Carl von Ossietzky
Universität
Oldenburg

Axiomatic profit and loss decomposition of path-independent instruments in continuous time and numerical approximations

Von der Fakultät V für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von

Herrn Hauke Stier
geboren am 31.07.1998 in Nordenham.

Betreuender Gutachter: Prof. Dr. Gero Junike

Zweite Gutachterin: Prof. Dr. Angelika May

Tag der Disputation: 20.03.2025

Zusammenfassung

(German)

Diese Arbeit besteht aus vier Abschnitten. Im ersten Abschnitt konstruieren wir eine Klasse von zeitkontinuierlichen Zerlegungen von pfadunabhängigen Finanzinstrumenten, wie zum Beispiel dem Preis einer Aktienoption mit den Risikofaktoren Zins oder Wechselkurs. Im Vergleich zu Zerlegungen, die die Risikofaktoren nur an diskreten Zeitpunkten auswerten, berücksichtigt diese Klasse die gesamte Information der Pfade und ist konsistent für verschiedene Zeitgitter. Wir zeigen, dass es eine eindeutige Zerlegung gibt, indem wir die drei Eigenschaften Exaktheit, Symmetrie und Normalisierung fordern. Diese eindeutige Zerlegung ist der stochastische Grenzwert einer zeitdiskreten Zerlegung und leidet unter dem Fluch der Dimensionen. Haben die Risikofaktoren keine gleichzeitigen Sprünge, so zeigen wir, dass der numerische Aufwand erheblich von exponentiellem zu linearem Wachstum reduziert werden kann.

Im ersten Abschnitt zeigen wir die stochastische Konvergenz von diskreten zu zeitkontinuierlichen Zerlegungen. In der Praxis können wir nur einen Pfad der Risikofaktoren beobachten. Um sicherzustellen, dass die Zerlegungen dieses Pfades konvergieren, zeigen wir im zweiten Abschnitt, dass es ein pfadabhängiges, nicht-äquidistantes Zeitgitter gibt, das fast sichere Konvergenz gewährleistet. Weiter stellen wir Bedingungen auf, die eine bestimmte Konvergenzgeschwindigkeit gewährleisten.

Zur Approximation von zeitkontinuierlichen nutzen wir zeitdiskrete Zerlegungen mit einem feinen Gitter. Zur Zerlegung von Optionen müssen dazu die Preise sehr häufig berechnet werden. Für eine schnelle Berechnung analysieren wir im dritten Abschnitt die COS Methode, eine Fourier-Approximationstechnik zur Bepreisung von Aktienoptionen. Die Methode erfordert die Existenz der charakteristischen Funktion des Finanzmodells. Da die klassische COS Methode für bestimmte Optionen in allgemeinen Dimensionen vergleichsweise langsam ist, führen wir die gedämpfte COS Methode ein. Wir beweisen die Konvergenz der Methode und zeigen, dass sie exponentiell konvergiert, wenn die charakteristische Funktion exponentiell abfällt. Um die COS Methode anzuwenden, müssen zwei Tuning Parameter festgelegt werden: ein Trunkierungsbereich für die Dichte und die Anzahl von Termen, um die Dichte durch eine Kosinusreihe zu approximieren. Für eine vordefinierte Fehlertoleranz leiten wir eine explizite Schranke für den Trunkierungsbereich und eine implizite Formel für die Anzahl der Terme her.

Die Berechnung der Tuning Parameter von Fourier-Preismethoden wie der COS-Methode ist für fortgeschrittene Modelle wie das Heston-Modell sehr rechenintensiv. Im vierten Abschnitt verwenden wir maschinelle Lernverfahren, um diese Tuning Parameter unabhängig von einer vordefinierten Fehlertoleranz vorherzusagen. Dieser Ansatz ermöglicht eine schnelle Bepreisung bei voller Fehlerkontrolle.

Abstract

(English)

This thesis consists of four essays. In the first chapter, we construct a class of time-continuous decompositions of path-independent instruments, such as the price of a stock option with the risk factors time decay or stock price. Compared to decompositions that evaluate the risk factors only at discrete time points, this class takes into account the whole information of the paths and is consistent across different time grids. We show that there is a unique decomposition by demanding the three properties exactness, symmetry and normalization. This unique decomposition is the stochastic limit of a time-discrete decomposition and it suffers from the curse of dimensionality. If the risk factors do not have simultaneous jumps, we show that the numerical effort can be significantly reduced from exponential to linear growth.

In the first essay, we show stochastic convergence of discrete to time-continuous decompositions. In practice, we can only observe one path of the risk factors. To ensure that the decompositions of this path converge, we show in the second essay, that there is a path-dependent, non-equidistant time grid that ensures almost sure convergence. Conditions that ensure a certain speed of convergence are also provided.

To approximate time-continuous decompositions, we use time-discrete decompositions with a fine grid. To decompose options, the corresponding prices need to be computed frequently. For a fast computation, in the third section we analyse the COS method, a Fourier approximation technique for pricing stock options, given that the characteristic function of the financial model exists. As the classical COS method is comparatively slow for certain options in general dimensions, we introduce the damped COS method. We prove the convergence of the method and show that it converges exponentially when the characteristic function decays exponentially. To apply the COS method, one must specify tuning parameters: a truncation range for the density and a number of terms to approximate the density by a cosine series. Given some error tolerance, we derive an explicit bound for the truncation range and an implicit formula for the number of terms. The computation of the tuning parameters of Fourier pricing methods such as the COS method is computationally expensive for advanced models such as the Heston model. In the fourth essay, we use machine learning techniques to predict these tuning parameters independently of a predefined error tolerance. This approach allows for fast pricing with full error control.

Contents

List of tables

List of figures

List of papers

Chapter 1 provides an introduction for this thesis, while Chapter 6 gives future research opportunities. Chapters 2, 4 and 5 each present a self-contained, submitted article, while Chapter 3 is an unpublished essay that extends the results of Chapter 2. This is a list of the corresponding articles and a detailed disclosure of the contributions of the author of this thesis:

Chapter 2: Junike, G., H. Stier and M. C. Christiansen (2024). "Profit and loss decomposition in continuous time and approximations", accepted in Finance and Stochastics. In: arXiv preprint arXiv:2212.06733.
Gero Junike and Hauke Stier contributed equally to the conception of the project, the derivation and interpretation of the main results, and the writing of the article. Marcus Christiansen provided advisory support and contributed significantly to interpreting the results.

Chapter 3: Unpublished essay by Hauke Stier.
Hauke Stier would like to thank Gero Junike and Marcus Christiansen for fruitful discussions.

Chapter 4: Junike, G. and H. Stier (2024). "From characteristic functions to multivariate distribution functions and European option prices by the damped COS method". In: arXiv preprint arXiv:2307.12843.
Both authors contributed equally to the conception of the project, the derivation and interpretation of the main results, and the writing of the article.

Chapter 5: Junike, G. and H. Stier (2024). "Enhancing Fourier pricing with machine learning". In: arXiv preprint arXiv:2412.05070.
Both authors contributed equally to the conception of the project, the derivation and interpretation of the main results, and the writing of the article.

There might be minor modifications in the chapter contents compared to the corresponding articles.

Chapter 1

Introduction

The analysis of profits and losses (P&L) between two reporting dates is a widespread task in risk management. Solvency 2, the regulation standard for insurance companies in the European Union, states that a profit and loss attribution must be performed in sufficient detail, when an internal model is used, see Article 240 in European Commission (2015). The total change in the P&L should be explained by the movements of the sources of risk or risk factors and should therefore be decomposed into the contributions of the risk factors. For example, consider a European investor who buys a put option on an American stock. The value of the investment is equal to the product of the option price Y in USD and the USDEUR exchange rate X. As a decomposition of the P&L at time 1 we look at two contributions δ_X and δ_Y of X and Y, such that

$$X(1)Y(1) - X(0)Y(0) = \delta_X + \delta_Y.$$

A widely used decomposition method is the sequential updating (SU) method, see Oaxaca (1973) and Blinder (1973). In a one-period setting, the SU decomposition is given by

$$\delta_X = X(1)Y(0) - X(0)Y(0), \quad \delta_Y = X(1)Y(1) - X(1)Y(0).$$

In this example, we update X first, followed by Y. Alternatively, Y could be updated first, so there are multiple SU decompositions, depending on a predefined update order. The SU decomposition can be generalized to a multi-period setting by applying the above calculations to multiple subintervals of the time horizon. The multi-period SU decomposition further depends on the choice of the time grid. To overcome the dependence of the update order, one can use the averaged SU (ASU) decomposition, which is the arithmetic mean of all SU decompositions with different update orders. To address the choice of the time grid, Jetses and Christiansen (2022) and Christiansen (2022) analyze the SU decomposition when the size of the time steps approaches zero. The corresponding

infinitesimal sequential updating (ISU) decomposition uses the whole information of the risk factors and is independent of the choice of the time grid. Jetses and Christiansen (2022) also define the infinitesimal averaged sequential updating (IASU) decomposition as the arithmetic mean of the different ISU decompositions.

In Chapter 2, we analyze the case where the P&L of path-independent instruments can be described by a twice differentiable function and introduce a class of time-continuous decompositions, based on stochastic integrals and Itô's formula, called Itô decompositions. The ISU decompositions and the IASU decomposition are included as special cases. We prove that there is a unique Itô decomposition that satisfies the three useful axioms of exactness, symmetry and normalization. We further show that it is indistinguishable from the IASU decomposition. We also analyze the SU decompositions with an increasingly fine time grid. As in Jetses and Christiansen (2022), we show that the SU/ASU decomposition converges in probability to the ISU/IASU decomposition, but we also allow risk factors with non-zero covariations. In practice, the IASU decomposition is difficult to compute, because for d risk factors there are $d!$ different SU decompositions. In this work, we show that the numerical effort can be significantly reduced from exponential to linear growth if the risk factors do not have simultaneous jumps. Since the ISU/IASU decompositions are defined by stochastic integrals, we further provide numerical experiments on the actual computation of the decompositions of call option with stochastic interest rates and a basket option with up to 30 risk factors, see Section 2.5.

In Chapter 3, we extend the results from Chapter 2, where we have shown that the SU decomposition converges in probability to the ISU decomposition if the time steps converge to zero. Unfortunately, we can only observe one path of the risk factors in practice. To ensure the convergence of the decompositions of this path, we use path-wise stochastic integration to show that one can choose a path-dependent, non-equidistant time grid, such that the SU decomposition converges to the ISU decomposition almost surely. We further analyze the order of convergence of the SU and ASU decompositions. We provide theoretical error bounds and perform numerical experiments.

Next, consider the example above of a European investor buying a put option on an American stock. Before applying the decomposition methods, we have to calculate the price of the put option, which is equal to

$$e^{-rT} \int_{\mathbb{R}} w(x)g(x)dx, \tag{1.0.1}$$

where g is the density of an appropriate financial model and w describes the payoff of the option. In many financial models the exact structure of the density g is unknown, but fortunately its Fourier transform $\hat{g}$ is given. There are various techniques to solve the

integral (1.0.1), for example Monte Carlo simulation, numerical quadrature and Fourier techniques based on $\hat{g}$, such as the COS method, see Fang and Oosterlee (2009a), the Carr-Madan formula, see Carr and Madan (1999) and the Lewis formula, see Eberlein et al. (2010).

Chapter 4 and 5 focus on the COS method, which can often compute prices faster than other methods. This is particularly important when performing a decomposition on a fine time grid to approximate time-continuous decompositions, as option prices need to be calculated frequently for this purpose. We introduce the main idea of the COS method for a one-dimensional underlying (the d-dimensional case is introduced in Chapter 4): Given the Fourier transform of the density of the log-returns of the underlying, the density is truncated on some finite interval $[-L, L]$ for $L > 0$. Then the truncated density is approximated by a finite Fourier-cosine series with N terms, i.e.

$$g(x) \approx g(x)1_{[-L,L]}(x) \approx \frac{c_0}{2} + \sum_{k=1}^{N} c_k \cos\left(k\pi \frac{x+L}{2L}\right),$$

where for $k = 0, \ldots, N$ the coefficients c_k are defined by

$$c_k = \frac{1}{L}\int_{\mathbb{R}} g(x) \cos\left(k\pi \frac{x+L}{2L}\right) dx = \frac{1}{L}\Re\left\{\hat{g}\left(\frac{k\pi}{2L}\right) e^{i\frac{k\pi}{2}}\right\}$$

with $\Re(z)$ denoting the real part of a complex number z. Now we can replace g in (1.0.1) by its approximation, which leads to

$$\int_{\mathbb{R}} w(x)g(x)dx \approx \frac{c_0 v_0}{2} + \sum_{k=1}^{N} c_k v_k$$

with

$$v_k = \int_{-L}^{L} w(x) \cos\left(k\pi \frac{x+L}{2L}\right) dx$$

for $k = 0, \ldots, N$. The COS method is particularly fast if the coefficients v_k can be obtained analytically.

To apply the COS method one has to specify two tuning parameters: the truncation range L and the number of terms N. In the one-dimensional case, Junike and Pankrashkin (2022) and Junike (2024) prove the convergence of the COS method and provide explicit bounds for both tuning parameters to ensure an error bound $\varepsilon > 0$ on the distance between the exact solution and the approximation by the COS method. The truncation range of a put option with strike K can be chosen by

$$L = \sqrt[n]{\frac{1}{i^n} \frac{\partial^n}{\partial u^n} \hat{g}(u)\Big|_{u=0}} \times \left(\frac{2Ke^{-rT}}{\varepsilon}\right)^{\frac{1}{n}}, \tag{1.0.2}$$

where $n \in \mathbb{N}$ is even. If g is also $s+1 \in \mathbb{N}$ times differentiable, then the number of terms can be chosen by

$$N = I \times \left(\frac{2^{s+\frac{5}{2}} L^{s+2}}{s\pi^{s+1}} \frac{12Ke^{-rT}}{\varepsilon} \right)^{\frac{1}{s}}, \tag{1.0.3}$$

where

$$I := \left(\frac{1}{2\pi} \int_{\mathbb{R}} |u|^{s+1} |\hat{g}(u)| du \right)^{\frac{1}{s}},$$

see Junike (2024) for details.

In Chapter 4, we analyze the COS method for a multidimensional underlying to price basket options. For many options, such as an arithmetic basket put option, w is not integrable and the coefficients v_k are not given analytically. In this case, the classical COS method is relatively slow because the coefficients v_k have to be calculated numerically. To price such options fast, we introduce the damped COS method. The idea of damping, going back at least to Carr and Madan (1999), is to multiply the payoff function w by an exponential function to make it integrable.
We prove the convergence of the (damped) COS method in the multidimensional case and, similar to Junike and Pankrashkin (2022), we derive an explicit formula for the truncation range. We also provide an implicit formula for the number of terms, which is a completely new approach compared to the one-dimensional case. We further analyze the order of convergence of the multidimensional COS method.

For the COS method, the formulas for the truncation range and the number of terms in Eqs. (1.0.2) and (1.0.3) depend on a higher derivative of the characteristic function and an integral, which are very time consuming to compute for advanced financial models. In Chapter 5, we propose a prediction of the derivative and the integral using machine learning methods to speed up the computation. Fortunately, the predicted values do not depend on the error tolerance. Therefore, we can use the predicted values and the error tolerance to compute the tuning parameters L and N. Then we use the COS method for a fast pricing of the options, resulting in a fast pricing approach with full error control.

In summary, we analyze time-continuous profit and loss decompositions and their approximations. We prove stochastic and almost sure convergence of time-discrete to time-continuous decompositions and provide conditions to reduce the numerical complexity. For a fast and exact computation of decompositions of stock options on a fine time grid, we analyze the COS method. We derive formulas for the tuning parameters of the COS method in the multidimensional case and use machine learning to speed up the computation of the

tuning parameters in one dimension. The structure of the thesis is visualized in Figure 1.1.

Chapter 6 presents ideas for future research. Chapters 2 to 5 each have an appropriate introduction with literature review and conclusion. The thesis is based on journal articles submitted during the doctoral studies, see the list of papers.

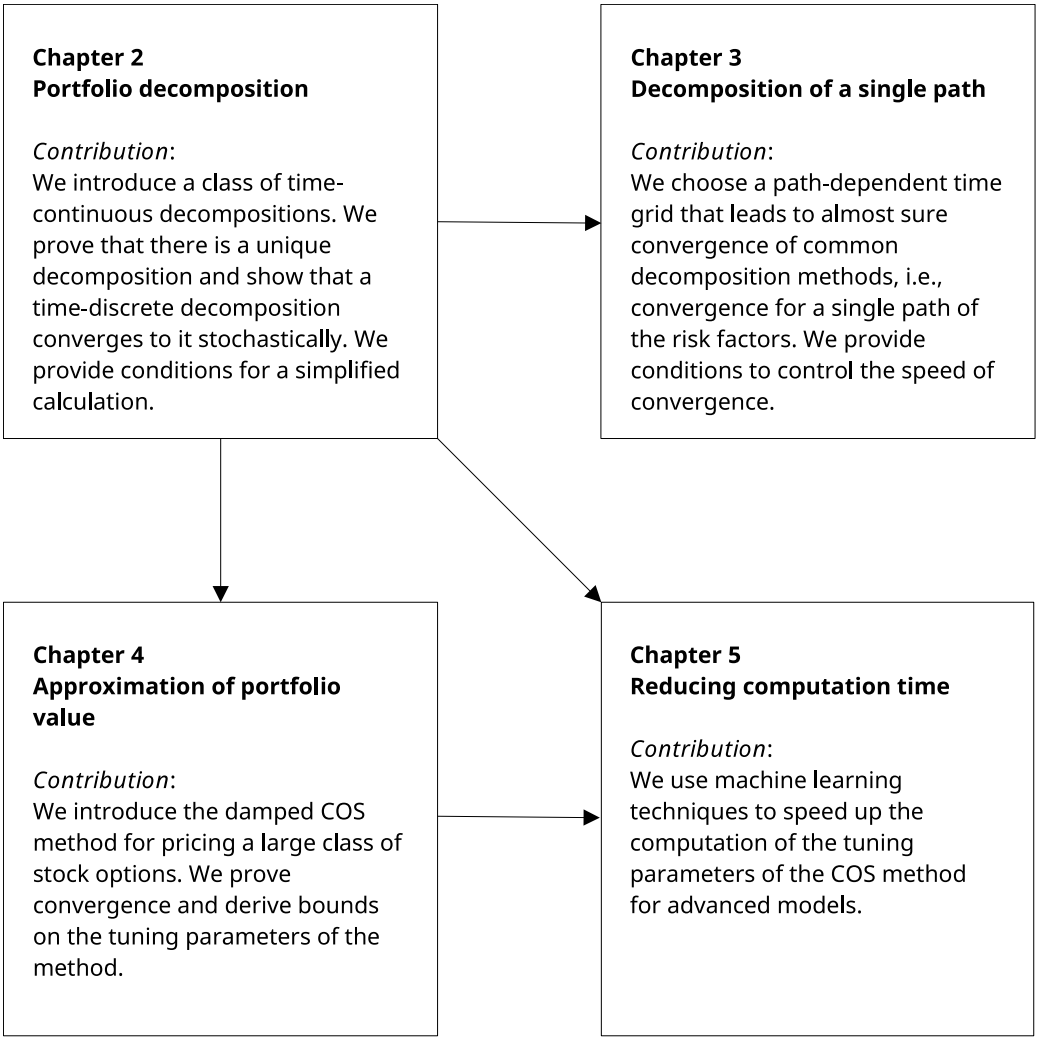

Figure 1.1: Structure of the thesis.

Chapter 2

Profit and loss decomposition in continuous time and approximations

Abstract

Financial institutions and insurance companies that analyze the evolution and sources of profits and losses often look at risk factors only at discrete reporting dates, ignoring the detailed paths. Continuous-time decompositions avoid this weakness and also make decompositions consistent across different reporting grids. We construct a large class of continuous-time decompositions from a new extended version of Itô's formula and uniquely identify a preferred decomposition from the axioms of exactness, symmetry and normalization. This unique decomposition turns out to be a stochastic limit of recursive Shapley values, but it suffers from a curse of dimensionality as the number of risk factors increases. We develop an approximation that breaks this curse when the risk factors almost surely have no simultaneous jumps.

Keywords: profit and loss attribution; sequential decompositions; change analysis; risk decomposition; Itô's formula

2.1 Introduction

Profit and loss (P&L) attribution, also known as change analysis, has a long history in risk management. P&L attribution is the process of analyzing the change between two valuation dates and explaining the evolution of the P&L by the movement of the sources (risk factors) between the two dates, see Candland and Latz (2014). In other words, the change in the P&L over time is *decomposed* in terms of the different risk factors to explain

how each factor *contributes* to the P&L. In the literature, there are many ways to obtain a P&L attribution. For example, consider a portfolio in EUR consisting of a long position in the S&P 500, Y for short. The P&L of such a portfolio is driven by two risk factors: Y and the USDEUR exchange rate, X for short. To decompose the P&L over one year, we look for two real numbers D_X and D_Y, such that

$$X(1)Y(1) - X(0)Y(0) \stackrel{!}{=} D_X + D_Y.$$

The numbers D_X and D_Y are interpreted as the contribution of X and Y to the P&L. In the literature we can find many desirable properties that a decomposition should possess, see Shubik (1962), Friedman and Moulin (1999) and Shorrocks (2013) among many others. The authors argue that a decomposition should be *symmetric*, i.e., the contributions of the risk factors should be independent of the way in which the risk factors are labeled or ordered. These authors also require that the sum of all contributions equals the P&L, such decompositions are called *exact*. Further, Christiansen (2022) argues that a decomposition should be *normalized*, i.e., if a risk factor remains constant, its contribution to the P&L should be zero. It is also desirable for a decomposition to consider the full path of each risk factor, i.e., to use all available information, see Mai (2023) and Flaig and Junike (2024).

Common decomposition principles

A common method for creating decompositions is to sequentially update the risk factors one by one while 'freezing' all other risk factors. This idea dates back at least to Oaxaca (1973) and Blinder (1973), who developed a *sequential updating* (SU) decomposition technique in a one-period setting. The SU decomposition works as follows when we update the risk factor X first:

$$D_X = X(1)Y(0) - X(0)Y(0), \quad D_Y = X(1)Y(1) - X(1)Y(0).$$

Alternatively, one may update Y first to obtain

$$D_X = X(1)Y(1) - X(0)Y(1), \quad D_Y = X(0)Y(1) - X(0)Y(0).$$

Each SU decomposition is exact, but if there are d risk factors, there are $d!$ different updating orders and therefore $d!$ different SU decompositions. Candland and Latz (2014) call the one-period SU decomposition *waterfall* and apply it to P&L attribution. See Fortin et al. (2011) for an overview on how the SU decomposition is used in various fields of economics.

The SU decomposition can also be defined in a multi-period setting by dividing the time horizon into sub-intervals and applying the SU decomposition recursively on each sub-interval. Jetses and Christiansen (2022) and Christiansen (2022) analyzed the limit of the SU decomposition when the mesh size of the time grid converges to zero. In the limit, the decomposition takes the whole path into account and the limiting SU decomposition is called the *infinitesimal sequential updating* (ISU) decomposition. The ISU decomposition is independent of any time grid, which is helpful "to prevent inconsistencies when using conflicting sub-intervals for different purposes", see p. 2 in Flaig and Junike (2024).

The *averaged sequential updating* (ASU) decomposition, also known as the *Shapley value*, is simply the arithmetic average of the $d!$ possible SU decompositions. It has many desirable properties, in particular: it is exact and symmetric. Shapley (1953) introduces the ASU decomposition for cooperative games. Shubik (1962) defines the ASU decomposition for cost functions. Sprumont (1998) and Friedman and Moulin (1999) provide an axiomatization of the ASU decomposition for cost functions. Jetses and Christiansen (2022) define the *infinitesimal averaged sequential updating* (IASU) decomposition as the average of the $d!$ possible ISU decompositions.

Main contributions

In this paper, we start directly in a time-continuous setting: If the portfolio is a twice differentiable function of the risk factors and the risk factors have continuous paths, Itô's formula provides a natural additive decomposition of the portfolio. Our main contributions are as follows: In order to treat risk factors with jumps, we prove an expanded version of Itô's formula and use it to define a large class of reasonable decompositions, which we call *Itô decompositions* and which include all $d!$ ISU and the IASU decompositions as special cases. We prove that there is a unique Itô decomposition (up to indistinguishability) that satisfies the three axioms of exactness, symmetry and normalization. We show that it is indistinguishable from the IASU decomposition. We further show that the IASU decomposition can be interpreted as the limiting case of the ASU decomposition: compared to Jetses and Christiansen (2022), who assume that the covariations between the risk factors are zero, we use much weaker assumptions to prove the convergence of the SU/ ASU decompositions to the ISU/ IASU decompositions.

In summary, we propose to use the IASU decomposition to obtain a P&L attribution because it considers the whole paths of the risk factors and satisfies the axioms of exactness, symmetry and normalization. However, in practical applications, the IASU decomposition

has two drawbacks: a) similar to the ASU decomposition, it suffers from the curse of dimensionality; b) the IASU decomposition is defined by stochastic integrals, which somehow must be approximated in practice. Naively approximating these integrals can lead to decompositions that are no longer exact. As another important contribution of this paper, we show that the IASU decomposition does not suffer from the curse of dimensionality if the risk factors do not have simultaneous jumps. In this case, the IASU decomposition is indistinguishable from the average of two (suitably selected) ISU decompositions. To avoid point b), we suggest approximating ISU/ IASU by SU/ ASU.

Up to now, most practitioners have applied an arbitrary SU decomposition in a one-period setting to obtain an annual P&L attribution, see Candland and Latz (2014). Working with real market data, Flaig and Junike (2024) empirically show that the SU decomposition depends significantly on the order or labeling of the risk factors, and that some SU decompositions may even ignore relevant risk factors, which may "lead to wrong trading and hedging decisions", see p. 2 in Flaig and Junike (2024).

Our theoretical analysis suggests using the average of only two SU decompositions[1] with a sufficiently fine time grid to obtain a P&L attribution, since such a decomposition is arbitrarily close to the IASU decomposition when the risk factors do not have simultaneous jumps. Thus, our analysis is highly relevant for practitioners: we recommend computing two SU decompositions instead of one and using a finer grid than just annual data to obtain a decomposition that is much closer to the IASU decomposition than a single SU decomposition. While the choice of the decomposition (the average of two SU decompositions) is theoretically justified, we have only numerical experiments available to estimate the time grid and we recommend using monthly or weekly data.

Further literature review

Is there any other way to break the curse of dimensionality? Christiansen (2022) proves that the ISU decomposition is symmetric if it is stable with respect to small perturbations in the empirical observation of the risk factors. In Appendix 2.A.4, we show that the ISU decomposition of a simple product of two correlated Brownian motions is not stable. This shows that stability is a rather strong assumption.

There are other decomposition principles as well: There is the so-called *one-at-a-time* (OAT) decomposition, which is also known as *bump and reset*, see Candland and Latz (2014). The OAT decomposition is closely related to the SU decomposition. It is symmetric, but

[1] Define one SU decomposition in any order, e.g., alphabetically ascending, and the other SU decomposition by the reverse order, e.g., alphabetically descending, see Theorem 2.3.10 for details.

generally not exact. Frei (2020) analyzes the limit of the OAT decomposition when the mesh size of the time grid converges to zero.

There are also completely different approaches: Fischer (2004) uses a conditional expectations approach. Rosen and Saunders (2010) use the Hoeffding method for a decomposition of credit risk portfolios. Frei (2020) uses the Euler principle for risk attribution. Ramlau-Hansen (1991) and Norberg (1999) decompose surplus in life insurance by heuristic integral representations, where the integrators are interpreted as the driving forces of change and determine the attribution. A similar idea is used in Schilling et al. (2020) based on the martingale representation theorem.

Contents

In Section 2.2, we establish some notation. In Section 2.3, we develop an extended version of Itô's formula and introduce the family of Itô decompositions. We show that the IASU decomposition is the only exact and symmetric Itô decomposition, and we break the curse of dimensionality of the IASU decomposition in Theorem 2.3.10. In Section 2.4, we prove that the IASU decomposition can be approximated by the ASU decomposition. In Section 2.5, we provide some numerical applications. Section 2.6 concludes.

2.2 Notation

Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions, i.e., $\mathcal{F}_0$ contains all null sets and $\mathbb{F}$ is right-continuous. Let $\mathcal{X}$ be the set of all $\mathbb{F}$-semimartingales. A so-called *risk basis* or *information basis* is a d-dimensional semimartingale $X \in \mathcal{X}^d$, and its d components are denoted as *risk factors* or *sources of risk*. For a stopping time s, we define the stopped semimartingale by $X^s = (X_1^s, ..., X_d^s)$. Equality of random variables is understood in the almost sure sense and equality of stochastic processes is understood up to indistinguishability. Let $\mathcal{C}_2$ be the set of twice differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}$. For $f \in \mathcal{C}_2$ and $i, j = 1, ..., d$, we write f_i and f_{ij} for the partial derivatives $\partial_i f$ and $\partial_i \partial_j f$. We call a map $F : \mathcal{X}^d \to \mathcal{X}$ *non-anticipative* if for any stopping time s it holds that

$$F\left(X^s\right)(t) = F(X)(\min(t, s)), \quad t \geq 0. \tag{2.2.1}$$

Such a non-anticipative mapping depends only on the information up to time t, i.e., on X^t. By $\mathcal{M}$ we denote some sub-space of all non-anticipative mappings. By $\mathcal{M}(\mathcal{C}_2)$ we denote the space of functionals $F : \mathcal{X}^d \to \mathcal{X}$ such that $F(X) = (f(X(t)))_{t\geq 0}$, $X \in \mathcal{X}^d$,

for some $f \in \mathcal{C}_2$, which are clearly non-anticipative. By σ_d we denote the set of all $d!$ permutations of $\{1, ..., d\}$. Let $id \in \sigma_d$ be the identity. In a slight abuse of notation, we define for $\pi \in \sigma_d$

$$\pi(x_1, ..., x_d) = \left(x_{\pi(1)}, ..., x_{\pi(d)}\right), \quad x \in \mathbb{R}^d,$$

and

$$\pi(X_1, ..., X_d) = \left(X_{\pi(1)}, ..., X_{\pi(d)}\right), \quad X \in \mathcal{X}^d.$$

For two one-dimensional semimartingales Z and Y and a càglàd process H, we denote by $\int_0^t H(s)dZ(s) := \int_{0+}^t H(s)dZ(s)$ the stochastic integral. In particular $\int_0^0 H(s)dZ(s) = 0$ by convention. We further set $Z(0-) = 0$,

$$Z(t-) = \lim_{\varepsilon \searrow 0} Z(t-\varepsilon), \quad t > 0, \quad \Delta Z(t) = Z(t) - Z(t-), \quad t \geq 0,$$

$$[Z, Y] = ZY - Z(0)Y(0) - \int_0^{\cdot} Z(u-)dY - \int_0^{\cdot} Y(u-)dZ$$

and

$$[Z, Y]^c = [Z, Y] - \sum_{0<s\leq \cdot} \Delta Z(s)\Delta Y(s).$$

By "$\xrightarrow{p}$" we denote the convergence in probability of a sequence of random variables. For $A \subset \{1, ..., d\}$ we define the projection

$$\begin{aligned} p_A : \mathbb{R}^d &\to \mathbb{R}^d \\ x &\mapsto (x_1 1_A(1), ..., x_d 1_A(d)), \end{aligned}$$

where the function $1_A(h)$ is one if $h \in A$ and zero otherwise.

2.3 Family of Itô decompositions

Similar to Shorrocks (2013) and Christiansen (2022), we define a *decomposition* as follows:

Definition 2.3.1.
A map

$$\begin{aligned} \delta : \mathcal{M} \times \mathcal{X}^d &\to \mathcal{X}^d \\ (F, X) &\mapsto (\delta_1(F, X), ..., \delta_d(F, X)) \end{aligned}$$

is called a *decomposition*.

We interpret $\delta_i(F,X)(t)$ as the contribution of X_i to the profit and loss $F(X)(t)-F(X)(0)$ in $(0,t]$. We recall the following three axioms from the literature:

i) A decomposition is called *exact* if for all $F \in \mathcal{M}$ and all $X \in \mathcal{X}^d$ the following equation holds:

$$F(X) - F(X)(0) = \delta_1(F,X) + ... + \delta_d(F,X).$$

ii) A decomposition is called *symmetric* if for all $\pi \in \sigma_d$, all $F \in \mathcal{M}$ and all $X \in \mathcal{X}^d$ the following implication holds:

$$F(X) = F(\pi(X)) \quad \Longrightarrow \quad \delta_i(F,X) = \delta_{\pi^{-1}(i)}(F,\pi(X)).$$

iii) A decomposition is called *normalized* if for all $0 \leq r < s < \infty$, all $i = 1,...,d$, all $F \in \mathcal{M}$ and all $X \in \mathcal{X}^d$ the following implication holds:

$$\begin{aligned} &X_i \text{ is indistinguishable from a constant process on } (r,s] \\ &\Longrightarrow \delta_i(F,X) \text{ is indistinguishable from a constant process on } (r,s]. \end{aligned}$$

Axiom i) ensures that a decomposition is able to fully explain the P&L, see Shorrocks (2013) and Christiansen (2022). Axiom ii) considers symmetric maps F and states that if F does not depend on the order or labeling of the risk factors, then the decomposition shall also be independent of the order or labeling of the risk factors. The symmetry axiom is motivated by the fact that $\delta_i(F,X)$ represents the contribution of X_i and that the term $\delta_{\pi^{-1}(i)}(F,\pi(X))$ also describes the contribution of

$$\pi(X)_{\pi^{-1}(i)} = (X_{\pi(1)},...,X_{\pi(d)})_{\pi^{-1}(i)} = X_i.$$

The symmetry axiom has already been mentioned in similar form in Friedman and Moulin (1999) and Shorrocks (2013). If the risk factor X_i remains constant during some time interval $(r,s]$, it does not contribute to $F(X)(s)-F(X)(r)$, so the contribution of X_i in $(r,s]$ should also be zero. This is exactly reflected by the normalization axiom, taken from Christiansen (2022).

Next, we indicate how Itô's formula helps to define decomposition principles: Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice continuously differentiable. For $i,j = 1,...,d$ let

$$I_i := \int_0^{\cdot} f_i(X(s-))dX_i(s), \quad I_{ij} := \int_0^{\cdot} f_{ij}(X(s-))d[X_i,X_j]^c(s), \tag{2.3.1}$$

$$S := \sum_{0<s\leq \cdot} \Big\{ f(X(s)) - f(X(s-)) - \sum_{i=1}^{d} f_i(X(s-))\Delta X_i(s) \Big\}. \tag{2.3.2}$$

Itô's formula states that for $t \geq 0$ it holds for any semimartingale X that

$$f(X(t)) - f(X(0)) = \sum_{i=1}^{d} I_i(t) + \frac{1}{2}\sum_{i=1}^{d} I_{ii}(t) + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{d} I_{ij}(t) + S(t). \tag{2.3.3}$$

Assume that X has continuous paths without interaction effects, i.e., $S = 0$ and $I_{ij} = 0$, $i \neq j$. Then, Eq. (2.3.3) provides a natural way to additively decompose the P&L $f(X(t)) - f(X(0))$:

By the normalization axiom, I_i and I_{ii} should be assigned to δ_i, which is interpreted as the contribution of X_i. To see this, assume that some δ_j would depend on I_i or I_{ii} for $i \neq j$. Assume that X_j is constant everywhere. According to the normalization axiom, we would then have $\delta_j = 0$. So, δ_j must not depend on I_i or on I_{ii}.

However, how to handle the interaction effects I_{ij}, $i \neq j$ and the jump part S is not so obvious. Therefore, in Proposition 2.3.3 we provide an extended version of Itô's formula. Based on Proposition 2.3.3, we define the large family of *Itô decompositions* in Definition 2.3.4 and show in Section 2.3 that the family of Itô decompositions contains many well-known decomposition principles as special cases. Within the family of Itô decompositions, we will identify a single decomposition that satisfies the axioms of exactness, symmetry and normalization. For $A \subseteq \{1, \ldots, d\}$, $i \in \{1, ..., d\}$ and $s > 0$ define

$$Y_i^A(s) := f\Big(X(s-) + p_A(\Delta X(s))\Big) - f\Big(X(s-) + p_{A\setminus\{i\}}(\Delta X(s))\Big) - f_i(X(s-))\Delta X_i(s)$$

and

$$S_i^A(X) := \sum_{0<s\leq\cdot} Y_i^A(s).$$

For $\pi \in \sigma_d$ define

$$S_i^\pi(X) := S_i^{\{j\,|\pi(j)\leq\pi(i)\}}(X). \tag{2.3.4}$$

To obtain $S_i^\pi(X)$, all time points s where X_i jumps are considered. All risk factors except X_i are fixed at s or $s-$, depending on the choice of π, and only X_i varies between $s-$ and s.

Lemma 2.3.2.
Let $i \in \{1, ..., d\}$, $X \in \mathcal{X}^d$ and $A \subset \{1, ..., d\}$. If $i \in A$, then $S_i^A(X)$ is a semimartingale with a.s. paths of finite variation on compacts.

Proof. Fix $X \in \mathcal{X}^d$. Let N be a null set such that $u \mapsto |X_i(u)(\omega)|$, $i = 1, ..., d$, is càdlàg for $\omega \in \Omega \setminus N$ and

$$\sum_{h,j=1}^{d} \sum_{0<s\leq t} |\Delta X_h(s)(\omega)\Delta X_j(s)(\omega)| < \infty, \quad \omega \in \Omega \setminus N, \quad t \geq 0. \tag{2.3.5}$$

Such N exists since X is a semimartingale. Let $\omega \in \Omega \setminus N$ and $t \geq 0$. Let $\mathcal{M}_\omega \subset \mathbb{R}^d$ be the closure of the set $\{X(u)(\omega),\, u \in [0,t]\}$, which is compact. The function f and its derivatives are continuous and reach a maximum and minimum on the convex hull of $\mathcal{M}_\omega$, which is compact by Carathéodory's theorem, see Section 2.3 in Grünbaum (2013). Hence, f and its derivatives are bounded on the convex hull of $\mathcal{M}_\omega$. Let $s \in (0,t]$. Let us develop f around $X(s-)(\omega)$ using a Taylor expansion. We have that

$$f\Big(X(s-)(\omega) + p_A(\Delta X(s)(\omega))\Big) = f(X(s-)(\omega)) + \sum_{h\in A} f_h(X(s-)(\omega))\Delta X_h(s)(\omega) + R(\omega),$$

where $R(\omega)$ is the remainder of the Taylor expansion, i.e., for some $\theta(\omega) \in [0,1]$ it holds that

$$R(\omega) = \frac{1}{2} \sum_{h,j\in A} f_{hj}\Big(X(s-)(\omega) + \theta(\omega)p_A(\Delta X(s)(\omega))\Big)\Delta X_h(s)(\omega)\Delta X_j(s)(\omega).$$

The term $f\left(X(s-)(\omega) + p_{A\setminus\{i\}}(\Delta X(s)(\omega))\right)$ can be treated similarly. Since $i \in A$, it holds for some $C(\omega) > 0$, which does not depend on s or $\theta(\omega)$, that

$$Y_i^A(s) \leq C(\omega) \sum_{h,j\in A} |\Delta X_h(s)(\omega)\Delta X_j(s)(\omega)|.$$

It follows by Inequality (2.3.5) that

$$\sum_{0<s\leq t} |Y_i^A(s)(\omega)| < \infty, \quad \omega \in \Omega \setminus N. \tag{2.3.6}$$

Since t was arbitrarily chosen, Inequality (2.3.6) implies that $u \mapsto S_i^A(X)(u)(\omega)$, $\omega \in \Omega \setminus N$, is càdlàg and of finite variation on compacts. Therefore, $S_i^A(X)$ is a semimartingale. □

Proposition 2.3.3.
Let $\pi \in \sigma_d$ and $f \in \mathcal{C}_2$ and $X \in \mathcal{X}^d$. For all $t \geq 0$ it holds that

$$f(X(t)) - f(X(0)) = \sum_{i=1}^{d} \Big\{ I_i(t) + \frac{1}{2} I_{ii}(t) + \frac{1}{2} \sum_{\substack{j=1 \\ i\neq j}}^{d} I_{ij}(t) + S_i^\pi(t) \Big\},$$

where I_i and I_{ij} are defined in Eq. (2.3.1) and S_i^π is defined in Eq. (2.3.4).

Proof. Since the series telescopes, we have that

$$\begin{aligned} &f(X(s)) - f(X(s-)) \\ &= \sum_{i=1}^{d} f\Big(X(s-) + p_{\{j\,|\pi(j)\leq\pi(i)\}}(\Delta X(s))\Big) - f\Big(X(s-) + p_{\{j\,|\pi(j)<\pi(i)\}}(\Delta X(s))\Big). \end{aligned}$$

By Inequality (2.3.6), it holds for any $t \geq 0$ that

$$\sum_{i=1}^{d} S_i^{\pi}(X)(t) = \sum_{0<s\leq t} \sum_{i=1}^{d} Y_i^{\{j\,|\pi(j)\leq\pi(i)\}}(s) = S(t), \tag{2.3.7}$$

where S is defined in Eq. (2.3.2). The claim follows then by the classical Itô's formula. □

Proposition 2.3.3 generalizes the classical Itô's formula, because for any $\pi \in \sigma_d$ it holds that $\sum_{i=1}^{d} S_i^{\pi}(X) = S$, see Eq. (2.3.7).

Definition 2.3.4.
Let $\lambda_{ij} \in [0,1]$ for $i,j = 1,...,d$. Let $\mu_\pi \in [0,1]$ for $\pi \in \sigma_d$. The decomposition

$$\delta^{\text{Itô}} : \mathcal{M}(\mathcal{C}_2) \times \mathcal{X}^d \to \mathcal{X}^d, \quad (F,X) \mapsto (\delta_1^{\text{Itô}}(F,X), ..., \delta_d^{\text{Itô}}(F,X)),$$

where

$$\delta_i^{\text{Itô}}(F,X) = I_i + \frac{1}{2} I_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{d} \lambda_{ij} I_{ij} + \sum_{\pi\in\sigma_d} \mu_\pi S_i^{\pi}(X), \quad i = 1,...,d,$$

is called *Itô decomposition with parameters* $(\lambda_{ij})_{i,j=1,...,d}$ *and* $(\mu_\pi)_{\pi\in\sigma_d}$.

The definition of the Itô decomposition is motivated by Proposition 2.3.3 and the normalization axiom: Below Eq. (2.3.3) we already argued that I_i and I_{ii} should be attributed to X_i in order to satisfy the normalization axiom. If parts of the interaction effect I_{ij} were assigned to the contribution of X_h for $h \notin \{i,j\}$, the decomposition would no longer be normalized. Therefore, only the risk factors X_i and X_j are assigned shares λ_{ij} and λ_{ji} of the interaction effect I_{ij}.

Note that $S_i^{\pi}(X)$ contains only jumps in the i-th component. If $S_i^{\pi}(X)$ were assigned to the contribution of some X_j, $j \neq i$, the normalization axiom would be violated if X_j is constant. Therefore, S_i^{π} should be assigned to the contribution of X_i. Since there are $d!$ different ways to decompose the jumps without violating neither the normalization axiom nor the exactness axiom, we propose to assign to X_i a weighted average of all $S_i^{\pi}(X)$, $\pi \in \sigma_d$.

Remark 2.3.5.
Since each Itô decomposition is linear in the first argument F, a portfolio can be decomposed by decomposing each individual instrument.

We recall some special members of the family of Itô decompositions, namely the IASU and the $d!$ different ISU decompositions, which were introduced in Jetses and Christiansen (2022). All Itô decompositions are normalized. We will prove that the IASU decomposition is the only Itô decomposition that is exact and symmetric. We will also see that the ISU decomposition is closely related to the IASU decomposition and that the IASU decomposition is the limiting case of the well-known ASU decomposition (also known as Shapley value), which is defined over a discrete time grid in Section 2.4.

Definition 2.3.6.
The *IASU (infinitesimal averaged updating) decomposition* $\delta^{\mathrm{IASU}} : \mathcal{M}(\mathcal{C}_2) \times \mathcal{X}^d \to \mathcal{X}^d$ is defined by

$$\delta_i^{\mathrm{IASU}}(F, X) = I_i + \frac{1}{2}\sum_{j=1}^{d} I_{ij} + \frac{1}{d!}\sum_{\pi\in\sigma_d} S_i^{\pi}(X), \quad i = 1, ..., d.$$

Remark 2.3.7.
The Itô decompositions are overparameterised: based on Eq. (2.A.3) in Lemma 2.A.2, we can represent the IASU decomposition as

$$\delta_i^{IASU}(F, X) = I_i + \frac{1}{2}\sum_{j=1}^{d} I_{ij} + \sum_{\substack{A\subseteq\{1,...,d\}\\ i\in A}} S_i^A(X)\xi_{i,A},$$

where

$$\xi_{i,A} := \sum_{\substack{\pi\in\sigma_d\\ \{j|\pi(j)\leq\pi(i)\}=A}} \frac{1}{d!} = \frac{(|A|-1)!(d-|A|)!}{d!}. \tag{2.3.8}$$

Hence, the computational cost to obtain δ_i^{IASU} can be reduced from $\mathcal{O}(d!)$ to $\mathcal{O}(2^{d-1})$ for $d \to \infty$.

Definition 2.3.8.
Let $\pi \in \sigma_d$. The *ISU (infinitesimal updating) decomposition* $\delta^{ISU,\pi} : \mathcal{M}(\mathcal{C}_2) \times \mathcal{X}^d \to \mathcal{X}^d$ *with updating order* π is defined by

$$\delta_i^{\mathrm{ISU},\pi}(F, X) = I_i + \frac{1}{2}I_{ii} + \sum_{\substack{j=1\\ \pi(j)<\pi(i)}}^{d} I_{ij} + S_i^{\pi}(X), \quad i = 1, ..., d.$$

Theorem 2.3.9.
Every Itô decomposition that is symmetric and exact is indistinguishable from the IASU decomposition. The IASU decomposition is related to the ISU decomposition by

$$\delta_i^{\mathrm{IASU}}(F,X) = \frac{1}{d!}\sum_{\pi\in\sigma_d}\delta_i^{\mathrm{ISU},\pi}(F,X), \quad i=1,...,d, \quad X\in\mathcal{X}^d, \quad F\in\mathcal{M}(\mathcal{C}_2). \tag{2.3.9}$$

Proof. The proof of Theorem 2.3.9 can be found in Appendix 2.A.2. □

The next theorem shows that the curse of dimensionality of the IASU decomposition can be broken if there are no simultaneous jumps.

Theorem 2.3.10.
Let $X\in\mathcal{X}^d$ and $F\in\mathcal{M}(\mathcal{C}_2)$. If $\Delta X_h\Delta X_j = 0$ for all $h,j\in\{1,\dots,d\}$ with $h\neq j$, then

$$\delta_i^{\mathrm{IASU}}(F,X) = \frac{1}{2}\big(\delta_i^{\mathrm{ISU},\pi}(F,X) + \delta_i^{\mathrm{ISU},\pi'}(F,X)\big), \quad i=1,...,d, \tag{2.3.10}$$

for any $\pi\in\sigma_d$ and $\pi' = d+1-\pi$.

Proof. Let $0<s<\infty$. In the case of $\Delta X_i(s)=0$, we have that

$$\begin{aligned}
&f\left(X(s-)+p_{\{j\,|\pi(j)\leq\pi(i)\}}(\Delta X(s))\right) - f\left(X(s-)+p_{\{j\,|\pi(j)<\pi(i)\}}(\Delta X(s))\right)\\
&= f\left(X(s-)+p_{\{j\,|\pi(j)<\pi(i)\}}(\Delta X(s))\right) - f\left(X(s-)+p_{\{j\,|\pi(j)<\pi(i)\}}(\Delta X(s))\right)\\
&= 0.
\end{aligned}$$

In the case of $\Delta X_i(s)\neq 0$, it holds that $X_j(s) = X_j(s-)$ for all $j\neq i$, and hence,

$$\begin{aligned}
&f\left(X(s-)+p_{\{j\,|\pi(j)\leq\pi(i)\}}(\Delta X(s))\right) - f\left(X(s-)+p_{\{j\,|\pi(j)<\pi(i)\}}(\Delta X(s))\right)\\
&= f\left(X(s)\right) - f\left(X(s-)\right).
\end{aligned}$$

Hence, for $\pi\in\sigma_d$ and $i=1,...,d$ it holds that

$$\delta_i^{\mathrm{ISU},\pi} = I_i + \frac{1}{2}I_{ii} + \sum_{\substack{j=1\\ \pi(j)<\pi(i)}}^{d} I_{ij} + \sum_{\substack{0<s\leq\cdot\\ \Delta X_i(s)\neq 0}} \{f\left(X(s)\right) - f\left(X(s-)\right) - f_i(X(s-))\Delta X_i(s)\}. \tag{2.3.11}$$

Due to Eqs. (2.3.9, 2.A.10), we have that

$$\delta_i^{\mathrm{IASU}}(F,X) = I_i + \frac{1}{2}\sum_{j=1}^{d} I_{ij} + \sum_{\substack{0<s\leq\cdot\\ \Delta X_i(s)\neq 0}} \{f\left(X(s)\right) - f\left(X(s-)\right) - f_i(X(s-))\Delta X_i(s)\}. \tag{2.3.12}$$

Let $\delta_i^{\mathrm{ISU},\pi}$ be the ISU decomposition with updating order $\pi \in \sigma_d$ and define $\pi'(i) = d + 1 - \pi(i)$, $i = 1, ..., d$. Note that

$$\sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} + \sum_{\substack{j=1 \\ \pi\prime(j)<\pi\prime(i)}}^{d} = \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} + \sum_{\substack{j=1 \\ \pi(j)>\pi(i)}}^{d} = \sum_{\substack{j=1 \\ i\neq j}}^{d}. \tag{2.3.13}$$

Eqs. (2.3.11, 2.3.12, 2.3.13) imply Eq. (2.3.10). □

Remark 2.3.11.
Theorem 2.3.10 can be generalized to the case where some but not all risk factors have simultaneous jumps. For example, suppose $d = 3$ and $\Delta X_1 \Delta X_j = 0$, $j \in \{2, 3\}$ but possibly $\Delta X_2 \Delta X_3 \neq 0$. It is then easy to see that Eq. (2.3.10) still holds. Or, if $d = 4$ and $\Delta X_1 \Delta X_j = 0$, $j \in \{2, 3, 4\}$, the IASU decomposition can be written as a weighted average of four ISU decompositions instead of eight ISU decompositions, which would be necessary if all risk factors had simultaneous jumps.

Corollary 2.3.12.
Let $X \in \mathcal{X}^d$ and $F \in \mathcal{M}(\mathcal{C}_2)$. If $[X_h, X_j] = 0$ for all $h, j \in \{1, \ldots, d\}$ with $h \neq j$, then

$$\delta_i^{\mathrm{IASU}}(F, X) = \delta_i^{\mathrm{ISU},\pi}(F, X), \quad i = 1, ..., d,$$

where $\pi \in \sigma_d$ is arbitrary.

Proof. The assumption $[X_i, X_j] = 0$ for $i \neq j$ implies $\Delta X_i \Delta X_j = \Delta[X_i, X_j] = 0$. Therefore, $S_i^{\pi_1} = S_i^{\pi_2}$, $\pi_1, \pi_2 \in \sigma_d$, see the proof of Theorem 2.3.10. The assertion follows directly from the Definitions 2.3.6 and 2.3.8. □

Example 2.3.13.
How does the IASU decomposition treat simultaneous jumps? Let $d = 2$ and assume that $X = (X_1, X_2)$ is a pure-jump semimartingale of finite variation. Then the IASU decomposition is given by

$$\delta_1^{\mathrm{IASU}}(F, X) = \frac{1}{2} \sum_{0<s\leq \cdot} \Big\{ \big\{ f\big(X(s), X_2(s-)\big) - f\big(X(s-)\big) \big\} + \big\{ f\big(X(s)\big) - f\big(X_1(s-), X_2(s)\big) \big\} \Big\},$$

$$\delta_2^{\mathrm{IASU}}(F, X) = \frac{1}{2} \sum_{0<s\leq \cdot} \Big\{ \big\{ f\big(X(s)\big) - f\big(X_1(s), X_2(s-)\big) \big\} + \big\{ f\big(X_1(s-), X_2(s)\big) - f\big(X(s-)\big) \big\} \Big\}.$$

The latter formulas are averages of sequential updates from time point $s-$ to time point s.

Example 2.3.14.
We decompose the P&L of the portfolio $P = X_1X_2$ of a foreign stock, where X_1 is the foreign exchange rate and X_2 is the stock price in the foreign currency. The instantaneous P&L of the foreign stock in domestic currency is given by

$$dP(t) = X_1(t-)dX_2(t) + X_2(t-)dX_1(t) + d[X_1, X_2](t),$$

i.e., it can be decomposed into the variation of the exchange rate, variation of the stock price and interaction effects, compare with Mai (2023). The IASU decomposition equally distributes the interaction effect between δ_1^{IASU} and δ_2^{IASU}. To see this, observe that

$$\begin{aligned}\delta_1^{\text{IASU}}(F, X) = &\int_0^{\cdot} X_2(s-)dX_1(s) + \frac{1}{2}[X_1, X_2]^c \\ &+ \frac{1}{2}\sum_{0<s\leq\cdot}\Big\{\{X_1(s)X_2(s-) - X_1(s-)X_2(s-)\} \\ &+ \{X_1(s)X_2(s) - X_1(s-)X_2(s)\} - 2X_2(s-)(X_1(s) - X_1(s-))\Big\} \\ = &\int_0^{\cdot} X_2(s-)dX_1(s) + \frac{1}{2}[X_1, X_2],\end{aligned}$$

where $F(X) = X_1X_2$. For δ_2^{IASU} the reasoning is similar.

2.4 SU and ASU decompositions and their limits

The time-dynamic SU and ASU decompositions are defined on discrete time grids, see Jetses and Christiansen (2022) and Christiansen (2022). A light introduction to the SU decomposition can be found in Candland and Latz (2014). In this section, we recall the definitions of these decompositions and we provide sufficient conditions such that the SU and the ASU decompositions converge to the ISU and IASU decompositions, respectively, as the mesh size of the discrete time grid converges to zero. We recall the following definition from p. 64 in Protter (2005).

Definition 2.4.1.
An infinite sequence of finite stopping times $0 = s_0 < s_1 < s_2 < ...$ such that $\sup_k s_k = \infty$ a.s. is called an *unbounded random partition*. A sequence $(\gamma_n)_{n\in\mathbb{N}}$ of unbounded random partitions $\gamma_n = \{0 = s_0^n < s_1^n < s_2^n < ...\}$ is said to *tend to the identity* if $\sup_k |s_{k+1}^n - s_k^n| \to 0$ a.s. for $n \to \infty$.

Definition 2.4.2.
Let $\gamma = \{0 = s_0 < s_1 < \dots\}$ be an unbounded random partition. The *SU (sequential updating) decomposition* $\delta^{SU,\pi,\gamma} : \mathcal{M} \times \mathcal{X}^d \to \mathcal{X}^d$ *with updating order* $\pi \in \sigma_d$ is defined by

$$\delta_i^{\mathrm{SU},\pi,\gamma}(F,X) = \sum_{l=0}^{\infty} \Big\{ F\left(X^{s_l} + p_{\{j\,|\pi(j)\le\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right) - F\left(X^{s_l} + p_{\{j\,|\pi(j)<\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right) \Big\}. \tag{2.4.1}$$

In words, divide the time horizon $[0,t]$ into finitely many sub-intervals, and to obtain the contribution of X_i, fix all risk factors at the beginning s_l or the end s_{l+1} of each sub-interval (depending on the updating order π) and allow only X_i to vary between s_l and s_{l+1}.

Proposition 2.4.3.
The decomposition $\delta^{\mathrm{SU},\pi,\gamma} : \mathcal{M} \times \mathcal{X}^d \to \mathcal{X}^d$ is well defined by Eq. (2.4.1) and exact. The sum in Eq. (2.4.1) evaluated at $t \in [0,\infty)$ is a.s. finite.

Proof. Let $X \in \mathcal{X}^d$, $F \in \mathcal{M}$, $\pi \in \sigma_d$, $n \in \mathbb{N}$, and $t \ge 0$. By $x \wedge y$ we denote the minimum of two real numbers x and y. Using Eq. (2.2.1) twice, we get

$$\begin{aligned}\delta_i^{\mathrm{SU},\pi,\gamma}(F,X)(t\wedge s_n) &= \sum_{l=0}^{\infty} \Big\{ F\left(X^{s_l\wedge s_n} + p_{\{j\,|\pi(j)\le\pi(i)\}}\left(X^{s_{l+1}\wedge s_n} - X^{s_l\wedge s_n}\right)\right)(t) \\ &\qquad - F\left(X^{s_l\wedge s_n} + p_{\{j\,|\pi(j)<\pi(i)\}}\left(X^{s_{l+1}\wedge s_n} - X^{s_l\wedge s_n}\right)\right)(t) \Big\} \qquad (2.4.2)\\ &= \sum_{l=0}^{n-1} \Big\{ F\left(X^{s_l} + p_{\{j\,|\pi(j)\le\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right)(t\wedge s_n) \\ &\qquad - F\left(X^{s_l} + p_{\{j\,|\pi(j)<\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right)(t\wedge s_n) \Big\} \qquad (2.4.3)\end{aligned}$$

since all addends with $l \ge n$ on the right hand-side of Eq. (2.4.2) are equal to zero. By Eq. (2.4.3), for each n, the process $\delta_i^{\mathrm{SU},\pi,\gamma}(F,X)$ stopped at s_n is a finite sum of semimartingales and hence a semimartingale. By Section 2 of Part II in Protter (2005) and since $s_n \to \infty$ a.s. for $n \to \infty$, the process $\delta_i^{\mathrm{SU},\pi,\gamma}(F,X)$ is a semimartingale and the decomposition $\delta^{\mathrm{SU},\pi,\gamma}$ is therefore well defined. The fact that $s_n \to \infty$ a.s. implies that the sum in Eq. (2.4.1) evaluated at t is a.s. finite. We show exactness: Let $x \in \mathbb{R}^d$. Since

$$p_{\{j\,|\pi(j)\le\pi(\pi^{-1}(d))\}}(x) = x \quad \text{and} \quad p_{\{j\,|\pi(j)<\pi(\pi^{-1}(1))\}}(x) = 0,$$

we have for any $t \in [0,\infty)$ and $n \in \mathbb{N}$ by Eq. (2.4.3) that

$$\begin{aligned}
&\sum_{i=1}^{d} \delta_i^{\mathrm{SU},\pi,\gamma}(F,X)(t \wedge s_n) \\
&= \sum_{\substack{i=1 \\ i \neq \pi^{-1}(d)}}^{d} \sum_{l=0}^{n-1} F\left(X^{s_l} + p_{\{j\,|\pi(j)\leq\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right)(t \wedge s_n) + \sum_{l=0}^{n-1} F\left(X^{s_{l+1}}\right)(t \wedge s_n) \\
&\quad - \sum_{\substack{i=1 \\ i \neq \pi^{-1}(1)}}^{d} \sum_{l=0}^{n-1} F\left(X^{s_l} + p_{\{j\,|\pi(j)<\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right)(t \wedge s_n) - \sum_{l=0}^{n-1} F\left(X^{s_l}\right)(t \wedge s_n).
\end{aligned}$$

For each $i \in \{1,...,d\} \setminus \{\pi^{-1}(d)\}$ there is exactly one $k \in \{1,...,d\} \setminus \{\pi^{-1}(1)\}$ such that

$$p_{\{j\,|\pi(j)\leq\pi(i)\}}(x) = p_{\{j\,|\pi(j)<\pi(k)\}}(x),$$

since $\pi(k) = \pi(i)+1$ if and only if $k = \pi^{-1}(\pi(i)+1)$. Thus, we get

$$\begin{aligned}
\sum_{i=1}^{d} \delta_i^{\mathrm{SU},\pi,\gamma}(F,X)(t \wedge s_n) &= \sum_{l=0}^{n-1} F\left(X^{s_{l+1}}\right)(t \wedge s_n) - \sum_{l=0}^{n-1} F\left(X^{s_l}\right)(t \wedge s_n) \\
&= F\left(X^{s_n}\right)(t \wedge s_n) - F\left(X^{s_0}\right)(t \wedge s_n) \\
&= F(X)(t \wedge s_n) - F(X)(0).
\end{aligned}$$

Since t and n were arbitrary and $s_n \to \infty$ a.s., the decomposition $\delta^{\mathrm{SU},\pi,\gamma}$ is exact. To see the last point, note that two processes with càdlàg paths are indistinguishable if they are modifications. □

Example 2.4.4.

Assume $d = 2$. The SU decomposition with respect to γ defines $d! = 2$ decompositions, namely $\delta^{\mathrm{SU},id,\gamma}(F,X)$ and $\delta^{\mathrm{SU},\varrho,\gamma}(F,X)$ with $\varrho(1) = 2$ and $\varrho(2) = 1$, by

$$\begin{aligned}
\delta_1^{\mathrm{SU},id,\gamma}(F,X) &= \sum_{l=0}^{\infty} \left\{F\left(X_1^{s_{l+1}}, X_2^{s_l}\right) - F\left(X_1^{s_l}, X_2^{s_l}\right)\right\}, \\
\delta_2^{\mathrm{SU},id,\gamma}(F,X) &= \sum_{l=0}^{\infty} \left\{F\left(X_1^{s_{l+1}}, X_2^{s_{l+1}}\right) - F\left(X_1^{s_{l+1}}, X_2^{s_l}\right)\right\}
\end{aligned}$$

and

$$\begin{aligned}
\delta_1^{\mathrm{SU},\varrho,\gamma}(F,X) &= \sum_{l=0}^{\infty} \left\{F\left(X_1^{s_{l+1}}, X_2^{s_{l+1}}\right) - F\left(X_1^{s_l}, X_2^{s_{l+1}}\right)\right\}, \\
\delta_2^{\mathrm{SU},\varrho,\gamma}(F,X) &= \sum_{l=0}^{\infty} \left\{F\left(X_1^{s_l}, X_2^{s_{l+1}}\right) - F\left(X_1^{s_l}, X_2^{s_l}\right)\right\}.
\end{aligned}$$

Definition 2.4.5.
Let $\gamma = \{0 = s_0 < s_1 < \dots\}$ be an unbounded random partition. The *ASU (averaged sequential updating) decomposition* $\delta^{ASU,\gamma} : \mathcal{M} \times \mathcal{X}^d \to \mathcal{X}^d$ is defined by

$$\delta_i^{\mathrm{ASU},\gamma}(F, X) = \frac{1}{d!} \sum_{\pi \in \sigma_d} \delta_i^{\mathrm{SU},\pi,\gamma}(F, X), \quad i = 1, ..., d.$$

Remark 2.4.6.
As in Shorrocks (2013), we observe that

$$\delta_i^{ASU,\gamma}(F, X) = \frac{1}{d!} \sum_{\pi \in \sigma_d} \delta_i^{\mathrm{SU},\pi,\gamma}(F, X) = \sum_{\substack{A \subseteq \{1,...,d\} \\ i \in A}} \delta_i^{\mathrm{SU},A,\gamma}(F, X) \xi_{i,A}$$

for $\xi_{i,A}$ defined in Eq. (2.3.8) and

$$\delta_i^{\mathrm{SU},A,\gamma}(F, X) := \sum_{l=0}^{\infty} \left\{ F\left(X^{s_l} + p_A\left(X^{s_{l+1}} - X^{s_l}\right)\right) - F\left(X^{s_l} + p_{A\setminus\{i\}}\left(X^{s_{l+1}} - X^{s_l}\right)\right) \right\}.$$

Thereby, the computational cost to obtain $\delta_i^{ASU,\gamma}$ can be reduced from $\mathcal{O}(d!)$ to $\mathcal{O}(2^{d-1})$.

Theorem 2.4.7.
Let $\pi \in \sigma_d$ and $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of unbounded random partitions tending to the identity. Let $F \in \mathcal{M}(\mathcal{C}_2)$, $X \in \mathcal{X}^d$, $t \geq 0$ and $i \in \{1, ..., d\}$. Then it holds for $n \to \infty$ that

$$\delta_i^{\mathrm{SU},\pi,\gamma_n}(F, X)(t) \xrightarrow{p} \delta_i^{\mathrm{ISU},\pi}(F, X)(t),$$
$$\delta_i^{\mathrm{ASU},\gamma_n}(F, X)(t) \xrightarrow{p} \delta_i^{\mathrm{IASU}}(F, X)(t).$$

Proof. The proof of Theorem 2.4.7 can be found in Appendix 2.A.3. □

The next example shows that the assumption $F \in \mathcal{M}(\mathcal{C}_2)$ in Theorem 2.4.7 is important to ensure convergence.

Example 2.4.8.
Let Z be a stochastic process with independent increments and $Z_0 = 0$. Jumps of Z shall only occur at fixed times $J = \{2 - l^{-1}, l \in \mathbb{N}\}$, and for each $l \in \mathbb{N}$, the process jumps by $\pm l^{-1}$ with equal probability for upward and downward movements. The process Z stays constant between jumps. Then, Z is a semimartingale, see Černỳ and Ruf (2021). Let

$$f(x_1, x_2) = |x_1 - x_2|,$$

so $f \notin \mathcal{C}_2$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a deterministic sequence of unbounded partitions $\gamma_n = \{0 = s_0^n < s_1^n < \cdots\}$ tending to the identity such that γ_n contains the first n smallest elements of J but the intersection with $(2 - n^{-1}, 2]$ is empty. Assume that $X = (Z, Z)$. Then, for $t \geq 2$ it follows that

$$\sum_{l=0}^{\infty} \left\{ f(X_1^{s_{l+1}^n}(t), X_2^{s_l^n}(t)) - f(X_1^{s_l^n}(t), X_2^{s_l^n}(t)) \right\} = \sum_{l=1}^{n} l^{-1},$$

which is divergent for $n \to \infty$, so the SU decomposition does not converge for the map $F(X)(t) := f(X(t))$, $t \geq 0$.

How can the IASU decomposition be computed efficiently in practice? If we naively approximate the integrals in Definition 2.3.6 numerically, then we may lose exactness of the decomposition, which is undesirable in many applications. Theorem 2.4.7 suggests using the ASU decomposition as an approximation of the IASU decomposition. However, this becomes computationally infeasible for moderately large d, since the computational cost to obtain $\delta_i^{ASU,\gamma}$ scales like $\mathcal{O}(2^{d-1})$. The next corollary provides an elegant solution when there are no simultaneous jumps.

Definition 2.4.9.
Let $\gamma = \{0 = s_0 < s_1 < \dots\}$ be an unbounded random partition. The *2SU (average of two sequential updating) decomposition* $\delta^{2SU,\pi,\gamma} : \mathcal{M} \times \mathcal{X}^d \to \mathcal{X}^d$ *with updating order* $\pi \in \sigma_d$ is defined by

$$\delta_i^{2\mathrm{SU},\pi,\gamma}(F, X) = \frac{1}{2}\left(\delta_i^{\mathrm{SU},\pi,\gamma}(F, X) + \delta_i^{\mathrm{SU},\pi',\gamma}(F, X)\right), \quad i = 1, ..., d,$$

where $\pi' = d + 1 - \pi$.

Corollary 2.4.10.
Let $\pi \in \sigma_d$ and $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of unbounded random partitions tending to the identity. Let $F \in \mathcal{M}(\mathcal{C}_2)$, $X \in \mathcal{X}^d$, $i \in \{1, ..., d\}$ and $t \geq 0$.

i) If $\Delta X_h \Delta X_j = 0$ for all $h, j \in \{1, \dots, d\}$ with $h \neq j$, then

$$\delta_i^{2\mathrm{SU},\pi,\gamma_n}(F, X)(t) \xrightarrow{p} \delta_i^{\mathrm{IASU}}(F, X)(t), \quad n \to \infty.$$

ii) If $[X_h, X_j] = 0$ for all $h, j \in \{1, \dots, d\}$ with $h \neq j$, then

$$\delta_i^{\mathrm{SU},\pi,\gamma_n}(F, X)(t) \xrightarrow{p} \delta_i^{\mathrm{IASU}}(F, X)(t), \quad n \to \infty.$$

Proof. If $\Delta X_h \Delta X_j = 0$, $h \neq j$, it holds by Theorem 2.3.10 that $\delta_i^{\mathrm{IASU}}(F,X) = \frac{1}{2}(\delta_i^{\mathrm{ISU},\pi}(F,X) + \delta_i^{\mathrm{ISU},\pi'}(F,X))$, which is the limit of $\delta_i^{\mathrm{2SU},\pi,\gamma_n}(F,X)$ by Theorem 2.4.7. If $[X_h, X_j] = 0$, $h \neq j$, apply Corollary 2.3.12 and Theorem 2.4.7. □

In particular, the 2SU decomposition with arbitrary updating order π is exact and approximates the IASU decomposition when the risk factors do not have simultaneous jumps. In this case, the computationally expensive averaging to obtain the ASU decomposition can be omitted and the computational complexity to approximate δ_i^{IASU} decreases from $\mathcal{O}(2^{d-1})$ to $\mathcal{O}(1)$. Theorem 2.4.7 and Corollary 2.4.10 are also illustrated in Figure 2.1.

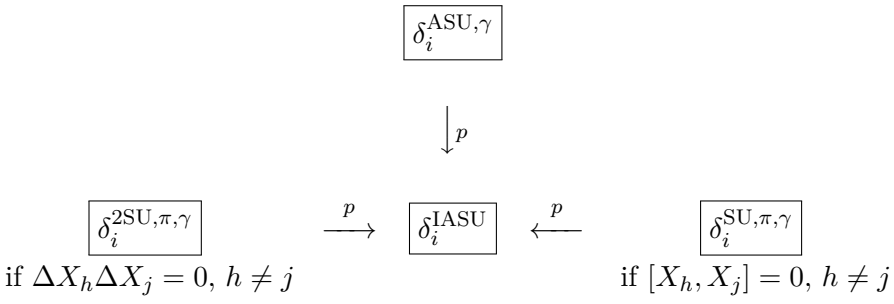

Figure 2.1: Overview of discrete approximations of the IASU decomposition.

Last, we define the OAT decomposition. To obtain the contribution of X_i, all risk factors are fixed at the origin and only X_i is allowed to change from the beginning of a sub-interval to the end of that sub-interval.

Definition 2.4.11.
Let $\gamma = \{0 = s_0 < s_1 < \dots\}$ be an unbounded random partition. The *OAT (one-at-a-time) decomposition* $\delta^{OAT,\gamma} : \mathcal{M} \times \mathcal{X}^d \to \mathcal{X}^d$ is defined by

$$\delta_i^{\mathrm{OAT},\gamma}(F,X) = \sum_{l=0}^{\infty} \left\{ F\left(X_1^{s_l}, ..., X_{i-1}^{s_l}, X_i^{s_{l+1}}, X_{i+1}^{s_l}, ..., X_d^{s_l}\right) - F\left(X^{s_l}\right) \right\}, \quad i = 1, ..., d.$$

Remark 2.4.12.
The OAT decomposition is symmetric but in general not exact. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of unbounded random partitions tending to the identity. For each $i \in \{1, ..., d\}$ choose a permutation $\pi_i \in \sigma_d$ such that $\pi_i(i) = 1$. Then $\delta_i^{\mathrm{OAT},\gamma_n}$ is indistinguishable from $\delta_i^{\mathrm{SU},\pi_i,\gamma_n}$. If $F \in \mathcal{M}(\mathcal{C}_2)$ then it holds by Theorem 2.4.7 for $t \geq 0$ that

$$\delta_i^{\mathrm{OAT},\gamma_n}(F,X)(t) \xrightarrow{p} \delta_i^{\mathrm{ISU},\pi_i}(F,X)(t), \quad i = 1, ..., d$$

for $n \to \infty$. Thus, by Corollary 2.3.12, the three decompositions principles OAT, SU (with arbitrary order $\pi \in \sigma_d$) and ASU are asymptotically indistinguishable if there are no interaction effects.

2.5 Applications

Investment portfolios of financial institutions or insurance companies may include instruments such as stocks, plain vanilla or callable bonds, convertible bonds, inflation-linked bonds, contingent convertible bonds (CoCos), basket options, foreign exchange options and structured products. These instruments often depend on multiple risk factors such as different foreign exchange rates, interest rates for different maturities, credit spreads, inflation rate, some trigger activations for CoCos, multiple equities and time decay. Candland and Latz (2014) also considered defaults and rating changes as risk factors.

In order to obtain a P&L attribution of such instruments, we propose the IASU decomposition because it is exact, symmetric and normalized, and it takes into account the whole paths of the risk factors, i.e., uses all available information. The last point also avoids inconsistencies when reporting a P&L attribution for different time grids, e.g., on an annual, quarterly, monthly and weekly basis. The IASU decomposition involves a stochastic integral. To approximate the IASU decomposition, we propose the ASU or 2SU decomposition with a sufficiently fine time grid, as such an approximation is always an exact decomposition. The use of the 2SU decomposition is theoretically justified when the risk factors do not have simultaneous jumps.

In Section 2.5.1, we provide an exemplary decomposition of a plain vanilla call option with stochastic interest rates on a foreign stock. A change in the P&L of this option can be explained by movements in the stock, the yield curve, the foreign exchange rate and time decay. Thus, there are $d = 4$ risk factors. We analyze the unexplained P&L of the OAT decomposition, the range of the SU and 2SU decompositions over all possible updating orders $\pi \in \sigma_d$ for different time grids, and the convergence of the ASU decomposition to the IASU decomposition.

Computing the ASU decomposition to approximate the IASU decomposition becomes infeasible when the number of risk factors d is moderately large: For example, a plain vanilla bond paying coupons may depend on d yield curves. A basket option may depend on d stocks. In practice, $d = 30$ is a common case for basket options, see Grzelak et al. (2023). In Section 2.5.2, we decompose a digital cash-or-nothing basket put option. We illustrate that it is impossible to obtain the ASU decomposition in reasonable time

when $d = 30$ and we show how the 2SU decomposition is able to break the curse of dimensionality.

2.5.1 Decomposing a call option with stochastic interest rates

In this section, we allocate the P&L of the price of a plain vanilla European call option with strike K and maturity $T = 10$ with stochastic interest rates and foreign exchange exposure. The stock price S is given by a Black-Scholes model with constant volatility $\sigma_S > 0$ and with stochastic interest rates r. The dynamics under the risk neutral measure are given by

$$dS(t) = r(t)S(t)dt + \sigma_S S(t)dB_S(t)$$

and

$$dr(t) = \kappa(\eta - r(t))dt + \sigma_r dB_r(t)$$

with constant volatility $\sigma_r > 0$, long term mean $\eta \in \mathbb{R}$ and speed of mean reversion $\kappa > 0$. Under the physical measure, the stock has drift $\mu_S \in \mathbb{R}$ and the foreign exchange rate Y is assumed to follow a geometric Brownian motion with drift $\mu_Y \in \mathbb{R}$ and volatility $\sigma_Y > 0$ driven by the Brownian motion B_Y. The Brownian motions are assumed to have correlations

$$dB_S(t)dB_r(t) = \rho_{Sr}dt, \quad dB_S(t)dB_Y(t) = \rho_{SY}dt, \quad \text{and} \quad dB_Y(t)dB_r(t) = \rho_{Yr}dt.$$

The time left to maturity is denoted by $\tau(t) = T - t$. The price of the plain vanilla call option $p_{call}(t)$ at time t is given by a twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, see Rabinovitch (1989), i.e.,

$$p_{call}(t) = f\big(S(t), r(t), Y(t), \tau(t)\big) =: F(S, r, Y, \tau)(t), \quad t \geq 0,$$

with

$$f(s, r, y, \tau) = ys\Phi\big(d_+(s, r, \tau)\big) - yKP(r, \tau)\Phi\big(d_-(s, r, \tau)\big),$$

where Φ denotes the distribution function of a standard normal distribution and

$$\begin{aligned} d_\pm(s, r, \tau) &= \frac{\log\left(\frac{s}{KP(r,\tau)}\right) \pm \frac{1}{2}v(\tau)}{\sqrt{v(\tau)}}, \\ v(\tau) &= \sigma_S^2\tau + \sigma_r^2\frac{\tau - 2g_\kappa(\tau) + g_{2\kappa}(\tau)}{\kappa^2} - 2\rho_{Sr}\sigma_S\sigma_r\frac{\tau - g_\kappa(\tau)}{\kappa}, \\ g_\kappa(\tau) &= \frac{1 - e^{-\kappa\tau}}{\kappa}. \end{aligned}$$

The bond price $P(r,\tau)$ is given by

$$P(r,\tau) = A(\tau)e^{-g_\kappa(\tau)r},$$

where

$$A(\tau) = \exp\left(\left(\eta + \frac{\sigma_r^2\lambda}{\kappa} - \frac{\sigma_r^2}{2\kappa^2}\right)(g_\kappa(\tau)-\tau) - \frac{1}{\kappa}\left(\frac{\sigma_r g_\kappa(\tau)}{2}\right)^2\right)$$

and λ denotes the market price of risk. For simplicity, we set the market price of risk to zero and hence assume that the dynamics of r under the physical and the risk neutral measure are identical. Section 24.2 in Björk (2009) describes how to estimate the parameters for r from market data. We simulate 1000 paths of the stock, interest rate and foreign exchange rate under the physical measure over one year. For each path, we decompose the price of the call option at time $t = 1$ with respect to the $d = 4$ risk factors $X := (S, r, Y, \tau)$. We use the following parameters:

$$K = S(0) = 100, \quad \mu_S = 0.05, \quad \sigma_S = 0.4, \quad Y(0) = 1.1, \quad \mu_Y = 0, \quad \sigma_Y = 0.05$$

and

$$r(0) = 0.08, \quad \kappa = 0.1, \quad \eta = 0.05, \quad \sigma_r = 0.01, \quad \rho_{Sr} = -0.7, \quad \rho_{SY} = -0.4, \quad \rho_{Yr} = 0.7.$$

By $\Delta F := F(X)(1) - F(X)(0)$, we denote the P&L of the option over one year. Figure 2.2 shows the relative unexplained P&L of the OAT decomposition, i.e.,

$$\frac{|\Delta F - \sum_{i=1}^d \delta_i^{OAT,\gamma}(F,X)(1)|}{|\Delta F|}.$$

We use as time grids γ annual, quarterly, monthly, weekly and daily time steps. As observed in Flaig and Junike (2024), we also see that the unexplained P&L of the OAT decomposition is significant for all time grids.

Figure 2.3 shows the relative range of the $d!$ SU decompositions for the risk factor S, i.e,

$$\max_{\pi\in\sigma_d}\left(\frac{\delta_1^{SU,\pi,\gamma}(F,X)(1)}{\delta_1^{IASU}(F,X)(1)}\right) - \min_{\pi\in\sigma_d}\left(\frac{\delta_1^{SU,\pi,\gamma}(F,X)(1)}{\delta_1^{IASU}(F,X)(1)}\right)$$

and the relative range of the $\frac{d!}{2}$ 2SU decompositions for the risk factor S. The limiting IASU decomposition is approximated by an ASU decomposition with $10,000$ time steps per year. We observe that the range is significant for the SU decompositions and insignificant for the 2SU decompositions.

The speed of convergence of the ASU to the IASU decomposition is illustrated in Figure 2.4 for the risk factor S, i.e., we show the convergence of

$$\frac{\delta_1^{ASU,\gamma}(F,X)(1)}{\delta_1^{IASU}(F,X)(1)}$$

to one when the partition γ tends to the identity. Figures 2.3 and 2.4 look similar for other risk factors.

In further numerical experiments, we calculate the relative difference between the ASU decomposition and the 2SU decompositions

$$\left| \frac{\delta_i^{2SU,\pi,\gamma}(F,X)(1) - \delta_i^{ASU,\gamma}(F,X)(1)}{\delta_i^{IASU,\gamma}(F,X)(1)} \right|$$

over all risk factors $i \in \{1,...,d\}$, time grids γ and updating orders $\pi \in \sigma_d$, and observe values of less than 0.6% in 95% of the simulations. In conclusion, we find that the ASU decomposition and the 2SU decompositions are strongly dependent on the time grid, but using monthly or weekly time steps instead of annual time steps significantly reduces the deviation of the ASU and 2SU decompositions from the IASU decomposition.

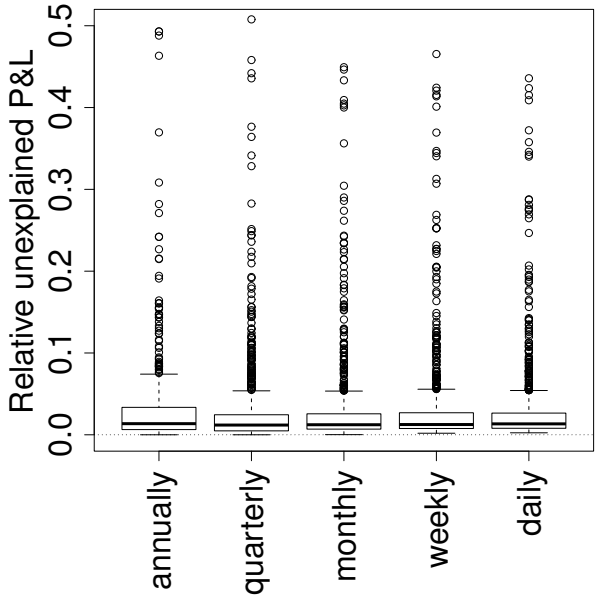

Figure 2.2: Relative unexplained P&L for the OAT decomposition of a plain vanilla call option in a foreign currency at time $t = 1$ for different time grids.

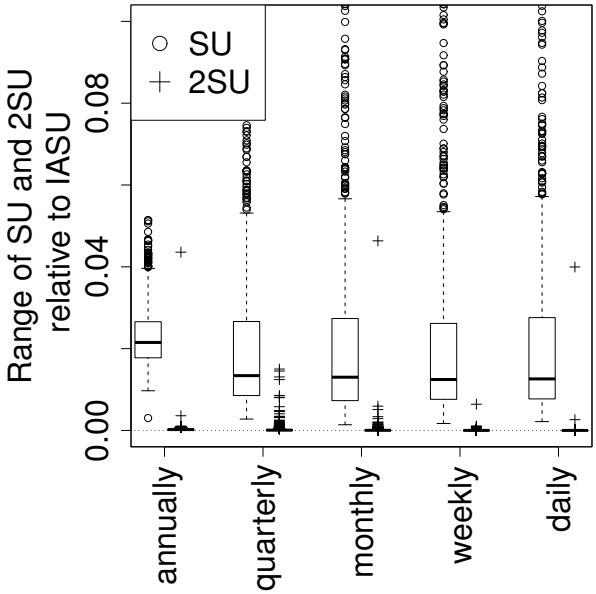

Figure 2.3: Relative range of all SU decompositions and 2SU decompositions for the risk factor S.

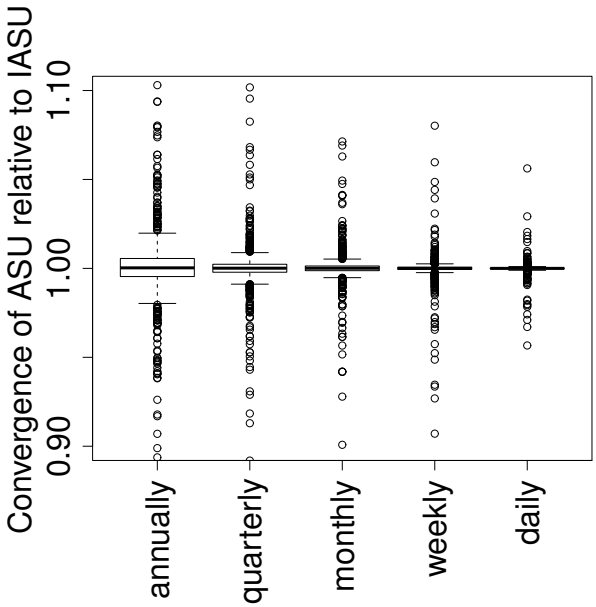

Figure 2.4: Convergence of the ASU decomposition to the IASU decomposition for the risk factor S.

2.5.2 Decomposing a basket option

In this section, we compare the computational cost of obtaining a one-year P&L attribution of a basket option using a naive SU decomposition with annual time grid with the computational cost of obtaining an ASU and a 2SU decomposition based on a monthly time grid, respectively. We consider d risk factors: time decay and $d-1$ different stocks. A digital cash-or-nothing basket put option pays \$1 at maturity T if $S_1(T) \leq K, \ldots, S_{d-1}(T) \leq K$ and zero otherwise. The stock prices are given by a Black-Scholes model. We set the interest rate r to zero. We set $S_i(0) = K = 100$, $i = 1, \ldots, d-1$ and $T = 2$. The price of the option at time $t \in [0, T)$ is equal to $\Phi\big(\log(K), \ldots, \log(K)\big)$, where Φ is the distribution function of a $d-1$ dimensional normal distribution with location

$$\Big(\log\left(S_1(t)\right) - (r - \frac{1}{2}\sigma^2)(T-t), \ldots, \log\left(S_{d-1}(t)\right) - (r - \frac{1}{2}\sigma^2)(T-t)\Big) \in \mathbb{R}^{d-1}$$

and covariance matrix $\Sigma(T-t)$, where we set $\sigma = 0.2$, $\rho = 0.5$ and

$$\Sigma_{ij} = \begin{cases} \sigma^2, & i = j \\ \rho\sigma^2, & i \neq j. \end{cases}$$

Basket options are often priced using Monte Carlo techniques, see Glasserman (2004). For moderate dimensions, many basket options can also be priced using faster Fourier techniques, see Eberlein et al. (2010) and Junike and Stier (2024). We compute Φ using a simple Monte Carlo simulation implemented in C++ with $100,000$ simulations. The experiments are performed on a laptop with Intel i7-11850H processor and 32 GB RAM.

Table 2.1 shows the CPU time needed to obtain Φ for $d \in \{4, 15, 30\}$. We measure CPU times by averaging over 100 runs. Since in some cases the arguments of Φ to obtain a SU decomposition with a certain update order π are the same for different contributions, we need to evaluate Φ only $dL + 1$ times, where L is the number of sub-intervals of $[0, T]$, to obtain the d individual contributions. For example, the 2SU and ASU decompositions with a monthly time grid require $(12d + 1) \cdot 2$ and $(12d + 1) \cdot 2^{d-1}$ evaluations of Φ, respectively.

Table 2.1 also shows the CPU time to compute the SU, ASU and 2SU decompositions. A naive SU decomposition based on an annual time grid is at most 24 times faster than a 2SU decomposition with a monthly time grid. The computational cost of the 2SU decomposition for each contribution is dimension independent, except for the longer time required to evaluate Φ. Compared to the ASU decomposition, the 2SU decomposition is

2^{d-2} times faster. The ASU decomposition cannot be computed in reasonable time for $d \geq 30$.

	Number of evaluations of Φ	$d = 4$	$d = 15$	$d = 30$
Evaluation of Φ	1	0.018 sec	0.15 sec	0.54 sec
SU with annual grid	$d + 1$	(0.09 sec)	(2.4 sec)	(16.7 sec)
2SU with monthly grid	$(12d + 1) \cdot 2$	(1.76 sec)	(54.3 sec)	(390 sec)
ASU with monthly grid	$(12d + 1) \cdot 2^{(d-1)}$	(7.06 sec)	(123.6 hours)	(3318.7 years)

Table 2.1: CPU time to compute the d contributions of the SU, ASU and 2SU decompositions of a basket option over one year using different time grids. The CPU time of Φ is obtained from a Monte Carlo simulation. The CPU times in brackets are estimated using the CPU time of Φ and the known complexities of the three decompositions.

Remark 2.5.1.
To reduce the computational time, it is possible to compute the d contributions for the SU, 2SU and ASU decompositions in parallel, which would reduce the numerical effort by a factor of d. Furthermore, the sums for the SU, 2SU and ASU decompositions can also be parallelized. For example, for the 2SU decomposition we need to perform $2(dL + 1)$ function evaluations to obtain all d contributions. If a function evaluation takes 0.54 sec in $d = 30$ dimensions as in the Table 2.1, the computation time for the 2SU decomposition with monthly time grid could be reduced from 390 sec to about 0.54 sec using 722 cores for parallelization.

2.6 Conclusions

We showed that the IASU decomposition is the only (up to indistinguishability) exact and symmetric decomposition in the family of Itô decompositions, which is a large class of normalized decompositions based on an extended version of Itô's formula. This axiomatic result, together with the fact that the IASU decomposition is grid-independent and considers the full paths of the risk basis, makes it a decomposition of choice from a theoretical perspective. In practice, the calculation of the IASU decomposition comes with two challenges: it involves stochastic integrals that must be approximated, and the computational effort explodes as the number of risk factors increases.

We have shown that the IASU decomposition can be approximated by the ASU decomposition (which is always exact and symmetric) if we use a sufficiently fine time grid, but the ASU decomposition also suffers from the curse of dimensionality as the number of risk factors increases. For applications where different risk factors may have interactions but almost surely do not have simultaneous jumps, we have shown that the IASU decomposition is indistinguishable from the average of two ISU decompositions, thus breaking the curse of dimensionality. Therefore, from a theoretical point of view, the 2SU decomposition with sufficiently fine time steps is an appropriate approximation of the IASU decomposition.

Based on our own numerical experiments and the empirical analysis of Flaig and Junike (2024), we recommend using monthly or even weekly time steps instead of annual time steps.

The additional computational cost of our two recommendations is moderate, but the theoretical properties of the decomposition are dramatically improved.

Acknowledgements

We thank Bernd Buchwald and Andreas Märkert from Hannover Re, Solveig Flaig from Deutsche Rück, Julien Hambuckers from University of Liège, Jan-Frederik Mai from XAIA Investment GmbH, Thorsten Schmidt from Universtität Freiburg and an anonymous referee for very helpful comments that improved this paper.

2.A Appendix

2.A.1 Auxiliary results

Lemma 2.A.1.
Let $i, j \in \{1, ..., d\}$. Let $\pi, \eta \in \sigma_d$ and $x \in \mathbb{R}^d$. Then it holds that

$$\eta^{-1}\Big(p_{\{j\,|\pi(j)\leq\pi(\eta^{-1}(i))\}}(\eta(x))\Big) = p_{\{j\,|\pi(\eta^{-1}(j))\leq\pi(\eta^{-1}(i))\}}(x). \tag{2.A.1}$$

Proof. Let $k \in \{j|\pi(j) \leq \pi(\eta^{-1}(i))\}$, which is equivalent to

$$\eta(k) \in \{j|\pi(\eta^{-1}(j)) \leq \pi(\eta^{-1}(i))\}.$$

Since $\left(\eta^{-1}(x)\right)_{\eta(k)} = x_k$ and $\left(\eta(x)\right)_k = x_{\eta(k)}$, we obtain that

$$\begin{aligned}\Big(\eta^{-1}\big(p_{\{j\,|\pi(j)\leq\pi(\eta^{-1}(i))\}}(\eta(x))\big)\Big)_{\eta(k)} &= \Big(p_{\{j\,|\pi(j)\leq\pi(\eta^{-1}(i))\}}(\eta(x))\Big)_k \\ &= \big(p_{\{j\,|\pi(\eta^{-1}(j))\leq\pi(\eta^{-1}(i))\}}(x)\big)_{\eta(k)},\end{aligned}$$

which leads to Eq. (2.A.1). □

Lemma 2.A.2.
Let $\eta \in \sigma_d$, $i \in \{1, ..., d\}$, $X \in \mathcal{X}^d$, $F \in \mathcal{M}(\mathcal{C}_2)$ and $(\mu_\pi)_{\pi\in\sigma_d} \subset [0, 1]$. If $F(\eta(X)) = F(X)$, then it holds that

$$\sum_{\pi\in\sigma_d} \mu_\pi S^\pi_{\eta^{-1}(i)}(\eta(X)) = \sum_{\substack{A\subseteq\{1,...,d\}\\ i\in A}} S^A_i(X)\xi_{i,A,\eta}$$

with

$$\xi_{i,A,\eta} := \sum_{\substack{\pi\in\sigma_d\\ \{j|\pi(\eta^{-1}(j))\leq\pi(\eta^{-1}(i))\}=A}} \mu_\pi. \tag{2.A.2}$$

In particular, for an Itô decomposition δ with parameters $(\lambda_{ij})_{i,j=1,...,d}$ and $(\mu_\pi)_{\pi\in\sigma_d}$, we have that

$$\delta_i(F, X) = I_i + \frac{1}{2}I_{ii} + \sum_{\substack{j=1\\ j\neq i}}^{d} \lambda_{ij} I_{ij} + \sum_{\substack{A\subseteq\{1,...,d\}\\ i\in A}} S^A_i(X)\xi_{i,A,id}. \tag{2.A.3}$$

Proof. Let $\eta \in \sigma_d$ and $F(X)(t) = f(X(t))$, $t \geq 0$, with $F(\eta(X)) = F(X)$ for $X \in \mathcal{X}^d$. Let $i \in \{1, ..., d\}$. By Eq. (2.A.1) it holds for $s > 0$ that

$$\begin{aligned}& f\Big(\eta(X(s-)) + p_{\{j\,|\pi(j)\leq\pi(\eta^{-1}(i))\}}(\Delta\eta(X)(s))\Big) \\ =\ & f\Big(\eta\Big[X(s-) + \eta^{-1}\big[p_{\{j\,|\pi(j)\leq\pi(\eta^{-1}(i))\}}(\eta(\Delta X(s)))\big]\Big]\Big) \\ \overset{(2.A.1)}{=}\ & f\Big(\eta\big[X(s-) + p_{\{j\,|\pi(\eta^{-1}(j))\leq\pi(\eta^{-1}(i))\}}(\Delta X(s))\big]\Big) \\ =\ & f\Big(X(s-) + p_{\{j\,|\pi(\eta^{-1}(j))\leq\pi(\eta^{-1}(i))\}}(\Delta X(s))\Big).\end{aligned}$$

The last equality follows from the symmetry of f. Similarly, if we replace "$\leq$" with "$<$", we get that

$$f\Big(\eta(X(s-)) + p_{\{j\,|\pi(j)<\pi(\eta^{-1}(i))\}}(\Delta\eta(X)(s))\Big)$$

$$= f\Big(X(s-) + p_{\{j\,|\pi(\eta^{-1}(j))<\pi(\eta^{-1}(i))\}}(\Delta(X)(s))\Big).$$

Let $\eta \in \sigma_d$ and $f \in \mathcal{C}_2$. If $f(x) = f(\eta(x))$, $x \in \mathbb{R}^d$, it is straightforward to see that for $x \in \mathbb{R}^d$ it holds that

$$f_i(x) = f_{\eta^{-1}(i)}(\eta(x)), \quad f_{ij}(x) = f_{\eta^{-1}(i)\eta^{-1}(j)}(\eta(x)) \quad \text{and} \quad (\eta(x))_{\eta^{-1}(i)} = x_i. \tag{2.A.4}$$

Therefore it follows that

$$S^{\pi}_{\eta^{-1}(i)}(\eta(X)) = S_i^{\pi\circ\eta^{-1}}(X). \tag{2.A.5}$$

Thus, similarly to Shorrocks (2013), for any re-ordering $\eta(X)$ of the risk basis X we can conclude that

$$\begin{aligned}
\sum_{\pi\in\sigma_d} \mu_\pi S^{\pi}_{\eta^{-1}(i)}(\eta(X)) &\overset{(2.A.5)}{=} \sum_{\pi\in\sigma_d} \mu_\pi S_i^{\pi\circ\eta^{-1}}(X) \\
&= \sum_{\substack{A\subseteq\{1,\dots,d\}\\ i\in A}} \sum_{\substack{\pi\in\sigma_d\\ \{j|\pi(\eta^{-1}(j))\le\pi(\eta^{-1}(i))\}=A}} \mu_\pi S_i^{\pi\circ\eta^{-1}}(X) \\
&= \sum_{\substack{A\subseteq\{1,\dots,d\}\\ i\in A}} S_i^A(X) \sum_{\substack{\pi\in\sigma_d\\ \{j|\pi(\eta^{-1}(j))\le\pi(\eta^{-1}(i))\}=A}} \mu_\pi \\
&= \sum_{\substack{A\subseteq\{1,\dots,d\}\\ i\in A}} S_i^A(X)\xi_{i,A,\eta}.
\end{aligned}$$

Eq. (2.A.3) follows directly for $\eta = id$. □

Lemma 2.A.3.
Let δ be an Itô decomposition with parameters $(\lambda_{ij})_{i,j=1,\dots,d}$ and $(\mu_\pi)_{\pi\in\sigma_d}$. Let $i \in \{1,\dots,d\}$. If δ is symmetric and exact, it follows that

$$\xi_{i,A,id} = \xi_{\eta^{-1}(i),\eta^{-1}(A),id} \tag{2.A.6}$$

for any $\eta \in \sigma_d$, where $\xi_{i,A,id}$ is defined in Eq. (2.A.2) and $\eta(A) := \{\eta(j) : j \in A\}$. Further, for any $a \in \{1,\dots,d\}$ it holds that

$$\sum_{j=1}^{d} \sum_{\substack{A\subseteq\{1,\dots,d\}\\ |A|=a,\, j\in A}} \xi_{j,A,id} = 1. \tag{2.A.7}$$

Proof. First we show Eq. (2.A.6). Let $A \subseteq \{1,\dots,d\}$ with $i \in A$. Let $\pi, \eta \in \sigma_d$. Because of

$$\{j|\pi(\eta^{-1}(j)) \le \pi(\eta^{-1}(i))\} = A \quad \Leftrightarrow \quad \{j|\pi(j) \le \pi(\eta^{-1}(i))\} = \eta^{-1}(A),$$

it holds that

$$\xi_{i,A,\eta} = \sum_{\substack{\pi\in\sigma_d \\ \{j|\pi(\eta^{-1}(j))\leq\pi(\eta^{-1}(i))\}=A}} \mu_\pi = \sum_{\substack{\pi\in\sigma_d \\ \{j|\pi(j)\leq\pi(\eta^{-1}(i))\}=\eta^{-1}(A)}} \mu_\pi = \xi_{\eta^{-1}(i),\eta^{-1}(A),id}. \quad (2.A.8)$$

Now let $f(x) = \prod_{j=1}^d x_j^2$ and $F(X)(t) = f(X(t))$, $t \geq 0$, so that $F(X) = F(\pi(X))$, $\pi \in \sigma_d$. For $B \subseteq \{1,\dots,d\}$ with $i \in B$ and $t \geq 0$, let

$$X_j(t) = \begin{cases} 1_{[1,\infty)}(t), & j \in B \\ 1_{[0,1)}(t), & j \notin B. \end{cases}$$

Then it follows that

$$f\Big(X(1-) + p_A(\Delta X(1))\Big) = \begin{cases} 1, & A = B \\ 0, & A \neq B \end{cases}$$

and therefore

$$S_i^A(X)(1) = \begin{cases} 1, & A = B \\ 0, & A \neq B \end{cases}$$

for $A \subseteq \{1,\dots,d\}$ with $i \in A$. For $\eta \in \sigma_d$ it follows by Lemma 2.A.2 that

$$\delta_{\eta^{-1}(i)}(F,\eta(X))(1) = \sum_{\substack{A\subseteq\{1,\dots,d\} \\ i\in A}} S_i^A(X)(1)\xi_{i,A,\eta} = \xi_{i,B,\eta}.$$

Since δ is symmetric, we have that

$$\xi_{\eta^{-1}(i),\eta^{-1}(B),id} \overset{(2.A.8)}{=} \xi_{i,B,\eta} = \delta_{\eta^{-1}(i)}(F,\eta(X))(1) = \delta_i(F,X)(1) = \xi_{i,B,id}.$$

Since B was arbitrary, we have just shown Eq. (2.A.6).

Now we iteratively show Eq. (2.A.7). Let $X_j(t) = 1_{[1,\infty)}(t)$, $t \geq 0$, $j = 1,\dots,d$ and let $f^a \in \mathcal{C}_2$ such that for $a \in \{1,\dots,d\}$

$$f^a(x) = \begin{cases} 1, & \sum_{j=1}^d x_j = a \\ 0, & \sum_{j=1}^d x_j \in (-\infty, a-1] \cup [a+1,\infty) \end{cases}$$

and $f_i^a(X) = 0$ if $\sum_{j=1}^d x_j \leq a-1$, $i = 1,\dots,d$. Let $F^a(X)(t) = f^a(X(t))$, $t \geq 0$. If $a = d$, then

$$S_j^A(X)(1) = \begin{cases} 1, & |A| = a \\ 0, & \text{otherwise} \end{cases}$$

for $j = 1, \dots, d$ and $A \subseteq \{1, \dots, d\}$ with $j \in A$. By exactness and Lemma 2.A.2 it follows that

$$\begin{aligned} 1 &= F^a(X)(1) - F^a(X)(0) \\ &= \sum_{j=1}^{d} \delta_j(F^a, X)(1) \\ &= \sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ j \in A}} S_j^A(X)(1)\xi_{j,A,id} \\ &= \sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ |A|=d,\, j \in A}} \xi_{j,A,id}. \end{aligned} \tag{2.A.9}$$

Now let $a = d - 1$, then

$$S_j^A(X)(1) = \begin{cases} 1, & |A| = a \\ -1, & |A| = a + 1 \\ 0, & \text{otherwise} \end{cases}$$

for $A \subseteq \{1, \dots, d\}$ with $j \in A$. Again, by exactness we have that

$$\begin{aligned} 0 =& F^a(X)(1) - F^a(X)(0) \\ =& \sum_{j=1}^{d} \delta_j(F^a, X) \\ =& \sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ j \in A}} S_j^A(X)(1)\xi_{j,A,id} \\ =& \sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ |A|=d-1,\, j \in A}} \xi_{j,A,id} - \sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ |A|=d,\, j \in A}} \xi_{j,A,id}. \end{aligned}$$

Using Eq. (2.A.9) we obtain that

$$\sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ |A|=d-1,\, j \in A}} \xi_{j,A,id} = 1.$$

Iteratively for any $a \in \{1, \dots, d\}$ it follows that

$$\sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ |A|=a,\, j \in A}} \xi_{j,A,id} = 1.$$

□

2.A.2 Proof of Theorem 2.3.9

Proof. First we show that the IASU decomposition is exact and symmetric and satisfies Eq. (2.3.9): By Proposition 2.3.3, it follows that δ^{IASU} is an exact Itô decomposition. Use Eq. (2.A.4) to see that the IASU decomposition is symmetric. If $d = 1$, Eq. (2.3.9) is trivially true. Assume $d \geq 2$. Fix $i \in \{1, ..., d\}$. Note that

$$\sum_{\pi \in \sigma_d} 1_{\{\pi(j) < \pi(i)\}} = \begin{cases} \frac{d!}{2}, & j \neq i \\ 0, & j = i. \end{cases}$$

It follows that

$$\begin{aligned} \frac{1}{d!} \sum_{\pi \in \sigma_d} \sum_{\substack{j=1 \\ \pi(j) < \pi(i)}}^{d} I_{ij} &= \sum_{j=1}^{d} \left\{ I_{ij} \frac{1}{d!} \sum_{\pi \in \sigma_d} 1_{\{\pi(j) < \pi(i)\}} \right\} \\ &= \frac{1}{2} I_{i1} + ... + \frac{1}{2} I_{i(i-1)} + \frac{1}{2} I_{i(i+1)} + ... + \frac{1}{2} I_{id} \\ &= \frac{1}{2} \sum_{j \neq i} I_{ij}. \end{aligned} \tag{2.A.10}$$

Eq. (2.A.10) implies Eq. (2.3.9). Now we show that all exact and symmetric Itô decompositions are indistinguishable from the IASU decomposition. Let δ be a symmetric and exact Itô decomposition with parameters $(\lambda_{ij})_{i,j=1,...,d}$ and $(\mu_\pi)_{\pi \in \sigma_d}$. Since the Itô decomposition is over-parameterised, we use the alternative parametrization according to Eq. (2.A.3). To prove that δ is indistinguishable from the IASU decomposition, we show that λ_{ij} and $\xi_{i,A,id}$ are equal to the coefficients $\frac{1}{2}$ and $\xi_{i,A}$ as defined in Eq. (2.3.8).

Suppose that $\lambda_{hk} \neq \frac{1}{2}$. Let $X \in \mathcal{X}^d$ have continuous paths with $X_i = 1$, $i \notin \{h, k\}$, and $[X_h, X_k] \neq 0$. Let $F(X) = \prod_{i=1}^{d} X_i$. Then $F(X) = F(\pi(X))$ for $\pi \in \sigma_d$. Note that $I_{kh} = I_{hk}$. As δ is exact, we have

$$\sum_{i=1}^{d} \delta_i(F, X) = I_h + I_k + \lambda_{hk} I_{hk} + \lambda_{kh} I_{kh} = F(X) - F(X)(0) = I_h + I_k + I_{hk},$$

hence $\lambda_{kh} = 1 - \lambda_{hk} \neq \lambda_{hk}$. Let $\pi \in \sigma_d$ such that $\pi^{-1}(h) = k$. Then, it follows that

$$\delta_{\pi^{-1}(h)}(F, \pi(X)) = \delta_k(F, \pi(X)) = I_h + \lambda_{kh} I_{kh} \neq I_h + \lambda_{hk} I_{hk} = \delta_h(F, X).$$

That means that δ is not symmetric, which is a contradiction to our assumption. So we necessarily have that $\lambda_{ij} = \frac{1}{2}$, $i, j = 1, ..., d$.

Now let $a \in \{1, \ldots, d\}$. For $i, j \in \{1, \ldots, d\}$, let $A, B \subseteq \{1, \ldots, d\}$ with $|A| = |B| = a$ and $i \in A$, $j \in B$. Then there is a permutation $\eta \in \sigma_d$ such that $\eta^{-1}(A) = B$ and $j = \eta^{-1}(i)$. By Eq. (2.A.6) it follows that

$$\xi_{i,A,id} = \xi_{j,B,id}. \tag{2.A.11}$$

Let $A_1, \dots, A_d \subseteq \{1, \dots, d\}$ with $j \in A_j$ and $|A_j| = a$, $j = 1, \dots, d$. Since

$$|\{A \subseteq \{1, \dots, d\} : j \in A, |A| = a\}| = \binom{d-1}{a-1}, \tag{2.A.12}$$

we obtain by Eqs. (2.A.7), (2.A.11) and (2.A.12) that

$$1 = \sum_{j=1}^{d} \sum_{\substack{A \subseteq \{1,\dots,d\} \\ |A|=a,\, j \in A}} \xi_{j,A,id} = \sum_{j=1}^{d} \binom{d-1}{a-1} \xi_{j,A_j,id} = d \binom{d-1}{a-1} \xi_{i,A,id}$$

for $A \subseteq \{1, \dots, d\}$ with $i \in A$ and $|A| = a$. Therefore we can conclude that

$$\xi_{i,A,id} = \frac{1}{d\binom{d-1}{|A|-1}} = \frac{(|A|-1)!(d-|A|)!}{d!}.$$

□

2.A.3 Proof of Theorem 2.4.7

Proof. Let $t > 0$. Fix some $i \in \{1, ..., d\}$ and some permutation π. Since $F \in \mathcal{M}(\mathcal{C}_2)$, by definition there is an $f \in \mathcal{C}_2$ such that $F(X)(t) = f(X(t))$, $t \geq 0$. We first show that $\delta^{\mathrm{SU},\pi,\gamma_n}(F, X)(t) \xrightarrow{p} \delta^{\mathrm{ISU},\pi}(F, X)(t)$ for $n \to \infty$. Let $\gamma_n = \{0 = s_0^n < s_1^n < ...\}$, $n \in \mathbb{N}$, be a sequence of unbounded random partitions tending to the identity. Let $\alpha > 0$ and

$$\mathcal{A}_\alpha := \{s \in (0, t] \,:\, \max_{j=1,\dots,d} |\Delta X_j(s)| > \alpha\}.$$

The set $\mathcal{A}_\alpha$ contains all time points in $[0, t]$ where at least one component of a path $u \mapsto X(u)$ has jumps greater than α. The SU decomposition $\delta_i^{\mathrm{SU},\pi,\gamma_n}$ with respect to γ_n can be written as

$$\begin{aligned}
\delta_i^{\mathrm{SU},\pi,\gamma_n}(F, X)(t) = \sum_{l \in \mathbb{A}_\alpha} \Big\{ & f\left(X^{s_l^n}(t) + p_{\{j\,|\pi(j) \leq \pi(i)\}}\left(X^{s_{l+1}^n}(t) - X^{s_l^n}(t)\right)\right) \\
& - f\left(X^{s_l^n}(t) + p_{\{j\,|\pi(j) < \pi(i)\}}\left(X^{s_{l+1}^n}(t) - X^{s_l^n}(t)\right)\right) \Big\} \\
+ \sum_{l \in \mathbb{A}_\alpha^c} \Big\{ & f\left(X^{s_l^n}(t) + p_{\{j\,|\pi(j) \leq \pi(i)\}}\left(X^{s_{l+1}^n}(t) - X^{s_l^n}(t)\right)\right) \\
& - f\left(X^{s_l^n}(t) + p_{\{j\,|\pi(j) < \pi(i)\}}\left(X^{s_{l+1}^n}(t) - X^{s_l^n}(t)\right)\right) \Big\},
\end{aligned} \tag{2.A.13}$$

where $\mathbb{A}_\alpha = \{l \in \mathbb{N}_0 : \mathcal{A}_\alpha \cap (s_l^n, s_{l+1}^n] \neq \emptyset\}$ and $\mathbb{A}_\alpha^c = \mathbb{N}_0 \setminus \mathbb{A}_\alpha$. The first sum at the right-hand side of Eq. (2.A.13) converges a.s. for $n \to \infty$ to

$$\sum_{s \in \mathcal{A}_\alpha} \left\{ f\left(X(s-) + p_{\{j \,|\pi(j) \leq \pi(i)\}}(\Delta X(s))\right) - f\left(X(s-) + p_{\{j \,|\pi(j) < \pi(i)\}}(\Delta X(s))\right) \right\}. \tag{2.A.14}$$

Using a Taylor expansion and the same arguments as in the proof of the classical Itô's formula, one can show that the second sum of the right-hand side of Eq. (2.A.13) converges in probability for $n \to \infty$ to

$$I_i(t) + \frac{1}{2} H_{ii}(t) + \sum_{\pi(j) < \pi(i)} H_{ij}(t) - \sum_{s \in \mathcal{A}_\alpha} \left\{ f_i\left(X(s-)\right) \Delta X_i(s) + \frac{1}{2} f_{ii}\left(X(s-)\right) \left(\Delta X_i(s)\right)^2 + \sum_{\substack{j=1 \\ \pi(j) < \pi(i)}}^{d} f_{ij}\left(X(s-)\right) \Delta X_i(s) \Delta X_j(s) \right\}, \tag{2.A.15}$$

where $H_{ij} = \int_0^\cdot f_{ij}(X(s-)) d[X_i, X_j](s)$. The sum of the Eqs. (2.A.14) and (2.A.15) is

$$I_i(t) + \frac{1}{2} H_{ii}(t) + \sum_{\pi(j) < \pi(i)} H_{ij}(t) + \tag{2.A.16}$$

$$\sum_{s \in \mathcal{A}_\alpha} \left\{ f\left(X(s-) + p_{\{j \,|\pi(j) \leq \pi(i)\}}(\Delta X(s))\right) - f\left(X(s-) + p_{\{j \,|\pi(j) < \pi(i)\}}(\Delta X(s))\right) - f_i\left(X(s-)\right) \Delta X_i(s) \right\} \tag{2.A.17}$$

$$- \sum_{s \in \mathcal{A}_\alpha} \frac{1}{2} f_{ii}\left(X(s-)\right) \left(\Delta X_i(s)\right)^2 \tag{2.A.18}$$

$$- \sum_{s \in \mathcal{A}_\alpha} \sum_{\substack{j=1 \\ \pi(j) < \pi(i)}}^{d} f_{ij}\left(X(s-)\right) \Delta X_i(s) \Delta X_j(s). \tag{2.A.19}$$

Since X is a semimartingale, and because of Lemma 2.3.2, we can see that the sums (2.A.17), (2.A.18) and (2.A.19) are absolutely convergent for $\alpha \to 0$ so that (2.A.16-2.A.19) converge for $\alpha \to 0$ to $\delta^{\mathrm{ISU},\pi}(F, X)(t)$, using that

$$I_{ij} = H_{ij} - \sum_{0 \leq s \leq \cdot} f_{ij}\left(X(s-)\right) \Delta X_i(s) \Delta X_j(s).$$

By Theorem 2.3.9 we get $\delta^{\mathrm{ASU},\gamma_n}(F, X)(t) \xrightarrow{p} \delta^{\mathrm{IASU},\pi}(F, X)(t)$ for $n \to \infty$. □

2.A.4 Stability

In this section, we use the notation of Christiansen (2022). For $i = 1, 2$, let $\tau_i : [0, \infty) \to [0, \infty)$ with $\tau_i(t) \leq t$ for all $t \geq 0$. The function

$$\tau(t) = \big(\tau_1(t), \tau_2(t)\big)$$

is called a *delay*. A delay is called *phased* if there is an unbounded partition $\{0 = s_0 < s_1 < ...\}$ of $[0, \infty)$ such that on each interval $(s_l, s_{l+1}]$, at most one component of τ is nonconstant. Let $(\tau^n)_{n\in\mathbb{N}}$ be a *refining sequence of delays that increase to identity* (rsdii), i.e.,

$$\tau_i^n\big([0,t]\big) \subset \tau_i^{n+1}\big([0,t]\big),\ n \in \mathbb{N}, \quad \text{and } \overline{\bigcup_{n\in\mathbb{N}} \tau_i^n\big([0,t]\big)} = [0,t], \quad i = 1, 2.$$

Let $\mathcal{T}$ be a set containing at least one phased rsdii. Let $X = (X_1, X_2)$ be a semimartingale, and define

$$X \diamond \tau := (X_1 \circ \tau_1, X_2 \circ \tau_2), \quad \tau \in \mathcal{T}.$$

Let

$$\mathbb{X} = \{X \diamond \tau\ :\ \tau \in \mathcal{T}\} \cup \{X\}.$$

Let $\mathbb{D}_0$ be the set of càdlàg processes starting in zero and let $\varrho : \mathbb{X} \to \mathbb{D}_0$. A mapping $\delta : \mathbb{X} \to \mathbb{D}_0^2$ is called *decomposition scheme of* ϱ. The mapping δ assigns to each $Y \in \mathbb{X}$ a decomposition of $\varrho(Y)$. The ISU decomposition scheme is abbreviated δ^{ISU}.

A decomposition scheme is called *stable at* X if

$$\delta(X \diamond \tau^n)(t-) \xrightarrow{p} \delta(X)(t-), \quad n \to \infty,$$

at each $t > 0$ for all rsdii $(\tau^n)_{n\in\mathbb{N}} \subset \mathcal{T}$.

Proposition 2.A.4.
Assume that $X = (X_1, X_2)$ with $X_1 = X_2 = B$ for a Brownian motion B. Let $\varrho(Y) = Y_1 Y_2$ be a simple product. Then, there is a set $\mathcal{T}$ of continuous phased rsdii such that the ISU decomposition δ^{ISU} of ϱ is not stable at X.

Proof. Suppose that $\mathcal{T}$ contains a continuous phased rsdii $(\tau^n) = (\tau_1^n, \tau_2^n), n \in \mathbb{N}$, with $\tau_1^n \leq \tau_2^n$, $n \in \mathbb{N}$. For a partition $(a_{l,i}^n, b_{l,i}^n]$, $l \in \mathbb{N}_0$, $i = 1, 2$ of $[0, \infty)$ such that $(\tau_j^n)_{j\neq i}$ is constant on $(a_{l,i}^n, b_{l,i}^n]$, let $\tau_1^n(a_{l,2}^n) = \tau_2^n(a_{l,2}^n)$, $n \in \mathbb{N}$, $l \in \mathbb{N}_0$. In addition, let $\mathcal{T}$ also contain

$(\tilde{\tau}^n)_{n\in\mathbb{N}} = ((\tau_2^n, \tau_1^n))_{n\in\mathbb{N}}$. Since $\tau_2^n(a_{l,1}^n) = \tau_2^n(b_{l,1}^n) = \tau_1^n(b_{l,1}^n)$ and by the multidimensional Taylor theorem,

$$\begin{aligned}\delta_1^{ISU}(X \diamond \tau^n)(t) &= \sum_l \Big(\varrho\big((X \diamond \tau^n)^{b_{l,1}^n \wedge t}\big) - \varrho\big((X \diamond \tau^n)^{a_{l,1}^n \wedge t}\big) \Big) \\ &= \sum_l \varrho_1\big((X \diamond \tau^n)^{a_{l,1}^n \wedge t}\big) \Big(X_1\big(\tau_1^n(b_{l,1}^n \wedge t)\big) - X_1\big(\tau_1^n(a_{l,1}^n \wedge t)\big) \Big).\end{aligned}$$

By the definitions of X_1, X_2 and ρ,

$$\begin{aligned}\delta_1^{ISU}(X \diamond \tau^n)(t) &= \sum_l B\big(\tau_2^n(a_{l,1}^n \wedge t)\big) \Big(B\big(\tau_1^n(b_{l,1}^n \wedge t)\big) - B\big(\tau_1^n(a_{l,1}^n \wedge t)\big) \Big) \\ &= \sum_l B\big(\tau_1^n(b_{l,1}^n \wedge t)\big) \Big(B\big(\tau_1^n(b_{l,1}^n \wedge t)\big) - B\big(\tau_1^n(a_{l,1}^n \wedge t)\big) \Big) \\ &= \sum_l B(t_l)\big(B(t_l \wedge t) - B(t_{l-1} \wedge t)\big) \\ &= 2\sum_l \frac{\big(B(t_l) + B(t_{l-1})\big)}{2}\big(B(t_l \wedge t) - B(t_{l-1} \wedge t)\big) \\ &\quad - \sum_l B(t_{l-1})\big(B(t_l \wedge t) - B(t_{l-1} \wedge t)\big)\end{aligned}$$

for $t_l^n := \tau_1^n(b_{l,1}^n) = \tau_1^n(a_{l+1,1}^n) = \tau_2^n(b_{l-1,2}^n) = \tau_2^n(a_{l,2}^n)$. Let $\int_0^t B_s \circ dB_s$ denote the Stratonovich integral and $\int_0^t B_s dB_s$ the Itô integral. It holds that

$$\delta_1^{ISU}(X \diamond \tau^n)(t) \xrightarrow{p} 2\int_0^t B_s \circ dB_s - \int_0^t B_s dB_s = \frac{1}{2}B_t^2 + \frac{1}{2}t$$

for $n \to \infty$. By the same arguments,

$$\begin{aligned}\delta_1^{ISU}(X \diamond \tilde{\tau}^n)(t) &= \sum_l \Big(\varrho\big((X \diamond \tilde{\tau}^n)^{b_{l,2}^n \wedge t}\big) - \varrho\big((X \diamond \tilde{\tau}^n)^{a_{l,2}^n \wedge t}\big) \Big) \\ &= \sum_l \varrho_1\big((X \diamond \tilde{\tau}^n)^{a_{l,2}^n \wedge t}\big) \Big(X_2\big(\tau_2^n(b_{l,2}^n \wedge t)\big) - X_2\big(\tau_2^n(a_{l,2}^n \wedge t)\big) \Big) \\ &= \sum_l B\big(\tau_1^n(a_{l,2}^n \wedge t)\big) \Big(B\big(\tau_2^n(b_{l,2}^n \wedge t)\big) - B\big(\tau_2^n(a_{l,2}^n \wedge t)\big) \Big) \\ &= \sum_l B\big(\tau_2^n(a_{l,2}^n \wedge t)\big) \Big(B\big(\tau_2^n(b_{l,2}^n \wedge t)\big) - B\big(\tau_2^n(a_{l,2}^n \wedge t)\big) \Big) \\ &= \sum_l B(t_l)\big(B(t_{l+1} \wedge t) - B(t_l \wedge t)\big) \\ &\xrightarrow{p} \int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t\end{aligned}$$

for $n \to \infty$. Therefore,

$$\operatorname{plim}_{n\to\infty} \delta_i^{ISU}(X \diamond \tau^n)(t) \neq \operatorname{plim}_{n\to\infty} \delta_i^{ISU}(X \diamond \tilde{\tau}^n)(t), \quad i = 1, 2,$$

for $t > 0$, and hence, the ISU decomposition of $\varrho(X)$ cannot be stable at X. □

Chapter 3

On the convergence of sequential updating methods for P&L attribution

3.1 Introduction

This chapter extends Junike et al. (2024), which is given in Chapter 2, who analyze the case that the P&L can be described by a twice differentiable function and that the risk factors are described by semimartingales with nontrivial interaction effects. They show that the SU decomposition converges in probability if the time steps converge to zero. But in practice, we can only observe one path of the risk factors. Based on path-wise integration methodology, see Karandikar (1995), we show that it is possible to choose a path-dependent non-equidistant time grid, such that the SU decomposition converges to the ISU decomposition almost surely.

Another open question is the order of convergence of the OAT, SU and ASU decompositions: How fast does the decomposition converge when the grid size gets smaller? We analyse the convergence of the decompositions depending on the choice of the path-dependent partition of the time horizon theoretically and undermine it with numerical experiments.

Hence we make the following contributions: First, we show that there is a sequence of path-dependent time grids, such that the SU decomposition converges almost surely. Additionally, we specify how to choose the time grid, such that the SU decomposition converges with a desired order of convergence. The results can be extended to the OAT and ASU decomposition.

3.2 Notation

We use the same notation and definitions as in Junike et al. (2024) and specify further relevant definitions: For $X \in \mathcal{X}^d$ and a stopping time s and a vector of stopping times $\tau = (\tau_1, \dots, \tau_d)$, we define stopped semimartingales:

$$X^s = (X_1^s, \dots, X_d^s) \text{ and } X^\tau = (X_1^{\tau_1}, \dots, X_d^{\tau_d}).$$

Let $|\cdot|$ denote the absolute value and $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^d$. Let the infimum of the empty set be infinity, i.e. $\inf\{\emptyset\} := \infty$. For $X \in \mathcal{X}$ and a sequence of finite stopping times $\gamma = (s_l)_{l \in \mathbb{N}_0}$ with $s_0 = 0$ and $s_l < s_{l+1}$ a.s. for all $l \in \mathbb{N}_0$, let

$$X^\gamma = X(0)1_{\{0\}} + \sum_{l=0}^{\infty} X^{s_l} 1_{(s_l, s_{l+1}]}$$

and

$$\delta X^l = (X^{s_{l+1}} - X^{s_l}). \tag{3.2.1}$$

3.3 Convergence of the SU decomposition

For twice differentiable functions, Junike et al. (2024) show the existence of the ISU decomposition as the stochastic limit of SU decompositions. In practice, it is typically only possible to observe a single path of the risk factors. To ensure that the SU decomposition of this path converges, we need the following theorem, which shows the convergence of the SU decomposition in an almost surely sense for general semimartingales.

Theorem 3.3.1.
Let $\pi \in \sigma_d$, $F \in \mathcal{M}(\mathcal{C}_2)$ and $X \in \mathcal{X}^d$. Let $t \geq 0$ and $i \in 1, \dots, d$. Then there exists a sequence $(\gamma_n)_{n \in \mathbb{N}}$ of unbounded random partitions tending to the identity, such that it holds for $n \to \infty$ that

$$\delta_i^{\mathrm{SU},\pi,\gamma_n}(F, X)(t) \xrightarrow{a.s.} \delta_i^{\mathrm{ISU},\pi}(F, X)(t).$$

Proof. The proof of Theorem 3.3.1 can be found in the appendix. □

Since the definitions of the OAT and the ASU decompositions are based on the SU decomposition, see Definition 4.5 and Remark 4.12 in Junike et al. (2024), we immediately get the following result.

Corollary 3.3.2.
Let $F \in \mathcal{M}(\mathcal{C}_2)$ and $X \in \mathcal{X}^d$. Let $t \geq 0$ and $i \in 1, \ldots, d$. Then there exist sequences $(\gamma_n)_{n\in\mathbb{N}}$,$(\mu_n)_{n\in\mathbb{N}}$ of unbounded random partitions tending to the identity, such that it holds for $n \to \infty$ that

$$\delta_i^{\mathrm{OAT},\gamma_n}(F,X)(t) \overset{a.s.}{\to} \delta_i^{\mathrm{IOAT}}(F,X)(t),$$
$$\delta_i^{\mathrm{ASU},\mu_n}(F,X)(t) \overset{a.s.}{\to} \delta_i^{\mathrm{IASU}}(F,X)(t).$$

Proof. The corollary is a direct consequence of Theorem 3.3.1. □

In practice, the ISU, IOAT and IASU decompositions usually need to be approximated numerically and the latter theorem motivates the use of the SU, OAT and ASU decompositions for this purpose. In the next theorem, we show how, in the case of continuous risk factors, the random partition can be chosen such that the error in terms of $E[\sup_{0\leq t\leq T} |\cdot|]$ is smaller than or equal to a given error tolerance $a > 0$. The result can be used for a sequence of error bounds $(a_n)_{n\in\mathbb{N}}$ to achieve convergence of the SU decomposition with a certain order, specified by the choice of $(a_n)_{n\in\mathbb{N}}$. We need the following assumption, cf. p. 112 in Protter (2005).

Assumption 3.3.3.
Let $Z \in \mathcal{X}$ be a continuous semimartingale and $Z = M + A$ be a decomposition of Z into a locally square integrable martingale M and a finite variation process A. We assume that $E[Z^2(T)] < \infty$, $E[[M,M](T)] < \infty$ and that A is of integrable variation, i.e.

$$E\big[\int_0^T |dA(s)|\big] < \infty. \tag{3.3.1}$$

Theorem 3.3.4.
Let $X \in \mathcal{X}^d$ be a d-dimensional continuous semimartingale with X_i, $i = 1, \ldots, d$ satisfying Assumption 3.3.3. Let $F \in \mathcal{M}(\mathcal{C}_2)$ with $F(X)(t) = f(X(t))$ for some $f \in \mathcal{C}_2$ and let f, f_i, f_{ij}, $i,j = 1,\ldots,d$ be Lipschitz continuous with joint Lipschitz constant $L > 0$. For any $a > 0$ let $\gamma = (s_l)_{l\in\mathbb{N}_0}$ be defined by $s_0 = 0$ and

$$\begin{aligned} s_{l+1} = \inf\big\{t \geq s_l : \|X(t) - X^{s_l}(t)\| \geq a \text{ or} \\ |f_{ij}(X(t))X_j(t) - f_{ij}(X^{s_l}(t))X_j^{s_l}(t)| \geq a,\ i,j = 1,\ldots,d\big\} \end{aligned} \tag{3.3.2}$$

for $l \geq 0$. Then for the SU and ISU decomposition with respect to γ and $\pi \in \sigma_d$, there is a constant $C > 0$ such that

$$E\Big[\sup_{0\leq t\leq T} |\delta_i^{SU,\pi,\gamma}(F,X)(t) - \delta_i^{ISU,\pi}(F,X)(t)|\Big] \leq Ca.$$

Proof. The proof of Theorem 3.3.4 is presented in the appendix. □

Corollary 3.3.5.
Under the assumptions of Theorem 3.3.4, there are constants $C_1, C_2 > 0$ such that

$$E\Big[\sup_{0\le t\le T}|\delta_i^{OAT,\gamma}(F,X)(t)-\delta_i^{IOAT}(F,X)(t)|\Big]\le C_1 a$$

and

$$E\Big[\sup_{0\le t\le T}|\delta_i^{ASU,\gamma}(F,X)(t)-\delta_i^{IASU}(F,X)(t)|\Big]\le C_2 a.$$

Proof. The corollary is a direct consequence of Theorem 3.3.4. □

The following example is intended to give an intuition of how the time grid must be selected depending on different market environments in order to fulfil the error bound in Theorem 3.3.4.

Example 3.3.6.
Let $d = 1$ and $X = \sigma B$ for a standard Brownian motion B and $\sigma > 0$. Let f be the identity. The random partition defined in Theorem 3.3.4 is equal to

$$s_{l+1} = \inf\{t \ge s_l : \|X(t) - X^{s_l}(t)\| = a\}$$

and by the optional stopping theorem, see Section 12.4 in Grimmett and Stirzaker (2001), it follows that

$$E[s_1] = \frac{a^2}{\sigma^2}.$$

That is, the higher the variance, the narrower the random partition of the time grid has to be chosen.

The following example illustrates the convergence of the SU decomposition. The order of convergence depends on the error bound according to Theorem 3.3.4.

Example 3.3.7.
Let X_1 be a Brownian motion and $X_2 = \rho X_1 + \sqrt{1-\rho^2}Z$ for a Brownian motion Z, independent of X_1. Let $X = (X_1, X_2)$ and $F(X) = X_1X_2$. The partition defined in Theorem 3.3.4 simplifies to

$$s_{l+1} = \inf\{t \ge s_l : \|X(t) - X^{s_l}(t)\| = a\},$$

since

$$f_{ij}(X) = \begin{cases} 1, & i \neq j \\ 0, & i = j. \end{cases}$$

For a sequence $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$ let

$$\mathrm{err}(n) := E\Big[\sup_{0\leq t\leq T} |\delta_1^{SU,\pi,\gamma_{a_n}}(F,X)(t) - \delta_1^{ISU,\pi}(F,X)(t)|\Big]$$

define the error between the SU and the ISU decomposition of $F(X)$ for some $\pi \in \sigma_d$ with respect to γ_{a_n}, where γ_{a_n} is the random partition defined in Eq. (3.3.2) with respect to a_n.

We simulate X_1 and X_2 with step size $1/1000$ and approximate the ISU decomposition with a SU decomposition with 1000 equidistant time steps and time horizon $T = 1$. In Figure 3.3.7 we can see a log-log plot of $\mathrm{err}(n)$ with respect to $a_n = 1/n^2$ for $n = 10, 20, ...$ for $\rho = 0$ and $\rho = 0.9$, respectively. Using linear regression we determine the slope of the two graphs. In both cases it is larger than 3. This underlines the results of Theorem 3.3.4, which provides a theoretical slope of 2.

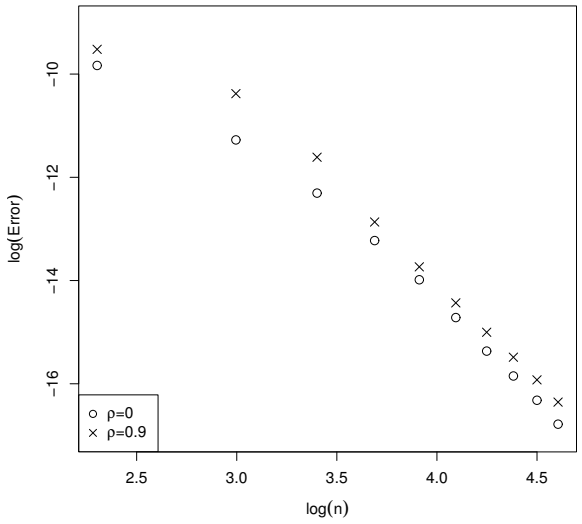

Figure 3.1: Convergence of the SU decomposition with respect to γ_{a_n} of a product of two Brownian motions with correlation parameter $\rho = 0$ and $\rho = 0.9$ with γ_{a_n} as in Eq. (3.3.2) with $a_n = 1/n^2$.

3.4 Conclusion

In this chapter, we analyzed the convergence properties of prevalent sequential updating decompositions. We extended the work of Junike et al. (2024), who only show the convergence in probability of the SU decomposition. We prove that by choosing a path-dependent, non-equidistant time grid, the SU/OAT/ASU decomposition converges almost surely to the ISU/IOAT/IASU decomposition. This extension addresses the limitations of the widespread use of equidistant time grids in practice when only a single path of the risk factors is given. In addition, we provide conditions for bounding the distance between the SU and the ISU decompositions. We illustrate the speed of convergence of the SU decomposition depending on the choice of the non-equidistant time grid in a numerical experiment.

3.A Appendix

Proof of Theorem 5

Proof. Consider two 1-dimensional semimartingales $Z, Y \in \mathcal{X}$ and define the sequence of stopping times $\gamma_n = \{0 = s_0^n < s_1^n < ...\}$ by $s_0^n = 0$ and

$$s_{l+1}^n = \min\left(\inf\left\{t \geq s_l^n : |Z(t) - Z^{s_l^n}(t)| + |(YZ)(t) - (YZ)^{s_l^n}(t)| \geq 2^{-n}\right\}, s_l^n + \frac{1}{n}\right).$$

Note that the condition $\min(\cdot, s_l^n + \frac{1}{n})$ is only necessary to make $(\gamma_n)_{n\in\mathbb{N}_0}$ tend to the identity, if X is constant on some subinterval. Theorem 2 in Karandikar (1995) shows that

$$\sum_{l=0}^{\infty} Z^{s_l^n}(t)\big(Y^{s_{l+1}^n}(t) - Y^{s_l^n}(t)\big) \overset{a.s.}{\to} \int_0^t Z(s-)dY(s)$$

for $n \to \infty$ and each $t \geq 0$. Similarly, we can conclude that

$$\begin{aligned}
&\sum_{l=0}^{\infty} Z^{s_l^n}(t)\big(Y^{s_{l+1}^n}(t) - Y^{s_l^n}(t)\big)^2 \\
&= \sum_{l=0}^{\infty} Z^{s_l^n}(t)\big((Y^{s_{l+1}^n})^2(t) - (Y^{s_l^n})^2(t)\big) - 2\sum_{l=0}^{\infty} Z^{s_l^n}(t)Y^{s_l^n}(t)\big(Y^{s_{l+1}^n}(t) - Y^{s_l^n}(t)\big)
\end{aligned}$$

converges almost surely to

$$\int_0^t Z(s-)dY^2(s) - 2\int_0^t Z(s-)Y(s-)dY(s) = \int_0^t Z(s-)d[Y,Y](s)$$

for $n \to \infty$. Similarly, for two semimartingales Y_1, Y_2 and a corresponding choice of γ_n, one can also show that

$$\sum_{l=0}^{\infty} Z^{s_l^n}(t)\big(Y_1^{s_{l+1}^n}(t) - Y_1^{s_l^n}(t)\big)\big(Y_2^{s_{l+1}^n}(t) - Y_2^{s_l^n}(t)\big) \overset{a.s.}{\to} \int_0^t Z(s-)d[Y_1, Y_2](s)$$

using the polarization identity

$$[Y_1, Y_2] = \frac{1}{2}([Y_1 + Y_2, Y_1 + Y_2] - [Y_1, Y_1] - [Y_2, Y_2]).$$

Theorem 4.7 in Junike et al. (2024) shows that the limit of SU decompositions exists for $F \in \mathcal{M}(\mathcal{C}_2)$ and is equal to the ISU decomposition. The proof uses a pathwise argumentation, except that Eq. (32) in their work only states the convergence in probability of step functions to Itô integrals. The above almost surely convergence results show that there is a sequence of unbounded random partitions $(\gamma_n)_{n\in\mathbb{N}}$, so that the SU decomposition with respect to $(\gamma_n)_{n\in\mathbb{N}}$ converges almost surely to the ISU decomposition. □

Proof of Theorem 3.3.4

Proof. We use the same ideas as in the proof of Theorem 2 in Karandikar (1995), but several new arguments are also needed. In the following we use the notation as in Eq. (3.2.1) to describe discrete increments. Let γ be defined as in Eq. (3.3.2). By the Taylor expansion, see Lemma 3.A.4, the SU decomposition with respect to γ is given by

$$\begin{aligned}\delta_i^{SU,\pi,\gamma}(F, X) = \sum_{l=0}^{\infty} \Big(& f_i(X^{s_l})\delta X_i^l + \frac{1}{2} f_{ii}(X^{s_l})(\delta X_i^l)^2 \\ & + \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} f_{ij}(X^{s_l})\delta X_i^l \delta X_j^l + R_l^{1,i} - R_l^{2,i}\Big)\end{aligned} \tag{3.A.1}$$

with

$$R_l^{1,i} = \frac{1}{2} \sum_{\substack{j=1 \\ \pi(j)\leq\pi(i)}}^{d} (f_{ij}(\xi_l^{1,i}) - f_{ij}(X^{s_l}))\delta X_i^l \delta X_j^l,$$

$$R_l^{2,i} = \frac{1}{2} \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} (f_{ij}(\xi_l^{2,i}) - f_{ij}(X^{s_l}))\delta X_i^l \delta X_j^l,$$

and

$$\xi_l^{1,i} = \theta_l^1 \left(X^{s_l} + p_{\{j\,|\pi(j)\leq\pi(i)\}}\,(X^{s_{l+1}} - X^{s_l})\right) + (1-\theta_l^1)X^{s_l}, \quad \theta_l^1 \in [0,1],$$

$$\xi_l^{2,i} = \theta_l^2 \left(X^{s_l} + p_{\{j\,|\pi(j)<\pi(i)\}}\,(X^{s_{l+1}} - X^{s_l})\right) + (1-\theta_l^2)X^{s_l}, \quad \theta_l^2 \in [0,1].$$

First, we analyse the remainder. For a single summand of $R_l^{1,i}$ it holds that

$$\begin{aligned}|f_{ij}(\xi_l^{1,i}(t)) - f_{ij}(X^{s_l}(t))| &\leq L\|\xi_l^{1,i}(t) - X^{s_l}(t)\| \\ &\leq L\|X^{s_l}(t) + p_{\{j\,|\pi(j)\leq\pi(i)\}}\left(X^{s_{l+1}} - X^{s_l}\right)(t) - X^{s_l}(t)\| \\ &\leq L\|X^{s_{l+1}}(t) - X^{s_l}(t)\| \\ &\leq La\end{aligned}$$

and analogously

$$|f_{ij}(\xi_l^{2,i}(t)) - f_{ij}(X^{s_l}(t))| \leq La.$$

Therefore we obtain

$$\begin{aligned}&E\Big[\sup_{0\leq t\leq T}|\sum_{l=0}^{\infty} R_l^{1,i}(t) - R_l^{2,i}(t)|\Big] \\ \leq\ &E\Big[\sup_{0\leq t\leq T}\sum_{l=0}^{\infty}|R_l^{1,i}(t)| + |R_l^{2,i}(t)|\Big] \\ \leq\ &La\sum_{\substack{j=1\\ \pi(j)\leq\pi(i)}}^{d} E\Big[\sup_{0\leq t\leq T}\sum_{l=0}^{\infty}|\delta X_i^l(t)\delta X_j^l(t)|\Big].\end{aligned}$$

For all $b, c \in \mathbb{R}$, it holds that $2bc = (b+c)^2 - b^2 - c^2$, hence

$$|bc| \leq \frac{1}{2}\left((b+c)^2 + b^2 + c^2\right). \tag{3.A.2}$$

Using Eq. (3.A.2) and Corollary 3.A.3 we obtain

$$\begin{aligned}&E\Big[\sup_{0\leq t\leq T}\Big|\sum_{l=0}^{\infty} R_l^{1,i}(t) - R_l^{2,i}(t)\Big|\Big] \\ \leq\ &\frac{1}{2}La\sum_{\substack{j=1\\ \pi(j)\leq\pi(i)}}^{d} E\Big[\sup_{0\leq t\leq T}\sum_{l=0}^{\infty}(\delta X_i^l(t) + \delta X_j^l(t))^2 + (\delta X_i^l(t))^2 + (\delta X_j^l(t))^2\Big] \\ \leq\ &\frac{1}{2}La\sum_{\substack{j=1\\ \pi(j)\leq\pi(i)}}^{d} E_{ij} \\ =\ &Ca\end{aligned}$$

with

$$\begin{aligned}E_{ij} :=&6E\Big[X_i^2(T)\Big] - 6E\Big[X_i^2(0)\Big] - 3E\Big[[M_i, M_i](T)\Big] + 6a\Big(\sqrt{E\Big[[M_i, M_i](T)\Big]} \\ &+ E\Big[\int_0^T |dA_i|\Big]\Big) + 6E\Big[X_j^2(T)\Big] - 6E\Big[X_j^2(0)\Big] - 3E\Big[[M_j, M_j](T)\Big]\end{aligned}$$

$$+ 6a\Big(\sqrt{E\Big[[M_j, M_j](T)\Big]} + E\Big[\int_0^T |dA_j|\Big]\Big) + 6a^2$$

and $C = \frac{1}{2}L\sum_{\pi(j)\leq\pi(i)} E_{ij} > 0$.

Now we treat the remaining sums in Eq. (3.A.1). Let

$$H_i = \sum_{l=0}^{\infty} f_i(X^{s_l})\delta X_i^l \quad \text{and} \quad H_{ij} = \sum_{l=0}^{\infty} f_{ij}(X^{s_l})\delta X_i^l \delta X_j^l.$$

By Lemma 3.A.1 and 3.A.2 it follows that

$$E\Big[\sup_{0\leq t\leq T}\Big|H_i(t) - \int_0^t f_i(X(s))dX_i(s)\Big|\Big] \leq C_i a$$

and

$$E\Big[\sup_{0\leq t\leq T}\Big|H_{ij}(t) - \int_0^t f_{ij}(X(s))d[X_i, X_j](s)\Big|\Big] \leq C_{ij} a$$

for some $C_i, C_{ij} > 0$, $i, j = 1, \ldots, d$. Then, the SU decomposition can be written as

$$\delta_i^{SU,\pi,\gamma}(F, X) = H_i + \frac{1}{2}H_{ii} + \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} H_{ij} + \sum_{l=0}^{\infty}(R_l^{1,i} - R_l^{2,i}).$$

The ISU decomposition is given by

$$\begin{aligned}\delta_i^{ISU,\pi}(F, X) = &\int_0^t f_i(X(s))dX_i(s) + \frac{1}{2}\int_0^t f_{ii}(X(s))d[X_i, X_i](s) \\ &+ \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} \int_0^t f_{ij}(X(s))d[X_i, X_j](s),\end{aligned}$$

see Definition 3.8 in Junike et al. (2024). Therefore we obtain

$$\begin{aligned}&E\Big[\sup_{0\leq t\leq T}\Big|\delta_i^{SU,\pi,\gamma}(F, X)(t) - \delta_i^{ISU,\pi}(F, X)(t)\Big|\Big] \\ \leq\ &E\Big[\sup_{0\leq t\leq T}\Big|H_i(t) - \int_0^t f_i(X(s))dX_i(s)\Big|\Big] \\ &+ \frac{1}{2}E\Big[\sup_{0\leq t\leq T}\Big|H_{ii}(t) - \int_0^t f_{ii}(X(s))d[X_i, X_i](s)\Big|\Big] \\ &+ \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} E\Big[\sup_{0\leq t\leq T}\Big|H_{ij}(t) - \int_0^t f_{ij}(X(s))d[X_i, X_j](s)\Big|\Big] \\ &+ E\Big[\sup_{0\leq t\leq T}\Big|\sum_{l=0}^{\infty} R_l^{1,i}(t) - R_l^{2,i}(t)\Big|\Big] \\ \leq\ &(C_i + \frac{1}{2}C_{ii} + \sum_{\substack{j=1 \\ \pi(j)<\pi(i)}}^{d} C_{ij} + C)a.\end{aligned}$$

□

Auxiliary results

Lemma 3.A.1.
Let $X \in \mathcal{X}^d$ be a d-dimensional semimartingale. Let $Z \in \mathcal{X}$ be a 1-dimensional semimartingale satisfying Assumption 3.3.3. Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz continuous with constant $L > 0$. For $a > 0$, let $\gamma := (s_l)_{l \in \mathbb{N}_0}$ be defined by $s_0 = 0$ and

$$s_{l+1} = \inf\{t \geq s_l : \|X(t) - X^{s_l}(t)\| \geq a\}$$

for $l \geq 0$. Let Y be defined as

$$Y = \sum_{l=0}^{\infty} f(X^{s_l}) \delta Z^l.$$

Then there is a constant $C > 0$, such that

$$E\Big[\sup_{0 \leq t \leq T} \Big| Y(t) - \int_0^t f(X(s-)) dZ(s) \Big| \Big] \leq Ca.$$

Proof. The proof is based on ideas of Theorem 2 in Karandikar (1995). It holds that

$$Y(t) = \int_0^t f(X^\gamma(s)) dZ(s).$$

By the choice of γ and the Lipschitz continuity of f, we have

$$\sup_{0 \leq t \leq T} |f(X^\gamma(t)) - f(X(t-))| \leq La.$$

Let $Z = M + A$ be the decomposition of Z into a locally square integrable martingale M with localizing sequence of stopping times $(\sigma_n)_{n \in \mathbb{N}}$ and a finite variation process A, see Theorem 1 of Section III in Protter (2005). We obtain that

$$\begin{aligned} & E\Big[\sup_{0 \leq t \leq T} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dZ(s) \Big| \Big] \\ \leq\ & E\Big[\sup_{0 \leq t \leq T} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dM(s) \Big| \Big] \\ & + E\Big[\sup_{0 \leq t \leq T} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dA(s) \Big| \Big]. \end{aligned} \tag{3.A.3}$$

First we treat the martingale part. It follows by Jensen's inequality, the monotonicity of $x \mapsto x^2$, $x \geq 0$ and by the monotone convergence theorem that

$$\begin{aligned} & \Big(E\Big[\sup_{0 \leq t \leq T} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dM(s) \Big| \Big] \Big)^2 \\ \leq\ & E\Big[\sup_{0 \leq t \leq T} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dM(s) \Big|^2 \Big] \end{aligned}$$

$$= \lim_{n\to\infty} E\Big[\sup_{0\le t\le \min(\sigma_n,T)} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dM(s)\Big|^2\Big]$$
$$\le \limsup_{n\to\infty} E\Big[\sup_{0\le t\le \min(\sigma_n,T)} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dM(s)\Big|^2\Big].$$

Furthermore, similar to Eq. (4) in Karandikar (1995) and by using the Lipschitz continuity, the latter term is bounded by

$$4 \limsup_{n\to\infty} E\Big[\int_0^{\min(\sigma_n,T)} (f(X^\gamma(s)) - f(X(s-)))^2\, d[M,M](s)\Big]$$
$$\le 4L^2a^2 \limsup_{n\to\infty} E[[M,M](\min(\sigma_n,T))]$$
$$= 4L^2a^2 E[[M,M](T)]. \tag{3.A.4}$$

For the finite variation part, it holds that

$$E\Big[\sup_{0\le t\le T} \Big| \int_0^t (f(X^\gamma(s)) - f(X(s-)))\, dA(s)\Big|\Big]$$
$$\le E\Big[\sup_{0\le t\le T} \int_0^t |f(X^\gamma(s)) - f(X(s-))||dA(s)|\Big] \tag{3.A.5}$$
$$\le LaE\Big[\sup_{0\le t\le T} \int_0^t |dA(s)|\Big]$$
$$\le LaE\Big[\int_0^T |dA(s)|\Big], \tag{3.A.6}$$

where $E[\int_0^T |dA(s)|] < \infty$ by Assumption 3.3.3. We obtain the required result from Eqs. (3.A.4) and (3.A.6) by setting

$$C = 2L\sqrt{E[[M,M](T)]} + LE[\int_0^T |dA(s)|]).$$

□

Lemma 3.A.2.

Let $Z_1, Z_2 \in \mathcal{X}$ be semimartingales satisfying Assumption 3.3.3 and let $X \in \mathcal{X}^d$ be a d-dimensional semimartingale. Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz continuous with constant $L > 0$. Let $a > 0$. Let $\gamma := (s_l)_{l\in\mathbb{N}_0}$ be defined by $s_0 = 0$ and

$$s_{l+1} = \inf\{t \ge s_l : \|X(t) - X^{s_l}(t)\| \ge a \text{ or}$$
$$|f(X(t))Z_i(t) - f(X^{s_l}(t))Z_i^{s_l}(t)| \ge a, \quad i = 1,2\}.$$

for $l \ge 0$. Let Y be defined as

$$Y = \sum_{l=0}^{\infty} f(X^{s_l})(Z_1^{s_{l+1}} - Z_1^{s_l})(Z_2^{s_{l+1}} - Z_2^{s_l}).$$

Then there is a constant $C > 0$ such that

$$E\Big[\sup_{0\le t\le T}\Big|Y(t) - \int_0^t f(X(s-))d[Z_1, Z_2](s)\Big|\Big] \le Ca.$$

Proof. Since

$$\begin{aligned} Y = &\sum_{l=0}^{\infty} f(X^{s_l})(Z_1^{s_{l+1}} Z_2^{s_{l+1}} - Z_1^{s_l} Z_2^{s_l}) \\ &- \sum_{l=0}^{\infty} f(X^{s_l}) Z_1^{s_l}(Z_2^{s_{l+1}} - Z_2^{s_l}) \\ &- \sum_{l=0}^{\infty} f(X^{s_l}) Z_2^{s_l}(Z_1^{s_{l+1}} - Z_1^{s_l}), \end{aligned}$$

it holds that

$$\begin{aligned} Y(t) = &\int_0^t f(X^\gamma(s))d(Z_1Z_2)(s) - \int_0^t f(X^\gamma(s))Z_1^\gamma(s)dZ_2(s) \\ &- \int_0^t f(X^\gamma(s))Z_2^\gamma(s)dZ_1(s). \end{aligned} \tag{3.A.7}$$

Since the product of semimartingales is again a semimartingale, we can apply Lemma 3.A.1 and obtain

$$E\Big[\sup_{0\le t\le T}\Big|\int_0^t \left(f(X^\gamma(s)) - f(X(s-))\right) d(Z_1Z_2)(s)\Big|\Big] \le C_1 a$$

for a $C_1 > 0$. By the choice of γ, it holds that

$$\sup_{0\le t\le T} |f(X^\gamma(t))Z_i^\gamma(t) - f(X(t-))Z_i(t-)| \le a, \quad i = 1, 2.$$

Analogously to the proof of Lemma 3.A.1, we obtain

$$E\Big[\sup_{0\le t\le T}\Big|\int_0^t \left(f(X^\gamma(s))Z_1^\gamma(s) - f(X(s-))Z_1(s-)\right) dZ_2(s)\Big|\Big] \le C_2 a$$

and

$$E\Big[\sup_{0\le t\le T}\Big|\int_0^t \left(f(X^\gamma(s))Z_2^\gamma(s) - f(X(s-))Z_2(s-)\right) dZ_1(s)\Big|\Big] \le C_3 a$$

for suitable constants $C_2, C_3 > 0$. Altogether, since $d[Z_1, Z_2] = d(Z_1Z_2) - Z_1dZ_2 - Z_2dZ_1$, it follows that

$$\begin{aligned} &E\Big[\sup_{0\le t\le T}\Big|Y(t) - \int_0^t f(X(s-))d[Z_1, Z_2](s)\Big|\Big] \\ \le\ &E\Big[\sup_{0\le t\le T}\Big|\int_0^t \left(f(X^\gamma(s)) - f(X(s-))\right) d(Z_1Z_2)(s)\Big|\Big] \end{aligned}$$

$$
\begin{aligned}
&+ E\Big[\sup_{0\le t\le T}\Big|\int_0^t (f(X^\gamma(s))Z_1^\gamma(s) - f(X(s-))Z_1(s-))\,dZ_2(s)\Big|\Big] \\
&+ E\Big[\sup_{0\le t\le T}\Big|\int_0^t (f(X^\gamma(s))Z_2^\gamma(s) - f(X(s-))Z_2(s-))\,dZ_1(s)\Big|\Big] \\
&\le Ca
\end{aligned}
$$

for $C = C_1 + C_2 + C_3$. □

Corollary 3.A.3.
Let $X \in \mathcal{X}^d$ be a d-dimensional continuous semimartingale with X_i, $i = 1, \dots, d$ satisfying Assumption 3.3.3. Let $\gamma = (s_l)_{l\in\mathbb{N}_0}$ be an unbounded random partition. Then it holds that

$$
\begin{aligned}
&E\Big[\sup_{0\le t\le T}\sum_{l=0}^{\infty}(\delta X_i^l(t) + \delta X_j^l(t))^2 + (\delta X_i^l(t))^2 + (\delta X_j^l(t))^2\Big] \\
&\le 6E\Big[X_i^2(T)\Big] - 6E\Big[X_i^2(0)\Big] - 3E\Big[[M_i,M_i](T)\Big] + 6a\Big(\sqrt{E\Big[[M_i,M_i](T)\Big]} \\
&\quad + E\Big[\int_0^T |dA_i|\Big]\Big) + 6E\Big[X_j^2(T)\Big] - 6E\Big[X_j^2(0)\Big] - 3E\Big[[M_j,M_j](T)\Big] \\
&\quad + 6a\Big(\sqrt{E\Big[[M_j,M_j](T)\Big]} + E\Big[\int_0^T |dA_j|\Big]\Big) + 6a^2. \qquad (3.A.8)
\end{aligned}
$$

Proof. For $\omega \in \Omega$ let $m(\omega) = k$, if $s_k(\omega) < t \le s_{k+1}(\omega)$. It follows that

$$
\begin{aligned}
&E\Big[\sup_{0\le t\le T}\sum_{l=0}^{\infty}(\delta X_i^l(t) + \delta X_j^l(t))^2 + (\delta X_i^l(t))^2 + (\delta X_j^l(t))^2\Big] \\
&\le E\Big[\sup_{0\le t\le T}\sum_{l=0}^{m-1}(\delta X_i^l(t) + \delta X_j^l(t))^2 + (\delta X_i^l(t))^2 + (\delta X_j^l(t))^2\Big] \\
&\quad + E\Big[\sup_{0\le t\le T}(\delta X_i^m(t) + \delta X_j^m(t))^2 + (\delta X_i^m(t))^2 + (\delta X_j^m(t))^2\Big] \\
&\le E\Big[\sup_{0\le t\le T}\sum_{l=0}^{\infty}(\delta X_i^l(T) + \delta X_j^l(T))^2 + (\delta X_i^l(T))^2 + (\delta X_j^l(T))^2\Big] \\
&\quad + E\Big[\sup_{0\le t\le T}(\delta X_i^m(t) + \delta X_j^m(t))^2 + (\delta X_i^m(t))^2 + (\delta X_j^m(t))^2\Big]. \qquad (3.A.9)
\end{aligned}
$$

Now we analyse the first term of Eq. (3.A.9). Using $(a+b)^2 \le 2(a^2+b^2)$ for $a, b \in \mathbb{R}$, we obtain

$$
\begin{aligned}
&E\Big[\sup_{0\le t\le T}\sum_{l=0}^{\infty}(\delta X_i^l(T) + \delta X_j^l(T))^2 + (\delta X_i^l(T))^2 + (\delta X_j^l(T))^2\Big] \\
&\le 3E\Big[\sum_{l=0}^{\infty}(\delta X_i^l(T))^2\Big] + 3E\Big[\sum_{l=0}^{\infty}(\delta X_j^l(T))^2\Big]. \qquad (3.A.10)
\end{aligned}
$$

For the first part we obtain

$$\begin{aligned}
&E\Big[\sum_{l=0}^{\infty}(\delta X_i^l(T))^2] \\
&= E\Big[\sum_{l=0}^{\infty}(X_i^{s_{l+1}}(T))^2 - (X_i^{s_l}(T))^2 - 2X_i^{s_l}(T)(X_i^{s_{l+1}}(T) - X_i^{s_l}(T))\Big] \\
&\leq E\Big[\int_0^T 1 dX_i^2\Big] + 2E\Big[\int_0^T X_i^\gamma dX_i\Big] \\
&\leq E\Big[X_i^2(T)\Big] - E\Big[X_i^2(0)\Big] + 2E\Big[\int_0^T (X_i^\gamma - X_i) dX_i\Big] + 2E\Big[\int_0^T X_i dX_i\Big]. \qquad (3.A.11)
\end{aligned}$$

Further, we analyse the third term of Eq. (3.A.11). It holds that

$$E\Big[\int_0^T (X_i^\gamma - X_i) dX_i\Big] \leq E\Big[\int_0^T (X_i^\gamma - X_i) dM_i\Big] + E\Big[\int_0^T (X_i^\gamma - X_i) dA_i\Big],$$

where it follows for the first part by the Itô isometry that

$$\begin{aligned}
&E\Big[\int_0^T (X_i^\gamma - X_i) dM_i\Big]^2 \\
&\leq E\Big[(\int_0^T (X_i^\gamma - X_i) dM_i)^2\Big] \\
&\leq E\Big[\int_0^T (X_i^\gamma - X_i)^2 d[M_i, M_i]\Big] \\
&\leq a^2 E\Big[[M_i, M_i](T)\Big]
\end{aligned}$$

and for the second part that

$$E\Big[\int_0^T (X_i^\gamma - X_i) dA_i\Big] \leq aE\Big[\int_0^T |dA_i|\Big].$$

For the other terms in Eq. (3.A.11) it holds that

$$2E\Big[\int_0^T X_i dX_i\Big] + E\Big[X_i^2(T)\Big] - E\Big[X_i^2(0)\Big] = 2E\Big[X_i^2(T)\Big] - 2E\Big[X_i^2(0)\Big] - E\Big[[X_i, X_i](T)\Big].$$

Since X_i has continuous paths, M_i and A_i are also continuous and $[X_i, X_i] = [M_i, M_i]$ a.s., see p. 131 in Protter (2005). In total, for the first term of Eq. (3.A.10) it follows that

$$\begin{aligned}
3E\Big[\sum_{l=0}^{\infty}(\delta X_i^l(T))^2\Big] \leq& 6E\Big[X_i^2(T)\Big] - 6E\Big[X_i^2(0)\Big] - 3E\Big[[M_i, M_i](T)\Big] \\
&+ 6a\Big(\sqrt{E\Big[[M_i, M_i](T)\Big]} + E\Big[\int_0^T |dA_i|\Big]\Big).
\end{aligned}$$

The second part on the right-hand side of Eq. (3.A.10) can be treated equivalently. For the second part of Eq. (3.A.9) we have that

$$\begin{aligned} & E\left[\sup_{0\leq t\leq T}(\delta X_i^m(t)+\delta X_j^m(t))^2+(\delta X_i^m(t))^2+(\delta X_j^m(t))^2\right] \\ \leq\ & E\left[\sup_{0\leq t\leq T}4a^2+a^2+a^2\right] \\ =\ & 6a^2, \end{aligned}$$

which leads to Eq. (3.A.8). □

Lemma 3.A.4.

(Taylor's theorem). Let $U \subset \mathbb{R}^d$ be open. Let $x, a \in U$ such that $\lambda x + (1-\lambda)a \in U$ for all $\lambda \in [0,1]$. Let $f : U \to \mathbb{R}$ be twice continuously differentiable. Then it holds that

$$f(x) = f(a) + \sum_{h=1}^{d} f_h(a)\,(x_h - a_h) + \frac{1}{2}\sum_{h,j=1}^{d} f_{hj}(a)\,(x_h - a_h)\,(x_j - a_j) + R,$$

where the remainder R can be expressed for some $\theta \in [0,1]$ and $\xi = \theta x + (1-\theta)a$ by

$$R = \frac{1}{2}\sum_{h,j=1}^{d} \left(f_{hj}(\xi) - f_{hj}(a)\right)(x_h - a_h)\,(x_j - a_j)\,.$$

Proof. Theorem 2 of Section 7 in Forster (2017) □

Chapter 4

From characteristic functions to multivariate distribution functions and European option prices by the damped COS method

Abstract

We provide a unified framework to obtain numerically certain quantities, such as the distribution function, absolute moments and prices of financial options, from the characteristic function of some (unknown) probability density function using the Fourier-cosine expansion (COS) method. The classical COS method is numerically very efficient in one-dimension, but it cannot deal very well with certain integrands in general dimensions. Therefore, we introduce the damped COS method, which can handle a large class of integrands very efficiently. We prove the convergence of the (damped) COS method and study its order of convergence. The method converges exponentially if the characteristic function decays exponentially. To apply the (damped) COS method, one has to specify two parameters: a truncation range for the multivariate density and the number of terms to approximate the truncated density by a cosine series. We provide an explicit formula for the truncation range and an implicit formula for the number of terms. Numerical experiments up to five dimensions confirm the theoretical results.

Keywords: Fourier-transform; numerical integration; inversion theorem; COS method; CDF; option pricing

4.1 Introduction

We aim to solve the following integral numerically:

$$\int_{\mathbb{R}^d} w(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}. \tag{4.1.1}$$

The function g is usually a density and the function w is called *function of interest.* Integrals as in (4.1.1) appear in a wide range of applications: The integral is equal to the cumulative distribution function (CDF) of the density g if w is an indicator function. CDFs appear in many scientific disciplines.

If w is the absolute value, the integral describes the absolute moment of the density g, which plays an important role in various disciplines but is not easy to obtain, see Von Bahr (1965), Brown (1972), Barndorff-Nielsen and Stelzer (2005) and references therein.

In a financial context, the function w might also describe some financial contract, which depends on several assets. The function g is then the density of the logarithmic returns of the assets and the integral describes the price of the contract.

In many cases, the precise structure of g is unknown, but the Fourier transform $\widehat{g}$ is often given in closed form. For example: while the joint density of the sum of two independent random variables can only be expressed as a convolution and is usually not given explicitly, the joint characteristic function is much simpler to obtain (it is just the product of the marginal characteristic functions). Moreover, the characteristic function of a Lévy process at a particular time-point is usually given explicitly thanks to the Lévy-Khinchin formula.

The integral in (4.1.1) can be solved numerically using various techniques, including (quasi) Monte Carlo simulation, numerical quadrature and Fourier inversion. Special Fourier-inversion methods exist in the case of a CDF, e.g., the Gil-Pelaez formula, see Gil-Pelaez (1951) and extensions, e.g., Schorr (1975), Waller et al. (1995), Hughett (1998) in one-dimension and Shephard (1991a), Shephard (1991b) in d dimensions.

The COS method, see Fang and Oosterlee (2009a) for $d = 1$ and Ruijter and Oosterlee (2012) for $d = 2$, is a Fourier inversion technique. The COS method has been applied extensively in computational finance, see Fang and Oosterlee (2009b), Fang and Oosterlee (2011), Grzelak and Oosterlee (2011), Zhang and Oosterlee (2013), Leitao et al. (2018), Liu et al. (2019a), Liu et al. (2019b), Oosterlee and Grzelak (2019) and Bardgett et al. (2019).

The COS method has also been applied to solve backward stochastic differential equations, see Ruijter and Oosterlee (2015) and Andersson et al. (2023).

Other Fourier techniques to solve integrals as in (4.1.1) can be found in Carr and Madan (1999), Lord et al. (2008), Eberlein et al. (2010), Ortiz-Gracia and Oosterlee (2013), Ortiz-Gracia and Oosterlee (2016) and Bayer et al. (2024). The COS method compares favorably to other Fourier inversion techniques, see Fang and Oosterlee (2009a).

The main idea of the COS method is to truncate the integration range in (4.1.1) to some finite hypercube and to approximate the density g on the finite truncation range by a classical Fourier-cosine expansion. Formulas for the truncation range and the number of terms of the Fourier cosine-series are given in Junike and Pankrashkin (2022) and in Junike (2024) for the one-dimensional case.

There is a clever trick to approximate the cosine coefficients for g in a very fast and robust way using $\widehat{g}$. The COS method is particularly fast when the Fourier-cosine coefficients of the function of interest w are given analytically. For instance, in multivariate dimensions, the Fourier-cosine coefficients of a CDF can be obtained analytically. However, in many cases, the Fourier-cosine coefficients are not given in closed form. Ruijter and Oosterlee (2012) propose in these cases to obtain the Fourier coefficients of the function of interest numerically by a discrete cosine transform, but this approach slows the COS method significantly. In this article, we introduce the *damped COS method*, which is able to avoid the expensive application of the discrete cosine transform if the Fourier transform of the function of interest is given in closed form. The Fourier transform $\widehat{w}$ is often known, e.g., if w describes a CDF, an absolute moment or a financial contract, see Section 4.5. The main idea is to damp w by multiplying it by an exponential function in order to make the damped function of interest integrable. The idea of introducing a damping factor dates back at least to Carr and Madan (1999). In multivariate dimensions, damping is also applied by Eberlein et al. (2010), where the optimal damping factor is determined by Bayer et al. (2023).

In moderate dimensions, the COS method is a fast, robust and straightforward-to-implement alternative to the d-dimensional Gil-Pelaez formula, see Shephard (1991a), Shephard (1991b) or the multivariate Lewis formula, see Eberlein et al. (2010), in particular if g is smooth and has semi-heavy tails. A useful feature of the COS method is the fact that in important cases all parameters necessary to tune the method can be obtained directly given some error tolerance.

This article makes the following main contributions: We prove the convergence of the multidimensional (damped) COS method, we analyze the order of convergence of the (damped) COS method, and we provide explicit and implicit formulas for the truncation range and the number of terms, respectively. The new approach to find the number of terms in d dimensions is completely distinct from the one-dimensional case discussed in Junike (2024). Unlike Ruijter and Oosterlee (2012), who analyze the classical COS method, we include in our analysis numerical uncertainty on the characteristic function $\widehat{g}$ and on the Fourier-cosine coefficients of the function of interest. This makes it possible to understand how approximations on $\widehat{g}$ and the Fourier-cosine coefficients of the function of interest affect the total error of the COS method.

This article is structured as follows: In Section 4.2 we fix some notation. In Section 4.3 we introduce the multidimensional (damped) COS method, prove its convergence, analyze the order of convergence and provide explicit and implicit formulas for the truncation range and the number of terms. In Section 4.4 we discuss some examples for g and $\widehat{g}$. In Section 4.5 we discuss some functions of interest, i.e., examples for w. In Section 4.6 we provide numerical experiments. Section 4.7 concludes.

4.2 Notation

Let $d \in \mathbb{N}$. Let $\mathcal{L}^1$ and $\mathcal{L}^2$ denote the sets of integrable and square integrable functions from $\mathbb{R}^d$ to $\mathbb{R}$ and by $\langle ., . \rangle$ and $\|.\|_2$ we denote the scalar product and the (semi)norm on $\mathcal{L}^2$, respectively. The supremum norm of a function $g : \mathbb{R}^d \to \mathbb{C}$ is defined by $\|g\|_\infty := \sup_{\boldsymbol{x} \in \mathbb{R}^d} |g(\boldsymbol{x})|$. By $\Re\{z\}$ and $\Im\{z\}$ we denote the real and imaginary parts of a complex number $z \in \mathbb{C}$. The complex unit is denoted by i. By Γ, we denote the Gamma function. The Euclidean norm and the maximum norm on $\mathbb{R}^d$ are denoted by $|.|$ and by $|.|_\infty$, respectively. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ we define

$$\boldsymbol{x} \geq \boldsymbol{y} :\Leftrightarrow x_1 \geq y_1, ..., x_d \geq y_d$$

and treat "$\leq$", "$<$", "$>$", "$=$" and "$\neq$" similarly. We set $\mathbb{R}^d_+ := \{\boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} > \boldsymbol{0}\}$. For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} \leq \boldsymbol{b}$, two complex vectors $\boldsymbol{z}, \boldsymbol{y} \in \mathbb{C}^d$ we define $\boldsymbol{z} + \boldsymbol{y} := (z_1 + y_1, ..., z_d + y_d) \in \mathbb{C}^d$ and treat $\boldsymbol{zy}$ and $\frac{\boldsymbol{z}}{\boldsymbol{y}}$ similarly. For $\lambda \in \mathbb{C}$, we further define

$$\begin{aligned}
\boldsymbol{z} \cdot \boldsymbol{y} :=& z_1 y_1 + ... + z_d y_d \in \mathbb{C}, \\
\lambda \boldsymbol{z} :=& (\lambda z_1, ..., \lambda z_d) \in \mathbb{C}^d, \\
[\boldsymbol{a}, \boldsymbol{b}] :=& [a_1, b_1] \times ... \times [a_d, b_d] \subset \mathbb{R}^d, \\
(-\boldsymbol{\infty}, \boldsymbol{b}] :=& (-\infty, b_1] \times ... \times (-\infty, b_d] \subset \mathbb{R}^d,
\end{aligned}$$

$$\exp(\boldsymbol{x}) := (\exp(x_1), ..., \exp(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d,$$
$$\log(\boldsymbol{x}) := (\log(x_1), ..., \log(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d_+.$$

For a subset $A \subset \mathbb{R}^d$, we define the indicator function $1_A(\boldsymbol{x})$ by one if $\boldsymbol{x} \in A$ and by zero otherwise. Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $\boldsymbol{N} = (N_1, ..., N_d) \in \mathbb{N}_0^d$ and a sequence $(a_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathbb{N}_0^d} \subset \mathbb{C}$, we define

$$\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}{}' a_{\boldsymbol{k}} := \sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}} \frac{1}{2^{\Lambda(\boldsymbol{k})}} a_{\boldsymbol{k}},$$

where $\Lambda(\boldsymbol{k})$ is the number of components of the vector $\boldsymbol{k}$ that are equal to zero, i.e., $\Lambda(\boldsymbol{k}) := \sum_{h=1}^d 1_{\{0\}}(k_h)$. For an integrable function $g : \mathbb{R}^d \to \mathbb{C}$ we define its *Fourier transform* by

$$\widehat{g}(\boldsymbol{u}) := \int_{\mathbb{R}^d} g(\boldsymbol{x}) e^{i\boldsymbol{u}\cdot\boldsymbol{x}} d\boldsymbol{x}, \quad \boldsymbol{u} \in \mathbb{R}^d. \tag{4.2.1}$$

This definition of the Fourier transform also appears in Definition 22.6 in Bauer (1996) and in Eberlein et al. (2010). Provided the integral in (4.2.1) exists, the domain of $\widehat{g}$ may also be extended to parts of the complex plane. If $g \geq 0$ and $\int g(\boldsymbol{x}) d\boldsymbol{x} = 1$, then g is called *density*, $\widehat{g}$ is called the *characteristic function* and the map $\boldsymbol{y} \mapsto \int_{(-\infty, \boldsymbol{y}]} g(\boldsymbol{x}) d\boldsymbol{x}$ is called the *cumulative distribution function* (CDF).

4.3 Damped COS method

Typically, the function of interest, w, is only locally integrable, but $w \notin \mathcal{L}^1$. We provide two examples: The integral in (4.1.1) is equal to the CDF of g evaluated at $\boldsymbol{y} \in \mathbb{R}^d$ if $w(\boldsymbol{x}) = 1_{(-\infty, \boldsymbol{y}]}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$. In a financial context, an arithmetic basket put option is defined by $w(\boldsymbol{x}) = \max(K - \sum_{h=1}^d e^{x_h}, 0)$, $\boldsymbol{x} \in \mathbb{R}^d$, where $K > 0$. To introduce the *damped COS method*, we will consider a damped function of interest that is assumed to be integrable. Note that for many models and function of interests, both $\widehat{g}$ and $\widehat{w}$ are given in closed form, see e.g., Ruijter and Oosterlee (2012), Eberlein et al. (2010) and references therein.

For a scaling factor $\lambda > 0$, a shift parameter $\boldsymbol{\mu} \in \mathbb{R}^d$ and a damping factor $\boldsymbol{\alpha} \in \mathbb{R}^d$, we define the *damped density* by

$$f(\boldsymbol{x}) = \lambda e^{\boldsymbol{\alpha}\cdot(\boldsymbol{x}+\boldsymbol{\mu})} g(\boldsymbol{x} + \boldsymbol{\mu}), \quad \boldsymbol{x} \in \mathbb{R}^d \tag{4.3.1}$$

and the *damped function of interest* by

$$v(\boldsymbol{x}) = \frac{1}{\lambda} e^{-\boldsymbol{\alpha}\cdot(\boldsymbol{x}+\boldsymbol{\mu})} w(\boldsymbol{x} + \boldsymbol{\mu}), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{4.3.2}$$

By definition, it follows that

$$\int_{\mathbb{R}^d} w(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{R}^d} v(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}. \tag{4.3.3}$$

Thanks to Proposition 4.3.2, f is a density centered at zero if we choose λ and $\boldsymbol{\mu}$ carefully and $\widehat{f}$ is given in closed form if $\widehat{g}$ is given in closed form.

Remark 4.3.1.
In some applications it might be useful to know that f is a density. However, in the proofs, we use only that f is nonnegative and integrable.

Proposition 4.3.2.
Let $g \in \mathcal{L}_1$ and $\boldsymbol{\alpha} \in \mathbb{R}^d$. Assume that g is a density and that $\boldsymbol{x} \mapsto |\boldsymbol{x}|e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})$ is integrable. Let $\lambda = (\widehat{g}(-i\boldsymbol{\alpha}))^{-1}$ then $\lambda \in (0,\infty)$. Choose $\boldsymbol{\mu} \in \mathbb{R}^d$ by

$$\mu_h = -\lambda i \left.\frac{\partial}{\partial u_h}\widehat{g}(\boldsymbol{u} - i\boldsymbol{\alpha})\right|_{\boldsymbol{u}=\boldsymbol{0}}, \quad h = 1, ..., d. \tag{4.3.4}$$

Define $f(\boldsymbol{x}) = \lambda e^{\boldsymbol{\alpha}\cdot(\boldsymbol{x}+\boldsymbol{\mu})}g(\boldsymbol{x}+\boldsymbol{\mu})$, $\boldsymbol{x} \in \mathbb{R}^d$. Then f is a density with characteristic function

$$\widehat{f}(\boldsymbol{u}) = \lambda e^{-i\boldsymbol{u}\cdot\boldsymbol{\mu}}\widehat{g}(\boldsymbol{u} - i\boldsymbol{\alpha}), \quad \boldsymbol{u} \in \mathbb{R}^d. \tag{4.3.5}$$

Further, the moments of f of first order are zero, i.e., $\int_{\mathbb{R}^d} f(\boldsymbol{x})x_h d\boldsymbol{x} = 0$, $h = 1, ..., d$.

Proof. Use $\int |\boldsymbol{x}|e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})d\boldsymbol{x} < \infty$ and split the integration range into $\mathbb{R}^d \setminus B_1$ and B_1, where B_1 is the unit ball, to see that $\boldsymbol{x} \mapsto e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})$ is integrable. Since $\lambda = (\int_{\mathbb{R}^d} e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})d\boldsymbol{x})^{-1}$ and g is a density we have $\lambda \in (0,\infty)$. By the definition of λ, f is a density. Since $f \in \mathcal{L}^1$, $\widehat{f}$ exists. A direct analysis shows (4.3.5). By Theorem 25.2 in Bauer (1996), the partial derivatives in Equation (4.3.4) exist and it holds that $\mu_h = \lambda \int_{\mathbb{R}^d} e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})x_h d\boldsymbol{x}$. Finally, we have that

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(\boldsymbol{x})x_h d\boldsymbol{x} \\
=&\int_{\mathbb{R}^d} \lambda e^{\boldsymbol{\alpha}\cdot(\boldsymbol{x}+\boldsymbol{\mu})}g(\boldsymbol{x}+\boldsymbol{\mu})x_h d\boldsymbol{x} \\
=&\lambda \int_{\mathbb{R}^d} e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})x_h d\boldsymbol{x} - \mu_h \lambda \int_{\mathbb{R}^d} e^{\boldsymbol{\alpha}\cdot\boldsymbol{x}}g(\boldsymbol{x})d\boldsymbol{x} = 0.
\end{aligned}$$

□

In some cases $\widehat{g}$ needs to be approximated numerically; e.g., in Duffie et al. (2003), $\widehat{g}$ is the solution to some ordinary differential equation, which itself needs to be solved numerically before applying the COS method. From now on, we assume that $f \in \mathcal{L}_1$ and that $\widehat{f}$ is

given explicitly or can be efficiently approximated numerically by some function ϑ and that v is (at least) locally integrable. At several places, we assume $v \in \mathcal{L}_1$, which can usually be achieved by setting $\boldsymbol{\alpha} \neq \mathbf{0}$. We describe the COS method in detail in order to approximate the right-hand side of Equation (4.3.3). Let $\boldsymbol{M} \in \mathbb{R}_+^d$ large enough so that

$$\int_{\mathbb{R}^d} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} \approx \int_{[-\boldsymbol{M},\boldsymbol{M}]} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}. \tag{4.3.6}$$

Let $\boldsymbol{L} \geq \boldsymbol{M}$. If f is centered at zero, we truncate f on $[-\boldsymbol{L}, \boldsymbol{L}]$ and approximate the truncated damped density using a Fourier-series. We intuitively have that

$$f \approx f 1_{[-\boldsymbol{L},\boldsymbol{L}]} \approx \sum_{\mathbf{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}{}' a_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]} \approx \sum_{\mathbf{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}{}' c_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]} \approx \sum_{\mathbf{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}{}' \tilde{c}_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]}, \tag{4.3.7}$$

where we define the basis functions

$$e_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{h=1}^{d} \cos\left(k_h \pi \frac{x_h + L_h}{2L_h}\right), \quad \boldsymbol{x} \in \mathbb{R}^d, \quad \boldsymbol{k} \in \mathbb{N}_0^d,$$

and the classical Fourier-cosine coefficients of $f 1_{[-\boldsymbol{L},\boldsymbol{L}]}$ are given for $\boldsymbol{k} \in \mathbb{N}_0^d$ by

$$\begin{aligned}
a_{\boldsymbol{k}} &= \frac{1}{\prod_{h=1}^d L_h} \int_{[-\boldsymbol{L},\boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x}\} \\
&\approx \frac{1}{\prod_{h=1}^d L_h} \int_{\mathbb{R}^d} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} && (4.3.8) \\
&= \frac{1}{2^{d-1} \prod_{h=1}^d L_h} \sum_{\boldsymbol{s}=(1,\pm 1,\ldots,\pm 1) \in \mathbb{R}^d} \Re\left\{\widehat{f}\left(\frac{\pi}{2}\frac{\boldsymbol{s}\boldsymbol{k}}{\boldsymbol{L}}\right) \exp\left(i\frac{\pi}{2}\boldsymbol{s}\cdot\boldsymbol{k}\right)\right\} =: c_{\boldsymbol{k}} && (4.3.9) \\
&\approx \frac{1}{2^{d-1} \prod_{h=1}^d L_h} \sum_{\boldsymbol{s}=(1,\pm 1,\ldots,\pm 1) \in \mathbb{R}^d} \Re\left\{\vartheta\left(\frac{\pi}{2}\frac{\boldsymbol{s}\boldsymbol{k}}{\boldsymbol{L}}\right) \exp\left(i\frac{\pi}{2}\boldsymbol{s}\cdot\boldsymbol{k}\right)\right\} =: \tilde{c}_{\boldsymbol{k}}. && (4.3.10)
\end{aligned}$$

Sometimes it is necessary to choose $\boldsymbol{L} > \boldsymbol{M}$ to ensure that $c_{\boldsymbol{k}}$ is close enough to $a_{\boldsymbol{k}}$. The key insight of the COS method is the fact that the integral at the right-hand side of Equation (4.3.8) can be solved explicitly[1]. If $\widehat{f}$ needs to be approximated by some function ϑ, we use $\tilde{c}_{\boldsymbol{k}}$ instead of $c_{\boldsymbol{k}}$.

The idea of the multidimensional COS method is to approximate f as in (4.3.7), and hence the right-hand side of Equation (4.3.6), by

$$\int_{[-\boldsymbol{M},\boldsymbol{M}]} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} \approx \sum_{\mathbf{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}{}' \tilde{c}_{\boldsymbol{k}} \underbrace{\int_{[-\boldsymbol{M},\boldsymbol{M}]} v(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x}}_{=: v_{\boldsymbol{k}}} \tag{4.3.11}$$

[1] We use $\prod_{h=1}^d \cos\theta_h = \frac{1}{2^{d-1}} \sum_{\boldsymbol{s}=(1,\pm 1,\ldots,\pm 1) \in \mathbb{R}^d} \cos(\boldsymbol{s}\cdot\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathbb{R}^d$, which follows by mathematical induction, from the fact that the cosine is an even function and from the trigonometric identities stated in Equations (4.3.17, 4.3.31) in Abramowitz and Stegun (1972).

$$\approx \sum_{\mathbf{0}\leq \boldsymbol{k}\leq \boldsymbol{N}}{}' \tilde{c}_{\boldsymbol{k}} \underbrace{\int_{\mathbb{R}^d} v(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x}}_{=:\tilde{v}_{\boldsymbol{k}}}. \tag{4.3.12}$$

Classical COS method: If $\boldsymbol{\alpha} = \mathbf{0}$, we speak of the classical COS method. In important cases,
Equation (4.3.11) can be solved analytically. Examples include, in one-dimension, CDFs, plain vanilla put and call options and digital options, see Fang and Oosterlee (2009a) and, in two dimensions, geometric basket options, call-on-maximum options and put-on-minimum options, see Ruijter and Oosterlee (2012). In general dimensions, the coefficients $v_{\boldsymbol{k}}$ of a CDF are given in closed form, see Example 4.5.1. In the case that the integral in Equation (4.3.11) cannot be solved directly (e.g. for arithmetic basket options), Ruijter and Oosterlee (2012) propose solving the integral in Equation (4.3.11) numerically to obtain $v_{\boldsymbol{k}}$, e.g., by the discrete cosine transform or some quadrature rule. However, solving the integral in (4.3.11) numerically for each $\boldsymbol{k}$ is expensive and slows down the COS method significantly.

Damped COS method: If $\boldsymbol{\alpha} \neq \mathbf{0}$, we speak of the damped COS method. Assume that v is integrable, which usually can be achieved by setting $\boldsymbol{\alpha} \neq \mathbf{0}$. Then we propose to approximate $v_{\boldsymbol{k}}$ by $\tilde{v}_{\boldsymbol{k}}$. This works if $\boldsymbol{M}$ is large enough. Similar to the solution presented in Equation (4.3.9), the coefficients $\tilde{v}_{\boldsymbol{k}}$ are given analytically:

$$\tilde{v}_{\boldsymbol{k}} = \frac{1}{2^{d-1}} \sum_{\boldsymbol{s}=(1,\pm 1,\ldots,\pm 1)\in\mathbb{R}^d} \Re\left\{ \widehat{v}\left(\frac{\pi}{2}\frac{\boldsymbol{s}\boldsymbol{k}}{\boldsymbol{L}}\right) \exp\left(i\frac{\pi}{2}\boldsymbol{s}\cdot\boldsymbol{k}\right)\right\}. \tag{4.3.13}$$

In the remainder of the article, we will prove conditions under which the integral (4.1.1) can be approximated by the (damped) COS method.

Remark 4.3.3.
In the special case that $\widehat{f}$ only takes real values, the computational cost of the COS method can be reduced by (about) a factor of one half, since $c_{\boldsymbol{k}} = 0$ if $\sum_{h=1}^{d} k_h$ is odd.

In order to prove the convergence of the COS method in Theorem 4.3.6 and Corollary 4.3.7, we need the concept of *COS-admissibility*, which is introduced in Definition 4.3.4 and extends Definition 1 in Junike and Pankrashkin (2022) to the multidimensional setting.

Definition 4.3.4.
Let $\boldsymbol{L} = (L_1, ..., L_d) \in \mathbb{R}_+^d$. A function $f \in \mathcal{L}^1$ is called *COS-admissible* if

$$B_f(\boldsymbol{L}) := \sum_{\boldsymbol{k} \in \mathbb{N}_0^d}{}' \frac{1}{\prod_{h=1}^d L_h} \left| \int_{\mathbb{R}^d \setminus [-\boldsymbol{L}, \boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2 \to 0, \quad \min_{h=1,...,d} L_h \to \infty.$$

By Proposition 4.3.5, it follows that bounded densities with existing moments are COS-admissible, which indicates that the class of d-dimensional, COS-admissible densities is large.

Proposition 4.3.5.
Assume $f \in \mathcal{L}^1 \cap \mathcal{L}^2$ with

$$\int_{\mathbb{R}^d} |\boldsymbol{x}|^{2d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} < \infty. \tag{4.3.14}$$

Then f is COS-admissible. Let $\boldsymbol{L} = (L_1, ..., L_d) \in \mathbb{R}_+^d$; then it holds that

$$B_f(\boldsymbol{L}) \leq \Xi \int_{\mathbb{R}^d \setminus [-\boldsymbol{L}, \boldsymbol{L}]} \prod_{h=1}^d \max\left\{ x_h^2 L_h^{-2}, 1 \right\} |f(\boldsymbol{x})|^2 d\boldsymbol{x} \tag{4.3.15}$$

$$\leq \frac{\Xi}{d \min\limits_{h=1,...,d} L_h^{2d}} \int_{\mathbb{R}^d \setminus [-\boldsymbol{L}, \boldsymbol{L}]} |\boldsymbol{x}|^{2d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} + \Xi \int_{\mathbb{R}^d \setminus [-\boldsymbol{L}, \boldsymbol{L}]} |f(\boldsymbol{x})|^2 d\boldsymbol{x}, \tag{4.3.16}$$

where $\Xi = \frac{\pi^2}{3} \sum_{h=1}^d \left(\frac{\pi^2}{3} + 1 \right)^{h-1}$.

Proof. Let $\boldsymbol{L} \in \mathbb{R}_+^d$ and $\boldsymbol{j} \in \mathbb{Z}^d$. It follows by Parseval's identity

$$\begin{aligned}
&\int_{[2\boldsymbol{j}\boldsymbol{L} - \boldsymbol{L}, 2\boldsymbol{j}\boldsymbol{L} + \boldsymbol{L}]} |f(\boldsymbol{x})|^2 d\boldsymbol{x} \\
&= \sum_{\boldsymbol{k} \in \mathbb{N}_0^d}{}' \frac{1}{\prod_{h=1}^d L_h} \left| \int_{[2\boldsymbol{j}\boldsymbol{L} - \boldsymbol{L}, 2\boldsymbol{j}\boldsymbol{L} + \boldsymbol{L}]} f(\boldsymbol{x}) \prod_{h=1}^d \underbrace{\cos\left(k_h \pi \frac{x_h - (2j_h L_h - L_h)}{2L_h} \right)}_{= (-1)^{j_h k_h} \cos\left(k_h \pi \frac{x_h + L_h}{2L_h} \right)} d\boldsymbol{x} \right|^2 \\
&= \sum_{\boldsymbol{k} \in \mathbb{N}_0^d}{}' \frac{1}{\prod_{h=1}^d L_h} \left| \int_{[2\boldsymbol{j}\boldsymbol{L} - \boldsymbol{L}, 2\boldsymbol{j}\boldsymbol{L} + \boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2.
\end{aligned} \tag{4.3.17}$$

By the Cauchy-Schwarz inequality, we obtain with $g(\boldsymbol{j}) := \prod_{h=1}^d \max\{|j_h|, 1\}$,

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d \setminus [-\boldsymbol{L}, \boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2 \\
&= \left| \sum_{\boldsymbol{j} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}} \frac{g(\boldsymbol{j})}{g(\boldsymbol{j})} \int_{[2\boldsymbol{j}\boldsymbol{L} - \boldsymbol{L}, 2\boldsymbol{j}\boldsymbol{L} + \boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2
\end{aligned}$$

$$\leq \Big(\underbrace{\sum_{\boldsymbol{j}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} \frac{1}{(g(\boldsymbol{j}))^2}}_{=\Xi} \Big) \sum_{\boldsymbol{j}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} (g(\boldsymbol{j}))^2 \left| \int_{[2\boldsymbol{j}\boldsymbol{L}-\boldsymbol{L},2\boldsymbol{j}\boldsymbol{L}+\boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2 . \tag{4.3.18}$$

The fact that $\Xi = \frac{\pi^2}{3} \sum_{h=1}^{d} \left(\frac{\pi^2}{3}+1\right)^{h-1}$ can be shown by mathematical induction over d. Then it follows that

$$\begin{aligned} B_f(\boldsymbol{L}) &\overset{(4.3.18)}{\leq} \Xi \sum_{\boldsymbol{j}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} (g(\boldsymbol{j}))^2 {\sum_{\boldsymbol{k}\in\mathbb{N}_0^d}}' \frac{1}{\prod_{h=1}^{d} L_h} \left| \int_{[2\boldsymbol{j}\boldsymbol{L}-\boldsymbol{L},2\boldsymbol{j}\boldsymbol{L}+\boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2 \\ &\overset{(4.3.17)}{=} \Xi \sum_{\boldsymbol{j}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} (g(\boldsymbol{j}))^2 \int_{[2\boldsymbol{j}\boldsymbol{L}-\boldsymbol{L},2\boldsymbol{j}\boldsymbol{L}+\boldsymbol{L}]} |f(\boldsymbol{x})|^2 d\boldsymbol{x}. \end{aligned}$$

For $\boldsymbol{j}\in\mathbb{Z}^d$ and $\boldsymbol{x}\in[2\boldsymbol{j}\boldsymbol{L}-\boldsymbol{L},2\boldsymbol{j}\boldsymbol{L}+\boldsymbol{L}]$, one has $|j_h| \leq \frac{|x_h|}{L_h}$, $h=1,...,d$. It follows that $(g(\boldsymbol{j}))^2 \leq \prod_{h=1}^{d} \max\left\{x_h^2 L_h^{-2}, 1\right\}$, which implies Inequality (4.3.15). By Young's inequality, it holds that

$$\prod_{h=1}^{d} \left(\max\left\{x_h^{2d} L_h^{-2d}, 1\right\}\right)^{\frac{1}{d}} \leq \frac{1}{d} \sum_{h=1}^{d} \max\left\{x_h^{2d} L_h^{-2d}, 1\right\} \leq \frac{|\boldsymbol{x}|^{2d}}{d \min\limits_{h=1,...,d} L_h^{2d}} + 1.$$

In the last inequality, we used $\max\{a,b\} \leq a+b$ for any $a,b\geq 0$ and $\sum_{h=1}^{d} x_h^{2d} \leq |\boldsymbol{x}|^{2d}$, which follows from the monotonicity of the p-norm. Hence, Inequality (4.3.16) holds. Assumption (4.3.14) and $f\in\mathcal{L}^2$ imply $B_f(\boldsymbol{L}) \to 0$, $\min_{h=1,...,d} L_h \to \infty$. □

The following theorem shows that multivariate densities can be approximated by a cosine expansion. The theorem also includes numerical uncertainty on the Fourier transform $\widehat{f}$.

Theorem 4.3.6.
Assume that $f\in\mathcal{L}^1\cap\mathcal{L}^2$ is COS-admissible. Let $\vartheta:\mathbb{R}^d\to\mathbb{C}$ and define $\tilde{c}_{\boldsymbol{k}}$ as in Equation (4.3.10). For any $\varepsilon>0$ there is a $\boldsymbol{L}\in\mathbb{R}_+^d$, a $\boldsymbol{N}\in\mathbb{N}^d$ and a $\gamma>0$ such that $\left\|\widehat{f}-\vartheta\right\|_\infty < \gamma$ implies

$$\left\| f - {\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}}' \tilde{c}_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]} \right\|_2 < \varepsilon.$$

Note that $\boldsymbol{N}$ depends on $\boldsymbol{L}$ and that γ depends on both $\boldsymbol{L}$ and $\boldsymbol{N}$.

Proof. Define $e_{\boldsymbol{k}}^{\boldsymbol{L}} = e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]}$. It holds for $\boldsymbol{l},\boldsymbol{k}\in\mathbb{N}_0^d$ that

$$\left\langle e_{\boldsymbol{k}}^{\boldsymbol{L}}, e_{\boldsymbol{l}}^{\boldsymbol{L}} \right\rangle = \begin{cases} 2^{\Lambda(\boldsymbol{k})} \prod_{h=1}^{d} L_h & ,\boldsymbol{k}=\boldsymbol{l}, \\ 0 & ,\text{otherwise}, \end{cases} \tag{4.3.19}$$

where Λ is defined Section 4.2. For any $\boldsymbol{L} \in \mathbb{R}_+^d$ and $\boldsymbol{N} \in \mathbb{N}^d$, it holds that

$$\left\| f - \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{N}}{}' \tilde{c}_{\boldsymbol{k}} e_{\boldsymbol{k}}^{\boldsymbol{L}} \right\|_2 \le \underbrace{\left\| f - f1_{[-\boldsymbol{L},\boldsymbol{L}]} \right\|_2}_{=:A_1(\boldsymbol{L})} + \underbrace{\left\| f1_{[-\boldsymbol{L},\boldsymbol{L}]} - \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{N}}{}' a_{\boldsymbol{k}} e_{\boldsymbol{k}}^{\boldsymbol{L}} \right\|_2}_{=:A_2(\boldsymbol{L},\boldsymbol{N})}$$
$$+ \underbrace{\left\| \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{N}}{}' (a_{\boldsymbol{k}} - c_{\boldsymbol{k}}) e_{\boldsymbol{k}}^{\boldsymbol{L}} \right\|_2}_{=:A_3(\boldsymbol{L},\boldsymbol{N})} + \underbrace{\left\| \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{N}}{}' (c_{\boldsymbol{k}} - \tilde{c}_{\boldsymbol{k}}) e_{\boldsymbol{k}}^{\boldsymbol{L}} \right\|_2}_{=:A_4(\boldsymbol{L},\boldsymbol{N})}.$$

Further,

$$\begin{aligned} A_3(\boldsymbol{L},\boldsymbol{N})^2 &= \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{N}} \sum_{\mathbf{0} \le \boldsymbol{l} \le \boldsymbol{N}} \frac{1}{2^{\Lambda(\boldsymbol{k})+\Lambda(\boldsymbol{l})}} (a_{\boldsymbol{k}} - c_{\boldsymbol{k}})(a_{\boldsymbol{l}} - c_{\boldsymbol{l}}) \left\langle e_{\boldsymbol{k}}^{\boldsymbol{L}}, e_{\boldsymbol{l}}^{\boldsymbol{L}} \right\rangle \\ &\le \sum_{\boldsymbol{k} \in \mathbb{N}_0^d}{}' \prod_{h=1}^{d} \{L_h\} \left| a_{\boldsymbol{k}} - c_{\boldsymbol{k}} \right|^2 = B_f(\boldsymbol{L}), \end{aligned}$$

see Definition 4.3.4. For $\varepsilon > 0$, choose $\boldsymbol{L} \in \mathbb{R}_+^d$ such that $A_1(\boldsymbol{L}) < \frac{\varepsilon}{4}$ and $B_f(\boldsymbol{L}) < (\frac{\varepsilon}{4})^2$. Hence, $A_3(\boldsymbol{L},\boldsymbol{N}) < \frac{\varepsilon}{4}$. Then choose $\boldsymbol{N} \in \mathbb{N}^d$ such that $A_2(\boldsymbol{L},\boldsymbol{N}) < \frac{\varepsilon}{4}$. Such $\boldsymbol{N}$ exists by classical Fourier analysis. By the definition of $c_{\boldsymbol{k}}$ and $\tilde{c}_{\boldsymbol{k}}$, see Equation (4.3.9), it follows that

$$|c_{\boldsymbol{k}} - \tilde{c}_{\boldsymbol{k}}| \le \frac{1}{2^{d-1} \prod_{h=1}^d L_h} \sum_{\boldsymbol{s}=(1,\pm 1,\ldots,\pm 1) \in \mathbb{R}^d} \left| \widehat{f}\left(\frac{\pi}{2} \frac{\boldsymbol{s}\boldsymbol{k}}{\boldsymbol{L}}\right) - \vartheta\left(\frac{\pi}{2} \frac{\boldsymbol{s}\boldsymbol{k}}{\boldsymbol{L}}\right) \right| \le \frac{\left\| \widehat{f} - \vartheta \right\|_\infty}{\prod_{h=1}^d L_h}.$$

Similarly to the analysis of A_3, we have

$$A_4(\boldsymbol{L},\boldsymbol{N})^2 \le \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{N}} \prod_{h=1}^{d} \{L_h\} \left| c_{\boldsymbol{k}} - \tilde{c}_{\boldsymbol{k}} \right|^2 \le \frac{\left\| \widehat{f} - \vartheta \right\|_\infty^2}{\prod_{h=1}^d L_h} \prod_{h=1}^{d} \{N_h + 1\}. \tag{4.3.20}$$

Choose $\gamma = \frac{\varepsilon}{4} \sqrt{\prod_{h=1}^d L_h} \left(\prod_{h=1}^d \{N_h + 1\} \right)^{-\frac{1}{2}}$. Then $\left\| \widehat{f} - \vartheta \right\|_\infty < \gamma$ implies $A_4(\boldsymbol{L},\boldsymbol{N}) < \frac{\varepsilon}{4}$, which concludes the proof. □

Corollary 4.3.7 provides sufficient conditions to ensure that the COS method approximates the integral (4.1.1) within a predefined error tolerance $\varepsilon > 0$, including numerical uncertainty on $\widehat{f}$ and numerical uncertainty on the cosine coefficients of the function of interest v: either because the $v_{\boldsymbol{k}}$ are approximated by solving the integral in Equation (4.3.11) numerically or because $v_{\boldsymbol{k}}$ are approximated by $\tilde{v}_{\boldsymbol{k}}$ defined in Equation (4.3.12).

Corollary 4.3.7. *(Convergence of the COS method)*
Let $f \in \mathcal{L}^1 \cap \mathcal{L}^2$ be COS-admissible and $v : \mathbb{R}^d \to \mathbb{R}$ be locally in $\mathcal{L}^2$; that is, $v1_{[-\boldsymbol{M},\boldsymbol{M}]} \in \mathcal{L}^2$

for any $\boldsymbol{M} \in \mathbb{R}_+^d$. Assume $vf \in \mathcal{L}^1$; then the integral of the product of f and v can be approximated by a finite sum as follows: Let $\varepsilon > 0$. Let $\boldsymbol{M} \in \mathbb{R}_+^d$ and $\xi > 0$ such that

$$\int_{\mathbb{R}^d \setminus [-\boldsymbol{M}, \boldsymbol{M}]} |v(\boldsymbol{x}) f(\boldsymbol{x})| \, d\boldsymbol{x} \leq \frac{\varepsilon}{3}, \quad \left\| v 1_{[-\boldsymbol{M}, \boldsymbol{M}]} \right\|_2 \leq \xi. \tag{4.3.21}$$

Let $\boldsymbol{L} \geq \boldsymbol{M}$ such that

$$\left\| f - f 1_{[-\boldsymbol{L}, \boldsymbol{L}]} \right\|_2 \leq \frac{\varepsilon}{12\xi} \tag{4.3.22}$$

and

$$B_f(\boldsymbol{L}) \leq \left(\frac{\varepsilon}{12\xi} \right)^2. \tag{4.3.23}$$

Choose $\boldsymbol{N} \in \mathbb{N}^d$ large enough, so that

$$\left\| f 1_{[-\boldsymbol{L}, \boldsymbol{L}]} - {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' a_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L}, \boldsymbol{L}]} \right\|_2 \leq \frac{\varepsilon}{12\xi}. \tag{4.3.24}$$

For some $\vartheta : \mathbb{R}^d \to \mathbb{C}$ assume

$$\left\| \widehat{f} - \vartheta \right\|_\infty \leq \frac{\varepsilon}{12\xi} \frac{\sqrt{\prod_{h=1}^d L_h}}{\sqrt{\prod_{h=1}^d \{N_h + 1\}}}. \tag{4.3.25}$$

Let $\eta > 0$ such that ${\sum}'_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}} |\tilde{c}_{\boldsymbol{k}}|^2 \leq \eta$. Let $(\tilde{v}_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathbb{N}_0^d} \subset \mathbb{R}$ such that

$${\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' |\tilde{v}_{\boldsymbol{k}} - v_{\boldsymbol{k}}|^2 \leq \frac{\varepsilon^2}{9\eta}. \tag{4.3.26}$$

Then it follows that

$$\left| \int_{\mathbb{R}^d} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' \tilde{c}_{\boldsymbol{k}} \tilde{v}_{\boldsymbol{k}} \right| \leq \varepsilon. \tag{4.3.27}$$

Proof. Define $e_{\boldsymbol{k}}^{\boldsymbol{L}} = e_{\boldsymbol{k}} 1_{[-\boldsymbol{L}, \boldsymbol{L}]}$. Let $A_1(\boldsymbol{L})$, $A_2(\boldsymbol{L}, \boldsymbol{N})$ and $A_4(\boldsymbol{L}, \boldsymbol{N})$ be as in the proof of Theorem 4.3.6. By Inequalities (4.3.20, 4.3.25) it follows that $A_4(\boldsymbol{L}, \boldsymbol{N}) \leq \frac{\varepsilon}{12\xi}$. Due to $v_{\boldsymbol{k}} = \langle v 1_{[-\boldsymbol{M}, \boldsymbol{M}]}, e_{\boldsymbol{k}}^{\boldsymbol{L}} \rangle$ and applying Theorem 4.3.6 and the Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
&\Big| \int_{\mathbb{R}^d} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' \tilde{c}_{\boldsymbol{k}} \tilde{v}_{\boldsymbol{k}} \Big| \\
&= \Big| \int_{\mathbb{R}^d \setminus [-\boldsymbol{M}, \boldsymbol{M}]} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} + \langle v 1_{[-\boldsymbol{M}, \boldsymbol{M}]}, f \rangle - {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' \tilde{c}_{\boldsymbol{k}} \langle v 1_{[-\boldsymbol{M}, \boldsymbol{M}]}, e_{\boldsymbol{k}}^{\boldsymbol{L}} \rangle - {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' \tilde{c}_{\boldsymbol{k}} (\tilde{v}_{\boldsymbol{k}} - v_{\boldsymbol{k}}) \Big| \\
&\leq \underbrace{\int_{\mathbb{R}^d \setminus [-\boldsymbol{M}, \boldsymbol{M}]} |v(\boldsymbol{x}) f(\boldsymbol{x})| d\boldsymbol{x}}_{=: D_1(\boldsymbol{M})} + \Big| \Big\langle v 1_{[-\boldsymbol{M}, \boldsymbol{M}]}, f - {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' \tilde{c}_{\boldsymbol{k}} e_{\boldsymbol{k}}^{\boldsymbol{L}} \Big\rangle \Big| + \underbrace{\sqrt{{\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' |\tilde{v}_{\boldsymbol{k}} - v_{\boldsymbol{k}}|^2 \, {\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}}' |\tilde{c}_{\boldsymbol{k}}|^2}}_{=: D_2(\boldsymbol{N}, \boldsymbol{L}, \boldsymbol{M})}
\end{aligned}$$

$$
\begin{aligned}
&< \frac{\varepsilon}{3} + \|v1_{[-\boldsymbol{M},\boldsymbol{M}]}\|_2 \left\| f - \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' \tilde{c}_{\boldsymbol{k}} e_{\boldsymbol{k}}^{\boldsymbol{L}} \right\|_2 + \frac{\varepsilon}{3} \\
&< \frac{\varepsilon}{3} + \xi \left(A_1(\boldsymbol{L}) + A_2(\boldsymbol{L},\boldsymbol{N}) + \sqrt{B_f(\boldsymbol{L})} + A_4(\boldsymbol{L},\boldsymbol{N}) \right) + \frac{\varepsilon}{3} \\
&\leq \frac{\varepsilon}{3} + \xi \left(\frac{\varepsilon}{12\xi} + \frac{\varepsilon}{12\xi} + \frac{\varepsilon}{12\xi} + \frac{\varepsilon}{12\xi} \right) + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}
$$

□

Junike and Pankrashkin (2022) and Junike (2024) assume that f has semi-heavy tails, i.e., f decays exponentially or faster. Here, we make the same assumption in multivariate dimensions in order to be able to estimate $\boldsymbol{M}$, $\boldsymbol{L}$ and $\boldsymbol{N}$.

Definition 4.3.8.
A function $f : \mathbb{R}^d \to \mathbb{R}$ *decays exponentially* if there are $C_1, C_2, m > 0$ such that for $|\boldsymbol{x}| > m$ it holds that $|f(\boldsymbol{x})| \leq C_1 e^{-C_2|\boldsymbol{x}|}$.

Lemma 4.3.9.
Let $f \in \mathcal{L}^1 \cap \mathcal{L}^2$. Let $\boldsymbol{M}, \boldsymbol{L} \in \mathbb{R}_+^d$ with $\boldsymbol{M} \leq \boldsymbol{L}$; then it holds that

$$
\|f1_{[-\boldsymbol{L},\boldsymbol{L}]} - \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' a_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]}\|_2^2 \leq \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} - \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |c_{\boldsymbol{k}}|^2 + G(\boldsymbol{L}),
$$

where

$$
G(\boldsymbol{L}) := B_f(\boldsymbol{L}) + 2\sqrt{B_f(\boldsymbol{L}) \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x}}. \tag{4.3.28}
$$

Proof. Let

$$
\phi_{\boldsymbol{k}} := \frac{1}{\prod_{h=1}^{d} L_h} \int_{\mathbb{R}^d \setminus [-\boldsymbol{L},\boldsymbol{L}]} f(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x}, \quad \boldsymbol{k} \in \mathbb{N}_0^d.
$$

It holds that $c_{\boldsymbol{k}} = a_{\boldsymbol{k}} + \phi_{\boldsymbol{k}}$. It follows by the Cauchy-Schwarz inequality that

$$
\begin{aligned}
\prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |c_{\boldsymbol{k}}|^2 &= \prod_{h=1}^{d} L_h \left(\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |a_{\boldsymbol{k}}|^2 + \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |\phi_{\boldsymbol{k}}|^2 + 2 \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |\phi_{\boldsymbol{k}}||a_{\boldsymbol{k}}| \right) \\
&\leq \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |a_{\boldsymbol{k}}|^2 + B_f(\boldsymbol{L}) + 2 \prod_{h=1}^{d} L_h \sqrt{\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |\phi_{\boldsymbol{k}}|^2 \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |a_{\boldsymbol{k}}|^2} \\
&\overset{(4.3.17)}{\leq} \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |a_{\boldsymbol{k}}|^2 + \underbrace{B_f(\boldsymbol{L}) + 2\sqrt{B_f(\boldsymbol{L}) \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x}}}_{=G(\boldsymbol{L})} \qquad (4.3.29) \\
&\overset{(4.3.17)}{\leq} \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} + G(\boldsymbol{L}). \qquad (4.3.30)
\end{aligned}
$$

Hence,

$$\begin{aligned}\|f1_{[-\boldsymbol{L},\boldsymbol{L}]} - \sum_{\boldsymbol{0}\le\boldsymbol{k}\le\boldsymbol{N}}{}' a_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]}\|_2^2 &\overset{(4.3.19)}{\le} \prod_{h=1}^{d} L_h \sum_{k_1>N_1 \text{ or...or } k_d>N_d}{}' |a_{\boldsymbol{k}}|^2 \\ &= \prod_{h=1}^{d} L_h \sum_{\boldsymbol{k}\in\mathbb{N}_0^d}{}' |a_{\boldsymbol{k}}|^2 - \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\le\boldsymbol{k}\le\boldsymbol{N}}{}' |a_{\boldsymbol{k}}|^2 \\ &\overset{(4.3.17,4.3.29)}{\le} \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} - \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\le\boldsymbol{k}\le\boldsymbol{N}}{}' |c_{\boldsymbol{k}}|^2 + G(\boldsymbol{L}).\end{aligned}$$

□

Theorem 4.3.10. *(Classical COS method: Find $\boldsymbol{M}$ and $\boldsymbol{L}$)*
Let $f \in \mathcal{L}^1 \cap \mathcal{L}^2$ be a nonnegative function satisfying Inequality (4.3.14). Let $v : \mathbb{R}^d \to \mathbb{R}$ be bounded with $\|v\|_\infty \in (0,\infty)$. Let $n \ge 2$ be some even number and assume the moments of f of $n^{th}-$order exist, i.e.,

$$m_h(n) := \int_{\mathbb{R}^d} x_h^n f(\boldsymbol{x}) d\boldsymbol{x} = i^{-n} \frac{\partial^n}{\partial u_h^n} \widehat{f}(\boldsymbol{u})\bigg|_{\boldsymbol{u}=\boldsymbol{0}} \in (0,\infty), \quad h = 1,...,d. \tag{4.3.31}$$

Assume that f decays exponentially. Let $\varepsilon > 0$ be small enough. Define

$$M_h := \left(\frac{3d\,\|v\|_\infty}{\varepsilon} m_h(n)\right)^{\frac{1}{n}}, \quad h = 1,...,d, \tag{4.3.32}$$

and $\boldsymbol{L} = \boldsymbol{M} = (M_1,...,M_d) \in \mathbb{R}_+^d$. There is a $\boldsymbol{N} \in \mathbb{N}_0^d$ such that

$$\left| \int_{\mathbb{R}^d} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - \sum_{\boldsymbol{0}\le\boldsymbol{k}\le\boldsymbol{N}}{}' c_{\boldsymbol{k}} v_{\boldsymbol{k}} \right| \le \varepsilon. \tag{4.3.33}$$

Corollary 4.3.11. *(Damped COS method: Find $\boldsymbol{M}$ and $\boldsymbol{L}$)*
Assume that all assumptions in Theorem 4.3.10 hold and that $v \in \mathcal{L}^1 \cap \mathcal{L}^2$, v satisfies Inequality (4.3.14) and v decays exponentially. Define $\tilde{v}_{\boldsymbol{k}}$ as in (4.3.12). There is a $\boldsymbol{N} \in \mathbb{N}_0^d$ such that

$$\left| \int_{\mathbb{R}^d} v(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - \sum_{\boldsymbol{0}\le\boldsymbol{k}\le\boldsymbol{N}}{}' c_{\boldsymbol{k}} \tilde{v}_{\boldsymbol{k}} \right| \le \varepsilon. \tag{4.3.34}$$

Corollary 4.3.12. *(Find $\boldsymbol{N}$)*
If the assumptions in Theorem 4.3.10, respectively Corollary 4.3.11, hold and if

$$\left| (2\pi)^{-d} \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{u})|^2 d\boldsymbol{u} - \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\le\boldsymbol{k}\le\boldsymbol{N}}{}' |c_{\boldsymbol{k}}|^2 \right| \le \frac{\varepsilon^2}{162 \left\| v 1_{[-\boldsymbol{M},\boldsymbol{M}]} \right\|_2^2} \tag{4.3.35}$$

for some $\boldsymbol{N} \in \mathbb{N}_0^d$, then Inequality (4.3.33), respectively Inequality (4.3.34), is satisfied.

Proof. We first prove Theorem 4.3.10: Equation (4.3.31) follows by Theorem 25.2 in Bauer (1996). For $h \in \{1, ..., d\}$ let $\pi_h : \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{x} \mapsto x_h$. Let λ^d be the Lebesgue measure on $\mathbb{R}^d$ and define the finite and positive measure $\mu := f\lambda^d$. By Markov's inequality, it follows that

$$\begin{aligned}\int_{\mathbb{R}^d \setminus [-\boldsymbol{M}, \boldsymbol{M}]} |v(\boldsymbol{x}) f(\boldsymbol{x})| \, d\boldsymbol{x} &\leq \|v\|_\infty \sum_{h=1}^{d} \mu\left(\left\{\boldsymbol{x} \in \mathbb{R}^d : |\pi_h(\boldsymbol{x})| \geq M_h\right\}\right) \\ &\leq \|v\|_\infty \sum_{h=1}^{d} \frac{m_h(n)}{M_h^n} = \frac{\varepsilon}{3}.\end{aligned}$$

The last equality follows by the definition of $\boldsymbol{M}$. Define $\xi := \|v\|_\infty \sqrt{2^d \prod_{h=1}^d M_h}$. It holds that $\left\|v 1_{[-\boldsymbol{M}, \boldsymbol{M}]}\right\|_2 \leq \xi$. Hence, the inequalities in (4.3.21) are satisfied. Next, we use the following auxiliary result: Let $s \geq 0$, $a > 0$ and $n \in \mathbb{N}_0$ and $d \in \mathbb{N}$. Then it holds by mathematical induction over n and Theorem 8.11 in Amann and Escher (2009) that

$$\int_{\{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}| > s\}} e^{-a|\boldsymbol{x}|} |\boldsymbol{x}|^n d\boldsymbol{x} = \frac{d\pi^{\frac{d}{2}}}{\Gamma\left(1 + \frac{d}{2}\right)} e^{-as} \frac{(n+d-1)!}{a^{n+d}} \sum_{k=0}^{n+d-1} \frac{(as)^k}{k!}. \tag{4.3.36}$$

For ε small enough, $\boldsymbol{L}$ is large enough. Using that f decays exponentially and applying Equation (4.3.36), we obtain with $\ell := \min_{h=1,...,d} L_h$ that

$$\left\|f - f 1_{[-\boldsymbol{L}, \boldsymbol{L}]}\right\|_2 \leq C_1 \sqrt{\int_{\{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}| > \ell\}} e^{-2C_2|\boldsymbol{x}|} d\boldsymbol{x}} \leq \frac{\varepsilon}{12\xi}. \tag{4.3.37}$$

The last inequality holds true if ε is small enough because, thanks to Inequality (4.3.36), the term in the middle of (4.3.37) decreases exponentially in ε, while the term at the right-hand side of (4.3.37) goes to zero like $\varepsilon^{1+\frac{d}{2n}}$ for $\varepsilon \searrow 0$. Hence, Inequality (4.3.22) holds. By Inequality (4.3.16) it holds that $B_f(\boldsymbol{L}) \leq \varepsilon^2 (12\xi)^{-2}$ if ε is small enough because $B_f(\boldsymbol{L})$ decreases exponentially in ε: to see this, use Inequality (4.3.37) and observe that the term $\int_{\mathbb{R}^d \setminus [-\boldsymbol{L}, \boldsymbol{L}]} |\boldsymbol{x}|^{2d} |f(\boldsymbol{x})|^2 d\boldsymbol{x}$ converges exponentially thanks to Inequality (4.3.36). Hence, Inequality (4.3.23) holds. By classical Fourier analysis, there is a $\boldsymbol{N} \in \mathbb{N}_0^d$ such that Inequality (4.3.24) is satisfied. By assumption we have $c_{\boldsymbol{k}} = \tilde{c}_{\boldsymbol{k}}$ and $v_{\boldsymbol{k}} = \tilde{v}_{\boldsymbol{k}}$. Inequalities (4.3.25) and (4.3.26) hold trivially. Apply Corollary 4.3.7 to finish the proof of Theorem 4.3.10.

We prove Corollary 4.3.11: We have to show that Inequality (4.3.26) holds to proof Corollary 4.3.11. Let $G(\boldsymbol{L})$ be as in Equality (4.3.28). Observe $G(\boldsymbol{L}) \to 0$, $\min_h L_h \to \infty$ because f is COS-admissible by Proposition 4.3.5. There is $\boldsymbol{P} \in \mathbb{R}_+^d$ and a $\gamma > 0$ such that $G(\boldsymbol{L}) \leq \gamma$ for all $\boldsymbol{L} \geq \boldsymbol{P}$. By Inequality (4.3.30), it follows for all $\boldsymbol{N} \in \mathbb{N}^d$ and all $\boldsymbol{L} \geq \boldsymbol{P}$ that

$$\sum_{\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{N}}{}' |c_{\boldsymbol{k}}|^2 \leq \frac{\int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} + \gamma}{\prod_{h=1}^d P_h} =: \eta < \infty. \tag{4.3.38}$$

It follows by Proposition 4.3.5 for all $\boldsymbol{N} \in \mathbb{N}^d$ that

$$\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |\tilde{v}_{\boldsymbol{k}} - v_{\boldsymbol{k}}|^2 \leq \sum_{\boldsymbol{k}\in\mathbb{N}_0^d}{}' \left| \int_{\mathbb{R}^d\setminus[-\boldsymbol{M},\boldsymbol{M}]} v(\boldsymbol{x}) e_{\boldsymbol{k}}(\boldsymbol{x}) d\boldsymbol{x} \right|^2 \leq \prod_{h=1}^{d} M_h B_v(\boldsymbol{M}) \leq \frac{\varepsilon^2}{9\eta} \tag{4.3.39}$$

the last inequality holds true if ε is small enough because the term $M_h B_v(\boldsymbol{M})$ decreases exponentially in ε since v decays exponentially, while the right-hand side of (4.3.39) goes to zero like ε^2 for $\varepsilon \searrow 0$.

We prove Corollary 4.3.12: Let $G(\boldsymbol{L})$ be defined as in Equation (4.3.28). By Lemma 4.3.9 and the Plancherel theorem, it follows that

$$\begin{aligned}
\|f1_{[-\boldsymbol{L},\boldsymbol{L}]} - \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' a_{\boldsymbol{k}} e_{\boldsymbol{k}} 1_{[-\boldsymbol{L},\boldsymbol{L}]}\|_2^2 &\leq \left| (2\pi)^{-d} \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{u})|^2 d\boldsymbol{u} - \prod_{h=1}^{d} L_h \sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}{}' |c_{\boldsymbol{k}}|^2 \right| + G(\boldsymbol{L}) \\
&\leq \frac{\varepsilon^2}{162 \left\| v1_{[-\boldsymbol{M},\boldsymbol{M}]} \right\|_2^2} + \frac{\varepsilon^2}{162 \left\| v1_{[-\boldsymbol{M},\boldsymbol{M}]} \right\|_2^2} \\
&= \left(\frac{\varepsilon}{9 \left\| v1_{[-\boldsymbol{M},\boldsymbol{M}]} \right\|_2} \right)^2 .
\end{aligned} \tag{4.3.40}$$

The last inequality holds because for $\varepsilon > 0$ small enough, $\boldsymbol{L}$ is large enough so that $G(\boldsymbol{L}) \leq \frac{\varepsilon^2}{162\|v1_{[-\boldsymbol{M},\boldsymbol{M}]}\|_2^2}$ since $G(\boldsymbol{L})$ decreases exponentially. Note that we may replace the term $\frac{\varepsilon}{12\xi}$ in Inequalities (4.3.22, 4.3.23, 4.3.24) in Corollary 4.3.7 by $\frac{\varepsilon}{9\left\|v1_{[-\boldsymbol{M},\boldsymbol{M}]}\right\|_2}$, since $c_{\boldsymbol{k}} = \tilde{c}_{\boldsymbol{k}}$. Apply Inequality (4.3.40) to conclude. □

Assume the density f in Corollary 4.3.12 is the density of a Lévy process at a particular time point and $\widehat{f}$ is real. The next Proposition 4.3.13 shows that the term $(2\pi)^{-d}\int_{\mathbb{R}^d}|\widehat{f}(\boldsymbol{u})|^2 d\boldsymbol{u}$ is then given in closed form if f is known. The densities of many Lévy processes are given explicitly or in terms of specialized functions, e.g., for the tempered Stable process, the Meixner process, the Normal Inverse Gaussian process, the Variance Gamma process and the Generalized Hyperbolic process, see Barndorff-Nielsen (1997), Madan et al. (1998), Schoutens (2003) and references therein. In the case that $\boldsymbol{\alpha} \neq \boldsymbol{0}$, the Fourier transform $\widehat{f}$ is usually not real. However, in some cases, such as the normal distribution, the integral $(2\pi)^{-d}\int_{\mathbb{R}^d}|\widehat{f}(\boldsymbol{u})|^2 d\boldsymbol{u}$ can still be obtained in closed form.

Proposition 4.3.13.

Let $(\boldsymbol{X}_t)_{t\geq 0}$ be a d-dimensional Lévy process. Assume that the characteristic function $\widehat{f}_{\boldsymbol{X}_t}$ of $\boldsymbol{X}_t$ is real for all $t > 0$ and that $\boldsymbol{X}_t$ has a density, denoted by $f_{\boldsymbol{X}_t}$. Let $T > 0$. Then

$$(2\pi)^{-d}\int_{\mathbb{R}^d}|\widehat{f}_{\boldsymbol{X}_T}(\boldsymbol{u})|^2 d\boldsymbol{u} = f_{\boldsymbol{X}_{2T}}(\boldsymbol{0}).$$

Proof. Using that $\boldsymbol{X}$ has independent and stationary increments and that $\widehat{f}_{X_T}$ is real, it follows that

$$(2\pi)^{-d}\int_{\mathbb{R}^d}|\widehat{f}_{\boldsymbol{X}_T}(\boldsymbol{u})|^2 d\boldsymbol{u} =(2\pi)^{-d}\int_{\mathbb{R}^d}\left(\widehat{f}_{\boldsymbol{X}_T}(\boldsymbol{u})\right)^2 d\boldsymbol{u} = (2\pi)^{-d}\int_{\mathbb{R}^d}\widehat{f}_{\boldsymbol{X}_{2T}}(\boldsymbol{u})d\boldsymbol{u} = f_{\boldsymbol{X}_{2T}}(\boldsymbol{0}).$$

□

Remark 4.3.14.
Provided the expression $(2\pi)^{-d}\int_{\mathbb{R}^d}|\widehat{f}(\boldsymbol{u})|^2 d\boldsymbol{u}$ can be obtained precisely, Inequality (4.3.35) makes it possible to define a stopping criterion for $\boldsymbol{N}$. In particular, Inequality (4.3.35) enables us to determine $\boldsymbol{N}$ *while* computing the coefficients $c_{\boldsymbol{k}}$: incrementally increase $\boldsymbol{N}$ and compute $|c_{\boldsymbol{k}}|$ and $|c_{\boldsymbol{k}}|^2$ simultaneously. Stop when Inequality (4.3.35) is met. However, since the right-hand side of Equation (4.3.35) converges to zero at least like $O\left(\varepsilon^2\right)$, rounding off errors makes it difficult to find $\boldsymbol{N}$ by Inequality (4.3.35) for very small ε. Using arbitrary-precision arithmetic instead of fixed-precision arithmetic should overcome this drawback.

The next theorem implies that the COS method converges exponentially if $\widehat{f}$ decays exponentially, i.e., if Inequality (4.3.41) holds for all $p > 0$. The cases (i) and (ii) in Theorem 4.3.15 treat the classical and the damped COS method, respectively. The bound for the order of convergence of the damped COS method is slightly better.

Theorem 4.3.15. *(Order of convergence)*
Assume $f \in \mathcal{L}^1 \cap \mathcal{L}^2$ satisfies Inequality (4.3.14) and decays exponentially. Assume v is bounded. Let $\gamma > 0$ and $\beta \in (0,1)$. For $n \in \mathbb{N}$, let $\boldsymbol{N} = (n,...,n) \in \mathbb{N}^d$ and $\boldsymbol{M} = \boldsymbol{L} = (\gamma n^\beta,...,\gamma n^\beta)$. Assume for some $p > \frac{d}{2}$ that

$$|\widehat{f}(\boldsymbol{u})| \leq O\left(|\boldsymbol{u}|_\infty^{-p}\right), \quad |\boldsymbol{u}|_\infty \to \infty. \tag{4.3.41}$$

(i) Define $v_{\boldsymbol{k}}$ as in Equation (4.3.11). Then it holds that

$$\left|\int_{\mathbb{R}^d} v(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x} - {\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}}' c_{\boldsymbol{k}}v_{\boldsymbol{k}}\right| \leq O\left(n^{-(1-\beta)p+\frac{d}{2}}\right), \quad n\to\infty.$$

(ii) Assume that $v \in \mathcal{L}^1 \cap \mathcal{L}^2$, v satisfies Inequality (4.3.14) and v decays exponentially. Define $\tilde{v}_{\boldsymbol{k}}$ as in Equation (4.3.12). Then it holds that

$$\left|\int_{\mathbb{R}^d} v(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x} - {\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}}' c_{\boldsymbol{k}}\tilde{v}_{\boldsymbol{k}}\right| \leq O\left(n^{-(1-\beta)(p-\frac{d}{2})}\right), \quad n\to\infty.$$

Proof. Let $A_1(\boldsymbol{L})$, $A_2(\boldsymbol{L},\boldsymbol{N})$, $D_1(\boldsymbol{M})$ and $D_2(\boldsymbol{N},\boldsymbol{L},\boldsymbol{M})$ be as in the proof of Corollary 4.3.7. Since $v_{\boldsymbol{k}} = \langle v1_{[-\boldsymbol{M},\boldsymbol{M}]}, e^{\boldsymbol{L}}_{\boldsymbol{k}}\rangle$ and similarly to the proof of Corollary 4.3.7 we have that

$$\begin{aligned}&\left|\int_{\mathbb{R}^d} v(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x} - {\sum_{\boldsymbol{0}\leq\boldsymbol{k}\leq\boldsymbol{N}}}' c_{\boldsymbol{k}}\tilde{v}_{\boldsymbol{k}}\right| \\ \leq& D_1(\boldsymbol{M}) + \|v1_{[-\boldsymbol{M},\boldsymbol{M}]}\|_2\left(A_1(\boldsymbol{L}) + A_2(\boldsymbol{L},\boldsymbol{N}) + \sqrt{B_f(\boldsymbol{L})}\right) + D_2(\boldsymbol{N},\boldsymbol{L},\boldsymbol{M}). \quad (4.3.42)\end{aligned}$$

We will analyze the order of convergence of each term at the right-hand side of Inequality (4.3.42): Since v is bounded and f decays exponentially, $D_1(\boldsymbol{M})$, $A_1(\boldsymbol{L})$ and $\sqrt{B_f(\boldsymbol{L})}$ decay exponentially, i.e., can be bounded by $O(\exp(-C_3 n^{\beta}))$, $n\to\infty$, for some C_3, see proof of Theorem 4.3.10. By Inequality (4.3.38), the term ${\sum}' |c_{\boldsymbol{k}}|^2$ is bounded. In case i), $D_2(\boldsymbol{N},\boldsymbol{L},\boldsymbol{M}) = 0$. In case ii), $D_2(\boldsymbol{N},\boldsymbol{L},\boldsymbol{M})$ decays exponentially, see proof of Corollary 4.3.11. Last, we treat $A_2(\boldsymbol{L},\boldsymbol{N})$. Let $j\in\{1,...,d\}$. Let n be large enough. Let $\boldsymbol{k}\in\mathbb{N}_0^d$ such that $k_j > n$. By Equation (4.3.9) and Inequality (4.3.41), there is a constant $a_1 > 0$ so that

$$|c_{\boldsymbol{k}}|^2 \overset{(4.3.9)}{\leq} \left(\frac{1}{2^{d-1}\prod_{h=1}^d L_h}\sum_{\boldsymbol{s}=(1,\pm1,...,\pm1)\in\mathbb{R}^d}\left|\widehat{f}\left(\frac{\pi}{2}\frac{\boldsymbol{s}\boldsymbol{k}}{\boldsymbol{L}}\right)\right|\right)^2 \overset{(4.3.41)}{\leq} a_1 n^{2\beta(p-d)}|\boldsymbol{k}|_\infty^{-2p}.$$

By mathematical induction over d and the applying the integral test of convergence, one can show that

$$\sum_{\boldsymbol{k}\in\mathbb{N}_0^d, k_j>n}|\boldsymbol{k}|_\infty^{-2p} \leq \frac{2^{d-1}}{(2p-d)n^{2p-d}}. \quad (4.3.43)$$

It follows by Inequality (4.3.43) for some $a_2 > 0$ that

$$\prod_{h=1}^d L_h \sum_{\boldsymbol{k}\in\mathbb{N}_0^d, k_j>n}|c_{\boldsymbol{k}}|^2 \leq a_2 n^{-(1-\beta)(2p-d)}. \quad (4.3.44)$$

Let $G(\boldsymbol{L})$ be defined as in Equality (4.3.28). By Equality (4.3.19), the Cauchy-Schwarz (CS) inequality and Inequality (4.3.30), we obtain

$$\begin{aligned}&A_2(\boldsymbol{L},\boldsymbol{N})^2 \\ \overset{(4.3.19)}{\leq}& \prod_{h=1}^d L_h {\sum_{k_1>N_1 \text{ or...or } k_d>N_d}}' |a_{\boldsymbol{k}} + c_{\boldsymbol{k}} - c_{\boldsymbol{k}}|^2 \\ \overset{(CS)}{\leq}& \prod_{h=1}^d L_h {\sum_{k_1>N_1 \text{ or...or } k_d>N_d}}' |c_{\boldsymbol{k}}|^2 + \prod_{h=1}^d L_h {\sum_{\boldsymbol{k}\in\mathbb{N}_0^d}}' |a_{\boldsymbol{k}} - c_{\boldsymbol{k}}|^2 \\ &+ 2\sqrt{\prod_{h=1}^d L_h {\sum_{\boldsymbol{k}\in\mathbb{N}_0^d}}' |c_{\boldsymbol{k}}|^2 \prod_{h=1}^d L_h {\sum_{\boldsymbol{k}\in\mathbb{N}_0^d}}' |a_{\boldsymbol{k}} - c_{\boldsymbol{k}}|^2}\end{aligned}$$

$$\overset{(4.3.30)}{\leq} \sum_{j=1}^{d} \Big(\prod_{h=1}^{d} L_h \sum_{\boldsymbol{k}\in\mathbb{N}_0^d, k_j>n} |c_{\boldsymbol{k}}|^2 \Big) + B_f(\boldsymbol{L}) + 2\sqrt{\Big(\int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 d\boldsymbol{x} + G(\boldsymbol{L}) \Big) B_f(\boldsymbol{L})}$$

$$\overset{(4.3.44)}{\leq} O\left(n^{-(1-\beta)(2p-d)} \right), \quad n \to \infty,$$

since $B_f(\boldsymbol{L})$ and $G(\boldsymbol{L})$, converge exponentially to zero. Since v is bounded, we have that $\|v1_{[-\boldsymbol{M},\boldsymbol{M}]}\|_2 \leq O\left(n^{\frac{d\beta}{2}}\right)$, $n \to \infty$. Noting that $\frac{-(1-\beta)(2p-d)+d\beta}{2} = -(1-\beta)p + \frac{d}{2}$, shows (i). It holds $\|v1_{[-\boldsymbol{M},\boldsymbol{M}]}\|_2 \leq \|v\|_2$ if $v \in \mathcal{L}_2$, which implies (ii). □

4.4 Characteristic functions

In this section, in Examples 4.4.1 and 4.4.2, we recall the normal and the Variance Gamma distributions from the literature. Remark 4.4.3 and Examples 4.4.4 and 4.4.5 provide a financial context.

Example 4.4.1. *(Normal distribution)*
Let $\boldsymbol{X}$ be a multivariate normal random variable with location $\boldsymbol{\eta} \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^d \times \mathbb{R}^d$. The random variable $\boldsymbol{X}$ has characteristic function $\widehat{g}(\boldsymbol{u}) = \exp\left(i\boldsymbol{\eta}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}\cdot\Sigma\boldsymbol{u}\right)$, $\boldsymbol{u} \in \mathbb{R}^d$, which can be extended to $\mathbb{C}^d$, i.e., $\widehat{g}(\boldsymbol{u} - i\boldsymbol{\alpha})$ exists for all $\boldsymbol{\alpha} \in \mathbb{R}^d$. By Proposition 4.3.2 we set $\lambda = \exp\left(-\boldsymbol{\eta}\cdot\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}\cdot\Sigma\boldsymbol{\alpha}\right)$ and $\boldsymbol{\mu} = \boldsymbol{\eta} + \Sigma\boldsymbol{\alpha}$. The characteristic function of the damped density f, defined in Equation (4.3.1), is given by $\widehat{f}(\boldsymbol{u}) = \exp\left(-\frac{1}{2}\boldsymbol{u}\cdot\Sigma\boldsymbol{u}\right)$. A straightforward computation shows that

$$(2\pi)^{-d} \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{u})|^2 d\boldsymbol{u} = \frac{2^{-d}}{\sqrt{\pi^d \det(\Sigma)}}.$$

Example 4.4.2. *(Variance Gamma distribution)*
Let $\boldsymbol{Z}$ be a d-dimensional, standard normal random variable. Let G be a Gamma distributed random variable, independent of $\boldsymbol{Z}$, with shape $a > 0$ and scale $s > 0$. Let $\boldsymbol{\eta}, \boldsymbol{\theta} \in \mathbb{R}^d$ and $\boldsymbol{\sigma} \in \mathbb{R}^d_+$. Consider $\boldsymbol{X} = \boldsymbol{\eta} + \boldsymbol{\theta}G + \sqrt{G}\boldsymbol{\sigma}\boldsymbol{Z}$. The distribution of $\boldsymbol{X}$ is denoted by $\mathrm{VG}(a, s, \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\sigma})$. Define $\Sigma \in \mathbb{R}^d \times \mathbb{R}^d$ such that $\Sigma_{ii} = \sigma_i^2$ and $\Sigma_{ij} = 0$ for $i \neq j$. Then $\boldsymbol{X}$ has the characteristic function

$$\widehat{g}(\boldsymbol{u}) = \exp\left(i\boldsymbol{\eta}\cdot\boldsymbol{u}\right)\left(1 - is\boldsymbol{\theta}\cdot\boldsymbol{u} + \frac{1}{2}s\boldsymbol{u}\cdot\Sigma\boldsymbol{u}\right)^{-a},$$

see Luciano and Schoutens (2006). The (extended) Fourier transform $\widehat{g}(\boldsymbol{u} - i\boldsymbol{\alpha})$ exists for all $\boldsymbol{\alpha} \in \mathbb{R}^d$ with $\zeta := 1 - s\boldsymbol{\theta}\cdot\boldsymbol{\alpha} - \frac{1}{2}s\boldsymbol{\alpha}\cdot\Sigma\boldsymbol{\alpha} > 0$, see Bayer et al. (2023). By Proposition

4.3.2 we set $\lambda = \exp(-\boldsymbol{\eta}\cdot\boldsymbol{\alpha})\,\zeta^a$ and $\boldsymbol{\mu} = \boldsymbol{\eta} + as\zeta^{-1}(\boldsymbol{\theta}+\Sigma\boldsymbol{\alpha})$. The characteristic function of the damped density f, defined in Equation (4.3.1), is given by

$$\widehat{f}(\boldsymbol{u}) = \exp\left(-i\frac{as}{\zeta}(\boldsymbol{\theta}+\Sigma\boldsymbol{\alpha})\cdot\boldsymbol{u}\right)\left(1 - i\frac{s}{\zeta}(\boldsymbol{\theta}+\Sigma\boldsymbol{\alpha})\cdot\boldsymbol{u} + \frac{1}{2}\frac{s}{\zeta}\boldsymbol{u}\cdot\Sigma\boldsymbol{u}\right)^{-a}.$$

Apply the Courant–Fischer–Weyl min-max principle[2] to see that $|\widehat{f}(\boldsymbol{u})| \leq O\left(|\boldsymbol{u}|_\infty^{-2a}\right)$ for $|\boldsymbol{u}|_\infty \to \infty$.

Remark 4.4.3.

In a financial context, we model d stock prices over time by a d-dimensional positive semimartingale $(\boldsymbol{S}(t))_{t\geq 0}$ on a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$. The filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions with $\mathcal{F}_0 = \{\Omega, \emptyset\}$. The logarithmic returns are defined by $\boldsymbol{X}(t) := \log(\boldsymbol{S}(t))$, $t \geq 0$. There is a bank account paying continuous compound interest $r \in \mathbb{R}$. There is a European option $w : \mathbb{R}^d \to \mathbb{R}$ with maturity $T > 0$ and payoff $w(\boldsymbol{X}(T))$ at time T. We denote by g the (risk-neutral) density of $\log(\boldsymbol{S}(T))$. The time-0 price of the European option is then given by $e^{-rT}\int_{\mathbb{R}} w(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}$. This integral can be approximated by the (damped) COS method.

Example 4.4.4. *(BS model)*

Let $\Sigma \in \mathbb{R}^d \times \mathbb{R}^d$ be a symmetric positive definite matrix. For the Black-Scholes (BS) model, the logarithmic returns $\boldsymbol{X}(T)$ are normally distributed with location $\boldsymbol{\eta} := \log(\boldsymbol{S}(0)) + (\boldsymbol{r} - \frac{1}{2}\mathrm{diag}(\Sigma))T$ and covariance matrix $T\Sigma$, where $\boldsymbol{r} = (r, ..., r) \in \mathbb{R}^d$ and $\mathrm{diag}(\Sigma) \in \mathbb{R}^d$ denotes the diagonal of Σ. We often use the following parameters: $\boldsymbol{K} = \boldsymbol{S}(0) = (100, ..., 100)$, $T = 1$, $r = 0$ and $\Sigma_{ii} = \sigma^2$, $\Sigma_{ij} = 0$, $i \neq j$, where $\sigma = 0.2$, i.e.,

$$\boldsymbol{\eta} := (4.58517, ..., 4.58517), \quad \boldsymbol{y} := \log(\boldsymbol{K}) = (4.60517, ..., 4.60517). \tag{4.4.1}$$

Example 4.4.5. *(VG model)*

Let $\nu > 0$, $\boldsymbol{\sigma} \in \mathbb{R}^d_+$ and $\boldsymbol{\theta} \in \mathbb{R}^d$. In the multivariate Variance Gamma (VG) model, see Luciano and Schoutens (2006), the logarithmic returns $\boldsymbol{X}(T)$ follow a VG$(\frac{T}{\nu}, \nu, \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\sigma})$ distribution, where

$$\eta_h := \log(S_h(0)) + \left(r + \frac{1}{\nu}\log\left(1 - \frac{1}{2}\sigma_h^2\nu - \theta_h\nu\right)\right)T, \quad h = 1, \ldots, d.$$

[2]We thank Alexey Chernov for pointing this out to us.

4.5 Functions of interest

Example 4.5.1. *(CDF)*
Let $w(\boldsymbol{x}) = 1_{(-\boldsymbol{\infty},\boldsymbol{y}]}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$ for some $\boldsymbol{y} \in \mathbb{R}^d$. The integral in (4.1.1) is equal to the CDF of the density g evaluated at $\boldsymbol{y}$. The coefficients $v_{\boldsymbol{k}}$, defined in Equation (4.3.11), can be obtained in closed form if $\boldsymbol{\alpha} = \boldsymbol{0}$. Let $\boldsymbol{M}, \boldsymbol{L} \in \mathbb{R}_+^d$ as in Section 4.3. Let $\gamma_h := \min(y_h - \mu_h, M_h)$, $h = 1, .., d$. It holds for $\boldsymbol{k} \in \mathbb{N}_0^d$ that $v_{\boldsymbol{k}} = 0$ if $\gamma_h < -M_h$ for any h, and otherwise

$$v_{\boldsymbol{k}} = \lambda^{-1} \prod_{\substack{h=1 \\ k_h=0}}^{d} \{\gamma_h + M_h\} \prod_{\substack{h=1 \\ k_h>0}}^{d} \left\{ \frac{2L_h}{\pi k_h} \Big(\sin\left(k_h \pi \frac{\gamma_h + L_h}{2L_h}\right) - \sin\left(k_h \pi \frac{-M_h + L_h}{2L_h}\right) \Big) \right\}.$$

It holds that $\|v\|_\infty \leq 1$ and $\left\| v 1_{[-\boldsymbol{M},\boldsymbol{M}]} \right\|_2^2 \leq 2^d \prod_{h=1}^d M_h$.

Next, we assume for some $\boldsymbol{\alpha} \in \mathbb{R}^d$ that the map $\boldsymbol{x} \mapsto w(\boldsymbol{x}) e^{-\boldsymbol{\alpha}\cdot\boldsymbol{x}}$ is integrable. The Fourier-transform of w then exists at all points $\boldsymbol{u} + i\boldsymbol{\alpha} \in \mathbb{C}^d$, $\boldsymbol{u} \in \mathbb{R}^d$. Let $\lambda > 0$ and $\boldsymbol{\mu} \in \mathbb{R}^d$. Let v be as in Equation (4.3.2). The Fourier-transform of v is given by $\widehat{v}(\boldsymbol{u}) = \lambda^{-1} e^{-i\boldsymbol{u}\cdot\boldsymbol{\mu}} \widehat{w}(\boldsymbol{u} + i\boldsymbol{\alpha})$. Hence, a closed form expression for $\widehat{w}$ is sufficient to obtain a closed form expression for $\widehat{v}$. We can then directly obtain $\tilde{v}_{\boldsymbol{k}}$, defined in Equation (4.3.12) via Equation (4.3.13). For many functions of interest, $\widehat{w}$ is known in closed form in d dimensions, as shown in the next examples. We include Example 4.5.2 to test the damped COS method for $\boldsymbol{\alpha} \neq \boldsymbol{0}$.

Example 4.5.2. *(CDF with damping)*
Let w as in Example 4.5.1. A simple calculation shows that the Fourier-transform of w exists for $\boldsymbol{z} \in \mathbb{C}^d$ such that $\Im\{z_h\} < 0$, $h = 1, ..., d$, and is given by $\widehat{w}(\boldsymbol{z}) = \prod_{h=1}^d \frac{K_h^{iz_h}}{iz_h}$. For $\lambda > 0$ and $\boldsymbol{\mu} \in \mathbb{R}^d$, let v be as in Equation (4.3.2). It holds for $\boldsymbol{\alpha} < \boldsymbol{0}$ that $\|v\|_\infty \leq \lambda^{-1} e^{-\boldsymbol{\alpha}\cdot\log(\boldsymbol{K})}$ and

$$\|v\|_2^2 = \lambda^{-2} \prod_{h=1}^{d} \frac{\exp\left(-2\alpha_h \big(\log(K_h)\big)\right)}{-2\alpha_h}.$$

Example 4.5.3. *(Absolute moment)*
Assume for simplicity that $d = 1$. Let $w(x) = |x| = w^+(x) + w^-(x)$, where $w^+(x) = \max(x, 0)$ and $w^-(x) = \max(-x, 0)$. The integral in (4.1.1) is equal to the absolute moment of the density g. The coefficients $v_{\boldsymbol{k}}$, defined in Equation (4.3.11), can be obtained in closed form for any $\alpha \in \mathbb{R}$ by a computer algebra system. One may approximate the positive and the negative part of w separately by the COS method, using $\alpha > 0$ for w^+ and $\alpha < 0$ for w^-, to ensure that the damped functions of interest, $v^\pm$, are bounded, respectively.

Example 4.5.4. *(Arithmetic basket put option)*
An arithmetic basket put option is defined by $w(\boldsymbol{x}) = \max(K - \sum_{h=1}^{d} e^{x_h}, 0)$, $\boldsymbol{x} \in \mathbb{R}^d$ for some $K > 0$; compare with Remark 4.4.3. The Fourier-transform of w exists for $\boldsymbol{z} \in \mathbb{C}^d$ such that $\Im\{z_h\} < 0$, $h = 1, ..., d$, and is given by

$$\widehat{w}(\boldsymbol{z}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{z}\cdot\boldsymbol{x}} w(\boldsymbol{x}) d\boldsymbol{x} = \frac{K^{(1+i\sum_{h=1}^{d} z_h)} \prod_{h=1}^{d} \Gamma(iz_h)}{\Gamma\left(i\sum_{h=1}^{d} z_h + 2\right)}. \tag{4.5.1}$$

Equation (4.5.1) follows by an elementary substitution[3] from Equation (5.14.1) in Olver et al. (2010) and is also mentioned in a similar form in Hubalek and Kallsen (2003). If $\boldsymbol{\alpha} < \boldsymbol{0}$, it holds that $\|v\|_\infty \leq \lambda^{-1} K^{1-\sum_{h=1}^{d} \alpha_h}$ and, using Equation (5.14.1) in Olver et al. (2010) once more, it follows that

$$\left\| v 1_{[-\boldsymbol{M},\boldsymbol{M}]} \right\|_2^2 \leq \|v\|_2^2 \leq \frac{K^{2-2\sum_{h=1}^{d} \alpha_h}}{\lambda^2} \frac{\prod_{h=1}^{d} \Gamma(-2\alpha_h)}{\Gamma(1 + \sum_{h=1}^{d} (-2\alpha_h))}.$$

If $d = 1$ and $\alpha = 0$, we have that $\|v\|_\infty \leq K$ and $\left\| v 1_{[-M,M]} \right\|_2^2 \leq 2MK^2$.

Example 4.5.5. *(Various other European options)*
A cash-or-nothing put option is defined by $w(\boldsymbol{x}) = 1_{[0,\boldsymbol{K}]}(e^{\boldsymbol{x}})$, $\boldsymbol{x} \in \mathbb{R}^d$ for some $\boldsymbol{K} \in \mathbb{R}_+^d$; compare with Remark 4.4.3. The option pays 1\$ at maturity if $\boldsymbol{S}(T) \leq \boldsymbol{K}$ and nothing otherwise. The integral $\int w(\boldsymbol{x}) g(\boldsymbol{x}) d\boldsymbol{x}$ is equal to $G(\log(\boldsymbol{K}))$, where G is the CDF of g. For put and call options on the maximum or minimum of d assets, see Eberlein et al. (2010); for spread options, see Hurd and Zhou (2010).

4.6 Numerical experiments

We provide several numerical experiments to solve the integral in (4.1.1) using the COS method. Reference values are obtained by Theorem 3.2 in Eberlein et al. (2010), who express the integral in (4.1.1) by another integral involving the Fourier-transforms $\widehat{g}$ and $\widehat{w}$. Eberlein et al. (2010) require a damping factor, which we set to $\boldsymbol{R} = (-4, \ldots, -4)$. To compute the reference values, we use the command *cubintegrate* with the method *cuhre* from the R-package *cubature* with relative tolerance 10^{-11}. We confirm all reference values using the COS method with $\boldsymbol{N} = (2000, \ldots, 2000)$ and a truncation range obtained from Equation (4.3.32) with $\varepsilon = 10^{-9}$ using $n = 8$ moments. For the normal distribution, in the uncorrelated case, reference values are also given in closed form for a CDF. All experiments

[3] We thank Friedrich Hubalek from TU Wien for pointing this out to us.

are performed on a modern laptop with Intel i7-11850H processor and 32 GB RAM. The COS method and Monte Carlo simulations are implemented in C++ using for-loops without parallelization. The memory requirements are minimal.

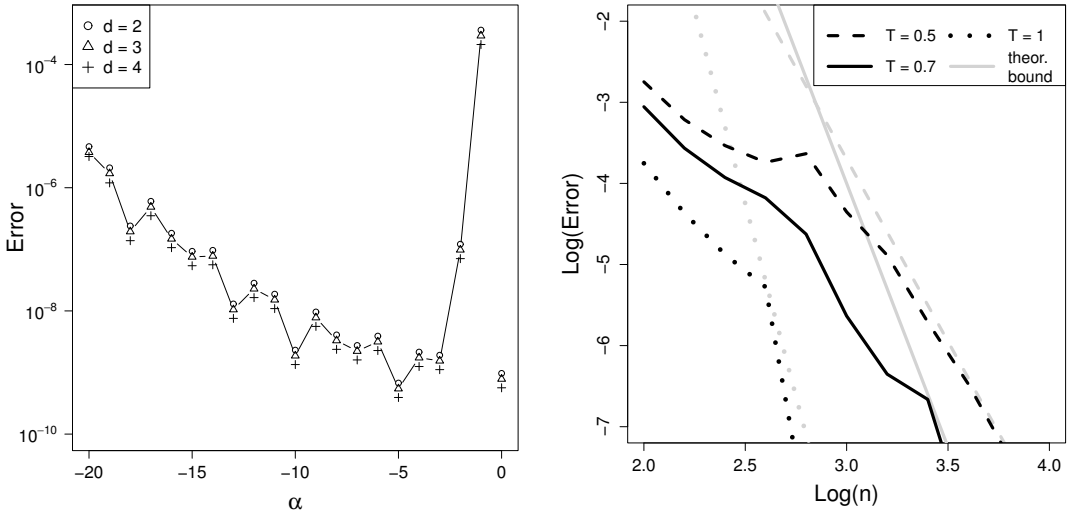

Figure 4.1: *Left*: Error of the CDF of the normal distribution for different damping factors with $\boldsymbol{\eta}$ and $\boldsymbol{y}$ as in (4.4.1) and $\Sigma_{ii} = \sigma^2$, $\Sigma_{ij} = 0$, $i \neq j$, where $\sigma = 0.2$. Further, $\boldsymbol{M} = \boldsymbol{L} = (20\sigma, ..., 20\sigma)$ and $\boldsymbol{N} = (70, ..., 70)$. Reference values are obtained by the closed form solution. *Right*: Logarithmic error of the price by the COS method for the VG model over the logarithmic number of terms for an arithmetic basket put option and $d = 2$. We choose $\boldsymbol{N} = (n, n)$ and $\boldsymbol{M} = \boldsymbol{L} = (\gamma n^\beta, \gamma n^\beta)$ with $\gamma = \beta = \frac{1}{2}$. We set $\boldsymbol{S}(0) = (50, 50)$, $K = 100$, $\boldsymbol{\sigma} = (0.2, 0.2)$, $\boldsymbol{\theta} = (-0.03, -0.03)$, $\nu = 0.1$, $r = 0$ and $\boldsymbol{\alpha} = (-4, -4)$. The theoretical bound from Theorem 4.3.15, i.e., a line with slope $-(1-\beta)(p-\frac{d}{2})$, is shown in gray. For the VG model, we have $p = \frac{2T}{\nu}$. Reference values are obtained by Eberlein et al. (2010) and are given by 3.8998, 4.6509 and 5.5951 for $T = 0.5$, $T = 0.7$ and $T = 1$, respectively.

First, we investigate the influence of the damping factor $\boldsymbol{\alpha}$ on the accuracy of the COS method to obtain the CDF of a normal distribution, where reference values can be obtained in closed form. According to Example 4.5.5, the CDF can also be interpreted as the price of a cash-or-nothing put option. Figure 4.1 shows the behavior of the COS method for different damping factors in dimensions $d \in \{2, 3, 4\}$. If $\boldsymbol{\alpha}$ is too close to zero, almost no damping takes place and the difference between $v_{\boldsymbol{k}}$ and $\tilde{v}_{\boldsymbol{k}}$ is large, which implies a relatively high error for the COS method. If $|\boldsymbol{\alpha}|$ is too big, $\|v\|_\infty$ and $\|v\|_2$ become very large and the truncation error increases. However, we observe in the example that

a wide range of damping factors work well in various dimensions. Further, fixing the number of terms $\boldsymbol{N}$ and the truncation range $\boldsymbol{L}$, the accuracy of the classical COS method with $\boldsymbol{\alpha} = \mathbf{0}$ and the damped COS method with $\boldsymbol{\alpha} \neq \mathbf{0}$ is very similar for some damping factors.

We illustrate the order of convergence of the COS method for an arithmetic basket put option in the VG model. We compare three different maturities. In Figure 4.1 we can see that the theoretical bound from Theorem 4.3.15 for the order of convergence is sharp and close to the empirical order of convergence.

4.6.1 On the choice of N

In this section, we approximate the CDF for the VG distribution and the price of an arithmetic put option in the BS model (normal distribution). We also discuss (see Table 4.1) arithmetic put options in one-dimension, which are referred to simply as *put options*. No damping is necessary to price these options using the COS method, see Fang and Oosterlee (2009a). The methodology can also be applied to other functions of interest. We compare different strategies to choose the number of terms $\boldsymbol{N}$. For $d = 1$ we also consider the bound for N from Theorem 3.8 in Junike (2024), which can be obtained as follows: If the (damped) density f is $J + 1$ times differentiable with bounded derivatives, the number of terms can be chosen by

$$N \geq \left(\frac{2^{s+\frac{5}{2}} \left\| f^{(s+1)} \right\|_\infty L^{s+2}}{s\pi^{s+1}} \frac{12 \left\| v \right\|_\infty}{\varepsilon} \right)^{\frac{1}{s}}, \tag{4.6.1}$$

where $s \in \{1, ..., J\}$. The term $\left\| f^{(s+1)} \right\|_\infty$ can be bounded by

$$\| f^{(s+1)} \|_\infty \leq \frac{1}{2\pi} \int_{\mathbb{R}} |u|^{s+1} |\varphi(u)| du. \tag{4.6.2}$$

For the BS model, the integral in Inequality (4.6.2) can be solved explicitly, and we choose $s = 40$. The density of the VG distribution is given in terms of the Bessel function of the third kind by Equation (23) in Madan et al. (1998). According to Küchler and Tappe (2008), the density of the VG distribution is $J + 1$ times continuously differentiable if J is equal to the largest natural number less than $\frac{2T}{\nu} - 2$. For the VG distribution, we use $s = J$.

In Table 4.1 one can see that Corollary 4.3.12 provides a sharper bound for $\boldsymbol{N}$ than Junike (2024). This is particularly noticeable for the VG distribution, which is less smooth

than the normal distribution. However, the formula in Junike (2024) is numerically more stable, compare with Remark 4.3.14. The number of terms obtained by Corollary 4.3.12 is roughly twice as large as the minimal number of terms necessary to stay below the error tolerance. The CPU time to solve the integral appearing in Corollary 4.3.12 is of similar magnitude in one-dimension compared to the CPU time of the COS method.

Choosing $\boldsymbol{N}$	N	CPU time COS	d	ε	$\boldsymbol{L}$	Ref. value	Function of interest/ model parameters
Minimal N	20	0.009	1	10^{-4}	0.9	0.79193	CDF of VG distribution with
Cor. 4.3.12	46	0.016					$\nu = 0.19$, $a = \frac{1}{\nu}$, $s = \nu$,
Junike (2024)	107	0.031(0.28)					$\theta = \eta = 0$, $\sigma = 0.13$
Minimal N	20	0.013	1	10^{-3}	1.3	2.5978	Put option in VG model with
Cor. 4.3.12	64	0.030(0.16)					$\sigma = 0.1213$, $\theta = -0.1436$,
Junike (2024)	152	0.065(0.26)					$\nu = 0.1686$
Minimal N	10	0.005	1	10^{-2}	1.2	3.9827	Put option in BS model with
Cor. 4.3.12	16	0.006					$\Sigma = 0.2^2$
Junike (2024)	20	0.007					
Minimal $\boldsymbol{N}$	40	2.32	2	10^{-2}	$(3.9, 7.9)$	10.5051	Basket put in BS model with
Cor. 4.3.12	72	7.04					$\Sigma_{11} = 0.2^2$, $\Sigma_{22} = 0.4^2$,
							$\Sigma_{12} = \Sigma_{21} = \frac{1}{2}\sqrt{\Sigma_{11}\Sigma_{22}}$

Table 4.1: Comparison of different strategies to choose $\boldsymbol{N}$ to obtain the CDF for the VG$(a, s, \eta, \theta, \sigma)$ distribution at $y = 0.1$ without damping and the price of an arithmetic basket put option in the BS and the VG models. For the (basket) put option, we set $\boldsymbol{S}(0) = (50, ..., 50)$, $K = 50d$, $T = 1$, $r = 0$ and $\boldsymbol{N} = (N, \dots, N)$. We obtain the truncation range $\boldsymbol{L} = (L, ..., L)$ from Inequality (4.3.32) using $n = 8$ moments. Damping is only necessary for the two-dimensional basket option, where we set $\boldsymbol{\alpha} = (-4, -4)$. Reference values are obtained by Eberlein et al. (2010). We average over ten runs to obtain the CPU time, which is measured in milliseconds. The CPU times to solve the integrals in Inequality (4.6.2) and Corollary 4.3.12 for the VG model, using R's *integrate* function with default values, are reported in brackets.

4.6.2 Comparison with Monte Carlo

We compare the COS method with a Monte Carlo (MC) simulation to obtain the CDF of the normal distribution, where reference values can be obtained in closed form. According to Example 4.5.5, the CDF can also be interpreted as the price of a cash-or-nothing put option. The computational complexity of a MC simulation with $U \in \mathbb{N}$ runs scales like $O(U)$. We estimate U by the central limit theorem using a statistical error of 0.99. The COS method consists of d-nested sums. According to Equation (4.3.9), the computational complexity of the COS method scales like $O\left(\prod_{h=1}^{d}\{N_h\}\right)$. A MC simulation converges relatively slowly, but it scarcely depends on the dimension. On the other hand, the complexity of the COS method grows exponentially in the dimension; however, the COS method also converges exponentially for the normal distribution. The choice between MC and the COS method depends both on the dimension and on the error tolerance ε: the higher d the better MC compares to the COS method, but the smaller ε, the faster the COS method performs.

d	N	L	U	τ_{COS}	τ_{MC}	Reference value
1	30	2.0	16481995016	8.9e-6	5.1e+3	0.539827
2	30	2.4	13700525367	3.7e-4	1.6e+4	0.291414
3	40	3.0	8795611829	4.9e-2	1.2e+4	0.157313
4	50	3.6	5156004587	1.1e+1	8.1e+3	0.084922
5	50	4.2	2902219256	1.4e+3	7.8e+3	0.045843

Table 4.2: CPU time of the COS method (τ_{COS}) and CPU time of a MC simulation (τ_{MC}) to obtain the CDF of a normal distribution. We set $\varepsilon = 10^{-5}$, $\boldsymbol{\alpha} = (-7, \dots, -7)$, $\Sigma_{ii} = \sigma^2$, $\Sigma_{ij} = 0$, $i \neq j$, where $\sigma = 0.2$, $T = 1$ and let $\boldsymbol{\eta}$ and $\boldsymbol{y}$ as in (4.4.1). We set $N_1 = \cdots = N_d = N$. We obtain the truncation range $\boldsymbol{L} = (L, ..., L)$ from Inequality (4.3.32) using $n = 8$ moments. The reference value can be obtained in closed form. CPU time is measured in seconds.

In several numerical experiments, we observe that the COS method is faster than MC for $d \leq 4$ and $\varepsilon \leq 10^{-3}$. For $d = 5$, the COS method outperforms MC for $\varepsilon \leq 10^{-5}$; otherwise, a MC simulation is faster. If $\varepsilon = 10^{-9}$ and $d = 4$, the COS method needs 220 terms in each dimension to stay below the error tolerance, and the CPU time is about one hour. We estimate that a MC simulation would take longer than $20,000$ years. Some experiments are reported in Table 4.2 and Figure 4.2.

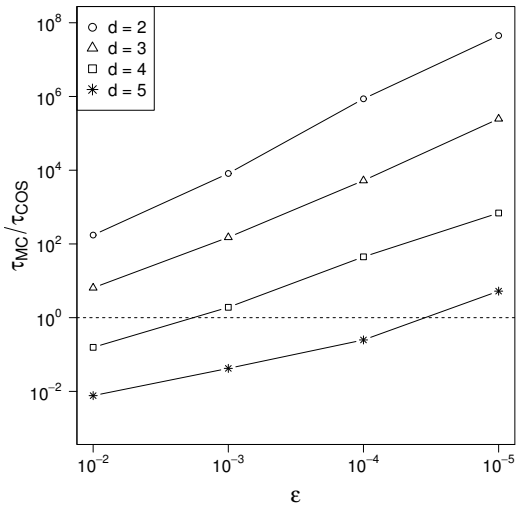

Figure 4.2: Ratio of the CPU time of the COS method (τ_{COS}) and the CPU time of a MC simulation (τ_{MC}). Parameters are as in Table 4.2.

4.7 Conclusions

In this article we introduced and discussed the damped COS method, which is a numerical tool to solve certain multidimensional integrals numerically, e.g., to obtain a CDF from a characteristic function or to price a financial contract. The (damped) COS method requires several parameters: In particular, one has to specify a truncation range $\boldsymbol{L}$ for the density f and the number of terms $\boldsymbol{N}$ of cosine functions to approximate the truncated density. Corollary 4.3.7 provides sufficient conditions on $\boldsymbol{L}$ and $\boldsymbol{N}$ to ensure the convergence of the COS method within a given error tolerance $\varepsilon > 0$. Theorem 4.3.10 and Corollary 4.3.12 provide formulas for the truncation range $\boldsymbol{L}$ and the number of terms $\boldsymbol{N}$, respectively. Theorem 4.3.15 provides an upper bound of the order of convergence of the COS method. Numerical experiments indicate that the bound is sharp. In particular, the (damped) COS method converges exponentially if the Fourier transform $\widehat{f}$ decays exponentially.

Chapter 5

Enhancing Fourier pricing with machine learning

Abstract

Fourier pricing methods such as the Carr-Madan formula or the COS method are classic tools for pricing European options for advanced models such as the Heston model. These methods require tuning parameters such as a damping factor, a truncation range, a number of terms, etc. Estimating these tuning parameters is difficult or computationally expensive. Recently, machine learning techniques have been proposed for fast pricing: they are able to learn the functional relationship between the parameters of the Heston model and the option price. However, machine learning techniques suffer from error control and require retraining for different error tolerances. In this research, we propose to learn the tuning parameters of the Fourier methods (instead of the prices) using machine learning techniques. As a result, we obtain very fast algorithms with full error control: Our approach works with any error tolerance without retraining, as demonstrated in numerical experiments using the Heston model.

Keywords: Machine learning; computational finance; option pricing; Fourier pricing; error control; Heston model

5.1 Introduction

Fourier methods, such as the Carr-Madan formula and the COS method, see Carr and Madan (1999) and Fang and Oosterlee (2009a), are widely used to price European options. In order to speed up option pricing, Liu et al. (2019a), Liu et al. (2019b), Yang et al. (2017)

and Sirignano and Spiliopoulos (2018) propose a prediction of option prices using neural networks. Ruf and Wang (2020) provide a comprehensive review of neural networks for option pricing. Liu et al. (2019a) and Liu et al. (2019b) use a parametric approach and consider an advanced stock price model, such as the Heston model, see Heston (1993). They use a set of market parameters, including strike price and maturity, as well as model parameters, to predict the corresponding option prices. De Spiegeleer et al. (2018) use machine learning techniques based on Gaussian process regression for prediction of option prices.

While De Spiegeleer et al. (2018), Liu et al. (2019a) and Liu et al. (2019b) were able to accelerate the existing Fourier methods to some extent, their approaches also exhibited certain limitations. Liu et al. (2019a) and Liu et al. (2019b) obtain a mean absolute error (MAE) of about 10^{-4}. De Spiegeleer et al. (2018) also obtain a MAE of about 10^{-4} and a maximum absolute error of approximately 10^{-3} on their sample. In De Spiegeleer et al. (2018), Table 2, the authors compare the numerical effort with the Carr-Madan formula and obtain an acceleration factor between 10 and 40 for European options.

However, the approaches described in Liu et al. (2019a), Liu et al. (2019b) and De Spiegeleer et al. (2018) suffer from a lack of error control: To achieve higher numerical pricing accuracy, deeper neural networks are necessary and the machine learning methods need to be retrained with more samples, which is very time-consuming and impractical in most situations.

In this paper, we propose an indirect use of machine learning methods to improve the accuracy and efficiency of existing pricing techniques with full error control. We focus on the COS method, but our approach is also applicable to other methods, i.e., we also discuss the Carr-Madan formula.

We describe the main idea of the COS method, details can be found, e.g., in Fang and Oosterlee (2009a), Oosterlee and Grzelak (2019) and Junike and Pankrashkin (2022): Given only the characteristic function of the log-returns of the underlying, the density of the log-returns is approximated in two steps: i) truncate the density on a finite interval $[a, b]$ and ii) approximate the truncated density by a finite Fourier-cosine approximation with N terms. There is a clever trick to obtain the cosine-coefficients of the truncated density efficiently from the characteristic function. The CPU time of the COS method depends linearly on the number of terms N. Note that the choice of the truncation range has a significant influence on the number of terms required to achieve a certain accuracy. There are explicit formulas for the truncation range and the number of terms depending on an error tolerance $\varepsilon > 0$, see Junike and Pankrashkin (2022) and Junike (2024). However, the

truncation range formula requires evaluating higher-order derivatives of the characteristic function, which can be very time-consuming, e.g., in the case of the Heston model. The formula for the number of terms requires integration of the product of the characteristic function and a polynomial, which is also very time consuming. Fortunately, the time-consuming part required to obtain $[a, b]$ and N *does not depend on the required error tolerance* ε.

In this paper, we use machine learning techniques to learn the n-th derivatives of the characteristic function evaluated at zero and learn the integral of the characteristic function times a polynomial, which is independent of the required error tolerance. Then, we use these predicted values and the error tolerance to obtain the truncation range and the number of terms. The COS method can then be applied to price European options.

Different traders may use different error tolerances, but our machine learning techniques do not require retraining. This error control is an advantage over direct prediction of option prices by machine learning techniques. The actual calculation of the option price using the COS method is then very fast.

The paper is structured as follows. Section 5.2 gives an overview of the Heston model, which will be used in the numerical experiments. In Section 5.3, we introduce the COS method and the Carr-Madan formula and machine learning techniques. Section 5.4 provides the numerical experiments to demonstrate the performance of the proposed method. Section 5.5 concludes the paper.

5.2 The Heston model

Consider a financial market with a riskless bank-account and a stock with deterministic price $S_0 > 0$ today and random price S_T at some future date $T > 0$. In the Heston model with parameters $\kappa > 0$, $\theta > 0$, $\xi > 0$, $\rho \in [-1, 1]$ and $v_0 > 0$, the stock price is described by the following system of differential equations

$$\frac{dS_t}{S_t} = rdt + \sqrt{v_t}dW_t, \;\; S_0 \geq 0 \tag{5.2.1}$$

$$dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t}dZ_t. \tag{5.2.2}$$

W and Z are correlated Brownian motions such that $\operatorname{cov}[dW_t dZ_t] = \rho dt$, see Heston (1993).

The CIR process, described by Equation (5.2.2), stays positive if $2\kappa\theta \geq \xi^2$, which is known as the *Feller condition*, see Andersen and Piterbarg (2007). The characteristic function of the log stock price, see Bakshi et al. (1997), is given by

$$\begin{aligned}\varphi_{\log(S_t)}(u) &= E\left[\exp(iu\log\left(S_t\right)\right] \\ &= \exp\left(iu\left(\log S_0 + rt\right)\right) \\ &\quad \times \exp(\theta\kappa\xi^{-2}\left(\left(\kappa - \rho\xi ui - d\right)t - 2\log\left(\frac{1-ge^{-dt}}{1-g}\right)\right)) \\ &\quad \times \exp\left(v_0\xi^{-2}\frac{\left(\kappa - \rho\xi iu - d\right)\left(1-e^{-dt}\right)}{1-ge^{-dt}}\right),\end{aligned}$$

where

$$\begin{aligned}d &= \left(\left(\rho\xi ui - \kappa\right)^2 - \xi^2\left(-iu - u^2\right)\right)^{\frac{1}{2}}, \\ g &= \frac{\kappa - \rho\xi ui - d}{\kappa - \rho\xi ui + d}.\end{aligned}$$

5.3 Algorithms: Numerical tools and machine learning

5.3.1 The Carr-Madan formula

Carr and Madan (1999) showed that the price of a European call option with strike K and time to maturity T is given by

$$e^{-\alpha\log(K)}e^{-rT}\frac{1}{\pi}\int_0^\infty \Re\left\{e^{-iv\log(K)}\frac{\varphi_{\log(S_T)}\left(v - i(\alpha+1)\right)}{\alpha^2 + \alpha - v^2 + i(2\alpha+1)v}\right\}dv, \tag{5.3.1}$$

where $\alpha > 0$ is a damping factor such that $E[S_T^{1+\alpha}] < \infty$ and $\varphi_{\log(S_T)}$ is the characteristic function of $\log(S_T)$. $\Re(z)$ denotes the real part of a complex number z and $i = \sqrt{-1}$ is the complex unit. The integral in Eq. (5.3.1) can be truncated to $(0, M)$, for some $M > 0$, and then be evaluated using, e.g., Simpson's rule with N grid points.

5.3.2 The COS method

We summarize the COS method. This section is based on Fang and Oosterlee (2009a), Junike and Pankrashkin (2022) and Junike (2024). Let μ be the expectation of $\log(S_T)$ under the risk-neutral measure and assume that the characteristic function φ_X of the centralized log-returns $X := \log(S_T) - \mu$ is given in closed-form. The function φ_X is

explicitly given for many models such as the Heston model. The price of a European put option with maturity $T > 0$ and strike $K > 0$ is given by

$$\int_{\mathbb{R}} e^{-rT} \max\left(K - e^{x+\mu}, 0\right) f(x)dx, \tag{5.3.2}$$

where f is the density of X. The price of a call option can be obtained by the put-call-parity. Very often, f is not explicitly given and the COS method can be used to approximate f and the price of the option.

For some $L > 0$, the density f is truncated and the truncated density is approximated by a cosine series expansion:

$$f(x) \approx \frac{c_0}{2} + \sum_{k=1}^{N} c_k \cos\left(k\pi \frac{x+L}{2L}\right), \quad x \in [-L, L], \tag{5.3.3}$$

where for $k = 0, 1, ..., N$, the coefficients c_k are defined by

$$c_k := \frac{1}{L} \int_{\mathbb{R}} f(x) \cos\left(k\pi \frac{x+L}{2L}\right) dx = \frac{1}{L} \Re\left\{ \varphi\left(\frac{k\pi}{2L}\right) e^{i\frac{k\pi}{2}} \right\}. \tag{5.3.4}$$

The second Equality in (5.3.4) follows from a simple analysis, see Fang and Oosterlee (2009a). The price of a European put option can be approximated by replacing f in (5.3.2) with its approximation (5.3.3), which gives

$$\int_{\mathbb{R}} e^{-rT} \max\left(K - e^{x+\mu}, 0\right) f(x)dx \approx \frac{c_0 v_0}{2} + \sum_{k=1}^{N} c_k v_k,$$

where

$$v_k := \int_{-L}^{L} e^{-rT} \max\left(K - e^{x+\mu}, 0\right) \cos\left(k\pi \frac{x+L}{2L}\right) dx, \quad k \in \{0, 1, 2, ...\}.$$

The coefficients c_k are given in closed form when φ_X is given analytically and the coefficients v_k can also be computed explicitly in important cases, e.g., for plain vanilla European put or call options and digital options, see Fang and Oosterlee (2009a). This makes the COS method numerically very efficient and robust.

We provide formulas for the coefficients v_k for a European put option: Let $d := \min\left(\log\left(K\right) - \mu, L\right)$. For a European put option, it holds that $v_k = 0$ if $d \leq -L$ and otherwise

$$v_k = e^{-rT} \left(K\Psi_0(k) - e^{\mu}\Psi_1(k)\right),$$

where

$$\Psi_0(k) = \begin{cases} d + L & , k = 0 \\ \frac{2L}{k\pi} \sin\left(k\pi \frac{d+L}{2L}\right) & , k > 0 \end{cases}$$

and

$$\Psi_1(k) = \frac{e^d\left(\frac{k\pi}{2L}\sin\left(k\pi\frac{d+L}{2L}\right) + \cos\left(k\pi\frac{d+L}{2L}\right)\right) - e^{-L}}{1+\left(\frac{k\pi}{2L}\right)^2}, \quad k \geq 0,$$

see Appendix A in Junike and Pankrashkin (2022). To price a call option it is numerically more stable to price a put option instead and use the put-call parity, see Fang and Oosterlee (2009a).

To apply the COS method, one has to specify the truncation range $[-L, L]$ and the number of terms N. For a given error tolerance ε small enough, both parameters can be chosen as follows to ensure an approximation error smaller than ε, see Junike and Pankrashkin (2022) and Junike (2024). If ε is small enough and f has semi-heavy tails, the truncation range of a put option can be chosen using Markov's inequality by

$$L = L(\varepsilon, \mu_n) = \mu_n \times \left(\frac{2Ke^{-rT}}{\varepsilon}\right)^{\frac{1}{n}}, \tag{5.3.5}$$

where $n \in \mathbb{N}$ is even and μ_n is the n-th root of the n-th moment of X, which can be obtained using a computer algebra system and the relation

$$\mu_n = \sqrt[n]{\frac{1}{i^n}\,\frac{\partial^n}{\partial u^n}\varphi_X(u)\Big|_{u=0}}. \tag{5.3.6}$$

Often, $n \in \{4, 6, 8\}$ is a reasonable choice, see Corollary 9 in Junike and Pankrashkin (2022). If f is also $s+1 \in \mathbb{N}$ times differentiable with bounded derivatives, then the number of terms can be chosen by

$$N = N(\varepsilon, I_s) = I_s \times \left(\frac{2^{s+\frac{5}{2}}L^{s+2}}{s\pi^{s+1}}\frac{12Ke^{-rT}}{\varepsilon}\right)^{\frac{1}{s}}, \tag{5.3.7}$$

where

$$I_s := \left(\frac{1}{2\pi}\int_{\mathbb{R}}|u|^{s+1}|\varphi_X(u)|du\right)^{\frac{1}{s}}, \tag{5.3.8}$$

see Equation (3.8) in Junike (2024). The last integral can be solved numerically by standard techniques and in some cases it is given explicitly. One should choose s such that the left-hand side of Inequality (5.3.7) is minimized. For the Heston model, s is set to $s = 20$ in Junike (2024). An implementation of the truncation range, the number of terms and the COS method for the Heston model can found in Appendix 5.A.3.

5.3.3 Machine learning techniques

Decision Tree: Decision trees (DT), see Breiman et al. (1984), operate by recursively partitioning the input data into subsets, thereby forming a tree-like structure, see Table 5.1 and Figure 5.1. At each internal node of the DT, the algorithm selects a feature and a threshold value to split the data into two subsets.

For example, in the first row of Table 5.1, all input values with maturity T less than or equal to 0.1019998 are assigned to node 1, all other values are assigned to node 2. The goal of these splits is to create child nodes with greater homogeneity. The recursive splitting process continues until a stopping criterion is met, such as a maximum tree depth or a minimum node size for splitting.

To build a DT for regression, the splitting is based on variance reduction. The algorithm selects the features and thresholds that most strongly reduce the variance at each node for splitting.

Given new samples, predictions are made at the leaf nodes, where the model assigns the average of the data points within the node. This simplicity and transparency make DT highly effective at handling complex data sets while maintaining interpretability.

nodeID	leaf node	variable	split value	left-child (if variable $<$ split value)	right-child (if variable $\geq$ split value)	prediction
0	No	T	0.101999	1	2	NA
1	Yes	NA	NA	NA	NA	46.988648
2	No	v_0	0.838772	3	4	NA
3	Yes	NA	NA	NA	NA	14.185356
4	Yes	NA	NA	NA	NA	2.344154

Table 5.1: Example of a DT.

Random Forest: Random forests (RF), see Breiman (2001) are an ensemble of DTs to improve the accuracy and robustness of predictions. Each DT in the RF is trained on a random subset of the data using bootstrap aggregation. At each node, a random subset of the features is used for the splitting. In a RF, each DT makes a prediction independently and the final output is determined by averaging the individual predictions of each single tree.

Neural Networks: A neural network (NN) consists of one or more layers, each consisting of a number of artificial neurons, see Goodfellow et al. (2016). A single neuron transforms its multidimensional input $x \in \mathbb{R}^n$ into a one-dimensional output. For some weights $w \in \mathbb{R}^{n+1}$, the weighted mean of the input is then transformed by an activation function $g : \mathbb{R} \to \mathbb{R}$, i.e., the output of a neuron is given by $g\left(\sum_{i=1}^{n} w_i x_i + w_{n+1}\right)$. Examples of activation functions are the ReLU function $g(y) = \max(y, 0)$ or the Sigmoid function $g(y) = \frac{1}{1+e^{-x}}$. In the first layer of the NN, the neurons receive the input data and the output of each neuron is passed to all neurons in the following layers until the last layer is reached.

At the start of training, the weights of the NN are randomly initialized. During the training phase, the weights are chosen in such a way that the functional relationship between input and output data is mapped as well as possible.

In this work, we test the following regularization techniques that can improve the robustness of the NN: Dropout means randomly deactivating some neurons. Gaussian noise is a regularization technique that adds normally distributed numbers with zero mean and small variance to each weight at each update step. Batch normalization standardizes the inputs of each layer. These and other regularization techniques are discussed in detail in, for example, Goodfellow et al. (2016).

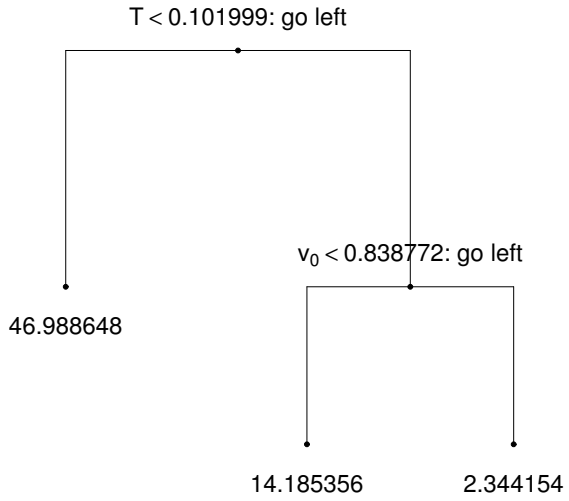

Figure 5.1: Example of a DT as in Table 5.1.

5.4 Numerical experiments

In this section, we use the machine learning techniques DT, NN and RF to predict the tuning parameters of the Carr-Madan formula and the COS method. For training, we randomly generate parameters of the Heston model. The ranges of the six parameters are shown in Table 5.2. The wide ranges of these parameters include parameters that are typically used for the Heston model, see Andersen (2008), Crisóstomo (2015), Cui et al. (2017), Engelmann et al. (2021), Fang and Oosterlee (2009a), Forde et al. (2012), Levendorskiĭ (2012) and Schoutens et al. (2003).

Parameter	Value range
speed of mean reversion κ	$[10^{-3}, 10]$
level of mean reversion θ	$[10^{-3}, 2]$
volatility of variance ξ	$[10^{-2}, 5]$
correlation coefficient ρ	$[-0.99, 0.99]$
initial variance v_0	$[10^{-3}, 2]$
time to maturity T	$[250^{-1}, 10]$

Table 5.2: Range of parameters of the Heston model, including parameters that are typically used.

For each sample (consisting of the five parameters for the Heston model and the maturity), we compute μ_4 and μ_8 and I_{20} for the entire data set, using Eqs. (5.3.6, 5.3.8). The derivatives of φ_X are calculated using a computer algebra system. As a side note: One may also approximate the moments as in Choudhury and Lucantoni (1996) to avoid the computation of the derivatives.

We exclude all the model parameters for which Eq. (5.3.6) gives negative results, assuming that the moments do not exist in these cases and we remove all parameters for which the Feller condition $2\kappa\eta \geq \xi^2$ is not satisfied.

In the following numerical experiments, we price a European call option with $S_0 = 100$, strike $K = 100$ and interest rate $r = 0$. We also tested other strikes, i.e., $K \in \{75, 125\}$ and obtained similar results. For each sample, we calculate a reference price. To obtain the reference prices we use the COS method with truncation range $L(\varepsilon, \mu_8)$ and number of terms $N(\varepsilon, I_{20})$, where we set $\varepsilon = 10^{-9}$. To confirm the prices we use the Carr-Madan formula with truncation range $M = 1024$, $N = 2^{20}$ and appropriate damping factors. We remove a few samples where the prices were too unstable and the COS method and the Carr-Madan

formula give completely different results. For all remaining options, the COS method and the Carr-Madan formula coincide at least up to seven decimal place.

We receive a cleaned data set of 250,000 samples. We take 100,000 samples for training and validation and use the remaining 50,000 samples as a test set. All experiments are run on a laptop with an Intel i7-11850H processor and 32 GB of RAM.

5.4.1 On the tuning parameters of the COS method

To apply the COS method, we use the formulas for the truncation range and the number of terms in Eq. (5.3.5) and (5.3.7). For the Heston model, it is time-consuming to compute μ_8 in Eq. (5.3.6) and to solve the integral I_{20} in Eq. (5.3.8). Therefore, we use the machine learning techniques DT, RF and NN for a fast estimation of μ_8 and I_{20}.

To identify an appropriate architecture for the different machine learning techniques, we perform a rough hyperparameter optimization. For the DT, we optimize over the maximum depth and the minimum node size. In addition, the number of DTs in the RF is optimized, resulting in the hyperparameters shown in Table 5.3. The R package *ranger* is used for both DT and RF. We consider a big DT (bDT) of arbitrary depth and a small DT (sDT) of depth 5. The sDT for μ_8 and the sDT for I_{20} are tabulated in Appendix 5.A.3 and 5.A.3 and could be implemented directly without using additional software packages.

Parameters	bDT for I_{20}	bDT for μ_8	RF for I_{20}	RF for μ_8	sDT for I_{20}	sDT for μ_8
Number of DT	1	1	500	600	1	1
Maximal tree depth	30	unlimited	50	90	5	5
Minimal node size to split at	8	6	1	1	5	5

Table 5.3: Selected hyperparameters of the DT and RF.

The architectural specifications of the NN are described in Table 5.4. The NN is trained with 100 epochs, a validation split of 0.2 and the mean squared error (MSE)

$$\text{MSE}(y, \tilde{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2, \quad y, \tilde{y} \in \mathbb{R}^d,$$

as the loss metric. For the starting values of the weights we use the He initialization, see He et al. (2015). For the NN, we use tensorflow via the keras package called from R.

Table 5.5 shows the MSE on the test set for the different machine learning techniques. It can be observed that for μ_8, the NN has a smaller MSE than the RF, while the bDT has a comparatively large MSE. With regard to I_{20}, the RF has the smallest MSE, while the MSE of the NN and the bDT are about 40% larger. The sDT has a significantly larger MSE for both μ_8 and I_{20}.

Parameters	Optimization range	NN for I_{20}	NN for μ_8
Hidden layers	$\{1, \dots, 4\}$	4	3
Neurons	$\{32, 64, 128, ..., 2048\}$	1024, 256, 256, 32	256, 128, 32
Activation function	ReLU, Leaky ReLU, Sigmoid, ELU, tanh	Sigmoid	Sigmoid
Dropout rate	$\{0, 0.1, 0.2, ..., 0.5\}$	0.2	0
Noise rate	$\{0.01, 0.02, ..., 0.1\}$	0.07	0.02
Optimizer	Adam, SGD, RMSProp	Adam	Adam
Batch normalization	yes, no	no	no
Batch size	$\{128, 256, 512, 1024\}$	512	256

Table 5.4: Selected hyperparameters of the NN.

	RF	NN	bDT	sDT
μ_8	0.0703	0.0058	0.2764	2.2390
I_{20}	33.6615	44.2353	49.0372	61.9859

Table 5.5: MSE of the prediction of μ_8 and I_{20} for different ML techniques on the test set.

Next, we calculate the price of the call option for different model parameters. We use the COS method with $L(\varepsilon, \mu_4)$ or $L(\varepsilon, \mu_8)$ and $N(\varepsilon, I_{20})$, where $\varepsilon \in \{10^{-1}, ..., 10^{-7}\}$.

The Table 5.6 shows the percentage of samples in the test set for which the required accuracy is achieved by obtaining μ_n and I_s directly from Eqs. (5.3.6, 5.3.8), which is very time-consuming, or by estimating μ_n and I_s via DTs, RF or a NN, which is very fast. The direct way of obtaining μ_8 and I_{20} and the estimation by the RF result in 100% accurate option prices on the test set for all ε. The NN also achieves a high accuracy of about

99.98% for all ε. This result could be further improved with a different NN architecture and additional training. It can be observed that a single bDT is also able to estimate I_{20} and μ_8 with sufficient accuracy to price the call option with different error bounds for at least 99.96% of the samples. And even a simple technique like the sDT already achieves an accuracy of at least 98% on the test set.

These very good results are a consequence of the fact that the formulas in Eqs. (5.3.5) and (5.3.7) are derived using many inequalities, thus overestimating the minimum truncation range L and the number of terms N needed to accurately price the option. Therefore, a rough estimate of μ_8 and I_{20} is sufficient for precise option pricing.

ε	Ana. calc. of μ_4 and num. integration of I_{20}	Ana. calc. of μ_8 and num. integration of I_{20}	μ_8 and I_{20} via RF	μ_8 and I_{20} via NN	μ_8 and I_{20} via bDT	μ_8 and I_{20} via sDT
10^{-1}	0.999%	100%	100%	100%	100%	99.904%
10^{-2}	0.999%	100%	100%	99.996%	99.998%	99.684%
10^{-3}	100%	100%	100%	99.994%	99.998%	99.320%
10^{-4}	100%	100%	100%	99.988%	99.986%	98.824%
10^{-5}	100%	100%	100%	99.986%	99.976%	98.512%
10^{-6}	100%	100%	100%	99.988%	99.970%	98.288%
10^{-7}	100%	100%	100%	99.990%	99.968%	98.192%

Table 5.6: Accuracy of the COS method for different error tolerances ε on the test set for a call option with $S_0 = 100$ and $K = 100$ with μ_4, μ_8 and I_{20} calculated directly and via DTs, RF and a NN.

The Table 5.7 illustrates the CPU time of the COS method, where L and N are obtained by different error tolerances. The COS method is implemented in C++ using for-loops without parallelization. It is well known, that $L(\varepsilon, \mu_8)$ is usually closer to the optimal truncation range than $L(\varepsilon, \mu_4)$, see Junike and Pankrashkin (2022). It is therefore not surprising that the average CPU time is about 10 times faster using the truncation range $L(\varepsilon, \mu_8)$ compared to $L(\varepsilon, \mu_4)$, see Table 5.7.

ε	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-7}$
$L(\varepsilon, \mu_4)$	$5.89 \cdot 10^{-5}$	$1.15 \cdot 10^{-4}$	$2.34 \cdot 10^{-4}$	$4.86 \cdot 10^{-4}$	$1.02 \cdot 10^{-3}$	$2.11 \cdot 10^{-3}$
$L(\varepsilon, \mu_8)$	$3.38 \cdot 10^{-5}$	$4.66 \cdot 10^{-5}$	$6.67 \cdot 10^{-5}$	$9.80 \cdot 10^{-5}$	$1.47 \cdot 10^{-4}$	$2.20 \cdot 10^{-4}$

Table 5.7: Average CPU time (in sec.) on the test set of the COS method with truncation range $L(\varepsilon, \mu_4)$ or $L(\varepsilon, \mu_8)$ and number of terms $N(\varepsilon, I_{20})$ to price a call option with $S_0 = 100$ and $K = 100$. Here, we only take into account the CPU time of the COS method ignoring the CPU time to estimate L and N.

Ana. calc. of μ_4 and num. integration of I_{20}	μ_8 and I_{20} via RF	μ_8 and I_{20} via NN	μ_8 and I_{20} via bDT	μ_8 and I_{20} via sDT
$1.122 \cdot 10^{-2}$	$6.921 \cdot 10^{-4}$	$7.056 \cdot 10^{-5}$	$2.607 \cdot 10^{-6}$	$2.036 \cdot 10^{-6}$

Table 5.8: Average CPU time on the test set in sec. for calculating μ_n and I_{20} directly or using machine learning techniques.

Let us set $\varepsilon = 10^{-4}$ and let us consider two scenarios: i) A trader estimates μ_4 and I_{20} directly. (Estimating μ_8 directly is too time consuming for the Heston model). ii) A trader estimates μ_8 and I_{20} using machine learning techniques. From Table 5.6, we can see that both approaches will price the options very accurately for different error tolerances and parameters of the Heston model. What is the impact on the total CPU time? As shown in Table 5.8, the CPU time to obtain μ_4 and I_{20} directly takes about 0.011sec. (Most of the time is used to estimate I_{20}, we used R's function *integrate* with default values for numerical integration). The computation of μ_4 and I_{20} dominates the total CPU time, since the pure application of the COS method takes about $2.34 \cdot 10^{-4}$ sec., see Table 5.7. On the other hand, the CPU time to estimate μ_8 and I_{20} using machine learning techniques is about a factor of 100 to $1,000$ times faster than the direct computation of μ_4 and I_{20}. The total CPU time of the COS method estimating μ_8 and I_{20} via a NN is about $1.4 \cdot 10^{-4}$ sec. In summary, approach ii) is almost 100 times faster than approach i).

5.4.2 On the tuning parameters of the Carr-Madan formula

In order to apply the Carr-Madan formula, one must specify three parameters, namely the damping factor $\alpha > 0$, the truncation range M and the number of grid points N. In

the following, we use a NN and a RF to estimate these parameters. We set $M = 1200$ and determine optimal parameters α and N for the entire training set, such that N is minimal to achieve an error bound of 10^{-7}. We then train a NN and a RF to learn these optimal parameters. Since the estimate $\hat{N}$ of the NN and the RF sometimes significantly underestimates the true N, we double the output of the NN and the RF to improve the accuracy of the Carr-Madan formula. This step was not necessary for the COS method, since the theoretical formulas for the truncation range and number of terms are larger than the minimal truncation range and number of terms.

To measure the accuracy of the Carr-Madan formula, we price a call option with $S_0 = K = 100$ and $r = 0$, using the predicted values for α and N of the NN and the RF. We obtain the required accuracy of $\varepsilon = 10^{-7}$ for 90.55% and 93.49% of the samples in the test set for the RF and the NN, respectively.

To compare these results, we also use standard parameters of the Carr-Madan formula: Carr and Madan (1999) suggest the default values $M = 1024$ and $N = 2^{12}$ as a rule of thumb. The Carr-Madan formula is very sensitive with respect to the damping factor, we choose $\alpha = 1.95$. For these default values, the accuracy of 10^{-7} is reached in only 18.33% of the samples in the test set (any other fixed α leads to an even lower proportion). Consequently, RFs and NNs are a useful tool for improving the accuracy of the Carr-Madan formula, since there is no single damping factor α and number of grid points N for all cases.

5.5 Conclusion

In this paper, we proposed an indirect use of machine learning to improve the efficiency and accuracy of the Carr-Madan formula and the COS method for option pricing. Junike and Pankrashkin (2022) and Junike (2024) provide explicit bounds on the truncation range and the number of terms to apply the COS method. These bounds ensure that the COS method prices a European option within a predefined error tolerance. It is generally time-consuming to obtain these bounds using classical numerical tools. In this paper, we instead estimate these bounds using machine learning techniques such as RF, DT and NN. We summarize the advantages:

- Compared to directly estimating the option prices using machine learning techniques as in Liu et al. (2019a), Liu et al. (2019b) and De Spiegeleer et al. (2018), our approach allows for full error control.

- Compared to estimating the bounds using classical numerical methods, our approach is much faster: about a factor 100.
- Compared to using a fast rule of thumb (as proposed in Fang and Oosterlee (2009a) and Carr and Madan (1999)) to estimate the tuning parameters of the COS method or the Carr-Madan formula, our approach is much more reliable. For the COS method, see Junike and Pankrashkin (2022) for examples where a rule of thumb based on cumulants leads to serious mispricing. For the Carr-Madan formula, see Section 5.4.2.

We tested RF, DT and NN to estimate the bounds to obtain the truncation range and the number of terms to apply the COS method. Among these techniques, the RF works best (accurate on 100% of the test set). The NN has a similar performance. But even a small DT gives very satisfactory results (accurate on 98.2% of the test set). Estimation of the tuning parameters of the Carr-Madan formula by a RF or a NN works in about 90% of all samples in a test set.

5.A Appendix

5.A.1 Decision tree of depth 5 to predict I_{20}

nodeID	leaf node	variable	split value	left-child (if variable $\leq$ split value)	right-child (if variable $>$ split value)	prediction
0	No	T	0.186064	1	2	NA
1	No	v_0	0.236779	3	4	NA
2	No	T	1.143101	5	6	NA
3	No	T	0.062439	7	8	NA
4	No	ξ	2.705391	9	10	NA
5	No	ξ	2.436885	11	12	NA
6	No	T	2.887055	13	14	NA
7	No	v_0	0.022762	15	16	NA
8	No	ρ	0.976444	17	18	NA
9	No	T	0.020387	19	20	NA
10	No	v_0	0.698183	21	22	NA
11	No	T	0.420034	23	24	NA
12	No	T	0.527950	25	26	NA

nodeID	leaf node	variable	split value	left-child (if variable $\leq$ split value)	right-child (if variable $>$ split value)	prediction
13	No	θ	0.784247	27	28	NA
14	No	θ	0.640466	29	30	NA
15	No	ρ	-0.468963	31	32	NA
16	No	ξ	2.258602	33	34	NA
17	No	ξ	2.470854	35	36	NA
18	Yes	NA	NA	NA	NA	429.628317
19	No	v_0	0.587902	37	38	NA
20	No	v_0	0.694761	39	40	NA
21	No	ρ	0.806959	41	42	NA
22	No	ρ	-0.965657	43	44	NA
23	No	ρ	0.960696	45	46	NA
24	No	θ	0.719155	47	48	NA
25	No	ρ	0.910680	49	50	NA
26	No	ρ	0.971400	51	52	NA
27	No	ξ	1.991873	53	54	NA
28	No	κ	3.484651	55	56	NA
29	No	T	5.496469	57	58	NA
30	No	T	5.071144	59	60	NA
31	Yes	NA	NA	NA	NA	487.342705
32	Yes	NA	NA	NA	NA	235.195790
33	Yes	NA	NA	NA	NA	102.893171
34	Yes	NA	NA	NA	NA	201.553675
35	Yes	NA	NA	NA	NA	36.203604
36	Yes	NA	NA	NA	NA	88.277112
37	Yes	NA	NA	NA	NA	62.208418
38	Yes	NA	NA	NA	NA	32.587266
39	Yes	NA	NA	NA	NA	25.738887
40	Yes	NA	NA	NA	NA	14.188984
41	Yes	NA	NA	NA	NA	64.373567
42	Yes	NA	NA	NA	NA	145.963858
43	Yes	NA	NA	NA	NA	133.842206
44	Yes	NA	NA	NA	NA	31.212500
45	Yes	NA	NA	NA	NA	9.945991

nodeID	leaf node	variable	split value	left-child (if variable ≤ split value)	right-child (if variable > split value)	prediction
46	Yes	NA	NA	NA	NA	38.703759
47	Yes	NA	NA	NA	NA	8.617465
48	Yes	NA	NA	NA	NA	4.853007
49	Yes	NA	NA	NA	NA	18.833877
50	Yes	NA	NA	NA	NA	47.724354
51	Yes	NA	NA	NA	NA	10.171048
52	Yes	NA	NA	NA	NA	45.038898
53	Yes	NA	NA	NA	NA	4.335898
54	Yes	NA	NA	NA	NA	7.306731
55	Yes	NA	NA	NA	NA	4.961622
56	Yes	NA	NA	NA	NA	2.733986
57	Yes	NA	NA	NA	NA	3.421163
58	Yes	NA	NA	NA	NA	2.172563
59	Yes	NA	NA	NA	NA	1.894587
60	Yes	NA	NA	NA	NA	1.144569

5.A.2 Decision tree of depth 5 to predict μ_8

nodeID	leaf node	variable	split value	left-child (if variable ≤ splitvalue)	right-child (if variable > splitvalue)	prediction
0	No	T	3.399204	1	2	NA
1	No	T	1.169626	3	4	NA
2	No	θ	0.959098	5	6	NA
3	No	T	0.428831	7	8	NA
4	No	θ	0.900109	9	10	NA
5	No	θ	0.498129	11	12	NA
6	No	κ	2.697343	13	14	NA
7	No	T	0.183570	15	16	NA
8	No	ξ	2.347370	17	18	NA
9	No	θ	0.417688	19	20	NA
10	No	ρ	-0.102140	21	22	NA
11	No	θ	0.273765	23	24	NA

nodeID	leaf node	variable	split value	left-child (if variable $\leq$ splitvalue)	right-child (if variable $>$ splitvalue)	prediction
12	No	ρ	-0.192910	25	26	NA
13	No	ξ	2.690779	27	28	NA
14	No	ρ	-0.116578	29	30	NA
15	No	T	0.074047	31	32	NA
16	No	ξ	2.265810	33	34	NA
17	No	θ	0.834463	35	36	NA
18	No	ρ	-0.186050	37	38	NA
19	No	θ	0.226219	39	40	NA
20	No	ρ	-0.354863	41	42	NA
21	No	ξ	2.903877	43	44	NA
22	No	T	2.237251	45	46	NA
23	No	θ	0.177632	47	48	NA
24	No	ρ	-0.286902	49	50	NA
25	No	ξ	2.631217	51	52	NA
26	No	T	6.095978	53	54	NA
27	No	ρ	-0.015948	55	56	NA
28	No	ρ	-0.210373	57	58	NA
29	No	ξ	3.133527	59	60	NA
30	No	T	6.186172	61	62	NA
31	Yes	NA	NA	NA	NA	0.366800
32	Yes	NA	NA	NA	NA	0.754696
33	Yes	NA	NA	NA	NA	1.068672
34	Yes	NA	NA	NA	NA	1.464186
35	Yes	NA	NA	NA	NA	1.367173
36	Yes	NA	NA	NA	NA	1.973656
37	Yes	NA	NA	NA	NA	3.110276
38	Yes	NA	NA	NA	NA	2.107476
39	Yes	NA	NA	NA	NA	1.381474
40	Yes	NA	NA	NA	NA	2.026955
41	Yes	NA	NA	NA	NA	3.456599
42	Yes	NA	NA	NA	NA	2.541016
43	Yes	NA	NA	NA	NA	3.922665
44	Yes	NA	NA	NA	NA	5.965712

nodeID	leaf node	variable	split value	left-child (if variable ≤ splitvalue)	right-child (if variable > splitvalue)	prediction
45	Yes	NA	NA	NA	NA	2.993397
46	Yes	NA	NA	NA	NA	3.799591
47	Yes	NA	NA	NA	NA	1.785174
48	Yes	NA	NA	NA	NA	2.496068
49	Yes	NA	NA	NA	NA	3.837462
50	Yes	NA	NA	NA	NA	3.023879
51	Yes	NA	NA	NA	NA	4.734888
52	Yes	NA	NA	NA	NA	6.274068
53	Yes	NA	NA	NA	NA	3.484621
54	Yes	NA	NA	NA	NA	4.394604
55	Yes	NA	NA	NA	NA	8.766430
56	Yes	NA	NA	NA	NA	5.524519
57	Yes	NA	NA	NA	NA	17.084012
58	Yes	NA	NA	NA	NA	10.369077
59	Yes	NA	NA	NA	NA	6.240159
60	Yes	NA	NA	NA	NA	8.173560
61	Yes	NA	NA	NA	NA	4.647802
62	Yes	NA	NA	NA	NA	5.944725

5.A.3 Simple implementation

The following algorithm implements the COS method in R for the Heston model to price European put and call options.

Algorithm 1 Implementation details of the COS method in the Heston model

```
#Characteristic function of log-returns in the Heston with parameters params.
#The characteristic function is taken from Schoutens et. al (2004).
psiLogST_Heston = function(u, mat, params, S0, r){
  kappa = params[1] #speed of mean reversion
  theta = params[2] #level of mean reversion
  xi = params[3] #vol of vol
  rho = params[4] #correlation vol stock
  v0 = params[5] #initial vol
  d = sqrt((rho * xi * u * 1i - kappa)^2 - xi^2 * (-1i * u - u^2))
```

```
  mytmp = kappa - rho * xi * u * 1i
  g = (mytmp - d) / (mytmp + d)
  expdmat = exp(-d * mat)
  tmp0 = 1i * u * (log(S0) + r * mat)
  tmp1 = (mytmp - d) * mat - 2 * log((1 - g * expdmat) / (1 - g))
  tmp2 = theta * kappa * xi^(-2) * tmp1
  tmp3 = v0 * xi^(-2) * (mytmp - d) * (1 - expdmat) / (1 - g * expdmat)
  exp(tmp0 + tmp2 + tmp3)
}
library(Deriv) #There are much faster alternatives like SageMath.
psiLogST_Heston1=Deriv(psiLogST_Heston, "u")

#mu is equal to E[log(S_T)]
mu = function(mat, params, S0, r){
  Re(-1i * psiLogST_Heston1(0, mat, params, S0, r))
}
#Characteristic function of centralized log-returns in the Heston model.
phi = function(u, mat, params, S0, r){
  psiLogST_Heston(u, mat, params, S0, r) * exp(-1i * u * mu(mat, params, S0, r))
}
#cosine coefficients of the density.
ck = function(L, mat, N, params, S0, r){
  k = 0:N
  return(1 / L * Re(phi(k * pi / (2 * L), mat, params, S0, r) * exp(1i * k * pi/2)))
}
#cosine coefficients of a put option, see Appendix Junike and Pankrashkin (2022).
vk = function(K, L, mat, N, params, S0, r){
  mymu = mu(mat, params, S0, r) #mu = E[log(S_T)]
  d = min(log(K) - mymu, L)
  if(d <= -L)
    return(rep(0, N + 1)) #Return zero vector
  k = 0:N
  psi0 = 2 * L / (k * pi) * (sin(k * pi * (d + L) / (2 * L)))
  psi0[1] = d + L
  tmp1 = k * pi / (2 * L) * sin( k * pi * (d + L) / (2 * L))
  tmp2 = cos(k * pi * (d + L) / (2 * L))
  tmp3 = 1 + (k * pi / (2 * L))^2
  psi1 = (exp(d) * (tmp1 + tmp2) - exp(-L)) / tmp3
  return(exp(-r * mat) * (K * psi0 - exp(mymu) * psi1))
}
#approximation of put option by COS method
put_COS = function(K, L, mat, N, params, S0, r){
  tmp = ck(L, mat, N, params, S0, r) * vk(K, L, mat, N, params, S0, r)
  tmp[1] = 0.5 * tmp[1] #First term is weighted by 1/2
  return(sum(tmp))
}
```

```
#approximation of call option by COS method using put-call parity
call_COS = function(K, L, mat, N, params, S0, r){
  return(put_COS(K, L, mat, N, params, S0, r) + S0 - K * exp(-r * mat))
}

#Derivatives of the characteristic function of the centralized log-returns in the Heston model.
phi1 = Deriv(phi, "u")
phi2 = Deriv(phi1, "u")
phi3 = Deriv(phi2, "u") #Takes very long but has to be done only once.
phi4 = Deriv(phi3, "u") #Takes very long but has to be done only once.
save(phi4, file = "phi4.RData") #save for later use. Load with load("phi4.RData").

#Price a put option in the Heston model by the COS method.
eps = 10^-6 #error tolerance
K = 90 #strike
S0 = 100 #current stock price
r = 0.1 #interest rates
params = c(0.6067, 0.0707, 0.2928, -0.7571, 0.0654)
mat = 0.7 #maturity
mu_n = abs(phi4(0, mat, params, S0, r)) #4-th moment of log-returns.
L = (2 * K * exp(-r * mat) * mu_n / eps)^(1 / 4) #Junike (2024, Eq. (3.10)).
s = 20 #number of derivatives to determine the number of terms
integrand = function(u){1 / (2 * pi) * abs(u)^(s + 1) * abs(phi(u, mat, params, S0, r))}
boundDeriv = integrate(integrand, -Inf, Inf)$value
tmp = 2^(s + 5 / 2) * boundDeriv * L^(s + 2) * 12 * K * exp(-r * mat)
N = ceiling((tmp / (s * pi^(s + 1) * eps))^(1 / s)) #Number of terms, Junike (2024, Sec. 6.1)
put_COS(K, L, mat, N, params, S0, r) #The price of put option is 2.773954.
```

Chapter 6

Future research

In Chapters 2 and 3, we analyzed a class of decompositions of path-independent instruments based on Itô's formula when the P&L can be described by a twice differentiable function. It would be interesting to analyze decompositions of more general functionals using the functional Itô formula, see Levental et al. (2013) and Cont and Fournié (2010). This would enable the decomposition of path-dependent options.
In Theorem 3.3.4 we provided conditions to ensure a certain speed of convergence of sequential updating decompositions when the risk factors are continuous. This could be generalized to semimartingales with jumps. The speed of convergence is determined by a bound on the expected value of the error. An interesting extension would be to derive a pathwise error bound.

In Chapters 4 and 5, we analyzed the COS method for multidimensional integrals. We used full grids to discretize the integration domain. It is an open research question to what extent the numerical effort can be reduced by using sparse grids instead of full grids.
In Chapter 5 we used machine learning techniques to model the formulas for the tuning parameters of the COS method in the one-dimensional case. It would be interesting to use the same approach for the formulas in general dimensions.

Bibliography

Abramowitz, M. and I. A. Stegun (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Vol. 55. US Government printing office.

Amann, H. and J. Escher (2009). *Analysis III*. Birkhäuser Verlag AG, Springer.

Andersen, L. B. G. (2008). "Efficient simulation of the Heston stochastic volatility model". In: *Journal of Computational Finance* 11(3), pp. 1–43.

Andersen, L. B. G. and V. V. Piterbarg (2007). "Moment explosions in stochastic volatility models". In: *Finance and Stochastics* 11(1), pp. 29–50.

Andersson, K., A. Andersson, and C. W. Oosterlee (2023). "Convergence of a robust deep FBSDE method for stochastic control". In: *SIAM Journal on Scientific Computing* 45(1), A226–A255.

Bakshi, G., C. Cao, and Z. Chen (1997). "Empirical performance of alternative option pricing models". In: *The Journal of Finance* 52(5), pp. 2003–2049.

Bardgett, C., E. Gourier, and M. Leippold (2019). "Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets". In: *Journal of Financial Economics* 131(3), pp. 593–618.

Barndorff-Nielsen, O. E. (1997). "Normal inverse Gaussian distributions and stochastic volatility modelling". In: *Scandinavian Journal of Statistics* 24(1), pp. 1–13.

Barndorff-Nielsen, O. E. and R. Stelzer (2005). "Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes". In: *Scandinavian Journal of Statistics* 32(4), pp. 617–637.

Bauer, H. (1996). *Probability Theory*. Vol. 23. Walter de Gruyter.

Bayer, C. et al. (2023). "Optimal Damping with Hierarchical Adaptive Quadrature for Efficient Fourier Pricing of Multi-Asset Options in Lévy Models". In: *Journal of Computational Finance* 27(3), pp. 43–86.

Bayer, C. et al. (2024). "Quasi-Monte Carlo for Efficient Fourier Pricing of Multi-Asset Options". In: *arXiv preprint arXiv:2403.02832*.

Björk, T. (2009). *Arbitrage theory in continuous time*. Oxford university press.

Blinder, A. S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates". In: *Journal of Human resources*, pp. 436–455.

Breiman, L. (2001). "Random forests". In: *Machine Learning* 45, pp. 5–32.

Breiman, L. et al. (1984). *Classification and regression trees*. Taylor & Francis.

Brown, B. M. (1972). "Formulae for absolute moments". In: *Journal of the Australian Mathematical Society* 13(1), pp. 104–106.

Candland, A. and C. Latz (2014). *Profit and Loss attribution. In Internal Models and Solvency II – From Regulation to Implementation*. Risk Books, London.

Carr, P. and D. Madan (1999). "Option valuation using the fast Fourier transform". In: *Journal of Computational Finance* 2(4), pp. 61–73.

Černỳ, A. and J. Ruf (2021). "Pure-jump semimartingales". In: *Bernoulli* 27(4), pp. 2624–2648.

Choudhury, G. L. and D. M. Lucantoni (1996). "Numerical computation of the moments of a probability distribution from its transform". In: *Operations Research* 44(2), pp. 368–381.

Christiansen, M. C. (2022). "On the decomposition of an insurer's profits and losses". In: *Scandinavian Actuarial Journal*, pp. 1–20.

Cont, R. and D.-A. Fournié (2010). "Change of variable formulas for non-anticipative functionals on path space". In: *Journal of Functional Analysis* 259(4), pp. 1043–1072.

Crisóstomo, R. (2015). "An analysis of the Heston stochastic volatility model: Implementation and calibration using matlab". In: *arXiv preprint arXiv:1502.02963*.

Cui, Y., S. del Baño Rollin, and G. Germano (2017). "Full and fast calibration of the Heston stochastic volatility model". In: *European Journal of Operational Research* 263(2), pp. 625–638.

De Spiegeleer, J. et al. (2018). "Machine learning for quantitative finance: fast derivative pricing, hedging and fitting". In: *Quantitative Finance* 18(10), pp. 1635–1643.

Duffie, D., D. Filipović, and W. Schachermayer (2003). "Affine processes and applications in finance". In: *The Annals of Applied Probability* 13(3), pp. 984–1053.

Eberlein, E., K. Glau, and A. Papapantoleon (2010). “Analysis of Fourier transform valuation formulas and applications”. In: *Applied Mathematical Finance* 17(3), pp. 211–240.

Engelmann, B., F. Koster, and D. Oeltz (2021). “Calibration of the Heston stochastic local volatility model: A finite volume scheme”. In: *International Journal of Financial Engineering* 8(01), p. 2050048.

European Commission (2015). “Commission delegated regulation (EU) 2015/35 of 10 October 2014 supplementing directive 2009/138/EC of the European parliament and of the council on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II)”. In: *Official Journal of European Union*.

Fang, F. and C. W. Oosterlee (2009a). “A novel pricing method for European options based on Fourier-cosine series expansions”. In: *SIAM Journal on Scientific Computing* 31(2), pp. 826–848.

Fang, F. and C. W. Oosterlee (2009b). “Pricing early-exercise and discrete barrier options by Fourier-cosine series expansions”. In: *Numerische Mathematik* 114(1), p. 27.

Fang, F. and C. W. Oosterlee (2011). “A Fourier-based valuation method for Bermudan and barrier options under Heston’s model”. In: *SIAM Journal on Financial Mathematics* 2(1), pp. 439–463.

Fischer, T. (2004). “On the decomposition of risk in life insurance”. In: *Technische Universität Darmstadt*.

Flaig, S. and G. Junike (2024). “Profit and loss attribution: an empirical study”. In: *European Actuarial Journal*, pp. 1–7.

Forde, M., A. Jacquier, and R. Lee (2012). “The small-time smile and term structure of implied volatility under the Heston model”. In: *SIAM Journal on Financial Mathematics* 3(1), pp. 690–708.

Forster, O. (2017). *Analysis 2: Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgleichungen*. Springer-Verlag.

Fortin, N., T. Lemieux, and S. Firpo (2011). “Chapter 1 - Decomposition Methods in Economics”. In: ed. by O. Ashenfelter and D. Card. Vol. 4. Handbook of Labor Economics. Elsevier, pp. 1–102.

Frei, C. (2020). “A new approach to risk attribution and its application in credit risk analysis”. In: *Risks* 8(2), p. 65.

Friedman, E. and H. Moulin (1999). “Three methods to share joint costs or surplus”. In: *Journal of economic Theory* 87(2), pp. 275–312.

Gil-Pelaez, J. (1951). “Note on the inversion theorem”. In: *Biometrika* 38(3-4), pp. 481–482.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Vol. 53. Stochastic Modelling and Applied Probability. Springer New York.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press.

Grimmett, G. R. and D. R. Stirzaker (2001). “Probability and Random Processes”. In.

Grünbaum, B. (2013). “Convex Polytopes”. In: *Graduate Texts in Mathematics*. Vol. 221. Springer: New York.

Grzelak, L. A., J. Jablecki, and D. Gatarek (2023). “Efficient pricing and calibration of high-dimensional basket options”. In: *International Journal of Computer Mathematics* 101(8), pp. 865–888.

Grzelak, L. A. and C. W. Oosterlee (2011). “On the Heston model with stochastic interest rates”. In: *SIAM Journal on Financial Mathematics* 2(1), pp. 255–286.

He, Kaiming et al. (2015). “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.

Heston, S. L. (1993). “A closed-form solution for options with stochastic volatility with applications to bond and currency options”. In: *The Review of Financial Studies* 6(2), pp. 327–343.

Hubalek, F. and J. Kallsen (2003). “Variance-optimal hedging and Markowitz-efficient portfolios for multivariate processes with stationary independent increments with and without constraints”. In: *TU München*.

Hughett, P. (1998). “Error bounds for numerical inversion of a probability characteristic function”. In: *SIAM journal on numerical analysis* 35(4), pp. 1368–1392.

Hurd, T. R. and Z. Zhou (2010). “A Fourier transform method for spread option pricing”. In: *SIAM Journal on Financial Mathematics* 1(1), pp. 142–157.

Jetses, J. and M. C. Christiansen (2022). “A general surplus decomposition principle in life insurance”. In: *Scandinavian Actuarial Journal*, pp. 1–25.

Junike, G. (2024). “On the number of terms in the COS method for European option pricing”. In: *Numerische Mathematik* 156(2), pp. 533–564.

Junike, G. and K. Pankrashkin (2022). “Precise option pricing by the COS method–How to choose the truncation range”. In: *Applied Mathematics and Computation* 421, p. 126935.

Junike, G. and H. Stier (2024). “From characteristic functions to multivariate distribution functions and European option prices by the damped COS method”. In: *arXiv preprint arXiv:2307.12843*.

Junike, G., H. Stier, and M. C. Christiansen (2024). “Profit and loss decomposition in continuous time and approximations ”. In: *arXiv preprint arXiv:2212.06733*.

Karandikar, R. L. (1995). “On pathwise stochastic integration”. In: *Stochastic Processes and their applications* 57(1), pp. 11–18.

Küchler, U. and S. Tappe (2008). “On the shapes of bilateral Gamma densities”. In: *Statistics & Probability Letters* 78(15), pp. 2478–2484.

Leitao, Á. et al. (2018). “On the data-driven COS method”. In: *Applied Mathematics and Computation* 317, pp. 68–84.

Levendorskiĭ, S. (2012). “Efficient pricing and reliable calibration in the Heston model”. In: *International Journal of Theoretical and Applied Finance* 15(07), p. 1250050.

Levental, S., M. Schroder, and S. Sinha (2013). “A simple proof of functional Itô’s lemma for semimartingales with an application”. In: *Statistics & Probability Letters* 83(9), pp. 2019–2026.

Liu, S., C. W. Oosterlee, and S. M. Bohte (2019a). “Pricing options and computing implied volatilities using neural networks”. In: *Risks* 7(1), pp. 1–22.

Liu, S. et al. (2019b). “A neural network-based framework for financial model calibration”. In: *Journal of Mathematics in Industry* 9(1), pp. 1–28.

Lord, R. et al. (2008). “A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes”. In: *SIAM Journal on Scientific Computing* 30(4), pp. 1678–1705.

Luciano, E. and W. Schoutens (2006). “A multivariate jump-driven financial asset model”. In: *Quantitative finance* 6(5), pp. 385–402.

Madan, D., P. P. Carr, and E. C. Chang (1998). “The Variance Gamma process and Option Pricing”. In: *Review of Finance* 2(1), pp. 79–105.

Mai, J. F. (2023). “Performance attribution with respect to interest rates, FX, carry, and residual market risks”. In: *arXiv preprint arXiv:2302.01010*.

Norberg, R. (1999). “A theory of bonus in life insurance”. In: *Finance and Stochastics* 3(4), pp. 373–390.

Oaxaca, R. (1973). “Male-Female Wage Differentials in Urban Labor Markets”. In: *International Economic Review*, pp. 693–709.

Olver, F. W. J. et al. (2010). *NIST handbook of mathematical functions*. Cambridge university press.

Oosterlee, C. W. and L. A. Grzelak (2019). *Mathematical modeling and computation in finance: with exercises and Python and MATLAB computer codes*. World Scientific.

Ortiz-Gracia, L. and C. W. Oosterlee (2013). “Robust pricing of European options with wavelets and the characteristic function”. In: *SIAM Journal on Scientific Computing* 35(5), B1055–B1084.

Ortiz-Gracia, L. and C. W. Oosterlee (2016). “A highly efficient Shannon wavelet inverse Fourier technique for pricing European options”. In: *SIAM Journal on Scientific Computing* 38(1), B118–B143.

Protter, P. (2005). “Stochastic Differential Equations”. In: *Stochastic Integration and Differential Equations*. Springer Berlin Heidelberg, pp. 249–361.

Rabinovitch, R. (1989). “Pricing Stock and Bond Options when the Default-Free Rate is Stochastic”. In: *Journal of Financial and Quantitative Analysis* 24(4), pp. 447–457.

Ramlau-Hansen, H. (1991). “Distribution of Surplus in Life Insurance”. In: *ASTIN Bulletin* 21(1), pp. 57–71.

Rosen, D. and D. Saunders (2010). “Risk factor contributions in portfolio credit risk models”. In: *Journal of Banking & Finance* 34(2), pp. 336–349.

Ruf, J. and W. Wang (2020). “Neural networks for option pricing and hedging: a literature review”. In: *Journal of Computational Finance*.

Ruijter, M. J. and C. W. Oosterlee (2012). “Two-dimensional Fourier cosine series expansion method for pricing financial options”. In: *SIAM Journal on Scientific Computing* 34(5), B642–B671.

Ruijter, M. J. and C. W. Oosterlee (2015). “A Fourier cosine method for an efficient computation of solutions to BSDEs”. In: *SIAM Journal on Scientific Computing* 37(2), A859–A889.

Schilling, K. et al. (2020). “Decomposing Dynamic Risks into Risk Components”. In: *Management Science* 66(12), pp. 5738–5756.

Schorr, B. (1975). “Numerical inversion of a class of characteristic functions”. In: *BIT Numerical Mathematics* 15(1), pp. 94–102.

Schoutens, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley Online Library.

Schoutens, W., E. Simons, and J. Tistaert (2003). “A perfect calibration! Now what?” In: *The best of Wilmott*, pp. 281–304.

Shapley, L. S. (1953). “A Value for n-Person Games”. In: *Contributions to the Theory of Games, Volume II*. Ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press: Princeton, pp. 307–318.

Shephard, N. G. (1991a). “From characteristic function to distribution function: a simple framework for the theory”. In: *Econometric theory* 7(4), pp. 519–529.

Shephard, N. G. (1991b). “Numerical integration rules for multivariate inversions”. In: *Journal of Statistical Computation and Simulation* 39(1-2), pp. 37–46.

Shorrocks, A. F. (2013). “Decomposition procedures for distributional analysis: a unified framework based on the Shapley value”. In: *Journal of Economic Inequality* 11(1), pp. 99–126.

Shubik, M. (1962). “Incentives, decentralized control, the assignment of joint costs and internal pricing”. In: *Management science* 8(3), pp. 325–343.

Sirignano, J. and K. Spiliopoulos (2018). “DGM: A deep learning algorithm for solving partial differential equations”. In: *Journal of Computational Physics* 375, pp. 1339–1364.

Sprumont, Y. (1998). “Ordinal Cost Sharing”. In: *Journal of Economic Theory* 81(1), pp. 126–162.

Von Bahr, B. (1965). “On the convergence of moments in the central limit theorem”. In: *The Annals of Mathematical Statistics*, pp. 808–818.

Waller, L. A., B. W. Turnbull, and J. M. Hardin (1995). “Obtaining distribution functions by numerical inversion of characteristic functions with applications”. In: *The American Statistician* 49(4), pp. 346–350.

Yang, Y., Y. Zheng, and T. Hospedales (2017). “Gated neural networks for option pricing: Rationality by design”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1.

Zhang, B. and C. W. Oosterlee (2013). “Efficient pricing of European-style Asian options under exponential Lévy processes based on Fourier cosine expansions”. In: *SIAM Journal on Financial Mathematics* 4(1), pp. 399–426.

Contact

Hauke Stier
Email: haukestier@web.de

Affidavit

(English)

I, Hauke Stier, hereby declare that the submitted thesis is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given. I confirm that I followed the guidelines of good scientific practice of the Carl von Ossietzky University Oldenburg.
I further declare that neither the submitted thesis nor parts of it have been presented as part of an examination degree to any other university.

Eidesstattliche Erklärung

(German)

Hiermit versichere ich, Hauke Stier, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis festgelegt sind, befolgt habe.
Ich versichere weiterhin, dass weder diese Arbeit noch Teile davon an einer anderen Universität eingereicht wurden.