Berlin-New York; Kexue Chubanshe (Science Press), Beijing, 1981. Translated from the Chinese.

Maria Rita Iacò, Stefan Thonhauser, and Robert F Tichy. Distribution functions, extremal limits and optimal transport. *Indag. Math.*, 26(5):823–841, 2015.

Yasuhiko Ikebe. The Galerkin method for the numerical solution of Fredholm integral equations of the second kind. *Siam Review*, 14(3):465–491, 1972.

Edgars Jakobsons and Ruodu Wang. Negative dependence in matrix arrangement problems. *Annals of Operations Research*, pages 1–22, 2017.

Olav Kallenberg. *Foundations of Modern Probability*. Springer Verlag, New York, 2002.

L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974. Pure and Applied Mathematics.

Sunil Kumar and Ian H Sloan. A new collocation-type method for Hammerstein integral equations. *Math. Comp.*, pages 585–593, 1987.

Frances Y Kuo. Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces. *J. Complexity*, 19(3):301–320, 2003.

Harold Kushner and Paul G Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24. Springer Science & Business Media, 2013.

Gunther Leobacher and Friedrich Pillichshammer. *Introduction to Quasi-Monte Carlo Integration and Applications*. Compact Textbook in Mathematics. Birkhäuser/Springer, Cham, 2014.

X Sheldon Lin, Gordon E Willmot, and Steve Drekic. The classical risk model with a constant dividend barrier: analysis of the Gerber–Shiu discounted penalty function. *Insurance Math. Econom.*, 33(3):551–566, 2003.

F. Lundberg. *Approximerad framställning av sannolikhetsfunktionen. Aterförsäkring av kollektivrisker*. Akad. Afhandling. Almqvist o. Wiksell, Uppsala, 1903.

Thibaut Lux, Antonis Papapantoleon, et al. Improved fréchet–hoeffding bounds on d-copulas and applications in model-free finance. *The Annals of Applied Probability*, 27 (6):3633–3671, 2017.

Jan-Frederik Mai and Matthias Scherer. *Financial Engineering with Copulas Explained*. Palgrave Macmillan, 2014.

Athena Makroglou. Numerical solution of some second order integro-differential equations arising in ruin theory. In *Proceedings of the 3rd Conference in Actuarial Science and Finance*, pages 2–5, 2004.

Albert W Marshall, Ingram Olkin, and Barry C Arnold. *Inequalities: Theory of Majorization and Its Applications*. New York, NY: Springer New York, 2011.

Piotr Mikusiński and Michael D Taylor. Some approximations of n-copulas. *Metrika*, 72 (3):385–414, 2010.

Piotr Mikusinski, Howard Sherwood, and Michael D Taylor. Shuffles of min. *Stochastica*, 13(1):61–74, 1992.

Roger B Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2007.

Harald Niederreiter. Low-discrepancy and low-dispersion sequences. *Journal of number theory*, 30(1):51–70, 1988.

Erich Novak and Henryk Woźniakowski. *Tractability of multivariate problems. Volume II: Standard information for functionals*, volume 12 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010.

Art B. Owen. Multidimensional variation for Quasi-Monte Carlo. In *Contemporary multivariate analysis and design of experiments*, volume 2 of *Ser. Biostat.*, pages 49–74. World Sci. Publ., Hackensack, NJ, 2005.

Florian Pausinger and Anne Marie Svane. A Koksma-Hlawka inequality for general discrepancy systems. *J. Complexity*, 31(6):773–797, 2015. ISSN 0885-064X.

Michael Preischl. Bounds on integrals with respect to multivariate copulas. *Dependence Modeling*, 4(1), 2016.

Michael Preischl. On majorization in dependence modeling. *arXiv preprint arXiv:1709.04728*, 2017.

Michael Preischl and Stefan Thonhauser. Optimal reinsurance for gerber–shiu functions in the cramér-lundberg model. *Insurance: Mathematics and Economics*, 87:82–91, 2019.

Michael Preischl, Stefan Thonhauser, and Robert F Tichy. Integral equations, quasi-monte carlo methods and risk modeling. In *Contemporary Computational Mathematics-A Celebration of the 80th Birthday of Ian Sloan*, pages 1051–1074. Springer, 2018.

Giovanni Puccetti and Ludger Rüschendorf. Computation of sharp bounds on the distribution of a function of dependent risks. *J. Comput. Appl. Math.*, 236(7):1833–1840, 2012.

Giovanni Puccetti and Ludger Rüschendorf. Computation of sharp bounds on the expected value of a supermodular function of risks with given marginals. *Comm. Statist. Simulation Comput.*, 44(3):705–718, 2015.

Giovanni Puccetti and Ruodu Wang. Extremal dependence concepts. *Statist. Sci.*, 30 (4):485–517, 11 2015. doi: 10.1214/15-STS525. URL `http://dx.doi.org/10.1214/15-STS525`.

Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugels. *Stochastic processes for insurance and finance*. John Wiley & Sons Ltd., Chichester, 1999.

Ludger Rüschendorf. Inequalities for the expectation of δ-monotone functions. *Probab. Theory Related Fields*, 54(3):341–349, 1980.

Arne Sandstrom. *Solvency: models, assessment and regulation*. Chapman and Hall/CRC, 2005.

Friedrich Schmid, Rafael Schmidt, Thomas Blumentritt, Sandra Gaißer, and Martin Ruppert. Copula-based measures of multivariate association. In *Copula Theory and its Applications*, pages 209–236. Springer, 2010.

Hanspeter Schmidli. Optimal proportional reinsurance policies in a dynamic setting. *Scandinavian Actuarial Journal*, 2001(1):55–68, 2001.

Hanspeter Schmidli. On minimizing the ruin probability by investment and reinsurance. *The Annals of Applied Probability*, 12(3):890–907, 2002.

Hanspeter Schmidli. *Stochastic control in insurance*. Springer Science & Business Media, 2008.

M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

Ian H Sloan. A quadrature-based approach to improving the collocation method. *Numer. Math*, 54(1):41–56, 1988.

Ian H Sloan and James N Lyness. The representation of lattice quadrature rules as multiple sums. *Math. Comp.*, 52(185):81–94, 1989.

Ian H Sloan and Henryk Woźniakowski. Tractability of multivariate integration for weighted Korobov classes. *J. Complexity*, 17(4):697–721, 2001.

Josef Stoer and Roland Bulirsch. *Numerische Mathematik. 2.* Springer-Lehrbuch. [Springer Textbook]. Springer, Berlin, fourth edition, 2000.

Qihe Tang and Li Wei. Asymptotic aspects of the gerber–shiu function in the renewal risk model using wiener–hopf factorization and convolution equivalence. *Insurance: Mathematics and Economics*, 46(1):19–31, 2010.

Robert F. Tichy. Über eine zahlentheoretische Methode zur numerischen Integration und zur Behandlung von Integralgleichungen. *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II*, 193(4-7):329–358, 1984. ISSN 0723-9319.

S Twomey. On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *J. ACM*, 10(1):97–101, 1963.

Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.

Bin Wang and Ruodu Wang. The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal.*, 102(10):1344–1360, 2011.

Ruodu Wang, Liang Peng, and Jingping Yang. Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417, 2013.

Howard R Waters. Some mathematical aspects of reinsurance. *Insurance: Mathematics and Economics*, 2(1):17–26, 1983.