Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Arbitrage-free methods to price European options under the SABR model

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Zaza van der Have
Delft, the Netherlands
January 2015

MSc THESIS APPLIED MATHEMATICS

"Arbitrage-free methods to price European options under the SABR model"

Zaza van der Have

Delft University of Technology

Daily supervisors

Dr. T.P.T. Dijkstra
Prof.dr.ir. C.W. Oosterlee

Responsible professor

Prof.dr.ir. C.W. Oosterlee

Other thesis committee members

Dr.ir. R.J. Fokkink
Ir. M.J. Ruijter

January 2015

Delft, the Netherlands

Abstract

In this thesis we discuss several methods to price European options under the SABR model. In general, methods given in literature are not free of arbitrage and/or inaccurate for long maturities. This led to the development of a new pricing approach. We extend the BCOS method from one dimension to two dimensions. This extension is necessary for application of a simplification of the BCOS method, the DCOS method, to the SABR model. In this pricing method we use the characteristic function of the discrete forward process and the Fourier-based COS method. It is possible to price European options under the SABR model for multiple strikes in one computation with the DCOS method. Besides valuing European options, we can also price Bermudan and discretely monitored barrier options with this pricing approach.

Preface

This thesis has been submitted for the degree of Master of Science in Applied Mathematics at Delft University of Technology, the Netherlands and was carried out for Rabobank in Utrecht.
I would like to thank some people who have contributed in the process of writing this thesis. First of all, I would like to acknowledge prof.dr.ir. Kees Oosterlee, dr. Tim Dijkstra and ir. Marjon Ruijter for their advice and guidance during the project and for being part of the examination committee. Furthermore, I would like to thank dr.ir. Robbert Fokkink from TU Delft for being part of the examination committee. I would also like to thank the whole Pricing Model Validation (PMV) team of Rabobank for the pleasant working environment and for interesting discussions. In particular, Anton van der Stoep for his feedback on my draft. Last but certainly not least, I would like to express my gratitude to my family and friends, and especially to Silvester Wulffers, who supported me through the duration of my study.

Zaza van der Have
Utrecht, January 2015

Contents

Chapter 1

Introduction

In this thesis, we will discuss different methods and in particular we propose a method for pricing European options under the so-called SABR model. First of all, we will introduce some financial terms.

In finance, a *derivative* is a contract with a value that is derived from the performance of an *underlying* entity, e.g. an interest rate or an asset. In this thesis, we abbreviate the underlying entity by the underlying. A *European call option* is a derivative which gives the owner (holder) of the option the right, but not the obligation, to buy an underlying asset or instrument at a certain *expiration time* T for a specified price K. This price K is called the *strike* of the option. The holder of a *European put option* has the right to sell an underlying asset or instrument at time T for strike K. The *payoff* of an option is its value at the time of maturity T, for example the payoff of a European call option is the maximum of the value of the underlying at time T minus the strike K and 0. The *forward* value of a derivative is the current agreed upon value of the derivative on a specified date in the future. In finance, *arbitrage* is the practice of taking advantage of incorrect pricing in the market, i.e. earn a higher return than the *risk-free* interest rate without taking risk. In this thesis, the risk-free interest rate is assumed to be deterministic. The *risk-neutral measure* $\mathbb{Q}$, is a probability measure under which the value of a derivative is equal to its expected discounted payoff.

In 2002, P.S. Hagan, D. Kumar, A.S. Lesniewski and D.E. Woodward introduced a new stochastic model to price and hedge European options [14]. This model is called the SABR model. S stands for "stochastic" and ABR represents the Greek characters α, β and ρ, which are parameters in the SABR model. In this thesis we will use σ instead of α, because σ is the general character used in the literature to indicate the volatility. The SABR model is developed to improve previous asset or interest rate models, e.g. the Black-Scholes model [6] and the local volatility model of Dupire [8], with the purpose that it matches market behaviour better than previous models.
In [14], a very convenient formula was given to calculate the so-called Black implied volatility of the SABR model. This formula, also known as the Hagan formula, leads however to arbitrage possibilities for low strikes, as one can show by examining the corresponding probability density function (PDF) [22]. Following the pricing approach of Hagan, Kumar, Lesniewski and Woodward, other methods have been developed for pricing European options under the SABR model. We discussed the Hagan formula and the methods of Obłój [19], Andreasen and Huge [1], and Balland and Tran [5] in a previous internship report [16]. In this thesis, we discuss the Hagan formula and two other methods to price options under the SABR model. The developers of

both pricing approaches [2, 15] claimed to reduce the arbitrage possibilities or even remove the arbitrage entirely. Besides arbitrage, we also check the accuracy of each method by comparing so-called volatility smiles obtained by the method to smiles obtained by a finite difference solver. The latter is a partial differential equation (PDE) solver and is very accurate. Because of this accuracy, this thesis uses the volatility smiles given by this finite difference reference solver as the reference smiles.

Antonov, Konikov and Spector introduced a pricing method with a mimicking model [2]. In their paper, they derived an exact formula to calculate the value of a European call option for the case that in the SABR model the correlation $\rho = 0$. Then, the authors discussed how one can map the SABR parameters for the general case where $\rho \neq 0$ to parameters of a mimicking model where the correlation is zero. Hagan, Kumar, Lesniewski and Woodward gave a PDE approach [15] to recover a PDE to determine the PDF. By numerical methods, one can approximate the PDF and this approximation can be used to calculate the price of a European option.

In particular, we propose a new method in this thesis. When the characteristic function of a model is known, we can use the (two-dimensional) COS method to price European options [9, 23]. Unfortunately, no analytical expression for the bivariate characteristic function of the SABR model is available. The BCOS method (Backward Stochastic Differential Equation COS method) of M.J. Ruijter and C.W. Oosterlee [25, 24] uses the characteristic function of the discrete forward process to price options. In this thesis, we expand the BCOS method from one dimension to two dimensions. Then, we use a simplified version of the two-dimensional BCOS method for pricing European options under the SABR model. The use of backward stochastic differential equations (BSDEs) does not appear in this simplified method, that is why we call it the DCOS method, where D stands for discrete. We discuss different methods to use the DCOS method to price European options under the SABR model. There are three main difficulties: how to choose the best discretization scheme, how to choose the number of time steps, and should we use logarithmic transformation(s). To gain more insight into the choice for the best discretization scheme we compare the characteristic function of the Heston model with the characteristic function of its discrete forward process.

This thesis is organized as follows. In Chapter 2 we introduce the SABR model and three methods for pricing European options under the SABR model; the Hagan formula, the option pricing method of Antonov, Konikov and Spector, and the arbitrage-free pricing method of Hagan, Kumar, Lesniewski and Woodward. Then, in Chapters 3 to 6, we describe the development of the new pricing method. The BCOS method is explained in Chapter 3. In Chapter 4 we expand the BCOS method from one dimension to two dimensions. We discuss different methods for pricing European options under the SABR model with the two-dimensional DCOS method in Chapters 5 and 6. A conclusion is given in Chapter 7.
In the appendices we provide additional information and derivations. In Appendices A and B, we give the derivation of some formulas used in the option pricing methods we discuss in Chapter 2. We give an introduction in Itô-Taylor expansion in Appendix C. In Appendix D we determine the Euler, Milstein and 2.0-weak-Taylor schemes and in Appendix E we provide the derivation of the characteristic function of the discrete forward process. Finally, in Appendix F we provide the adjusted-Predictor-Corrector schemes for the Heston and the SABR model.

Chapter 2

SABR model and its pricing methods

In their model [14], Hagan, Kumar, Lesniewski and Woodward assumed that both the forward F, e.g. the forward swap rate, and the volatility of the forward σ are stochastic over time. The SABR model is given by the following forward stochastic differential equations (FSDEs):

$$\begin{aligned} \mathrm{d}F_t &= \sigma_t \, (F_t)^\beta \, \mathrm{d}W_t^1, \qquad F_0 = f, & (2.1) \\ \mathrm{d}\sigma_t &= \nu \sigma_t \, \mathrm{d}W_t^2, \qquad \sigma_0 = \alpha, & (2.2) \\ \mathrm{d}W_t^1 \, \mathrm{d}W_t^2 &= \rho \, \mathrm{d}t, \end{aligned}$$

where W^1 and W^2 are standard Brownian motions under the forward measure and the parameters β, ν and ρ are constants. The exponent $0 \leq \beta \leq 1$, the vol-vol (volatility of the volatility) $\nu > 0$ and the correlation $-1 \leq \rho \leq 1$ are constraints for the SABR parameters.

This chapter is organized as follows. Hagan's formula and its main advantages and disadvantages are considered in Section 2.1. In Section 2.2, we explain the option pricing method of Antonov, Konikov and Spector. Finally, the arbitrage-free pricing method of Hagan, Kumar, Lesniewski and Woodward is given in Section 2.3.

2.1 The Hagan formula

From FSDEs (2.1) and (2.2), Hagan, Kumar, Lesniewski and Woodward derived a formula to calculate the Black implied volatility. This so-called Hagan formula is given by [21]

$$\sigma_B(0, T, K, f, \alpha) = A \left(\frac{\widetilde{z}}{y(\widetilde{z})} \right) B, \tag{2.3}$$

with

$$\begin{aligned} A &= \frac{\alpha}{(fK)^{\frac{(1-\beta)}{2}} \left[1 + \frac{(1-\beta)^2}{24} \log(f/K)^2 + \frac{(1-\beta)^4}{1920} \log(f/K)^4 \right]}, & (2.4a) \\ B &= 1 + \left(\frac{(1-\beta)^2}{24} \frac{\alpha^2}{(fK)^{1-\beta}} + \frac{\rho \beta \nu \alpha}{4 (fK)^{\frac{1-\beta}{2}}} + \frac{2 - 3\rho^2}{24} \nu^2 \right) T, & (2.4b) \\ \widetilde{z} &= \frac{\nu}{\alpha} (fK)^{\frac{1-\beta}{2}} \log(f/K), & (2.4c) \end{aligned}$$

$$y(\widetilde{z}) \quad = \quad \log\left(\frac{\sqrt{1-2\rho\widetilde{z}+\widetilde{z}^2}+\widetilde{z}-\rho}{1-\rho}\right). \tag{2.4d}$$

One can obtain

$$\lim_{z\to 0}\left(\frac{z}{y(z)}\right)=1. \tag{2.5}$$

The values of European options[1] with this pricing approach are given by the Black-Scholes formulas [6]

$$V^C_{\mathrm{BS}}(t_0=0,T,K,f,\sigma_B) \quad = \quad D(T)\left[f\cdot N(d_+)-K\cdot N(d_-)\right], \tag{2.6}$$

$$V^P_{\mathrm{BS}}(t_0=0,T,K,f,\sigma_B) \quad = \quad D(T)\left[K\cdot N(-d_-)-f\cdot N(-d_+)\right], \tag{2.7}$$

with

$$d_\pm=\frac{\log(f/K)\pm\frac{1}{2}\left(\sigma_B\right)^2 T}{\sigma_B\sqrt{T}}. \tag{2.8}$$

Here $N(\cdot)$ denotes the standard normal cumulative distribution function, V^C is the value of a European call at time 0 and V^P is the value of a European put at time 0. f is the value of the forward at time 0. T is the exercise date and, because we set today's time t_0 at 0, T is also the time to maturity. $D(T)$ is the discount factor, the value today of €1 delivered at date T, and K is the strike.

The volatility σ_B that is used in these formulas, also called the Black implied volatility, is calculated with the Hagan formula (2.3). A plot of the Black implied volatility σ_B against the strike K is called a volatility smile. The implementation of the Hagan formula and the Black-Scholes pricing formulas is relatively simple. This ease of implementation is a great advantage for the Hagan formula to extract the Black implied volatility of options.

2.1.1 Arbitrage

A necessary condition for non-arbitrage is the put-call parity. If a model is not compatible with the put-call parity,

$$V^C(0,T,K,f,\sigma_B)-V^P(0,T,K,f,\sigma_B)=D\left(T\right)\left[f-K\right], \tag{2.9}$$

then the model is not free of arbitrage. In [14], financial derivatives are priced with the Black-Scholes pricing formulas. The Black-Scholes formulas are compatible with the put-call parity, so the Hagan formula (2.3) is too.

It is well-known that the second partial derivative of the European call price with respect to the strike is equal to the discounted conditional probability density function $Q_{F_T}\left(T,F|f,\alpha\right)$ of the forward $F_T=F$ at time of maturity T, given today's value of the forward $F_0=f$ and volatility $\sigma_0=\alpha$, [22]

$$D(T)Q_{F_T}\left(T,F|f,\alpha\right)=\left.\frac{\partial^2 V^C}{\partial K^2}\right|_{K=F}. \tag{2.10}$$

When we extract the Black implied volatility of an option using the Hagan formula, we may obtain, depending on the specific parameters, negative probabilities for low strikes and occasionally also for high strikes. By definition, a probability density can never be negative, so here

[1] We do not specify the underlying in our examples, e.g. the underlying F could be a forward swap rate, such an option is then called a swaption.

we observe arbitrage. We give an example below.

Example 1 We take as parameters $\beta = 0.7$, $\alpha = 0.05735$, $\rho = -0.48$ and $\nu = 0.47$. Also, we set $T = 10$, $f = 0.05735$ and the constant risk-free interest rate $r = 0$. Central differences are used to approximate the second derivative in (2.10). The PDF corresponding to this example is shown in Figure 2.1. For low strikes the PDF is negative and by (2.10) the corresponding values of European call options are non-convex for those strikes. So, pricing options with the Hagan formula (2.3) is not free of arbitrage in this example. We give an arbitrage possibility below.

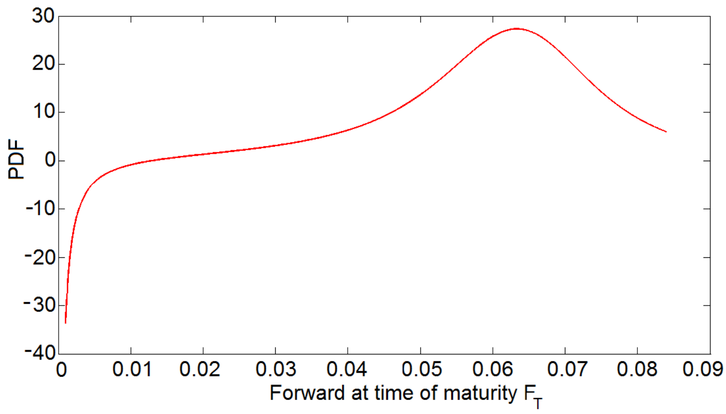

Figure 2.1: The incorrect conditional PDF of F_T given $F_0 = f$.

Because of the concavity in the prices for European call options for low strikes we observe

$$V^C(0, T, 0.0095, f, \sigma_B) + V^C(0, T, 0.0105, f, \sigma_B) < 2V^C(0, T, 0.01, f, \sigma_B)$$

We buy a European call option with strike value 0.0095, a European call option with strike value 0.0105, and we sell two European call options with strike value 0.01. All options have the same underlying F_t and the same time of maturity T.
At time T, we observe the following cases:

- When $F_T \leq 0.0095$ or $F_T > 0.0105$ our payoff P is given by $P = 0$.
- When $0.0095 < F_T \leq 0.01$ our payoff P is given by $P = F_T - 0.0095 > 0$.
- When $0.01 < F_T \leq 0.0105$ our payoff P is given by $P = 0.0105 - F_T \geq 0$.

So, the payoff is always non-negative, our initial earning is positive and we have risk-free interest rate $r = 0$. We started at time 0 with no money and at time T we have with certainty a positive amount of money:

$$V^C(0, T, 0.0095, f, \sigma_B) + V^C(0, T, 0.0105, f, \sigma_B) - 2V^C(0, T, 0.01, f, \sigma_B) + P > 0,$$

which is an example of arbitrage.

2.1.2 Accuracy of the Hagan formula

Besides the presence of arbitrage, Hagans formula has another disadvantage. The authors in [2, 14] observed that the Hagan formula is not accurate for long maturities T. For maturities longer than 10 years the error in the Black implied volatility for (2.3) can be 100 basis points

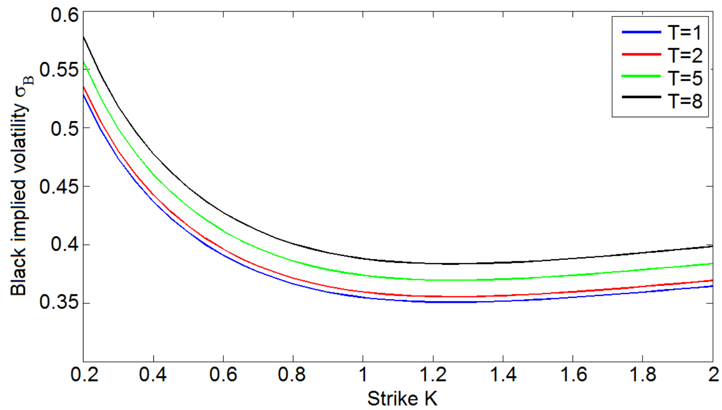

Figure 2.2: Volatility smiles calculated with the Hagan formula.

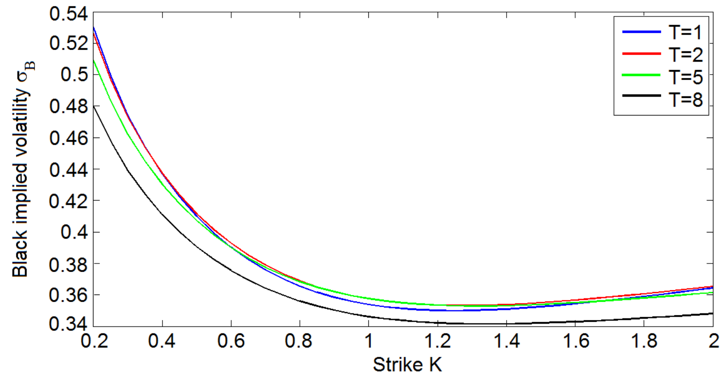

Figure 2.3: Volatility smiles calculated with the finite difference reference solver.

(BPS) or more [2], even for options for which the strike K is equal to the current value of the forward f, also called at the money (ATM) options.

In Figures 2.2 and 2.3 we can observe that the volatility smile given by the Hagan formula (2.3) can move in the wrong direction when the time to maturity T increases. For these figures, we used the parameters $\alpha = 0.35$, $\beta = 0.8$, $\rho = 0$, $\nu = 0.4$, $r = 0$ and $f = 1$. The volatility smiles for different times to maturity T are shown in these figures. In Figure 2.2 it is shown that, for this example, the level of the volatility smile raises when T increases. The solution given by the finite difference reference solver is shown in Figure 2.3. As explained in Section 1, in this thesis we consider the volatility smiles given by this solver as the reference smiles. As we can see in Figure 2.3 the curvature decreases and the level of the smile lowers when time to maturity T increases. This incorrect behaviour of the Hagan smile in Figure 2.2 does not only happen for this example, but it happens in many cases. This is an indication for the fact that the Hagan formula (2.3) is only accurate for small time to maturity.

2.2 The method of Antonov, Konikov and Spector

In both [2] and [3], A. Antonov, M. Konikov and M. Spector introduced a method to price European options with the SABR model. In this section, we describe this pricing approach and some of its advantages and disadvantages. In Section 2.2.1 an analytical formula to price European call options when $\rho = 0$ is given. We give a method to map the parameters of the correlated SABR model to parameters of an uncorrelated model in Section 2.2.2. In Section 2.2.3 we consider the (absence of) arbitrage in the method of Antonov, Konikov and Spector. Finally, we describe some of the advantages and disadvantages of this pricing approach in Section 2.2.4.

2.2.1 Call price for the zero correlation SABR

Consider the FSDEs of the SABR model (2.1) and (2.2) with $\mathrm{d}W_t^1\,\mathrm{d}W_t^2 = 0$ and $\beta \in [0, 1)$. In [2, 3] an analytical formula is given to calculate the forward value of a European call option in this zero correlation case, which we denote by V_{ZC}^C, is derived

$$\begin{aligned} V_{\mathrm{ZC}}^C(0, T, K, f, \alpha, \beta, \nu) &= (f-K)^+ + \frac{2}{\pi}\sqrt{Kf}\left[\int_{s_-}^{s_+} \frac{\sin(\eta\phi(s))}{\sinh(s)} G\left(T\nu^2, s\right) \mathrm{d}s \right. \\ &+ \left. \sin(\eta\pi)\int_{s_+}^{\infty} \frac{e^{-\eta\psi(s)}}{\sinh(s)} G\left(T\nu^2, s\right)\, \mathrm{d}s\right], \end{aligned} \tag{2.11}$$

where

$$G(t, s) = \frac{2e^{-\frac{t}{8}}}{t\sqrt{\pi t}} \int_s^{\infty} u\, e^{-\frac{u^2}{2t}} \sqrt{\cosh(u) - \cosh(s)}\ \mathrm{d}u, \tag{2.12a}$$

$$\phi(s) = 2\arctan\left(\sqrt{\frac{\sinh^2(s) - \sinh^2(s_-)}{\sinh^2(s_+) - \sinh^2(s)}}\right), \tag{2.12b}$$

$$\psi(s) = 2\,\mathrm{arctanh}\left(\sqrt{\frac{\sinh^2(s) - \sinh^2(s_+)}{\sinh^2(s) - \sinh^2(s_-)}}\right), \tag{2.12c}$$

$$s_\pm = \mathrm{arcsinh}\left(\frac{\nu|q \pm q_0|}{\alpha}\right), \tag{2.12d}$$

$$q = \frac{K^{1-\beta}}{1-\beta}, \qquad q_0 = \frac{f^{1-\beta}}{1-\beta}, \qquad \eta = \left|\frac{1}{2(\beta-1)}\right|. \tag{2.12e}$$

As mentioned, the price of a European put can be calculated from the price of a European call by the using put-call parity (2.9). The analytical formula (2.11) consists of two double integrals. These integrals can be calculated numerically. This integration is slower but more accurate than the easier to employ Hagan formula (2.3). This improvement in accuracy is especially visible for long time to maturity and/or low strikes. To improve calculation speed the authors in [2] gave an approximation for the function $G(t, s)$.

$$G(t, s) \approx \sqrt{\frac{\sinh(s)}{s}} e^{-\frac{s^2}{2t} - \frac{t}{8}} \left[R(t, s) + \delta R(t, s)\right], \tag{2.13}$$

where

$$\begin{aligned} R(t, s) &= 1 + \frac{3tg(s)}{8s^2} - \frac{5t^2\left(-8s^2 + 3g^2(s) + 24g(s)\right)}{128s^4}, \\ &+ \frac{35t^3\left(-40s^2 + 3g^3(s) + 24g^2(s) + 120g(s)\right)}{1024s^6}, \end{aligned} \tag{2.14a}$$

$$\delta R(t, s) = e^{\frac{t}{8}} - \frac{3072 + 384t + 24t^2 + t^3}{3072}, \tag{2.14b}$$

$$g(s) = s\coth(s) - 1. \tag{2.14c}$$

For completeness, we give the derivation of approximation (2.13) in Appendix A.

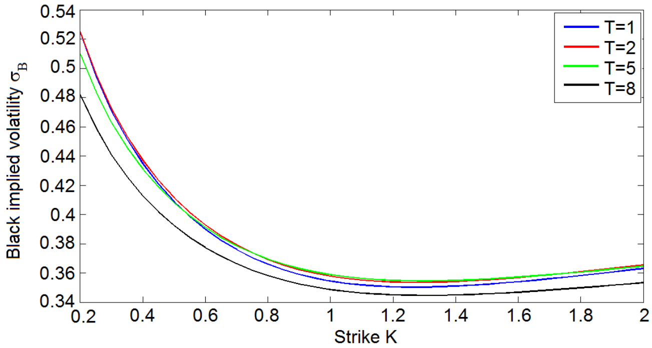

Figure 2.4: Volatility smiles calculated with (2.11).

The authors in [2] provided an exact solution for the zero correlation SABR model. Comparing Figures 2.3 and 2.4 confirms this claim for the example given in Section 2.1.2, as indeed the volatility smile moves in the correct direction in Figure 2.4.

2.2.2 Mapping to zero correlation case

In practice, the correlation in the SABR model is of course nonzero, i.e. $\mathrm{d}W_1\,\mathrm{d}W_2 = \rho\,\mathrm{d}t$. Antonov, Konikov and Spector explained how they used mapping techniques to produce a mimicking model. In this case, the general SABR-model is mimicked by a SABR model with zero correlation. For this mimicking model one can calculate the option price with formula (2.11) and this price can be used as an approximation of the option price under the original SABR model with correlation ρ.

The FSDEs of the mimicking model read:

$$\begin{aligned} \mathrm{d}\widetilde{F}_t &= \widetilde{\sigma}_t \left(\widetilde{F}_t\right)^{\widetilde{\beta}} \mathrm{d}\widetilde{W}_t^1, \qquad \widetilde{F}_0 = f, && (2.15)\\ \mathrm{d}\widetilde{\sigma}_t &= \widetilde{\nu}\widetilde{\sigma}_t\,\mathrm{d}\widetilde{W}_t^2, \qquad \widetilde{\sigma}_0 = \widetilde{\alpha}, && (2.16)\\ \mathrm{d}\widetilde{W}_t^1\,\mathrm{d}\widetilde{W}_t^2 &= 0. \end{aligned}$$

The authors in [2] fixed $\widetilde{\beta}$ and $\widetilde{\nu}$ and they approximated $\widetilde{\alpha}$ from these two parameters as[2]

$$\widetilde{\alpha} = \widetilde{\alpha}^{(0)} + T\widetilde{\alpha}^{(1)} + \mathcal{O}\left(T^2\right), \tag{2.17}$$

where

$$\widetilde{\alpha}^{(0)} = \frac{2\Phi\delta\widetilde{q}\widetilde{\nu}}{\Phi^2-1}, \qquad \Phi = \left(\frac{\sigma_{\min}+\rho\alpha+\nu\delta q}{(1+\rho)\alpha}\right)^{\frac{\widetilde{\nu}}{\nu}}, \qquad \delta\widetilde{q} = \frac{K^{1-\widetilde{\beta}}-f^{1-\widetilde{\beta}}}{1-\widetilde{\beta}}, \tag{2.18a}$$

$$\widetilde{\alpha}^{(1)} = \widetilde{\alpha}^{(0)}\widetilde{\nu}^2 \left[\frac{\frac{1}{2}(\beta-\widetilde{\beta})\log(Kf)+\frac{1}{2}\log(\alpha\sigma_{\min})-\frac{1}{2}\log\left(\widetilde{\alpha}^{(0)}\sqrt{\delta\widetilde{q}^2\widetilde{\nu}^2+\left(\widetilde{\alpha}^{(0)}\right)^2}\right)-\mathcal{B}_{\min}}{\frac{\Phi^2-1}{\Phi^2+1}\log(\Phi)}\right], \tag{2.18b}$$

$$\sigma_{\min} = \sqrt{\nu^2\delta q^2+2\rho\nu\delta q\alpha+\alpha^2}, \qquad \delta q = \frac{K^{1-\beta}-f^{1-\beta}}{1-\beta}, \tag{2.18c}$$

[2]Note there is an error in the formula for $\widetilde{\alpha}^{(1)}$ given in [2], the correct formula is given in [3]

$$\mathcal{B}_{\min} = -\frac{\beta\rho}{2(1-\beta)\sqrt{1-\rho^2}}\left(\pi - \varphi_0 - \arccos(\rho) - I\right), \quad \varphi_0 = \arccos\left(-\frac{\delta q\nu + \alpha\rho}{\sigma_{\min}}\right), \tag{2.18d}$$

$$I = \begin{cases} \frac{2}{\sqrt{1-L^2}}\left[\arctan\left(\frac{u_0+L}{\sqrt{1-L^2}}\right) - \arctan\left(\frac{L}{\sqrt{1-L^2}}\right)\right], & \text{for} \quad L < 1, \\ \frac{1}{\sqrt{L^2-1}}\log\left(\frac{u_0\left(L+\sqrt{L^2-1}\right)+1}{u_0\left(L-\sqrt{L^2-1}\right)+1}\right), & \text{for} \quad L > 1, \end{cases} \tag{2.18e}$$

$$u_0 = \frac{\delta q\nu\rho + \alpha - \sigma_{\min}}{\delta q\nu\sqrt{1-\rho^2}}, \qquad L = \frac{\sigma_{\min}(1-\beta)}{K^{1-\beta}\nu\sqrt{1-\rho^2}}. \tag{2.18f}$$

In [2], Antonov, Konikov and Spector claimed that good choices for $\widetilde{\beta}$ and $\widetilde{\nu}$ are given by

$$\widetilde{\beta} = \beta \qquad \text{and} \qquad \widetilde{\nu}^2 = \nu^2 - \frac{3}{2}\left[\nu^2\rho^2 + \alpha\nu\rho(1-\beta)f^{\beta-1}\right]. \tag{2.19}$$

They did not give a proof of the exactness of these choices and they mentioned that they based their choices primarily on numerical experiments. In some cases, mapping the SABR model to the mimicking model with this heuristic choices of $\widetilde{\beta}$ and $\widetilde{\nu}$ is however impracticable, e.g. $\widetilde{\nu}$ is not a real number if $\rho \in \left(\sqrt{2/3}, 1\right]$ or $\rho \approx -1$.

We approximate the forward price of a European call option by using formula (2.11), i.e.

$$V^C_{\text{ZCmap}}(0, T, K, f, \alpha, \beta, \nu, \rho) \approx V^C_{\text{ZC}}(0, T, K, f, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\nu}), \tag{2.20}$$

where we denote the forward value of a European call option calculated by the mapping procedure of [2] by V^C_{ZCmap}.

2.2.3 Arbitrage

The authors in [2, 3] claimed that their approach is nearly arbitrage-free. This section gives a view on the (absence of) arbitrage in the zero correlation pricing method (Section 2.2.3) and the mapping to the zero correlation approach (Section 2.2.3) of Antonov, Konikov and Spector. From now on, we call this method Antonov's method or Antonov's pricing approach.

Arbitrage in zero correlation model

Theoretically, Antonov's pricing approach for the zero correlation model is arbitrage-free, because its developers gave an analytical formula for pricing in the zero correlation model. This analytical formula is the exact solution to the model and therefore it leads to arbitrage-free option prices. Antonov's method is also compatible with the put-call parity (2.9), because the method uses the parity to calculate put prices from call prices. Practically, numerical errors in approximating the double integrals in (2.11) can however introduce arbitrage.

In Section 2.1.1, we provide an example of arbitrage as a result of pricing with the Hagan formula (2.3). The PDFs of both Hagan's formula and Antonov's pricing approach, where $\rho = 0$ and where the other parameters are the same as in Section 2.1.1, are shown in Figure 2.5. This figure shows that for this example Antonov's method is, neglecting numerical errors, free of arbitrage, because the PDF is non-negative and the integral over the entire space equals one.

Arbitrage in mapping to the zero correlation model

The analysis of arbitrage in the case of the mapping to the zero correlation model is more involved. The mapping parameter $\widetilde{\alpha}$ is strike dependent, i.e. for every strike value K the parameter $\widetilde{\alpha}$ is different, while the mapping parameters $\widetilde{\beta}$ and $\widetilde{\nu}$ do not depend on the strike K.

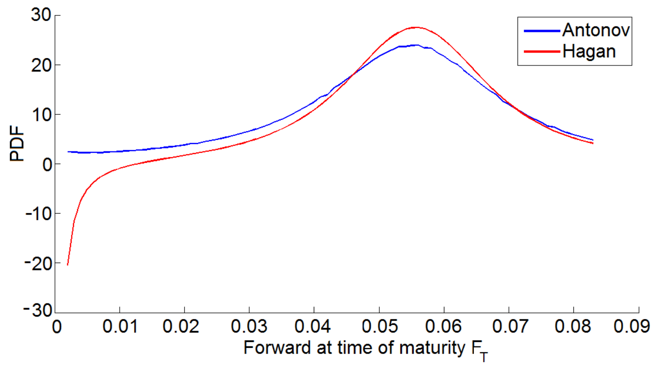

Figure 2.5: The (incorrect) conditional PDF of F_T given $F_0 = f$.

For $\rho \approx 0$ it holds that $\widetilde{\alpha}(K)$ is approximately constant and for large values of $|\rho|$ the function $\widetilde{\alpha}(K)$ behaves more like a higher degree polynomial. Figure 2.6 shows the function $\widetilde{\alpha}(K)$ for different values of ρ, while the other parameters are the same as in Section 2.2.3.

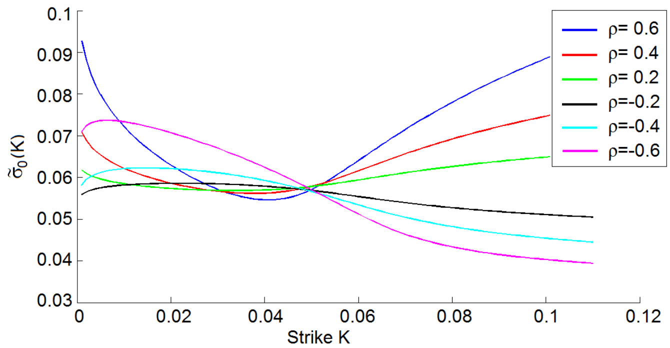

Figure 2.6: $\widetilde{\alpha}$ for different values of ρ.

The non-constant function $\widetilde{\alpha}(K)$ can result in arbitrage possibilities. Function $\widetilde{\alpha}(K)$ has specifically influence on $\frac{\partial^2 V_{ZC}^C}{\partial K^2}$ when both $\frac{\mathrm{d}\widetilde{\alpha}}{\mathrm{d}K}$ and $\frac{\mathrm{d}^2\widetilde{\alpha}}{\mathrm{d}K^2}$ are large. As in (2.10) we have[3]

$$\left.\frac{\partial^2 V_{ZC}^C}{\partial K^2}\right|_{K=F_T} = Q_{F_T}(T, F|f, \alpha). \tag{2.21}$$

This implies that $\widetilde{\alpha}(K)$ has influence on PDF $Q_{F_T}(T, F|f, \alpha)$ when both $\left|\frac{\mathrm{d}\widetilde{\alpha}}{\mathrm{d}K}\right|$ and $\left|\frac{\mathrm{d}^2\widetilde{\alpha}}{\mathrm{d}K^2}\right|$ are large.

If $\widetilde{\alpha}(K)$ is constant, Antonov's pricing approach represents the arbitrage-free zero correlation case (2.11). For small values of $|\rho|$, Antonov's pricing approach remains in general arbitrage-free, because for these values of ρ it follows that $\left|\frac{\mathrm{d}\widetilde{\alpha}}{\mathrm{d}K}\right|$ and $\left|\frac{\mathrm{d}^2\widetilde{\alpha}}{\mathrm{d}K^2}\right|$ are also small. On the other hand, large values of $|\rho|$ can lead to non-convex option prices (mostly for high strikes). This is a consequence of the fact that $\left|\frac{\mathrm{d}\widetilde{\alpha}}{\mathrm{d}K}\right|$ and $\left|\frac{\mathrm{d}^2\widetilde{\alpha}}{\mathrm{d}K^2}\right|$ are often relatively large for large values of $|\rho|$ and therefore they have a significant influence on $Q_{F_T}(T, F|f, \alpha)$.

These arbitrage possibilities are thus generally present for large $|\rho|$, small β and long time to

[3]Since formula (2.20) is for the forward option value we can omit the discount factor

maturity T, as is shown in Figure 2.7. The result in this figure are based on the parameters $f = 0.05735$, $\beta = 0.1$, $\alpha = 0.05735$, $\rho = -0.7$, $\nu = 0.47$ and $T = 20$.

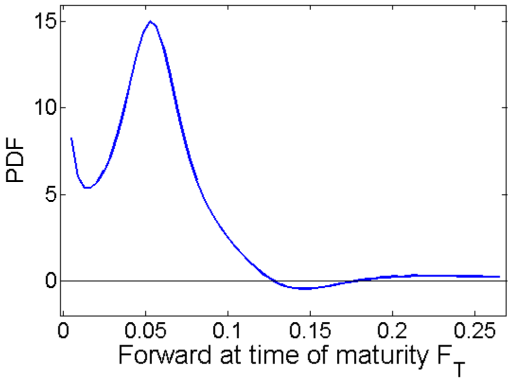

Figure 2.7: Incorrect PDF determined with Antonov's method.

2.2.4 Brief analysis

In this section we give a view on some properties of Antonov's option pricing approach in relation to the Hagan formula (2.3) for pricing European options under the SABR model.

- Antonov's [2] pricing approach is near arbitrage-free, wheras option pricing with the Hagan formula often implies arbitrage for low strikes as explained in Section 2.1.1.
- Antonov's method gives an analytical formula for option pricing in the zero correlation case. This analytical formula is an exact solution for the model and can be easily computed. On the other hand, the Hagan formula is in most cases not accurate for long time to maturity.
- Pricing with Antonov's method is slower than with the Hagan formula. Also, the mapping parameter $\widetilde{\alpha}$ is strike-dependent when considering the mapping. This makes the pricing procedure expensive when $\rho \neq 0$.
- The Hagan formula is applicable for every value of β, α, ρ and ν. Antonov's method is not applicable for several values of the parameters, e.g. mapping parameter $\widetilde{\nu} \notin \mathbb{R}$ for high values of $|\rho|$.
- The mapping parameters given in Section 2.2.2 are based on heuristics and it is not proven that these parameters are the optimal choices.
- The method of Antonov, Konikov and Spector is in general more accurate for longer time to maturity than Hagan's formula. A requirement for this is that the mapping parameters exist. The derivation of parameter $\widetilde{\alpha}$ is accurate to $\mathcal{O}(T)$, so this parameter is less accurate for long maturities T. Still, Antonov's method behaves better than Hagan's formula for long time to maturity, especially when the correlation parameter ρ is small.

2.3 Hagan's arbitrage-free approach

Hagan, Kumar, Lesniewski and Woodward improved their formula (2.3) to an arbitrage-free option pricing approach. In 2013 they introduced another pricing method for the SABR model

[15]. In this thesis, we call this method Hagan's arbitrage-free pricing approach. The authors in [15] reduced the SABR model from two dimensions to one dimension and derived a PDE for the conditional PDF $Q_{F_T}(T, F|f, \alpha)$, which is exact up to $\mathcal{O}(T)$. Hagan's arbitrage-free method is a small maturity method, just as the original Hagan formula. The PDE can be solved numerically with a finite difference scheme, e.g. the Crank-Nicolson scheme. In Section 2.3.1, we give the PDE and the finite difference scheme for this PDE is given in Section 2.3.2. Next, Section 2.3.3 shows under which conditions this pricing approach is free of arbitrage. We give the formulas to price European options with this method in Section 2.3.4 and we discusses the accuracy of these formulas in Section 2.3.5. Finally, some advantages and disadvantages of this pricing approach are given in Section 2.3.6.

2.3.1 The PDE for the conditional probability density function

The idea of Hagan's arbitrage-free method [15] is to derive a PDE for the conditional PDF $Q_{F_t}(t, F|f, \alpha)$ with singular perturbation methods. In this section, we abbreviate $Q_{F_t}(t, F|f, \alpha)$ by $Q(t, F)$. Like the Hagan formula, the computed $Q(t, F)$ is accurate for small time to maturity. When $Q(t, F)$ is computed for some $t \in [0, T]$, one can determine the forward value of European options with time to maturity t by using the following formulas

$$V^C(0, t, K, f, \alpha) = \int_K^\infty (F-K)Q(t, F)\ \mathrm{d}F, \quad V^P(0, t, K, f, \alpha) = \int_{-\infty}^K (K-F)Q(t, F)\ \mathrm{d}F. \tag{2.22}$$

This pricing method is free of arbitrage when $Q(t, F)$ is a probability density function and when the put-call parity (2.9) holds. Therefore, the requirements for this method to be arbitrage-free are

- $Q(t, F) \geq 0$,
- $\int_{-\infty}^\infty Q(t, F)\ \mathrm{d}F = 1$,
- $V^C(0, t, K, f, \alpha) - V^P(0, t, K, f, \alpha) = f - K$. When $\int_{-\infty}^\infty Q(t, F)\,\mathrm{d}F = 1$,

$$\begin{aligned} V^C(0, t, K, f, \alpha) - V^P(0, t, K, f, \alpha) &= \int_K^\infty (F-K)Q(t, F)\ \mathrm{d}F - \int_{-\infty}^K (K-F)Q(t, F)\ \mathrm{d}F \\ &= \int_{-\infty}^\infty FQ(t, F)\ \mathrm{d}F - K, \end{aligned}$$

  which results in the last requirement $\int_{-\infty}^\infty FQ(t, F)\ \mathrm{d}F = f$.

The authors in [15] derived a PDE for $Q(t, F)$ and they wanted to solve this PDE on a finite domain $[F_{\min}, F_{\max}]$ where $0 \leq F_{\min} < f < F_{\max}$. $F_{\min} = 0$ is often a good choice. Hagan, Kumar, Lesniewski and Woodward used a finite difference method to solve the PDE on a two-dimensional grid. They defined

$$Q(t, F) = \begin{cases} 0, & \text{for } F < F_{\min}, \\ Q^L(t)\delta\left(F - F_{\min}\right), & \text{at } F = F_{\min}, \\ Q^c(t, F), & \text{for } F_{\min} < F < F_{\max}, \\ Q^R(t), \delta\left(F - F_{\max}\right) & \text{at } F = F_{\max}, \\ 0, & \text{for } F > F_{\max}. \end{cases} \tag{2.23}$$

Substitution of (2.23) in the three requirements above results in the following no-arbitrage requirements:

$$Q^L(t), Q^c(t, F), Q^R(t) \quad \geq \quad 0, \tag{2.24}$$

$$Q^L(t) + \int_{F_{\min}}^{F_{\max}} Q^c(t,F)\ \mathrm{d}F + Q^R(t) = 1, \tag{2.25}$$

$$F_{\min}Q^L(t) + \int_{F_{\min}}^{F_{\max}} FQ^c(t,F)\ \mathrm{d}F + F_{\max}Q^R(t) = f. \tag{2.26}$$

Just as in their original pricing method (2.3), the authors in [15] used singular perturbation techniques to determine the following PDE

$$\frac{\partial Q^c(t,F)}{\partial t} = \frac{\partial^2}{\partial F^2}\left(M(t,F)Q^c(t,F)\right), \tag{2.27}$$

where

$$M(t,F) = \frac{1}{2}\alpha^2\left(1+2\rho\nu z(F)+\nu^2 z^2(F)\right)\exp\left(\rho\nu\alpha\Gamma(F)t\right)F^{2\beta}, \tag{2.28}$$

$$z(F) = \begin{cases} \frac{F^{1-\beta}-f^{1-\beta}}{\alpha(1-\beta)}, & \text{for } 0 \le \beta < 1, \\ \frac{1}{\alpha}(\log(F)-\log(f)), & \text{for } \beta = 1, \end{cases} \tag{2.29}$$

$$\Gamma(F) = \begin{cases} \frac{F^{\beta}-f^{\beta}}{F-f}, & \text{for } F \neq f, \\ \beta F^{\beta-1}, & \text{for } F = f. \end{cases} \tag{2.30}$$

These equations come with the boundary conditions,

$$\lim_{F\downarrow F_{\min}} M(t,F)Q^c(t,F) = 0, \qquad \lim_{F\uparrow F_{\max}} M(t,F)Q^c(t,F) = 0, \tag{2.31}$$

and for $0 < t < T$, we find

$$\frac{\mathrm{d}Q^L(t)}{\mathrm{d}t} = \lim_{F\downarrow F_{\min}} \frac{\partial}{\partial F}\left(M(t,F)Q^c(t,F)\right), \quad \frac{\mathrm{d}Q^R(t)}{\mathrm{d}t} = \lim_{F\uparrow F_{\max}} \frac{\partial}{\partial F}\left(M(t,F)Q^c(t,F)\right). \tag{2.32}$$

The initial conditions are

$$Q^L(0) = 0, \ \ Q^R(0) = 0, \text{ at } F = F_{\max}, \tag{2.33}$$

$$\lim_{t\downarrow 0} Q^c(t,F) = \delta(F-f), \text{ for } F_{\min} < F < F_{\max}. \tag{2.34}$$

The derivation of (2.27) is given in Appendix B.

2.3.2 Finite difference scheme

The authors in [15] employed the Crank-Nicolson scheme to solve (2.27). In [11], Le Floc'h and Kennedy used alternative schemes for solving this PDE. Le Floc'h and Kennedy also compared the results and properties of these different schemes. We will now follow [15].

Define a two-dimensional grid, where N and J are the number of time steps and the number of steps in the forward F, respectively. Let $\Delta t = T/N$ define the size of a time step and make sure that J is chosen such that $F_{\max} = F_{\min} + Jh$ and $f = F_{\min} + (j_0 - 1/2)h$ for some $j_0 \in \{1, 2, \ldots J\}$. Define $F_j = F_{\min} + (j - 1/2)h$ for $j = 1, \ldots J$ and let cell j be defined by $[F_j - h/2, F_j + h/2]$.

Let $Q_j^n = Q^c(n\Delta t, F_j)$ be the PDF of $F = F_j$ at time $n\Delta t$, where $n = 0, 1, \ldots, N$. Assume that the probability hQ_j^n is spread uniformly in each cell j and let $M_j^n = M(n\Delta t, F_j)$, $Q_L^n = Q^L(n\Delta t)$, and $Q_R^n = Q^R(n\Delta t)$. Using the Crank-Nicolson scheme we obtain the following discretization of

(2.27):

For $j = 2, \ldots, J-1$,

$$\begin{aligned} Q_j^{n+1} &= Q_j^n + \frac{\Delta t}{2h^2}\left\{M_{j+1}^{n+1}Q_{j+1}^{n+1} - 2M_j^{n+1}Q_j^{n+1} + M_{j-1}^{n+1}Q_{j-1}^{n+1}\right\} \\ &+ \frac{\Delta t}{2h^2}\left\{M_{j+1}^{n}Q_{j+1}^{n} - 2M_j^{n}Q_j^{n} + M_{j-1}^{n}Q_{j-1}^{n}\right\}, \end{aligned} \tag{2.35}$$

and for the boundaries

$$Q_1^{n+1} = Q_1^n + \frac{\Delta t}{2h^2}\left\{M_2^{n+1}Q_2^{n+1} - 3M_1^{n+1}Q_1^{n+1} + M_2^nQ_2^n - 3M_1^nQ_1^n\right\}, \tag{2.36}$$

$$Q_J^{n+1} = Q_J^n + \frac{\Delta t}{2h^2}\left\{M_{J-1}^{n+1}Q_{J-1}^{n+1} - 3M_J^{n+1}Q_J^{n+1} + M_{J-1}^nQ_{J-1}^n - 3M_J^nQ_J^n\right\}, \tag{2.37}$$

and

$$Q_L^{n+1} = Q_L^n + \frac{\Delta t}{h}\left\{M_1^{n+1}Q_1^{n+1} + M_1^nQ_1^n\right\}, \tag{2.38}$$

$$Q_R^{n+1} = Q_R^n + \frac{\Delta t}{h}\left\{M_J^{n+1}Q_J^{n+1} + M_J^nQ_J^n\right\}, \tag{2.39}$$

where the initial conditions are given by

$$Q_L^0 = 0, \qquad Q_j^0 = \begin{cases} 0, & \text{for } j \neq j_0, \\ 1/h, & \text{for } j = j_0, \end{cases} \qquad Q_R^0 = 0. \tag{2.40}$$

To advance from time $n\Delta t$ to time $(n+1)\Delta t$ one can solve the system (B.36) with the tridiagonal matrix algorithm, also known as the Thomas algorithm.

2.3.3 Arbitrage

As explained in Section 2.3.1, there are three requirements necessary for the discretization to be free of arbitrage (2.24), (2.25) and (2.26). Requirements (2.25) and (2.26) hold for the Crank-Nicolson scheme, which can easily be shown by induction. Requirement (2.24) only holds for specific choices of J and N. Here, we give an analysis for the system (B.36) and derive an intuitive bound[4] for N compared to J by means of induction. This analysis uses the claim that $M(F,t) \geq 0$ for all $F \geq 0$ and $t \geq 0$. In Appendix B we give a proof of this claim.

The derivation of this intuitive bound is also based on induction.
One can easily observe that for $n = 0$ it holds that

$$Q_L^n, Q_j^n, Q_R^n \geq 0, \qquad \text{for all } j \in \{1, 2, \ldots, J\}. \tag{2.41}$$

Let us assume, for some $n \in \{1, 2, \ldots N-1\}$,

$$Q_L^n, Q_j^n, Q_R^n \geq 0, \qquad \text{for all } j \in \{1, 2, \ldots, J\}, \tag{2.42}$$

then Q_j^{n+1} can be found by solving system (B.36). Let (2.43) be the abbreviation of (B.36), where $\mathbf{A}$ is the concerning $J \times J$ matrix and $\widetilde{Q}^n$ is as defined in Appendix B.

$$\mathbf{A} \cdot Q^{n+1} = \widetilde{Q}^n \tag{2.43}$$

[4] The exact bound depends on α, β, ν, ρ, f and T.

One can observe that $\mathbf{A}_{ii} > 0$ and $\mathbf{A}_{ij} \leq 0$ for all $i, j \in \{1, \ldots, J\}$, where $j \neq i$, because $M(t, F)$ is a non-negative function. Also, when $\frac{\Delta t}{h^2}$ is small enough, then not only is A irreducibly diagonal dominant, but also holds $\widetilde{Q}^n \geq 0$ there. $\mathbf{A}$ is an M-matrix [27, Theorem 2.10] and therefore $\mathbf{A}^{-1} \geq 0$. It holds that $Q^{n+1} \geq 0$ when both $\widetilde{Q}^n \geq 0$ and $\mathbf{A}^{-1} \geq 0$. Now, we can observe that $Q_L^{n+1} \geq 0$ and $Q_R^{n+1} \geq 0$. By induction requirement (2.24) always holds when $\frac{\Delta t}{h^2}$ is chosen small enough.

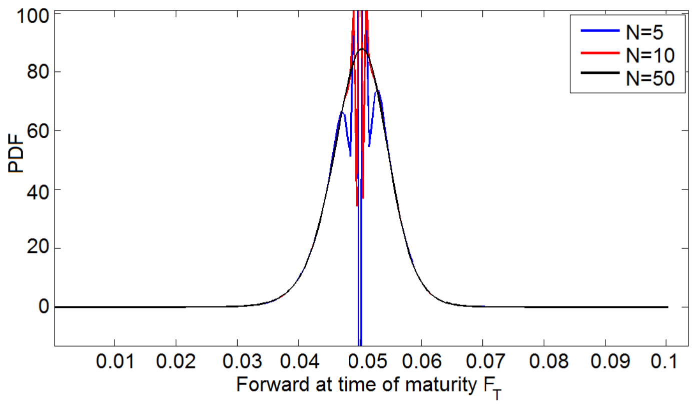

Figure 2.8: The (incorrect) conditional PDF of F_T given $F_0 = f$ for different N.

Requirement (2.24) does not hold when $\frac{\Delta t}{h^2}$ is too large, as shown in Figure 2.8. For the results in this figure we set the parameters $T = 0.5$, $f = 0.05$, $\beta = 0.5$, $\alpha = 0.03$, $\rho = -0.25$, $\nu = 0.6$, $F_{\min} = 0$, $j_0 = 100$ and $J = 500$. For small N there are oscillations in the PDF. For $N = 5$ these oscillations lead to negative densities and therefore to arbitrage. For $N = 10$ there are also oscillations, but no negative densities. In this case there is no arbitrage, but the PDF is not a smooth function like for $N = 50$.

2.3.4 Pricing formulas

One can price European options by substituting the approximations for $Q^c(T, F)$, $Q^L(T)$ and $Q^R(T)$ in formula (2.22), which gives for $F_{\min} < K < F_{\max}$:

$$V_H^C(0, T, K, f, \alpha) = (F_{\max} - K)\, Q_R^N + \sum_{j=j_k+1}^{J} h(F_j - K) Q_j^N + \frac{1}{2} Q_{j_k}^N (F_{\min} + j_k h - K)^2, \quad (2.44)$$

$$V_H^P(0, T, K, f, \alpha) = (K - F_{\min})\, Q_L^N + \sum_{j=1}^{j_k-1} h(K - F_j) Q_j^N + \frac{1}{2} Q_{j_k}^N (K - F_{\min} - j_k h + h)^2, \quad (2.45)$$

where $j_k \in \{1, \ldots, J\}$ such that $F_{\min} + (j_k - 1)h < K \leq F_{\min} + j_k h$ and

$$\begin{array}{lll} V_H^C(0, T, K, f, \alpha) = 0, & V_H^P(0, T, K, f, \alpha) = K - f, & \text{when } K > F_{\max}, \\ V_H^C(0, T, K, f, \alpha) = f - K, & V_H^P(0, T, K, f, \alpha) = 0, & \text{when } K < F_{\min}, \end{array} \quad (2.46)$$

where we denote the forward values of a European call option and a European put option calculated with Hagan's arbitrage-free pricing approach [15] by V_H^C and V_H^P, respectively.

2.3.5 Accuracy of Hagan's arbitrage-free method

The PDE (2.27) for the conditional PDF $Q(T, F)$ is exact up to $\mathcal{O}(T)$ [15]. Hagan's arbitrage-free pricing approach is thus only accurate for small maturities, as is it shown in Figure 2.9. The results in this figure are based on the parameters from Section 2.1.2 and comparison of this figure with Figure 2.3 shows that the smile for $T = 1$ is accurate. However when T increases, the smile given by Hagan's arbitrage-free method moves in the wrong direction. The Hagan formula exhibits this same behaviour.

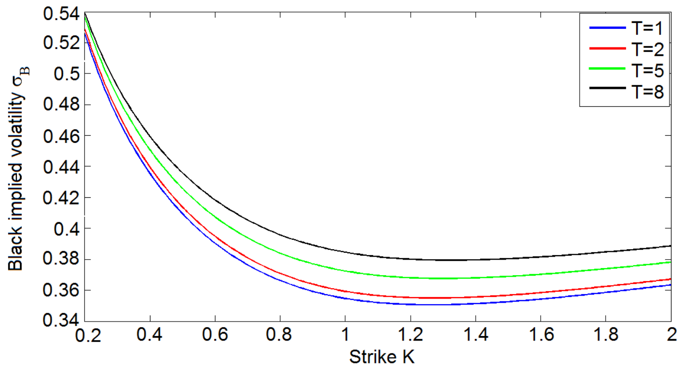

Figure 2.9: Volatility smiles calculated with Hagan's arbitrage-free pricing approach.

2.3.6 Brief analysis

In this section we summarize some properties of the arbitrage-free option pricing method by Hagan, Kumar, Lesniewski and Woodward. We compare it with their original method described in Section 2.1.

- Both methods are small maturity methods, which means that the methods are not accurate for long time to maturity.
- Hagan's arbitrage-free method is arbitrage-free when $\frac{\Delta t}{h^2}$ is small enough. So, it is possible to obtain arbitrage-free prices by decreasing the size of a time step Δt or increasing discretization step h. The Hagan formula (2.3) is not always free of arbitrage as it is shown in Section 2.1.1.
- Pricing with the Hagan arbitrage-free method is slower than with the Hagan formula.
- For Hagan's arbitrage-free pricing approach one has to solve a linear system on a grid for every time step. For increased accuracy the size of the grid should increase. As a consequence also the number of time steps has to be increased to ensure arbitrage-free prices. As a result, both the required computer memory and the CPU time will increase.

We wish to provide an arbitrage-free method to price European options under the SABR model that is accurate. Since the pricing methods discussed in this chapter are or not arbitrage-free, or not accurate for long time to maturity, or both, we develop a new pricing method in Chapters 4-6.

Chapter 3

The BCOS method

M.J. Ruijter and C.W. Oosterlee developed a Fourier method [25] to solve backward stochastic differential equations (BSDEs) using the characteristic function of the underlying process. This method is called the one-dimensional BCOS method (Backward Stochastic Differential Equation COS method). When the underlying forward stochastic differential equation (FSDE) can be written as

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad X_0 = x, \quad t \geq 0, \tag{3.1}$$

and the corresponding characteristic function cannot easily be derived, we can use the characteristic function of a discrete forward process to approximate the solution [24]. The underlying FSDE (3.1) is approximated by different Taylor schemes, such as the Euler, Milstein and Order 2.0 weak Taylor schemes.

In this section we explain the BCOS method where the characteristic function is approximated by the characteristic function of the discrete forward process and the contents of this chapter is quite similar to the method in [24]. In Section 3.1 we describe the discretization of the forward process X_t by different Taylor schemes and give the corresponding characteristic functions. We give an introduction of the COS method in Section 3.2. In Section 3.3 we give the BCOS method to price financial derivatives for which the underlying is of the form (3.1). Formulas to approximate some conditional expectations are given in Section 3.4. In Section 3.5 we present briefly how to apply the BCOS method for a European option (without early-exercise dates). In this section we also explain how to proceed if there are early-exercise dates in the option contract. In Section 3.6 an error analysis is given. Finally, we give two examples of pricing with the BCOS method in Section 3.7. We will employ the one-dimensional BCOS method to price options under the SABR model in Chapter 4.

3.1 The discrete forward process and its characteristic function

Just like Ruijter and Oosterlee [24], we define[1] a time-grid $t_j = j\Delta t$ for $j = 0, 1, \ldots, M$, with fixed time steps $\Delta t = \frac{T}{M}$. We write $X_m = X_{t_m}$, $W_m = W_{t_m}$ and $\Delta W_{m+1} = W_{m+1} - W_m$. The discrete forward process is denoted by $X_m^\Delta = X_{t_m}^\Delta$, where $X_0^\Delta = X_0$. To determine the values X_{m+1}^Δ, for $m = 0, \ldots, M-1$, we use one of the following Taylor schemes: Euler, Milstein, or 2.0 weak Taylor.

[1] From now on we redefine some symbols, e.g. M, N and z which are defined differently in Chapter 2.

According to Kloeden and Platen [18] the definitions of the order of strong convergence and the order of weak convergence are:

Definition 3.1. *The approximating process* X^Δ *converges in the strong sense with order* $\gamma_1 \in (0,\infty]$ *if there exists a finite constant* C *and a positive constant* δ *such that*

$$\mathbb{E}\left[\left|X_M - X_M^\Delta\right|\right] \leq C(\Delta t)^{\gamma_1}, \tag{3.2}$$

for any time discretization with maximum step size $\Delta t \in (0,\delta)$.

Definition 3.2. *The approximating process* X^Δ *converges in the weak sense with order* $\gamma_2 \in (0,\infty]$ *if for any polynomial* g *there exists a finite constant* C *and a positive constant* δ *such that*

$$\left|\mathbb{E}\left[g\left(X_M\right)\right] - \mathbb{E}\left[g\left(X_M^\Delta\right)\right]\right| \leq C(\Delta t)^{\gamma_2}, \tag{3.3}$$

for any time discretization with maximum step size $\Delta t \in (0,\delta)$.

The Euler scheme
The Euler approximation for FSDE (3.1) has the form

$$X_{m+1}^\Delta = X_m^\Delta + \mu(X_m^\Delta)\Delta t + \sigma(X_m^\Delta)\Delta W_{m+1}, \tag{3.4}$$

for $m = 0,\ldots,M-1$. The order of strong convergence is $\gamma_1 = 0.5$ and the order of weak convergence is $\gamma_2 = 1$.

The Milstein scheme
The Milstein approximation for FSDE (3.1) has the form

$$X_{m+1}^\Delta = X_m^\Delta + \mu(X_m^\Delta)\Delta t + \sigma(X_m^\Delta)\Delta W_{m+1} + \frac{1}{2}\sigma(X_m^\Delta)\frac{\mathrm{d}\sigma(X_m^\Delta)}{\mathrm{d}X_m^\Delta}\left((\Delta W_{m+1})^2 - \Delta t\right), \tag{3.5}$$

for $m = 0,\ldots,M-1$ and where $\gamma_1 = 1$ and $\gamma_2 = 1$.

The weak Taylor scheme of order 2.0
The weak Taylor scheme of order 2.0 for FSDE (3.1), given $X_m^\Delta = x$, has the form [18]

$$\begin{aligned} X_{m+1}^\Delta &= x + \mu(x)\Delta t + \sigma(x)\Delta W_{m+1} + \frac{1}{2}\sigma(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}\left((\Delta W_{m+1})^2 - \Delta t\right) \\ &+ \frac{\mathrm{d}\mu(x)}{\mathrm{d}x}\sigma(x)\Delta Z_{m+1} + \frac{1}{2}\left(\mu(x)\frac{\mathrm{d}\mu(x)}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2\mu(x)}{\mathrm{d}x^2}\sigma^2(x)\right)(\Delta t)^2 \\ &+ \left(\mu(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x} + \frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2\sigma(x)}{\mathrm{d}x^2}\right)(\Delta W_{m+1}\Delta t - \Delta Z_{m+1}), \end{aligned} \tag{3.6}$$

where $\Delta Z_{m+1} = \frac{1}{2}\left(\Delta W_{m+1}\Delta t + \zeta_{m+1}(\Delta t)^{3/2}\right)$, $\zeta_{m+1} \sim \mathcal{N}(0,1/3)$ and $m = 0,\ldots,M-1$. For the weak Taylor of order 2.0 scheme it holds that $\gamma_1 = 1$ and $\gamma_2 = 2$.

We observe that

$$\mathbb{E}[\Delta Z_{m+1}] = 0, \quad \mathrm{Var}\left(\Delta Z_{m+1}\right) = \frac{1}{3}\left(\Delta t\right)^3 \quad \text{and} \quad \mathrm{Cov}\left(\Delta W_{m+1}, \Delta Z_{m+1}\right) = \frac{1}{2}\left(\Delta t\right)^2. \tag{3.7}$$

If we replace ΔZ_{m+1} by $\Delta Z_{m+1} = \frac{1}{2}\Delta W_{m+1}\Delta t$ as the authors in [24] suggested, then

$$\mathbb{E}[\Delta Z_{m+1}] = 0, \quad \mathrm{Var}\left(\Delta Z_{m+1}\right) = \frac{1}{4}\left(\Delta t\right)^3 \quad \text{and} \quad \mathrm{Cov}\left(\Delta W_{m+1}, \Delta Z_{m+1}\right) = \frac{1}{2}\left(\Delta t\right)^2. \tag{3.8}$$

This replacement has the same moments in first order and simplifies the scheme. Ruijter and Oosterlee called this new scheme 2.0-weak-Taylor and they observe that $\gamma_1 = 1$ and $\gamma_2 = 2$ for the 2.0-weak-Taylor scheme.

Similarly as in [24], we can write the discretization schemes in general form, as follows:

$$X^{\Delta}_{m+1} = x + m(x)\Delta t + s(x)\Delta W_{m+1} + \kappa(x)(\Delta W_{m+1})^2, \qquad X^{\Delta}_{m} = x. \tag{3.9}$$

For the Euler scheme, we find

$$m(x) = \mu(x), \quad s(x) = \sigma(x), \quad \kappa(x) = 0, \tag{3.10}$$

for the Milstein scheme, we have

$$m(x) = \mu(x) - \frac{1}{2}\sigma(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}, \quad s(x) = \sigma(x), \quad \kappa(x) = \frac{1}{2}\sigma(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}, \tag{3.11}$$

and for the 2.0-weak-Taylor scheme, we see that

$$m(x) = \mu(x) - \frac{1}{2}\sigma(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x} + \frac{1}{2}\left(\mu(x)\frac{\mathrm{d}\mu(x)}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2\mu(x)}{\mathrm{d}x^2}\sigma^2(x)\right)\Delta t, \tag{3.12a}$$

$$s(x) = \sigma(x) + \frac{1}{2}\left(\frac{\mathrm{d}\mu(x)}{\mathrm{d}x}\sigma(x) + \mu(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x} + \frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2\sigma(x)}{\mathrm{d}x^2}\right)\Delta t, \tag{3.12b}$$

$$\kappa(x) = \frac{1}{2}\sigma(x)\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}. \tag{3.12c}$$

For the discretization schemes above we can determine a characteristic function, which is given in the lemma below.

Lemma 3.1. *The characteristic function of X^{Δ}_{m+1}, given $X^{\Delta}_{m} = x$, is given by*

$$\begin{aligned}\phi_{X^{\Delta}_{m+1}}\left(u \,\middle|\, X^{\Delta}_{m} = x\right) &= \mathbb{E}\left[\exp\left(iuX^{\Delta}_{m+1}\right)\,\middle|\, X^{\Delta}_{m} = x\right] \\ &= \exp\left(iux + ium(x)\Delta t - \frac{\frac{1}{2}u^2 s^2(x)\Delta t}{1 - 2iu\kappa(x)\Delta t}\right)(1 - 2iu\kappa(x)\Delta t)^{-\frac{1}{2}}. \end{aligned}\tag{3.13}$$

Proof. For $\kappa(x) = 0$,

$$\begin{aligned}\phi_{X^{\Delta}_{m+1}}\left(u \,\middle|\, X^{\Delta}_{m} = x\right) &= \mathbb{E}\left[\exp\left(iuX^{\Delta}_{m+1}\right)\,\middle|\, X^{\Delta}_{m} = x\right] \\ &= \mathbb{E}\left[\exp\left(iux + ium(x)\Delta t + ius(x)\Delta W_{m+1}\right)\,\middle|\, X^{\Delta}_{m} = x\right] \\ &= \exp(iux + ium(x)\Delta t)\mathbb{E}\left[\exp\left(ius(x)\Delta W_{m+1}\right)\right], \end{aligned}\tag{3.14}$$

where $\Delta W_{m+1} \sim \mathcal{N}(0, \Delta t)$. This implies

$$\begin{aligned}\phi_{X^{\Delta}_{m+1}}\left(u \,\middle|\, X^{\Delta}_{m} = x\right) &= \exp(iux + ium(x)\Delta t)\phi_{\mathcal{N}(0,\Delta t)}(us(x)) \\ &= \exp\left(iux + ium(x)\Delta t - \frac{1}{2}u^2 s^2(x)\Delta t\right). \end{aligned}\tag{3.15}$$

For $\kappa(x) \neq 0$, we find

$$\begin{aligned}&\phi_{X^{\Delta}_{m+1}}\left(u \,\middle|\, X^{\Delta}_{m} = x\right) = \mathbb{E}\left[\exp\left(iuX^{\Delta}_{m+1}\right)\,\middle|\, X^{\Delta}_{m} = x\right] \\ &\quad = \mathbb{E}\left[\exp\left(iux + ium(x)\Delta t + ius(x)\Delta W_{m+1} + iu\kappa(x)(\Delta W_{m+1})^2\right)\right]\end{aligned}$$

$$= \mathbb{E}\left[\exp\left(iux + ium(x)\Delta t + iu\kappa(x)\left(\Delta W_{m+1} + \frac{1}{2}\frac{s(x)}{\kappa(x)}\right)^2 - iu\frac{1}{4}\frac{s^2(x)}{\kappa(x)}\right)\right]$$

$$= \exp\left(iux + ium(x)\Delta t - iu\frac{1}{4}\frac{s^2(x)}{\kappa(x)}\right)\mathbb{E}\left[\exp\left(\kappa(x)iu\left(\Delta W_{m+1} + \frac{1}{2}\frac{s(x)}{\kappa(x)}\right)^2\right)\right], \quad (3.16)$$

where $\Delta W_{m+1}+\frac{1}{2}\frac{s(x)}{\kappa(x)} \sim \mathcal{N}\left(\frac{1}{2}\frac{s(x)}{\kappa(x)}, \Delta t\right)$ or equivalent by $\frac{1}{\Delta t}\left(\Delta W_{m+1} + \frac{1}{2}\frac{s(x)}{\kappa(x)}\right)^2 \sim \chi_1'^2\left(\frac{1}{4}\frac{s^2(x)}{\kappa^2(x)\Delta t}\right)$ which denotes the noncentral chi-sqaured distribution with one degree of freedom and noncentrality parameter $\frac{1}{4}\frac{s^2(x)}{\kappa^2(x)\Delta t}$. Hence,

$$\phi_{X^{\Delta}_{m+1}}\left(u\,|X^{\Delta}_m = x\right) = \exp\left(iux + ium(x)\Delta t - iu\frac{1}{4}\frac{s^2(x)}{\kappa(x)}\right)\phi_{\chi_1'^2\left(\frac{1}{4}\frac{s^2(x)}{\kappa^2(x)\Delta t}\right)}(u\kappa(x)\Delta t)$$

$$= \exp\left(iux + ium(x)\Delta t - iu\frac{1}{4}\frac{s^2(x)}{\kappa(x)}\right)\exp\left(\frac{1}{4}\frac{s^2(x)}{\kappa(x)}\frac{iu}{1-2iu\kappa(x)\Delta t}\right)(1-2iu\kappa(x)\Delta t)^{-\frac{1}{2}}$$

$$= \exp\left(iux + ium(x)\Delta t - \frac{\frac{1}{2}u^2 s^2(x)\Delta t}{1-2iu\kappa(x)\Delta t}\right)(1-2iu\kappa(x)\Delta t)^{-\frac{1}{2}}. \quad (3.17)$$

□

3.2 COS method

We wish to derive the value $V(0,T,X_0)$ of an option at time 0 with as underlying X_t and expiration date T. The payoff of the option at time T is given by $V(T,T,X_T) = g(X_T)$ for some function g, and we assume that there are no early-exercise dates. The value of the option $V(0,T,X_0)$ is given by the risk-neutral valuation formula:

$$V(0,T,X_0) = e^{-rT}\mathbb{E}\left[g(X_T)|X_0\right] = e^{-rT}\int_{\mathbb{R}} g(X)Q_{X_T}(T,X|X_0)\,\mathrm{d}X, \quad (3.18)$$

where r is the risk-free interest rate and $Q_{X_T}(T,X|X_0)$ is the probability density of $X_T = X$ given X_0.

The value of the option $V(0,T,X_0)$ can be obtained by the COS method of F. Fang and C.W. Oosterlee [9]. As they explained, the density $Q_{X_T}(T,X|X_0)$ decays to zero as $X \to \pm\infty$. Therefore, we can integrate over a finite interval $[a,b] \subseteq \mathbb{R}$ instead of integrating over the entire space, to approximate the option value

$$V(0,T,X_0) \approx e^{-rT}\int_a^b g(X)Q_{X_T}(T,X|X_0)\,\mathrm{d}X. \quad (3.19)$$

Suppose this choice of $[a,b]$ is not only chosen such that the option value is approximated well, but is also chosen such that the characteristic function of f can be approximated by integrating over $[a,b]$ instead of integrating over the entire space,

$$\phi_{X_T}(w|X_0) = \int_{\mathbb{R}} e^{iwX}Q_{X_T}(T,X|X_0)\,\mathrm{d}X \approx \int_a^b e^{iwX}Q_{X_T}(T,X|X_0)\,\mathrm{d}X. \quad (3.20)$$

Using the Fourier-cosine expansion of the density results in

$$Q_{X_T}(T,X|X_0) = {\sum_{k=0}^{\infty}}' A_k\cos\left(k\pi\frac{X-a}{b-a}\right), \quad (3.21)$$

where $\sum'$ indicates that the first term in the series summation is weighted by one-half and

$$A_k = \frac{2}{b-a}\int_a^b Q_{X_T}(T,X|X_0)\cos\left(k\pi\frac{X-a}{b-a}\right)\,\mathrm{d}X. \tag{3.22}$$

Rewriting A_k and substitution of the approximation of formula (3.20) leads to

$$\begin{aligned}
A_k &= \frac{2}{b-a}\int_a^b Q_{X_T}(T,X|X_0)\cos\left(k\pi\frac{X-a}{b-a}\right)\,\mathrm{d}X \\
&= \frac{2}{b-a}\int_a^b Q_{X_T}(T,X|X_0)\Re\left[\exp\left(ik\pi\frac{X-a}{b-a}\right)\right]\,\mathrm{d}X \\
&= \frac{2}{b-a}\Re\left[\exp\left(-ik\pi\frac{a}{b-a}\right)\int_a^b Q_{X_T}(T,X|X_0)\exp\left(ik\pi\frac{X}{b-a}\right)\,\mathrm{d}X\right] \\
&\approx \frac{2}{b-a}\Re\left[\exp\left(-ik\pi\frac{a}{b-a}\right)\phi_{X_T}\left(\frac{k\pi}{b-a}\middle| X_0\right)\right].
\end{aligned} \tag{3.23}$$

Now, we truncate the series summation by setting N as the number of Fourier coefficients employed, which leads to

$$Q_{X_T}(T,X|X_0) \approx {\sum_{k=0}^{N-1}}' \widetilde{A}_k \cos\left(k\pi\frac{X-a}{b-a}\right), \tag{3.24}$$

where

$$\widetilde{A}_k = \frac{2}{b-a}\Re\left[\exp\left(-ik\pi\frac{a}{b-a}\right)\phi_{X_T}\left(\frac{k\pi}{b-a}\middle| X_0\right)\right]. \tag{3.25}$$

Finally, we can derive the option pricing formula of the COS method by using Fubini's theorem

$$\begin{aligned}
V(0,T,X_0) &\approx e^{-rT}\int_a^b g(X)Q_{X_T}(T,X|X_0)\,\mathrm{d}X \\
&\approx e^{-rT}\int_a^b g(X){\sum_{k=0}^{N-1}}' A_k\cos\left(k\pi\frac{X-a}{b-a}\right)\mathrm{d}X \\
&= e^{-rT}{\sum_{k=0}^{N-1}}' A_k\int_a^b g(X)\cos\left(k\pi\frac{X-a}{b-a}\right)\mathrm{d}X \\
&\approx e^{-rT}{\sum_{k=0}^{N-1}}' \mathcal{V}_k\Re\left[\phi_{X_T}\left(\frac{k\pi}{b-a}\middle| X_0\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right],
\end{aligned} \tag{3.26}$$

where

$$\mathcal{V}_k = \frac{2}{b-a}\int_a^b g(X)\cos\left(k\pi\frac{X-a}{b-a}\right)\mathrm{d}X. \tag{3.27}$$

Often $\mathcal{V}_k$ in (3.27) is governed by an analytical expression.

3.3 BCOS method

Just as in the previous section, we wish to derive the value $V(0,T,X_0)$ of a derivative at time 0 with as underlying X_t and time to maturity T, where the FSDE of X_t is given by (3.1). The payoff of the option at time T is given by $V(T,T,X_T) = g(X_T)$ for some function g. We assume that we work in a complete market and therefore we can make a self-financing portfolio

Y_t consisting of a_t assets and bonds with risk-free return rate r, such that $Y_T = g(X_T)$. For $0 \le t \le T$, we have

$$\mathrm{d}Y_t = r(Y_t - a_t X_t)\,\mathrm{d}t + a_t\,\mathrm{d}X_t = (rY_t + (\mu(X_t) - rX_t)\,a_t)\,\mathrm{d}t + \sigma(X_t)a_t\,\mathrm{d}W_t. \tag{3.28}$$

If we set $Z_t = \sigma(X_t)a_t$, then (Y_t, Z_t) solves the BSDE

$$\mathrm{d}Y_t = -f(t, X_t, Y_t, Z_t)\,\mathrm{d}t + Z_t\,\mathrm{d}W_t, \qquad Y_T = g(X_T), \tag{3.29}$$

$$f(t,x,y,z) = -ry - \frac{\mu(x) - rx}{\sigma(x)}z. \tag{3.30}$$

The functions $\sigma : \mathbb{R} \to \mathbb{R}$ and $\mu : \mathbb{R} \to \mathbb{R}$ in FSDE (3.1) are assumed to be twice differentiable with respect to x, satisfy a Lipschitz condition in x and satisfy a linear growth condition in x. The function $f : [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is assumed to be uniformly continuous with respect to x and satisfies a Lipschitz condition in (y,z), with Lipschitz constant L_f and the function $g : \mathbb{R} \to \mathbb{R}$ is assumed to be uniformly continuous with respect to x. Also, there exists a constant C such that $|f(t,x,y,z)| + |g(x)| \le C(1 + |x|^p + |y| + |z|)$, $p \ge \frac{1}{2}$ holds [2]. Y_t is a self-financing portfolio, and therefore the option value is given by $V(0, T, X_0) = Y_0$.

We make a time grid of $M+1$ time points, where $t_j = j\Delta t$ and $\Delta t = \frac{T}{M}$. Integrating gives

$$Y_0 = g(X_T) + \int_0^T f(t, X_t, Y_t, Z_t)\ \mathrm{d}t - \int_0^T Z_t\ \mathrm{d}W_t. \tag{3.31}$$

At time t_m we observe that

$$Y_m = Y_{m+1} + \int_{t_m}^{t_{m+1}} f(t, X_t, Y_t, Z_t)\ \mathrm{d}t - \int_{t_m}^{t_{m+1}} Z_t\ \mathrm{d}W_t. \tag{3.32}$$

We take conditional expectations at both sides of the equation and apply the θ-method, for some $\theta \in [0,1]$ ($\theta = 1/2$ corresponds to the Trapezium Rule)

$$Y_m = \mathbb{E}_m[Y_{m+1}] + \int_{t_m}^{t_{m+1}} \mathbb{E}_m[f(t, X_t, Y_t, Z_t)]\ \mathrm{d}t \tag{3.33}$$

$$\approx \mathbb{E}_m[Y_{m+1}] + \Delta t\theta f(t_m, X_m, Y_m, Z_m) + \Delta t(1-\theta)\mathbb{E}_m[f(t_{m+1}, X_{m+1}, Y_{m+1}, Z_{m+1})]. \tag{3.34}$$

Analogously, we find by multiplication with ΔW_{m+1} and taking the conditional expectation:

$$\begin{aligned} 0 &= \mathbb{E}_m[Y_{m+1}\Delta W_{m+1}] + \int_{t_m}^{t_{m+1}} \mathbb{E}_m[f(t, X_t, Y_t, Z_t)(W_t - W_{t_m})]\ \mathrm{d}t \\ &- \int_{t_m}^{t_{m+1}} \mathbb{E}_m[Z_t]\ \mathrm{d}t \end{aligned} \tag{3.35}$$

$$\begin{aligned} &\approx \mathbb{E}_m[Y_{m+1}\Delta W_{m+1}] + \Delta t(1-\theta)\mathbb{E}_m[f(t_{m+1}, X_{m+1}, Y_{m+1}, Z_{m+1})\Delta W_{m+1}] - \Delta t\theta Z_m \\ &- \Delta t(1-\theta)\mathbb{E}_m[Z_{m+1}]. \end{aligned} \tag{3.36}$$

Using one of the approximation schemes of Section 3.1, and formulas (3.34) and (3.36) gives, for $m = M-1, \ldots, 0$,

$$Y_m^\Delta = \mathbb{E}_m[Y_{m+1}^\Delta] + \Delta t\theta f\left(t_m, X_m^\Delta, Y_m^\Delta, Z_m^\Delta\right) + \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^\Delta, Y_{m+1}^\Delta, Z_{m+1}^\Delta\right)\right],$$

[2] The conditions on the functions f and g guarantee the existence of a unique solution (Y,Z) to BSDE (3.29) [24]

$$\begin{aligned} Z_m^\Delta &= \frac{1}{\Delta t\theta}\mathbb{E}_m\left[Y_{m+1}^\Delta \Delta W_{m+1}\right] + \frac{1-\theta}{\theta}\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^\Delta, Y_{m+1}^\Delta, Z_{m+1}^\Delta\right)\Delta W_{m+1}\right] \\ &- \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^\Delta\right], \end{aligned} \tag{3.37, 3.38}$$

$$Y_M^\Delta = g\left(X_M^\Delta\right), \quad \text{and} \quad Z_M^\Delta = \sigma(x)\left.\frac{\mathrm{d}g(x)}{\mathrm{d}x}\right|_{x=X_M^\Delta}. \tag{3.39}$$

We observe that Y_m^Δ and Z_m^Δ depend on the value X_m^Δ, so when $X_m^\Delta = x$, then for $m = M-1,\ldots,0$

$$\begin{aligned} Y_m^\Delta(x) &= \mathbb{E}_m\left[Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right] + \Delta t\theta f\left(t_m, x, Y_m^\Delta(x), Z_m^\Delta(x)\right) \\ &+ \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right)\right], \end{aligned} \tag{3.40}$$

$$\begin{aligned} Z_m^\Delta(x) &= \frac{1}{\Delta t\theta}\mathbb{E}_m\left[Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right] \\ &+ \frac{1-\theta}{\theta}\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right)\Delta W_{m+1}\right], \end{aligned} \tag{3.41}$$

where $X_{m+1}^{\Delta,m,x}$ denotes the value of X_{m+1}^Δ given $X_m^\Delta = x$. $Y_m^\Delta(x)$ is implicit for $\theta > 0$ and can be determined by performing P Picard iterations [25], starting with initial guess $\mathbb{E}\left[Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right]$. The next section explains how we can approximate the conditional expectations in formulas (3.40) and (3.41). The value of the option is approximated by $V(0,T,X_0) = Y_0^\Delta(X_0)$.

3.4 Approximation of the conditional expectations

For each $m \in \{M-1,\ldots,0\}$, we wish to approximate the following conditional expectations: $\mathbb{E}_m\left[Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right]$, $\mathbb{E}_m\left[Z_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right]$, $\mathbb{E}_m\left[Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right]$,
$\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right)\right]$ and
$\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right)\Delta W_{m+1}\right]$.
We generalize this by deriving an equation for the conditional expectations $\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\right]$ and $\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right]$ for a general function $h(t,x)$.

Using the COS approximation with the characteristic function of the discrete scheme (3.9) we find

$$\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\right] \approx \sum_{k=1}^{N-1}{}' \mathcal{H}_k(t_{m+1})\Re\left\{\phi_{X_{m+1}^\Delta}\left(\frac{k\pi}{b-a}\middle| X_m^\Delta = x\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}, \tag{3.42}$$

and

$$\begin{aligned} &\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right] \\ &\approx \sum_{k=1}^{N-1}{}' \mathcal{H}_k(t_{m+1})\Re\left\{\mathbb{E}\left[\exp\left(ik\pi\frac{X_{m+1}^{\Delta,m,x}}{b-a}\right)\Delta W_{m+1}\right]\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}. \end{aligned} \tag{3.43}$$

Using the fact that $\Delta W_{m+1} \sim \mathcal{N}(0,\Delta t)$ and integration by parts gives

$$\mathbb{E}\left[\exp\left(iuX_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\exp\left(iux + ium(x)\Delta t + ius(x)\Delta W_{m+1} + iu\kappa(x)\left(\Delta W_{m+1}\right)^2\right)\Delta W_{m+1}\right] \\
&= \frac{1}{\sqrt{2\pi\Delta t}}\int_{-\infty}^{\infty} \exp\left(iux + ium(x)\Delta t + ius(x)y + iu\kappa(x)y^2\right) \mathrm{d}\left(-\Delta t e^{-\frac{y^2}{2\Delta t}}\right) \\
&= -\sqrt{\frac{\Delta t}{2\pi}} \lim_{R\to\infty}\left[\exp\left(iux + ium(x)\Delta t + ius(x)y + iu\kappa(x)y^2 - \frac{y^2}{2\Delta t}\right)\right]_{-R}^{R} \\
&+ \frac{1}{\sqrt{2\pi\Delta t}}\int_{-\infty}^{\infty} \Delta t e^{-\frac{y^2}{2\Delta t}} \mathrm{d}\left(\exp\left(iux + ium(x)\Delta t + ius(x)y + iu\kappa(x)y^2\right)\right) \\
&= 0 + \sqrt{\frac{\Delta t}{2\pi}}\int_{-\infty}^{\infty} \left(ius(x) + 2iu\kappa(x)y\right)\exp\left(iux + ium(x)\Delta t + ius(x)y + iu\kappa(x)y^2\right) e^{-\frac{y^2}{2\Delta t}}\ \mathrm{d}y \\
&= ius(x)\Delta t\mathbb{E}\left[\exp\left(iuX_{m+1}^{\Delta,m,x}\right)\right] + 2iu\kappa(x)\Delta t\mathbb{E}\left[\exp\left(iuX_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right]. \qquad (3.44)
\end{aligned}
$$

The last term can be evaluated in an analogously way, and so on, which results in

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(iuX_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right] &= ius(x)\Delta t\sum_{n=0}^{\infty}\left(2\kappa(x)\Delta t\right)^n (iu)^n\, \mathbb{E}\left[\exp\left(iuX_{m+1}^{\Delta,m,x}\right)\right] \\
&= ius(x)\Delta t\sum_{n=0}^{\infty}\left(2\kappa(x)\Delta t\right)^n (iu)^n\, \phi_{X_{m+1}^{\Delta}}\left(u\,\middle|X_m^{\Delta} = x\right). \qquad (3.45)
\end{aligned}
$$

We take the first two terms and leave only an $\mathcal{O}\left((\Delta t)^3\right)$ error, which gives

$$
\begin{aligned}
\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right] &\approx \sum_{k=1}^{N-1}{}' \mathcal{H}_k(t_{m+1})\Re\left\{\left[i\frac{k\pi}{b-a}s(x)\Delta t\phi_{X_{m+1}^{\Delta}}\left(\frac{k\pi}{b-a}\middle| X_m^{\Delta} = x\right)\right.\right. \\
&- \left.\left. 2s(x)\kappa(x)\left(\Delta t\right)^2\left(\frac{k\pi}{b-a}\right)^2\phi_{X_{m+1}^{\Delta}}\left(\frac{k\pi}{b-a}\middle| X_m^{\Delta} = x\right)\right]\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}, \qquad (3.46)
\end{aligned}
$$

where

$$
\mathcal{H}_k(t_{m+1}) = \frac{2}{b-a}\int_a^b h(t_{m+1}, x)\cos\left(k\pi\frac{x-a}{b-a}\right)\mathrm{d}x. \qquad (3.47)
$$

Let $\mathcal{Y}_k^{\Delta}(t_m)$, $\mathcal{Z}_k^{\Delta}(t_m)$ and $\mathcal{F}_k^{\Delta}(t_m)$ denote the Fourier-cosine coefficients of, respectively, $Y_m^{\Delta}(x)$, $Z_m^{\Delta}(x)$ and $f\left(t_m, x, Y_m^{\Delta}(x), Z_m^{\Delta}(x)\right)$, i.e.

$$
\begin{aligned}
\mathcal{Y}_k^{\Delta}(t_m) &= \frac{2}{b-a}\int_a^b Y_m^{\Delta}(x)\cos\left(k\pi\frac{x-a}{b-a}\right)\ \mathrm{d}x, \qquad (3.48) \\
\mathcal{Z}_k^{\Delta}(t_m) &= \frac{2}{b-a}\int_a^b Z_m^{\Delta}(x)\cos\left(k\pi\frac{x-a}{b-a}\right)\ \mathrm{d}x, \qquad (3.49) \\
\mathcal{F}_k^{\Delta}(t_m) &= \frac{2}{b-a}\int_a^b f\left(t_m, x, Y_m^{\Delta}(x), Z_m^{\Delta}(x)\right)\cos\left(k\pi\frac{x-a}{b-a}\right)\ \mathrm{d}x, \qquad (3.50)
\end{aligned}
$$

and at time of maturity T

$$
\begin{aligned}
\mathcal{Y}_k^{\Delta}(t_M) &= \frac{2}{b-a}\int_a^b g(x)\cos\left(k\pi\frac{x-a}{b-a}\right)\ \mathrm{d}x, \qquad (3.51) \\
\mathcal{Z}_k^{\Delta}(t_M) &= \frac{2}{b-a}\int_a^b \sigma(x)\frac{\mathrm{d}g(x)}{\mathrm{d}x}\cos\left(k\pi\frac{x-a}{b-a}\right)\ \mathrm{d}x, \qquad (3.52)
\end{aligned}
$$

$$\mathcal{F}_k^{\Delta}(t_M) \;=\; \frac{2}{b-a}\int_a^b f\left(t_m, x, g(x), \sigma(x)\frac{\mathrm{d}g(x)}{\mathrm{d}x}\right)\cos\left(k\pi\frac{x-a}{b-a}\right)\,\mathrm{d}x. \tag{3.53}$$

If the above integrals cannot be computed analytically, we can approximate them by computing the function on an x-grid and using the discrete Fourier-cosine transform or another numerical integration method.

3.5 BCOS method summarized

We define a time-grid $t_j = j\Delta t$ for $j = 0, 1, \ldots M$, with fixed time steps $\Delta t = \frac{T}{M}$. The algorithm reads:

Initial step: Compute the terminal coefficients $\mathcal{Y}_k^{\Delta}(t_M)$, $\mathcal{Z}_k^{\Delta}(t_M)$ and $\mathcal{F}_k^{\Delta}(t_M)$ with formulas (3.51), (3.52) and (3.53).

Loop: For $m = M-1$ to $m = 1$
approximate the necessary conditional expectations with Section 3.4 and compute functions $Y_m^{\Delta}(x)$, $Z_m^{\Delta}(x)$ and $f\left(t_m, x, Y_m^{\Delta}(x), Z_m^{\Delta}(x)\right)$ for $x \in [a,b]$ with formulas (3.40) and (3.41). Thereafter, compute/approximate the corresponding Fourier-cosine coeffcients $\mathcal{Y}_k^{\Delta}(t_m)$, $\mathcal{Z}_k^{\Delta}(t_m)$ and $\mathcal{F}_k^{\Delta}(t_m)$ with formulas (3.48), (3.49) and (3.50). Those integrals can be approximated by computing the function on an x-grid and using the discrete Fourier-cosine transform or another numerical integration method.

Terminal step: Compute $Y_0^{\Delta}(X_0)$ and $Z_0^{\Delta}(X_0)$.

We approximate the value $V(0,T,X_0)$ of the option by $V(0,T,X_0) = Y_0^{\Delta}(X_0)$ and the corresponding Delta by

$$\Delta_0 = \left.\frac{\partial V(0,T,x)}{\partial x}\right|_{x=X_0} = \frac{Z_0^{\Delta}(X_0)}{\sigma(X_0)}, \tag{3.54}$$

which corresponds to the amount of assets in the self-financing portfolio Y_t at time $t = 0$.

Remark 3.1. *If an option is governed by early-exercise dates τ_j for $j = 1, 2, \ldots, n$, then we choose Δt such that each of the early-exercise dates corresponds to a point in our time-grid. We replace formula (3.40) by*

$$Y_m^{\Delta}(x) = \begin{cases} \max\Big\{g(x), \mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\right] + \Delta t\theta f\left(t_m, x, Y_m^{\Delta}(x), Z_m^{\Delta}(x)\right) \\ \quad + \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\right)\right]\Big\}, & \textit{for } t_m = \tau_j, \\ \\ \mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\right] + \Delta t\theta f\left(t_m, x, Y_m^{\Delta}(x), Z_m^{\Delta}(x)\right) \\ \quad + \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\right)\right], & \textit{for } t_m \neq \tau_j, \end{cases}$$

and formula (3.41) by

$$Z_m^{\Delta}(x) = \begin{cases} \sigma(x)\frac{\mathrm{d}g(x)}{\mathrm{d}x}, & \textit{for } t_m = \tau_j,\ Y_m^{\Delta}(x) = g(x), \\ \\ \frac{1}{\Delta t\theta}\mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\Delta W_{m+1}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\right] \\ \quad + \frac{1-\theta}{\theta}\mathbb{E}_m\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right), Z_{m+1}^{\Delta}\left(X_{m+1}^{\Delta,m,x}\right)\right)\Delta W_{m+1}\right], & \textit{otherwise,} \end{cases}$$

for $m = M-1, ..., 0$. In Section 3.7 we give an example of approximating the value of a Bermudan option by using the BCOS method and in Section 4.8 we explain more about path-dependent options, such as Bermudan options.

3.6 Error analysis

In this section we perform an error analysis for the BCOS method under the risk-neutral $\mathbb{Q}$-measure, because we will not use the physical $\mathbb{P}$-measure in the development of the method to price European options under the SABR model in Chapters 4-6. This means basically that we only analyze the error in the option price Y_m and not the error in the Delta Z_m. Therefore, we abbreviate $f(t,x,y,z)$ to $f(t,x,y)$. Nevertheless, we give the formulas for pricing under the $\mathbb{P}$-measure in this chapter and in Chapter 4, to make these formulas available for general use[3].

The error in option value Y_m consists of four different components, namely the error as a result of the Δ-time-discretization, the θ-method-discretization, the COS method, and the Picard iterations. Just like Ruijter and Oosterlee [24], we perform the error analysis for the 2.0-weak-Taylor scheme and parameter $\theta = \frac{1}{2}$. In Sections 3.6.1 and 3.6.2, we discuss the local errors as a result of the Δ-time-discretization and the θ-method-discretization, respectively. In Section 3.6.3 we give a short overview of other errors related to the BCOS method. Finally, we consider the global error in Section 3.6.4. For the error analysis we use the notation and results of Appendix C.

3.6.1 Local error Δ-time-discretization

Let $X_{m+1}^{m,x}$ and $X_{m+1}^{\Delta,m,x}$ denote, respectively, the values of X_{m+1} and X_{m+1}^{Δ} given $X_m = x$, where $X_{m+1}^{\Delta,m,x}$ is derived by the 2.0-weak-Taylor scheme.

Lemma 3.2. *For a sufficiently smooth function $h(t,x)$ we have the following local weak error*

$$\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{m,x}\right) - h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\right] = \mathcal{O}\left((\Delta t)^3\right). \tag{3.55}$$

Proof. The 2.0-weak-Taylor scheme corresponds to

$$\begin{aligned} X_{m+1}^{\Delta,m,x} &= \sum_{\widehat{\alpha}\in\mathcal{A}} c_{\widehat{\alpha}}(x) I_{\widehat{\alpha},t_m,t_{m+1}} + c_{(0,1)}(x)\left(\frac{1}{2}\Delta W_{m+1}\Delta t - I_{(0,1),t_m,t_{m+1}}\right) \\ &+ c_{(1,0)}(x)\left(\frac{1}{2}\Delta W_{m+1}\Delta t - I_{(1,0),t_m,t_{m+1}}\right), \end{aligned} \tag{3.56}$$

where $c(x) = x$ and $\mathcal{A} = \{v, (0), (1), (0,0), (0,1), (1,0), (1,1)\}$. By Lemma C.2, we have for all $l \in \mathbb{N}$

$$\mathbb{E}\left[\left(X_{m+1}^{m,x} - X_{m+1}^{\Delta,m,x}\right)^l\right] = \mathcal{O}\left((\Delta t)^3\right). \tag{3.57}$$

A Taylor series expansion of function h around $X_{m+1}^{\Delta,m,x} = X_{m+1}^{m,x}$ gives

$$h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!}\left(X_{m+1}^{m,x} - X_{m+1}^{\Delta,m,x}\right)^l \frac{\partial^l}{\left(\partial X_{m+1}^{m,x}\right)^l} h\left(t_{m+1}, X_{m+1}^{m,x}\right). \tag{3.58}$$

[3] In risk management the possibility of valuing under the $\mathbb{P}$-measure is important, e.g. for value at risk (VaR) and for credit valuation adjustment (CVA).

Combining formulas (3.57) and (3.58) implies

$$\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{m,x}\right) - h\left(t_{m+1}, X_{m+1}^{\Delta,m,x}\right)\right] \quad = \quad \mathcal{O}\left((\Delta t)^3\right). \tag{3.59}$$

□

3.6.2 Local error θ-method-discretization

We observe from formula (3.33) that given $X_m = x$,

$$\begin{aligned} Y_m(x) &= \mathbb{E}_m[Y_{m+1}(X_{m+1})] + \int_{t_m}^{t_{m+1}} \mathbb{E}_m[f(t, X_t, Y_t(X_t))]\ \mathrm{d}t \\ &= \mathbb{E}_m[Y_{m+1}(X_{m+1})] + \frac{1}{2}\Delta t\left(f(t_m, x, Y_m(x)) + \mathbb{E}_m[f(t_{m+1}, X_{m+1}, Y_{m+1}(X_{m+1}))]\right) \\ &+ \mathcal{R}_m^y(x), \end{aligned} \tag{3.60}$$

which gives for the θ-method-discretization error

$$\begin{aligned} \mathcal{R}_m^y(x) &= \int_{t_m}^{t_{m+1}} \mathbb{E}_m[f(t, X_t, Y_t(X_t))]\ \mathrm{d}t \\ &- \frac{1}{2}\Delta t\left(f(t_m, x, Y_m(x)) + \mathbb{E}_m[f(t_{m+1}, X_{m+1}, Y_{m+1}(X_{m+1}))]\right). \end{aligned} \tag{3.61}$$

Lemma 3.3. *For a sufficiently smooth function $f(t,x,y)$ and payoff function $g(x)$ we have the following θ-method-discretization error*

$$\mathcal{R}_m^y(x) = \mathcal{O}\left((\Delta t)^3\right). \tag{3.62}$$

Proof. For a generally smooth function $h(t,x)$, we find with Theorem C.1, $\mathcal{A} = \{v, (0), (1)\}$ and $\overline{\mathcal{A}} = \{(0,0,0), (1,0,0), (0,1,0), (1,1,0)\}$

$$\begin{aligned} \int_{t_m}^{t_{m+1}} h\left(t, X_t^{m,x}\right) \mathrm{d}t &= \int_{t_m}^{t_{m+1}} h(t_m, x) + h_{(0)}(t_m, x) I_{(0),t_m,t} + h_{(1)}(t_m, x) I_{(1),t_m,t} \\ &+ \sum_{\widehat{\alpha} \in \mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_.^{m,x}\right)\right]_{t_m,t}\ \mathrm{d}t \\ &= h(t_m, x) I_{(0),t_m,t_{m+1}} + h_{(0)}(t_m, x) I_{(0,0),t_m,t_{m+1}} + h_{(1)}(t_m, x) I_{(1,0),t_m,t_{m+1}} \\ &+ \sum_{\widehat{\alpha} \in \overline{\mathcal{A}}} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}-}\left(., X_.^{m,x}\right)\right]_{t_m,t_{m+1}}, \end{aligned} \tag{3.63}$$

which implies

$$\begin{aligned} &\mathbb{E}\left[\int_{t_m}^{t_{m+1}} h\left(t, X_t^{m,x}\right) \mathrm{d}t - \frac{1}{2}\Delta t\left(h(t_m, x) + h\left(t_{m+1}, X_{m+1}^{m,x}\right)\right)\right] \\ &= h(t_m, x)\Delta t + \frac{1}{2} h_{(0)}(t_m, x)(\Delta t)^2 + \mathcal{O}\left((\Delta t)^3\right) \\ &-\frac{1}{2}\Delta t\left(h(t_m, x) + h(t_m, x) + h_{(0)}(t_m, x)\Delta t + \frac{1}{2} h_{(0,0)}(t_m, x)(\Delta t)^2 + \mathcal{O}\left((\Delta t)^3\right)\right) \\ &= \mathcal{O}\left((\Delta t)^3\right). \end{aligned} \tag{3.64}$$

□

3.6.3 Fourier errors

The COS method introduces an error which consists of four parts

- The integration range truncation error ε_1, introduced in formula (3.19)

$$\varepsilon_1 = e^{-rT} \int_{\mathbb{R}\setminus[a,b]} g(X) Q_{X_T}(T, X|X_0)\, \mathrm{d}X. \tag{3.65}$$

- The series truncation error ε_2 introduced in formula (3.24), i.e.

$$\varepsilon_2 = e^{-rT} \sum_{k=N}^{\infty} A_k \int_a^b g(X) \cos\left(k\pi \frac{X-a}{b-a}\right) \mathrm{d}X. \tag{3.66}$$

- The error ε_3 related to the approximation of A_k in formula (3.23):

$$\varepsilon_3 = e^{-rT} {\sum_{k=0}^{N-1}}' \int_{\mathbb{R}\setminus[a,b]} Q_{X_T}(T, X|X_0) \cos\left(k\pi \frac{X-a}{b-a}\right) \mathrm{d}X \ \mathcal{V}_k. \tag{3.67}$$

- The error ε_4 related to the use of discrete cosine transform for formulas (3.48) and (3.50), when those integrals cannot be computed analytically:

$$\varepsilon_4 = e^{-rT} {\sum_{k=0}^{N-1}}' \Re\left[\phi_{X_T}\left(\frac{k\pi}{b-a}\middle| X_0\right) \exp\left(-ik\pi\frac{a}{b-a}\right)\right] \left[\mathcal{V}_k - \frac{2}{N_1}\sum_{n=0}^{N_1-1} g(\widetilde{x}_n) \cos\left(\frac{k\pi}{N_1}(n+1/2)\right)\right], \tag{3.68}$$

where $\widetilde{x}_n = a + (n+1/2)\frac{b-a}{N_1}$. In this paper we choose to set N_1 equal to the number of Fourier cosine coefficients N.

Fang and Oosterlee discussed the errors ε_1, ε_2 and ε_3 in Section 4 of their paper [9] and Atkinson discussed the error ε_4 in Chapter 5 of his book [4].

Also the Picard iterations introduce an error. The authors in [25] briefly discussed this error in Section 4.5 of their paper and they obtained that the iterations converge when Δt is small.

3.6.4 Global error

In this section we give the global error of the BCOS method, where we omit the errors introduced by the COS method and by the Picard iterations, as they can be made arbitrarily small depending on parameters. Just as the authors in [24], we define

$$e_m^y\left(X_m, X_m^\Delta\right) := Y_m(X_m) - Y_m^\Delta\left(X_m^\Delta\right), \tag{3.69}$$
$$e_m^f\left(X_m, X_m^\Delta\right) := f(t_m, X_m, Y_m(X_m)) - f\left(t_m, X_m^\Delta, Y_m^\Delta\left(X_m^\Delta\right)\right), \tag{3.70}$$

and

$$e_m^y(x) := Y_m(x) - Y_m^\Delta(x), \tag{3.71}$$
$$e_m^f(x) := f(t_m, x, Y_m(x)) - f\left(t_m, x, Y_m^\Delta(x)\right). \tag{3.72}$$

We can rewrite the error $e_m^y\left(X_m, X_m^\Delta\right)$ to

$$e_m^y\left(X_m, X_m^\Delta\right) = Y_m(X_m) - Y_m^\Delta(X_m) + Y_m^\Delta(X_m) - Y_m^\Delta\left(X_m^\Delta\right)$$

$$= e_m^y(X_m) + Y_m^\Delta(X_m) - Y_m^\Delta\left(X_m^\Delta\right). \tag{3.73}$$

The triangle inequality implies

$$\begin{aligned}\left|\mathbb{E}_0\left[e_m^y\left(X_m, X_m^\Delta\right)\right]\right| &\leq \left|\mathbb{E}_0\left[e_m^y(X_m)\right]\right| + \left|\mathbb{E}_0\left[Y_m^\Delta(X_m) - Y_m^\Delta\left(X_m^\Delta\right)\right]\right| \\ &\leq \mathbb{E}_0\left[\left|e_m^y(X_m)\right|\right] + \left|\mathbb{E}_0\left[Y_m^\Delta(X_m) - Y_m^\Delta\left(X_m^\Delta\right)\right]\right|. \end{aligned} \tag{3.74}$$

If $Y_m(x)$ is a sufficiently smooth function, we find with Lemma 3.2 that

$$\left|\mathbb{E}_0\left[e_m^y\left(X_m, X_m^\Delta\right)\right]\right| \leq \mathbb{E}_0\left[\left|e_m^y(X_m)\right|\right] + \mathcal{O}\left((\Delta t)^3\right). \tag{3.75}$$

The following lemma gives a bound on the error $e_m^y(X_m)$,

Lemma 3.4.

$$\mathbb{E}_0\left[\left|e_m^y(X_m)\right|\right] = \mathcal{O}\left((\Delta t)^2\right). \tag{3.76}$$

Proof. Formulas (3.40) and (3.60) give

$$\begin{aligned} e_m^y(x) &= Y_m(x) - Y_m^\Delta(x) \\ &= \mathbb{E}\left[Y_m\left(X_{m+1}^{m,x}\right)\right] + \frac{1}{2}\Delta t\left(f(t_m, x, Y_m(x)) + \mathbb{E}\left[f\left(t_{m+1}, X_{m+1}^{m,x}, Y_{m+1}\left(X_{m+1}^{m,x}\right)\right)\right]\right) + \mathcal{R}_m^y(x) \\ &- \mathbb{E}\left[Y_m^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right] - \frac{1}{2}\Delta t\left(f(t_m, x, Y_m^\Delta(x)) + \mathbb{E}\left[f\left(t_{m+1}, X_{m+1}^{\Delta,m,x}, Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right)\right]\right) \\ &= \mathbb{E}\left[e_{m+1}^y\left(X_{m+1}^{m,x}, X_{m+1}^{\Delta,m,x}\right)\right] + \frac{1}{2}\Delta t\left(e_m^f(x) + \mathbb{E}\left[e_{m+1}^f\left(X_{m+1}^{m,x}, X_{m+1}^{\Delta,m,x}\right)\right]\right) + \mathcal{R}_m^y(x). \end{aligned} \tag{3.77}$$

With Lemma 3.2, we find

$$\begin{aligned} \mathbb{E}\left[e_{m+1}^y\left(X_{m+1}^{m,x}, X_{m+1}^{\Delta,m,x}\right)\right] &= \mathbb{E}\left[Y_{m+1}\left(X_{m+1}^{m,x}\right) - Y_{m+1}^\Delta\left(X_{m+1}^{m,x}\right)\right] \\ &+ \mathbb{E}\left[Y_{m+1}^\Delta\left(X_{m+1}^{m,x}\right) - Y_{m+1}^\Delta\left(X_{m+1}^{\Delta,m,x}\right)\right] \\ &= \mathbb{E}\left[e_{m+1}^y\left(X_{m+1}^{m,x}\right)\right] + \mathcal{O}\left((\Delta t)^3\right) \end{aligned} \tag{3.78}$$

and, analogously,

$$\mathbb{E}\left[e_{m+1}^f\left(X_{m+1}^{m,x}, X_{m+1}^{\Delta,m,x}\right)\right] = \mathbb{E}\left[e_{m+1}^f\left(X_{m+1}^{m,x}\right)\right] + \mathcal{O}\left((\Delta t)^3\right). \tag{3.79}$$

The function $f(t,x,y)$ is Lipschitz in y with Lipschitz constant L_f, so

$$\left|e_m^f(x)\right| = \left|f(t_m, x, Y_m(x)) - f\left(t_m, x, Y_m^\Delta(x)\right)\right| \leq L_f\left|Y_m(x) - Y_m^\Delta(x)\right| = L_f\left|e_m^y(x)\right|, \tag{3.80}$$

and

$$\mathbb{E}\left[\left|e_{m+1}^f\left(X_{m+1}^{m,x}\right)\right|\right] \leq L_f\mathbb{E}\left[\left|e_{m+1}^y\left(X_{m+1}^{m,x}\right)\right|\right]. \tag{3.81}$$

Then combining all the results and Lemma 3.3 gives us the following result:

$$\begin{aligned} \left|e_m^y(x)\right| &\leq \mathbb{E}\left[\left|e_{m+1}^y\left(X_{m+1}^{m,x}, X_{m+1}^{\Delta,m,x}\right)\right|\right] + \frac{1}{2}\Delta t\left(\left|e_m^f(x)\right| + \mathbb{E}\left[\left|e_{m+1}^f\left(X_{m+1}^{m,x}, X_{m+1}^{\Delta,m,x}\right)\right|\right]\right) \\ &+ \left|\mathcal{R}_m^y(x)\right| \\ &\leq \mathbb{E}\left[\left|e_{m+1}^y\left(X_{m+1}^{m,x}\right)\right|\right] + \frac{1}{2}\Delta t\left(\left|e_m^f(x)\right| + \mathbb{E}\left[\left|e_{m+1}^f\left(X_{m+1}^{m,x}\right)\right|\right]\right) + \mathcal{O}\left((\Delta t)^3\right) \end{aligned}$$

$$\leq \mathbb{E}\left[\left|e^y_{m+1}\left(X^{m,x}_{m+1}\right)\right|\right] + \frac{1}{2}\Delta t L_f\left(\left|e^y_m(x)\right| + \mathbb{E}\left[\left|e^y_{m+1}\left(X^{m,x}_{m+1}\right)\right|\right]\right) + \mathcal{O}\left((\Delta t)^3\right), \quad (3.82)$$

which implies, for $\Delta t \leq 2/L_f$,

$$\left|e^y_m(x)\right| \leq \frac{1+\frac{1}{2}\Delta t L_f}{1-\frac{1}{2}\Delta t L_f}\mathbb{E}\left[\left|e^y_{m+1}\left(X^{m,x}_{m+1}\right)\right|\right] + \mathcal{O}\left((\Delta t)^2\right). \quad (3.83)$$

Since $\frac{1+\frac{1}{2}\Delta t L_f}{1-\frac{1}{2}\Delta t L_f} = 1 + \mathcal{O}(\Delta t)$ and $\mathbb{E}\left[\left|e^y_M\left(X^{m,x}_M\right)\right|\right] = 0$, we can observe, by iterating formula (3.83), that

$$\left|e^y_m(x)\right| = \mathcal{O}\left((\Delta t)^2\right). \quad (3.84)$$

□

Substitution of Lemma 3.4 into formula (3.75) gives us

$$\left|\mathbb{E}_0\left[e^y_m\left(X_m, X^\Delta_m\right)\right]\right| = \mathcal{O}\left((\Delta t)^2\right). \quad (3.85)$$

3.7 Examples

In this section we give two examples, one in which the value of a European call option under the CEV model is derived and one where the value of a Bermudan put option under the Black-Scholes model is computed. In both examples we price under the $\mathbb{Q}$-measure, which makes the BCOS method easier, because then $\mu(X_t) = rX_t$, where r is the risk-free interest rate. Like Ruijter and Oosterlee [24], we set $N = 2^9$ and the domain $[a,b]$ is determined as

$$[a,b] = \left[X_0 + c_1 - L\sqrt{c_2}, X_0 + c_1 + L\sqrt{c_2}\right], \quad (3.86)$$

where $c_1 = \mu(X_0)T$, $c_2 = \sigma^2(X_0)T$ and $L = 10$.

Remark 3.2. *If X_T is constrained, then one can adjust the interval $[a,b]$ to these constraints, i.e. if $X_T \geq 0$ then boundary a can be set equal to $\max\{0, X_0 + c_1 - L\sqrt{c_2}\}$. Because of an absorption condition, we have to add an atom at 0 on the conditional probability density of X_T given X_0. In this thesis, we choose to omit constraints that depend on the underlying. In the current market some underlyings, i.e. interest rates, can even become negative, while some underlyings, i.e. stock prices, can not be negative.*

Example 1 We derive the value of a European call option, where the underlying under the risk-neutral $\mathbb{Q}$-measure follows a CEV process, i.e.

$$\mathrm{d}X_t = rX_t\,\mathrm{d}t + \sigma\left(X_t\right)^\gamma \mathrm{d}W_t, \qquad X_0 = x, \quad t \geq 0, \quad (3.87)$$

where the risk-free interest rate r and the volatility $\sigma \geq 0$ are constants. The Euler approximation for FSDE (3.87) gives

$$X^\Delta_{m+1} = X^\Delta_m + rX^\Delta_m\Delta t + \sigma\left(X^\Delta_m\right)^\gamma \Delta W_{m+1}, \qquad X^\Delta_0 = x, \quad (3.88)$$

and Lemma 3.1 states that the characteristic function of X^Δ_{m+1}, given $X^\Delta_m = x_m$, is given by

$$\phi_{X^\Delta_{m+1}}\left(u\left|X^\Delta_m = x_m\right.\right) = \exp\left(iux_m + iurx_m\Delta t - \frac{1}{2}u^2\sigma^2\left(x_m\right)^{2\gamma}\Delta t\right). \quad (3.89)$$

Here, we choose the following parameter values: $x = 100,\ K = 100,\ T = 1,\ \sigma = 0.2,\ r = 0.1$ and $\gamma = 0.5$.
We obtain

$$\mathrm{d}Y_t = -rY_t\,\mathrm{d}t + Z_t\,\mathrm{d}W_t, \qquad Y_T = \max(K - X_T). \tag{3.90}$$

Notice that $f(t, x, y, z) = -ry$, is independent of z, and as a result we do not have to calculate the functions $Z^{\Delta}_m(x)$ and therefore we are allowed to use $\theta = 0$.

We use $\theta = 0$, $N = 2^9$ and $[a, b] = [90, 130]$. Also, we approximate each Fourier-cosine coefficient by the discrete Fourier-cosine transform. $V(0, T, x) = 9.5162582$ is the reference value of the call option from [17]. Table 3.1 shows the results of the BCOS method and we observe linear convergence. We can see that with 100 time steps the absolute error is less than one basis point.

Time steps M	1	10	100	1000	10.000	100.000
abs error	0.4679	4.9657e-02	4.9964e-03	4.9986e-04	4.9894e-05	4.8942e-06

Table 3.1: BCOS method where the FSDE is discretized with the Euler scheme.

Example 2 We derive the value of a Bermudan put option, where the underlying under the risk-neutral $\mathbb{Q}$-measure follows the Black-Scholes process

$$\mathrm{d}X_t = rX_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \qquad X_0 = x, \quad t \geq 0, \tag{3.91}$$

where the risk-free interest rate r and the volatility $\sigma \geq 0$ are constants. The Euler approximation for FSDE (3.91) gives us:

$$X^{\Delta}_{m+1} = X^{\Delta}_m + rX^{\Delta}_m\Delta t + \sigma X^{\Delta}_m \Delta W_{m+1}, \qquad X^{\Delta}_0 = x, \tag{3.92}$$

and Lemma 3.1 gives the characteristic function of X^{Δ}_{m+1}, given $X^{\Delta}_m = x_m$, is

$$\phi_{X^{\Delta}_{m+1}}\left(u \,\middle|\, X^{\Delta}_m = x_m\right) = \exp\left(iux_m + iurx_m\Delta t - \frac{1}{2}u^2\sigma^2\left(x_m\right)^2\Delta t\right). \tag{3.93}$$

We use the following parameter values: $x = 100,\ K = 110,\ T = 1,\ \sigma = 0.2$ and $r = 0.1$.
The holder of a Bermudan option has the right to exercise the option at a set of predetermined dates. We assume that the option has 10 exercise dates, $t_j = \frac{j}{10}$ for $j = 1, \ldots, 10$. The number of time steps M should be equal to a multiple of the number of exercise dates, so that each exercise date corresponds to a point in our time grid. We reproduced this example from Section 4.1 in [10]. The authors in [10] gave as reference value $V(0, T, x) = 10.479520123$. Table 3.2 shows the results of the BCOS method, where we approximate each Fourier-cosine coefficient and we use $\theta = 0$, $N = 2^9$, and $[a, b] = [-90, 310]$. We observe that the absolute error goes to zero for $M \to \infty$. Just as in the previous example, the BCOS method with the Euler scheme gives highly satisfactory results for $M = 100$. The 2.0-weak-Taylor scheme can also be used and fewer time steps are then necessary to reach the same accuracy.

Number of time steps M	10	50	100	500	1000
abs error	1.5558e-02	2.3886-03	1.0696e-03	6.5052 e-05	5.7383e-05

Table 3.2: BCOS method where the FSDE is discretized with the Euler scheme.

Chapter 4

Extension of the BCOS method to two dimensions

In this chapter, we expand the BCOS method of Ruijter and Oosterlee [24] from one dimension to two dimensions. This is necessary for application of the BCOS method to the SABR model. We assume that the underlying system of FSDEs, for $t \geq 0$, can be written as

$$\mathrm{d}X_t^1 = \mu_1(\mathbf{X}_t)\,\mathrm{d}t + \sigma_1(\mathbf{X}_t)\,\mathrm{d}W_t^1, \qquad X_0^1 = x_1, \tag{4.1}$$

$$\mathrm{d}X_t^2 = \mu_2(\mathbf{X}_t)\,\mathrm{d}t + \rho\sigma_2(\mathbf{X}_t)\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\sigma_2(\mathbf{X}_t)\,\mathrm{d}W_t^2, \qquad X_0^2 = x_2, \tag{4.2}$$

where W^1 and W^2 are uncorrelated standard Brownian motions and $\mathbf{X}_t = \left(X_t^1, X_t^2\right)$.

In this chapter we explain the two-dimensional BCOS method where the bivariate characteristic function is approximated by the bivariate characteristic function of the discrete forward process. In Section 4.1, the discretization of the forward process $\mathbf{X}_t$ by different Taylor schemes and the corresponding bivariate characteristic function is given. We propose an adjusted-Predictor-Corrector scheme in Section 4.2. In Section 4.3 we determine the accuracy of the bivariate characteristic function of the discretized Heston model. This is helpful for understanding the errors made with the two-dimensional BCOS method. In Section 4.4, an introduction in the two-dimensional COS method is given. In Section 4.5 we give the two-dimensional BCOS method to price derivatives for which the underlying follows FSDEs (4.1) and (4.2). We derive formulas to approximate some conditional expectations in Section 4.6. In Section 4.7 we describe how to apply the BCOS method for an option without path-dependency. In Section 4.8 we discuss the change in procedure of the BCOS method for some path-dependent options. An error analysis is given in Section 4.9. Finally, we give examples of pricing with the two-dimensional BCOS method in Section 4.10.

4.1 The discrete forward process and its characteristic function

We again define a time-grid $t_j = \Delta t$ for $j = 0, 1, \ldots, M$, with fixed time steps $\Delta t = \frac{T}{M}$. For $j = 1, 2$, we write $\mathbf{X}_m = \mathbf{X}_{t_m}$, $X_m^j = X_{t_m}^j$, $W_m^j = W_{t_m}^j$ and $\Delta W_{m+1}^j = W_{m+1}^j - W_m^j$. The discrete forward process is denoted by $\mathbf{X}_m^\Delta = \mathbf{X}_{t_m}^\Delta$, $X_m^{j,\Delta} = X_{t_m}^{j,\Delta}$, where $\mathbf{X}_0^\Delta = \left(X_0^{1,\Delta}, X_0^{2,\Delta}\right) = \left(X_0^1, X_0^2\right)$. To determine the values of $X_{m+1}^{j,\Delta}$, for $m = 0, \ldots, M-1$, we use one of the following Taylor schemes: the Euler, Milstein, or 2.0 weak Taylor schemes.

As in the one-dimensional case, we can write the discretization schemes in general form, as follows

$$\begin{aligned}
X^{1,\Delta}_{m+1} &= x_1 + m_1(\mathbf{x})\Delta t + s_1^{W^1}(\mathbf{x})\Delta W^1_{m+1} + s_1^{W^2}(\mathbf{x})\Delta W^2_{m+1} + \kappa_1^{W^1,W^2}(\mathbf{x})\Delta W^1_{m+1}\Delta W^2_{m+1} \\
&+ \kappa_1^{W^1}(\mathbf{x})\left(\Delta W^1_{m+1}\right)^2 + \kappa_1^{W^2}(\mathbf{x})\left(\Delta W^2_{m+1}\right)^2 + v_1(\mathbf{x})V^{1,2}_{m+1}, \qquad (4.3) \\
X^{2,\Delta}_{m+1} &= x_2 + m_2(\mathbf{x})\Delta t + s_2^{W^1}(\mathbf{x})\Delta W^1_{m+1} + s_2^{W^2}(\mathbf{x})\Delta W^2_{m+1} + \kappa_2^{W^1,W^2}(\mathbf{x})\Delta W^1_{m+1}\Delta W^2_{m+1} \\
&+ \kappa_2^{W^1}(\mathbf{x})\left(\Delta W^1_{m+1}\right)^2 + \kappa_2^{W^2}(\mathbf{x})\left(\Delta W^2_{m+1}\right)^2 + v_2(\mathbf{x})V^{1,2}_{m+1}, \qquad (4.4)
\end{aligned}$$

where $\mathbf{X}^\Delta_m = \mathbf{x} = (x_1, x_2)$, ΔW^1_{m+1} and ΔW^2_{m+1} are uncorrelated and both normally distributed with mean zero and variance Δt, and $V^{1,2}_{m+1}$ is an independent random variable with probability $P\left(V^{1,2}_{m+1} = \pm\Delta t\right) = \frac{1}{2}$.

For the Euler scheme, we have

$$\begin{array}{llll}
m_1(\mathbf{x}) = \mu_1(\mathbf{x}), & s_1^{W^1}(\mathbf{x}) = \sigma_1(\mathbf{x}), & s_1^{W^2}(\mathbf{x}) = 0, & \kappa_1^{W^1,W^2}(\mathbf{x}) = 0, \\
\kappa_1^{W^1}(\mathbf{x}) = 0, & \kappa_1^{W^2}(\mathbf{x}) = 0, & v_1(\mathbf{x}) = 0, & \\
 & & & \\
m_2(\mathbf{x}) = \mu_2(\mathbf{x}), & s_2^{W^1}(\mathbf{x}) = \rho\sigma_2(\mathbf{x}), & s_2^{W^2}(\mathbf{x}) = \sqrt{1-\rho^2}\sigma_2(\mathbf{x}), & \kappa_2^{W^1,W^2}(\mathbf{x}) = 0, \\
\kappa_2^{W^1}(\mathbf{x}) = 0, & \kappa_2^{W^2}(\mathbf{x}) = 0, & v_2(\mathbf{x}) = 0. &
\end{array} \qquad (4.5)$$

The order of strong convergence is $\gamma_1 = 0.5$ and the order of weak convergence is $\gamma_2 = 1$.

For the Milstein scheme, we find,

$$\begin{array}{ll}
m_1(\mathbf{x}) = \mu_1(\mathbf{x}) - \frac{1}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right), & s_1^{W^1}(\mathbf{x}) = \sigma_1(\mathbf{x}), \\
\kappa_1^{W^1,W^2}(\mathbf{x}) = \frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, & s_1^{W^2}(\mathbf{x}) = 0, \\
\kappa_1^{W^1}(\mathbf{x}) = \frac{1}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right), & \kappa_1^{W^2}(\mathbf{x}) = 0, \\
v_1(\mathbf{x}) = -\frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, & \\
 & \\
m_2(\mathbf{x}) = \mu_2(\mathbf{x}) - \frac{1}{2}\left(\rho\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), & s_2^{W^1}(\mathbf{x}) = \rho\sigma_2(\mathbf{x}), \\
\kappa_2^{W^1,W^2}(\mathbf{x}) = \frac{\sqrt{1-\rho^2}}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + 2\rho\sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), & s_2^{W^2}(\mathbf{x}) = \sqrt{1-\rho^2}\sigma_2(\mathbf{x}), \\
\kappa_2^{W^1}(\mathbf{x}) = \frac{\rho}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), & \kappa_2^{W^2}(\mathbf{x}) = \frac{1-\rho^2}{2}\sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}, \\
v_2(\mathbf{x}) = \frac{\sqrt{1-\rho^2}}{2}\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}. &
\end{array} \qquad (4.6)$$

The order of weak convergence is $\gamma_2 = 1$. The order of strong convergence is $\gamma_1 = 1$ when FSDEs (4.1)-(4.2) satisfy the following commutativity condition [18, Chapter 10 (3.13)]:

$$\sqrt{1-\rho^2}\sigma_2(x_1,x_2)\frac{\partial\sigma_1(x_1,x_2)}{\partial x_2} = 0 \quad \text{and} \quad \sqrt{1-\rho^2}\sigma_1(x_1,x_2)\frac{\partial\sigma_2(x_1,x_2)}{\partial x_1} = 0 \quad \forall x_1, x_2 \in \mathbb{R}^2. \qquad (4.7)$$

For the 2.0-weak-Taylor scheme, it follows that

$$\begin{aligned}
m_1(\mathbf{x}) &= \mu_1(\mathbf{x}) - \frac{1}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right) \\
&+ \frac{1}{2}\left(\mu_1(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_1} + \mu_2(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_2} + \frac{1}{2}\sigma_1^2(\mathbf{x})\frac{\partial^2\mu_1(\mathbf{x})}{(\partial x_1)^2}\right.
\end{aligned}$$

$$\begin{aligned}
&+ \quad \rho\sigma_1(\mathbf{x})\,\sigma_2(\mathbf{x})\,\frac{\partial^2\mu_1(\mathbf{x})}{\partial x_1\partial x_2}+\frac{1}{2}\sigma_2^2(\mathbf{x})\,\frac{\partial^2\mu_1(\mathbf{x})}{(\partial x_2)^2}\Bigg)\,\Delta t, && (4.8a)\\
s_1^{W^1}(\mathbf{x}) &= \sigma_1(\mathbf{x})+\frac{1}{2}\Bigg(\sigma_1(\mathbf{x})\,\frac{\partial\mu_1(\mathbf{x})}{\partial x_1}+\rho\sigma_2(\mathbf{x})\,\frac{\partial\mu_1(\mathbf{x})}{\partial x_2}+\mu_1(\mathbf{x})\,\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1}+\mu_2(\mathbf{x})\,\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\\
&+ \frac{1}{2}\sigma_1^2(\mathbf{x})\,\frac{\partial^2\sigma_1(\mathbf{x})}{(\partial x_1)^2}+\rho\sigma_1(\mathbf{x})\,\sigma_2(\mathbf{x})\,\frac{\partial^2\sigma_1(\mathbf{x})}{\partial x_1\partial x_2}+\frac{1}{2}\sigma_2^2(\mathbf{x})\,\frac{\partial^2\sigma_1(\mathbf{x})}{(\partial x_2)^2}\Bigg)\,\Delta t, && (4.8b)\\
s_1^{W^2}(\mathbf{x}) &= \frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\,\frac{\partial\mu_1(\mathbf{x})}{\partial x_2}\Delta t, \qquad \kappa_1^{W^1,W^2}(\mathbf{x})=\frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\,\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, && (4.8c)\\
\kappa_1^{W^1}(\mathbf{x}) &= \frac{1}{2}\left(\sigma_1(\mathbf{x})\,\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1}+\rho\sigma_2(\mathbf{x})\,\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right), \qquad \kappa_1^{W^2}(\mathbf{x})=0, && (4.8d)\\
v_1(\mathbf{x}) &= -\frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\,\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, && (4.8e)
\end{aligned}$$

$$\begin{aligned}
m_2(\mathbf{x}) &= \mu_2(\mathbf{x})-\frac{1}{2}\left(\rho\sigma_1(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}+\sigma_2(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right)\\
&+ \frac{1}{2}\Bigg(\mu_1(\mathbf{x})\,\frac{\partial\mu_2(\mathbf{x})}{\partial x_1}+\mu_2(\mathbf{x})\,\frac{\partial\mu_2(\mathbf{x})}{\partial x_2}+\frac{1}{2}\sigma_1^2(\mathbf{x})\,\frac{\partial^2\mu_2(\mathbf{x})}{(\partial x_1)^2}\\
&+ \rho\sigma_1(\mathbf{x})\,\sigma_2(\mathbf{x})\,\frac{\partial^2\mu_2(\mathbf{x})}{\partial x_1\partial x_2}+\frac{1}{2}\sigma_2^2(\mathbf{x})\,\frac{\partial^2\mu_2(\mathbf{x})}{(\partial x_2)^2}\Bigg)\,\Delta t, && (4.8f)\\
s_2^{W^1}(\mathbf{x}) &= \rho\sigma_2(\mathbf{x})+\frac{1}{2}\sigma_1(\mathbf{x})\,\frac{\partial\mu_2(\mathbf{x})}{\partial x_1}\Delta t\\
&+ \frac{\rho}{2}\Bigg(\sigma_2(\mathbf{x})\,\frac{\partial\mu_2(\mathbf{x})}{\partial x_2}+\mu_1(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}+\mu_2(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}+\frac{1}{2}\sigma_1^2(\mathbf{x})\,\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_1)^2}\\
&+ \rho\sigma_1(\mathbf{x})\,\sigma_2(\mathbf{x})\,\frac{\partial^2\sigma_2(\mathbf{x})}{\partial x_1\partial x_2}+\frac{1}{2}\sigma_2^2(\mathbf{x})\,\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_2)^2}\Bigg)\,\Delta t, && (4.8g)\\
s_2^{W^2}(\mathbf{x}) &= \sqrt{1-\rho^2}\sigma_2(\mathbf{x})+\frac{\sqrt{1-\rho^2}}{2}\Bigg(\sigma_2(\mathbf{x})\,\frac{\partial\mu_2(\mathbf{x})}{\partial x_2}+\mu_1(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}+\mu_2(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\\
&+ \frac{1}{2}\sigma_1^2(\mathbf{x})\,\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_1)^2}+\rho\sigma_1(\mathbf{x})\,\sigma_2(\mathbf{x})\,\frac{\partial^2\sigma_2(\mathbf{x})}{\partial x_1\partial x_2}+\frac{1}{2}\sigma_2^2(\mathbf{x})\,\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_2)^2}\Bigg)\,\Delta t, && (4.8h)\\
\kappa_2^{W^1,W^2}(\mathbf{x}) &= \frac{\sqrt{1-\rho^2}}{2}\left(\sigma_1(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}+2\rho\sigma_2(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), && (4.8i)\\
\kappa_2^{W^1}(\mathbf{x}) &= \frac{\rho}{2}\left(\sigma_1(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}+\rho\sigma_2(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), && (4.8j)\\
\kappa_2^{W^2}(\mathbf{x}) &= \frac{1-\rho^2}{2}\sigma_2(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}, \qquad v_2(\mathbf{x})=\frac{\sqrt{1-\rho^2}}{2}\sigma_1(\mathbf{x})\,\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}. && (4.8k)
\end{aligned}$$

The order of strong convergence is $\gamma_1 = 1$ when the FSDEs (4.1)-(4.2) satisfy the commutativity condition (4.7). The order of weak convergence is $\gamma_2 = 2$ [18].
The derivation of these Taylor schemes is given in Appendix D.

Remark 4.1. *For the strong convergence $\gamma_1 = 1$, the Milstein scheme and the 2.0-weak-Taylor scheme the FSDEs* (4.1)-(4.2) *have to satisfy the commutativity condition (4.7). We will observe in section 4.9 that the convergence of the BCOS method depends only on the order of weak convergence and not on the order of strong convergence, which implies that for application of the BCOS method it does not matter whether the FSDEs* (4.1)-(4.2) *satisfy the commutativity condition.*

Lemma 4.1. *The bivariate characteristic function of* $\mathbf{X}_{m+1}^{\Delta}$*, given* $\mathbf{X}_m^{\Delta} = \mathbf{x} = (x_1, x_2)$*, is given by*

$$\begin{aligned}
\phi_{\mathbf{X}_{m+1}^{\Delta}}(u_1, u_2|\mathbf{x}) &= \frac{\cosh(ic_6\Delta t)\exp(iu_1[x_1 + m_1(\mathbf{x})\Delta t] + iu_2[x_2 + m_2(\mathbf{x})\Delta t])}{\sqrt{(1-2ic_4\Delta t)(1-2ic_5\Delta t) + c_3^2(\Delta t)^2}} \qquad (4.9)\\
&\cdot \exp\left(-\frac{\Delta t}{2}\frac{c_1^2 + c_2^2 + \left[4(c_2^2c_4^2 + c_1^2c_5^2) - 4c_1c_2c_3(c_4+c_5) + (c_1^2+c_2^2)c_3^2\right](\Delta t)^2}{1 + \left(2c_3^2 + 4c_4^2 + 4c_5^2\right)(\Delta t)^2 + \left(c_3^2 - 4c_4c_5\right)^2(\Delta t)^4}\right)\\
&\cdot \exp\left(i(\Delta t)^2\left(\frac{-c_1^2c_4 - c_2^2c_5 - c_1c_2c_3 + (c_3^2 - 4c_4c_5)(c_1^2c_5 - c_1c_2c_3 + c_2^2c_4)(\Delta t)^2}{1 + \left(2c_3^2 + 4c_4^2 + 4c_5^2\right)(\Delta t)^2 + \left(c_3^2 - 4c_4c_5\right)^2(\Delta t)^4}\right)\right),
\end{aligned}$$

where

$$\begin{aligned}
c_1 &= u_1 s_1^{W^1}(\mathbf{x}) + u_2 s_2^{W^1}(\mathbf{x}), & c_4 &= u_1\kappa_1^{W^1}(\mathbf{x}) + u_2\kappa_2^{W^1}(\mathbf{x}),\\
c_2 &= u_1 s_1^{W^2}(\mathbf{x}) + u_2 s_2^{W^2}(\mathbf{x}), & c_5 &= u_1\kappa_1^{W^2}(\mathbf{x}) + u_2\kappa_2^{W^2}(\mathbf{x}), \qquad (4.10)\\
c_3 &= u_1\kappa_1^{W^1,W^2}(\mathbf{x}) + u_2\kappa_2^{W^1,W^2}(\mathbf{x}), & c_6 &= u_1 v_1(\mathbf{x}) + u_2 v_2(\mathbf{x}).
\end{aligned}$$

For the Euler scheme it follows that

$$\begin{aligned}
\phi_{\mathbf{X}_{m+1}^{\Delta}}(u_1, u_2|\mathbf{x}) &= \exp\Big(iu_1x_1 + iu_2x_2 + iu_1\mu_1(\mathbf{x})\Delta t + iu_2\mu_2(\mathbf{x})\Delta t - \frac{1}{2}u_1^2\sigma_1^2(\mathbf{x})\Delta t\\
&- \frac{1}{2}u_2^2\sigma_2^2(\mathbf{x})\Delta t - u_1u_2\rho\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\Delta t\Big). \qquad (4.11)
\end{aligned}$$

The proof of Lemma 4.1 is given in Appendix E.

4.2 An adjusted-Predictor-Corrector scheme

Besides the discretization schemes given in Section 4.1, we also consider whether we can use a predictor-corrector method. Following Kloeden and Platen [18], the family of predictor-corrector Euler schemes with $\gamma_1 = 0.5$ as the order of strong convergence [7] and $\gamma_2 = 1$ as the order of weak convergence [18] is given by

$$\begin{aligned}
X_{m+1}^{1,\Delta} &= x_1 + \left[\theta_1\bar{\mu}_1\left(\bar{X}_{m+1}^{1,\Delta}, \bar{X}_{m+1}^{2,\Delta}\right) + (1-\theta_1)\bar{\mu}_1(x_1, x_2)\right]\Delta t\\
&+ \left[\eta_1\sigma_1\left(\bar{X}_{m+1}^{1,\Delta}, \bar{X}_{m+1}^{2,\Delta}\right) + (1-\eta_1)\sigma_1(x_1, x_2)\right]\Delta W_{m+1}^1, \qquad (4.12)\\
X_{m+1}^{2,\Delta} &= x_2 + \left[\theta_2\bar{\mu}_2\left(\bar{X}_{m+1}^{1,\Delta}, \bar{X}_{m+1}^{2,\Delta}\right) + (1-\theta_2)\bar{\mu}_2(x_1, x_2)\right]\Delta t \qquad (4.13)\\
&+ \left[\eta_2\sigma_2\left(\bar{X}_{m+1}^{1,\Delta}, \bar{X}_{m+1}^{2,\Delta}\right) + (1-\eta_2)\sigma_2(x_1, x_2)\right]\left(\rho\Delta W_{m+1}^1 + \sqrt{1-\rho^2}\Delta W_{m+1}^2\right),
\end{aligned}$$

where $\theta_1, \theta_2, \eta_1, \eta_2 \in [0, 1]$ and

$$\begin{aligned}
\bar{X}_{m+1}^{1,\Delta} &= x_1 + \mu_1(x_1, x_2)\Delta t + \sigma_1(x_1, x_2)\Delta W_{m+1}^1, \qquad (4.14)\\
\bar{X}_{m+1}^{2,\Delta} &= x_2 + \mu_2(x_1, x_2)\Delta t + \sigma_2(x_1, x_2)\left(\rho\Delta W_{m+1}^1 + \sqrt{1-\rho^2}\Delta W_{m+1}^2\right), \qquad (4.15)\\
\bar{\mu}_1(x, y) &= \mu_1(x, y) - \eta_1\left(\sigma_1(x, y)\frac{\partial\sigma_1(x, y)}{\partial x} + \rho\sigma_2(x, y)\frac{\partial\sigma_1(x, y)}{\partial y}\right), \qquad (4.16)\\
\bar{\mu}_2(x, y) &= \mu_2(x, y) - \eta_2\left(\rho\sigma_1(x, y)\frac{\partial\sigma_2(x, y)}{\partial x} + \sigma_2(x, y)\frac{\partial\sigma_2(x, y)}{\partial y}\right). \qquad (4.17)
\end{aligned}$$

Often, the discretized FSDEs $X_{m+1}^{1,\Delta}$ and $X_{m+1}^{2,\Delta}$ are not in the general form (4.3)-(4.4). We use a Taylor series expansion to obtain discretized FSDEs in this form. We call the resulting scheme the *adjusted-Predictor-Corrector scheme*. An example is given below.

Example 1 For the SABR model with FSDEs,

$$\begin{aligned} \mathrm{d}F_t &= \sigma_t (F_t)^\beta \,\mathrm{d}W_t^1, && (4.18)\\ \mathrm{d}\sigma_t &= \rho\nu\sigma_t \,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t \,\mathrm{d}W_t^2, && (4.19)\end{aligned}$$

and $\left(F_m^\Delta, \sigma_m^\Delta\right) = (f, \alpha)$. We find

$$\begin{aligned} F_{m+1}^\Delta &= f + \left[\eta_1\bar{\alpha}\bar{f}^\beta + (1-\eta_1)\alpha f^\beta\right]\Delta W_{m+1}^1\\ &- \eta_1\left[\theta_1\left(\beta\bar{\alpha}^2\bar{f}^{2\beta-1} + \rho\nu\bar{\alpha}\bar{f}^\beta\right) + (1-\theta_1)\left(\beta\alpha^2 f^{2\beta-1} + \rho\nu\alpha f^\beta\right)\right]\Delta t, && (4.20)\\ \sigma_{m+1}^\Delta &= \alpha - \eta_2\nu^2\left[\theta_2\bar{\alpha} + (1-\theta_2)\alpha\right]\Delta t + \nu\left[\eta_2\bar{\alpha} + (1-\eta_2)\alpha\right]\left(\rho\Delta W_{m+1}^1 + \sqrt{1-\rho^2}\Delta W_{m+1}^2\right)\\ &= \alpha - \eta_2\nu^2\alpha\Delta t + \left(\nu\alpha - \theta_2\eta_2\nu^3\alpha\Delta t\right)\left(\rho\Delta W_{m+1}^1 + \sqrt{1-\rho^2}\Delta W_{m+1}^2\right)\\ &+ \eta_2\nu^2\alpha\left[\rho^2\left(\Delta W_{m+1}^1\right)^2 + 2\rho\sqrt{1-\rho^2}\Delta W_{m+1}^1\Delta W_{m+1}^2 + \left(1-\rho^2\right)\left(\Delta W_{m+1}^2\right)^2\right], && (4.21)\end{aligned}$$

where the predictors $\bar{f}$ and $\bar{\alpha}$ are given by

$$\bar{f} = f + \alpha f^\beta \Delta W_{m+1}^1 \quad \text{and} \quad \bar{\alpha} = \alpha + \nu\alpha\left[\rho\Delta W_{m+1}^1 + \sqrt{1-\rho^2}\Delta W_{m+1}^2\right]. \tag{4.22}$$

Because of the complexity of F_{m+1}^Δ, it would be complicated, if not impossible, to find an analytical expression for the bivariate characteristic function of the discrete processes: $\phi_{\left(F_{m+1}^\Delta, \sigma_{m+1}^\Delta\right)}(u_1, u_2 | f, \alpha)$. Our idea is to use a Taylor series expansion around $\left(\Delta W_{m+1}^1, \Delta W_{m+1}^2\right) = (0,0)$, which gives

$$\begin{aligned} F_{m+1}^\Delta = f + \alpha f^\beta \Delta W_{m+1}^1 &+ \eta_1\left[\beta\alpha^2 f^{2\beta-1} + \rho\nu\alpha f^\beta\right]\left(\Delta W_{m+1}^1\right)^2\\ &+ \eta_1\sqrt{1-\rho^2}\nu\alpha f^\beta\Delta W_{m+1}^1\Delta W_{m+1}^2\\ - \eta_1\left[\beta\alpha^2 f^{2\beta-1} + \rho\nu\alpha f^\beta\right]\Delta t &- \eta_1\theta_1\sqrt{1-\rho^2}\left[2\beta\nu\alpha^2 f^{2\beta-1} + \rho\nu^2\alpha f^\beta\right]\Delta t\Delta W_{m+1}^2\\ &- \eta_1\theta_1\left[3\rho\beta\nu\alpha^2 f^{2\beta-1} + \beta(2\beta-1)\alpha^3 f^{3\beta-2} + \rho^2\nu^2\alpha f^\beta\right]\Delta t\Delta W_{m+1}^1\\ - \eta_1\theta_1&\left[2\rho^2\beta\nu^2\alpha^2 f^{2\beta-1} + \left(2\beta^3 - 3\beta^2 + \beta\right)\alpha^4 f^{4\beta-3} + \frac{1}{2}\rho\beta(9\beta-5)\nu\alpha^3 f^{3\beta-2}\right]\Delta t\left(\Delta W_{m+1}^1\right)^2\\ &- \eta_1\theta_1\sqrt{1-\rho^2}\beta\left[3\rho\nu^2\alpha^2 f^{2\beta-1} + 2(2\beta-1)\nu\alpha^3 f^{3\beta-2}\right]\Delta t\Delta W_{m+1}^1\Delta W_{m+1}^2\\ &- \eta_1\theta_1\left(1-\rho^2\right)\beta\nu^2\alpha^2 f^{2\beta-1}\Delta t\left(\Delta W_{m+1}^2\right)^2 + \mathcal{O}\left(\left(\Delta W_{m+1}^1 + \Delta W_{m+1}^2\right)^3\right) && (4.23)\end{aligned}$$

Now, we can determine the corresponding bivariate characteristic function by using Lemma 4.1, where

$$\begin{aligned} m_1(f,\alpha) &= -\eta_1\left[\beta\alpha^2 f^{2\beta-1} + \rho\nu\alpha f^\beta\right], && (4.24a)\\ s_1^{W^1}(f,\alpha) &= \alpha f^\beta - \eta_1\theta_1\left[3\rho\beta\nu\alpha^2 f^{2\beta-1} + \beta(2\beta-1)\alpha^3 f^{3\beta-2} + \rho^2\nu^2\alpha f^\beta\right]\Delta t, && (4.24b)\\ s_1^{W^2}(f,\alpha) &= -\eta_1\theta_1\sqrt{1-\rho^2}\left[2\beta\nu\alpha^2 f^{2\beta-1} + \rho\nu^2\alpha f^\beta\right]\Delta t, && (4.24c)\\ \kappa_1^{W^1,W^2}(f,\alpha) &= \eta_1\sqrt{1-\rho^2}\left[\nu\alpha f^\beta - \theta_1\beta\left\{3\rho\nu^2\alpha^2 f^{2\beta-1} + 2(2\beta-1)\nu\alpha^3 f^{3\beta-2}\right\}\Delta t\right], && (4.24d)\end{aligned}$$

$$\begin{aligned}\kappa_1^{W^1}(f,\alpha) &= \eta_1\left[\beta\alpha^2 f^{2\beta-1} + \rho\nu\alpha f^{\beta} - \theta_1\left\{2\rho^2\beta\nu^2\alpha^2 f^{2\beta-1} + \left(2\beta^3 - 3\beta^2 + \beta\right)\alpha^4 f^{4\beta-3}\right.\right.\\ &+ \left.\left.\frac{1}{2}\rho\beta(9\beta-5)\nu\alpha^3 f^{3\beta-2}\right\}\Delta t\right], \quad (4.24e)\\ \kappa_1^{W^2}(f,\alpha) &= -\eta_1\theta_1\left(1-\rho^2\right)\beta\nu^2\alpha^2 f^{2\beta-1}\Delta t, \qquad v_1(f,\alpha) = 0, \quad (4.24f)\end{aligned}$$

$$\begin{aligned} &m_2(f,\alpha) = -\eta_2\nu^2\alpha, && s_2^{W^1}(f,\alpha) = \rho\nu\alpha - \theta_2\eta_2\rho\nu^3\alpha\Delta t,\\ &s_2^{W^2}(f,\alpha) = \sqrt{1-\rho^2}\nu\alpha - \theta_2\eta_2\sqrt{1-\rho^2}\nu^3\alpha\Delta t, && \kappa_2^{W^1,W^2}(f,\alpha) = 2\eta_2\rho\sqrt{1-\rho^2}\nu^2\alpha,\\ &\kappa_2^{W^1}(f,\alpha) = \eta_2\rho^2\nu^2\alpha, && \kappa_2^{W^2}(f,\alpha) = \eta_2(1-\rho^2)\nu^2\alpha,\\ &v_2(f,\alpha) = 0. \end{aligned} \quad (4.24g)$$

In Appendix F, we give the adjusted-Predictor-Corrector schemes for the Heston and the SABR model.

4.3 The characteristic function of the Heston model

Whereas we cannot easily analyze errors made with the discrete characteristic function for the SABR model, but we can analyze these errors for the Heston model. We therefore compare the bivariate characteristic function of the Heston model with the bivariate characteristic function of its discretization in this section. The FSDEs of the Heston model, where X_t^1 and X_t^2 denote the log forward process and the volatility process, respectively, are given by

$$\mathrm{d}X_t^1 = -\frac{1}{2}X_t^2\,\mathrm{d}t + \sqrt{X_t^2}\,\mathrm{d}W_t^1, \quad (4.25)$$

$$\mathrm{d}X_t^2 = \rho\gamma\sqrt{X_t^2}\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\gamma\sqrt{X_t^2}\,\mathrm{d}W_t^2. \quad (4.26)$$

This is an affine model, and therefore we can derive an analytical formula for its bivariate characteristic function [12]. The characteristic function of $\mathbf{X}_{m+1}^{\Delta}$, given $\mathbf{X}_m^{\Delta} = \mathbf{x} = (x_1, x_2)$, is then given by

$$\phi_{\mathbf{X}_{m+1}}(u_1,u_2|\mathbf{x}) = \exp\left(iu_1x_1 + \frac{1}{\gamma^2}\left[\zeta\tan\left(\arctan\left[\frac{iu_2\gamma^2 + iu_1\rho\gamma}{\zeta}\right] + \frac{\Delta t}{2}\zeta\right) - iu_1\rho\gamma\right]x_2\right), \quad (4.27)$$

where

$$\zeta = \sqrt{-iu_1\gamma^2 + u_1^2(\rho^2-1)\gamma^2}. \quad (4.28)$$

A Taylor series expansion around $\sqrt{\Delta t} = 0$ gives

$$\begin{aligned}\phi_{\mathbf{X}_{m+1}}(u_1,u_2|\mathbf{x}) &= \exp\left(iu_1x_1 + iu_2x_2 - \frac{iu_1 + u_1^2 + u_2^2\gamma^2 + 2u_1u_2\rho\gamma}{2}x_2\Delta t\right)\\ &\cdot \exp\left(-\frac{i}{4}\left[u_1^3\rho\gamma + u_1^2u_2(2\rho^2+1)\gamma^2 + 3u_1u_2^2\rho\gamma^3 + u_2^3\gamma^4\right]x_2(\Delta t)^2\right)\\ &\cdot \exp\left(\frac{1}{4}\left[u_1^2\rho\gamma + u_1u_2\gamma^2\right]x_2(\Delta t)^2 + \mathcal{O}\left((\Delta t)^3\right)\right). \quad (4.29)\end{aligned}$$

For each scheme (Euler, Milstein, 2.0-weak-Taylor, and adjusted-Predictor-Corrector) we will use Lemma 4.1 to determine the characteristic function of the discretized forward process. The Taylor series expansion of the characteristic function of $\mathbf{X}_{m+1}^{\Delta}$, given $\mathbf{X}_m^{\Delta} = \mathbf{x} = (x_1, x_2)$, around $\sqrt{\Delta t} = 0$ is given by

$$\phi_{\mathbf{X}_{m+1}^{\Delta}}(u_1,u_2|\mathbf{x}) = \exp\left(iu_1x_1 + iu_2x_2 + \frac{2i\left[u_1m_1(\mathbf{x}) + u_2m_2(\mathbf{x}) + c_4 + c_5\right] - c_1^2 - c_2^2}{2}\Delta t\right)$$

$$\cdot \quad \exp\left(-\frac{c_3^2+2c_4^2+2c_5^2+c_6^2+2i\left[c_1^2c_4+c_2^2c_5+c_1c_2c_3\right]}{2}(\Delta t)^2+\mathcal{O}\left((\Delta t)^3\right)\right). \tag{4.30}$$

For the Euler scheme, we find

$$\phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x}) = \exp\left(iu_1x_1+iu_2x_2-\frac{iu_1+u_1^2+u_2^2\gamma^2+2u_1u_2\rho\gamma}{2}x_2\Delta t\right). \tag{4.31}$$

For the Milstein scheme, we have

$$\begin{aligned}\phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x}) &= \exp\left(iu_1x_1+iu_2x_2-\frac{iu_1+u_1^2+u_2^2\gamma^2+2u_1u_2\rho\gamma}{2}x_2\Delta t\right)\\ &\cdot \exp\left(-\frac{i}{4}\left[u_1^3\rho\gamma+u_1^2u_2(1+2\rho^2)\gamma^2+3u_1u_2^2\rho\gamma^3+u_2^3\gamma^4\right]x_2(\Delta t)^2\right)\\ &\cdot \exp\left(-\frac{1}{16}\left[u_1^2(1+\rho^2)\gamma^2+2u_1u_2\rho\gamma^3+u_2^2\gamma^4\right](\Delta t)^2+\mathcal{O}\left((\Delta t)^3\right)\right). \end{aligned}\tag{4.32}$$

For the 2.0-weak-Taylor scheme, it follows that

$$\begin{aligned}\phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x}) &= \exp\left(iu_1x_1+iu_2x_2-\frac{iu_1+u_1^2+u_2^2\gamma^2+2u_1u_2\rho\gamma}{2}x_2\Delta t\right)\\ &\cdot \exp\left(-\frac{i}{4}\left[u_1^3\rho\gamma+u_1^2u_2(1+2\rho^2)\gamma^2+3u_1u_2^2\rho\gamma^3+u_2^3\gamma^4\right]x_2(\Delta t)^2\right)\\ &\cdot \exp\left(\frac{1}{4}\left[u_1^2\rho\gamma+u_1u_2\gamma^2\right]x_2(\Delta t)^2+\mathcal{O}\left((\Delta t)^3\right)\right). \end{aligned}\tag{4.33}$$

For the adjusted-Predictor-Corrector scheme, see Section 4.2 and Appendix F, we finally find

$$\begin{aligned}\phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x}) &= \exp\left(iu_1x_1+iu_2x_2-\frac{iu_1+u_1^2+u_2^2\gamma^2+2u_1u_2\rho\gamma}{2}x_2\Delta t\right)\\ &\cdot \exp\left(-\frac{i}{2}\left[u_1^3\eta_1\rho\gamma+u_1^2u_2(\eta_2\rho^2+\eta_1(1+\rho^2))\gamma^2+u_1u_2^2(\eta_1\rho+2\eta_2\rho)\gamma^3\right.\right.\\ &+ \left.\left. u_2^3\eta_2\gamma^4\right]x_2(\Delta t)^2+\frac{1}{2}\left[u_1^2\theta_1\rho\gamma+u_1u_2\theta_1\gamma^2\right]x_2(\Delta t)^2\right)\\ &\cdot \exp\left(-\frac{1}{8}\left[u_1^2\eta_1^2(1+\rho^2)\gamma^2+4u_1u_2\eta_1\eta_2\rho\gamma^3+2u_2^2\eta_2^2\gamma^4\right](\Delta t)^2+\mathcal{O}\left((\Delta t)^3\right)\right), \end{aligned}\tag{4.34}$$

where $\theta_1, \eta_1, \eta_2 \in [0,1]$.

So, the bivariate characteristic function of the Heston model discretized with the Euler, Milstein or adjusted-Predictor-Corrector scheme, independent of the choices for θ_1, η_1 and η_2, is exact up to $\mathcal{O}(\Delta t)$, i.e.

$$\phi_{\mathbf{X}_{m+1}}(u_1,u_2|\mathbf{x}) = \phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x})+\mathcal{O}\left((\Delta t)^2\right) \tag{4.35}$$

and for the Heston model discretized with the weak-2.0-Taylor scheme it is even exact up to $\mathcal{O}\left((\Delta t)^2\right)$:

$$\phi_{\mathbf{X}_{m+1}}(u_1,u_2|\mathbf{x}) = \phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x})+\mathcal{O}\left((\Delta t)^3\right). \tag{4.36}$$

For the Heston model, the discretization with the weak-2.0-Taylor scheme gives the best approximation for the characteristic function. We expect similar results for the SABR model.

4.4 Two-dimensional COS method

We derive the value $V(0,T,\mathbf{X}_0)$ of an option at time 0 with as underlying $\mathbf{X}_t$ and exercise date T. The payoff of the option at time T is given by $V(T,T,\mathbf{X}_T) = g(\mathbf{X}_T)$ for some function g and we assume that there are no early-exercise dates. The value of the option $V(0,T,\mathbf{X}_0)$ is then given by the risk-neutral valuation formula,

$$V(0,T,\mathbf{X}_0) = e^{-rT} \int\int_{\mathbb{R}^2} g(X_1,X_2) Q_{\mathbf{X}_T}(T,X_1,X_2|\mathbf{X}_0)\ \mathrm{d}X_1\,\mathrm{d}X_2, \tag{4.37}$$

where r is the risk-free interest rate and $Q_{\mathbf{X}_T}(T,X_1,X_2|\mathbf{X}_0)$ is the conditional density function of $X_T^1 = X_1$ and $X_T^2 = X_2$ given $\mathbf{X}_0$. We abbreviate (X_1,X_2) by $\mathbf{X}$.

The value of the option $V(0,T,\mathbf{X}_0)$ can be derived by using the two-dimensional COS method of M.J. Ruijter and C.W. Oosterlee [23]. As in Section 3.2, we truncate the integration domain from $\mathbb{R}^2$ to some finite domain $[a_1,b_1]\times[a_2,b_2] \subseteq \mathbb{R}^2$, which leads to

$$V(0,T,\mathbf{X}_0) \approx e^{-rT} \int_{a_2}^{b_2}\int_{a_1}^{b_1} g(\mathbf{X}) Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0)\ \mathrm{d}X_1\,\mathrm{d}X_2. \tag{4.38}$$

The two-dimensional Fourier-cosine expansion of the density results in

$$Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0) = \sum_{k_1=0}^{\infty}{}' \sum_{k_2=0}^{\infty}{}' A_{k_1,k_2} \cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{X_2-a_2}{b_2-a_2}\right), \tag{4.39}$$

where

$$A_{k_1,k_2} = \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\int_{a_2}^{b_2}\int_{a_1}^{b_1} Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0)\cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{X_2-a_2}{b_2-a_2}\right)\mathrm{d}X_1\,\mathrm{d}X_2. \tag{4.40}$$

Rewriting A_{k_1,k_2} gives $2A_{k_1,k_2} = A^+_{k_1,k_2} + A^-_{k_1,k_2}$, where

$$A^{\pm}_{k_1,k_2} = \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\int_{a_2}^{b_2}\int_{a_1}^{b_1} Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0)\cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1} \pm k_2\pi\frac{X_2-a_2}{b_2-a_2}\right)\mathrm{d}X_1\,\mathrm{d}X_2. \tag{4.41}$$

We truncate the series summations by setting N_1 and N_2 as the numbers of Fourier coefficients, which leads to

$$Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0) \approx \sum_{k_1=0}^{N_1-1}{}' \sum_{k_2=0}^{N_2-1}{}' A_{k_1,k_2} \cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{X_2-a_2}{b_2-a_2}\right). \tag{4.42}$$

Finally, we approximate $A^{\pm}_{k_1,k_2}$ by

$$\begin{aligned}
A^{\pm}_{k_1,k_2} &\approx \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\int\int_{\mathbb{R}^2} Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0)\cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1} \pm k_2\pi\frac{X_2-a_2}{b_2-a_2}\right)\mathrm{d}X_1\,\mathrm{d}X_2 \\
&= \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\Re\left\{\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1} \mp ik_2\pi\frac{a_2}{b_2-a_2}\right)\right. \\
&\quad \left.\int\int_{\mathbb{R}^2} Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0)\exp\left(ik_1\pi\frac{X_1}{b_1-a_1} \pm ik_2\pi\frac{X_2}{b_2-a_2}\right)\mathrm{d}X_1\,\mathrm{d}X_2\right\} \\
&= \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\Re\left\{\phi_{\mathbf{X}_T}\left(\frac{k_1\pi}{b_1-a_1}, \pm\frac{k_2\pi}{b_2-a_2}\middle|\,\mathbf{X}_0\right)\right.
\end{aligned}$$

$$\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}\mp ik_2\pi\frac{a_2}{b_2-a_2}\right)\Bigg\}. \tag{4.43}$$

This results in the following option pricing formula of the two-dimensional COS method:

$$\begin{aligned}
&V\left(0,T,\mathbf{X}_0\right)\\
&\approx e^{-rT}\int_{a_2}^{b_2}\int_{a_1}^{b_1} g(\mathbf{X})Q_{\mathbf{X}_T}(T,\mathbf{X}|\mathbf{X}_0)\;\mathrm{d}X_1\,\mathrm{d}X_2\\
&= e^{-rT}\int_{a_2}^{b_2}\int_{a_1}^{b_1} g(\mathbf{X})\sum_{k_1=0}^{N_1-1}{}'\sum_{k_2=0}^{N_2-1}{}' A_{k_1,k_2}\cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{X_2-a_2}{b_2-a_2}\right)\;\mathrm{d}X_1\,\mathrm{d}X_2\\
&= e^{-rT}\frac{b_1-a_1}{2}\frac{b_2-a_2}{2}\sum_{k_1=0}^{N_1-1}{}'\sum_{k_2=0}^{N_2-1}{}' A_{k_1,k_2}\mathcal{V}_{k_1,k_2}\\
&= e^{-rT}\frac{b_1-a_1}{2}\frac{b_2-a_2}{2}\sum_{k_1=0}^{N_1-1}{}'\sum_{k_2=0}^{N_2-1}{}' \frac{1}{2}\left(A^+_{k_1,k_2}+A^-_{k_1,k_2}\right)\mathcal{V}_{k_1,k_2}\\
&\approx e^{-rT}\sum_{k_1=0}^{N_1-1}{}'\sum_{k_2=0}^{N_2-1}{}'\frac{1}{2}\Re\left\{\phi_{\mathbf{X}_T}\left(\frac{k_1\pi}{b_1-a_1},\frac{k_2\pi}{b_2-a_2}\middle|\,\mathbf{X}_0\right)\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}-ik_2\pi\frac{a_2}{b_2-a_2}\right)\right.\\
&\left.+\,\phi_{\mathbf{X}_T}\left(\frac{k_1\pi}{b_1-a_1},-\frac{k_2\pi}{b_2-a_2}\middle|\,\mathbf{X}_0\right)\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}+ik_2\pi\frac{a_2}{b_2-a_2}\right)\right\}\mathcal{V}_{k_1,k_2},
\end{aligned} \tag{4.44}$$

where

$$\mathcal{V}_{k_1,k_2}=\frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\int_{a_2}^{b_2}\int_{a_1}^{b_1} g(\mathbf{X})\cos\left(k_1\pi\frac{X_1-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{X_2-a_2}{b_2-a_2}\right)\;\mathrm{d}X_1\,\mathrm{d}X_2. \tag{4.45}$$

4.5 Two-dimensional BCOS method

Just as in Section 4.4, we derive the value $V\left(0,T,\mathbf{X}_0\right)$ of a derivative at time 0 with as underlying $\mathbf{X}_t$ and exercise date T, where the FSDEs of $\mathbf{X}_t=\left(X_t^1,X_t^2\right)$ are given by (4.1) and (4.2). The payoff of the option at time T is given by $V\left(T,T,\mathbf{X}_T\right)=g\left(\mathbf{X}_T\right)$ for some function g and we assume that there are no early-exercise dates. We also assume that we are working in a complete market and therefore we can make a self-financing portfolio Y_t consisting of a_t^1 assets of X_t^1, a_t^2 assets of X_t^2 and bonds with risk-free return rate r, such that $Y_T=g\left(\mathbf{X}_T\right)$, and:

$$\begin{aligned}
\mathrm{d}Y_t &= r\left(Y_t-a_t^1X_t^1-a_t^2X_t^2\right)\mathrm{d}t+a_t^1\,\mathrm{d}X_t^1+a_t^2\,\mathrm{d}X_t^2\\
&= \left[rY_t+\left(\mu_1\left(\mathbf{X}_t\right)-rX_t^1\right)a_t^1+\left(\mu_2\left(\mathbf{X}_t\right)-rX_t^2\right)a_t^2\right]\mathrm{d}t\\
&+ \left[\sigma_1\left(\mathbf{X}_t\right)a_t^1+\rho\sigma_2\left(\mathbf{X}_t\right)a_t^2\right]\mathrm{d}W_t^1+\sqrt{1-\rho^2}\sigma_2\left(\mathbf{X}_t\right)a_t^2\,\mathrm{d}W_t^2,
\end{aligned} \tag{4.46}$$

for $0\le t\le T$. If we set $Z_t^1=\sigma_1\left(\mathbf{X}_t\right)a_t^1$ and $Z_t^2=\sigma_2\left(\mathbf{X}_t\right)a_t^2$, then (Y,Z^1,Z^2) solves the BSDE

$$\mathrm{d}Y_t=-f\left(t,X_t^1,X_t^2,Y_t,Z_t^1,Z_t^2\right)\mathrm{d}t+\left(Z_t^1+\rho Z_t^2\right)\mathrm{d}W_t^1+\sqrt{1-\rho^2}Z_t^2\,\mathrm{d}W_t^2, \tag{4.47}$$

$$f(t,x_1,x_2,y,z_1,z_2)=-ry-\frac{\mu_1(x_1,x_2)-rx_1}{\sigma_1(x_1,x_2)}z_1-\frac{\mu_2(x_1,x_2)-rx_2}{\sigma_2(x_1,x_2)}z_2, \tag{4.48}$$

where $Y_T=g(\mathbf{X}_T)$. The functions $\sigma_1,\sigma_2,\mu_1,\mu_2:\mathbb{R}^2\to\mathbb{R}$ in FSDEs (4.1) and (4.2) are assumed to be twice differentiable with respect to x_1 and x_2, Lipschitz in x_1 and x_2 and satisfy a linear growth condition in x_1 and x_2. The function $f : [0,T] \times \mathbb{R}^5 \to \mathbb{R}$ is assumed to be uniformly continuous with respect to x_1 and x_2 and satisfies a Lipschitz condition in (y, z_1, z_2), with Lipschitz constant L_f and the function $g : \mathbb{R}^2 \to \mathbb{R}$ is assumed to be uniformly continuous with respect to x_1 and x_2. Also, there exists a constant C such that $|f(t, \mathbf{x}, y, \mathbf{z})| + |g(\mathbf{x})| \leq C(1 + \|\mathbf{x}\|^p + |y| + \|\mathbf{z}\|)$, $p \geq \frac{1}{2}$[1]. Y_t is a self-financing portfolio, and therefore the option value is given by $V(0, T, \mathbf{X}_0) = Y_0$.

We define again a time grid of $M+1$ time points, where $t_j = j\Delta t$ and $\Delta t = \frac{T}{M}$. We define $\mathbf{\Lambda}_t := \left(X_t^1, X_t^2, Y_t, Z_t^1, Z_t^2\right)$, $\mathbf{\Lambda}_m = \mathbf{\Lambda}_{t_m}$ and $\mathbf{\Lambda}_m^\Delta = \left(X_m^{1,\Delta}, X_m^{2,\Delta}, Y_m\Delta, Z_m^{1,\Delta}, Z_m^{2,\Delta}\right)$. Integrating (4.48) gives us:

$$Y_0 = g(\mathbf{X}_T) + \int_0^T f(t, \mathbf{\Lambda}_t)\ \mathrm{d}t - \int_0^T \left(Z_t^1 + \rho Z_t^2\right)\ \mathrm{d}W_t^1 - \sqrt{1-\rho^2}\int_0^T Z_t^2\ \mathrm{d}W_t^2. \tag{4.49}$$

At time t_m this gives the recursion:

$$Y_m = Y_{m+1} + \int_{t_m}^{t_{m+1}} f(t, \mathbf{\Lambda}_t)\ \mathrm{d}t - \int_{t_m}^{t_{m+1}} \left(Z_t^1 + \rho Z_t^2\right)\ \mathrm{d}W_t^1 - \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}} Z_t^2\ \mathrm{d}W_t^2. \tag{4.50}$$

We take conditional expectations at both sides of the equation and use numerical integration to approximate the integral, for some $\theta \in [0,1]$

$$\begin{aligned} Y_m &= \mathbb{E}_m\left[Y_{m+1}\right] + \mathbb{E}_m\left[\int_{t_m}^{t_{m+1}} f(t, \mathbf{\Lambda}_t)\ \mathrm{d}t\right] && (4.51)\\ &\approx \mathbb{E}_m\left[Y_{m+1}\right] + \mathbb{E}_m\left[\Delta t\theta f(t_m, \mathbf{\Lambda}_m) + \Delta t(1-\theta) f(t_{m+1}, \mathbf{\Lambda}_{m+1})\right] \\ &= \mathbb{E}_m\left[Y_{m+1}\right] + \Delta t\theta f(t_m, \mathbf{\Lambda}_m) + \Delta t(1-\theta)\mathbb{E}_m\left[f(t_{m+1}, \mathbf{\Lambda}_{m+1})\right]. && (4.52)\end{aligned}$$

Multiplying equation (4.50) with ΔW_{m+1}^1 gives

$$\begin{aligned} Y_m \Delta W_{m+1}^1 &= Y_{m+1}\Delta W_{m+1}^1 + \int_{t_m}^{t_{m+1}} f(t, \mathbf{\Lambda}_t)\ \mathrm{d}t\Delta W_{m+1}^1 - \int_{t_m}^{t_{m+1}} \left(Z_t^1 + \rho Z_t^2\right)\ \mathrm{d}W_t^1 \Delta W_{m+1}^1 \\ &- \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}} Z_t^2\ \mathrm{d}W_t^2 \Delta W_{m+1}^1. && (4.53)\end{aligned}$$

Again, we take conditional expectations at both sides of the equation and use numerical integration. This gives

$$\begin{aligned} 0 &\approx \mathbb{E}_m\left[Y_{m+1}\Delta W_{m+1}^1\right] + \mathbb{E}_m\left[\left(\Delta t\theta f(t_m, \mathbf{\Lambda}_m) + \Delta t(1-\theta) f(t_{m+1}, \mathbf{\Lambda}_{m+1})\right)\Delta W_{m+1}^1\right] \\ &- \mathbb{E}_m\left[\left(\theta\Delta W_{m+1}^1\left(Z_m^1 + \rho Z_m^2\right) + (1-\theta)\Delta W_{m+1}^1\left(Z_{m+1}^1 + \rho Z_{m+1}^2\right)\right)\Delta W_{m+1}^1\right] \\ &- \sqrt{1-\rho^2}\mathbb{E}_m\left[\theta\Delta W_{m+1}^2\left(Z_m^2 + Z_{m+1}^2\right)\Delta W_{m+1}^1\right] \\ &= \mathbb{E}_m\left[Y_{m+1}\Delta W_{m+1}^1\right] + \Delta t(1-\theta)\mathbb{E}_m\left[f(t_{m+1}, \mathbf{\Lambda}_{m+1})\Delta W_{m+1}^1\right] \\ &- \Delta t\theta\left(Z_m^1 + \rho Z_m^2\right) - \Delta t(1-\theta)\mathbb{E}_m\left[Z_{m+1}^1 + \rho Z_{m+1}^2\right]. && (4.54)\end{aligned}$$

Analogously, we can find

$$\begin{aligned} 0 &\approx \mathbb{E}_m\left[Y_{m+1}\Delta W_{m+1}^2\right] + \Delta t(1-\theta)\mathbb{E}_m\left[f(t_{m+1}, \mathbf{\Lambda}_{m+1})\Delta W_{m+1}^2\right] \\ &- \Delta t\theta\sqrt{1-\rho^2}Z_m^2 - \Delta t(1-\theta)\sqrt{1-\rho^2}\mathbb{E}_m\left[Z_{m+1}^2\right]. && (4.55)\end{aligned}$$

[1]The conditions on the functions f and g guarantee the existence of a unique solution (Y, Z_1, Z_2) to BSDE (4.47), [24]

Using one of the approximation schemes of Section 4.1, and formulas (4.54) and (4.55) gives, for $m = M-1, ..., 0$,

$$
\begin{aligned}
Y_m^{\Delta} &= \mathbb{E}_m\left[Y_{m+1}^{\Delta}\right] + \Delta t\theta f\left(t_m, \mathbf{\Lambda}_m^{\Delta}\right) + \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\right)\right], && (4.56)\\
Z_m^{1,\Delta} &= \frac{1}{\Delta t\theta}\mathbb{E}_m\left[Y_{m+1}^{\Delta}\Delta W_{m+1}^{1}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^{1,\Delta} + \rho Z_{m+1}^{2,\Delta}\right]\\
&+ \frac{1-\theta}{\theta}\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\right)\Delta W_{m+1}^{1}\right] - \rho Z_m^{2,\Delta}, && (4.57)\\
Z_m^{2,\Delta} &= \frac{1}{\Delta t\theta\sqrt{1-\rho^2}}\mathbb{E}_m\left[Y_{m+1}^{\Delta}\Delta W_{m+1}^{2}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^{2,\Delta}\right]\\
&+ \frac{1-\theta}{\theta\sqrt{1-\rho^2}}\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\right)\Delta W_{m+1}^{2}\right], && (4.58)\\
Y_M^{\Delta} &= g\left(\mathbf{X}_M^{\Delta}\right), \qquad Z_M^{1,\Delta} = \sigma_1(\mathbf{x})\frac{\mathrm{d}g(\mathbf{x})}{\mathrm{d}x_1}\bigg|_{\mathbf{x}=\mathbf{X}_M^{\Delta}} \text{ and } Z_M^{2,\Delta} = \sigma_2(\mathbf{x})\frac{\mathrm{d}g(\mathbf{x})}{\mathrm{d}x_2}\bigg|_{\mathbf{x}=\mathbf{X}_M^{\Delta}}. && (4.59)
\end{aligned}
$$

We observe that Y_m^{Δ}, $Z_m^{1,\Delta}$ and $Z_m^{2,\Delta}$ depend on the value $\mathbf{X}_m^{\Delta}$, so when $\mathbf{X}_m^{\Delta} = \mathbf{x}$, for $m = M-1, \ldots, 0$, we have

$$
\begin{aligned}
Y_m^{\Delta}(\mathbf{x}) &= \mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right] + \Delta t\theta f\left(t_m, \mathbf{\Lambda}_m^{\Delta}(\mathbf{x})\right)\\
&+ \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right)\right], && (4.60)\\
Z_m^{1,\Delta}(\mathbf{x}) &= \frac{1}{\Delta t\theta}\mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^{1}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^{1,\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right) + \rho Z_{m+1}^{2,\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right]\\
&+ \frac{1-\theta}{\theta}\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right)\Delta W_{m+1}^{1}\right] - \rho Z_m^{2,\Delta}(\mathbf{x}), && (4.61)\\
Z_m^{2,\Delta}(\mathbf{x}) &= \frac{1}{\Delta t\theta\sqrt{1-\rho^2}}\mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^{2}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z_{m+1}^{2,\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right]\\
&+ \frac{1-\theta}{\theta\sqrt{1-\rho^2}}\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right)\Delta W_{m+1}^{2}\right], && (4.62)
\end{aligned}
$$

where $\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}$ denotes the value of $\mathbf{X}_{m+1}^{\Delta}$ given $\mathbf{X}_m^{\Delta} = \mathbf{x}$. $Y_m^{\Delta}(\mathbf{x})$ is implicit for $\theta > 0$ and can be determined by performing P Picard iterations, starting with initial guess $\mathbb{E}\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right]$. The value of the option can be approximated by $V(0,T,X_0) = Y_0^{\Delta}(X_0)$.

4.6 Approximation of the conditional expectations

For each $m \in \{M-1, \ldots, 0\}$, we wish to approximate the conditional expectations that appear in formulas (4.60), (4.61) and (4.62):

$$
\begin{aligned}
&\mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right], \quad \mathbb{E}_m\left[Z_{m+1}^{1,\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right], \quad \mathbb{E}_m\left[Z_{m+1}^{2,\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right],\\
&\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right)\right], \quad \mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^{1}\right], \quad \mathbb{E}_m\left[Y_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^{2}\right],\\
&\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right)\Delta W_{m+1}^{1}\right] \quad \text{and} \quad \mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}_{m+1}^{\Delta}\left(\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right)\Delta W_{m+1}^{2}\right].
\end{aligned}
$$

As in the 1D case, we generalize this by deriving an equation for the conditional expectations $\mathbb{E}\left[h\left(t_{m+1}, \mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right]$ and $\mathbb{E}\left[h\left(t_{m+1}, \mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^{j}\right]$ for a general function $h(t,\mathbf{x})$ and where $j \in \{1,2\}$.

Let

$$\mathcal{H}_{k_1,k_2}(t_{m+1}) = \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\int_{a_2}^{b_2}\int_{a_1}^{b_1} h(t_{m+1},x_1,x_2)\cos\left(k_1\pi\frac{x_1-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{x_2-a_2}{b_2-a_2}\right)\,\mathrm{d}x_1\,\mathrm{d}x_2. \tag{4.63}$$

Using the two-dimensional COS method, we find

$$\begin{aligned}
&\mathbb{E}\left[h\left(t_{m+1},\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right]\\
&\approx \sum_{k_1=0}^{N_1-1}{}' \sum_{k_2=0}^{N_2-1}{}' \frac{1}{2}\Re\left\{\phi_{\mathbf{X}_{m+1}^{\Delta}}\left(\frac{k_1\pi}{b_1-a_1},\frac{k_2\pi}{b_2-a_2}\middle|\,\mathbf{X}_m^{\Delta}=\mathbf{x}\right)\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}-ik_2\pi\frac{a_2}{b_2-a_2}\right)\right.\\
&\left.+\,\phi_{\mathbf{X}_{m+1}^{\Delta}}\left(\frac{k_1\pi}{b_1-a_1},-\frac{k_2\pi}{b_2-a_2}\middle|\,\mathbf{X}_m^{\Delta}=\mathbf{x}\right)\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}+ik_2\pi\frac{a_2}{b_2-a_2}\right)\right\}\mathcal{H}_{k_1,k_2}\left(t_{m+1}\right)
\end{aligned} \tag{4.64}$$

and

$$\begin{aligned}
&\mathbb{E}\left[h\left(t_{m+1},\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^1\right]\\
&= \sum_{k_1=0}^{N_1-1}{}' \sum_{k_2=0}^{N_2-1}{}' \frac{1}{2}\Re\left\{\mathbb{E}\left[\exp\left(ik_1\pi\frac{X_{m+1}^{1,\Delta,m,\mathbf{x}}}{b_1-a_1}+ik_2\pi\frac{X_{m+1}^{2,\Delta,m,\mathbf{x}}}{b_2-a_2}\right)\Delta W_{m+1}^1\right]\right.\\
&\cdot\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}-ik_2\pi\frac{a_2}{b_2-a_2}\right)+\mathbb{E}\left[\exp\left(ik_1\pi\frac{X_{m+1}^{1,\Delta,m,\mathbf{x}}}{b_1-a_1}-ik_2\pi\frac{X_{m+1}^{2,\Delta,m,\mathbf{x}}}{b_2-a_2}\right)\Delta W_{m+1}^1\right]\\
&\left.\cdot\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1}+ik_2\pi\frac{a_2}{b_2-a_2}\right)\right\}\mathcal{H}_{k_1,k_2}\left(t_{m+1}\right).
\end{aligned} \tag{4.65}$$

For convenience, we abbreviate the following expression

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(ik_1\pi\frac{X_{m+1}^{1,\Delta,m,\mathbf{x}}}{b_1-a_1}+ik_2\pi\frac{X_{m+1}^{2,\Delta,m,\mathbf{x}}}{b_2-a_2}\right)\Delta W_{m+1}^1\right]\\
&\quad= \mathbb{E}\left[\exp\left(i\left(d_1+d_2v+d_3w+d_4vw+d_5v^2+d_6w^2\right)\right)v\right],
\end{aligned} \tag{4.66}$$

where

$$\begin{aligned}
d_1 &= \tfrac{k_1\pi}{b_1-a_1}\left(x_1+m_1(\mathbf{x})\Delta t\right)+\tfrac{k_2\pi}{b_2-a_2}\left(x_2+m_2(\mathbf{x})\Delta t\right), & d_4 &= \tfrac{k_1\pi}{b_1-a_1}\kappa_1^{W^1,W^2}(\mathbf{x})+\tfrac{k_2\pi}{b_2-a_2}\kappa_2^{W^1,W^2}(\mathbf{x}),\\
d_2 &= \tfrac{k_1\pi}{b_1-a_1}s_1^{W^1}(\mathbf{x})+\tfrac{k_2\pi}{b_2-a_2}s_2^{W^1}(\mathbf{x}), & d_5 &= \tfrac{k_1\pi}{b_1-a_1}\kappa_1^{W^1}(\mathbf{x})+\tfrac{k_2\pi}{b_2-a_2}\kappa_2^{W^1}(\mathbf{x}),\\
d_3 &= \tfrac{k_1\pi}{b_1-a_1}s_1^{W^2}(\mathbf{x})+\tfrac{k_2\pi}{b_2-a_2}s_2^{W^2}(\mathbf{x}), & d_6 &= \tfrac{k_2\pi}{b_2-a_2}\kappa_2^{W^2}(\mathbf{x}),\\
v &= \Delta W_{m+1}^1, & w &= \Delta W_{m+1}^2,
\end{aligned} \tag{4.67}$$

where v and w are uncorrelated and both are normally distributed with mean zero and variance Δt. Using integration by parts gives us

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(i\left(d_1+d_2v+d_3w+d_4vw+d_5v^2+d_6w^2\right)\right)v\right]\\
&\quad=\frac{\exp(id_1)}{2\pi\Delta t}\int_{-\infty}^{\infty}\exp(i(d_3w+d_6w^2))\exp\left(-\frac{w^2}{2\Delta t}\right)\int_{-\infty}^{\infty}\exp\left(i\left((d_2+d_4w)v+d_5v^2\right)\right)\\
&\qquad\mathrm{d}\left(-\Delta t\exp\left(-\frac{v^2}{2\Delta t}\right)\right)\mathrm{d}w\\
&\quad=\frac{\exp(id_1)}{2\pi\Delta t}\int_{-\infty}^{\infty}\exp(i(d_3w+d_6w^2))\exp\left(-\frac{w^2}{2\Delta t}\right)
\end{aligned}$$

$$
\begin{aligned}
&\cdot \left\{ -\Delta t \lim_{R\to\infty} \left[\exp\left(i\left((d_2+d_4 w)v + d_5 v^2\right)\right) \exp\left(-\frac{v^2}{2\Delta}\right)\right] \right. \\
&\left. + i\Delta t \int_{-\infty}^{\infty} (d_2 + d_4 w + 2d_5 v) \exp\left(i\left((d_2+d_4 w)v + d_5 v^2\right)\right) \exp\left(-\frac{v^2}{2\Delta t}\right) \mathrm{d}v \right\} \mathrm{d}w \\
&= \frac{\exp(id_1)}{2\pi\Delta t} \int_{-\infty}^{\infty} \exp(i(d_3 w + d_6 w^2)) \exp\left(-\frac{w^2}{2\Delta t}\right) \\
&\cdot i\Delta t \int_{-\infty}^{\infty} (d_2 + d_4 w + 2d_5 v) \exp\left(i\left((d_2+d_4 w)v + d_5 v^2\right)\right) \exp\left(-\frac{v^2}{2\Delta t}\right) \mathrm{d}v\, \mathrm{d}w \\
&= i\Delta t \mathbb{E}\left[\exp\left(i\left(d_1 + d_2 v + d_3 w + d_4 vw + d_5 v^2 + d_6 w^2\right)\right)(d_2 + d_4 w + 2d_5 v)\right]. \qquad (4.68)
\end{aligned}
$$

Analogously, we find

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(i\left(d_1 + d_2 v + d_3 w + d_4 vw + d_5 v^2 + d_6 w^2\right)\right) w\right] \\
&\quad = i\Delta t \mathbb{E}\left[\exp\left(i\left(d_1 + d_2 v + d_3 w + d_4 vw + d_5 v^2 + d_6 w^2\right)\right)(d_3 + d_4 v + 2d_6 w)\right]. \qquad (4.69)
\end{aligned}
$$

This implies

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(i\left(d_1 + d_2 v + d_3 w + d_4 vw + d_5 v^2 + d_6 w^2\right)\right) v\right] \\
&= \left(id_2\Delta t - d_3 d_4 (\Delta t)^2 - 2d_2 d_5 (\Delta t)^2\right) \mathbb{E}\left[\exp\left(i\left(d_1 + d_2 v + d_3 w + d_4 vw + d_5 v^2 + d_6 w^2\right)\right)\right] \\
&+\mathcal{O}\left((\Delta t)^3\right) \\
&= \left(id_2\Delta t - d_3 d_4 (\Delta t)^2 - 2d_2 d_5 (\Delta t)^2\right) \mathbb{E}\left[\exp\left(ik_1\pi\frac{X_{m+1}^{1,\Delta,m,\mathbf{x}}}{b_1-a_1} + ik_2\pi\frac{X_{m+1}^{2,\Delta,m,\mathbf{x}}}{b_2-a_2}\right)\right] \\
&+\mathcal{O}\left((\Delta t)^3\right). \qquad (4.70)
\end{aligned}
$$

Repeating this analysis for $\mathbb{E}\left[\exp\left(ik_1\pi\frac{X_{m+1}^{1,\Delta,m,\mathbf{x}}}{b_1-a_1} - ik_2\pi\frac{X_{m+1}^{2,\Delta,m,\mathbf{x}}}{b_2-a_2}\right) W_{m+1}^1\right]$ results in

$$
\begin{aligned}
&\mathbb{E}\left[h\left(t_{m+1}, \mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^1\right] \\
&= {\sum_{k_1=0}^{N_1-1}}' {\sum_{k_2=0}^{N_2-1}}' \frac{1}{2}\Re\left\{\left(id_2^+\Delta t - d_3^+ d_4^+ (\Delta t)^2 - 2d_2^+ d_5^+ (\Delta t)^2\right)\phi_{\mathbf{X}_{m+1}^\Delta}\left(\frac{k_1\pi}{b_1-a_1}, \frac{k_2\pi}{b_2-a_2}\middle| \mathbf{X}_m^\Delta = \mathbf{x}\right)\right. \\
&\cdot \exp\left(-ik_1\pi\frac{a_1}{b_1-a_1} - ik_2\pi\frac{a_2}{b_2-a_2}\right) + \left(id_2^-\Delta t - d_3^- d_4^- (\Delta t)^2 - 2d_2^- d_5^- (\Delta t)^2\right) \qquad (4.71) \\
&\left.\cdot \phi_{\mathbf{X}_{m+1}^\Delta}\left(\frac{k_1\pi}{b_1-a_1}, -\frac{k_2\pi}{b_2-a_2}\middle| \mathbf{X}_m^\Delta = \mathbf{x}\right)\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1} + ik_2\pi\frac{a_2}{b_2-d_2}\right)\right\}\mathcal{H}_{k_1,k_2}(t_{m+1}),
\end{aligned}
$$

where

$$
\begin{aligned}
d_1^\pm &= \tfrac{k_1\pi}{b_1-a_1}(x_1 + m_1(\mathbf{x})\Delta t) \pm \tfrac{k_2\pi}{b_2-a_2}(x_2 + m_2(\mathbf{x})\Delta t), & d_4^\pm &= \tfrac{k_1\pi}{b_1-a_1}\kappa_1^{W^1,W^2}(\mathbf{x}) \pm \tfrac{k_2\pi}{b_2-a_2}\kappa_2^{W^1,W^2}(\mathbf{x}), \\
d_2^\pm &= \tfrac{k_1\pi}{b_1-a_1}s_1^{W^1}(\mathbf{x}) \pm \tfrac{k_2\pi}{b_2-a_2}s_2^{W^1}(\mathbf{x}), & d_5^\pm &= \tfrac{k_1\pi}{b_1-a_1}\kappa_1^{W^1}(\mathbf{x}) \pm \tfrac{k_2\pi}{b_2-a_2}\kappa_2^{W^1}(\mathbf{x}), \\
d_3^\pm &= \tfrac{k_1\pi}{b_1-a_1}s_1^{W^2}(\mathbf{x}) \pm \tfrac{k_2\pi}{b_2-a_2}s_2^{W^2}(\mathbf{x}), & d_6^\pm &= \pm\tfrac{k_2\pi}{b_2-a_2}\kappa_2^{W^2}(\mathbf{x}).
\end{aligned} \qquad (4.72)
$$

Analogously, we find

$$
\mathbb{E}\left[h\left(t_{m+1}, \mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\Delta W_{m+1}^2\right]
$$

$$= \sum_{k_1=0}^{N_1-1}{}' \sum_{k_2=0}^{N_2-1}{}' \frac{1}{2} \Re \left\{ \left(i d_3^+ \Delta t - d_2^+ d_4^+ (\Delta t)^2 - 2 d_3^+ d_6^+ (\Delta t)^2 \right) \phi_{\mathbf{X}_{m+1}^\Delta} \left(\frac{k_1 \pi}{b_1 - a_1}, \frac{k_2 \pi}{b_2 - a_2} \middle| \mathbf{X}_m^\Delta = \mathbf{x} \right) \right.$$
$$\cdot \exp\left(-i k_1 \pi \frac{a_1}{b_1 - a_1} - i k_2 \pi \frac{a_2}{b_2 - a_2} \right) + \left(i d_3^- \Delta t - d_2^- d_4^- (\Delta t)^2 - 2 d_3^- d_6^- (\Delta t)^2 \right) \tag{4.73}$$
$$\left. \cdot \phi_{\mathbf{X}_{m+1}^\Delta} \left(\frac{k_1 \pi}{b_1 - a_1}, -\frac{k_2 \pi}{b_2 - a_2} \middle| \mathbf{X}_m^\Delta = \mathbf{x} \right) \exp\left(-i k_1 \pi \frac{a_1}{b_1 - a_1} + i k_2 \pi \frac{a_2}{b_2 - a_2} \right) \right\} \mathcal{H}_{k_1,k_2}(t_{m+1}).$$

Let $\mathcal{Y}_{k_1,k_2}^\Delta(t_m)$, $\mathcal{Z}_{k_1,k_2}^{1,\Delta}(t_m)$, $\mathcal{Z}_{k_1,k_2}^{2,\Delta}(t_m)$ and $\mathcal{F}_{k_1,k_2}^\Delta(t_m)$ denote the Fourier-cosine coefficients of respectively $Y_m^\Delta(\mathbf{x})$, $Z_m^{1,\Delta}(\mathbf{x})$, $Z_m^{2,\Delta}(\mathbf{x})$ and $f\left(t_m, \mathbf{\Lambda}_m^\Delta(\mathbf{x})\right)$, i.e.

$$\mathcal{Y}_{k_1,k_2}^\Delta(t_m) = \frac{2}{b_1 - a_1} \frac{2}{b_2 - a_2} \int_{a_2}^{b_2} \int_{a_1}^{b_1} Y_m^\Delta(x_1, x_2) \cos\left(k_1 \pi \frac{x_1 - a_1}{b_1 - a_1} \right) \cos\left(k_2 \pi \frac{x_2 - a_2}{b_2 - a_2} \right) \mathrm{d}x_1 \, \mathrm{d}x_2, \tag{4.74}$$

$$\mathcal{Z}_{k_1,k_2}^{1,\Delta}(t_m) = \frac{2}{b_1 - a_1} \frac{2}{b_2 - a_2} \int_{a_2}^{b_2} \int_{a_1}^{b_1} Z_m^{1,\Delta}(x_1, x_2) \cos\left(k_1 \pi \frac{x_1 - a_1}{b_1 - a_1} \right) \cos\left(k_2 \pi \frac{x_2 - a_2}{b_2 - a_2} \right) \mathrm{d}x_1 \, \mathrm{d}x_2, \tag{4.75}$$

$$\mathcal{Z}_{k_1,k_2}^{2,\Delta}(t_m) = \frac{2}{b_1 - a_1} \frac{2}{b_2 - a_2} \int_{a_2}^{b_2} \int_{a_1}^{b_1} Z_m^{2,\Delta}(x_1, x_2) \cos\left(k_1 \pi \frac{x_1 - a_1}{b_1 - a_1} \right) \cos\left(k_2 \pi \frac{x_2 - a_2}{b_2 - a_2} \right) \mathrm{d}x_1 \, \mathrm{d}x_2, \tag{4.76}$$

$$\mathcal{F}_{k_1,k_2}^\Delta(t_m) = \frac{2}{b_1 - a_1} \frac{2}{b_2 - a_2} \int_{a_2}^{b_2} \int_{a_1}^{b_1} f\left(t_m, \mathbf{\Lambda}_m^\Delta(x_1, x_2)\right) \cos\left(k_1 \pi \frac{x_1 - a_1}{b_1 - a_1} \right) \cos\left(k_2 \pi \frac{x_2 - a_2}{b_2 - a_2} \right) \mathrm{d}x_1 \, \mathrm{d}x_2, \tag{4.77}$$

and at time of maturity T, we have
$Y_M^\Delta(x_1, x_2) = g(x_1, x_2)$, $Z_M^{1,\Delta}(x_1, x_2) = \sigma_1(x_1, x_2) \frac{\mathrm{d}g(x_1,x_2)}{\mathrm{d}x_1}$, $Z_M^{2,\Delta}(x_1, x_2) = \sigma_2(x_1, x_2) \frac{\mathrm{d}g(x_1,x_2)}{\mathrm{d}x_2}$, and $f\left(t_M, \mathbf{\Lambda}_M^\Delta(x_1, x_2)\right) = f\left(t_M, x_1, x_2, g(x_1, x_2), \sigma_1 \frac{\mathrm{d}g(x_1,x_2)}{\mathrm{d}x_1}, \sigma_2 \frac{\mathrm{d}g(x_1,x_2)}{\mathrm{d}x_2}\right)$. When the above integrals cannot be computed analytically, we can approximate them by computing the function on an $\mathbf{x}$-grid and using the two-dimensional discrete Fourier-cosine transform or another numerical integration method.

4.7 Two-dimensional BCOS method summarized

We define a time-grid $t_j = j\Delta t$ for $j = 0, 1, \ldots M$, with fixed time steps $\Delta t = \frac{T}{M}$. Also, we define function f by formula (4.48).

Initial step: Compute the terminal coefficients $\mathcal{Y}_{k_1,k_2}^\Delta(t_M)$, $\mathcal{Z}_{k_1,k_2}^{1,\Delta}(t_M)$, $\mathcal{Z}_{k_1,k_2}^{2,\Delta}(t_M)$ and $\mathcal{F}_{k_1,k_2}^\Delta(t_M)$ with formulas (4.74), (4.75), (4.76) and (4.77).

Loop: For $m = M - 1$ to $m = 1$
approximate the necessary conditional expectations with Section 4.6 and compute functions $Y_m^\Delta(\mathbf{x})$, $Z_m^{1,\Delta}(\mathbf{x})$, $Z_m^{2,\Delta}(\mathbf{x})$ and $f\left(t_m, t_m, \mathbf{\Lambda}_m^\Delta(\mathbf{x})\right)$ for $\mathbf{x} \in [a_1, b_1] \times [a_2, b_2]$ with formulas (4.60), (4.61) and (4.62).

Thereafter, compute/approximate the corresponding Fourier-cosine coefficients $\mathcal{Y}^{\Delta}_{k_1,k_2}(t_m)$, $\mathcal{Z}^{1,\Delta}_{k_1,k_2}(t_m)$, $\mathcal{Z}^{2,\Delta}_{k_1,k_2}(t_m)$ and $\mathcal{F}^{\Delta}_{k_1,k_2}(t_m)$ with formulas (4.74), (4.75), (4.76) and (4.77). Those integrals can be approximated by computing the function on an $\mathbf{x}$-grid and using the two-dimensional discrete Fourier-cosine transform or another numerical integration method.

Terminal step: Compute $Y^{\Delta}_0(\mathbf{X}_0)$, $Z^{1,\Delta}_0(\mathbf{X}_0)$ and $Z^{2,\Delta}_0(\mathbf{X}_0)$.

We approximate the value $V(0,T,\mathbf{X}_0)$ of the option by $V(0,T,\mathbf{X}_0) = Y^{\Delta}_0(\mathbf{X}_0)$ and the corresponding Deltas by

$$\Delta^1_0 = \left.\frac{\partial V(0,T,\mathbf{x})}{\partial x_1}\right|_{\mathbf{x}=\mathbf{X}_0} = \frac{Z^{1,\Delta}_0(\mathbf{X}_0)}{\sigma_1(\mathbf{X}_0)} \quad \text{and} \quad \Delta^2_0 = \left.\frac{\partial V(0,T,\mathbf{x})}{\partial x_2}\right|_{\mathbf{x}=\mathbf{X}_0} = \frac{Z^{2,\Delta}_0(\mathbf{X}_0)}{\sigma_2(\mathbf{X}_0)}, \tag{4.78}$$

which corresponds to the amount of assets in the self-financing portfolio Y_t at time $t=0$.

4.8 Path-dependent options

In this section we describe the change in procedure of the BCOS method for pricing some path-dependent options. Section 4.8.1 describes the procedure for a Bermudan option and Section 4.8.2 gives the procedure for a discretely monitored barrier option.

4.8.1 Bermudan option

In section 3.5, we described the change in the procedure for pricing a Bermudan option with the one-dimensional BCOS method. These changes are similar for the two-dimensional BCOS method. As mentioned, a Bermudan option can be exercised at predetermined dates. Let n be the number of early-exercise dates and let τ_j denotes the early-exercise dates for $j = 1,2,...,n$, where $0 \leq \tau_1 < \tau_2 < ... < \tau_n = T$. We choose Δt such that each of the early-exercise dates corresponds to a point in our time-grid[2]. We replace formulas (4.60), (4.61) and (4.62) by, respectively,

$$Y^{\Delta}_m(\mathbf{x}) = \begin{cases} \max\left\{g(\mathbf{x}), \widetilde{Y}^{\Delta}_m(\mathbf{x})\right\}, & \text{for } t_m = \tau_j, \\ \widetilde{Y}^{\Delta}_m(\mathbf{x}), & \text{for } t_m \neq \tau_j, \end{cases} \tag{4.79}$$

$$Z^{1,\Delta}_m(\mathbf{x}) = \begin{cases} \sigma_1(\mathbf{x})\frac{\partial g(\mathbf{x})}{\partial x_1}, & \text{for } t_m = \tau_j,\ Y^{\Delta}_m(\mathbf{x}) = g(\mathbf{x}), \\ \widetilde{Z}^{1,\Delta}_m(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.80}$$

$$Z^{2,\Delta}_m(\mathbf{x}) = \begin{cases} \sigma_2(\mathbf{x})\frac{\partial g(\mathbf{x})}{\partial x_2}, & \text{for } t_m = \tau_j,\ Y^{\Delta}_m(\mathbf{x}) = g(\mathbf{x}), \\ \widetilde{Z}^{2,\Delta}_m(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.81}$$

where

$$\begin{aligned} \widetilde{Y}^{\Delta}_m(\mathbf{x}) &= \mathbb{E}_m\left[Y^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right] + \Delta t\theta f\left(t_m, \mathbf{\Lambda}^{\Delta}_m(\mathbf{x})\right) \\ &+ \Delta t(1-\theta)\mathbb{E}_m\left[f\left(t_{m+1}, \mathbf{\Lambda}^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right)\right], \end{aligned} \tag{4.82}$$

$$\widetilde{Z}^{1,\Delta}_m(\mathbf{x}) = \frac{1}{\Delta t\theta}\mathbb{E}_m\left[Y^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\Delta W^1_{m+1}\right] - \frac{1-\theta}{\theta}\mathbb{E}_m\left[Z^{1,\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right) + \rho Z^{2,\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right]$$

[2] If desired, it is possible to choose Δt non-constant

$$+ \quad \frac{1-\theta}{\theta}\mathbb{E}_m\left[f\left(t_{m+1},\mathbf{\Lambda}^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right)\Delta W^1_{m+1}\right]-\rho Z^{2,\Delta}_m(\mathbf{x}), \tag{4.83}$$

$$\widetilde{Z}^{2,\Delta}_m(\mathbf{x}) = \frac{1}{\Delta t\theta\sqrt{1-\rho^2}}\mathbb{E}_m\left[Y^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\Delta W^2_{m+1}\right]-\frac{1-\theta}{\theta}\mathbb{E}_m\left[Z^{2,\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right]$$

$$+ \quad \frac{1-\theta}{\theta\sqrt{1-\rho^2}}\mathbb{E}_m\left[f\left(t_{m+1},\mathbf{\Lambda}^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right)\Delta W^2_{m+1}\right]. \tag{4.84}$$

4.8.2 Discretely monitored barrier options

The value of a barrier option depends on the value of the underlying at the observation dates. Whether, or not the value of the underlying hits the predetermined barrier at one of the observation dates determines the value of the barrier option. A knock-in barrier option has no value before the underlying hits the barrier, while a knock-out barrier option has no value after the underlying hits the barrier. In this section we consider barrier options of European and Bermudan style. The early-exercise dates of the Bermudan option can be different from the observation dates. Without loss of generality, we consider that the barrier B is for the underlying X^1_t. Let n_1 be the number of observation dates and let τ_j denotes the observation dates for $j = 1, 2, ..., n_1$, where $0 \le \tau_1 < \tau_2 < ... < \tau_{n_1} \le T$. Let n_2 denote the number of early-exercise dates and let ζ_i denote the early-exercise dates for $i = 1, ..., n_2$, where $0 \le \zeta_1 < \zeta_2 < ... < \zeta_{n_2} = T$. We choose Δt such that each of the observation dates and each of the early-exercise dates corresponds to a point in our time-grid.

For the original European or Bermudan option we obtain $\widehat{Y}^{\Delta}_m(\mathbf{x})$, $\widehat{Z}^{1,\Delta}_m(\mathbf{x})$ and $\widehat{Z}^{2,\Delta}_m(\mathbf{x})$ with the formulas (4.60), (4.61) and (4.62) if the option is European style, or with formulas (4.79), (4.80) and (4.81) if the option is Bermudan style. Also, define $\widetilde{Y}^{\Delta}_m(\mathbf{x})$, $\widetilde{Z}^{1,\Delta}_m(\mathbf{x})$ and $\widetilde{Z}^{2,\Delta}_m(\mathbf{x})$ the same as in section 4.8.1.

Knock-out barrier option
For an up-and-out option we define the event $E_t = \left\{X^1_t \ge B\right\}$ and for an down-and-out option we define the event $E_t = \left\{X^1_t \le B\right\}$. We replace formulas (4.60), (4.61) and (4.62) by, respectively,

$$Y^{\Delta}_m(\mathbf{x}) = \begin{cases} 0, & \text{for } t_m = \tau_j, E_{t_m}, \\ \max\left\{g(\mathbf{x}), \widetilde{Y}^{\Delta}_m(\mathbf{x})\right\}, & \text{for } t_m = \zeta_i, \neg\left(t_m = \tau_j, E_{t_m}\right), \\ \widetilde{Y}^{\Delta}_m(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.85}$$

$$Z^{1,\Delta}_m(\mathbf{x}) = \begin{cases} 0, & \text{for } t_m = \tau_j, E_{t_m}, \\ \sigma_1(\mathbf{x})\frac{\partial g(\mathbf{x})}{\partial x_1}, & \text{for } t_m = \zeta_i,\ Y^{\Delta}_m(\mathbf{x}) = g(\mathbf{x}), \\ \widetilde{Z}^{1,\Delta}_m(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.86}$$

$$Z^{2,\Delta}_m(\mathbf{x}) = \begin{cases} 0, & \text{for } t_m = \tau_j, E_{t_m}, \\ \sigma_2(\mathbf{x})\frac{\partial g(\mathbf{x})}{\partial x_2}, & \text{for } t_m = \zeta_i,\ Y^{\Delta}_m(\mathbf{x}) = g(\mathbf{x}), \\ \widetilde{Z}^{2,\Delta}_m(\mathbf{x}), & \text{otherwise}. \end{cases} \tag{4.87}$$

So, at each observation date we observe if the underlying hits the barrier or not.
Knock-in barrier option
For an up-and-in option we define the event $E_t = \left\{X^1_t \ge B\right\}$ and for an down-and-in option we define the event $E_t = \left\{X^1_t \le B\right\}$. We replace formulas (4.60), (4.61) and (4.62) by, respectively,

$$Y^{\Delta}_m(\mathbf{x}) = \begin{cases} \widehat{Y}^{\Delta}_m(\mathbf{x}), & \text{for } t_m = \tau_j, E_{t_m}, \\ \widetilde{Y}^{\Delta}_m(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.88}$$

$$Z_m^{1,\Delta}(\mathbf{x}) = \begin{cases} \widehat{Z}_m^{1,\Delta}(\mathbf{x}), & \text{for } t_m = \tau_j, E_{t_m}, \\ \widetilde{Z}_m^{1,\Delta}(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.89}$$

$$Z_m^{2,\Delta}(\mathbf{x}) = \begin{cases} \widehat{Z}_m^{2,\Delta}(\mathbf{x}), & \text{for } t_m = \tau_j, E_{t_m}, \\ \widetilde{Z}_m^{2,\Delta}(\mathbf{x}), & \text{otherwise}, \end{cases} \tag{4.90}$$

where $Y_M^\Delta(\mathbf{X}_M^\Delta) = Z_M^{1,\Delta}(\mathbf{X}_M^\Delta) = Z_M^{2,\Delta}(\mathbf{X}_M^\Delta) = 0$ if $\tau_{n_1} < T$.

So we have two BSDEs, one for the value of the original European or Bermudan option, and one represents the value of the barrier option. At each observation date we observe if the underlying hits the barrier or not. The value of the barrier option equals the value of the original option at an observation date, when the underlying hits the barrier.

4.9 Error analysis

Analogously to the analysis in Section 3.6, we perform an error analysis for the two-dimensional BCOS method under the risk-neutral $\mathbb{Q}$-measure. Again, we only analyze the error in the option price Y_m and not the error in the Deltas Z_m^1 and Z_m^2. Therefore, we abbreviate $f(t, x_1, x_2, y, z_1, z_2)$ by $f(t, x_1, x_2, y)$. The error analysis for the two-dimensional BCOS-method is not significantly more challenging then the analysis for the one-dimensional method.

The error in option value Y_m consists of four different components, namely the error as a result of: the Δ-time-discretization, the θ-method-discretization, the COS method, and the Picard iterations. Like in Section 3.6, we perform the error analysis for the 2.0-weak-Taylor scheme scheme and parameter $\theta = \frac{1}{2}$. The analysis of the Fourier errors introduced by the two-dimensional COS method [23] and the Picard iterations [4, 25] is analogously to Section 3.6.3. In Sections 4.9.1 and 4.9.2, we discuss the local errors as a result of the Δ-time-discretization and the θ-method-discretization respectively. Finally, we look at the global error in Section 4.9.3. For the error analysis we use the notation and results of Appendix C.

4.9.1 Local error Δ-time-discretization

Let $\mathbf{X}_{m+1}^{m,\mathbf{x}}$ and $\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}$ denote respectively the values of $\mathbf{X}_{m+1}$ and $\mathbf{X}_{m+1}^{\Delta}$ given $\mathbf{X}_m = \mathbf{x} = (x_1, x_2)$ and where $\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}$ is derived with the 2.0-weak-Taylor scheme.

Lemma 4.2. *For a sufficiently smooth function $h(t, \mathbf{x})$ we have the following local weak error*

$$\mathbb{E}\left[h\left(t_{m+1}, \mathbf{X}_{m+1}^{m,\mathbf{x}}\right) - h\left(t_{m+1}, \mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right] = \mathcal{O}\left((\Delta t)^3\right). \tag{4.91}$$

Proof. For $j = 1, 2$, the 2.0-weak-Taylor scheme corresponds to

$$\begin{aligned} X_{m+1}^{j,\Delta,m,\mathbf{x}} &= \sum_{\widehat{\alpha}\in\mathcal{A}} c_{\widehat{\alpha}}(x_j) I_{\widehat{\alpha},t_m,t_{m+1}} + c_{(0,1)}(x_j)\left(\frac{1}{2}\Delta W_{m+1}^1 \Delta t - I_{(0,1),t_m,t_{m+1}}\right) \\ &+ c_{(1,0)}(x_j)\left(\frac{1}{2}\Delta W_{m+1}^1 \Delta t - I_{(1,0),t_m,t_{m+1}}\right) + c_{(0,2)}(x_j)\left(\frac{1}{2}\Delta W_{m+1}^2 \Delta t - I_{(0,2),t_m,t_{m+1}}\right) \\ &+ c_{(2,0)}(x_j)\left(\frac{1}{2}\Delta W_{m+1}^2 \Delta t - I_{(2,0),t_m,t_{m+1}}\right) \\ &+ c_{(1,2)}(x_j)\left(\frac{1}{2}\Delta W_{m+1}^1 \Delta W_{m+1}^2 + V_{m+1}^{1,2} - I_{(1,2),t_m,t_{m+1}}\right) \end{aligned}$$

$$+ \quad c_{(2,1)}(x_j)\left(\frac{1}{2}\Delta W^1_{m+1}\Delta W^2_{m+1} - V^{1,2}_{m+1} - I_{(2,1),t_m,t_{m+1}}\right), \tag{4.92}$$

where $c(x_j) = x_j$ and $\mathcal{A} = \{\widehat{\alpha} \in \mathcal{M} : l(\widehat{\alpha}) \leq 2\}$. Now, Lemma C.4 gives for all $l_1, l_2 \in \mathbb{Z}_{\geq 0}$

$$\mathbb{E}\left[\left(X^{1,m,\mathbf{x}}_{m+1} - X^{1,\Delta,m,\mathbf{x}}_{m+1}\right)^{l_1}\left(X^{2,m,\mathbf{x}}_{m+1} - X^{2,\Delta,m,\mathbf{x}}_{m+1}\right)^{l_2}\right] = \mathcal{O}\left((\Delta t)^3\right). \tag{4.93}$$

The Taylor series expansion of function h around $\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1} = \mathbf{X}^{m,\mathbf{x}}_{m+1}$ gives

$$\begin{aligned} h\left(t_{m+1}, \mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right) &= \sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\left\{\frac{(-1)^{(l_1+l_2)}}{l_1! l_2!}\left(X^{1,m,\mathbf{x}}_{m+1} - X^{1,\Delta,m,\mathbf{x}}_{m+1}\right)^{l_1}\left(X^{2,m,\mathbf{x}}_{m+1} - X^{2,\Delta,m,\mathbf{x}}_{m+1}\right)^{l_2}\right. \\ &\quad \left. \cdot \frac{\partial^{l_1+l_2}}{\left(\partial X^{1,m,\mathbf{x}}_{m+1}\right)^{l_1}\left(\partial X^{2,m,\mathbf{x}}_{m+1}\right)^{l_2}} h\left(t_{m+1}, \mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right\}. \end{aligned} \tag{4.94}$$

Combining formulas (4.94) and (4.93) implies

$$\mathbb{E}\left[h\left(t_{m+1}, \mathbf{X}^{m,\mathbf{x}}_{m+1}\right) - h\left(t_{m+1}, \mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right] = \mathcal{O}\left((\Delta t)^3\right). \tag{4.95}$$

□

4.9.2 Local error θ-method-discretization

We observe from formula (4.51) that, given $\mathbf{X}_m = \mathbf{x} = (x_1, x_2)$,

$$\begin{aligned} Y_m(\mathbf{x}) &= \mathbb{E}_m\left[Y_{m+1}\left(\mathbf{X}_{m+1}\right)\right] + \int_{t_m}^{t_{m+1}} \mathbb{E}_m\left[f\left(t, X^1_t, X^2_t, Y_t\left(\mathbf{X}_t\right)\right)\right] \mathrm{d}t \\ &= \mathbb{E}_m\left[Y_{m+1}\left(\mathbf{X}_{m+1}\right)\right] + \frac{1}{2}\Delta t f\left(t_m, x_1, x_2, Y_m(\mathbf{x})\right) \\ &+ \frac{1}{2}\Delta t \mathbb{E}_m\left[f\left(t_{m+1}, X^1_{m+1}, X^2_{m+1}, Y_{m+1}\left(\mathbf{X}_{m+1}\right)\right)\right] + \mathcal{R}_m(\mathbf{x}). \end{aligned} \tag{4.96}$$

So, the θ-method-discretization error is given by

$$\begin{aligned} \mathcal{R}^y_m(\mathbf{x}) &= \int_{t_m}^{t_{m+1}} \mathbb{E}_m\left[f\left(t, X^1_t, X^2_t, Y_t\left(\mathbf{X}_t\right)\right)\right] \mathrm{d}t \\ &- \frac{1}{2}\Delta t\left(f(t_m, x, Y_m(\mathbf{x})) + \mathbb{E}_m\left[f\left(t_{m+1}, X^1_{m+1}, X^2_{m+1}, Y_{m+1}\left(\mathbf{X}_{m+1}\right)\right)\right]\right). \end{aligned} \tag{4.97}$$

Lemma 4.3. *For a sufficiently smooth function $f(t, x_1, x_2, y)$ and payoff function $g(\mathbf{x})$ we have the following θ-method-discretization error*

$$\mathcal{R}^y_m(\mathbf{x}) = \mathcal{O}\left((\Delta t)^3\right). \tag{4.98}$$

Proof. For a generally smooth function $h(t, \mathbf{x})$, we find with Theorem C.1, $\mathcal{A} = \{v, (0), (1), (2)\}$ and $\overline{\mathcal{A}} = \{(0,0,0), (0,1,0), (0,2,0), (1,0,0), (2,0,0), (1,2,0), (2,1,0)\}$

$$\begin{aligned} \int_{t_m}^{t_{m+1}} h\left(t, \mathbf{X}^{m,\mathbf{x}}_t\right) \mathrm{d}t &= \int_{t_m}^{t_{m+1}} h(t_m, \mathbf{x}) + h_{(0)}(t_m, \mathbf{x}) I_{(0),t_m,t} + h_{(1)}(t_m, \mathbf{x}) I_{(1),t_m,t} \\ &+ h_{(2)}(t_m, \mathbf{x}) I_{(2),t_m,t} + \sum_{\widehat{\alpha} \in \mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., \mathbf{X}^{m,\mathbf{x}}_{.}\right)\right]_{t_m,t} \mathrm{d}t \end{aligned}$$

$$
\begin{aligned}
&= h(t_m,\mathbf{x})I_{(0),t_m,t_{m+1}} + h_{(0)}(t_m,\mathbf{x})I_{(0,0),t_m,t_{m+1}} + h_{(1)}(t_m,\mathbf{x})I_{(1,0),t_m,t_{m+1}} \\
&+ h_{(2)}(t_m,\mathbf{x})I_{(2,0),t_m,t_{m+1}} + \sum_{\widehat{\alpha}\in\overline{\mathcal{A}}} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}-}\left(.,\mathbf{X}_{.}^{m,\mathbf{x}}\right)\right]_{t_m,t_{m+1}}. \qquad (4.99)
\end{aligned}
$$

This implies

$$
\begin{aligned}
&\mathbb{E}\left[\int_{t_m}^{t_{m+1}} h\left(t,\mathbf{X}_t^{m,\mathbf{x}}\right)\mathrm{d}t - \frac{1}{2}\Delta t\left(h(t_m,\mathbf{x}) + h\left(t_{m+1},\mathbf{X}_{m+1}^{m,\mathbf{x}}\right)\right)\right] \\
&= h(t_m,\mathbf{x})\Delta t + \frac{1}{2}h_{(0)}(t_m,\mathbf{x})(\Delta t)^2 + \mathcal{O}\left((\Delta t)^3\right) \\
&- \frac{1}{2}\Delta t\left(h(t_m,\mathbf{x}) + h(t_m,\mathbf{x}) + h_{(0)}(t_m,\mathbf{x})\Delta t + \mathcal{O}\left((\Delta t)^2\right)\right) \\
&= \mathcal{O}\left((\Delta t)^3\right). \qquad (4.100)
\end{aligned}
$$

□

4.9.3 Global error

In this section we detail the global error of the two-dimensional BCOS method, where we omit the errors introduced by the 2D COS method and by the Picard iterations. We define

$$
\begin{aligned}
e_m^y\left(\mathbf{X}_m,\mathbf{X}_m^\Delta\right) &= Y_m(\mathbf{X}_m) - Y_m^\Delta\left(\mathbf{X}_m^\Delta\right), \qquad (4.101)\\
e_m^f\left(\mathbf{X}_m,\mathbf{X}_m^\Delta\right) &= f(t_m,X_m^1,X_m^2,Y_m(\mathbf{X}_m)) - f\left(t_m,X_m^{1,\Delta},X_m^{2,\Delta},Y_m^\Delta\left(\mathbf{X}_m^\Delta\right)\right), \qquad (4.102)
\end{aligned}
$$

and

$$
\begin{aligned}
e_m^y(\mathbf{x}) &= Y_m(\mathbf{x}) - Y_m^\Delta(\mathbf{x}), \qquad (4.103)\\
e_m^f(\mathbf{x}) &= f(t_m,x_1,x_2,Y_m(\mathbf{x})) - f\left(t_m,x_1,x_2,Y_m^\Delta(\mathbf{x})\right). \qquad (4.104)
\end{aligned}
$$

We rewrite the error $e_m^y\left(\mathbf{X}_m,\mathbf{X}_m^\Delta\right)$ to

$$
\begin{aligned}
e_m^y\left(\mathbf{X}_m,\mathbf{X}_m^\Delta\right) &= Y_m(\mathbf{X}_m) - Y_m^\Delta(\mathbf{X}_m) + Y_m^\Delta(\mathbf{X}_m) - Y_m^\Delta\left(\mathbf{X}_m^\Delta\right) \\
&= e_m^y(\mathbf{X}_m) + Y_m^\Delta(\mathbf{X}_m) - Y_m^\Delta\left(\mathbf{X}_m^\Delta\right). \qquad (4.105)
\end{aligned}
$$

The triangle inequality implies

$$
\left|\mathbb{E}_0\left[e_m^y\left(\mathbf{X}_m,\mathbf{X}_m^\Delta\right)\right]\right| \le \mathbb{E}_0\left[\left|e_m^y(\mathbf{X}_m)\right|\right] + \left|\mathbb{E}_0\left[Y_m^\Delta(\mathbf{X}_m) - Y_m^\Delta\left(\mathbf{X}_m^\Delta\right)\right]\right|. \qquad (4.106)
$$

If $Y_m(\mathbf{x})$ is a sufficiently smooth function, we find with Lemma 4.2,

$$
\left|\mathbb{E}_0\left[e_m^y\left(\mathbf{X}_m,\mathbf{X}_m^\Delta\right)\right]\right| \le \mathbb{E}_0\left[\left|e_m^y(\mathbf{X}_m)\right|\right] + \mathcal{O}\left((\Delta t)^3\right). \qquad (4.107)
$$

The following Lemma gives a bound on the error $e_m^y(\mathbf{X}_m)$.

Lemma 4.4.

$$
\mathbb{E}_0\left[\left|e_m^y(\mathbf{X}_m)\right|\right] = \mathcal{O}\left((\Delta t)^2\right). \qquad (4.108)
$$

Proof. Formulas (4.60) and (4.96) give us

$$
\begin{aligned}
e_m^y(\mathbf{x}) &= Y_m(\mathbf{x}) - Y_m^\Delta(\mathbf{x}) \\
&= \mathbb{E}\left[e_{m+1}^y\left(\mathbf{X}_{m+1}^{m,\mathbf{x}},\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right] + \frac{1}{2}\Delta t\left(e_m^f(\mathbf{x}) + \mathbb{E}\left[e_{m+1}^f\left(\mathbf{X}_{m+1}^{m,\mathbf{x}},\mathbf{X}_{m+1}^{\Delta,m,\mathbf{x}}\right)\right]\right)
\end{aligned}
$$

$$+ \quad \mathcal{R}^y_m(\mathbf{x}). \tag{4.109}$$

With Lemma 4.2, we find

$$\begin{aligned} \mathbb{E}\left[e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}, \mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right] &= \mathbb{E}\left[Y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right) - Y^{\Delta}_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right] \\ &+ \mathbb{E}\left[Y^{\Delta}_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right) - Y^{\Delta}_{m+1}\left(\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right] \\ &= \mathbb{E}\left[e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right] + \mathcal{O}\left((\Delta t)^3\right) \end{aligned} \tag{4.110}$$

and analogously

$$\mathbb{E}\left[e^f_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}, \mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right] = \mathbb{E}\left[e^f_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right] + \mathcal{O}\left((\Delta t)^3\right). \tag{4.111}$$

The function $f(t, x_1, x_2, y)$ is Lipschitz in y with Lipschitz constant L_f, which gives

$$\left|e^f_m(\mathbf{x})\right| \leq L_f \left|e^y_m(\mathbf{x})\right|, \tag{4.112}$$

$$\mathbb{E}\left[\left|e^f_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right|\right] \leq L_f \mathbb{E}\left[\left|e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right|\right]. \tag{4.113}$$

Then, combining all results and Lemma 4.3 gives us:

$$\begin{aligned} |e^y_m(\mathbf{x})| &\leq \mathbb{E}\left[\left|e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}, \mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right|\right] + \frac{1}{2}\Delta t\left(\left|e^f_m(\mathbf{x})\right| + \mathbb{E}\left[\left|e^f_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}, \mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}\right)\right|\right]\right) \\ &+ |\mathcal{R}^y_m(\mathbf{x})| \\ &\leq \mathbb{E}\left[\left|e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right|\right] + \frac{1}{2}\Delta t L_f\left(|e^y_m(\mathbf{x})| + \mathbb{E}\left[\left|e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right|\right]\right) + \mathcal{O}\left((\Delta t)^3\right). \end{aligned} \tag{4.114}$$

This implies for, $\Delta t \leq 2/L_f$,

$$|e^y_m(\mathbf{x})| \leq \frac{1 + \frac{1}{2}\Delta t L_f}{1 - \frac{1}{2}\Delta t L_f}\mathbb{E}\left[\left|e^y_{m+1}\left(\mathbf{X}^{m,\mathbf{x}}_{m+1}\right)\right|\right] + \mathcal{O}\left((\Delta t)^2\right). \tag{4.115}$$

Since $\frac{1+\frac{1}{2}\Delta t L_f}{1-\frac{1}{2}\Delta t L_f} = 1 + \mathcal{O}(\Delta t)$ and $\mathbb{E}\left[\left|e^y_M\left(\mathbf{X}^{m,\mathbf{x}}_M\right)\right|\right] = 0$, we can observe by iterating formula (4.115) that

$$|e^y_m(\mathbf{x})| = \mathcal{O}\left((\Delta t)^2\right). \tag{4.116}$$

□

Substitution of Lemma 4.4 into formula (4.107) gives

$$\left|\mathbb{E}_0\left[e^y_m\left(\mathbf{X}_m, \mathbf{X}^{\Delta}_m\right)\right]\right| = \mathcal{O}\left((\Delta t)^2\right). \tag{4.117}$$

Remark 4.2. *We find first order convergence by performing a similar error analysis for the Euler, Milstein and adjusted-Predictor-Corrector schemes for* $\theta = \frac{1}{2}$.

4.10 Example

In this section we give an example in which we determine the value of a European call option under the SABR model. In the example we price under the $\mathbb{Q}$-measure, which simplifies the two-dimensional BCOS method. We set $N_1 = N_2 = N$ and we also choose N as the size of the $\mathbf{x}$-grid in both dimensions. Independent of the choice of the Taylor scheme, we define the domain $[a_1, b_1] \times [a_2, b_2]$ by

$$[a_1,b_1]\times[a_2,b_2] = \left[c_1^1 - L\sqrt{c_2^1 + \sqrt{c_4^1}}, c_1^1 + L\sqrt{c_2^1 + \sqrt{c_4^1}}\right] \times \left[c_1^2 - L\sqrt{c_2^2 + \sqrt{c_4^2}}, c_1^2 + L\sqrt{c_2^2 + \sqrt{c_4^2}}\right], \tag{4.118}$$

where $L = 10$ and c_n^j is the n-th cumulant of $X_T^{j,\Delta}$ with only one Euler step [9, 24]. There holds $c_1^1 = X_0^1 + \mu_1(\mathbf{X}_0)T$, $c_2^1 = \sigma_1^2(\mathbf{X}_0)T$, $c_4^1 = 0$, $c_1^2 = X_0^2 + \mu_2(\mathbf{X}_0)T$, $c_2^2 = \sigma_2^2(\mathbf{X}_0)T$ and $c_4^2 = 0$.

Example 2 We use the two-dimensional BCOS method to detemine the value of a European call option under the SABR model. Remember from formulas (2.1) and (2.2) that

$$\begin{aligned} \mathrm{d}F_t &= \sigma_t (F_t)^\beta \,\mathrm{d}W_t^1, & F_0 = f_0, \\ \mathrm{d}\sigma_t &= \rho\nu\sigma_t \,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t \,\mathrm{d}W_t^2, & \sigma_0 = \alpha, \end{aligned}$$

where W^1 and W^2 are uncorrelated Brownian motions. This system of FSDEs corresponds to formulas (4.47) and (4.48), so we can use the two-dimensional BCOS method to derive the value of the option. The payoff function of this option is given by

$$g(X_T) = \max(X_T - K, 0). \tag{4.119}$$

We work under the risk-neutral measure, so $f(t, x_1, x_2, y, z_1, z_2) = -ry$ is independent of z_1 and z_2. Therefore, we do not have to calculate the functions $Z_m^{1,\Delta}(\mathbf{x})$ and $Z_m^{2,\Delta}(\mathbf{x})$ and we are allowed to use $\theta = 0$.

We have the following parameter values:

$$f_0 = 2,\ K = 1.9,\ T = 0.5,\ \alpha = 0.2,\ \beta = 0.5,\ \rho = 0,\ \nu = 0.3,\ r = 0,\ N = 2^7.$$

We use $[a_1, b_1] \times [a_2, b_2] \approx [0, 4] \times [-0.2243, 0.6243]$ and we approximate each Fourier-cosine coefficient by the two-dimensional discrete Fourier-cosine transform. For the adjusted-Predictor-Corrector scheme we take $\eta_1 = \theta_1 = \eta_2 = \theta_2 = \frac{1}{2}$. Using Antonov's method, we find the value of the call option $V(0, T, f_0, \alpha) = 0.13903754$, which is our reference value. The results of the 2D BCOS method are shown in Table 4.1.

Number of time steps M	1	5	10	50	100
Absolute error	5.2178e-04	1.5384e-04	7.9643e-05	1.5932e-05	7.7008e-06

Table 4.1: Absolute error using 2D BCOS method where the FSDEs are discretized with the Euler scheme.

When we consider different values for β and different discretization schemes, we find the results in Table 4.2. We observe that the 2.0-weak-Taylor scheme performs the best for this example. Even for only one time step, the results for all schemes are satisfactory.

	Euler			Milstein		
M	1	5	10	1	5	10
$\beta = 0.1$	2.0739e-04	1.8461e-05	5.2542e-06	6.8269e-05	9.3294e-06	6.8178e-06
$\beta = 0.3$	1.0261e-04	5.6768e-05	3.4169e-05	1.1306e-04	1.6551e-05	1.1381e-05
$\beta = 0.5$	5.2178e-04	1.5384e-04	7.9643e-05	1.6382e-04	1.4253e-05	5.5868e-06
$\beta = 0.7$	1.0979e-03	3.1321e-04	1.6478e-04	2.6224e-04	2.6913e-05	1.4108e-05
$\beta = 0.9$	1.8368e-03	5.4024e-04	2.8307e-04	4.1953e-04	3.0748e-05	1.3998e-05
	2.0-weak-Taylor			adjusted-Predictor-Corrector		
M	1	5	10	1	5	10
$\beta = 0.1$	6.2035e-05	7.9196e-06	6.1018e-06	2.0874e-04	4.6906e-05	2.1370e-05
$\beta = 0.3$	8.9721e-05	1.1381e-05	8.7613e-06	2.3975e-04	5.3688e-05	2.3745e-05
$\beta = 0.5$	1.1978e-04	4.6900e-06	7.5319e-07	3.0434e-04	7.8035e-05	4.0513e-05
$\beta = 0.7$	2.0416e-04	1.4538e-05	7.8685e-06	4.0929e-04	1.0576e-04	5.2183e-05
$\beta = 0.9$	3.8073e-04	2.2627e-05	9.9133e-06	6.4697e-04	1.8260e-04	9.2731e-05

Table 4.2: Absolute error using 2D BCOS method where the FSDEs are discretized with different schemes.

Chapter 5

One time step DCOS method

We wish to propose a method for calculating the forward value of European options under the SABR model under the risk neutral measure, which simplifies the 2D BCOS method significantly as we observed in example 2 of Chapter 4. The use of BSDEs, which is mainly the basis of the BCOS method, does not appear in this simplified method. Therefore we call it the *DCOS method*, where D stands for discrete. In Table 4.2, we observed promising results for the DCOS method with only one time step, which is why we propose to price European options under the SABR model with the one time step DCOS method. In this chapter, we discuss this method and its advantages and disadvantages. Section 5.1 gives the pricing formulas of this method and Section 5.2 discusses the (dis)advantages. In Section 5.3, we use the logarithmic transform of the forward before we apply the DCOS method and in Section 5.4 we discuss the accuracy. Finally, in Section 5.5 we summarize our results.

5.1 The method

The underlying system of FSDEs is given by formulas (2.1) and (2.2)

$$\begin{aligned} \mathrm{d}F_t &= \sigma_t \left(F_t\right)^{\beta} \mathrm{d}W_t^1, & F_0 &= f, \\ \mathrm{d}\sigma_t &= \rho\nu\sigma_t \,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t \,\mathrm{d}W_t^2, & \sigma_0 &= \alpha, \end{aligned}$$

where W^1 and W^2 are uncorrelated standard Brownian motions.

We work in one instead of two dimensions, when we apply the DCOS method with only one time step to calculate the call price of a European option under the SABR model. This is, because our payoff function only depends on the forward value at time of maturity (F_T) and not on the volatility value at time of maturity (σ_T). By the same reasoning, we observe that pricing European options under the SABR model with the one time step DCOS method and the Euler scheme is independent of the correlation parameter ρ. Since we wish to propose a pricing method which depends on parameter ρ, we exclude the Euler scheme. Applying the theory of Chapters 3 and 4, we find the following formula to calculate the forward value of a European call option

$$V^C_{\mathrm{dcos}\,1}(0,T,K,f,\alpha) = {\sum_{k=0}^{N-1}}' \Re\left\{\phi_{F_T^\Delta}\left(\frac{k\pi}{b-a}\middle| f,\alpha\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}\mathcal{V}_k(T), \tag{5.1}$$

where

$$[a,b] = \left[f - 10\alpha f^{\beta}\sqrt{T}, f + 10\alpha f^{\beta}\sqrt{T}\right], \tag{5.2}$$

$$
\begin{aligned}
\mathcal{V}_k(T) &= \frac{2}{b-a}\int_a^b (F-K)^+ \cos\left(k\pi\frac{F-a}{b-a}\right)\, \mathrm{d}F, \\
&= \begin{cases} \frac{2}{b-a}\left(\frac{1}{2}K^2+\frac{1}{2}b^2-Kb\right), & \text{if } k=0 \text{ and } a\le K\le b, \\ \frac{2(b-a)}{k^2\pi^2}\left((-1)^k-\cos\left(k\pi\frac{K-a}{b-a}\right)\right), & \text{if } k\neq 0 \text{ and } a\le K\le b. \end{cases}
\end{aligned} \tag{5.3}
$$

Also, we find with Lemma 4.1:

$$
\begin{aligned}
\phi_{F_T^{\Delta}}\left(u|\, f,\alpha\right) &= \frac{\cosh\left(iv\Delta t\right)}{\sqrt{1-2ic_4\Delta t+c_3^2\left(\Delta t\right)^2}}\exp\left(iu\left[f+m_1\left(f,\alpha\right)\Delta t\right]\right) \\
&\cdot \quad \exp\left(-\frac{\Delta t}{2}\frac{c_1^2+c_2^2+\left[4c_2^2c_4^2-4c_1c_2c_3c_4+(c_1^2+c_2^2)c_3^2\right]\left(\Delta t\right)^2}{1+\left(2c_3^2+4c_4^2\right)\left(\Delta t\right)^2+c_3^4\left(\Delta t\right)^4}\right) \\
&\cdot \quad \exp\left(i\left(\Delta t\right)^2\left(\frac{-c_1^2c_4-c_1c_2c_3+(c_2^2c_4-c_1c_2c_3)c_3^2\left(\Delta t\right)^2}{1+\left(2c_3^2+4c_4^2\right)\left(\Delta t\right)^2+c_3^4\left(\Delta t\right)^4}\right)\right),
\end{aligned} \tag{5.4}
$$

where

$$
\begin{array}{lll}
c_1 = us_1^{W^1}\left(f,\alpha\right), & c_2 = us_1^{W^2}\left(f,\alpha\right), & c_3 = u\kappa_1^{W^1,W^2}\left(f,\alpha\right), \\
c_4 = u\kappa_1^{W^1}\left(f,\alpha\right), & v = uv_1(f,\alpha). &
\end{array} \tag{5.5}
$$

For the Milstein scheme, we observe

$$
\begin{array}{ll}
m_1\left(f,\alpha\right) = -\frac{\beta}{2}\alpha^2 f^{2\beta-1}-\frac{\varrho}{2}\nu\alpha f^{\beta}, & s_1^{W^1}\left(f,\alpha\right) = \alpha f^{\beta}, \\
s_1^{W^2}\left(f,\alpha\right) = 0, & \kappa_1^{W^1,W^2}\left(f,\alpha\right) = \frac{\sqrt{1-\rho^2}}{2}\nu\alpha f^{\beta}, \\
\kappa_1^{W^1}\left(f,\alpha\right) = \frac{\beta}{2}\alpha^2 f^{2\beta-1}+\frac{\varrho}{2}\nu\alpha f^{\beta}, & v_1(f,\alpha) = -\frac{\sqrt{1-\rho^2}}{2}\nu\alpha f^{\beta}.
\end{array} \tag{5.6}
$$

For the 2.0-weak-Taylor scheme, we have

$$
\begin{array}{ll}
m_1\left(f,\alpha\right) = -\frac{\beta}{2}\alpha^2 f^{2\beta-1}-\frac{\varrho}{2}\nu\alpha f^{\beta}, & s_1^{W^1}\left(f,\alpha\right) = \alpha f^{\beta}+\frac{1}{4}\beta(\beta-1)\alpha^2 f^{3\beta-2}T+\frac{\varrho}{2}\beta\alpha^2\nu f^{2\beta-1}T, \\
s_1^{W^2}\left(f,\alpha\right) = 0, & \kappa_1^{W^1,W^2}\left(f,\alpha\right) = \frac{\sqrt{1-\rho^2}}{2}\nu\alpha f^{\beta}, \\
\kappa_1^{W^1}\left(f,\alpha\right) = \frac{\beta}{2}\alpha^2 f^{2\beta-1}+\frac{\varrho}{2}\nu\alpha f^{\beta}, & v_1(f,\alpha) = -\frac{\sqrt{1-\rho^2}}{2}\nu\alpha f^{\beta}.
\end{array} \tag{5.7}
$$

For the adjusted Predictor-Corrector scheme, we choose to ignore the $\mathcal{O}\left(T\right)$ terms in $\kappa_1^{W^1}$, $\kappa_1^{W^2}$ and $\kappa_1^{W^1,W^2}$, because these terms introduce a significant error for large time to maturity T. For this scheme it follows

$$
m_1\left(f,\alpha\right) = -\eta_1\left[\beta\alpha^2 f^{2\beta-1}+\rho\nu\alpha f^{\beta}\right], \tag{5.8a}
$$

$$
s_1^{W^1}\left(f,\alpha\right) = \alpha f^{\beta}-\eta_1\theta_1\left[3\rho\beta\nu\alpha^2 f^{2\beta-1}+\beta(2\beta-1)\alpha^3 f^{3\beta-2}+\rho^2\nu^2\alpha f^{\beta}\right]T, \tag{5.8b}
$$

$$
s_1^{W^2}\left(f,\alpha\right) = -\eta_1\theta_1\sqrt{1-\rho^2}\left[2\beta\nu\alpha^2 f^{2\beta-1}+\rho\nu^2\alpha f^{\beta}\right]T, \tag{5.8c}
$$

$$
\kappa_1^{W^1,W^2}(f,\alpha) = \eta_1\sqrt{1-\rho^2}\nu\alpha f^{\beta}, \tag{5.8d}
$$

$$
\kappa_1^{W^1}(f,\alpha) = \eta_1\left[\beta\alpha^2 f^{2\beta-1}+\rho\nu\alpha f^{\beta}\right], \tag{5.8e}
$$

$$
v_1(f,\alpha) = 0. \tag{5.8f}
$$

5.2 Advantages and disadvantages

Pricing European options by this method has some advantages. It is a fast method and, in contrast to pricing with Hagan's formula, we do not obtain negative densities. These non-negative densities are the result of the mapping of the SABR model to its corresponding discretization model. The characteristic function of this discretized model is known, which ensures that there are no negative densities and the integral over the entire space equals 1 when N, a and b are chosen carefully.

Unfortunately, our new method has disadvantages too. The results in Figures 5.1 and 5.2 are based on the parameters from Section 2.1.2 and $N = 2^{11}$. When we compare these figures to Figure 2.3, we observe that the smiles are inaccurate. For $T = 1$ the volatility smile is inaccurate for out of the money options and when T increases the smile moves in the wrong direction. We can improve the accuracy by increasing the number of time steps, as we will show in Chapter 6.

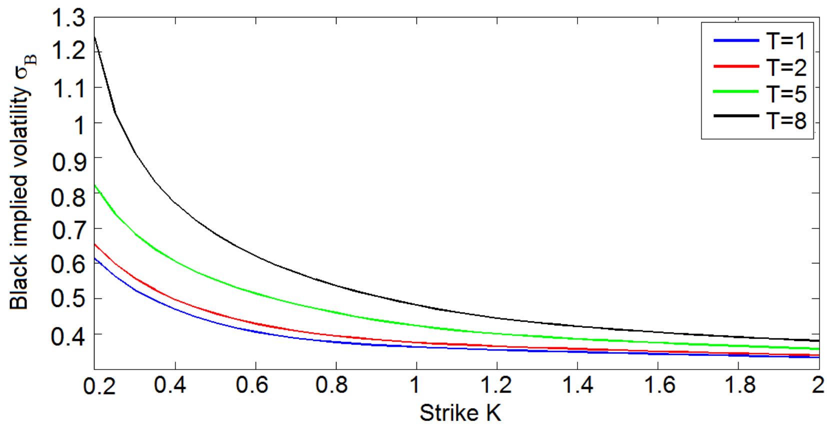

Figure 5.1: Volatility smiles calculated with the one time step DCOS method and the Milstein scheme. The smiles for the 2.0-weak-Taylor scheme look similar.

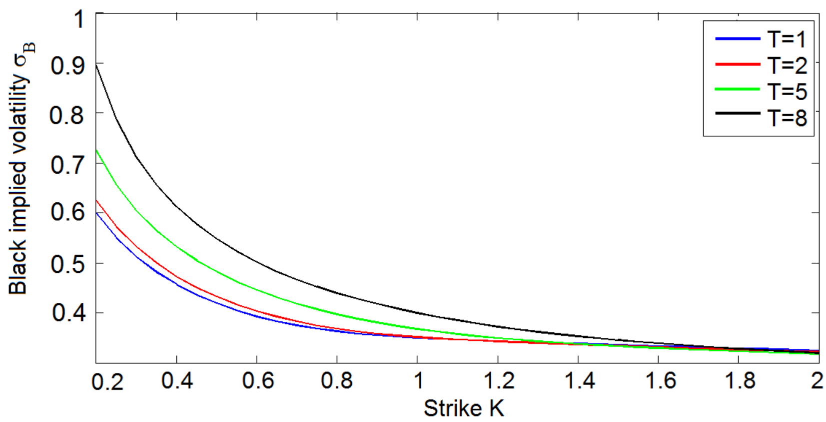

Figure 5.2: Volatility smiles calculated with the one time step DCOS method and the adjusted Predictor-Corrector scheme, where $\theta_1 = 0.5$ and $\eta_1 = 0.5$.

Figure 5.3 shows the conditional probability density functions, $Q_{F_T}(T, F|F_0 = f, \sigma_0 = \alpha)$, corresponding to the volatility smiles of Figure 5.1. The conditional PDFs for the 2.0-weak-Taylor scheme, look quite similar. We can observe the following properties

- As we mentioned before, we have no negative densities.
- $P(F_T < 0|f, \alpha) > 0$, especially when the time to maturity is large, which implies underpricing European call options for especially small strikes. This is a result of the discretization. In Section 5.3 we suggest a solution to prevent this.
- The PDF is peaked for larger time to maturity. Such peaks should be avoided, because the PDF has to be a differentiable function when $T > 0$. Andreasen and Huge [1] noticed similar behaviour when they were developing their pricing method. The reason for the peak occurring is the same as for the peaks we observe: it is the result of performing one time step instead of multiple. The authors in [1] tried to remove the peak by an adjustment to the method, but this adjustment is not beneficial for the accuracy. For their method we observed in [16] that it is better, with respect to the accuracy, to use more time steps than using the adjustment.

Remark 5.1. *$P(F_T < 0|f, \alpha) > 0$ is not always a disadvantage, for example in the case that the model is applied to interest rates where the rates may become negative [13]. On the other hand, since we are price under the SABR model, we prefer absorption at zero as [2, 5, 15].*

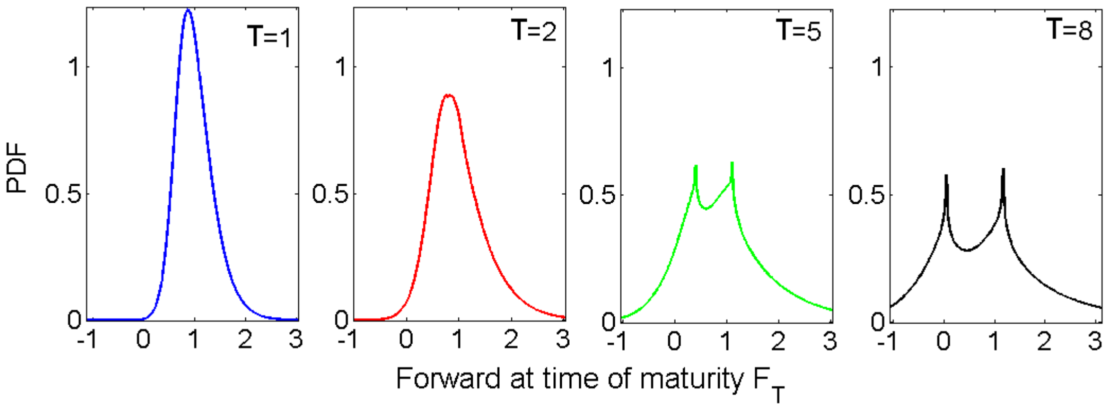

Figure 5.3: PDFs calculated with the one time step DCOS method and the Milstein scheme.

Figure 5.4 shows the conditional probability density functions, $Q(F_T|F_0 = f, \sigma_0 = \alpha)$, corresponding to the volatility smiles of Figure 5.2. We observe that, for this example, the adjusted-Predictor-Corrector scheme does not suffer from peaked densities. Unfortunately, this scheme does not exhibit this behaviour in general. For example, we also observe peaked densities, when we change the exponent parameter to $\beta = 0.1$ or when we do not ignore the $\mathcal{O}(T)$ terms in $\kappa_1^{W^1}$, $\kappa_1^{W^2}$ and $\kappa_1^{W^1,W^2}$.

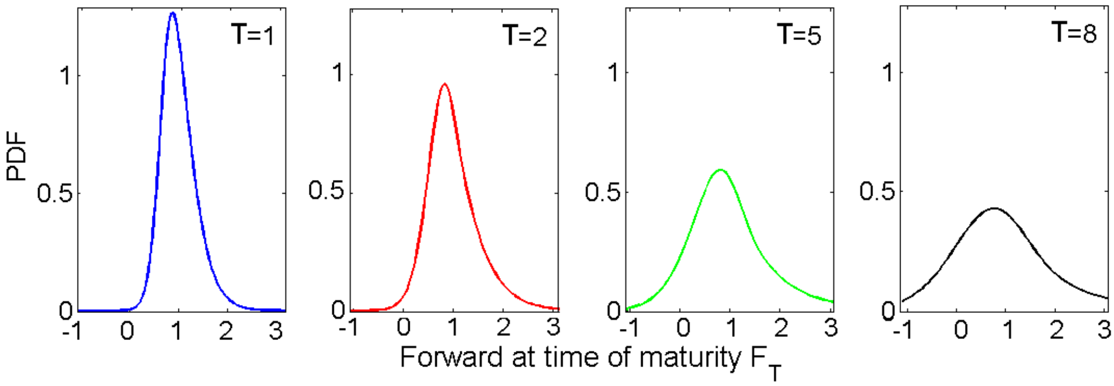

Figure 5.4: PDFs calculated with the one time step DCOS method and the adjusted Predictor-Corrector scheme.

5.3 Using the log transform for the forward variable

To avoid the positive probability of a negative F_T, we apply the DCOS method with one time step to the SABR model under the logarithmic transform for the forward[1]. We define $X_t^1 = \log(F_t)$ and using Itô's Lemma [26, Formula (4.4.23)], we find

$$\begin{aligned} \mathrm{d}X_t^1 &= -\frac{1}{2F_t^2}\sigma_t^2 F_t^{2\beta}\,\mathrm{d}t + \frac{1}{F_t}\sigma_t F_t^{\beta}\,\mathrm{d}W_t^1 \\ &= -\frac{1}{2}\sigma_t^2 \exp(2(\beta-1)X_t^1)\,\mathrm{d}t + \sigma_t \exp((\beta-1)X_t^1)\,\mathrm{d}W_t^1, \quad (5.9) \\ \mathrm{d}\sigma_t &= \rho\nu\sigma_t\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t\,\mathrm{d}W_t^2, \quad (5.10) \end{aligned}$$

where $X_0^1 = x_1 = \log(f)$ and $\sigma_0 = \alpha$.

This results in the following pricing formula

$$V_{\mathrm{dcos}\,2}^C(0,T,K,x_1,\alpha) = \sum_{k=0}^{N-1}{}' \Re\left\{\phi_{X_T^{1,\Delta}}\left(\frac{k\pi}{b-a}\middle| x_1,\alpha\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}\mathcal{V}_k(T), \quad (5.11)$$

where

$$\begin{aligned} x &= \log(f), \quad (5.12) \\ [a,b] &= \left[x_1 - \frac{1}{2}\alpha^2 f^{2\beta-2}T - 10\alpha f^{\beta-1}\sqrt{T}, x_1 - \frac{1}{2}\alpha^2 f^{2\beta-2}T + 10\alpha f^{\beta-1}\sqrt{T}\right], \quad (5.13) \\ \mathcal{V}_k(T) &= \frac{2}{b-a}\int_a^b (\exp(X)-K)^+\cos\left(k\pi\frac{X-a}{b-a}\right)\,\mathrm{d}X, \\ &= \begin{cases} \frac{2}{b-a}\left(\exp(b)-K-Kb+K\log(K)\right), & \text{if } k=0 \text{ and } a\le\log(K)\le b, \\ \frac{2(b-a)}{k^2\pi^2+(b-a)^2}\left((-1)^k\exp(b) - K\cos\left(k\pi\frac{\log(K)-a}{b-a}\right)\right) & \\ \quad +\frac{2K(b-a)^2}{k^3\pi^3+k\pi(b-a)^2}\sin\left(k\pi\frac{\log(K)-a}{b-a}\right), & \text{if } k\neq 0 \text{ and } a\le\log(K)\le b. \end{cases} \quad (5.14) \end{aligned}$$

Also, we find

$$\phi_{X_T^{1,\Delta}}(u|\, x_1,\alpha) = \frac{\cosh(iv\Delta t)}{\sqrt{1-2ic_4\Delta t + c_3^2(\Delta t)^2}}\exp\left(iu\left[x_1 + m_1(x_1,\alpha)\Delta t\right]\right)$$

[1] We can also use the log transform for the volatility. This transformation only affects the pricing formula of the adjusted-Predictor-Corrector method, but the results regarding to the accuracy are not significantly different.

$$\cdot \quad \exp\left(-\frac{\Delta t}{2}\frac{c_1^2+c_2^2+\left[4c_2^2c_4^2-4c_1c_2c_3c_4+(c_1^2+c_2^2)c_3^2\right](\Delta t)^2}{1+\left(2c_3^2+4c_4^2\right)(\Delta t)^2+c_3^4(\Delta t)^4}\right)$$

$$\cdot \quad \exp\left(i\,(\Delta t)^2\left(\frac{-c_1^2c_4-c_1c_2c_3+(c_2^2c_4-c_1c_2c_3)c_3^2(\Delta t)^2}{1+\left(2c_3^2+4c_4^2\right)(\Delta t)^2+c_3^4(\Delta t)^4}\right)\right), \tag{5.15}$$

where

$$c_1 = us_1^{W^1}(x_1,\alpha), \qquad c_2 = us_1^{W^2}(x_1,\alpha), \qquad c_3 = u\kappa_1^{W^1,W^2}(x_1,\alpha),$$
$$c_4 = u\kappa_1^{W^1}(x_1,\alpha), \qquad v = uv_1(x_1,\alpha).$$

We have for the Milstein scheme

$$m_1(x_1,\alpha) = -\frac{\beta}{2}\alpha^2\exp(2(\beta-1)x_1)-\frac{\rho\nu}{2}\alpha\exp((\beta-1)x_1), \tag{5.16a}$$
$$s_1^{W^1}(x_1,\alpha) = \alpha\exp((\beta-1)x_1), \tag{5.16b}$$
$$s_1^{W^2}(x_1,\alpha) = 0, \tag{5.16c}$$
$$\kappa_1^{W^1,W^2}(x_1,\alpha) = \frac{\sqrt{1-\rho^2}}{2}\nu\alpha\exp((\beta-1)x_1), \tag{5.16d}$$
$$\kappa_1^{W^1}(x_1,\alpha) = \frac{1}{2}\left(\alpha^2(\beta-1)\exp(2(\beta-1)x_1)+\rho\nu\alpha\exp((\beta-1)x_1)\right), \tag{5.16e}$$
$$v_1(x_1,\alpha) = -\frac{\sqrt{1-\rho^2}}{2}\nu\alpha\exp\left((\beta-1)x_1\right). \tag{5.16f}$$

For the 2.0-weak-Taylor scheme, we have

$$\begin{aligned} m_1(x_1,\alpha) &= -\frac{\beta}{2}\alpha^2\exp(2(\beta-1)x_1)-\frac{\rho\nu}{2}\alpha\exp((\beta-1)x_1)-\frac{3-5\beta+2\beta^2}{4}\alpha^4\exp(4(\beta-1)x_1)T \\ &- \frac{\nu^2}{4}\alpha^2\exp(2(\beta-1)x_1)T-\rho\nu\alpha^3(\beta-1)\exp(3(\beta-1)x_1)T, \end{aligned} \tag{5.17a}$$
$$\begin{aligned} s_1^{W^1}(x_1,\alpha) &= \alpha\exp((\beta-1)x_1)+\frac{(\beta-1)(\beta-4)}{4}\alpha^3\exp(3(\beta-1)x_1)T \\ &+ \frac{\beta-2}{2}\rho\alpha^2\nu\exp(2(\beta-1)x_1)T, \end{aligned} \tag{5.17b}$$
$$s_1^{W^2}(x_1,\alpha) = -\frac{\sqrt{1-\rho^2}}{2}\nu\alpha^2\exp(2(\beta-1)x_1)T, \tag{5.17c}$$
$$\kappa_1^{W^1,W^2}(x,\alpha) = \frac{\sqrt{1-\rho^2}}{2}\nu\alpha\exp((\beta-1)x_1), \tag{5.17d}$$
$$\kappa_1^{W^1}(x,\alpha) = \frac{1}{2}\left(\alpha^2(\beta-1)\exp(2(\beta-1)x_1)+\rho\nu\alpha\exp((\beta-1)x_1)\right), \tag{5.17e}$$
$$v_1(x_1,\alpha) = -\frac{\sqrt{1-\rho^2}}{2}\nu\alpha\exp\left((\beta-1)x_1\right). \tag{5.17f}$$

For the adjusted-Predictor-Corrector scheme, where we choose again to ignore the $\mathcal{O}(T)$ terms in $\kappa_1^{W^1}$, $\kappa_1^{W^2}$ and $\kappa_1^{W^1,W^2}$, it follows that

$$\begin{aligned} m_1(x_1,\alpha) &= -\theta_1\left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)\alpha^2A^2(x_1)+\eta_1\rho\nu\alpha A(x_1)\right] \\ &- (1-\theta_1)\left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)\alpha^2\exp\left(2(\beta-1)x_1\right)+\eta_1\rho\nu\alpha\exp((\beta-1)x_1)\right], \end{aligned} \tag{5.18a}$$
$$s_1^{W^1}(x_1,\alpha) = \eta_1\alpha A(x_1)+(1-\eta_1)\alpha\exp((\beta-1)x_1)$$

$$
\begin{aligned}
&- \theta_1 \left[(1+2\eta_1(\beta-1)) \left[(\beta-1)\alpha^3 \exp((\beta-1)x_1) + \rho\nu\alpha^2\right] A^2(x_1)\right. \\
&+ \left.\eta_1\rho\nu \left[(\beta-1)\alpha^2 \exp((\beta-1)x) + \rho\nu\alpha\right] A(x_1)\right] T, \quad (5.18b)
\end{aligned}
$$

$$
\begin{aligned}
s_1^{W^2}(x_1,\alpha) &= -\theta_1\sqrt{1-\rho^2}\left[(1+2\eta_1(\beta-1))\,\nu\alpha^2 A^2(x_1) + \eta_1\rho\nu^2\alpha A(x_1)\right] T, && (5.18c) \\
\kappa_1^{W^1,W^2}(x_1,\alpha) &= \eta_1\sqrt{1-\rho^2}\nu\alpha \exp\left((\beta-1)x_1\right), && (5.18d) \\
\kappa_1^{W^1}(x_1,\alpha) &= \eta_1\left[(\beta-1)\alpha^2 \exp\left((\beta-1)x_1\right) + \rho\nu\alpha\right] \exp\left((\beta-1)x_1\right), && (5.18e) \\
v_1(x_1,\alpha) &= 0, && (5.18f)
\end{aligned}
$$

where

$$
A(x_1) = \exp\left((\beta-1)x_1 - \frac{1}{2}(\beta-1)\exp\left(2(\beta-1)x_1\right) T\right). \quad (5.19)
$$

5.4 Accuracy

We proposed to use the log transform to ensure that there is zero probability of a negative forward value F_T at time of maturity. Figure 5.5 shows that indeed $P(F_T < 0|f,\alpha) = 0$. Unfortunately, we observe now overpricing instead of underpricing for small strikes. The 2.0-weak-Taylor scheme behaves similarly and the adjusted-Predictor-Corrector scheme behaves even worse.

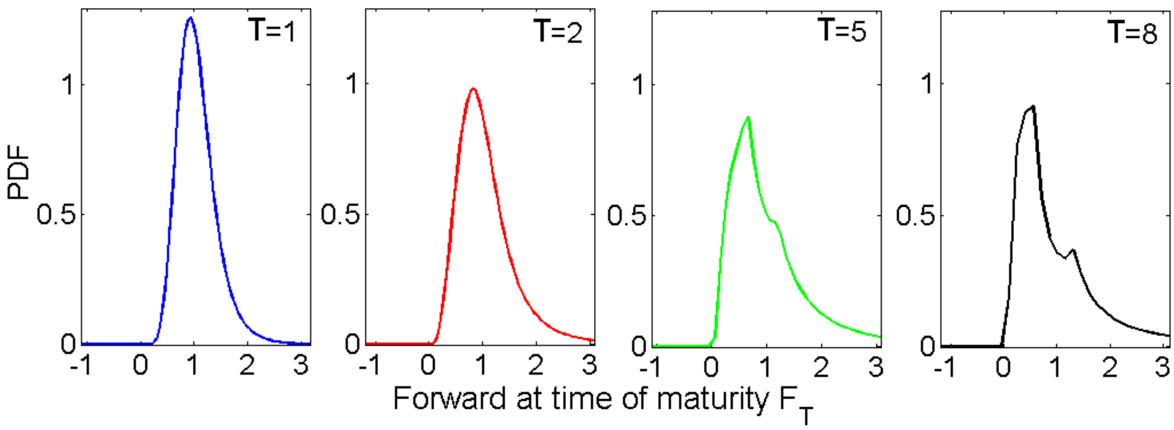

Figure 5.5: PDFs calculated with the one time step DCOS method and the Milstein scheme.

Like in Section 5.1, the probability density function is non-negative and peaked for larger time to maturity T. Also, we find that pricing with formula (5.11) leads to lower accuracy and for the 2.0-weak-Taylor scheme even to unrealistic volatility smiles, as we can see in Figures 5.6, 5.7 and 5.8.

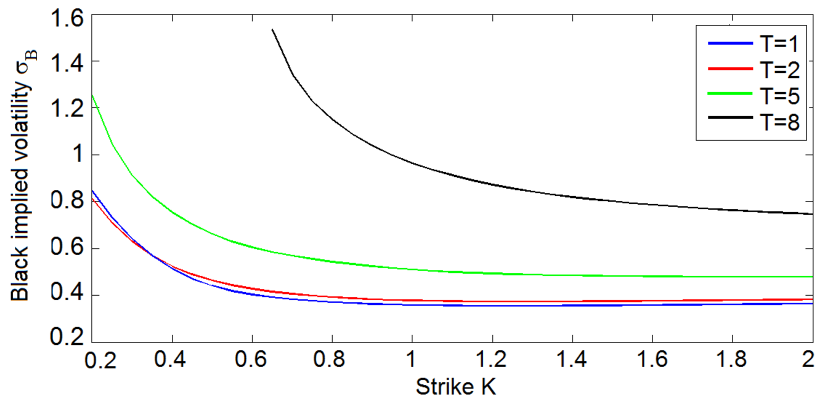

Figure 5.6: Volatility smiles calculated with the one time step DCOS method and the Milstein scheme.

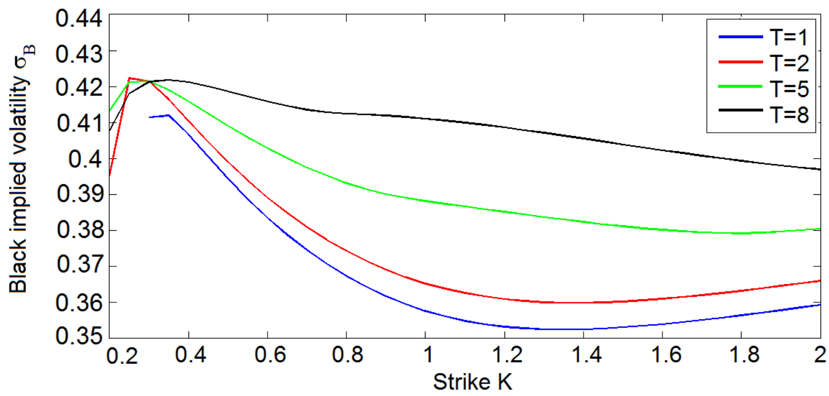

Figure 5.7: Volatility smiles calculated with the one time step DCOS method and the 2.0-weak-Taylor scheme.

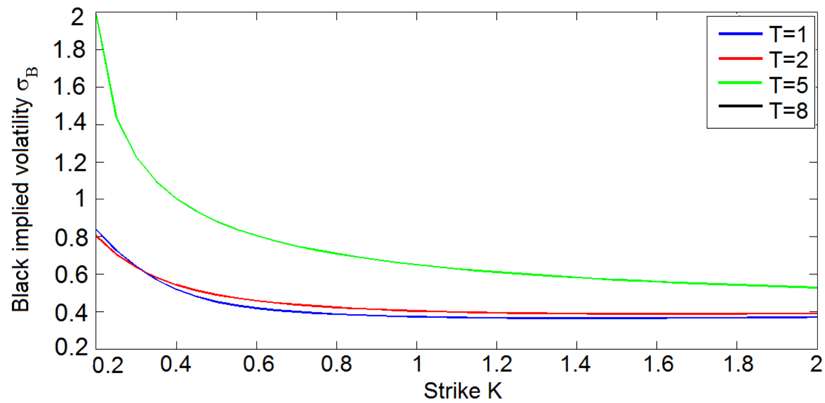

Figure 5.8: Volatility smiles calculated with the one time step DCOS method and the adjusted-Predictor-Corrector scheme, where $\theta_1 = 0.5$ and $\eta_1 = 0.5$.

5.5 Brief analysis

In this section we look at some properties of option pricing with the formulas (5.1) or (5.11) compared with the Hagan formula (2.3) as described in Section 2.

- Neglecting numerical errors introduced by the COS method, formulas (5.1) and (5.11) are free of arbitrage, because we are pricing European options where the underlying follows the discretized FSDE and the characteristic function is known for the discretized FSDE.
- None of the methods is accurate, especially for larger time to maturity.
- Pricing with formula (5.1) or (5.11) ensures non-negative probabilities in contrast to pricing with the Hagan formula (2.3), as is shown in Section 2.1.1.
- Pricing with the Hagan formula gives continuous densities in contrast to pricing with formula (5.1) or (5.11), because those formulas generate peaked densities for larger time to maturity.
- Pricing with formula (5.1) or (5.11) is fast, e.g. pricing a European call option with $N = 2^{11}$, the Euler scheme and formula (5.1) takes less than 0.005 seconds.
- For small strikes, formula (5.1) generates too small values and formula (5.11) generates too high prices for European call options.

As we mentioned in Chapter 2, we wish to provide an arbitrage-free method to price European options under the SABR model that is accurate. Since the pricing formulas (5.1) and (5.11) are not accurate for small strikes, we propose in Chapter 6 to apply the DCOS method with multiple time steps to price European options under the SABR model.

Chapter 6

The DCOS method applied to the SABR model

As we observed in Chapter 5, taking only one time step is not sufficient for pricing European options under the SABR model with the two-dimensional DCOS method. In this chapter we apply the DCOS method to the SABR model with multiple time steps[1]. Under the SABR model, the volatility σ_t is independent of the forward value F_t, which simplifies formulas (4.3) and (4.4) for the Euler, Milstein, 2.0-weak-Taylor and adjusted-Predictor-Corrector schemes. In Section 6.1 we give three pricing formulas for this method and in Section 6.2 we discuss the results. In Section 6.3 we explain a transformation to price European options under the SABR model for multiple strikes in one computation. In Section 6.4, we discuss the advantages and disadvantages of the best performing pricing formula and schemes. Finally, we compare our proposed method with the Hagan formula in Section 6.5.

6.1 The method

We derive the forward value $V^C(0, T, K)$ of a European call option under the SABR model, with exercise date T and strike value K. We consider the following three cases:

Case 1 F_t denotes the forward process and σ_t denotes the volatility process. Following (2.1) and (2.2), the underlying system of FSDEs is given by

$$\mathrm{d}F_t = \sigma_t (F_t)^{\beta} \, \mathrm{d}W_t^1, \qquad F_0 = f, \tag{6.1}$$

$$\mathrm{d}\sigma_t = \rho\nu\sigma_t \, \mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t \, \mathrm{d}W_t^2, \qquad \sigma_0 = \alpha, \tag{6.2}$$

where W^1 and W^2 are uncorrelated standard Brownian motions.

Case 2 Defining the log forward process by $X_t^1 = \log(F_t)$ and σ_t denotes the volatility process. Following (5.9) and (5.10), results in the following underlying system of FSDEs

$$\mathrm{d}X_t^1 = -\frac{1}{2}\sigma_t^2 \exp\left(2(\beta-1)X_t^1\right) \mathrm{d}t + \sigma_t \exp\left((\beta-1)X_t^1\right) \mathrm{d}W_t^1, \qquad X_0^1 = x_1, \tag{6.3}$$

[1]The SABR model does not satisfy the commutativity condition (4.7), so we do not have the certainty that the Milstein scheme or the 2.0-weak-Taylor scheme have strong order of convergence $\gamma_1 = 1$. Importantly, as we mentioned in Remark 4.1, the order of convergence of the DCOS method depends on the order of weak convergence γ_2 and not on the order of strong convergence γ_1.

$$\mathrm{d}\sigma_t = \rho\nu\sigma_t\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t\,\mathrm{d}W_t^2, \qquad \sigma_0 = \alpha, \tag{6.4}$$

where $x_1 = \log(f)$ and where W^1 and W^2 are uncorrelated standard Brownian motions.

Case 3 We use the logarithmic transformation for both the forward and the volatility. We define $X_t^1 = \log(F_t)$ and $X_t^2 = \log(\sigma_t)$, the underlying system of FSDEs is given by

$$\mathrm{d}X_t^1 = -\frac{1}{2}\exp\left(2X_t^2 + 2(\beta-1)X_t^1\right)\mathrm{d}t + \exp\left(X_t^2 + (\beta-1)X_t^1\right)\mathrm{d}W_t^1, \qquad X_0^1 = x_1, \tag{6.5}$$

$$\mathrm{d}X_t^2 = -\frac{1}{2}\nu^2\,\mathrm{d}t\rho\nu\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\,\mathrm{d}W_t^2, \qquad X_0^2 = x_2, \tag{6.6}$$

where $(x_1, x_2) = (\log(f), \log(\alpha))$ and where W^1 and W^2 are uncorrelated standard Brownian motions.

Just as in Chapter 4, we define the time-grid $t_j = \Delta t$ for $j = 0, 1, \ldots, M$, with fixed time steps $\Delta t = \frac{T}{M}$. For $j = 1, 2$, we again write $F_m = F_{t_m}$, $\sigma_m = \sigma_{t_m}$, $X_m^1 = X_{t_m}^1$, $X_m^2 = X_{t_m}^2$, $W_m^j = W_{t_m}^j$ and $\Delta W_{m+1}^j = W_{m+1}^j - W_m^j$. We denote the discrete processes by $\left(F_m^\Delta, \sigma_m^\Delta\right) = \left(F_{t_m}^\Delta, \sigma_{t_m}^\Delta\right)$, $\left(X_m^{1,\Delta}, \sigma_m^\Delta\right) = \left(X_{t_m}^{1,\Delta}, \sigma_{t_m}^\Delta\right)$ and $\left(X_m^{1,\Delta}, X_m^{2,\Delta}\right) = \left(X_{t_m}^{1,\Delta}, X_{2,t_m}^\Delta\right)$, for case 1, case 2 and case 3, respectively. The initial values of the discrete processes are given by $\left(F_0^\Delta, \sigma_0^\Delta\right) = (f, \alpha)$, $\left(X_0^{1,\Delta}, \sigma_0^\Delta\right) = (x_1, \alpha)$ and $\left(X_0^{1,\Delta}, X_0^{2,\Delta}\right) = (x_1, x_2)$. To advance from time m to time $m+1$, for $m = 0, \ldots, M-1$, we use one of the following schemes: the Euler, Milstein, 2.0-weak-Taylor, or adjusted-Predictor-Corrector scheme.

The forward value of a European call option at time 0 with underlying F_t, strike value K and expiration date T is given by:
For Case 1:

$$V_{\mathrm{DCOS\,1}}^C(0, T, K, f, \alpha) = V_0^\Delta(f, \alpha), \tag{6.7}$$

where the functions $V_m^\Delta(F_m, \sigma_m)$ can be recovered recursively backwards in time for all $(F_m, \sigma_m) \in [a_1, b_1] \times [a_2, b_2] = \left[f - 10\alpha f^\beta\sqrt{T}, f + 10\alpha f^\beta\sqrt{T}\right] \times \left[\alpha - 10\nu\alpha\sqrt{T}, \alpha + 10\nu\alpha\sqrt{T}\right]$ and $m = M-2, \ldots, 0$,

$$\begin{aligned}
V_m^\Delta(F_m, \sigma_m) &= \mathbb{E}_m\left[V_{m+1}^\Delta(F_{m+1}, \sigma_{m+1})\middle| F_m, \sigma_m\right] \\
&= \sum_{k_1=0}^{N_1-1}{}' \sum_{k_2=0}^{N_2-1}{}' \frac{1}{2}\Re\left\{\phi_{\left(F_{m+1}^\Delta, \sigma_{m+1}^\Delta\right)}\left(\frac{k_1\pi}{b_1-a_1}, \frac{k_2\pi}{b_2-a_2}\middle| F_m, \sigma_m\right)\right. \\
&\cdot\ \exp\left(-ik_1\pi\frac{a_1}{b_1-a_1} - ik_2\pi\frac{a_2}{b_2-a_2}\right) + \phi_{\left(F_{m+1}^\Delta, \sigma_{m+1}^\Delta\right)}\left(\frac{k_1\pi}{b_1-a_1}, -\frac{k_2\pi}{b_2-a_2}\middle| F_m, \sigma_m\right) \\
&\cdot\ \left.\exp\left(-ik_1\pi\frac{a_1}{b_1-a_1} + ik_2\pi\frac{a_2}{b_2-a_2}\right)\right\}\mathcal{V}_{k_1,k_2}(t_{m+1}).
\end{aligned} \tag{6.8}$$

The characteristic function $\phi_{\left(F_{m+1}^\Delta, \sigma_{m+1}^\Delta\right)}(u_1, u_2|F_m, \sigma_m)$ is determined with Lemma 4.1 for any of the discussed discretization schemes and we have

$$\mathcal{V}_{k_1,k_2}(t_{m+1}) = \frac{2}{b_1-a_1}\frac{2}{b_2-a_2}\int_{a_2}^{b_2}\int_{a_1}^{b_1} V_{m+1}^\Delta(F, \sigma)\cos\left(k_1\pi\frac{F-a_1}{b_1-a_1}\right)\cos\left(k_2\pi\frac{\sigma-a_2}{b_2-a_2}\right)\,\mathrm{d}F\,\mathrm{d}\sigma. \tag{6.9}$$

We approximate the above double integrals by computing the function $V^{\Delta}_{m+1}(F_{m+1},\sigma_{m+1})$ on a two-dimensional grid and using the (2D) discrete Fourier-cosine transform. As a special case, for $m = M-1$, we find with formulas (5.1) and (5.3):

$$V^{\Delta}_{M-1}(F_{M-1},\sigma_{M-1}) = \sum_{k=0}^{N_1-1}{}' \Re\left\{\phi_{(F^{\Delta}_M,\sigma^{\Delta}_M)}\left(\frac{k\pi}{b_1-a_1},0\middle| F_{M-1},\sigma_{M-1}\right)\exp\left(-ik\pi\frac{a_1}{b_1-a_1}\right)\right\}\mathcal{V}_k(t_M), \tag{6.10}$$

where

$$\begin{aligned}\mathcal{V}_k(t_M) &= \frac{4}{b_1-a_1}\int_{a_1}^{b_1}(F-K)^+\cos\left(k\pi\frac{F-a_1}{b_1-a_1}\right)\,\mathrm{d}F\\ &= \begin{cases}\frac{2}{b_1-a_1}\left(\frac{1}{2}K^2+\frac{1}{2}b_1^2-Kb_1\right), & \text{if } k=0 \text{ and } a_1\le K\le b_1,\\ \frac{2(b_1-a_1)}{k^2\pi^2}\left((-1)^k-\cos\left(k\pi\frac{K-a_1}{b_1-a_1}\right)\right), & \text{if } k\neq 0 \text{ and } a_1\le K\le b_1.\end{cases}\end{aligned} \tag{6.11}$$

Analogously, we have for Case 2,

$$V^C_{\mathrm{DCOS}\,2}(0,T,K,x_1,\alpha) = V^{\Delta}_0(x_1,\alpha), \tag{6.12}$$

and for Case 3,

$$V^C_{\mathrm{DCOS}\,3}(0,T,K,x_1,x_2) = V^{\Delta}_0(x_1,x_2), \tag{6.13}$$

where the functions $V^{\Delta}_m(\cdot,\cdot)$ can be recovered recursively, backwards in time, for all $m = M-2,\ldots,0$. The integration domains are given by

$$\begin{aligned}[a_1,b_1]\times[a_2,b_2] = \Big[&x_1-\alpha^2\exp(2(\beta-1)x_1)T-10\alpha\exp((\beta-1)x_1)\sqrt{T}\;,\\ &x_1+\alpha^2\exp(2(\beta-1)x_1)T+10\alpha\exp((\beta-1)x_1)\sqrt{T}\Big]\times\left[\alpha-10\nu\alpha\sqrt{T},\alpha+10\nu\alpha\sqrt{T}\right],\\ [a_1,b_1]\times[a_2,b_2] = \Big[&x_1-\exp(2x_2+2(\beta-1)x_1)T-10\exp(x_2+(\beta-1)x_1)\sqrt{T}\;,\\ &x_1+\exp(2x_2+2(\beta-1)x_1)T+10\exp(x_2+(\beta-1)x_1)\sqrt{T}\Big]\\ &\times\left[x_2-\nu^2T/2-10\nu\sqrt{T},\alpha-\nu^2T/2+10\nu\sqrt{T}\right],\end{aligned}$$

for Case 2 and Case 3, respectively. For both cases, formula (5.14) leads to the following formula for $\mathcal{V}_k(t_{m+1})$, for the special case $m = M-1$,

$$\begin{aligned}\mathcal{V}_k(t_M) &= \frac{2}{b_1-a_1}\int_{a_1}^{b_1}(\exp(X)-K)^+\cos\left(k\pi\frac{X-a_1}{b_1-a_1}\right)\,\mathrm{d}X,\\ &= \begin{cases}\frac{2}{b_1-a_1}\left(\exp(b_1)-K-Kb_1+K\log(K)\right), & \text{if } k=0 \text{ and } a_1\le\log(K)\le b_1,\\ \frac{2(b_1-a_1)}{k^2\pi^2+(b_1-a_1)^2}\left((-1)^k\exp(b_1)-K\cos\left(k\pi\frac{\log(K)-a_1}{b_1-a_1}\right)\right) & \\ \quad+\frac{2K(b_1-a_1)^2}{k^3\pi^3+k\pi(b_1-a_1)^2}\sin\left(k\pi\frac{\log(K)-a_1}{b_1-a_1}\right), & \text{if } k\neq 0 \text{ and } a\le\log(K)\le b_1.\end{cases}\end{aligned} \tag{6.14}$$

6.2 Results

In this section we present some results for the pricing formulas (6.7), (6.12) and (6.13) of Cases 1,2 and 3, respectively, by means of an example. We use the parameters of the example of Section 2.1.2, i.e. $\alpha = 0.35$, $\beta = 0.8$, $\rho = 0$, $\nu = 0.4$, also we take $f = 2$, $T = 1$, $N = 2^7$

and strikes $K = 1.8, 1.95, \ldots 3, 3.15$. As explained in Section 4.10 that N denotes the number of Fourier-cosine coefficients and the size of the grid in both dimensions. In this example, we determine the Black implied volatility smiles by the 2D DCOS for Case 1 with the Euler scheme and number of time steps $M = 20, 50, 100$. These volatility smiles and the mean of the absolute errors in basis points are given in Figure 6.1. When M increases, the volatility smile determined with formula (6.7) converges to the reference smile, which is determined by Antonov's pricing approach (2.11).

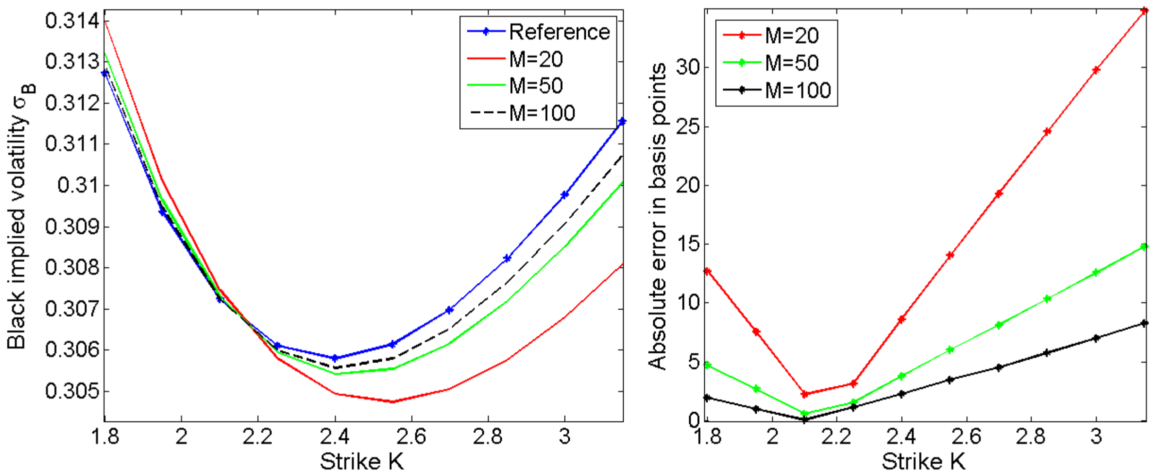

Figure 6.1: The volatility smiles and the absolute error in Black implied volatility in basis points.

Besides formula (6.7) for the Euler scheme, we also determine the Black implied volatilities with the formulas (6.12) and (6.13), for the other schemes: the Milstein, 2.0-weak-Taylor and adjusted-Predictor-Corrector schemes. Figure 6.2 shows the mean of the absolute errors in basis points for this example. Where Cases 1, 2 and 3 refer to the pricing formulas (6.7), (6.12) and (6.13), respectively. Formula (6.13) gives the best results for the Euler, Milstein and 2.0-weak-Taylor schemes, while formula (6.12) shows the fastest convergence for the adjusted-Predictor-Corrector scheme. In Section 4.9, we performed an error analysis for the 2.0-weak-Taylor scheme. In our analysis we observed second-order weak convergence for the option value, and analogously we obtained at least first-order weak convergence for the other schemes. In Figure 6.2 we observe the fastest convergence for the 2.0-weak-Taylor and the slowest convergence for the Euler scheme. This example indicates that 2.0-weak-Taylor performs best with respect to convergence.

We find it remarkable that formula (6.13) performs best for the Euler, Milstein and 2.0-weak-Taylor scheme and worst for the adjusted-Predictor-Corrector scheme. We believe this is a consequence of the errors introduced by the truncation of the Taylor series expansions. In formula (4.23) the truncation for the adjusted-Predictor-Corrector scheme in Case 1 is given.
We observe convergence of order 1.75 for the 2.0-weak-Taylor scheme, while we expected second order convergence. We also notice that the error introduced by the choice of the number of grid-points N is visible for the Milstein and the 2.0-weak-Taylor schemes when the absolute error is less than one basis point in Case 1 and less than 0.3 basis points in Cases 2 and 3. We believe that we do not observe the second order convergence for the 2.0-weak-Taylor scheme due to this same error. Because of the results in Figure 6.2, we advise to use pricing formula (6.13) with the Euler or the 2.0-weak-Taylor scheme. We will discuss this choice further in Section 6.4.

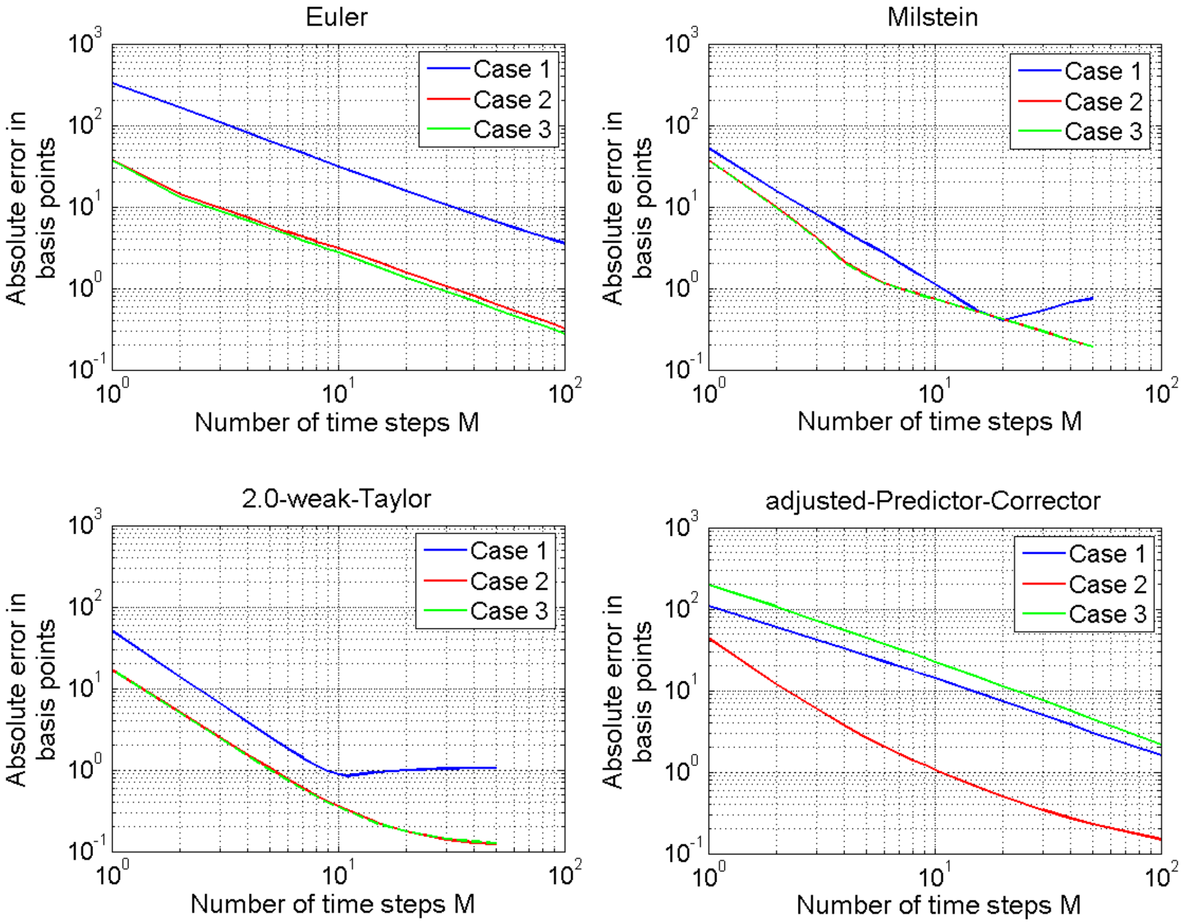

Figure 6.2: Mean of the absolute error in Black implied volatility with respectively the Euler, Milstein, 2.0-weak-Taylor and adjusted-Predictor-Corrector schemes, and formulas (6.7), (6.12) and (6.13).

6.3 Multiple strikes

It is possible to price European options under the SABR model for multiple strikes at once, by using the scaling symmetry of H. Park [20]. The forward value of a European call option under the SABR model with strike value $K > 0$ and time to maturity T is given by

$$V^C(0,T,K,f,\alpha) = \mathbb{E}\left[(F_T - K)^+ \middle| f, \alpha \right], \tag{6.15}$$

where the underlying system of FSDEs is given by (6.1) and (6.2). We use the following scaling transformations

$$\widehat{F}_t = \frac{K_0}{K} F_t \qquad \text{and} \qquad \widehat{\sigma}_t = \left(\frac{K_0}{K}\right)^{1-\beta} \sigma_t, \tag{6.16}$$

and we observe that, for any $K_0 \in \mathbb{R}_{>0}$, it holds that

$$\begin{aligned} V^C(0,T,K,f,\alpha) &= \mathbb{E}\left[(F_T - K)^+ \middle| f, \alpha \right] \\ &= \mathbb{E}\left[\left(\frac{K}{K_0} \widehat{F}_T - K \right)^+ \middle| f, \alpha \right] \\ &= \frac{K}{K_0} \mathbb{E}\left[\left(\widehat{F}_T - K_0 \right)^+ \middle| f, \alpha \right]. \end{aligned} \tag{6.17}$$

By Itô's Lemma it follows that

$$\begin{aligned}
\mathrm{d}\widehat{F}_t &= \frac{K_0}{K}\sigma_t (F_t)^\beta \,\mathrm{d}W_t^1 \\
&= \frac{K_0}{K}\left(\frac{K_0}{K}\right)^{\beta-1} \widehat{\sigma}_t \left(\frac{K}{K_0}\widehat{F}_t\right)^\beta \mathrm{d}W_t^1 \\
&= \widehat{\sigma}_t \left(\widehat{F}_t\right)^\beta \mathrm{d}W_t^1, \qquad (6.18) \\
\mathrm{d}\widehat{\sigma}_t &= \rho\left(\frac{K_0}{K}\right)^{1-\beta} \nu\sigma_t \,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\left(\frac{K_0}{K}\right)^{1-\beta} \nu\sigma_t \,\mathrm{d}W_t^2 \\
&= \rho\nu\widehat{\sigma}_t \,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\widehat{\sigma}_t \,\mathrm{d}W_t^2. \qquad (6.19)
\end{aligned}$$

$\widehat{F}_t$ and $\widehat{\sigma}_t$ follow a SABR process, with the same parameters β, ν and ρ as F_t and σ_t. Combining this observation and result (6.17) implies

$$V^C(T,K,f,\alpha) = \frac{K}{K_0}V^C\left(T, K_0, \frac{K_0}{K}f, \left(\frac{K_0}{K}\right)^{1-\beta}\alpha\right). \qquad (6.20)$$

This results in the following relations for the pricing formulas (6.7), (6.12) and (6.13):

$$\begin{aligned}
V^C_{\mathrm{DCOS}\,1}(0,T,K,f,\alpha) &\approx \frac{K}{K_0}V^C_{\mathrm{DCOS}\,1}\left(0,T,K_0,\frac{K_0}{K}f, \left(\frac{K_0}{K}\right)^{1-\beta}\alpha\right), \qquad (6.21) \\
V^C_{\mathrm{DCOS}\,2}(0,T,K,x_1,\alpha) &\approx \frac{K}{K_0}V^C_{\mathrm{DCOS}\,2}\left(0,T,K_0,\log\left(\frac{K_0}{K}\right)+x_1, \left(\frac{K_0}{K}\right)^{1-\beta}\alpha\right), \qquad (6.22) \\
V^C_{\mathrm{DCOS}\,3}(0,T,K,x_1,x_2) &\approx \frac{K}{K_0}V^C_{\mathrm{DCOS}\,3}\left(0,T,K_0,\log\left(\frac{K_0}{K}\right)+x_1, (1-\beta)\log\left(\frac{K_0}{K}\right)+x_2\right). \qquad (6.23)
\end{aligned}$$

We use those approximations to price European options with multiple strikes in one computation, by pricing the corresponding European option for one general strike K_0 for different initial values. Such a strike K_0 can be for example the ATM strike value or the mean of the minimal and the maximal strike value. We advise to use this scaling symmetry for strike values that are close to each other, because the multiple initial conditions increase the size of the domain $[a_1,b_1]\times[a_2,b_2]$. If the size of the domain is increased significantly, we also have to increase the number of Fourier coefficients and the number of grid-points to obtain the same accuracy.

6.4 Advantages and disadvantages

In Section 6.2 we advised to use pricing formula (6.13), i.e. Case 3, with the Euler or the 2.0-weak-Taylor schemes. We explain this choice and give some advantages and disadvantages for both pricing methods in this section.

- Formula (6.13) is more accurate than pricing formula (6.7) for the Euler, Milstein and 2.0-weak-Taylor schemes. An example is shown in Figure 6.2.
- In formula (6.13) we use the logarithmic transformation of the volatility. In Chapter 5 we observed that such a transformation avoids a positive conditional probability of a negative volatility at expiration time, i.e. the transformation ensures $P(\sigma_T < 0 | F_0 = f, \sigma_0 = \alpha) = 0$, while for formulas (6.7) and (6.12) this probability may be positive.

- The 2.0-weak-Taylor scheme exhibits the fastest convergence. As we can see in Figure 6.2, the error for the 2.0-weak-Taylor scheme with only five time steps is less than one basis point, while we need 28 time steps for the Euler scheme to obtain this accuracy.

- The discretized SABR model has a significantly easier characteristic function for the Euler scheme compared to the other schemes, which ensures that the CPU time per time step is the lowest for the Euler scheme. As a consequence, the Euler scheme often uses less CPU time than the 2.0-weak-Taylor scheme to gain the same accuracy. The characteristic functions for the Milstein, 2.0-weak-Taylor and adjusted-Predictor-Corrector schemes have the same order of size. In Table 6.1 we give an overview of the number of time steps and the CPU time we need for the Euler and 2.0-weak-Taylor schemes to obtain a certain accuracy for the example given in Section 6.2, where we use formula (6.23).

	M	CPU	Absolute error in BPS
Euler	28	65.41s	< 1
2.0-weak-Taylor	5	140.86s	< 1
Euler	56	134.82s	< 0.5
2.0-weak-Taylor	8	412.51s	< 0.5

Table 6.1: Number of time steps and CPU time needed to obtain an absolute error of 1 or 0.5 basis points in Black implied volatility for the Euler and 2.0-weak-Taylor schemes.

- We observe from Table 6.1 that the CPU time we need to obtain an accuracy of 1 basis point is 65.41 seconds for the Euler scheme and 140.86 seconds for the 2.0-weak-Taylor scheme (in Matlab on an i5-4670 CPU @ 3.40GHz, 4 Cores). Those high computation times are a disadvantage of the 2D DCOS method. We however expect that we can reduce those computation times significantly by using the GPU, because for every time step we do a large number of parallel computations. This is left for future research.

- When we neglect errors introduced by the COS method, i.e. by the choice of N and the domain $[a_1, b_1] \times [a_2, b_2]$, our pricing method is free of arbitrage, because as a result of the chosen discretization scheme variable F_T, given $F_0 = f$ and $\sigma_0 = \alpha$, is a random variable with a certain distribution, e.g. F_T is normally distributed with mean f and variance $\alpha f^\beta \Delta t$ for formula (6.7), the Euler scheme and one time step. For multiple time steps the distribution of variable F_T is more involved and we apply the DCOS method to determine the distribution, this ensures non-negative conditional probabilities and its integration over the entire space equals one. Also, this ensures compatibility with the put-call parity (2.9). The COS method may introduce errors which lead to arbitrage possibilities. When the domain $[a_1, b_1] \times [a_2, b_2]$ is chosen incorrectly, the integration of the conditional PDF over the entire space is significantly less than one. When the number of grid-points in both dimensions and the number of Fourier-cosine coefficients N is chosen too small, it is possible to obtain oscillations in the conditional PDF and/or its integral over the entire space is unequal to one. So, we recommend to choose the domain $[a_1, b_1] \times [a_2, b_2]$ and the number N carefully.
  To prevent arbitrage-possibilities for the DCOS method due to an incorrect choice for $[a_1, b_1] \times [a_2, b_2]$ and N, we advise to calculate the price of a European put from the price of a European call by using the put-call parity (2.9).

6.5 Brief analysis

We proposed to use pricing formula (6.13) with the Euler or the 2.0-weak-Taylor scheme. In this section we summarize some properties of those two pricing methods and we compare them with the Hagan formula. (2.3).

- The Hagan formula is not accurate for long time to maturity, while pricing European options with pricing formula (6.13) for the Euler or the 2.0-weak-Taylor scheme can be as accurate as desired by choosing N, M and the domain $[a_1, b_1] \times [a_2, b_2]$ carefully.
- The Hagan formula often implies arbitrage for low strikes, whereas option pricing with the DCOS method is free of arbitrage when domain $[a_1, b_1] \times [a_2, b_2]$ and the number N are chosen carefully.
- Pricing with the DCOS method is significantly slower than pricing with the Hagan formula. We are sure to reduce the computation times in the future by using GPU programming. Still, we think it is not feasible to have an accurate pricing method that is as fast as the almost instantaneous Hagan formula, but we believe it is possible to have an accurate pricing method that is as fast as Hagan's arbitrage-free pricing approach.

Chapter 7

Conclusion

In this chapter we present a conclusion of the thesis in section 7.1 and we give recommendations for further research in section 7.2.

7.1 Conclusion

In 2002, P.S. Hagan, D. Kumar, A.S. Lesniewski and D.E. Woodward introduced the SABR model and gave a formula to calculate the Black implied volatility of the SABR model [14]. This formula, also known as the Hagan formula, leads however to arbitrage possibilities for low strikes. Besides the presence of arbitrage, the Hagan formula has another disadvantage. The formula is not accurate for long maturities. Following the pricing approach of Hagan, Kumar, Lesniewski and Woodward, other methods have been developed for pricing European options under the SABR model. We discussed Antonov's method [2, 3] and Hagan's arbitrage free pricing approach [15]. Unfortunately, Antonov's method is not free of arbitrage, while Hagan's arbitrage free pricing approach is not accurate for long maturities. This led to the development of a new method.

M.J. Ruijter and C.W. Oosterlee developed a Fourier method [25, 24] to solve BSDEs using the characteristic function of the one-dimensional underlying process. This method is called the one-dimensional BCOS method. When the characteristic function of the underlying process cannot easily be derived, we can use the characteristic function of a discrete forward process to approximate the solution, where we approximate the underlying FSDE by the Euler scheme, the Milstein scheme or the 2.0-weak-Taylor scheme.

We extended the BCOS method from one dimension to two dimensions to solve BSDEs with a two-dimensional underlying process. This extension is necessary for application of a simplification of the BCOS method, the DCOS method, to the SABR model. Since no analytical expression for the bivariate characteristic function of the SABR model is available, we use the bivariate characteristic function of the discretized SABR model, where the underlying FSDEs are approximated by one of the following Taylor schemes: the Euler, Milstein, adjusted-Predictor-Corrector or 2.0-weak-Taylor scheme. Besides European options, some path-dependent options can be priced with the DCOS method. The change in procedure for the DCOS method is small for pricing Bermudan or discretely monitored barrier options.
Because of the the scaling symmetry of H. Park [20], it is possible to price European options under the SABR model for multiple strikes in one computation. We advise to use this scaling symmetry for strike values that are close to each other to retain the same accuracy.

We recommend to use the Euler scheme or the 2.0-weak-Tayor scheme for pricing European options under the SABR model with the DCOS method. The 2.0-weak-Tayor scheme has second order convergence, while for the other schemes only first order convergence can be achieved. As a consequence, for the 2.0-weak-Taylor scheme we need the smallest number of time steps to gain a certain accuracy. Unfortunately, the DCOS method is not fast yet. The characteristic function of the SABR model, discretized with the Euler scheme, is significantly simpler than the one determined with one of the other schemes, which ensures that the Euler scheme uses the least amount of CPU time per time step.
Also, we suggest to use the logarithmic transformations for both the forward and the volatility processes before applying the DCOS method. This ensures not only faster convergence for European option prices under the SABR model, but it also prevents the occurrence of a positive probability for a negative volatility.

7.2 Outlook

In this section we present three suggestions for future research.

Firstly, we would like to speed up the DCOS method. We are sure that we can reduce the required computation time significantly by using GPU programming in C or C++, because the DCOS method uses a large number of parallel computations.

When the DCOS method is fast, we can extent the DCOS method to a method that valuates options for which the underlying process has time-dependent parameters. Because of this time-dependency, we expect that more time steps are needed to gain a certain accuracy than without the time-dependency. The DCOS method is feasible for time-dependent parameters when it is fast.

In this thesis, we discussed the DCOS method for European, Bermudan and discretely monitored barrier options. As a final suggestion, we recommend to investigate other path-dependent options under the SABR model, e.g. Asian and American options.

Bibliography

[1] J. Andreasen and B. Huge. Expanded forward volatility. *Risk*, pages 101–107, January 2013.

[2] A. Antonov, M. Konikov, and M. Spector. SABR spreads its wings. *Risk*, pages 58–63, August 2013.

[3] A. Antonov and M. Spector. Advanced analytics for the SABR model. *SSRN paper*, March 2012.

[4] K.E. Atkinson. *An Introduction to Numerical Analysis*. John Wiley and Sons, second edition, 1989.

[5] P. Balland and Q. Tran. SABR goes normal. *Risk*, pages 76–81, June 2013.

[6] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, pages 673–654, May-June 1973.

[7] N. Bruti-Liberati and E. Platen. Strong Predictor-Corrector Euler Methods for Stochastic Differential Equations, July 2008.

[8] B. Dupire. Pricing with a smile. *Risk*, pages 18–20, January 1994.

[9] F. Fang and C.W. Oosterlee. A Novel Pricing Method for European Option Based on Fourier-Cosine Series Expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2008.

[10] F. Fang and C.W. Oosterlee. Pricing Early-Exercise and Discrete Barrier Options by Fourier-Cosine Series Expansions. *Numerische Mathematik*, 114(1):27–62, 2009.

[11] F. le Floc'h and G. Kennedy. Finite Difference Techniques for Arbitrage Free SABR. *SSRN paper*, January 2014.

[12] L.A. Grzelak. *Equity and Foreign Exchange Hybrid Models for Pricing Long-Maturity Financial Derivatives*. PhD thesis, TU Delft, 2011.

[13] L.A. Grzelak and C.W. Oosterlee. An Arbitrage-Free Hagan Implied Density via the Stochastic Collocation Method. *SSRN paper*, November 2014.

[14] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward. Managing Smile Risk. *Wilmott Magazine*, pages 84–108, September 2002.

[15] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward. Arbitrage-Free SABR. *Wilmott Magazine*, pages 60–75, January 2014.

[16] Z. van der Have. Pricing methods under the SABR model. *Internship report*, June 2014.

[17] J.C. Hull. *Options, Futures, and Other Derivatives.* Pearson Prentice Hall, sixth edition, 2006.

[18] P.E. Kloeden and E. Platen. *Numerical Solution of Stocastic Differential Equations.* Springer, first edition, 1992.

[19] J. Obłój. Fine-tune your smile correction to Hagan et al. *Wilmott magazine*, pages 102–109, May 2008.

[20] H. Park. Effcient Valuation Method for the SABR Model. *SSRN paper*, November 2013.

[21] R. Rebonato, K. McKay, and R. White. *The SABR/LIBOR Market Model.* Wiley, 2009.

[22] F.D. Rouah. Using the Risk Neutral Density to Verify No Arbitrage in Implied Volatility.

[23] M.J. Ruijter and C.W. Oosterlee. Two-dimensional Fourier Cosine Series Expansion Method for Pricing Financial Options. *SIAM Journal on Scientific Computing*, 34(5):B642–B671, 2012.

[24] M.J. Ruijter and C.W. Oosterlee. Numerical Fourier Method and second-order Taylor Scheme for Backward SDEs in Finance. *SSRN paper*, September 2014.

[25] M.J. Ruijter and C.W. Oosterlee. A Fourier-cosine method for an efficient computation of solutions to BSDEs. *To be published in SIAM Journal on Scientific Computing*, 2015.

[26] S.E. Shreve. *Stochastic Calculus for Finance II.* Springer, first edition, 2008.

[27] C. Vuik and D.J.P. Lahaye. Scientific Computing (wi4201). Technical report, TU Delft, 2013.

Appendix A

Derivation of the approximation for $G(t, s)$

The authors in [2] gave the following formula

$$G(t,s) \approx \sqrt{\frac{\sinh(s)}{s}} e^{-\frac{s^2}{2t}-\frac{t}{8}} \left[R(t,s) + \delta R(t,s)\right], \tag{A.1}$$

where

$$\begin{aligned} R(t,s) &= 1 + \frac{3tg(s)}{8s^2} - \frac{5t^2\left(-8s^2+3g^2(s)+24g(s)\right)}{128s^4} \\ &+ \frac{35t^3\left(-40s^2+3g^3(s)+24g^2(s)+120g(s)\right)}{1024s^6}, \end{aligned} \tag{A.2a}$$

$$\delta R(t,s) = e^{\frac{t}{8}} - \frac{3072+384t+24t^2+t^3}{3072}, \tag{A.2b}$$

$$g(s) = s\coth(s) - 1, \tag{A.2c}$$

as an approximation for

$$G(t,s) = \frac{2e^{-\frac{t}{8}}}{t\sqrt{\pi t}} \int_s^\infty u\, e^{-\frac{u^2}{2t}} \sqrt{\cosh(u)-\cosh(s)}\ \mathrm{d}u. \tag{A.3}$$

The derivation of the leading term in the approximation is given in [3], where $R(t,s) = 1$. In this appendix, we give the derivation of (A.1). This derivation uses integration by the following substitution:

$$u = \sqrt{s^2+w^2} \quad \text{and} \quad \mathrm{d}u = \frac{w}{\sqrt{s^2+w^2}}\mathrm{d}w. \tag{A.4}$$

A Taylor series expansion around $w = 0$ gives

$$u = \sqrt{s^2+w^2} = s + \frac{w^2}{2s} - \frac{w^4}{8s^3} + \frac{w^6}{16s^5} - \frac{5w^8}{128s^7} + \mathcal{O}\left(s^{-9}\right), \tag{A.5}$$

and

$$\begin{aligned} \cosh(u) &= \cosh\left(s + \frac{w^2}{2s} - \frac{w^4}{8s^3} + \frac{w^6}{16s^5} - \frac{5w^8}{128s^7} + \mathcal{O}\left(s^{-9}\right)\right) \\ &= \frac{1}{2}e^s e^{\frac{w^2}{2s}-\frac{w^4}{8s^3}+\frac{w^6}{16s^5}-\frac{5w^8}{128s^7}+\mathcal{O}(s^{-9})} + \frac{1}{2}e^{-s}e^{-\frac{w^2}{2s}+\frac{w^4}{8s^3}-\frac{w^6}{16s^5}+\frac{5w^8}{128s^7}+\mathcal{O}(s^{-9})} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}e^{s}\left[1+\frac{w^2}{2s}-\frac{w^4}{8s^3}+\frac{w^6}{16s^5}-\frac{5w^8}{128s^7}+\frac{1}{2}\left(\frac{w^4}{4s^2}+\frac{w^8}{64s^6}-\frac{w^6}{8s^4}+\frac{w^8}{16s^6}\right)\right.\\
&+ \frac{1}{3!}\left(\frac{w^6}{8s^3}-\frac{3w^8}{32s^5}+\frac{3w^{10}}{128s^7}+\frac{3w^{10}}{64s^7}\right)+\frac{1}{4!}\left(\frac{w^8}{16s^4}-\frac{4w^{10}}{64s^6}\right)+\frac{1}{5!}\left(\frac{w^{10}}{32s^5}-\frac{5w^{12}}{128s^7}\right)\\
&+ \left.\frac{w^{12}}{6!64s^6}+\frac{w^{14}}{7!128s^7}+\mathcal{O}\left(s^{-8}\right)\right]+\frac{1}{2}e^{-s}\left[1-\frac{w^2}{2s}+\frac{w^4}{8s^3}-\frac{w^6}{16s^5}+\frac{5w^8}{128s^7}\right.\\
&+ \frac{1}{2}\left(\frac{w^4}{4s^2}+\frac{w^8}{64s^6}-\frac{w^6}{8s^4}+\frac{w^8}{16s^6}\right)-\frac{1}{3!}\left(\frac{w^6}{8s^3}-\frac{3w^8}{32s^5}+\frac{3w^{10}}{128s^7}+\frac{3w^{10}}{64s^7}\right)\\
&+ \left.\frac{1}{4!}\left(\frac{w^8}{16s^4}-\frac{4w^{10}}{64s^6}\right)-\frac{1}{5!}\left(\frac{w^{10}}{32s^5}-\frac{5w^{12}}{128s^7}\right)+\frac{w^{12}}{6!64s^6}-\frac{w^{14}}{7!128s^7}+\mathcal{O}\left(s^{-8}\right)\right]\\
&= \cosh(s)\left[1+\frac{w^4}{8s^2}+\left(-\frac{w^6}{16}+\frac{w^8}{4!16}\right)\frac{1}{s^4}+\left(\frac{5w^8}{128}-\frac{w^{10}}{4!16}+\frac{w^{12}}{6!64}\right)\frac{1}{s^6}\right]\\
&+ \sinh(s)\left[\frac{w^2}{2s}+\left(-\frac{w^4}{8}+\frac{w^6}{48}\right)\frac{1}{s^3}+\left(\frac{w^6}{16}-\frac{w^8}{64}+\frac{w^{10}}{5!32}\right)\frac{1}{s^5}\right.\\
&+ \left.\left(-\frac{5w^8}{128}+\frac{3w^{10}}{256}-\frac{w^{12}}{4!128}+\frac{w^{14}}{7!128}\right)\frac{1}{s^7}+\mathcal{O}\left(s^{-8}\right)\right]. \qquad \text{(A.6)}
\end{aligned}
$$

We substitute this result in $\sqrt{\cosh(u)-\cosh(s)}$ and remove all terms w^n under the square root sign, where $n>8$. Thereafter, a Taylor series expansion around $\frac{1}{s}=0$ gives

$$
\begin{aligned}
\sqrt{\cosh(u)-\cosh(s)} &= |w|\sqrt{\frac{\sinh(s)}{2s}}\sqrt{1+\left(-\frac{w^2}{4}+\frac{w^4}{24}\right)\frac{1}{s^2}+\left(\frac{w^4}{8}-\frac{w^6}{32}\right)\frac{1}{s^4}-\frac{5w^6}{64s^6}}\\
&\overline{\;+\coth(s)\left[\frac{w^2}{4s}+\left(-\frac{w^4}{8}+\frac{w^6}{192}\right)\frac{1}{s^3}+\frac{5w^6}{64s^5}\right]+\mathcal{O}\left(s^{-4}\right)}\\
&= |w|\sqrt{\frac{\sinh(s)}{2s}}\left[1+\frac{a_1}{2s}+\frac{4a_2-a_1^2}{8s^2}+\frac{(a_1^3-4a_1a_2+8a_3)}{16s^3}\right.\\
&+ \left(\frac{9a_1^2a_2}{2}-6a_1a_3-3a_2^2+12a_4\right)\frac{1}{4!s^4}+\left(\frac{45a_1a_2^2}{2}-30a_1a_4-30a_2a_3+60a_5\right)\frac{1}{5!s^5}\\
&+ \left.\left(45a_2^3-180a_2a_4+360a_6\right)\frac{1}{6!s^6}+\mathcal{O}\left(s^{-4}\right)\right], \qquad \text{(A.7)}
\end{aligned}
$$

where $a_1=\frac{w^2\coth(s)}{4}$, $a_2=-\frac{w^2}{4}+\frac{w^4}{24}$, $a_3=\left(-\frac{w^4}{8}+\frac{w^6}{192}\right)\coth(s)$, $a_4=\frac{w^4}{8}-\frac{w^6}{32}$, $a_5=\frac{5w^6\coth(s)}{64}$ and $a_6=-\frac{5w^6}{64}$. Rewriting (A.7) and removing all terms w^n in the term $[\cdots]$, for $n>6$, leads to

$$
\sqrt{\cosh(u)-\cosh(s)} = |w|\sqrt{\frac{\sinh(s)}{2s}}\left[1+A_1w^2+A_2w^4+A_3w^6+\mathcal{O}\left(s^{-4}\right)\right], \qquad \text{(A.8)}
$$

where

$$
\begin{aligned}
A_1 &= \frac{\coth(s)}{8s}-\frac{1}{8s^2}, && \text{(A.9a)}\\
A_2 &= -\frac{\coth^2(s)}{128s^2}+\frac{1}{48s^2}-\frac{3\coth(s)}{64s^3}+\frac{7}{128s^4}, && \text{(A.9b)}\\
A_3 &= \frac{\coth^3(s)}{1024s^3}+\frac{5\coth^2(s)}{1024s^4}-\frac{5}{384s^4}+\frac{27\coth(s)}{1024s^5}-\frac{33}{1024s^6}. && \text{(A.9c)}
\end{aligned}
$$

Using the moments of a normal distribution with mean 0 and variance t, one can obtain

$$\int_0^\infty e^{-\frac{w^2}{2t}} w^n \, \mathrm{d}w = n!! \sqrt{\pi/2}\, t^{\frac{n+1}{2}}, \qquad \text{for } n = \text{even}, \tag{A.10}$$

where $n!!$ denotes the double factorial, for even n it holds that $n!! = \prod_{i=1}^{n/2}(2i-1)$.

Now, the approximation (A.1) for (A.3) can be found

$$\begin{aligned} G(t,s) &= \frac{2e^{-\frac{t}{8}}}{t\sqrt{\pi t}} \int_s^\infty u\, e^{-\frac{u^2}{2t}} \sqrt{\cosh(u) - \cosh(s)}\, \mathrm{d}u \\ &= \frac{\sqrt{2} e^{-\frac{s^2}{2t} - \frac{t}{8}}}{t\sqrt{\pi t}} \sqrt{\frac{\sinh(s)}{s}} \int_0^\infty w^2 e^{-\frac{w^2}{2t}} \left(1 + A_1 w^2 + A_2 w^4 + A_3 w^6 + \mathcal{O}\left(s^{-4}\right)\right) \mathrm{d}w \\ &= e^{-\frac{s^2}{2t} - \frac{t}{8}} \sqrt{\frac{\sinh(s)}{s}} \left[1 + 3tA_1 + 15t^2 A_2 + 105 t^3 A_3 + \mathcal{O}\left(s^{-4}\right)\right] \\ &= \sqrt{\frac{\sinh(s)}{s}} e^{-\frac{s^2}{2t} - \frac{t}{8}} \left[R(t,s) + \mathcal{O}\left(s^{-4}\right)\right]. \end{aligned} \tag{A.11}$$

Here, $R(t,s)$ is as in formula (A.1). The authors in [2, 3] added $\delta R(t,s)$ as a correction to ensure $G(t,0) = 1$.

Appendix B

Hagan's arbitrage-free method

In this appendix we give the derivation of PDE (2.27), the proof of $M(t,F) \geq 0$ (2.28), and the Crank-Nicolson scheme to solve (2.27) in matrix form.

Derivation of Hagan PDE (2.27)
Using perturbation techniques the authors in [15] derived PDE (2.27). The start of this derivation is analogously to the start of the derivation of the Hagan formula which is given in [16]. Assume that both the volatility σ and the vol-vol ν are small and write $F_t \to \widehat{F}_t$, $\sigma_t \to \epsilon\widehat{\sigma}_t$ and $\nu \to \epsilon\nu$, where $\epsilon > 0$ is small. The system of FSDEs (2.1)-(2.2) becomes

$$\mathrm{d}\widehat{F}_t = \epsilon\widehat{\sigma}_t \left(\widehat{F}_t\right)^{\beta} \mathrm{d}W_t^1, \tag{B.1}$$

$$\mathrm{d}\widehat{\sigma}_t = \epsilon\nu\widehat{\sigma}_t \,\mathrm{d}W_t^2, \tag{B.2}$$

$$\mathrm{d}W_t^1 \,\mathrm{d}W_t^2 = \rho \,\mathrm{d}t.$$

Define the probability density function $p\left(t,f,\alpha;s,F,\sigma\right)$ by[1]

$$p\left(t,f,\alpha;s,F,\sigma\right)\mathrm{d}F\,\mathrm{d}\sigma = \mathrm{prob}\left\{F < \widehat{F}_s < F + \mathrm{d}F, \sigma < \widehat{\sigma}_s < \sigma + \mathrm{d}\sigma \,\middle|\, \widehat{F}_t = f,\ \widehat{\sigma}_t = \alpha\right\}, \tag{B.3}$$

where $p\left(t,f,\sigma;t,F,\sigma\right) = \delta\left(F-f\right)\delta\left(\sigma-\alpha\right)$.

As given in [16], p satisfies the Fokker-Planck equation which leads to the PDE

$$\begin{cases} \frac{\partial p}{\partial s} = \frac{1}{2}\epsilon^2\sigma^2\frac{\partial^2}{\partial F^2}\left[F^{2\beta}p\right] + \frac{1}{2}\epsilon^2\nu^2\frac{\partial^2}{\partial\sigma^2}\left[\sigma^2 p\right] + \epsilon^2\nu\rho\frac{\partial^2}{\partial F\partial\sigma}\left[\sigma^2F^{\beta}p\right], & \text{for } s > t, \\ p = \delta\left(F-f\right)\delta\left(\sigma-\alpha\right), & \text{at } s = t. \end{cases} \tag{B.4}$$

Define for $k = 0, 1, 2, \ldots$

$$Q^{(k)} = Q^{(k)}\left(t,f,\alpha;s,F\right) = \int_{-\infty}^{\infty} \sigma^k p\left(t,f,\alpha;s,F,\sigma\right)\,\mathrm{d}\sigma. \tag{B.5}$$

We showed in [16, Appendix D] that $Q^{(2)}$ satisfies the Kolmogorov backward equation. With a similar proof, one can show that $Q^{(k)}$ satisfies the Kolmogorov backward equation for all $k \in \mathbb{N}\cup\{0\}$, which leads to the PDE

$$\begin{cases} \frac{\partial Q^{(k)}}{\partial t} = -\frac{1}{2}\epsilon^2\alpha^2 f^{2\beta}\frac{\partial^2 Q^{(k)}}{\partial f^2} - \frac{1}{2}\epsilon^2\nu^2\alpha^2\frac{\partial^2 Q^{(k)}}{\partial\alpha^2} - \epsilon^2\nu\rho\alpha^2 f^{\beta}\frac{\partial^2 Q^{(k)}}{\partial f\partial\alpha}, & \text{for } t < s, \\ Q^{(k)} = \alpha^k\delta\left(F-f\right), & \text{at } t = s. \end{cases} \tag{B.6}$$

[1] In Chapter 4 we denoted this probability density function by $Q_{F_s,\sigma_s}\left(s-t,F,\sigma|f,\alpha\right)$. In this appendix we choose to use another notation to avoid confusion.

Using the PDE for p, interchanging integration and differentiation and [16, Formula (D.5)] gives the following relation between $Q^{(0)}$ and $Q^{(2)}$, for $s > t$,

$$\frac{\partial p}{\partial s} = \frac{1}{2}\epsilon^2\sigma^2\frac{\partial^2}{\partial F^2}\left[F^{2\beta}p\right] + \frac{1}{2}\epsilon^2\nu^2\frac{\partial^2}{\partial \sigma^2}\left[\sigma^2 p\right] + \epsilon^2\nu\rho\frac{\partial^2}{\partial F\partial\sigma}\left[\sigma^2F^\beta p\right]$$

$$\Rightarrow$$

$$\int_{-\infty}^{\infty}\frac{\partial p}{\partial s}\ \mathrm{d}\sigma = \frac{1}{2}\epsilon^2\int_{-\infty}^{\infty}\sigma^2\frac{\partial^2}{\partial F^2}\left[F^{2\beta}p\right]\ \mathrm{d}\sigma + \frac{1}{2}\epsilon^2\nu^2\int_{-\infty}^{\infty}\frac{\partial^2}{\partial\sigma^2}\left[\sigma^2 p\right]\ \mathrm{d}\sigma + \epsilon^2\nu\rho\int_{-\infty}^{\infty}\frac{\partial^2}{\partial F\partial\sigma}\left[\sigma^2F^\beta p\right]\ \mathrm{d}\sigma$$

$$\Rightarrow$$

$$\frac{\partial}{\partial s}\left(\int_{-\infty}^{\infty} p\ \mathrm{d}\sigma\right) = \frac{1}{2}\epsilon^2\frac{\partial^2}{\partial F^2}\left(\int_{-\infty}^{\infty}\sigma^2F^{2\beta}p\ \mathrm{d}\sigma\right) + 0 + 0$$

$$\Rightarrow$$

$$\frac{\partial Q^{(0)}}{\partial s} = \frac{1}{2}\epsilon^2\frac{\partial^2}{\partial F^2}\left[F^{2\beta}Q^{(2)}\right]. \tag{B.7}$$

The goal of the authors in [15] was to write $Q^{(2)}$ in terms of $Q^{(0)}$. For this analysis we write $\tau = s - t$ and $\tau_{ex} = T - t$, and we take

$$z = \frac{1}{\epsilon\alpha}\int_f^F \frac{1}{\overline{f}^\beta}\ \mathrm{d}\overline{f}, \tag{B.8}$$

and define

$$B(\epsilon\alpha z) = f^\beta. \tag{B.9}$$

One can notice that $B(0) = F^\beta$.

For the partial derivatives it holds that

$$\begin{aligned}
\frac{\partial}{\partial f} &\rightarrow \frac{\partial}{\partial z}\frac{\partial z}{\partial f} &&= \frac{-1}{\epsilon\alpha B(\epsilon\alpha z)}\frac{\partial}{\partial z},\\
\frac{\partial^2}{\partial f^2} &\rightarrow \left(\frac{-1}{\epsilon\alpha B(\epsilon\alpha z)}\frac{\partial}{\partial z}\right)^2 &&= \frac{1}{\epsilon^2\alpha^2B(\epsilon\alpha z)^2}\frac{\partial^2}{\partial z^2} - \frac{B'(\epsilon\alpha z)}{\epsilon\alpha B(\epsilon\alpha z)^3}\frac{\partial}{\partial z},\\
\frac{\partial}{\partial\alpha} &\rightarrow \frac{\partial}{\partial\alpha} + \frac{\partial}{\partial z}\frac{\partial z}{\partial\alpha} &&= \frac{\partial}{\partial\alpha} - \frac{z}{\alpha}\frac{\partial}{\partial z},\\
\frac{\partial^2}{\partial\alpha^2} &\rightarrow \left(\frac{\partial}{\partial\alpha} - \frac{z}{\alpha}\frac{\partial}{\partial z}\right)^2 &&= \frac{\partial^2}{\partial\alpha^2} + \frac{2z}{\alpha^2}\frac{\partial}{\partial z} - \frac{2z}{\alpha}\frac{\partial^2}{\partial z\partial\alpha} + \frac{z^2}{\alpha^2}\frac{\partial^2}{\partial z^2},\\
\frac{\partial^2}{\partial f\partial\alpha} &\rightarrow \frac{-1}{\epsilon\alpha B(\epsilon\alpha z)}\frac{\partial}{\partial z}\left(\frac{\partial}{\partial\alpha} - \frac{z}{\alpha}\frac{\partial}{\partial z}\right) &&= \frac{1}{\epsilon\alpha B(\epsilon\alpha z)}\left(-\frac{\partial^2}{\partial z\partial\alpha} + \frac{1}{\alpha}\frac{\partial}{\partial z} + \frac{z}{\alpha}\frac{\partial^2}{\partial z^2}\right),
\end{aligned}$$

which leads to the PDE

$$\begin{cases}
\frac{\partial Q^{(k)}}{\partial\tau} &= \left[\frac{1}{2} + z\epsilon\rho\nu + \frac{1}{2}\epsilon^2\nu^2z^2\right]\frac{\partial^2Q^{(k)}}{\partial z^2} + \left[-\frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)} + \epsilon^2\nu^2z + \epsilon\nu\rho\right]\frac{\partial Q^{(k)}}{\partial z}\\
&\quad - \left[\epsilon\nu\rho\alpha + z\epsilon^2\nu^2\alpha\right]\frac{\partial^2Q^{(k)}}{\partial z\partial\alpha} + \frac{1}{2}\epsilon^2\nu^2\alpha^2\frac{\partial^2Q^{(k)}}{\partial\alpha^2}, & \text{for } \tau > 0,\\
Q^{(k)} &= \frac{\alpha^{k-1}}{\epsilon F^\beta}\delta(z), & \text{at } \tau = 0.
\end{cases} \tag{B.10}$$

We define $\widehat{Q}^{(k)}(\tau, z, \alpha) = \frac{\epsilon F^\beta}{\alpha^{k-1}}Q^{(k)}(\tau, z, \alpha)$, then

$$\begin{cases}
\frac{\partial Q^{(k)}}{\partial\tau} &= \frac{\alpha^{k-1}}{\epsilon F^\beta}\frac{\partial\widehat{Q}^{(k)}}{\partial\tau},\\
\frac{\partial^2Q^{(k)}}{\partial z\partial\alpha} &= \frac{(k-1)\alpha^{k-2}}{\epsilon F^\beta}\frac{\partial\widehat{Q}^{(k)}}{\partial z} + \frac{\alpha^{k-1}}{\epsilon F^\beta}\frac{\partial^2\widehat{Q}^{(k)}}{\partial z\partial\alpha},\\
\frac{\partial^2Q^{(k)}}{\partial\alpha^2} &= \frac{(k-2)(k-1)\alpha^{k-3}}{\epsilon F^\beta}\widehat{Q}^{(k)} + \frac{2(k-1)\alpha^{k-2}}{\epsilon F^\beta}\frac{\partial\widehat{Q}^{(k)}}{\partial\alpha} + \frac{\alpha^{k-1}}{\epsilon F^\beta}\frac{\partial^2\widehat{Q}^{(k)}}{\partial\alpha^2}.
\end{cases} \tag{B.11}$$

$\widehat{Q}^{(k)}(\tau,z,\alpha)$ is the solution of the PDE

$$\begin{cases} \frac{\partial \widehat{Q}^{(k)}}{\partial \tau} & = \left[\frac{1}{2}+z\epsilon\rho\nu+\frac{1}{2}\epsilon^2\nu^2z^2\right]\frac{\partial^2\widehat{Q}^{(k)}}{\partial z^2}-\left[\frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}+\left(\epsilon^2\nu^2 z+\epsilon\nu\rho\right)(k-2)\right]\frac{\partial\widehat{Q}^{(k)}}{\partial z} \\ & -\left[\epsilon\nu\rho\alpha+z\epsilon^2\nu^2\alpha\right]\frac{\partial^2\widehat{Q}^{(k)}}{\partial z\partial\alpha}+\frac{1}{2}\epsilon^2\nu^2\left[\alpha^2\frac{\partial^2 Q^{(k)}}{\partial\alpha^2}+(k-2)(k-1)\widehat{Q}^{(k)}\right. \\ & \left.+\,2(k-1)\alpha\frac{\partial\widehat{Q}^{(k)}}{\partial\alpha}\right], & \text{for } \tau>0, \\ \widehat{Q}^{(k)} & =\delta(z), & \text{at } \tau=0. \end{cases} \tag{B.12}$$

As the authors in [15] noticed, α does not enter the PDE for $\widehat{Q}^{(k)}$ until $\mathcal{O}(\epsilon)$ approximation, so $\frac{\partial\widehat{Q}^{(k)}}{\partial\alpha}$, $\frac{\partial^2\widehat{Q}^{(k)}}{\partial z\alpha}$ and $\frac{\partial^2\widehat{Q}^{(k)}}{\partial\alpha^2}$ are all $\mathcal{O}(\epsilon)$. The PDE can be reduced to

$$\begin{cases} \frac{\partial \widehat{Q}^{(k)}}{\partial \tau} & = \left[\frac{1}{2}+z\epsilon\rho\nu+\frac{1}{2}\epsilon^2\nu^2z^2\right]\frac{\partial^2\widehat{Q}^{(k)}}{\partial z^2}-\left[\frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}+\left(\epsilon^2\nu^2 z+\epsilon\nu\rho\right)(k-2)\right]\frac{\partial\widehat{Q}^{(k)}}{\partial z} \\ & -\epsilon\nu\rho\alpha\frac{\partial^2\widehat{Q}^{(k)}}{\partial z\partial\alpha}+\frac{1}{2}\epsilon^2\nu^2(k-2)(k-1)\widehat{Q}^{(k)}+\mathcal{O}\left(\epsilon^3\right), & \text{for } \tau>0, \\ \widehat{Q}^{(k)} & =\delta(z), & \text{at } \tau=0. \end{cases}$$

The aim of this analysis is to write $Q^{(2)}$ in terms of $Q^{(0)}$, and one can observe that when $\tau>0$ and when $k=0$ or $k=2$ it holds that

$$\begin{aligned} \frac{\partial\widehat{Q}^{(0)}}{\partial\tau} &= \left[\frac{1}{2}+z\epsilon\rho\nu+\frac{1}{2}\epsilon^2\nu^2z^2\right]\frac{\partial^2\widehat{Q}^{(0)}}{\partial z^2}+\left[-\frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}+2\left(\epsilon^2\nu^2z+\epsilon\nu\rho\right)\right]\frac{\partial\widehat{Q}^{(0)}}{\partial z} \\ &- \epsilon\nu\rho\alpha\frac{\partial^2\widehat{Q}^{(0)}}{\partial z\partial\alpha}+\epsilon^2\nu^2\widehat{Q}^{(0)}+\mathcal{O}\left(\epsilon^3\right) \qquad \text{(B.13)} \\ &= \frac{\partial^2}{\partial z^2}\left[\left(\frac{1}{2}+z\epsilon\rho\nu+\frac{1}{2}\epsilon^2\nu^2z^2\right)\widehat{Q}^{(0)}\right]-\frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}\frac{\partial\widehat{Q}^{(0)}}{\partial z}-\epsilon\nu\rho\alpha\frac{\partial^2\widehat{Q}^{(0)}}{\partial z\partial\alpha}+\mathcal{O}\left(\epsilon^3\right), \\ \frac{\partial\widehat{Q}^{(2)}}{\partial\tau} &= \left[\frac{1}{2}+z\epsilon\rho\nu+\frac{1}{2}\epsilon^2\nu^2z^2\right]\frac{\partial^2\widehat{Q}^{(2)}}{\partial z^2}-\frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}\frac{\partial\widehat{Q}^{(2)}}{\partial z}-\epsilon\nu\rho\alpha\frac{\partial^2\widehat{Q}^{(2)}}{\partial z\partial\alpha}+\mathcal{O}\left(\epsilon^3\right), \qquad \text{(B.14)} \end{aligned}$$

with initial conditions

$$\widehat{Q}^{(0)}(\tau,z,\alpha)=\delta(z) \quad \text{and} \quad \widehat{Q}^{(2)}(\tau,z,\alpha)=\delta(z), \qquad \text{at } \tau=0. \tag{B.15}$$

The PDEs for $\widehat{Q}^{(0)}$ and $\widehat{Q}^{(2)}$ are very similar and both functions have the same initial condition. The next step is to transform $\widehat{Q}^{(0)}$ in a few steps such that this final transformed quantity satisfies the PDE of $\widehat{Q}^{(2)}$ in $\mathcal{O}\left(\epsilon^2\right)$ accuracy, because then this transformed quantity equals $\widehat{Q}^{(2)}$ up to $\mathcal{O}\left(\epsilon^2\right)$.

For the first transformation, we define

$$\widehat{U}(\tau,z,\alpha)=\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\widehat{Q}^{(0)}(\tau,z,\alpha). \tag{B.16}$$

For the partial derivatives it holds that

$$\begin{cases} \left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial\widehat{Q}^{(0)}}{\partial\tau} & =\frac{\partial\widehat{U}}{\partial\tau}, \\ \left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial\widehat{Q}^{(0)}}{\partial z} & =\frac{\partial\widehat{U}}{\partial z}-\frac{2\epsilon\rho\nu+2\epsilon^2\nu^2z}{1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2}\widehat{U}, \\ \left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial^2\widehat{Q}^{(0)}}{\partial z\partial\alpha} & =\frac{\partial^2\widehat{U}}{\partial z\partial\alpha}-\frac{2\epsilon\rho\nu+2\epsilon^2\nu^2z}{1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2}\frac{\partial\widehat{U}}{\partial\alpha}, \\ \frac{\partial^2\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\widehat{Q}^{(0)}}{\partial z^2} & =\frac{\partial^2\widehat{U}}{\partial z^2}. \end{cases} \tag{B.17}$$

We obtain with the geometric series, i.e.

$$\frac{1}{1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2}=1-2z\epsilon\rho\nu+\mathcal{O}\left(\epsilon^2\right). \tag{B.18}$$

The PDE of $\widehat{U}$ is now given by

$$\begin{aligned}
\frac{\partial \widehat{U}}{\partial \tau} &= \frac{1}{2}\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial^2 \widehat{U}}{\partial z^2} - \frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}\frac{\partial \widehat{U}}{\partial z} + \frac{\epsilon\alpha B'(\epsilon\alpha z)(2\epsilon\rho\nu+2\epsilon^2\nu^2 z)}{2B(\epsilon\alpha z)\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)}\widehat{U} \\
&- \epsilon\rho\nu\alpha\frac{\partial^2 \widehat{U}}{\partial z\partial\alpha} + \frac{\epsilon\rho\nu\alpha(2\epsilon\rho\nu+2\epsilon^2\nu^2 z)}{\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)}\frac{\partial \widehat{U}}{\partial \alpha} + \mathcal{O}\left(\epsilon^3\right) \\
&= \frac{1}{2}\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial^2 \widehat{U}}{\partial z^2} - \frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}\frac{\partial \widehat{U}}{\partial z} + \frac{\epsilon^2\alpha B'(\epsilon\alpha z)\rho\nu}{B(\epsilon\alpha z)}\widehat{U} \\
&- \epsilon\rho\nu\alpha\frac{\partial^2 \widehat{U}}{\partial z\partial\alpha} + 2\epsilon^2\rho^2\nu^2\alpha\frac{\partial \widehat{U}}{\partial \alpha} + \mathcal{O}\left(\epsilon^3\right). \qquad \text{(B.19)}
\end{aligned}$$

We observe that α does not enter the PDE for $\widehat{U}$ up to $\mathcal{O}(\epsilon)$, so $\frac{\partial \widehat{U}}{\partial \alpha}$ is $\mathcal{O}(\epsilon)$ accurate. The PDE can be reduced to

$$\begin{cases}
\frac{\partial \widehat{U}}{\partial \tau} = \frac{1}{2}\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial^2 \widehat{U}}{\partial z^2} - \frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}\frac{\partial \widehat{U}}{\partial z} + \frac{\epsilon^2\alpha B'(\epsilon\alpha z)\rho\nu}{B(\epsilon\alpha z)}\widehat{U} & \\
\quad -\epsilon\rho\nu\alpha\frac{\partial^2 \widehat{U}}{\partial z\partial\alpha} + \mathcal{O}\left(\epsilon^3\right), & \text{for } \tau>0, \\
\widehat{U} = \delta(z), & \text{at } \tau=0.
\end{cases} \qquad \text{(B.20)}$$

For the second transformation we define

$$\Gamma = -\frac{B'(\epsilon\alpha z)}{B(\epsilon\alpha z)}, \qquad \text{(B.21)}$$

$$U(\tau,z,\alpha) = e^{\epsilon^2\rho\nu\alpha\Gamma\tau}\widehat{U}(\tau,z,\alpha). \qquad \text{(B.22)}$$

One can notice that

$$\begin{aligned}
\frac{\partial\Gamma}{\partial z} &= \mathcal{O}(\epsilon), & \frac{\partial^2\Gamma}{\partial z^2} &= \mathcal{O}\left(\epsilon^2\right), \\
\frac{\partial\Gamma}{\partial \alpha} &= \mathcal{O}(\epsilon), & \frac{\partial^2\Gamma}{\partial z\partial\alpha} &= \mathcal{O}\left(\epsilon^2\right).
\end{aligned} \qquad \text{(B.23)}$$

For the partial derivatives it holds that

$$\begin{cases}
e^{\epsilon^2\rho\nu\alpha\Gamma\tau}\frac{\partial \widehat{U}}{\partial \tau} = \frac{\partial U}{\partial \tau} - \epsilon^2\rho\nu\alpha\Gamma U, \\
e^{\epsilon^2\rho\nu\alpha\Gamma\tau}\frac{\partial \widehat{U}}{\partial z} = \frac{\partial U}{\partial z} + \mathcal{O}\left(\epsilon^3\right), \\
e^{\epsilon^2\rho\nu\alpha\Gamma\tau}\frac{\partial^2 \widehat{U}}{\partial z^2} = \frac{\partial^2 U}{\partial z^2} + \mathcal{O}\left(\epsilon^3\right), \\
e^{\epsilon^2\rho\nu\alpha\Gamma\tau}\frac{\partial^2 \widehat{U}}{\partial z\partial\alpha} = \frac{\partial^2 U}{\partial z\partial\alpha} + \mathcal{O}\left(\epsilon^3\right).
\end{cases} \qquad \text{(B.24)}$$

This leads to the PDE

$$\begin{cases}
\frac{\partial U}{\partial \tau} = \frac{1}{2}\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2\right)\frac{\partial^2 U}{\partial z^2} - \frac{\epsilon\alpha B'(\epsilon\alpha z)}{2B(\epsilon\alpha z)}\frac{\partial U}{\partial z} - \epsilon\rho\nu\alpha\frac{\partial^2 U}{\partial z\partial\alpha} + \mathcal{O}\left(\epsilon^3\right), & \text{for } \tau>0, \\
U = \delta(z), & \text{at } \tau=0.
\end{cases} \qquad \text{(B.25)}$$

This PDE equals (B.14) through $\mathcal{O}\left(\epsilon^2\right)$. From this, one can conclude that

$$\begin{aligned}
\widehat{Q}^{(2)}(\tau,z,\alpha) &= U(\tau,z,\alpha) + \mathcal{O}\left(\epsilon^3\right) \\
&= \widehat{Q}^{(0)}(\tau,z,\alpha)e^{\epsilon^2\rho\nu\alpha\Gamma\tau}\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2+\mathcal{O}\left(\epsilon^3\right)\right). \qquad \text{(B.26)}
\end{aligned}$$

Using

$$\widehat{Q}^{(0)} = \epsilon F^\beta \alpha Q^{(0)} \qquad \text{and} \qquad \widehat{Q}^{(2)} = \frac{\epsilon F^\beta}{\alpha}Q^{(2)}, \qquad \text{(B.27)}$$

one can obtain

$$Q^{(2)}(t,f,\alpha;s,F) = \alpha^2 Q^{(0)}(t,f,\alpha;s,F)e^{\epsilon^2\rho\nu\alpha\Gamma(s-t)}\left(1+2z\epsilon\rho\nu+\epsilon^2\nu^2z^2+\mathcal{O}\left(\epsilon^3\right)\right). \qquad \text{(B.28)}$$

From the geometric series, [16, Formula (D.14)] and forward differences, one can obtain that

$$\begin{aligned}
\Gamma &= -\frac{B'(\epsilon\alpha z)}{B(\epsilon\alpha z)} && = -\frac{B(\epsilon\alpha z) - B(0)}{\epsilon\alpha z\ B(\epsilon\alpha z)} + \mathcal{O}(\epsilon) \\
&= -\frac{f^\beta - F^\beta}{\left(\int_f^F \frac{1}{\overline{f}^\beta}\ \mathrm{d}\overline{f}\right)\ f^\beta} + \mathcal{O}(\epsilon) && = -\frac{f^\beta - F^\beta}{\left(\frac{F-f}{f^\beta} + \mathcal{O}(\epsilon)\right) f^\beta} + \mathcal{O}(\epsilon) \\
&= \frac{F^\beta - f^\beta}{F - f} + \mathcal{O}(\epsilon).
\end{aligned} \tag{B.29}$$

Substitution of (B.28) in (B.7) leads to the following PDE

$$\frac{\partial Q^{(0)}}{\partial s} = \frac{1}{2}\epsilon^2\alpha^2\frac{\partial^2}{\partial F^2}\left[F^{2\beta}Q^{(0)}e^{\epsilon^2\rho\nu\alpha\Gamma(s-t)}\left(1 + 2z\epsilon\rho\nu + \epsilon^2\nu^2 z^2\right)\right], \qquad \text{for } s > t, \tag{B.30}$$

which is accurate up to $\mathcal{O}\left(\epsilon^2\right)$ and where

$$z = \frac{1}{\epsilon\alpha}\int_f^F \frac{1}{\overline{f}^\beta}\ \mathrm{d}\overline{f} = \frac{F^{1-\beta} - f^\beta}{\epsilon\alpha(1-\beta)}, \tag{B.31}$$

$$\Gamma = \frac{F^\beta - f^\beta}{F - f}. \tag{B.32}$$

Rewriting $\epsilon\alpha \to \alpha$ and $\epsilon\nu \to \nu$ gives (2.27).

Proof of $M(t, F) \geq 0$
Let $t \geq 0$ fixed. We observe that $M(t, F)$ is continuous in F and for the initial forward value f it holds that $M(t, f) > 0$.
When $F = 0$ or when $1 + 2\rho\nu z(F) + \nu^2 z^2(F) = 0$ it holds that $M(t, F) = 0$. With the quadratic formula

$$1 + 2\rho\nu z(F) + \nu^2 z^2(F) = 0 \quad \text{when } z(F) = \frac{-2\rho\nu \pm 2\nu\sqrt{\rho^2 - 1}}{2\nu^2}, \tag{B.33}$$

there is only a real solution when $|\rho| = 1$, then $z(F) = -\frac{\rho}{\nu}$.

For the case $-1 < \rho < 1$ it holds that $M(t, F) > 0$ for all $F > 0$, because $M(t, F)$ is continuous, $M(t, f) > 0$ and $M(t, F) \neq 0$ for all $F > 0$.

When $\rho = 1$, one can obtain that $M(t, F)$ is a product of non-negative functions,

$$\begin{aligned}
M(t, F) &= \frac{1}{2}\alpha^2\left(1 + 2\nu z(F) + \nu^2 z^2(F)\right)\exp\left(\rho\nu\alpha\Gamma(F)t\right)F^{2\beta} \\
&= \frac{1}{2}\alpha^2\left(1 + \nu z(F)\right)^2\exp\left(\rho\nu\alpha\Gamma(F)t\right)F^{2\beta} \\
&\geq 0, \qquad \text{for } F \geq 0.
\end{aligned} \tag{B.34}$$

Analogously, when $\rho = -1$, we find

$$\begin{aligned}
M(t, F) &= \frac{1}{2}\alpha^2\left(1 - 2\nu z(F) + \nu^2 z^2(F)\right)\exp\left(\rho\nu\alpha\Gamma(F)t\right)F^{2\beta} \\
&= \frac{1}{2}\alpha^2\left(1 - \nu z(F)\right)^2\exp\left(\rho\nu\alpha\Gamma(F)t\right)F^{2\beta}
\end{aligned}$$

$$\geq \quad 0, \qquad\qquad \text{for } F \geq 0. \tag{B.35}$$

The Crank-Nicolson scheme to solve (2.27) in matrix form is given by

$$\begin{bmatrix} 1+\frac{3\Delta t}{2h^2}M_1^{n+1} & -\frac{\Delta t}{2h^2}M_2^{n+1} & & & 0 \\ -\frac{\Delta t}{2h^2}M_1^{n+1} & 1+\frac{\Delta t}{h^2}M_2^{n+1} & -\frac{\Delta t}{2h^2}M_3^{n+1} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{\Delta t}{2h^2}M_{J-2}^{n+1} & 1+\frac{\Delta t}{h^2}M_{J-1}^{n+1} & -\frac{\Delta t}{2h^2}M_J^{n+1} \\ 0 & & & -\frac{\Delta t}{2h^2}M_{J-1}^{n+1} & 1+\frac{3\Delta t}{2h^2}M_J^{n+1} \end{bmatrix} \begin{bmatrix} Q_1^{n+1} \\ Q_2^{n+1} \\ \vdots \\ Q_{J-1}^{n+1} \\ Q_J^{n+1} \end{bmatrix} = \begin{bmatrix} \widetilde{Q}_1^n \\ \widetilde{Q}_2^n \\ \vdots \\ \widetilde{Q}_{J-1}^n \\ \widetilde{Q}_J^n \end{bmatrix} \tag{B.36}$$

where

$$\widetilde{Q}_1^n = Q_1^n + \frac{\Delta t}{2h^2}\left\{M_2^n Q_2^n - 3M_1^n Q_1^n\right\}, \tag{B.37}$$

$$\widetilde{Q}_j^n = Q_j^n + \frac{\Delta t}{2h^2}\left\{M_{j+1}^n Q_{j+1}^n - 2M_j^n Q_j^n + M_{j-1}^n Q_{j-1}^n\right\}, \quad \text{for } j = 2, \ldots, J-1, \tag{B.38}$$

$$\widetilde{Q}_J^n = Q_J^n + \frac{\Delta t}{2h^2}\left\{M_{J-1}^n Q_{J-1}^n - 3M_J^n Q_J^n\right\}. \tag{B.39}$$

To advance from time $n\Delta t$ to time $(n+1)\Delta t$ one can solve this system with the tridiagonal matrix algorithm.

Appendix C

Itô-Taylor expansion

This appendix gives formulas for the Itô-Taylor expansion in one and in two dimensions.

One-dimensional Itô-Taylor expansion
Assume we have an FSDE is of the form (3.1). First we introduce some notation [18, 24]:
We call a row vector

$$\widehat{\alpha} = \left(j_1, j_2, ..., j_{l(\widehat{\alpha})}\right), \tag{C.1}$$

where $j_i \in \{0, 1\}$ for $i \in \{1, ..., l\}$, a multi-index of length

$$l(\widehat{\alpha}) \in \{1, 2, ...\}. \tag{C.2}$$

Also,

$$\widehat{\alpha}- = \left(j_1, j_2, ..., j_{l(\widehat{\alpha})-1}\right) \quad \text{and} \quad -\widehat{\alpha} = \left(j_2, j_3, ..., j_{l(\widehat{\alpha})}\right). \tag{C.3}$$

The multi-index of length zero is denoted by v,

$$l(v) = 0. \tag{C.4}$$

The set of all multi-indices is denoted by $\mathcal{M}$,

$$\mathcal{M} = \{(j_1, j_2, ..., j_l) : j_i \in \{0, 1\}, i \in \{1, ..., l\}, \text{ for } l = 1, 2, 3, ...\} \cup v. \tag{C.5}$$

We call $\mathcal{A} \subseteq \mathcal{M}$ a hierarchical set if $\mathcal{A}$ is non-empty, if the multi-indices in $\mathcal{A}$ are uniformly bounded in length, i.e. $\sup_{\widehat{\alpha}\in\mathcal{A}} l(\widehat{\alpha}) < \infty$, and if $-\widehat{\alpha} \in \mathcal{A}$ for each $\widehat{\alpha} \in \mathcal{A} \setminus \{v\}$.
The remainder set $\mathcal{B}(\mathcal{A})$ of $\mathcal{A}$ is given by

$$\mathcal{B}(\mathcal{A}) = \{\widehat{\alpha} \in \mathcal{M} \setminus \mathcal{A} : -\widehat{\alpha} \in \mathcal{A}\}. \tag{C.6}$$

We define

$$L^0 = \frac{\partial}{\partial t} + \mu(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2} \quad \text{and} \quad L^1 = \sigma(x)\frac{\partial}{\partial x}. \tag{C.7}$$

Let $h(t, x_t)$ be a general sufficiently smooth function, we denote

$$h_{\widehat{\alpha}} = L^{j_1}L^{j_2}...L^{j_{l(\widehat{\alpha})}}h. \tag{C.8}$$

Also, we define the multiple Itô integral $I_{\widehat{\alpha}}\left[h(., x.)\right]_{s,t}$ recursively by

$$I_{\widehat{\alpha}}\left[h(., x.)\right]_{s,t} = \begin{cases} h(t, x_t), & \text{if } l(\alpha) = 0, \\ \int_s^t I_{\widehat{\alpha}-}\left[h(., x.)\right]_{s,u} \mathrm{d}u, & \text{if } l(\widehat{\alpha}) \geq 1, j_{l(\widehat{\alpha})} = 0, \\ \int_s^t I_{\widehat{\alpha}-}\left[h(., x.)\right]_{s,u} \mathrm{d}W_u, & \text{if } l(\widehat{\alpha}) \geq 1, j_{l(\widehat{\alpha})} = 1, \end{cases} \tag{C.9}$$

and we abbreviate

$$I_{\widehat{\alpha},s,t} = I_{\widehat{\alpha}}[1]_{s,t}. \tag{C.10}$$

Theorem C.1. *Let $X_{m+1}^{m,x}$ denote the value of X_{m+1} given $X_m = x$ and let $\mathcal{A} \subseteq \mathcal{M}$ be a hierarchical set. The Itô-Taylor expansion for a general sufficiently smooth function $h(t, x_t)$ is given by [18, 24]*

$$h\left(t_{m+1}, X_{m+1}^{m,x}\right) = \sum_{\widehat{\alpha}\in\mathcal{A}} h_\alpha(t_m, x) I_{\widehat{\alpha},t_m,t_{m+1}} + \sum_{\alpha\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_{.}^{m,x}\right)\right]_{t_m,t_{m+1}}. \tag{C.11}$$

By using Theorem C.1 we find the following conditional expectations of the Itô-Taylor expansion for a sufficiently smooth function $h(t, x)$

$$\begin{aligned}
\mathbb{E}\left[h\left(t_{m+1}, X_{m+1}^{m,x}\right)\right] &= h(t_m, x) + h_{(0)}(t_m, x)\Delta t + \frac{1}{2} h_{(0,0)}(t_m, x)\left(\Delta t\right)^2 + \mathcal{O}\left(\left(\Delta t\right)^3\right), &\text{(C.12)}\\
\mathbb{E}\left[\left(h\left(t_{m+1}, X_{m+1}^{m,x}\right)\right)^2\right] &= h^2(t_m, x) + \left[h_{(1)}^2(t_m, x) + 2h(t_m, x)h_{(0)}(t_m, x)\right]\Delta t\\
&+ \left[h_{(0)}^2(t_m, x) + \frac{1}{2}h_{(1,1)}^2(t_m, x) + h(t_m, x)h_{(0,0)}(t_m, x)\right]\left(\Delta t\right)^2\\
&+ h_{(1)}(t_m, x)\left[h_{(0,1)}(t_m, x) + h_{(1,0)}(t_m, x)\right]\left(\Delta t\right)^2 + \mathcal{O}\left(\left(\Delta t\right)^3\right). &\text{(C.13)}
\end{aligned}$$

We can observe that for $\mathcal{A} = \{v, (0), (1), (0,0), (0,1), (1,0), (1,1)\}$ and sufficiently smooth function $h(t, x)$ the following holds

$$\mathbb{E}\left[\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_{.}^{m,x}\right)\right]_{t_m,t_{m+1}}\right] = \mathcal{O}\left(\left(\Delta t\right)^3\right), \tag{C.14}$$

and

$$\mathbb{E}\left[\left(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_{.}^{m,x}\right)\right]_{t_m,t_{m+1}}\right)^2\right] = \mathcal{O}\left(\left(\Delta t\right)^3\right). \tag{C.15}$$

Similarly, we can observe that for all $l \in \mathbb{N}_{\geq 2}$ [24], it follows that

$$\mathbb{E}\left[\left(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_{.}^{m,x}\right)\right]_{t_m,t_{m+1}}\right)^l\right] = \mathcal{O}\left(\left(\Delta t\right)^{1.5l}\right). \tag{C.16}$$

We summarize our results in the following lemma

Lemma C.1. *Let $X_{m+1}^{m,x}$ denote the value of X_{m+1} given $X_m = x$ and let $\mathcal{A} = \{v, (0), (1), (0,0), (0,1), (1,0), (1,1)\}$. For a sufficiently smooth function $h(t, x)$ it holds for all $l \in \mathbb{N}_{\geq 2}$ that*

$$\mathbb{E}\left[\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_{.}^{m,x}\right)\right]_{t_m,t_{m+1}}\right] = \mathcal{O}\left(\left(\Delta t\right)^3\right), \tag{C.17}$$

$$\mathbb{E}\left[\left(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_{.}^{m,x}\right)\right]_{t_m,t_{m+1}}\right)^l\right] = \mathcal{O}\left(\left(\Delta t\right)^{1.5l}\right). \tag{C.18}$$

For the 2.0-weak-Taylor scheme we approximate $I_{(0,1),t_m,t_{m+1}}$ and $I_{(1,0),t_m,t_{m+1}}$ by $\frac{1}{2}\Delta t\Delta W_{m+1}$. This replacement was chosen such that it has the same moments in first order, i.e.

$$\begin{aligned}
\mathbb{E}\left[I_{(0,1),t_m,t_{m+1}}\right] &= \mathbb{E}\left[\frac{1}{2}\Delta t\Delta W_{m+1}\right], && \text{(C.19)}\\
\mathbb{E}\left[I_{(1,0),t_m,t_{m+1}}\right] &= \mathbb{E}\left[\frac{1}{2}\Delta t\Delta W_{m+1}\right], && \text{(C.20)}\\
\text{Cov}\left(I_{(0,1),t_m,t_{m+1}}, \Delta W_{m+1}\right) &= \text{Cov}\left(\frac{1}{2}\Delta t\Delta W_{m+1}, \Delta W_{m+1}\right), && \text{(C.21)}\\
\text{Cov}\left(I_{(1,0),t_m,t_{m+1}}, \Delta W_{m+1}\right) &= \text{Cov}\left(\frac{1}{2}\Delta t\Delta W_{m+1}, \Delta W_{m+1}\right). && \text{(C.22)}
\end{aligned}$$

Now, we find by using Lemma C.1 and [18, Chapter 5]:

Lemma C.2. *Let $X^{m,x}_{m+1}$ and $X^{\Delta,m,x}_{m+1}$ denote the values of respectively X_{m+1} and X^{Δ}_{m+1} given $X_m = x$, where $X^{\Delta,m,x}_{m+1}$ is derived with the 2.0-weak-Taylor scheme. Also, we define $\mathcal{A} = \{v, (0), (1), (0,0), (0,1), (1,0), (1,1)\}$. For a sufficiently smooth function $h(t,x)$ it holds that for all $l \in \mathbb{N}$*

$$\begin{aligned}
&\mathbb{E}\left[\left(h\left(t_{m+1}, X^{\Delta,m,x}_{t_{m+1}}\right) - h\left(t_{m+1}, X^{m,x}_{t_{m+1}}\right)\right)^l\right]\\
&= \mathbb{E}\left[\left(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(.,X^{m,x}_{.}\right)\right]_{t_m,t_{m+1}} + h_{(0,1)}(t_m,x)\left(\frac{1}{2}\Delta W_{m+1}\Delta t - I_{(0,1),t_m,t_{m+1}}\right)\right.\right.\\
&+ \left.\left. h_{(1,0)}(t_m,x)\left(\frac{1}{2}\Delta W_{m+1}\Delta t - I_{(1,0),t_m,t_{m+1}}\right)\right)^l\right]\\
&= \mathcal{O}\left((\Delta t)^3\right). \qquad \text{(C.23)}
\end{aligned}$$

Because of the replacement $I_{(0,1),t_m,t_{m+1}}$ and $I_{(1,0),t_m,t_{m+1}}$ by $\frac{1}{2}\Delta t\Delta W_{m+1}$, we loss the higher order of accuracy for $l \in \mathbb{N} \geq 3$.

Two-dimensional Itô-Taylor expansion

Assume our system of FSDEs is of the form (4.1) and (4.2). In two dimensions, the set of all multi-indices is denoted by $\mathcal{M}$,

$$\mathcal{M} = \{(j_1, j_2, ..., j_l) : j_i \in \{0,1,2\}, i \in \{1,...,l\}, \text{ for } l = 1,2,3,...\} \cup v, \qquad \text{(C.24)}$$

and

$$\begin{aligned}
L^0 &= \frac{\partial}{\partial t} + \mu_1(x_1,x_2)\frac{\partial}{\partial x_1} + \mu_2(x_1,x_2)\frac{\partial}{\partial x_2}\\
&+ \frac{1}{2}\left[\sigma_1^2(x_1,x_2)\frac{\partial^2}{\partial x_1^2} + 2\rho\sigma_1(x_1,x_2)\sigma_2(x_1,x_2)\frac{\partial^2}{\partial x_1\partial x_2} + \sigma_2^2(x_1,x_2)\frac{\partial^2}{\partial x_2^2}\right], && \text{(C.25)}\\
L^1 &= \sigma_1(x_1,x_2)\frac{\partial}{\partial x_1} + \rho\sigma_2(x_1,x_2)\frac{\partial}{\partial x_2}, && \text{(C.26)}\\
L^2 &= \sqrt{1-\rho^2}\sigma_2(x_1,x_2)\frac{\partial}{\partial x_2}. && \text{(C.27)}
\end{aligned}$$

The multiple Itô integral $I_{\widehat{\alpha}}\left[h(.,x.)\right]_{s,t}$ is defined recursively by

$$I_{\widehat{\alpha}}\left[h(.,x.)\right]_{s,t} = \begin{cases} h(t,x_t), & \text{if } l(\widehat{\alpha}) = 0, \\ \int_s^t I_{\widehat{\alpha}-}\left[h(.,x.)\right]_{s,u} \mathrm{d}u, & \text{if } l(\widehat{\alpha}) \geq 1, j_{l(\widehat{\alpha})} = 0, \\ \int_s^t I_{\widehat{\alpha}-}\left[h(.,x.)\right]_{s,u} \mathrm{d}W_u^1, & \text{if } l(\widehat{\alpha}) \geq 1, j_{l(\widehat{\alpha})} = 1, \\ \int_s^t I_{\widehat{\alpha}-}\left[h(.,x.)\right]_{s,u} \mathrm{d}W_u^2, & \text{if } l(\widehat{\alpha}) \geq 1, j_{l(\widehat{\alpha})} = 2, \end{cases} \tag{C.28}$$

and we abbreviate

$$I_{\widehat{\alpha},s,t} = I_{\widehat{\alpha}}[1]_{s,t}. \tag{C.29}$$

Theorem C.2. *Let* $\mathbf{X}_{m+1}^{m,\mathbf{x}} = \left(X_{m+1}^{1,m,\mathbf{x}}, X_{m+1}^{2,m,\mathbf{x}}\right)$ *denote the value of* $\mathbf{X}_m$ *given* $\mathbf{X}_m = \mathbf{x} = (x_1, x_2)$ *and let* $\mathcal{A} \subseteq \mathcal{M}$ *be a hierarchical set. The Itô-Taylor expansion for a general sufficiently smooth function* $h(t,\mathbf{x})$ *is given by [18]*

$$h\left(t_{m+1}, X_{m+1}^{1,m,\mathbf{x}}, X_{m+1}^{2,m,\mathbf{x}}\right) = \sum_{\widehat{\alpha}\in\mathcal{A}} h_{\widehat{\alpha}}(t_m,\mathbf{x}) I_{\widehat{\alpha},t_m,t_{m+1}} + \sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_.^{1,m,\mathbf{x}}, X_.^{2,m,\mathbf{x}}\right)\right]_{t_m,t_{m+1}}. \tag{C.30}$$

Similar to Appendix C, we find:

Lemma C.3. *Let* $\mathbf{X}_{m+1}^{m,\mathbf{x}}$ *denote the value of* $\mathbf{X}_{m+1}$ *given* $\mathbf{X}_m = \mathbf{x} = (x_1, x_2)$ *and let* $\mathcal{A} = \{\widehat{\alpha} \in \mathcal{M} : l(\widehat{\alpha}) \leq 2\}$*. For sufficiently smooth functions* $h(t,\mathbf{x})$ *and* $\widehat{h}(t,\mathbf{x})$ *it holds that for all* $l_1, l_2 \in \mathbb{Z}_{\geq 0}$ *with* $l_1 + l_2 \geq 2$*,*

$$\mathbb{E}\left[\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_.^{m,\mathbf{x}}\right)\right]_{t_m,t_{m+1}}\right] = \mathcal{O}\left((\Delta t)^3\right), \tag{C.31}$$

$$\mathbb{E}\left[\left(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(., X_.^{m,\mathbf{x}}\right)\right]_{t_m,t_{m+1}}\right)^{l_1}\left(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[\widehat{h}_{\widehat{\alpha}}\left(., X_.^{m,\mathbf{x}}\right)\right]_{t_m,t_{m+1}}\right)^{l_2}\right] = \mathcal{O}\left((\Delta t)^{1.5(l_1+l_2)}\right). \tag{C.32}$$

For the 2.0-weak-Taylor scheme we approximate $I_{(0,j),t_m,t_{m+1}}$ and $I_{(j,0),t_m,t_{m+1}}$ by $\frac{1}{2}\Delta t \Delta W_{m+1}^j$, for $j = 1, 2$. Also, we replace $I_{(1,2),t_m,t_{m+1}}$ and $I_{(2,1),t_m,t_{m+1}}$, respectively, by $\frac{1}{2}\left(\Delta W_{m+1}^1 \Delta W_{m+1}^2 + V_{m+1}^{1,2}\right)$ and $\frac{1}{2}\left(\Delta W_{m+1}^2 \Delta W_{m+1}^1 + V_{m+1}^{2,1}\right)$, where $V_{m+1}^{1,2}$ is an independent random variable with $P\left(V_{m+1}^{1,2} = \pm\Delta t\right) = \frac{1}{2}$ and $V_{m+1}^{2,1} = -V_{m+1}^{1,2}$ [18]. These replacements were chosen so that for all $i, j, k, l = 1, 2$ with $i \neq j$, it follows that[1]

$$\mathbb{E}\left[I_{(0,j),t_m,t_{m+1}}\right] = \mathbb{E}\left[\frac{1}{2}\Delta t \Delta W_{m+1}^j\right], \tag{C.33}$$

$$\mathbb{E}\left[I_{(j,0),t_m,t_{m+1}}\right] = \mathbb{E}\left[\frac{1}{2}\Delta t \Delta W_{m+1}^j\right], \tag{C.34}$$

$$\mathrm{Cov}\left(I_{(0,j),t_m,t_{m+1}}, \Delta W_{m+1}^k\right) = \mathrm{Cov}\left(\frac{1}{2}\Delta t \Delta W_{m+1}^j, \Delta W_{m+1}^k\right), \tag{C.35}$$

$$\mathrm{Cov}\left(I_{(j,0),t_m,t_{m+1}}, \Delta W_{m+1}^k\right) = \mathrm{Cov}\left(\frac{1}{2}\Delta t \Delta W_{m+1}^j, \Delta W_{m+1}^k\right), \tag{C.36}$$

[1] These equalities can easily be verified in a numerical experiment

$$\mathbb{E}\left[I_{(i,j),t_m,t_{m+1}}\right] = \mathbb{E}\left[\frac{1}{2}\left(\Delta W^i_{m+1}\Delta W^j_{m+1} + V^{i,j}_{m+1}\right)\right], \tag{C.37}$$

$$\mathbb{E}\left[\left(I_{(i,j),t_m,t_{m+1}}\right)^2\right] = \mathbb{E}\left[\frac{1}{4}\left(\Delta W^i_{m+1}\Delta W^j_{m+1} + V^{i,j}_{m+1}\right)^2\right], \tag{C.38}$$

$$\mathbb{E}\left[I_{(i,j),t_m,t_{m+1}}\Delta W^k_{m+1}\right] = \mathbb{E}\left[\frac{1}{2}\left(\Delta W^i_{m+1}\Delta W^j_{m+1} + V^{i,j}_{m+1}\right)\Delta W^k_{m+1}\right], \tag{C.39}$$

$$\mathbb{E}\left[I_{(i,j),t_m,t_{m+1}}\Delta W^k_{m+1}\Delta W^l_{m+1}\right] = \mathbb{E}\left[\frac{1}{2}\left(\Delta W^i_{m+1}\Delta W^j_{m+1} + V^{i,j}_{m+1}\right)\Delta W^k_{m+1}\Delta W^l_{m+1}\right], \tag{C.40}$$

$$\mathbb{E}\left[I_{(i,j),t_m,t_{m+1}}I_{(0,k),t_m,t_{m+1}}\right] = \mathbb{E}\left[\frac{1}{2}\left(\Delta W^i_{m+1}\Delta W^j_{m+1} + V^{i,j}_{m+1}\right)I_{(0,k),t_m,t_{m+1}}\right], \tag{C.41}$$

$$\mathbb{E}\left[I_{(i,j),t_m,t_{m+1}}I_{(k,0),t_m,t_{m+1}}\right] = \mathbb{E}\left[\frac{1}{2}\left(\Delta W^i_{m+1}\Delta W^j_{m+1} + V^{i,j}_{m+1}\right)I_{(k,0),t_m,t_{m+1}}\right]. \tag{C.42}$$

Now, we find by using Lemma C.3 and [18, Chapter 5] the following lemma:

Lemma C.4. *Let* $\mathbf{X}^{m,\mathbf{x}}_{m+1}$ *and* $\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}$ *denote the values of respectively* $\mathbf{X}_{m+1}$ *and* $\mathbf{X}^{\Delta}_{m+1}$ *given* $\mathbf{X}_m = \mathbf{x}$*, where* $\mathbf{X}^{\Delta,m,\mathbf{x}}_{m+1}$ *is derived with the 2.0-weak-Taylor scheme. Also, we define* $\mathcal{A} = \{\widehat{\alpha} \in \mathcal{M} : l(\widehat{\alpha}) \leq 2\}$*. For sufficiently smooth functions* $h(t,\mathbf{x})$ *and* $\widehat{h}(t,\mathbf{x})$ *it holds that for all* $l_1, l_2 \in \mathbb{Z}_{\geq 0}$

$$\begin{aligned}
&\mathbb{E}\left[\left(h\left(t_{m+1},\mathbf{X}^{\Delta,m,\mathbf{x}}_{t_{m+1}}\right) - h\left(t_{m+1},\mathbf{X}^{m,\mathbf{x}}_{t_{m+1}}\right)\right)^{l_1}\left(\widehat{h}\left(t_{m+1},\mathbf{X}^{\Delta,m,\mathbf{x}}_{t_{m+1}}\right) - \widehat{h}\left(t_{m+1},\mathbf{X}^{m,\mathbf{x}}_{t_{m+1}}\right)\right)^{l_2}\right] \\
&= \mathbb{E}\Bigg[\Bigg(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[h_{\widehat{\alpha}}\left(.,\mathbf{X}^{m,\mathbf{x}}_{.}\right)\right]_{t_m,t_{m+1}} + h_{(0,1)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^1_{m+1}\Delta t - I_{(0,1),t_m,t_{m+1}}\right) \\
&+ h_{(1,0)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^2_{m+1}\Delta t - I_{(1,0),t_m,t_{m+1}}\right) + h_{(0,2)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^2_{m+1}\Delta t - I_{(0,2),t_m,t_{m+1}}\right) \\
&+ h_{(2,0)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^2_{m+1}\Delta t - I_{(2,0),t_m,t_{m+1}}\right) \\
&+ h_{(1,2)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^1_{m+1}\Delta W^2_{m+1} + V^{1,2}_{m+1} - I_{(1,2),t_m,t_{m+1}}\right) \\
&+ h_{(2,1)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^1_{m+1}\Delta W^2_{m+1} + V^{2,1}_{m+1} - I_{(2,1),t_m,t_{m+1}}\right)\Bigg)^{l_1} \\
&\cdot\Bigg(\sum_{\widehat{\alpha}\in\mathcal{B}(\mathcal{A})} I_{\widehat{\alpha}}\left[\widehat{h}_{\widehat{\alpha}}\left(.,\mathbf{X}^{m,\mathbf{x}}_{.}\right)\right]_{t_m,t_{m+1}} + \widehat{h}_{(0,1)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^1_{m+1}\Delta t - I_{(0,1),t_m,t_{m+1}}\right) \\
&+ \widehat{h}_{(1,0)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^2_{m+1}\Delta t - I_{(1,0),t_m,t_{m+1}}\right) + \widehat{h}_{(0,2)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^2_{m+1}\Delta t - I_{(0,2),t_m,t_{m+1}}\right) \\
&+ \widehat{h}_{(2,0)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^2_{m+1}\Delta t - I_{(2,0),t_m,t_{m+1}}\right) \\
&+ \widehat{h}_{(1,2)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^1_{m+1}\Delta W^2_{m+1} + V^{1,2}_{m+1} - I_{(1,2),t_m,t_{m+1}}\right) \\
&+ \widehat{h}_{(2,1)}(t_m,\mathbf{x})\left(\frac{1}{2}\Delta W^1_{m+1}\Delta W^2_{m+1} + V^{2,1}_{m+1} - I_{(2,1),t_m,t_{m+1}}\right)\Bigg)^{l_2}\Bigg] \\
&= \mathcal{O}\left((\Delta t)^3\right).
\end{aligned} \tag{C.43}$$

Appendix D

Taylor schemes

In this appendix we derive the Euler, Milstein and 2.0-weak-Taylor schemes. We have the following FSDEs

$$
\begin{aligned}
\mathrm{d}X_t^1 &= \mu_1(\mathbf{X}_t)\,\mathrm{d}t + \sigma_1(\mathbf{X}_t)\,\mathrm{d}W_t^1, & X_0^1 &= x_1, \quad &\text{(D.1)}\\
\mathrm{d}X_t^2 &= \mu_2(\mathbf{X}_t)\,\mathrm{d}t + \rho\sigma_2(\mathbf{X}_t)\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\sigma_2(\mathbf{X}_t)\,\mathrm{d}W_t^2, & X_0^2 &= x_2. \quad &\text{(D.2)}
\end{aligned}
$$

Integrating gives

$$
\begin{aligned}
X_{m+1}^1 &= X_m^1 + \int_{t_m}^{t_{m+1}} \mu_1(\mathbf{X}_t)\,\mathrm{d}t + \int_{t_m}^{t_{m+1}} \sigma_1(\mathbf{X}_t)\,\mathrm{d}W_t^1, &\text{(D.3)}\\
X_{m+1}^2 &= X_m^2 + \int_{t_m}^{t_{m+1}} \mu_2(\mathbf{X}_t)\,\mathrm{d}t + \rho\int_{t_m}^{t_{m+1}} \sigma_2(\mathbf{X}_t)\,\mathrm{d}W_t^1\\
&+ \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}} \sigma_2(\mathbf{X}_t)\,\mathrm{d}W_t^2. &\text{(D.4)}
\end{aligned}
$$

Let $h(x, y)$ be a function whose first and second partial derivatives are defined and are continuous. The two-dimensional Itô Lemma [26, Formula (4.6.8)] gives

$$
\begin{aligned}
\mathrm{d}h(\mathbf{X}_t) &= \frac{\partial h(\mathbf{X}_t)}{\partial X_t^1}\mathrm{d}X_t^1 + \frac{\partial h(\mathbf{X}_t)}{\partial X_t^2}\mathrm{d}X_t^2 + \frac{1}{2}\frac{\partial^2 h(\mathbf{X}_t)}{\left(\partial X_t^1\right)^2}\left(\mathrm{d}X_t^1\right)^2 + \frac{\partial^2 h(\mathbf{X}_t)}{\partial X_t^1 \partial X_t^2}\mathrm{d}X_t^1\,\mathrm{d}X_t^2\\
&+ \frac{1}{2}\frac{\partial^2 h(\mathbf{X}_t)}{\left(\partial X_t^2\right)^2}\left(\mathrm{d}X_t^2\right)^2. &\text{(D.5)}
\end{aligned}
$$

Integrating gives [26, Formula (4.6.10)], for all $t \geq t_m$

$$
\begin{aligned}
h(\mathbf{X}_t) &= h(\mathbf{X}_m) + \int_{t_m}^{t}\Bigg(\mu_1(\mathbf{X}_u)\frac{\partial h(\mathbf{X}_u)}{\partial X_u^1} + \mu_2(\mathbf{X}_u)\frac{\partial h(\mathbf{X}_u)}{\partial X_u^2} + \frac{1}{2}\sigma_1(\mathbf{X}_u)^2\frac{\partial^2 h(\mathbf{X}_u)}{\left(\partial X_u^1\right)^2}\\
&+ \rho\sigma_1(\mathbf{X}_u)\sigma_2(\mathbf{X}_u)\frac{\partial^2 h(\mathbf{X}_u)}{\partial X_u^1\partial X_u^2} + \frac{1}{2}\sigma_2^2(\mathbf{X}_u)\frac{\partial^2 h(\mathbf{X}_u)}{\left(\partial X_u^2\right)^2}\Bigg)\mathrm{d}u\\
&+ \int_{t_m}^{t}\sigma_1(\mathbf{X}_u)\frac{\partial h(\mathbf{X}_u)}{\partial X_u^1} + \rho\sigma_2(\mathbf{X}_u)\frac{\partial h(\mathbf{X}_u)}{\partial X_u^2}\mathrm{d}W_u^1\\
&+ \sqrt{1-\rho^2}\int_{t_m}^{t}\sigma_2(\mathbf{X}_u)\frac{\partial h(\mathbf{X}_u)}{\partial X_u^2}\mathrm{d}W_u^2. &\text{(D.6)}
\end{aligned}
$$

We assume that μ_1, μ_2, σ_1 and σ_2 are functions whose first and second partial derivatives are defined and are continuous. Substitution of result (D.6), for $\mu_1(\mathbf{X}_t)$, $\mu_2(\mathbf{X}_t)$, $\sigma_1(\mathbf{X}_t)$, and $\sigma_2(\mathbf{X}_t)$, in formula (D.3) gives

$$
\begin{aligned}
X^1_{m+1} &= X^1_m + \int_{t_m}^{t_{m+1}} \mu_1(\mathbf{X}_m)\,\mathrm{d}t + \int_{t_m}^{t_{m+1}} \sigma_1(\mathbf{X}_m)\,\mathrm{d}W^1_t \\
&+ \int_{t_m}^{t_{m+1}}\int_{t_m}^{t} \Bigg(\mu_1(\mathbf{X}_u)\frac{\partial \mu_1(\mathbf{X}_u)}{\partial X^1_u} + \mu_2(\mathbf{X}_u)\frac{\partial \mu_1(\mathbf{X}_u)}{\partial X^2_u} + \frac{1}{2}\sigma_1^2(\mathbf{X}_u)\frac{\partial^2 \mu_1(\mathbf{X}_u)}{(\partial X^1_u)^2} \\
&+ \rho\sigma_1(\mathbf{X}_u)\sigma_2(\mathbf{X}_u)\frac{\partial^2 \mu_1(\mathbf{X}_u)}{\partial X^1_u \partial X^2_u} + \frac{1}{2}\sigma_2^2(\mathbf{X}_u)\frac{\partial^2 \mu_1(\mathbf{X}_u)}{(\partial X^2_u)^2}\Bigg)\,\mathrm{d}u\,\mathrm{d}t \\
&+ \int_{t_m}^{t_{m+1}}\int_{t_m}^{t} \sigma_1(\mathbf{X}_u)\frac{\partial \mu_1(\mathbf{X}_u)}{\partial X^1_u} + \rho\sigma_2(\mathbf{X}_u)\frac{\partial \mu_1(\mathbf{X}_u)}{\partial X^2_u}\,\mathrm{d}W^1_u\,\mathrm{d}t \\
&+ \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}}\int_{t_m}^{t} \sigma_2(\mathbf{X}_u)\frac{\partial \mu_1(\mathbf{X}_u)}{\partial X^2_u}\,\mathrm{d}W^2_u\,\mathrm{d}t \\
&+ \int_{t_m}^{t_{m+1}}\int_{t_m}^{t} \Bigg(\mu_1(\mathbf{X}_u)\frac{\partial \sigma_1(\mathbf{X}_u)}{\partial X^1_u} + \mu_2(\mathbf{X}_u)\frac{\partial \sigma_1(\mathbf{X}_u)}{\partial X^2_u} + \frac{1}{2}\sigma_1^2(\mathbf{X}_u)\frac{\partial^2 \sigma_1(\mathbf{X}_u)}{(\partial X^1_u)^2} \\
&+ \rho\sigma_1(\mathbf{X}_u)\sigma_2(\mathbf{X}_u)\frac{\partial^2 \sigma_1(\mathbf{X}_u)}{\partial X^1_u \partial X^2_u} + \frac{1}{2}\sigma_2^2(\mathbf{X}_u)\frac{\partial^2 \sigma_1(\mathbf{X}_u)}{(\partial X^2_u)^2}\Bigg)\,\mathrm{d}u\,\mathrm{d}W^1_t \\
&+ \int_{t_m}^{t_{m+1}}\int_{t_m}^{t} \sigma_1(\mathbf{X}_u)\frac{\partial \sigma_1(\mathbf{X}_u)}{\partial X^1_u} + \rho\sigma_2(\mathbf{X}_u)\frac{\partial \sigma_1(\mathbf{X}_u)}{\partial X^2_u}\,\mathrm{d}W^1_u\,\mathrm{d}W^1_t \\
&+ \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}}\int_{t_m}^{t} \sigma_2(\mathbf{X}_u)\frac{\partial \sigma_1(\mathbf{X}_u)}{\partial X^2_u}\,\mathrm{d}W^2_u\,\mathrm{d}W^1_t. \qquad \text{(D.7)}
\end{aligned}
$$

We find with iterative use of formula $(D.6)$ that

$$
\begin{aligned}
X^1_{m+1} &= X^1_m + \mu_1(\mathbf{X}_m)\Delta t + \sigma_1(\mathbf{X}_m)\Delta W^1_{m+1} \\
&+ \Bigg(\mu_1(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X^1_m} + \mu_2(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X^2_m} + \frac{1}{2}\sigma_1^2(\mathbf{X}_m)\frac{\partial^2 \mu_1(\mathbf{X}_m)}{(\partial X^1_m)^2} \\
&+ \rho\sigma_1(\mathbf{X}_m)\sigma_2(\mathbf{X}_m)\frac{\partial^2 \mu_1(\mathbf{X}_m)}{\partial X^1_m \partial X^2_m} + \frac{1}{2}\sigma_2^2(\mathbf{X}_m)\frac{\partial^2 \mu_1(\mathbf{X}_m)}{(\partial X^2_m)^2}\Bigg)\int_{t_m}^{t_{m+1}}\int_{t_m}^{t}\mathrm{d}u\,\mathrm{d}t \\
&+ \left(\sigma_1(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X^1_m} + \rho\sigma_2(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X^2_m}\right)\int_{t_m}^{t_{m+1}}\int_{t_m}^{t}\mathrm{d}W^1_u\,\mathrm{d}t \\
&+ \sqrt{1-\rho^2}\sigma_2(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X^2_m}\int_{t_m}^{t_{m+1}}\int_{t_m}^{t}\mathrm{d}W^2_u\,\mathrm{d}t \\
&+ \Bigg(\mu_1(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X^1_m} + \mu_2(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X^2_m} + \frac{1}{2}\sigma_1^2(\mathbf{X}_m)\frac{\partial^2 \sigma_1(\mathbf{X}_m)}{(\partial X^1_m)^2} \\
&+ \rho\sigma_1(\mathbf{X}_m)\sigma_2(\mathbf{X}_m)\frac{\partial^2 \sigma_1(\mathbf{X}_m)}{\partial X^1_m \partial X^2_m} + \frac{1}{2}\sigma_2^2(\mathbf{X}_m)\frac{\partial^2 \sigma_1(\mathbf{X}_m)}{(\partial X^2_m)^2}\Bigg)\int_{t_m}^{t_{m+1}}\int_{t_m}^{t}\mathrm{d}u\,\mathrm{d}W^1_t \\
&+ \left(\sigma_1(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X^1_m} + \rho\sigma_2(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X^2_m}\right)\int_{t_m}^{t_{m+1}}\int_{t_m}^{t}\mathrm{d}W^1_u\,\mathrm{d}W^1_t \qquad \text{(D.8)} \\
&+ \sqrt{1-\rho^2}\sigma_2(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X^2_m}\int_{t_m}^{t_{m+1}}\int_{t_m}^{t}\mathrm{d}W^2_u\,\mathrm{d}W^1_t + \mathcal{O}\left(\sum_{j_1,j_2,j_3=0}^{2} I_{(j_1,j_2,j_3),t_m,t_{m+1}}\right),
\end{aligned}
$$

where $I_{(j_1,j_2,j_3),t_m,t_{m+1}}$ is defined as in (C.29), for example

$$I_{(0,2,1),t_m,t_{m+1}} = \int_{t_m}^{t_{m+1}} \int_{t_m}^{t} \int_{t_m}^{u} \mathrm{d}s \, \mathrm{d}W_u^2 \, \mathrm{d}W_t^1. \tag{D.9}$$

We can also obtain this result by using Theorem C.2 and $\mathcal{A} = \{\widehat{\alpha} \in \mathcal{M} : l(\widehat{\alpha}) \leq 2\}$.

Like Ruijter and Oosterlee [24], we replace $\int_{t_m}^{t_{m+1}} \int_{t_m}^{t} \mathrm{d}W_u^j \, \mathrm{d}t$ and $\int_{t_m}^{t_{m+1}} \int_{t_m}^{t} \mathrm{d}u \, \mathrm{d}W_t^j$ by $\frac{1}{2}\Delta W_{m+1}^j \Delta t$, and
$\int_{t_m}^{t_{m+1}} \int_{t_m}^{t} \mathrm{d}W_u^j \, \mathrm{d}W_t^j$ by $\frac{1}{2}\left(\left(\Delta W_{m+1}^j\right)^2 - \Delta t\right)$.

Let $V_{m+1}^{i,j}$ independent random variables, with $P\left(V_{m+1}^{i,j} = \pm\Delta t\right) = \frac{1}{2}$ and $V_{m+1}^{i,j} = -V_{m+1}^{j,i}$, for $i, j \in \{1, 2\}$ and $i \neq j$. According to Kloeden and Platen [18], the replacement $\int_{t_m}^{t_{m+1}} \int_{t_m}^{t} \mathrm{d}W_u^i \, \mathrm{d}W_t^j$ by $\frac{1}{2}\left(\Delta W_{m+1}^i \Delta W_{m+1}^j + V_{m+1}^{i,j}\right)$, is correct when the diffusion matrix satisfies the commutativity condition

$$\sqrt{1-\rho^2}\sigma_2(x_1, x_2) \frac{\partial \sigma_1(x_1, x_2)}{\partial x_2} = 0 \quad \text{and} \quad \sqrt{1-\rho^2}\sigma_1(x_1, x_2) \frac{\partial \sigma_2(x_1, x_2)}{\partial x_1} = 0 \quad \forall x_1, x_2 \in \mathbb{R}^2. \tag{D.10}$$

Now, we have

$$\begin{aligned}
X_{m+1}^1 \approx\; & X_m^1 + \mu_1(\mathbf{X}_m)\Delta t + \sigma_1(\mathbf{X}_m)\Delta W_{m+1}^1 \\
+\; & \frac{1}{2}\left(\mu_1(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X_m^1} + \mu_2(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X_m^2} + \frac{1}{2}\sigma_1^2(\mathbf{X}_m)\frac{\partial^2 \mu_1(\mathbf{X}_m)}{(\partial X_m^1)^2}\right. \\
+\; & \left.\rho\sigma_1(\mathbf{X}_m)\sigma_2(\mathbf{X}_m)\frac{\partial^2 \mu_1(\mathbf{X}_m)}{\partial X_m^1 \partial X_m^2} + \frac{1}{2}\sigma_2^2(\mathbf{X}_m)\frac{\partial^2 \mu_1(\mathbf{X}_m)}{(\partial X_m^2)^2}\right)(\Delta t)^2 \\
+\; & \frac{1}{2}\left(\sigma_1(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X_m^1} + \rho\sigma_2(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X_m^2}\right)\Delta W_{m+1}^1 \Delta t \\
+\; & \frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{X}_m)\frac{\partial \mu_1(\mathbf{X}_m)}{\partial X_m^2}\Delta W_{m+1}^2 \Delta t \\
+\; & \frac{1}{2}\left(\mu_1(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X_m^1} + \mu_2(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X_m^2} + \frac{1}{2}\sigma_1^2(\mathbf{X}_m)\frac{\partial^2 \sigma_1(\mathbf{X}_m)}{(\partial X_m^1)^2}\right. \\
+\; & \left.\rho\sigma_1(\mathbf{X}_m)\sigma_2(\mathbf{X}_m)\frac{\partial^2 \sigma_1(\mathbf{X}_m)}{\partial X_m^1 \partial X_m^2} + \frac{1}{2}\sigma_2^2(\mathbf{X}_m)\frac{\partial^2 \sigma_1(\mathbf{X}_m)}{(\partial X_m^2)^2}\right)\Delta W_{m+1}^1 \Delta t \\
+\; & \frac{1}{2}\left(\sigma_1(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X_m^1} + \rho\sigma_2(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X_m^2}\right)\left(\left(\Delta W_{m+1}^1\right)^2 - \Delta t\right) \\
+\; & \frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{X}_m)\frac{\partial \sigma_1(\mathbf{X}_m)}{\partial X_m^2}\left(\Delta W_{m+1}^1 \Delta W_{m+1}^2 - V_{m+1}^{1,2}\right).
\end{aligned} \tag{D.11}$$

For the Euler scheme we only use the the first line of (D.11), this results in

$$X_{m+1}^{1,\Delta} = X_m^{1,\Delta} + \mu_1\left(\mathbf{X}_m^\Delta\right)\Delta t + \sigma_1\left(\mathbf{X}_m^\Delta\right)\Delta W_{m+1}^1. \tag{D.12}$$

For the Milstein scheme we remove the terms $(\Delta t)^2$, $\Delta W_{m+1}^1 \Delta t$, $\Delta W_{m+1}^2 \Delta t$ of (D.11), we find

$$\begin{aligned}
X_{m+1}^{1,\Delta} =\; & X_m^{1,\Delta} + \mu_1\left(\mathbf{X}_m^\Delta\right)\Delta t + \sigma_1\left(\mathbf{X}_m^\Delta\right)\Delta W_{m+1}^1 \\
+\; & \frac{1}{2}\left(\sigma_1\left(\mathbf{X}_m^\Delta\right)\frac{\partial \sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}} + \rho\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial \sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}}\right)\left(\left(\Delta W_{m+1}^1\right)^2 - \Delta t\right)
\end{aligned}$$

$$+ \quad \frac{\sqrt{1-\rho^2}}{2}\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}}\left(\Delta W_{m+1}^1\Delta W_{m+1}^2 - V_{m+1}^{1,2}\right). \tag{D.13}$$

For the 2.0-weak-Taylor scheme, we obtain

$$\begin{aligned}
X_{m+1}^{1,\Delta} &= X_m^{1,\Delta} + \mu_1\left(\mathbf{X}_m^\Delta\right)\Delta t + \sigma_1\left(\mathbf{X}_m^\Delta\right)\Delta W_{m+1}^1 \\
&+ \frac{1}{2}\left(\mu_1\left(\mathbf{X}_m^\Delta\right)\frac{\partial\mu_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}} + \mu_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\mu_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}} + \frac{1}{2}\sigma_1^2\left(\mathbf{X}_m^\Delta\right)\frac{\partial^2\mu_1\left(\mathbf{X}_m^\Delta\right)}{\left(\partial X_m^{1,\Delta}\right)^2}\right. \\
&+ \left.\rho\sigma_1\left(\mathbf{X}_m^\Delta\right)\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial^2\mu_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}\partial X_m^{2,\Delta}} + \frac{1}{2}\sigma_2^2\left(\mathbf{X}_m^\Delta\right)\frac{\partial^2\mu_1\left(\mathbf{X}_m^\Delta\right)}{\left(\partial X_m^{2,\Delta}\right)^2}\right)(\Delta t)^2 \\
&+ \frac{1}{2}\left(\sigma_1\left(\mathbf{X}_m^\Delta\right)\frac{\partial\mu_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}} + \rho\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\mu_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}}\right)\Delta W_{m+1}^1\Delta t \\
&+ \frac{\sqrt{1-\rho^2}}{2}\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\mu_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}}\Delta W_{m+1}^2\Delta t \\
&+ \frac{1}{2}\left(\mu_1\left(\mathbf{X}_m^\Delta\right)\frac{\partial\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}} + \mu_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}} + \frac{1}{2}\sigma_1^2\left(\mathbf{X}_m^\Delta\right)\frac{\partial^2\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\left(\partial X_m^{1,\Delta}\right)^2}\right. \\
&+ \left.\rho\sigma_1\left(\mathbf{X}_m^\Delta\right)\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial^2\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}\partial X_m^{2,\Delta}} + \frac{1}{2}\sigma_2^2\left(\mathbf{X}_m^\Delta\right)\frac{\partial^2\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\left(\partial X_m^{2,\Delta}\right)^2}\right)\Delta W_{m+1}^1\Delta t \\
&+ \frac{1}{2}\left(\sigma_1\left(\mathbf{X}_m^\Delta\right)\frac{\partial\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{1,\Delta}} + \rho\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}}\right)\left(\left(\Delta W_{m+1}^{1,\Delta}\right)^2 - \Delta t\right) \\
&+ \frac{\sqrt{1-\rho^2}}{2}\sigma_2\left(\mathbf{X}_m^\Delta\right)\frac{\partial\sigma_1\left(\mathbf{X}_m^\Delta\right)}{\partial X_m^{2,\Delta}}\left(\Delta W_{m+1}^1\Delta W_{m+1}^2 - V_{m+1}^{1,2}\right).
\end{aligned} \tag{D.14}$$

Analogously, we can find the schemes for $X_{m+1}^{2,\Delta}$ and we can write the discretization schemes in the general form

$$\begin{aligned}
X_{m+1}^{1,\Delta} &= x_1 + m_1(\mathbf{x})\Delta t + s_1^{W^1}(\mathbf{x})\Delta W_{m+1}^1 + s_1^{W^2}(\mathbf{x})\Delta W_{m+1}^2 + \kappa_1^{W^1,W^2}(\mathbf{x})\Delta W_{m+1}^1\Delta W_{m+1}^2 \\
&+ \kappa_1^{W^1}(\mathbf{x})\left(\Delta W_{m+1}^1\right)^2 + \kappa_1^{W^2}(\mathbf{x})\left(\Delta W_{m+1}^2\right)^2 + v_1(\mathbf{x})V_{m+1}^{1,2}, & \text{(D.15)} \\
X_{m+1}^{2,\Delta} &= x_2 + m_2(\mathbf{x})\Delta t + s_2^{W^1}(\mathbf{x})\Delta W_{m+1}^1 + s_2^{W^2}(\mathbf{x})\Delta W_{m+1}^2 + \kappa_2^{W^1,W^2}(\mathbf{x})\Delta W_{m+1}^1\Delta W_{m+1}^2 \\
&+ \kappa_2^{W^1}(\mathbf{x})\left(\Delta W_{m+1}^1\right)^2 + \kappa_2^{W^2}(\mathbf{x})\left(\Delta W_{m+1}^2\right)^2 + v_2(\mathbf{x})V_{m+1}^{1,2}, & \text{(D.16)}
\end{aligned}$$

where $\mathbf{X}_m^\Delta = \mathbf{x} = (x_1, x_2)$.

For the Euler scheme, it follows that

$$\begin{aligned}
&m_1(\mathbf{x}) = \mu_1(\mathbf{x}), \quad s_1^{W^1}(\mathbf{x}) = \sigma_1(\mathbf{x}), \quad s_1^{W^2}(\mathbf{x}) = 0, \quad \kappa_1^{W^1,W^2}(\mathbf{x}) = 0, \\
&\kappa_1^{W^1}(\mathbf{x}) = 0, \quad \kappa_1^{W^2}(\mathbf{x}) = 0, \quad v_1(\mathbf{x}) = 0, \\
&m_2(\mathbf{x}) = \mu_2(\mathbf{x}), \quad s_2^{W^1}(\mathbf{x}) = \rho\sigma_2(\mathbf{x}), \quad s_2^{W^2}(\mathbf{x}) = \sqrt{1-\rho^2}\sigma_2(\mathbf{x}), \quad \kappa_2^{W^1,W^2}(\mathbf{x}) = 0, \\
&\kappa_2^{W^1}(\mathbf{x}) = 0, \quad \kappa_2^{W^2}(\mathbf{x}) = 0, \quad v_2(\mathbf{x}) = 0.
\end{aligned} \tag{D.17}$$

For the Milstein scheme, we find

$$
\begin{aligned}
&m_1(\mathbf{x}) = \mu_1(\mathbf{x}) - \tfrac{1}{2}\left(\sigma_1(\mathbf{x})\tfrac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\tfrac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right), && s_1^{W^1}(\mathbf{x}) = \sigma_1(\mathbf{x}),\\
&\kappa_1^{W^1,W^2}(\mathbf{x}) = \tfrac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\tfrac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, && s_1^{W^2}(\mathbf{x}) = 0,\\
&\kappa_1^{W^1}(\mathbf{x}) = \tfrac{1}{2}\left(\sigma_1(\mathbf{x})\tfrac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\tfrac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right), && \kappa_1^{W^2}(\mathbf{x}) = 0,\\
&v_1(\mathbf{x}) = -\tfrac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\tfrac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, &&\\
&m_2(\mathbf{x}) = \mu_2(\mathbf{x}) - \tfrac{1}{2}\left(\rho\sigma_1(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \sigma_2(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), && s_2^{W^1}(\mathbf{x}) = \rho\sigma_2(\mathbf{x}),\\
&\kappa_2^{W^1,W^2}(\mathbf{x}) = \tfrac{\sqrt{1-\rho^2}}{2}\left(\sigma_1(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + 2\rho\sigma_2(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), && s_2^{W^2}(\mathbf{x}) = \sqrt{1-\rho^2}\sigma_2(\mathbf{x}),\\
&\kappa_2^{W^1}(\mathbf{x}) = \tfrac{\rho}{2}\left(\sigma_1(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), && \kappa_2^{W^2}(\mathbf{x}) = \tfrac{1-\rho^2}{2}\sigma_2(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_2},\\
&v_2(\mathbf{x}) = \tfrac{\sqrt{1-\rho^2}}{2}\sigma_1(\mathbf{x})\tfrac{\partial\sigma_2(\mathbf{x})}{\partial x_1}. &&
\end{aligned} \tag{D.18}
$$

For the 2.0-weak-Taylor scheme, we have

$$
\begin{aligned}
m_1(\mathbf{x}) &= \mu_1(\mathbf{x}) - \frac{1}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right)\\
&+ \frac{1}{2}\left(\mu_1(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_1} + \mu_2(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_2} + \frac{1}{2}\sigma_1^2(\mathbf{x})\frac{\partial^2\mu_1(\mathbf{x})}{(\partial x_1)^2}\right.\\
&+ \left.\rho\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\frac{\partial^2\mu_1(\mathbf{x})}{\partial x_1\partial x_2} + \frac{1}{2}\sigma_2^2(\mathbf{x})\frac{\partial^2\mu_1(\mathbf{x})}{(\partial x_2)^2}\right)\Delta t,
\end{aligned} \tag{D.19a}
$$

$$
\begin{aligned}
s_1^{W^1}(\mathbf{x}) &= \sigma_1(\mathbf{x}) + \frac{1}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_2} + \mu_1(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \mu_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right.\\
&+ \left.\frac{1}{2}\sigma_1^2(\mathbf{x})\frac{\partial^2\sigma_1(\mathbf{x})}{(\partial x_1)^2} + \rho\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\frac{\partial^2\sigma_1(\mathbf{x})}{\partial x_1\partial x_2} + \frac{1}{2}\sigma_2^2(\mathbf{x})\frac{\partial^2\sigma_1(\mathbf{x})}{(\partial x_2)^2}\right)\Delta t,
\end{aligned} \tag{D.19b}
$$

$$
s_1^{W^2}(\mathbf{x}) = \frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\frac{\partial\mu_1(\mathbf{x})}{\partial x_2}\Delta t, \qquad \kappa_1^{W^1,W^2}(\mathbf{x}) = \frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}, \tag{D.19c}
$$

$$
\kappa_1^{W^1}(\mathbf{x}) = \frac{1}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2}\right), \qquad \kappa_1^{W^2}(\mathbf{x}) = 0, \tag{D.19d}
$$

$$
v_1(\mathbf{x}) = -\frac{\sqrt{1-\rho^2}}{2}\sigma_2(\mathbf{x})\frac{\partial\sigma_1(\mathbf{x})}{\partial x_2} \tag{D.19e}
$$

$$
\begin{aligned}
m_2(\mathbf{x}) &= \mu_2(\mathbf{x}) - \frac{1}{2}\left(\rho\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right)\\
&+ \frac{1}{2}\left(\mu_1(\mathbf{x})\frac{\partial\mu_2(\mathbf{x})}{\partial x_1} + \mu_2(\mathbf{x})\frac{\partial\mu_2\left(\mathbf{X}_m^{\Delta}\right)}{\partial x_2} + \frac{1}{2}\sigma_1^2(\mathbf{x})\frac{\partial^2\mu_2(\mathbf{x})}{(\partial x_1)^2}\right.\\
&+ \left.\rho\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\frac{\partial^2\mu_2(\mathbf{x})}{\partial x_1\partial x_2} + \frac{1}{2}\sigma_2^2(\mathbf{x})\frac{\partial^2\mu_2(\mathbf{x})}{(\partial x_2)^2}\right)\Delta t,
\end{aligned} \tag{D.19f}
$$

$$
\begin{aligned}
s_2^{W^1}(\mathbf{x}) &= \rho\sigma_2(\mathbf{x}) + \frac{1}{2}\sigma_1(\mathbf{x})\frac{\partial\mu_2(\mathbf{x})}{\partial x_1}\Delta t\\
&+ \frac{\rho}{2}\left(\sigma_2(\mathbf{x})\frac{\partial\mu_2(\mathbf{x})}{\partial x_2} + \mu_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \mu_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2} + \frac{1}{2}\sigma_1^2(\mathbf{x})\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_1)^2}\right.\\
&+ \left.\rho\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\frac{\partial^2\sigma_2(\mathbf{x})}{\partial x_1\partial x_2} + \frac{1}{2}\sigma_2^2(\mathbf{x})\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_2)^2}\right)\Delta t,
\end{aligned} \tag{D.19g}
$$

$$
\begin{aligned}
s_2^{W^2}(\mathbf{x}) &= \sqrt{1-\rho^2}\sigma_2(\mathbf{x}) + \frac{\sqrt{1-\rho^2}}{2}\left(\sigma_2(\mathbf{x})\frac{\partial\mu_2(\mathbf{x})}{\partial x_2} + \mu_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \mu_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right. \\
&+ \left.\frac{1}{2}\sigma_1^2(\mathbf{x})\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_1)^2} + \rho\sigma_1(\mathbf{x})\sigma_2(\mathbf{x})\frac{\partial^2\sigma_2(\mathbf{x})}{\partial x_1\partial x_2} + \frac{1}{2}\sigma_2^2(\mathbf{x})\frac{\partial^2\sigma_2(\mathbf{x})}{(\partial x_2)^2}\right)\Delta t, \qquad \text{(D.19h)} \\
\kappa_2^{W^1,W^2}(\mathbf{x}) &= \frac{\sqrt{1-\rho^2}}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + 2\rho\sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), \qquad \text{(D.19i)} \\
\kappa_2^{W^1}(\mathbf{x}) &= \frac{\rho}{2}\left(\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1} + \rho\sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}\right), \qquad \text{(D.19j)} \\
\kappa_2^{W^2}(\mathbf{x}) &= \frac{1-\rho^2}{2}\sigma_2(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_2}, \quad v_2(\mathbf{x}) = \frac{\sqrt{1-\rho^2}}{2}\sigma_1(\mathbf{x})\frac{\partial\sigma_2(\mathbf{x})}{\partial x_1}. \qquad \text{(D.19k)}
\end{aligned}
$$

Appendix E

Characteristic function

apc In this appendix, we derive the characteristic function of $\mathbf{X}^{\Delta}_{m+1}$, given $\mathbf{X}^{\Delta}_{m} = \mathbf{x}$, where

$$\begin{aligned} X^{1,\Delta}_{m+1} &= x_1 + m_1(\mathbf{x})\Delta t + s_1^{W^1}(\mathbf{x})\Delta W^1_{m+1} + s_1^{W^2}(\mathbf{x})\Delta W^2_{m+1} + \kappa_1^{W^1,W^2}(\mathbf{x})\Delta W^1_{m+1}\Delta W^2_{m+1} \\ &+ \kappa_1^{W^1}(\mathbf{x})\left(\Delta W^1_{m+1}\right)^2 + \kappa_1^{W^2}(\mathbf{x})\left(\Delta W^2_{m+1}\right)^2 + v_1(\mathbf{x})V^{1,2}_{m+1}, \quad (E.1) \\ X^{2,\Delta}_{m+1} &= x_2 + m_2(\mathbf{x})\Delta t + s_2^{W^1}(\mathbf{x})\Delta W^1_{m+1} + s_2^{W^2}(\mathbf{x})\Delta W^2_{m+1} + \kappa_2^{W^1,W^2}(\mathbf{x})\Delta W^1_{m+1}\Delta W^2_{m+1} \\ &+ \kappa_2^{W^1}(\mathbf{x})\left(\Delta W^1_{m+1}\right)^2 + \kappa_2^{W^2}(\mathbf{x})\left(\Delta W^2_{m+1}\right)^2 + v_2(\mathbf{x})V^{1,2}_{m+1}. \quad (E.2) \end{aligned}$$

Also, ΔW^1_{m+1} and ΔW^2_{m+1} are uncorrelated and both are normally distributed with mean zero and variance Δt and $V^{1,2}_{m+1}$ is an independent random variable with $P\left(V^{1,2}_{m+1} = \pm\Delta t\right) = \frac{1}{2}$.

The characteristic function of $\mathbf{X}^{\Delta}_{m+1}$, given $\mathbf{X}^{\Delta}_{m} = \mathbf{x}$, is given by

$$\begin{aligned} \phi_{\mathbf{X}^{\Delta}_{m+1}}(u_1,u_2|\mathbf{x}) &= \mathbb{E}\left[\exp\left(iu_1 X^{1,\Delta}_{m+1} + iu_2 X^{2,\Delta}_{m+1}\right)\Big| \mathbf{X}^{\Delta}_{m} = \mathbf{x}\right] \\ &= \mathbb{E}\Big[\exp\Big(iu_1\Big[x_1 + m_1(\mathbf{x})\Delta t + s_1^{W^1}(\mathbf{x})\Delta W^1_{m+1} + s_1^{W^2}(\mathbf{x})\Delta W^2_{m+1} \\ &+ \kappa_1^{W^1,W^2}(\mathbf{x})\Delta W^1_{m+1}\Delta W^2_{m+1} + \kappa_1^{W^1}(\mathbf{x})\left(\Delta W^1_{m+1}\right)^2 + \kappa_1^{W^2}(\mathbf{x})\left(\Delta W^2_{m+1}\right)^2 \\ &+ v_1(\mathbf{x})V^{1,2}_{m+1}\Big] + iu_2\Big[x_2 + m_2(\mathbf{x})\Delta t + s_2^{W^1}(\mathbf{x})\Delta W^1_{m+1} \\ &+ s_2^{W^2}(\mathbf{x})\Delta W^2_{m+1} + \kappa_2^{W^1,W^2}(\mathbf{x})\Delta W^1_{m+1}\Delta W^2_{m+1} + \kappa_2^{W^1}(\mathbf{x})\left(\Delta W^1_{m+1}\right)^2 \\ &+ \kappa_2^{W^2}(\mathbf{x})\left(\Delta W^2_{m+1}\right)^2 + v_2(\mathbf{x})V^{1,2}_{m+1}\Big]\Big)\Big]. \quad (E.3) \end{aligned}$$

We further abbreviate

$$\begin{array}{ll} c_1 = u_1 s_1^{W^1}(\mathbf{x}) + u_2 s_2^{W^1}(\mathbf{x}), & c_4 = u_1\kappa_1^{W^1}(\mathbf{x}) + u_2\kappa_2^{W^1}(\mathbf{x}), \\ c_2 = u_1 s_1^{W^2}(\mathbf{x}) + u_2 s_2^{W^2}(\mathbf{x}), & c_5 = u_1\kappa_1^{W^2}(\mathbf{x}) + u_2\kappa_2^{W^2}(\mathbf{x}), \\ c_3 = u_1\kappa_1^{W^1,W^2}(\mathbf{x}) + u_2\kappa_2^{W^1,W^2}(\mathbf{x}), & c_6 = u_1 v_1(\mathbf{x}) + u_2 v_2(\mathbf{x}), \\ v = \Delta W^1_{m+1}, & w = \Delta W^2_{m+1}, \\ V = V^{1,2}_{m+1}. & \end{array} \quad (E.4)$$

Recall from Section 3.1 that the characteristic function of the non-central chi-squared distribution with one degree of freedom and non-centrality parameter λ reads

$$\phi_{\chi_1'^2(\lambda)}(u) = \exp\left(\frac{i\lambda u}{1-2iu}\right)\frac{1}{\sqrt{1-2iu}}. \quad (E.5)$$

We can rewrite the characteristic function to

$$
\begin{aligned}
\phi_{\mathbf{X}^{\Delta}_{m+1}}\left(u_1,u_2|\mathbf{x}\right) &= \exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \mathbb{E}\left[\exp\left(i\left(c_1v+c_2w+c_3vw+c_4v^2+c_5w^2+c_6V\right)\right)\right]\\
&= \exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \mathbb{E}\left[\exp\left(i\left(c_1v+c_2w+c_3vw+c_4v^2+c_5w^2\right)\right)\right]\mathbb{E}\left[\exp\left(ic_6V\right)\right]\\
&= \cosh\left(ic_6\Delta t\right)\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \mathbb{E}\left[\exp\left(i\left(c_1v+c_2w+c_3vw+c_4v^2+c_5w^2\right)\right)\right]. \qquad \text{(E.6)}
\end{aligned}
$$

We assume $c_4 \neq 0$ and $c_5 \neq 0$, which gives

$$
\begin{aligned}
&\phi_{\mathbf{X}^{\Delta}_{m+1}}\left(u_1,u_2|\mathbf{x}\right) = \cosh\left(ic_6\Delta t\right)\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \mathbb{E}\left[\exp\left(i\left(c_2w+c_5w^2+c_4\left(v+\frac{c_1+c_3w}{2c_4}\right)^2-\frac{\left(c_1+c_3w\right)^2}{4c_4}\right)\right)\right]\\
&= \frac{\cosh\left(ic_6\Delta t\right)}{2\pi\Delta t}\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(i\left(c_2w+c_5w^2+c_4\left(v+\frac{c_1+c_3w}{2c_4}\right)^2-\frac{\left(c_1+c_3w\right)^2}{4c_4}\right)\right)\\
&\cdot\ \exp\left(-\frac{v^2+w^2}{2\Delta t}\right)\mathrm{d}v\,\mathrm{d}w\\
&= \frac{\cosh\left(ic_6\Delta t\right)}{2\pi\Delta t}\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \int_{-\infty}^{\infty}\exp\left(i\left(c_2w+c_5w^2-\frac{\left(c_1+c_3w\right)^2}{4c_4}\right)\right)\exp\left(-\frac{w^2}{2\Delta t}\right)\int_{-\infty}^{\infty}\exp\left(ic_4\left(v+\frac{c_1+c_3w}{2c_4}\right)^2\right)\\
&\cdot\ \exp\left(-\frac{v^2}{2\Delta t}\right)\mathrm{d}v\,\mathrm{d}w\\
&= \frac{\cosh\left(ic_6\Delta t\right)}{\sqrt{2\pi\Delta t}}\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \int_{-\infty}^{\infty}\exp\left(i\left(c_2w+c_5w^2-\frac{\left(c_1+c_3w\right)^2}{4c_4}\right)\right)\exp\left(-\frac{w^2}{2\Delta t}\right)\phi_{\chi_1'^2\left(\frac{(c_1+c_3w)^2}{4c_4^2\Delta t}\right)}\left(c_4\Delta t\right)\mathrm{d}w\\
&= \frac{\cosh\left(ic_6\Delta t\right)}{\sqrt{2\pi\Delta t(1-2ic_4\Delta t)}}\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)\\
&\cdot\ \int_{-\infty}^{\infty}\exp\left(i\left(c_2w+c_5w^2-\frac{\left(c_1+c_3w\right)^2}{4c_4}\right)\right)\exp\left(-\frac{w^2}{2\Delta t}\right)\exp\left(\frac{i(c_1+c_3w)^2}{4c_4(1-2ic_4\Delta t)}\right)\mathrm{d}w\\
&= \frac{\cosh\left(ic_6\Delta t\right)\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right]+iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)}{\sqrt{2\pi\Delta t(1-2ic_4\Delta t)}}\\
&\cdot\ \exp\left(-\frac{(\frac{1}{2}+ic_4)c_1^2\left(\Delta t\right)^2}{1+4c_4^2\left(\Delta t\right)^2}+\frac{\mu^2}{2\sigma^2}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right)\\
&\cdot\ \exp\left(i\left(\left(c_2-\frac{2c_1c_3c_4\left(\Delta t\right)^2}{1+4c_4^2\left(\Delta t\right)^2}\right)w+\left(c_5-\frac{c_3^2c_4\left(\Delta t\right)^2}{1+4c_4^2\left(\Delta t\right)^2}\right)w^2\right)\right)\mathrm{d}w, \qquad \text{(E.7)}
\end{aligned}
$$

where

$$
\sigma^2 = \frac{\Delta t(1+4c_4^2\left(\Delta t\right)^2)}{1+4c_4^2\left(\Delta t\right)^2+\left(\Delta t\right)^2c_3^2} \qquad \text{and} \qquad \mu = -\frac{\sigma^2c_1c_3\Delta t}{1+4c_4^2\left(\Delta t\right)^2}. \qquad \text{(E.8)}
$$

We abbreviate

$$c_7 = c_2 - \frac{2c_1c_3c_4\,(\Delta t)^2}{1+4c_4^2\,(\Delta t)^2} \qquad \text{and} \qquad c_8 = c_5 - \frac{c_3^2c_4\,(\Delta t)^2}{1+4c_4^2\,(\Delta t)^2}. \tag{E.9}$$

So, the integral is given by

$$\int_{-\infty}^{\infty} \exp\left(ic_7w + ic_8w^2\right)\exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right)\mathrm{d}w = \sigma\sqrt{2\pi}\,\mathbb{E}\left[\exp\left(ic_7\left[\sigma W+\mu\right] + ic_8\left[\sigma W+\mu\right]^2\right)\right], \tag{E.10}$$

where $W \sim \mathcal{N}(0,1)$. Rewriting gives that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(i(c_7\left[\sigma W+\mu\right] + c_8\left[\sigma W+\mu\right]^2)\right)\right]\\
&= \exp\left(i\left(c_7\mu + c_8\mu^2 - \frac{(c_7+2c_8\mu)^2}{4c_8}\right)\right)\mathbb{E}\left[\exp\left(ic_8\sigma^2\left(W + \frac{c_7+2c_8\mu}{2c_8\sigma}\right)^2\right)\right]\\
&= \exp\left(i\left(c_7\mu + c_8\mu^2 - \frac{(c_7+2c_8\mu)^2}{4c_8}\right)\right)\phi_{\chi_1'^2\left(\left(\frac{c_7+2c_8\mu}{2c_8\sigma}\right)^2\right)}\left(c_8\sigma^2\right)\\
&= \exp\left(i\left(c_7\mu + c_8\mu^2 - \frac{(c_7+2c_8\mu)^2}{4c_8}\right)\right)\exp\left(\frac{i\left(c_7+2c_8\mu\right)^2}{4c_8(1-2ic_8\sigma^2)}\right)\frac{1}{\sqrt{1-2i\sigma^2c_8}}.
\end{aligned} \tag{E.11}$$

So, the characteristic function of $\mathbf{X}_{m+1}^{\Delta}$, given $\mathbf{X}_m^{\Delta} = \mathbf{x}$, is given by

$$\begin{aligned}
\phi_{\mathbf{X}_{m+1}^{\Delta}}(u_1,u_2|\mathbf{x}) \;=\;& \frac{\sigma\cosh\left(ic_6\Delta t\right)\exp\left(i\left(u_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right] + u_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right] + c_7\mu + c_8\mu^2\right)\right)}{\sqrt{\Delta t(1-2ic_4\Delta t)(1-2i\sigma^2c_8)}}\\
&\cdot\;\exp\left(-\frac{c_1^2\Delta t}{2(1+4c_4^2\left(\Delta t\right)^2)} + \frac{\mu^2}{2\sigma^2} - \frac{\left(c_7+2c_8\mu\right)^2\sigma^2}{2(1+4c_8^2\sigma^4)}\right)\\
&\cdot\;\exp\left(i\left(-\frac{c_4c_1^2\left(\Delta t\right)^2}{1+4c_4^2\left(\Delta t\right)^2} - \frac{\left(c_7+2c_8\mu\right)^2c_8\sigma^4}{1+4c_8^2\sigma^4}\right)\right).
\end{aligned} \tag{E.12}$$

We can simplify this to

$$\begin{aligned}
\phi_{\mathbf{X}_{m+1}^{\Delta}}(u_1,u_2|\mathbf{x}) \;=\;& \frac{\cosh\left(ic_6\Delta t\right)\exp\left(iu_1\left[x_1+m_1\left(\mathbf{x}\right)\Delta t\right] + iu_2\left[x_2+m_2\left(\mathbf{x}\right)\Delta t\right]\right)}{\sqrt{(1-2ic_4\Delta t)(1-2ic_5\Delta t) + c_3^2\left(\Delta t\right)^2}}\\
&\cdot\;\exp\left(-\frac{\Delta t}{2}\frac{c_1^2+c_2^2+\left[4(c_2^2c_4^2+c_1^2c_5^2) - 4c_1c_2c_3(c_4+c_5) + (c_1^2+c_2^2)c_3^2\right]\left(\Delta t\right)^2}{1+\left(2c_3^2+4c_4^2+4c_5^2\right)\left(\Delta t\right)^2 + \left(c_3^2-4c_4c_5\right)^2\left(\Delta t\right)^4}\right)\\
&\cdot\;\exp\left(i\left(\Delta t\right)^2\left(\frac{-c_1^2c_4 - c_2^2c_5 - c_1c_2c_3 + (c_3^2-4c_4c_5)(c_1^2c_5 - c_1c_2c_3 + c_2^2c_4)\left(\Delta t\right)^2}{1+\left(2c_3^2+4c_4^2+4c_5^2\right)\left(\Delta t\right)^2 + \left(c_3^2-4c_4c_5\right)^2\left(\Delta t\right)^4}\right)\right).
\end{aligned} \tag{E.13}$$

In a similar way, we found that the above characteristic function is true for $c_4 = 0$ and/or $c_5 = 0$ as well.

Appendix F

Adjusted-Predictor-Corrector schemes

For the *Heston* model, where X_t^1 denotes the *log forward process* and no mean reverting for the volatility X_t^2 is included, we have the FSDEs,

$$\mathrm{d}X_t^1 = -\frac{1}{2}X_t^2\,\mathrm{d}t + \sqrt{X_t^2}\,\mathrm{d}W_t^1, \tag{F.1}$$

$$\mathrm{d}X_t^2 = \rho\gamma\sqrt{X_t^2}\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\gamma\sqrt{X_t^2}\,\mathrm{d}W_t^2. \tag{F.2}$$

The corresponding bivariate characteristic function of $\left(X_{m+1}^{1,\Delta}, X_{m+1}^{2,\Delta}\right)$ is given for $\left(X_m^{1,\Delta}, X_m^{2,\Delta}\right) = (x_1, x_2)$, by using Lemma 4.1 where

$$\begin{array}{ll}
m_1(x_1,x_2) = -\frac{1}{2}x_2 - \frac{1}{2}\eta_1\rho\gamma, & s_1^{W^1}(x_1,x_2) = \sqrt{x_2} - \frac{1}{2}\theta_1\rho\gamma\sqrt{x_2}\Delta t, \\
\kappa_1^{W^1,W^2}(x_1,x_2) = \frac{1}{2}\eta_1\sqrt{1-\rho^2}\gamma, & s_1^{W^2}(x_1,x_2) = -\frac{1}{2}\theta_1\sqrt{1-\rho^2}\gamma\sqrt{x_2}\Delta t, \\
\kappa_1^{W^1}(x_1,x_2) = \frac{1}{2}\eta_1\rho\gamma, & \kappa_1^{W^1}(x_1,x_2) = 0, \\
v_1(x_1,x_2) = 0, & \\
 & \\
m_2(x_1,x_2) = -\frac{1}{2}\eta_2\gamma^2, & s_2^{W^1}(x_1,x_2) = \rho\gamma\sqrt{x_2}, \\
\kappa_2^{W^1,W^2}(x_1,x_2) = \eta_2\rho\sqrt{1-\rho^2}\gamma^2, & s_2^{W^2}(x_1,x_2) = \sqrt{1-\rho^2}\gamma\sqrt{x_2}, \\
\kappa_2^{W^1}(x_1,x_2) = \frac{1}{2}\eta_2\rho^2\gamma^2, & \kappa_2^{W^2}(x_1,x_2) = \frac{1}{2}\eta_2(1-\rho^2)\gamma^2, \\
v_2(x_1,x_2) = 0. &
\end{array} \tag{F.3}$$

For the *SABR* model, where F_t denotes the *forward process* and σ_t denotes the *volatility process*, we have the FSDEs,

$$\mathrm{d}F_t = \sigma_t (F_t)^\beta\,\mathrm{d}W_t^1, \tag{F.4}$$

$$\mathrm{d}\sigma_t = \rho\nu\sigma_t\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\sigma_t\,\mathrm{d}W_t^2. \tag{F.5}$$

The corresponding bivariate characteristic function of $\left(F_{m+1}^\Delta, \sigma_{m+1}^\Delta\right)$ is given for $\left(F_m^\Delta, \sigma_m^\Delta\right) = (f, \alpha)$, by using Lemma 4.1 where

$$m_1(f,\alpha) = -\eta_1\left[\beta\alpha^2 f^{2\beta-1} + \rho\nu\alpha f^\beta\right], \tag{F.6a}$$

$$s_1^{W^1}(f,\alpha) = \alpha f^\beta - \eta_1\theta_1\left[3\rho\beta\nu\alpha^2 f^{2\beta-1} + \beta(2\beta-1)\alpha^3 f^{3\beta-2} + \rho^2\nu^2\alpha f^\beta\right]\Delta t, \tag{F.6b}$$

$$s_1^{W^2}(f,\alpha) = -\eta_1\theta_1\sqrt{1-\rho^2}\left[2\beta\nu\alpha^2 f^{2\beta-1} + \rho\nu^2\alpha f^\beta\right]\Delta t, \tag{F.6c}$$

$$\begin{aligned}
\kappa_1^{W^1,W^2}(f,\alpha) &= \eta_1\sqrt{1-\rho^2}\left[\nu\alpha f^{\beta}-\theta_1\beta\left\{3\rho\nu^2\alpha^2 f^{2\beta-1}+2(2\beta-1)\nu\alpha^3 f^{3\beta-2}\right\}\Delta t\right], && \text{(F.6d)}\\
\kappa_1^{W^1}(f,\alpha) &= \eta_1\left[\beta\alpha^2 f^{2\beta-1}+\rho\nu\alpha f^{\beta}-\theta_1\left\{2\rho^2\beta\nu^2\alpha^2 f^{2\beta-1}+\left(2\beta^3-3\beta^2+\beta\right)\alpha^4 f^{4\beta-3}\right.\right.\\
&+ \left.\left.\frac{1}{2}\rho\beta(9\beta-5)\nu\alpha^3 f^{3\beta-2}\right\}\Delta t\right], && \text{(F.6e)}\\
\kappa_1^{W^2}(f,\alpha) &= -\eta_1\theta_1\left(1-\rho^2\right)\beta\nu^2\alpha^2 f^{2\beta-1}\Delta t, && \text{(F.6f)}\\
v_1(f,\alpha) &= 0, && \text{(F.6g)}
\end{aligned}$$

$$\begin{aligned}
&m_2(f,\alpha) = -\eta_2\nu^2\alpha, && s_2^{W^1}(f,\alpha) = \rho\nu\alpha-\theta_2\eta_2\rho\nu^3\alpha\Delta t,\\
&s_2^{W^2}(f,\alpha) = \sqrt{1-\rho^2}\nu\alpha-\theta_2\eta_2\sqrt{1-\rho^2}\nu^3\alpha\Delta t, && \kappa_2^{W^1,W^2}(f,\alpha) = 2\eta_2\rho\sqrt{1-\rho^2}\nu^2\alpha,\\
&\kappa_2^{W^1}(f,\alpha) = \eta_2\rho^2\nu^2\alpha, && \kappa_2^{W^2}(f,\alpha) = \eta_2(1-\rho^2)\nu^2\alpha,\\
&v_2(f,\alpha) = 0.
\end{aligned} \quad \text{(F.6h)}$$

For the *SABR* model, where X_t^1 denotes the *log forward process* and σ_t denotes the *volatility process*, we have the FSDEs

$$\begin{aligned}
\mathrm{d}X_t^1 &= -\frac{1}{2}\sigma_t^2\exp\left(2(\beta-1)X_t^1\right)\mathrm{d}t+\sigma_t\exp\left((\beta-1)X_t^1\right)\mathrm{d}W_t^1, && \text{(F.7)}\\
\mathrm{d}\sigma_t &= \rho\nu\sigma_t\,\mathrm{d}W_t^1+\sqrt{1-\rho^2}\nu\sigma_t\,\mathrm{d}W_t^2. && \text{(F.8)}
\end{aligned}$$

We find the corresponding bivariate characteristic function of $\left(X_{m+1}^{1,\Delta},\sigma_{m+1}^{\Delta}\right)$, given $\left(X_m^{1,\Delta},\sigma_m^{\Delta}\right)=(x_1,\alpha)$, by using Lemma 4.1 where

$$\begin{aligned}
m_1(x_1,\alpha) &= -\theta_1\left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)\alpha^2 A^2(x_1)+\eta_1\rho\nu\alpha A(x_1)\right] && \text{(F.9a)}\\
&- (1-\theta_1)\left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)\alpha^2\exp\left(2(\beta-1)x_1\right)+\eta_1\rho\nu\alpha\exp((\beta-1)x_1)\right],\\
s_1^{W^1}(x_1,\alpha) &= \eta_1\alpha A(x_1)+(1-\eta_1)\alpha\exp((\beta-1)x_1)\\
&- \theta_1\left[(1+2\eta_1(\beta-1))\left[(\beta-1)\alpha^3\exp((\beta-1)x_1)+\rho\nu\alpha^2\right]A^2(x_1)\right.\\
&+ \left.\eta_1\rho\nu\left[(\beta-1)\alpha^2\exp((\beta-1)x_1)+\rho\nu\alpha\right]A(x_1)\right]\Delta t, && \text{(F.9b)}\\
s_1^{W^2}(x_1,\alpha) &= -\theta_1\sqrt{1-\rho^2}\left[(1+2\eta_1(\beta-1))\nu\alpha^2 A^2(x_1)+\eta_1\rho\nu^2\alpha A(x_1)\right]\Delta t, && \text{(F.9c)}\\
\kappa_1^{W^1,W^2}(x_1,\alpha) &= \eta_1\sqrt{1-\rho^2}\nu\alpha A(x_1)\\
&- \theta_1\sqrt{1-\rho^2}\alpha^2\left[(1+2\eta_1(\beta-1))\left[2(\beta-1)\nu\alpha\exp((\beta-1)x_1)+\rho\nu^2\right]A^2(x_1)\right.\\
&+ \left.\eta_1\rho(\beta-1)\nu^2\exp((\beta-1)x_1)A(x_1)\right]\Delta t, && \text{(F.9d)}\\
\kappa_1^{W^1}(x_1,\alpha) &= \eta_1\left[(\beta-1)\alpha^2\exp\left((\beta-1)x_1\right)+\rho\nu\alpha\right]A(x_1)\\
&- \theta_1\left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)\left[\rho^2\nu^2\alpha^2+4\rho(\beta-1)\nu\alpha^3\exp\left((\beta-1)x_1\right)\right.\right.\\
&+ \left.2(\beta-1)^2\alpha^4\exp\left(2(\beta-1)x_1\right)\right]A^2(x_1) && \text{(F.9e)}\\
&+ \left.\eta_1\rho\nu\left(\rho(\beta-1)\nu\alpha^2\exp((\beta-1)x_1)+\frac{1}{2}(\beta-1)^2\alpha^3\exp\left(2(\beta-1)x_1\right)\right)A(x_1)\right]\Delta t,\\
\kappa_1^{W^2}(x_1,\alpha) &= -\theta_1\left(\frac{1}{2}+\eta_1(\beta-1)\right)(1-\rho^2)\nu^2\alpha^2 A^2(x_1)\Delta t, && \text{(F.9f)}\\
v_1(x_1,\alpha) &= 0, && \text{(F.9g)}
\end{aligned}$$

$$m_2(x_1,\alpha) = -\eta_2\nu^2\alpha, \qquad s_2^{W^1}(x_1,\alpha) = \rho\nu\alpha - \theta_2\eta_2\rho\nu^3\alpha\Delta t,$$
$$s_2^{W^2}(x_1,\alpha) = \sqrt{1-\rho^2}\nu\alpha - \theta_2\eta_2\sqrt{1-\rho^2}\nu^3\alpha\Delta t, \qquad \kappa_2^{W^1,W^2}(x_1,\alpha) = 2\eta_2\rho\sqrt{1-\rho^2}\nu^2\alpha,$$
$$\kappa_2^{W^1}(x_1,\alpha) = \eta_2\rho^2\nu^2\alpha, \qquad \kappa_2^{W^2}(x_1,\alpha) = \eta_2(1-\rho^2)\nu^2\alpha,$$
$$v_2(x_1,\alpha) = 0, \tag{F.9h}$$

where

$$A(x_1) = \exp\left((\beta-1)x_1 - \frac{1}{2}(\beta-1)\exp\left(2(\beta-1)x_1\right)\Delta t\right). \tag{F.10}$$

For the *SABR* model, where X_t^1 denotes the *log forward process* and X_t^2 denotes the *log volatility process*, we have the FSDEs

$$\mathrm{d}X_t^1 = -\frac{1}{2}\exp\left(2X_t^2 + 2(\beta-1)X_t^1\right)\mathrm{d}t + \exp\left(X_t^2 + (\beta-1)X_t^1\right)\mathrm{d}W_t^1, \tag{F.11}$$

$$\mathrm{d}X_t^2 = -\frac{1}{2}\nu^2\,\mathrm{d}t + \rho\nu\,\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\nu\,\mathrm{d}W_t^2. \tag{F.12}$$

We determine the corresponding bivariate characteristic function of $\left(X_{m+1}^{1,\Delta}, X_{m+1}^{2,\Delta}\right)$, given $\left(X_m^{1,\Delta}, X_m^{2,\Delta}\right) = (x_1,x_2)$, by using Lemma 4.1 where

$$\begin{aligned}
m_1(x_1,x_2) &= -\theta_1\left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)A^2(x_1,x_2) + \eta_1\rho\nu A(x_1,x_2)\right] - (1-\theta_1) \\
&\cdot \left[\left(\frac{1}{2}+\eta_1(\beta-1)\right)\exp\left(2x_2+2(\beta-1)x_1\right) + \eta_1\rho\nu\exp\left(x_2+(\beta-1)x_1\right)\right],
\end{aligned} \tag{F.13a}$$

$$\begin{aligned}
s_1^{W^1}(x_1,x_2) &= \eta_1 A(x_1,x_2) + (1-\eta_1)\exp\left(x_2+(\beta-1)x_1\right) \\
&- \theta_1\left[\rho\nu + (\beta-1)\exp\left(x_2+(\beta-1)x_1\right)\right] \\
&\cdot \left[\left(1+2\eta_1(\beta-1)\right)A^2(x_1,x_2) + \eta_1\rho\nu A(x_1,x_2)\right]\Delta t,
\end{aligned} \tag{F.13b}$$

$$s_1^{W^2}(x_1,x_2) = -\theta_1\sqrt{1-\rho^2}\nu\left[\left(1+2\eta_1(\beta-1)\right)A^2(x_1,x_2) + \eta_1\rho\nu A(x_1,x_2)\right]\Delta t, \tag{F.13c}$$

$$\begin{aligned}
\kappa_1^{W^1,W^2}(x_1,x_2) &= \eta_1\sqrt{1-\rho^2}\nu A(x_1,x_2) - \theta_1\sqrt{1-\rho^2}\left[\rho\nu^2 + (\beta-1)\nu\exp\left(x_2+(\beta-1)x_1\right)\right] \\
&\cdot \left[2\left(1+2\eta_1(\beta-1)\right)A^2(x_1,x_2) + \eta_1\rho\nu A(x_1,x_2)\right]\Delta t,
\end{aligned} \tag{F.13d}$$

$$\begin{aligned}
\kappa_1^{W^1}(x_1,x_2) &= \eta_1\left[\rho\nu + (\beta-1)\exp\left(x_2+(\beta-1)x_1\right)\right]A(x_1,x_2) \\
&- \theta_1\left[\rho\nu + (\beta-1)\exp\left(x_2+(\beta-1)x_1\right)\right]^2 \\
&\cdot \left[\left(1+2\eta_1(\beta-1)\right)A^2(x_1,x_2) + \frac{1}{2}\eta_1\rho\nu A(x_1,x_2)\right]\Delta t,
\end{aligned} \tag{F.13e}$$

$$\kappa_1^{W^2}(x_1,x_2) = -\theta_1(1-\rho^2)\nu^2\left[\left(1+2\eta_1(\beta-1)\right)A(x_1,x_2)^2 + \frac{1}{2}\eta_1\rho\nu A(x_1,x_2)\right]\Delta t, \tag{F.13f}$$

$$v_1(x_1,x_2) = 0, \tag{F.13g}$$

$$\begin{aligned}
&m_2(x_1,x_2) = -\tfrac{1}{2}\nu^2, \qquad s_2^{W^1}(x_1,x_2) = \rho\nu, \qquad s_2^{W^2}(x_1,x_2) = \sqrt{1-\rho^2}\nu, \\
&\kappa_2^{W^1,W^2}(x_1,x_2) = 0, \qquad \kappa_2^{W^1}(x_1,x_2) = 0, \qquad \kappa_2^{W^2}(x_1,x_2) = 0, \\
&v_2(x_1,x_2) = 0,
\end{aligned} \tag{F.13h}$$

where

$$A(x_1,x_2) = \exp\left(x_2 + (\beta-1)x_1 - \frac{1}{2}\nu^2\Delta t - \frac{1}{2}(\beta-1)\exp\left(2x_2+2(\beta-1)x_1\right)\Delta t\right). \tag{F.14}$$