Florian Nachbagauer, BSc

On refined QMC methods for finance

Master's Thesis

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Financial and Actuarial Mathematics

submitted to

Graz University of Technology

Supervisor:
Priv.-Doz. Dipl.-Ing. Dr.techn. Stefan Michael Thonhauser

Institute of Statistics

Graz, March 2018

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

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 master's thesis.

Datum/Date	Unterschrift/Signature

Zusammenfassung

Monte Carlo und Quasi-Monte Carlo (QMC) Methoden sind ein essentielles Werkzeug, sowohl in der Finanz-, als auch in der Versicherungsmathematik, in Anbetracht der Notwendigkeit der Bewertung von Lebensversicherungen mit eingebetteten Optionen. Viele Probleme resultieren in hoch-dimensionalen Integralen, für welche herkömmliche Quadratur-Regeln nicht geeignet sind. Die klassische Fehlerabschätzung für QMC Methoden, bekannt als die Koksma-Hlawka Ungleichung, setzt einen Integranden mit endlicher Variation voraus, was in vielen praktischen Fällen nicht gegeben ist. Trotzdem werden QMC Methoden erfolgreich auch auf Integrale angewendet, welche die üblichen Bedingungen nicht erfüllen. In dieser Arbeit werden Transformationen, welche beschränkte Integranden liefern, sowie Möglichkeiten zur Glättung von nicht differenzierbaren Integranden präsentiert, was uns schließlich ermöglicht, Integranden mit endlicher Variation zu konstruieren. Wir werden zeigen, dass bestimmte Fälle von nicht differenzierbaren Funktionen endliche Variation besitzen. In anderen Fällen werden wir die Funktion in einen glatten Anteil und in einen kleinen, nicht glatten Rest zerlegen. Auf dem glatten Anteil lässt sich die Koksma-Hlawka Ungleichung anwenden, während wir die Fehlerabschätzung auf dem nicht glatten Rest mittels gewöhnlicher Monte Carlo Integration und einem Konfidenzintervall durchführen. Die verschiedenen Methoden werden mit Ergebnissen aus MATLAB-Berechnungen veranschaulicht.

Abstract

Monte Carlo and quasi-Monte Carlo (QMC) methods are an essential tool in mathematical finance as well as in actuarial mathematics, considering the need for the valuation of life insurance contracts with embedded options. Common problems often result in high-dimensional integrals, where traditional quadrature rules fail. The classical error estimate for QMC methods, known as the Koksma-Hlawka inequality, requires the integrand to have finite variation, which is not given in many practical cases. Nonetheless, QMC methods are successfully applied to integrals, which do not fulfil the usual conditions. In the context of this work, we will discuss transformations, which yield bounded integrands as well as methods for smoothing non-differentiable integrands, which ultimately allows us to construct integrands with finite variation. For special cases of non-differential functions, we will prove, that they have finite variation. In other cases, we will decompose those functions in a smooth part with finite variation and a small, non-smooth rest. While the Koksma-Hlawka inequality is applicable on the smooth part, ordinary Monte Carlo can be used in combination with a confidence interval, to estimate the error on the non-smooth rest. The different methods considered in this work are illustrated with results obtained from calculations performed in MATLAB.

Contents

Abstract vi

Notation and Abbreviations ix

Motivation and Overview xi

1 Introduction to Monte Carlo Methods 1
- 1.1 Ordinary Monte Carlo Method . . . 1
- 1.2 Random Number Generation . . . 4
  - 1.2.1 Inverse Transform Method . . . 4
  - 1.2.2 The Normal Distribution . . . 4
  - 1.2.3 The Box-Muller Method . . . 5
  - 1.2.4 Generating Multivariate Normals . . . 5
- 1.3 Introduction to QMC . . . 6
- 1.4 Discrepancy . . . 8
- 1.5 Variation in the Sense of Hardy and Krause . . . 13
- 1.6 Koksma-Hlawka Inequality . . . 20
- 1.7 Randomized QMC . . . 21
  - 1.7.1 Additive Shift . . . 22
  - 1.7.2 Digital Shift . . . 23
  - 1.7.3 Scrambled Nets . . . 23

2 Introduction to Mathematical Finance 25
- 2.1 The General Diffusion Model . . . 25
- 2.2 The Fundamental Theorems of Asset Pricing . . . 29
- 2.3 The Black-Scholes Model . . . 32

3 Practical Aspects 37
- 3.1 Transformation to the Unit Cube . . . 38
  - 3.1.1 Linear Transformation . . . 38
  - 3.1.2 Truncation Error . . . 39
  - 3.1.3 Bounds on the Derivatives of the Normal Density . . . 41
  - 3.1.4 Inverse Normal CDF Transformation . . . 43
  - 3.1.5 Logit Transformation . . . 44
- 3.2 Factorization of the Covariance Matrix . . . 48

4 Non-Differentiable Functions **51**
4.1 Finite Variation of the Maximum Function 51
4.2 Approximation with Smooth Functions 56
4.2.1 The Indicator Function 56
4.2.2 The Positive-Part Function 57
4.2.3 Min, Max and Minmax-Functions 58
4.2.4 Indicator Function on a Polyhedron 58
4.2.5 Indicator Function on a Set with Restrictions 59
4.2.6 Convolution with a Bump Function 59
4.3 A Cliquet-Type Payoff 62

5 Practical Examples **67**
5.1 A Comparison between MC, QMC and RQMC 67
5.2 A Smoothed Indicator Function 69
5.3 Cliquet-Type Payoff 73
5.4 Transformations and Matrix Factorizations 75

6 Conclusion **79**

Notation and Abbreviations

In this work we will use the following notations and abbreviations:

MC ... Monte Carlo

QMC ... Quasi-Monte Carlo

RQMC ... Randomized Quasi-Monte Carlo

R.V. ... random variable

a.s. ... almost surely

PDF ... probability density function

CDF ... cumulative distribution function

iid ... independent and identical distributed

$\xrightarrow{\boldsymbol{d}}$ ... $X_N \xrightarrow[N\to\infty]{d} X$, convergence in distribution

$\boldsymbol{I_s}$ ... the $s \times s$ identity matrix

$\boldsymbol{\mathcal{I}_s}$ ... the index set $\{1, 2, \ldots, s\}$

$\mathbb{1}$ ... the 1-vector $(1, 1, \ldots, 1)^T$, the number of entries should be clear from the context

$\boldsymbol{U(0,1)}$ or $U([0,1])$ ... the uniform distribution on $[0,1]$, s.t. $X \sim U(0,1)$ if and only if its CDF is $F_X(t) = \mathbb{P}[X \leq t] = t$ for $t \in [0,1]$

$\boldsymbol{U([0,1]^s)}$ ... the multivariate uniform distribution on $[0,1]^s$, s.t. the vector $X = (X_1, \ldots, X_s) \sim U([0,1]^s)$ if and only if $X_j \overset{iid}{\sim} U(0,1)$ for all $j \in \mathcal{I}_s$

$\boldsymbol{a \leq b}$ for **vectors** $a, b \in \mathbb{R}^s$ ... we say $a \leq b$ if and only if $a_j \leq b_j$ for all $j \in \mathcal{I}_s$

$\boldsymbol{[a,b]}$ for **vectors** $a \leq b \in \mathbb{R}^s$ ... the hyperrectangle which is the Cartesian product $[a,b] = [a_1, b_1] \times \ldots \times [a_s, b_s]$

$\boldsymbol{\partial^u}$ for a set $u \subseteq \mathcal{I}_s$ ... the differential operator $\partial^u = \prod_{j \in u} \frac{\partial}{\partial x_j}$ which takes the first derivative with respect to each component variable

$\boldsymbol{\lfloor x \rfloor}$ for any $x \in \mathbb{R}$ ... the largest integer $n \leq x$, $\lfloor x \rfloor := \max\{n \in \mathbb{N} | n \leq x\}$

$\boldsymbol{(\cdot)^+}$ ... the positive part function, $(x)^+ = \max\{x, 0\}$

$(\cdot)^-$ ... the negative part function, $(x)^- = \max\{-x, 0\}$

$(\boldsymbol{a \pm \varepsilon})$ ... the open interval $(a - \varepsilon, a + \varepsilon)$

$[\boldsymbol{a \pm \varepsilon}]$ ... the closed interval $[a - \varepsilon, a + \varepsilon]$

Motivation and Overview

Monte Carlo methods play an essential role in today's context of scientific computing, especially in mathematical finance. Basically, every method involving the use of random numbers to estimate some quantity is called a Monte Carlo method. Our primary focus lies on the calculation of an integral over an s-dimensional domain. Because an analytic expression for the antiderivative cannot be found in most cases, one has to resort to numerical methods. For high dimensional integrals, which are common in the field of financial and actuarial mathematics, traditional quadrature rules often fail because of the "curse of dimensionality", while Monte Carlo methods are still able to produce good results in many cases, with a convergence rate of $O(N^{-1/2})$. A higher rate of convergence can be achieved with the use of "low-discrepancy sequences" instead of random numbers, which is then called quasi-Monte Carlo method. The elements of a low-discrepancy sequence are not random (and also not independent), but they have very good distributional properties, which results in a better convergence rate. There are some restrictions to the applicability of quasi-Monte Carlo methods, e.g. the dimension of the problem has to be bounded, and it must be known a priori, which does not have to be the case for "ordinary" Monte Carlo. What makes quasi-Monte Carlo special is, that we have a deterministic error estimate, as in contrary to ordinary Monte Carlo, where we have only a probabilistic error estimate (i.e. a confidence interval). This estimate, known as the Koksma-Hlawka inequality, consists of two factors, namely the discrepancy, which is a measure for the uniformity of the sequence used for the integration nodes, and the variation of the integrand, or more specifically, the variation in the sense of Hardy and Krause. Both of these terms, the discrepancy and the variation, are rather hard to calculate, or even to find a good estimate. Another difficulty is, that the variation of the integrand is unbounded for many integrands resulting from problems in mathematical finance, and thus the error estimate is useless. The reason for this is twofold: Low-discrepancy sequences are always defined on the s-dimensional unit cube $[0,1]^s$, while the problems considered in this work result in integrals over the domain $\mathbb{R}^s$, and because of that, one needs to apply some transformation. Depending on the choice of transformation, the resulting integrand can be bounded or unbounded, and an unbounded integrand always implies unbounded variation. The second reason is, that most payoff-functions, which define the structure of our integrands, are not differentiable, because they involve some maximum-function etc. Although differentiability is only a sufficient, but not a necessary criterion for the variation to be finite, the most convenient way for estimating the variation of a function is by estimating its derivatives, which is not possible, if the function is not differentiable. Despite these theoretical difficulties, Monte Carlo and quasi-Monte Carlo methods are successfully applied in practice, and the question, why quasi-Monte Carlo is performing so well, is a current field of research.

In this work, we will discuss different transformations and present methods for approx-

imating non-differentiable functions often involved in mathematical finance with smooth functions with finite variation. Also, we are going show, that special cases of non-differentiable functions do have finite variation. Chapter 1 aims to give an introduction on Monte Carlo methods in general and the theory behind quasi-Monte Carlo methods and the error estimation. Chapter 2 is a formal introduction into the modelling world of mathematical finance and the Black-Scholes model in particular, which is the main source for the problems considered in Chapter 5. In Chapter 3, we discuss advantages and disadvantages of different transformations and the different ways to incorporate dependency. The topic of Chapter 4 is non-differentiable functions, with the focus on ways to estimate the variation and ways for smoothing functions with unbounded variation, thus making the variation finite. In Chapter 5, we present some practical and numerical examples, with a comparison of the performance of the different methods.

Chapter 1

Introduction to Monte Carlo Methods

As already said in the motivation, our primary focus lies on the calculation of an integral over an s-dimensional domain. Monte Carlo methods make the use of random numbers in order to approximate such an integral numerically. This chapter should explain the basics of Monte Carlo and quasi-Monte Carlo methods and is structured in the following manner: In Section 1.1, we will give an overview about the ordinary Monte Carlo method, which uses pseudo-random numbers, with the main sources being the introductory chapters in Glasserman [5], Leobacher & Pillichshammer [15] and Dick & Pillichshammer [3]. Section 1.2 deals with the generation of random numbers, which lies at the heart of every Monte Carlo simulation. Beginning with Section 1.3, our main focus will be the quasi-Monte Carlo method, which uses specially constructed numbers instead of pseudo-random numbers to achieve better rates of convergence. Sections 1.4 to 1.6 are about the error analysis of QMC methods and preliminary definitions.

1.1 Ordinary Monte Carlo Method

Our aim is to to approximate the integral with some quadrature rule

$$I_N(f) \approx \int_{[0,1]^s} f(x)dx,$$

which is defined as

$$I_N(f) := \frac{1}{N}\sum_{n=1}^{N} w_n f(x_n)$$

with weights w_n and integration nodes x_n.

In one dimension, the midpoint rule or trapezoidal rule work for most cases. Additionally, if the integrand is sufficiently smooth, a higher order of convergence can be achieved e.g. using Simpson's rule. For example, with the trapezoidal rule, the numeric estimate converges with a speed of $O(N^{-2})$ for a twice differentiable function f, which means, the integration error is $|I_N(f) - \int_0^1 f(x)dx| = O(N^{-2})$. For a four times differentiable function f , the Simpson's rule has a convergence of $|I_N(f) - \int_0^1 f(x)dx| = O(N^{-4})$. Another option would be the Gauss quadrature, which yields very good results for functions that

are well approximated by polynomials, because with N integration nodes, it is possible to calculate the integral of a polynomial with a degree of $2N-1$ or less *exactly*.

However, as the dimension s increases, these methods show a big deficiency in the order of convergence. In higher dimensions, these rules are constructed as the s-fold Cartesian product of the one dimensional quadrature points, hence they form a regular lattice in the unit cube. For the total number of quadrature points we have $N = m^s$, where m is the number of points in each coordinate dimension. The error of the trapezoidal rule is then $O(m^{-2})$, but written in terms of N, the error is

$$\left| I_N(f) - \int_{[0,1]^s} f(x)dx \right| = O(N^{-2/s}),$$

which is, even for moderate s, a poor rate of convergence, as the problem dimension can be in the hundreds with practical problems. The so called "curse of dimensionality" refers to this difficulty when dealing with high dimensional problems. Because of this, we might want to try a different approach, utilizing results from probability theory.

Proposition 1.1.1 (**Strong Law of Large Numbers** [6]). *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables and assume that $\mathbb{E}[|X_1|] < \infty$. Let $\overline{X}_N$ be their arithmetic mean, $\overline{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$. We have*

$$\lim_{N\to\infty} \overline{X}_N = \mathbb{E}[X_1]$$

with probability 1.

Consider now a random variable X, uniformly distributed in $[0,1]^s$. The fact

$$\alpha := \mathbb{E}[f(X)] = \int_{[0,1]^s} f(x)dx \tag{1.1.1}$$

makes it possible, to draw iid-samples $X_1, X_2, \ldots$ from the $U([0,1]^s)$-distribution and use their arithmetic mean to estimate the integral

$$\hat{\alpha}_N := \frac{1}{N}\sum_{n=1}^{N} f(X_n) \approx \int_{[0,1]^s} f(x)dx. \tag{1.1.2}$$

This Monte Carlo estimator is unbiased because of equation (1.1.1) and Proposition 1.1.1 guarantees convergence.

Let

$$\sigma_f^2 := \int_{[0,1]^s} (f(x)-\alpha)^2 dx \tag{1.1.3}$$

be the **variance** of f. If σ_f^2 is finite, which is equivalent to the condition that f is square-integrable, we have

$$\mathrm{Var}[\hat{\alpha}_N] = \frac{\sigma_f^2}{N}.$$

A proof can be found in [3], Theorem 1.5. Applying Jensen's inequality, for the expected error, one derives

$$\mathbb{E}\left[\left| \int_{[0,1]^s} f(x)dx - \hat{\alpha}_N \right|\right] \leq \sqrt{\mathrm{Var}[\hat{\alpha}_N]} = \frac{\sigma_f}{\sqrt{N}},$$

which is of the order $O(N^{-\frac{1}{2}})$, because σ_f^2 does not depend on N. This rate of convergence holds for all dimensions s and is superior to the trapezoidal rule already for $s = 4$, but that does not mean, that ordinary Monte Carlo breaks the curse of dimension, because σ_f is usually not independent of s. A remaining problem is the computation of σ_f^2, because it is again an s-dimensional integral, which also requires the knowledge of α. We are however safe, if N is large enough, to use the sample variance

$$s_{f,N}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(f(X_n) - \hat{\alpha}_N)^2 \tag{1.1.4}$$

as a replacement for σ_f^2, because $\mathbb{E}[s_{f,N}^2] = \sigma_f^2$ and $s_{f,N}^2 \xrightarrow[N\to\infty]{\text{a.s.}} \sigma_f^2$. For the actual computation of $s_{f,N}^2$, the same realizations of X_n can be used as in the computation of $\hat{\alpha}_N$. Now we are going to present two well-known results from probability theory, in order to derive a confidence interval with the Monte Carlo estimate.

Proposition 1.1.2 (**Central Limit Theorem** [6]). *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}[|X_1|]^2 < \infty$, and let $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \mathrm{Var}[X_1]$ be their expectation and variance. We have*

$$\frac{\overline{X}_N - \mu}{\frac{\sigma}{\sqrt{N}}} \xrightarrow[N\to\infty]{d} N(0,1).$$

Proposition 1.1.3 (**Slutsky**). *Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be sequences of random variables. If there exists a random variable X and a constant c, such that*

$$X_N \xrightarrow[N\to\infty]{d} X \quad \text{and} \quad Y_N \xrightarrow[N\to\infty]{d} c,$$

we have

$$X_N Y_N \xrightarrow[N\to\infty]{d} cX.$$

With the help of the above propositions, we are able to derive

$$\frac{\hat{\alpha}_N - \alpha}{\frac{s_{f,N}}{\sqrt{N}}} = \frac{\hat{\alpha}_N - \alpha}{\frac{\sigma_f}{\sqrt{N}}} \cdot \underbrace{\frac{\frac{\sigma_f}{\sqrt{N}}}{\frac{s_{f,N}}{\sqrt{N}}}}_{\to 1 \text{ a.s.}} \xrightarrow[N\to\infty]{d} N(0,1),$$

thus, for "large" N, one can say,

$$\hat{\alpha}_N - \alpha \overset{\text{approx.}}{\sim} N(0, \frac{s_{f,N}^2}{N}). \tag{1.1.5}$$

We are now able to construct a level-δ confidence interval for the the actual value α as

$$\Big[\hat{\alpha}_N - z_{\delta/2}\frac{s_{f,N}}{\sqrt{N}}, \hat{\alpha}_N + z_{\delta/2}\frac{s_{f,N}}{\sqrt{N}}\Big],$$

with $z_{\delta/2} = \Phi^{-1}(1-\delta/2)$ being the quantile of the normal distribution. There is literature dealing also with non-asymptotic error estimates, e.g. see Graham and Talay [6].

Despite the easy applicability of ordinary Monte Carlo methods, a convergence rate of $O(1/\sqrt{N})$ is still to slow for many practical cases. Beginning with Section 1.3, we will see quasi-Monte Carlo (QMC) methods, which achieve convergence rate of at least $O((\log N)^s/N)$.

1.2 Random Number Generation

In order to apply any Monte Carlo method, we need some method, which supplies us with an arbitrary amount of random numbers. There are well established algorithms, with implementations in almost any programming language, which shall meet our needs. Although we treat the numbers generated by such algorithms as random, they are not truly random, but actually purely deterministic, hence such algorithms are called **pseudo-random number generators**. For generating random numbers uniformly distributed on $[0, 1]$, one of the most widespread algorithms is the Mersenne Twister (MT19937), developed 1997 by Matsumoto and Nishimura [18]. An introduction and overview of different random number generators can be found in Chapter 2 of [5].

As nearly all random number generators are constructed to generate uniformly distributed samples, the question arises, how to get from the uniform distribution to some other distribution F. The most straightforward way to do this is the inverse transform method, which we shall see in the following section.

1.2.1 Inverse Transform Method

A real random variable X with cumulative distribution function $F : \mathbb{R} \to [0, 1]$ fulfils $\mathbb{P}[X \leq x] = F(x)$ per definition. For the sake of simplicity let's assume that F is strictly increasing, which is equivalent to its probability density function f being positive. In this case the inverse F^{-1} is well-defined on the interval $[0, 1]$. Let $V \sim U([0, 1])$ be given, i.e. $\mathbb{P}[V \leq v] = v$. By defining X as

$$X = F^{-1}(V)$$

we can see, that X has the desired distributional property via

$$\begin{aligned} \mathbb{P}[X \leq x] &= \mathbb{P}[F^{-1}(V) \leq x] \\ &= \mathbb{P}[V \leq F(x)] \\ &= F(x). \end{aligned} \tag{1.2.1}$$

In cases where F is not strictly increasing, but just non-decreasing, the classical inverse F^{-1} does not exist. By defining

$$F^{-1}(u) := \inf\{x : F(x) \geq u\},$$

this problem can be avoided, and the equation (1.2.1) holds again, as can be seen in [5]. So in order to sample from any distribution F, we first generate samples from the uniform distribution and as a second step, apply the the inverse c.d.f. F^{-1} on them, to receive samples with the desired distribution.

1.2.2 The Normal Distribution

Definition 1.2.1. A random variable X is **normally distributed** with parameters μ and $\sigma > 0$, if it has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2} \Big).$$

Because neither the CDF of a normal distribution, nor its inverse can be written in terms of elementary functions, numeric approximations have to be used. However, available approximations are highly accurate and aren't more complex than the evaluation of functions e.g. involving exp or log, although these are elementary functions.

Definition 1.2.2. A random vector $X = (X_1, \dots, X_s)$ is **(multivariate) normally distributed** with mean $\mu \in \mathbb{R}^s$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{s\times s}$, if it has the joint PDF

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^s |\det(\Sigma)|}} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\Big).$$

A random vector $Y = (Y_1, \dots, Y_s)$ is said to be (multivariate) **standard** normally distributed, if it is normally distributed with μ being the s-dimensional zero-vector, and Σ being the $(s \times s)$-identity matrix. In this case, the PDF is just the s-fold product of the PDFs of one-dimensional standard normal distributions.

1.2.3 The Box-Muller Method

For generating pairs of samples from the normal distribution, a somewhat more elegant way than the inverse transform method can be obtained by the following observation [15]:

Consider a two-dimensional standard normal vector $X = (X_1, X_2)$, i.e., $\mu = (0, 0)$ and $\Sigma = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Via substitution to polar coordinates we have

$$\begin{aligned}
\mathbb{P}[\|X\|_2 \le t] &= \frac{1}{2\pi} \iint_{\|X\|_2 \le t} \exp(-\frac{1}{2}(x_1^2 + x_2^2))dx_1 dx_2 \\
&= \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{r=0}^{t} \exp(-\frac{1}{2}r^2) r dr d\varphi \\
&= 1 - \exp(-t^2/2).
\end{aligned}$$

This is an easily invertible CDF for the radius (the euclidean norm) of X, with the inverse being $F_R^{-1}(u) = \sqrt{-2\log(1-u)}$. From the rotational symmetry of the PDF it follows, that for a given radius, each polar angle has the same probability, e.g., it is uniformly distributed in $[0, 2\pi)$. Finally, the method can be described by the following steps:

1. Generate two independent $U([0, 1))$-variables V and W
2. Set $R = \sqrt{-2\log(1-V)}$
3. Set $X_1 = R\cos(2\pi W)$ and $X_2 = R\sin(2\pi W)$

1.2.4 Generating Multivariate Normals

Generating a s-dimensional standard normal distributed vector is easy, because one just has to generate s one-dimensional independent and identical distributed (iid) samples. If some dependency between the components of the vector is desired, things are bit different. In this case, the proposition below should prove helpful.

Proposition 1.2.3. *Let $Y = (Y_1, \dots, Y_s)$ be standard normally distributed, $\mu \in \mathbb{R}^s$, $L \in \mathbb{R}^{s\times s}$ be a regular matrix and $\Sigma = LL^T$. Then $X := LY + \mu$ is normally distributed with mean μ and covariance-matrix Σ.*

Proof. First observe, that the PDF of Y is simply $f_Y(y) = \frac{1}{\sqrt{(2\pi)^s}} \exp(-\frac{1}{2}y^T y)$. Let $\phi(y) := Ly + \mu$ and $A \subset \mathbb{R}^s$. The transformation ϕ is invertible, because L is regular and the inverse is $\phi^{-1}(x) = L^{-1}(x - \mu)$. The determinant of its Jacobian equals

$$|\det J_{\phi^{-1}}(x)| = |\det L^{-1}| = \frac{1}{|\det L|} = \frac{1}{\sqrt{|\det \Sigma|}}.$$

Studying the probability of the event $[X \in A]$ leads to

$$\begin{aligned}\mathbb{P}[X \in A] &= \mathbb{P}[\phi(Y) \in A] \\ &= \mathbb{P}[Y \in \phi^{-1}(A)] \\ &= \int_{\phi^{-1}(A)} f_Y(y)dy,\end{aligned}$$

(applying the change of variables theorem here)

$$\begin{aligned}&= \int_A f_Y(\phi^{-1}(x))|\det J_{\phi^{-1}}(x)|dx \\ &= \int_A \frac{1}{\sqrt{(2\pi)^s|\det \Sigma|}} \exp\Big(-\frac{1}{2}(x-\mu)^T \underbrace{(L^{-1})^T L^{-1}}_{=\Sigma^{-1}}(x-\mu)\Big),\end{aligned}$$

which yields, that the PDF of X is exactly of the form

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^s|\det(\Sigma)|}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big),$$

therefore the distribution of X is as stated in the proposition. □

So if one wants to generate an s-dimensional $N(\mu, \Sigma)$-distributed sample x, one first has to find a matrix L such that $LL^T = \Sigma$, then generate a vector $y = (y_1, \ldots, y_s)$ consisting of s independent and standard normally distributed samples and finally setting $x = Ly + \mu$.

If Σ is positive definite, there always exists a matrix L fulfilling $LL^T = \Sigma$. The most common methods for finding such a matrix are the Cholesky factorization and the principal component decomposition, with implementation in almost every programming language used for numerical computations. When applying the Cholesky factorization, the resulting factor L is a lower triangular matrix, which has a numerical advantage because half of its entries are 0. With the principal component decomposition, the original matrix is decomposed into three factors $\Sigma = V\Delta V^T$. Here the matrix Δ is a diagonal matrix with the eigenvalues of Σ as entries, and V is an orthogonal matrix. Via setting $L = V\Delta^{1/2}$, one has again $LL^T = V\Delta^{1/2}\Delta^{1/2}V^T = \Sigma$. This decomposition has some statistical interpretation about the influence of each component via the corresponding eigenvalues, that is occasionally useful [5]. Also, for given L with $LL^T = \Sigma$ and any orthogonal matrix Q the new matrix $\tilde{L} = LQ$ fulfils $\tilde{L}\tilde{L}^T = LQQ^TL^T = \Sigma$.

1.3 Introduction to QMC

In ordinary Monte Carlo, we generate a uniformly distributed pseudo-random sequence $(x_n)_{n\leq N}$ to estimate the true value of the integral via

$$\int_{[0,1]^s} f(x)dx \approx \frac{1}{N}\sum_{n=1}^{N} f(x_n). \tag{1.3.1}$$

If we take a look at Figure 1.3.1 on the left, one can clearly see, that there are on the one hand areas, where many points are clustered, and on the other hand areas, that are rather sparsely populated.

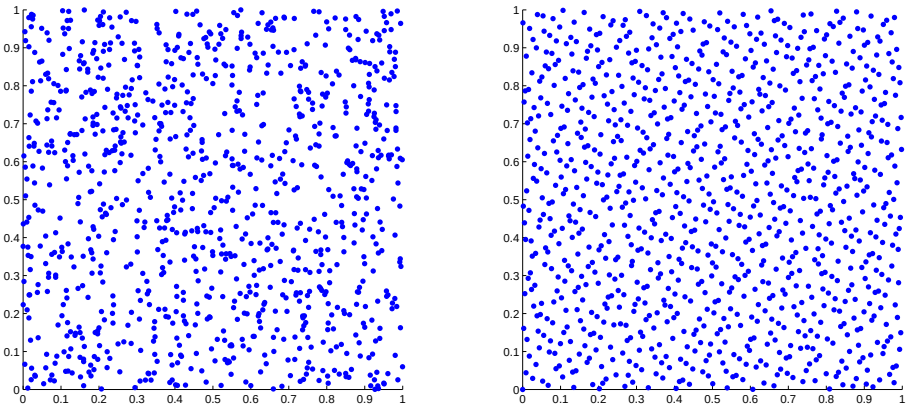

Figure 1.3.1: 1000 Pseudo-random numbers (left) vs. the first 1000 points of the Halton sequence (right).

This phenomenon lies in the nature of random numbers, but is in our case rather unwanted. The question arises, how to construct sequences, that are more equally distributed over the whole area, e.g. like the Halton sequence in Figure 1.3.1 on the right. This is where quasi-Monte Carlo comes into play. QMC deals exactly with the construction of such sequences and the error analysis when using those sequences to approximate the integral (1.3.1). As it turns out, this is in many cases very efficient and beats ordinary MC in terms of convergence. One thing we need to keep in mind is, that those sequences aren't random anymore but purely deterministic, which is why we don't automatically get an error estimator via the sample variance (1.1.4). We have, however, a deterministic error bound at hand, if the integrand meets some smoothness condition, as we will see in Chapter 1.6. This error bound, the Koksma-Hlawka inequality, is a product of two terms, one is a measure for the "uniformity" of the sequence and the other depends only on the variation of the integrand, thus we can study the properties of the sequence and the integrand completely separate. Classical references regarding QMC are Kuipers & Niederreiter [12] and Drmota & Tichy [4], among the more recent literature one should name Dick & Pillichshammer [3]. It remains to define the terms uniformity and variation used here, which will be done in the following paragraph, respectively in Section 1.5. All the definitions and results in this and the following section can be found in Leobacher & Pillichshammer [15] with more details and proofs. We need them here to get an overview and because they are often referred to in this work.

Consider an infinite sequence of points $\mathcal{S} = (x_n)_{n\in\mathbb{N}_0}$ in $[0,1)^s$. Let us denote

$$A([a,b),\mathcal{S},N) = \#\{n \in \mathbb{N}_0 : 0 \leq n \leq N-1 \text{ and } x_n \in [a,b)\}$$

as the number of indices within the first N points of $\mathcal{S}$, that lie in the hyper-rectangle

$[a,b) \subseteq [0,1)^s$. Let λ_s be the s-dimensional Lebesgue measure, s.t. $\lambda_s([a,b)) = \prod_{i=1}^{s}(b_i - a_i)$.

Definition 1.3.1. An infinite sequence $\mathcal{S}$ is said to be **uniformly distributed modulo one** (or equidistributed), if for every interval of the form $[a,b) \subseteq [0,1)^s$ we have

$$\lim_{N\to\infty} \frac{A([a,b),\mathcal{S},N)}{N} = \lambda_s([a,b)). \tag{1.3.2}$$

In other terms, the ratio of the number of points within the interval to the total number of points should behave like the volume of the interval. The following theorem states a particularly useful property of sequences which are uniformly distributed modulo one.

Theorem 1.3.2 ([15])**.** *A sequence $(x_n)_{n\in\mathbb{N}_0}$ in $[0,1)^s$ is uniformly distributed modulo one if and only if for every Riemann integrable function $f : [0,1]^s \to \mathbb{R}$ we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) = \int_{[0,1]^s} f(x)dx. \tag{1.3.3}$$

A consequence of Theorem 1.3.2 is, that sequences which satisfy (1.3.2) are, at least theoretically, viable choices as integration nodes. Still, we need some measure to quantify the "uniformity" of such a sequence, in order to compare different sequences and estimate the convergence rate in (1.3.3). The discrepancy as such a measure is introduced in the following section. But before we proceed, let's look at an example.

Definition 1.3.3. Let $b \in \mathbb{N}, b \geq 2$. The **b-adic radical inverse function** is defined as $\phi_b : \mathbb{N}_0 \to [0,1]$,

$$\phi_b(n) = \sum_{j=0}^{\infty} \frac{n_j}{b^{j+1}}$$

for $n \in \mathbb{N}_0$ with b-adic digit expansion $n = \sum_{j=0}^{\infty} n_j b^j$, where $n_j \in \{0,1,\ldots,b-1\}$ and only a finite number of $n_j \neq 0$.

The **van der Corput sequence in base b** is defined as $(x_n)_{n\in\mathbb{N}_0}$ with $x_n = \phi_b(n)$.

Proposition 1.3.4 ([15])**.** *The van der Corput sequence in base b is uniformly distributed modulo one.*

1.4 Discrepancy

Now we will introduce the discrepancy as a sequence's measure of uniformity, or rather, the deviation from uniformity.

Definition 1.4.1. Let $\mathcal{P}$ be a N-element pointset in $[0,1)^s$ (where duplicates are allowed). The **extreme discrepancy** D_N of this point set is defined as

$$D_N(\mathcal{P}) = \sup_{\substack{a,b\in[0,1]^s \\ a\leq b}} \left| \frac{A([a,b),\mathcal{P},N)}{N} - \lambda_s([a,b)) \right|.$$

For an infinite sequence $\mathcal{S}$ the discrepancy $D_N(\mathcal{S})$ is the discrepancy of the first N elements of $\mathcal{S}$.

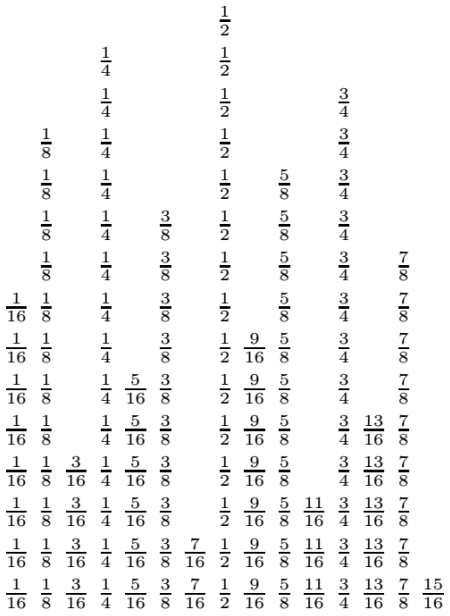

Figure 1.3.2: The n-th line in this picture, taken from [5], represent the first n elements of the van der Corput sequence in base 2

In words, the (extreme) discrepancy measures the worst case deviation of the fraction of points within a hyperrectangle and its volume for all hyperrectangles in $[0,1]^s$. One could also consider other shapes than hyperrectangles, e.g. balls, rotated rectangles or all convex sets in $[0,1]^s$. These sorts of discrepancies are much harder to estimate, Matoušek [17] deals with this topic. It seems somehow intuitive, that a sequence $\mathcal{S}$ is uniformly distributed modulo one if and only if $\lim_{N\to\infty} D_N(\mathcal{S}) = 0$ as it is indeed shown in [15], Theorem 2.15. Sequences with lower overall discrepancy result in a faster convergence in equation (1.3.3). In most cases it's enough to consider a simplified version of the discrepancy, where the supremum is taken only of hyperrectangles with a fixed corner in the origin.

Definition 1.4.2. Let $\mathcal{P}$ be a N-element pointset in $[0,1)^s$. The **star discrepancy** D_N^* of this point set is defined as

$$D_N^*(\mathcal{P}) = \sup_{a\in[0,1]^s} \left| \frac{A([0,a),\mathcal{P},N)}{N} - \lambda_s([0,a)) \right|.$$

For an infinite sequence $\mathcal{S}$ the star discrepancy $D_N^*(\mathcal{S})$ is the star discrepancy of the first N elements of $\mathcal{S}$.

If we consider the dimension s as fixed, the star discrepancy and the extreme discrepancy are bounded by each other times a constant, as the following proposition states:

Proposition 1.4.3 ([15])**.** *For every N-element point set $\mathcal{P}$ in $[0,1)^s$ we have*

$$D_N^*(\mathcal{P}) \le D_N(\mathcal{P}) \le 2^s D_N^*(\mathcal{P}).$$

We have also lower bounds for the discrepancy:

Proposition 1.4.4 ([15]). *For every N-element point set $\mathcal{P}$ in $[0,1)^s$ we have*

$$D_N(\mathcal{P}) \geq \frac{1}{N} \quad \text{and} \quad D_N^*(\mathcal{P}) \geq \frac{1}{2^s N}.$$

When thinking of point sets with low discrepancy, many would first come up with the **regular lattice**, which is defined as the m^s-element set

$$\Gamma_{m,s} := \left\{ \left(\frac{2n_1+1}{2m}, \ldots, \frac{2n_s+1}{2m} \right) : n_1, \ldots, n_s \in \{0, \ldots, m-1\} \right\}. \tag{1.4.1}$$

As it turns out, the regular lattice does not perform very good in terms of discrepancy, except in dimension 1, where it has the lowest possible discrepancy $\frac{1}{N}$. In higher dimensions, we shall see sequences with much lower discrepancy.

Proposition 1.4.5 ([15]). *Let $s, m \in \mathbb{N}, m \geq 2$. For the star discrepancy of the regular lattice $\Gamma_{m,s}$ with $N = m^s$ elements in $[0,1)^s$ it holds that*

$$D_N^*(\Gamma_{m,s}) = 1 - \left(1 - \frac{1}{2m}\right)^s.$$

Another reason, why the regular lattice is bad, can be seen in Figure 1.4.1: Consider the integration of an additive function $f : [0,1]^2 \to \mathbb{R}$ where $f(x,y) = f_1(x) + f_2(y)$. When projecting the 16 (blue) points of the regular lattice on the left to each of the coordinate axes (red and green points), many of them coincide with each other and the effective number of points used to estimate the integral is only 4. This is a waste of information and computational resources, because if you position those 16 points in a smarter way, like on the right, the projections still result in 16 distinct points on both of the coordinate axes.

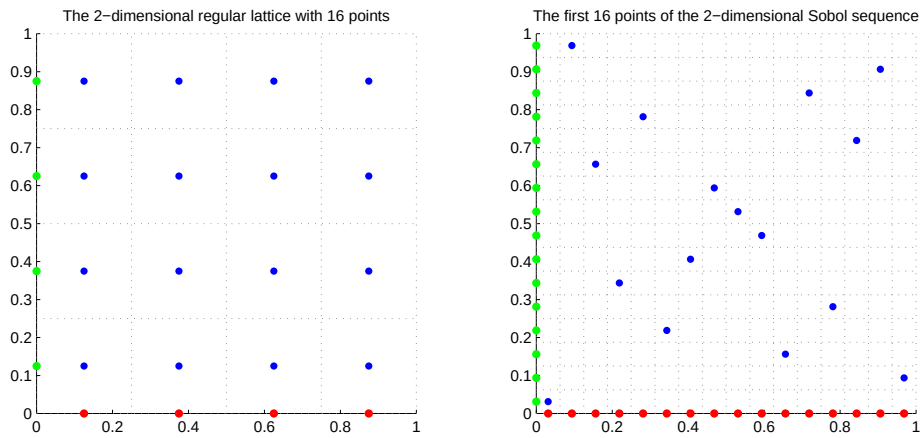

Figure 1.4.1: The regular lattice (left) vs. the Sobol sequence, each with 16 points.

The following result is interesting for analysing the asymptotic behaviour of the discrepancy, or to get an lower bound for the number of points required to achieve some given accuracy:

Proposition 1.4.6 (**Roth** [15]). *For every dimension $s \in \mathbb{N}$ there exists a constant $c_s > 0$ with the following property: for every N-element point set $\mathcal{P}$ in $[0,1)^s$,*

$$D_N(\mathcal{P}) \geq D_N^*(\mathcal{P}) \geq c_s \frac{(\log N)^{\frac{s-1}{2}}}{N}.$$

As a first example of a uniformly distributed modulo one sequence in more dimensions we will introduce the Halton sequence. This sequence is very popular because of its easy implementation.

Definition 1.4.7. Let $s \in \mathbb{N}$ and let $b_1, \ldots, b_s \geq 2$ be integers. The **Halton sequence** in bases $b_1, \ldots, b_s$ is the sequence $\mathcal{S}_{b_1,\ldots,b_s} = (x_n)_{n \in \mathbb{N}_0}$ whose n-th element is given by

$$x_n := (\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_s}(n)),$$

where ϕ_{b_i} is the radical inverse function from Definition 1.3.3.

Proposition 1.4.8 ([15]). *Let $s \in \mathbb{N}$ and let $b_1, \ldots, b_s \geq 2$ be pairwise coprime integers. For the star discrepancy of the Halton sequence $\mathcal{S}_{b_1,\ldots,b_s}$ we have*

$$D_N^*(\mathcal{S}_{b_1,\ldots,b_s}) \leq \frac{1}{N}\left(\prod_{j=1}^{s} \frac{b_j \log(b_j N)}{\log b_j}\right).$$

Hence asymptotically for $N \to \infty$ we have $D_N^*(\mathcal{S}_{b_1,\ldots,b_s}) = O(\frac{(\log N)^s}{N})$.

Unfortunately the discrepancy does not perform very well for growing dimension. Also, if some of the bases used are really close together relative to their size, e.g. 41 and 43, the resulting points are poorly distributed and can have strong dependence, as one can see in Figure 1.4.2. Because of that, we will introduce another type of point-sets and sequences,

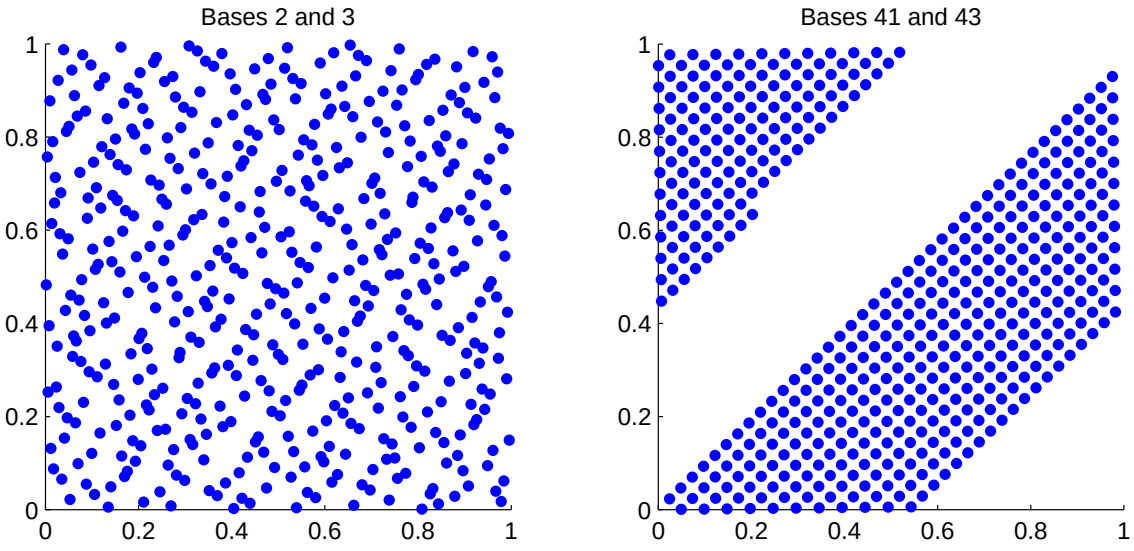

Figure 1.4.2: The first 500 points of a 2-dimensional Halton sequence with different bases.

which have a very low star discrepancy, namely the (t, m, s)-nets and (t, s)-sequences. These point-sets are constructed in a way, that their discrepancy is 0 for a large class of intervals, hoping that this construction yields a low overall star discrepancy. First we need to define the class of elementary intervals.

Definition 1.4.9. Let $b \in \mathbb{N}, b \geq 2$. An **elementary interval** in base b is an interval of the form

$$J = \prod_{j=1}^{s} \left[\frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right),$$

where $d_j \in \mathbb{N}_0$ and $a_j \in \{0, 1, \ldots, b^{d_j} - 1\}$ for all $j = 1, 2, \ldots, s$.

Definition 1.4.10. Let $m, s, b \in \mathbb{N}, b \geq 2$, and let $t \in \{0, \ldots, m\}$. A **(t, m, s)-net in base b** is a b^m-element point set $\mathcal{P}$ in $[0, 1)^s$ which satifies

$$\frac{A(J, \mathcal{P}, N)}{N} = \lambda_s(J)$$

for every s-dimensional elementary interval J in base b having volume b^{-m+t}. The parameter t is called the **quality parameter** of the net. Furthermore, $\mathcal{P}$ is called a **strict** (t, m, s)-net in base b, if t is the smallest number $u \in \{0, \ldots, m\}$ such that $\mathcal{P}$ is a (u, m, s)-net in base b.

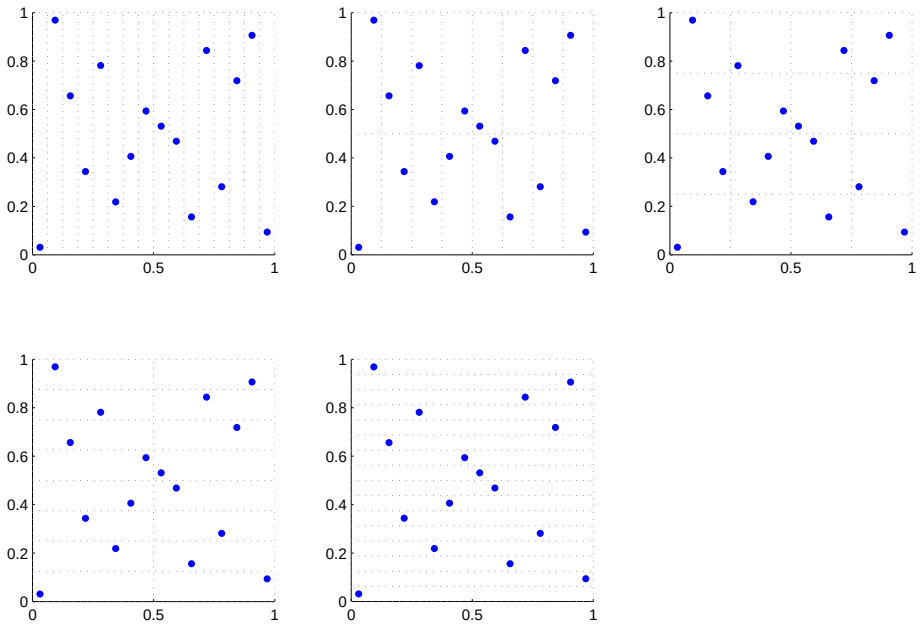

Figure 1.4.3: A $(0, 4, 2)$-net in base 2: every elementary interval with volume $2^{-4} = \frac{1}{16}$ contains exactly one point.

Definition 1.4.11. Let $s, b \in \mathbb{N}, b \geq 2$, and let $t \in \mathbb{N}_0$. An infinite sequence $(x_n)_{n \in \mathbb{N}_0}$ of points in $[0, 1)^s$ is called a **(t, s)-sequence in base b**, if for all integers $m > t$ and $k \geq 0$, the point set $\{x_{kb^m}, x_{kb^m+1}, \ldots, x_{kb^m+b^m-1}\}$ forms a (t, m, s)-net in base b. The parameter t is again called the quality parameter, and a (t, s)-sequence is called a strict (t, s)-sequence, if it is not a $(t - 1, s)$-sequence.

Best known examples of (t, s)-sequences are the ones first constructed by I.M. Sobol in 1967 and H. Faure 1982. Faure sequences are $(0, d)$-sequences and thus optimize the quality parameter t. However, they require a base at least as large as the smallest prime $\geq s$. Sobol sequences use base 2 regardless of the dimension (which has computational as well as uniformity advantages) but their t parameter grows with the dimension s. Today, it seems most convenient to construct (t, s)-sequences using H. Niederreiter's approach via digital nets, where both sequences mentioned before appear as special cases ([15] and [5]). We don't want to go into detail to describe the implementation and instead refer to the literature, e.g. [15] for the basics of digital nets, or Bratley and Fox 1988 [1] respectively Joe and Kuo 2003 [9] for the an algorithm generating Sobol points up to 1111 dimensions. An implementation of the latter algorithm is available in MATLAB, which we will use in the practical part of this work.

Proposition 1.4.12 ([15]). *For the star discrepancy of a (t, m, s)-net $\mathcal{P}$ in base b we have*

$$D^*_{b^m}(\mathcal{P}) \leq \frac{1}{b^{m-t}} \sum_{k=1}^{s-1} \binom{m-t}{k} (b-1)^k.$$

Proposition 1.4.13 ([15]). *For the star discrepancy of a (t, s)-sequence $\mathcal{S}$ in base b we have*

$$D^*_N(\mathcal{S}) \leq \frac{b^t(b-1)}{N} \sum_{m=0}^{r} \sum_{k=0}^{s-1} \binom{m-t}{k} (b-1)^k,$$

where $r = \lfloor \frac{\log N}{\log b} \rfloor$.

1.5 Variation in the Sense of Hardy and Krause

Essential for the error estimation of the QMC integral is the right measure for the variation of the integrand. It is not obvious how to generalize the total variation of an one-dimensional function to more dimensions. In this chapter we will introduce the variation in the sense of Hardy and Krause, which will appear in the formula for the error estimation in the next section. Most definitions and results in this section are based on Owen [20]. We are now going to set up some new notation and definitions, which we use only in contexts of the Hardy-Krause-variation.

Definition 1.5.1. With $\mathcal{I}_s$ we denote the index set $\{1, 2, \ldots, s\}$. For $u, v \subseteq \mathcal{I}_s$ write $|u|$ for the cardinality of u, and $u - v$ for the set difference $u \setminus v$. For integers $j \leq k$, the set $\{j, j+1, \ldots, k\}$ is written as $j\!:\!k$. A unary minus denotes the complement with respect to $\mathcal{I}_s$, s.t. $-u = \mathcal{I}_s - u$. For $u \in \mathcal{I}_s, x \in \mathbb{R}^s$, the expression x_u denotes the $|u|$-tuple consisting of the components x_j for $j \in u$. Suppose that $u, v \subseteq \mathcal{I}_s$ and $x, z \in \mathbb{R}^s$ with $u \cup v = \mathcal{I}_s$ and $u \cap v = \emptyset$. The symbol $x_u : z_v$ represents the point y with $y_j = x_j$ if $j \in u$ and $y_j = z_j$ for $j \in v$. If t is a scalar value, $x_u : t_v$ is the point y with $y_j = x_j$ if $j \in u$ and $y_j = t$ for $j \in v$. The 3-fold ":"-notation is defined in an analogue way: For a partition $u \uplus v \uplus w = \mathcal{I}_s$, the expression $x_u : y_v : z_w$ denotes the point ξ with $\xi_j = x_j$ if $j \in u$, $\xi_j = y_j$ if $j \in v$ and $\xi_j = z_j$ if $j \in w$.

Let $f : \mathbb{R}^s \to \mathbb{R}$ and $u \subset \mathcal{I}_s$. With $f(x_u; y_{-u})$ we denote the function $g : \mathbb{R}^{|u|} \to \mathbb{R}$, which is defined as $g(z_u) = f(z_u : y_{-u})$ for $z_u \in \mathbb{R}^{|u|}$.

Let $a, b \in \mathbb{R}$, $a \leq b$ and $n \in \mathbb{N}$. A set $\mathcal{Y} = \{y^{(0)}, y^{(1)}, \ldots, y^{(n)}\} \subset \mathbb{R}$ with $a = y^{(0)} < y^{(1)} < \ldots < y^{(n)} < b$ is called **ladder on $[a, b]$**. For each $y \in \mathcal{Y}$, define its **successor** y^+ as

$$y^+ := \begin{cases} y^{(k+1)} & \text{if } y = y^{(k)} \text{ with } k < n, \\ b & \text{if } y = y^{(n)}. \end{cases}$$

For $a, b \in \mathbb{R}^s$ with $a \leq b$ let $\mathcal{Y}_j$ be a ladder on $[a_j, b_j]$ for all $j = 1, \ldots, s$. The Cartesian product $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_s$ is called a **multidimensional ladder on $[a, b]$**. The successor y^+ of $y = (y_1, \ldots, y_s) \in \mathcal{Y}$ is defined by its component-wise successors, s.t. $y^+ = (y_1^+, \ldots, y_s^+)$.

Remark 1.5.2. Whenever it is clear from the context, we may leave the $[a, b]$ out of the notation, like in the short versions of the following definitions.

Definition 1.5.3. Let $a, b \in \mathbb{R}^s$ with $a \leq b$, $f : [a, b] \to \mathbb{R}$. The ***s*-fold alternating sum of f over $[a, b]$** is

$$\Delta f = \Delta(f; a, b) = \sum_{v \subseteq \mathcal{I}_s} (-1)^{|v|} f(a_v : b_{-v}).$$

For $u \subseteq \mathcal{I}_s$ and a given hyperrectangle $[a, b] \in \mathbb{R}^s$ define the operator Δ_u via

$$(\Delta_u f)(x_{-u}) = (\Delta_u(f; a_u, b_u))(x_{-u}) = \sum_{v \subseteq u} (-1)^{|v|} f(a_v : b_{(u-v)} : x_{-u}).$$

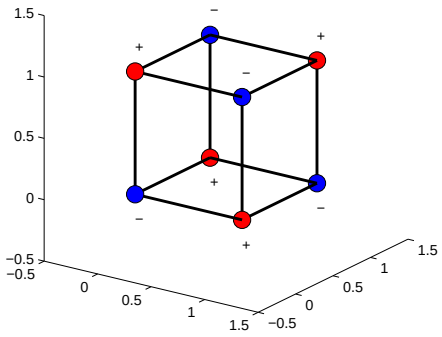

Figure 1.5.1: When calculating $\Delta(f; 0, 1)$ for an $f : [0, 1]^3 \to \mathbb{R}$, the function has to be evaluated at the corners of $[0, 1]^3$ with the corresponding signs.

Remark 1.5.4. When speaking of $(\Delta_u f)(x_{-u})$, we say the components in u have been differentiated, while we still can plug in any value for the components in $-u$, hence $(\Delta_u f)$ is a function in x_{-u}. For $x_{-u} = b_{-u}$, the definition of Δ_u here coincides with the definition of Δ_u in [20], Section 4. Observe, that $\Delta_{\mathcal{I}_s} = \Delta$, $\Delta_\emptyset f = f$, and for $u \uplus v = w$ we have also $\Delta_w = \Delta_v \Delta_u$.

Proof. The point $\Delta_w = \Delta_v \Delta_u$ is not that obvious, therefore we are going to prove it. Let $u, v, w \subseteq \mathcal{I}_s$, with $u \cup v = w$ and $u \cap v = \emptyset$. Set

$$g(x_{-u}) := (\Delta_u f)(x_{-u}) = \sum_{p \subseteq u} (-1)^{|p|} f(a_p : b_{u-p} : x_{-u}).$$

We have

$$\begin{aligned}(\Delta_v g)(y_{-u-v}) &= \sum_{q\subseteq v}(-1)^{|q|} g(\underbrace{a_q : b_{v-q} : y_{-u-v}}_{=:x_{-u}(q,v)}) \\ &= \sum_{q\subseteq v}(-1)^{|q|}\sum_{p\subseteq u}(-1)^{|p|} f(a_p : b_{u-p} : x_{-u}(q,v)) \\ &= \sum_{\substack{q\subseteq v\\ p\subseteq u}}(-1)^{|q|+|p|} f(a_{p+q} : b_{u+v-q-p} : y_{-u-v}) \\ &\big[w = v+u, \quad r = p+q\big] \\ &= \sum_{r\subseteq w}(-1)^{|r|} f(a_r : b_{w-r} : y_{-w}) = (\Delta_w f)(y_{-w}).\end{aligned}$$

□

Example 1.5.5. Consider a function $f : [0,1]^2 \to \mathbb{R}$ in two variables, which we write as $f(x,y)$. We also write $\Delta_x = \Delta_{\{1\}}$, $\Delta_y = \Delta_{\{2\}}$ and $\Delta_{xy} = \Delta_{\{1,2\}}$. We want to calculate $\Delta_{xy}(f; [x_1,x_2]\times[y_1,y_2])$, but first observe, that

$$(\Delta_y f)(x) = f(x,y_2) - f(x,y_1),$$

which is a function still depending on the variable x. Applying Δ_x, we obtain

$$\begin{aligned}\Delta_x((\Delta_y f)(x)) &= \Delta_x(f(x,y_2)) - \Delta_x(f(x,y_1)) \\ &= \big[f(x_2,y_2) - f(x_1,y_2)\big] - \big[f(x_2,y_1) - f(x_1,y_1)\big] \\ &= \Delta_{xy} f.\end{aligned}$$

Lemma 1.5.6 (Product rule for the Δ_u-operator)**.** *Let $a,b\in\mathbb{R}^s$ with $a\le b$, $f,g:[a,b]\to\mathbb{R}$ and $u\subseteq\mathcal{I}_s$. We have*

$$(\Delta_u(fg))(x_{-u}) = \sum_{v\subseteq u}(\Delta_v f)(b_{u-v} : x_{-u})\cdot(\Delta_{u-v} g)(a_v : x_{-u}). \tag{1.5.1}$$

Proof. We will use an induction, with the assumption that (1.5.1) holds for all u with $|u|\le m$.

Induction basis $m=1$: Let $j\in\mathcal{I}_s$, set $u=\{j\}$.

$$\begin{aligned}(\Delta_u(fg))(x_{-u}) &= f(b_u : x_{-u})g(b_u : x_{-u}) - f(a_u : x_{-u})g(a_u : x_{-u}) \\ &= f(b_u : x_{-u})g(b_u : x_{-u}) - f(b_u : x_{-u})g(a_u : x_{-u}) \\ &\quad + f(b_u : x_{-u})g(a_u : x_{-u}) - f(a_u : x_{-u})g(a_u : x_{-u}) \\ &= \underbrace{f(b_u : x_{-u})(\Delta_u g)(x_{-u})}_{\text{summand for } v=\emptyset} + \underbrace{(\Delta_u f)(x_{-u})g(a_u : x_{-u})}_{\text{summand for } v=u}.\end{aligned}$$

Induction step $m\to m+1$:
Let $u\subset\mathcal{I}_s$ with $|u| = m < s$, $d\in\mathcal{I}_s - u$ and $\tilde{u} = u\uplus\{d\}$. Using the properties from

Remark 1.5.4, we have

$$\begin{aligned}
(\Delta_{\tilde{u}}(fg))(x_{-\tilde{u}}) &= (\Delta_{\{d\}}(\Delta_u(fg)))(x_{-\tilde{u}}) \\
&= (\Delta_u(fg))(b_{\{d\}} : x_{-\tilde{u}}) - (\Delta_u(fg))(a_{\{d\}} : x_{-\tilde{u}}) \\
&= \sum_{v \subseteq u} (\Delta_v f)(b_{u-v+\{d\}} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_v : b_{\{d\}} : x_{-\tilde{u}}) \\
&\quad - \sum_{v \subseteq u} (\Delta_v f)(a_{\{d\}} : b_{u-v} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_{v+\{d\}} : x_{-\tilde{u}}) \\
&= \sum_{v \subseteq u} \Big[(\Delta_v f)(b_{u-v+\{d\}} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_v : b_{\{d\}} : x_{-\tilde{u}}) \\
&\quad - (\Delta_v f)(b_{u-v+\{d\}} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_{v+\{d\}} : x_{-\tilde{u}}) \Big] \\
&\quad + \sum_{v \subseteq u} \Big[(\Delta_v f)(b_{u-v+\{d\}} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_{v+\{d\}} : x_{-\tilde{u}}) \\
&\quad - (\Delta_v f)(a_{\{d\}} : b_{u-v} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_{v+\{d\}} : x_{-\tilde{u}}) \Big] \\
&= \sum_{v \subseteq u} (\Delta_v f)(b_{u-v+\{d\}} : x_{-\tilde{u}}) \cdot (\Delta_{u-v+\{d\}} g)(a_v : x_{-\tilde{u}}) \\
&\quad + \sum_{v \subseteq u} (\Delta_{v+\{d\}} f)(b_{u-v} : x_{-\tilde{u}}) \cdot (\Delta_{u-v} g)(a_{v+\{d\}} : x_{-\tilde{u}}).
\end{aligned}$$

In the first sum, set $\tilde{v} = v$, hence $u - v + \{d\} = \tilde{u} - \tilde{v}$, and restrict the summation elements to $\tilde{v} \subseteq \tilde{u}$ with $d \notin \tilde{v}$. In the second sum, set $\tilde{v} = v + \{d\}$, hence $u - v = \tilde{u} - \tilde{v}$ and restrict the summation elements to $\tilde{v} \subseteq \tilde{u}$ with $d \in \tilde{v}$. Now the summands are exactly the same, and both sums can be combined to one sum, which yields

$$(\Delta_{\tilde{u}}(fg))(x_{-\tilde{u}}) = \sum_{\tilde{v} \subseteq \tilde{u}} (\Delta_{\tilde{v}} f)(b_{\tilde{u}-\tilde{v}} : x_{-\tilde{u}}) \cdot (\Delta_{\tilde{u}-\tilde{v}} g)(a_{\tilde{v}} : x_{-\tilde{u}}).$$

The cardinality of $\tilde{u}$ is $m + 1$, and since the element d was arbitrary from $\mathcal{I}_s - u$ the induction step is complete.

□

Definition 1.5.7. Let $a, b \in \mathbb{R}^s$ with $a \leq b$ and $\mathcal{Y}$ be a (multidimensional) ladder on $[a, b]$. The **variation of f over $\mathcal{Y}$** is defined as

$$V_{\mathcal{Y}} = \sum_{y \in \mathcal{Y}} |\Delta(f; y, y^+)|.$$

Definition 1.5.8. Let $a, b \in \mathbb{R}^s$ with $a \leq b$ and $\mathbb{Y}$ be the set of all ladders on $[a, b]$. The **variation** of f on the hyperrectangle $[a, b]$, in the sense of **Vitali**, is

$$V(f) = V_{[a,b]}(f) = \sup_{\mathcal{Y} \in \mathbb{Y}} V_{\mathcal{Y}}.$$

If $V(f) < \infty$, we also write $f \in BV$.

Definition 1.5.9. The **variation** of f on the hyperrectangle $[a, b]$, in the sense of **Hardy and Krause**, is

$$V_{HK}(f) = V_{HK}(f; a, b) = \sum_{\emptyset \neq u \subseteq \mathcal{I}_s} V_{[a_u, b_u]}(f(x_u; b_{-u})).$$

If $V_{HK}(f) < \infty$, we also write $f \in BVHK$.

Proposition 1.5.10 ([20])**.** *Let $\mathcal{Y}$ and $\tilde{\mathcal{Y}}$ be ladders on the hyperrectangle $[a, b]$ with $\mathcal{Y} \subseteq \tilde{\mathcal{Y}}$. Then $V_{\mathcal{Y}}(f; a, b) \leq V_{\tilde{\mathcal{Y}}}(f; a, b)$.*

Remark 1.5.11. This result permits us, to replace the supremum over all ladders in Definition 1.5.8 by the supremum over a subset $\tilde{\mathbb{Y}} \subseteq \mathbb{Y}$ of ladders, because if for every $\mathcal{Y} \in \mathbb{Y}$ there is a $\tilde{\mathcal{Y}} \in \tilde{\mathbb{Y}}$ with $\mathcal{Y} \subseteq \tilde{\mathcal{Y}}$, both suprema are equivalent. If $[a, b] = [0, 1]^s$, one can for example restrict oneself to ladders, which are the s-fold Cartesian product of one one-dimensional ladder. As a proof, consider some general ladder $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_s$. Construct a new one-dimensional ladder from the the union, s.t. $\bar{\mathcal{Y}}_1 = \bigcup_{j=1}^{s} \mathcal{Y}_j$ and choose the new refined ladder as $\tilde{\mathcal{Y}} = \bar{\mathcal{Y}}_1 \times \ldots \times \bar{\mathcal{Y}}_1$.

Proposition 1.5.12 ([20])**.** *Suppose that $f(x)$ is defined on the hyperrectangle $[a, b]$ and $f(x)$ does not depend on x_u for a non-empty $u \in \mathcal{I}_s$. Then $V(f) = 0$.*

Definition 1.5.13. For $a, b, \in \mathbb{R}^s$, $j \in \mathcal{I}_s$ and $c \in [a_j, b_j]$, the **coordinate split of $[\boldsymbol{a}, \boldsymbol{b}]$** is the set $\{L, R\}$ of hyperrectangles defined as.

$$L = \{x \in [a, b] | x_j \leq c\} \text{ and}$$
$$R = \{x \in [a, b] | x_j \geq c\}.$$

A **split of the hyperrectangle $[\boldsymbol{a}, \boldsymbol{b}]$** is a set $\{[a_i, b_i] | 1 \leq i \leq m < \infty\}$, where $\bigcup_{i=1}^{m} [a_i, b_i] = [a, b]$ and $[a_i, b_i) \cap [a_j, b_j) = \emptyset$, if $i \neq j$.

Proposition 1.5.14 ([20])**.** *Let $\{[a_i, b_i] | 1 \leq i \leq m < \infty\}$ be a split of $[a, b]$ and $f : [a, b] \to \mathbb{R}$. We have*

$$V_{[a,b]}(f) = \sum_{i=1}^{m} V_{[a_i, b_i]}(f).$$

Proposition 1.5.15 ([20])**.** *Let $f(x)$ be defined on the hyperrectangle $[a, b]$. Let $\tilde{f}(x)$ be defined on the hyperrectangle $[\tilde{a}, \tilde{b}]$ by $\tilde{f}(x) = f(\tilde{x})$ where $\tilde{x}_j = \phi_j(x_j)$ with ϕ_j is a strictly monotone increasing invertible function from $[\tilde{a}_j, \tilde{b}_j]$ onto $[a_j, b_j]$, s.t. $\phi_j(\tilde{a}_j) = a_j$ and $\phi_j(\tilde{b}_j) = b_j$. We have*

$$V_{[\tilde{a}, \tilde{b}]}(\tilde{f}) = V_{[a,b]}(f)$$

and also

$$V_{HK}(\tilde{f}; \tilde{a}, \tilde{b}) = V_{HK}(f; a, b).$$

Proposition 1.5.16 ([20])**.** *Let f and g be functions on the hyperrectangle $[a, b]$. If $f, g \in BVHK$, then $f + g$, $f - g$, and fg are in $BVHK$. If $f \in BVHK$ with $|f| > C > 0$ then $1/f \in BVHK$. If $f, g \in BV$, then $f + g$ and $f - g$ are in BV, but fg is in general not in BV. For $\alpha, \beta \in \mathbb{R}$ we have $V_{[a,b]}(\alpha + \beta f) = |\beta| V_{[a,b]}(f)$ and $V_{HK}(\alpha + \beta f) = |b| V_{HK}(f)$.*

Because the definition of the variation in the sense of Hardy and Krause is a bit unwieldy, an alternative method of calculation would be desirable. The following results provide such an alternative. If the derivative $\frac{\partial^s}{\partial x_1 \cdots \partial x_s} f(x)$ exists, it is easy to check that

$$\int_{[a,b]} \frac{\partial^s}{\partial x_1 \cdots \partial x_s} f(x) dx = \Delta(f; a, b). \tag{1.5.2}$$

To simplify the notation, for $u \subseteq \mathcal{I}_s$ we define $\partial^u = \prod_{j \in u} \frac{\partial}{\partial x_j}$.

Proposition 1.5.17 ([20]). *If $\partial^{\mathcal{I}_s} f$ is continuous on $[a, b]$ then*

$$V(f) = \int_{[a,b]} |\partial^{\mathcal{I}_s} f(x)| dx.$$

Especially, if $\partial^{\mathcal{I}_s} f$ is continuous, we have

$$V_{[a,b]}(f) \leq Vol([a, b]) \max_{x \in [a,b]} |\partial^{\mathcal{I}_s} f(x)|. \tag{1.5.3}$$

Remark 1.5.18. Using Proposition 1.5.17, we can calculate the Hardy and Krause variation of a function via

$$V_{HK}(f) = \sum_{\emptyset \neq u \subseteq \mathcal{I}_s} \int_{[a_u, b_u]} \left| \partial^u f(x_u : b_{-u}) \right| dx_u, \tag{1.5.4}$$

which can in many cases be easier than the direct calculation through the definition. A disadvantage is of course, that this formula only holds, if the function is s-times continuously differentiable.

Example 1.5.19. For

$$f(x_1, x_2) = I_{x_1 + x_2 \leq 1}(x_1, x_2)$$

we have $V_{[0,1]^2}(f) = \infty$.

Proof. Choose $n \in \mathbb{N}$ and let $\mathcal{Y}_n^1 = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}\}$ be an equidistant ladder in one dimension. We define a two-dimensional ladder via $\mathcal{Y}_n = \mathcal{Y}_n^1 \times \mathcal{Y}_n^1$. Let

$$S_n = \left\{ \left(\frac{i}{n}, \frac{n-i-1}{n} \right) \middle| i \in \{0, 1, \ldots, n-1\} \right\}$$

and observe, that $S_n \subset \mathcal{Y}_n$, as well as $|S_n| = 1$. For $y = (\frac{i}{n}, \frac{n-i-1}{n}) \in S_n$, we have $|\Delta(f; y, y^+)| = 1$, because

$$\begin{aligned} |\Delta(f; y, y^+)| &= \Big| f\Big(\frac{i+1}{n}, \frac{n-i}{n} \Big) + f\Big(\frac{i}{n}, \frac{n-i-1}{n} \Big) \\ &\quad - f\Big(\frac{i+1}{n}, \frac{n-i-1}{n} \Big) - f\Big(\frac{i}{n}, \frac{n-i}{n} \Big) \Big| \\ &= |0 + 1 - 1 - 1| = 1. \end{aligned} \tag{1.5.5}$$

Finally we have

$$V_{[0,1]^2}(f) \geq V_{\mathcal{Y}_n}(f) = \sum_{y \in \mathcal{Y}_n} |\Delta(f; y, y^+)| \geq \sum_{y \in \mathcal{S}_n} \underbrace{|\Delta(f; y, y^+)|}_{=1} = n \to \infty,$$

because this holds for all n. □

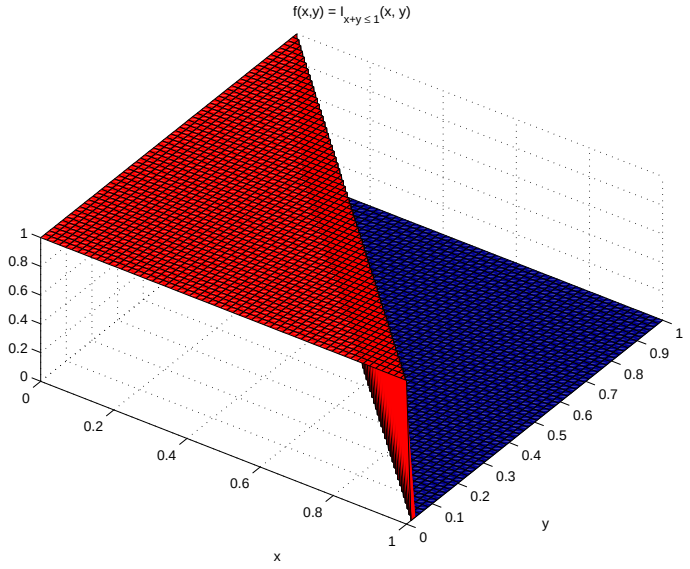

Figure 1.5.2: A plot of Example 1.5.19

Example 1.5.20. For

$$f(x) = \min\Big(\sum_{i=1}^{s} x_i, \frac{s}{2}\Big) \tag{1.5.6}$$

we have $V_{[0,1]^s}(f) < \infty$ if $s \leq 2$ and $V_{[0,1]^s}(f) = \infty$ if $s \geq 3$.

Proof. We perform the proof only for dimension $s = 3$: Again we define a ladder $\mathcal{Y}_n = \mathcal{Y}_n^1 \times \mathcal{Y}_n^1 \times \mathcal{Y}_n^1$, with $\mathcal{Y}_n^1$ being the same as in the previous example and the additional assumption, that n is an even number. Let

$$S_n = \Big\{\Big(\frac{i}{n}, \frac{j}{n}, \frac{3}{2} - \frac{i+j+1}{n}\Big)\Big| i \in \{\frac{n}{2}, \dots, n-1\}, j \in \{0, 1, \dots, \frac{n}{2} - 1\}\Big\},$$

the cardinality of which is $|S_n| = \frac{n^2}{4}$, and again, we have $S_n \subset \mathcal{Y}_n$. For $y \in S_n$ one can easily show in a similar way to the previous example, that $|\Delta(f; y, y^+)| = \frac{1}{n}$, which gives us

$$V_{[0,1]^3}(f) \geq \sum_{y \in \mathcal{Y}_n} |\Delta(f; y, y^+)| \geq \sum_{y \in \mathcal{S}_n} \underbrace{|\Delta(f; y, y^+)|}_{=1/n} = \frac{n^2}{4}\frac{1}{n} = \frac{n}{4} \to \infty.$$

□

Remark 1.5.21. As you see in the previous examples, if the integrand is continuous, but the subdomain, where the integrand is not differentiable is not part of a hyperplane $H_{j,c} := \{x \in \mathbb{R}^s | x_j = c\}$ (i.e., an $(s-1)$-dimensional subspace of $\mathbb{R}^s$, where one component x_j is fixed at the constant value c) for some index j and constant c, it can be, that the variation is infinite. If the non-differentiable domain would be part of some hyperplane $H_{j,c}$, one could split the integration domain in the sense of Definition 1.5.13 along $x_j = c$, and thus by using Proposition 1.5.14, the variation can be calculated separately on each (hyper)rectangle of the split. Within the hyperrectangles, the integrand is differentiable, hence the variation is finite. With a similar reasoning, one can see, that a indicator

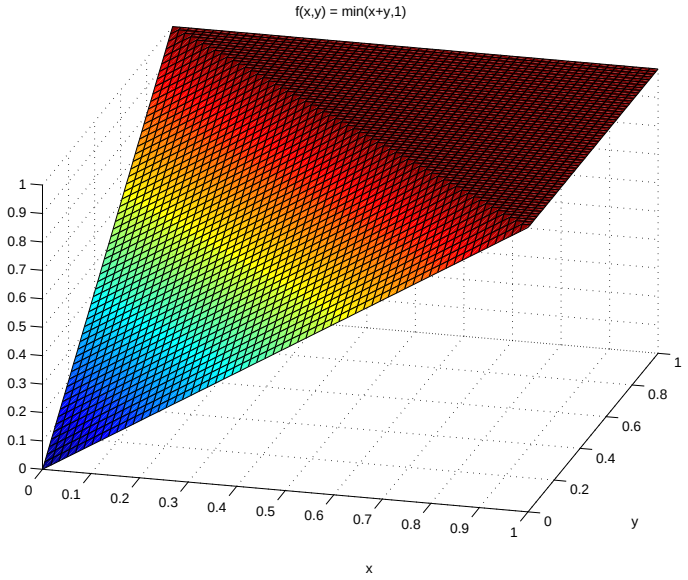

Figure 1.5.3: A plot of Example 1.5.20 for $s = 2$

function on a rectangle (although discontinuous) has finite variation, if and only if the rectangle is axis-parallel.

When first dealing with multidimensional variation, one might be surprised, how so many well-behaving functions have infinite variation. In the case of Example 1.5.19, one can blame the discontinuity, but then, even Example 1.5.20 has infinite variation, although it is even Lipschitz-continuous. The key issue in the previous example is, that the number of cubes in the ladder, which have a non-zero delta, grows quadratically, while the Δf itself decays only linearly. Clearly, if the delta of every single cube would decay fast enough, the variation would be finite, which is of course the case for s-times differentiable functions. But, just because the function is not differentiable, this doesn't mean, that the variation has to be infinite. In the case of Example 1.5.20, it depends on the number of dimensions. The following example of a non-differentiable function has finite variation, regardless of the number of dimensions:

Example 1.5.22. For

$$f(x) = \max_{i=1,\dots,s} x_i \tag{1.5.7}$$

we have $V_{[0,1]^s}(f) = 1$, which we are going to prove in Chapter 4.

1.6 Koksma-Hlawka Inequality

Definition 1.6.1. Let $\mathcal{P} = (x_0, \cdots, x_{N-1})$ be an N-element point set in $[0,1)^s$ and $f : [0,1]^s \to \mathbb{R}$. With $e(f, \mathcal{P})$ we denote the integration error

$$e(f, \mathcal{P}) = \int_{[0,1]^s} f(x)dx - \frac{1}{N} \sum_{n=0}^{N-1} f(x_n).$$

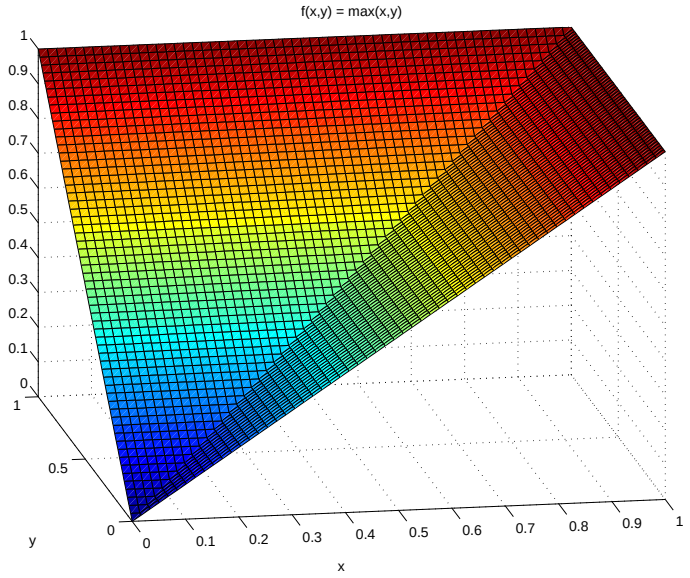

Figure 1.5.4: A plot of Example 1.5.22 for $s = 2$

Theorem 1.6.2 (**Koksma-Hlawka inequality**). *Let $\mathcal{P}$ be an N-element point set in $[0,1)^s$ and $f : [0,1]^s \to \mathbb{R}$. We have*

$$|e(f,\mathcal{P})| \leq V_{HK}(f) D_N^*(\mathcal{P}).$$

Proof. See e.g. Kuipers and Niederreiter (1974) [12].

1.7 Randomized QMC

Although the Koksma-Hlawka inequality by itself is sharp, the discrepancy and variation are both very hard to calculate, and the discrepancy estimates we presented in Section 1.4 are very rough for growing dimension s, therefore they are often not feasible in practice. Estimation of the variation of the integrand requires a deep understanding of its structure and can often only be done with symbolic analysis, as in contrary to ordinary MC methods, where one can basically plug in any function and receive a numerical error estimate via the sample variance for free.

For these reasons, so called *randomized QMC* (RQMC) methods were introduced, which combine the best of both worlds, namely:

- Fast convergence, like $O(N^{-1+\delta})$ for many functions, even if they are not smooth.
- The estimator is unbiased, as in contrary to QMC, where the estimator is only asymptotically unbiased.
- Cheap error estimation, because a variance can be calculated from the randomized results.

While the Koksma-Hlawka inequality and its variants provide the theoretical foundations for the convergence rate of QMC methods, practical performance is then measured

via the variance of the RQMC estimate. At the time of the writing of this work, researchers are still working on theoretical explanations for the observed convergence rates with unbounded and/or non-smooth integrands, which often outperform the rates predicted by the existing theory.

1.7.1 Additive Shift

The strategy is to first construct a deterministic low-discrepancy point set, and then apply a randomization, which on the one hand retains the good distributional properties of the point set and on the other hand has enough randomness to yield an unbiased estimator [3].

Let's start with the most simple randomization technique, which we shall refer to as *additive shift*, but first we need to define the fractional part in our context.

Definition 1.7.1. For any number $x \in \mathbb{R}$ define its **fractional part** as

$$\text{frac}(x) = x - \lfloor x \rfloor.$$

For a vector $x = (x_1, \dots, x_s) \in \mathbb{R}^s$ define

$$\text{frac}(x) = (\text{frac}(x_1), \dots, \text{frac}(x_s)).$$

Let $N, q \in \mathbb{N}$ and $\mathcal{P} = \{x_0, \dots, x_{N-1}\} \subset [0,1)^s$ be a N-element point set used for QMC integration (e.g. from a Halton or Sobol sequence). We generate q s-dimensional random numbers $u_0, \dots, u_{q-1} \overset{iid}{\sim} U([0,1]^s)$ and for each $k = 0, \dots, q-1$ define a new, *shifted* sequence with $\mathcal{P}^{(k)} = \{y_0^{(k)}, \dots, y_{N-1}^{(k)}\}$, where

$$y_n^{(k)} := \text{frac}(x_n + u_k). \tag{1.7.1}$$

The fractional part is required here, so that the shifted points do not "fall" out of the interval $[0,1]^s$. Observe, that each shifted point $y_n^{(k)}$ is uniformly distributed in $[0,1]^s$ for all $n = 0, \dots, N-1$ and $k = 0, \dots, q-1$.

With

$$I_N^{(k)}(f) = \frac{1}{N} \sum_{n=0}^{N-1} f(y_n^{(k)}),$$

we denote the k-th **RQMC repeat**.

We have

$$\mathbb{E}[I_N^{(k)}(f)] = \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}[f(y_n^{(k)})] = \int_{[0,1]^s} f(x)dx \tag{1.7.2}$$

because of $y_n^{(k)} \overset{iid}{\sim} U([0,1]^s)$, therefore this estimator is unbiased. As the final **RQMC estimate**, we take the mean of all repeats,

$$\hat{I}_{N,q}(f) = \frac{1}{q} \sum_{k=0}^{q-1} I_N^{(k)}(f).$$

To construct a confidence interval and measure the error, we also need an estimate for $\text{Var}[\hat{I}_{N,q}(f)]$. With

$$s^2_{f,RQMC} = \frac{1}{q(q-1)} \sum_{k=0}^{q-1} (I_N^{(k)}(f) - \hat{I}_{N,q}(f))^2 \tag{1.7.3}$$

we have an unbiased estimator for the variance[1] [2].

The number of repeats, q, is usually chosen fixed around 10-50, while N is taken in the thousands or more. For a fair comparison between MC and RQMC, or between s_f^2 and $s_{f,RQMC}^2$ respectively, one should take $N^{MC} = q \cdot N^{RQMC}$, s.t. the number of function evaluations is the same in both cases ([2]).

If $\mathcal{P}$ was a (t, m, s)-net, its shifted versions $\mathcal{P}^{(k)}$ are no (t, m, s)-nets any more. In the next sections, we shall see randomizations, which preserve the (t, m, s)-net structures of the sequence.

1.7.2 Digital Shift

Definition 1.7.2. Let $b \in \mathbb{N}$, $b \geq 2$ and $x, u \in [0, 1)$ with the base b representations

$$x = (0.x_1x_2\ldots)_b \text{ and } u = (0.u_1u_2\ldots)_b.$$

The **digit-wise addition operator** in base b is defined as

$$y = x \oplus_b u,$$

where $y = (0.y_1y_2...)_b$ with the digits in base b being

$$y_i = x_i + u_i \mod b.$$

For vectors $x = (x_1, \ldots, x_s)$ and $u = (u_1, \ldots, u_s) \in [0, 1^s$ define

$$x \oplus_b u = (x_1 \oplus_b u_1, \ldots, x_s \oplus_b u_s).$$

The randomization technique using digital shift works completely analogue to the additive shift, just replace Equation (1.7.1) by

$$y_n^{(k)} := x_n \oplus_b u_k. \tag{1.7.4}$$

With this construction, we again obtain an unbiased estimator (see [3], Section 13.1). But what is most important here is, that the digital shift preserves the (t, m, s)-net structure, i.e., if $\mathcal{P} = \{x_0, \ldots, x_{N-1}\}$ is a (t, m, s)-net in base b, then $\mathcal{P}^{(k)} = \{y_0^{(k)}, \ldots, y_{N-1}^{(k)}\}$ defined via Equation (1.7.4) is again a (t, m, s)-net with probability 1 (a proof can be found e.g. in [3], Lemma 4.67). This way, all the theoretical results for (t, m, s)-nets are still valid for the digitally shifted point set.

Remark 1.7.3. For the Sobol sequence, which is a (t, s)-sequence in base 2, the digit-wise addition can be easy implemented with bitwise XOR-operation.

1.7.3 Scrambled Nets

A more sophisticated way of randomizations are the **scrambled nets** introduced by Owen 1997 [19], therefore often also referred to as **Owen's scrambling**. This algorithm uses a hierarchy of permutations [5], for a single coordinate dimension this works as follows: Divide the unit interval into b partitions of length $\frac{1}{b}$ and randomly permute those partitions.

[1] In the source, $s_{f,RQMC}^2$ is said to be an unbiased estimate for the mean square error, which coincides with the variance in this context

Further divide each partition into b subintervals of length $\frac{1}{b^2}$ and permute each of those b partitions within itself, where each partition get its own random and independent permutation. Continue this procedure, s.t. at the j-th step, b^{j-1} partitions are constructed, each consisting of b subintervals of length $\frac{1}{b^j}$, and permute each partition independently. For a s-dimensional point set, apply this procedure to each coordinate, using independent sets of coordinates for each dimension.

To make this description more formal, consider a point $x \in [0, 1)$ in one dimension, with base b expansion $x = (0.x_1x_2\ldots)_b$. This point is mapped to a new point $y = (0.y_1y_2\ldots)_b$ the following way:

- Choose a random permutation π on $\{0, \ldots, b-1\}$ and set $y_1 = \pi(x_1)$.
- For each of the b possible values of x_1, fix a random permutation π_{x_1} and set $y_2 = \pi_{x_1}(x_2)$.
- For each of the b^2 possible value combinations of x_1 and x_2, fix a random permutation π_{x_1,x_2} and set $y_3 = \pi_{x_1,x_2}(x_3)$.
- ...
- In step j, fix b^{j-1} permutations for all possible combinations of $x_1, x_2, \ldots, x_{j-1}$ and set $y_j = \pi_{x_1,x_2,\ldots,x_{j-1}}(x_j)$.

This algorithm yields again an unbiased estimator and also preserves the (t, m, s)-net property with probability 1, as can be seen e.g. in [3]. What makes scrambling particularly interesting, is the increased order of convergence. If the integrand is sufficiently smooth, a convergence rate of $O(N^{-3/2+\delta})$ can be achieved (see e.g. Theorem 6.25 in [2]). A sufficient condition on the integrand would be having square integrable partial mixed derivatives up to order one in each variable.

Because Owen's scrambling algorithm needs to generate and store all the necessary permutations, the number of which grows exponentially with the depth of the algorithm, this is again not feasible in practice. Matoušek proposed a simplified version in [16], where he greatly reduces the number of permutations involved, while still preserving the main properties of Owen's scrambling. For the base b expansions of $x = (0.x_1x_2\ldots)_b$ and $y = (0.y_1y_2\ldots)_b$ he only considers permutations of the form

$$y_j = \sum_{i=1}^{j} h_{ij}x_i + g_j \mod b, \tag{1.7.5}$$

where the g_j's and the h_{ij}'s with $i < j$ are chosen randomly and independently from $\{0, 1, ..., b-1\}$, and the h_{jj}'s are chosen randomly and independently from $\{1, 2, ..., b-1\}$. Note, that Equation (1.7.5) also incorporates the permutation's dependence on the previous digits $x_1, \ldots, x_{j-1}$ and therefore avoids the exponential number of permutations required for the "full scrambling". This method, also called **"affine linear scrambling"** or "affine matrix scrambling" (just write the above equation in matrix notation) is available in MATLAB, with the references being [16] and [8].

Chapter 2

Introduction to Mathematical Finance

Since the aim of this work is to apply the Monte Carlo and quasi-Monte Carlo methods, discussed in the previous Chapter 1, to problems arising from mathematical finance, we first introduce a general market model with a diffusion process and later state the Black-Scholes model as a special case of the general model. The Sections 2.1 and 2.2 follow to a great part the reasoning of Shreve [22]. The Black-Scholes model discussed in Section 2.3 is the main source for the problems in the practical part of this work. We assume the reader to have knowledge about probability theory and stochastic calculus.

2.1 The General Diffusion Model

Definition 2.1.1 (**Market Model 1**)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathcal{F}(t)$, $0 \leq t \leq T$, such that $\mathcal{F}(s) \subseteq \mathcal{F}(t) \subseteq \mathcal{F}$ for $0 \leq s \leq t \leq T$. We assume to have an adapted **interest rate process** $R(t)$, with which we define the **discount process** as

$$D(t) = e^{-\int_0^t R(u)du}.$$

Furthermore we assume to have an adapted, d-dimensional Brownian motion $W(t) = (W_1(t), \ldots, W_d(t))$ and model the price of m stocks as

$$dS_i(t) = \alpha_i(t)S_i(t)dt + S_i(t)\sum_{i=1}^{d} \sigma_{ij}(t)dW_j(t), \quad i = 1, \ldots, m \tag{2.1.1}$$

where the mean rate of return vector $(\alpha_1(t), \ldots, \alpha_m(t))$ and the $m \times d$ volatility matrix $(\sigma_{ij(t)})$ are also assumed to be adapted processes.

In this model, we are free to invest in a bank account with continuous interest rate $R(t)$, as well as in any of the m stocks with price $S_1(t), \ldots, S_m(t)$. One unit of the bank account is assumed to start with the value 1 at time $t = 0$, and its value at time t is given as

$$e^{\int_0^t R(u)du} = \frac{1}{D(t)}.$$

It is assumed, that the market is always liquid, that we have no transaction costs and that short positions are possible.

Proposition 2.1.2. *The solution of equation* (2.1.1) *is given as*

$$S_i(t) = S_i(0) \cdot \exp\left(\int_0^t \alpha_i(u) - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2(u)du + \sum_{j=1}^{d} \int_0^t \sigma_{ij}(u)dW_j(u) \right)$$

Proof. We consider only the one-dimensional case ($m = d = 1$). Let $f(x) = e^x$, and $X(t) = \int_0^t \alpha_1(u) - \frac{1}{2}\sigma_{11}^2(u)du + \int_0^t \sigma_{11}(u)dW_1(u)$. Our candidate for the solution has then the form $S_1(t) = e^{X(t)} = f(X(t))$. Applying the Itô-formula, which can be found in [22], yields

$$\begin{aligned} dS_1(t) = df(X(t)) &= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))(dX(t))^2 \\ &= e^{X(t)}\Big((\alpha_1(t) - \frac{1}{2}\sigma_{11}^2(t))dt + \sigma_{11}(t)dW_1(t) + \frac{1}{2}\sigma_{11}^2 dt\Big) \\ &= S_1(t)\Big(\alpha_1(t)dt + \sigma_{11}dW_1(t)\Big). \end{aligned}$$

□

The main application of models like this, is to determine the fair price of financial derivatives, which often cannot be observed on the market. In this work, we will focus on the pricing of European options in path-independent, as well as path-dependent settings. A European option is a contract, that pays some amount $V(T)$ at a fixed time T (the "maturity"), and this amount is explicitely defined through a payoff-function h. Thus, $V(T)$ has to be a $\mathcal{F}(T)$-measurable random variable. We consider only two cases:

- $V(T) = h(S_1(T), \ldots, S_m(T))$ in the path-independent setting, i.e., the payoff depends only on the final value of one or more stocks, or
- $V(T) = h(S(t_1), S(t_2), \ldots, S(t_n))$ in the path-dependent setting, which means, the payoff depends on the history of the value of one stock at fixed observation points $0 < t_1 < t_2 < \ldots < t_n = T$. Since we don't observe the stock's value at all times $0 < t < T$, this is only an approximation, hence we call this a **discretely sampled** path-dependent option.

A European call option on a single asset S has the payoff-function $h(S(T)) = (S(T) - K)^+$, where K is the strike price defined in the contract. Today's approach to option pricing is through the so-called "risk-neutral measure", which we will define in the following paragraphs, but in order to be able to do this, we will first recall some useful tools from probability theory and stochastic calculus.

Proposition 2.1.3 (**Change of measure** [22])**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* Z *be an almost surely nonnegative random variable with* $\mathbb{E}[Z] = 1$*. For* $A \in \mathcal{F}$ *define*

$$\tilde{\mathbb{P}}[A] = \int_A Z(\omega)d\,\mathbb{P}(\omega). \tag{2.1.2}$$

Then $\tilde{\mathbb{P}}$ *is a probability measure. Furthermore, if* X *is a nonnegative R.V., then*

$$\tilde{\mathbb{E}}[X] = \mathbb{E}_{\tilde{\mathbb{P}}}[X] = \mathbb{E}[XZ].$$

If Z *is almost surely strictly positive, we also have*

$$\mathbb{E}[Y] = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right].$$

Definition 2.1.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathcal{F}(t)$, defined for $0 \leq t \leq T$ with $\mathcal{F}(T) \subseteq \mathcal{F}$, where T is a fixed final time. Suppose further that Z is an almost surely positive R.V. satisfying $\mathbb{E}[Z] = 1$ and we define $\tilde{\mathbb{P}}$ by (2.1.2). We call the process

$$Z(t) = \mathbb{E}[Z|\mathcal{F}(t)], \quad 0 \leq t \leq T \tag{2.1.3}$$

the **Radon-Nikodým derivative process**.

Theorem 2.1.5 (**Girsanov** [22])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathcal{F}(t)$, defined for $0 \leq t \leq T$, where T is a fixed final time. Let $W(t) = (W_1(t), \ldots, W_d(t))$ be an adapted, d-dimensional Brownian motion, and let $\Theta(t) = (\Theta_1(t), \ldots, \Theta_d(t))$ be an adapted d-dimensional process. Define*

$$Z(t) = \exp\bigg(-\int_0^t \Theta(u) \cdot dW(u) - \frac{1}{2}\int_0^t \|\Theta(u)\|^2 du\bigg), \tag{2.1.4}$$

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)du, \tag{2.1.5}$$

and assume that $\mathbb{E}\left[\int_0^T \|\Theta(u)\|^2 du\right] < \infty$. Set $Z = Z(T)$. Then $\mathbb{E}[Z] = 1$, and under the probability measure $\tilde{\mathbb{P}}$ given as in (2.1.2)*, the process $\tilde{W}(t)$ is a d-dimensional Brownian motion.*

Consider now the Market Model 1 in one dimension (m=1) with one driving Brownian motion (d=1), i.e.

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t), \tag{2.1.6}$$

with $\sigma(t) > 0$ almost surely. Because of $dD(t) = -R(t)D(t)dt$ and Itô's product rule, we have for the differential of the discounted stock price

$$\begin{aligned} d(D(t)S(t)) &= D(t)dS(t) + S(t)dD(t) + \underbrace{dS(t)dD(t)}_{=0} \\ &= D(t)(\alpha(t)S(t)dt + \sigma(t)S(t)dW(t)) + S(t)(-R(t))D(t)dt \\ &= (\alpha(t) - R(t))D(t)S(t)dt + \sigma(t)D(t)S(t)dW(t) \qquad (2.1.7) \\ &= \sigma(t)D(t)S(t)\bigg(\frac{\alpha(t) - R(t)}{\sigma(t)}dt + dW(t)\bigg). \qquad (2.1.8) \end{aligned}$$

Let us set

$$\Theta(t) = \frac{\alpha(t) - R(t)}{\sigma(t)} \tag{2.1.9}$$

and with that, define

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)du, \tag{2.1.10}$$

so we can rewrite (2.1.8) to

$$d(D(t)S(t)) = \sigma(t)D(t)S(t)d\tilde{W}(t). \tag{2.1.11}$$

If we now define a new measure $\tilde{\mathbb{P}}$ as in Girsanov's theorem, with the $\Theta(t)$ as in (2.1.9) and the additional assumption $\mathbb{E}\left[\int_0^T \Theta^2(d)du\right] < \infty$, then $\tilde{W}(t)$ is a Brownian motion under $\tilde{\mathbb{P}}$. In integral notation, the equation (2.1.11) writes as

$$D(t)S(t) = \underbrace{D(0)}_{=1}S(0) + \int_0^t \sigma(u)D(u)S(u)d\tilde{W}(u), \tag{2.1.12}$$

which means, it is an Itô integral, hence, the discounted stock price is a martingale under $\tilde{\mathbb{P}}$. We call $\tilde{\mathbb{P}}$ the risk-neutral measure, but let us define this more formally:

Definition 2.1.6. A probability measure $\tilde{\mathbb{P}}$ is called a **risk-neutral measure**, if

1. $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent (i.e. $\tilde{\mathbb{P}}[A] = 0 \iff \mathbb{P}[A] = 0$), and
2. under $\tilde{\mathbb{P}}$, the discounted stock prices $D(t)S_i(t)$ are martingales for $i = 1, \dots, m$.

The process $\Theta(t) = \frac{\alpha(t)-R(t)}{\sigma(t)}$ is called **market price of risk** process and it requires the volatility $\sigma(t)$ to be strictly positive. The risk-neutral measure eliminates the deterministic drift term - the market price of risk - from the differential equation (e.g. (2.1.8)), which explains its name.

Now let us consider the evolution of a portfolio's value under the risk-neutral measure. Let $X(t)$, $0 \le t \le T$, denote the portfolio process, with initial capital $X(0)$. Suppose at each time t, the investor holds $\Delta(t)$ shares of stock. The position $\Delta(t)$ can be random, but must be adapted to the filtration. The remainder of the portfolio value, $X(t) - \Delta(t)S(t)$, is invested in the bank account. This means, that the differential of the portfolio consists of a capital gain $\Delta(t)dS(t)$ due to changes in the stock price, and interest earnings $R(t)(X(t) - \Delta(t)S(t))dt$ from the bank account. Hence we have

$$dX(t) = \Delta(t)dS(t) + R(t)(X(t) - \Delta(t)S(t))dt.$$

Expanding the term for $dS(t)$ yields

$$\begin{aligned} dX(t) &= \Delta(t)\Big(\alpha(t)S(t)dt + \sigma(t)S(t)dW(t)\Big) + R(t)\Big(X(t) - \Delta(t)S(t)\Big)dt \\ &= R(t)X(t)dt + \Delta(t)(\alpha(t) - R(t))S(t)dt + \Delta(t)\sigma(t)S(t)dW(t) \\ &= R(t)X(t)dt + \Delta(t)\sigma(t)S(t)\Big(\Theta(t)dt + dW(t)\Big). \end{aligned}$$

Again, we apply Itô's product rule to the differential of the discounted portfolio process:

$$\begin{aligned} d(D(t)X(t)) &= X(t)dD(t) + D(t)dX(t) + \underbrace{dX(t)dD(t)}_{=0} \\ &= X(t)(-R(t))D(t)dt + D(t)R(t)X(t)dt \\ &\quad + \Delta(t)\sigma(t)D(t)S(t)\Big(\Theta(t)dt + dW(t)\Big) \\ &= \Delta(t)\sigma(t)D(t)S(t)d\tilde{W}(t) \qquad (2.1.13) \\ &= \Delta(t)d(D(t)S(t)). \qquad (2.1.14) \end{aligned}$$

From equation (2.1.13) we can now see, via an argument analogue to equation (2.1.12), that the discounted portfolio process is a martingale under $\tilde{\mathbb{P}}$, if the investment strategy $\Delta(t)$ meets some technical assumptions, e.g. progressive measurability and square integrability.

Let $V(T)$ be an $\mathcal{F}(T)$-measurable R.V. representing the payoff of a derivative security at time T. The payoff can be path-dependent, but we assume it to meet the integrability condition $\mathbb{E}[V^2(T)] < \infty$. Our aim is to find a **hedge**, which is a portfolio process $X(t)$, $0 \le t \le T$, with the corresponding investment strategy $\Delta(t)$ and initial capital $X(0)$, s.t.

$$X(T) = V(T) \quad \text{almost surely.}$$

So regardless, how the payoff $V(T)$ will be, our hedging portfolio X should yield the exact same payoff at time T. We shall see later, that this investment strategy actually exists. Once we know the hedging portfolio (sometimes also replicating portfolio), we have

$$D(t)X(t) = \tilde{\mathbb{E}}[D(T)X(T)|\mathcal{F}(t)] = \tilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}(t)],$$

because $D(t)X(t)$ is a martingale under $\tilde{\mathbb{P}}$. The value $X(t)$ of the hedging portfolio is the capital needed at time t in order to replicate the payoff $V(T)$ of the derivative security at time T. Therefore, we can say, it is also the price $V(t)$ of the derivative security at time t, and rewrite the above equation to

$$D(t)V(t) = \tilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}(t)], \quad 0 \leq t \leq T.$$

We can divide the equation by $D(t)$, because it is $\mathcal{F}(t)$-measurable, and get

$$V(t) = \tilde{\mathbb{E}}\left[\frac{D(T)}{D(t)}V(T)|\mathcal{F}(t)\right] = \tilde{\mathbb{E}}\left[e^{-\int_t^T R(u)du}V(T)|\mathcal{F}(t)\right]. \tag{2.1.15}$$

With this formula, we are now able to determine the price of a derivative security with a payoff only at terminal time T in a one-dimensional diffusion model, provided, the investment strategy $\Delta(t)$, $0 \leq t \leq T$, exists. The argumentation here is based on Shreve [22]. The same formula holds in the multidimensional case with completely analogue arguments and under the assumption, that we can find a risk-neutral measure and a hedging investment strategy. We shall refer to equation (2.1.15) as the **risk-neutral pricing formula**.

2.2 The Fundamental Theorems of Asset Pricing

In this section, we will discuss a bit more about, how a market model is qualified to be used for asset pricing, which requirements it should fulfil and what it should not do. Two key features stand out in this context: The absence of arbitrage and completeness.

Definition 2.2.1. An **arbitrage** is a portfolio value process $X(t)$ satisfying $X(0) = 0$ and also satisfying for some time $T > 0$

$$\mathbb{P}[X(t) \geq 0] = 1, \quad \mathbb{P}[X(T) > 0] > 0.$$

In other words, it is an investment strategy starting with zero capital, which never makes losses, but it makes profits with a positive probability. It should be somehow clear, that a sensible model must not allow arbitrage. If arbitrage is possible over a longer period of time, everybody would be able to generate earnings in arbitrary height, which is not realistic. Although there are sometimes possibilities for arbitrage in real markets, it is assumed, that if people exploit this possibility, the market automatically rebalances its prices through the supply and demand principle and eliminates the arbitrage.

Definition 2.2.2. A market model is **complete**, if every derivative security can be hedged.

In order to come up with the risk-neutral pricing formula, we assumed, that a hedge exists. Therefore, completeness of a market model is particularly useful, in the sense that,

since we have a hedge for every derivative security, we can apply the risk-neutral pricing formula to every derivative security within the model.

Now we present two characterizations about when a model fulfils these desired properties.

Theorem 2.2.3 (**First fundamental theorem of asset pricing** [22])**.** *If a market model has a risk-neural probability measure, then it does not admit arbitrage.*

Theorem 2.2.4 (**Second fundamental theorem of asset pricing** [22])**.** *Consider a market model that has a risk-neutral probability measure. The model is complete if and only if the risk-neutral probability measure is unique.*

In our one-dimensional market model 1, we constructed the risk-neutral measure via inserting the $\Theta(t)$ from (2.1.9) into Girsanov's theorem. The risk-neutral measure is unique, because the $\Theta(t)$ is unique (It can be shown that for every measure equivalent to $\mathbb{P}$ one can find a $\Theta(t)$, i.e., two distinct measures have two distinct $\Theta(t)$'s in Theorem 2.1.5). Hence this model is complete and free of arbitrage. We will now investigate the multidimensional case ($m > 1$ and $d > 0$). One can derive a formula for the differentials of the discounted stocks analogue to equation (2.1.7):

$$d(D(t)S_i(t)) = D(t)S_i(t)\Big((\alpha_i(t) - R(t))dt + \sum_{j=1}^{d} \sigma_{ij}(t)dW_j(t)\Big), \quad i = 1, \dots, m. \tag{2.2.1}$$

In order to make the discounted stock prices martingales, we would like to rewrite (2.2.1) as

$$d(D(t)S_i(t)) = D(t)S_i(t)\sum_{j=1}^{d} \sigma_{ij}(t)\Big(\underbrace{\Theta_j(t)dt + dW_j(t)}_{=d\tilde{W}_j(t)}\Big), \quad i = 1, \dots, m \tag{2.2.2}$$

and use Girsanov's theorem, such that $\tilde{W}(t)$ is again a Brownian motion under the new measure. We need to find the process $\Theta(t) = (\Theta_1(t), \dots, \Theta_d(t))$, s.t.

$$\alpha_i(t) - R(t) = \sum_{j=1}^{d} \sigma_{ij}(t)\Theta_j(t), \quad i = 1, \dots, m \tag{2.2.3}$$

for the equations (2.2.1) and (2.2.2) to be equivalent. This is a system of m equations in the d unknown processes $\Theta_1(t), \dots, \Theta_d(t)$, which are called the *market price of risk equations*. To ensure the existence and uniqueness of a solution to (2.2.3), we add conditions to the market model 1 and call it market model 1a.

Definition 2.2.5. Let the **market model 1a** be defined in the same way as market model 1, with the following additional assumptions:

- $m = d$, s.t. the number of stocks is the same as the number of driving Brownian motions. We shall refer to both parameters with d.
- The matrix $\sigma(t) = (\sigma_{ij}(t))_{i=1,\dots,d;j=1,\dots,d}$ is invertible for all $0 \leq t \leq T$.

With these assumptions, we have exactly one solution to (2.2.3) and, with the help of Girsanov's theorem, a uniquely defined risk-neutral measure. This has the implication, through the first and second fundamental theorem of asset pricing, that the market model 1a is free of arbitrage and complete.

The key tool to find the hedge of a derivative security is the following theorem:

Theorem 2.2.6 (**Martingale representation** [22])**.** *Let T be a fixed positive time, and assume that $\mathcal{F}(t)$, $0 \leq t \leq T$, is the filtration generated by the d-dimensional Brownian motion $W(t)$,$0 \leq t \leq T$. Let $M(t)$, $0 \leq t \leq T$, be a martingale with respect to this filtration under $\mathbb{P}$ with $\mathbb{E}[M^2(T)] < \infty$. Then there is an adapted, d-dimensional process $\Gamma(u) = (\Gamma_1(u), \ldots, \Gamma_d(u))$, $0 \leq u \leq T$, such that*

$$M(t) = M(0) + \int_0^t \Gamma(u) \cdot dW(u), \quad 0 \leq t \leq T.$$

If, in addition, we assume the notation and assumptions of Theorem 2.1.5 and if $\tilde{M}(t)$, $0 \leq t \leq T$, is a $\tilde{\mathbb{P}}$-martingale with $\mathbb{E}[\tilde{M}^2(T)] < \infty$, then there is an adapted, d-dimensional process $\tilde{\Gamma}(u) = (\tilde{\Gamma}_1(u), \ldots, \tilde{\Gamma}_d(u))$ such that

$$\tilde{M}(t) = \tilde{M}(0) + \int_0^t \tilde{\Gamma}(u) \cdot d\tilde{W}(u), \quad 0 \leq t \leq T.$$

Assume we have a derivative security with payoff $V(T)$ at the maturity T. It's discounted price process $D(t)V(t)$, $0 \leq t \leq T$, obtained through the risk-neutral pricing formula, is a martingale under $\tilde{\mathbb{P}}$. Because of the martingale representation theorem, we know, that there is a process $\tilde{\Gamma}(u) = (\tilde{\Gamma}_1(u), \ldots, \tilde{\Gamma}_d(u))$ such that

$$D(t)V(t) = V(0) + \sum_{j=1}^{d} \int_0^t \tilde{\Gamma}_j(u) d\tilde{W}_j(u), \quad 0 \leq t \leq T. \tag{2.2.4}$$

Let $X(t)$ be a portfolio value process starting with captial $X(0)$, and let $\Delta_i(t)$ denote the share of the i-th stock $S_i(t)$ at time $0 \leq t \leq T$ for $i = 1, \ldots, d$. Analogue to equation (2.1.14), we have

$$\begin{aligned} d(D(t)X(t)) &= \sum_{i=1}^{d} \Delta_i(t) d(D(t)S_i(t)) \\ &= \sum_{i=1}^{d} \sum_{j=1}^{d} \Delta_i(t) D(t) S_i(t) \sigma_{ij}(t) d\tilde{W}_j(t), \end{aligned}$$

and the same equation in integral notation with swapping the sums writes as

$$D(t)X(t) = X(0) + \sum_{j=1}^{d} \int_0^t \sum_{i=1}^{d} \Delta_i(u) D(u) S_i(u) \sigma_{ij}(u) d\tilde{W}_j(u). \tag{2.2.5}$$

If we want to achieve $X(t) = V(t)$, we need to have

$$\tilde{\Gamma}_j(t) = \sum_{i=1}^{d} \Delta_i(t) D(t) S_i(t) \sigma_{ij}(t), \quad j = 1, \ldots, d.$$

So here we have d equations in the d unknown processes $\Delta_1(t), \dots, \Delta_d(t)$, $0 \le t \le T$, which are solvable under the assumptions of market model 1a. The problem here is, that the martingale representation theorem is not constructive, it gives us only the existence of the $\Gamma_1(u), \dots, \tilde{\Gamma}_d(u)$, hence we have only the existence of the hedging strategy $\Delta_1(t), \dots, \Delta_d(t)$. We shall see a concrete formula for the hedging strategy in the next section about the Black-Scholes model.

2.3 The Black-Scholes Model

Definition 2.3.1. The d-dimensional **Black-Scholes model** is defined in the same way as the market model 1a from Definition 2.2.5, but with the processes $\alpha_i(t)$, $\sigma_{ij}(t)$ and $R(t)$ being deterministic and constant over time. Thus we write them without the dependence on t, and call $\alpha = (\alpha_1, \dots, \alpha_d)$ the mean rate of return vector, $\sigma = (\sigma_{ij})_{i=1,\dots,d;j=1,\dots d}$ the volatility matrix, which has to be invertible, and R the continuous interest rate.

In this model, the discounting factor $D(t)$ becomes

$$D(t) = e^{-Rt}.$$

The solution to the now simplified stochastic differential equation (2.1.1) becomes

$$S_i(t) = S_i(0) \cdot \exp\left(t\Big(\alpha_i - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2\Big) + \sum_{j=1}^{d}\sigma_{ij}W_j(t)\right), \quad i = 1, \dots, d,$$

which writes completely without (stochastic) integrals.

Proposition 2.3.2. *Let $X(t) = (X_1(t), \dots, X_d(t))$ where $X_i(t) = \log \frac{S_i(t)}{S_i(0)}$. Then $X(t)$ is multivariate normally distributed, with*

$$\mathbb{E}[X_i(t)] = t(\alpha_i - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2) \quad \textit{and} \quad \mathrm{Cov}[X_i(t), X_k(t)] = t\sum_{j=1}^{d}\sigma_{ij}\sigma_{kj}.$$

With $\Sigma = \sigma\sigma^T$, this can be written as

$$X(t) \sim N\Big(t(\alpha - \frac{1}{2}\,\mathrm{diag}(\Sigma)), t\Sigma\Big).$$

Proof. Observe, that

$$X_i(t) = t(\alpha_i - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2) + \sum_{j=1}^{d}\sigma_{ij}W_j(t).$$

Because of $\mathbb{E}[W_j(t)] = 0$, the part $\mathbb{E}[X_i(t)] = t(\alpha_i - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2)$ can be easily seen. We know, that $W_j(t)$ and $W_l(t)$ are independent for $j \neq l$ and $N(0,t)$-distributed, hence we

have

$$\begin{aligned}\mathrm{Cov}[X_i(t), X_k(t)] &= \mathbb{E}\left[(X_i(t) - \mathbb{E}[X_i(t)])(X_k(t) - \mathbb{E}[X_k(t)])\right] \\ &= \mathbb{E}\Big[\sum_{j=1}^{d} \sigma_{ij} W_j(t) \sum_{l=1}^{d} \sigma_{kl} W_l(t)\Big] \\ &= \mathbb{E}\Big[\sum_{j=1}^{d} \sigma_{ij}\sigma_{kj} W_j^2(t)\Big] \\ &= \sum_{j=1}^{d} \sigma_{ij}\sigma_{kj} t,\end{aligned}$$

because the summands with $j \neq l$ are 0 under the expectation, and $\mathbb{E}[W_j^2(t)] = \mathrm{Var}[W_j(t)] = t$. □

Remark 2.3.3. The process $X(t)$ from Proposition 2.3.2 is called **log return**.

Proposition 2.3.4. *The price of an European option with payoff* $V(T) = h(S_1(T), \ldots, S_d(T))$ *is*

$$V(0) = e^{-RT}\tilde{\mathbb{E}}[h(S_1(T), \ldots, S_d(T))] = e^{-RT} \int_{\mathbb{R}^d} h(S(0) \cdot e^x) f_X(x) dx,$$

where $f_X(x)$ *is the PDF of the log return, i.e. the PDF of a multivariate normal distribution with mean and covariance matrix as in Proposition 2.3.2, and the short-hand notation* $S(0) \cdot e^x$ *denotes the element-wise product and exponential function, s.t.*

$$S(0) \cdot e^x = (S_1(0)e^{x_1}, \ldots, S_d(0)e^{x_d}),$$

Consider the log return process in the one-dimensional case,

$$X(t) = t(R - \frac{\sigma^2}{2}) + \sigma W(t). \tag{2.3.1}$$

Since the risk-neutral measure is used here, we have $\alpha = R$. We are interested in the covariance of the value of X at two different times, e.g. u and v. We know from the properties of the Brownian motion, that $\mathrm{Cov}[W(u), W(v)] = \min(u, v)$, hence we have

$$\mathrm{Cov}[X(u), X(v)] = \mathrm{Cov}[\sigma W(u), \sigma W(v)] = \sigma^2 \min(u, v).$$

For d observation times $0 < t_1 < \ldots < t_d$, the covariance matrix of the vector $(X(t_1), \ldots, X(t_d))$ has the form

$$\Sigma = \sigma^2 \cdot \begin{pmatrix} t_1 & t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & t_2 & \ldots & t_2 \\ t_1 & t_2 & t_3 & \ldots & t_3 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1 & t_2 & t_3 & \ldots & t_d \end{pmatrix}. \tag{2.3.2}$$

For the mean we have

$$\tilde{\mathbb{E}}[X(t_i)] = t_i(R - \frac{\sigma^2}{2}), \quad i = 1, \ldots, d. \tag{2.3.3}$$

Proposition 2.3.5. *We consider a discretely sampled path-dependent option in the one-dimensional Black-Scholes model, with payoff only at maturity T and the observation points being $0 < t_1 < \ldots < t_d = T$. Let the payoff function be $V(T) = h(S(t_1), \ldots, S(t_d))$. The price of this option is*

$$V(0) = e^{-RT}\tilde{\mathbb{E}}[h(S(t_1), \ldots, S(t_d))] = e^{-RT}\int_{\mathbb{R}^d} h(S(0)e^x) f_X(x)dx,$$

where $f_X(x)$ is the PDF of the log return $(X(t_1), \ldots, X(t_d))$, i.e., a multivariate normal distribution with mean as in equation (2.3.3) *and covariance matrix as in equation* (2.3.2).

We are now going to show, how to construct a hedge in the one-dimensional Black-Scholes model. Since the Brownian motion is a Markov process and the only source of uncertainty, we can replace the condition on $\mathcal{F}(t)$ in the risk-neutral pricing formula with the condition that the stock price process' value is known, thus

$$V(t) = e^{-R(T-t)}\tilde{\mathbb{E}}[V(T)|\mathcal{F}(t)] = e^{-R(T-t)}\tilde{\mathbb{E}}[V(T)|S(t) = x] =: F(t,x), \text{ for some } x > 0.$$

The so defined function $F(t,x)$ now represents the value of the option at time t, given the stock price at this moment is $S(t) = x$, hence we have $V(t) = F(t, S(t))$. We want a similar function for the discounted option value, where the second argument is the discounted stock price, so we define

$$\tilde{F}(t,x) = e^{-Rt}F(t, e^{Rt}x).$$

This new function $\tilde{F}(t,x)$ now fulfils

$$D(t)V(t) = \tilde{F}(t, D(t)S(t)).$$

Expanding this equation with Itô's formula (see e.g. [22] for more details), assuming regularity of $\tilde{F}$, yields

$$\begin{aligned} d(D(t)V(t)) &= d\tilde{F}(t, D(t)S(t)) \\ &= \frac{\partial \tilde{F}}{\partial t}(t, D(t)S(t))dt + \frac{\partial \tilde{F}}{\partial x}(t, D(t)S(t))d(D(t)S(t)) \\ &\quad + \frac{1}{2}\frac{\partial^2 \tilde{F}}{\partial x^2}(t, D(t)S(t))(d(D(t)S(t))^2 \\ &= U(t)dt + \frac{\partial \tilde{F}}{\partial x}(t, D(t)S(t))D(t)S(t)d\sigma\tilde{W}(t). \end{aligned}$$

But we know, that $D(t)V(t)$ is a martingale and has no deterministic drift term, hence $U(t) = 0$ and we have

$$D(t)V(t) = V(0) + \int_0^t \frac{\partial \tilde{F}}{\partial x}(u, D(u)S(u))D(t)S(u)\sigma d\tilde{W}(u). \tag{2.3.4}$$

Compare this equation (2.3.4) with equation (2.2.5) and we can directly see, that

$$\Delta(t) = \frac{\partial \tilde{F}}{\partial x}(t, D(t)S(t)).$$

In the multidimensional case, the x in $F(t,x)$ as well as in $\tilde{F}(t,x)$ would be a vector with d elements, and an analogue calculation with Itô's formula would give us

$$D(t)V(t) = V(0) + \sum_{i=1}^{d} \int_0^t \sum_{j=1}^{d} \frac{\partial \tilde{F}}{\partial x_i}(u, D(u)S(u))D(t)S_i(u)\sigma_{ij}d\tilde{W}_j(u). \tag{2.3.5}$$

After comparison with (2.2.5), we see that

$$\Delta_i(t) = \frac{\partial \tilde{F}}{\partial x_i}(t, D(t)S(t)), \quad i = 1, \ldots, d.$$

Thus for options, where we know $F(t,x)$ in a closed form, we can construct $\tilde{F}(t,x)$ as above and calculate the derivatives to get the hedging strategy $(\Delta_i(t))_{i=1,\ldots,d}$. In other cases, this can be done approximatively with numerical methods.

Chapter 3

Practical Aspects

For the purpose of this work, we want to write all our problems in a standardized form. As it is common in QMC literature, we denote the dimension of the integral with s. For almost all cases in this section, X denotes a (general) multivariate normally distributed R.V., whereas Y denotes multivariate standard normally distributed R.V.. In both cases of options we mentioned in the Section 2.3, we had an integral of the form

$$\int_{\mathbb{R}^s} h(Ce^x) f_X(x) dx, \tag{3.0.1}$$

where $Ce^x = (C_1 e^{x_1}, \dots, C_s e^{x_s})$ for a vector of constants C, and f_X being the PDF of a multivariate normal distribution $N(\mu, \Sigma)$. All of our problems considered in Chapter 5 can be written in the form of (3.0.1). In order to successfully apply a QMC-method to this integral, we first need to decide on the following points:

- How to transform it from $\mathbb{R}^s$ to $[0,1]^s$,
- in which of the functions h and f we should incorporate the covariance, and
- how to factorize the covariance matrix.

Although the theoretical value of the integral is always the same, regardless how we rewrite or transform it, it can make a huge difference in the quality of the result and the convergence of the QMC-method. We will mainly discuss three methods of transformation:

- Restricting the integral to the domain $[-c, c]^s$ and transform it linearly to $[0,1]^s$. By restricting we also make an additional error, which needs to be estimated.
- Transform each coordinate with the inverse normal CDF $x_i = \Phi^{-1}(z_i)$.
- Transform each coordinate with the inverse logistic function $x_i = \text{logit}(z_i) = \log \frac{z_i}{1-z_i}$

As to the question, where to incorporate the covariance, we set $g(x) = h(Ce^x)$ and consider our problem integral (3.0.1), which writes now as

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^s} g(x) f_X(x) dx. \tag{3.0.2}$$

As X is in general not standard normally distributed, it's components can have dependence, and the covariance matrix hides in the PDF:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^s |\det(\Sigma)|}} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \Big).$$

But if we have a factorization L of Σ at hand, s.t. $LL^T = \Sigma$ and use the transformation $x = Ly + \mu$, one could also write

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^s} g(Ly + \mu) f_Y(y) dy, \tag{3.0.3}$$

where $f_Y(y)$ is the PDF of a standard normal distribution. This representation can have some numeric advantages, especially if one uses some special factorizations of L, which we will see in Section 3.2. On the other hand, if $g(x)$ has finite variation, but is not differentiable, it can happen, that $g(Ly + \mu)$ has infinite variation (because the multiplication with L could cause some rotation), which is of course bad for our theoretical error estimation.

3.1 Transformation to the Unit Cube

3.1.1 Linear Transformation

The most simple way of transforming the integrand is to restrict our problem (3.0.1) to a finite domain, which should be big enough to capture the essential part of the integrand, and transform it linearly to the unit cube. We assume

$$\int_{\mathbb{R}^s} g(x) f_X(x) dx \approx \int_{[-c,c]^s} g(x) f_X(x) dx = (2c)^s \int_{[0,1]^s} g(\tau(z)) f_X(\tau(z)) dz, \tag{3.1.1}$$

with the transformation being

$$\tau(z) = 2cz - c\mathbf{1}.$$

The determinant of its Jacobian equals

$$|\det J_\tau(x)| = (2c)^s,$$

which explains the factor in front of the last integral in (3.1.1). An advantage of this linear transformation is, that the structure of the integrand is not changed, it is just rescaled, therefore calculating or estimating derivatives is equally easy (or hard) as for the original integrand. Of course, we introduce an additional error through the truncation of the integral domain, which is a disadvantage. In the following two sections, we provide methods to estimate the truncation error and to estimate the derivatives of the normal PDF, which can be used to further estimate the variation of the PDF. The availability of these methods motivate the use of the linear transformation.

Example 3.1.1. Consider an European Call option with maturity $T = 1$ years, strike $K = 0.7$ and starting value $S(0) = 1$. Let the parameters for the Black-Scholes model be $R = \alpha = 0$ and $\sigma = 20\%$ p.a. We write $S(T) = S(0)e^{X(T)}$, and we know the distribution of $X(T)$ under the risk-neutral measure, which is

$$X(T) \sim N(-\frac{\sigma^2}{2}T, \sigma^2 T) = N(-0.02, 0.04).$$

Now we have to calculate

$$\int_{-\infty}^{\infty} h(S(0)e^x) f_X(x) dx = \int_{-\infty}^{\infty} (e^x - 0.7)^+ \frac{1}{\sqrt{2\pi}0.2} e^{-\frac{(x+0.02)^2}{0.08}} dx, \tag{3.1.2}$$

which is actually not very hard to do in one dimension. We put this example here to show plots of the integrand, and of the effect of the different transformations on the integrand.

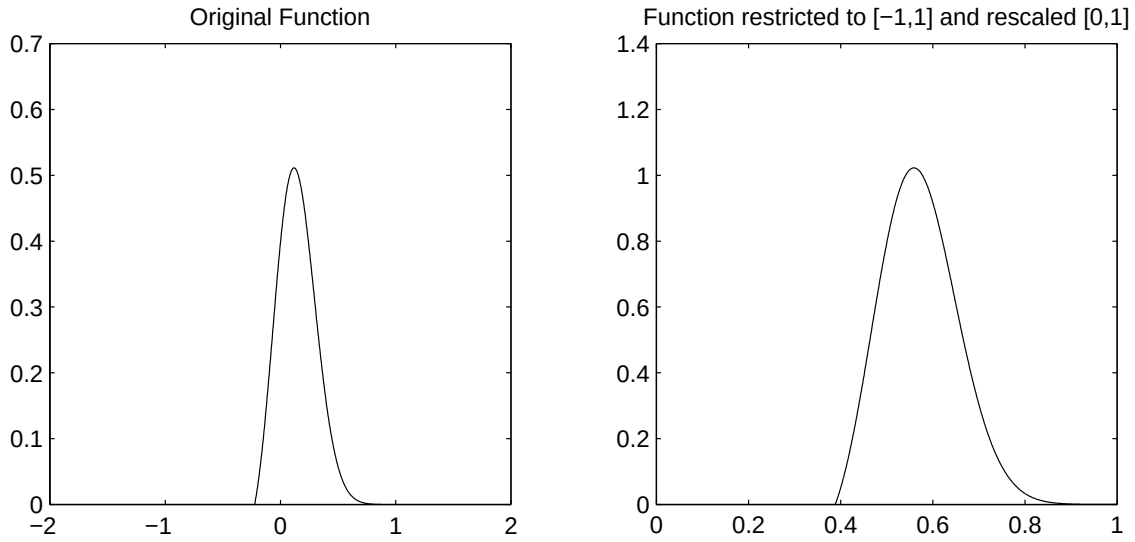

Figure 3.1.1: To the left, we see the original integrand from (3.1.2). For $x \leq \log K \approx -0.36$, the function is 0, whereas for $x > \log K$, the function value is always greater than 0, although it converges pretty fast. Hence it seems sensible, to truncate the integration domain at ± 1, s.t. that the truncation error is contained. To the right, we have the integrand transformed from $[-1, 1]$ to $[0, 1]$. In case of this simple payoff function, one can of course also choose the integration domain as $[\log K, 1]$, s.t. no integration nodes are used outside the support of the integrand. But for the most higher dimensional cases, this cannot be that easily observed as for this plot of a one-dimensional function.

3.1.2 Truncation Error

Since we need to truncate the integral domain in order to use the linear transformation, we also need to estimate the error introduced by the truncation. We do not want to have to do this estimated for every integrand separately, hence we try to do it for a class of integrands which contains many real world examples. We consider the class of derivative securities, which have a payoff function fulfilling

$$0 \leq h(S_1, \ldots, S_s) \leq \kappa_1 S_1 + \ldots + \kappa_s S_s,$$

for some positive constants $\kappa_1, \ldots, \kappa_s > 0$. The variables $S_1, \ldots, S_s$ here can either denote the $S_1(T), \ldots, S_s(T)$ as in Proposition 2.3.4 or the $S(t_1), \ldots S(t_s)$ as in Proposition 2.3.5. Under this assumptions, the function g fulfils

$$g(x) = h(Ce^x) \leq \kappa_1 e^{x_1} + \ldots + \kappa_s e^{x_s}. \tag{3.1.3}$$

We truncate each component variable at $\pm c$, hence our integration domain is $[-c, c]^s = \{x \in \mathbb{R}^s | \|x\|_\infty \leq c\}$. We start with the representation (3.1.1), which we write as the sum

$$\int_{\mathbb{R}^s} g(x) f_X(x) dx = \int_{\|x\|_\infty \leq c} \ldots dx + \int_{\|x\|_\infty > c} \ldots dx,$$

where the first term is the integral we actually calculate with some QMC-method, and the second term is the truncation error, which we will denote as $TE(c)$ from now on. Using

our assumption on the payoff, we estimate the truncation error by

$$TE(c) := \int_{\|x\|_\infty > c} g(x) f_X(x) dx \leq \sum_{i=1}^{s} \kappa_i \int_{\|x\|_\infty > c} e^{x_i} f_X(x) dx.$$

Each summand here can again be written as

$$\kappa_i \int_{\|x\|_\infty > c} e^{x_i} f_X(x) dx = \kappa_i \Bigg(\underbrace{\int_{\mathbb{R}^d} e^{x_i} f_X(x) dx}_{=:E_i} - \underbrace{\int_{\|x\|_\infty \leq c} e^{x_i} f_X(x) dx}_{=:R_i} \Bigg),$$

where the term E_i can be easily calculated, because

$$E_i = \mathbb{E}[e^{X_i}] = e^{\mu_i + \Sigma_{ii}/2}.$$

Since we want to estimate $TE(c)$ from above, we need to estimate the terms R_i from below. Let L be a decomposition of Σ with $LL^T = \Sigma$ and L_i the i-th line of L. We are here completely free in the choice of the decomposition. We use the transformation $x = Ly + \mu$, which yields

$$R_i = \int_{\|Ly+\mu\|_\infty \leq c} e^{L_i y + \mu_i} f_Y(y) dy,$$

where f_Y is the PDF of the multivariate standard normal distribution, which can be factored into the densities of univariate $N(0,1)$-distributions. If $\|y + L^{-1}\mu\|_\infty \leq \frac{c}{\|L\|_\infty}$ it follows from the definition of the operator norm, that

$$\|Ly + \mu\|_\infty = \|L(y + L^{-1}\mu)\|_\infty \leq \|L\|_\infty \cdot \|y + L^{-1}\mu\|_\infty \leq \|L\|_\infty \frac{c}{\|L\|_\infty}.$$

Therefore

$$\{y \in \mathbb{R}^s | \|y + L^{-1}\mu\|_\infty \leq \frac{c}{\|L\|_\infty}\} \subseteq \{y \in \mathbb{R}^s | \|Ly + \mu\|_\infty \leq c\},$$

and

$$\begin{aligned}
R_i &\geq \int_{\|y+L^{-1}\mu\|_\infty \leq \frac{c}{\|L\|_\infty}} e^{L_i y + \mu_i} f_Y(y) dy \\
&= e^{\mu_i} \prod_{j=1}^{s} \left(\int_{-\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j}^{+\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j} e^{L_{ij} y_j} f_{Y_j}(y_j) dy_j \right) \\
&\quad f_{Y_j} \text{ is just a } N(0,1)\text{-PDF, so via completing the square in the exponent, we get} \\
&= e^{\mu_i} \prod_{j=1}^{s} \left(\int_{-\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j}^{+\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j} e^{\frac{L_{ij}^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_j - L_{ij})^2}{2}} dy_j \right) \\
&\quad \text{substituting } z_j = y_j - L_{ij} \text{ yields} \\
&= e^{\mu_i} \prod_{j=1}^{s} \left(e^{\frac{L_{ij}^2}{2}} \int_{-\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j - L_{ij}}^{+\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j - L_{ij}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_j^2}{2}} dz_j \right) \\
&= e^{\mu_i + \frac{\Sigma_{ii}}{2}} \prod_{j=1}^{s} \left(\Phi\Big(\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j - L_{ij} \Big) - \Phi\Big(-\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j - L_{ij} \Big) \right),
\end{aligned}$$

also using the fact that $\sum_{j=1}^{s} L_{ij}^2 = \Sigma_{ii}$ To summarize our results, our estimate for the truncation error is

$$TE(c) \leq \sum_{i=1}^{s} \kappa_i e^{\mu_i + \Sigma_{ii}/2}(1 - D_i), \tag{3.1.4}$$

with

$$D_i = \prod_{j=1}^{s} \Big(\Phi\Big(\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j - L_{ij} \Big) - \Phi\Big(-\frac{c}{\|L\|_\infty} - (L^{-1}\mu)_j - L_{ij} \Big) \Big).$$

Obviously, as $c \to \infty$, we have $TE(c) \to 0$. Since the choice of the decomposition $LL^T = \Sigma$ is free, one can choose it to maximize the D_i-terms in order to minimize $TE(c)$, but this would be another question by itself.

3.1.3 Bounds on the Derivatives of the Normal Density

We also want to point out an idea for estimating the mixed derivatives (up to order one in each variable) of the normal density combined with the linear transformation, in order to estimate the variation in the sense of Hardy and Krause. The estimate presented here is very rough and it may be, that there are sharper bounds in the literature.

We consider just the PDF part of our integrand from equation (3.1.1),

$$\omega(z) = (2c)^s f_X(\tau(z))$$

with the transformation being $\tau(z) = 2cz - c\mathbf{1}$. We define some shorthand-notations, but keep in mind, that the letters still denote functions in z:

$$Q := -\frac{1}{2}(\tau(z) - \mu)^T \Sigma^{-1} (\tau(z) - \mu)$$

and

$$K = \frac{(2c)^s}{(2\pi)^{s/2}\sqrt{|\det \Sigma|}},$$

because then we can write the integrand as

$$\omega(z) = Ke^Q.$$

Lets start with the derivatives of the exponent,

$$Q_j := \frac{\partial}{\partial z_j} Q = -(\Sigma^{-1})_j (\tau(z) - \mu) 2c,$$

and

$$Q_{jk} := \frac{\partial^2}{\partial z_j \partial z_k} Q = -(\Sigma^{-1})_{jk} 4c^2.$$

Keeping this notation, we determine the mixed derivatives of the integrand $\omega(z)$:

$$\begin{aligned}
\partial^{\{1\}}\omega(z) &= Ke^Q Q_1, \\
\partial^{\{1,2\}}\omega(z) &= Ke^Q Q_2 Q_1 + Ke^Q Q_{12} = Ke^Q[Q_{12} + Q_1 Q_2], \\
\partial^{\{1,2,3\}}\omega(z) &= Ke^Q[Q_1 Q_2 Q_3 + Q_{12} Q_3 + Q_{13} Q_2 + Q_1 Q_{23}], \\
\partial^{\{1,2,3,4\}}\omega(z) &= Ke^Q[Q_1 Q_2 Q_3 Q_4 \\
&\qquad + Q_{12} Q_3 Q_4 + Q_{13} Q_2 Q_4 + Q14 Q_2 Q_3 \\
&\qquad + Q_1 Q_{23} Q_4 + Q_1 Q_{24} Q_3 + Q_1 Q_2 Q_{34} \\
&\qquad + Q_{12} Q_{34} + Q_{13} Q_{24} + Q_{14} Q_{23}].
\end{aligned}$$

One can already observe the pattern here to write the general formula. For a set $u \subseteq \mathcal{I}_s$, let $\mathcal{P}(u)$ denote the set of all partitions P of u, whose elements p are subsets of u with only 1 or 2 elements. We have

$$\partial^u \omega(z) = \Big(\prod_{i \in u} \frac{\partial}{\partial z_i} \Big) \omega(z) = K e^Q \sum_{P \in \mathcal{P}(u)} \prod_{p \in P} Q_p. \tag{3.1.5}$$

Lets find bounds for Q_j and Q_{jk}:

$$\max_j |Q_j| = \left\| \begin{pmatrix} Q_1 \\ \vdots \\ Q_d \end{pmatrix} \right\|_\infty \leq 2c \|\Sigma^{-1}\|_\infty \|\tau(z) - \mu\|_\infty,$$

and

$$\|\tau(z) - \mu\|_\infty \leq \max_i \Big(\max(|\tau_i(0) - \mu_i|, |\tau_i(1) - \mu_i|) \Big) = \max_i \Big(\max(|c + \mu_i|, |c - \mu_i|) \Big),$$

since $\tau(z)$ is linear, and the extrema of a linear function always occur on the boundary. For the second derivative of the exponent, we have

$$\sqrt{|Q_{jk}|} \leq 2c \max_{j \neq k} \sqrt{|(\Sigma^{-1})_{jk}|}.$$

We define q as

$$q := 2c \max \Big(\max_{j \neq k} \sqrt{|(\Sigma^{-1})_{jk}|}, \|\Sigma^{-1}\|_\infty \max_i \big(\max(|c + \mu_i|, |c - \mu_i|) \big) \Big),$$

because with this, we have $|Q_j| \leq q$, $|Q_{jk}| \leq q^2$ and $\prod_{p \in P} Q_p \leq q^{|u|}$. So for the derivatives of the actual integrand, we have

$$|\partial^u \omega(z)| \leq K \underbrace{e^Q}_{\leq 1} |\mathcal{P}(u)| q^{|u|}. \tag{3.1.6}$$

It would also be nice to have a formula for $|\mathcal{P}(u)|$, which we obtain by the following observation: The number of partitions of an n-element set into subsets of size 1 and 2 is equal to the number of involutions on the set $\{1, 2, \ldots, n\}$. An involution is a permutation, that is self-inverse. For a permutation Π to be self-inverse, it either has to map an element on itself, s.t. $\Pi(i) = i$, or if it maps an element to another element, $\Pi(j) = k$, then is also has to map the other direction, s.t. $\Pi(k) = j$, in other words, it switches the pair (k, j). Now it should be obvious, that each involution corresponds to exactly one partition with subsets of size 1 and 2: The elements of the 1-element subsets are mapped on themselves, and the elements of the 2-element subsets are swapped with each other. While we do not dig more into combinatorics at this point, we might refer to the literature, e.g. [10] and just state the results here.

Let Z_n be the number of involutions on $\{1, 2, \ldots, n\}$. It fulfils the recurrence relation $Z_n = Z_{n-1} + (n-1) Z_{n-2}$. Z_n can be also calculated from the formula

$$Z_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} (2k-1)!! = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{2^k (n-2k)! k!}.$$

The first few numbers are $Z_1 = 1$, $Z_2 = 2$, it goes on with 4, 10, 26, 76, ..., hence Z_n grows really fast. We can now replace the $|\mathcal{P}(u)|$ in equation (3.1.6) with $Z_{|u|}$. Nonetheless, this estimate is really rough and pessimistic and the practical use may be limited because of the fast growth in $|u|$.

3.1.4 Inverse Normal CDF Transformation

The transformation with the inverse CDF of the normal distribution seems natural for integrands with a normal density function. Before we apply the transformation, we have to decide, if we want to do it on the representation as in (3.0.3) or the representation as in (3.0.2). Let $\Phi(y)$ be the CDF of the standard normal distribution, and $\Phi^{-1}(z)$ be it's inverse. For vectors y and z, we define both $\Phi(y) = (\Phi(y_1), \dots, \Phi(y_s))$ and $\Phi^{-1}(z) = (\Phi^{-1}(z), \dots, \Phi^{-1}(z))$ element-wise. Let's start with the representation (3.0.3), where f_Y was the PDF of a standard normal R.V. and use the substitution $z_i = \Phi(y_i)$. We can rewrite the differential as $dz = f_Y(y)dy$, so we have

$$\int_{\mathbb{R}^s} g(Ly + \mu) f_Y(y) dy = \int_{[0,1]^s} g(L\Phi^{-1}(z) + \mu) dz. \tag{3.1.7}$$

Because of the choice of transformation, the PDF factor of the integrand has disappeared. Sometimes, this can be an advantage. For a unbounded function g however, the integrand is now also unbounded because $\Phi^{-1}(z)$ is unbounded on $[0,1]^s$, therefore its variation is also infinite. E.g., for a European call option with $h(S(T)) = (S(T) - K)^+$, we have an unbounded integrand, whereas for a binary option, the payoff is in $\{0,1\}$ and the resulting integrand also stays bounded. We also want to take a look at the case, in which we apply the transform to the representation (3.0.2), where $f_X(x)$ was the PDF of a general (not standard) normal distributed R.V. We substitute $z_i = \Phi(x_i)$, and receive

$$\int_{\mathbb{R}^s} g(x) f_X(x) dx = \int_{\mathbb{R}^s} g(x) \frac{f_X(x)}{f_Y(x)} f_Y(x) dx. = \int_{[0,1]^s} g(\Phi^{-1}(z)) \frac{f_X(\Phi^{-1}(z))}{f_Y(\Phi^{-1}(z))} dz. \tag{3.1.8}$$

The function $f_Y(x)$ here is again the PDF of a standard normal R.V., and the idea presented here is called *Importance Sampling*. One is not restricted, to use the standard normal PDF here for f_Y. A sensible choice is, to use a distribution f_Y, that has an easily invertible CDF, i.e., we have F_Y^{-1} available. Depending on the integrand, it could also make sense, to choose different transformations for each coordinate. In many cases, importance sampling yields better results, because one can choose, to put more weight on certain areas of the integration domain and less weight on areas, where e.g. the integrand is almost constant.

Example 3.1.1 revisited We begin with the same original problem, but now we apply the inverse normal CDF transformation. Consider first the transformation from (3.1.7):

$$\int_{-\infty}^{\infty} (e^x - 0.7)^+ \frac{1}{\sqrt{2\pi}0.2} e^{-\frac{(x+0.02)^2}{0.08}} dx = \int_0^1 (e^{0.2\Phi^{-1}(z) - 0.02} - 0.7)^+ dz \tag{3.1.9}$$

The other variant, from equation (3.1.8) yields:

$$\begin{aligned} &\int_{-\infty}^{\infty} (e^x - 0.7)^+ \frac{1}{\sqrt{2\pi}0.2} e^{-\frac{(x+0.02)^2}{0.08}} dx \\ &= \int_0^1 (e^{\Phi^{-1}(z)} - 0.7)^+ \frac{1}{0.2} \exp\left(\frac{\Phi^{-1}(z)^2}{2} - \frac{(\Phi^{-1}(z) + 0.02)^2}{0.08} \right) dz \end{aligned} \tag{3.1.10}$$

In Figure 3.1.2, we have plots of both variants.

Remark 3.1.2. It is of course also possible, to transform pairs of coordinate dimensions via the Box-Muller method.

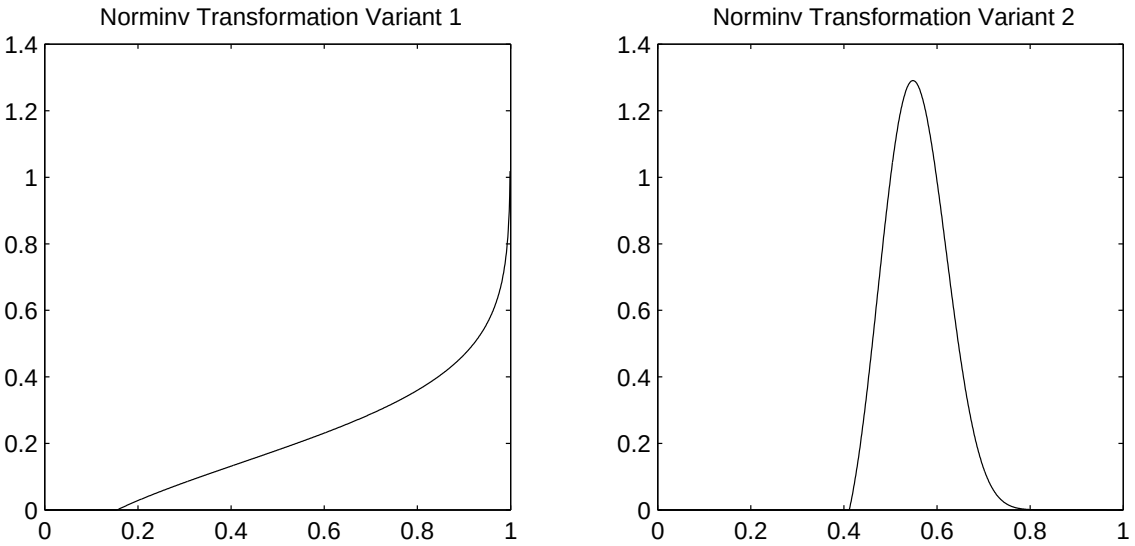

Figure 3.1.2: To the left, we have the integrand from (3.1.9). For $z \to 1$, the integrand tends to infinity, which is rather unwanted behaviour, because then, all sorts of theoretical error estimates are useless. Despite that, QMC- and RQMC-methods can still yield viable results, if one takes care, that the point 1 is excluded from the integration nodes. The function on the right, which is based on (3.1.10), is bounded and converges towards 0 for $z \to 1$, hence there is no truncation error to take care of, and the variation is also finite, s.t. theoretical error estimation is possible. It the one-dimensional case, it is easy to check, that the integrand stays bounded, if and only if $\sigma < 0$.

3.1.5 Logit Transformation

A very useful transformation is the logit function

$$x_i = F_Y^{-1}(z_i) = \lambda_i \operatorname{logit}(z_i) = \lambda_i \log \frac{z_i}{1 - z_i},$$

with a scaling parameter $\lambda_i > 0$. For the differentials, we have then

$$dx_i = \frac{\lambda_i}{z_i(1 - z_i)} dz_i, \quad \text{resp.} \quad dx = \left(\prod_{i=1}^{s} \frac{\lambda_i}{z_i(1 - z_i)} \right) dz.$$

The inverse transformation is the logistic function

$$z_i = \frac{e^{\frac{x_i}{\lambda_i}}}{1 + e^{\frac{x_i}{\lambda_i}}} = \frac{1}{1 + e^{-\frac{x_i}{\lambda_i}}},$$

which is also the CDF of the logistic distribution, hence it's also a special case of importance sampling. To shorten up the notation, define

$$\tau(z) = (\tau_1(z_1), \ldots, \tau_s(z_s)) = \left(\lambda_1 \log \frac{z_1}{1 - z_1}, \ldots, \lambda_s \log \frac{z_s}{1 - z_s} \right), \tag{3.1.11}$$

and the resulting transformed integral from (3.0.2) is then

$$\int_{\mathbb{R}^s} g(x) f_X(x) dx = \int_{[0,1]^s} g(\tau(z)) f_X(\tau(z)) \prod_{i=1}^{s} \frac{\lambda_i}{z_i(1 - z_i)} dz. \tag{3.1.12}$$

We also give here the integrand, which results from applying the logit transformation to representation (3.0.3):

$$\int_{\mathbb{R}^s} g(Ly+\mu)f_Y(y)dy = \int_{[0,1]^s} g(L\tau(z)+\mu)f_Y(\tau(z))\prod_{i=1}^{s}\frac{\lambda_i}{z_i(1-z_i)}dz. \tag{3.1.13}$$

The usefulness of this transformations lies in the fact, that the integrand stays bounded, if the function g does not grow "too fast". Under some weak conditions, the integrand and all it's derivatives are bounded and approach 0 at the boundary of $[0,1]^s$, which we are going to show in the next proposition.

Proposition 3.1.3. *Let $g(x):\mathbb{R}^s\to\mathbb{R}$, $f_X(x)$ be the PDF of an s-dimensional normally distributed R.V. X, and $\tau(z)$ be defined as in equation (3.1.11). We assume g to be piecewise continuously differentiable with a finite number of pieces, and that for every $M\subseteq\mathcal{I}_s$, there exists constants $C_1^M,\dots,C_s^M$, s.t.*

$$\partial^M g(x) \le C_1^M e^{x_1}+\ldots+C_s^M e^{x_s}$$

holds almost everywhere. In this case, we have

$$\lim_{z_i\to 1}\partial^M\left(g(\tau(z))f_X(\tau(z))\prod_{i=1}^{s}\tau_i'(z_i)\right)=0,$$

as well as

$$\lim_{z_i\to 0}\partial^M\left(g(\tau(z))f_X(\tau(z))\prod_{i=1}^{s}\tau_i'(z_i)\right)=0$$

for every $M\subseteq\mathcal{I}_s$ and every $i\in\mathcal{I}_s$, if the other components z_j for $j\neq i$ are fixed with $z_j\in(0,1)$.

Remark 3.1.4. The $\tau_i'(z_i)=\frac{\lambda_i}{z_i(1-z_i)}$, so the function considered here is exactly the integrand in (3.1.12).

Proof. First observe, that we have the product rule $\partial^M(a(x)b(x))=\sum_{u\subseteq M}\partial^M a(x)\partial^{M-u}b(x)$, and the 3-fold product rule

$$\partial^M(a(x)b(x)c(x))=\sum_{u\uplus v\uplus w=M}\partial^u a(x)\partial^v b(x)\partial^w c(x).$$

Then, let's recall the notation from Section 3.1.3: We write the PDF part as $f_X(\tau(z))=K_0e^Q$ with $K_0=\frac{1}{(2\pi)^{s/2}\sqrt{|\det\Sigma|}}$,

$$Q:=-\frac{1}{2}(\tau(z)-\mu)^T\Sigma^{-1}(\tau(z)-\mu),$$

but the $\tau(z)$ is from (3.1.11) this time. The derivatives of Q are now $Q_j=\partial^{\{j\}}Q=-(\Sigma^{-1})_j(\tau(z)-\mu)\tau_j'(z_j)$ and $Q_{jk}=\partial^{\{j,k\}}Q=-(\Sigma^{-1})_{jk}\tau_j'(z_j)\tau_k'(z_k)$, if $j\neq k$. Keeping in mind that the Q's are slightly different to Section 3.1.3, the formula from equation (3.1.5)

$$\partial^u(f_X(\tau(z))=K_0e^Q\sum_{P\in\mathcal{P}(u)}\prod_{p\in P}Q_p$$

still holds ($\mathcal{P}(u)$ denotes the set of all partitions P of u, whose elements p are subsets of u with only 1 or 2 elements). For $M \subseteq \mathcal{I}_s$, let's determine an expression for the derivative of the complete integrand:

$$\begin{aligned}
\partial^M \Big(g(\tau(z)) f_X(\tau(z)) \prod_{j=1}^{s} \tau_j'(z_j) \Big) &= \sum_{u \uplus v \uplus w = M} \partial^u (g(\tau(u)) \partial^v (f_X(\tau(z)) \partial^w \prod_{j=1}^{s} \tau_j'(z_j) \\
&= \sum_{u \uplus v \uplus w = M} (\partial^u g)(\tau(u)) \Big(\prod_{j \in u} \tau_j'(z_j) \Big) \Big(\prod_{j \in w} \tau_j''(z_j) \Big) \Big(\prod_{j \in \underbrace{\mathcal{I}_s - w}_{=u+v}} \tau_j'(z_j) \Big) K_0 e^Q \sum_{P \in \mathcal{P}(v)} \prod_{p \in P} Q_p \\
&= \sum_{u \uplus v \uplus w = M} \sum_{P \in \mathcal{P}(v)} K_0 e^Q (\partial^u g)(\tau(u)) \underbrace{\Big(\prod_{j \in u} (\tau_j'(z_j))^2 \Big)}_{:=(T_u')^2} \underbrace{\Big(\prod_{j \in v} \tau_j'(z_j) \Big)}_{:=T_v'} \underbrace{\Big(\prod_{j \in w} \tau_j''(z_j) \Big)}_{:=T_w''} \prod_{p \in P} Q_p \\
&\leq \sum_{u \uplus v \uplus w = M} \sum_{P \in \mathcal{P}(v)} \underbrace{K_0 e^Q \Big(C_1^u e^{\tau_1(z_1)} + \ldots + C_s^u e^{\tau_s(z_s)} \Big) (T_u')^2 T_v' T_w'' \prod_{p \in P} Q_p}_{:=S_{u,v,w,P}}
\end{aligned}$$

For each summand $S_{u,v,w,P}$ here, we have to proof, that

$$\lim_{z_i \to 1} S_{u,v,w,P} = 0 \quad \text{and} \quad \lim_{z_i \to 0} S_{u,v,w,P} = 0.$$

Also, we have four cases to differentiate:

1. $i \in u$,
2. $i \in w$,
3. $i \in v$ and $i \in p$ with $|p| = 1$, or
4. $i \in v$ and $i \in p$ with $|p| = 2$.

We will carry out the proof only for $\lim_{z_i \to 1}$ in the case $i \in v$ and $i \in p$ with $|p| = 2$, because all the other cases work in a similar fashion, thus we leave them to the reader. All the components z_j, $j \neq i$ are fixed within $(0, 1)$, thus we can assume everything, that does not depend on z_i, to be constant:

- $K_0 e^Q = K_0 \exp\Big(-\frac{\Sigma_{ii}}{2} \tau_i^2(z_i) + K_1 \tau_i(z_i) + K_2 \Big)$,
- $C_1^u e^{\tau_1(z_1)} + \ldots + C_s^u e^{\tau_s(z_s)} = C_i^u e^{\tau_i(z_i)} + K_3 \leq 2 C_i^u e^{\tau_i(z_i)}$ for z_i close enough to 1,
- $(T_u')^2 T_w'' = K_4$, the term has no dependency on z_i at all,
- $T_v' = K_5 \tau_i'(z_i)$, and finally
- $\prod_{p \in P} Q_p = K_6 \tau_i'(z_i)$.

We also want to note, that $\Sigma_{ii} = \mathrm{Var}[X_i] > 0$ is always positive, as well as the transformation's scaling parameter λ_i is always chosen to be positive. Putting together our pieces

(with $\tilde{K}_0 = K_0 e^{K_2} 2C_i^u K_4 K_5 K_6$), we have,

$$\begin{aligned}
\lim_{z_i \to 1} S_{u,v,w,P} &= \lim_{z_i \to 1} \tilde{K}_0 \exp\Big(-\frac{\Sigma_{ii}}{2}\tau_i^2(z_i) + K_1\tau_i(z_i)\Big) e^{\tau_i(z_i)} (\tau_i'(z_i))^2 \\
&= \lim_{z_i \to 1} \tilde{K}_0 \exp\Big(-\frac{\Sigma_{ii}}{2}\tau_i^2(z_i) + (K_1+1)\tau_i(z_i) + 2\log\tau_i'(z_i)\Big) \\
&= \lim_{z_i \to 1} \tilde{K}_0 \exp\Big(-\frac{\Sigma_{ii}\lambda_i}{2}\log^2\frac{z_i}{1-z_i} + \tilde{K}_1 \log\frac{z_i}{1-z_i} + 2\log\frac{\lambda_i}{z_i(1-z_i)}\Big) \\
&= \lim_{z_i \to 1} \tilde{K}_0 \exp\Big(-\frac{\Sigma_{ii}\lambda_i}{2}\log^2\frac{z_i}{1-z_i} + \tilde{K}_1 \log\frac{z_i}{1-z_i} + 2\log\frac{z_i}{1-z_i} + 2\log\frac{\lambda_i}{z_i^2}\Big) \\
&= \lim_{z_i \to 1} \tilde{K}_0 \exp\Bigg(\underbrace{\log\frac{z_i}{1-z_i}}_{\to\infty} \Big(\underbrace{-\frac{\Sigma_{ii}\lambda_i}{2}}_{<0} \underbrace{\log\frac{z_i}{1-z_i}}_{\to\infty} + \tilde{K}_1 + 2\Big) + \underbrace{2\log\frac{\lambda_i}{z_i^2}}_{\to 2\log\lambda_i}\Bigg) \\
&= 0,
\end{aligned}$$

because the exponent tends towards $-\infty$.

□

Example 3.1.1 revisited 2 Once again we start with the same problem, but this time we apply the logit-transformation, like in (3.1.13):

$$\begin{aligned}
&\int_{-\infty}^{\infty} (e^x - K)^+ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&\quad = \int_{-\infty}^{\infty} (e^{\sigma y + \mu} - K)^+ \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&\quad = \int_0^1 (e^{\sigma\tau(z)+\mu} - K)^+ \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau(z)^2}{2}} \frac{1}{z(1-z)} dz, \qquad (3.1.14)
\end{aligned}$$

$$y = \tau(z) = \log\frac{z}{1-z}.$$

We also try representation (3.1.12), but now we use a scaling parameter $\neq 1$, and we set it $\lambda = \sigma = 0.2$, to be equal to the standard deviation.

$$\begin{aligned}
&\int_{-\infty}^{\infty} (e^x - K)^+ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&\quad = \int_0^1 (e^{\tau(z)} - K)^+ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\tau(z)-\mu)^2}{2\sigma^2}} \frac{\sigma}{z(1-z)} dz, \qquad (3.1.15)
\end{aligned}$$

$$x = \tau(z) = \sigma\log\frac{z}{1-z}.$$

As we see in the Figure 3.1.3, both integrands are very similar, and both cover the interval $[0,1]$ better than the other transformations. Depending on the parameters of the original integrand, these variants can yield significantly different integrands, and one has to decide in the particular case which, one is better suited. But both have in common, that there is no truncation error, and the theoretical error estimates are applicable, because the variation is finite.

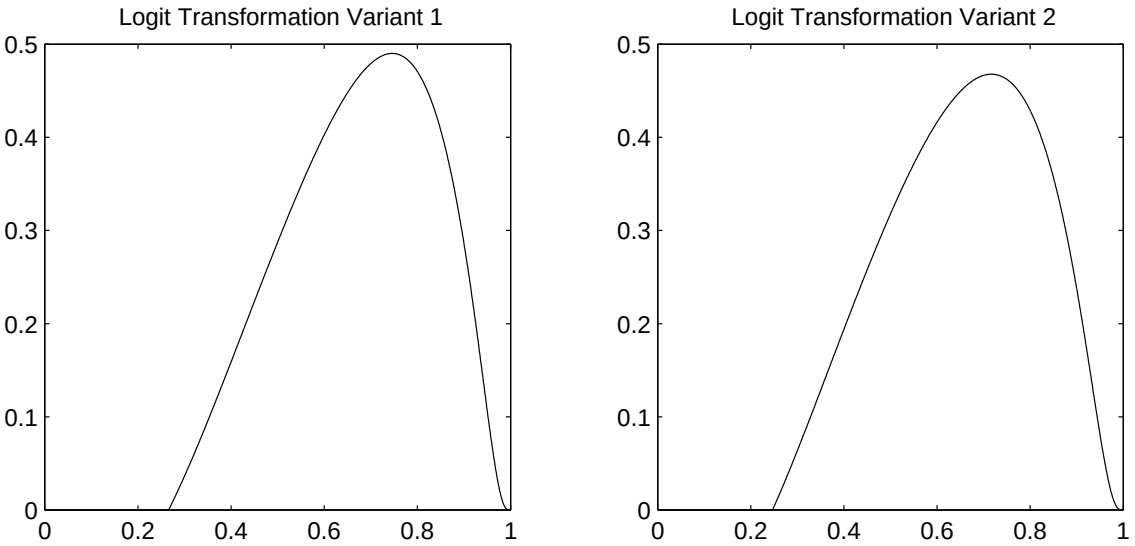

Figure 3.1.3: On the left, we have the plot of (3.1.14), and on the right, the plot of (3.1.15).

The transformations mentioned here are only examples, and one is free to use any transformation one finds suitable. The paper [13] by Kuo et al. deals especially with this topic, and although it is mainly about a different class of integrands, most points also apply in our case. E.g., it could be worthwhile, to apply a linear transformation on the integrand in $\mathbb{R}^s$, s.t., the support is centered around the origin, and the component variables are scaled properly with respect to each other, before mapping it to the unit cube with some distribution.

3.2 Factorization of the Covariance Matrix

For many high dimensional problems, QMC-methods perform better than one would expect, and that is, because these problems have a low effective dimension. E.g., a additive function $f(x) = f_1(x_1) + \ldots + f_s(x_s)$ has effective dimension 1, because it can be written as the sum of s one-dimensional functions, therefore calculating the integral of such a function is just as hard as calculating s one-dimensional integrals. Every s-dimensional function can be decomposed as a sum of 2^s terms ([13])

$$f(z) = \sum_{u \subseteq \mathcal{I}_s} f_u(z_u), \tag{3.2.1}$$

s.t. $\int_0^1 f_u(z_u) dz_j = 0$ for all $j \in u$. Here, f_u is a function depending only on the component variables z_j for $j \in u$ (see also the notation in Section 1.5). We also have ([13])

$$\sigma_f^2 = \sum_{u \subseteq \mathcal{I}_s} \sigma_{f_u}^2,$$

which is why this is known as the ANOVA (analysis of variance) decomposition. Now, if there is a $k < s$, and f can be written as

$$f(z) = \sum_{\substack{u \subseteq \mathcal{I}_s \\ |u| \leq k}} f_u(z_u),$$

and k is the smallest number s.t. this equation still holds, then f is said to have the effective dimension k. In this case, the behaviour of f is completely explained by the interactions of order $\leq k$. We also say, f has an effective low dimension, if σ_f^2 is to a great part explained by the sum of the $\sigma_{f_u}^2$ terms for $|u|$ significantly smaller than s, and while the rest of the terms maybe not strictly 0, they only explain a small part, e.g. 1% of σ_f^2.

The ANOVA decomposition itself can only be estimated and we want to refer to the appendix of [13] for more details. However, we want to discuss a few simple ideas, that can greatly reduce the number of important variables and thus efficiently lower the effective dimension of a problem. These ideas also can be found in Leobacher and Pillichshammer [15] in more detail.

Consider a path dependent option of the form as in Proposition 2.3.5, with s equidistant observation times $0 < \frac{T}{s} < \frac{2T}{s} < \ldots < T$. The covariance matrix of the log returns $X(t_i) = \log \frac{S(t_i)}{S(0)}$ then takes the form

$$\Sigma = \sigma^2 \frac{T}{s} \cdot \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 2 & \ldots & 2 \\ 1 & 2 & 3 & \ldots & 3 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 2 & 3 & \ldots & s \end{pmatrix}.$$

If we were to write our problem in the representation (3.0.3), we have to find a decomposition L fulfilling $LL^T = \Sigma$, regardless of the transformation used to get from $\mathbb{R}^s$ to $[0,1]^s$. The most straight forward method for finding such would be the Cholesky decomposition of Σ, which we can write immediately:

$$L_c = \sigma \sqrt{\frac{T}{s}} \cdot \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{pmatrix}.$$

Note also, that the matrix multiplication $L_c y$, which has to be done for every function evaluation in the QMC-method, is actually equivalent to the cumulative sum of y times a factor, i.e.

$$L_c y = \sigma \sqrt{\frac{T}{s}} \cdot (y_1, y_1 + y_2, \ldots, y_1 + y_2 + \ldots + y_s),$$

and thus can be computed in $O(s)$ time, instead of $O(s^2)$, as would be the naive matrix-vector multiplication. In terms of reducing the effective dimension however, other decompositions might prove to be more efficient.

The PCA decomposition for example, which is actually a decomposition into three factor matrices

$$\Sigma = V \Delta V^T.$$

The middle matrix Δ here is a diagonal matrix containing the eigenvalues of Σ, and V is an orthogonal matrix. Hence we just can take the square root and set $L_p = V\sqrt{\Delta}$, so we have $L_p L_p^T = \Sigma$. We may assume the eigenvalues to be sorted from the largest to the smallest. It is usually the case, that first eigenvalue is much larger than all the others, and that

only the first few eigenvalues are significant at all. Thus, using this factorization, much weight is put on a small number of coordinate dimensions, and little weight is put on the rest, therefore the effective dimension should be reduced significantly. An increased rate of convergence can be seen for this method in the practical section of this work. A drawback for large s is, that the matrix multiplication $L_p y$ takes $O(s^2)$ time, but there is also a way to avoid this. We know, that there is an orthogonal matrix P, such that $L_c P = L_p$, and the multiplication Py corresponds to the discrete sine transform (DST) [15]. The DST can be done in $O(s \log s)$ time, thus, the multiplication $L_c P y$ can be bypassed via a combination of DST and a cumulative sum in at most $O(s \log s)$ time.

As a third option, we just want to mention the Brownian bridge construction, where the first variable determines the Brownian motion at the last timestep W_T, the second variable determines $W_{T/2}$ conditioning on already known value of W_T, and so on. More precisely, the construction works in the following way: Suppose $Y_1, \ldots, Y_s \overset{iid}{\sim} N(0,1)$ and $s = 2^r$ for some r.

- Step 0: Set $W_0 = 0$ and $W_T = \sqrt{T} Y_1$.
- Step 1: Set $W_{T/2} = \frac{1}{2}(W_0 + W_T) + \sqrt{\frac{T}{4}} Y_2$
- Step 2: Set $W_{T/4} = \frac{1}{2}(W_0 + W_{T/2}) + \sqrt{\frac{T}{8T}} Y_3$ and $W_{3T/4} = \frac{1}{2}(W_{T/2} + W_T) + \sqrt{\frac{T}{8}} Y_4$
- Step k: For $i = 1$ to 2^{k-1}: Set $W_{\frac{2i-1}{2^k}T} = \frac{1}{2}(W_{\frac{i-1}{2^{k-1}}T} + W_{\frac{i}{2^{k-1}}T}) + \sqrt{\frac{T}{2^{k+1}}} Y_{(i+2^{k-1})}$
- Stop after step r.

The value of $X(t)$ can then simply be observed from equation (2.3.1). The corresponding decomposition of Σ is given as $L_H = L_c H$, and the orthogonal matrix H corresponds to the inverse Haar transform. Also for this case, the potentially slow matrix multiplication is of course not necessary, because the Brownian bridge construction can be executed in linear time.

Chapter 4

Non-Differentiable Functions

4.1 Finite Variation of the Maximum Function

We want to show here, that the integrand resulting from an option pricing problem with a payoff given as

$$h(S_1, \dots, S_s) = (\max\{S_1, \dots, S_s\} - K)^+$$

has finite variation, if we use the right representation in combination with the logit transformation. First, observe the identity

$$(\max\{S_1, \dots, S_s\} - K)^+ = \max\{(S_1 - K)^+, \dots, (S_s - K)^+\}.$$

Setting $g(x) = h(e^x)$, (we do not strictly need the constant C from equation (3.0.1), because we can incorporate it in the PDF part, if we adapt the expectation of X accordingly) our problem integral then writes as

$$\int_{\mathbb{R}^s} g(x) f_X(x) dx = \int_{[0,1]^s} g(\tau(z)) f_X(\tau(z)) \prod_{i=1}^{s} \frac{1}{z_i(1-z_i)} dz = \int_{[0,1]^s} \tilde{g}(z)\tilde{f}(z) dz,$$

with $\tau(z) = (\tau_1(z_1), \dots \tau_1(z_s))$ being the same transformation for each coordinate, defined as in (3.1.11) with scaling parameter $\lambda = 1$, i.e. $\tau_1(z_i) = \log \frac{z_i}{1-z_i}$. We set

$$\tilde{g}(z) = g(\tau(z)) = \max_{i=1,\dots,s} (e^{\tau_i(z_i)} - K)^+ \quad \text{and} \quad \tilde{f}(z) = f_X(\tau(z)) \prod_{i=1}^{s} \frac{1}{z_i(1-z_i)}, \qquad (4.1.1)$$

to shorten up the notation.

Lemma 4.1.1. *The variation in the sense of Hardy and Krause of the product of the functions $\tilde{g}$ and $\tilde{f}$ given in* (4.1.1) *on $[0,1]^s$ is finite, i.e., $V_{HK}(\tilde{g}\tilde{f}; [0,1]^s) < \infty$.*

Proof. As a first step, we are going to show, that

$$\omega(x) = \max_{i=1,\dots,s} x_i,$$

has finite variation on the interval $[0, b]^s$ for any $b > 0$.

Let $\mathcal{Y}_n^1 = \{y^{(0)}, y^{(1)}, \dots, y^{(n)}\} \subset \mathbb{R}$ with $0 = y^{(0)} < y^{(1)} < \dots < y^{(n)} < b$ be a ladder on $[0, b]$. We construct a s-dimensional ladder $\mathcal{Y}_n = \prod_{i=1}^{s} \mathcal{Y}_n^1$, which consists of the

same points in each dimension. Investigating the variation on ladders of this form is no restriction, because any ladder on $[0,b]^s$ can be refined to a ladder of this form, see also Proposition 1.5.10 and Remark 1.5.11. For a non-empty subset of indices $I \subseteq \mathcal{I}_s$, define

$$B_I := \{x \in [0,b]^s : x_i = x_j \text{ if } i,j \in I, x_i > x_j \text{ if } i \in I, j \notin I\},$$

which is the set, where the maximum is in the i-th component for every $i \in I$ ($\max_{x \in B_I} = x_i$ for all $i \in I$). $B_{\mathcal{I}_s}$ is just the diagonal of the s-dimensional hypercube. Observe, that

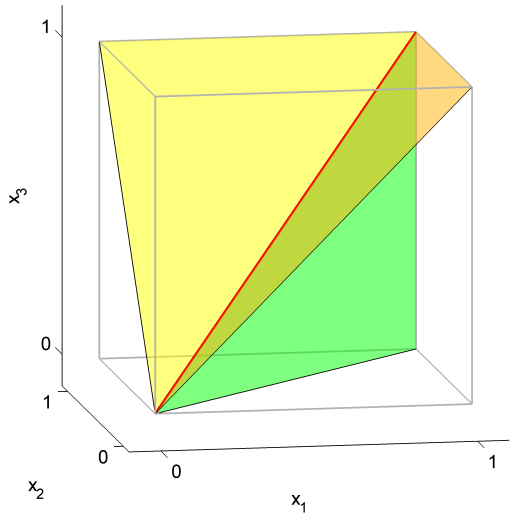

Figure 4.1.1: The B_I's of the unit cube in 3 dimensions:
$B_{\{1,2,3\}}$ is the diagonal in red, $B_{\{1,2\}}$ is the green triangle (without the red diagonal), $B_{\{2,3\}}$ is the yellow triangle, $B_{\{1,3\}}$ is the orange triangle, $B_{\{1\}}$ is the space limited by (but not including) the green and orange triangle, $B_{\{2\}}$ is the space limited by the green and yellow triangle, and $B_{\{3\}}$ is the space limited by the yellow and orange triangle.

the B_I's form a partition of the unit cube, i.e.

$$[0,b]^s = \biguplus_{\emptyset \neq I \subseteq \mathcal{I}_s} B_I.$$

We define $M_I := B_I \cap \mathcal{Y}_n$, such that the M_I's form a partition of $\mathcal{Y}_n$. Considering the variation on the ladder $\mathcal{Y}_n$ as in Definition 1.5.7, we have

$$V_{\mathcal{Y}_n}(\omega) = \sum_{\emptyset \neq I \subseteq \mathcal{I}_s} V_{M_I}(\omega),$$

with

$$V_{M_I}(\omega) = \sum_{y \in M_I} |\Delta(\omega; y, y^+)|.$$

First, let's look at V_{M_I} for $I \neq \mathcal{I}_s$. The function ω does not depend on the variable(s) x_j, where $j \notin I$. It follows from Prop. 1.5.12, that the alternating sum is zero, and therefore $V_{M_I} = 0$. The case $V_{M_{\mathcal{I}_s}}$ remains to be investigated. The Δ-operator writes as

$$\Delta(\omega; y, y_+) = \sum_{v \subseteq \mathcal{I}_s} (-1)^{|v|} \omega(y_v : y^+_{-v}).$$

As all $y \in M_{\mathcal{I}_s}$ lie on the diagonal, all components have the same value, so let's write $y = (y^1, y^1, \dots, y^1)$ only in terms of the first component y^1.

$$\omega(y_v : y^+_{-v}) = \begin{cases} y_1 & \text{if } v = \mathcal{I}_s, \\ y_1^+ & \text{else.} \end{cases}$$

This means, almost all the terms y_1^+ cancel out except one, and the result is

$$|\Delta(\omega; y, y^+)| = y_1^+ - y_1. \tag{4.1.2}$$

Taking the sum yields

$$V_{\mathcal{Y}_n} = V_{M_{\mathcal{I}_s}} = \sum_{y \in M_{\mathcal{I}_s}} |\Delta(\omega; y, y^+)| = \sum_{y \in M_{\mathcal{I}_s}} y_1^+ - y_1 = b - 0,$$

because this is a telescopic sum. The variation (in the sense of Vitali) is now

$$V_{[0,b]^s}(\omega) = \sup_n V_{\mathcal{Y}_n} = b.$$

When calculating the Vitali-variations of lower order, at least one component is fixed at the upper interval boundary b, therefore the maximum is always constant at b and the variation is 0. This yields

$$V_{HK}(\omega; [0,b]^s) = b.$$

As the second step, we are going to prove that $\tilde{g}$ has finite variation on the interval $[0,b]^s$ for any fixed $0 < b < 1$. Remember the definition of $\tilde{g}$,

$$\tilde{g}(z) = g(\tau(z)) = \max_{i=1,\dots,s} (e^{\tau_i(z_i)} - K)^+,$$

but for now, we assume, that $K = 0$, and define $\phi_1(z_i) = e^{\tau_1(z_i)} = \frac{z_i}{1-z_i}$ and $\phi(z) = (\phi_1(z_1), \dots, \phi_1(z_s))$, such that we can write

$$\tilde{g}(z) = \max_{i=1,\dots,s} e^{\tau_i(z_i)} = \omega(\phi(z)).$$

We are now able to directly apply Proposition 1.5.15, and with $\phi_1(0) = 0$, we receive

$$V_{[0,b]^s}(\tilde{g}) = V_{[\phi(0),\phi_1(b)]^s}(\omega) = \frac{b}{1-b},$$

as well as

$$V_{HK}(\tilde{g}; [0,b]^s) = \frac{b}{1-b}.$$

As we see, for $b \to 1$, the term $\frac{b}{1-b}$ tends to ∞. For the last step, we have to take into account the PDF-factor in the integrand. From Proposition 3.1.3 we know, that $\tilde{g}(z)\tilde{f}(z) \to 0$ for $z_i \to 1$ or $z_i \to 0$, which lets us hope, that the variation stays bounded. We have to apply the product rule for the Δ-operator (Lemma 1.5.1) which adds some complexity, but luckily, most of the terms in the sum are 0. In our case, the product rule writes as

$$\Delta(\tilde{g}\tilde{f}; y, y^+) = \sum_{v \subseteq \mathcal{I}_s} (\Delta_v(\tilde{g}; y, y^+))(y^+_{-v}) \cdot (\Delta_{-v}(\tilde{f}; y, y^+))(y_v) \quad \text{for } y \in \mathcal{Y}_n.$$

Let $\emptyset \neq I \subseteq \mathcal{I}_s$. Like before, we calculate the variation separately on each M_I:

$$V_{M_I}(\tilde{g}\tilde{f}) = \sum_{y \in M_I} \left| \sum_{v \subseteq \mathcal{I}_s} (\Delta_v(\tilde{g}; y, y^+))(y^+_{-v}) \cdot (\Delta_{-v}(\tilde{f}; y, y^+))(y_v) \right|. \tag{4.1.3}$$

Choose any fixed $i_0 \in I$. We will show, that

$$|(\Delta_v(\tilde{g}; y, y^+))(y^+_{-v})| = \begin{cases} \phi_1(y^+_{i_0}) - \phi_1(y_{i_0}) & \text{if } v = I, \\ 0 & \text{else,} \end{cases} \tag{4.1.4}$$

because:

Case $v = I$: The maximum can only be taken in components from I, the other components have a smaller value per definition of M_I. The result is $\phi_1(y^+_{i_0}) - \phi_1(y_{i_0})$ completely analogue to (4.1.2).

Case $I \setminus v \neq \emptyset$: There exists at least one component $j \in I \setminus v$ that is not differentiated, s.t.

$$(\Delta_v(\tilde{g}; y, y^+))(y^+_{-v}) = \sum_{u \subseteq v} (-1)^{|u|} \tilde{g}(y_u : y^+_{(v-u)+(-v)}) = \sum_{u \subseteq v} (-1)^{|u|} \underbrace{\phi_1(y^+_j)}_{\text{const. for all } u} = 0$$

Case $v \setminus I \neq \emptyset$: There exists at least one component $j \in v \setminus I$ that $\tilde{g}$ does not depend on, therefore the alternating sum is 0.

Using (4.1.4), equation (4.1.3) simplifies to

$$\begin{aligned} V_{M_I}(\tilde{g}\tilde{f}) &= \sum_{y \in M_I} \left| (\Delta_I(\tilde{g}; y, y^+))(y^+_{-I}) \cdot (\Delta_{-I}(\tilde{f}; y, y^+))(y_I) \right| \\ &= \sum_{y \in M_I} (\phi_1(y^+_{i_0}) - \phi_1(y_{i_0}))|(\Delta_{-I}(\tilde{f}; y, y^+))(y_I)| \end{aligned} \tag{4.1.5}$$

Now let's rewrite M_I in the form

$$\begin{aligned} M_I &= \left\{ (y_1, \ldots, y_s) \in \mathcal{Y}_n | y_j \in \mathcal{Y}^1_n \text{ for } j \notin I, y_i = y_l > \max_{j \notin I} y_j \text{ for } i, l \in I \right\} \\ &= \{ (y_{-I} : t_I) \in \mathcal{Y}_n | y_{-I} \in (\mathcal{Y}_n)_{-I}, t \in \mathcal{Y}^1_n, t > \max_{j \notin I} y_j \}, \end{aligned}$$

and by using the above interpretation of the set M_I, we can write the sum from (4.1.5) as

$$V_{M_I}(\tilde{g}\tilde{f}) = \sum_{y_{-I} \in (\mathcal{Y}_n)_{-I}} \sum_{\substack{t \in \mathcal{Y}^1_n \\ t > \max_{j \notin I} y_j}} (\phi_1(t^+) - \phi_1(t))|(\Delta_{-I}(\tilde{f}; y_{-I}, y^+_{-I}))(t_I)|$$

We use (1.5.3) here to estimate the Δ:

$$\leq \sum_{y_{-I} \in (\mathcal{Y}_n)_{-I}} \sum_{\substack{t \in \mathcal{Y}^1_n \\ t > \max_{j \notin I} y_j}} (\phi_1(t^+) - \phi_1(t)) \max_{z_{-I} \in [y_{-I}, y^+_{-I}]} |\partial^{-I} \tilde{f}(z_{-I} : t_I)| Vol([y_{-I}, y^+_{-I}])$$

$$\leq \sum_{y_{-I} \in (\mathcal{Y}_n)_{-I}} Vol([y_{-I}, y^+_{-I}]) \sum_{t \in \mathcal{Y}^1_n} \max_{z_{-I} \in [y_{-I}, y^+_{-I}]} |\partial^{-I} \tilde{f}(z_{-I} : t_I)| (\phi_1(t^+) - \phi_1(t)).$$

Let

$$\bar{f}^I(t) := \max_{z_{-I}\in[0,1]^{|-I|}} |\partial^{-I}\tilde{f}(z_{-I} : t_I)|,$$

which takes the scalar value t in all components of I, and which is finite, because $\tilde{f}$ is continuously differentiable with bounded derivatives (see Prop. 3.1.3). We continue with the above reasoning:

$$\begin{aligned} V_{M_I}(\tilde{g}\tilde{f}) &\leq \sum_{y_{-I}\in(\mathcal{Y}_n)_{-I}} Vol([y^{-I}, y_+^{-I}]) \sum_{t\in\mathcal{Y}_n^1} \bar{f}^I(t)(\phi_1(t^+) - \phi_1(t)), \\ &\xrightarrow[n\to\infty]{} Vol([0,b]^{|-I|}]) \int_0^b \bar{f}^I(t) d\phi_1(t) \\ &= b^{s-|I|} \int_0^b \bar{f}^I(t)\phi_1'(t)dt. \end{aligned}$$

Because the integrand approaches 0 for $t \to 1$, it is bounded on $[0,1]$ (this claim can be proved with a reasoning analogue to Prop. 3.1.3), and

$$V_{M_I}(\tilde{g}\tilde{f}) \leq \underbrace{\int_0^1 \bar{f}^I(t)\phi_1'(t)dt}_{=:C_I} < \infty$$

for all $b < 1$, hence we can make the limit $b \to 1$ and still stay bounded.

$$V_{[0,1]^s}(\tilde{g}\tilde{f}) = \sup_n V_{\mathcal{Y}_n}(\tilde{g}\tilde{f}) = \sup_n \sum_{\emptyset\neq I\subseteq\mathcal{I}_s} V_{M_I}(\tilde{g}\tilde{f}) \leq \sum_{\emptyset\neq I\subseteq\mathcal{I}_s} C_I$$

Since $\tilde{g}\tilde{f}$ is constant 0 at the boundary of $[0,1]^s$, the Vitali-variations of lower order are again 0, and

$$V_{HK}(\tilde{g}\tilde{f}) \leq \sum_{\emptyset\neq I\subseteq\mathcal{I}_s} C_I < \infty.$$

We assumed here, that $K = 0$, so it still remains, to argue, that this also works for $K > 0$. First of all, remember the definition of the Hardy and Krause-variation as a sum of Vitali-variations:

$$V_{HK}(\tilde{g}\tilde{f}; [0,1]^s) = \sum_{\emptyset\neq u\subseteq\mathcal{I}_s} V_{[0,1]^{|u|}}(\tilde{g}\tilde{f}(x_u; 1_{-u})).$$

In this sum, all the summands $u \subsetneq \mathcal{I}_s$ have at least one component fixed at 1, which means $\tilde{g}\tilde{f}(x_u; 1_{-u})$ is constant 0, and therefore $V_{[0,1]^{|u|}}(\tilde{g}\tilde{f}(x_u; 1_{-u})) = 0$ for $u \subsetneq \mathcal{I}_s$. For the remaining summand with $u = \mathcal{I}_s$, we divide the unit cube into 2^s hyperrectangles R_v: For $v \subseteq \mathcal{I}_s$ we define

$$R_v = \{x \in [0,1]^s : x_j \geq \frac{K}{1+K} \text{ for all } j \in v \text{ and } x_j \leq \frac{K}{1+K} \text{ for all } j \in -v\}.$$

The set $\{R_v | v \subseteq \mathcal{I}_s\}$ forms a split of $[0,1]^s$ in the sense of Definition 1.5.13, hence we have $V_{[0,1]^s}(\tilde{g}\tilde{f}) = \sum_{v\in\mathcal{I}_s} V_{R_v}(\tilde{g}\tilde{f})$. Observe, that $e^{\tau_1(K/(1+K))} = K$ and that $(e^{\tau_i(z_i)} - K)^+ = 0$ if $z_i \leq K/(1+K)$, and thus, for $z \in R_v$ we can write $\tilde{g}$ as

$$\tilde{g}(z) = \max_{i=1,\dots,s} (e^{\tau_i(z_i)} - K)^+ = \underbrace{\max_{i\in v} e^{\tau_i(z_i)}}_{\tilde{g}_1(z)} - K.$$

Now, on $R_{\mathcal{I}_s}$, we know, that $\tilde{g}\tilde{f} = \tilde{g}_1\tilde{f} - K\tilde{f}$. We have already shown in the case $K = 0$, that $V_{R_{\mathcal{I}_s}}(\tilde{g}_1\tilde{f}) \leq V_{[0,1]^s}(\tilde{g}_1\tilde{f}) < \infty$, and we know, that $V_{R_v}(K\tilde{f}) < \infty$ because $\tilde{f}$ is continuously differentiable on $[0,1]^s$ with bounded derivatives. Since the sum and the difference of two functions with bounded (HK-)variation also have bounded (HK-)variation (Proposition 1.5.16), we also have $V_{R_{\mathcal{I}_s}}(\tilde{g}\tilde{f}) < \infty$. For the cases $v \subsetneq \mathcal{I}_s$, we also have $V_{R_v}(\tilde{g}\tilde{f}) < \infty$, which can be shown with very similar arguments.

□

4.2 Approximation with Smooth Functions

For non-differentiable functions, which are bounded, but have infinite variation, we propose an alternative approach: Approximate a given integrand f with a smooth function f_β, i.e., a function that is infinitely times differentiable, s.t. its variation is finite. This way we can apply a QMC-method, and estimate the error via the Koksma-Hlawka inequality (Theorem 1.6.2). For the residual $f - f_\beta$, which should be small by construction, we use ordinary MC- or RQMC-methods and estimate the error with a confidence interval. This could be interpreted as a variance reduction technique. We want to decompose the integral of f as

$$\int_{[0,1]^s} f(x)dx = \underbrace{\int_{[0,1]^s} f_\beta(x)dx}_{\text{QMC with Koksma-Hlawka}} + \underbrace{\int_{[0,1]^s} f(x) - f_\beta(x)dx}_{\text{MC with confidence interval}} . \tag{4.2.1}$$

The parameter β denotes the smoothness, a small β means, that the derivatives of f_β are smaller and thus also the variation, whereas a large β means, that f_β is closer to the original function f, and we have ultimately $\lim_{\beta\to\infty} f_\beta(x) = f(x)$ pointwise for $x \in [0,1]^s$. We will give a few recipes for smoothing functions that are common in option pricing.

4.2.1 The Indicator Function

We start here with the indicator function $I_{(0,\infty)}(x) : \mathbb{R} \to [0,1]$ as a building block for more complex functions. Let

$$SI(x) := \begin{cases} \frac{1}{1+e^{\frac{-4x}{1-x^2}}} & \text{if } |x| < 1, \\ 1 & \text{if } x \geq 1 \text{ and} \\ 0 & \text{if } x \leq -1, \end{cases}$$

and for $\beta > 0$

$$SI_\beta(x) := SI(\beta x).$$

We have the following properties:

- $SI \in C^\infty(\mathbb{R})$ and $SI_\beta \in C^\infty(\mathbb{R})$,
- SI_β and $I_{(0,\infty)}$ differ only on the interval $(-\frac{1}{\beta}, \frac{1}{\beta})$, and
- $SI_\beta \to I_{(0,\infty)}$ pointwise for $\beta \to \infty$.

4.2.2 The Positive-Part Function

Consider the positive part $f(x) = x^+ = \max(x, 0)$. Define

$$m(x) := \begin{cases} (x + 1 - SI(x))SI(x) & \text{if } |x| < 1, \\ x & \text{if } x \geq 1 \text{ and }, \\ 0 & \text{if } x \leq -1, \end{cases}$$

and

$$m_\beta(x) := \frac{1}{\beta} m(\beta x) \quad \text{for } \beta > 0.$$

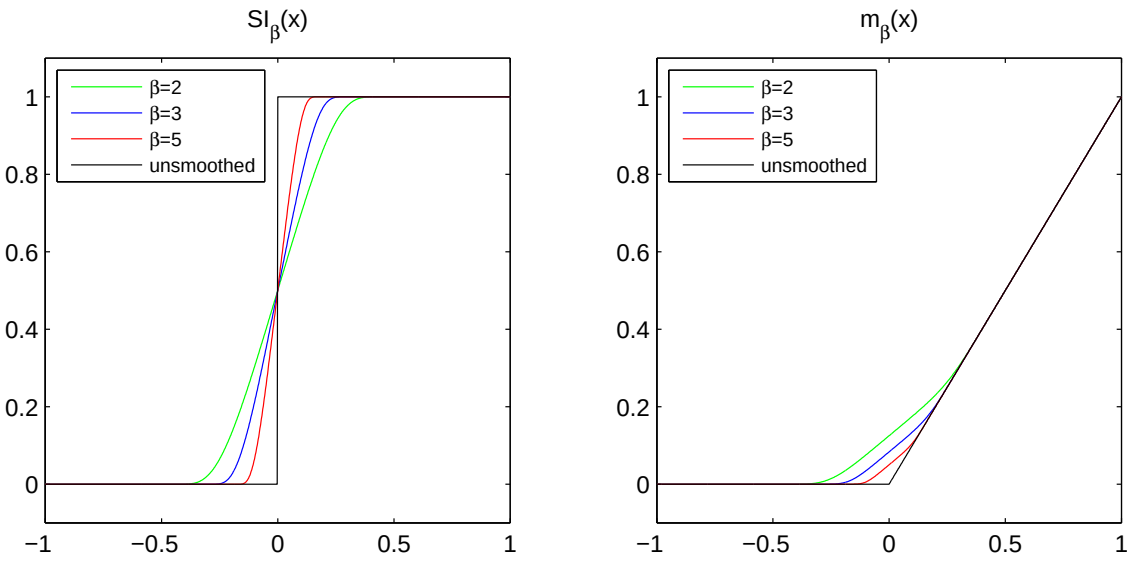

Figure 4.2.1

We have again similar properties:

- $m_\beta \in C^\infty(\mathbb{R})$,
- $m_\beta(x)$ and $(x)^+$ differ only on the interval $(-\frac{1}{\beta}, \frac{1}{\beta})$ with the maximum difference being $\frac{1}{4\beta}$, and
- $m_\beta \to (\cdot)^+$ uniformly for $\beta \to \infty$, because the maximum difference also converges to 0.

Remark 4.2.1. Some more observations: For $|x| \geq \frac{1}{\beta}$ and $k \geq 2$ we have

$$\left| \frac{d^k}{dx^k} m(x) \right| = 0.$$

Let's assume, we have constants C_k, $k \geq 1$ given, such that

$$\left| \frac{d^k}{dx^k} m(x) \right| \leq C_k \quad \text{for } |x| \leq \frac{1}{\beta}. \tag{4.2.2}$$

For k not too large, those constants can be obtained with programs that support symbolic calculations. We have then

$$\left|\frac{d^k}{dx^k}m_\beta(x)\right| = \beta^{k-1}\left|\frac{d^k}{dx^k}m(\beta x)\right| \leq \beta^{k-1}C_k I_{(-1/\beta,1/\beta)}(x), \quad k \geq 2.$$

Remark 4.2.2. By looking at the graphs of the first four derivatives of m, we estimated the constants in (4.2.2) and obtained

$$\begin{aligned} C_1 &= 1, \\ C_2 &= 1.4, \\ C_3 &= 7.6 \quad \text{and} \\ C_4 &= 70. \end{aligned} \tag{4.2.3}$$

4.2.3 Min, Max and Minmax-Functions

We these building blocks, it's not hard to construct smooth versions of the functions $\max(\cdot, k)$, $\min(\cdot, K)$ and the min-max-function $Mm(\cdot, k, K) := \max(\min(\cdot, K), k)$. Observe, that

- $\max(x, k) = x + (k - x)^+$,
- $\min(x, K) = x - (x - K)^+$, and
- $Mm(x, k, K) = x + (k - x)^+ - (x - K)^+$.

For the smooth versions, replace the positive part $(\cdot)^+$ with the previously defined function $m_\beta(x)$, and we have

- $Smax_\beta(x, k) = x + m_\beta(k - x)$,
- $Smin_\beta(x, K) = x - m_\beta(x - K)$, and
- $Smm_\beta(x, k, K) = x + m_\beta(k - x) - m_\beta(x - K)$.

For $Smm_\beta(x, k, K)$ assume, that $K - k \geq \frac{2}{\beta}$, because then, the smoothed areas around k and K don't overlap and we have (from the properties of m_β) for the k-th derivative

$$|Smm_\beta^{(k)}(x)| \leq \beta^{k-1}C_k\Big(I_{(k\pm\frac{1}{\beta})}(x) + I_{(K\pm\frac{1}{\beta})}(x)\Big) \quad \text{for } k \geq 2. \tag{4.2.4}$$

4.2.4 Indicator Function on a Polyhedron

Let a_i for $i = 1, \ldots m$ the rows of A and $b \in \mathbb{R}^d$. The indicator function, $I_{Ax\leq b}(x)$, on the polyhedron $\{x \in \mathbb{R}^d | Ax \leq b\}$ can be approximated by taking the product of indicator functions on single restrictions, s.t.

$$SP_\beta := \prod_{i=1}^{m} SI(\beta(b_i - a_i^T x)). \tag{4.2.5}$$

4.2.5 Indicator Function on a Set with Restrictions

The restrictions do not need to be linear. We can take any set of restrictions of the following form: Let $a_i : \mathbb{R}^d \to \mathbb{R}$ for $i = 1, \ldots, m$ be a set of differentiable functions. The indicator function on the set $B = \{x^d \in \mathbb{R} | a_i(x) \leq b_i \text{ for } i = 1, \ldots, m\}$ can be approximated by

$$SB_\beta := \prod_{i=1}^{m} SI(\beta(b_i - a_i(x))).$$

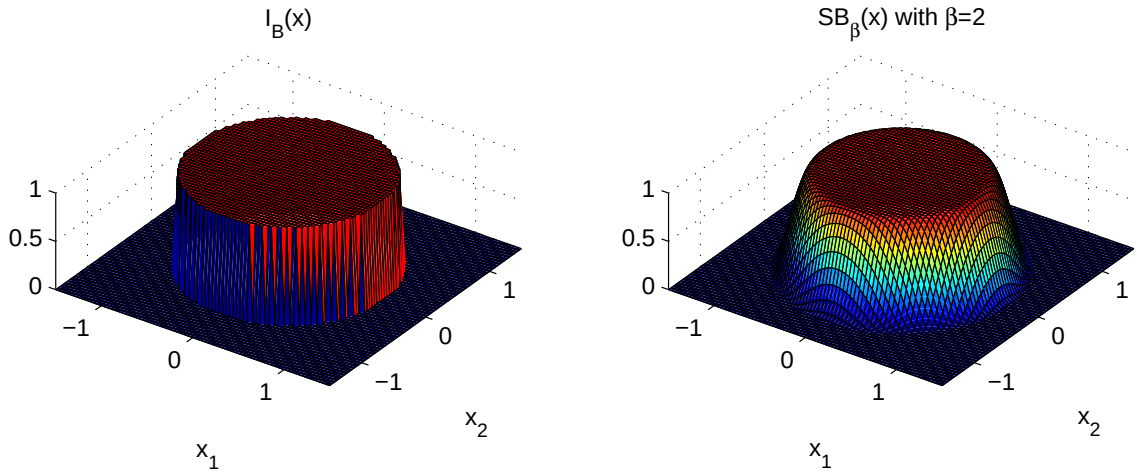

Figure 4.2.2: The indicator function $I_B(x) = I_{\|x \leq 1\|}(x)$ and it's smoothed variant $SB_\beta(x)$

4.2.6 Convolution with a Bump Function

A theoretically easy way to smooth any bounded, function is to convolve it with a bump function, which is an infinitely times differentiable function with compact support. This has the great advantage, that estimating the variation is very easy because of the properties of the differential operator in combination with a convolution. The major drawback of this variant is, that the convolution is again an s-dimensional integral, and one cannot just calculate it numerically, because it still has a variable in it. Despite that, we do present this idea here anyway, because we will construct a fairly simple error estimate for (possibly non-differentiable) Lipschitz-continuous functions, without the need to calculate the smoothed function explicitly.

Consider the function

$$\phi_\beta(x) := \frac{d}{dx} SI_\beta(x) \tag{4.2.6}$$

in the one-dimensional case, and

$$\psi_\beta(x) := \prod_{i=1}^{s} \phi_\beta(x_i) \tag{4.2.7}$$

in the multi-dimensional case. We have $\psi_\beta \in C_0^\infty(\mathbb{R}^s)$, which means it is infinitely times differentiable and has compact support, with $supp(\psi_\beta) = \left[-\frac{1}{\beta}, \frac{1}{\beta}\right]^s = \{x \in \mathbb{R}^s | \|x\|_\infty \leq \frac{1}{\beta}\}$.

Also, since

$$\int_{\mathbb{R}^s} \psi_\beta(x)dx = 1,$$

ψ_β fulfils all the properties of a mollifier. Let $f : R^s \to \mathbb{R}$ be a bounded function with $f \in L^1(\mathbb{R}^s)$ and $supp(f) \subseteq [0,1]^s$. We define f_β as

$$f_\beta(x) := (f * \psi_\beta)(x) = \int_{[0,1]^s} f(y)\psi_\beta(x-y)dy. \tag{4.2.8}$$

From the properties of the convolution we have $(f * \psi_\beta) \in C_0^\infty(\mathbb{R}^s)$ with $supp(f * \psi_\beta) \subseteq \left[-\frac{1}{\beta}, 1+\frac{1}{\beta}\right]^s =: U_\beta$. Observe the identity for the integrals

$$\begin{aligned}
\int_{U_\beta} f_\beta(x)dx &= \int_{\mathbb{R}^s} (f * \psi_\beta)(x)dx \\
&= \int_{\mathbb{R}^s} \int_{[0,1]^s} f(y)\psi_\beta(x-y)dydx \\
&= \int_{[0,1]^s} f(y) \underbrace{\int_{\mathbb{R}^d} \psi_\beta(x-y)dx}_{=1} dy \\
&= \int_{[0,1]^s} f(y)dy.
\end{aligned}$$

We will now try to estimate the Hardy and Krause-variation of f_β on U_β using formula (1.5.4), and again it suffices to consider the Vitali-variation, because the other terms are 0:

$$\begin{aligned}
V_{HK}(f_\beta; U_\beta) &= \sum_{\emptyset \neq u \subseteq \mathcal{I}_s} \int_{[-\frac{1}{\beta}, 1+\frac{1}{\beta}]^{|u|}} \left|\partial^u f_\beta(x_u : (1+1/\beta)_{-u})\right| dx_u \\
&= \int_{[-\frac{1}{\beta}, 1+\frac{1}{\beta}]^s} \left|\partial^{\mathcal{I}_s} f_\beta(x)\right| dx + \sum_{\emptyset \neq u \subsetneq \mathcal{I}_s} \int_{[-\frac{1}{\beta}, 1+\frac{1}{\beta}]^{|u|}} \underbrace{\left|\partial^u f_\beta(x_u : (1+1/\beta)_{-u})\right|}_{=0} dx_u.
\end{aligned}$$

Exploiting the properties of the differentiation of a convolution yields

$$\begin{aligned}
\int_{U_\beta} \left|\partial^{\mathcal{I}_s} f_\beta(x)\right| dx &= \int_{\mathbb{R}^s} |\partial^{\mathcal{I}_s}(f * \psi_\beta)(x)| dx \\
&= \int_{\mathbb{R}^s} \left| \int_{[0,1]^s} f(y)\partial^{\mathcal{I}_s}\psi_\beta(x-y)dy \right| dx \\
&\leq \int_{[0,1]^s} \underbrace{|f(y)|}_{\leq C_f} \underbrace{\int_{\mathbb{R}^s} |\partial^{\mathcal{I}_s}\psi_\beta(x-y)|dx}_{=C_{\psi,\beta},\text{ independent of } y} dy \\
&\leq C_f \cdot C_{\psi,\beta},
\end{aligned}$$

which is very nice, because the only thing we need to know from f is some upper bound C_f. The $C_{\psi,\beta}$ is independent of the integrand f and can be calculated very easily in advance. Since $\psi_\beta(x) = \prod_{i=1}^s \phi_\beta(x_i)$, we have

$$\left|\partial^{\mathcal{I}_s}\psi_\beta(x)\right| = \left|\prod_{i=1}^s \phi'_\beta(x_i)\right| = \prod_{i=1}^s |\phi'_\beta(x_i)|.$$

We know, that $\phi'_\beta(x_i) \geq 0$ if $x_i \leq 0$ and that $\phi'_\beta(-x_i) = -\phi'_\beta(x_i)$, hence we can explicitly calculate the integral

$$\begin{aligned}\int_{\mathbb{R}^s} |\partial^{\mathcal{I}_s}\psi_\beta(x)|dx &= \int_{\mathbb{R}^s} \prod_{i=1}^s |\phi'_\beta(x_i)|dx = \prod_{i=1}^s \int_{-\infty}^{\infty} |\phi'_\beta(x_i)|dx_i \\ &= \prod_{i=1}^s 2\int_{-\infty}^{0} \phi'_\beta(x_i)dx_i = \prod_{i=1}^s 2\phi_\beta(0)dx_i = (2\beta)^s, \end{aligned} \tag{4.2.9}$$

and as the final estimate for the variation in the sense of Hardy and Krause, we receive

$$V_{HK}(f_\beta; U_\beta) \leq (2\beta)^s \sup_{x\in[0,1]^s} |f(x)|$$

The drawback here is, that the definition of f_β in (4.2.8) is an s-dimensional integral in y, and still contains variable x, which is why we cannot just calculate this integral numerically (at least not for a dimension s large enough, to be interesting in our context). Fast and explicit evaluation of the function one wants to integrate is essential for all kinds of numerical integration, which is not possible with this approach.

Under the additional assumption, that f is Lipschitz-continuous w.r.t. the $\|\cdot\|_\infty$-norm, i.e., there exists a constant $L_f > 0$, s.t.

$$|f(x) - f(y)| \leq L_f\|x - y\|_\infty \quad \forall x, y \in \mathbb{R}^s, \tag{4.2.10}$$

we are able to present an error estimate. Observe

$$\begin{aligned}&\left|I_{\mathcal{P}}(f) - \int_{[0,1]^s} f(x)dx\right| \leq \\ &\underbrace{\left|I_{\mathcal{P}}(f) - I_{\mathcal{P}}(f_\beta)\right|}_{\leq \|f-f_\beta\|_\infty Vol(U_\beta)} + \underbrace{\left|I_{\mathcal{P}}(f_\beta) - \int_{U_\beta} f_\beta(x)dx\right|}_{\leq V_{HK}(f_\beta)D^*_N(\mathcal{P})} + \underbrace{\left|\int_{U_\beta} f_\beta(x)dx - \int_{[0,1]^s} f(x)dx\right|}_{=0}\end{aligned}$$

For all of these estimates, we do not need to explicitly calculate the convolution $f_\beta = f * \psi_\beta$. But we know, that $f_\beta(x)$ is a weighted mean of the values of f in an $\frac{1}{\beta}$ neighbourhood of x (w.r.t. the $\|\cdot\|_\infty$-norm), which means

$$|f(x) - f_\beta(x)| \leq \sup_{\|x-y\|_\infty \leq \frac{1}{\beta}} |f(x) - f(y)| \leq L_f \frac{1}{\beta} \quad \text{for all } x \in U_\beta,$$

utilizing the Lipschitz-continuity of f. We are now able to state the following lemma:

Lemma 4.2.3. *Let $f : \mathbb{R}^s \to R$ be bounded and Lipschitz-continuous, s.t. (4.2.10) holds, and let $supp(f) \subseteq [0,1]^s$. Let $\mathcal{P}$ be an N-element point set in $U_\beta = \left[-\frac{1}{\beta}, 1+\frac{1}{\beta}\right]^s$. We have*

$$\left|I_{\mathcal{P}}(f) - \int_{[0,1]^s} f(x)dx\right| \leq L_f \frac{1}{\beta}\Big(1 + \frac{2}{\beta}\Big)^s + (2\beta)^s \sup_{x\in[0,1]^s} |f(x)| D^*_N(\mathcal{P}).$$

Remark 4.2.4. The conditions, that f is (Lipschitz-)continuous and that it has compact support implies, that f has to be 0 at the boundary of $[0,1]$, which is, if we were to apply the transformations discussed in Section 3.1, a restriction on the choice of the transformation.

4.3 A Cliquet-Type Payoff

Let $G_i(x)$, $i = 1, \ldots, s$ be the distribution functions of lognormal distributions, i.e.

$$G_i(x) = \Phi\Big(\frac{\ln x - \mu_i}{\sigma_i}\Big),$$

with $\Phi(x)$ being the standard normal CDF. We have the inverse relationship $G_i^{-1}(y) = e^{\mu_i + \sigma_i \Phi^{-1}(y)}$.

We consider an option with the payoff function

$$h(S(T)) = Mm\Big(\sum_{i=1}^{s} Mm(S_i(T), floor_i, cap_i), floor_g, cap_g\Big) \tag{4.3.1}$$

in the Black-Scholes model. Payoffs of this form appear e.g. in the work of Ralf Korn [11]. We assume the stock price processes to be independent from each other, i.e. $S_i(t)$ and $S_j(t)$ are independent for $i \neq j$. A way for introducing dependency is presented at the end of this section. To shorten notation, we set

$$f_i := floor_i, \quad c_i := cap_i, \quad f_g := floor_g \quad \text{and} \quad c_g := cap_g.$$

We transform our integral with the inverse CDF:

$$I = e^{-RT} \int_{\mathbb{R}^d} h(S(0)e^x) f_X(x) dx = e^{-RT} \int_{[0,1]^s} h(e^{\mu_1 + \sigma_1 \Phi(y_1)}, \ldots, e^{\mu_s + \sigma_s \Phi(y_s)}) dy, \tag{4.3.2}$$

and with the G_i functions defined above (incorporate the $S_i(0)$ in the μ_i), we have

$$\begin{aligned} I &= e^{-RT} \int_{[0,1]^s} h(G_1^{-1}(y_1), \ldots, G_s^{-1}(y_s)) dy \\ &= e^{-RT} \int_{[0,1]^s} Mm\Big(\sum_{i=1}^{s} Mm(G_i^{-1}(y_i), f_i, c_i), f_g, c_g\Big) dy. \end{aligned}$$

Thus our integrand is the function

$$H(y) = Mm\Big(\sum_{i=1}^{s} Mm(G_i^{-1}(y_i), f_i, c_i), f_g, c_g\Big). \tag{4.3.3}$$

We set $u_i = G_i(c_i)$ and $d_i = G_i(f_i)$. Obviously, the minmax function Mm in the given form is not differentiable and it's variation is unbounded (see e.g. Example 1.5.20), therefore we use the Smm_β function instead for the QMC integral. But we just replace the outer Mm function, we do not need to replace the inner Mm function, because in this form, the inner Mm function (within the sum) only produces axis-parallel kinks where the function is not differentiable, and axis-parallel kinks don't destroy the boundedness of the HK-variation (see Remark 1.5.21). Our smoothed integrand is then

$$H_\beta(y) = Smm_\beta\Big(\sum_{i=1}^{s} Mm(G_i^{-1}(y_i), f_i, c_i), f_g, c_g\Big).$$

For the error estimation of the integral, we have the deterministic error bound

$$\left|\int_{[0,1]^s} H(x)dx - I_{\mathcal{P}}(H_\beta)\right| \le \left|\int_{[0,1]^s} H(x)dx - \int_{[0,1]^d} H_\beta(x)dx\right| + \left|\int_{[0,1]^s} H_\beta(x)dx - I_{\mathcal{P}}(H_\beta)\right|$$
$$\le \int_{[0,1]^s} |H(x) - H_\beta(x)|dx + V_{HK}(H_\beta)\cdot D_N^*(\mathcal{P}). \tag{4.3.4}$$

We are going to analyse the Hardy and Krause variation of H_β, but before we look at the Vitali variation. We define $\gamma_i(x_i)$ as

$$\begin{aligned}\gamma_i(x_i) &:= \frac{d}{dx_i} Mm(G_i^{-1}(x_i), f_i, c_i)\\ &= (G_i^{-1})'(x_i)\cdot I_{[f_i,c_i]}((G_i^{-1})'(x_i))\\ &= (G_i^{-1})'(x_i)\cdot I_{[d_i,u_i]}(x_i).\end{aligned} \tag{4.3.5}$$

Calculating the variation via formula (1.5.2) gives us

$$\begin{aligned}V(H_\beta) &= \int_{[0,1]^s} |\partial^{\mathcal{I}_s} H_\beta(x)|dx\\ &= \int_0^1 \cdots \int_0^1 \left| Smm_\beta^{(s)}\Big(\sum_{i=1}^s Mm(G_i^{-1}(x_i), f_i, c_i), f_g, c_g\Big)\prod_{i=1}^s \gamma_i(x_i)\right| dx_1\cdots dx_n\end{aligned}$$

Since $\gamma_i(x_i)$ is 0 outside of $[d_i, u_i]$, we restrict the integration limits, and the local minmax-functions can be left out.

$$= \int_{d_s}^{u_s} \cdots \int_{d_1}^{u_1} \left| Smm_\beta^{(s)}\Big(\sum G_i^{-1}(x_i), f_g, c_g\Big)\right| \prod_{i=1}^s \gamma_i(x_i)dx_1\cdots dx_s$$

For the next step we use estimate (4.2.4):

$$\le \beta^{s-1}C_s \int_{d_s}^{u_s} \cdots \int_{d_1}^{u_1} \Big(I_{(f_g\pm\frac{1}{\beta})}(\sum\ldots) + I_{(c_g\pm\frac{1}{\beta})}(\sum\ldots)\Big)\prod_{i=1}^s \gamma_i(x_i)dx_1\cdots dx_s$$

$$\le \beta^{s-1}C_s \int_{d_s}^{u_s} \cdots \int_{d_2}^{u_2} \left(\int_{G_1(f_g-\sum_{i=2}^s G_i^{-1}(x_i)-1/\beta)}^{G_1(f_g-\sum_{i=2}^s G_i^{-1}(x_i)+1/\beta)} \gamma_1(x_1)dx_1 \right.$$
$$\left. + \int_{G_1(c_g-\sum_{i=2}^s G_i^{-1}(x_i)-1/\beta)}^{G_1(c_g-\sum_{i=2}^s G_i^{-1}(x_i)+1/\beta)} \gamma_1(x_1)dx_1 \right) \prod_{i=2}^s \gamma_i(x_i)dx_2\cdots dx_s$$

The integration limits explain as following: For the indicator function $I_{(f_g\pm\frac{1}{\beta})}$ to be not equal 0, we need to have

$$\sum_{i=1}^s G_i^{-1}(x_i) \in \Big[f_g \pm \frac{1}{\beta}\Big],$$

and this is equivalent to

$$G_1^{-1}(x_1) \in \Big[f_g - \sum_{i=2}^s G_i^{-1}(x_i) \pm \frac{1}{\beta}\Big],$$

and thus also to

$$x_i \in \Big[G_1\Big(f_g - \sum_{i=2}^s G_i^{-1}(x_i) - \frac{1}{\beta}\Big), G_1\Big(f_g - \sum_{i=2}^s G_i^{-1}(x_i) + \frac{1}{\beta}\Big)\Big].$$

Estimating the integral yields

$$\begin{aligned}
&\int_{G_1(f_g-\sum_{i=2}^s G_i^{-1}(x_i)-1/\beta)}^{G_1(f_g-\sum_{i=2}^s G_i^{-1}(x_i)+1/\beta)} \gamma_1(x_1)dx_1 \le \int_{G_1(f_g-\sum_{i=2}^s G_i^{-1}(x_i)-1/\beta)}^{G_1(f_g-\sum_{i=2}^s G_i^{-1}(x_i)+1/\beta)} (G_1^{-1})'(x_i)dx_1 \\
&= G_1^{-1}\Big(G_1\Big(f_g - \sum_{i=2}^s G_i^{-1}(x_i) + \frac{1}{\beta}\Big)\Big) - G_i^{-1}\Big(G_1\Big(f_g - \sum_{i=2}^s G_i^{-1}(x_i) - \frac{1}{\beta}\Big)\Big) \\
&= \Big(f_g - \sum_{i=2}^s G_i^{-1}(x_i) + \frac{1}{\beta}\Big) - \Big(f_g - \sum_{i=2}^s G_i^{-1}(x_i) - \frac{1}{\beta}\Big) \\
&= \frac{2}{\beta}.
\end{aligned}$$

The same relations hold for the global cap c_g. Continuing with $V(H_\beta)$ we have now

$$V(H_\beta) \le \beta^{s-1} C_s \int_{d_s}^{u_s} \cdots \int_{d_2}^{u_2} \Big[\frac{2}{\beta} + \frac{2}{\beta}\Big] \prod_{i=2}^s \gamma_i(x_i)dx_1 \cdots dx_s \tag{4.3.6}$$

$$= 4\beta^{s-2} C_s \prod_{i=2}^s \Big(\int_{d_i}^{u_i} \gamma_i(x_i)dx_i \Big) \tag{4.3.7}$$

$$= 4\beta^{s-2} C_s \prod_{i=2}^s \Big(G_i^{-1}(u_i) - G_i^{-1}(d_i) \Big) \tag{4.3.8}$$

$$= 4\beta^{s-2} C_s \prod_{i=2}^s (c_i - f_i) \tag{4.3.9}$$

The choice of the index 1 for the innermost integral, and therefore the index missing in this final formula in the product, was arbitrary. It makes sense to choose the index i which maximizes $c_i - f_i$, in order to make the product as small as possible by leaving out the biggest factor. The Vitali-variations $V_{[0_u,1_u]}(H_\beta(x_u; 1_{-u}))$ of lower order ($u \subsetneq \mathcal{I}_s$, $|u| \ge 2$), which we also need to be calculated as part of the variation in the sense of Hardy and Krause, can be estimated with the same formula by using the following observation: For the components $j \notin u$, the x_j is fixed at $x_j = 1$ and we have $Mm(G_j^{-1}(x_j), f_j, c_j) = c_j$, and thus

$$H_\beta(x_u; 1_{-u}) = Smm_\beta\Big(\sum_{j \notin u} c_j + \sum_{i \in u} Mm(G_i^{-1}(y_i), f_i, c_i), f_g, c_g \Big).$$

If one shifts a function by an additive constant $a = \sum_{j \notin u} c_j$, the variation does not change, hence we are allowed to switch to the function

$$H_\beta^u(x_u) := H_\beta(x_u; 1_{-u}) - a = Smm_\beta\Big(\sum_{i \in u} Mm(G_i^{-1}(y_i), f_i, c_i), f_g - a, c_g - a \Big).$$

This new function $H_\beta^u(x_u)$ is of the same type as $H_\beta(x)$, only with an adapted global cap and floor, and only depending on $|u|$ variables instead of s. So the formula (4.3.9) can still be applied, but s has to be replaced with $|u|$, and the product has to go over all indices $i \in u \setminus u_0$, where $u_0 \in u$ can be arbitrarily chosen. The Vitali-variations of order one, where $u = \{j\}$ has just one element, can be immediately estimated by

$$V_{[a_j,b_j]}(H_\beta(x_{\{j\}}; b_{-\{j\}}) \le c_j - f_j.$$

This leads to the following lemma:

Lemma 4.3.1. *For $u \subseteq \mathcal{I}_s$, $|u| \geq 2$, we define $\kappa(u) := \arg\max_{i \in u}(c_i - f_i)$. Let $\beta > 0$ and C_k be the constants from* (4.2.2)*. We have*

$$V_{HK}(H_\beta) \leq \sum_{i=1}^{s}(c_i - f_i) + \sum_{\substack{u \subseteq \mathcal{I}_s \\ |u| \geq 2}} 4\beta^{|u|-2} C_{|u|} \prod_{i \in u \setminus \kappa(u)} (c_i - f_i).$$

As a second step, we want to estimate the smoothing error, the integral of the difference between the original and the smoothed function:

$$SE_\beta = \int_{[0,1]^s} |H(x) - H_\beta(x)| dx = \int_B |H(x) - H_\beta(x)| \leq Vol(B) \cdot \frac{1}{4\beta}, \tag{4.3.10}$$

with

$$B = \{x \in [0,1]^s | H(x) \neq H_\beta(x)\}.$$

Because $Vol(B) \leq 1$, an easy estimate would be $SE_\beta \leq \frac{1}{4\beta}$. But our hope is, that $Vol(B) = O(\beta^{-1})$, because then we would have $SE_\beta = O(\beta^{-2})$. We define the set E as

$$E = \{x \in [0,1]^s | \forall i = 1, \ldots, s : x_i \geq u_i \vee x_i \leq d_i\},$$

which denotes the corners of the hypercube, where in each coordinate the function $G_i^{-1}(x_i)$, is either larger or equal than c_i or less or equal than f_i. With

$$R = [0,1]^s \setminus E$$

we denote the rest. We investigate $B \cap R$ and $B \cap E$ separately.

First let $x \in B \cap R$: There exists an $j \in \{1, \ldots, s\}$, s.t. $d_j \leq x_j \leq u_j$, which is equivalent to $Mm(G_j^{-1}(x_j), f_j, c_j) = G_j^{-1}(x_i)$. The functions $H(x)$ and $H_\beta(x)$ differ only if

$$\left| \sum_{i=1}^{s} Mm(G_i^{-1}(x_i), f_i, c_i) - f_g \right| \leq \frac{1}{\beta} \text{ or } \left| \sum_{i=1}^{s} Mm(G_i^{-1}(x_i), f_i, c_i) - c_g \right| \leq \frac{1}{\beta}.$$

We continue with the left formula with f_g, because the other case is completely analogue. The condition

$$\left| G_j^{-1}(x_j) + \sum_{\substack{i=1 \\ i \neq j}}^{s} Mm(G_i^{-1}(x_i), f_i, c_i) - f_g \right| \leq \frac{1}{\beta}$$

is equivalent to the left case above, and it is also equivalent to

$$G_j\Big(f_g - \sum_{\substack{i=1 \\ i \neq j}}^{s} Mm(G_i^{-1}(x_i), f_i, c_i) - \frac{1}{\beta}\Big) \leq x_j \leq G_j\Big(f_g - \sum_{\substack{i=1 \\ i \neq j}}^{s} Mm(G_i^{-1}(x_i), f_i, c_i) + \frac{1}{\beta}\Big).$$

Because G_j is Lipschitz-continuous, we can estimate the size of this interval for x_j:

$$G_j\Big(f_g - \sum \ldots + \frac{1}{\beta}\Big) - G_j\Big(f_g - \sum \ldots - \frac{1}{\beta}\Big) \leq L_{G_j} \cdot \frac{2}{\beta}.$$

Since the same calculation holds for the interval around c_g, the variable x_j now has to be in one of two intervals of the size $L_{G_j} \frac{2}{\beta}$. For the volume of $B \cap R$, we are not allowed to

fix a component x_j, hence we have to take the maximum of the Lipschitz-constants for our estimate to receive:

$$Vol(B \cap R) \leq \max_{j=1,\dots,s} L_{G_j} \frac{4}{\beta}.$$

For the set $B \cap E$ first observe, that $H(x)$ as well as $H_\beta(x)$ is constant there, because the local caps and floors are reached. Now if in a certain corner, the values of $\sum_{i=1}^{s} Mm(G_i^{-1}(y_i), f_i, c_i)$ is within a $1/\beta$-neighbourhood of f_g or c_g, it is still in domain of $Smm_\beta(\cdot, f_g, c_g)$, where it differs from $Mm(\cdot, f_g, c_g)$. Then, as $H(x)$ and $H_\beta(x)$ are constant in this corner, they differ on the whole domain of the corner. This is a case that hopefully doesn't happen in many of the 2^s corner domains of the unit-cube, and then, maybe the volume of the corner is not that big. This has to be investigated in the concrete case of the function. For the estimate, we write a sum over all of the corners, and hope, that most of the summands, or even all, are actually 0:

$$C_c = Vol(B \cap E) = \sum_{v \subseteq \mathcal{I}_s} \left(I_{(\sum_{i \in v} c_i + \sum_{i \in -v} f_i) \in (f_g \pm 1/\beta) \cup (c_g \pm 1/\beta)} \prod_{i \in v} (1 - u_i) \prod_{i \in -v} d_i \right).$$

So here v are the components of the corner, where the $x_i \geq u_i$, and $-v$ are the components with $x_i \leq d_i$. The indicator indicates, if the function value in the corner is within a $\frac{1}{\beta}$ neighbourhood of f_g or c_g, and the products denote the volume of the concrete corner. Now we have

$$Vol(B) \leq \max_{j=1,\dots,s} L_{G_j} \frac{4}{\beta} + C_c, \tag{4.3.11}$$

and in most cases, C_c should be small.

We assumed independence between the $S_i(T), i = 1, \dots, s$, such that for the inverse transformation applied in (4.3.2) we only had to transform each coordinate "on its own", without any dependence on the other coordinate variables. If one wants to introduce dependency, the variant of equation (3.1.8) would be suitable, because it doesn't rotate the argument of the payoff function factor, thus the HK-variation stays bounded.

Chapter 5

Practical Examples

5.1 A Comparison between MC, QMC and RQMC

We consider the function $f : [0,1]^3 \to \mathbb{R}$, which is defined as

$$f(x) = \psi_2(x - 1/2 \cdot \mathbb{1}) = \prod_{i=1}^{3} \phi_2(x_i - 1/2),$$

where the functions ψ_2 and ϕ_2 are defined as in (4.2.7) respectively (4.2.6) with smoothing parameter $\beta = 2$. Using the definition of f and (4.2.9), one can easily observe, that

$$\int_{[0,1]^3} f(x)dx = 1 \quad \text{and} \quad V_{HK}(f) = (2\beta)^3 = 64.$$

The main reasons for the usage of this function here is, that we know the true value of the integral, and the true value of the variation, s.t. we can compare the error estimates and the actual error with each other. The graphic in Figure 5.1.1 depicts a plot of the integrand in dependency of the first 2 components, and component x_3 is fixed at 0.5.

In Figure 5.1.2 we have a comparison of the different errors, the estimates on the left, and the actual error on the right. A Sobol-sequence was used for the (R)QMC-calculation. For the QMC error estimate, the Koksma-Hlawka inequality (Theorem 1.6.2) was used, where the discrepancy term was estimated with Proposition 1.4.13. The t-parameter in the discrepancy estimate was assumed to be $t = 0$ [21]. For the RQMC-calculation, the affine linear scrambling from Section 1.7.3 was used as a randomization technique, with $q = 64$ repeats. The error estimates for MC and RQMC are 99%-confidence intervals, where the variances were calculated with (1.1.4) and (1.7.3).

First observe, that for the ordinary Monte Carlo method (green), the error estimate and the actual error coincide to a great degree, and both behave similar to $N^{-\frac{1}{2}}$. Secondly, the QMC error estimate (blue) starts with a value, which is higher than the estimates of the other methods by a factor of 100(!). Although it decreases with a steeper curve than MC, the order of convergence significantly less than N^{-1} (which is also due to the rough discrepancy estimate), and it does by far not catch up with the error estimate of the MC method for $N = 10^6$. This represents a big gap to the actual error of the QMC-method, which clearly decreases at a rate of at least N^{-1}. Finally, the RQMC-error and its estimate behave roughly the same way, they start with values similar to the MC-method, but decrease considerably faster with an order of convergence of $O(N^{-1})$, which is clearly

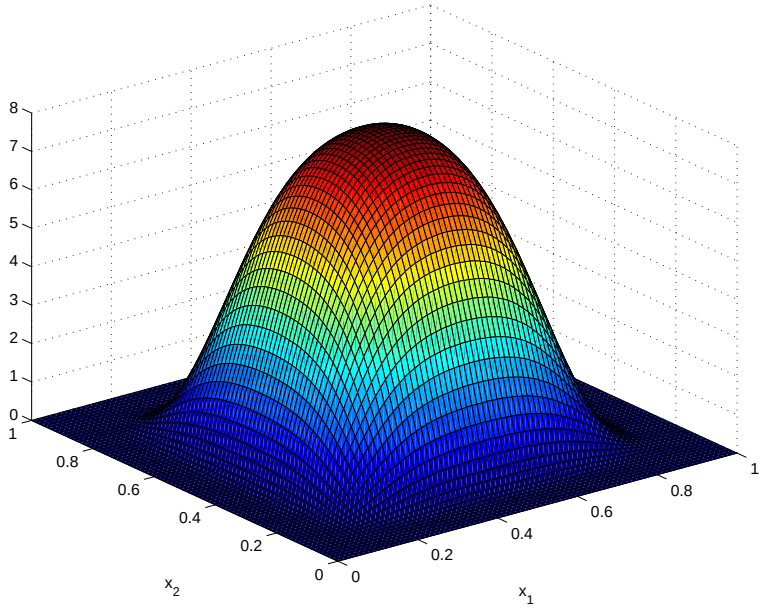

Figure 5.1.1: The integrand with $x_3 = 0.5$ fixed

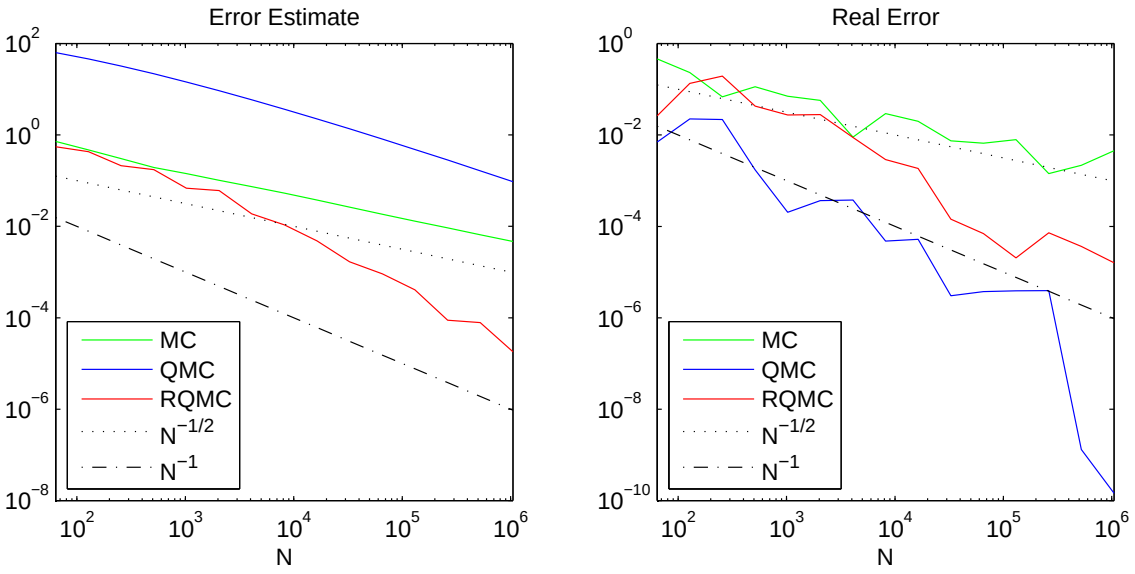

Figure 5.1.2: Comparison of the different errors in dependency of N

visible in the plots. One can conclude from this, that the suitability of the Koksma-Hlawka inequality for practical purposes is questionable, because even in this toy example in only 3 dimensions, there is a huge gap between the estimate and the real error. As the variation in this example grows exponentially with the number of dimensions, the situation for higher dimensions is even worse. However, this should not diminish the applicability of QMC, which yields excellent results. It seems, that if one wants to have a realistic error estimate, one has to accept a tradeoff between the good results of the QMC-method for the results of the RQMC-method, where we have the same order of convergence, but the actual error

is higher.

The last plot in Figure 5.1.3 show the actual estimates of the integral in combination with the confidence interval. As you can see, the width of the confidence interval decreases much faster for RQMC than for MC. We didn't include the QMC-error estimate in this plot, because it's larger by orders of magnitudes.

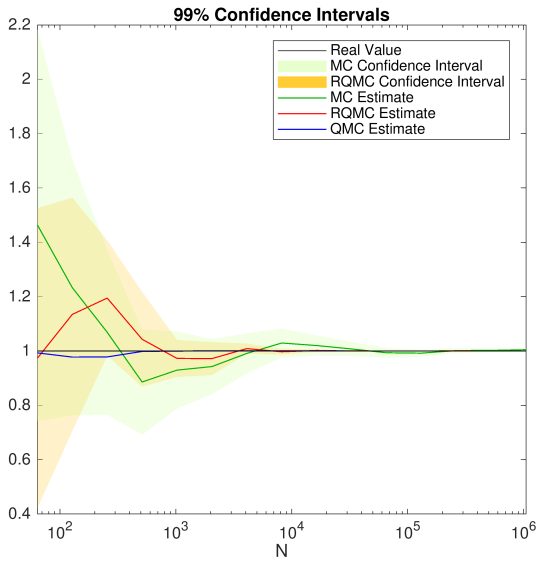

Figure 5.1.3: Confidence Intervals in dependency of N

5.2 A Smoothed Indicator Function

We consider a fictive binary option on two assets with a payoff that has infinite variation. The time horizon is fixed at $T = 1$, the both assets have the starting value $S_{1,2}(0) = 2$. The payoff should be 1, if $S_1(T) \leq 3$, $S_2(T) \leq 3$, and $S_1(T) + S_2(T) \leq 4$, and 0 otherwise. With $S = (S_1(T), S_2(T))^T$, we can write this payoff as

$$h(S) = I_{AS \leq b}(S), \quad \text{with} \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 3 \\ 3 \\ 4 \end{pmatrix}.$$

A plot of the payoff is given in Figure 5.2.1. As you see, the condition $S_1(T) + S_2(T) \leq 4$ introduces an non-continuous and non-differentiable edge, which is also not axis-parallel, and therefore causes the variation to be infinite. We assume S to follow the Black-Scholes model, with $R = 0$, $\sigma = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and for the integrand, we choose the logit-transform with the representation as it is in (3.1.12). Thus our integrand is

$$Q(z) = g(\tau(z)) f_X(\tau(z)) \prod_{i=1}^{s} \frac{\lambda_i}{z_i(1 - z_i)} = I_{AS \leq b}(e^{\tau(z)}) f_X(\tau(z)) \prod_{i=1}^{s} \frac{\lambda_i}{z_i(1 - z_i)},$$

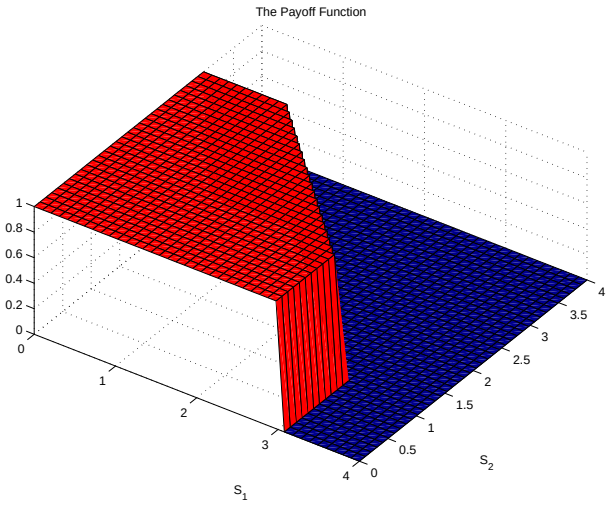

Figure 5.2.1: The Payoff Function

with f_X being the PDF of a multivariate normal distribution $X \sim N(\mu, \Sigma)$, where $\Sigma = \sigma\sigma^T = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and

$$\mu = \begin{pmatrix} \log(S_1(0)) - \Sigma_{11}/2 \\ \log(S_2(0)) - \Sigma_{22}/2 \end{pmatrix} = \begin{pmatrix} 0.1931 \\ 0.1931 \end{pmatrix},$$

thus, we incorporated the starting value of the assets in the PDF part of the integrand. For the scaling parameters of the logit-transform, we set $\lambda_1 = \lambda_2 = 0.65$, which is heuristically chosen s.t. the integrand fills out the whole integration domain reasonably well and the maximum value of the integrand is small (see also [13]). Now since we know, that the integrand has infinite variation because of the term $I_{AS\leq b}(\cdot)$, we want to do a decomposition in a smoothed part and a small rest, s.t. we can execute the idea presented in (4.2.1), namely to apply QMC to the smoothed part and MC to the rest. For that purpose, we replace $I_{AS\leq b}$ with the smooth approximation SP_β, as given in (4.2.5), and we receive our smoothed integrand

$$Q_\beta(z) = SP_\beta(e^{\tau(z)}) f_X(\tau(z)) \prod_{i=1}^{s} \frac{\lambda_i}{z_i(1-z_i)},$$

which is depicted in Figure 5.2.2 for different values of β. The smoothed integrands should now have a finite HK-variation, and it would be nice, to have an estimate of the variation. We recall the definition of the HK-variation,

$$\begin{aligned} V_{HK}(Q_\beta) &= \sum_{\emptyset \neq u \subseteq \mathcal{I}_s} V_{[0_u, 1_u]}(Q_\beta(x_u; 1_{-u})) \\ &= V_{[0,1]}(\underbrace{Q_\beta(x_{\{1\}}; 1_{\{2\}})}_{=0}) + V_{[0,1]}(\underbrace{Q_\beta(x_{\{2\}}; 1_{\{1\}})}_{=0}) \\ &\quad + V_{[0,1]^2}(Q_\beta), \end{aligned}$$

thus it is sufficient to consider the Vitali-variation of Q_β. We estimate $V_{[0,1]^2}(Q_\beta)$ numerically with $V_{\mathcal{Y}_n}(Q_\beta)$ over the ladder

$$\mathcal{Y}_n = \left\{\frac{0}{n}, \frac{1}{n}, \cdots \frac{n-1}{n}\right\} \times \left\{\frac{0}{n}, \frac{1}{n}, \cdots \frac{n-1}{n}\right\},$$

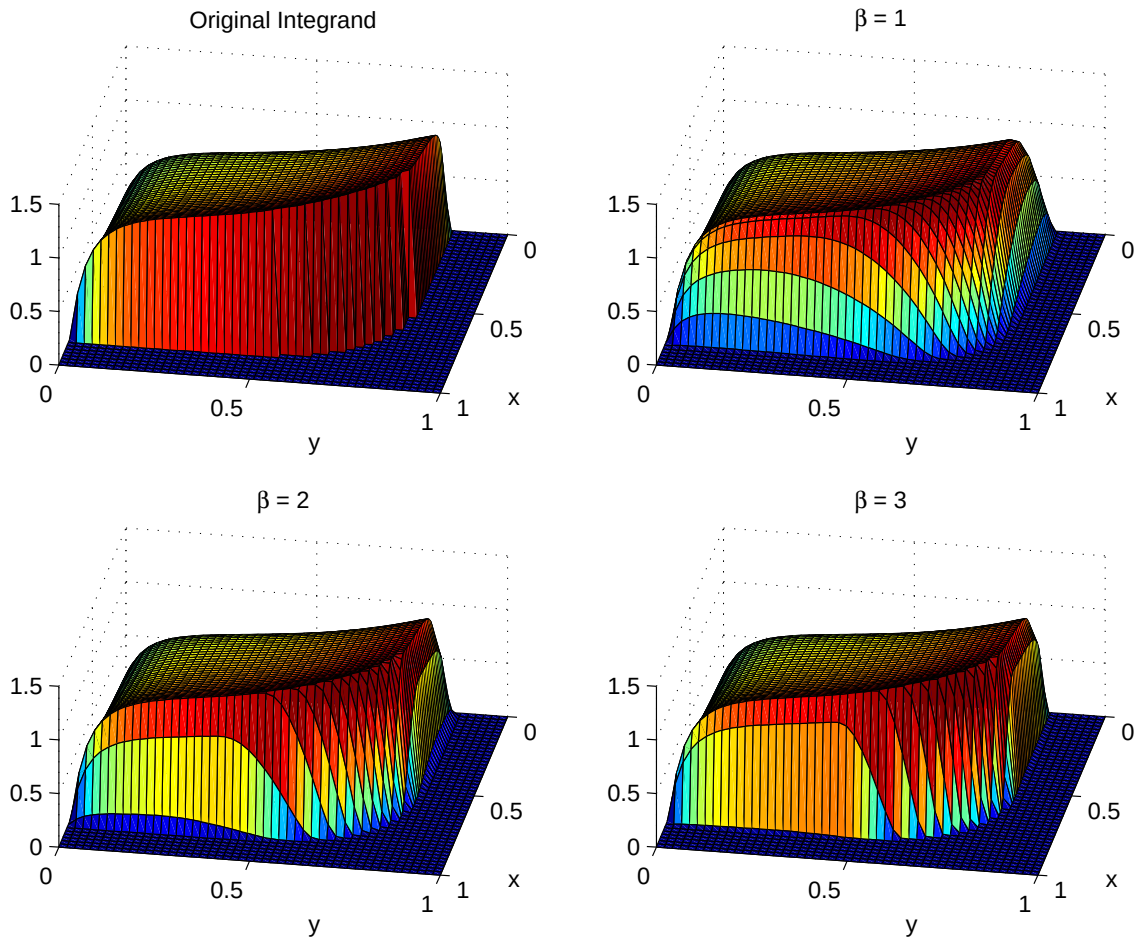

Figure 5.2.2: The original integrand and its smoothed version for different smoothing parameters

and for growing $n \to \infty$, one can clearly see, that $V_{y_n}(Q_\beta)$ converges towards some value. At least for $\beta \in \{1, 2, 3\}$, it suffices use n up to 4096, as one can see in Table 5.2.1. So to be safe, we assume

$$\begin{aligned} V_{HK}(Q_1) &\leq 10.3, \\ V_{HK}(Q_2) &\leq 16.5 \text{ and} \\ V_{HK}(Q_3) &\leq 22.3. \end{aligned}$$

Let $\mathcal{P} = (x_0, \cdots, x_{N-1})$ the first N points of a 2-dimensional Sobol-sequence, and $y_0, \cdots, y_{N-1} \overset{iid}{\sim} U([0,1]^s)$. We now decompose our integral, and approximate one part with QMC and the other with MC, s.t. we have

$$\begin{aligned} \int_{[0,1]^s} Q(z)dz &= \int_{[0,1]^s} Q_\beta(z)dz + \int_{[0,1]^s} Q(z) - Q_\beta(z)dz \\ &\approx \frac{1}{N}\sum_{n=0}^{N-1} Q_\beta(x_n) + \frac{1}{N}\sum_{n=0}^{N-1}(Q(y_n) - Q_\beta(y_n)). \end{aligned}$$

n \beta	1	2	3
16	9.085173984	12.46328455	13.33008065
32	9.952346893	15.04226792	18.59282626
64	10.16048973	16.0081651	21.05236183
128	10.21684768	16.3347254	21.90322294
256	10.23135494	16.4329586	22.15252346
512	10.23592695	16.45643278	22.21395454
1024	10.23682124	16.46359306	22.23089217
2048	10.237049	16.46508687	22.23478838
4096	10.23710588	16.46546155	22.23585557

Table 5.2.1: The convergence of $V_{\mathcal{Y}_n}(Q_\beta)$

With that, our integration error E is also decomposed into $E = E_{QMC} + E_{MC}$, with

$$E_{QMC} = \int_{[0,1]^s} Q_\beta(z)dz - \frac{1}{N}\sum_{n=0}^{N-1} Q_\beta(x_n)$$

and

$$E_{MC} = \int_{[0,1]^s} Q(z) - Q_\beta(z)dz - \frac{1}{N}\sum_{n=0}^{N-1}(Q(x_n) - Q_\beta(x_n)).$$

With Theorem 1.6.2, we have for the QMC-error the upper bound given as

$$|E_{QMC}| = |e(Q_\beta, \mathcal{P})| \leq V_{HK}(Q_\beta) \cdot D_N^*(\mathcal{P}),$$

and for the MC-error, by using Equation (1.1.5), we receive

$$E_{MC} \overset{\text{approx.}}{\sim} N\Big(0, \frac{s^2_{Q-Q_\beta}}{N}\Big).$$

From the previous equation, we deduce, that

$$|E_{MC}| \leq z_{\delta/2} \cdot \frac{s_{Q-Q_\beta}}{\sqrt{N}}$$

is fulfilled with the probability $1 - \delta$. The discrepancy of the Sobol-sequence used for the QMC-calculation can again be estimated via Proposition 1.4.13. We have now all the data we need for a comparison of the results. If we look at the results for $N = 2^{20} = 1048576$, one can see, that the combined error of the decomposition method is still larger than that of the pure MC. For the choice of $\beta's$ used here, the QMC error fraction is the dominating part of the total error. All of the final values lie within the error-bounds of every other method, which indicates the correctness of the different methods. It would make sense, to choose the β even smaller, s.t. the QMC error fraction gets smaller in comparison to the MC error fraction. The decomposition method in general can be seen as a variance reduction technique and what is actually gained from this decomposition is a reduction of variance, because the QMC result has no variance in it, and it's error estimate is deterministic. The part of the calculation, that is afflicted by randomness is reduced, thus the result and its error bound can be seen as the safer one.

N			256	1024	4096	16384	65536	262144	1048576
Pure MC		value	0,62386	0,61802	0,63228	0,63187	0,62930	0,62791	0,62755
		error	0,05898	0,02972	0,01491	0,00743	0,00371	0,00186	0,00093
Pure QMC		value	0,62750	0,62444	0,62538	0,62643	0,62665	0,62671	0,62676
Decomposition Method	beta = 1	value	0,61727	0,62358	0,62959	0,62529	0,62686	0,62670	0,62686
		error	1,82143	0,67111	0,23292	0,07738	0,02501	0,00795	0,00251
	beta = 2	value	0,61911	0,62427	0,62810	0,62561	0,62699	0,62674	0,62685
		error	2,90696	1,06869	0,36947	0,12222	0,03920	0,01230	0,00380
	beta = 3	value	0,62100	0,62426	0,62756	0,62597	0,62703	0,62673	0,62686
		error	3,92468	1,44171	0,49781	0,16444	0,05262	0,01644	0,00505

Components of the Decomposition Method	beta = 1	QMC value	0,61586	0,61568	0,61602	0,61605	0,61604	0,61604	0,61604
		MC value	0,00141	0,00790	0,01358	0,00925	0,01081	0,01066	0,01082
		QMC error	1,81055	0,66387	0,22883	0,07544	0,02405	0,00747	0,00227
		MC error	0,01088	0,00724	0,00409	0,00194	0,00096	0,00048	0,00024
	beta = 2	QMC value	0,62162	0,62339	0,62430	0,62408	0,62408	0,62409	0,62409
		MC value	- 0,00251	0,00088	0,00380	0,00153	0,00290	0,00265	0,00277
		QMC error	2,90039	1,06348	0,36658	0,12085	0,03852	0,01196	0,00363
		MC error	0,00657	0,00522	0,00289	0,00137	0,00068	0,00034	0,00017
	beta = 3	QMC value	0,62292	0,62436	0,62574	0,62558	0,62558	0,62558	0,62558
		MC value	- 0,00192	- 0,00010	0,00182	0,00039	0,00146	0,00114	0,00128
		QMC error	3,91992	1,43730	0,49543	0,16333	0,05206	0,01616	0,00491
		MC error	0,00476	0,00441	0,00237	0,00111	0,00056	0,00028	0,00014

Table 5.2.2: A comparison between pure MC, pure QMC, and the decomposition method for an integrand with infinite variation. The MC error estimates are 95%-confidence intervals.

For most practical situations, the great effort in comparison to a RQMC method is probably not worth it, given that estimation of the variation in the way we did it here is not possible for higher dimensions, as well as the fact, that the possibly rough discrepancy estimate may lead to an unreasonably high error estimate, as we have already seen in Figure 5.1.2.

5.3 Cliquet-Type Payoff

We consider an option of the form (4.3.1) on three assets, with the following parameters:

- $T = 1$,
- $S_1(0) = S_2(0) = S_3(0) = 2$,
- $floor_1 = floor_2 = floor_3 = 1$,
- $cap_1 = cap_2 = cap_3 = 3$,
- $floor_g = 7.5$ and
- $cap_g = 4.5$.

We assume the Black-Scholes model with parameters

- $R = 0$,
- $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$ and no dependency between the assets.

Thus, the terminal-value of both our assets has the form $S_i(1) = e^X$, where $X \sim N(\mu_i, \sigma_i^2)$, with $\mu_i = \log(S_i(0)) - \frac{\sigma_i^2}{2} = 0.5681$. Now we have all ingredients for our integrand, which is then

$$H(y) = Mm\Big(\sum_{i=1}^{s} Mm(G_i^{-1}(y_i), f_i, c_i), f_g, c_g\Big)$$

(the *cap*'s and *floor*'s are abbreviated with c and f). This integrand is depicted in Figure 5.3.1 on the top right, while on the top left, we have the integrand, which would result, if no global cap and no global floor was used. As one can see, all of the kinks, where this function is not differentiable, are parallel to the x- or y-axis, thus the finite variation is preserved for this part, because of Remark 1.5.21. Now when applying the global cap and floor with the outer Min-Max-function (in the top right plot), additional non-differentiable kinks are introduced, which run about diagonally (not exactly - they are slightly curved), which causes the variation to be infinite, if the number of dimensions is greater equal 3. So in order to bring down the variation to a finite level, the outer Min-Max-function is replaced by the smooth approximation given in Section 4.2.3, which can be seen in the lower left plot of Figure 5.3.1. In the lower right plot, we have the difference between the original integrand and the smoothed variant. Observe, that the difference exists only on two small stripes, and the maximum difference is $\frac{1}{4\beta} = 0.125$. We are now going to estimate the variation and the smoothing error. For the variation, we use Lemma 4.3.1:

$$V_{HK}(H_\beta) \leq \sum_{i=1}^{3}(c_i - f_i) + \sum_{\substack{u\subseteq\{1,2,3\}\\ |u|\geq 2}} 4\beta^{|u|-2} C_{|u|} \prod_{i\in u\setminus\kappa(u)} (c_i - f_i).$$

The term $(c_i - f_i)$ is always 2 in our case, $C_2 = 1.4$ and $C_3 = 7.6$, so we receive

$$\begin{aligned} V_{HK}(H_\beta) &\leq 3\cdot 2 + 3\cdot 4\cdot \beta^{2-2}\cdot 1.4\cdot 2 + 4\beta^{3-1}\cdot 7.6\cdot 2^2 \\ &= 39.6 + 121.6\beta. \end{aligned}$$

In order to estimate the smoothing error SE_β, we need the Lipschitz-constants of G_i. Since the G_i's are CDF's of the lognormal-distribution, it's Lipschitz-constants are the maximum of the PDF's of the lognormal-distribution with given μ_i and σ_i. By looking at a plot of the lognormal-PDF with our given parameters, we can easily say $L_{G_1} = L_{G_2} = L_{G_3} = 5.2$. In the sense of Equation (4.3.11), we have

$$Vol(B) \leq 5.2 \cdot \frac{4}{\beta},$$

because $C_c = 0$. Via using (4.3.10), we have then

$$SE_\beta \leq Vol(B) \cdot \frac{1}{4\beta} = 5.2 \cdot \frac{1}{\beta^2}.$$

Combining our results, for the error in the sense of Equation (4.3.4), we have

$$E_\beta = \left| \int_{[0,1]^s} H(x)dx - I_{\mathcal{P}}(H_\beta) \right| \leq (39.6 + 121.6\beta)\cdot D_N^*(\mathcal{P}) + 5.2\cdot\frac{1}{\beta^2}. \tag{5.3.1}$$

If we assume, that we are given the pointset $\mathcal{P}$ with $|\mathcal{P}| = N$ and a discrepancy estimate, one could e.g. determine the parameter β, s.t. (5.3.1) is minimized. Consider a Sobol-pointset in 3 dimensions with N=2^{20}. Our often used discrepancy estimate from Proposition 1.4.13 yields $D_N^*(\mathcal{P}) \leq 0.0014887$. The total error is then minimized for $\beta = 3.8586$, with the error being $E_\beta = 1.1067$

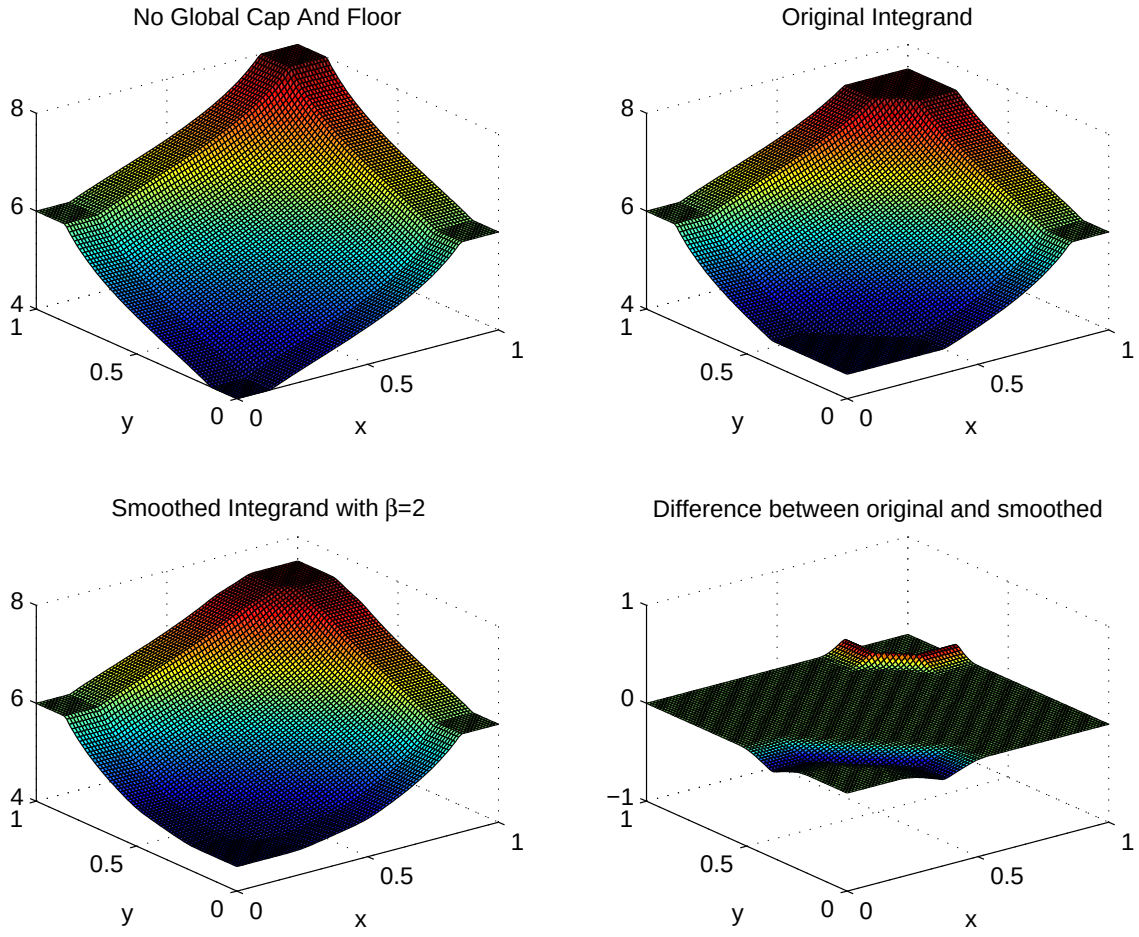

Figure 5.3.1: Different variants of the integrand. The component x_3 is fixed at $x_3 = G_3(2)$, s.t. $S_3(1) = 2$.

5.4 Transformations and Matrix Factorizations

Here we want to give a short comparison between different matrix factorizations and transformations from $\mathbb{R}^s$ to $[0,1]^s$ we discussed in Chapter 3. We consider a discretely sampled Asian option in the sense of Proposition 2.3.5, with the payoff given as

$$h(S(t_1), S(t_2), \ldots, S(t_s)) = \left(\frac{1}{s} \sum_{i=1}^{s} S(t_i) - K \right)^+.$$

We assume the Black-Scholes model, with $T = 1$ year, strike $K = 100$, volatility $\sigma = 0.17$, interest rate $R = 0.03$ and a starting price of $S_0 = 100$. We transform our integrand with the inverse normal CDF in the form of (3.1.7) and the logit-transform in the form of (3.1.13). We heuristically chose the scaling parameters $\lambda = 0.65$ for the logit-transform, the choice is motivated by the same argument as in Section 5.2. For the factorization L of Σ, we use the Cholesky decomposition and the PCA decomposition, as discussed in Section 3.2. We have done calculations for $s = 3$, 12, 52 and 120, which can be seen in Figures 5.4.1 and 5.4.2. The number of RQMC repeats is always fixed at $q = 64$, while the N in the plots is the number N_{RQMC} of integration nodes used in each RQMC repeat.

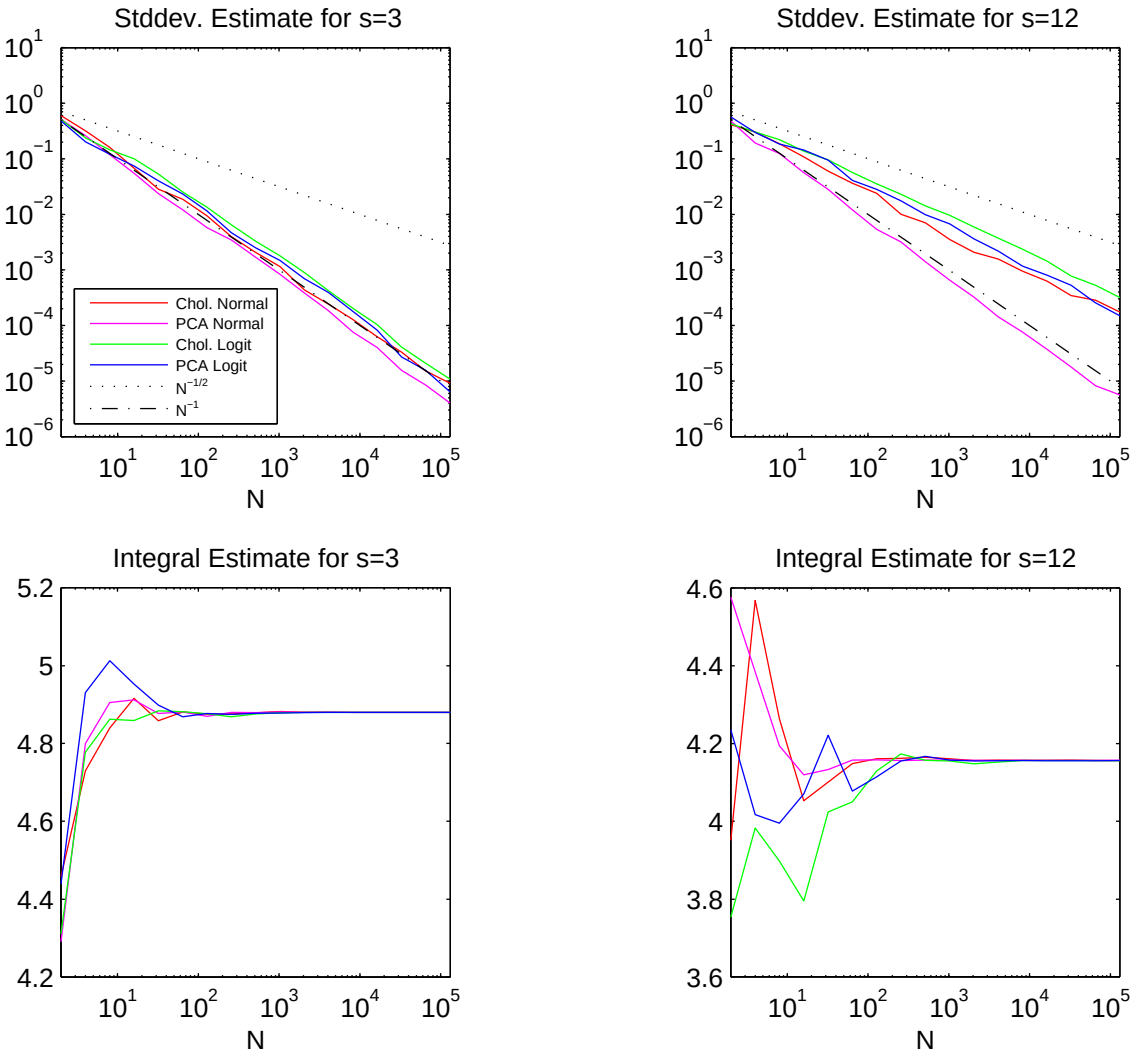

Figure 5.4.1: Standard deviation and integral estimates for $s = 3$ and 12

As one can clearly see, in the case $s = 3$, all of our four different methods converge with $O(N^{-1})$, although the "PCA normal" is slightly better than the rest. For a rising number of dimensions, the convergence speed is reduced in all cases, except for the "PCA normal" method. It is apparent, that for a large number of dimensions, both methods using the logit-transform perform far worse than the "PCA normal" method, and also significantly worse than the "Cholesky normal" method, with a convergence rate of "only" $O(N^{-1/2})$. A calculation for $s = 360$ was also carried out, and the results from the methods involving the logit-transform were not usable, at least for $N^{17} = 131072$, which was the maximum number of points used for a single RQMC repeat. The results of the methods involving the inverse normal transformation for $s = 360$ however were quite similar to the results for $s = 120$. We assume, the bad performance of the logit-transformations is mainly due to the density term in (3.1.13), which is

$$f_Y(\tau(z)) \prod_{i=1}^{s} \frac{\lambda_i}{z_i(1-z_i)} = \prod_{i=1}^{s} f_{Y_i}(\tau_i(z_i)) \frac{\lambda_i}{z_i(1-z_i)}.$$

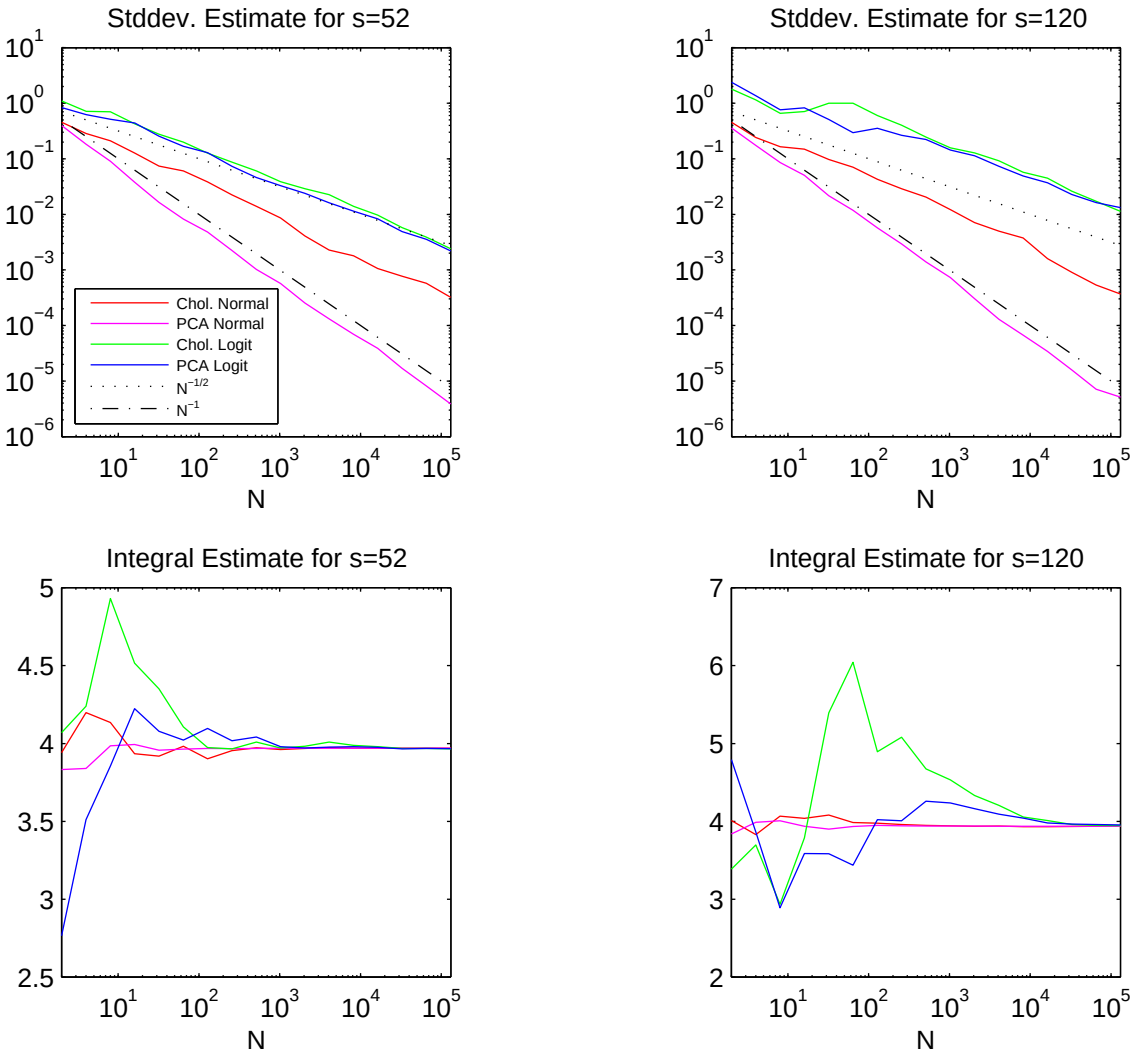

Figure 5.4.2: Standard deviation and integral estimates for $s = 52$ and 120

While this term converges to 0 towards the boundaries of $[0,1]^s$, it has a huge peak in the center, which grows exponentially with s. Hence the variation in the sense of Hardy and Krause is enormous, and also the variance in the sense of (1.1.3). This term also adds a big amount of complexity to the integrand, it depends equally on all variables and this dependency cannot be reduced by the PCA decomposition in this representation, because the matrix L only appears in the other factor of the integrand. These results partially match the findings of Kuo et al. [13], where it is reported that for a certain class of integrands, the normal-transformation always outperforms the logit-transformation, while for another class, this is not the case. The interesting fact here in our example is though, that in trying, to improve the theoretical error estimate by using a transformation s.t. the integrand becomes bounded, the actual numeric results get worse.

Chapter 6

Conclusion

From the context of this work and other QMC literature, it should be clear, that, if applicable, QMC methods perform at least as good as MC methods, and better in many cases. With applicability here we do not mean boundedness and smoothness, but rather that the problem dimension is finite and known. Through randomization we obtain error estimates without any restrictions on the integrand and without any knowledge of its structure, except of course a formula for evaluating the integrand at any given point. While the error estimate of the Koksma-Hlawka inequality is not viable for most practical purposes, because even for well behaving functions with a small number of variables, the error estimate is still an order above the MC error estimate for a reasonably large amount of integration nodes, its benefit is of theoretical nature, because the enhanced convergence rates of QMC methods can be put on theoretical foundations, at least for functions with finite variation. With the application of the logit transformation for unbounded payoffs on the one hand, and the smoothing of non-differentiable kinks, which caused infinite variation on the other hand, we were able to construct integrands with finite variation. For these integrals we have an convergence rate of at least $O(\frac{(\log N)^s}{N})$, which is the convergence rate of the discrepancy of the Halton sequence from Prop. 1.4.8, and an expectedly better rate with the use of the Sobol sequence. However, in the case of the practical example from Section 5.4, we observed a better convergence with the use of the inverse normal CDF transformation, despite unboundedness of the integrand, which may indicate, that the classical Koksma-Hlawka inequality with the variation in the sense of Hardy and Krause is not the best suited error analysis tool for these types of integrands. It may be worthwhile to consider the function space setting introduced in [14], which deals with the integration of functions on $\mathbb{R}^s$ that are weighted with some probability density, but with the use of randomly shifted lattice rules and the corresponding error analysis, which was outside the scope of this work. Under the category further reading, we also would like to mention the paper [7] about non-differentiable integrands, with the central statement being, that most terms in the ANOVA decomposition (3.2.1) (namely the ones of lower order and most importance) are actually reasonably smooth, which could explain, that QMC also works well such integrands, which are not smooth on the first sight.

Bibliography

[1] Paul Bratley and Bennett L. Fox. Algorithm 659: Implementing sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.*, 14(1):88–100, March 1988.

[2] Josef Dick, Frances Y. Kuo, and Ian H. Sloan. High-dimensional integration: the quasi-Monte Carlo way. *Acta Numer.*, 22:133–288, 2013.

[3] Josef Dick and Friedrich Pillichshammer. *Digital nets and sequences.* Cambridge University Press, Cambridge, 2010. Discrepancy theory and quasi-Monte Carlo integration.

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

[5] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York).* Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.

[6] Carl Graham and Denis Talay. *Stochastic simulation and Monte Carlo methods*, volume 68 of *Stochastic Modelling and Applied Probability.* Springer, Heidelberg, 2013. Mathematical foundations of stochastic simulation.

[7] Michael Griebel, Frances Y. Kuo, and Ian H. Sloan. The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition. *Math. Comp.*, 82(281):383–400, 2013.

[8] Hee Sun Hong and Fred J. Hickernell. Algorithm 823: implementing scrambled digital sequences. *ACM Trans. Math. Software*, 29(2):95–109, 2003.

[9] Stephen Joe and Frances Y. Kuo. Remark on Algorithm 659: implementing Sobol's quasirandom sequence generator. *ACM Trans. Math. Software*, 29(1):49–57, 2003.

[10] Donald E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition [of MR0445948].

[11] Ralf Korn. A real-life MC-simulation application: Chance-risk classification of pension products. Presentation at Graz Summer School on Applications of Quasi-Monte Carlo methods, June 2017.

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

[13] F. Y. Kuo, W. T. M. Dunsmuir, I. H. Sloan, M. P. Wand, and R. S. Womersley. Quasi-Monte Carlo for highly structured generalised response models. *Methodol. Comput. Appl. Probab.*, 10(2):239–275, 2008.

[14] Frances Y. Kuo, Ian H. Sloan, Grzegorz W. Wasilkowski, and Benjamin J. Waterhouse. Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands. *J. Complexity*, 26(2):135–160, 2010.

[15] Gunther Leobacher and Friedrich Pillichshammer. *Introduction to quasi-Monte Carlo integration and applications*. Compact Textbooks in Mathematics. Birkhäuser/Springer, Cham, 2014.

[16] Jiří Matoušek. On the L_2-discrepancy for anchored boxes. *J. Complexity*, 14(4):527–556, 1998.

[17] Jiří Matoušek. *Geometric discrepancy*, volume 18 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1999. An illustrated guide.

[18] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, January 1998.

[19] Art B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.*, 34(5):1884–1910, 1997.

[20] 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.

[21] Rudolf Schürer and Wolfgang Schmid. Mint, the online database for optimal parameters of (t, m, s)-nets, (t, s)-sequences, orthogonal arrays, linear codes, and ooas. `http://mint.sbg.ac.at/`, Feb 2018.

[22] Steven E. Shreve. *Stochastic calculus for finance. II.* Springer Finance. Springer-Verlag, New York, 2004. Continuous-time models.