Mathematical Methods for the efficient Assessment of Market and Credit Risk

Vom Fachbereich Mathematik
der Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

von

Oliver Reiß

1. Gutachter: Prof. Dr. Jürgen Franke

2. Gutachter: Prof. Dr. Vladimir Spokoiny

Datum der Disputation: 15. September 2003

D 386

Zusammenfassung

Diese Dissertation beschreibt verbesserte Methoden und Algorithmen zur Bewertung von Markt– und Kreditrisiken und präsentiert deren Anwendung im Rahmen von sowohl standard als auch innovativen Marktmodellen. Die Methoden und Algorithmen zur Analyse des Marktrisikos komplexer Portfolios benötigen eine genaue Kenntnis der Optionssensitivitäten, der sogenannten ,,Griechen". Eine Analyse der Symmetrien in den Finanzmärkten zeigt Beziehungen zwischen den Griechen auf, die ihre effiziente Berechnung ermöglichen. Diese Relationen werden hauptsächlich im Black Scholes Modell erläutert, jedoch gelten einige dieser Beziehungen auch in allgemeineren Modellen, zum Beispiel im Heston Model.
Im allgemeinen werden Portfolios von vielen Basisinstrumenten beeinflusst, daher ist es notwendig, ihre Abhängigkeiten zu charakterisieren. Es ist üblich, solche Abhängigkeiten durch Korrelationsmatrizen zu beschreiben. Jedoch sind die geschätzten Korrelationsmatrizen in der Praxis durch statistische Fehler gestört oder sind aufgrund fehlender Daten singulär. Ein schneller Algorithmus wird vorgestellt, der eine verallgemeinerte Cholesky Zerlegung einer gestörten Korrelationsmatrix liefert. Der Vorteil der verallgemeinerten Methode besteht im Gegensatz zum standard Cholesky Algorithmus darin, dass sie auch auf semi–positive, Rank defiziente Matrizen angewandt werden kann. Außerdem liefert sie eine approximative Zerlegung, falls die Eingangsmatrix indefinit ist. Ein Vergleich mit anderen bekannten Algorithmen, die ähnliche Eigenschaften aufweisen, wird vorgenomen und es stellt sich heraus, dass der neue Algorithmus in den Situationen zu empfehlen ist, in denen der Rechenaufwand besonders kritisch ist.
Die Bestimmung der Gewinn– und Verlustverteilung (P&L) mit Hilfe der klassischen Fourier Invertierung der entsprechenden charakteristischen Funktion ist eine wichtige Methode, die jedoch bei nicht–integrablen charakteristischen Funktionen divergiert. In dieser Arbeit werden zwei Sätze aufgezeigt, die die Fourier Invertierung nicht–integrabler charakteristischer Funktionen erlauben, indem die unbekannte Verteilung durch eine Verteilung mit bekannter Dichte und charakteristischer Funktion approximiert wird. Es wird außerdem gezeigt, dass der FFT Algorithmus dazu geeignet ist die Invertierung durchzuführen, wenn die zugehörige Dichte einen kompakten Träger besitzt.

Diese Methoden spielen eine entscheidende Rolle bei der effizienten Bestimmung der P&L großer Portfolios. Der sogenannte ,,Delta Gamma Normal" Ansatz wurde der Industriestandard für die Einschätzung von Marktrisiken. Es wird gezeigt, dass die Anwendung der vorgestellten Methoden zur substantiellen Reduzierung des Rechenaufwandes dieses Ansatzes führt. Die gleichen Optimierungen lassen sich auch auf das Delta Gamma Student Modell übertragen.
Ein standard Modell zur Bestimmung der P&L eines Kreditportfolios ist CreditRisk$^+$. Dessen Verteilung ist diskret und kann mit Hilfe der Wahrscheinlichkeit erzeugenden Funktion bestimmt werden. Dafür wird eine numerisch stabile Methode vorgestellt und als Alternative wird ein neuer, auf der Fourier Invertierung beruhender Algorithmus, präsentiert. Ferner wird eine Verallgemeinerung des CreditRisk$^+$ Modells entwickelt, die auch Marktrisiken berücksichtigt und deren zugehörige Verteilung mit den vorgestellten Fourier Methoden effizient bestimmt werden kann.

Abstract

The central theme in this thesis concerns the development of enhanced methods and algorithms for appraising market and credit risks and their application within the context of standard and more advanced market models. Generally, methods and algorithms for analysing market risk of complex portfolios involve detailed knowledge of option sensitivities, the so–called "Greeks". Based on an analysis of symmetries in financial market models, relations between option sensitivities are obtained, which can be used for the efficient valuation of the Greeks. Mainly, the relations are derived within the Black Scholes model, however, some relations are also valid for more general models, for instance the Heston model.
Portfolios are usually influenced by lots of underlyings, so it is necessary to characterise the dependencies of these basic instruments. It is usual to describe such dependencies by correlation matrices. However, estimations of correlation matrices in practice are disturbed by statistical noise and usually have the problem of rank deficiency due to missing data. A fast algorithm is presented which performs a generalized Cholesky decomposition of a perturbed correlation matrix. In contrast to the standard Cholesky algorithm, an advantage of the generalized method is that it works for semi–positive, rank deficient matrices as well. Moreover, it gives an approximative decomposition when the input matrix is indefinite. A comparison with known algorithms with similar features is performed and it turns out, that the new algorithm can be recommended in situations where computation time is the critical issue.
The determination of a profit and loss distribution by Fourier inversion of its characteristic function is a powerful tool, but it can break down when the characteristic function is not integrable. In this thesis, methods for Fourier inversion of non–integrable characteristic functions are studied. In this respect, two theorems are obtained which are based on a suitable approximation of the unknown distribution with known density and characteristic function. Further it will be shown, that straightforward Fast Fourier inversion works, when the according density lives on a bounded interval.

The above techniques are of crucial importance to determine the profit and loss distribution (P&L) of large portfolios efficiently. The so–called Delta Gamma normal approach has become industrial standard for the estimation of market risk. It is shown, that the performance of the Delta Gamma normal approach can be improved substantially by application of the developed methods. The same optimization procedure also applies to the Delta Gamma Student model.
A standard tool for computing the P&L distribution of a loan portfolio is the CreditRisk$^+$ model. Basically, the CreditRisk$^+$ distribution is a discrete distribution which can be computed from its probability generating function. For this a numerically stable method is presented and as an alternative, a new algorithm based on Fourier inversion is proposed. Finally, an extension of the CreditRisk$^+$ model to market risk is developed, which distribution can be obtained efficiently by the presented Fourier inversion methods as well.

Table of Contents

Introduction

During the last decades, there has been an impressive development in financial markets, especially the turnover in exchange traded financial derivatives has been grown tremendously.

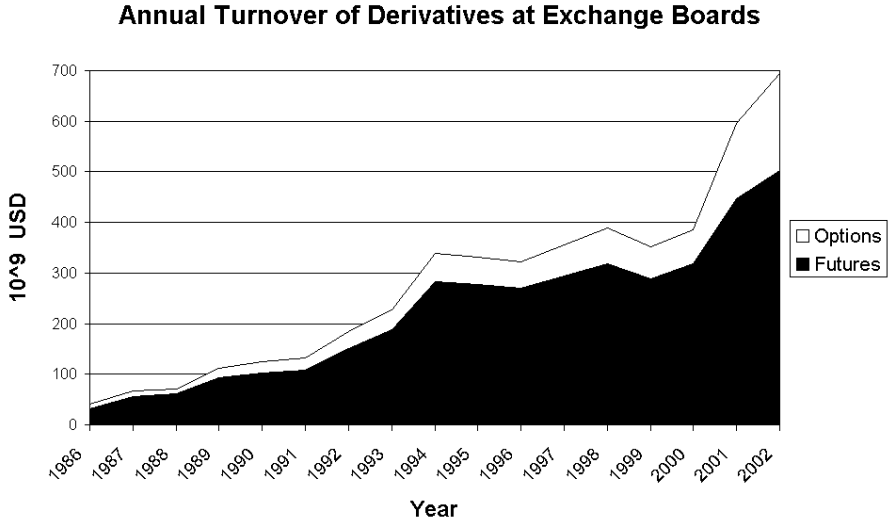

Figure 0.1: Annual turnover of options and futures traded at exchange boards. The data are estimated and provided by the Bank for International Settlements, `http://www.bis.org`.

Due to this development a broad range of derivatives is liquidly traded. An advantage of derivative instruments is, that they can be used by any company or private investor to get an insurance against market fluctuations. On the other hand, financial products become more and more difficult and typical portfolios include lots of such instruments. So it has become difficult to get an adequate assessment of financial risk, the possibility to lose money due to uncertainty. There are spectacular examples of bankrupts which occured, because the risk has been estimated wrongly. One well–known example is the collapse of the *Long Term Capital Management* Fund (LTCM) in 1998, which occured just when I started my research in the area of mathematical finance. This crisis was so serious that the federal reserve bank had to intervene to avoid a breakdown of the global financial markets. So there is a need to monitor financial risks reliably.
Under the leading of the Bank of International Settlements a first agreement has been founded in 1988 [4] (with several modifications thereafter), which demands that each bank has to deposit its financial risk with capital resources. In this context the risk measure "Value at Risk" (VaR) has been established, which is defined as the qth loss–quantile of the profit and loss distribution of a portfolio at some future

time horizon T. Usually, q is one percent and the length of the future period T is given by ten trading days, hence with a probability of 99% the amount of money lost within ten days will not exceed VaR. Meanwhile there is some criticism on this risk measure, because it does not tell something about the losses, which occur in the worst one percent cases. However, the VaR is quite good for regularity purposes, since it is relatively easy to see ex post by the supervising institutions, whether the reported risk was correctly determined.
Partially due to lots of mathematical complications (e.g. complex derivatives, large dimensionality, estimation of model parameters) the Basel agreement was changed in 1995 [5] so wisely, that banks get the possibility to use their own models, the so–called "internal models", for determination of their financial risk. This international agreement was put into national legislation in almost all nations, for example in Germany 1997 (became effective in October 1998) [25] in the "Grundsätze über das Eigenmittel und die Liquidität der Institute" which is an addendum of the "Gesetz über das Kreditwesen".

In general, *"risk"* is the possibility to sustain a downside from an activity or an omission due to uncertainty. In the context of financial markets the activity is an investment, the possible downside is loss of money and the uncertainty is the unknown behaviour of asset prices in the future. But how did it come to such an explosive development in derivative instruments, which led to very complicated questions for risk management? Besides the human greed of trying to gain profit by speculations, in 1973 Black and Scholes [10] initiated the development of a mathematical theory for pricing and hedging derivate products.
Long before Black and Scholes, a first approach was given by Louis M. Bachelier in 1900, who suggested to describe asset prices by Brownian motions [3], but his work was ignored for some decades. Later in the 1950s Jimmie Savage rediscovered Bachelier's ideas and made them known [9]. From an economic point of view Bachelier's approach has two disadvantages. Firstly, stock prices can become negative in this model, which is unrealistic in practice. Secondly, rather than absolute returns, relative returns of assets are of interest. Therefore Osborne suggested to model relative returns of a stock by a Gaussian process with identically, independent distributed increments, leading to a geometric Brownian motion process for the asset price itself [65]. Osborne's model has become a standard model in the financial community under another name: *the Black Scholes model*.
With their famous work "The Pricing of Options and Corporate Liabilities" [10] Black and Scholes constitute the field of mathematical finance. They showed by arbitrage arguments that the price of a European call option is uniquely determined as the initial investment of a self–financing portfolio in the stock and a riskless bond, such that the option payoff is perfectly replicated at maturity by this portfolio. This approach was such epoch making, that Scholes and Merton, who put the Black and Scholes method on a more rigorous base [60], received the Nobel price of economics[1] *for a new method to determine the value of derivatives* in 1997 (Black died in 1995). Black and Scholes used a special market model for their analysis and so the question arose, whether an extension of their approach could be applied to more general

[1]The official title of this price is "The Bank of Sweden Price in Economic Sciences in Memory of Alfred Nobel". Source: The Nobel e–Museum `http://www.nobel.se`.

market models. Harrison and Pliska extended in [48, 49] the Black and Scholes arguments to general complete markets. In fact, they discovered a mathematical existence and uniqueness theorem as interpretation of economic properties of a rather general market model. In their setup they consider a system of price processes X on a probability space $(\Omega, P, \mathcal{F})$. They showed, that the economic arbitrage property, that one cannot gain a free lunch from a zero investment is equivalent with the existence of a measure Q equivalent to P such that X is a martingale under Q. Moreover, if Q is unique, then any contingent claim can be replicated by a certain combination of the basic assets and the market is called "complete". Generalizations of this *Fundamental Theorem of Asset Pricing* can be found, for instance, in [21, 22, 23].

While the mathematical theory for the pricing and hedging of financial derivatives has been settled meanwhile, the question how to measure risk of a large portfolio both reliably and efficiently still remains. In principle, the profit and loss distribution contains all the information. However, for risk management purposes only a certain functional, or risk measure of the profit and loss distribution is relevant. In practice, this measure is usually the VaR. The VaR is most popular, because it is required by the supervising authorities. In spite of its popularity, the VaR–measure has the disadvantage of being non–subadditive in general: The VaR of two merged portfolios might be larger than the sum of the VaRs of the individual components. In this respect, Artzner, Delbaen, Eber and Heath established in 1999 a system of axioms for a *coherent measure of risk* [2], which fulfills economic reasonable conditions.

In this thesis *Mathematical Methods for the Efficient Assessment of Market and Credit Risk* are presented. These methods can advantageously be applied to portfolios, which are typically large and contain rather complex structured products. In this context an important issue is to model the financial markets as realistic as possible, however, such that their computational treatment remains feasible. Since the obtained results are very important for risk controlling departments, this work is more or less intended as a reference guide, therefore relevant, known mathematical results are included for convenience. The thesis is divided into two parts. Part 1 concentrates on *Mathematical Methods* whereas part 2 deals with *Applications in Risk Management*. In the appendix some *Properties of common continuous Distributions* are recalled.

In the first chapter the *Efficient Computation of Option Price Sensitivities*, the so–called "Greeks", is studied. Relationships between Greeks of financial derivatives are established by utilizing symmetries in financial markets models. These relations are very useful to determine the Greeks efficiently. An interesting special case is the European contingent claim in the Black Scholes model. It is shown in theorem 1.11 that knowledge of two rather arbitrary Greeks is enough to compute all the other commonly used ones. Several examples are given. Further it is shown that some relations are valid beyond the Black Scholes model, e.g. in the Heston stochastic volatility model.

In chapter 2 *A generalized non–square Cholesky Decomposition Algorithm* is developed. This algorithm (algorithm 2.7) is in fact a generalization of the known $\mathrm{LDL}^{\mathrm{T}}$ decomposition of semi–positive symmetric matrices. For positive symmetric matrices algorithm 2.7 yields the usual Cholesky decomposition. For semi–positive symmetric matrices S it essentially returns a decomposition $S = CC^{\top}$, with C being a non–

square column–regular matrix, with rank equal to the rank of S. For an indefinite symmetric input matrix the presented algorithm returns the Cholesky decomposition of a semi–positive approximant and an upper bound of the approximation error is established in theorem 2.9. A canonical application of the presented algorithm – not only in the context of mathematical finance – is the decomposition of a disturbed or rank deficient correlation matrix. The algorithm will be compared with several other algorithms in this context which are known in the literature. From this comparison it turns out that the presented method is advantageous in computationally demanding situations.

In chapter 3 the *Fourier Inversion of Characteristic Functions* is studied. For the convenience of the reader basic properties of the Fourier transform and features of characteristic functions are recalled. The characteristic function contains all information about the distribution of a real–valued random variable. It is well known that if the characteristic function is integrable, the random variable has a density which can be obtained by Fourier inversion of the characteristic function. However, if the characteristic function is not integrable a straightforward Fourier inversion does not work. For this situation, two inversion theorems (theorems 3.22 and 3.25) are obtained, which are based on the idea of approximating an unknown distribution by a known distribution with the same first moments. Another, at a first glance surprising result is that one can even obtain an approximative density from a non–integrable characteristic function by the well–known *Fast Fourier Transform* (FFT) algorithm, if the underlying random variable has a distribution with bounded support. This result is explained in section 3.5, where also some techniques for smoothing FFT results are suggested.

The fourth chapter is concerned with *Assessment of Market Risk using Quadratic Approximations*. The mathematical methods developed in part 1, are bundled to improve the Delta Gamma normal approach for the estimation of market risk. As a result, the crucial Fourier inversion algorithm in section 4.1 to determine the profit and loss density has been substantially improved with regard to computational efficiency. By a study of two simple examples it is shown that the usual quadratic Taylor approximation can give very bad results. An alternative approach is suggested, which is motivated by a certain interpolation. In practice asset returns are heavy–tailed rather than normal (or lognormal) distributed. In this context, a quite interesting method is the Delta Gamma t_f approach presented in [43]. It turns out that all improvements developed for the Delta Gamma normal algorithm can successfully be applied to the Delta Gamma t_f model also.

Chapter 5 is dedicated to *CreditRisk$^+$ and an Extension to Incorporate Market Risk*. The computation of the loss distribution in the CreditRisk$^+$ model is treated in section 5.2 by Fourier inversion methods, in particular without the introduction of a basic loss unit as in classical approach [19]. It turns out, that this method is quite flexible in the sense that it can be used also for several generalizations of the standard CreditRisk$^+$ model, for instance in the case of dependent sector variables. In the classical approach of the CreditRisk$^+$ model, the distribution is obtained by introduction of a basic loss unit, leading to an integer–valued loss variable which may be computed via its probability generating function. In [19] it is proposed to compute the probabilities by Panjer's recursion, which turned out to be numerically

instable, however. Recently, a new method for computing the respective probabilities by "nested calculation" is suggested in [38]. But, in [38] the stability of this algorithm is left as an open problem. In section 5.2, the numerical stability of this algorithm is proven. The CreditRisk$^+$ model describes the loan portfolio at one certain future time. In section 5.4 this model is extended as a process of cumulative losses which can be studied in continuous time. This is considered to be an important step since it enables the combination of market and credit risk. In this respect in section 5.5 a model is constructed which covers both the CreditRisk$^+$ model in continuous time and the Delta normal market risk model as special cases. Furthermore, an algorithm to compute the profit and loss distribution of this combined model, which unifies market and credit risk, is presented.

The research on plain credit risk modelling continues to be an important issue. The new Basel Capital accord [6] intends that in the future banks have to deposit their credit risk with own equity, where the amount essentially depends on the credit nominal and the rating of the obligor. As a shortcoming of this plan, the diversification of a loan portfolio is not taken into account. The methods resulting from this thesis and further research on credit risk models could be a key for suggesting more adequate rules to the banking supervision.
The presented mathematical methods to assess financial risks may contribute to improve risk management techniques in financial institutions. The model presented in section 5.5 permits an efficient valuation of the profit and loss of a portfolio, which is influenced by market risk as well as credit risk. Although yet only plain credit risk portfolios are traded, e.g. asset–backed securities, it seems very likely that options on more complex portfolios will be traded in the future, and so an increasing demand for models which unify market and credit risk can be expected.

Part I

Mathematical Methods

Chapter 1

Efficient Computation of Option Price Sensitivities[1]

No practitioner software can survive without providing derivatives of options prices with respect to underlying market or model parameters. The common option price sensitivities, the so–called Greeks, are listed. The computation of these sensitivities is often cumbersome and this chapter provides methods to avoid differentiation as much as possible. The results are applied to European style options and multi–asset options (rainbow options) as well as to plain vanilla options priced in Heston's stochastic volatility model. In all this examples time–consuming computations of derivatives can be avoided.

Due to model–independent, fundamental symmetries in financial markets, many Greeks are related among each other. Homogeneity properties of financial markets such as the *homogeneity of time* and the *homogeneity of the price level* of a financial product are exploited. As a consequence, a mathematical rigorous definition of the terms "strike" and "barrier", which are loosely used in practice, are established. Based on this definition, a natural representation for strike–defined options is introduced, which can help to save computations in the case that a closed form solution for the option price is given.

The basic market model widely used in practice is the Black–Scholes model with stocks (currencies) paying a continuous dividend yield (a foreign interest rate) and with a riskless cash bond. This model supports the homogeneity properties that are valid in general, and its structure is so simple that one can concentrate on the essential elements. In this framework there are additional, model dependent relations for the Greeks of European options, e.g. the partial differential equation the value function must satisfy and relations implied by the assumed distribution of the underlying.

As a special case, the Greeks of European options in the one–dimensional Black–Scholes model are taken under consideration. It turns out, that one only needs to know two not quite arbitrary Greeks in order to calculate all the other Greeks without differentiation. A detailed analysis of the combinations of which two Greeks are sufficient is given in theorem 1.11. By the example of a path–independent barrier call the power of this insight is presented.

[1] This chapter is a collection from the papers I wrote together with Uwe Wystup (Commerzbank AG) and which are published in [94], [95] and [96].

Another interesting example is a European derivative security depending on two assets. For such rainbow options the analysis of the risk due to changes in the correlation of the two assets is very important and in turns out, that this sensitivity is related to simultaneous changes of the two underlying securities, the cross–gamma. To illustrate the relations which hold in this case, the call (put) on the minimum (maximum) of two assets is studied.
Since there are fundamental relations which hold in general, it is worth to treat one example beyond the Black–Scholes model. By the analysis of the Greeks of the plain vanilla options in Heston's stochastic volatility model [50], the extension of this work to more general models is also discussed.
Related work about the computation of Greeks includes Carr [18], who views Greeks as values of suitable derivative securities and then computes them using risk neutral valuation. The computation of Greeks using Monte Carlo simulation has been discussed by Broadie and Glasserman [12] and Glasserman and Zhao [44].
The homogeneity relations discussed in this chapter have several advantages:

1. Time saving in computing derivatives.
2. Robust implementation compared to Greeks via difference quotients and computation of Greeks for Monte Carlo based values.
3. Checks on the quality and consistency of Greeks produced by finite–difference, tree, or Monte Carlo methods.
4. Evidence of relationships among Greeks that wouldn't be seen merely by looking at difference quotients.

To elaborate these statements, recall the computation of Greeks in a recombining tree (see figure 1.1). Following Pelsser and Vorst [66], one can compute some Greeks by the construction of *one* binomial tree.

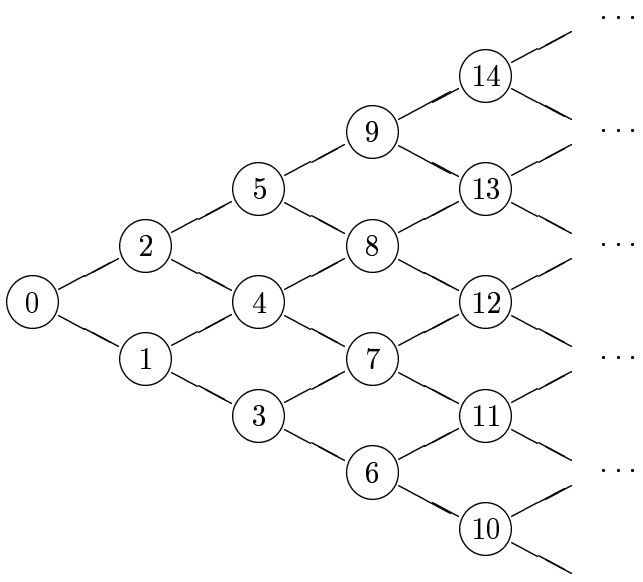

Figure 1.1: A usual recombining binomial tree. Assume, that the node 4 represents the actual spot and valuation date, then some Greeks can be computed within one tree. These Greeks are Delta (use knots 3, 5), Gamma (3, 4, 5), Theta (0, 12) and Charm (1, 2, 7, 8).

As it will be shown in theorem 1.11, the knowledge of the Delta and the Gamma

is sufficient to compute all other common Greeks of a European option. Hence for European options all Greeks can be determinated by construction of one binomial tree only and so the homogeneity relations provide a method for a fast computation of the sensitivities. Since Theta can be determined either from the binomial tree or by application of the theorem 1.11, there is also the possibility to check the quality of the Greeks.
The notations of the market parameters and the commonly used Greeks and their symbols are listed. The list of the Greeks is not claimed to be complete, since one can always define more derivatives of the option price function.

Notation of Market Parameters

S	stock price or stock price process
r	risk free interest rate
q	dividend yield (continuously paid)
σ	volatility of one stock, or volatility matrix of several stocks
ς	correlation in the two-asset market model
t	date of evaluation ("today")
T	date of maturity
$\tau = T - t$	time to maturity of an option
$v(\ldots)$	value of an option
K	strike of an option
B	level of an option
v_x	partial derivation of v with respect to x

The Greeks

Delta	Δ	v_S
Gamma	Γ	v_{SS}
Theta	Θ	v_t
Rho	ρ	v_r
Rho q	ρ_q	v_q
Vega	Φ	v_σ
Kappa (correlation sensitivity)	κ	v_ς
Leverage/Gearing	λ, Ω	$\frac{S}{v} v_S$
Vomma	Φ'	$v_{\sigma\sigma}$
Speed		v_{SSS}
Charm		v_{St}
Color		v_{SSt}
Cross		$v_{S\sigma}$
Dual Theta	DualΘ	v_T
Strike Delta	Δ^K	v_K
Strike Gamma	Γ^K	v_{KK}
Barrier Delta	Δ^B	v_B
Barrier Gamma	Γ^B	v_{BB}
Beta[2]	β_{ij}	$\frac{\sigma_i}{\sigma_j}\varsigma_{ij}$

[2]Indeed, this Greek is not a derivative of the option value, but this definition of "Beta" is quite common, especially between a stock and an index.

1.1 Fundamental Properties

1.1.1 Homogeneity of Time

In most cases the price of the option is not a function of both the current time t *and* the maturity time T, but rather *only* a function of the time to maturity $\tau = T - t$ implying the relations

$$\Theta = v_t = -v_\tau = -v_T = -\text{Dual}\Theta. \tag{1.1}$$

This relationship naturally extends to the situation of options depending on several intermediate times such as compound or Bermuda options.

1.1.2 Scale-Invariance of Time

The principle of the scale–invariance of time holds in general. In a market model, parameters may be quoted on an annual basis. This idea is presented in a Black–Scholes framework, in which the volatility is such a model parameter. The same idea can easily be applied to other market models.
One may want to measure time in units other than years in which case interest rates and volatilities, which are normally quoted on an annual basis, must be changed according to the following rules for all $a > 0$:

$$\begin{aligned} \tau &\rightarrow \frac{\tau}{a} \\ r &\rightarrow ar \\ q &\rightarrow aq \\ \sigma &\rightarrow \sqrt{a}\sigma \end{aligned} \tag{1.2}$$

The option's value must be invariant under this rescaling, hence

$$v(\tau, r, q, \sigma, \ldots) = v(\frac{\tau}{a}, ar, aq, \sqrt{a}\sigma, \ldots) \tag{1.3}$$

Differentiate this equation with respect to a and obtain for $a = 1$

$$0 = \tau\Theta + r\rho + q\rho_q + \frac{1}{2}\sigma\Phi, \tag{1.4}$$

a general relation between the Greeks *theta, rho, rhoq* and *vega*. The same principle also holds for a derivative based on multiple underlyings:

$$\begin{aligned} &v(\tau, r, q_1, ..., q_n, \sigma_{11}, ..., \sigma_{nn}) = \\ &v(\frac{\tau}{a}, ar, aq_1, ..., aq_n, \sqrt{a}\sigma_{11}, ..., \sqrt{a}\sigma_{nn}) \end{aligned} \tag{1.5}$$

This proves the

Theorem 1.1 *(scale invariance of time)*

$$0 = \tau\Theta + r\rho + \sum_{i=1}^{n} q_i\rho_{q_i} + \frac{1}{2}\sum_{i,j=1}^{n} \Phi_{ij}\sigma_{ij}, \tag{1.6}$$

where Φ_{ij} denotes the differentiation of v with respect to σ_{ij}.

1.1.3 Scale Invariance of Prices

The general idea is that the value of securities may be measured in a different unit, just like values of European stocks are now measured in Euro instead of local currencies. Option contracts usually depend on strikes and barriers. Rescaling can have different effects on the value of the option.
Essentially one may consider several types of homogeneity classes. Let $v(S,K)$ be the value function of an option, where S is the spot (or a vector of spots) and K the strike or barrier (or a vector of strikes or barriers).

Definition 1.2 (homogeneity classes) *A value function is K-homogeneous of degree n if for all $a>0$ holds:*

$$v(aS,aK) \quad = \quad a^n v(S,K). \tag{1.7}$$

We call an option whose value function is K–homogeneous of degree 1 a strike–defined option *and an option whose value function is K–homogeneous of degree 0 a* barrier–defined option.

The value function of a European call or put option with strike K is then K-homogeneous of degree 1, a digital option which pays a fixed amount if the stock price is higher than a barrier B is B-homogeneous of degree 0. The path-independent barrier call option paying $(S-k)^+ I_{\{S>K\}}$ is (k,K)-homogeneous of degree 1. A power call with cap paying $\min(C,((S-K)^+)^2)$ has a homogeneity structure of $v(aS,aK,a^2C)=a^2v(S,K,C)$.
Such a scale invariance can be used to determine some relations among the Greeks. This is explained by two examples, first the strike–defined option and second the barrier–defined option. The generalization to options with more parameters like the mentioned path–independent barrier call or power–call can easily be done. For the barrier call one can use the results for the multi–dimensional strike–defined option, what is explained below in 1.3.2.

Example: The barrier–defined option

For a barrier–defined value function holds for all $a,b>0$

$$v(S,B) \quad = \quad v(abS,abB). \tag{1.8}$$

Take the derivative with respect to a at $a=1$:

$$0 \quad = \quad v_S(bS,bB)bS + v_B(bS,bB)bB. \tag{1.9}$$

One may set $b=1$ to get the relation

$$\Delta x + \Delta^B B = 0. \tag{1.10}$$

Differentiation of equation (1.9) with respect to b and evaluate the result at $b=1$:

$$0 \quad = \quad v_{SS}S^2 + 2v_{SB}SB + v_{BB}B^2. \tag{1.11}$$

On the other hand we can differentiate the relation between delta and barrier–delta with respect to B and get

$$v_{SB}S + B\Gamma^B + \Delta^B \quad = \quad 0. \tag{1.12}$$

Together with equation (1.11) one can conclude

$$S^2\Gamma + S\Delta = B^2\Gamma^B + B\Delta^B. \tag{1.13}$$

Example: The strike–defined option

For a strike-defined value function holds for all $a, b > 0$

$$abv(S,K) = v(abS, abK). \tag{1.14}$$

Differentiation with respect to a at $a = 1$ yields:

$$bv(S,K) = bSv_S(bS,bK) + bkv_K(bx,bk). \tag{1.15}$$

Differentiate with respect to b and obtain for $b = 1$:

$$\begin{aligned} v(S,K) &= Sv_S + Sv_{SS}S + Sv_{SK}K + Kv_K + Kv_{KS}S + Kv_{KK}K && (1.16)\\ &= S\Delta + S^2\Gamma + 2SKv_{SK} + K\Delta^K + K^2\Gamma^K. && (1.17)\end{aligned}$$

Equation (1.15) evaluated at $b = 1$:

$$v = S\Delta + K\Delta^K. \tag{1.18}$$

Differentiate this equation with respect to K and obtain

$$\begin{aligned} \Delta^K &= Sv_{KS} + \Delta^K + K\Gamma^K, && (1.19)\\ KSv_{SK} &= -K^2\Gamma^K. && (1.20)\end{aligned}$$

Together with equation (1.17) one concludes:

$$S^2\Gamma = K^2\Gamma^K. \tag{1.21}$$

Conclusion for higher dimensions

In general one obtains in a similar manner

Theorem 1.3 *(price homogeneity)*

$$v = \sum_{i=1}^{n} x_i\Delta_i + \sum_{j=1}^{m} k_j\Delta_j^k \tag{1.22}$$

$$\sum_{i,j=1}^{n} x_ix_j\Gamma_{ij} = \sum_{i,j=1}^{m} k_ik_j\Gamma_{ij}^k \tag{1.23}$$

for strike-defined options and

$$0 = \sum_{i=1}^{n} x_i\Delta_i + \sum_{j=1}^{m} l_j\Delta_j^l \tag{1.24}$$

$$\sum_{i,j=1}^{n} x_ix_j\Gamma_{ij} + \sum_{i=1}^{n} x_i\Delta_i = \sum_{i,j=1}^{m} l_il_j\Gamma_{ij}^l + \sum_{i=1}^{m} l_i\Delta_i^l \tag{1.25}$$

for barrier–defined options.

1.1.4 The Natural Price Representation for Strike–defined Options

Let $v(S,K) \in \mathbf{C}^2$ be the value of a strike–defined option, hence $v(S,K)$ is homogeneous of degree 1. Furthermore, let the option price formula has the structure

$$v(S,K) = Sf(S,K) + Kg(S,K) \tag{1.26}$$

From the Euler theorem on homogeneous functions [13, p. 262] it is known:

$$v(S,K) = Sv_S(S,K) + Kv_K(S,K) \tag{1.27}$$

This raises the question:

Can we conclude $f(S,K) = v_S(S,K)$ and $g(S,K) = v_K(S,K)$?

The answer is: *No, not in general.*

Definition 1.4 (Natural Representation) *Let $v(S,K)$ be $\mathbf{C}^2$, i.e. a two times continuously differentiable function, and let v be a homogeneous function of degree* 1. *Then the representation*

$$v(S,K) = Sf(S,K) + Kg(S,K) \tag{1.28}$$

is called a natural representation *if and only if*

$$S^2\partial_S f(S,K) = K^2\partial_K g(S,K) \tag{1.29}$$

Theorem 1.5 (Existence and Uniqueness of the Natural Representation) *Let $v(S,K)$ be $\mathbf{C}^2$ and homogeneous of degree* 1. *Then the natural representation of $v(S,K)$ exists, is unique and given by:*

$$v(S,K) = Sv_S(S,K) + Kv_K(S,K) \tag{1.30}$$

Proof:
The representation $v = Sv_S + Kv_K$ exists by assumptions on v and is a natural representation, since one obtains by application of $S\partial_S$ and $K\partial_K$ on the stated representation:

$$S^2 v_{SS} + KSv_{KS} = 0 \tag{1.31}$$
$$K^2 v_{KK} + KSv_{KS} = 0 \tag{1.32}$$

The difference yields $S^2 v_{SS} = K^2 v_{KK}$, hence $v = Sv_S + Kv_K$ is a natural representation.
To prove the uniqueness, let $v(S,K) = xf(S,K) + kg(S,K)$ be a natural representation. Since v is homogeneous of degree 1, f and g are homogeneous of degree 0 and hence by the Euler theorem:

$$Sg_S(S,K) + Kg_K(S,K) = 0 \tag{1.33}$$

Together with $S^2 f_S(S,K) = K^2 g_K(S,K)$ one can conclude:

$$Sf_S(S,K) + Kg_S(S,K) = 0 \tag{1.34}$$

On the other hand, differentiation of $v(S,K) = Sf(S,K) + Kg(S,K)$ with respect to S yields

$$v_S(S,K) = f(S,K) + Sf_S(S,K) + Kg_S(S,K) \tag{1.35}$$

Hence $f(S,K) = v_S(S,K)$ and therefore the natural representation is unique. □

The question rised at the beginning can now be answered more exactly. If the price formula of a strike–defined option is given in the natural price representation, then one can read off the Deltas directly from this formula. Of course, one has to differentiate in order to check, whether a pricing formula is given in its natural representation, but usually this differentiation is done much easier then computing the Deltas. The application of this technique will be presented in 1.5.2.
Of course, the natural representation can easily be extended to higher dimensions, i.e. strike–defined options with several strikes or several underlyings. But in practice, the case with one strike and one underlying is most relevant since it is unusual to have a closed form solution for the option price in higher dimensions. Therefore options with only one underlying and one strike have been focused in this section.

1.2 European Options in the Black–Scholes Model

The n-dimensional Black–Scholes model is given by

$$\begin{aligned} dS_i(t) &= S_i(t)[(r-q_i)\,dt + \sigma_i\,dW_i(t)], \quad i=1,\dots,n && (1.36)\\ \mathbf{Cov}(W_i(t),W_j(t)) &= \varsigma_{ij}t, && (1.37)\end{aligned}$$

where r is the risk-free rate, q_i the dividend rate of asset i or foreign interest rate of exchange rate i, σ_i the volatility of asset i and $(W_1,\dots,W_n)$ a standard Brownian motion (under the risk-neutral measure) with correlation matrix ς.
Let v denote today's value of the European option with payoff $f(S_1(T),\dots,S_n(T))$ at maturity T. Then it is known that v satisfies the *Black–Scholes partial differential equation*, which is a relation among the Greeks:

$$0 = -v_\tau - rv + \sum_{i=1}^{n} s_i(r-q_i)v_{s_i} + \frac{1}{2}\sum_{i,j=1}^{n}\sigma_i\sigma_j\varsigma_{ij}s_is_jv_{s_is_j}, \qquad (1.38)$$

where s_i denotes $S_i(0)$.

1.2.1 Relations among Greeks Based on the Lognormal Distribution

The value function v has a representation given by the n-fold integral

$$v = e^{-r\tau}\int f\left(\dots,S_i(0)e^{\sigma_i\sqrt{\tau}x_i+\mu_i\tau},\dots\right)g(\vec{x},\varsigma)\,d\vec{x}, \qquad (1.39)$$

where $\mu_i = r - q_i - \frac{1}{2}\sigma_i^2$ and $g(\vec{x},\varsigma)$ is the n-variate standard normal density with correlation matrix ς. Since one does not want to assume differentiability of the payoff f, one defines a change in the variables $y_i := S_i(0)e^{\sigma_i\sqrt{\tau}x_i+\mu_i\tau}$, which leads to

$$v = e^{-r\tau}\int f(\dots,y_i,\dots)g\left(\dots,\frac{\ln\frac{y_i}{S_i(0)}-\mu_i\tau}{\sigma_i\sqrt{\tau}},\dots,\varsigma\right)\frac{d\vec{y}}{\prod y_i\sigma_i\sqrt{\tau}}. \qquad (1.40)$$

Properties of the Normal Distribution

Some properties of the multivariate normal density function g will be collected. Suppose that the vector X of n random variables with means zero and unit variances has a nonsingular normal multivariate distribution with probability density function

$$g(x_1, \ldots, x_n; \varsigma) = (2\pi)^{-\frac{1}{2}n} |\mathbf{C}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{C}\mathbf{x}\right). \tag{1.41}$$

Here $\mathbf{C}$ is the inverse of ς, the covariance matrix of X. Then the following identity published in [67] can be proved easily by writing the density in terms of its characteristic function.

Theorem 1.6 *(Plackett's Identity)*

$$\frac{\partial g}{\partial \varsigma_{ij}} = \frac{\partial^2 g}{\partial x_i \partial x_j}. \tag{1.42}$$

In the two–dimensional case this reads as[3]

$$\frac{\partial n_2(x, y; \varsigma)}{\partial \rho} = \frac{\partial^2 n_2(x, y; \varsigma)}{\partial x \partial y}, \tag{1.43}$$

which can readily be extended to the corresponding cumulative distribution function:

$$\frac{\partial \mathcal{N}_2(x, y; \varsigma)}{\partial \rho} = \frac{\partial^2 \mathcal{N}_2(x, y; \varsigma)}{\partial x \partial y} = n_2(x, y; \varsigma). \tag{1.44}$$

Correlation Risk and Cross–Gamma

Using the abbreviation $g_{jk} := \frac{\partial^2 g}{\partial x_j \partial x_k}$, the cross–gamma and correlation risk are

$$\frac{\partial^2 v}{\partial S_j(0) \partial S_k(0)} = e^{-r\tau} \frac{1}{S_j(0) S_k(0) \sigma_j \sigma_k \tau} \int f(\ldots, y_i, \ldots) g_{jk} \frac{d\vec{y}}{\prod y_i \sigma_i \sqrt{\tau}}, \tag{1.45}$$

$$\frac{\partial v}{\partial \varsigma_{jk}} = e^{-r\tau} \int f(\ldots, y_i, \ldots) g_{\rho_{jk}} \frac{d\vec{y}}{\prod y_i \sigma_i \sqrt{\tau}}. \tag{1.46}$$

Invoking Plackett's identity (1.42) saying that $g_{\varsigma_{jk}} = g_{jk}$ leads to

Theorem 1.7 *(cross–gamma–correlation risk relationship)*

$$\frac{\partial v}{\partial \varsigma_{jk}} = S_j(0) S_k(0) \sigma_j \sigma_k \tau \frac{\partial^2 v}{\partial S_j(0) \partial S_k(0)}. \tag{1.47}$$

Interest Rate Risk and Delta

A similar computation yields

Theorem 1.8 *(delta–rho relationship)*

$$\frac{\partial v}{\partial q_j} = -S_j(0) \tau \frac{\partial v}{\partial S_j(0)}, \tag{1.48}$$

$$\frac{\partial v}{\partial r} = -\tau \left(v - \sum_{j=1}^{n} S_j(0) \frac{\partial v}{\partial S_j(0)} \right). \tag{1.49}$$

[3] See A.8.1 for the definition of the bivariate normal density n_2 and distribution function $\mathcal{N}_2$.

Volatility Risk and Gamma

The first and second derivative of the density g satisfy:

$$g_j = -g \sum_{i=1}^{n} x_i C_{ij}, \tag{1.50}$$

$$g_{jk} = g \sum_{i=1}^{n} x_i C_{ij} \sum_{i=1}^{n} x_i C_{ik} - g C_{kj}. \tag{1.51}$$

For the j-th vega one finds thus

$$\sigma_j \frac{\partial v}{\partial \sigma_j} = e^{-r\tau} \int f \cdot g \cdot \left(\sum_{i=1}^{n} x_i C_{ij} x_j^- - 1 \right) \frac{d\vec{y}}{\prod y_i \sigma_i \sqrt{\tau}}, \tag{1.52}$$

$$x_j^- := \frac{\ln \frac{y_i}{S_i(0)} - (r - q_i + \frac{1}{2}\sigma_i^2)\tau}{\sigma_i \sqrt{\tau}} = x_j - \sigma_j \sqrt{\tau}, \tag{1.53}$$

where the arguments of f and g are omitted to simplify the notation.
For the cross–gammas one obtains:

$$\sigma_j \sigma_k S_j(0) S_k(0) \tau \frac{\partial^2 v}{\partial S_j(0) \partial S_k(0)} = e^{-r\tau} \int f \cdot g \cdot B_{jk} \frac{d\vec{y}}{\prod y_i \sigma_i \sqrt{\tau}}, \tag{1.54}$$

$$B_{jk} := \sum_{i=1}^{n} x_i C_{ij} \sum_{i=1}^{n} x_i C_{ik} - C_{kj} - \sum_{i=1}^{n} x_i C_{ij} \sigma_k \sqrt{\tau} \delta_{jk}. \tag{1.55}$$

Multiply by ρ_{jk}, sum over k and remember that ς is the inverse of C to obtain:

$$\sum_{k=1}^{n} \rho_{jk} \sigma_j \sigma_k S_j(0) S_k(0) \tau \frac{\partial^2 v}{\partial S_j(0) \partial S_k(0)} = e^{-r\tau} \int f \cdot g \cdot D_j \frac{d\vec{y}}{\prod y_i \sigma_i \sqrt{\tau}}, \tag{1.56}$$

$$D_j := \sum_{i=1}^{n} x_i C_{ij} x_j - 1 - \sum_{i=1}^{n} x_i C_{ij} x_j + \sum_{i=1}^{n} x_i C_{ij} x_j^-. \tag{1.57}$$

In summary, this proves the

Theorem 1.9 *(gamma–vega relationship)*

$$\sigma_j \frac{\partial v}{\partial \sigma_j} = \sum_{k=1}^{n} \rho_{jk} \sigma_j \sigma_k S_j(0) S_k(0) \tau \frac{\partial^2 v}{\partial S_j(0) \partial S_k(0)}. \tag{1.58}$$

Remark 1.10 *In one dimension, the gamma–vega and delta–rho relationships are also mentioned in [75]. Shaw shows there that $v_\sigma - \sigma\tau S^2(t) v_{S(t)S(t)}$ satisfies the Black–Scholes partial differential equation and is hence identically zero for path-independent options. Note that the gamma–vega and the delta–rho relationships do not hold for barrier options, that are path–dependent options.*

1.3 The One–Dimensional Case

1.3.1 Results for European Claims in the Black–Scholes Model

The relationships among the Greeks for European options in the one–dimensional Black–Scholes model are listed below.

0	$=$	$\tau\Theta + r\rho + q\rho_q + \frac{1}{2}\sigma\Phi$	scale invariance of time	(1.59)
v	$=$	$S\Delta + K\Delta^K$	price homogeneity (strike)	(1.60)
$S^2\Gamma$	$=$	$K^2\Gamma^K$	price homogeneity (strike)	(1.61)
$S\Delta$	$=$	$-B\Delta^B$	price homogeneity (barrier)	(1.62)
0	$=$	$B^2\Gamma^B + B\Delta^B - S\Delta - S^2\Gamma$	price homogeneity (barrier)	(1.63)
ρ	$=$	$-\tau(v - S\Delta)$	delta–rho relationship	(1.64)
ρ	$=$	$-\tau v - \rho_q$	rates symmetry	(1.65)
rv	$=$	$\Theta + (r-q)S\Delta + \frac{1}{2}\sigma^2 S^2\Gamma$	Black–Scholes PDE	(1.66)
qv	$=$	$\Theta + (q-r)K\Delta^K + \frac{1}{2}\sigma^2 K^2\Gamma^K$	dual Black–Scholes (strike)	(1.67)
rv	$=$	$\Theta + (q-r+\sigma^2)B\Delta^B + \frac{1}{2}\sigma^2 B^2\Gamma^B$	dual Black–Scholes (barrier)	(1.68)
ρ_q	$=$	$-\tau S\Delta$	delta–rho relationship	(1.69)
ρ	$=$	$-\tau K\Delta^K$	by (1.60) and (1.71)	(1.70)
Φ	$=$	$\sigma\tau S^2\Gamma$	gamma–vega relationship	(1.71)

An interpretation of equation (1.71) can be found in [79]. There are surely more relations one can prove, but the next theorem will give a deeper insight into the relations among the Greeks of European options.

Theorem 1.11 *If the price and two Greeks g_1, g_2 of a European option are given with*

$$g_1 \in G_1 := \{\Delta, \Delta^K, \Delta^B, \rho, \rho_q\}, \tag{1.72}$$

$$g_2 \in G_2 := \{\Gamma, \Gamma^K, \Gamma^B, \Phi, \Theta\}, \tag{1.73}$$

then all the other Greeks ($\in G_1 \cup G_2$) can be calculated. Furthermore, if Θ and another Greek from G_2 is given, it is also possible, to determine all other Greeks.

Proof:
The relations (1.59) to (1.66) are independent of each other. and equations (1.67) to (1.71) are conclusions. The next table provides an overview of all these relations. A X or an O denote that the Greek appears in the relation. Relations marked by X provide a connection between the Greeks of G_1 and G_2. The relations marked with O concern only the Greeks of one set.

		$Greeks \in G_1$					$Greeks \in G_2$				
equation	v	Δ	Δ^K	Δ^B	ρ	ρ_q	Γ	Γ^K	Γ^B	Φ	Θ
(1.59)					X	X				X	X
(1.60)	O	O	O								
(1.61)							O	O			
(1.62)		O		O							
(1.63)		X		X			X		X		
(1.64)	O	O			O						
(1.65)	O				O	O					
(1.66)	X	X					X				X
(1.67)	X		X					X			X
(1.68)	X			X					X		X
(1.69)		O				O					
(1.70)			O		O						
(1.71)							O			O	

Let us now assume the option price and one Greek from the set G_1 are given. Then a look at the table shows that all Greeks of the set G_1 can be evaluated. If all Greeks of the set G_1 are known and additionally one Greek of the set G_2 is given, all other Greeks can be determined. On the other hand, only eight equations are independent, so the knowledge of two Greeks is also the minimum knowledge one needs to determine all ten Greeks. This is the proof of the first statement.
If Θ and another Greek from G_2 is given, then it is always possible to determine one Greek of the set G_1 and one applies the part of this theorem already proved. If Γ, Γ^K or Γ^B is given, one can use one of the Black–Scholes equations (1.66) to (1.68). If vega Φ is given, one can use (1.71) to get Γ. □

Concluding this section, the example of a path–independent barrier call is studied to illustrate, how the theorem 1.11 can be used to organize the calculation of the Greeks.

1.3.2 A Path–Independent Barrier Call

Value

The payoff of a path–independent down–and–out barrier call is given by

$$f(S_T, k, K) \quad = \quad (S-k)^+ \cdot I_{\{S_T > K\}} \tag{1.74}$$

Let $k < K$ – otherwise it would be a plain vanilla call – and therefore the payoff can be written as $(S_T - k)I_{\{S_T>K\}}$. Intuitively one would call K a barrier, but due to the scaling behavior $f(aS_T, ak, aK) = af(S_T, k, K)$ and the definition in section 1.1.3, the parameter k and K are strikes. Hence the path–independent barrier call is a strike–defined option. Using the abbreviation

$$d_\pm \quad := \quad \frac{\ln(\frac{S_0}{K}) + (r-q)\tau \pm \frac{1}{2}\sigma^2\tau}{\sqrt{\sigma^2\tau}}, \tag{1.75}$$

the value of a path–independent down–and–out barrier call is given by [84]:

$$\begin{aligned} v(S_0,k,K) &= e^{-r\tau}\int_K^\infty \frac{s-k}{s\sqrt{2\pi\sigma^2\tau}}\exp\left(-\frac{(\ln(\frac{s}{S_0})-(r-q)\tau+\frac{1}{2}\sigma^2\tau)^2}{2\sigma^2\tau}\right)ds \\ &= S_0e^{-q\tau}\mathcal{N}(d_+)-ke^{-r\tau}\mathcal{N}(d_-). \end{aligned} \tag{1.76}$$

The common Greeks of this option will be determined by using the theorem 1.11. One can see, that this theorem is quite useful for an efficient computation of the Greeks.

Greeks

Delta. Since differentiation cannot be avoided entirely, let us compute the derivative with respect to k, which is obviously

$$v_k = -e^{-r\tau}\mathcal{N}(d_-). \tag{1.77}$$

Differentiation of the integral representation of v with respect to K yields

$$\begin{aligned} v_K &= e^{-r\tau}\frac{k-K}{K\sqrt{2\pi\sigma^2\tau}}\exp\left(-\frac{(\ln(\frac{K}{S_0})-(r-q)\tau+\frac{1}{2}\sigma^2\tau)^2}{2\sigma^2\tau}\right) \\ &= \frac{k-K}{K}\frac{1}{\sqrt{\sigma^2\tau}}e^{-r\tau}n(d_-). \end{aligned} \tag{1.78}$$

In Theorem 1.11 only one strike was assumed. In this example there are two strikes, and therefore one needs two Greeks from the set G_1 to determine all other Greeks of this set. From the price homogeneity the relation

$$v = S_0v_{S_0}+kv_k+Kv_K \tag{1.79}$$

in known, whence one obtains for the spot delta

$$v_{S_0} = e^{-q\tau}\mathcal{N}(d_+)+\frac{K-k}{S_0}\frac{1}{\sqrt{\sigma^2\tau}}e^{-r\tau}n(d_-). \tag{1.80}$$

Rho. Using the relations (1.64) and (1.69) one obtains

$$v_r = \tau ke^{-r\tau}\mathcal{N}(d_-)+\tau\frac{K-k}{\sqrt{\sigma^2\tau}}e^{-r\tau}n(d_-), \tag{1.81}$$

$$v_q = -\tau S_0e^{-q\tau}\mathcal{N}(d_+)-\tau\frac{K-k}{\sqrt{\sigma^2\tau}}e^{-r\tau}n(d_-). \tag{1.82}$$

Gamma. Now all Greeks $\in G_1$ are calculated. To determine some other Greeks without differentiation we need at least one Greek of the set G_2. In the theorem above was assumed, that the option has only one strike, but this option depends on two strikes. So one has to differentiate thrice to get all dual gammas.

$$v_{kk} = 0 \tag{1.83}$$

$$v_{kK} = \frac{1}{K}\frac{1}{\sqrt{\sigma^2\tau}}e^{-r\tau}n(d_-) \tag{1.84}$$

$$v_{KK} = -\frac{k}{K^2}\frac{e^{-r\tau}}{\sqrt{\sigma^2\tau}}n(d_-)+\frac{k-K}{K^2}\frac{e^{-r\tau}}{\sigma^2\tau}n(d_-)d_- \tag{1.85}$$

The extension of (1.61) to the case of one stock and two strikes is the equation (1.23) with $n = 1$ and $m = 2$. In this example this relation is given by

$$S_0^2\Gamma = k^2\Gamma^{kk} + 2kK\Gamma^{kK} + K^2\Gamma^{KK}. \tag{1.86}$$

From this relation, which follows from the homogeneity of v, one obtains for the spot gamma without differentiation

$$v_{S_0S_0} = \frac{ke^{-r\tau}}{S_0^2\sqrt{\sigma^2\tau}}n(d_-) + \frac{k-K}{S_0^2}\cdot\frac{e^{-r\tau}}{\sigma^2\tau}n(d_-)d_-. \tag{1.87}$$

Vega. From (1.71) one gets

$$v_\sigma = \sqrt{\tau}ke^{-r\tau}n(d_-) - (K-k)e^{-r\tau}\frac{1}{\sigma}n(d_-)d_-. \tag{1.88}$$

Theta. By the scale invariance of time (1.59) one obtains

$$\begin{aligned} v_t = -v_\tau &= -rke^{-r\tau}\mathcal{N}(d_-) + qS_0e^{-q\tau}\mathcal{N}(d_+) \\ &\quad -(r-q)\frac{K-k}{\sqrt{\sigma^2\tau}}e^{-r\tau}n(d_-) - \frac{\sigma}{2\sqrt{\tau}}ke^{-r\tau}n(d_-) \\ &\quad +\frac{1}{2\tau}(K-k)e^{-r\tau}n(d_-)d_- \end{aligned} \tag{1.89}$$

1.4 A European Claim in the Two–Dimensional Black–Scholes Model

1.4.1 Pricing of a European Rainbow Option

Rainbow options are financial instruments which depend on several risky assets. Many of them are very sensitive to changes of correlation. Let us call kappa (κ) the derivative of the option value v with respect to the correlation ς. Extensive computational effort is needed to compute kappa, even in a simple framework, but in the Black–Scholes model with two stocks and one cash bond, the cross–gamma–correlation risk relationship (1.47) can easily be used to find kappa.
Let the stock price processes S_1 and S_2 be described by

$$\ln\frac{S_1(\tau)}{S_1(0)} = (r - q_1 - \frac{1}{2}\sigma_1^2)\tau + \sigma_1 W_\tau^1, \tag{1.90}$$

$$\ln\frac{S_2(\tau)}{S_2(0)} = (r - q_2 - \frac{1}{2}\sigma_2^2)\tau + \sigma_2\varsigma W_\tau^1 + \sigma_2\sqrt{1-\varsigma^2}W_\tau^2. \tag{1.91}$$

W^1 and W^2 are two independent Brownian motions under the risk neutral measure. The probability density for the distribution of $S_1(\tau)$ is denoted by $h_1(x)$ and is given by the log-normal density

$$h_1(x) = \frac{1}{\sqrt{2\pi\sigma_1^2\tau}}\frac{1}{x}\exp\left(-\frac{A^2}{2\sigma_1^2\tau}\right), \tag{1.92}$$

$$A := \ln\left(\frac{x}{S_1(0)}\right) - r\tau + q_1\tau + \frac{1}{2}\sigma_1^2\tau. \tag{1.93}$$

The equation for the second stock price process can be written as

$$\begin{aligned}\ln\frac{S_2(\tau)}{S_2(0)} &= (r-q_2-\frac{1}{2}\sigma_2^2)\tau+\frac{\sigma_2\varsigma}{\sigma_1}\left(\ln\left(\frac{S_1(\tau)}{S_1(0)}\right)-(r-q_1-\frac{1}{2}\sigma_1^2)\tau\right)\\ &\quad+\sigma_2\sqrt{1-\varsigma^2}W_\tau^2. \qquad (1.94)\end{aligned}$$

The conditional distribution of $S_2(\tau)$ given $S_1(\tau)$ is thus log-normal with density

$$h_{2|1}(y|x) = \frac{1}{y\sqrt{2\pi\sigma_2^2(1-\varsigma^2)\tau}}\exp\left(-\frac{B^2}{2\sigma_2^2(1-\varsigma^2)\tau}\right), \qquad (1.95)$$

$$B := \left[\ln\left(\frac{y}{S_2(0)}\right)-r\tau+q_2\tau+\frac{1}{2}\sigma_2^2\tau-\frac{\sigma_2\varsigma}{\sigma_1}A\right]. \qquad (1.96)$$

The joint distribution of $S_1(\tau)$ and $S_2(\tau)$ is given by the product of h_1 and h_2

$$h(x,y) = h_1(x)\cdot h_{2|1}(y|x). \qquad (1.97)$$

A European option with maturity τ and payoff $f(S_1(\tau),S_2(\tau))$ will be priced by

$$v = e^{-r\tau}\int_0^\infty\int_0^\infty h(x,y)\cdot f(x,y)dxdy. \qquad (1.98)$$

This integral has exactly the structure of the integrals studied in section 1.2.1. Using the results provided above, one can collect several relationships for the Greeks in the two-dimensional case. Additional, the fundamental symmetry "scale invariance of time" is valid too. Since we concentrate on European options, the two dimensional Black–Scholes partial differential equation also holds.

1.4.2 Relations among the Greeks

The relationships among the Greeks found in n dimensions will be specialized to the case $n=2$. To simplify the notations, the abbreviations $\Delta_i := \partial_{S_i(0)}v$ and $\Gamma_{ij} := \partial_{S_i(0)}\partial_{S_j(0)}v$ are introduced. Some results are:

$$\begin{aligned}
0 &= \rho_{q_1}+S_1(0)\tau\Delta_1, && (1.99)\\
0 &= \rho_{q_2}+S_2(0)\tau\Delta_2, && (1.100)\\
0 &= q_1\rho_{q_1}+q_2\rho_{q_2}+\frac{1}{2}\sigma_1\Phi_1+\frac{1}{2}\sigma_2\Phi_2+r\rho_r+\tau\Theta, && (1.101)\\
0 &= \Theta-rv+(r-q_1)S_1(0)\Delta_1+(r-q_2)S_2(0)\Delta_2\\
&\quad+\frac{1}{2}\sigma_1^2S_1(0)^2\Gamma_{11}+\varsigma\sigma_1\sigma_2S_1(0)S_2(0)\Gamma_{12}+\frac{1}{2}\sigma_2^2S_2(0)^2\Gamma_{22}, && (1.102)\\
\kappa &= \sigma_1\sigma_2\tau S_1(0)S_2(0)\Gamma_{12}, && (1.103)\\
0 &= \varsigma\kappa-\sigma_1\Phi_1+\sigma_1^2\tau S_1(0)^2\Gamma_{11}, && (1.104)\\
0 &= \varsigma\kappa-\sigma_2\Phi_2+\sigma_2^2\tau S_2(0)^2\Gamma_{22}, && (1.105)\\
0 &= \sigma_1\Phi_1-\sigma_2\Phi_2-\sigma_1^2\tau S_1(0)^2\Gamma_{11}+\sigma_2^2\tau S_2(0)^2\Gamma_{22}, && (1.106)\\
\rho_r &= -\tau\left(v-S_1(0)\Delta_1-S_2(0)\Delta_2\right), && (1.107)\\
0 &= \tau v+\rho_{q_1}+\rho_{q_2}+\rho_r. && (1.108)
\end{aligned}$$

Of course one can get more relations by combining the relations above. The relations, which are chosen to be presented, are either similar to the one-dimensional case or have another natural interpretation:

- (1.99) and (1.100). These relations are a justification for the rough way to deal with dividends. One subtracts the dividends from the actual spot price and prices the option with this price and without dividends. This relation is not effected by the two-dimensionality of the problem.

- (1.101). This is the two-dimensional version of the general invariance under time scaling.

- (1.102). This is the Black–Scholes differential equation. This relation must hold, because we concentrated on European claims. It turns out, that the dynamic of an option price is described by the market model and that the price of the option is defined as a boundary problem.

- (1.103). This is the cross–gamma correlation–risk relationship; it is remarkable, that this relationship has such a simple structure.

- (1.104) and (1.105). These are the gamma–vega relationships. Notice that one can determine κ only by knowledge of some derivatives with respect to parameters which concern only one stock. Of course, there is no difference between the first and the second stock. These relations are valid in the one-dimensional case with $\kappa \equiv 0$.

- (1.106) follows from (1.103).

- (1.107). This is the delta–rho relationship. The interest rate risk is well known to be the negative product of duration and the amount of money invested. The term in the parentheses is exactly the amount of money one would have to invest in the cash bond in order to delta–hedge the option.

- (1.108). This relation is the two-dimensional rates symmetry, an extension of equation (1.65). It follows from (1.99), (1.100) and (1.107).

In the following one example will be treated in full detail. Further examples such as outside barrier options and spread options are available in [84].

1.4.3 European Options on the Minimum or Maximum of Two Assets

Let us consider the payoff

$$[\phi\,(\eta \min(\eta S_1(T), \eta S_2(T)) - K)]^+ . \tag{1.109}$$

This is a European call ($\phi = +1$) or put ($\phi = -1$) on the minimum ($\eta = +1$) or maximum ($\eta = -1$) of the two assets $S_1(T)$ and $S_2(T)$ with strike K. Its value

function has been published in [78] and can be written using the bivariate normal distribution function $\mathcal{N}_2(\cdot,\cdot;\cdot)$, see A.8.1, as

$$\begin{aligned} v(t,S_1(t),S_2(t),K,T,q_1,q_2,r,\sigma_1,\sigma_2,\varsigma,\phi,\eta) \qquad (1.110)\\ = \phi\Big[S_1(t)e^{-q_1\tau}\mathcal{N}_2(\phi d_1,\eta d_3;\phi\eta\varsigma_1)\\ +S_2(t)e^{-q_2\tau}\mathcal{N}_2(\phi d_2,\eta d_4;\phi\eta\varsigma_2)\\ -Ke^{-r\tau}\left(\frac{1-\phi\eta}{2}+\phi\mathcal{N}_2(\eta(d_1-\sigma_1\sqrt{\tau}),\eta(d_2-\sigma_2\sqrt{\tau});\varsigma)\right)\Big],\\ \sigma^2 := \sigma_1^2+\sigma_2^2-2\varsigma\sigma_1\sigma_2, \qquad (1.111)\\ \varsigma_1 := \frac{\varsigma\sigma_2-\sigma_1}{\sigma}, \qquad (1.112)\\ \varsigma_2 := \frac{\varsigma\sigma_1-\sigma_2}{\sigma}, \qquad (1.113)\\ d_1 := \frac{\ln(S_1(t)/K)+(r-q_1+\frac{1}{2}\sigma_1^2)\tau}{\sigma_1\sqrt{\tau}}, \qquad (1.114)\\ d_2 := \frac{\ln(S_2(t)/K)+(r-q_2+\frac{1}{2}\sigma_2^2)\tau}{\sigma_2\sqrt{\tau}}, \qquad (1.115)\\ d_3 := \frac{\ln(S_2(t)/S_1(t))+(q_1-q_2-\frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \qquad (1.116)\\ d_4 := \frac{\ln(S_1(t)/S_2(t))+(q_2-q_1-\frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}. \qquad (1.117) \end{aligned}$$

Greeks

Delta. Space homogeneity implies that

$$v = S_1(t)\frac{\partial v}{\partial S_1(t)}+S_2(t)\frac{\partial v}{\partial S_2(t)}+K\frac{\partial v}{\partial K}. \qquad (1.118)$$

Using this equation one only has to differentiate twice in order to get all deltas. It turns out, that the value function is given in the natural representation, which is presented in the appendix, and one is allowed to read off the deltas:

$$\begin{aligned} \frac{\partial v}{\partial S_1(t)} &= \phi e^{-q_1\tau}\mathcal{N}_2(\phi d_1,\eta d_3;\phi\eta\varsigma_1) \qquad (1.119)\\ \frac{\partial v}{\partial S_2(t)} &= \phi e^{-q_2\tau}\mathcal{N}_2(\phi d_2,\eta d_4;\phi\eta\varsigma_2) \qquad (1.120)\\ \frac{\partial v}{\partial K} &= -\phi e^{-r\tau}\left(\frac{1-\phi\eta}{2}+\phi\mathcal{N}_2(\eta(d_1-\sigma_1\sqrt{\tau}),\eta(d_2-\sigma_2\sqrt{\tau});\varsigma)\right) \qquad (1.121) \end{aligned}$$

Gamma. Computing the gammas is actually the last situation where differentiation is needed. Use the identities

$$\begin{aligned} \frac{\partial}{\partial x}\mathcal{N}_2(x,y;\varsigma) &= n(x)\mathcal{N}\left(\frac{y-\varsigma x}{\sqrt{1-\varsigma^2}}\right), \qquad (1.122)\\ \frac{\partial}{\partial y}\mathcal{N}_2(x,y;\varsigma) &= n(y)\mathcal{N}\left(\frac{x-\varsigma y}{\sqrt{1-\varsigma^2}}\right), \qquad (1.123) \end{aligned}$$

and obtain

$$\frac{\partial^2 v}{\partial (S_1(t))^2} = \frac{\phi e^{-q_1\tau}}{S_1(t)\sqrt{\tau}} \left[\frac{\phi}{\sigma_1} n(d_1) \mathcal{N}\left(\eta\sigma \frac{d_3 - d_1\varsigma_1}{\sigma_2\sqrt{1-\varsigma^2}} \right) - \frac{\eta}{\sigma} n(d_3) \mathcal{N}\left(\phi\sigma \frac{d_1 - d_3\varsigma_1}{\sigma_2\sqrt{1-\varsigma^2}} \right) \right], \tag{1.124}$$

$$\frac{\partial^2 v}{\partial (S_2(t))^2} = \frac{\phi e^{-q_2\tau}}{S_2(t)\sqrt{\tau}} \left[\frac{\phi}{\sigma_2} n(d_2) \mathcal{N}\left(\eta\sigma \frac{d_4 - d_2\varsigma_2}{\sigma_1\sqrt{1-\varsigma^2}} \right) - \frac{\eta}{\sigma} n(d_4) \mathcal{N}\left(\phi\sigma \frac{d_2 - d_4\varsigma_2}{\sigma_1\sqrt{1-\varsigma^2}} \right) \right], \tag{1.125}$$

$$\frac{\partial^2 v}{\partial S_1(t) \partial S_2(t)} = \frac{\phi\eta e^{-q_1\tau}}{S_2(t)\sigma\sqrt{\tau}} n(d_3) \mathcal{N}\left(\phi\sigma \frac{d_1 - d_3\varsigma_1}{\sigma_2\sqrt{1-\varsigma^2}} \right). \tag{1.126}$$

Kappa. The sensitivity with respect to correlation is directly related to the cross–gamma

$$\frac{\partial v}{\partial \varsigma} = \sigma_1\sigma_2\tau S_1(t) S_2(t) \frac{\partial^2 v}{\partial S_1(t) \partial S_2(t)}. \tag{1.127}$$

Vega. Using the equations (1.104) and (1.105), one obtains the following formulas for the vegas:

$$\frac{\partial v}{\partial \sigma_1} = \frac{\varsigma v_\varsigma + \sigma_1^2 \tau (S_1(t))^2 v_{S_1(t)S_1(t)}}{\sigma_1} \tag{1.128}$$

$$= S_1(t) e^{-q_1\tau} \sqrt{\tau} \left[\varsigma_1 \phi\eta n(d_3) \mathcal{N}\left(\phi\sigma \frac{d_1 - d_3\varsigma_1}{\sigma_2\sqrt{1-\varsigma^2}} \right) + n(d_1) \mathcal{N}\left(\eta\sigma \frac{d_3 - d_1\varsigma_1}{\sigma_2\sqrt{1-\varsigma^2}} \right) \right], \tag{1.129}$$

$$\frac{\partial v}{\partial \sigma_2} = \frac{\varsigma v_\varsigma + \sigma_2^2 \tau (S_2(t))^2 v_{S_2(t)S_2(t)}}{\sigma_2} \tag{1.130}$$

$$= S_2(t) e^{-q_2\tau} \sqrt{\tau} \left[\varsigma_2 \phi\eta n(d_4) \mathcal{N}\left(\phi\sigma \frac{d_2 - d_4\varsigma_2}{\sigma_1\sqrt{1-\varsigma^2}} \right) + n(d_2) \mathcal{N}\left(\eta\sigma \frac{d_4 - d_2\varsigma_2}{\sigma_1\sqrt{1-\varsigma^2}} \right) \right]. \tag{1.131}$$

Rho. Looking at (1.99), (1.100) and (1.107) the rhos are given by

$$\frac{\partial v}{\partial q_1} = -S_1(t)\tau \frac{\partial v}{\partial S_1(t)}, \tag{1.132}$$

$$\frac{\partial v}{\partial q_2} = -S_2(t)\tau \frac{\partial v}{\partial S_2(t)}, \tag{1.133}$$

$$\frac{\partial v}{\partial r} = -K\tau \frac{\partial v}{\partial K}. \tag{1.134}$$

Theta. Among the various ways to compute theta one may use the one based on (1.101).

$$\frac{\partial v}{\partial t} = -\frac{1}{\tau} \left[q_1 v_{q_1} + q_2 v_{q_2} + r v_r + \frac{\sigma_1}{2} v_{\sigma_1} + \frac{\sigma_2}{2} v_{\sigma_2} \right]. \tag{1.135}$$

1.5 Application to other Market Models

1.5.1 Beyond Black–Scholes

Up to now, the ideas for an efficient computation of the Greeks has been illustrated in the Black–Scholes model and in some part the specific properties of this model have been used. Nevertheless, there are some properties, which are so fundamental, that they should hold in any realistic market model. These fundamental properties are the homogeneity of time, the scale invariance of time and the scale invariance of prices. For every market model one uses, one should ascertain, that the model fulfills these properties. An example for a market model with a non–deterministic volatility is Heston's stochastic volatility model [50]. This model also fulfills the fundamental properties and the ideas for the efficient computation of the Greeks in this framework will be studied.
In such a more general model one needs to clarify the notion of vega. A change of volatility could mean a change of the entire underlying volatility process. If the pricing formula depends on input parameters such as initial volatility, volatility of volatility, mean reversion of volatility, then one can consider derivatives with respect to such parameters. It turns out that the presented strategy to compute Greeks can still be applied successfully in a stochastic volatility model.

1.5.2 Heston's Stochastic Volatility Model

$$dS_t = S_t\left[(r-q)dt + \sqrt{\sigma_t}dW_t^{(1)}\right], \tag{1.136}$$

$$d\sigma_t = \alpha(\bar{\sigma}-\sigma_t)\,dt + \gamma\sqrt{\sigma_t}dW_t^{(2)}, \tag{1.137}$$

$$\mathbf{Cov}\left[dW_t^{(1)}, dW_t^{(2)}\right] = \varsigma\,dt. \tag{1.138}$$

The model for the variance σ_t is the same as the one used by *Cox, Ingersoll and Ross* for the short term interest rate, see [20]. $\bar{\sigma} > 0$ is the long term variance, $\alpha > 0$ is the rate of mean-reversion and γ is the volatility of the volatility process. One also has to introduce a parameter λ and the quantity $\lambda\sigma$ is called the market price of volatility risk. Heston provides a closed-form solution for European vanilla options paying

$$\left[\phi\left(S_T - K\right)\right]^+. \tag{1.139}$$

As usual, the binary variable ϕ takes the value $+1$ for a call and -1 for a put, K is the strike (in units of the domestic currency), q is the dividend yield (or foreign risk free rate) of the asset (or exchange rate) S, r denotes the (domestic) risk free rate and T is the expiration time.

Abbreviations

$$u_1 := \frac{1}{2} \tag{1.140}$$

$$u_2 := -\frac{1}{2} \tag{1.141}$$

$$
\begin{aligned}
b_1 &:= \alpha + \lambda - \gamma\varsigma && (1.142)\\
b_2 &:= \alpha + \lambda && (1.143)\\
d_j &:= \sqrt{(\varsigma\gamma\varphi i - b_j)^2 - \gamma^2(2u_j\varphi i - \varphi^2)} && (1.144)\\
g_j &:= \frac{b_j - \varsigma\gamma\varphi i + d_j}{b_j - \varsigma\gamma\varphi i - d_j} && (1.145)\\
\tau &:= T - t && (1.146)\\
D_j(\tau,\varphi) &:= \frac{b_j - \varsigma\gamma\varphi i + d_j}{\gamma^2}\left[\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}}\right] && (1.147)\\
C_j(\tau,\varphi) &:= (r-q)\varphi i\tau \\
&\quad + \frac{\alpha\bar{\sigma}}{\gamma^2}\left\{(b_j - \varsigma\gamma\varphi i + d)\tau - 2\ln\left[\frac{1 - g_j e^{d_j\tau}}{1 - e^{d_j\tau}}\right]\right\} && (1.148)\\
f_j(x,\sigma,t,\varphi) &:= e^{C_j(\tau,\varphi) + D_j(\tau,\varphi)\sigma + i\varphi x} && (1.149)\\
P_j(x,\sigma,\tau,y) &:= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-i\varphi y} f_j(x,\sigma,\tau,\varphi)}{i\varphi}\right] d\varphi && (1.150)\\
p_j(x,\sigma,\tau,y) &:= \frac{1}{\pi}\int_0^\infty \Re\left[e^{-i\varphi y} f_j(x,\sigma,\tau,\varphi)\right] d\varphi && (1.151)\\
P_+(\phi) &:= \frac{1-\phi}{2} + \phi P_1(\ln S_t, \sigma_t, \tau, \ln K) && (1.152)\\
P_-(\phi) &:= \frac{1-\phi}{2} + \phi P_2(\ln S_t, \sigma_t, \tau, \ln K) && (1.153)
\end{aligned}
$$

This notation is motivated by the fact that the numbers P_j are the cumulative distribution functions (in the variable y) of the log-spot price after time τ starting at x with some volatility σ. The functions p_j are the respective densities.

Value

The value function for European vanilla options is given by

$$v = \phi\left[S_t e^{-q\tau} P_+(\phi) - K e^{-r\tau} P_-(\phi)\right] \tag{1.154}$$

The value function takes the form of the Black–Scholes formula for vanilla options. The probabilities $P_\pm(\phi)$ correspond to $\mathcal{N}(\phi d_\pm)$ in the constant volatility case.

Greeks

One can use the homogeneity of prices to obtain the deltas. But of course, one has to show, that the price formula (1.154) is given in its natural representation. Therefore, the following strategy is used. First, assume that equation (1.154) gives the natural price representation, which is defined in 1.1.4. Under this assumption one can read off the deltas, and from the deltas one may derive the gammas by differentiation. Using the definition of the natural representation (1.29) one can show, that the assumption on the representation of equation (1.154) was true and hence the determined Deltas and Gammas are correct.

Spot delta.

$$\Delta = \frac{\partial v}{\partial S_t} = \phi e^{-q\tau} P_+(\phi) \tag{1.155}$$

Dual delta.

$$\Delta^K = \frac{\partial v}{\partial K} = -\phi e^{-r\tau} P_-(\phi) \tag{1.156}$$

Gamma. Under the condition, that the deltas are correct, we obtain for the gammas by differentiation:

Spot Gamma.

$$\Gamma = \frac{\partial \Delta}{\partial S_t} = \frac{\partial \Delta}{\partial x}\frac{\partial x}{\partial S_t} = \frac{e^{-q\tau}}{S_t} p_1(\ln S_t, \sigma_t, \tau, \ln K) \tag{1.157}$$

Dual Gamma.

$$\Gamma^K = \frac{\partial \Delta^K}{\partial K} = \frac{\partial \Delta^K}{\partial y}\frac{\partial y}{\partial K} = \frac{e^{-r\tau}}{K} p_1(\ln S_t, \sigma_t, \tau, \ln K) \tag{1.158}$$

Proof of the natural representation assumption From 1.1.4 it is known, that the initial guess for the Deltas is correct, if the relation

$$S_t^2 \Gamma \;=\; K^2 \Gamma^K \tag{1.159}$$

holds. In fact, one may plug in the Gammas into this equation

$$S_t e^{-q\tau} p_1(\ln S_t, \sigma_t, \tau, \ln K) = K e^{-r\tau} p_2(\ln S_t, \sigma_t, \tau, \ln K), \tag{1.160}$$

and this statement is true. So the calculation for the Deltas and Gammas has been finished.

Rho. Rho is connected to Delta via equations (1.69) and (1.71).

$$\frac{\partial v}{\partial r} \;=\; \phi K e^{-r\tau} \tau P_-(\phi), \tag{1.161}$$

$$\frac{\partial v}{\partial q} \;=\; -\phi S_t e^{-q\tau} \tau P_+(\phi). \tag{1.162}$$

Theta. Theta can be computed using the partial differential equation for the Heston vanilla option

$$\begin{aligned} rv \;=\;& v_t + (r-q)Sv_S + \frac{1}{2}\sigma S^2 v_{SS} \\ & + \frac{1}{2}\gamma\sigma v_{\sigma\sigma} + \gamma\sigma S v_{\sigma S} + [\alpha(\bar{\sigma} - \sigma) - \lambda] v_\sigma \end{aligned} \tag{1.163}$$

where the derivatives with respect to initial variance σ must be evaluated numerically.

1.6 Conclusion

As one can see, the use of very general fundamental symmetries in financial markets can help to compute analytical formulas of Greeks for analytically known value functions of options in one and higher dimensional markets. The method helps saving computation time for the mathematician who has to differentiate complicated formulas as well as for the computer, because analytical results for Greeks are usually faster to evaluate than finite differences involving at least twice the computation of the option's value. Knowing how the Greeks are related among each other can speed up finite–difference-, tree–, or Monte Carlo–based computation of Greeks or lead at least to a quality check. Many of the results are valid beyond the Black–Scholes model, as an example an analysis of plain vanilla options in Heston's stochastic volatility model has been performed.

For European options in the Black–Scholes model, which is the most common model in practice, it turned out, that the knowledge of two (not quite arbitrary) Greeks and the option price is enough, to determine all other common Greeks without any further differentiation. Most remarkably some relations in this case are based on properties of the normal distribution refreshing the active interplay between mathematics and financial markets.

The computation of the Greeks is quite important, not only for trading aspects, but also for risk management purposes. The relevance of the Greeks in risk management will be explained in chapter 4, where a standard approach to assess market risk is studied in detail. This so called Delta–Gamma normal approach requires the knowledge of all "Deltas" and "Gammas". Since all first derivatives are called "Delta" and all second derivatives are called "Gamma" in the context of risk management, the ideas of this chapter for an efficient computation of the Greeks turn out to be very fruitful.

Chapter 2

A generalized non–square Cholesky Decomposition Algorithm[1]

In financial modelling it is quite common to describe certain risk factors by (log–) normal random variables and hence the dependence between these risk factors is simply given by a correlation matrix. In order to decompose these risk factors into independent risk contributions there is the problem to find the Cholesky factorization of the correlation matrix. But also beyond financial applications there is the need to compute a Cholesky decomposition of a semi–positive symmetric matrix.
The usual Cholesky decomposition (see e.g. [68]) is well known and widely used in practice. It is a fast and numerical stable algorithm, if the given symmetric input matrix S is positive, i.e. it holds $x^\top Sx > 0$ for all vectors $x \neq 0$. But the usual Cholesky decomposition algorithm fails, if the input matrix is not positive but semi–positive, i.e. it holds $x^\top Sx \geq 0$ for all vectors – although a Cholesky decomposition exists, which can be determined by LDL^T decomposition.
The LDL^T algorithm will be extended in such a way, that it coincides for a semi–positive input with the LDL^T decomposition. Based on this decomposition, a non–square Cholesky decomposition for a rank deficient semi–positive symmetric matrix S can be introduced, hence one obtains a decomposition $S = CC^\top$ with a non–square matrix C. Some properties of semi–positive matrices and the LDL^T decomposition are recalled, including an analysis of the existence and uniqueness of certain versions of the LDL^T decomposition. To obtain a numerical stable LDL^T algorithm a pivoting strategy based on [45] is recommended.
There is no Cholesky decomposition if the given symmetric matrix is not semi–positive and in this case one may aim to approximate this matrix by a semi–positive one and to compute the Cholesky decomposition of the approximation. This can be accomplished by the generalized Cholesky decomposition algorithm which is based on the LDL^T decomposition. This algorithm performs an additional rescaling which guarantees, that any indefinite symmetric matrix with non–negative diagonal elements will be approximated by changing off–diagonal elements only. Hence, if the diagonal elements of the input matrix are non–negative, the approximation has the same diagonal elements – a property which is quite important in the context of per-

[1] This chapter is an enhancement of my paper [89].

turbed correlation matrices. The error of this approximation is studied and upper bounds for each step of the algorithm are established.
There are other LDL^T based algorithms which return a positive approximation of a symmetric matrix. One algorithm has been provided by Gill, Murray and Wright [39] from the context of numerical optimization and refinements of this algorithm have been made by Eskow and Schnabel [73, 74]. The generalized Cholesky decomposition algorithm is compared with this two approximation approaches. In contrast to the algorithms mentioned before, the new algorithm gives a decomposition, which has a lower rank. This gives the important opportunity to introduce a dimension reduction, if possible, and hence this algorithm may help to save computation time.
The application of the presented algorithm in the context of perturbed correlation matrices is studied. There are two standard techniques to approximate a unit–diagonal symmetric matrix by a correlation matrix, i.e. a symmetric, semi–positive, unit–diagonal matrix. One method is based on spectral decomposition and the other one is the linear shrinking approach. A comparison of these three algorithms with respect to the computation time, the rank of the approximation and the approximation error concludes this chapter.

2.1 LDL^T for Semi–positive Matrices

Let us recall some basics about semi–positive matrices in the following lemma.

Lemma 2.1 *Let S be a semi–positive symmetric $N \times N$ matrix. Then the following statements hold:*

1. $S_{ii} \geq 0$
2. $|S_{ij}| \leq \sqrt{S_{ii} S_{jj}}$
3. *If the $N \times N$ matrix*

$$S = \left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array} \right) \tag{2.1}$$

is semi–positive, then the following statements are true:

- *$\tilde{S}$ is semi–positive*
- *If $S_{11} > 0$ then $\tilde{S} - \frac{1}{S_{11}} s\, s^\top$ is semi–positive*
- *If $S_{11} = 0$ then $s = 0$.*

Proof:

1. Set $x_k = \delta_{ki}$ and obtain:

$$0 \leq x^\top S x = \sum_{k,l} \delta_{ik} S_{kl} \delta_{li} = S_{ii} \tag{2.2}$$

2. Fix k, l with $k \neq l$ and set $x_i = \delta_{ik} + \alpha\delta_{il}$. Then holds for all $\alpha \in \mathbb{R}$:

$$0 \quad \leq \quad x^\top S x = \sum_{i,j} (\delta_{ki} + \alpha\delta_{li}) S_{ij} (\delta_{jk} + \alpha\delta_{jl}) = S_{kk} + 2\alpha S_{kl} + \alpha^2 S_{ll} \tag{2.3}$$

   Hence $S_{kl}^2 - S_{kk} S_{ll} \leq 0$.

3. Set $x = (\alpha, y)^\top$. Then holds for all $\alpha \in \mathbb{R}$, $y \in \mathbb{R}^{N-1}$:

$$0 \leq x^\top S x = \alpha^2 S_{11} + 2\alpha s^\top y + y^\top \tilde{S} y \tag{2.4}$$

   With $\alpha = 0$ one obtains the semi–positivity of $\tilde{S}$. Since (2.4) holds for all α we may conclude $s^\top y - S_{11} y^\top \tilde{S} y \leq 0$ for all y and obtain the other two statements.

□

There are several representations of the LDL$^{\mathrm{T}}$ decomposition, three versions are stated in the next theorem.

Theorem 2.2 *Let S be a symmetric, semi–positive $N \times N$ matrix.*

1. *Then there exists a decomposition*

$$S \quad = \quad L \cdot D \cdot L^\top \tag{2.5}$$

   where D is a diagonal $N \times N$ matrix and L is a unit left-triangular $N \times N$ matrix.

2. *Define the head of a column i in a matrix L by*

$$h_L(i) \quad := \quad \min\{j | L_{ji} \neq 0\}. \tag{2.6}$$

   If M is the rank of $S \neq \mathbf{0}$, then there exists a unique decomposition

$$S \quad = \quad L D L^\top \tag{2.7}$$

   where D is a positive diagonal $M \times M$ matrix and L is a $N \times M$ matrix with the properties:

   - *The head in each column is 1, i.e. $L_{h_L(i)i} = 1$.*
   - *The head of the* ith *column stands below the head of the* $(i-1)$th *column, i.e. $h_L(i) > h_L(i-1)$.*

3. *Furthermore, if M is the rank of S with $S \neq 0$, then there exists a permutation P such that*

$$P S P^\top \quad =: \quad U = L D L^\top \tag{2.8}$$

   with a positive diagonal $M \times M$ matrix D and L is a $N \times M$ matrix with $L_{ii} = 1$ and the decomposition of U is unique.

Proof:
The statements will simultaneously be proved by complete induction. For $N = 1$ a $\mathrm{LDL}^{\mathrm{T}}$ decomposition is given by $S = (1)(S_{11})(1)$, hence $D_{11} = S_{11}$. For $S \neq \mathbf{0}$ this decomposition is obviously unique, hence the three statements hold.
Let us assume, that the theorem holds for dimension $N-1$. Then for the decomposition of a semi–positive $N \times N$ matrix with $N \geq 2$,

$$S = \left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array}\right) \tag{2.9}$$

we distinguish three cases:

1. Case: $S_{11} \neq 0$ and $\tilde{S} \neq \frac{ss^\top}{S_{11}}$
   Statement 1. By the induction assumption there exists a $\mathrm{LDL}^{\mathrm{T}}$ decomposition

$$\tilde{S} - \frac{ss^\top}{S_{11}} = \tilde{L}\tilde{D}\tilde{L}^\top \tag{2.10}$$

   and so a $\mathrm{LDL}^{\mathrm{T}}$ decomposition is given by

$$\left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{s}{S_{11}} & \tilde{L} \end{array}\right) \left(\begin{array}{c|c} S_{11} & 0 \\ \hline 0 & \tilde{D} \end{array}\right) \left(\begin{array}{c|c} 1 & \frac{s^\top}{S_{11}} \\ \hline 0 & \tilde{L}^\top \end{array}\right) \tag{2.11}$$

   Statement 2. Since $\mathrm{rank}(S) \geq 2$, any decomposition has to satisfy the equation:

$$\left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array}\right) = \left(\begin{array}{c|c} L_{11} & 0 \\ \hline l & \tilde{L} \end{array}\right) \left(\begin{array}{c|c} D_{11} & 0 \\ \hline 0 & \tilde{D} \end{array}\right) \left(\begin{array}{c|c} 1 & l^\top \\ \hline 0 & \tilde{L}^\top \end{array}\right) \tag{2.12}$$

$$= \left(\begin{array}{c|c} L_{11}^2 D_{11} & L_{11} D_{11} l^\top \\ \hline L_{11} D_{11} l & D_{11} l l^\top + \tilde{L}\tilde{D}\tilde{L}^\top \end{array}\right) \tag{2.13}$$

   Since L_{11} is either 0 or 1, we find the necessary conditions $L_{11} = 1$, $D_{11} = S_{11}$ and $l = \frac{1}{S_{11}} s$. The remaining equation $\tilde{S} - \frac{ss^\top}{S_{11}} = \tilde{L}\tilde{D}\tilde{L}^\top$ has a unique solution under the structure conditions on $\tilde{L}$ by induction assumption and the matrix

$$L = \left(\begin{array}{c|c} 1 & 0 \\ \hline l & \tilde{L} \end{array}\right) \tag{2.14}$$

   also fulfills the structure conditions.

   Statement 3. By induction assumption there exists a permutation $\tilde{P}$ such that $\tilde{P}(\tilde{S} - \frac{ss^\top}{S_{11}})\tilde{P}^\top = \tilde{L}\tilde{D}\tilde{L}^\top$. Then define

$$P := \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{P} \end{array}\right) \tag{2.15}$$

   and any decomposition must satisfy

$$PSP^\top = \left(\begin{array}{c|c} S_{11} & s^\top \tilde{P}^\top \\ \hline \tilde{P}s & \tilde{P}\tilde{S}\tilde{P}^\top \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline l & \tilde{L} \end{array}\right) \left(\begin{array}{c|c} D_{11} & 0 \\ \hline 0 & \tilde{D} \end{array}\right) \left(\begin{array}{c|c} 1 & l^\top \\ \hline 0 & \tilde{L}^\top \end{array}\right)$$

$$= \left(\begin{array}{c|c} D_{11} & D_{11} l^\top \\ \hline D_{11} l & D_{11} l l^\top + \tilde{L}\tilde{D}\tilde{L}^\top \end{array}\right) \tag{2.16}$$

   Since we have the necessary conditions $D_{11} = S_{11}$ and $\tilde{P}s = D_{11} l$, the given decomposition of $PSP^\top$ is unique.

2. Case: $S_{11} \neq 0$ and $\tilde{S} = \frac{ss^\top}{S_{11}}$

Statement 1. The proof of the first case also holds in this case.

Statement 2. The rank of S is 1 and any decomposition has to satisfy the equation

$$\left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \frac{ss^\top}{S_{11}} \end{array}\right) = \begin{pmatrix} L_{11} \\ l \end{pmatrix} (D_{11})(L_{11}\ l^\top) \tag{2.17}$$

which has under the condition $L_{11} \in \{0, 1\}$ the unique solution

$$D_{11} = S_{11} \qquad L_{11} = 1 \qquad l = \frac{1}{S_{11}} s \tag{2.18}$$

and the structure condition is also fulfilled.

Statement 3. Let P be the unit matrix, then any decomposition has to satisfy

$$\left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \frac{ss^\top}{S_{11}} \end{array}\right) = \begin{pmatrix} 1 \\ l \end{pmatrix} (D_{11})(1\ l^\top) = \left(\begin{array}{c|c} D_{11} & D_{11} l^\top \\ \hline D_{11} l & D_{11} l l\top \end{array}\right) \tag{2.19}$$

with the unique solution $D_{11} = S_{11}$ and $l = s/S_{11}$.

3. Case: $S_{11} = 0$

Statement 1. By induction assumption there exists a decomposition of the semi–positive matrix $\tilde{S} = \tilde{L}\tilde{D}\tilde{L}^\top$. Since $s = 0$ in this case by lemma 2.1 a $\mathrm{LDL^T}$ decomposition of S is given by

$$\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \tilde{S} \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline l & \tilde{L} \end{array}\right) \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \tilde{D} \end{array}\right) \left(\begin{array}{c|c} 1 & l^\top \\ \hline 0 & \tilde{L}^\top \end{array}\right) \tag{2.20}$$

This decomposition is not unique, since any $l \in \mathbb{R}^{N-1}$ fits.

Statement 2. In this situation the rank of S is equal to the rank of $\tilde{S}$ and therefore any decomposition of type 2 must have the structure

$$\left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \tilde{S} \end{array}\right) = \begin{pmatrix} l^\top \\ \tilde{L} \end{pmatrix} \tilde{D} \begin{pmatrix} l & \tilde{L}^\top \end{pmatrix} \tag{2.21}$$

The condition of the (1,1) element yields $0 = l\tilde{D}l^\top$. Since $\tilde{D}$ is positive and diagonal we obtain $l = 0$. The remaining condition is $\tilde{S} = \tilde{L}\tilde{D}\tilde{L}^\top$ which has a unique solution under the structure conditions by induction assumption and the matrix L also fulfills these conditions.

Statement 3. Since $S \neq \mathbf{0}$ there is a j such that $S_{jj} \neq 0$. Let P be the permutation matrix which permutes 1 with j. Then there is a unique decomposition of the matrix $U = PSP^\top$, because U belongs to either of the cases studied above.

□

In principle, this proof yields almost the $\mathrm{LDL}^{\mathrm{T}}$ algorithm, but to obtain a numerical stable algorithm, one must take care of some points. First of all, it is very critical to decide numerically, whether a value is zero or not. So one defines an $\epsilon > 0$ and every number smaller than ϵ is assumed to be zero. For the choice of ϵ one can take the machine precision for example. If one works with the IEEE representation of floating point numbers, the machine precision is a relative precision; for example the relative precision of a 8 byte double is about 1e-15. On the other hand several (at most N) computations are needed to determine D_{ii} and therefore one must also consider the roundoff for this number. So I suggest to take

$$\epsilon = \text{relative precision} \cdot \max_i S_{ii} \cdot N \tag{2.22}$$

In general, the elements D_{ii} are not bounded from below by a positive number and therefore one has to use a pivoting procedure, which will be performed by symmetric permutations on the matrix S. The permutation is chosen such, that the largest diagonal element gets at the leading position. If the pivot is less then ϵ, the corresponding column of L (the arbitrary l) is set to zero.
The following recursive algorithm for the $\mathrm{LDL}^{\mathrm{T}}$ decomposition of semi–positive matrices performs the permutations as in the third part of the theorem, but to keep the algorithm more simple, it only deals with square matrices. In order to obtain the decomposition as in part 3 of the theorem, define M as the number of non–zero diagonal elements of D and cut the last $N - M$ columns of L and the last $N - M$ columns and rows of D.

Algorithm 2.3 ($\mathrm{LDL}^{\mathrm{T}}$ for semi–positive matrices)

1. *If the dimension of S is 1, define $P = (1)$, $L = (1)$ and $D = S$. Stop.*

2. *Choose $\hat{P}$ as the permutation between 1 and j, where j is such, that $S_{jj} \geq S_{ii}$ holds for all $i = 1, \ldots, N$. Define the $N \times N$ matrix U by*

$$\hat{P}S\hat{P}^\top =: U = \left(\begin{array}{c|c} U_{11} & u^\top \\ \hline u & \tilde{U} \end{array}\right) \tag{2.23}$$

3a. *If $U_{11} > \epsilon$ compute the LDL^{T} decomposition of the $(N-1) \times (N-1)$ matrix $\tilde{U} - \frac{1}{U_{11}} uu^\top$:*

$$\tilde{P}(\tilde{U} - \frac{1}{U_{11}} uu^\top)\tilde{P}^\top = \tilde{L}\tilde{D}\tilde{L}^\top \tag{2.24}$$

Define

$$L := \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{1}{U_{11}}\tilde{P}u & \tilde{L} \end{array}\right), \quad D := \left(\begin{array}{c|c} U_{11} & 0 \\ \hline 0 & \tilde{D} \end{array}\right), \quad P := \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{P} \end{array}\right)\hat{P} \tag{2.25}$$

Stop.

3b. *If $U_{11} \leq \epsilon$, then define $P = \hat{P}$, $D = \mathbf{0}$ and $L = \mathbf{1}$, where $\mathbf{1}$ denotes the unit matrix. Stop.*

From the algorithm 2.3 one can easily obtain the asymptotic computation time which is given by $\frac{1}{6}N^3 + \mathcal{O}(N^2)$ operations, where one operation consists of one multiplication, one subtraction and the expense for the loop management. Note that a matrix multiplication costs $N^3 + \mathcal{O}(N^2)$ of such operations and hence this algorithm is as fast as the usual Cholesky decomposition.

Remark 2.4 *In step 3 the algorithm makes a case distinction whether U_{11} is numerically zero or not. Under the additional setting $\frac{uu^\top}{U_{11}} = 0$ and $\frac{1}{U_{11}}\tilde{P}u = 0$ in the case $U_{11} \approx 0$, the part 3a is equivalent to 3b. You will not often find an equation of the form "Division by 0 = 0", but in the context of Singular Value Decomposition (SVD), one also gets such "equality" [68]. This is a remarkable connection between the SVD and the LDL$^{\mathrm{T}}$ decomposition.*

Remark 2.5 *The LDL$^{\mathrm{T}}$ algorithm can be used for several tasks in linear algebra:*

- *It can be used to determine the rank of a semi–positive symmetric matrix.*
- *One can obtain the pseudo–inverse of S (the inverse if S is positive) by $S^+ = P(L^{-1})^\top \tilde{D} L^{-1} P^\top$ where $\tilde{D}$ is diagonal with $\tilde{D}_{ii} = \frac{1}{D_{11}}$ if $D_{11} > 0$ and $\tilde{D}_{ii} = 0$ otherwise. So it is enough to invert the triangular matrix L.*
- *The determinant of S is given by* $\det S = \prod_{i=1}^{N} D_{ii}$.
- *With $U := DL^\top$ this is the LU decomposition of $PSP^\top$.*
- *Define $C = L\sqrt{D}$, then $PSP^\top = CC^\top$. So the LDL$^{\mathrm{T}}$ algorithm provides a non–square Cholesky decomposition, which is the most significant application of this algorithm. Hence in the following we may call the LDL$^{\mathrm{T}}$ decomposition also a Cholesky decomposition.*

2.2 A Generalized LDL$^{\mathrm{T}}$ based Cholesky Algorithm

For semi–positive matrices the existence of the LDL$^{\mathrm{T}}$ decomposition has been shown and an algorithm to perform this factorization has been presented. In the general case, i.e. for non semi–positive input, the LDL$^{\mathrm{T}}$ decomposition may not exist:

Lemma 2.6 *In general, there is no LDL$^{\mathrm{T}}$ decomposition for symmetric matrices.*

Proof:
Consider the following example:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 1 & 0 \\ l & 1 \end{pmatrix} \cdot \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \cdot \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \tag{2.26}$$

The first column of this matrix equation lead to the following two equations:

$$0 = d_1 \tag{2.27}$$
$$1 = d_1 l \tag{2.28}$$

Obviously, this system has no solution for d_1. □

Due to the non–existence of a general $\mathrm{LDL}^{\mathrm{T}}$ decomposition the generalization of the algorithm is intended in the following manner. If the symmetric input matrix S is semi–positive the usual $\mathrm{LDL}^{\mathrm{T}}$ decomposition shall be performed and if S is not semi–positive, the $\mathrm{LDL}^{\mathrm{T}}$ decomposition of a semi–positive approximation of S shall be returned. The approximation shall be done in such a way, that only off–diagonal elements of S will be changed and therefore it is necessary that the input matrix S has only non–negative diagonal elements, since there are no semi–positive matrices with negative diagonal elements (see lemma 2.1). The corresponding algorithm will be presented and an upper bound for the approximation error will be established.

2.2.1 The Algorithm

Now the generalized Cholesky algorithm is presented, which yields in fact the $\mathrm{LDL}^{\mathrm{T}}$ decomposition, generalized in the manner mentioned before. The idea of the algorithm is based on the boundary for the absolute value of an off–diagonal element of a semi–positive matrix, which is given by lemma 2.1. Therefore one checks at each step if the elements of the first column fulfill this condition and each element which exceeds the boundary will be set to the nearest value in this range.
For the following algorithm, the symmetric $N \times N$ input matrix S must have non–negative diagonal elements. Again, one also has to define an $\epsilon > 0$, for example by (2.22), to decide whether a value is numerical zero or not.

Algorithm 2.7 (Generalized Cholesky Decomposition)

1. *If the dimension of S is 1, define $P = (1)$, $L = (1)$ and $D = S$. Stop.*

2. *Choose $\hat{P}$ as the permutation between 1 and j, where j is such that $S_{jj} \geq S_{ii}$ holds for all $i = 1, \ldots, N$. Define the $N \times N$ matrix U by*

$$\hat{P} S \hat{P}^{\top} \quad =: \quad U = \left(\begin{array}{c|c} U_{11} & u^{\top} \\ \hline u & \tilde{U} \end{array}\right) \tag{2.29}$$

3. *Rescale the vector u, that is define the vector $\hat{u}$:*

$$\hat{u}_j \quad := \quad \mathrm{sgn}(u_j) \min\left(|u_j|, \sqrt{U_{11} U_{jj}}\right) \tag{2.30}$$

where $\mathrm{sgn}(x) = 1$ *iff* $x > 0$, $\mathrm{sgn}(x) = -1$ *iff* $x < 0$ *and* $\mathrm{sgn}(x) = 0$ *iff* $x = 0$.

4a. *If $U_{11} > \epsilon$ compute the LDL^{T} decomposition of the $(N-1) \times (N-1)$ matrix $\tilde{U} - \frac{1}{U_{11}} \hat{u}\hat{u}^{\top}$:*

$$\tilde{P}(\tilde{U} - \frac{1}{U_{11}} \hat{u}\hat{u}^{\top})\tilde{P}^{\top} = \tilde{L}\tilde{D}\tilde{L} \tag{2.31}$$

Define

$$L := \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{1}{U_{11}} \tilde{P}\hat{u} & \tilde{L} \end{array}\right) \quad D := \left(\begin{array}{c|c} U_{11} & 0 \\ \hline 0 & \tilde{D} \end{array}\right) \quad P := \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{P} \end{array}\right) \hat{P} \tag{2.32}$$

Stop.

4b. *If $U_{11} \leq \epsilon$, then define $P = \hat{P}$, $D = \mathbf{0}$ and $L = \mathbf{1}$.*

If the input matrix S is semi–positive, one would never detect a violation of the boundary for an off–diagonal element and in this situation step 3 of the algorithm is given by $\hat{u} := u$. Hence the algorithm yields the $\mathrm{LDL}^{\mathrm{T}}$ decomposition for semi–positive input. Since only off–diagonal elements are changed by the rescaling (step 3), the diagonal elements are unchanged. Note that this rescaling guarantees, that the diagonal elements of the input matrix in the next recursion step remain non–negative. Since the computational effort of the rescaling technique is of order $\mathcal{O}(N)$ at each step, its total costs amount to $\mathcal{O}(N^2)$, hence it does not effect the asymptotic computation time. The total computational effort of the algorithm 2.7 is again given by $\frac{1}{6}N^3 + \mathcal{O}(N^2)$ operations, where one operation consists of one multiplication, one subtraction and the expense for the loop management.

2.2.2 On the Bound of the Approximation error

In the case of a semi–positive input, the algorithm 2.7 yields an exact decomposition. Only for non–semi–positive inputs there is the need to discuss about an approximation error. So let us introduce the negativity of a matrix, which is the important value in this discussion.

Definition 2.8 (Negativity of a square matrix) *Let A be a $N \times N$ matrix. The negativity of A is defined by*

$$\Lambda(A) \;:=\; \min\{\lambda \geq 0 \mid A + \lambda \mathbf{1} \textit{ is semi–positive}\} \tag{2.33}$$

The negativity is a non–negative, positive homogeneous and sub–additive functional on square matrices; i.e. let A and B be $N \times N$ matrices and $\alpha > 0$, then holds $\Lambda(A) \geq 0$, $\Lambda(\alpha A) = \alpha \Lambda(A)$ and $\Lambda(A+B) \leq \Lambda(A) + \Lambda(B)$. Obviously, the negativity of any semi–positive matrix is 0 and if a symmetric matrix S is not semi–positive, then $\Lambda(S)$ is the absolute value of the smallest (most negative) eigenvalue of S.
Since the approximation error of the algorithm 2.7 is 0 for semi–positive matrices, one might expect that the approximation error is bounded in some way by the negativity. Indeed, this is the statement of the

Theorem 2.9 *Let S be a symmetric $N \times N$ matrix with non–negative diagonal elements. The error due to a semi–positive approximation by algorithm 2.7 is bounded by the negativity $\lambda = \Lambda(S)$ of the input matrix S in the sense, that for the approximation error and the negativity of the remaining matrix holds:*

$$\left|\left| \left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array}\right) - \left(\begin{array}{c|c} S_{11} & \hat{s}^\top \\ \hline \hat{s} & \tilde{S} \end{array}\right) \right|\right|_F^2 \;\leq\; 2\lambda(N-1)(2S_{11}+\lambda) \tag{2.34}$$

$$\Lambda\left(\left(\begin{array}{c|c} S_{11} & \hat{s}^\top \\ \hline \hat{s} & \tilde{S} \end{array}\right)\right) \;\leq\; \lambda + \sqrt{\lambda(N-1)(2S_{11}+\lambda)} \tag{2.35}$$

$$\Lambda\left(\tilde{S} - \frac{1}{S_{11}}\hat{s}\hat{s}^\top\right) \;\leq\; \lambda N \qquad \textit{if } S_{11} > 0 \tag{2.36}$$

where S_{11} is the largest diagonal element of S, $\hat{s}$ is the rescaled row and $||A||_F$ denotes the Frobenius norm of a matrix A, which is defined by $\sqrt{\sum_{ij} A_{ij}^2}$.

Before this theorem will be proven, some basic relations for general symmetric matrices are stated in the next lemma.

Lemma 2.10 *Let S be a symmetric matrix with the usual notation*

$$S = \left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array}\right) \tag{2.37}$$

and let $\lambda \in \mathbb{R}$. Then holds: $S + \lambda\mathbf{1}$ is semi–positive if and only if

- $S_{11} + \lambda \geq 0$.
- $s_i^2 \leq (S_{11} + \lambda)(\tilde{S}_{ii} + \lambda)$.
- *The matrix*

$$\begin{cases} \tilde{S} - \frac{ss^\top}{S_{11}+\lambda} + \lambda\mathbf{1} & \text{if } S_{11} + \lambda > 0 \\ \tilde{S} + \lambda\mathbf{1} & \text{if } S_{11} + \lambda = 0 \end{cases}$$

  is semi–positive.

Proof:
If $S + \lambda\mathbf{1}$ is semi–positive, apply lemma 2.1 on this matrix and the three statements follow directly.
To show the other direction, first assume that $S_{11} + \lambda > 0$. Define $x^\top = (\alpha, y)$ with arbitrary $\alpha \in \mathbb{R}$ and $y \in \mathbb{R}^{N-1}$. Then holds:

$$\begin{aligned} x^\top(S + \lambda\mathbf{1})x &= \begin{pmatrix} \alpha \\ y \end{pmatrix}^\top \left(\begin{array}{c|c} S_{11} + \lambda & s^\top \\ \hline s & \tilde{S} + \lambda\mathbf{1} \end{array}\right) \begin{pmatrix} \alpha \\ y \end{pmatrix} \\ &= \alpha^2(S_{11} + \lambda) + 2\alpha s^\top y + y^\top(\tilde{S} + \lambda\mathbf{1})y \qquad (2.38) \\ &= (S_{11} + \lambda)(\alpha + \frac{s^\top y}{S_{11} + \lambda})^2 + y^\top \left(\tilde{S} - \frac{ss^\top}{S_{11} + \lambda} + \lambda\mathbf{1}\right) y \qquad (2.39) \\ &\geq 0 \qquad (2.40) \end{aligned}$$

In the case that $S_{11} + \lambda = 0$, one concludes by the second condition that $s_i = 0$. Since $\tilde{S} + \lambda\mathbf{1}$ is semi–positive by the third condition, the matrix $S + \lambda\mathbf{1}$ is semi–positive too. □

Proof of Theorem 2.9:
Let λ be the negativity of the matrix S. $\hat{s}_i$ differs from s_i only if a rescaling has been done and in this case one concludes via lemma 2.10:

$$\left(s_i - \mathrm{sgn}(s_i)\sqrt{S_{11}\tilde{S}_{ii}}\right)^2 \leq \lambda(S_{11} + S_{ii}) + \lambda^2 \tag{2.41}$$

Since this holds for all i, we obtain the important inequality:

$$||s - \hat{s}||^2 \leq \lambda(N-1)(2S_{11} + \lambda) \tag{2.42}$$

The approximation error (2.34):
The left hand side of this equation is $2||s-\hat{s}||^2$, hence the statement follows directly from equation (2.42).
The rescaling impact on the negativity (2.35):
Note that

$$\Lambda\left(\left(\begin{array}{c|c} 0 & b^\top \\ \hline b & 0 \end{array}\right)\right) = ||b|| \tag{2.43}$$

where $||b||$ denotes the Euklidian norm of the vector b, since the eigenvalues of this matrix are given by $\pm||b||$ each with multiplicity 1, and by 0 with multiplicity $N-2$. Together with the sub–additivity of Λ one thus obtains:

$$\begin{aligned}
\Lambda\left(\left(\begin{array}{c|c} S_{11} & (\hat{s}-s+s)^\top \\ \hline \hat{s}-s+s & \tilde{S} \end{array}\right)\right) &\leq \Lambda\left(\left(\begin{array}{c|c} S_{11} & s^\top \\ \hline s & \tilde{S} \end{array}\right)\right) \\
&\quad +\Lambda\left(\left(\begin{array}{c|c} 0 & (\hat{s}-s)^\top \\ \hline \hat{s}-s & 0 \end{array}\right)\right) && (2.44)\\
&\leq \lambda + ||\hat{s}-s|| && (2.45)\\
&\leq \lambda + \sqrt{\lambda(N-1)(2S_{11}+\lambda)} && (2.46)
\end{aligned}$$

The recursion impact on the negativity (2.36):
From the Cauchy Schwarz inequality one concludes, that the matrix $||s||\mathbf{1} - ss^\top$ is semi–positive:

$$y^\top(||s||\mathbf{1} - ss^\top)y = ||s||^2||y||^2 - (s^\top y)^2 \geq 0 \tag{2.47}$$

Since S_{11} is the largest diagonal element, it holds $|\hat{s}_i| \leq S_{11}$ and hence the matrix $(N-1)S_{11}^2\mathbf{1} - ss^\top$ is semi–positive. To prove the last statement of the theorem, one has to show, that the matrix $\tilde{S} - \frac{\hat{s}\hat{s}^\top}{S_{11}} + N\lambda\mathbf{1}$ is semi–positive. In fact it is, since for all y holds:

$$\begin{aligned}
y^\top\left(\tilde{S} - \frac{\hat{s}\hat{s}^\top}{S_{11}} + N\lambda\mathbf{1}\right)y &\geq y^\top\left(\tilde{S} - \frac{\hat{s}\hat{s}^\top}{S_{11}+\lambda} + \lambda\mathbf{1}\right)y \\
&\quad +\lambda y^\top\left((N-1)\mathbf{1} - \frac{\hat{s}\hat{s}^\top}{S_{11}^2}\right)y \geq 0 && (2.48)
\end{aligned}$$

□

2.3 Comparison with known LDL$^{\mathrm{T}}$ based Approximation Algorithms

There are two other approaches of approximating a symmetric indefinite matrix by a positive symmetric matrix based on the LDL$^{\mathrm{T}}$ decomposition. The first method (**GMW**) is the algorithm by Gill, Murray and Wright [39], which comes from the context of numerical optimization. The idea is to choose a $\epsilon > 0$ and perform the usual LDL$^{\mathrm{T}}$ decomposition step until a diagonal element is smaller than ϵ. Once

a diagonal element less than ϵ occurs, it is set to ϵ and then the LDL^T algorithm proceeds. The value of ϵ has to be carefully chosen to keep the algorithm numerical stable.
In [73] Eskow and Schnabel give an improvement (**ES**) of the **GMW** algorithm. They also gave an error bound for the approximation error with respect to the spectral norm $||S - S'||_2$ using the Gerschgorin circle theorem (see e.g. [14]). In [74] Schnabel and Eskow enhanced their algorithm, which could run into computational difficulties with the effect, that the approximation error was much bigger in comparison to the error made by the algorithm **GMW**.
Both algorithms have the basic idea to approximate with respect to the spectral norm[2] and therefore to change (increase) only the diagonal elements, if necessary. In opposite to algorithm **2.7**, the algorithms **GMW** and **ES** return the LDL^T decomposition of a positive approximation if the input matrix is semi–positive but not strict positive, so they are not able to decompose rank deficient semi–positive matrices. Another difference between **GMW** and **ES** on the one hand and **2.7** on the other hand consists in the elements changed in the case of an approximation, since **GMW** and **ES** increase only diagonal elements and **2.7** changes off–diagonal elements only. To compare the different algorithms consider the same example which has been analyzed in [39] and [73]:

$$S = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 1 \end{pmatrix} \tag{2.49}$$

S is indefinite since its eigenvalues are given by (5.113, 0.089, -2.202). Let us apply the algorithm **2.7** on S and compare the result with the results from the algorithms **GMW** and **ES**. One obtains the following approximations:

$$S_{\mathbf{GMW}} = \begin{pmatrix} 3.771 & 1.000 & 2.000 \\ 1.000 & 6.015 & 3.000 \\ 2.000 & 3.000 & 3.242 \end{pmatrix} \qquad S_{\mathbf{ES}} = \begin{pmatrix} 3.000 & 1.000 & 2.000 \\ 1.000 & 3.220 & 3.000 \\ 2.000 & 3.000 & 3.220 \end{pmatrix}$$

$$S_{\mathbf{2.7}} = \begin{pmatrix} 1.000 & 1.000 & 1.000 \\ 1.000 & 1.000 & 1.000 \\ 1.000 & 1.000 & 1.000 \end{pmatrix}$$

In the following table the approximation errors of these algorithms with respect to several matrix norms are listed[3]:

Norm	**GMW**	**ES**	**2.7**
$\|\|S - S_*\|\|_2$	5.105	2.220	2.236
$\|\|S - S_*\|\|_{1,\infty}$	5.105	2.220	3.000
$\|\|S - S_*\|\|_F$	6.153	3.722	3.162

[2]The spectral norm of a matrix S is given by $||S||_2 := \sqrt{\text{largest eigenvalue of } S^\top S}$.
[3]The definition of the 1, ∞ norm is given by $||S||_1 := \max_j \sum_i |S_{ij}|$ and $||S||_\infty := \max_i \sum_j |S_{ij}|$.
Since S is symmetric in this case, it obviously holds $||S||_1 = ||S||_\infty$ what explains the notation $||S||_{1,\infty}$ in the table. S_* stands for the result of the corresponding algorithm.

In this example, the algorithm **GMW** is the worst one with respect to all three norms. The algorithm **ES** outperforms **2.7** with respect to the 1-norm, but on the other hand the new algorithm is better with respect to the Frobenius norm. Note that the algorithm **2.7** gives a similar spectral norm error like the algorithm **ES**, which is designed to minimize the approximation error with respect to the spectral norm.
Another great benefit from the algorithm **2.7** is, that the corresponding $\mathrm{LDL}^{\mathrm{T}}$ decomposition gives a dimension reduction. So from the algorithm **2.7** one obtains:

$$L_{\mathbf{2.7}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad D_{\mathbf{2.7}} = (1) \tag{2.50}$$

Note that although $S_{\mathbf{2.7}}$ is degenerate, it is in this example the best possible approximation with respect to the Frobenius norm under the condition, that the diagonal elements have to be preserved.
The property of the algorithm **2.7** that the diagonal elements will not be changed, is quite important for some applications, one may think of correlation matrices, for example. In such situations the algorithms **GMW** and **ES** cannot be applied. Of course, there are more techniques — thereunder diagonal preserving methods — to approximate an indefinite symmetric matrix by a semi–positive one. Some common methods are the subject of the next section.

2.4 Standard Techniques to obtain semi–positive Approximations

In practice, one often knows, that a given matrix must be semi–positive by theory, but due to numerical roundoffs or estimation errors one gets an indefinite, symmetric matrix. In this section some known standard techniques to obtain semi–positive approximations of symmetric matrices are recalled. There are some methods which yield the best approximation with respect to a certain matrix norm without any further constraint. Additional to this approach there are techniques which deal with the auxiliary condition, that the matrix and its semi–positive approximation must be unit–diagonal. The standard methods in this context are presented.

2.4.1 Minimal Approximations

Best Approximation in the Frobenius Norm

The most familiar approach to get a semi–definite approximation of an indefinite symmetric matrix S is given by a spectral decomposition. Let

$$S = Q^{\top} D Q \tag{2.51}$$

with Q orthogonal and D diagonal. Define D^* and $S_{\mathbf{FN}}$ by

$$D^* := \mathrm{diag}\left(\max(0, D_{11}), \ldots, \max(0, D_{NN})\right) \tag{2.52}$$

$$S_{\mathbf{FN}} := Q^{\top} D^* Q \tag{2.53}$$

then the following statement holds:

Theorem 2.11 *$S_{\mathbf{FN}}$ is the unique best semi–positive approximant of S with respect to the Frobenius norm, i.e. for all positive symmetric $S' \neq S$ holds:*

$$||S_{FN} - S||_F < ||S' - S||_F \tag{2.54}$$

For the proof see N.J. Higham [52], theorem 2.1. The computation of this approximant costs approximately $5N^3$ operations due to the computation of the eigenvalues and eigenvectors, where one operation consists of one multiplication, one addition and the loop management. Recall that in this respect the multiplication of two square $N \times N$ matrices cost $1 \cdot N^3$ operations.

Best Approximation in the spectral norm

Properties of an optimal semi–positive approximation with respect to the spectral norm can be found in Halmos [47], Higham [52].

Theorem 2.12 *Let S be a symmetric matrix and $\lambda = \Lambda(S)$ its negativity. Then the matrix*

$$S_{\mathbf{SN}} \quad := \quad S + \lambda \mathbf{1} \tag{2.55}$$

is a best semi–positive approximation of S with respect to the spectral norm with

$$||S_{\mathbf{SN}} - S||_2 = \lambda \tag{2.56}$$

A straight forward way to detect the negativity is to perform an eigenvalue decomposition. On the other hand, one could also use some bisection algorithm to detect λ. For each steps one must decide, whether $S + \lambda\mathbf{1}$ is semi–positive or not and the $\mathrm{LDL}^{\mathrm{T}}$ algorithm can be used for this task. For the details see again the paper of Higham, where also an upper bound for λ is given, which may be used to start the bisection approach.
The optimal approximation of S with respect to the spectral norm is not unique, what can be seen by the following example. Let

$$S = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}, \qquad S_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad S_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{2.57}$$

then S_1 is the approximant according to the theorem. On the other hand holds:

$$||S_1 - S'||_2 = ||S_2 - S'||_2 = 2 \tag{2.58}$$

2.4.2 Unit–diagonal Approximations

If the symmetric matrix S is for example a perturbed correlation matrix, hence $S_{ii} = 1$, one likes to get a semi–positive approximation S^* with $S^*_{ii} = 1$. Some standard algorithms to get such approximations, which may be quite important in some situations, are presented.

Hypersphere Decomposition

Let the symmetric $N \times N$ matrix S be unit diagonal. The aim is to find a semi–positive, unit–diagonal approximation of S. Since any semi–positive matrix S^* can be decomposed into $BB^\top$ by an lower triangular matrix B, it is enough to find the appropriate B to approximate S. The condition, that $S^*_{ii} = 1$ implies, that the rows of B contain unit vectors. An elegant method by Rousseeuw and Molenberghs [71] to describe the matrix B is to use angular coordinates $\Theta_{i,j}$:

$$B_{ij} = \begin{cases} 0 & \text{if } i < j \\ 1 & \text{if } i = j = 1 \\ \cos\Theta_{i,1} & \text{if } i = j > 1 \\ \left(\prod_{l=1}^{i-j} \sin\Theta_{i,l}\right) \cos\Theta_{i,i-j+1} & \text{if } i > j > 1 \\ \prod_{l=1}^{i-j} \sin\Theta_{i,l} & \text{if } i > 1 \text{ and } j = 1 \end{cases} \tag{2.59}$$

So for the ith row one needs $i-1$ angles to describe this row and the whole matrix B is described by $\frac{(N-1)N}{2}$ angles. Using this parametrization it is convenient to define an error measure, for example by the Frobenius norm:

$$f(\Theta) := \sum_{i,j=1}^{N} (S^*_{ij}(\Theta) - S_{ij})^2 \tag{2.60}$$

This method provides a way to find the best unit–diagonal, semi–positive approximation of S with respect to the Frobenius norm, since one can find the optimal angular combination by an unconstrained minimum search using the steepest descend method or even more sophisticated methods like conjugate gradients.
Rebonato and Jäckel pointed out in [69], that the hypersphere decomposition approach has the advantage that one may use other error measures than (2.60). But for the Frobenius norm it is known, that the best approximation is unique and therefore the minimum search will be stable.
However, one has to calibrate $\frac{(N-1)N}{2}$ parameters and for each iteration of the minimizing routine one has to evaluate a matrix product. So one minimizes in a $\mathcal{O}(N^2)$ dimensional space and the evaluation of the function costs about $\mathcal{O}(N^3)$. Therefore the computation time of the hypersphere decomposition algorithm is of $\mathcal{O}(N^5)$ in general, so it is a rather costly procedure for larger dimensions.
Therefore some other techniques have been developed, which do not aim an optimal solution of the problem. The condition "$||S - S^*|| = $ minimal" is replaced by the relaxed condition "$||S - S^*|| = $ small" and these methods can thus be seen as *quick and dirty* algorithms. The *rescaling* approach uses any semi–positive approximation, which may be not unit–diagonal, and rescales this approximation such that it becomes unit–diagonal. The *linear shrinking* yields an approximation by a convex combination of the unit–diagonal matrix and the unit matrix.

Rescaling of a semi–positive Approximation

Let S' be a semi–positive approximation of the symmetric matrix S, which is not unit–diagonal. Such approximation could be obtained for example by spectral de-

composition (2.53) or using theorem 2.12. In a next step one can perform the following rescaling procedure, to obtain a semi–positive, unit–diagonal approximation S^*:

$$S^*_{ij} := \begin{cases} \frac{S'_{ij}}{\sqrt{S'_{ii}S'_{jj}}} & \text{if } S'_{ii}S'_{jj} > 0 \\ 1 & \text{if } i = j \text{ and } S_{ii} = 0 \\ 0 & \text{else} \end{cases} \tag{2.61}$$

This rescaling technique needs $\mathcal{O}(N^2)$ operations and the total computation time of this technique depends on the costs to obtain S'. Mostly, this technique is applied on the best approximation with respect to the Frobenius norm $S_{\mathbf{FN}}$ and one obtains a unit–diagonal, semi–positive approximation based on spectral decomposition by this rescaling technique.

Linear Shrinking Approach

This method and some generalizations can be found in [26]. For a unit–diagonal symmetric matrix S one defines

$$\mu := \max\{m \in [0,1] \mid mS + (1-m)\mathbf{1} \text{ is semi–positive}\} \tag{2.62}$$

and a unit–diagonal, semi–positive approximation of S is given by

$$S_{\mathbf{LS}} := \mu S + (1-\mu)\mathbf{1} \tag{2.63}$$

where μ is determined by the negativity of S and is given by $\mu = \frac{1}{1+\Lambda(S)}$. To obtain the negativity of a symmetric matrix S one could compute all eigenvalues of S. If one proceeds so, the asymptotic computation time is given by $\frac{2}{3}N^3$ operations due to the Householder reduction, since performing the QR algorithm to diagonalize a tridiagonal matrix costs only $\mathcal{O}(N^2)$ operations if one does not compute the eigenvectors.

2.5 Application to perturbed correlation matrices

Consider a perturbed correlation matrix S, hence a unit–diagonal symmetric matrix. Under some circumstances this matrix may have negative eigenvalues. In this case one would prefer to approximate this matrix by a semi–positive unit–diagonal matrix [71].
There are three algorithms with an computational effort of $\mathcal{O}(N^3)$ which can be used to find a desired approximation. Let S be a perturbed correlation matrix. Then one can perform a spectral decomposition and can rescale the thus obtained best approximation with respect to the Frobenius norm to find the first approximation $S_{\mathbf{RSD}}$. The second method is to use the linear shrinking approach which yields $S_{\mathbf{LS}}$ and thirdly one can apply the algorithm 2.7 on S to determine the approximation $S_{\mathbf{2.7}}$, which is unit–diagonal too.

2.5.1 Comparison of three diagonal preserving algorithms

Let S' be a (randomly chosen) correlation matrix. Dice an $\epsilon \in [0, 0.2]$ and a matrix T by the law

$$T_{ij} := \begin{cases} 0 & \text{if } i = j \\ \pm 1 \text{ each with probability} \frac{1}{2} & \text{if } i > j \\ T_{ji} & \text{if } i < j \end{cases} \tag{2.64}$$

Then one may define a perturbed correlation matrix by $S := S' + \epsilon T$. The three algorithms are applied on each such diced 15×15 matrices S to recover S' and the approximations $S_{\mathbf{RSD}}$, $S_{\mathbf{LS}}$ and $S_{\mathbf{2.7}}$ are obtained.

In figure 2.1, the deviation with respect to the Frobenius norm between the results of the algorithms and the original correlation matrix S' is ploted against the negativity of S for 200 diced matrices:

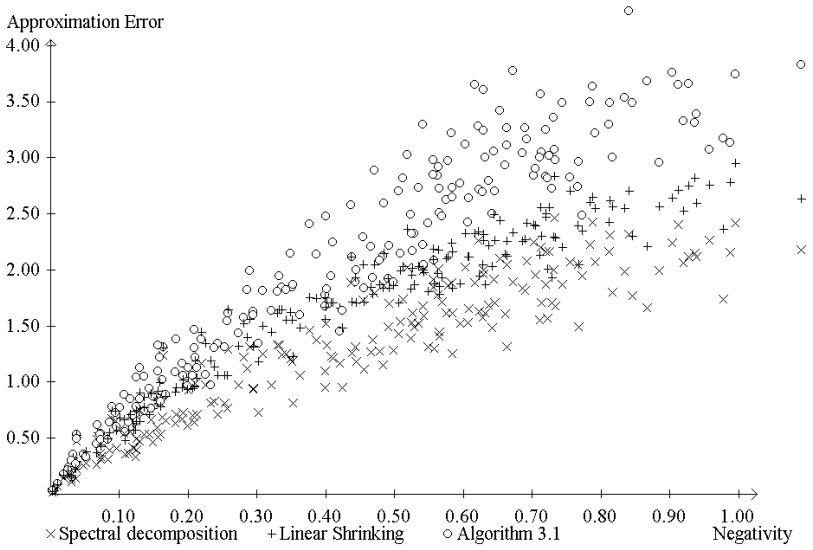

Figure 2.1: The approximation errors of the three algorithms with respect to the negativity of the perturbed matrix. In average, the Spectral Decomposition works best, the Linear Shrinking method is medium and the algorithm 2.7 yields the worst approximations.

For some applications the rank of the approximation may influence the computational effort of following tasks and a approximation with a lower rank could help to save computation time in such situations [89]. So for each of the matrices studied above, the rank of the approximation is plotted in dependence of the negativity of the input matrix. The result is shown in figure 2.2.

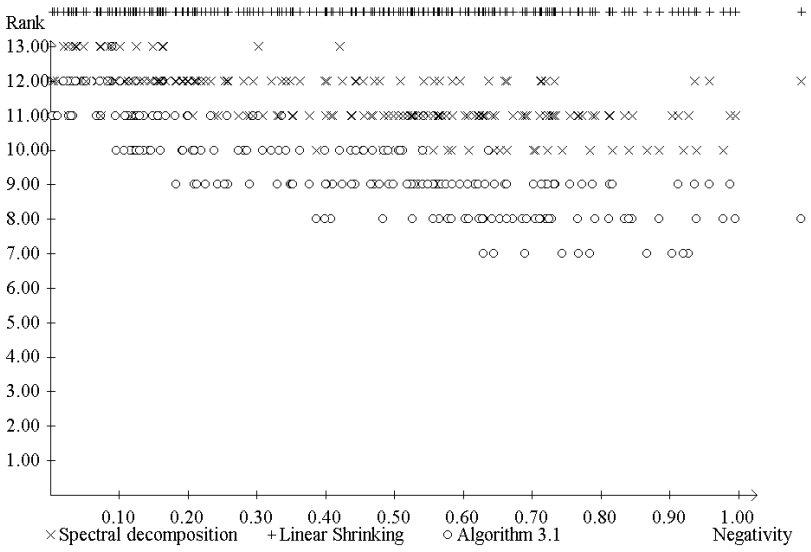

Figure 2.2: The Linear Shrinking method yields almost no dimension reduction, the Spectral Decomposition algorithm allows a medium rank reduction. The highest rank reduction is performed by the algorithm 2.7. The higher the negativity of S, the higher is the resulting rank reduction in average.

2.6 Conclusion

A generalized Cholesky decomposition algorithm based on the LDL^T factorization has been introduced. This algorithm can be used to decompose even semi–positive and rank deficient symmetric matrices, a situation in which the usual Cholesky algorithm fails. The asymptotical computational effort is identical to effort of the usual Cholesky decomposition algorithm and hence the use of the generalized Cholesky algorithm is recommended.

The presented algorithm 2.7 can also be used to find a semi–positive approximation of a symmetric matrix with non–negative diagonal elements. The returned approximation is a semi–positive matrix with an unchanged diagonal, and in general the approximant is rank deficient. The property, that the diagonal elements of the input matrix remain unchanged in the case of an approximation is quite important in the study of perturbed correlation matrices. Hence a new approach in this context has been established and the generalized Cholesky algorithm has been compared with two other standard approaches. The results of this comparison are listed in the following table.

	RSD	Linear Shrinking	Algorithm 2.7
Asymptotic computation time in matrix multiplications	≈ 5	$\frac{2}{3}$	$\frac{1}{6}$
Rank reduction	medium	none	best
Approximation error in average	smallest	medium	largest
Optimal use	if the approximation error is most critical.	A compromise between computation time and approximation error. In situations where the computation time of further processing depends on the rank of the approximant, this method shall not be used.	if computation time is most critical, especially if the rank reduction can be used to save time in further processing.

The feature, that the generalized Cholesky algorithm may return a non–square decomposition can be used to save computation time in several applications. In the context of Monte Carlo simulations for example, there may be the task to sample a N dimensional vector of multivariate normals with a given correlation matrix S. If the rank of S is $M < N$, then one only needs to generate M independent normal random numbers to generate one such N–dimensional vector per Monte Carlo iteration.

In applications for risk management, one can expect a large dimension reduction in general. In practice, there are round about 1000 risk factors like stock prices or foreign exchange rates which has to be taken into account. Since the dependence of this risk factors is often described by a correlation matrix based on the estimation of a one year time series (approximately 250 trading days) and the rank of the correlation matrix can not exceed the length of the time series, a serious dimension reduction can be achieved. As it is described in chapter 4, the computational effort of the Delta Gamma normal approximation to assess market risk partly consists of an eigenvalue decomposition of a symmetric matrix, whose dimension is given by the rank of the correlation matrix. In this respect, the generalized Cholesky decomposition algorithm improves the total computation time of the Delta Gamma normal method considerably.

Chapter 3

Fourier Inversion of Characteristic Functions[1]

Usually, the distribution of a real–valued random variable is given by its distribution function, or, if the distribution function is differentiable the distribution is described by the probability density. But the distribution of a real–valued random variable can also be described by the characteristic function. If the distribution has a density, then the corresponding characteristic function is the Fourier transform of the density. In any case, the characteristic function can be expressed by an expectation.

In this chapter, the inverse mapping is studied. If the characteristic function is integrable, then the corresponding distribution has a density which can be obtained by the Fourier inversion integral. But characteristic must not be integrable and in such situations the Fourier inversion integral does not exist. Also in this case, the distribution is uniquely described by the characteristic function and several methods to determine the distribution from the characteristic function are presented.

Two methods are presented where a Fourier inversion of the difference between two distributions can be performed. These techniques are based on the two theorems presented in section 3.3 and 3.4. For practical reasons, one usually compares an unknown distribution with a normal or the Gamma distribution. If also the first two moments of the unknown distribution are known, the comparison should be done with a distribution which has the same expectation and variance. Less theoretical, but quite interesting from a practitioners point of view, the Fourier inversion of a (not necessary integrable) characteristic function by the Fast Fourier Transform (FFT) is studied and yields an approximating density of the unknown distribution.

The presented techniques are quite important in financial mathematics, examples are given in chapters 4 and 5. In these examples, the use of the characteristic function is the key for successful computations. Since this work does not aim only mathematicians, but also practitioners with a less mathematical background who want to implement efficient methods for risk management, the basic properties of the Fourier transform and characteristic functions are recalled. For this reason also proofs of the basic properties are given, although they can be found in several textbooks, for example in [35].

[1] Parts of this chapter have been presented in [90] and will be published in [91].

3.1 A Survey on the Fourier Transformation

3.1.1 Basics on the Fourier Transformation

Definition 3.1 *A function $f : \mathbb{R} \to \mathbb{R}$ is called $\mathbf{L}^1$, if*

$$\int_{-\infty}^{\infty} |f(x)|\, dx < \infty \tag{3.1}$$

The set of integrable functions is the domain of the Fourier transformation, which is defined as follows.

Definition 3.2 *Let f be $\mathbf{L}^1$. Then the Fourier transform of f exists and is defined by:*

$$(\mathcal{F}f)(s) \;:=\; \int_{-\infty}^{\infty} e^{isx} f(x)\, dx \tag{3.2}$$

Lemma 3.3 *Let $f \in \mathbf{L}^1$. Then the Fourier transform of f is bounded:*

$$|(\mathcal{F}f)(s)| \;\leq\; \int_{-\infty}^{\infty} f(x)\, dx \qquad\qquad \forall\ s \tag{3.3}$$

Proof:

$$|(\mathcal{F}f)(s)| \;=\; \left| \int_{-\infty}^{\infty} e^{isx} f(x) dx \right| \leq \int_{-\infty}^{\infty} |e^{isx}||f(x)| dx = \int_{-\infty}^{\infty} |f(x)| dx \tag{3.4}$$

□

Lemma 3.4 *Let f be $\mathbf{L}^1$. Then its Fourier transform $\mathcal{F}f$ is continuous.*

Proof:
Let $\epsilon > 0$. Then holds

$$|(\mathcal{F}f)(s+\epsilon) - (\mathcal{F}f)(s)| \;\leq\; \left| \int_{-\infty}^{\infty} f(x)(e^{i(s+\epsilon)x} - e^{isx})\, dx \right| \tag{3.5}$$

$$\leq\; 2 \int_{\mathbb{R}\setminus[-\frac{1}{\sqrt{\epsilon}}, \frac{1}{\sqrt{\epsilon}}]} |f(x)|\, dx + \sqrt{\epsilon} \int_{-\frac{1}{\sqrt{\epsilon}}}^{\frac{1}{\sqrt{\epsilon}}} |f(x)|\, dx \tag{3.6}$$

since $|e^{ix} - 1| \leq |x|$ for all $x > 0$. In the limit $\epsilon \to 0$, the last expression tends to zero because f is integrable.

□

Lemma 3.5 (Convolution Property) *Let f, g be $\mathbf{L}^1$ and such, that the function*

$$h(x) \;:=\; \int_{-\infty}^{\infty} f(y)g(x-y)\,dy \tag{3.7}$$

is well defined and $h \in \mathbf{L}^1$. Then holds:

$$(\mathcal{F}h)(s) \;=\; (\mathcal{F}f)(s)\cdot(\mathcal{F}g)(s) \tag{3.8}$$

Proof:

$$(\mathcal{F}h)(s) \;=\; \int_{-\infty}^{\infty} dx\, e^{isx} \int_{-\infty}^{\infty} dy\, f(y)g(x-y) = \int_{-\infty}^{\infty} dy f(y) \int_{-\infty}^{\infty} dx e^{isx} g(x-y) \tag{3.9}$$

$$= \int_{-\infty}^{\infty} dy f(y) e^{isy} (\mathcal{F}g)(s) = (\mathcal{F}f)(s)\cdot(\mathcal{F}g)(s) \tag{3.10}$$

□

Lemma 3.6 *For all $f, g \in \mathbf{L}^1$, the* Parseval relation *holds:*

$$\int_{-\infty}^{\infty} g(s)e^{-its}(\mathcal{F}f)(s)\,ds \;=\; \int_{-\infty}^{\infty} f(x)(\mathcal{F}g)(x-t)\,dx \tag{3.11}$$

Proof:
Multiply the definition of the Fourier transform by $g(s)e^{-its}$ and integrate with respect to s. Since $f, g \in \mathbf{L}^1$ and their Fourier transforms are bounded, both integrals exists.
□

The Parseval relation seems to be quite unimpressive, but it is a powerful tool to prove some results on the inverse of the Fourier transformation, which is studied next.

3.1.2 The Inverse of the Fourier Transformation

Lemma 3.7 *Let $f, g \in \mathbf{L}^1$. Then the Fourier transforms of f and g are equal if and only if*

$$\int_{-\infty}^{\infty} |f(x) - g(x)|\,dx = 0 \tag{3.12}$$

Proof:
By the Parseval relation holds for any $h \in \mathbf{L}^1$:

$$\int_{-\infty}^{\infty} h(s)e^{-its}(\mathcal{F}(f-g))(s)\,ds \;=\; \int_{-\infty}^{\infty} (f(x)-g(x))(\mathcal{F}h)(x-t)\,dx \tag{3.13}$$

If f, g are equal with respect to the $\mathbf{L}^1$ norm, the right hand side is zero for all h, hence $\mathcal{F}(f-g)$ has a zero $\mathbf{L}^1$ norm. Because $\mathcal{F}f$ and $\mathcal{F}g$ are continuous, they are equal. On the other hand, if $\mathcal{F}f \equiv \mathcal{F}g$, $f-g$ must have a vanishing $\mathbf{L}^1$ norm. □

By this lemma it is obvious, that the Fourier transform is invertible. This statements holds in general and in certain situations an explicit inversion formula can be given. This is the subject of the next theorem, which is the most important theorem in the context of the Fourier transform.

Theorem 3.8 (Fourier inversion) *Let the continuous function f be $\mathbf{L}^1$ and let its Fourier transform $(\mathcal{F}f)$ be $\mathbf{L}^1$ too. Then holds:*

$$f(x) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}(\mathcal{F}f)(s)\,ds \tag{3.14}$$

Proof:
Let $\beta > 0$ and define $g(x) = \frac{\beta}{\sqrt{2\pi}} e^{-\frac{\beta^2 x^2}{2}}$, which is the density of the normal distribution with mean 0 and variance $1/\beta^2$. This is an integrable function and its Fourier transform is given by $(\mathcal{F}g)(s) = e^{-\frac{s^2}{2\beta^2}}$. Plug these two functions into the Parseval relation, divide by $\sqrt{2\pi\beta^2}$ and obtain:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{\beta^2 s^2}{2}} e^{-its}(\mathcal{F}f)(s)\,ds \;=\; \int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x-t)^2}{2\beta^2}}\,dx \tag{3.15}$$

Taking the limit $\beta \to 0$ yields the Fourier inversion formula, since $\lim_{\beta\to 0} e^{-\frac{\beta^2 s^2}{2}} = 1$ and on the right hand side, the integral of f with the normal density with mean t and vanishing variance β^2 yields $f(t)$. □

Remark 3.9 *In this theorem, it was assumed that f is continuous. This condition may be relaxed and f may have a countable number of discontinuities, such that the left and the right limit exists. Then holds if f and $\mathcal{F}f$ are $\mathbf{L}^1$:*

$$\frac{f(x-)+f(x+)}{2} \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}(\mathcal{F}f)(s)\,ds \tag{3.16}$$

Hence the Fourier inversion integral yields not the function f, but it returns a function $\tilde{f}$ such that the difference between f and $\tilde{f}$ with respect to the $\mathbf{L}^1$–norm vanishes and that the function $\tilde{f}$ fulfills the Dirichlet conditions *on any finite interval.*

So the inversion of the Fourier transformation is – up to a factor and a sign – the Fourier transformation itself, if the Fourier transform of f is integrable. But $\mathcal{F}f$ may not be integrable and in such situations the Fourier inversion integral does not exist. This problem is the mathematical center of this chapter.

3.1.3 The Tail of a Fourier Transform

Lemma 3.10 (Riemann–Lebesgue) *Let $f \in \mathbf{L}^1$. Then holds*

$$\lim_{s\to\pm\infty} \int_{-\infty}^{\infty} e^{isx} f(x)\, dx = 0 \tag{3.17}$$

Proof:
Since f is $\mathbf{L}^1$, there is for all $\epsilon > 0$ a finite step function $\tilde{f}$, such that

$$\int_{-\infty}^{\infty} |f(x) - \tilde{f}(x)|\, dx < \epsilon, \quad \text{hence} \quad \left| \int_{-\infty}^{\infty} e^{isx} f(x) dx - \int_{-\infty}^{\infty} e^{isx} \tilde{f}(x) dx \right| < \epsilon \tag{3.18}$$

and for each finite step of $\tilde{f}$ holds:

$$\lim_{s\to\pm\infty} \int_{x_n}^{x_{n+1}} \tilde{f}_n e^{isx}\, dx \;=\; \lim_{s\to\pm\infty} \tilde{f}_n \left[\frac{1}{is} e^{isx}\right]_{x_n}^{x_{n+1}} = 0 \tag{3.19}$$

□

Theorem 3.11 *Let $f \in \mathbf{L}^1$. If f has a nth derivative $f^{(n)}$ and if $f^{(n)}$ is $\mathbf{L}^1$, then holds:*

$$\lim_{s\to\pm\infty} |s|^n (\mathcal{F}f)(s) = 0 \tag{3.20}$$

Proof:
For $n = 0$ the statements holds due to the lemma of Riemann and Lebesgue (lemma 3.10). If $f'(x)$ is $\mathbf{L}^1$, then an integration by part shows:

$$(\mathcal{F}f)(s) \;=\; \int_{-\infty}^{\infty} e^{isx} f(x) dx = \left[\frac{e^{isx}}{is} f(x)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{1}{is} e^{isx} f'(x) dx \tag{3.21}$$

$$= \frac{i}{s} \int_{-\infty}^{\infty} e^{isx} f'(x) dx \tag{3.22}$$

For $n > 1$ one has to perform the partial integration n times. Since $\int_{-\infty}^{\infty} e^{isx} f^{(n)} dx$ vanished for $s \to \pm\infty$ due to lemma 3.10, the assertion is established. □

Lemma 3.12 *Let f be $\mathbf{L}^1$, $f \in \mathbf{C}^2$ and $f'' \in \mathbf{L}^1$. Then $\mathcal{F}f$ is integrable and there is a $\theta(s) \in \mathbf{L}^1$ such that*

$$f(x) \;=\; \int_{-\infty}^{\infty} e^{isx} \theta(s)\, ds \tag{3.23}$$

Proof:
By lemma 3.3 $(\mathcal{F}f)(s)$ is bounded and it decays at least with s^{-2} for $s \to \pm\infty$ by theorem 3.11. Hence $(\mathcal{F}f)(s)$ is $\mathbf{L}^1$ and $2\pi\theta(-s) = (\mathcal{F}f)(s)$ by theorem 3.8. □

3.2 Introduction to the Theory of Characteristic Functions

The characteristic function is a powerful tool in stochastics. This function exists for any real–valued random variable X and there are several relations which are very useful for computations in practice. Furthermore, the characteristic function contains all information about the distribution and in some cases, the density of X can be determined by Fourier inversion. Furthermore the characteristic function can easily be used to determine the moments or the cumulants of a distribution.
The main properties of characteristic functions are stated and some examples are given to illustrate these properties by some common distributions, which are described in the appendix A.

3.2.1 Definition and Relation to Fourier Transform

Definition 3.13 *Let X be a real-valued random variable. Then the characteristic function of X is defined by*

$$\Phi_X(s) := \mathbf{E}[e^{isX}] \tag{3.24}$$

where i denotes the imaginary unit.

Note, that the characteristic function exists for any real–valued random variable. From the definition, one can directly obtain the properties, which are listed in following lemma.

Lemma 3.14 *Let X, Y be two independent, real-valued random variables and Φ_X, Φ_Y their characteristic functions. Then holds for all $a, s \in \mathbb{R}$:*

$$\begin{aligned}
|\Phi_X(s)| &\leq 1 && (3.25)\\
\Phi_X(-s) &= \overline{\Phi_X(s)} && (3.26)\\
\Phi_{X+Y}(s) &= \Phi_X(s) \cdot \Phi_Y(s) && (3.27)\\
\Phi_{X+a}(s) &= e^{isa} \cdot \Phi_X(s) && (3.28)\\
\Phi_{a\cdot X}(s) &= \Phi_X(a \cdot s) && (3.29)
\end{aligned}$$

The next lemma, which can directly be derived from the definition, provides the relation between the Fourier transform and the characteristic function.

Lemma 3.15 *Let f be a probability density of the real–valued random variable X, hence $f(x) \geq 0$ and $\int f(x)dx = 1$. Then the characteristic function of X is given by the Fourier transform of f:*

$$\Phi_X(s) = (\mathcal{F}f)(s) \tag{3.30}$$

Theorem 3.16 *Two real–valued random variables X, Y have the same characteristic functions if and only if they have the same distribution.*

$$\mathbf{E}[f(X)] = \mathbf{E}[f(Y)] \tag{3.31}$$

Proof:
If X, Y have the same distribution, then $\mathbf{E}[e^{isX}] = \mathbf{E}[e^{isY}]$, hence $\Phi_X(s) = \Phi_Y(s)$. To show the other direction, one may define for all $\alpha \in \mathbb{R}$ and $\sigma > 0$:

$$f_{\alpha,\sigma}(x) \;:=\; \frac{1}{1 + e^{\frac{x-\mu}{\sigma}} + e^{\frac{\mu - x - \frac{1}{\sigma}}{\sigma}}} \tag{3.32}$$

and $f_{\alpha,\sigma}(x)$ is bounded, $\mathbf{L}^1$, $\mathbf{C}^2$ and $f''_{\alpha,\sigma}(x)$ is $\mathbf{L}^1$. Hence by lemma 3.12 there is a $\theta_{\alpha,\sigma}(s) \in \mathbf{L}^1$ such that

$$f_{\alpha,\sigma}(x) \;=\; \int_{-\infty}^{\infty} \theta_{\alpha,\sigma}(s) e^{isx} ds \tag{3.33}$$

Hence one obtains

$$\mathbf{E}[f_{\alpha,\mu}(X)] \;=\; \int_{-\infty}^{\infty} \theta_{\alpha,\sigma}(s)\mathbf{E}[e^{isX}]ds = \int_{-\infty}^{\infty} \theta_{\alpha,\sigma}(s)\Phi_X(s)ds \tag{3.34}$$

$$= \int_{-\infty}^{\infty} \theta_{\alpha,\sigma}(s)\Phi_Y(s)ds = \mathbf{E}[f_{\alpha,\sigma}(Y)] \tag{3.35}$$

This holds for all $\sigma > 0$. In the limit $\sigma \to 0$ holds

$$\lim_{\sigma \to 0} f_{\alpha,\sigma}(x) \;=\; \begin{cases} 1 & x < \mu \\ \frac{1}{2} & x = \mu \\ 0 & x > \mu \end{cases} \tag{3.36}$$

and hence

$$\lim_{\sigma \to 0} \mathbf{E}[f_{\alpha,\sigma}(X)] \;=\; \frac{1}{2}P[X < \alpha] + \frac{1}{2}P[X \leq \alpha] \tag{3.37}$$

$$= \frac{1}{2}P[Y < \alpha] + \frac{1}{2}P[Y \leq \alpha] \tag{3.38}$$

□

Hence one might expect that there is an inverse of lemma 3.15 and one obtains it by lemma 3.4 and theorem 3.8:

Theorem 3.17 *Is the characteristic function Φ_X of a random variable X integrable, then X has a bounded, continuous density f which is given by*

$$f(x) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \Phi_X(s)\, ds \tag{3.39}$$

This Fourier inversion theorem is very important. Assume, that a characteristic function $\Phi_X(s)$ of a random variable X is given and that $\Phi_X(s) \in \mathbf{L}^1$. Then this theorem tells us not only, that X has a continuous density $f(x)$, but it also gives a formula to compute the density.
In the case that the characteristic function is not integrable, one is faced with the problem of a Fourier inversion of a non–integrable function. Assume that the characteristic function Φ_X of a real–valued random variable X is given and that Φ_X is not integrable. Then the following question remains:

How to obtain the distribution of X from its characteristic function?

Three answers to this questions will be given in sections 3.3, 3.4 and 3.5.

Examples for the Decay Behaviour of Characteristic Functions

Since the density of a random variable and the corresponding characteristic function are related by a Fourier transformation, it is clear from 3.1.3 that the smoothness of the distribution affects the tail behaviour of the characteristic function and vice versa. This relation will be illustrated by some well–known distributions.
The density of the standard normal distribution is given by $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and all derivatives have the structure $p(x)e^{-x^2}$, where $p(x)$ denotes a polynomial in x. Hence all derivatives of the normal density are integrable and so $\lim_{s\to\pm\infty}|s|^n\Phi(s)=0$ for all n. This can also be seen from the characteristic function $e^{-s^2/2}$.
An opposite example is the chi–square distribution with one degree of freedom. The density of this distribution is given by $f(x)=\frac{e^{-x/2}}{\sqrt{2\pi x}}$ and is integrable. The first derivative of this density is given by $f'(x)=\frac{e^{x/2}(1+x)}{2\sqrt{2\pi x^3}}$ and this function is not integrable, because of the $x^{-3/2}$ behaviour near 0. The characteristic function of the $\chi^2(1)$ distribution is given by $\Phi(s)=\frac{1}{\sqrt{1-2is}}$, hence $\Phi(\pm\infty)=0$ but $\lim_{s\to\pm\infty}|s|\Phi(s)$ does not exist.
In the previous example the density was not bounded, but one can also give an example with an bounded density, where the characteristic function is not integrable. Consider for example the uniform distribution on $[a,b]$ with $a<b$. The corresponding characteristic function is given by $\frac{e^{isb}-e^{isa}}{(b-a)is}$ and is obviously not $\mathbf{L}^1$.

3.2.2 The Characteristic Function and Moments

If a distribution is studied, one is mostly interested in the moments and the absolute moments of the random variable X, and if they exist they will be denoted by

$$m_n=\mathbf{E}[X^n] \qquad\qquad M_n=\mathbf{E}[\,|X|^n\,] \tag{3.40}$$

The moments of a random variable X and its characteristic function are closely related:

Theorem 3.18 *Let X be a real-valued random variable and Φ_X its characteristic function. If the nth absolute moment M_n exists, then the characteristic function is n times continuous differentiable and is given by*

$$\Phi_X^{(n)}(s) \;=\; i^n\mathbf{E}[e^{isX}X^n] \tag{3.41}$$

Especially, $\Phi_X^{(n)}(0)=i^nm_n$, where m_n denotes the nth moment of X.

Proof:
From the definition follows, that $\Phi_X^{(n)}(s)=\mathbf{E}[e^{isX}(iX)^n]$ if the right hand side exists, and that is guaranteed by the existence of the nth absolute moment. □

An example for a distribution where even the first moment does not exist and where hence the characteristic function is not differentiable at 0 is given by the Cauchy distribution, which has the density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. The corresponding characteristic function is $\Phi(s) = e^{-|s|}$, which is obviously not differentiable at $s = 0$.

Lemma 3.19 *For all x and $n = 1, 2, \ldots$ holds:*

$$\left| e^{ix} - \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} \right| \leq \frac{|x|^n}{n!} \tag{3.42}$$

Proof:
Define the term within absolute signs by $t_n(x) = e^{ix} - \sum_{k=0}^{n-1} \frac{(ix)^k}{k!}$. Then one can show that

$$t_{n+1}(x) \;=\; i \int_0^x t_n(y)dy \tag{3.43}$$

and the lemma can be proved by induction. For $n = 1$ one obtains

$$|t_1(x)| \;=\; \left| i \int_0^x e^{iy} dy \right| \leq |x| \tag{3.44}$$

and if the lemma holds for n, one may conclude:

$$|t_{n+1}(x)| \;=\; \left| \int_0^x t_n(y)dy \right| \leq \int_0^{|x|} \frac{y^n}{n!} dy = \frac{|x|^{n+1}}{(n+1)!} \tag{3.45}$$

□

Theorem 3.20 *If all moments of the random variable X exists and if*

$$\lambda \;:=\; \limsup_{n\to\infty} \sqrt[n]{\frac{M_n}{n!}} = \limsup_{n\to\infty} \frac{e}{n} \sqrt[n]{M_n} < \infty \tag{3.46}$$

then the characteristic function is analytical at $s = 0$ and the Taylor series

$$\Phi_X(s) \;=\; \sum_{n=0}^{\infty} \frac{m_n}{n!} (is)^n \tag{3.47}$$

converges for $|s| < \frac{1}{\lambda}$.

Proof:
The difference between the Taylor series and the characteristic function can be estimated by the help of lemma 3.19 and the Jensen inequality:

$$\left| \Phi_X(s) - \sum_{k=0}^{n-1} \frac{m_k}{k!} (is)^k \right| \;\leq\; \mathbf{E}\left[e^{isX} - \sum_{k=0}^{n-1} \frac{1}{k!} (isX)^k \right] \leq \mathbf{E}\left[\frac{|sX|^n}{n!} \right] \tag{3.48}$$

In the limit $n \to \infty$, the right hand side tends to zero for $|s| < \frac{1}{\lambda}$, since by the Cauchy–Hadamard theorem λ is the radius of convergence of the power series $\sum_{n=0}^{\infty} \frac{M_n s^n}{n!}$. The equivalence of the two definitions of λ follows from the Stirling formula. □

Hence, if the moments do not increase too fast, the moments define the characteristic function and hence the distribution uniquely. An example of a distribution, which is *not* uniquely defined by its moments is the lognormal distribution, since the moments of this distribution have the asymptotic behaviour of e^{n^2}. Hence the characteristic function of the lognormal distribution is infinite often continuous differentiable, but it is not analytical at 0.

Remark 3.21 *The criteria of the theorem which guarantees that the moments of a random variable determine its distribution is sufficient, sufficient and necessary is the Carleman's condition [17]:*

$$\sum_{k=0}^{\infty} \frac{1}{\sqrt[2k]{M_{2k}}} = \infty \tag{3.49}$$

3.2.3 Relation to some Generating Functions

The Moment Generating Function

If the moment generating function of a random variable X exists, it is defined by

$$M_X(t) \;:=\; \mathbf{E}[e^{tX}] \tag{3.50}$$

and the relation to the characteristic function is given by

$$M_X(t) = \Phi_X(-it) \tag{3.51}$$

The name of this function is based on the relation

$$\mathbf{E}[X^n] \;=\; \frac{d^n M_X}{dt^n}(0) \tag{3.52}$$

Note that the characteristic function exists for real arguments in general, but only if it can be continued on the imaginary axis, the moment generating function exist. The Cauchy distribution is an example where no moment generating function exists.

The Cumulant Generating Function

Let X be a real-valued random variable with existing moment generating function $M_X(t)$. Then the cumulant generating function is defined by

$$C_X(t) := \ln M_X(t) = \ln \Phi_X(-it) \tag{3.53}$$

If existing, the nth cumulant is defined by

$$\kappa_n \;:=\; \frac{d^n C_X}{dt^n}(0) \tag{3.54}$$

and the following relation between the moments and the cumulants hold for all t:

$$\exp\left(\sum_{n=1}^{\infty}\kappa_n\frac{t^n}{n!}\right) = 1+\sum_{n=1}^{\infty}m_n\frac{t^n}{n!} \tag{3.55}$$

It will be shown in lemma 5.7 and remark 5.8 that this equation implies the following relation between the moments and the cumulants:

$$m_n = \sum_{j=1}^{n}\binom{n-1}{j-1}m_{n-j}\kappa_j \tag{3.56}$$

The name "cumulant" is based on the following property. Let $X_1,\dots X_N$ be independent random variables, then holds:

$$\kappa_n\left(\sum_{j=1}^{N}X_j\right) = \sum_{j=1}^{N}\kappa_n(X_j) \tag{3.57}$$

The first cumulant is the expectation, the second cumulant is the variance. For the higher cumulants there is not such a clear interpretation.

The Probability Generating Function

Let X be a random variable with values in $\mathbb{N}$. Then the probability generating function is defined by

$$G_X(z) := \mathbf{E}[z^X] \tag{3.58}$$

and this function exists at least for $|z|<1$. The name of this function is based on the relation

$$P[X=n] = \frac{1}{n!}\frac{d^nG_X}{dz^n}(0) \tag{3.59}$$

The following relation to the characteristic function holds:

$$\Phi_X(s) = G_X(e^{is}) \tag{3.60}$$

3.3 The Distribution Function by Fourier Inversion

In the previous introduction to characteristic functions it is shown, that the characteristic function contains all information about the distribution of a random variable. If the characteristic function is integrable, the density can directly be obtained by the Fourier inversion formula. But if the characteristic function is not integrable, the Fourier inversion integral does not exist. The following theorem provides a method to obtain the distribution function under the condition that the characteristic function decays for large $|s|$ like $s^{-\epsilon}$ with $\epsilon>0$ and that the second moment of the random variable exists.

Theorem 3.22 *Let $\Phi_X(s)$ be the characteristic function of the random variable X such, that $\lim_{s\to\pm\infty} s^\epsilon\, \Phi_X(s) = 0$ for a $\epsilon > 0$. Let further Φ_X be two times differentiable at $s = 0$. Then the distribution function $F(x)$ of the random variable X can be obtained by*

$$F(x) \;=\; G(x) - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\Phi_X(s) - \Phi_Y(s)}{is} e^{-isx}\, ds \tag{3.61}$$

where G is the distribution function of any random variable Y with characteristic function Φ_Y such, that $\mathbf{E}[Y^2]$ exists and that $\Phi_Y(s)$ decays with $o(s^{-\epsilon})$.

The proof of this theorem is given below. The idea of this inversion theorem is to approximate X by a random variable Y and to compute the difference of their distributions by a Fourier integral. The decay properties of the characteristic functions guarantee the the existence of the integral.
For applications one chooses a common distribution for Y, for example the normal distribution or the Gamma distribution. Since the integral yields the difference between the distributions F and G, the choice of the distribution G depends on how the distribution of X will look like. Of course, it is favourable to choose the distribution of Y such that the first and the second moment of X and Y are equal.

Remark 3.23 *Even if $\mathbf{E}[X]$ does not exist, the theorem could be applied. But then one has to find such a approximating distribution, that the integrand is integrable and bounded at $s = 0$. Possibly an approximation by the Cauchy distribution may work.*

Remark 3.24 *If one sets $G(x) = F(x-h)$ in the theorem, hence $\Phi_Y(s) = \Phi_X(s)e^{ihs}$, then one obtains*

$$F(x+h) - F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_X(s)\frac{1 - e^{-ihs}}{is} e^{-isx} ds \tag{3.62}$$

which is the Levy inversion formula for continuous distribution functions.

3.3.1 Proof of Theorem 3.22

First, the integral in equation (3.61) is well defined. Since the second moment of X and Y exist, the characteristic functions may be expanded near 0 by

$$\Phi_X(s) \;=\; 1 + i\mathbf{E}[X]s + \mathcal{O}(s^2) \tag{3.63}$$
$$\Phi_Y(s) \;=\; 1 + i\mathbf{E}[Y]s + \mathcal{O}(s^2) \tag{3.64}$$

therefore

$$\lim_{s\to 0} \frac{\Phi_X(s) - \Phi_Y(s)}{is} = \mathbf{E}[X] - \mathbf{E}[Y] \tag{3.65}$$

and so the integrand is bounded. Since $\Phi_X(s)$ and $\Phi_Y(s)$ decay with $0(s^{-(1+\epsilon)})$ for large $|s|$, the fraction $\frac{\Phi_X(s)-\Phi_Y(s)}{is}$ is an integrable function.

By definition of the characteristic function holds

$$\Phi_X(s) - \Phi_Y(s) = \int_{-\infty}^{\infty} e^{isx} d(F-G)(x) \tag{3.66}$$

$$= \left[e^{isx}(F-G)(x)\right]_{-\infty}^{\infty} - is\int_{-\infty}^{\infty} e^{isx}(F(x)-G(x))dx \tag{3.67}$$

$$= -is\int_{-\infty}^{\infty} e^{isx}(F(x)-G(x))dx \tag{3.68}$$

Hence $\frac{\Phi_X(s)-\Phi_Y(s)}{-is} = (\mathcal{F}(F-G))(s)$ and one may apply the Fourier inversion formula:

$$F(x) - G(x) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\Phi_X(s)-\Phi_Y(s)}{-is} e^{-isx}\, ds \tag{3.69}$$

3.4 Fourier Inversion for Non–Decaying Characteristic Functions

The idea of this method is not to obtain the density or the distribution function of X, but the integral of the distribution function. Even if the density is only defined in terms of Dirac δ functions, this function will be continuous and hence it is a good candidate for numerical computations. In order to use this method, one needs to know the expectation of the unknown distribution, since the idea is to approximate the unknown distribution by a distribution with the same mean. Then the Fourier inversion can be done using the following

Theorem 3.25 *Let F, G be the distribution functions of two random variables X and Y which have both an existing third absolute moment. Let $\Phi_X(s)$ and $\Phi_Y(s)$ denote their characteristic functions and assume that $\mathbf{E}[X] = \mathbf{E}[Y]$. Further define*

$$\hat{F}(x) := \int_{-\infty}^{x} F(y)\, dy \qquad \text{and} \qquad \hat{G}(x) := \int_{-\infty}^{x} G(y)\, dy \tag{3.70}$$

Then the following inversion formula holds:

$$\hat{F}(x) = \hat{G}(x) - \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\frac{\Phi_X(s)-\Phi_Y(s)}{s^2}\, ds \tag{3.71}$$

The proof of this theorem is given below. The application of this theorem is analogous to the use of theorem 3.22. Consider a random variable X with a given characteristic function $\Phi_X(s)$ and expectation $\mathbf{E}[X]$, then one chooses another random variable Y with the same mean. The usual distributions one chooses for Y are the normal or the Gamma distribution, since these distribution have a closed form solution for the characteristic function and the integral of the distribution function, see appendix A. Of course, it is recommended again to fit also the variance of X and Y, if the second moment of X is known.

3.4.1 Proof of Theorem 3.25

Lemma 3.26 *Let X, Y be two random variables with existing third absolute moment and $\mathbf{E}[X] = \mathbf{E}[Y]$. Then the function*

$$\frac{\Phi_X(s) - \Phi_Y(s)}{s^2} \tag{3.72}$$

is integrable.

Proof:
Since the third moments of X and Y exist, one may expand the characteristic functions at $s = 0$:

$$\Phi_X(s) = 1 + i\mathbf{E}[X]s - \mathbf{E}[X^2]s^2 + \mathcal{O}(s^3) \tag{3.73}$$

$$\Phi_Y(s) = 1 + i\mathbf{E}[Y]s - \mathbf{E}[Y^2]s^2 + \mathcal{O}(s^3) \tag{3.74}$$

From $\mathbf{E}[X] = \mathbf{E}[Y]$, one obtains $\lim_{s\to 0} \frac{\Phi_X(s)-\Phi_Y(s)}{s^2} = \mathbf{E}[Y^2 - X^2]$ and the function $\frac{\Phi_X(s)-\Phi_Y(s)}{s^2}$ is bounded. Since the absolute value of a characteristic function is bounded by 1, the function $\frac{\Phi_X(s)-\Phi_Y(s)}{s^2}$ decays at least with s^{-2} for $s \to \pm\infty$. Hence $\frac{\Phi_X(s)-\Phi_Y(s)}{s^2}$ is integrable. □

Lemma 3.27 *Let F and G be the distribution functions of X and Y, let $\hat{F}$ and $\hat{G}$ be defined by (3.70) and let $\mathbf{E}[X] = \mathbf{E}[Y]$. Then holds:*

$$\lim_{x\to\pm\infty} (\hat{F} - \hat{G})(x) = 0 \tag{3.75}$$

Proof:
From the existence of the expectation one can conclude:

$$\int_x^\infty y \; dF(y) = o(1) \qquad \text{as } x \to \infty \tag{3.76}$$

Since $o(1) = \int_x^\infty y \; dF(y) \geq x \int_x^\infty \; dF(y) = x(1 - F(x))$ holds:

$$F(x) = 1 - o(\frac{1}{x}) \qquad \text{as } x \to \infty \tag{3.77}$$

The same asymptotic behaviour also holds for G. Since both distributions have the same mean, holds:

$$0 = \lim_{x\to\infty} \int_{-\infty}^x y \; d(F - G)(y) \tag{3.78}$$

$$= \lim_{x\to\infty} [y(F(y) - G(y))]_{-\infty}^x - \int_{-\infty}^x F(y) - G(y) \; dy \tag{3.79}$$

$$= \lim_{x\to\infty} x(F(x) - G(x)) - (\hat{F}(x) - \hat{G}(x)) \tag{3.80}$$

$$= \lim_{x\to\infty} \hat{G}(x) - \hat{F}(x) \tag{3.81}$$

The case $x \to -\infty$ is trivial, since $\lim_{x\to-\infty} \hat{F}(x) = \lim_{x\to-\infty} \hat{G}(x) = 0$. □

Proof of Theorem 3.25:
Let F, G fulfill the conditions of the theorem. By definition of the characteristic function holds:

$$\Phi_X(s) - \Phi_Y(s) = \int_{-\infty}^{\infty} e^{isx}\, d(F-G)(x) \tag{3.82}$$

$$= \left[(F(x)-G(x))e^{isx}\right]_{-\infty}^{\infty} - is\int_{-\infty}^{\infty}(F(x)-G(x))e^{isx}dx \tag{3.83}$$

Since $\lim_{x\to\pm\infty} F(x) - G(x) = 0$ and e^{isx} is bounded, one obtains:

$$\Phi_X(s) - \Phi_Y(s) = -is\int_{-\infty}^{\infty}(F(x)-G(x))e^{isx}dx \tag{3.84}$$

Another integration by parts yield together with lemma 3.27:

$$\Phi_X(s) - \Phi_Y(s) = -is\left[\left(\hat{F}(x)-\hat{G}(x)\right)e^{isx}\right]_{-\infty}^{\infty} + (is)^2\int_{-\infty}^{\infty}\left(\hat{F}(x)-\hat{G}(x)\right)e^{isx}dx \tag{3.85}$$

$$\frac{\Phi_X(s)-\Phi_Y(s)}{(is)^2} = \int_{-\infty}^{\infty}\left(\hat{F}(x)-\hat{G}(x)\right)e^{isx}dx \tag{3.86}$$

So the function $\frac{\Phi_X(s)-\Phi_Y(s)}{(is)^2}$, which is integrable by lemma 3.26, is the Fourier transform of $\hat{F}(x) - \hat{G}(x)$. Hence the Fourier inversion formula (3.14) yields

$$\hat{F}(x) - \hat{G}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\Phi_f(s)-\Phi_g(s)}{(is)^2}e^{-isx}ds \tag{3.87}$$

□

3.5 Fourier Inversion by Fast Fourier Transformation

Let $\Phi_X(s)$ be the known characteristic function of a real-valued random variable X. Maybe, this random variable has no density, but let us assume that the distribution is "almost" continuous, that is, that X may be approximated by an random variable which has a density f. Further it is assumed, that f vanishes outside a given interval $[a, b]$. Then one may apply the following rule to obtain f:

Let $\Phi_X(s)$ be the characteristic function a random–variable X, and assume that the distribution of X can approximately be described by a density f which lives on a finite interval $[a,b]$. Then f can be obtained by Fourier inversion of the characteristic function using the Fast Fourier Transformation (FFT) algorithm.

This verbal statement will now be explained and justified. To describe the approximating density f of a random variable X one may require, that the following condition holds for each g in a suitable class of functions $\mathcal{G}$:

$$\int f(x)g(x)dx \quad = \quad \mathbf{E}[g(X)] \tag{3.88}$$

and a proper choice of $\mathcal{G}$ is given by

$$\mathcal{G} \quad := \quad \{g(x) = e^{isx} | s \in \mathcal{S}\} \tag{3.89}$$

where $\mathcal{S}$ is a finite set of real numbers. Note, that the functions g are linear independent, which is necessary to expect a suitable approximation. The expectation in (3.88) for these functions g is the characteristic function evaluated at some points s with $s \in \mathcal{S}$. Since the characteristic function is known, the density f is the only unknown in equation (3.88) and hence one can use this equation to determine f. It is natural to expect that f will be a good approximation.

Now the structure of the FFT algorithm will be used, which is described for example in [68]. Since the unknown density f vanishes outside a given interval $[a,b]$, its Fourier transform can be numerically computed by the following algorithm. Let S be the number of sample points (usually a power of 2) and define $\Delta x := \frac{b-a}{S-1}$ and $\Delta s := \frac{2\pi}{S\Delta x}$. Further define three S dimensional vectors and a $S \times S$ matrix M for $j,k = 0, \ldots S-1$ by:

$$x_k \quad := \quad a + k\Delta x \tag{3.90}$$

$$f_k \quad := \quad f(x_k) \tag{3.91}$$

$$s_k \quad := \quad \begin{cases} k\Delta s & \text{if } k < \frac{S}{2} \\ (k-S)\Delta s & \text{else} \end{cases} \tag{3.92}$$

$$M_{jk} \quad := \quad \exp\left(2\pi i \frac{jk}{S}\right) \tag{3.93}$$

The set $\mathcal{S}$ is given by $\mathcal{S} := \{s_k | k = 0, \ldots S-1\}$ and the Fourier transform of f at the points s_k will be denoted by $\Phi_k := \Phi_X(s_k)$.

The Fourier integral can be computed by the approximation

$$\int_{-\infty}^{\infty} e^{isx} f(x)dx \quad \approx \quad \Delta x e^{isa} \sum_{k=0}^{S-1} e^{isk\Delta x} f(a + k\Delta x) \tag{3.94}$$

Using the vector and matrix notations, this equation yields[2]:

$$\Phi_k = \Delta x e^{ias_k} \sum_{j=0}^{S-1} M_{kj} f_j \tag{3.95}$$

Having chosen a suitable interval $[a,b]$ and the number of Fourier steps S, the density f_k is the only unknown in equation (3.95). To determine f_k one has to solve the system of linear equations, which can be solved easily, since the inverse of M_{jk} is given by:

$$M_{jk}^{-1} = \frac{1}{S} \exp\left(-2\pi i \frac{jk}{S}\right) \tag{3.96}$$

A short calculation shows, that this is in fact the inverse:

$$\sum_{k=0}^{S-1} M_{jk} M_{kl}^{-1} = \frac{1}{S} \sum_{k=0}^{S-1} \left(e^{2\pi i \frac{j-l}{S}}\right)^k = \begin{cases} 1 & j = l \\ \frac{1}{S} \frac{1-e^{2\pi i (j-l)}}{1-e^{2\pi i (j-l)/S}} = 0 & j \neq l \end{cases} \tag{3.97}$$

Hence the inversion formula based on the linear equations is given by

$$f_j = \frac{1}{S\Delta x} \sum_{k=0}^{S-1} \exp\left(-2\pi i \frac{jk}{S}\right) e^{-ias_k} \Phi_k \tag{3.98}$$

Since M^{-1} has a quite similar structure as M, one can compute the matrix vector multiplication by the FFT algorithm to save computation time. So even if the characteristic function is not integrable, the Fourier inversion by FFT works, since the algorithm does not compute an integral, but it solves a set of linear equations which describe the Fourier transformation of the unknown density f.

3.5.1 Smoothing the FFT Result

In the most cases, one expects that the density f is a smooth function, but the result of the FFT algorithm may have a sawtooth pattern. This pattern is in fact a perturbation with the so–called Nyquist critical frequency $\frac{S}{2}\Delta s = \frac{\pi}{\Delta x}$, which corresponds with the cut–off of the domain of the characteristic function. To eliminate this perturbation, one can apply the following algorithm, which performs a midpoint interpolation and so the number of sample points will be reduced by 1:

Algorithm 3.28 (Elimination of a periodic perturbation) *Let f be a density sampled on equidistant points x_k, $k = 0, \ldots, S-1$. Define*

$$\tilde{f}_k := \frac{1}{2}(f(x_k) + f(x_{k-1})) \qquad k = 1, \ldots, S-1 \tag{3.99}$$

Then $\tilde{f}$ is a smooth version of the density f and $\tilde{f}$ is sampled on the points $x_k - \frac{1}{2}\Delta x$ with $k = 1, \ldots, S-1$.

[2]Using the special form of M_{jk} one can valuate the matrix vector product by $\mathcal{O}(S \ln S)$ instead of the usual $\mathcal{O}(S^2)$ operations, which is the reason why this kind of Fourier transformation is called "Fast".

Another smoothing technique is needed at some positions, where the density tends to infinity which can happen in two situations. Either the density may not exist at a certain point x because $P[X = x] > 0$, hence the density is a Dirac δ function at x. Or the density tends to infinity at x although $P[X = x] = 0$, an example for such an density gives the χ^2 distribution with one degree of freedom. Then the FFT result may be very oscillatory near x and for some k one may obtain $f_k < 0$. But a density must be non–negative! In such situations one the next trick will help:

Algorithm 3.29 (Smoothing near singularities) *Let f be a density sampled on equidistant points x_k, $k = 0, \ldots, S-1$. If there are inner points such that the density is negative, hence $f_k < 0$ for a k in $1, \ldots, S-2$, then perform the following algorithm for each such k:*

$$\begin{aligned} left &:= (f_{k-1})^+ \\ right &:= (f_{k+1})^+ \\ f_{k-1} &:= f_{k-1} + \frac{left}{left + right} f_k \\ f_{k+1} &= f_{k+1} + \frac{right}{left + right} f_k \\ f_k &:= 0 \end{aligned}$$

For applications in practice, these smoothing methods which have both the computational effort of $\mathcal{O}(S)$ are rather fast and hence it is recommended to smooth a density by three steps:

1. Apply the smoothing algorithm 3.28.
2. Perform a second smoothing by algorithm 3.29 on the result.
3. Finally apply the first smoothing algorithm (algorithm 3.28) again.

Since the application of the first smoothing algorithm shifts the x-domain by $\frac{1}{2}\Delta x$, this three step algorithm effects a shift of the domain by Δx.

3.6 Conclusion

In the first part an introduction to two mathematical mappings have been given. The first mapping is the Fourier transform, which maps $\mathbf{L}^1$ –functions to functions $\mathbb{R} \to \mathbb{C}$. The second mapping is the characteristic function, which maps real–valued random variables also to functions $\mathbb{R} \to \mathbb{C}$. For random variables with a density, the relation between this two mappings is given by the fact, that the characteristic function is the Fourier transform of the density. And if the characteristic function of a random variable is $\mathbf{L}^1$, then this random variable has a density which can be determined by the Fourier inversion integral.
In applications, the characteristic function of a random variable may not be integrable and hence the standard Fourier inversion integral does not exist. In such situations, three other methods have been provided to obtain the distribution of a random variable from its characteristic function. One method is based on the

well–known FFT algorithm. Since this algorithm is used in several applications, one might expect that it is contained in the most implementations of a mathematical library. Hence this method to invert a characteristic function could become quite popular for practitioners.
Two other methods are based on the theorems presented in 3.3 and 3.4. These aim to determine the difference between two distributions by Fourier inversion. Their advantage is, that they are independent of a certain Fourier transform algorithm and hence one may use more sophisticated integration methods than FFT for the valuation of the integral, e.g [63, 64].
The presented relations, ideas and Fourier inversion methods are valid in a general context and can fruitfully be used whenever one has to determine the distribution of a random variable from its characteristic function. In the second part of this thesis these techniques are used to solve problems which occur in mathematical finance. In chapter 4 the application of the Fourier inversion in the context of the Delta Gamma normal approximation is presented. In chapter 5 the Fourier inversion turns out to provide a numerical stable algorithm for the CreditRisk$^+$ model and for some generalizations of CreditRisk$^+$. In both contexts, the Fourier inversion techniques presented here are a key to solve mathematical challenging tasks in risk management.

Part II

Applications in Risk Management

Chapter 4

Assessment of Market Risk using Quadratic Approximations

In 1995 the Basel Committee on Banking Supervision published its proposal for market risk capital requirements [5]. Since these rules became effective, financial institutions are allowed to use internal models to measure their financial risk based on market fluctuations. In order to quantify this market risk, the so–called "Value at Risk" (VaR) became the most important risk measure. This risk measure is also used for supervising purposes, because banks have to underlay their risk with equity, and the amount is principally determined by the VaR.
The VaR of a portfolio is defined as a q quantile of the portfolio's profit and loss ($P\&L$) over a certain time horizon T:

Definition 4.1 (Value at Risk) *Let $P\&L$ be the random variable which states the profit and loss of a portfolio over an certain time horizon T. For a given quantile q the VaR is defined by*

$$VaR := \inf\{x | P[P\&L \leq -x] < q\} \tag{4.1}$$

Hence within the time T, usually ten trading days, this portfolio will not lose more than the VaR with probability $1-q$ and a typical value for q is 1%. There is one very nice aspect of this risk measure from the supervision point of view. The reported Value at Risk can be compared with the actual loss and the actual loss should exceed the reported Value at Risk only some rather few days a year. If it exceeds to often, then it is quite obvious that there is something wrong with the Value at Risk computation.
But on the other hand the Value at Risk gives no information, what happens if one of the q worst scenarios occur. This is one criticism of the risk measure "Value at Risk" [54]. An axiomatic approach on risk measures is presented in [2] and a *coherent measure of risk* must fulfill four axioms:

Definition 4.2 (coherent measure of risk) *Let X and Y be two real–valued random variables over the same probability space Ω. Then the mapping ρ from real–valued random variables on Ω to $\mathbb{R}$ is a coherent measure of risk if and only if*

$$\rho(X+Y) \leq \rho(X) + \rho(Y) \tag{4.2}$$

$$X \leq Y \implies \rho(X) \geq \rho(Y) \tag{4.3}$$

$$\rho(\alpha X) = \alpha\rho(X) \qquad \forall\ \alpha \geq 0 \tag{4.4}$$
$$\rho(X+\alpha) = \rho(X) - \alpha \tag{4.5}$$

In this definition X and Y represent the profit and loss of two portfolios. The first condition, the subadditivity of ρ, is the mathematical description of investment diversification. If the outcome of the portfolio X is never better then the outcome of the portfolio Y, then the monotonicity guarantees that X is rated to be more risky than Y. The positive homogeneity of a coherent risk measure corresponds to the scaling invariance of prices, which has been discussed in a different context in chapter 1. The last relation is the so–called *translation invariance*, which says, that adding a non–risky investment reduces the risk exactly by this amount. The most popular coherent risk measure is the *worst conditional expectation* (WCE). If one considers market risk, the $P\&L$ is a continuously distributed random variable and the WCE equals the *Conditional Value at Risk*, which is also called *Tail Conditional Expectation*:

Definition 4.3 (Conditional Value at Risk) *Let X be the profit and loss distribution of a portfolio with regard to a certain time horizon and let VaR be the Value at Risk of the portfolio to a quantile q. Then the Conditional Value at Risk (CVaR) with respect to the quantile q is defined by:*

$$CVaR := \mathbf{E}[X | X \leq -VaR] \tag{4.6}$$

The VaR gives the amount of money one will lose at least in the q worst cases, but the CVaR gives the expectation of the loss in these scenarios, hence one gets an impression about the size of such losses.

If one knows the distribution of a portfolio, then it is quite easy to determine the Value at Risk, the Conditional Value at Risk or any other risk measure. In order to obtain the profit and loss distribution one has to introduce some assumptions on the possible future behaviour of the underlying assets of the portfolio. In general it is almost impractical to compute the $P\&L$, because portfolios are very complex composed.
One quite popular approach in practice to overcome this problem is the Delta Gamma normal method, which assumes that the change of a portfolio's worth can be described by a quadratic form of normal distributed risk factors. There are several numerical methods to compute the Value at Risk based on these assumptions, for example the Cornish–Fischer expansion or saddlepoint approaches [61]. Another technique is the Fourier inversion algorithm which is studied in this chapter and which has the advantage, that the whole profit and loss distribution will be obtained and not only the tail behaviour.
In this chapter, the efficient implementation of the Fourier inversion algorithm for the Delta Gamma normal approach is presented and the computational effort of this method is studied. This algorithm is not quite new, but usual descriptions of this algorithm do not care about some very important points. There are pitfalls in this algorithm and the motivation is to warn the reader and to provide ways to avoid these traps. If one takes care of these points one will be rewarded by a fast and numerical stable algorithm.

It turns out, that the mathematical methods presented in the first part of this thesis are quite essential in this context. As the name of the method indicates, one has to determine derivatives of the portfolio and in this respect, the relations presented in chapter 1 are quite useful. The key to reduce the computation time in practice is the fast Cholesky decomposition algorithm introduced in chapter 2, which can also decompose rank–deficient, semi–positive symmetric matrices. It will be shown, that the linear algebra operations in the Delta Gamma normal algorithm dominate the computation time and for large portfolios the computation time corresponds to the effort of two matrix multiplications. Finally one has to perform a Fourier inversion of a characteristic function, which may not be integrable. In this respect, the Fourier methods presented in chapter 3 are quite advantageous.

After the presentation of the algorithm in section 4.1, the Delta Gamma normal approach will be discussed in detail. By the study of two examples it will be shown, that the usual way to obtain the quadratic approximation by Taylor expansion may be quite good, but it can also be rather bad. An alternative way to obtain a quadratic approximation is suggested, which bases on a simple integration theorem. Also the normal distribution assumption will be discussed and the Delta Gamma t_f is recalled, which has been proposed by Glasserman, Heidelberger and Shahabuddin [43]. Their approach provides a first method for an efficient determination of the Value at Risk for large portfolios, where the risk factors are heavy–tailed distributed. It turns out, that the developed optimization techniques for the Delta Gamma normal approach can also be applied on the Delta Gamma Student approach.

4.1 The Fourier Inversion Algorithm for Delta Gamma Normal

The Delta Gamma normal approach to asses market risk is based on two assumptions, which give the name to this approximation. The first assumption is, that the portfolio may be approximated by a quadratic form[1] and the second one is, that the risk factors are assumed to be normal distributed. These approximations were firstly suggested by Wilson [82] and a first reasonable numerical method to determine the Value at Risk based on this approximation was presented by Fallon [34] using Cornish–Fischer expansions. For a comparison of several numerical approaches see [55].

Here the Fourier inversion algorithm is presented which determines not only the Value at Risk of the portfolio but also the profit and loss distribution ($P\&L$). Some aspects of this algorithm are studied in [56, 72, 83], but for an implementation in practice, there are several pitfalls in this algorithm. In this presentation of the Fourier inversion algorithm each of these traps is indicated and a way to overcome possible problems is given. Furthermore, the asymptotic computational effort for a large number of risk factors N is analyzed.

[1] Note that all first derivatives are called *Delta* and all second derivatives are called *Gamma* and the derivatives are taken with respect to all risk factors like prices, interest rates or implied volatilities. This nomination differs from the nomination in chapter 1, where Delta (Gamma) was the first (second) derivative with respect to the underlying asset price only.

4.1.1 The Quadratic Approximation

The change of a portfolio value, the so–called profit or loss over a fixed time horizon T is denoted by $P\&L$. It is assumed, that the $P\&L$ may be modelled by

$$P\&L \quad = \quad \Theta + \Delta^\top X + X^\top \Gamma X \tag{4.7}$$

where X is a N dimensional normal distributed random variable with mean 0 and covariance matrix Σ and the $N \times N$ matrix Γ is symmetric. Θ is a scalar and Δ is a N dimensional vector.

How to obtain the Quadratic Approximation

A quite popular assumption in the practice of risk management is, that the underlying price processes like stock prices or foreign exchange rates are lognormal distributed. Then the price process at a fixed time T are modelled by

$$S_T^i \quad = \quad S_0^i \exp(X^i) \qquad\qquad i = 1, \ldots, N \tag{4.8}$$

where X is multivariate normal distributed with mean μ and covariance matrix Σ. Let $P(t,s) : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ denote the portfolio function, hence the value of the portfolio at a certain time t and a certain state s of the market. By a Taylor expansion of the portfolio function at the point $(0, S_0)$ one obtains:

$$\begin{aligned} P(T,S_T) \quad &= \quad P(0,S_0) + \frac{\partial P}{\partial t}T + \sum_{i=1}^{N} \frac{\partial P}{\partial s^i}(S_T^i - S_0^i) \\ &\quad + \sum_{i,j=1}^{N} \frac{\partial P}{\partial s^i}\frac{\partial P}{\partial s^j}(S_T^i - S_0^i)(S_T^j - S_0^j) \end{aligned} \tag{4.9}$$

Since T is small in practice, X^i is in average of order $\mathcal{O}(\sqrt{T})$, one may approximate equation (4.8) by

$$S_T^i \quad = \quad S_0^i \left(1 + X^i + \frac{1}{2}(X^i)^2\right) + \text{error, which is in average } \mathcal{O}(T^{\frac{3}{2}}) \tag{4.10}$$

Using this approximation, one obtains the quadratic approximation of the portfolio:

$$\begin{aligned} P(T,S_T) \quad &= \quad P(0,S_0) + \frac{\partial P}{\partial t}T + \sum_{i=1}^{N} S_0^i \frac{\partial P}{\partial s^i}X^i + \sum_{i,j=1}^{N} S_0^i S_0^j \frac{\partial P}{\partial s^i}\frac{\partial P}{\partial s^j}X^i X^j \\ &\quad + \frac{1}{2}\sum_{i=1}^{N} S_0^i \frac{\partial P}{\partial s^i}(X^i)^2 \end{aligned} \tag{4.11}$$

In this representation the normal variable X has a mean μ. By the transformation $Y = X - \mu$ one easily obtains a quadratic approximation of the portfolio where the random variables are normal distributed with mean 0 and covariance Σ. Hence for the further analysis it is assumed without loss of generality, that $\mathbf{E}[X] = 0$.
Note, that one needs to compute the derivatives of the portfolio. In this respect the methods presented in chapter 1 are very fruitful to save computation time, if one

proceeds to find a quadratic approximation on this way. However, the analysis of the Delta Gamma normal algorithm does not imply, that the quadratic approximation is based on a Taylor expansion – any other reasonable quadratic approximation may also be used. See section 4.2 for a discussion of the Taylor approximation in the context of the Delta Gamma normal approach.

Remark 4.4 *There is often the opinion, that the Delta normal approximation is as good as the Delta Gamma normal approximation, if the portfolio does not contain any options. But one can easily see from the last term in equation (4.11), that there is a difference to the Delta normal approximation, even if all second derivatives vanish.*

4.1.2 Decoupling of the Risk Factors

Usually the covariance matrix Σ is estimated from time series of perhaps 250 days. The number of risk factors is much bigger (perhaps about 1000), so the estimated covariance matrix will be singular. A fast algorithm to obtain the Cholesky decomposition even in this case is presented in chapter 2. Hence one finds a permutation P and a lower triangular matrix C such that:

$$\Sigma = P^\top C C^\top P \tag{4.12}$$

Now the relation $X = P^\top C S$ holds in distribution, where S is a vector of iid. normal distributed random numbers with mean 0 and variance 1. Hence the profit and loss may be written in terms of S:

$$\begin{aligned} P\&L &= \Theta + \Delta^\top P^\top C S + S^\top C^\top P \Gamma P^\top C S && (4.13)\\ &= \Theta + \tilde{\Delta}^\top S + S^\top \tilde{\Gamma} S && (4.14) \end{aligned}$$

where

$$\begin{aligned} \tilde{\Delta} &:= C^\top P \Delta && (4.15)\\ \tilde{\Gamma} &:= C^\top P \Gamma P^\top C && (4.16) \end{aligned}$$

For the analysis of the computational effort, let us define a basic operation *op* consisting of one addition, one multiplication and the loop management. Recall that the multiplication of two $N \times N$ matrices costs $N^3 + \mathcal{O}(N^2)$ *op* by this convention. It is known from chapter 2 that the Cholesky decomposition costs $\frac{1}{6}N^3 + \mathcal{O}(N^2)$ *op*. After this decomposition one has to compute $\tilde{\Delta}$ and $\tilde{\Gamma}$. The effort of computing $\tilde{\Delta}$ and $P\Gamma P^\top$ is of order $\mathcal{O}(N^2)$, hence for the asymptotic computation time one only has to consider the two matrix multiplications with C and $C^\top$. Since C is lower triangular, the first matrix multiplication costs asymptotical $\frac{1}{2}N^3$ *op*. The second matrix multiplication is even cheaper, because one knows that the resulting matrix is symmetric and hence the computational effort is given by $\frac{1}{3}N^3$ *op*. So the total effort of this step of the algorithm amounts to $1 \cdot N^3 + \mathcal{O}(N^2)$ *op*.

Remark 4.5 *Do to the possible rank defect of the correlation matrix, one shall consider the case that the rank of the covariance matrix is $M < N$. In this situation the*

generalized Cholesky decomposition introduced in chapter 2 helps to save a lot of computation time. The effort of the decomposition is again given by $N^3 + \mathcal{O}(N^2)$ op*, but since* C *is a* $M \times N$ *matrix, the first matrix multiplication costs* $\frac{1}{2}MN^2 + \mathcal{O}(N^2)$ op *and the second matrix multiplication efforts asymptotically* $\frac{1}{2}M^2(N - \frac{1}{3}M)$ op*. Hence the total effort is of order* $\frac{1}{6}(N+M)^3 - \frac{1}{3}M^3$op*. Note, that in all further computations, the dimension of the problem is* M *(instead of* N*), hence the computational effort in the following steps is also reduced.*

4.1.3 Eigenvalue Decomposition

First, a theoretical description is given, which tells *what* one has to do and afterwards some hints are given on *how* to do it. Usual descriptions on the Delta Gamma normal approximation do not handle this difference carefully enough. This is the reason why the Fourier inversion algorithm of the Delta Gamma normal approach is sometimes called slow.
Let Q be an orthogonal $N \times N$ matrix such that the matrix $\Lambda := Q^\top \tilde{\Gamma} Q$ is diagonal. Recall that the columns of Q contain the eigenvectors of $\tilde{\Gamma}$. Let U be a vector with N iid N(0,1) distributed random numbers, then one may set $S = QU$ since QU are again iid N(0,1) distributed. Hence the profit and loss may be written:

$$P\&L = \Theta + b^\top U + U^\top \Lambda U \qquad \text{where} \tag{4.17}$$

$$b := Q^\top \tilde{\Delta} \tag{4.18}$$

In practice, one might tend to calculate exactly this way. One takes some standard linear algebra package and computes all eigenvalues and eigenvectors. But one only needs the eigenvalues and this vector b. So computing all eigenvectors is too much overhead. To understand this, one has to take a closer look into the linear algebra algorithms.

How to perform the eigenvalue decomposition

The standard procedure to compute eigenvalues and eigenvectors of a symmetric matrix is the so called QR–algorithm [68]. This algorithm can be split into two steps, first one has to perform the Householder reduction and then one has to iterate some Givens rotations. Both steps have the idea to define in the ith step a matrix Q_i and to compute

$$M_i := Q_i^\top M_{i-1} Q_i \qquad \text{and} \qquad E_i := Q_i E_{i-1} \tag{4.19}$$

To initialize this algorithm, define $M_0 = \tilde{\Gamma}$ and $E_0 = \mathbf{1}$. One performs this algorithm until the matrix M_n is (almost) diagonal. The matrix E_n is an orthogonal matrix which columns contain the eigenvectors of M_0.
For the problem one is faced with in the context of the Delta Gamma approximation, it is enough to perform a similar, but significant faster algorithm. Define $M_0 = \tilde{\Gamma}$ and $b_0 = \tilde{\Delta}$. Then choose the same orthogonal matrices Q_i and compute

$$M_i := Q_i^\top M_{i-1} Q_i \qquad \text{and} \qquad b_i := Q_i b_{i-1} \tag{4.20}$$

until M_n is numerically diagonal. The advantage of this algorithm is, that in each step only one matrix and one vector has to be updated instead of two matrices.

For a more detailed analysis of the computation time it is important to analyze the two phases in the QR algorithm. At first, $N-2$ Householder reductions have to be performed. The costs to update the vector b at each step is given by $\mathcal{O}(N)$ *op* and negligible, since the effort for the whole Householder reduction of one matrix is well–known and given by $1 \cdot N^3 + \mathcal{O}(N^2)$ *op*. So the expense of the usual Householder reduction with updating the matrix E instead of updating only a vector b is given by $2 \cdot N^3 + \mathcal{O}(N^2)op$, and there is a benefit of a factor two in this phase by using the idea of updating the vector only.
The second phase of the QR algorithm consists of several Givens rotations until the matrix M is "diagonal enough". Therefore one cannot say how much rotations will be made, but the number of rotations one needs is of order $\mathcal{O}(N)$. Updating the matrix M_i is with costs of $\mathcal{O}(N)$ *op* rather fast, since the matrix itself is tridiagonal after the Householder reduction and updating the vector costs $\mathcal{O}(N)op$ too. So the second phase has a total computation time of $\mathcal{O}(N^2)op$. Updating the matrix E_i in the usual QR algorithm would cost $\mathcal{O}(N^2)$ in each step and so the computation time of the second phase in the usual QR algorithm would be of order $\mathcal{O}(N^3)op$, if one would determine all eigenvectors. So the idea of updating a vector only is very profitable in this phase, since the computation time of this algorithm is only quadratic in the number of risk factors. Hence the whole eigenvalue decomposition step can be performed with $N^3 + \mathcal{O}(N^2)$ operations *op*.

4.1.4 The Characteristic Function

From equation (4.17) one obtains

$$P\&L \quad = \quad \Theta + \sum_{j=1}^{N} \left(b_j U_j + \lambda_j U_j^2 \right) =: \Theta + \sum_{j=1}^{N} Y_j \tag{4.21}$$

where λ_i are the diagonal elements of Λ and b_i are the components of the vector b. The U_i are iid. N(0,1) random numbers. In order to obtain the characteristic function of $P\&L$, one may use the fact, that the addends Y_i are independent. For this reason it is essential to know the characteristic function of Y_i, which is a scaled, non–central chi–square distributed random variable with one degree of freedom.

Lemma 4.6 *Let U be univariate standard normal distributed. Then the characteristic function of $Y := bU + \lambda U^2$ is given by*

$$\Phi_Y(s) \quad = \quad \frac{1}{\sqrt{1 - i2\lambda s}} \exp\left(-\frac{b^2 s^2}{2(1 - i2\lambda s)} \right) \tag{4.22}$$

where the square root is taken such, that $\Re\sqrt{1 - i2\lambda s} > 0$ holds.

Proof:
Solving the corresponding integral by completion of the square yields:

$$\Phi_Y(s) \quad = \quad \mathbf{E}[e^{isbU + is\lambda U^2}] = \int_{-\infty}^{\infty} e^{isbx + is\lambda x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \tag{4.23}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(1-2is\lambda)(x-\frac{isb}{1-2is\lambda})^2 - \frac{s^2b^2}{2(1-2is\lambda)}\right) dx \quad (4.24)$$

$$= \exp\left(-\frac{s^2b^2}{2(1-2is\lambda)}\right) \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty-\frac{isb}{1-2is\lambda}}^{\infty-\frac{isb}{1-2is\lambda}} e^{-\frac{1}{2}(1-2is\lambda)x^2} dx \quad (4.25)$$

$$= \exp\left(-\frac{s^2b^2}{2(1-2is\lambda)}\right) \cdot \frac{1}{\sqrt{1-2is\lambda}} \quad (4.26)$$

To show the last equality, substitute $t = \frac{x}{\sqrt{1-2is\lambda}}$ where $\Re\sqrt{1-2is\lambda} > 0$. Then one obtains an integral along a straight line parallel to the real axis. Due to the choice of the sign of the square root, this integral is equal to the integral along the real axis, which is well–known, since it is the integral over the normal density. □

Since the characteristic function of a sum of independent random numbers is given by the product of the characteristic functions, one obtains

Theorem 4.7 *The characteristic function of the $P\&L - \Theta$ is given by*

$$\Phi_{P\&L-\Theta}(s) = \prod_{j=1}^{N} \frac{1}{\sqrt{1-i2\lambda_j s}} \cdot \exp\left(-\sum_{j=1}^{N} \frac{b_j^2 s^2}{2(1-i2\lambda_j s)}\right) \quad (4.27)$$

where $\Re\sqrt{1-i2\lambda_j s} > 0$ for each j.

Hence, after performing some numerical linear algebra operations to obtain the values b_j and λ_j, one has a closed form solution for the characteristic function of the $P\&L$ distribution. Also the expectation and the variance of the change of the portfolio value are given by b_j and λ_j:

Theorem 4.8 *The expectation and variance of the profit and loss distribution are given by:*

$$\mathbf{E}[P\&L] = \Theta + \sum_{j=1}^{N} \lambda_j \quad (4.28)$$

$$\mathbf{Var}[P\&L] = \sum_{j=1}^{N} 2\lambda_j^2 + b_j^2 \quad (4.29)$$

Proof:
Since $P\&L - \Theta$ can be written as a sum of independent Y_j, one obtains:

$$\mathbf{E}[P\&L - \Theta] = \mathbf{E}[\sum_{j=1}^{N} Y_j] = \sum_{j=1}^{N} \mathbf{E}[Y_j] \quad (4.30)$$

$$\mathbf{Var}[P\&L - \Theta] = \mathbf{Var}[\sum_{j=1}^{N} Y_j] = \sum_{j=1}^{N} \mathbf{Var}[Y_j] \quad (4.31)$$

Due to the representation $Y_j = b_j U + \lambda_j U^2$ where U is $N(0,1)$ distributed, one concludes:

$$\begin{aligned}
\mathbf{E}[Y_j] &= b_j \mathbf{E}[U] + \lambda_j \mathbf{E}[U^2] = \lambda_j && (4.32)\\
\mathbf{E}[Y_j^2] &= b_j^2 \mathbf{E}[U^2] + 2 b_j \lambda_j \mathbf{E}[U^3] + \lambda_j^2 \mathbf{E}[U^4] = b_j^2 + 3\lambda_j^2 && (4.33)\\
\mathbf{Var}[Y_j] &= \mathbf{E}[Y_j^2] - \mathbf{E}[Y_j]^2 = b_j^2 + 2\lambda_j^2 && (4.34)
\end{aligned}$$

□

On the numerical computation of the characteristic function

To compute the characteristic function of the profit and loss without complex multiplications, one can rewrite the representation given in theorem 4.7. First, the square root can be rewritten as an exponential:

$$\frac{1}{\sqrt{1 - i2\lambda_j s}} = \exp\left(-\frac{1}{4}\ln(1 + 4s^2\lambda_j^2) + \frac{i}{2}\arctan(2\lambda_j s)\right) \tag{4.35}$$

where the co–domain of the arctan is $]-\frac{\pi}{2}, \frac{\pi}{2}[$. On the other hand, the fraction in the exponential may be rewritten as

$$\frac{b_j^2 s^2}{2(1 - i2\lambda_j s)} = \frac{b_j^2 s^2}{2(1 + 4\lambda_j^2 s^2)} + i\frac{b_j^2 \lambda_j s^3}{1 + 4\lambda_j^2 s^2} \tag{4.36}$$

Hence one obtains the following representation of the characteristic function, which can be evaluated by only one evaluation using the complex numbers, namely the valuation of the complex exponential function. All other operations are computations on $\mathbb{R}$ and hence easy to implement:

$$\begin{aligned}
\Phi_{P\&L} = &\exp\left(-\left(\sum_{j=1}^{N}\frac{1}{4}\ln(1 + 4s^2\lambda_j^2) + \frac{b_j^2 s^2}{2(1 + 4\lambda_j^2 s^2)}\right)\right) \times \\
&\exp\left(i\left(\sum_{j=1}^{N}\frac{1}{2}\arctan(2\lambda_j s) - \frac{b_j^2 \lambda_j s^3}{1 + 4\lambda_j^2 s^2}\right) - i\Theta s\right) \qquad (4.37)
\end{aligned}$$

4.1.5 Fourier Inversion

From chapter 3 one has the insight that any distribution is determined by its characteristic function and also some methods to compute the distribution has been provided. Since the characteristic function of the P&L is known, the application of the Fourier inversion techniques presented in 3.3 and 3.5 turn out to be very fruitful in the context of the Delta Gamma normal approximation.

Comparison with normal distribution

In practical situations, the number of risk factors N is rather high and therefore it is reasonable to assume that there are many risk factors which significantly contribute

to the $P\&L$. With regard to the law of large numbers, one would then expect that the $P\&L$ is almost normal distributed.
Hence if one would like to perform the Fourier inversion based on theorem 3.22, it is natural to choose the normal distribution as the comparing distribution. It is recommended to fit the mean μ and the variance σ^2 of this normal distribution to the first to moments of the $P\&L$:

$$\mu = \Theta + \sum_{j=1}^{N} \lambda_j \qquad \text{and} \qquad \sigma^2 = \sum_{j=1}^{N} 2\lambda_j^2 + b_j^2 \tag{4.38}$$

Hence the distribution function of the P&L is given by

$$F_{P\&L}(x) \;=\; \mathcal{N}(\frac{x-\mu}{\sigma}) - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\Phi_{P\&L}(s) - e^{i\mu s} e^{-\frac{1}{2}\sigma^2 s^2}}{is} e^{-isx}\, ds \tag{4.39}$$

Since the integrand is bounded and decays at least with $s^{-\frac{3}{2}}$, this integral always exist and may be computed by any integration method. The Value at Risk (VaR) to a certain probability q can be obtained by a bisection algorithm, since it is the unique root of the equation

$$F_{P\&L}(\text{VaR}) \;=\; q \tag{4.40}$$

Fast Fourier Transformation

One may also compute the density of the profit and loss distribution using the method presented in 3.5. Hence one performs the Fast Fourier Transformation algorithm on the characteristic function $\Phi_{P\&L}$. Since the mean μ and the variance σ^2 of the P&L are known, one has a good estimation of the sampling interval for the density. It is quite useful to choose the corresponding interval to be $[\mu - \alpha\sigma, \mu + \alpha\sigma]$ and typical values for α lie round about 10.
After the Fourier inversion the density $f_{P\&L}$ of the profit and loss is given on an equidistant grid which corresponds to the interval $[\mu - \alpha\sigma, \mu + \alpha\sigma]$. To obtain the q–Value at Risk, one has to determine VaR such, that

$$\int_{-\infty}^{\text{VaR}} f_{P\&L}(x)dx \;=\; q \tag{4.41}$$

This integration can be done by summing up the function values times the sampling distance Δx until the sum is q. This gives the the position of VaR. Since one has the density of the profit and loss distribution, one can also determine other risk measures, for example the conditional value at risk CVaR, which is given by

$$\text{CVaR} \;=\; \frac{1}{q} \int_{-\infty}^{\text{VaR}} x f_{P\&L}(x)\, dx \tag{4.42}$$

The computation time for the whole algorithm is summarized. For the linear algebra operations (Decoupling of the risk factors, Eigenvalue decomposition) the computational effort is at most $2N^3 + \mathcal{O}(N^2)$ *op*, in the case that the covariance matrix has

full rank. Let S be the number of sample points for the Fast Fourier Transformation, then the valuation of the characteristic function costs $\mathcal{O}(NS)$ operations and the Fourier transformation costs $\mathcal{O}(S \ln S)$. Hence the total costs are given by

$$\text{Comp. Time} \quad = \quad 2N^3 op. + \mathcal{O}(N^2) op. + \mathcal{O}(NS) op. + \mathcal{O}(S \ln S) op. \qquad (4.43)$$

So for usual applications $N \geq 100$ and $S \sim 2^{12}$ the linear algebra operations dominate the computation time. But if all presented improvements on the linear algebra operations are taken into account, the computation time is given by the effort to multiply two times two $N \times N$ matrices. Hence this algorithm is applicable even for very large portfolios.

4.2 Discussion of the Quadratic Approximation

In the previous section an algorithm to determine the profit and loss distribution of a portfolio over a certain time horizon has been presented. This algorithm is based on the assumption, that the $P\&L$ can be written as a quadratic equation of normal distributed random variables with a given covariance.

The standard approach to obtain the quadratic approximation of the portfolio function is to determine the Taylor polynomial of second degree at the actual state of the market. One reason why this method is quite popular is, that one usually determines the first two derivatives of an instrument, if one computes its price, see the introduction of chapter 1. In general this is only an approximation which is quite good in the neighbourhood of the current state of the market, but this approximation may be rather bad far away.

Two examples will be presented to underline this possible problem with the Taylor expansion. There are portfolios for which this approximation is reliable, but one can easily construct portfolios, for which the Taylor expansion yields dramatical wrong results.

4.2.1 Two Examples

First Example: A Stock and long Puts

Let us consider a portfolio which consists of:

- 1 stock with today's spot price 100 EUR and a volatility of 30%.
- 2 put options with a maturity of 3 months and a strike price of 90 EUR.
- 2 put options with a maturity of 3 months and a strike price of 110 EUR.

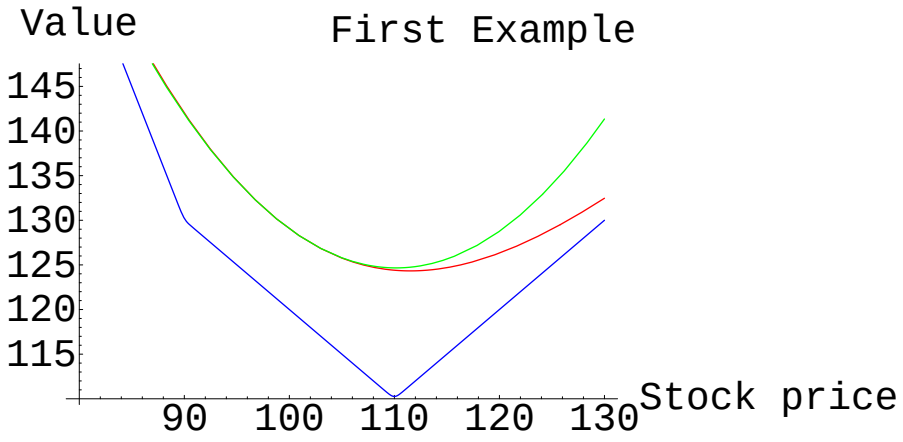

Figure 4.1: In the picture the payoff of the portfolio value, the quadratic approximation based on the first and second derivative, and the payoff of the portfolio in three months is plotted.

For this portfolio, the Value at Risk for a time horizon of one month is determined by the Delta normal approach, the Delta Gamma normal approach and the correct Value at Risk has been determined by an extensive Monte Carlo sampling, using the exact payoff function. The interest rate is assumed to be 0. The results are:

	Delta	Gamma	1% VaR	5% VaR
Delta normal	-0.86	not used	20.2 EUR	15.1 EUR
Delta Gamma normal	-0.86	0.085	8.1 EUR	8.0 EUR
Monte Carlo	not used	not used	7.8 EUR	7.7 EUR

Table 4.1: The Greeks and the Value at Risk of the first example, where the quadratic approximation is obtained by the Taylor approximation.

In this example, one can see that the Delta Gamma normal approach is much better than the Delta normal approach. The results obtained by the Delta Gamma normal method are comparable to the Monte Carlo results.

Second Example: Delta Hedged x^3 shaped Portfolio

Now the portfolio is changed a little bit. Instead of the long position of the put options with strike 90 EUR, this position becomes a short position, hence the portfolio is given by:

- 1 stock with today's spot price 100 EUR and a volatility of 30%.
- - 2 put options with a maturity of 3 months and a strike price of 90 EUR.
- 2 put options with a maturity of 3 months and a strike price of 110 EUR.

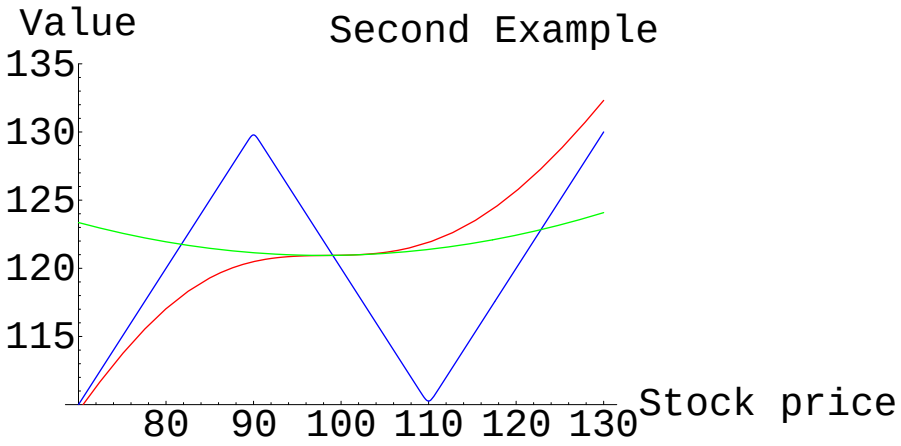

Figure 4.2: In the picture the payoff of the portfolio value, the quadratic approximation based on the first and second derivative, and the payoff of the portfolio in three months is plotted.

Again the one month Value at Risk is determined by three methods:

	Delta	Gamma	1% VaR	5% VaR
Delta normal	0.01	not used	0.45 EUR	0.39 EUR
Delta Gamma normal	0.01	0.006	0.24 EUR	0.24 EUR
Monte Carlo	not used	not used	1.89 EUR	0.96 EUR

Table 4.2: The Greeks and the Value at Risk of the second example, where the quadratic approximation is obtained by the Taylor approximation.

As one can see from figure 4.2 or from the Greeks in the table, the portfolio is Delta Gamma hedged. Hence the methods which base on Taylor approximations of the portfolio cannot detect the risk, which obviously becomes present if the stock falls. So an improvement of the quadratic approximation is necessary.

4.2.2 An Alternative Quadratic Approximation

The second example illustrates that the Delta Gamma normal approach to determine the Value at Risk can be rather bad. The problem is, that the Taylor expansion is the best approximation in an small area around one point, but it can diverge a good deal from a global point of view. Hence one may not approximate the portfolio by the Taylor polynomial of degree two, but by another quadratic form which fits better from a global point of view.
For the whole section the term *polynomial of degree n* includes also polynomials with an smaller degree, hence the x^n coefficient may be zero. With this convention, one can prove the

Lemma 4.9 *Let the three points (x_1, y_1), (x_2, y_2) and (x_3, y_3) be such that the condition $x_1 \neq x_2 \neq x_3 \neq x_1$ holds. Then there is one and only one polynomial of degree 2 which goes though these points and this polynomial is given by*

$$p(x) = y_1 \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + y_2 \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + y_3 \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} \tag{4.44}$$

Proof:
By construction of the polynomial holds $p(x_i) = y_i$ for $i = 1, 2, 3$ and $p(x)$ is a polynomial of degree 2. Any polynomial of degree 2 can uniquely be represented by $ax^2 + bx + c$ and the condition on a, b, c that this polynomial goes through the three points can be written as

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \tag{4.45}$$

The determinant of the matrix is given by $(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$ and is hence not zero by assumption. So a, b, c are uniquely determined and $p(x)$ is unique. □

Theorem 4.10 *Let $n_{\mu,\sigma}(x)$ be the normal density with mean μ and variance $\sigma^2 > 0$, and let f be a polynomial of degree 5. Define p to be the polynomial of degree two which goes through the three points $(x_i, f(x_i))$ with*

$$x_1 := \mu - \sigma\sqrt{3}, \qquad x_2 := \mu, \qquad x_3 := \mu + \sigma\sqrt{3} \tag{4.46}$$

Then holds:

$$\int_{-\infty}^{\infty} f(x) n_{\mu,\sigma}(x) dx = \int_{-\infty}^{\infty} p(x) n_{\mu,\sigma}(x) dx \tag{4.47}$$

Proof:
Without loss of generality one may assume that $\mu = 0$ and $\sigma = 1$, because the substitution $t = \frac{x-\mu}{\sigma}$ yields again the integral of a polynomial times the normal density.
Let $p_f(x)$ denote the polynomial of degree 2 with the condition $p_f(x_i) = f(x_i)$ for $i = 1, 2, 3$. From the construction of this polynomial it is clear, that the mapping $f \to p_f$ is linear. Since the integral is also linear, it is enough to show, that the theorem holds for $f(x) = x^k$ with $k = 0, 1, 2, 3, 4, 5$. For $k = 0, 1, 2$ the second order polynomial is the function itself and hence the theorem holds. The cases $k = 3, 4, 5$ can easily be verified:

$$\int_{-\infty}^{\infty} x^3 n_{0,1}(x) dx = 0 = \int_{-\infty}^{\infty} 3x\, n_{0,1}(x) dx = \int_{-\infty}^{\infty} p_{x^3}(x)\, n_{0,1}(x) dx \tag{4.48}$$

$$\int_{-\infty}^{\infty} x^4 n_{0,1}(x) dx = 3 = \int_{-\infty}^{\infty} 3x^2\, n_{0,1}(x) dx = \int_{-\infty}^{\infty} p_{x^4}(x)\, n_{0,1}(x) dx \tag{4.49}$$

$$\int_{-\infty}^{\infty} x^5 n_{0,1}(x) dx = 0 = \int_{-\infty}^{\infty} 9x\, n_{0,1}(x) dx = \int_{-\infty}^{\infty} p_{x^5}(x)\, n_{0,1}(x) dx \tag{4.50}$$

□

This theorem provides a quadratic approximation of a polynomial of degree 5, such that the integral with respect to the normal density equals for the two functions. Even if the theorem does not hold for general functions (e.g. $f(x) = x^6$), one can define the quadratic approximation for any function, since p is given by three points. The integral of the approximation with respect to the normal density is exact for polynomials of degree 5, hence one might expect that this quadratic approximation is more reasonable for a global approximation with respect to the normal density than the Taylor polynomial of second degree. So for quadratic approximations of a portfolio one should also consider this approximation instead of a Taylor expansion. In order to obtain this approximation one has to evaluate the portfolio for several states of the market, but the possibly better quadratic approximation justifies this effort.

From a quadratic approximation of a portfolio one can easily determine the first and the second derivative of this polynomial and these derivatives are the Delta and the Gamma of the portfolio for the quadratic approximation. With these Greeks one can perform the Delta Gamma normal algorithm exactly on the same way described before. Since stocks are usually assumed to be lognormal distributed, one can apply the idea of the theorem as follows. Let S_t be the actual stock price and T the time horizon for the value at risk computation. Then define the quadratic approximation as the function which coincides with the portfolio value at $t+T$ in the three points

$$x_1 = \mathbf{E}[S_{t+T}]e^{-\sigma\sqrt{3T}} \qquad x_2 = \mathbf{E}[S_{t+T}] \qquad x_3 = \mathbf{E}[S_{t+T}]e^{+\sigma\sqrt{3T}} \tag{4.51}$$

Hence the idea of a more global approximation also applies for the lognormal distribution.

4.2.3 Revisiting the Examples

To illustrate the idea of the approximation based on three sample points, the two examples from section 4.2.1 are computed again, but now the quadratic approximation is obtained by the technique described just before.

The First Example

The quadratic approximation of the first portfolio is shown in the picture. Since the payoff is x^2 shaped, there is no big difference to the Taylor approximation:

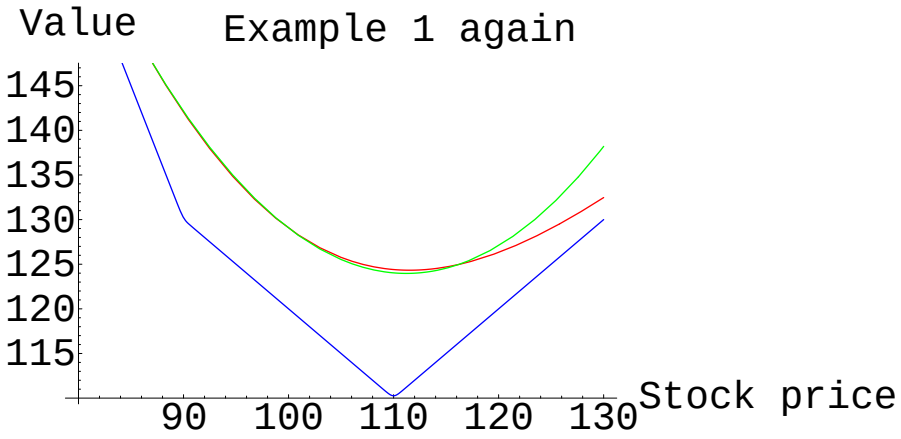

Figure 4.3: In the picture the value of the portfolio against the stock price is shown. Also the quadratic approximation based on three sample points and the payoff in three months is shown.

Again, the Value at Risk for a time horizon of one month is determined by the Delta normal approach, the Delta Gamma normal approach and the correct Value at Risk has been determined by an extensive Monte Carlo sampling. The results are comparable to the results obtained by the Taylor approximation of the portfolio:

	Delta	Gamma	1% VaR	5% VaR
Delta normal	-0.90	not used	21.0 EUR	15.7 EUR
Delta Gamma normal	-0.90	0.081	8.8 EUR	8.7 EUR
Monte Carlo	not used	not used	7.8 EUR	7.7 EUR

Table 4.3: The Greeks and the Value at Risk of the first example, where the quadratic approximation is obtained by the three suggested sampling points.

The Second Example

The quadratic approximation of the second portfolio obtained from the three sample points looks quite different from the second order Taylor approximation:

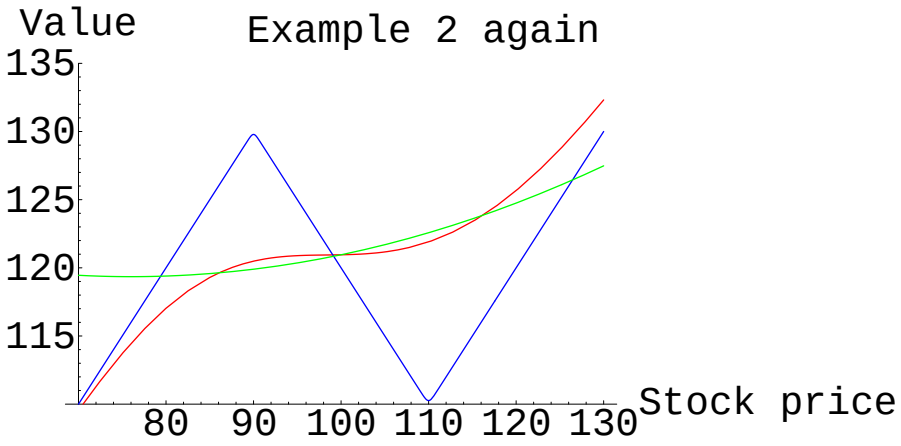

Figure 4.4: In the picture the value of the second portfolio against the stock price is shown. Also the quadratic approximation based on three sample points and the payoff in three months is shown.

Again the 1 month Value at Risk can determined by three methods:

	Delta	Gamma	1% VaR	5% VaR
Delta normal	0.13	not used	2.9 EUR	2.1 EUR
Delta Gamma normal	0.13	0.006	1.4 EUR	1.3 EUR
Monte Carlo	not used	not used	1.9 EUR	0.96 EUR

Table 4.4: The Greeks and the Value at Risk of the second example, where the quadratic approximation is obtained by the three suggested sampling points.

In this case, the Delta Gamma normal approach, where the quadratic approximation is based on three sampling points, yields a better result for the Value at Risk than the Taylor approximation, which is obvious if one compares table 4.2 with table 4.4. The reason for this improvement is that the x^3 shaped portfolio payoff can at least in some sense be approximated by a linear function.

4.3 Discussion of the Normal Distribution Assumption

Beside the assumption that the portfolio may be approximated by a quadratic form, the Delta Gamma normal approach also presumes, that the (relative) returns of the financial underlyings are normal distributed. This distribution assumption will be analysed in brief based on a time series of daily data from January 2001 to December 2002 of the four underlyings

- DAX (Deutscher Aktienindex),

- Dow Jones, more precisely the Dow Jones Industrial Average,
- stock of the Deutsche Telekom AG (quoted at Xetra), and
- the Euro US-$ exchange rate.

Hence two asset indices, one single stock and one foreign exchange rate are studied. From the time series of each asset S_{T_i}, a time series of returns R_i is introduced by:

$$R_i \quad := \quad \ln(S_{T_{i+1}}) - \ln(S_{T_i}) \tag{4.52}$$

For each of the four series of returns the mean and the variance are determined by the standard estimators:

$$\mathbf{E}[R] \quad = \quad \frac{1}{N}\sum_{i=1}^{N} R_i \tag{4.53}$$

$$\mathbf{Var}[R] \quad = \quad \frac{1}{N-1}\left(\sum_{i=1}^{N} R_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} R_i\right)^2\right) \tag{4.54}$$

Since in financial markets all terms are usually expressed on an annual basis, one may introduce the annualized mean by $250 * \mathbf{E}[R]$ and the volatility $\sqrt{250 * \mathbf{Var}[R]}$. The number 250 represents the average number of trading days a year. A first comparison between the empirical distribution of the R_i and the normal distribution with the same mean and variance can be performed by printing quantile–quantile plots.

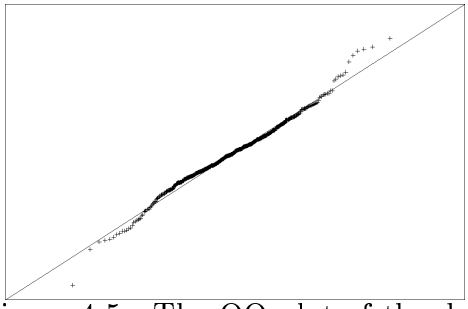

Figure 4.5: The QQ plot of the daily DAX returns and the normal distribution with an annualized mean of -40% and a volatility of 35%.

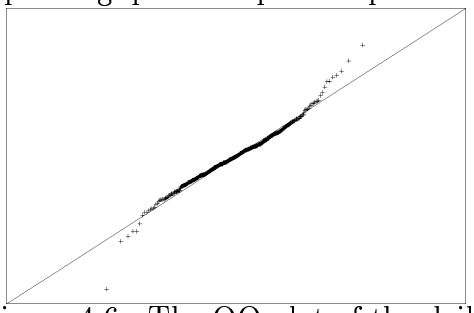

Figure 4.6: The QQ plot of the daily Dow Jones Industrial Average returns and the normal distribution with an annualized mean of -12% and a volatility of 23%.

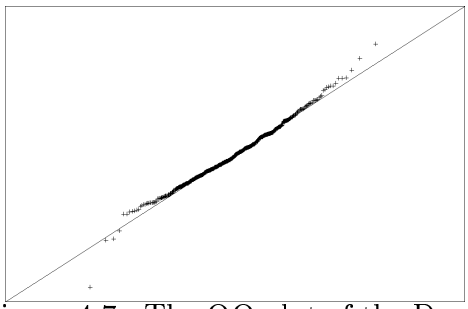

Figure 4.7: The QQ plot of the Deutsche Telekom stock returns and the normal distribution with mean -46% and a volatility of 58%.

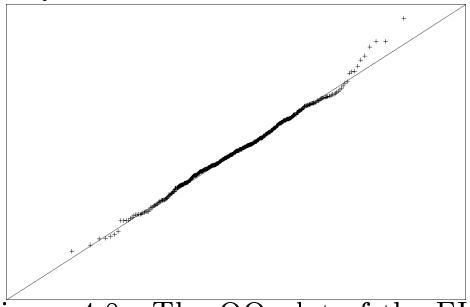

Figure 4.8: The QQ plot of the EURO US-$ returns and the normal distribution with an annualized mean of 5% and a volatility of 10%.

These four pictures show that the returns of the financial assets are almost normal distributed. However, especially at the tails, the normal distribution assumption is violated and distributions with heavy–tails seem to be more realistic.
More detailed studies on the returns of financial data have been done and the presence of heavy–tails has been confirmed. Therefore the use of other than the normal or lognormal distributions in finance are discussed, for example in [30] the class of hyperbolic distributions in finance is analyzed, and the impact of the hyperbolic distribution assumption for Value at Risk computations is studied in [7]. In [31] also the fit of financial returns to generalized hyperbolic distributions has been studied. The modelling of a financial price process by a generalized inverse Gaussian diffusion and the impact of portfolio optimization is the subject of [33].
There is also an approach which is independent from a certain distribution assumption. If one concentrates on the computation of the value at risk only, one is interested in the tail of a distribution and the extreme value theory may be applied. An approach based on this idea is the *peaks over threshold (POT)* method, which is described e.g. in [11, 58].
Another very interesting approach using a fat tail distribution for risk management is suggested by Glasserman, Heidelberger and Shahabuddin in [43], where the risk factors are assumed to be multivariate student distributed. In this approach it is also assumed, that the portfolio may be written as a quadratic form of the risk factors. This method is shortly discussed, because the improving techniques for the Delta Gamma normal approach also apply in this setting.

4.3.1 The Delta Gamma t_f Approach

In the Delta Gamma t_f approach, the profit and loss is modelled by

$$P\&L \quad = \quad \Theta + \Delta X + X^\top \Gamma X \tag{4.55}$$

where Θ, Δ and Γ describe the quadratic form of the portfolio as in equation (4.7). In this setting X is multivariate $t_f(0, \Sigma)$ distributed. By a decoupling of the risk factors using the generalized non–square Cholesky decomposition algorithm presented in chapter 2, one can represent X in distribution by

$$X \quad = \quad \frac{P^\top C S}{\sqrt{Y/f}} \tag{4.56}$$

where S is a vector of iid. distributed N(0,1) variables and Y is a random variable, which is independent from S and $\chi^2(f)$ distributed. Hence the profit and loss can be written as

$$P\&L \quad = \quad \Theta + \sqrt{\frac{f}{Y}} \tilde{\Delta}^\top S + \frac{f}{Y} S^\top \tilde{\Gamma} S \tag{4.57}$$

By exactly the same transformation described in the *Eigenvalue Decomposition* step of the Delta Gamma normal algorithm one obtains

$$P\&L \quad = \quad \Theta + \sqrt{\frac{f}{Y}} b^\top U + \frac{f}{Y} U^\top \Lambda U \tag{4.58}$$

where Λ is a diagonal matrix. Until here there is no essential difference to the Delta Gamma normal approach. Hence the computation time for the operations which have been performed is again given by $2N^3 + \mathcal{O}(N^2)op$. Even if the following part to determine the value at risk becomes more costly now, the Delta Gamma t_f method can also be applied for very large portfolios.
The addends in the sum

$$P\&L = \Theta + \sum_{j=1}^{N} b_j \sqrt{\frac{f}{Y}} U_j + \frac{f}{Y} \lambda_j U_j^2 \tag{4.59}$$

are uncorrelated but they are dependent! Hence the trick that one can write the characteristic function as a product of the characteristic functions of each single addend does not work here. In this situation one may define for each $z \in \mathbb{R}$ the random variable Z_z by

$$Z_z := \frac{Y}{f}(P\&L - z) = \frac{Y}{f}(\Theta - z) + \sum_{j=1}^{N} \left(\sqrt{\frac{Y}{f}} b_j U_j + \lambda_j U_j^2 \right) \tag{4.60}$$

Since $Y > 0$ almost sure, one easily obtains the relation $P[P\&L \leq z] = P[Z_z \leq 0]$. Hence the value at risk with respect to the quantile q of the portfolio is given by the number z for which the relation $P[Z_z \leq 0] = q$ holds. This equation can be numerically solved by a bisection algorithm for example and one only needs an algorithm to compute $P[Z_z \leq 0]$.
The computation of $P[Z_z \leq 0]$ can be done via the characteristic function of Z_z, which is given by

$$\Phi_{Z_z}(s) = \mathbf{E}[e^{isZ_z}] = \mathbf{E}\left[\mathbf{E}[e^{isZ_z}|Y]\right] \tag{4.61}$$

$$= \mathbf{E}\left[\left(\prod_{j=1}^{N} \frac{1}{\sqrt{1 - i2\lambda_j s}} \right) \exp\left(-Y \left(\frac{i(z-\Theta)s}{f} + \sum_{j=1}^{N} \frac{b_j^2 s^2}{2f(1 - i2\lambda_j s)} \right) \right) \right] \tag{4.62}$$

This relation directly follows from the characteristic function of the Delta Gamma normal approach, see theorem 4.7. Since Y is $\chi^2(f)$ distributed, one obtains for the remaining expectation

$$\Phi_{Z_z}(s) = \left(\prod_{j=1}^{N} \frac{1}{\sqrt{1 - i2\lambda_j s}} \right) \left(1 + \frac{2i(z-\Theta)s}{f} + \sum_{j=1}^{N} \frac{b_j^2 s^2}{f(1 - i2\lambda_j s)} \right)^{-\frac{f}{2}} \tag{4.63}$$

By one of the Fourier inversion techniques discussed in chapter 3 one obtains the distribution of Z_z and hence one also knows $P[Z_z \leq 0]$. Note that for the bisection one has to evaluate the characteristic functions for several z, but the product and sum in the expression for Φ_{Z_z} do not contain z and hence the valuation of the characteristic function for several values of z can be done efficiently.

4.4 Conclusion

A very efficient algorithm to obtain the profit and loss distribution of a portfolio within the Delta Gamma normal approach has been presented. A careful analysis of the linear algebra operations and the use of the generalized non–square Cholesky decomposition algorithm show, that the computation time for large portfolios is equivalent with the time needed to multiply two times two $N \times N$ matrices, where N denotes the number of risk factors.
Another aspect of the presented algorithm is the Fourier inversion which is discussed in detail in chapter 3. The algorithm yields the whole profit and loss distribution and hence this method can be used to determine not only the value at risk, but almost all measures of risk, since these a usually some functionals of the distribution.
In order to obtain the quadratic approximation of a portfolio, it is a very popular method in practice to take take Taylor expansions of the portfolio; but also an other method based on an integration theorem has been presented. There are even more approaches to determine the quadratic approximation [77], and a discussion of some methods can be found in [62]. Two simple portfolios are studied and the second example was constructed in such a way, that its payoff is x^3 shaped. For such a portfolio a Taylor based quadratic approximation is very bad and the approximation based on three sampling points yields much better results. As a recommendation for risk manager, one should run the Delta Gamma normal approach with both kinds of quadratic approximations. A significant difference in the results is a clear indication that the quadratic approximation of the portfolio is rather bad. A more carefully analysis of the risk of such a portfolio has to be performed, for example by Monte Carlo methods.
In [43] the Delta Gamma t_f approach was presented. It has been shown, that the optimization methods introduced in the context of the Delta Gamma normal approach also apply in this context. Hence there is a first heavy–tailed approach to assess the market risk of large portfolios with almost closed form solutions. From the point of view of a risk manager it could be rather interesting how much the value at risk or other risk ratios change, for either student distributed risk factors with f degrees of freedom or normal distributed risk factors, which is the special case for $f \to \infty$. Due to different values of f one possibly obtains different values for the value at risk or the conditional value at risk and the size of these deviations is an indication, how big the influence of a wrong model for the risk computations will be.
For real world portfolios Delta Gamma normal or Delta Gamma t_f are only approximations, and as it was shown in the second example for the normal case, that the value at risk computations may be rather wrong, if the portfolio cannot be approximated by a quadratic form properly. For such portfolios Monte Carlo based approaches have been proposed. In order to use such methods efficiently, it is very useful to have an approximative solution, which can be used for variance reduction [40, 41, 42]. In this context the presented fast and numerical stable algorithms are a key for more improved techniques to assess market risk.

Chapter 5

CreditRisk$^+$ and an Extension to Incorporate Market Risk[1]

A quite popular model to describe credit risk is CreditRisk$^+$, which has been presented by Credit Suisse First Boston in 1997 [19]. This is a two state model which is quite natural from point of view of a buy–and–hold investor who is only interested in the states *default* and *no default*. In the original description it was assumed that each loss given default is an integer multiple of a so-called basic loss unit and the mathematical description of the model is based on probability generating functions. The first numerical method to handle this model was based on Panjer recursions, but these may be instable. Other techniques for the CreditRisk$^+$ model have been developed, e.g. the *saddlepoint approximation* [46] or the *nested calculation* [38]. In section 5.2.1 the numerical stability of the nested calculation will be proven.
The CreditRisk$^+$ model will be described in terms of characteristic functions instead of probability functions. Since the proper choice of the basic loss unit may be crucial, it is an advantage of this alternative presentation that no basic loss unit has to be introduced. The result of this analysis is the characteristic function of the loss of a loan portfolio. In order to obtain the distribution of the credit loss, a Fourier inversion of the characteristic function has to be performed. The characteristic function is not integrable, but two of the techniques discussed in chapter 3 may be applied to obtain the distribution of the credit loss. Hence an alternative algorithm is established, which can be used for efficient computations in the framework of CreditRisk$^+$.
Since the result of this algorithm is the loss distribution, one can easily compute some functionals of the loss. For risk management purposes, the main functionals are some risk measures like the Value at Risk or the Worst Conditional Expectation. But of course, almost any other functional can be determined if the distribution of the loan portfolio is known.
In the CreditRisk$^+$ model the sector variables are assumed to be independent Gamma distributed. From a practitioners point of view, this assumption is too rigorous, since one would like to introduce the sectors due to general classes of business and typical sectors could be *construction*, *banking*, *utility industry*, *transportation*, etc. Also economic sectors like gross domestic product or the business activity of certain countries or currency areas can be taken into account. Using such sectors, it is

[1] Parts of this chapter have been presented in [86, 90] and will be published in [87, 91, 92].

quite easy to estimate the sector affiliations of each obligor, but these sectors are obviously not independent. One way to abstain from the independence assumption is presented in [15] and this study showed that there is an effect of dependent sector variables on the variance of the credit loss. A more general approach to handle dependent sector variables will be presented in section 5.3. Since it is difficult in general to describe the dependency of random variables completely, it is suggested for the application in practice to introduce dependent lognormal sector variables, because the dependency of lognormal random variables can easily be described by a covariance matrix. The resulting model is more complex and the valuation of the characteristic function of the credit loss can be done by Monte Carlo only, but the computational effort is tolerable.

The CreditRisk$^+$ model gives an answer to the question of the size of the losses *at the end* of a fixed period, but one would like to analyse the process of the losses *during* this period. Therefore a model with lognormal sector processes is introduced in section 5.4. The description of the default events in this model arise quite natural according to the CreditRisk$^+$ model.

With a model which allows a time continuous description of credit risk, one has made a large step to combine market and credit risk. The next step is done in section 5.5, where a a portfolio valuation in a combined model is introduced. In this model, the credit spreads are modelled by the sector processes. The profit and loss distribution of this portfolio can be obtained with the same techniques presented in the previous sections. One can consider two special cases of this model. First, if there is no market risk one obtains the model presented in section 5.4. In the special case, that the portfolio contains no credit risk, one gets the well known Delta normal approach to determine market risk, hence a link between credit risk and market risk has been established.

5.1 The CreditRisk$^+$ Model

In this section the CreditRisk$^+$ model introduced by Credit Suisse First Boston in 1997 [19] is presented. Usually the descriptions of this model are based on probability generating functions and therefore the introduction of a basic loss unit is necessary. As it is explained in chapter 3, the characteristic function is a general tool to describe real–valued random distributions. So the CreditRisk$^+$ model is presented in terms of characteristic functions and no basic loss unit has to be introduced. Also a specific risk sector, the so–called "idiosyncratic risk", is introduced as proposed in the CreditRisk$^+$ manual [19][A 12.3].

5.1.1 Introduction of the Model

The aim of the following analysis is to characterize the losses which may occur in a loan portfolio with N obligors. The credit loss of the portfolio is given by

$$X := \sum_{j=1}^{N} I^j L^j \tag{5.1}$$

where L^j denotes the loss if the jth obligor defaults. L^j is the so–called *loss given default* and it assumed, that L^j is deterministic. I^j is the default indicator of the jth obligor, where $I^j = 1$ indicates the default of the obligor and $I^j = 0$ its survival at some fixed time horizon T.
To model the default events, the time of default of obligor j is given by Y^j and Y^j is exponential distributed with intensity λ^j [2]. These exponential random variables are for given λ^j independent, hence

$$I^j = \begin{cases} 1 & \text{if } Y^j \leq T \\ 0 & \text{else} \end{cases} \tag{5.2}$$

The reliability of obligor j is described by a rating parameter p^j and it will be shown in lemma 5.2, that p^j is – at least in approximation for small p^j – the probability that the obligor j will default within one year.
In order to model dependencies between the obligors, K sectors are introduced. The affiliation of each obligor to these sectors is given by a_k^j for $(k = 1, \ldots K)$ and a_0^j denotes the so–called *idiosyncratic risk*. These affiliations have to fulfill the conditions:

$$a_k^j \geq 0 \qquad \forall\ j = 1, \ldots, N\ ;\ k = 0, \ldots K \tag{5.3}$$

$$\sum_{k=0}^{K} a_k^j = 1 \qquad \forall\ j = 1, \ldots, N \tag{5.4}$$

The sectors in this model are described by independent Gamma distributed random variables R^k with expectation 1 and variance σ_k^2. The parameter $\sigma_k > 0$ is the volatility of the kth sector. With these ingredients, the default intensity of the jth obligor is defined by

$$\lambda^j := p^j \left(a_0^j + \sum_{k=1}^{K} a_k^j R^k \right) =: p^j \tilde{\lambda}^j \tag{5.5}$$

Remark 5.1 *By definition, the following statements hold:*

$$\tilde{\lambda}^j > 0 \qquad \textit{almost sure} \tag{5.6}$$

$$\mathbf{E}[\tilde{\lambda}^j] = 1 \tag{5.7}$$

Lemma 5.2 *For small p^j, the one year default probability is given by p^j:*

$$P[Y^j \leq 1] = p^j + \mathcal{O}\left((p^j)^2\right) \tag{5.8}$$

Proof:
The default probability of obligor j conditioned on the state of the sector variables R is given by:

$$P[Y^j \leq 1 | R] = 1 - e^{-\lambda^j} = 1 - e^{-p^j \tilde{\lambda}^j} \tag{5.9}$$

One can take the expectation over R and expand the exponential function:

$$P[Y^j \leq 1] = \mathbf{E}[P[Y^j \leq 1|R]] = \mathbf{E}[1 - e^{-p^j \tilde{\lambda}^j}] \tag{5.10}$$

$$= p^j \mathbf{E}[\tilde{\lambda}^j] + \mathcal{O}\left((p^j)^2\right) = p^j + \mathcal{O}\left((p^j)^2\right) \tag{5.11}$$

□

[2] To define the exponential distribution it is assumed that $\lambda^j > 0$. But all following computations go through, if one defines $I^j = 0$ if $\lambda^j = 0$.

Remark 5.3 *For a bank, typical values are $N \approx 10,000$ or even bigger, $K \lesssim 100$ and the losses given default L^j can range from a few hundred Euro for credit card exposure up to more than a billion Euro for large company loans. A typical value for the time horizon T in this context is 1 year.*

Assumption 5.4 *In the whole chapter it is assumed, that the rating parameters p^j are small. Hence there is a low probability for each obligor to default within one year.*

5.1.2 The Characteristic Function of the Loss Distribution

Lemma 5.5 *For all $x \geq 0$ and $s \in \mathbb{R}$ holds:*

$$\left| \left(e^{is}(1-e^{-x}) + e^{-x}\right) - \left(e^{-x(1-e^{is})}\right) \right| \quad \leq \quad 3x^2 \tag{5.12}$$

Proof:
Fix $s \in \mathbb{R}$ and define the function

$$f_s(x) \quad := \quad e^{is}(1-e^{-x}) + e^{-x} - e^{-x(1-e^{is})} \tag{5.13}$$

$f_s(x)$ is two times continuous differentiable; the first two derivatives are given by:

$$f_s'(x) \quad = \quad e^{is}e^{-x} - e^{-x} + (1-e^{is})e^{-x(1-e^{is})} \tag{5.14}$$

$$f_s''(x) \quad = \quad -e^{is}e^{-x} + e^{-x} - (1-e^{is})^2 e^{-x(1-e^{is})} \tag{5.15}$$

Since $f_s(0) = 0$, $f_s'(0) = 0$ and $|f_s''(x)| \leq 6$ for all $x \geq 0$, one obtains the statement of the lemma by the Taylor theorem. □

This lemma will be used to compute the characteristic function of X using the assumption that the annual default probabilities of each obligor are small. The characteristic function of X conditioned on R is given by:

$$\Phi_{X|R}(s) \quad = \quad \prod_{j=1}^{N} \Phi_{L^j I^j|R}(s) = \prod_{j=1}^{N} \Phi_{I^j|R}(L^j s) \tag{5.16}$$

For the characteristic function of I^j conditioned on R one obtains:

$$\Phi_{I^j|R}(s) \quad = \quad \mathbf{E}[e^{isI^j}|R] = \int_0^T e^{is1}\lambda_j e^{-\lambda^j t}dt + \int_T^\infty e^{is0}\lambda^j e^{-\lambda^j t}dt \tag{5.17}$$

$$= \quad e^{is}(1-e^{-\lambda^j T}) + e^{-\lambda^j T} \tag{5.18}$$

Since $\lambda^j T$ is proportional to p^j and hence small, one may approximate by lemma 5.5:

$$\Phi_{I^j|R}(s) \quad \approx \quad \exp\left(\lambda^j T(e^{is}-1)\right) \tag{5.19}$$

This is the characteristic function of the Poisson distribution. Let $f^j(s)$ denote the error of this approximation, then this error is bounded by $|f^j(s)| \leq 3(\lambda^j T)^2$. One thus obtains together with (5.5) and (5.16):

$$\Phi_{X|R}(s) \quad = \quad \exp\left(\sum_{j=1}^{N} p^j \left(a_0^j + \sum_{k=1}^{K} a_k^j R^k\right) T(e^{iL^j s} - 1)\right) + F(s) \tag{5.20}$$

The error term $F(s)$ comes from the products of the characteristic functions $\Phi_{I^j|R}(s)$, each with an bounded error $f^j(s)$. Since characteristic functions are bounded by 1, the overall error $F(s)$ of the conditioned characteristic function $\Phi_{X|R}(s)$ is bounded by powers of $(\lambda^j T)^2$. The conditioning on R will be solved by taking the expectation and terms of order $\mathcal{O}((p^j)^2)$ will be neglected[3]. Recall, that R^k are independent gamma distributed with mean 1 and variance σ_k^2, hence:

$$\begin{aligned}
\Phi_X(s) &= \mathbf{E}\left[\Phi_{X|R}(s)\right] && (5.21)\\
&= \mathbf{E}\left[\exp\left(\sum_{j=1}^{N} p^j \left(a_0^j + \sum_{k=1}^{K} a_k^j R^k\right) T(e^{iL^j s}-1)\right)\right] && (5.22)\\
&= e^{\sum_{j=1}^{N} a_0^j p^j T(e^{iL^j s}-1)} \mathbf{E}\left[\exp\left(\sum_{k=1}^{K} R^k \sum_{j=1}^{N} p^j a_k^j T(e^{iL^j s}-1)\right)\right] && (5.23)\\
&= e^{\sum_{j=1}^{N} a_0^j p^j T(e^{iL^j s}-1)} \prod_{k=1}^{K} \mathbf{E}\left[\exp\left(\sum_{j=1}^{N} p^j a_k^j T(e^{iL^j s}-1) R^k\right)\right] && (5.24)
\end{aligned}$$

By well known properties of the gamma distribution one obtains

$$\Phi_X(s) = \exp\left(\sum_{j=1}^{N} a_0^j p^j T(e^{iL^j s}-1)\right) \prod_{k=1}^{K} \left(\frac{1}{1+\sigma_k^2 \sum_{j=1}^{N} p^j a_k^j T(1-e^{iL^j s})}\right)^{\frac{1}{\sigma_k^2}} \quad (5.25)$$

Using the main branch of the logarithm, one can rewrite the characteristic function:

$$\Phi_X(s) = \exp\left(\sum_{j=1}^{N} a_0^j p^j T(e^{iL^j s}-1) - \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \ln\left[1+\sigma_k^2 T \sum_{j=1}^{N} a_k^j p^j (1-e^{iL^j s})\right]\right) \quad (5.26)$$

Remark on Independent Random Loss Given Defaults

In the CreditRisk$^+$ model, the loss given defaults are deterministic. Under the assumption, that the default probabilities are small, which is made anyway, one can easily include stochastic, but independent random L^j. Neglecting all terms of order $\mathcal{O}((\lambda^j)^2)$, one obtains from (5.19) the characteristic function of the loss of the jth obligor conditioned on the sector variables and the loss given default:

$$\Phi_{L^j I^j|R}(s) = \exp\left(\lambda^j T(e^{iL^j s}-1)\right) \approx 1+\lambda^j T(e^{iL^j s}-1) \quad (5.27)$$

Taking the expectation over the j loss given default yields:

$$\begin{aligned}
\Phi_{L^j I^j|R}(s) &= \mathbf{E}[1+\lambda^j T(e^{iL^j s}-1)|R] = 1+\lambda^j T(\Phi_{L^j}(s)-1) && (5.28)\\
&\approx \exp\left(\lambda^j T(\Phi_{L^j}(s)-1)\right) && (5.29)
\end{aligned}$$

[3]The error analysis holds for any distribution of R as long as all moments of R exist. In the case of the Gamma distribution, one can show more easily, that the error is of order $\mathcal{O}((p^j)^2)$. However, for generalizations of the CreditRisk$^+$ model it is useful to have an error analysis which is independent of certain properties of the distribution.

and $\Phi_{L^j}(s)$ denotes the characteristic function of the j loss given default. Using this approximation in the calculations of the CreditRisk$^+$ model, one thus obtains the following approximation of the characteristic function for the credit loss X with independent stochastic loss given defaults:

$$\begin{aligned}\Phi_X(s) &= \exp\left(\sum_{j=1}^{N} a_0^j p^j T(\Phi_{L^j}(s)-1)\right. \\ &\quad \left. -\sum_{k=1}^{K}\frac{1}{\sigma_k^2}\ln\left[1+\sigma_k^2 T\sum_{j=1}^{N} a_k^j p^j(1-\Phi_{L^j}(s))\right]\right) \qquad (5.30)\end{aligned}$$

Of course, one can apply the Fourier inversion methods presented in chapter 3 on this characteristic function. But in the following it is assumed that the loss given defaults are deterministic.

5.1.3 The First Moments of the Credit Loss

The cumulant generating function $C_X(s)$ of a random variable X is defined by $C_X(s) := \ln \mathbf{E}[e^{sX}]$, if this expectation exists and then holds the general relationship $C(s) = \ln \Phi(-is)$ between the cumulant generation function and the characteristic function. Hence, the cumulant generating function of the credit loss is given by:

$$C_X(s) = \sum_{j=1}^{N} a_0^j p^j T(e^{L^j s}-1) - \sum_{k=1}^{K}\frac{1}{\sigma_k^2}\ln\left[1+\sigma_k^2 T\sum_{j=1}^{N} a_k^j p^j(1-e^{L^j s})\right] \qquad (5.31)$$

The cumulant generating function can be used to prove the

Lemma 5.6 *The first two moments of X are given by:*

$$\mathbf{E}[X] = T\sum_{j=1}^{N} p^j L^j \qquad (5.32)$$

$$\mathbf{Var}[X] = T\sum_{j=1}^{N} p^j (L^j)^2 + T^2\sum_{k=1}^{K}\sigma_k^2\left(\sum_{j=1}^{N} a_k^j p^j L^j\right)^2 \qquad (5.33)$$

$$\mathbf{E}[X^2] = T\sum_{j=1}^{N} p^j (L^j)^2 + T^2\sum_{k=1}^{K}\sigma_k^2\left(\sum_{j=1}^{N} a_k^j p^j L^j\right)^2 + T^2\left(\sum_{j=1}^{N} p^j L^j\right)^2 \qquad (5.34)$$

Proof:
Compute the first and second derivative of $C_X(s)$ to determine the first two cumulants of X:

$$C'(s) = T\sum_{j=1}^{N} a_0^j p^j e^{L^j s} L^j + \sum_{k=1}^{K}\frac{T\sum_{j=1}^{N} a_k^j p^j e^{L^j s} L^j}{1+\sigma_k^2 T\sum_{j=1}^{N} a_k^j p^j (1-e^{L^j s})} \qquad (5.35)$$

$$C''(s) = T\sum_{j=1}^{N} a_0^j p^j e^{L^j s}(L^j)^2 + \sum_{k=1}^{K} \frac{T\sum_{j=1}^{N} a_k^j p^j e^{L^j s}(L^j)^2}{1+\sigma_k^2 T\sum_{j=1}^{N} a_k^j p^j (1-e^{L^j s})} + \sum_{k=1}^{K} \frac{\sigma_k^2 \left(T\sum_{j=1}^{N} a_k^j p^j e^{L^j s} L^j\right)^2}{\left(1+\sigma_k^2 T\sum_{j=1}^{N} a_k^j p^j (1-e^{L^j s})\right)^2} \tag{5.36}$$

Note, that the first cumulant is the expectation and the second cumulant is the variance. So the valuation of these expressions at $s = 0$ yields:

$$\mathbf{E}[X] = T\sum_{j=1}^{N} a_0^j p^j L^j + T\sum_{j=1}^{N}\sum_{k=1}^{K} a_k^j p^j L^j = T\sum_{j=1}^{N} p^j L^j \tag{5.37}$$

$$\mathbf{Var}[X] = T\sum_{j=1}^{N} p^j (L^j)^2 + T^2 \sum_{k=1}^{K} \sigma_k^2 \left(\sum_{j=1}^{N} a_k^j p^j L^j\right)^2 \tag{5.38}$$

$$\mathbf{E}[X^2] = \mathbf{Var}[X] + \mathbf{E}[X]^2 \tag{5.39}$$

□

Using the same idea one can of course also compute the next cumulants, but their explicit knowledge is not so important for the following analysis. Note, that the introduction of the idiosyncratic risk does not effect the mean of X, but there is an influence on the variance of the resulting distribution; it reduces the variance.

5.2 Obtaining the Credit Loss Distribution

There are at least two possibilities to obtain the distribution of X. The first method is based on nested calculations proposed by Giese [38]. This approach is based on calculations using the probability generating function and hence a basic loss unit has to be introduced. Giese claimed, that the nested calculation method is numerical stable, but a proof of the stability remained to be done and is presented below.
The second way to obtain the distribution is to perform a Fourier inversion on the characteristic function based on the methods presented in chapter 3. The advantage of this technique is, that no basic loss unit has to be introduced.

5.2.1 Series Expansion of the Probability Generating Function[4]

The nested calculation method is based on probability generating functions and hence the credit loss has to take values in $\mathbb{N}$. Therefore a basic loss unit ν is introduced and each loss given default L^j is expressed by integer multiples l^j of the basic

[4]The numerical stability of the nested calculation has been obtained together with Hermann Haaf and John Schoenmakers in [86] and will also be published in [87].

loss unit ν:

$$L^j = l^j \nu \qquad l^j = 1, 2, \ldots \tag{5.40}$$

Then the credit loss of the loan portfolio takes values which are integer multiples of ν and hence the probability generating function $G_X(z)$ of X exists. Due to the relation $G_X(z) = \Phi_X(-i \ln z)$ one obtains:

$$G_{X/\nu}(z) = \exp\left(\sum_{j=1}^{N} a_0^j p^j T(z^{l^j} - 1) - \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \ln\left[1 + \sigma_k^2 T \sum_{j=1}^{N} a_k^j p^j (1 - z^{l^j})\right]\right) \tag{5.41}$$

The idea of the nested calculation approach is to find the series expansion of $G_{X/\nu}(z)$ at $z = 0$, since this expansion has the structure

$$G_{X/\nu}(z) = \sum_{n=0}^{\infty} P[X = n\nu]\, z^n \tag{5.42}$$

Thus the series expansion of $G_{X/\nu}$ gives the (discrete) distribution of the credit loss. Before the steps of the nested calculation algorithm are presented, one lemma has to be introduced, which is important for the algorithm.

Lemma 5.7 *Let the following the power series converge for $|z| < \epsilon$ with $\epsilon > 0$:*

$$M(z) := \sum_{j=0}^{\infty} m_j z^j \tag{5.43}$$

$$C(z) := \sum_{j=0}^{\infty} \kappa_j z^j \tag{5.44}$$

Let further hold $M(z) = \exp(C(z))$ for all $|z| < \epsilon$. Then holds:

$$m_0 = \exp(\kappa_0) \tag{5.45}$$

$$m_n = \sum_{k=1}^{n} \frac{k}{n} m_{n-k} \kappa_k \qquad n = 1, 2, \ldots \tag{5.46}$$

And the inverse relations are given by

$$\kappa_0 = \ln(m_0) \tag{5.47}$$

$$\kappa_n = \frac{1}{m_0}\left(m_n - \sum_{k=1}^{n-1} \frac{k}{n} m_{n-k} \kappa_k\right) \tag{5.48}$$

Proof:
The relation $m_0 = \exp(\kappa_0)$ follows directly from $M(0) = \exp(C(0))$. For all n holds:

$$M^{(n)}(0) = n!\, m_n \qquad \text{and} \qquad C^{(n)}(0) = n!\, \kappa_n \tag{5.49}$$

On the other hand, for $n \geq 1$ one obtains:

$$M^n(z) = \frac{d^n}{dz^n}\exp(C(z)) = \left(\frac{d}{dz}\right)^{n-1} M(z)\cdot C'(z) \tag{5.50}$$

By the product rule of Leibniz one obtains:

$$M^n(z) = \sum_{k=0}^{n-1}\binom{n-1}{k} M(z)^{(n-1-k)}(z)C^{(1+k)}(z) \tag{5.51}$$

Evaluation at $z = 0$ yields the second assertion, since the faculties cancel out. The inverse relations are just the inverse of the first two equations. □

Remark 5.8 *This lemma also proves the well–known relation between moments and cumulants. Let X be a random variable such that all moments exist and let the moments fulfill the Carleman condition. Then the moment generating function and the cumulant generating function are analytical at $z = 0$ and their series expansions are given by*

$$M_X(z) = 1 + \sum_{n=1}^{\infty}\frac{m_n}{n!}z^n \qquad \text{and} \qquad C_X(z) = \sum_{n=1}^{\infty}\frac{\kappa_n}{n!}z^n \tag{5.52}$$

where m_n denotes the nth moment and κ_n the nth cumulant. Due to the general relation $M_X(z) = \exp(C_X(z))$ one may apply the previous lemma and obtains:

$$m_n = \sum_{j=1}^{n}\binom{n-1}{j-1} m_{n-j}\kappa_j \tag{5.53}$$

The Nested Calculation Algorithm

The algorithm aims to determine a series expansion. For the algorithmic implementation one has to fix a number M, which is the number of coefficients to determine. So the relations $P(z) = \exp(Q(z))$ or $Q(z) = \ln(P(z))$ between two polynomials P and Q are valid up to terms of order $\mathcal{O}(z^M)$. In applications, a conservative upper bound for M is given by the biggest possible loss measured in basic loss units, hence $M = \sum_{j=1}^{N} l^j$. If one is only interested in the losses up to a certain quantile q, there is also the possibility for an adaptive choice of M[5]. The algorithm consists of four steps, which are shortly presented and analyzed:

1. Step: First define polynomials $P_{1,k}(z)$ by

$$P_{1,k}(z) := 1 - \sigma_k^2\sum_{j=1}^{N} a_k^j p^j T(z^{l^j} - 1) \tag{5.54}$$

$P_{1,k}(z)$ is a polynomial and its computation costs $\mathcal{O}(KN)$ operations. Note that the z^0 coefficient is larger then 1 and that all other coefficients are less or equal than zero, since $l^j \geq 1$.

[5]Therefore first choose $M = 1$, perform the 4 steps and store the intermediate results, then increase M by 1 and perform the 4 steps again until $\sum_{n=0}^{M-1} P[X = n\nu] > q$

2. Step: Compute $P_{2,k}(z) = \ln(P_{1,k}(z))$ using lemma 5.7. The corresponding equations for the coefficients are

$$\kappa_0 = \ln(m_0) \tag{5.55}$$

$$\kappa_n = \frac{1}{m_0}\left(m_n - \sum_{j=1}^{n-1}\binom{j}{n}\kappa_j m_{n-j}\right)) \tag{5.56}$$

The post–condition of the first step states $m_0 \geq 1$ and $m_n < 0$ for $n > 0$ and hence one obtains the inequalities $\kappa_0 > 0$ and $\kappa_n \leq 0$ for $n > 0$. This computation is numerical stable, because in equation (5.56) only non–positive values are added up and so there is no cancelling effect in the numerics. The number of operations one needs to perform this step is given by $\mathcal{O}(KM^2)$.

3. Step: Determine

$$P_3(z) := \sum_{j=1}^{N} a_0^j p^j T(z^{l^j} - 1) - \sum_{k=1}^{K} \frac{1}{\sigma_k^2} P_{2,k}(z) \tag{5.57}$$

Now the term which describes the idiosyncratic risk is added to the negative sum of polynomials computed in the second step. Let κ_n denote the z^n coefficients of $P_3(z)$ now, then one obtains the inequalities $\kappa_0 < 0$ and $\kappa_n \leq 0$. Note that only summations of numbers with the same sign have to be computed and hence no canceling effects occur. The computational effort of this step is given by $\mathcal{O}(N + KM)$.

4. Step: At last, determine the polynomial expression for $G_X(z) = \exp(P_3(z))$ using lemma 5.7. Let m_n denote now the coefficients in the power series for $G_X(z)$, which can be obtained by the relations

$$m_0 = \exp(\kappa_0) \tag{5.58}$$

$$m_n = \sum_{k=1}^{n} \frac{k}{n} m_{n-k}\kappa_k \tag{5.59}$$

Since $\kappa_0 \leq 0$ and $\kappa_n \geq 0$ of all $n > 0$, one easily obtains that $m_0 \geq 0$, and that m_n can be computed by a sum of positive numbers, which is again numerical stable. For the computation of this last step one has to perform $\mathcal{O}(M^2)$ operations.

Recall that the m_n from step 4 are the result of the whole computation, since $m_n = P[X = n\nu]$. The analysis of each step of this algorithm shows, that there is never a summation of two numbers with opposite sign. Hence the algorithm is numerical stable, but the stability of this algorithm is closely related with the structure of the probability generating function, hence this algorithm could fail if its idea is applied on other probability generating functions.
The total computational effort of this algorithm is given by $\mathcal{O}(KM^2 + KN)$. In practice, especially for homogeneous portfolios, M has to be chosen of order $\mathcal{O}(N)$ and hence for large loan portfolios this algorithm is quadratic in the number of obligors.

5.2.2 The Credit Loss Distribution by Fourier Inversion

Another technique to obtain the distribution of the credit loss is based on Fourier inversion, which has been studied in chapter 3. Again one obtains a numerical stable algorithm to determine the whole distribution. The advantage of this approach is that no artificial basic loss unit has to be introduced and that the numerical stability of these methods do not depend on a special structure of the characteristic function and can therefore also be used for generalizations of the CreditRisk$^+$ model. There are two algorithms available, one based on theorem 3.25 and the other is based on the general remarks on the Fast Fourier Transformation stated in section 3.5.

Fourier Inversion using Theorem 3.25

The theorem 3.25 exactly applies for the situation one is faced with in the context of the CreditRisk$^+$ model. It can be used to determine the distribution of a real-valued random variable X if the characteristic function and the first two moments are given. For an efficient use of that theorem it is necessary to find an approximative distribution with the same mean and variance.
The approximative distribution must have three moments and the first two moments have to be adjusted to the expectation and the variance of the credit loss X, which are given by lemma 5.6. Since X is a non-negative random variable, the comparative distribution shall live on $\mathbb{R}^+$. Examples of such distributions are the lognormal distribution or the gamma distribution and their generalizations.
In [16] the limit of a large loan portfolio is studied. It turned out, that the limit distribution is of the credit loss X in CreditRisk$^+$ is the gamma distribution. Since the number of obligors is usually high, an approximation by the gamma distribution is reasonable. Additionally, all expressions which occur in the application of the Gamma distribution in theorem 3.25 can be expressed by closed form solutions. Therefore the approximation of X by the Gamma distribution is recommended.
Since the expectation and the variance of the credit loss X are known by lemma 5.6, the parameters λ and a of the Gamma distribution with the same first two moments can be determined by

$$\lambda = \frac{\mathbf{E}[X]}{\mathbf{Var}[X]} \qquad\qquad a = \frac{\mathbf{E}[X]^2}{\mathbf{Var}[X]} \tag{5.60}$$

Let $F(x)$ denote the distribution function of X and $\hat{F}$ the integral of the distribution function. Then theorem 3.25 yields the equation

$$\hat{F}(x) \;=\; xP(a,\lambda x) - \frac{a}{\lambda}P(a+1,\lambda x) - \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\frac{\Phi_X(s) - \frac{\lambda^a}{(\lambda - is)^a}}{s^2}\,ds \tag{5.61}$$

where Γ denotes the Gamma function and P denotes the incomplete gamma function (see e.g. [1, 68]). Due to the special choice of λ and a, the integrand is zero for $s \to 0$. Hence the integral is well defined and converges, because the integrand decays with s^{-2} as $s \to \pm\infty$. In opposite to the FFT method, this integral may be evaluated by any integration method and hence one is independent of a certain integration algorithm. So there is the advantage that one can use more sophisticated integration

tools which may have a better rate of convergence. Hence the computation time will be reduced if the integration algorithm requires a smaller number of sample points. The function $\hat{F}$ is continuous and so the numerical valuation of $\hat{F}$ will be stable. To obtain the distribution function of X, differentiate $\hat{F}(x)$ with respect to x.

Remark 5.9 *Of course, one can also use other densities for the approximation than the Gamma distribution. But since $X \geq 0$ by definition, one should concentrate on distributions on $\mathbb{R}^+$. So the lognormal distribution or the generalized gamma distribution (without shift and fitted to the first three moments) will also work, but the handling of these distributions is more expensive. The advantage in using such distribution could be, that the valuation of the Fourier integral becomes more accurate, or faster if a smaller number of sample points suffices.*

Fourier Inversion by FFT

Although the credit loss is a discrete random variable, it is obvious that there is an approximative density for this distribution if the number of obligors is large. This density can be determined using the FFT algorithm.
The loss of a loan portfolio lies naturally between 0, the case that no obligor defaults, and $\sum_{j=1}^{N} L^j$, which correspond to the worst case that all obligors default. Hence the density f lives on $[0, \sum_{j=1}^{N} L^j]$. Let S be the number of sample points, then the distances between two adjacent sampling points are given by:

$$\Delta x = \frac{1}{S-1} \sum_{j=1}^{N} L^j \qquad \text{and} \qquad \Delta s = \frac{2\pi}{S \Delta x} \tag{5.62}$$

The choice of S depends on the accuracy one needs for further computations using the density. If g is a smooth function, more precisely if the first derivative of g is bounded, the error of the expectation of $g(X)$ is of order Δx:

$$\left| \mathbf{E}[g(X)] - \sum_{k=0}^{S-1} g(k\Delta x) f(k\Delta x) \right| \quad = \quad \mathcal{O}(\Delta x) \tag{5.63}$$

In practice, S will be taken as a power of 2, hence $S = 2^n$ and typical values for n lie round about 10. If a proper value of S has been chosen, one has to evaluate the characteristic function at $s_j = (j - \frac{S}{2})\Delta s$, since these values are the input for the FFT algorithm.
The formula (5.26) is numerical more stable than the representation (5.25), since one has to evaluate a sum of complex numbers, each with negative real part instead of a product of complex numbers. As it is shown in 3.5, the Fourier inversion of this function using the FFT algorithm yields the density of X.
In order to save computation time, one should recall that $\Phi_X(-s) = \overline{\Phi_X(s)}$. The effort to compute the characteristic function at all sample points is given by $\mathcal{O}(SNK)$ and the Fourier inversion by FFT costs $\mathcal{O}(S \ln S)$ operations. So the overall computation time is given by

$$\text{Computational Effort} \quad = \quad \mathcal{O}(SKN + S \ln S) \tag{5.64}$$

Hence for large loan portfolios this Fourier inversion method is faster then the nested calculation approach.

5.3 Lognormal or other Sector Variables

In the CreditRisk$^+$ model the sector variables R^k are assumed to be independent gamma distributed. The reason for this assumption is based on the simple calculations resulting from this supposition and not on statistical evidence. Hence one may suggest another distribution for the sector variables and one suggestion is to use dependent lognormal sectors. The advantage of this choice is, that the dependence of the sectors can easily be described by a correlation matrix. To estimate these correlations one can use as a first approximation the correlations between the corresponding asset indices. On the other hand, the characteristic function of the credit loss can not be obtained by a closed form solution any more, but it can be numerically computed by a Monte Carlo method.
To show, that the changeover from the Gamma distribution to the lognormal distribution has no dramatic impact on the distribution, the case of independent lognormal sector variables is shortly discussed. In this case, the difference between this model and the CreditRisk$^+$ model is rather small, because the gamma distribution and the lognormal distribution each with mean 1 and variance σ^2 look quite similar. It will be shown in 5.3.2, that the expectation and the variance of the credit loss are also equal in this situation.

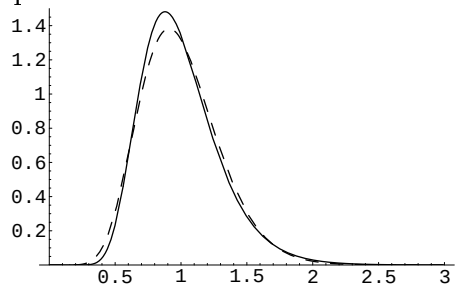

Figure 1: The lognormal density (solid) and the density of the gamma distribution (dashed); Both with mean 1 and $\sigma = 0.3$.

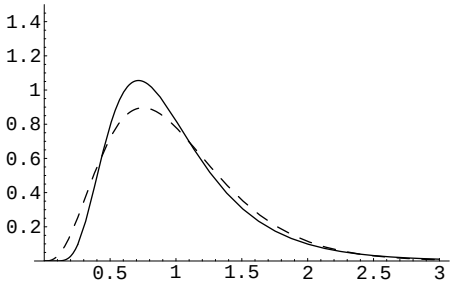

Figure 2: The lognormal density (solid) and the density of the gamma distribution (dashed); Both with mean 1 and $\sigma = 0.5$.

However, the computations in this section are valid for any jointly distributed sector variables R^k – as long as all pairwise covariances exist. The basic idea is to integrate in the space of the characteristic functions by Monte Carlo and to perform a Fourier inversion afterwards. A related numerical treatment is also suggested in [59] and Merino and Nyfeler illustrate there the numerical stability of such an approach exemplarily.

5.3.1 Introduction of the Model

This model is similar to the CreditRisk$^+$ model presented in section 5.1, but now the sector variables are dependent distributed, such that $\mathbf{Cov}[R^k, R^l]$ exists and $\mathbf{E}[R^k] = 1$ holds. Since the other ingredients of the model are unchanged, one can inherit the characteristic function of the credit losses conditioned on R, see (5.20):

$$\Phi_{X|R}(s) = \exp\left(\sum_{j=1}^{N} p^j T \left(a_0^j + \sum_{k=1}^{K} a_k^j R^k\right)\left(e^{iL^j s} - 1\right)\right) \tag{5.65}$$

In order to obtain the characteristic function of the credit loss X one has to take the expectation over R. One approach to solve this problem is to use a Monte Carlo method to compute the characteristic function of X. But before the Monte Carlo approach is studied, the mean and the variance of X under dependent sector variables will be determined. Since there is no closed form solution for the characteristic function $\Phi_X(s)$ one has to change the way to get the moments: Invert the order of differentiation and integration.

5.3.2 The First Moments of the Credit Loss

Theorem 5.10 *The first two moments of X are given by:*

$$\mathbf{E}[X] = T\sum_{j=1}^{N} p^j L^j \tag{5.66}$$

$$\mathbf{E}[X^2] = T\sum_{j=1}^{N} p^j (L^j)^2 + T^2\left[\sum_{j=1}^{N} p^j L^j\right]^2 + T^2\sum_{j,n=1}^{N} p^j L^j p^n L^n \sum_{k,l=1}^{K} a_k^j a_l^n \mathbf{Cov}[R^k, R^l] \tag{5.67}$$

$$\mathbf{Var}[X] = T\sum_{j=1}^{N} p^j (L^j)^2 + T^2\sum_{j,n=1}^{N} p^j L^j p^n L^n \sum_{k,l=1}^{K} a_k^j a_l^n \mathbf{Cov}[R^k, R^l] \tag{5.68}$$

Proof:

In general holds:

$$\Phi_X(s) = \mathbf{E}[\Phi_{X|R}(s)] \tag{5.69}$$

To determine the nth moment one needs to compute:

$$\mathbf{E}[X^n] = \lim_{s\to 0} \frac{1}{i^n}\frac{d^n}{ds^n}\mathbf{E}[\Phi_{X|R}(s)] \tag{5.70}$$

In fact, there are three limits in this expression: the expectation (integration), a differentiation and the valuation at one point. From (5.65) it is clear, that $\Phi_{X|R}(s)$ is analytical and that it may written as a power series in s. Since power series converge on a compact interval uniformly, the exchange of the limits is allowed:

$$\mathbf{E}[X^n] = \mathbf{E}[\lim_{s\to 0} \frac{1}{i^n}\frac{d^n}{ds^n}\Phi_{X|R}(s)] \tag{5.71}$$

The first two derivatives of $\Phi_{X|R}(s)$ are

$$\Phi'_{X|R}(s) = \Phi_{X|R}(s)\cdot\sum_{j=1}^{N} p^j T\left(a_0^j + \sum_{k=1}^{K} a_k^j R^k\right) e^{iL^j s}(iL^j) \tag{5.72}$$

$$\Phi''_{X|R}(s) = \Phi_{X|R}(s)\cdot\sum_{j=1}^{N} p^j T\left(a_0^j + \sum_{k=1}^{K} a_k^j R^k\right) e^{iL^j s}(iL^j)^2 \tag{5.73}$$

$$+\Phi_{X|R}(s)\cdot\left[\sum_{j=1}^{N} p^j T\left(a_0^j + \sum_{k=1}^{K} a_k^j R^k\right) e^{iL^j s}(iL^j)\right]^2 \tag{5.74}$$

Valuation at $s = 0$ yields:

$$\frac{1}{i}\Phi'_{X|R}(0) = T\sum_{j=1}^{N} p^j L^j (a_0^j + \sum_{k=1}^{K} a_k^j R^k) \tag{5.75}$$

$$\frac{1}{i^2}\Phi''_{X|R}(0) = T\sum_{j=1}^{N} p^j (L^j)^2 (a_0^j + \sum_{k=1}^{K} a_k^j R^k) + \left[T\sum_{j=1}^{N} p^j L^j (a_0^j + \sum_{k=1}^{K} a_k^j R^k)\right]^2 \tag{5.76}$$

Take the expectation over R and recall that $\mathbf{E}[R^k] = 1$. To simplify the expressions, the relation $a_0^j + \sum_{k=1}^{K} a_k^j = 1$ has been used:

$$\mathbf{E}[X] = T\sum_{j=1}^{N} p^j L^j \tag{5.77}$$

$$\mathbf{E}[X^2] = T\sum_{j=1}^{N} p^j (L^j)^2 + T^2 \left[\sum_{j=1}^{N} p^j L^j\right]^2 + T^2 \sum_{j,n=1}^{N} p^j L^j p^n L^n \sum_{k,l=1}^{K} a_k^j a_l^n \mathbf{Cov}[R^k, R^l] \tag{5.78}$$

□

Remark 5.11 *Note that in the case of uncorrelated sector variables the first two moments are identical to the moments in the CreditRisk$^+$ model, since the variance of the random variables R^k is given by $\sigma_k^2 = \mathbf{Cov}[R^k, R^k]$.*

5.3.3 The Characteristic Function by a Monte Carlo Approach

From equation (5.65) the characteristic function of X conditioned on R is known and hence the characteristic function of X can be written as an expectation:

$$\Phi_X(s) = \mathbf{E}[\Phi_{X|R}(s)] \tag{5.79}$$

$$= e^{\left(\sum_{j=1}^{N} p^j T a_0^j (e^{iL^j s} - 1)\right)} \mathbf{E}\left[\exp\left(\sum_{k=1}^{K} R^k \sum_{j=1}^{N} T p^j a_k^j (e^{iL^j s} - 1)\right)\right] \tag{5.80}$$

$$= \exp(\xi_0(s))\ \mathbf{E}\left[\exp\left(\sum_{k=1}^{K} R^k \xi_k(s)\right)\right] \qquad \text{where} \tag{5.81}$$

$$\xi_k(s) := \sum_{j=1}^{N} T p^j a_k^j (e^{iL^j s} - 1) \qquad \text{for } k = 0, \ldots K \tag{5.82}$$

First one has to decide, for which s one wants to compute the characteristic function $\Phi_X(s)$ and S denotes the number of sample points. Since the functions $\xi_k(s)$ are

deterministic functions, this functions can be valuated before starting the Monte Carlo loop for each regarded s. The advantage of this organisation of the Monte Carlo procedure is, that each diced tuple of R can be used for each s to compute the expectation, hence the computation time for generation of the random numbers is independent of S. In the case that R^k are dependent lognormal distributed, the effort to dice one sample is given by $\mathcal{O}(K^2)$, however, the same effort will mostly hold for other multivariate distributions. If M is the number of Monte Carlo iterations, the computational effort of this method is given by:

$$\underbrace{\mathcal{O}(KNS)}_{\text{Computing } \xi_k} + \underbrace{\mathcal{O}(MK^2)}_{\text{Dicing } R^k} + \underbrace{\mathcal{O}(MSK)}_{\text{Valuation of the Monte Carlo sum}} \tag{5.83}$$

Note, that in practice N and M are larger than K, S and that the computational effort does not contain a term NM. So the effort of this Monte Carlo method is tolerable, in contrast to a direct simulation. A direct Monte Carlo approach on the model for X would effort for each Monte Carlo simulation first to sample the sector variables R, then compute the default intensities and sample the time of default of each obligor. Hence the direct simulation approach is not feasible due to a computational effort of $\mathcal{O}(KMN)$.
The proposed algorithm admits two possibilities for parallel computing. The first possibility is the usual Monte Carlo parallelism of valuating the expression parallel for several diced random numbers. Second, one can also compute the expression for each s independently for a fixed dice of random numbers R. Since one is interested in an expectation and not in the simulation of a stochastic process, it may be favourable to use Quasi-Monte-Carlo, due to a potentially much better convergence.
Since the first two moments of X are known and a numerical algorithm to obtain the characteristic function of X is established, one can compute the distribution of X by one of the Fourier inversion techniques presented in section 5.2.2.

5.4 A Time–continuous Model

In the previous models the distribution of the loss for a fixed date T was studied. In this part, a model is introduced, which allows the credit risk analysis in continuous time. This model is abutted to the model with dependent sector variables; but to describe the sectors continuous time processes are involved rather than discrete random variables.

5.4.1 Introduction of the Model

Let us fix a time horizon $T_\infty > 1$ and introduce K sector processes R_t^k on $t \in [0, T_\infty]$ which are assumed to be geometric Brownian motions with:

$$R_0^k = 1 \tag{5.84}$$
$$\mathbf{E}[R_t^k] = 1 \tag{5.85}$$
$$\mathbf{Cov}[\ln R_t^k, \ln R_t^l] = C_{kl} t \tag{5.86}$$

As before, consider sector affiliations a_k^j, such that the following relations hold:

$$a_k^j \geq 0 \qquad \forall\ j = 1, \ldots, N;\ k = 0, \ldots, K \tag{5.87}$$

$$\sum_{k=0}^{K} a_k^j = 1 \qquad \forall\ j = 1, \ldots, N \tag{5.88}$$

Again there is a rating parameter p^j for each obligor, which is the probability that the jth obligor defaults within one year and it is assumed that p^j is rather small. For the interpretation of p^j as the annual default probability, the proof of lemma 5.2 holds. In contrast to the previous models, the default intensity becomes a stochastic process now:

$$\lambda_t^j := p^j \left(a_0^j + \sum_{k=1}^{K} a_k^j R_t^k \right) =: p^j \tilde{\lambda}_t^j \tag{5.89}$$

The default event of the jth obligor is denoted by T^j and T^j is exponentially distributed with the intensity process λ_t^j. Also the processes I_t^j are introduced:

$$I_t^j := \begin{cases} 0 & \text{if } t < T^j \\ 1 & \text{if } t \geq T^j \end{cases} \tag{5.90}$$

Conditioned on R, these processes are independent from the sector processes and

$$P[T^j < t] = P[I_t = 1] = 1 - \exp\left(- \int_0^t \lambda_s ds \right) \tag{5.91}$$

If the jth obligor defaults, the loss due to this default is given by L^j and hence the process of cumulated defaults is given by

$$X_t := \sum_{j=1}^{N} L^j I_t^j \tag{5.92}$$

Remark 5.12 (Default indicator and Poisson process) *Let Y_t be a Poisson process with time dependent default intensity $\lambda_t \geq 0$, that is a process with right continuous paths and existing left limits such that for all $t > s \geq 0$ holds:*

$$Y_0 = 0 \tag{5.93}$$

$$Y_t - Y_s \in \mathbb{N} \tag{5.94}$$

$$P[Y_t - Y_s = n] = \frac{1}{n!} \left(\int_s^t \lambda_u du \right)^n \cdot \exp\left(- \int_s^t \lambda_u du \right) \tag{5.95}$$

$$Y_t - Y_s \quad \textit{is independent of } Y_{[0,s]} \tag{5.96}$$

Then the following relation holds for the default indicator process

$$I_t = \begin{cases} 0 & \textit{if } Y_t = 0 \\ 1 & \textit{if } Y_t > 0 \end{cases} \tag{5.97}$$

5.4.2 The Characteristic Function by Monte Carlo

Let $\mathcal{R}_t$ be the natural filtration induced by the processes R^k. If one defines the sector variables in section 5.3 to be $\frac{1}{t}\int_0^t R^k_\tau \, d\tau$, one obtains the characteristic function of X_t conditioned on $\mathcal{R}_t$:

$$\Phi_{X_t|\mathcal{R}}(s) \;=\; \exp\left(\sum_{j=1}^{N}(e^{iL^j s}-1)p^j(a_0^j t + \sum_{k=1}^{K} a_k^j \int_0^t R^k_\tau d\tau)\right) \tag{5.98}$$

Since the sector processes R^k_t are not independent, the expectation to get the characteristic function $\Phi_{X_t}(s)$ can be valuated by Monte Carlo only:

$$\begin{aligned}
\Phi_{X_t}(s) &= \mathbf{E}[\Phi_{X_t|\mathcal{R}}(s)] && (5.99)\\
&= \exp\left(\sum_{j=1}^{N}(e^{iL^j s}-1)p^j a_0^j t\right)\mathbf{E}\left[\exp\left(\sum_{k=1}^{K}\sum_{j=1}^{N}(e^{iL^j s}-1)p^j a_k^j \int_0^t R^k_\tau d\tau\right)\right] && (5.100)\\
&= \exp(\xi_0(s)t)\;\mathbf{E}\left[\exp\left(\sum_{k=1}^{K}\xi_k(s)\int_0^t R^k_\tau d\tau\right)\right] \qquad \text{where} && (5.101)\\
\xi_k(s) &:= \sum_{j=1}^{N}(e^{iL^j s}-1)p^j a_k^j \qquad \text{for } k=0,\ldots K
\end{aligned}$$

Let $\mathcal{S}$ be the set of values of s for which one wants to valuate the characteristic function. Then it is possible to compute the expressions $\xi_k(s)$ $s \in \mathcal{S}$ before starting the Monte Carlo procedure. For each Monte Carlo iteration one has to sample the random numbers

$$I^k \;\sim\; \int_0^t R^k_\tau d\tau \tag{5.102}$$

with dependent geometric Brownian motions R^k_τ. Using one draw of I^k one computes $\sum_{k=1}^{K}\exp(\xi_k(s)I^k)$ for each $s \in \mathcal{S}$ and the total cost of this Monte Carlo method is comparable with the method in section 5.3.3. There is only a larger effort in dicing I^k, since one has to sample paths and to integrate them instead of dicing one lognormal random number, but the asymptotic costs are again given by (5.83).

Remark 5.13 *To sample I^k one can proceed as follows. Let B be the Cholesky decomposition of the log–covariance matrix C and W^l_t independent Brownian motions. Then R^k_t is given by*

$$R^k_t \;=\; \exp\left(\sum_{l=1}^{K} B_{kl}W^l_t - \frac{1}{2}C_{kk}t\right) \tag{5.103}$$

To compute the integral one divides the interval $[0,t]$ into D intervals of equal length and computes the integral by the approximation

$$I^k \;\approx\; \frac{t}{2D}\sum_{j=1}^{D}(R^k_{\frac{(j-1)t}{D}} + R^k_{\frac{jt}{D}}) \tag{5.104}$$

There is a lot of literature on the properties of the integral of the geometric Brownian motion, see for example [28, 29, 36, 85].

In order to obtain the distribution of X one can use the Fourier inversion techniques presented in section 5.2.2. Usually one would use the FFT based method in this context, since then the set $\mathcal{S}$ is clearly given by the discretization of the characteristic function. To use the Fourier inversion method based on theorem 3.25 one needs to know the first two moments of X, which will be computed now to round off the discussion.

5.4.3 Expectation and Variance of the Credit Loss Process

Lemma 5.14 *Let R_t^k and R_t^l be two geometric Brownian motions such that $(\ln R_t^k + \frac{1}{2}C_{kk}t)$ and $(\ln R_t^l + \frac{1}{2}C_{ll}t)$ are normal distributed with mean 0 and covariance $C_{kl}t$. Then holds:*

$$\mathbf{E}[\left[\int_0^t R_\tau^k d\tau\right] = t \tag{5.105}$$

$$\mathbf{E}\left[\int_0^t R_\tau^k d\tau \int_0^t R_\theta^l d\theta\right] = \frac{2}{(C_{kl})^2}\left(e^{C_{kl}t} - 1 - C_{kl}t\right) \tag{5.106}$$

Proof:
To prove the first equation, one only has to use Fubini's theorem. For the second statement, let $\tau \leq \theta$:

$$\mathbf{E}[R_\tau^k R_\theta^l] = \mathbf{E}[R_\tau^k R_\tau^l \frac{R_\theta^l}{R_\tau^l}] = \mathbf{E}[R_\tau^k R_\tau^l] = \exp(C_{kl}\tau) \tag{5.107}$$

Hence,

$$\mathbf{E}\left[\int_0^t R_\tau^k d\tau \int_0^t R_\theta^l d\theta\right] = \int_0^t d\tau \int_0^t d\theta \mathbf{E}\left[R_\tau^k R_\theta^l\right] \tag{5.108}$$

$$= 2\int_0^t d\tau \int_\tau^t d\theta e^{C_{kl}\tau} = 2\int_0^t d\tau e^{C_{kl}\tau}(t-\tau) \tag{5.109}$$

$$= \frac{2}{(C_{kl})^2}\left(e^{C_{kl}t} - 1 - C_{kl}t\right) \tag{5.110}$$

□

This lemma and theorem 5.10 prove the

Theorem 5.15 *The expectation and variance of X_t are given by*

$$\mathbf{E}[X_t] = t\sum_{j=1}^N p^j L^j \tag{5.111}$$

$$\mathbf{Var}[X_t] = t\sum_{j=1}^N p^j (L^j)^2 + \sum_{j,n=1}^N p^j L^j p^n L^n \sum_{k,l=1}^K a_k^j a_l^n \left(\frac{2(e^{C_{kl}t} - 1 - C_{kl}t)}{(C_{kl})^2} - t^2\right) \tag{5.112}$$

5.5 Combining Market Risk and Credit Risk

Up to now only credit risk has been analyzed in this paper. In this section the previous model will be extended in such a way, that it incorporates market risk. The idea of the approach presented here differs from other models, which combine market and credit risk (see e.g. [27]). It starts from the CreditRisk$^+$ model which is widen in such a way, that when there is no credit risk one obtains the well known Delta normal approach to assess market risk.
Whenever one talks about market risk, there is a portfolio depending on market risk factors and one has the possibility to evaluate a portfolio for a given state of these factors. Here a model of a portfolio which combines market risk and credit risk is introduced. Finally, the computation of the profit and loss distribution (P&L) of this portfolio is aspired and an approximative solution of this task can be given.
In addition to the assumption that default probabilities are small, it is also assumed – according to the Delta normal model – that the effect of the (normal distributed) market fluctuations may be linearized. An algorithm to compute the P&L is presented which is based on these assumptions. The computational effort of this algorithm is tolerable for a medium number of obligors and the Monte Carlo part is independent of the number of market risk factors.

5.5.1 A Portfolio with Market Risk and Credit Risk

The credit risk driving factors are the default indicators I_t^j and the sector processes R_t^k of the credit risk model presented in section 5.4. Due to changes of the sector processes R_t^k, the obligors default intensities alter and this effects the probability of default of each obligor. In fact, this is the so–called spread risk, which is modelled by the sector processes in this way.
Additional to the sector variables, there are market risk factors M_t^l, $l = 1, \ldots, \tilde{K}$. Examples of market risk factors are returns of stock prices, foreign exchange rates or interest rates. Let M_t^l are dependent Brownian motions with drift 0 and different volatilities. To describe the dependencies between R_t^k and M_t^l a covariance matrix C of the following structure is introduced:

$$C = \begin{pmatrix} C^1 & C^{2\top} \\ C^2 & C^3 \end{pmatrix} \tag{5.113}$$

C^1 is a $K \times K$ matrix, C^3 is a $\tilde{K} \times \tilde{K}$ matrix and C^2 is a $\tilde{K} \times K$ matrix and the entries of these matrices are given by

$$C_{kl}^1 = \frac{1}{t}\mathbf{Cov}(\ln R_t^k, \ln R_t^l) \tag{5.114}$$

$$C_{kl}^2 = \frac{1}{t}\mathbf{Cov}(M_t^k, \ln R_t^l) \tag{5.115}$$

$$C_{kl}^3 = \frac{1}{t}\mathbf{Cov}(M_t^k, M_t^l) \tag{5.116}$$

Let $N^j(M_t, t)$ denote the nominal amount of the jth obligor at time t, discounted by the risk free interest rate. For simplicity, let us assume that there is only one and fixed settlement date T^j for each obligor. The probability, that the jth obligor

survives up to T^j under the condition that no default has occured up to time t is given by $P[I^j_{T^j} = 0 | I^j_t = 0]$. These are functions of the state of the sector processes:

$$q^j(r,t) := P[I^j_{T^j} = 0 | R_t = r \wedge I^j_t = 0] \tag{5.117}$$

$$= \mathbf{E}\left[\exp\left(-p^j \int_t^{T^j} (a^j_0 + \sum_{k=1}^{K} a^j_k R^k_\tau) d\tau \right) \middle| R^k_t = r^k \right] \tag{5.118}$$

Since any claim fraught with credit risk is discounted by this probability, the value process of the portfolio can be written:

$$V_t = \sum_{j=1}^{N} N^j(M_t, t) q^j(R_t, t)(1 - I^j_t) \tag{5.119}$$

5.5.2 Obtaining an Approximative Profit and Loss Distribution

Besides the assumption that p^j is small, also the following linearization is used:

Assumption 5.16 (Delta Approach) *Let us assume, that the nominal functions may be approximated by linear functions:*

$$N^j(M_t, t) \approx N^j(M_0, 0) + \frac{\partial N^j}{\partial t}(M_0, 0)t + \frac{\partial N^j}{\partial M_t}(M_0, 0) \cdot M_t \tag{5.120}$$

$$=: \overset{0}{N^j}(t) + \overset{\Delta}{N^j} \cdot M_t \tag{5.121}$$

Using this assumption, the value process is given by:

$$V_t = \sum_{j=1}^{N} (\overset{0}{N^j}(t) + \overset{\Delta}{N^j} \cdot M_t)\, q^j(R_t, t)\, (1 - I^j_t) \tag{5.122}$$

In the next step the dependence between M_t and R_t will be analyzed. Let B be the Cholesky decomposition of the covariance matrix C, with the structure:

$$C = \begin{pmatrix} C^1 & C^{2\top} \\ C^2 & C^3 \end{pmatrix} = \begin{pmatrix} B^1 & 0 \\ B^2 & B^3 \end{pmatrix} \begin{pmatrix} B^1 & 0 \\ B^2 & B^3 \end{pmatrix}^\top = BB^\top \tag{5.123}$$

Then the usual method to sample $\ln R_t$ and M_t is to choose two vectors W^1_t, W^2_t of independent Brownian motions and to use the transformation

$$\begin{pmatrix} \ln R^k_t + \frac{1}{2} C_{kk} t \\ M_t \end{pmatrix} = \begin{pmatrix} B^1 & 0 \\ B^2 & B^3 \end{pmatrix} \begin{pmatrix} W^1_t \\ W^2_t \end{pmatrix} \tag{5.124}$$

Hence the state of the sector process is given by

$$R^k_t = \exp\left((B^1 W^1_t)^k - \frac{1}{2} C^1_{kk} t \right) \tag{5.125}$$

and M_t can be expressed by

$$M_t = B^2 W_t^1 + B^3 W_t^2 \tag{5.126}$$

Hence, the portfolio has the representation

$$V_t = \sum_{j=1}^{N} q^j(R_t,t)(1-I_t^j)\left(\overset{0}{N}{}^j(t) + B^{2\top}\overset{\Delta}{N}{}^j \cdot W_t^1 + B^{3\top}\overset{\Delta}{N}{}^j \cdot W_t^2\right) \tag{5.127}$$

The P&L may be determined by Fourier inversion of the characteristic function of V_t. A first step to compute the characteristic function is to determine the characteristic function conditioned on the state of R_t and I_t:

$$\begin{aligned}\Phi_{V_t|(R_t,I_t)}(s) &= \mathbf{E}[e^{isV_t}|R_t,I_t] &(5.128)\\ &= \exp\left(is\sum_{j=1}^{N} q^j(R_t,t)(1-I_t^j)(\overset{0}{N}{}^j(t) + \overset{\Delta}{N}{}^j \cdot B^2 W_t^1)\right)\times \\ &\quad \exp\left(-\frac{1}{2}s^2\left\|\sum_{j=1}^{N} q^j(R_t,t)(1-I_t^j)B^{3\top}\overset{\Delta}{N}{}^j\right\|^2 t\right) &(5.129)\end{aligned}$$

As in the whole paper, the default probability for each obligor is assumed to be rather small. Hence by lemma 5.5 one may approximate:

$$q^j(r,t) \approx \exp\left(-p^j(T^j-t)(a_0^j + \sum_{k=1}^{K} a_k^j r^k)\right) \tag{5.130}$$

The characteristic function of V_t can now be evaluated by a Monte Carlo method using the representation

$$\Phi_{V_t}(s) = \mathbf{E}\left[\Phi_{V_t|(\mathcal{R}_t,\mathcal{I}_t)}(s)\right] \tag{5.131}$$

For one Monte Carlo sample one has to sample the paths W_t^1 of independent Brownian motions. By equation (5.125) one can determine R_t^k and then by equation (5.130) one obtains $q^j(R_t,t)$. Then for each obligor, one has to sample the binary random variable

$$I_t^j = \begin{cases} 0 & \text{with probability } 1-\exp\left(-p^j\int_0^t (a_0^j + \sum_{k=1}^{K} a_k^j R_\tau^k)d\tau\right) \\ 1 & \text{with probability } \exp\left(-p^j\int_0^t (a_0^j + \sum_{k=1}^{K} a_k^j R_\tau^k)d\tau\right)\end{cases} \tag{5.132}$$

Of course, with one sample of I_t and R_t one can evaluate the expression under the expectation for several s. If one proceeds so, the numerical effort of this method is obviously of order $\mathcal{O}(M\cdot(K+N+N\cdot K+S))$ where M denotes the number of Monte Carlo valuations and S is the number of sample points of $\Phi_{V_t}(s)$. Hence this method can be used in practice if the number of obligors is not too big; but since the effort is independent of the number of market risk factors $\tilde{K}$, this approach can be applied even for large $\tilde{K}$. In order to obtain the density of the portfolios profit and loss distribution one can use the Fourier inversion technique based on the Fast Fourier Transformation described in sections 3.5 and 5.2.2.

5.6 Conclusion and Prospects on further Research

The CreditRisk$^+$ model has been presented in terms of characteristic functions instead of the usual approach by probability generating functions. One advantage of this description is, that no basic loss unit has to be introduced, since the proper choice of the basic loss unit may be critical, if the loss given defaults have different orders of magnitude.
There are two approaches to determine the credit loss distribution. One method is the new recursion scheme *nested calculation* proposed by Giese [38]. Now a proof of the numerical stability of this algorithm has been established. But one has to introduce a basic loss unit, since this method is based on the probability generating function. Another approach is the Fourier inversion of the characteristic function based on the general results concerning the Fourier inversion of characteristic functions presented in chapter 3. These general techniques can successfully be applied in the context of CreditRisk$^+$. The Fourier inversion method to obtain the distribution of the credit loss is also numerical stable and it is faster than the nested calculation scheme for portfolios with a large number of obligors.
All computations in the context of the CreditRisk$^+$ model or its generalizations base on the assumption, that the default probability of each obligors is small. In practice, there are less credit–worthy obligors and hence a generalization of CreditRisk$^+$ for non–investment grade obligors is a challenging task for future research.

The presented generalization of the CreditRisk$^+$ model combines market and credit risk. In the special case, that there is only one obligor who never defaults, the previous analysis of the P&L distribution gives the so–called "Delta normal" approach, which is the most simple and well known idea to deal with market risk alone. In the case, that there is no market risk, the model conforms with the credit risk model presented in section 5.4. This underlines that the presented model is a natural generalization to describe market risk and credit risk.
In order to find other and more general models which combine market risk and credit risk it is necessary to find other credit risk models which allow the handling of a stochastic loss given default. The first approach is to assume, that the sector variables and the loss given defaults are independent, as it is discussed in the *remark on independent random loss given defaults* in section 5.1.2. This approach has been analyzed by Bürgisser, Kurth and Wagner [16]. They also give a remark how to overcome this assumption, but a deeper analysis remains to be done. In [27] another approach is presented to model market and credit risk. Duffie and Pan split the portfolio into a *value component* and a *default component* and they treat each component separately.
But for applications in risk management it seems to be necessary to model dependencies between the sectors and the loss given default. For example, assume that you hold put options on the stock of a bank which are sold by this bank. In the case of default of this bank the stock will fall, but your (theoretical) win on the puts is worthless since the bank is unable to pay them out. The model presented in section 5.5 provides a first approach to deal with such situations, which may be a basis for further research on more realistic but still efficient methods.

Part III

Appendix

Appendix A

Properties of common continuous Distributions

The basic properties of some common distributions are recalled for the convenience of the reader. This appendix covers the distributions which are relevant for the present work. There are lots of distributions which are surely also common, but which are not used in the previous chapters and which are therefore not collected here. Since the presented results are well–known, the proofs are omitted and can be found in several textbooks.
The properties analyzed for a random variable X, which is usually defined by its density $\rho(x)$, are its expectation $\mathbf{E}[X]$, its variance $\mathbf{Var}[X]$ and its skewness $\mathbf{Skew}[X]$. Also the higher moments $\mathbf{E}[X^n]$ and the absolute moments $\mathbf{E}[\,|X|^n\,]$ are studied. The distribution function $P[X \leq x_0]$ and the characteristic function $\Phi_X(s)$ are also given. If some expressions have no closed solution or became to ugly, their presentation has been omitted.
An algorithm to sample a random variable according to each distribution is given. These algorithms assume, that one can sample the uniform distribution on $[0,1]$ or on another closed interval. Some of these algorithms use the sampling of another distribution than the uniform, but then one can find an algorithm to sample this other distribution also in this appendix.

A.1 The normal distribution $N(\mu, \sigma^2)$

Definition A.1 *The standard normal density and its distribution function are defined by:*

$$n(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{A.1}$$

$$\mathcal{N}(x) := \int_{-\infty}^{x} n(t)dt \tag{A.2}$$

Definition A.2 *The density of the normal distribution is given by*

$$\rho_{N(\mu,\sigma^2)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} n(\frac{x-\mu}{\sigma}) \qquad \sigma > 0 \tag{A.3}$$

A random number X with this density is called $N(\mu, \sigma^2)$ distributed and the parameters μ and σ have a clear interpretation:

Lemma A.3 *Let X be a $N(\mu, \sigma^2)$ distributed random number. Then holds:*

$$\begin{aligned}
\mathbf{E}[X] &= \mu && \text{(A.4)}\\
\mathbf{Var}[X] &= \sigma^2 && \text{(A.5)}\\
\mathbf{Skew}[X] &= \mathbf{E}[\left(\frac{X-\mu}{\sigma}\right)^3] = 0 && \text{(A.6)}\\
\mathbf{E}[X^2] &= \sigma^2 + \mu^2 && \text{(A.7)}\\
\mathbf{E}[(X-\mu)^n] &= \begin{cases} (n-1)!!\ \sigma^n & \text{if } n \text{ even}^1 \\ 0 & \text{if } n \text{ odd} \end{cases} && \text{(A.8)}\\
\mathbf{E}[\,|X-\mu|^n\,] &= \frac{\sqrt{2}^n}{\sqrt{\pi}}\Gamma(\frac{n+1}{2})\sigma^n && \text{(A.9)}\\
P[X \le x_0] &= \mathcal{N}(\frac{x_0-\mu}{\sigma}) && \text{(A.10)}\\
\hat{G}(x) &:= \int_{-\infty}^{x} \mathcal{N}(\frac{y-\mu}{\sigma})dy = (x-\mu)\mathcal{N}(\frac{x-\mu}{\sigma}) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} && \text{(A.11)}\\
\Phi_X(s) &= \exp\left(i\mu s - \frac{\sigma^2 s^2}{2}\right) && \text{(A.12)}\\
\Phi_{X^2}(s) &= \frac{1}{\sqrt{1-2\sigma^2 is}} \exp\left(i\mu^2 s + \frac{\sigma^2\mu^2 s^2}{2(1-2\sigma^2 is)}\right) && \text{(A.13)}
\end{aligned}$$

In the last equation the square root has to be taken such, that $\Re\sqrt{1-2\sigma^2 is} > 0$. Hence $\Phi_{X^2}(s)$ is continuous with $\Phi_{X^2}(0) = 1$.

Lemma A.4 *Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ be independent. Then holds*

$$(X_1 + X_2) \;\sim\; N(\mu_1 + \mu_2\ ,\ \sigma_1^2 + \sigma_2^2) \qquad \text{(A.14)}$$

Lemma A.5 *Let U_1 and U_2 be independent uniform distributed on [-1,1] and such that $U_1^2 + U_2^2 \le 1$. Then*

$$X_1 := U_1\sqrt{-2\frac{\ln(U_1^2+U_2^2)}{U_1^2+U_2^2}} \quad \text{and} \quad X_2 := U_2\sqrt{-2\frac{\ln(U_1^2+U_2^2)}{U_1^2+U_2^2}} \qquad \text{(A.15)}$$

are independent $N(0,1)$ distributed. If X is $N(0,1)$ distributed, then

$$Y \;:=\; \mu + \sigma X \qquad \text{(A.16)}$$

is $N(\mu, \sigma^2)$ distributed.

[1] !! denotes the double factorial with $(-1)!! = 1!! = 1$ and $(2n+1)!! = (2n+1)\cdot(2n-1)!!$.

A.2 The lognormal distribution $Ln(\mu, \sigma^2)$

Definition A.6 *A non–negative random variable X is lognormal distributed, in signs $X \sim Ln(\mu, \sigma^2)$, if its density is given by:*

$$\rho_{Ln(\mu,\sigma)}(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \qquad x, \sigma > 0 \quad \text{(A.17)}$$

The name of the lognormal distribution is based on the fact, that the logarithm of a lognormal distribution is normal distributed:

Lemma A.7 *Let X be $Ln(\mu, \sigma^2)$ distributed. Then the following statements hold:*

$$\ln(X) \sim N(\mu, \sigma^2) \quad \text{(A.18)}$$

$$\mathbf{E}[X] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad \text{(A.19)}$$

$$\mathbf{Var}[X] = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \quad \text{(A.20)}$$

$$\mathbf{E}[X^n] = \exp\left(n\mu + \frac{n^2\sigma^2}{2}\right) \quad \text{(A.21)}$$

$$P[X \leq x_0] = \mathcal{N}\left(\frac{\ln(x_0) - \mu}{\sigma}\right) \quad \text{(A.22)}$$

There is no closed form solution for the characteristic function $\mathbf{E}[e^{isX}]$, which is infinitely often differentiable but not analytical at $s = 0$.

Lemma A.8 *Let N be a normal distributed random number with mean μ and variance σ^2. Then the variable*

$$X := \exp(N) \quad \text{(A.23)}$$

is $Ln(\mu, \sigma)$ distributed.

A.3 The Chi–square distribution

A.3.1 The central χ^2 distribution $\chi^2(f)$

Definition A.9 *A non–negative random variable X is called chi–square distributed with $f > 0$ degrees of freedom, $X \sim \chi^2(f)$, if its density is given by*

$$\rho_{\chi^2(f)}(x) := \frac{x^{\frac{f}{2}-1}}{2^{\frac{f}{2}}\Gamma(\frac{f}{2})} e^{-\frac{x}{2}} \qquad x \geq 0 \quad \text{(A.24)}$$

Some properties of this distribution are listed below.

Lemma A.10 *Let X be $\chi^2(f)$ distributed. Then holds:*

$$\begin{aligned}
\mathbf{E}[X] &= f && \text{(A.25)}\\
\mathbf{Var}[X] &= 2f && \text{(A.26)}\\
\mathbf{Skew}[X] &= \frac{2\sqrt{2}}{\sqrt{f}} && \text{(A.27)}\\
\mathbf{E}[X^n] &= \frac{2^n\Gamma(\frac{f}{2}+n)}{\Gamma(\frac{f}{2})} = \prod_{j=0}^{n-1}(f+2j) && \text{(A.28)}\\
P[X \le x_0] &= P(\frac{f}{2}, \frac{x_0}{2}) && \text{(A.29)}\\
\mathbf{E}[e^{-\alpha X}] &= \frac{1}{\sqrt{1+2\alpha}^f} \qquad \Re\alpha > -\frac{1}{2} && \text{(A.30)}\\
\Phi_{\chi^2(f)}(s) &= \frac{1}{\sqrt{1-2is}^f} && \text{(A.31)}
\end{aligned}$$

$P(a,x)$ denotes the incomplete gamma function. This notation agrees with the notation in [1, 68] where numerical algorithms and some properties of P can be found. The square root in the expression for the characteristic function has to be taken such, that $\Re\sqrt{1-2is} > 0$ holds.

Lemma A.11 *Let N_i, $i = 1,\ldots,f$ be independent standard normal distributed, hence $N_i \sim N(0,1)$ and define*

$$X := \sum_{i=1}^{f} N_i^2 \qquad \text{(A.32)}$$

Then X is $\chi^2(f)$ distributed.

Using this lemma, one can easily simulate $\chi^2(f)$ distributed random numbers. Note that $\chi^2(2)$ is an exponential distribution with intensity $\frac{1}{2}$ (see A.5). Based on this idea, the next lemma provides a more efficient way to generate χ^2 distributed random numbers with an even degree of freedom:

Lemma A.12 *Let $f = 2k$ with $k \in \mathbb{N}$. Let U_i with $i = 1,\ldots k$ be independent random numbers, uniform distributed on $]0,1]$. Then*

$$-2\sum_{i=1}^{k} \ln(U_i) \qquad \text{(A.33)}$$

is a $\chi^2(f)$ distributed random number.

A.3.2 The non–central χ^2 distribution $\chi^2(f,\lambda)$

Definition A.13 *Let N_j with $j = 1,\ldots f$ be independent standard normal random variables and define X by*

$$X := \sum_{j=1}^{f} (N_j + \mu_j)^2 \qquad \text{(A.34)}$$

The distribution of X is the non–central χ^2 distribution, $X \sim \chi^2(f,\lambda)$ where f denotes the degrees of freedom and the non–centrality parameter $\lambda \geq 0$ is defined by

$$\lambda := \sum_{j=1}^{f} \mu_j^2 \tag{A.35}$$

Obviously, the non–central χ^2 distribution is a generalization of the central χ^2 distribution and the relation $\chi^2(f) = \chi^2(f,0)$ holds. The definition of $\chi^2(f,\lambda)$ provides a method to sample random numbers with this distribution; the main properties are listed in the next lemma.

Lemma A.14 *Let $X \sim \chi^2(f,\lambda)$. Then holds:*

$$\mathbf{E}[X] = f + \lambda \tag{A.36}$$

$$\mathbf{Var}[X] = 2f + 4\lambda \tag{A.37}$$

$$\mathbf{Skew}[X] = \frac{f+3\lambda}{\sqrt{\frac{f}{2}+\lambda}^3} \tag{A.38}$$

$$\Phi_{\chi^2(f,\lambda)}(s) = \frac{1}{\sqrt{1-2is}^f} \exp\left(i\lambda s + \frac{\lambda s^2}{2(1-2is)}\right) \tag{A.39}$$

$$\rho_{\chi^2(f,\lambda)}(x) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} \frac{x^{\frac{f}{2}+j-1} e^{-\frac{x}{2}}}{2^{\frac{f}{2}+j-1}\Gamma(\frac{f}{2}+j)} \qquad x \geq 0 \tag{A.40}$$

$$P[X \leq x_0] = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} P\left(\frac{x_0}{2}, \frac{f+2j}{2}\right) \qquad x_0 \geq 0 \tag{A.41}$$

A.4 The Cauchy distribution $C(\mu,\sigma)$

Definition A.15 *The density of the Cauchy distribution with location parameter μ and scaling parameter $\sigma > 0$ is given by:*

$$\rho_{C(\mu,\sigma)}(x) := \frac{1}{\pi\sigma} \cdot \frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2} \tag{A.42}$$

Lemma A.16 *Let X be $C(\mu,\sigma)$ distributed. Then holds:*

$$P[X \leq x_0] = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x_0-\mu}{\sigma}\right) \tag{A.43}$$

$$\Phi_{C(\mu,\sigma)}(s) = e^{i\mu s - \sigma|s|} \tag{A.44}$$

The density is symmetric about μ and decays so slowly that the moments of X do not exist and the characteristic function is not differentiable at $s = 0$.

This distribution is often used as comparing density to generate random numbers with respect to an other density by the rejection algorithm (see A.6.1). The reason why the Cauchy distribution is quite popular in this context is, that it fulfills two essential requirements. First, since the distribution is heavy tailed, a small multiple of the Cauchy distribution will dominate the other distribution. On the other hand, one needs to simulate random numbers with respect to the comparing density, for example $C(\mu,\sigma)$ distributed numbers, which can be done easily:

Lemma A.17 *Let U be a uniform distributed random number on $[0,1]$. Then*

$$X := \mu + \sigma \tan(\pi U) \tag{A.45}$$

is $C(\mu,\sigma)$ distributed.

A.5 The exponential distribution $E(\lambda)$

Definition A.18 *The density of the exponential distribution with intensity $\lambda > 0$ is given by*

$$\rho_{E(\lambda)}(x) := \lambda e^{-\lambda x} \qquad x \geq 0 \tag{A.46}$$

Lemma A.19 *Let X be an exponential distributed random number with intensity λ. Then the following relations hold:*

$$\mathbf{E}[X] = \frac{1}{\lambda} \tag{A.47}$$

$$\mathbf{Var}[X] = \frac{1}{\lambda^2} \tag{A.48}$$

$$\mathbf{Skew}[X] = 2 \tag{A.49}$$

$$\mathbf{E}[X^n] = \frac{n!}{\lambda^n} \tag{A.50}$$

$$\Phi_X(s) = \frac{\lambda}{\lambda - is} \tag{A.51}$$

$$P[X \leq T] = 1 - e^{-\lambda T} \qquad T \geq 0 \tag{A.52}$$

The next lemma is provides a way to generate $E(\lambda)$ distributed random numbers.

Lemma A.20 *Let U be a uniform distributed random number on $]0,1[$. Then the variable*

$$X := -\frac{1}{\lambda}\ln(U) \tag{A.53}$$

is exponentially distributed with intensity λ.

A.6 The Gamma distribution $\Gamma(\lambda, a)$

Definition A.21 *The density of the gamma distribution is given by:*

$$\rho_{\Gamma(\lambda,a)}(x) := \frac{\lambda}{\Gamma(a)} (\lambda x)^{a-1} e^{-\lambda x} \qquad \lambda, a > 0 \quad x \geq 0 \tag{A.54}$$

The gamma distribution is a generalization of the exponential distribution, which will be obtained for $a = 1$. For $\lambda = \frac{1}{2}$ the Gamma distribution is a special case of the χ^2 distribution, because $\chi^2(f) = \Gamma(\frac{1}{2}, \frac{f}{2})$.

Lemma A.22 *Let X be a $\Gamma(\lambda, a)$ distributed. Then holds:*

$$\mathbf{E}[X] = \frac{a}{\lambda} \tag{A.55}$$

$$\mathbf{Var}[X] = \frac{a}{\lambda^2} \tag{A.56}$$

$$\mathbf{Skew}[X] = \frac{2}{\sqrt{a}} \tag{A.57}$$

$$\mathbf{E}[X^n] = \frac{a \cdot (a+1) \cdot \ldots \cdot (a+n)}{\lambda^n} \tag{A.58}$$

$$\mathbf{E}[e^{-\alpha X}] = \frac{\lambda^a}{(\lambda+\alpha)^a} \tag{A.59}$$

$$\Phi_X(s) = \frac{\lambda^a}{(\lambda - is)^a} \tag{A.60}$$

$$P[X < x_0] = P(a, \lambda x) \qquad x_0 \geq 0 \tag{A.61}$$

$$\hat{G}(x) := \int_0^x P(a, \lambda y) dy = xP(a, \lambda x) - \frac{a}{\lambda} P(a+1, \lambda x) \tag{A.62}$$

P denotes the incomplete gamma function (see A.3.1).

Remark A.23 *There exists a gamma distribution for any mean μ and positive variance σ^2, and the corresponding parameters a and λ are given by*

$$a = \frac{\mu^2}{\sigma^2} \qquad \lambda = \frac{\mu}{\sigma^2} \tag{A.63}$$

The next lemma helps to generate $\Gamma(\lambda, a)$ distributed random numbers:

Lemma A.24 *Let Y be $\Gamma(1, a)$ distributed and define*

$$X := \frac{Y}{\lambda} \tag{A.64}$$

Then X is $\Gamma(\lambda, a)$ distributed.

Hence, in order to sample $\Gamma(\lambda, a)$ distributed random numbers it is enough to provide an algorithm to sample $\Gamma(1, a)$. To generate such random numbers, one may use the so–called "rejection method". This general method and its application to $\Gamma(1, a)$ will be presented now.

A.6.1 The Rejection method and how to sample $\Gamma(1, a)$

The rejection method is a general method to sample random numbers with respect to a given density $\rho(x)$. As precondition, one has to find an integrable function $g(x)$ which fulfills the two conditions:

1. It must hold $\rho(x) \leq g(x)$ for all x.
2. One has to know, how to sample random numbers with respect to the density

$$\tilde{g} := \frac{g(x)}{\int g(x) dx} \tag{A.65}$$

If such a g has been found, the rejection algorithm is quite simple:

Algorithm A.25 (Rejection–algorithm)

1. *Sample a $\tilde{g}$ distributed number X.*
2. *Sample (independent) a uniform [0,1] distributed number U.*
3. *If $\rho(X) \leq Ug(X)$ then return X, otherwise go to step 1 again.*

The computation time of this algorithm is given by the product of the time needed to generate one sample of X and U and the average number of loop passages, which is given by:

$$\mathbf{E}\left[\frac{\text{Samples of } \tilde{g}}{\text{Samples of } \rho}\right] = \int g(x)dx \tag{A.66}$$

How to simulate $\Gamma(1,a)$ for $a > 1$

Due to the second condition on $g(x)$, there are only a few functions used in practice. Quite popular is the choice based on the Cauchy distribution, hence

$$g(x) = \frac{c}{1 + \left(\frac{x-\mu}{\sigma}\right)^2} \tag{A.67}$$

and $\tilde{g}(x)$ is given by the density of $C(\mu,\sigma)$. For the Gamma distribution with $\lambda = 1$ and $a > 1$ one can set the parameters of the function g to

$$\mu = a - 1 \tag{A.68}$$
$$\sigma = \sqrt{2a-1} \tag{A.69}$$
$$c = \frac{e^{1-a}(a-1)^{(a-1)}}{\Gamma(a)} \tag{A.70}$$

and then $\rho_{\Gamma(1,a)}$ is dominated by g and one can easily sample Cauchy $C(\mu,\sigma)$ distributed random numbers, see lemma A.17. The average number of iterations is given by $\pi\sqrt{2a-1}e^{1-a}(a-1)^{a-1}/\Gamma[a] < \pi$.

How to simulate $\Gamma(1,a)$ for $0 < a \leq 1$

The same idea, but another g can be used to sample $\Gamma(1,a)$ with $0 < a \leq 1$. The density $\rho_{\Gamma(a,1)}$ is in this case bounded by

$$g(x) = \frac{1}{\Gamma(a)} \cdot \begin{cases} x^{a-1} & \text{if } 0 < x < 1 \\ e^{1-x} & \text{if } x \geq 1 \end{cases} \tag{A.71}$$

Normalization yields:

$$\tilde{g}(x) = \frac{a}{a+1} \cdot \begin{cases} x^{a-1} & \text{if } 0 < x < 1 \\ e^{1-x} & \text{if } x \geq 1 \end{cases} \tag{A.72}$$

The distribution function and its inverse are given by

$$\tilde{G}(x) = \begin{cases} \frac{1}{(a+1)}x^a & \text{if } 0 < x < 1 \\ 1 - \frac{a}{a+1}e^{1-x} & \text{if } x \geq 1 \end{cases} \tag{A.73}$$

$$\tilde{G}^{-1}(u) = \begin{cases} ((a+1)u)^{\frac{1}{a}} & \text{if } u \leq \frac{1}{1+a} \\ 1 - \ln\left((1-u)\frac{a+1}{a}\right) & \text{if } u > \frac{1}{1+a} \end{cases} \tag{A.74}$$

To sample $\tilde{g}$ distributed numbers one only has to apply $\tilde{G}^{-1}$ on uniform $[0,1]$ distributed random numbers. Hence one can use the rejection algorithm again and in this case the average number of iterations is given by $\frac{(a+1)}{a\Gamma(a)} \leq 2$.

A.6.2 The generalized Gamma distribution $\Gamma(\lambda, a, x_0, \beta)$

Definition A.26 *The density of the generalized Gamma distribution* $\Gamma(\lambda, a, x_0, \beta)$ *with* $\lambda, a, \beta > 0$ *and* $x_0 \in \mathbb{R}$ *is given by*

$$\rho_{\Gamma(\lambda,a,x_0,\beta)}(x) := \frac{\lambda}{\beta\Gamma(a)}[\lambda(x-x_0)]^{\frac{a}{\beta}-1}e^{-[\lambda(x-x_0)]^{\frac{1}{\beta}}} \qquad x \geq x_0 \tag{A.75}$$

The special case $\Gamma(\lambda, a, x_0, 1)$ *is called the* shifted Gamma distribution.

The following relation to the Gamma distribution shows that the generalized Gamma distribution is in fact a generalization of the Gamma distribution and provides also a method to sample $\Gamma(\lambda, a, x_0, \beta)$ distributed random numbers.

Lemma A.27 *Let* Y *be* $\Gamma(1, a)$ *distributed and define*

$$X := \frac{1}{\lambda}Y^\beta + x_0 \tag{A.76}$$

with $\lambda, \beta > 0$. *Then* X *is* $\Gamma(\lambda, a, x_0, \beta)$ *distributed.*

Obviously, the relation $\Gamma(\lambda, a, 1, 0) = \Gamma(\lambda, a)$ holds. The main properties of the generalized Gamma distribution are:

Lemma A.28 *Let* X *be* $\Gamma(\lambda, a, x_0, \beta)$ *distributed. Then holds:*

$$\mathbf{E}[X] = x_0 + \frac{\Gamma(a+\beta)}{\lambda\Gamma(a)} \tag{A.77}$$

$$\mathbf{Var}[X] = \frac{1}{\lambda^2}\left(\frac{\Gamma(a+2\beta)}{\Gamma(a)} - \frac{\Gamma(a+\beta)^2}{\Gamma(a)^2}\right) \tag{A.78}$$

$$\mathbf{E}[(X-x_0)^n] = \frac{\Gamma(a+n\beta)}{\lambda^n\Gamma(a)} \tag{A.79}$$

$$P[X \leq x] = P\left(a, [\lambda(x-x_0)]^{\frac{1}{\beta}}\right) \qquad x \geq x_0 \tag{A.80}$$

There is no closed form solution for the characteristic function in general, in the special case of the shifted Gamma distribution, hence $\beta = 1$, *the characteristic function is given by:*

$$\Phi_{\Gamma(\lambda,a,x_0,1)} = \frac{\lambda^a e^{isx_0}}{(\lambda - is)^a} \tag{A.81}$$

A.7 The Student distribution $t_f(\mu, \sigma^2)$

Definition A.29 *Let X be a real-valued random variable with the density*

$$\rho_{t_f(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{\pi f \sigma^2}} \frac{\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f}{2})} \left(1 + \frac{(x-\mu)^2}{f\,\sigma^2}\right)^{-\frac{f+1}{2}} \tag{A.82}$$

where $\sigma, f > 0$. Then X is Student distributed, $X \sim t_f(\mu, \sigma^2)$.

Note, that this definition incorporates a location parameter μ and a scaling parameter σ. In several textbooks these parameters are not given and that density can be obtained by setting $\mu = 0$ and $\sigma = 1$. The Cauchy distribution is a special case of the Student distribution, which is obtained in the case $f = 1$. The main properties of this distribution are listed below.

Lemma A.30 *Let X be $t_f(\mu, \sigma^2)$ distributed. Then holds:*

$$\mathbf{E}[X] = \mu \quad \text{if } f > 1 \tag{A.83}$$

$$\mathbf{Var}[X] = \sigma^2 \frac{f}{f-2} \quad \text{if } f > 2 \tag{A.84}$$

$$\mathbf{Skew}[X] = 0 \quad \text{if } f > 3 \tag{A.85}$$

$$\mathbf{E}[(X-\mu)^n] = \begin{cases} \sigma^n \frac{\sqrt{f}^n}{2\sqrt{\pi}} \frac{\Gamma(\frac{f-n}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{f}{2})} & \text{if } n \text{ even} \\ 0 & \text{else} \end{cases} \quad \text{if } f > n \tag{A.86}$$

$$\mathbf{E}[|X-\mu|^n] = \sigma^n \frac{\sqrt{f}^n}{\sqrt{\pi}} \frac{\Gamma(\frac{f-n}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{f}{2})} \quad \text{if } f > n \tag{A.87}$$

$$\Phi_X(s) = \frac{2e^{i\mu s}}{\Gamma(\frac{f}{2})} \left(\frac{\sqrt{f}\sigma|s|}{2}\right)^{\frac{f}{2}} K_{\frac{f}{2}}(\sqrt{f}\sigma|s|) \tag{A.88}$$

where $K_\nu(x)$ denotes the modified Bessel function of the third kind[2]*.*

Lemma A.31 *Let $f > 1$ and define*

$$\tilde{\mu} := \mu\sqrt{\frac{2}{f}} \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})} \tag{A.89}$$

Let $N \sim N(\tilde{\mu}, \sigma^2)$ and $X \sim \chi^2(f)$ where N and X are independent. Then the random variable

$$T := \frac{N}{\sqrt{X/f}} \tag{A.90}$$

is Student distributed, $T \sim t_f(\mu, \sigma^2)$.

[2]This denotation follows [1]. $K_\lambda(x)$ is sometimes called the modified Bessel function of the second kind. The difference in the nomination depends on whether there is a distinction in positive and negative orders of the modified Bessel function of the first kind (into first and second kind) or not.

To sample $t_f(\mu, \sigma^2)$ distributed random numbers for $f \geq 1$ one can use either the last lemma or the rejection algorithm (see A.6.1) with the Cauchy distribution as comparing distribution:

Lemma A.32 *Let $f \geq 1$, $\sigma > 0$ and $\mu \in \mathbb{R}$. Then holds for all $x \in \mathbb{R}$:*

$$\rho_{t_f(\mu,\sigma^2)}(x) \leq \frac{\sqrt{\pi}\,\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f}{2})}\,\rho_{C(\mu,\sigma\sqrt{f})}(x) \tag{A.91}$$

If one samples $t_f(\mu, \sigma^2)$ distributed random variables with $f > 1$ using the rejection method based on this lemma, the average number of iterations is bounded by

$$\frac{\sqrt{\pi}\,\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f}{2})} \leq \sqrt{\frac{\pi f}{2}} \approx 1.253\sqrt{f} \tag{A.92}$$

A.8 The multivariate normal distribution $N_d(\mu, \Sigma)$

Definition A.33 *Let μ be a d dimensional vector and Σ a symmetric semi–positive $d \times d$ matrix. Define a $d \times d$ matrix A such that $\Sigma = AA^\top$ holds and a d dimensional vector Y with independent $N(0,1)$ distributed components. Then the vector X defined by*

$$X := \mu + AY \tag{A.93}$$

is multivariate normal distributed, $X \sim N_d(\mu, \Sigma)$.

For the existence of the Cholesky decomposition $AA^\top$ of the matrix Σ and an efficient algorithm to obtain A see chapter 2. This definition of the multivariate normal distribution provides also an algorithm do sample $N_d(\mu, \Sigma)$ distributed random numbers. The vector μ is the mean and the matrix Σ is the covariance of the distribution:

Lemma A.34 *Let $X \sim N_d(\mu, \Sigma)$. Then holds:*

$$\mathbf{E}[X_k] = \mu_k \tag{A.94}$$

$$\mathbf{Cov}(X_k, X_l) = \mathbf{E}[(X_k - \mu_k)(X_l - \mu_l)] = \Sigma_{kl} \tag{A.95}$$

Lemma A.35 *If the symmetric matrix Σ is positive, the multivariate normal distribution $N_d(\mu, \Sigma)$ has a density which is given by:*

$$\rho_{N_d(\mu,\Sigma)}(x) = \frac{1}{\sqrt{2\pi}^d}\frac{1}{\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right) \tag{A.96}$$

Lemma A.36 *Let X be $N_d(\mu, AA^\top)$ distributed and $V \in \mathbb{R}^d$. Then holds:*

$$\mathbf{E}\left[e^{isV\cdot X}\right] = e^{isV\cdot\mu - \frac{1}{2}s^2||A^\top V||^2} \tag{A.97}$$

A.8.1 The bivariate normal distribution $N_2(\mu, \Sigma)$

A special case of the multivariate normal distribution is the bivariate normal distribution, hence $d = 2$. In this case, the dependence between the two normal variables is described by one real number, the correlation ρ.

Definition A.37 *The standard bivariate normal density and distribution functions are defined by:*

$$n_2(x, y; \rho) := \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) \tag{A.98}$$

$$\mathcal{N}_2(x, y; \rho) := \int_{-\infty}^{x}\int_{-\infty}^{y} n_2(u, v; \rho)\, du\, dv \tag{A.99}$$

In [53] a numerical algorithm to compute the function $\mathcal{N}_2$ is presented and a source code for this function is available at `http://www.MathFinance.de/frontoffice.html`. An example where this function occurs in the pricing formula of an option is given by the Rainbow option example in 1.4.1.

Lemma A.38 *Let* $\Sigma = \left(\begin{array}{c|c} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \hline \sigma_1\sigma_2\rho & \sigma_2^2 \end{array}\right)$ *with* $\sigma_1, \sigma_2 > 0$ *and* $|\rho| < 1$. *Then the density of* $N_2(\mu, \Sigma)$ *is given by*

$$\rho_{N_2(\mu,\Sigma)}(x_1, x_2) = n_2\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}; \rho\right) \tag{A.100}$$

Lemma A.39 *Let* $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ *and the correlation between* X_1 *and* X_2 *is given by* ρ. *Then holds:*

$$(X_1 + X_2) \sim N\left(\mu_1 + \mu_2\ ,\ \sigma_1^2 + 2\sigma_1\sigma_2\rho + \sigma_2^2\right) \tag{A.101}$$

A.9 Multivariate lognormal distribution $Ln(\mu, \Sigma)$

Definition A.40 *Let* Σ *be a semi–positive* $d \times d$ *matrix and* $\mu \in \mathbb{R}^d$. *Let further* N *be* $N_d(\mu, \Sigma)$ *distributed and define*

$$X_k := \exp(N_k) \qquad k = 1, \ldots, d \tag{A.102}$$

Then X *is multivariate lognormal distributed,* $X \sim Ln(\mu, \Sigma)$.

This distribution is very important in financial mathematics, since it is a quite usual assumption, that stocks or foreign exchange rates are dependent lognormal distributed. For a discussion of this assumption see 4.3. In chapter 5 the lognormal distribution with mean 1 is also suggested to describe the sector variables in the CreditRisk$^+$ model.

Lemma A.41 *Let* $X \sim Ln(\mu, \Sigma)$. *Then holds:*

$$\mathbf{E}[X_k] = \exp\left(\mu_k + \frac{1}{2}\Sigma_{kk}\right) \tag{A.103}$$

$$\mathbf{E}[X_k X_l] = \exp\left(\mu_k + \mu_l + \frac{1}{2}(\Sigma_{kk} + 2\Sigma_{kl} + \Sigma_{ll})\right) \tag{A.104}$$

$$\mathbf{Cov}(X_k, X_l) = \exp\left(\mu_k + \mu_l + \frac{1}{2}(\Sigma_{kk} + \Sigma_{ll})\right)\left(\exp(\Sigma_{kl}) - 1\right) \tag{A.105}$$

A.10 The multivariate Student distribution $t_f(\tilde{\mu}, \Sigma)$

Definition A.42 *Let $\tilde{\mu}$ be a d dimensional vector, Σ a semi–positive $d \times d$ matrix and $f > 0$. Define $N \sim N_d(\tilde{\mu}, \Sigma)$ and $X \sim \chi^2(f)$ independent from N. Then the d dimensional random variable*

$$Y := \frac{1}{\sqrt{X/f}} N \tag{A.106}$$

is multivariate Student distributed, $Y \sim t_f(\tilde{\mu}, \Sigma)$.

Lemma A.43 *Let X be $t_f(\tilde{\mu}, \Sigma)$ distributed. Then holds:*

$$\mathbf{E}[X_k] = \tilde{\mu}_k \sqrt{\frac{f}{2}} \frac{\Gamma(\frac{f-1}{2})}{\Gamma(\frac{f}{2})} \qquad \text{if } f > 1 \tag{A.107}$$

$$\mathbf{E}[X_k X_l] = \frac{f}{f-2} (\Sigma_{kl} + \tilde{\mu}_k \tilde{\mu}_l) \qquad \text{if } f > 2 \tag{A.108}$$

Acknowledgements

At first I would like to thank John Schoenmakers (Weierstrass Institute) for his permanent dedication to discuss posed problems and I'm grateful to him for many valuable comments and suggestions. Further I enjoyed the close and pleasant cooperation with Uwe Wystup (Commerzbank) in the research on symmetries in financial markets. My thanks also go to Hermann Haaf (Commerzbank) for rising interesting questions regarding the CreditRisk$^+$ model.

I would like to thank Vladimir Spokoiny (Weierstrass Institute) and my colleagues at the Weierstrass institute for the convenient working atmosphere. A special thank goes to Jürgen Franke (University of Kaiserslautern), who supported me kindly during the PhD proceeding.

I am also grateful for the financial and scientific support within the following projects. I was partially affiliated to the Delft University of Technology, Netherlands, with support of the NWO. There has been a closed cooperation with the bmb+f project "Effiziente Methoden zur Bestimmung von Risikomaßen" (Efficient Methods for the Valuation of Risk Measures), which is also supported by Bankgesellschaft Berlin AG. My Research was also supported by the DFG Research Center "Mathematics for key technologies" (FZT 86) in Berlin.

Besides my honest thanks for the scientific and financial support, I am also very grateful to my friends for their private assistance. My thanks also go the musical circle KSO of the Collegium musicum of the Technische Universität Berlin and the Freie Universität Berlin, especially to the viola group and the members of the "Oberseminar".

Bibliography

[1] Milton Abramowitz and Irene A. Stegun (Eds.): *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, Dover Publications, Inc., New York, ninth printing, 1972.

[2] Philippe Artzner, Freddy Delbaen, Jean–Marc Eber and David Heath: Coherent Measures of Risk, *Mathematical Finance* Vol. **9**, No. 3, pp. 203–228, (1999).

[3] Louis M. Bachelier: Théorie de la Spéculation, *Ann. Sci. École Norm. Sup.* Vol. **17**, pp. 21–86, (1900), Reprint: Theory of Speculation in *The Random Character of Stock Market Prices*, Paul H. Cootner (Ed.), Cambridge, Mass. (1964).

[4] Bank for International Settlements: International convergence of capital measurements and capital standards (Basel Capital Accord), *Basel Committee Publications*, No. **4**, (July 1988), available at `http://www.bis.org`.

[5] Bank for International Settlements: An internal model–based approach to market risk capital requirements, *Basel Committee Publications* No. **17**, (April 1995), available at `http://www.bis.org`.

[6] Bank for International Settlements: *The New Basel Capital Accord*, Consultative Document of the Bank for International Settlements, January 2001, available at `http://www.bis.org`.

[7] Christian Bauer: Value at Risk using Hyperbolic Distributions, *Journal of Economics and Business*, Vol. **52**, pp. 455–467, (2000).

[8] Verlag C.H.Beck: *Bankrecht*, 28. Auflage, Deutscher Taschenbuch Verlag, 1999.

[9] Peter L. Bernstein: *Capital Ideas: The Improbable Origins of Modern Wall Street*, The Free Press, New York, NY, 1993.

[10] Fischer Black and Myron Scholes: The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, Vol. **81**, No. 3, pp. 637–654, (1973).

[11] Milan Borkovec and Claudia Klüppelberg: Extremwerttheorie für Fianzzeitreihen – ein unverzichtbares Werkzeug im Risikomanagement, in *Handbuch Risikomanagement*, B. Rudolph, L. Johanning (Eds.), Uhlenbruch Verlag, Bad Soden, 2000, pp. 219–241.

[12] M. Broadie and P. Glasserman: Estimating Security Price Derivatives using Simulation, *Management Science*, Vol. **42**, pp. 269–285, (1996).

[13] I.N. Bronstein and K.A. Semendjajew: *Taschenbuch der Mathematik*, 25. Auflage, G. Grosche, V. Ziegler, D. Ziegler (Eds.), B.G. Teubner Verlagsgesellschaft Stuttgart, Leipzig and Verlag Nauka Moskau, 1991.

[14] Richard A. Brualdi and Stephen Mellendorf: Regions in the Complex Plane Containing the Eigenvalues of a Matrix, *Amer. Math. Monthly*, Vol. **101**, pp. 975–985, (1994).

[15] Peter Bürgisser, Alexander Kurth, Armin Wagner and Michael Wolf: Integrating Correlations, *RISK Magazine : Risk*, Vol. **12**, No. 7, (July 1999).

[16] Peter Bürgisser, Alexander Kurth and Armin Wagner: Incorporating Severity Variations into Credit Risk, *Journal of Risk*, Vol. **3**, No. 4, (Summer 2001).

[17] T. Carleman: Sur le problème des moments, *Comptes rendus hebdomadaires des sèances de l'Acadèmie des Sciences*, **174**, pp. 1680–1682, (1922).

[18] P. Carr: Deriving Derivatives of Derivative Securities, *The Journal of Computational Finance*, Vol. **4**, No. 2, pp. 5–29, (2001).

[19] Credit Suisse First Boston: CreditRisk$^+$: A Credit Risk Management Framework, available at `http://www.csfb.com/creditrisk`, (1997).

[20] J.C. Cox, J.E. Ingersoll and S.A. Ross: A Theory of the Term Structure of Interest Rates, *Econometrica* **53**, pp. 385–407, (1985).

[21] Robert C. Dalang, Andrew Morton and Walter Willinger: Equivalent Martingale Measures and No–Arbitrage in Stochastic Securities Market Models, *Stochastics and Stochastic Reports*, Vol. **29**, pp. 185–201, (1990).

[22] Freddy Delbaen and Walter Schachermayer: A general version of the fundamental theorem of asset pricing, *Mathematische Annalen*, Vol. **300**, pp. 463–520, (1994).

[23] Freddy Delbaen and Walter Schachermayer: The fundamental theorem of asset pricing for unbounded stochastic processes, *Mathematische Annalen*, Vol. **312**, pp. 215–250, (1998).

[24] Peter Deuflhard and Andeas Hohmann: *Numerische Mathematik 1: Eine algorithmisch orientierte Einführung*, Walter de Gruyter, Berlin New York, 2. Auflage, 1993.

[25] Deutsche Bundesbank: Monatsbericht Mai 1998. available at `http://www.bundesbank.de`.

[26] Susan J. Devlin, R. Gnanadesikan and J.R.Kettenring: Robust estimation and outlier detection with correlation coefficients, *Biometrika*, Vol. **62**, No. 3, pp. 531–545, (1975).

[27] Darrell Duffie and Jun Pan: Analytical value-at-risk with jumps and credit risk, *Finance and Stochastics*, Vol **5**, No. 2, pp. 155–180, (2001).

[28] Daniel Dufresne: The Distribution of a Perpetuity, with Applications to Risk Theory and Pension Funding, *Scand. Actuarial J.*, pp. 39–79, (1990).

[29] Daniel Dufresne: The Integral of geometric Brownian motion, *Adv. Appl. Prob.*, Vol. **33**, pp. 223–241, (2001).

[30] Ernst Eberlein and Ulrich Keller: Hyperbolic Distributions in Finance, *Bernoulli*, Vol. **1**, No. 3, pp. 281–299, (1995).

[31] Ernst Eberlein and Karsten Prause: The generalized hyperbolic model: financial derivatives and risk measures, in *Mathematical Finance - Bachelier Congress 2000*, H. Geman, D. Madan, S. Pliska and T. Vorst (Eds.) Springer Verlag, 2002.

[32] R. Eller and H. Deutsch: *Derivate und interne Modelle: Modernes Risikomanagement*, Schäffer Poeschel Verlag, Stuttgart, 1998.

[33] Susanne Emmer, Claudia Klüppelberg and Ralf Korn: Optimal Portfolios with bounded Capital-at-Risk, *Mathematical Finance*, Vol. **11**, pp. 365–384, (2001).

[34] William Fallon: Calculating Value-at-Risk, *Working paper of the Wharton Financial Institutions Center* No. 96-49, (1996).

[35] William Feller: *An Introduction to Probability Theory and Its Applications*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc. New York. Volume I, Third Edition, 1968 – Volume II, 1966.

[36] Hélyette Geman and Marc Yor: Bessel Processes, Asian Options, and Perpetuities, *Mathematical Finance*, Vol. **3**, No. 4, pp. 349–375, (October 1993).

[37] R. Geske: The Valuation of Compound Options, *Journal of Financial Economics*, Vol. **7**, pp. 63–81, (1979).

[38] Götz Giese: Enhancing CreditRisk$^+$, *Risk*, Vol. **16**, No 4, (2003).

[39] Philip E. Gill, Walter Murray and Margaret H. Wright: *Practical Optimization*, Academic Press, Eighth Printing, 1989.

[40] Paul Glasserman, Philip Heidelberger and Perwez Shahabuddin: Importance Sampling and Stratification for Value-at-Risk, in *Proceedings of the Sixth International Conference on Computational Finance 1999*, Yaser S. Abu–Mostafa, Blake Le Baron, Anrew W. Lo, Andreas S. Weigend, (Eds.), pp. 7–24, MIT Press, Cambridge MA, 2000.

[41] Paul Glasserman, Philip Heidelberger and Perwez Shahabuddin: Variance Reduction Techniques for Estimating Value-at-Risk, *Management Science*, Vol. **46**, No. 10, pp. 1349–1364, (2000).

[42] Paul Glasserman, Philip Heidelberger and Perwez Shahabuddin: Efficient Monte Carlo Methods for Value-at-Risk, in *Mastering Risk Volume 2: Applications*, Carol Alexander (Ed.), pp. 7–20, Financial Times Prentice Hall, London, 2001.

[43] Paul Glasserman, Philip Heidelberger and Perwez Shahabuddin: Portfolio Value-at-Risk with Heavy-tailed Risk Factors, *Mathematical Finance* Vol. **12**, No. 3, pp. 239–269, (2002).

[44] Paul Glasserman and X. Zhao: Fast Greeks by Simulation in Forward LIBOR Models, *The Journal of Computational Finance*, Vol. **3**, No. 1, pp 5–39, (1999).

[45] Gene H. Golub and Charles F. van Loan: *Matrix Computations*, third edition, Johns Hopkins University Press, Baltimore, 1996.

[46] Michael B. Gordy: Saddlepoint Approximation of CreditRisk$^+$, *Journal of Banking and Finance*, Vol **26**, No. 7, pp.1337 – 1355, (August 2002).

[47] P.R.Halmos: Positive approximants of operators, *Indiana Univ. Math. J.*, **21**, pp. 951–960, (1972).

[48] J. Michael Harrison and Stanley R. Pliska: Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stochastic Processes and their Applications* Vol. **11**, pp. 215–260, (1981).

[49] J. Michael Harrison and Stanley R. Pliska: A Stochastic Calculus Model of Continuous Trading: Complete Markets, *Stochastic Processes and their Applications* Vol. **15**, pp. 313–316, (1983).

[50] S. Heston: A Closed–Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *The Review of Financial Studies*, Vol. **6**, No. 2, (1993).

[51] R. Heynen and H. Kat: Crossing Barriers, *Risk*, Vol. **7**, No. 6, pp. 46–51, (1994).

[52] Nicholas J. Higham: Computing a Nearest Symmetric Positive Semidefinite Matrix, *Linear Algebra and Its Applications*, Vol. **103**, pp. 103–118, (1988).

[53] John C. Hull: *Options, Futures, and other Derivatives*, Prentice Hall, London, third edition, 1997.

[54] Stefan R. Jaschke: Quantile-VaR is the Wrong Measure to Quantify Market Risk for Regulatory Purposes, *SFB 373 Discussion Paper*, No. 2001–55, Humboldt University Berlin, (2001).

[55] Stefan R. Jaschke: The Cornish Fischer Expansion in the Context of Delta-Gamma-Normal Approximations, *Journal of Risk*, Vol. **4**, No. 4, pp. 33–52, (2002).

[56] Stefan R. Jaschke: Maths of all angles, *GARP Risk Review* No. 6, May/June 2002.

[57] V. A. Kholodnyi and J. F. Price: *Foreign Exchange Option Symmetry*, World Scientific Publishing, Singapore, 1998.

[58] Claudia Klüppelberg: Risk management with extreme value theory, in *Extreme Values in Finance, Telecommunications and the Environment*, B. Finkenstädt and H. Rootzen (Eds.), Chapman & Hall / CRC, Boca Raton, To appear 2003.

[59] Sandro Merino and Mark Nyfeler: Calculating portfolio loss, *Risk*, Vol. **15**, No. 8, pp. 82–86, (2002).

[60] Robert C. Merton: Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science*, Vol. **4**, pp. 141–183, (1973).

[61] Jorge Mina and Andrew Ulmer: Delta–Gamma Four Ways, Working paper, RiskMetrics Group, 1999.

[62] Jorge Mina: Calculating VaR through Quadratic Approximations, *Journal of Risk Finance*, Vol. **2**, No. 2, pp. 49–55, (2001).

[63] Masatake Mori and Takuya Ooura: The double exponential formula for oscillatory functions over the half infinite interval, *J. Comput. Appl. Math.* Vol. **38**, pp. 353–360, (1991).

[64] Masatake Mori and Hidetosi Takahasi: Double exponential formulas for numerical integration, *Publ. RIMS. Kyoto Univ.*, Vol. **10**, pp. 721–741, (1974).

[65] M.F.M. Osborne: Brownian Motion in the Stock Market, *Operations Research*, Vol. **7**, No. 2, pp. 145–173, (1959).

[66] Antoon Pelsser and Ton C.F. Vorst: The Binomial Model and the Greeks, *Journal of Derivatives*, Vol. **1** , No. 3, pp. 45–49, (Spring 1994).

[67] R.L. Plackett: A Reduction Formula for Normal Multivariate Integrals, *Biometrika*, Vol. **41**, pp. 351–360, (1954).

[68] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery: *Numerical Recipes in C : The Art of Scientific Computing*, Second Edition, Cambridge University Press, Cambridge, 1992.

[69] Riccardo Rebonato and Peter Jäckel: The most general methodology to create a valid correlation matrix for risk management and option pricing purposes, Working paper, October 1999, available at `http://www.rebonato.com`.

[70] Horst Rinne: *Taschenbuch der Statistik*, Verlag Harri Deutsch, Thun und Frankfurt am Main, 1997.

[71] Peter J. Rousseeuw and Geert Molenberghs: Transformation of non positive semidefinite correlation matrices, *Commun. Statist. - Theory Meth.*, Vol. **22**, pp. 965–984, (1993).

[72] C. Rouvinez Going Greek with VAR, *Risk*, Vol. **10**, No. 2, pp. 57–65, (1997).

[73] Robert B. Schnabel and Elizabeth Eskow: A new modified Cholesky Factorization, *SIAM J. Sci. Stat. Comput.*, Vol **11**, No 6, pp. 1136–1158, (1990).

[74] Robert B. Schnabel and Elizabeth Eskow: A revised modified Cholesky factorization algorithm, *SIAM J. Optim.*, Vol **9**, No 4, pp. 1135–1148, (1999).

[75] W. Shaw: *Modelling Financial Derivatives with Mathematica*, Cambridge University Press, 1998.

[76] Steven E. Shreve: Stochastic Calculus and Finance, Lecture notes (1996), Carnegie Mellon University, available at
`http://www.cs.cmu.edu/~chal/shreve.html`.

[77] Gerold Studer: Market risk computation for nonlinear portfolios, *Journal of Risk*, Vol **1**, No 4, (1999).

[78] R. Stulz: Options on the Minimum or Maximum of Two Assets, *Journal of Financial Economics*, Vol. **10**, pp. 161–185, (1982).

[79] N. Taleb: *Dynamic Hedging*, Wiley, New York, 1996.

[80] Lloyd N. Trefethen and David Bau: *Numerical Linear Algebra*, siam Society for Industrial and Applied Mathematics, 1997.

[81] Christoph W. Ueberhuber: *Numerical Computation 1: Methods,Software, and Analysis*, Springer-Verlag Berlin Heidelberg, 1997.

[82] Thomas Wilson: Plugging the GAP, *Risk*, Vol. **7**, No. 10, pp. 74–80, (1994).

[83] Thomas Wilson: Value at Risk, in *Risk Management and Analysis*, C. Alexander (Ed.), Wiley, Chichester, England, 1999.

[84] Uwe Wystup: *Vanilla Options*, Formula Catalogue, available at
`http://www.mathfinance.de`.

[85] Marc Yor: On some exponential functionals of Brownian motion, *Adv. Appl. Prob.* Vol. **24** No. 3, pp. 509 – 531, (1992).

Publications

[86] Hermann Haaf, Oliver Reiß and John Schoenmakers: Numerically Stable Computation of CreditRisk$^+$, *WIAS–Preprint* No. 846, (2003).

[87] Hermann Haaf, Oliver Reiß and John Schoenmakers: Numerically Stable Computation of CreditRisk$^+$, in *CreditRisk$^+$ in the banking industry*, Matthias Gundlach, Frank Lehrbass (Eds.), Springer Verlag, Berlin Heidelberg, to appear in 2004.

[88] Grigori N. Milstein, Oliver Reiß and John Schoenmakers: Monte Carlo methods for pricing and hedging American options, *WIAS–Preprint* No. 850, (2003).

[89] Oliver Reiß: A generalized non–square Cholesky decomposition algorithm with applications to finance, *WIAS–Preprint* No. 760, (2002).

[90] Oliver Reiß: Fourier Inversion algorithms for generalized CreditRisk$^+$ models and an extension to incorporate market and credit risk, *WIAS–Preprint* No. 817, (2003).

[91] Oliver Reiß: Fourier Inversion Techniques for CreditRisk$^+$, in *CreditRisk$^+$ in the banking industry*, Matthias Gundlach, Frank Lehrbass (Eds.), Springer Verlag, Berlin Heidelberg, to appear in 2004.

[92] Oliver Reiß: A Generalization of CreditRisk$^+$ with an Extension to Incorporate Market Risk, *CreditRisk$^+$ in the banking industry*, Matthias Gundlach, Frank Lehrbass (Eds.), Springer Verlag, Berlin Heidelberg, to appear in 2004.

[93] Oliver Reiß, John Schoenmakers and Martin Schweizer: Endogenous interest rate dynamics in asset markets, *WIAS–Preprint* No. 652, (2001).

[94] Oliver Reiß and Uwe Wystup: Efficient computation of option price sensitivities using homogeneity and other tricks, *WIAS–Preprint* No. 584, (2000).

[95] Oliver Reiß and Uwe Wystup: Efficient computation of option price sensitivities using homogeneity and other tricks, *The Journal of Derivatives*, Vol. **9**, No. 2, pp. 41–53, (2001).

[96] Oliver Reiß and Uwe Wystup: Efficient computation of option price sensitivities using homogeneity and other tricks, in *Foreign Exchange Risk: Models, Instruments and Strategies*, Jürgen Hakala and Uwe Wystup (Eds.), Risk Books, London, 2002.

PhD - Declaration

Die vorliegende Arbeit wurde eigenständig verfasst und alle benutzten Hilfsmittel wurden genannt. Diese oder eine ähnliche Arbeit wurde nicht als Prüfungsarbeit für eine staatliche oder akademische Prüfung eingereicht. Soweit Teile dieser Arbeit veröffentlicht sind oder Teile nach heutigem Stand veröffentlicht werden, ist das in dieser Dissertationsschrift vermerkt.

Berlin, Juni 2003

This work has been written by myself and I have mentioned all resources which have been used. This or a similar work has not been submitted for an official or academic exam. As far as some parts of this work have been published or will be published by actual state, this is indicated in this thesis.

Berlin, June 2003

Lebenslauf

Oliver Reiß
Geboren am 2. Oktober 1972 in Münster (Westfalen), Deutschland.
Nationalität: Deutsch

6.Aug. 1979 – 6.Jul. 1983	Besuch der "städtische Grundschule Stiftberg" in Herford.
22.Aug. 1983 – 11.Jun. 1992	Besuch des "Königin Mathilde Gymnasium der Stadt Herford", abgeschlossen mit der "allgemeinen Hochschulreife".
WS 1992/93 – SS 1998	Studium der Physik an der Universität Osnabrück; die Studienschwerpunkte waren mathematische und theoretische Physik sowie die Informatik. Das Thema der Diplomarbeit war "Relativistische Behandlung von Bewegungen mit Zwangsbedingungen".
9.Apr. 1998	Erhalt des Hochschulabschlusses "Diplom Physiker" mit der Gesamtnote "Sehr gut".
Seit September 1998	Wissenschaftlicher Mitarbeiter in der Forschungsgruppe "Stochastische Algorithmen und nichtparametrische Statistik" am Weierstraß–Institut für Angewandte Analysis und Stochastik in Berlin. Meine Tätigkeiten umfassen die Forschung auf verschiedenen Gebieten der angewandten Finanzmathematik und die Unterstützung der Beratungsaktivitäten des Instituts in diesem Bereich.

Berlin, Juni 2003

Curriculum Vitae

Oliver Reiß
Born at 2nd of October, 1972 in Münster (Westfalen), Germany.
Nationality: German

6.Aug. 1979 – 6.Jul. 1983	Attendance of the "städtische Grundschule Stiftberg" in Herford.
22.Aug. 1983 – 11.Jun. 1992	Attendance of the "Könige Mathilde Gymnasium der Stadt Herford", finished with the "allgemeinen Hochschulreife" (general qualification for university entrance).
WS 1992/93 – SS 1998	Study of physics at the university of Osnabrück; the major fields of study were mathematical and theoretical physics, and computer science. The title of the diploma thesis was "Relativistische Behandlung von Bewegungen mit Zwangsbedingungen" (Relativistic handling of movements under constraints).
9.Apr. 1998	Receiving of the university degree "Diplom Physiker" (master of physics) with the final grade "Sehr gut".
Since September 1998	Scientific Assistant at the Weierstrass Institute for Applied Analysis and Stochastics in Berlin, associated to the research group "Stochastic Algorithms and Non-parametric Statistics". My activities cover the research on several aspects of applied mathematical finance and the assistance of the institutes consulting activities in this field.

Berlin, June 2003