Technische Universität München

Fakultät für Mathematik

Polynomial Preserving Processes and Application to Finance

Master's Thesis by Wahid Khosrawi-Sardroudi

Supervisor: Prof. Dr. Claudia Klüppelberg

Advisor: Prof. Dr. Damir Filipovic (École polytechnique fédérale de Lausanne)

Submission date: December 1'st, 2014

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Ecublens, November 30'th 2014

Acknowledgements

I would like to thank my advisor Prof. Dr. Damir Filipovic (EPFL) for the chance to write my thesis under his supervision. I am in particular grateful for the chance of experiencing current research. I would further like to thank Prof. Dr. Claudia Klüppelberg for introducing me to Prof. Filipovic. I would also like to thank Prof. Dr. Daniel Kressner, Prof. Dr. Martin Larsson, Dr. Sergio Pulido and Prof. Dr. André Uschmajew for their constant help and support.
Finally I want to thank my parents for their love and for being there for me whenever I needed them.

Zusammenfassung

Das Thema dieser Masterarbeit sind Polynomerhaltende prozesse and deren Anwendung in der Finanzmathematik. Polynomerhaltende Prozesse wie von Filiovic und Larsson definiert sind solche Diffusionsprozesse, deren Generator invariant auf beliebigem Polynomraum ist. Dies ermoeglicht die Berechnung der (gemischten) Momente in dem ein Matrixexponential bestimmt wird. Anders als affine Prozesse, die einen Spezialfall von polynomerhaltenden prozessen bilden, kann die characteristische Funktion nicht mehr als Loesung einer gewoehnlichen Differentialgleichung dargestellt werde. Daher ist es noetig auf weitestgehend Moment basierenden Approximationsmethoden zurueckzugreifen. Dies geschieht im zweiten Teil mit dem Ziel der Bewertung Europaeischer Optionen. Insbesondere fuer das power Heston model, ein Model das wir vorstellen und besprechen werden.

Contents

List of Figures

List of Tables

Introduction

Continuous time market models have been applied successfully in the last decades to financial mathematics, starting from Bachelier. In particular the contributions of Black, Scholes and Merton need to be mentioned and their infamous Black-Scholes-Merton model. However, the assumptions in their models can not explain certain observations made about financial markets and hence more sophisticated models need to be applied and studied.

One of the mentioned shortcomings of the Black-Scholes-Merton model is that it assumes a constant volatility, which is in contradiction to the observed volatility smile. One way to deal with this problem are stochastic volatility models. These are multi factor models, where one or several factors describe the volatility of another factor. Probably the most famous and most widely used of these models is the Heston model. This model has the advantage of having a characteristic function that is known in closed form, see [17]. Therefore Fourier inversion techniques based on Plancherel's theorem can be applied for a wide range of applications such as pricing European options. There are several extensions to this model to include other phenomena observed in financial markets such as jumps.

The Heston model itself belongs to a wide class of market models based on affine processes. These processes allow the computation of the characteristic function by solving a system of Riccati equations. Affine models have been successfully applied in finance, e.g. to model term structures, see e.g. [9] and [7].

There are several reasons for leaving the affine case for a more general setup. One would be the need for non trivial dynamics on compact state spaces, see the introductory part in [12]. In this thesis we want to present a bigger class of stochastic processes, namely polynomial preserving processes or to be more precise, diffusions as we only present the theory as proposed by Filipovic and Larsson in [12]. Polynomial preserving processes have appeared in the literature for a long time, see Wong 1964 in [32], Filipovic and Larsson gave the first existence and uniqueness analysis for polynomial preserving diffusions in more then one dimension. Several authors have already studies polynomial preserving processes, among them Mazet in [22], Zhou in [33] or Forman and Sørensen in [15].

A first rigorous treatment of the time homogenous Markov process case with jumps was provided by Cuchiero and Cuchiero et al. in [6, 5]. The concept there was to consider a Markovian setting and conditions that would allow for the computation of moments by means of matrix exponentiation. The insight was that if the Markov semigroup is invariant on any polynomial space, then under certain assumptions, there exists a representing matrix of the generator whose matrix exponential is the representing matrix of the semigroup. Filipovic and Larsson take another approach. First they consider only the pure Itô diffusion case. Further, they define the polynomial preserving property not via the invariance of the semigroup on polynomial spaces, but rather by starting from the generator. The results that then allow for the computation of moments do not rely on a Markov setting, but rather standard stochastic calculus. This gives them enough flexibility to link the question of existence and uniqueness of polynomial preserving diffusions to the associated martingale problem.

In this thesis we want to present their concept and demonstrate applications to finance. These applications will be the pricing of European options in a stochastic volatility framework. In particular we will discuss the Power Heston model, see [11], and proof its weak convergence for any fixed time to the Heston model.

This thesis is structured as follows. In the first part, we begin by introducing the concept of polynomial preserving diffusions as presented by Filipovic and Larsson. We present certain results such as the computation of moments. We continue by giving examples of such diffusions and introduce the previously mentioned power Heston model. We finish the first part by showing that the characteristic function in the power Heston model is entire, a conjecture made by Filipovic, and that prices of European type options converge to those in the Heston model as a certain parameter increases.

In the second part we present two applications, one is the approximation of conditional transition densities. There we present explicit numerical examples for the power Heston model and discuss those. Finally we will give a short presentation of the concepts introduced in [10].

Part I

Polynomial preserving diffusions

Chapter 1

Definition and properties of polynomial preserving diffusions

1.1. Definitions

We begin by presenting the mathematical setup as introduced by Filipovic and Larsson in [12]. Unlike Cuchiero [5] and Cuchiero et al. [6], the starting point of the investigation is not the semigroup of a time homogenous Markov process. Filipovic and Larsson start their investigations by considering the partial differential operator

$$\mathcal{G}f = \frac{1}{2}Tr(a\nabla^2 f) + \nabla f^\top b,$$

defined on $\mathcal{C}^2(\mathbb{R}^d)$ with $a : \mathbb{R}^d \to \mathbb{S}^d$, $b : \mathbb{R}^d \to \mathbb{R}^d$ continuous functions. Let $E \subset \mathbb{R}^d$ arbitrary. We want $\mathcal{G}$ to be the generator of an E valued diffusion, hence we assume

$$a(x) \in \mathbb{S}^d_+ \quad \text{for all } x \in E.$$

Filipovic and Larsson make use of the duality between the (weak) solution of an stochastic differential equation with generator $\mathcal{G}$ and the martingale problem associated to $\mathcal{G}$. This allows to combine methods from each point of view such as local times for stochastic processes to prove their results regarding existence and uniqueness of solutions. We follow their example and give a short review.

Let W_t be a $d-$dimensional standard Brownian motion. Consider the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{SDE}$$

with $\sigma : \mathbb{R} \to \mathbb{R}^{d\times m}$ any continuous map such that $\sigma\sigma^\top = a$. Further we require that the solution X takes values in E.

Following Rogers and Williams in [26] (Theorem V.20.1), the law of X can equivalently be described as the solution $\mathbb{P}$ to the martingale problem that we will now define.

Let $\Omega = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ be the space of $\mathbb{R}^d$ valued continuous functions. We equip Ω with the topology induced by the locally uniform convergence. Let $\mathcal{F}$ be the Borel σ-field on Ω with respect to this topology. The coordinate process is denoted by $X = (X_t)_{t\geq 0}$, i.e. X is a random variable on the above probability space and its value at time t is $X_t(\omega), \omega \in \Omega$. Let $(\mathcal{F}_t)_{t\geq 0}$ be the canonical right-continuous filtration of the coordinate process. If μ is a probability measure on the Borel σ-field $\mathcal{B}(\mathbb{R}^d)$ supported on E, we say that a probability measure $\mathbb{P}$ is a solution to the martingale problem for $(\mathcal{G}, E)$ with initial law μ if

$$P(X_o \in \cdot) = \mu, \quad \mathbb{P}(X_t \in E) = 1 \, \forall t \geq 0,$$

and the process (MP)

$$N_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_s)\, ds$$

is a martingale for all $f \in \mathcal{C}_c^\infty$. Uniqueness holds for the martingale problem for $(\mathcal{G}, E)$ with initial law μ if at most one solution exists.

Remark 1.1. *For the initial value X_0 in* (SDE) *deterministic, μ is the point measure on that initial value. Further uniqueness in the martingale problem is equivalent to uniqueness in law for E valued solutions in* (SDE)*. For results and definition of unique solutions to stochastic differential equations, see e.g. [26], chapter V.3.*

Filipovic and Larsson continue by showing that existence to (MP) for $(\mathcal{G}, E)$ for arbitrary initial law implies that $\mathcal{G}$ is well defined on $\mathcal{C}^2(E)$, see Lemma 2.1 in [12]. We can now state the definition of polynomial preserving diffusions. Let $Pol_n(E)$ denote the set of d-variate polynomials on E with at most degree n.

Definition 1.2. *An E valued solution to* (SDE) *is called polynomial preserving diffusion if the operator $\mathcal{G}$ is polynomial preserving, i.e. $\mathcal{G}(Pol_n(E)) \subset Pol_n(E)$ for all $n \geq 0$.*

The definition of polynomial preserving diffusions by their generators allows for a simple and intuitive characterization.

Lemma 1.3 (Lemma 2.3 in [12])**.** *The following are equivalent*

1. *$\mathcal{G}$ is polynomial preserving.*

2. $\mathcal{G}(Pol_n(E)) \subset Pol_n(E)$ *for* $n \in \{1, 2\}$.

3. *The components of the coefficients* a *and* b *restricted to* E *are in* $Pol_2(E)$ *and* $Pol_1(E)$*, respectively.*

In particular *3.* is a useful characterization, since it allows to check for the polynomial preserving property by simply looking at the generator coefficients. For the proof of the lemma we refer the reader to [12]. Filipovic and Larsson further make the convention that whenever a and b have components in $Pol_2(E)$ and $Pol_1(E)$, they assume they lie globally in $Pol_2(\mathbb{R}^d)$ and $Pol_1(\mathbb{R}^d)$, respectively, whenever $\mathcal{G}$ is polynomial preserving.

An important observation is the following. The above definition does not require $\mathcal{G}$ to be the generator of a Markov process. Likewise uniqueness is not required. There are two reasons stated in [12]. First, existence of a solution in (MP) does by itself not require the solution to be a Markov processes. In the terminology of Rogers and Williams that would require (MP) to be well posed. And two, in the context of Markov processes the polynomial preserving property holds if and only if the corresponding semigroup $(P_t)_{t\geq 0}$ is invariant on $Pol_n(E)$ $\forall n$. Further down we will see, that this property can be established without a Markov setting. Only Itô calculus for (SDE) is necessary to establish this result, see Theorem 1.5.

1.2. Computation of power moments and moment determinacy

In this chapter we will present the results most important for our numerical approximation methods in the second part of this thesis. To be more precise, we show the results in [12] regarding the computation of power moments of polynomial preserving processes by means of matrix exponentiation.

Let N_n denote the dimension of $Pol_n(E)$. If E has non empty interior, combinatorial arguments show that

$$N_n = \binom{n+d}{n}.$$

Let $\beta_n = \{h_1, \ldots, h_{N_n}\}$ be any basis of $Pol_n(E)$. Define

$$H_n(x) = (h_1(x), \ldots, h_{N_n}(x))^\top.$$

In particular, any $p \in Pol_n(E)$ can by uniquely identified with an $\vec{p} \in \mathbb{R}^{N_n}$ such that

$$p(x) = H_n(x)^\top \vec{p}. \tag{1.1}$$

Coming back to $\mathcal{G}$ and noticing that by the invariance property and linearity of the operator there exists a unique representing matrix G_n for $\mathcal{G}|_{Pol_n(E)}$ with respect to the basis β, we can deduce

$$\mathcal{G}p(x) = H_n(x)^\top G_n \vec{p} = \vec{p}^\top G_n^\top H_n(x). \tag{1.2}$$

Remark 1.4. *The above equation can be used to compute the representing matrix for given basis β, since one only needs to compute the action of $\mathcal{G}$ on the basis polynomials and express the results as linear combination of the basis polynomials.*

We now state the theorem that allows the computation of conditional moments.

Theorem 1.5 (Theorem 3.2 in [12])**.** *Suppose $\mathcal{G}$ is polynomial preserving, and let $\mathbb{P}$ be a solution to the martingale problem for $(\mathcal{G}, E)$. If $\mathbb{E}\left[||X_0||^{2n}\right] < \infty$ for $n \in \mathbb{N}$, then for any $p \in Pol_n(E)$ with coordinate representation $\vec{p} \in \mathbb{R}^{N_n}$, it holds*

$$\mathbb{E}\left[p(X_t)|\mathcal{F}_t\right] = H_n(X_t) \top e^{(T-t)G_n}\vec{p},$$

for $t \leq T$.

Proof. We only give a sketch of the proof as given in [12]. Let X be a weak solution to (SDE). Let $p \in Pol_n(E)$ with coordinate vector $\vec{p} \in \mathbb{R}^{N_n}$. By Itô's formula it holds

$$p(X_u) = p(X_t) + \int_t^u \mathcal{G}p(X_s)\, ds + \int_t^u \nabla p(X_s)^\top \sigma(X_s) dW_s.$$

The next step is to argue why the stochastic integral above is a true martingale, see Lemma 3.1 in [12]. Using linearity of expectation and integration together with (1.1) and (1.2) it holds

$$\begin{aligned}\vec{p}^\top \mathbb{E}\left[H_n(X_u)|\mathcal{F}_t\right] = \mathbb{E}\left[p(X_u)|\mathcal{F}_t\right] &= p(X_t) + \mathbb{E}\left[\int_t^u \mathcal{G}p(X_s)\, ds|\mathcal{F}_t\right] \\ &= \vec{p}^\top H_n(X_t) + (G_n\vec{p})^\top \mathbb{E}\left[\int_t^u H_n(X_s)\, ds|\mathcal{F}_t\right].\end{aligned}$$

The next step in [12] is to justify Fubini's theorem. As a consequence this yields

$$\vec{p}^\top F(u) = \vec{p}^\top H_n(X_t) + \vec{p}^\top G_n^\top \int_t^u F(s)\, ds,$$

where $t \leq u \leq T$ and $F(u) = \mathbb{E}\left[H_n(X_u)|\mathcal{F}_t\right]$. Choosing unit vectors for $\vec{p}$ this gives a system of linear integral equations for $F(u)$ with unique solution $F(u) = e^{(u-t)G_n^\top} H_n(X_t)$. In conlusion

$$\mathbb{E}\left[p(X_T)|\mathcal{F}_t\right] = F(T)^\top \vec{p} = H_n(X_t)^\top e^{(T-t)G_n}\vec{p}.$$

□

As mentioned before, we want to stress the point that the proof does not require a Markovian setting. The retrieval of the semigroup from the infinitesimal generator is usually known in the context of Feller processes, or as in Cuchiero et al. for Markov processes with a true martingale assumption. Here, we only relied on basic stochastic calculus.

We conclude with the following corollary (the proof is in [12]). But first recall multi-index notation, i.e. for $\boldsymbol{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, $x \in \mathbb{R}^d$ we write for the $\boldsymbol{k}$- power moment $x^{\boldsymbol{k}} := x_1^{k_1} \cdots x_d^{k_d}$. Further the length of $\boldsymbol{k}$ is defined by $|\boldsymbol{k}| := k_1 + \cdots + k_d$.

Corollary 1.6 (Corollary 3.3 in [12])**.** *Under the assumptions of Theorem 1.5, in addition to X_0 being deterministic, it holds that the expectation of any power moment of any finite dimensional distribution of X, i.e. for all multi-indices $\boldsymbol{k}(1), \dots, \boldsymbol{k}(m)$ and $0 \leq t_1 < \cdots < t_m$ the expectation*

$$\mathbb{E}\left[X_{t_1}^{\boldsymbol{k}(1)} \cdots X_{t_m}^{\boldsymbol{k}(m)}\right],$$

is uniquely determined by $\mathcal{G}$.

Clearly, the condition $\mathbb{E}\left[||X_0||^{2n}\right] < \infty$ above holds for all $n \geq 0$ for deterministic intitial value of the stochastic differential equation, or in the martingale problem terminology, for degenerated initial law μ concentrated on one point $x_0 \in E$.

Filipovic and Larsson continue by providing sufficient conditions for finite exponential moments. They use the fact that for affine diffusions this is given, so they split the quadratic part of the coefficient a of $\mathcal{G}$ from the affine part. Formally this can be stated as follows. Due to Lemma 1.3, the components of a can be written as

$$a_{ij}(x) = \alpha_{ij} + \sum_{k=1}^{d} a_{ij}^k x_k + \sum_{k,l=1}^{d} A_{ij}^{kl} x_k x_l,$$

where the coefficients of these polynomials are real numbers. Now take the monomials of degree two above and define the matrix $A(x)$ componentwise by

$$A_{ij}(x) = \sum_{k,l=1}^{d} A_{ij}^{kl} x_k x_l.$$

Theorem 1.7 (Theorem 3.4 in [12])**.** *Let $\mathcal{G}$ be polynomial preserving, $\mathbb{P}$ be a solution to the martingale problem for $(\mathcal{G}, E)$ with initial value $x_0 \in E$, i.e. deterministic initial law concentrated on x_0. Let $T \geq 0$ and A as above. If*

$$\sup_{x \in E} ||A(x)|| < \infty,$$

the for all ϵ sufficiently small, we have

$$\mathbb{E}\left[\exp\left(\epsilon||X_T||\right)\right] \leq \exp\left(c_1 + c_2||x_0||\right),$$

where c_1 and c_2 are some nonnegative constants.

The theorem stated above will be useful when checking for moment determinacy of the process. We do not give proof of this theorem but refer to the original paper.

Filipovic and Larsson continue presenting results regarding uniqueness results for the martingale problem (or equivalently to ((SDE)). We only want to present one of their results and demonstrate the proof at this point. For the other results we refer to the original paper.

Theorem 1.8 (Theorem 4.1 in [12])**.** *Suppose $\mathcal{G}$ is polynomial preserving, $\mathbb{P}$ a solution to the martingale problem for initial value $x_0 \in E$. If for each $t \geq 0$ there exists $\epsilon > 0$ such that $\mathbb{E}\left[\exp(\epsilon||X_t||)\right] < \infty$, then the solution $\mathbb{P}$ is unique.*

Proof. The proof is motivated on the classical moment problem. First notice that if a distribution has analytical characteristic function around zero, then all moments exist and the distribution is moment determinant which follows from the power series representation of the characteristic function (or moment generating function), see e.g. [20] and [19]. By assumption we have for the component processes $\mathbb{E}\left[\exp(\epsilon|X_{it})|\right] < \infty$ for all $t \geq 0$. Hence the characteristic function of each is analytical on $B_\epsilon(0)$ ([20] chapter on analytical characteristic functions or Theorem 1.7.1 in [31]) and by that moment determinant. By [25] (Theorem 3) , marginal moment determinancy of $\mathbb{R}^d$ implies moment determinacy of the measure itself. In conclusion the power moments of the finite dimensional distributions determine the law uniquely. The theorem is proved with Corollary 1.6. □

Remark 1.9. *Theorem 4.2 in [12] states a general result on polynomial preserving operators $\mathcal{G}$ in one dimension, name that uniqueness for the martingale problem for $(\mathcal{G}, E)$ holds if $d = 1$ and $\mathcal{G}$ polynomial preserving.*

Chapter 2

Examples of Polynomial Preserving Diffusions

We now conclude this part by presenting in short a few examples of polynomial preserving diffusions. In particular we will be interested in polynomial preserving processes that describe stochastic volatility models for financial markets.

In the sequel, denote by W_t a standard d-dimensional Brownian motion. When there is no ambiguity, we ommit the dimension d. For $d > 1$ we write the i-th component process as W_{it}. Polynomial preserving diffusions can be regarded as generalizations of affine diffusions in the sense that affine diffusions are polynomial preserving diffusions as well, while the other implication is not true anymore, see e.g. the example of the Jacobi process further down

2.1. Basic examples of polynomial preserving diffusions

Example 2.1 (Vasicek model)**.** *A common model for stochastic interest rates is the Vasicek model. Its dynamics are given by*

$$dr_t = a(b - r_t)dt + \sigma dW_t,$$

with constants a, b, r and σ. One shortcoming of this model is the fact that it can take negative values. Here, r_t is a Gaussian process with closed form expectation and variance.

Example 2.2 (The Cox-Ingersoll-Ross model)**.** *A common model for interest rates and stochastic volatilities. Its dynamics are given by*

$$dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}dW_t.$$

One advantage over the Vasicek model is that if a and b are chosen such that $ab > 0$, then r_t takes no negative values. In the further down presented Heston model this process will be used to model the volatility process.

Example 2.3 (Jacobi process)**.** *The Jacobi process is a polynomial preserving, but not affince diffusion with the property of beeing compactly supported. Its dynamics are given by*

$$dV_t = \kappa(\theta - V_t)dt + \sigma_v\sqrt{V_t(1 - V_t/\bar{v})}dW_t.$$

Provided $0 < \theta < \bar{v}, \kappa > 0$, this process does not die out and is supported on $[0, \bar{v}]$. In the power Heston model further down, the Jacobi process will be used as volatility process to construct a compactly supported stochastic volatility model.

2.2. Polynomial preserving stochastic volatility models

We now want to present the models most interesting for our application aims in finance, namely the pricing of European type options. One of the most prominent shortcomings in the Black-Scholes (-Merton) model is that it can not describe the volatility smile. One possible solution to this are stochastic volatility models as we present in the sequel.

Example 2.4 (The Heston model)**.** *The maybe most commonly used stochastic volatility model in practice is the Heston model. One advantage is that it is an affine diffusion whose characteristic function is known in closed form. Hence Fourier inversion methods, see e.g. [9], can be applied to price options. Denoting the volatility process as V_t and the log price process as L_t, the model dynamics are given by*

$$\begin{aligned} dV_t &= (b + \beta V_t)dt + \sigma\sqrt{V_t}dW_{1t} \\ dX_t &= (r - \frac{1}{2}V_t)dt + \sqrt{V_1}\left(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}\right), \end{aligned}$$

where $(W_{1,t}, W_{2,t})$ is a two dimensional standard Brownian motion, $\sigma > 0$, b and β chosen such that V_t is supported on the non negative (or positive, see Feller condition) real line, $\rho \in [-1, 1]$ usually negative (leverage effect) and r the risk less rate. The mean vector $c(x)$ and the diffusion matrix $a(x)$ ($x \in \mathbb{R}^2$) are given by

$$c(x) = \begin{bmatrix} b + \beta x_1 \\ r - \frac{1}{2}x_1 \end{bmatrix}, \quad a(x) = \begin{bmatrix} \sigma^2 x_1 & \sigma\rho x_1 \\ \sigma\rho x_1 & x_1 \end{bmatrix}.$$

We will now present the model of main interest in this thesis, the power Heston model as introduced in [11]. The idea is to replace the volatility process in the Heston model with a Jacobi process. As we will see this will have the advantage

of making the characteristic function of the log price process entire. This implies that the power series expansion of the characteristic function around zero, whose coefficients are given by the moments one can compute for polynomial preserving diffusions, has global radius of convergence, unlike the Heston model whose strip of analyticity is not the whole complex plane, see [9].

Simply replacing the volatility process will however cause the model to not be polynomial preserving anymore, unless the correlation ρ is zero. Filipoviv et al. present a solution to this problem in [11]. The dynamics of this solution are given by

Example 2.5 (The Power Heston model)**.**

$$dv_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t(1 - \frac{V_t}{\bar{v}_n})}dW_{1t}$$
$$dX_t = r - \frac{1}{2}v_t dt + \rho\sqrt{V_t(1 - \frac{V_t}{\bar{v}_n})}dW_{1t} + \sqrt{(1-\rho^2)V_t + \rho^2\frac{V_t^2}{\bar{v}_n}}dW_{2t}.$$

In this solution, the instantaneous correlation between V_t and X_t are

$$\frac{d\langle V, X\rangle_t}{\sqrt{d\langle V\rangle_t d\langle X\rangle_t}} = \rho\sqrt{1 - V_t/\bar{v}}.$$

Hence the correlation between stock return and squared volatility change is close to ρ for $V_t << \bar{v}$. The correlation vanishes as V_t goes to zero.

We now want to prove a certain result, which compares the Power Heston model to the Heston model. Before we do so, we want to make the parametrization comparable by fixing $b = \theta\kappa$ and $\beta = -\kappa$.

This way the mean vector $c(x)$ is the same as in the Heston model. The diffusion matrix now reads

$$a(x) = \begin{bmatrix} \sigma^2 x_1(1 - \frac{x_1}{\bar{v}_n}) & \sigma\rho x_1(1 - \frac{x_1}{\bar{v}_n}) \\ \sigma\rho x_1(1 - \frac{x_1}{\bar{v}_n}) & x_1 \end{bmatrix}.$$

Using this parametrization, we can now proof the following result.

Lemma 2.6. *Let $\bar{v}_n$ be a sequence of positive real numbers such that $\bar{v}_n \to \infty$ as $n \to \infty$. Let X^n be the log prices process in the power Heston model with $\bar{v} = \bar{v}_n$. The coordinate process is denoted by X_t^n.*

Further, let X^H be the log price process in the Heston model and let initial values (deterministic) and the parameters ρ, b, β and σ be the same in both models. Then for any fixed time t, it holds $X_t^n \Rightarrow X_t^H$ (converges weakly). As a consequence, European type option prices in the power Heston model converge to those in the Heston model as $n \to \infty$.

Proof. We will use the fact that the Heston model is moment determinant since the characteristic function is analytic around zero, see [9]. Therefore, by the method of moments (see e.g. Billingsley [4]) it suffices to show that any moment in the power Heston model converges to the corresponding moment in the Heston model as $\bar{v}_n \to \infty$.

denote by $\mathcal{G}^n$ the generator of the power Heston model with $\bar{v} = \bar{v}_n$. Likewise let $\mathcal{G}$ be the generator of the Heston model. In the same manner denote by $a^n(x)$ and $a(x)$ the respective diffusion matrices. It follows

$$a^n(x) = a(x) - \frac{1}{\bar{v}_n}\begin{bmatrix}\sigma^2 x_1^2 & \sigma\rho x_1^2 \\ \sigma\rho x_1^2 & 0\end{bmatrix}.$$

We will write r for the residual matrix on the right side above. It follows

$$\begin{aligned}\mathcal{G}^n f &= c^\top \nabla f + \frac{1}{2} Tr(a^n \nabla^2 f) \\ &= c^\top \nabla f + \frac{1}{2}\left(Tr(a\nabla^2 f) - \frac{1}{\bar{v}_n} Tr(r\nabla^2 f)\right) = \mathcal{G} - \frac{1}{2\bar{v}_n} Tr(r\nabla^2 f).\end{aligned}$$

Denote by G_n and G the representing matrix of $\mathcal{G}^n$ and $\mathcal{G}$ on $Pol_n(\mathbb{R}^2)$. In particular, $-1/2Tr(r\nabla^2)$ has a representing matrix on $Pol_n(\mathbb{R}^2)$ as well, we denote it with R. It follows from above that

$$G_n = G - \frac{1}{\bar{v}_n} R.$$

Since R does not depend on $\bar{v}_n$, the matrix $R/\bar{v}_n$ converges to the zero matrix. By continuity of the matrix exponential, we conclude that

$$e^{\tau G_n} \to e^{\tau G}, \quad \text{as } \bar{v}_n \to \infty,$$

for any real number τ. Theorem 1.5 now implies that all moments up to degree n converge to those in the Heston model. Since n was arbitrary, we have shown moment convergence, in particular for the moments of the log price, which proves the lemma. □

We will finish this chapter by showing a last property of the power Heston model, namely that the (joint) characteristic function is an entire function. We begin by establishing certain results.

Theorem 2.7 (Theorem 1.7.1 in [31])**.** *Let $F(x)$ be a distribution function with characteristic function $f(z), z \in i\mathbb{R}$. If*

$$\int_{\mathbb{R}} e^{r|x|} dF(x) < \infty$$

for all $0 \le r < R, R > 0$, then $f(z)$ is an analytic function in the circle $z < R$. The converse statement is true as well.

As a consequence we have the following corollary.

Corollary 2.8. *Let X be a real valued random variable, $\epsilon > 0$ and*

$$f(u) = \mathbb{E}[e^{uX}] < \infty \tag{2.1}$$

for all $u \in (-\epsilon, \epsilon)$. Then $f(z), z \in \mathbb{C}$ is analytic on the circle $|z| < \epsilon$.

Proof. Denote by X^+ the positive part of X and the negative part by X^-. In particular we then have $X = X^+ - X^-$ and $|X| = X^+ + X^-$. For $u < 0$ we can obviously follow $\mathbb{E}\left[\exp(u|X|)\right] < \infty$. Therefore assume for now $u > 0$. Then

$$\begin{aligned} \mathbb{E}\left[e^{u|X|}\right] = \mathbb{E}\left[e^{uX^+}e^{uX^-}\right] &= \mathbb{E}\left[e^{uX^+}\mathbb{1}_{\{X\geq 0\}}\right] + \mathbb{E}\left[e^{uX^-}\mathbb{1}_{\{X<0\}}\right] \\ &\leq \mathbb{E}\left[e^{uX^+}\right] + \mathbb{E}\left[e^{uX^-}\right]. \end{aligned} \tag{2.2}$$

We will now show that each summand is finite. We will use the fact that (2.1) is true for negative $u > -\epsilon$ as well. We have

$$\infty > \mathbb{E}\left[e^{uX}\right] = \mathbb{E}\left[e^{uX^+ - uX^-}(\mathbb{1}_{\{X\geq 0\}} + \mathbb{1}_{\{X<0\}})\right] = \mathbb{E}\left[e^{uX^+}\right] + \mathbb{E}\left[e^{(-u)X^-}\right].$$

This is true for both, positive and negative u. Therefore the RHS of (2.2) is finite. By theorem 2.7 the corollary is proved. □

Remark 2.9. *By corollary 2.8 we follow that if* (2.1) *is true for all $u \in \mathbb{R}$, than the characteristic function of X is an entire function.*

We show the characteristic function is entire in a slightly more general setting. Consider the stochastic differential equation with smooth coefficient functions b and a,

$$\begin{aligned} dX_t &= (r - \frac{1}{2}V_t)dt + \rho\sqrt{V_t(1 - \frac{V_t}{\bar{v}})}dW_{1t} + \sqrt{(1-\rho^2)V_t + \frac{\rho^2}{\bar{v}}V_t^2}dW_{2t} \\ dV_t &= b(V_t)dt + a(V_t)dW_{1t} \end{aligned} \tag{2.3}$$

with initial conditions x_0 and $0 \leq v_0 \leq \bar{v}$ where W_{1t} and W_{2t} are independent Brownian motions, $\bar{v} > 0$, constant riskless interest rate r and $\rho \in [-1, 1]$.

Lemma 2.10. *Consider model* (2.3)*. Assume that $(V_s)_{0\leq s\leq t}$ is compactly supported and almost surely non negative such that $P(V_s \leq \bar{v}, \forall 0 \leq s \leq t) = 1$ for some $t > 0$. Then*

$$\mathbb{E}[e^{uX_t}] < \infty \quad , \forall u \in \mathbb{R}$$

and as a consequence the characteristic function of X_t is an entire function.

Proof. Since $(V_s)_{0\leq s\leq t}$ is compactly supported with $(V_s)_{0\leq s\leq t} \leq \bar{v}$ almost surely, we can conclude that the processes $\tilde{V}_t := \rho\sqrt{V_t(1-\frac{V_t}{\bar{v}})}$ and $\hat{V}_t := \sqrt{(1-\rho^2)V_t + \frac{\rho^2}{\bar{v}}V_t^2}$ are compactly supported as well with upper bounds $\tilde{A}$ and $\hat{A}$.

We want to show that X_t has finite moment generating function on the whole real line. Note that we have

$$X_t = x_0 + \int_0^t (r - \frac{1}{2}V_s)ds + \int_0^t \tilde{V}_s dW_{1s} + \int_0^t \hat{V}_s dW_{2s}.$$

Plugging in (2.2) into the mgf and using the tower property of conditional expectation we get

$$\begin{aligned}
\mathbb{E}\left[e^{uX_t}\right] &= \mathbb{E}[e^{u(x_0+\int_0^t (r-\frac{1}{2}V_s)ds+\int_0^t \tilde{V}_s dW_{1s}+\int_0^t \hat{V}_s dW_{2s})}]) \\
&= \mathbb{E}\left[\mathbb{E}[e^{u(x_0+\int_0^t (r-\frac{1}{2}V_s)ds+\int_0^t \tilde{V}_s dW_{1s}+\int_0^t \hat{V}_s dW_{2s})}]|W_{1s}, 0\leq s\leq t\right] \\
&= \mathbb{E}\left[e^{ux_0}e^{u\int_0^t (r-\frac{1}{2}V_s)ds}e^{u\int_0^t \tilde{V}_s dW_{1s}}e^{\frac{u}{2}\int_0^t \hat{V}_s^2 ds}\right]
\end{aligned}$$

Since $V_t, \tilde{V}_t$ and $\hat{V}_t$ are each compactly supported processes, we can follow that the deterministic integrals over them are bounded. Therefore it suffices to show that $\mathbb{E}[e^{u\int_0^t \tilde{V}_s dW_{1s}}]$ is finite. Since $(\tilde{V}_s)_{s\leq t}$ is bounded, so is (u is fixed) $(u\tilde{V}_s)_{s\leq t}$. Denote that bound with $B = |u|\tilde{A}$ and define

$$M_t := \int_0^t u\tilde{V}_s dW_{1s}.$$

That means we need to show $\mathbb{E}[e^{M_t}] < \infty$. Note that M_t is a local martingale. Since $\langle M\rangle_t$ is of bounded variation we have that $M_t - \frac{1}{2}\langle M\rangle_t$ is a semi martingale. Next define

$$Z_t := \exp(M_t - \frac{1}{2}\langle M\rangle_t).$$

By Ito's formula we get

$$dZ_t = uZ_t\tilde{V}_t dW_{1t} \quad , Z_0 = 1.$$

Once again we can conclude that Z_t is a local martingale. Now notice that since $\frac{1}{2}\langle M\rangle_t = \frac{1}{2}u^2\int_0^t \tilde{V}_s^2 dt \leq \frac{1}{2}|u|B^2 t$ we can write

$$e^{M_t} \leq e^{M_t - \frac{1}{2}\langle M\rangle_t + \frac{1}{2}|u|B^2 t} = e^{\frac{1}{2}|u|B^2 t}e^{M_t - \frac{1}{2}\langle M\rangle_t}$$

and it suffices to show that

$$\mathbb{E}\left[e^{M_t - \frac{1}{2}\langle M\rangle_t}\right] = \mathbb{E}\left[Z_t\right] < \infty.$$

Let T_n be a stopping time making $Z_{t\wedge T_n}$ a true martingale. Since $Z_0 = 1$ we conclude that $\mathbb{E}[Z_{t\wedge T_n}] = 1$ for all n. Since T_n goes to infinity with probability one we get by Fatou's lemma:

$$\mathbb{E}[Z_t] = \mathbb{E}\left[\lim_{n\to\infty} Z_{t\wedge T_n}\right] \leq \liminf_{n\to\infty} \mathbb{E}[Z_{t\wedge T_n}] = 1.$$

Hence we have deduced that X_t has finite mgf over the whole real line. By corollary 2.8 we can conclude that the characteristic function of X_t is an entire function. □

Remark 2.11. *Since the radius of convergence of an analytic function around a point $z \in \mathbb{C}$ is given by the distance to the closest singularity, we have that the power series representation around zero of the characteristic function of X_t converges globally hence uniformly on any compact set.*

Part II

Applications to finance

Chapter 3

Density Approximation for Polynomial Preserving Processes

When interested in pricing European type options on assets at maturity T, these prices can be computed performing quadrature (numerical integration) if the density at time T under a risk neutral measure is known. Typically however this is not the case and hence approximation methods are required. In case of affine processes, the characteristic function can be computed by solving a potentially high dimensional Riccati equation, see e.g. [9]. Therefore Fourier inversion methods can be applied to compute prices. Fourier inversion however can be a very challenging task. Therefore alternative methods are desirable.

In this chapter we want to present a methodology presented by Filipovic et al. in [13]. The idea is to expand the unknown density g by projecting on to orthogonal polynomials with respect to an weighted $\mathcal{L}_w$ space where w is a weight function satisfying a certain assumption, see further below. In [13], the authors make use of the fact that one can relate properties of density functions with properties of the characteristic function. As mentioned, the characteristic function for affine processes is a solution to a Riccati equation. For polynomial preserving processes this tool is not available anymore, hence we will not present the results specifically regarding affine processes.

Density approximation methods are widely used in a wide field of applications and there are several different approaches to this. In case of the auxiliary density w above being the standard normal density, the method presented in [13] is in fact the Gram-Charlier expansion of the density g. Another method is the saddle point approximation, which does not rely on moments directly. Instead cumulants that are closely related to moments are used to describe the density as a linear combination of normal densities and its derivatives. An overview and discussion of available approximation methods can e.g. be found in [24].

3.1. Basic methodology

We start by presenting the theoretical framework and results as presented in [13]. We adapt notation and restrict to the results relevant for polynomial preserving diffusions. Throughout this chapter let g denote a probability density function on $\mathbb{R}^d$. Using multi-index notation for the order $\alpha \in \mathbb{N}_0^d$, we define the power moments of g by

$$\mu_\alpha := \int_{\mathbb{R}^d} \xi^\alpha g(\xi)\, d\xi,$$

and we assume these moments are finite and known in closed form for all orders , i.e. in our context by means of computing a matrix exponential. As is standard for multi-index notation we define the length for an order the same as the degree for multivariate monomials with exponent α, i.e.

$$|\alpha| := \sum_{i=1}^{d} \alpha_i.$$

We want to approximate the density g with a truncated generalized Fourier expansion in a weighted $\mathcal{L}^2$ space. This weight is an auxiliary probability density on $\mathbb{R}^d$ denoted by w. Define $\mathcal{L}^2_w$ to be the space of (equivalence classes of) $\mathcal{L}^2_w$ integrable functions f on $\mathbb{R}^d$ with respect to the norm

$$||f||_{\mathcal{L}^2_w} = \left(\int_{\mathbb{R}^d} |f(\xi)|^2 w(\xi)\, d\xi \right)^{\frac{1}{2}}.$$

This norm is induced by the inner product

$$\langle f, h \rangle_{\mathcal{L}^2_w} := \int_{\mathbb{R}^d} f(\xi) h(\xi) w(\xi)\, d\xi.$$

Our goal is to approximate the typically unknown density g by expanding the likelihood ratio g/w in terms of orthogonal polynomials associated to the weight w. As done in [13], we start by making the following assumptions. For sufficient conditions we refer the reader to the original paper [13] itself.

Assumption 1. *There exists a set of polynomials* $\{H_\alpha | \alpha \in \mathbb{N}_0^d\}$ *with* $deg(H_\alpha) = |\alpha|$ *that is orthogonal and normalized in* $\mathcal{L}^2_w$*, i.e.*

$$\langle H_\alpha, H_{\alpha'} \rangle_{\mathcal{L}^2_w} = \delta_{\alpha,\alpha'} = \mathbb{1}_{\{\alpha=\alpha'\}}.$$

In particular, this implies $H_0 = 1$*, where* 0 *is here to be understood as the* d*-tupel of zeros.*

Assumption 2. *The likelihood ratio is in $\mathcal{L}^2_w$, i.e.*

$$||\frac{g}{w}||^2_{\mathcal{L}^2_w} = \int_{\mathbb{R}^d} \frac{g^2(\xi)}{w(\xi)}\, d\xi < \infty.$$

As a consequence, the coefficients

$$c_\alpha := \langle \frac{g}{w}, H_\alpha \rangle_{\mathcal{L}^2_w} = \int_{\mathbb{R}^d} H_\alpha(\xi) g(\xi)\, d\xi$$

are well defined. They can be computed as a linear combination of the known moments of g where the coefficients are the coefficients of H_α in the monomial base. Alternatively, for $d = 1$, they can be computed by a triangle recursion scheme as presented in Appendix Listing A.1.

Our final goal is to approximate the integral

$$\int_{\mathbb{R}^d} f(\xi) g(\xi)\, d\xi,$$

where f is usually the payoff at maturity of an European type (basket) option and g the density at maturity of the underlying assets. Using the fact that $\{H_\alpha | \alpha \in \mathbb{N}^d_0\}$ forms an orthonormal base, we can re-wright the above equation and get

$$\int_{\mathbb{R}^d} f(\xi) \frac{g(\xi)}{w(\xi)} w(\xi)\, d\xi = \sum_{|\alpha|=0}^{\infty} \langle f, H_\alpha \rangle_{\mathcal{L}^2_w}\ \langle \frac{g}{w}, H_\alpha \rangle_{\mathcal{L}^2_w}, \tag{3.1}$$

where in the following we assume integrability of f such that the above statements are well defined.

We can now approximate the above integral by truncating the Fourier series. Introducing the functions

$$g^J(x) := w(x) \left(\sum_{|\alpha|=0}^{J} c_\alpha H_\alpha(x) \right),$$

for an order $J \geq 0$, this approximation is of the form

$$\int_{\mathbb{R}^d} f(\xi) g^J(\xi)\, d\xi = \sum_{|\alpha|=0}^{J} c_\alpha \langle f, H_\alpha \rangle_{\mathcal{L}^2_w}.$$

One way to compute the above approximations is to explicitly construct the functions g^J and then compute the left hand side by quadrature. The other one would

be to compute the factors of the right hand side (one as the linear combination of known moments and one by in general quadrature) and summing up the products.

Noticing that $H_0 = 1$ together with orthogonality (and normality) of $\{H_\alpha | \alpha \in \mathbb{N}_0^d\}$ and standard $\mathcal{L}_w^2$ theory, see e.g. [14], yields the following Theorem.

Theorem 3.1 (Theorem 2.1 in [13])**.** *Under the two assumptions above, for all $J \geq 0$, the functions g^J satisfy*

$$\int_{\mathbb{R}} g^J(\xi)\, d\xi = 1 \tag{3.2}$$

$$\begin{gathered} \lim_{J\to\infty} ||\frac{g^J}{w} - \frac{g}{w}||^2_{\mathcal{L}_w^2} = \\ \lim_{J\to\infty} \int_{\mathbb{R}^d} |g^J(\xi) - g(\xi)|^2 \frac{d\xi}{w(\xi)} = 0 \end{gathered} \tag{3.3}$$

Remark 3.2. *In light of* (3.2) *one might think of the functions g^J as density functions. However since they can take negative values we call them, as done in [13], pseudo densities.*

The next Lemma shows a property of the coefficients c_α in relation to the moments of w and g. This will, as we will see in the concrete numerical example for the power Heston model further down, be useful when implementing the density approximation. The idea is to choose the auxiliary density as close as possible to g, i.e one would like the likelihood ration to be close to one. This should be the case when as many of the first c_α as possible are zero. For the proof we refer to [13].

Lemma 3.3 (Moment Matching Principle, Lemma 2.2 in [13])**.** *Denote by λ_α the power moment of order $\alpha \in \mathbb{N}_0^d$ for the auxiliary density w. Suppose that for $n \geq 1$ we have $\mu_\alpha = \lambda_\alpha$ for all $|\alpha| \leq n$. Then $c_\alpha = 0$ for $1 \leq \alpha \leq n$.*

From the definition of the pseudo densities g^J, one can see that the polynomial part can be seen as the projection of the true density on the subspace of all polynomials up to degree J with respect to the $\mathcal{L}_w^2$ inner product. This manifests in the following Lemma.

Lemma 3.4. *For all $J \geq 0$, the pseudo densities satisfy*

$$\sum_{|\alpha|=0}^{J} c_\alpha \langle \xi^\gamma, H_\alpha \rangle_{\mathcal{L}_w^2} = \int_{\mathbb{R}^d} \xi^\gamma g^J(\xi)\, d\xi = \int_{\mathbb{R}^d} \xi^\gamma g(\xi)\, d\xi,$$

for all power moments $\gamma \in \mathbb{N}_0^d$ with $|\gamma| \leq J$.

Proof. The first equality has already been stated and is just a re-writing of the definition of c_α and the fact that $\{H_\alpha | \alpha \in \mathbb{N}_0^d\}$ is a orthonormal basis of $\mathcal{L}_w^2$. We proof the second equality. It suffices to show

$$\langle \xi^\gamma, H_\alpha \rangle_{\mathcal{L}_w^2} = 0$$

for all γ with $|\gamma| < |\alpha|$. Since ξ^γ is a polynomial of degree $|\gamma|$, there exist coefficients $a_\lambda \in \mathbb{R}$ for $\lambda \in \mathbb{N}_0^d$ with $\xi^\gamma = \sum_{|\lambda|=0}^{|\gamma|} a_\lambda H_\lambda(\xi)$. The above statement follows by orthogonality.

□

Filipovic et al. continue by presenting sufficient conditions for the two assumptions above. For our following numerical examples, only one is relevant. For the other conditions, we therefore refer to [13].

Lemma 3.5 (Lemma 3.1 in [13])**.** *Suppose that the density function w has finite exponential moment*

$$\int_{\mathbb{R}^d} \exp(\epsilon_0 ||\xi||) w(\xi)\, d\xi < \infty$$

for some $\epsilon_0 > 0$. Then the set of polynomials is dense in $\mathcal{L}_w^2$ with respect to $||\cdot||_{\mathcal{L}_w^2}$.

Remark 3.6. *The above Lemma ensures moment determinacy of the measure $w(x)dx$ and can be extended to general non negative Borel measures, see [8] Theorem 3.2.17 and Theorem 3.2.18 and the discussions made in between. Dunkl and Xu further state that in one dimension it holds that if a measure μ for which all moments exist is moment determinant, then the set of polynomials is dense in $\mathcal{L}_\mu^2$. For dimensions $d \geq 2$ this is not true anymore as shown by Berg and Thill in [3].*

In the next section we will discuss a concrete example, namely the log price in the power Heston model, and present results for a numerical implementation. As it turns out the auxiliary densities we will use are certain normal densities. The above lemma ensures the density of polynomials for this choice as normal densities have analytic characteristic function.

3.2. Concrete example: The Power Heston Model

We continue by applying the above described method to approximate the transition density of the log price in the power Heston model at maturity T. As a first step we present results on the density as shown in [11]. In particular we will present auxiliary densities for which the Assumptions 1 and 2 are satisfied. Following that we present a numerical implementation and discuss results.

3.2.1. Log Price Density in the Power Heston Model

In the following we denote by g_t the true transition density at time t of the log price process X_t in the power Heston model. At maturity T, we sometimes write $g_T = g$. The volatility process is denoted by V_t. Further let (W_{1t}, W_{2t}) be a two dimensional standard Brownian motion. The dynamics of this model are given in (2.5) in the first part chapter 2. We present the findings and proofs in [11]. Note that since we are only interested in the transition density of the log price, we have a one dimensional setting, i.e $d = 1$ and $\alpha \in \mathbb{N}_0$.

Lemma 3.7. *Let $T > 0$, and assume that*

$$\Sigma_T^2 := \int_0^T \left((1-\rho^2)V_t + \rho^2 V_t^2/\bar{v}\right) dt. \tag{3.4}$$

Further, define

$$\mathbb{E}\left[\frac{1}{\sqrt{\Sigma_T^2}}\right] < \infty. \tag{3.5}$$

Then the distribution of X_T admits a continuous density $g(x)$ on $\mathbb{R}$ that satisfies, for any $\epsilon < 1/(2\bar{v}T)$,

$$\int_{\mathbb{R}} e^{\epsilon x^2} g(x)\, dx < \infty. \tag{3.6}$$

Proof. Rewriting the stochastic differential as an integral equation for both processes and grouping, we infer that the log price X_T can be written as

$$X_T = M_T + \int_0^T \sqrt{(1-\rho^2)V_t + \rho^2 V_t^2/\bar{v}}\, dW_{2t}$$

where we denote

$$M_T = X_0 + \int_0^T (r - V_t/2)\, dt + \frac{\rho}{\sigma}\left(V_T - V_0 - \int_0^T \kappa\left(\theta - V_t\right) dt\right).$$

Conditioning on the process $\{V_t,\, t \in [0,T]\}$, it follows that the random variable X_T is Gaussian with mean M_T and variance Σ_T^2, given in (3.4). Its density is

$$G_T(x) = \frac{1}{\sqrt{2\pi\Sigma_T^2}} \exp\left[-\frac{(x-M_T)^2}{2\Sigma_T^2}\right].$$

Assumption (3.5) together with $G_T(x) \le 1/\sqrt{2\pi\Sigma_T^2}$ for all x, and dominated convergence, implies that the function

$$g_T(x) = \mathbb{E}\left[G_T(x)\right]$$

is continuous in $x \in \mathbb{R}$. Fubini's theorem implies that, for any bounded measurable function $f(x)$,

$$\mathbb{E}\left[f(X_T)\right] = \mathbb{E}\left[\int_{\mathbb{R}} f(x)\, G_T(x)\, dx\right] = \int_{\mathbb{R}} f(x)\, g_T(x)\, dx.$$

Hence the distribution of X_T admits the continuous density $g(x) = g_T(x)$.

Let $\epsilon < 1/(2\bar{v}T)$, and define $\Delta = 1/(2\bar{v}T) - \epsilon > 0$. The mean and variance of $G_T(x)$ admit the uniform bounds

$$|M_T| \le K, \quad \Sigma_T^2 \le \bar{v}T$$

where $K = K(T)$ is a finite constant. Completing the square implies

$$\begin{aligned} e^{\epsilon x^2} G_T(x) &\le \frac{1}{\sqrt{2\pi\Sigma_T^2}} \exp\left[\epsilon x^2 - \frac{(x - M_T)^2}{2\bar{v}T}\right] \\ &= \frac{1}{\sqrt{2\pi\Sigma_T^2}} \exp\left[-\left(\sqrt{\Delta}\, x - \frac{M_T}{2\bar{v}T\sqrt{\Delta}}\right)^2 + \frac{\epsilon M_T^2}{2\bar{v}T\Delta}\right] \\ &\le \frac{1}{\sqrt{2\pi\Sigma_T^2}} \exp\left[-\frac{\left(x - \frac{M_T}{2\bar{v}T\sqrt{\Delta}}\right)^2}{2\frac{1}{2\Delta}}\right] \exp\left[\frac{\epsilon K^2}{2\bar{v}T\Delta}\right] \end{aligned}$$

Integration then gives

$$\int_{\mathbb{R}} e^{\epsilon x^2} G_T(x)\, dx \le \frac{1}{\sqrt{2\Delta\Sigma_T^2}} \exp\left(\frac{\epsilon K^2}{2\bar{v}T\Delta}\right).$$

Hence (3.6) follows by taking expectation on both sides, changing the order of integration of the nonnegative random variable, and using (3.5). □

Remark 3.8. *We have already seen that the power Heston model has an entire characteristic function for $\bar{v} < \infty$. This was shown using methods of stochastic calculus. Under the assumptions of the Lemma above, the same follows immediately as a consequence of the density function since the lemma implies*

$$\hat{g}(z) = \int_{\mathbb{R}} e^{zx} g(x)\, dx < \infty, \quad \forall z \in \mathbb{R}.$$

As in Theorem 2.7 and Corollary 2.8, this is sufficient (and necessary) for $\hat{g}$ to be entire.

Corollary 3.9. *Under the assumption of Lemma 3.7, for any $\epsilon < 1/(4\bar{v}T)$,*

$$\int_{\mathbb{R}} e^{\epsilon x^2} g(x)^2\, dx < \infty. \tag{3.7}$$

As a consequence, the likelihood ratio function $g(x)/w(x)$ lies in the weighted Lebesgue space $\mathcal{L}^2_w = \left\{ f \mid \int_{\mathbb{R}} f(x)^2 \, w(x) dx < \infty \right\}$ for any Gaussian density $w(x)$ with variance

$$\sigma_w^2 > \bar{v}T/2. \tag{3.8}$$

Proof. Since $e^{\epsilon x^2/2} g(x)$ is a continuous and integrable function, it is uniformly bounded in $x \in \mathbb{R}$. Now (3.7) follows from (3.6). The last claim follows from it since

$$\left\| \frac{g}{w} \right\|^2_{\mathcal{L}^2_w} = \int_{\mathbb{R}} \frac{g(x)^2}{w(x)} \, dx < \infty,$$

for $1/(2\sigma_w^2) < 1/(4\bar{v}T)$. □

Due to Corollary 3.9, we have a family of auxiliary variances, for which Assumption 2 is satisfied. Since the family of orthogonal polynomials associated to the normal density are the generalized Hermit polynomials, they are our choice for the orthonormal system of polynomials in Assumption 1. For a collection of theoretical and practical results on Hermite polynomials, see the Appendix.

It is yet unclear which normal density to choose from the set of admissible ones. Our next step is to discuss this issue.

3.2.2. Choice of the auxiliary variance

In light of Lemma 3.9 we fix the auxiliary density $w(x)$ to be the density function of the normal distribution, i.e.

$$w(x) = \phi^{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{\sigma^2} \right].$$

The moment matching principle (Lemma 3.3) suggests that we could expect optimal results for choosing $\phi^{\mu,\sigma}$ to match the first two moments of the true density g. We want to investigate this. We begin by fixing a parametrization for which the variance of the true density is not in violation to condition (3.8). This parametrization is given in Table 3.1. It makes sense to assume that the auxiliary mean should be matched to catch most of the 'randomness' described by g. Appendix Figure 1 suggest this to be true. Therefore we want to focus on the choice of the variance.

We start by constructing pseudo densities of the same order J under the same parametrization. For these constructions we use different auxiliary densities, with same mean (the true one) but different variances. Figure 3.1 suggests that the moment matching auxiliary density performs best. Our numerical examples have shown that the order J of the approximation influences the tolerance of how close

lnS_0	v_0	$\bar{v}$	b	β	σ	ρ	K	T
0	0.04	0.08	0.04	-1	0.2	-0.5	1	1

Table 3.1: Parameters in Heston and power Heston model for numerical examples. In the power Heston model, the mean is -0.02 and the variance is 0.0407.

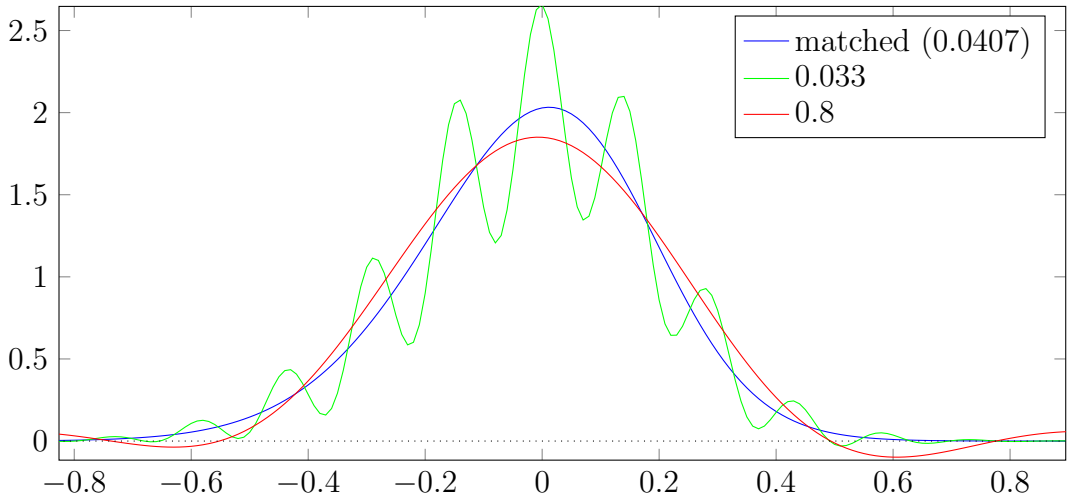


Figure 3.1: Density plot for order $J = 60$. In all cases the first moment of the auxiliary density is matched. Plots for different variances of the auxiliary density. Parametrization of the process is given in Table 3.1. Moment matching does not violate condition (3.8).

one needs to be to the true variance before observing unreasonable behavior such as high oscillations or negative density values. The bigger the order J, the smaller the tolerance.

Figure 3.1 shows that the moment matching case does not show odd behavior such as unreasonable oscillations or negative density values, while the other pseudo densities show these effects even though we already are considering the order $J = 60$. As we will see later, in cases where the true variance would be in violation to (3.8), augmenting the auxiliary variance can help improve performance, however under the restriction that one limits the approximation to smaller orders. This makes sense in light of the observation that the tolerance of how far one is from the true variance decreases as J increases.

In conclusion the auxiliary density should match the first two moments of g whenever possible, i.e. without violating (3.8). In the sequel we will always use moment matched auxiliary densities unless explicitly stated otherwise. Further down we will make further investigations in the role of (3.8). We will also make the observation that the parameter σ if big, even if (3.8) is satisfied in the moment matching case, can cause the approximation to show unexpected behavior.

3.2.3. The Sequence of Pseudo Densities

In the sequel we will use the parametrization given in Table 3.1 for the numerical examples in this subsection. Further the moment matching case will be applied. This has the benefit of not violating (3.8) so we can exclude errors that might come from such a violation.

Since round off errors and therefore other numerical issues will always play a role where computers are used due to the fact that computers can only represent a finite amount of numbers, it is necessary to check whether there is an order J from which the errors become dominant in the sense that higher order approximations start to become more inaccurate then previous orders. This subsection is devoted to make a more systematic analysis regarding this question by not just simply looking at plotted results.

Unlike true affine diffusions like the Heston model, the characteristic function of polynomial preserving diffusions can not be computed by (numerically) solving a system of Riccati equations (see e.g. (make collection for affine processes)) and then compare option prices for the pseudo densities against those one can compute by Fourier inversion techniques. It is also not possible to simply let $\bar{v}$ go to infinity and compare option prices to the Heston model, since this would lead to a violation of (3.8).

Applying moment based approximations for the characteristic function such as Edgeworth type expansions are only approximations in nature and it is questionable whether they can reach the accuracy of density expansions themselves.

We take another approach. Denote by w a normal density satisfying (3.8). Theorem 3.1 implies

$$\int_{\mathbb{R}} |g^{\alpha}(\xi) - g^{\alpha-1}(\xi)|^2 \frac{d\xi}{w(\xi)} = c_{\alpha}^2 \to 0, \text{ as } \alpha \to \infty. \tag{3.9}$$

It would be wrong to check convergence of the sequence $g^{\alpha}, \alpha < J$ to the pseudo density g^{J} for J big since (3.9) implies that higher orders inherit the errors of previous orders.

Now recall that c_{α} is a linear combination of $\mu_g(0), \ldots, \mu_g(\alpha)$, the moments of the log price distribution at maturity. Since c_{α} converges to zero for $\alpha \to \infty$, one can expect problems when $|\mu_g(\alpha)|$ explodes. Round off errors become an issue when one takes linear combinations of numbers whose absolute values are big relative to the final quantity one computes. Therefore we want to analyze the coefficients c_{α}, and since they are the only non closed form quantities, it makes sense to assume that they will be the cause for numerical issues.

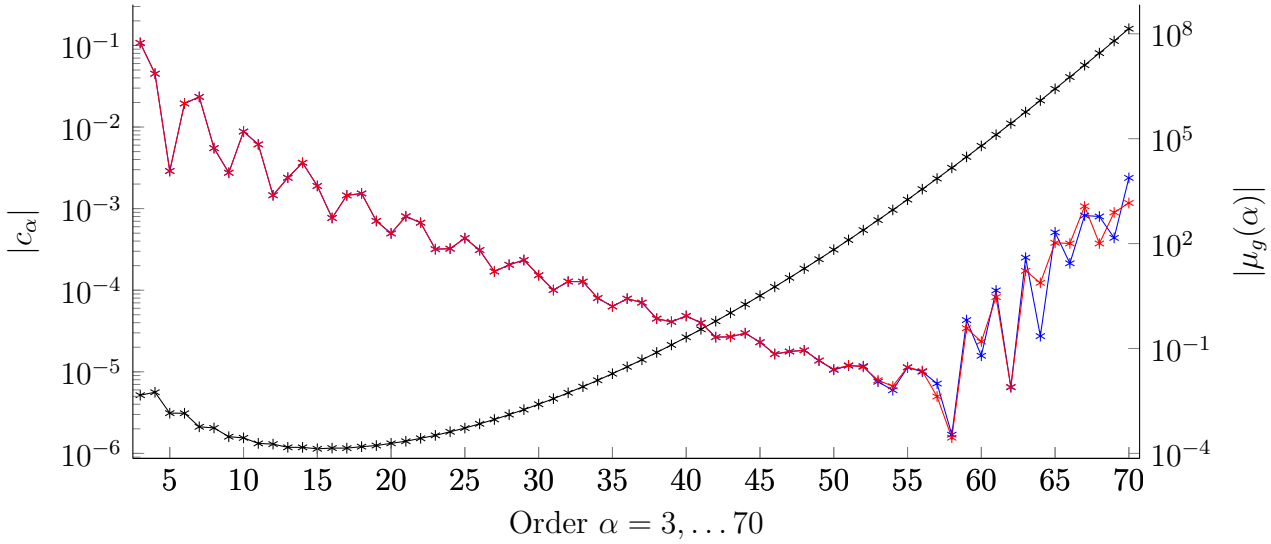


Figure 3.2: Plot of absolute values of the Hermite moments. Blue line is computed using the triangle recursion, red by Orthogonal polynomial construction. Black line are absolute values of moments, scale on the right.

We fix the order $J = 70$ and compute the coefficients c_α up to this order. We compute them once by explicitly constructing the polynomials H_k and once using the introduced triangle scheme introduced in the Appendix. While both methods produce very similar values, we expect (even if not definitive prove) accurate results. When the values for c_α begin to differ, at least one of the methods becomes inaccurate. We further want to check whether the point where this happens is the same as when moments become big. We present results in Figure 3.2.

Indeed, the range of orders where the two methods compute different results is about the same as when the absolute values of the coefficients c_α begin to grow (something we would not expect due to (3.9). This range, being around $\alpha = 57$, coincides with moments of magnitude 10^5, while to coefficients that are being computed at that order are in the range of 10^-5. Hence, round off errors could explain the divergence of the two different computation methods and why the c_α begin to grow.

In order to avoid these errors we therefore continue by disregarding all c_α with $\alpha > 55$ and use the order $J = 55$ as reference.

A good sign for $J = 55$ being an order sufficiently large to achieve a high accuracy would be if option prices as a function of J would not change much compared to the price for $J = 55$ for a longer sequence of prices. Let f be the payoff of an European type option such that $f \leq K$ for some constant $K > 0$. It holds

$$\left| \int_{\mathbb{R}} f(e^\xi)(g(\xi) - g^J(\xi))\, d\xi \right| \leq K \int_{\mathbb{R}} |g(\xi) - g^J(\xi)|\, d\xi. \tag{3.10}$$

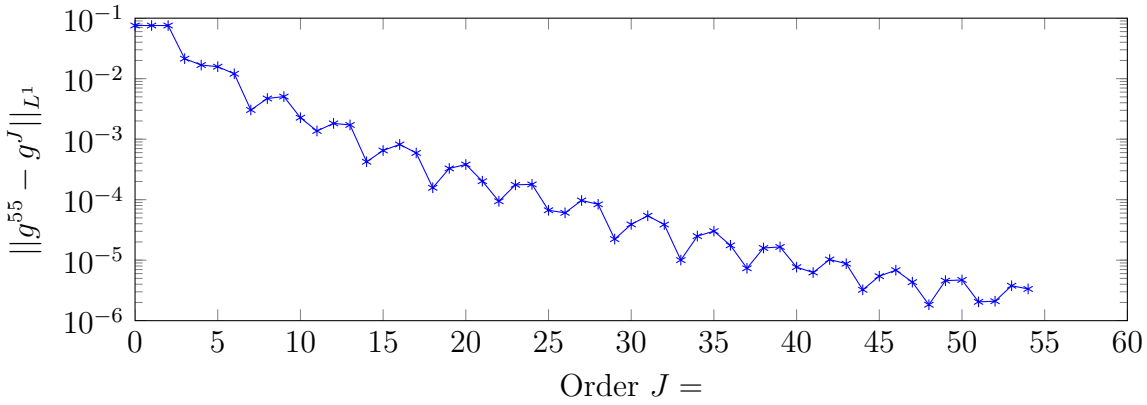


Figure 3.3: Plot of L^1 norms of $g^{55} - g^J$ for $J = 0, .., 54$. Parametrization is given in Table 3.1.

So the right choice for error measurement for pricing would be the L^1 norm, and since g and by construction the pseudo densities are L^1 functions, this can be done without problems. We therefore take g^{55} as replacement for the true density in (3.10) and let J go from 0 to 54. Figure 3.3 shows the numerical results. Note that already from g^J with $40 \leq J \leq 54$ we see that the difference to g^{55} is around 10^{-5}.

Before continuing to use option prices as a accuracy measure for the computed pseudo densities, we want to discuss the impact of the parameters on the quality of the pseudo densities. In particular we will test the importance of (3.8) and the special role the parameter σ seems to play.

3.2.4. Parameters and calibration in the Power Heston Model

A well known result in financial mathematics is that under certain market assumptions the no arbitrage price of financial derivatives can be computed by taking expectation under a risk neutral measure, see e.g. [28]. However, when there is no unique risk neutral measure, there are several no arbitrage prices and the one compatible with observed market prices is chosen by means of calibration. In the example of the Power Heston model this means that one wants to find the parametrization which can explain best the observed prices of liquid financial assets[1] at the time.

From a mathematical point of view this calibration is an (in general constrained nonlinear) optimization procedure over the set of admissible parameters. The target function in each optimization step is computed by computing asset prices under the current set of parameters. Since we want to compute these prices by approximating

[1]The liquidity is to ensure market efficiency and hence correct prices in the market itself.

the density of the log price process at maturity it is necessary to know whether the current set of parameters will produce reliable approximations.

Of particular interest will be the question whether condition (3.8) is only of theoretical nature or whether it is crucial that this condition is satisfied for the density approximations to perform well, since they pose a relevant constrain on the set of admissible parameters. In addition it would mean that if this condition is crucial, then one can not simply approximate the Heston Model by letting $\bar{v}$ go to infinity, causing a violation of condition (3.8) for any given choice of the other parameters (assuming of course $T > 0$).

3.2.5. The Condition of Minimal Variance for the Auxiliary Density

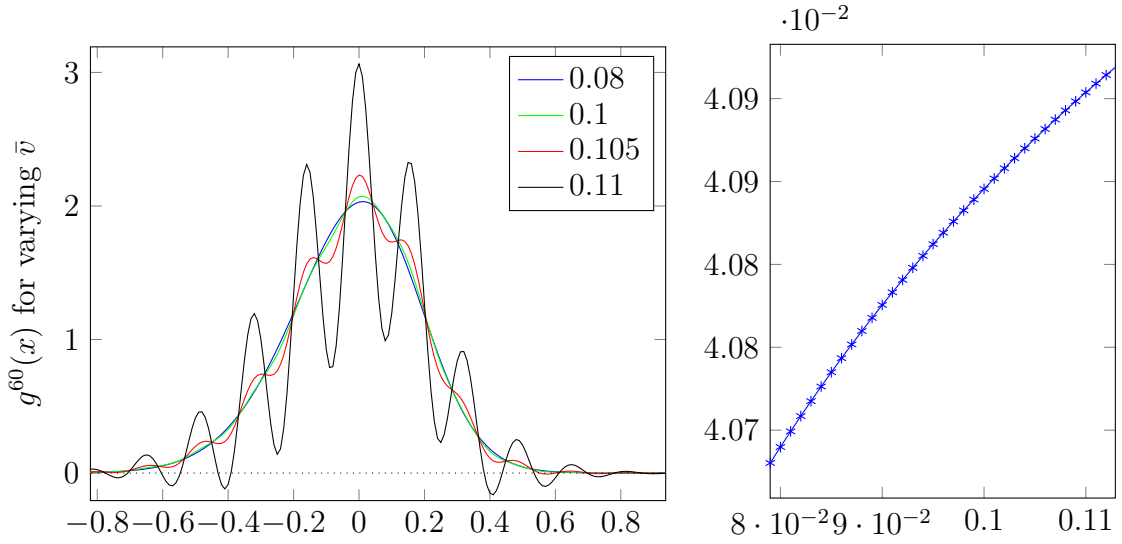
We begin by investigating condition (3.8). We start with the same choice of parameters as in Figure 3.1, and we always match the first two moments of the auxiliary density with those of the true transition density g. Recall that in this parametrization the true variance (and hence the variance of the auxiliary density) satisfies (3.8). We fix the approximation order $J = 60$ and we compare the plots for various $\bar{v}$. Only for $\bar{v} = 0.08$ the true variance is in accordance with (3.8), for the other ones the violation is monotonically increasing. Since we have $T = 1$, condition (3.8) reads $\bar{v}/2$. Results are shown in Figure 3.4.

As one can see, there seems to be some tolerance to condition (3.8) before the pseudo densities show odd behavior. However our numerical test showed that this tolerance, much similar as to the moment matching principle, decreases as J increases. In conclusion one can say that condition (3.8) is a reasonable constraint on the parameters when applying the presented transition density approximation method.

3.2.6. Parameters and the condition

After having investigated the impact of condition (3.8), we would like to check if there is a need to restrict the parameters to not have an ill conditioned problem. Our test however showed that the same volatility related parametrization can be altered without affecting accuracy, with one exception.

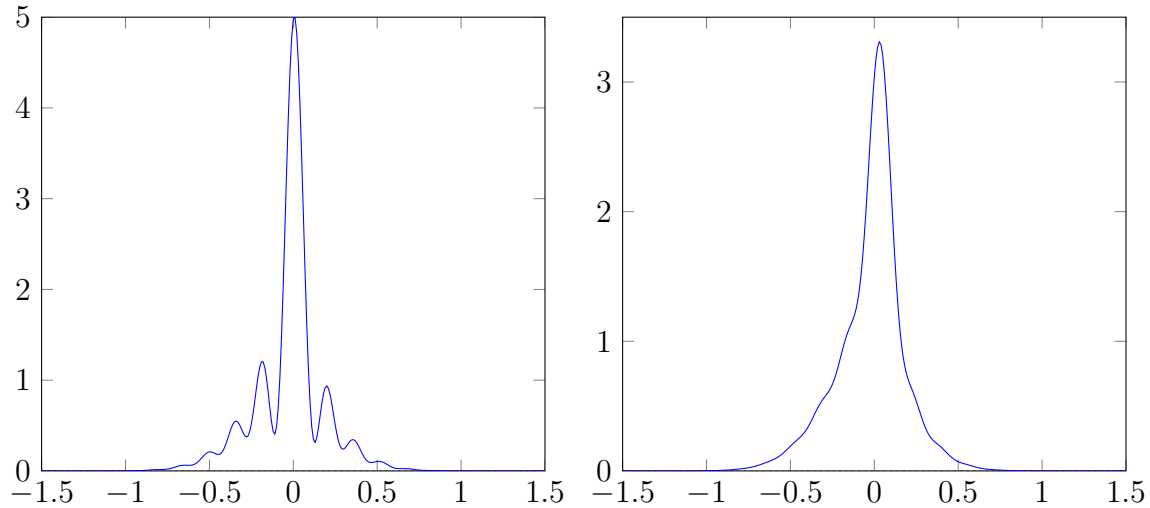
The volatility of the volatility process σ seems to play a special role, see Figure 3.5. One possible explanation could be that 3.5 in Lemma 3.4 is violated. However, since to this point we do not have sufficient conditions for this assumption, we can not check if this really is the case. We further observed, using a different parametrization, that the choice of σ in order to show unexpected behavior, depends on the other parameters. However in all cases, we got similar results as in Figure 3.5.



(a) Density plots for different $\bar{v}$.

(b) $\sigma_g^2 = \sigma_w^2$ as function of $\bar{v}$.

Figure 3.4: Plots to investigate the importance of the condition for minimal auxiliary variance. Only $\bar{v} = 0.08$ does not violate the condition. Density plots are in a) and the true variance as a function of $\bar{v}$ is in b). Auxiliary densities always match the first two moments of the true density.



(a) Density plot for $\sigma = 5$, $J = 60$.

(b) density plot for $\sigma = 1$, $J = 60$.

Figure 3.5: Density plots in a) and b) for parameters given in Table 3.1, only σ was changed as indicated. Moment matching for the auxiliary density is applied.

3.3. Computing option prices in the power Heston model

We proceed by presenting approximations to option prices in the power Heston model using the density approximation approach. We price European call options at the money with maturity $T = 1$. Unless stated otherwise, we use the parametrization given in Table 3.1.

As before, denote by g the true density of the log price at maturity. If f is the payoff function of an European type option with the terminal stock price as argument, the price of the option under the risk neutral measure with density g is given by

$$\Pi = \int_{\mathbb{R}} f(e^x) g(x)\, dx.$$

Given g, Π can be computed using numerical integration (in more than one dimension (cubature) this task itself might however already proof challenging). Since g is not known in analytical closed form, approximations are necessary as presented above.

A remarkable insight in [11] is that for European call options, numerical integration can be avoided by introducing a recurrence scheme to compute the Fourier coefficients $\langle f, H_\alpha \rangle_{\mathcal{L}^2_w}$ in (3.1). We present this in the following.

3.3.1. Avoiding Numerical Integration

We want to compute the Fourier coefficients

$$f_\alpha := \langle f, H_\alpha \rangle_{\mathcal{L}^2_{\phi^{\mu,\sigma}}} = \int_{\mathbb{R}} f(\xi) H_\alpha(\xi) \phi^{\mu,\sigma}(\xi)\, d\xi \tag{3.11}$$

up to a given order J, i.e. for all $\alpha \leq J$. Assuming f to be piecewise affine linear, as it is for call options, a naive approach would be to see the generalized and normalized Hermite polynomials H_α as a linear combination of monomials. In the sequel we only consider European call options. Therefore, defining k to be the log strike, it would suffice to compute expressions of the form

$$\int_k^\infty \xi^\alpha \phi^{\mu,\sigma}(\xi)\, d\xi, \tag{3.12}$$

such that f_α could be computed as a linear combination of those. Using a change of variable $\xi = \sigma x + \mu$, the expression in (3.12) can once again be computed by computing a linear combination of the form

$$\int_{\frac{k-\mu}{\sigma}}^\infty \xi^\alpha \phi(\xi)\, d\xi,$$

with ϕ being the $\mathcal{N}(0,1)$ density function. Integrating by part and using $\frac{d}{d\xi}\phi(\xi) = -\xi\phi(\xi)$, one can derive the following recurrence,

$$\int_{\frac{k-\mu}{\sigma}}^{\infty} \xi^{\alpha}\phi(\xi)\,d\xi = \left(\frac{k-\mu}{\sigma}\right)^{\alpha-1}\phi\left(\frac{k-\mu}{\sigma}\right) + \int_{\frac{k-\mu}{\sigma}}^{\infty} \xi^{\alpha-2}\phi(\xi)\,d\xi. \tag{3.13}$$

The two initial values can be expresses in closed form using the cumulative normal distribution function. However this recurrence comes with two main problems. The first is that the computation of f_α would require several linear combinations of expressions as in (3.13) for different $\alpha' \leq \alpha$. This can become computationally expansive and difficult to implement. The second is that for higher orders, typical values in (3.13) are very big compared to f_α. Therefore numerical instability becomes an issue.

Filipovic and Pulido avoid these issues in [11] by not deriving a recurrence in the monomial basis, but using the known 3 term recurrence for the generalized and normalized Hermite polynomials to derive a recurrence that computes the coefficients f_α directly. In what follows, we present their results.

We deviate slightly from previous notation. Denote by $H_\alpha^{\mu,\sigma}$ the generalized and normalized Hermite polynomial of degree α associated to $\phi^{\mu,\sigma}$. For simplicity of notation write $H_\alpha^{0,1} = H_\alpha$. For all $n, m \in \mathbb{N}_0$ (see the Appendix) it holds

$$\int_{\mathbb{R}} H_n^{\mu,\sigma}(x)H_m^{\mu,\sigma}(x)\phi^{\mu,\sigma}(x)\,dx = \delta_{n,m}.$$

Combining Lemma A.2 with Lemma A.4, it follows

$$H_n(x) = \frac{1}{\sqrt{n}}\left(xH_{n-1}(x) - H'_{n-1}(x)\right). \tag{3.14}$$

Using (3.14) together with $\phi'(x) = -x\phi(x)$ and integrating by parts yields

$$\int_k^{\infty} H_n(x)e^{\sigma x}\phi(x)\,dx = \frac{1}{\sqrt{n}}\left(H_{n-1}(k)e^{\sigma k}\phi(k) + \sigma\int_k^{\infty} H_{n-1}(x)e^{\sigma x}\phi(x)\,dx\right).$$

Defining $I_n(k;\sigma) := \int_k^{\infty} H_n(x)e^{\sigma x}\phi(x)\,dx$, the above recurrence can be rewritten as

$$I_n(k;\sigma) = \frac{1}{\sqrt{n}}\left(H_n - 1(k)e^{\sigma k}\phi(k) + \sigma I_{n-1}(k;\sigma)\right).$$

In particular we have

$$I_0(k;\sigma) = \int_k^{\infty} e^{\sigma x}\phi(x)\,dx = e^{\frac{\sigma^2}{2}}\int_{k-\sigma}^{\infty}\phi(x)\,dx = e^{\frac{\sigma^2}{2}}\left(1 - \Phi(k-\sigma)\right),$$

where Φ is the cumulative distribution function of the standard normal distribution. For $n \geq 1$, define

$$J_n(x) := \frac{H_n - 1(x)\phi(x)}{\sqrt{n}},$$

and

$$J_0(x) := 1 - \Phi(x).$$

We will now use the above definitions to derive a recursive algorithm for the coefficients f_n in (3.11).

As always, we assume $r = 0$. For European call options with strike $K = e^k$ it holds

$$f_n = \int_k^\infty (e^x - K)\, H_n^{\mu,\sigma}(x)\phi^{\mu,\sigma}(x)\, dx.$$

Since $H^{\mu,\sigma}(x) = H_n(\frac{x-\mu}{\sigma}$, see the Appendix, making the change of variable $y = \frac{x-\mu}{\sigma}$ yields

$$f_n = \int_{\frac{k-\mu}{\sigma}}^\infty \left(e^{\mu+\sigma y} - K\right) H_n(y)\phi(y)\, dy.$$

Consequently, with the above definitions we have

$$f_n = e^\mu I_n(\frac{k-\mu}{\sigma}; \sigma) - K J_n(\frac{k-\mu}{\sigma}).$$

In conclusion, the coefficients f_n can be computed recursively without any quadrature. Since the coefficients c_n are computed without quadrature as well, prices of European call options can be approximated by truncation in (3.1) without any numerical integration. Unlike the initial approach, this recursion constitutes a better conditioned problem, since the values of f_n are, due to the normalization, not as big as before and in particular they are computed by the recursion directly.

3.3.2. Explicit construction of the pseudo density

We continue by shortly presenting the explicit construction of the pseudi density. Introduce $p(x) := \sum_k^J c_k H_k$ for some order $J \geq 0$.We approximate prices by using quadrature on

$$\Pi \approx \int_{\mathbb{R}} f(e^x)(w(x)p(x))\, dx$$

directly, but not by computing the coefficients of p and then using the Horner scheme since this would be numerically unstable. Instead realize that p is a linear combination of polynomials satisfying a three term recursion. Therefore the Clenshaw algorithm is numerically more stable and should be used instead to increase stability and therefore accuracy of the computed prices.

For the Clenshaw algorithm and a discussion, we refer to [29].

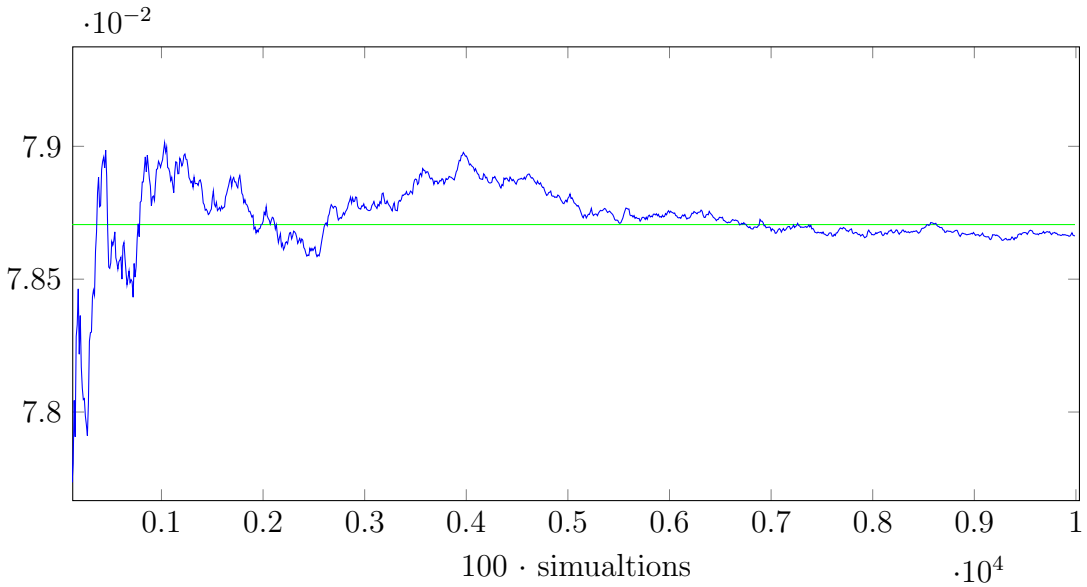


Figure 3.6: Running mean of the MC simulated option prices. Each point corresponds to 100 simulations. Overall 10e6 simulations. Euler scheme for spatial discretization. Equidistant for time with time step Δ_t =10e-5. Parameters are given in Table 3.1.

3.3.3. Computed prices

Since the above presented approaches for option pricing are only approximations in nature, it is necessary to have a benchmark to analyze the accuracy. We therefore perform a Monte Carlo simulation and use the computed Monte Carlo price as benchmark. The parameters for option (European call) and model are given in Table 3.1, and for obvious reasons the same parameters will be used for the Monte Carlo simulation and the density approximation. In particular this allows for the moment matching case for the auxiliary density.

Next, we specify the Monte Carlo simulation. We make an equidistant time discretization of the interval $[0, 1]$ with time step $\Delta_t = 10^{-5}$. For the spatial discretization we use an Euler scheme. We make overall 10^6 simulations. If the simulated variance leaves the support of the variance process V_t, we set the next value to be the value of the border of the support it crossed. We show the running mean of the simulation in Figure 3.6. For comparison we include the price using the pure recursion approximation (the pure recursive method and the explicit construction of g^J produce prices that are indistinguishable in Figure 3.6).

We continue by using the Monte Carlo prices as benchmark. More specifically, we compute prices using the pure recursive method and the explicit pseudo density construction for different orders J. These prices are then compared to the Monte Carlo result by means of relative errors and absolute differences in Black Scholes

J	Abs.$^{\text{rec}}$	Rel.$^{\text{rec}}$	Rel.$^{g^J}$	Abs.$^{g^J}$	$\Delta_\sigma^{\text{rec}}$	$\Delta_\sigma^{g^J}$
0	1.8634e-03	2.3689e-02	8.2229e-01	6.4682e-02	4.6941e-03	1.6377e-01
10	4.5534e-05	5.7886e-04	5.7599e-03	4.5307e-04	1.1469e-04	1.1413e-03
20	4.3605e-05	5.5434e-04	7.3507e-04	5.7821e-05	1.0983e-04	1.4564e-04
30	4.5128e-05	5.7370e-04	5.0685e-04	3.9869e-05	1.1367e-04	1.0043e-04
40	4.5025e-05	5.7240e-04	5.3467e-04	4.2057e-05	1.1341e-04	1.0594e-04
55	4.5034e-05	5.7251e-04	5.6214e-04	4.4218e-05	1.1344e-04	1.1138e-04

Table 3.2: Monte Carlo Simulated price as benchmark for density approximation prices for different orders J. Priced option is a European call. Relative errors are computed with respect to the Monte Carlo Price. Subscript "rec" indicates that the pure recursive method has been used, "g^J" the explicit pseudo density construction. The absolute differences between Black Scholes implied volatilities (using the Matlab function blsimpv with error tolerance 1e-12) are denoted by Δ_σ. Parameters for model and option given in Table 3.1. Moment matching for w is applied. Used discretization for Monte Carlo is equidistant in time with time step Δ_t= 10e-5. Euler scheme for spatial discretization. Overall 10e6 simulations. Differences between the pure recursive method and the explicit construction of g^J are small since the computed prices differ only by 8.2e-07.

implied volatilities. Results are presented in Table 3.2.

Table 3.2 suggests that computed prices converge to a price different from the Monte Carlo Price. This makes sense since the Monte Carlo price is by itself only an approximation. But since we do not know the price of convergence, hence the need for approximations, it would be good to know how close we already are to the true price. Similar to what we have done before, we compute the prices for the orders up until $J = 55$ and check how close the prices already are for smaller orders to the price for $J = 55$. Results are shown in Figure 3.7. With the remarks about accuracy made above we can assume the price for $J = 55$ to be without numerical errors. Hence, since the Figure suggests almost exponential converge as J grows, the price for order $J = 55$ should already be very close to the true price.

3.4. Convergence to the Heston model

We conclude this chapter by making a numerical weak convergence study of the power Heston model to the Heston model by computing European option prices. First we describe out methodology. For option and both models we fix the parameters given in Table 3.1. The only parameter that will vary in the sequel is $\bar{v}$. Further we do not apply moment matching in general anymore. We introduce a minimal auxiliary variance, and only when the true variance is above this minimal variance, we apply moment matching. In the other case we set the minimal variance as auxiliary variance. In what follows, unless stated otherwise, we always fix the order $J = 55$ due to the accuracy remarks made above.

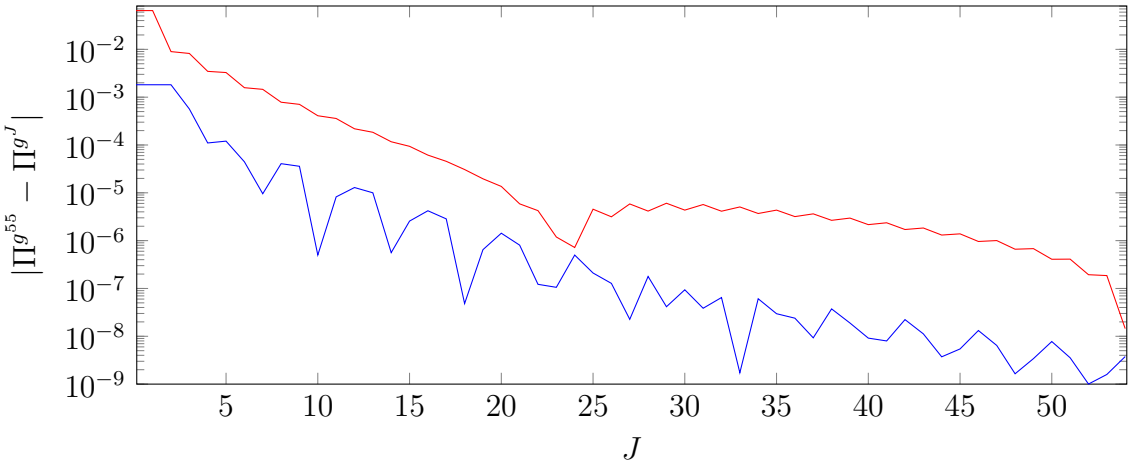


Figure 3.7: Plot shows price differences for different orders J against the order $J = 55$. Parameters given in Table 3.1. Blue line is computed using the pure recursion method, red line with the explicit pseudo density construction.

$\bar{v}_n$	0.08	1	10	100	1000
$\|\|\vec{\mu_n} - \vec{\mu}\|\|$	1.5826e-01	8.2694e-02	1.1761e-02	1.2229e-03	1.2277e-04

Table 3.3: Convergence of the first 20 moments of the power Heston model to the Heston model. Parameters are given in Table 3.1.

As we have seen, the power Heston model converges weakly to the Heston model. In particular, since both models are moment determinant, the moments of the Power Heston model converge to those of the Heston model as $\bar{v} \to \infty$. Our first step is to compute the difference of moments for various $\bar{v} = \bar{v}_n$. We start by considering the vector $\vec{\mu_n} = (\mu_0, \mu_1, ..., \mu_{20})$ consisting of the Power Heston moments using $\bar{v}_n$. Denote by $\vec{\mu}$ the corresponding moment vector for the Heston model. We know that

$$\lim_{n\to\infty} ||\vec{\mu_n} - \vec{\mu}|| = 0.$$

Using the same parameters as in Table 3.1, numerical results for these norms are presented in Table 3.3.

We see that the difference in norms is not negligible even for $\bar{v} = 1000$. We made Monte Carlo simulations for the Heston model with the same parameters and observed that for 10e6 simulations and Δ_t =5e-5, no sample path of the volatility process has exceeded the value 0.4, which is far below the value 1000. Together with Appendix Figures 2 and 3 this suggests that simulated sample paths when using the same seed should be very similar for both models which in return should result in very similar option prices in the Monte Carlo simulations. We test this for different choices of $\bar{v}$ and include a simulation of the power Heston model with a new seed. Results and used simulation parameters are given in Table 3.4.

$\bar{v}_n$	$\lvert\Pi_H - \Pi_{PH}^{\text{same seed}}\rvert$	$\lvert\Pi_H - \Pi_{PH}^{\text{new seed}}\rvert$	$\Delta_\sigma^{\text{same seed}}$	$\Delta_\sigma^{\text{new seed}}$
0.08	1.4186e-03	1.3404e-03	3.5730e-03	3.3760e-03
1	1.3304e-04	3.3014e-04	3.3505e-04	8.3144e-04
10	1.4124e-05	4.8810e-04	3.5570e-05	1.2293e-03
100	1.3222e-06	3.4946e-04	3.3296e-06	8.8003e-04
1000	1.3174e-07	1.1076e-04	3.3178e-07	2.7895e-04

Table 3.4: Comparison of option prices by MC simulation. Comparison between Heston and Power Heston model. The Heston price is denoted with Π_H. The Power Heston price for the same seed as for the Heston model is $\Pi_{PH}^{\text{same seed}}$ and the one for a new seed $\Pi_{PH}^{\text{new seed}}$. Further the Black Scholes implied volatilities are compared where the absolute differences are denoted by $\Delta_\sigma^{\text{same seed}}$ and $\Delta_\sigma^{\text{new seed}}$ respectively. Simulation parameters are $\Delta_t = 1/20000$. The average price in the three simulations for the Heston Model is $\bar{\Pi}_H \approx 0.15691$, with implied Black Scholes volatility $\sigma_{impl}^{BS} \approx 0.39589$. Implied volatilities computed with an error tolerance of 1e-12, using the build in Matlab function blsimpv.

These simulations give us an idea of how far the prices in the two models should be apart. This will be useful as a sanity check when we compute and compare prices using density approximation.

The next step is to apply the density approximation for the different choices of $\bar{v}$. We use the same choices as for the Monte Carlo simulations. Further above, we have seen that condition (3.8) is important to ensure convergence of the densities. Hence the augmentation of $\bar{v}$ will cause problems. For this reason we will use the previously mentioned minimal auxiliary variance. For $\bar{v} = 100$, condition (3.8) would be satisfied for $\sigma_w > 50$. This choice however would not lead to desirable results as we can not go up to arbitrary orders due to round off errors and computational costs to compensate the distance to the true variance (the true variance of the Power Heston model is for all considered $\bar{v}$ between 0.0406 and 0.042, hence only for $\bar{v} = 0.08$ we can apply moment matching). We need good results for $J = 55$. Therefore we can only proceed heuristically in what follows, since we do not have theoretical proof for the density convergence if the auxiliary variance is in violation to (3.8). We plot densities for different choices of σ_w, coming from different minimal auxiliary variances, see Figure 3.8. We observe that it seems optimal to choose $\sigma_w \geq 0.2$ to avoid negative density values. We do this and continue with computing prices and compare them.

We compute the price of the European call option in the Heston model using Fourier inversion. For the prices in the power Heston model, we use the density approximation. It will not matter which of the two presented methods we use since both produce very similar prices. Hence we choose the pure recursive method. Results are displayed in Table 3.5.

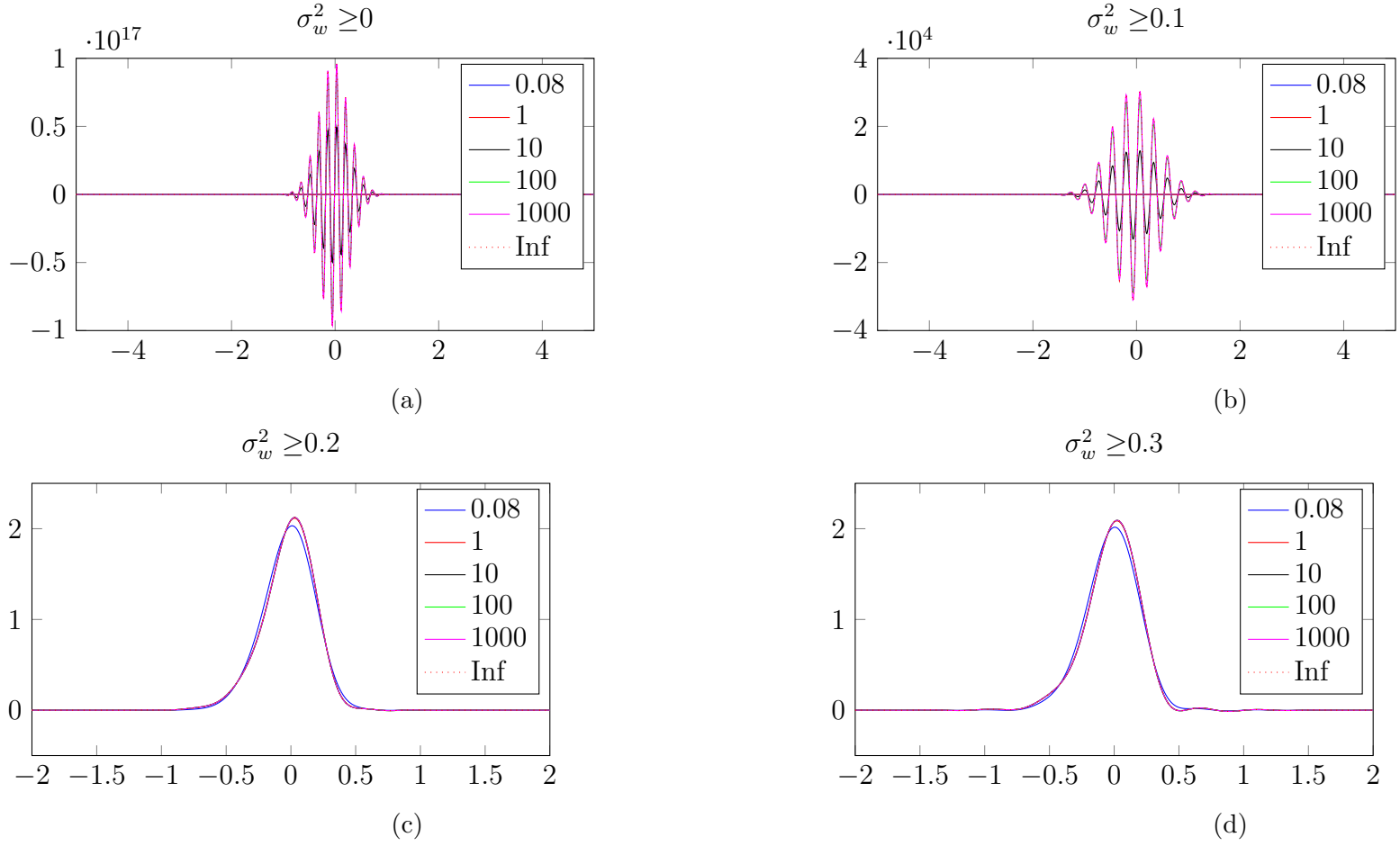


Figure 3.8: Plots for different choices of minimal auxiliary variances.

$\bar{v}_n$	Abs. error	Rel. error	Δ_σ	$\bar{v}_n$	Abs. error	Rel. error	Δ_σ
0.8	1.4371e-03	1.8599e-02	3.6195e-03	100	1.7581e-05	2.2754e-04	4.4277e-05
1	1.1667e-04	1.5100e-03	2.9383e-04	1000	1.8816e-05	2.4352e-04	4.7388e-05
10	5.2432e-06	6.7859e-05	1.3205e-05	∞	1.8954e-05	2.4530e-04	4.7734e-05

Table 3.5: Comparison of power Heston prices to Heston prices. Heston computed by Fourier inversion. All parameters except $\bar{v}$ are given in Table 3.1.

An interesting observation in Table 3.5 is that the best approximation to the Heston model is achieved for $\bar{v} = 10$. One possible explanation is that for bigger values of $\bar{v}$, the density approximation itself becomes too inaccurate due to condition (3.8). The above made Monte Carlo simulations on the other hand suggest that for $\bar{v} = 10$ we can not expect a price closer to the Heston price as we computed with the density approximation.

3.5. Conclusion

The conditional transition density approximation approach we presented seems to show good results for the Power Heston model. However the restriction on the choice of parameters pose a serious constraint. In particular, it is important that (3.8) is satisfied. We have further observed that it will not make sense to increases accuracy by augmenting the order J since for higher orders numerical errors seem to become too big of an issue. One solution to this problem could be by working in a multi precision framework. This however would come with the price of computational cost.

We have further seen that European option in the power Heston model can be (approximatly) priced without the need of any quadrature. This was demonstrated by using the known Hermite recursion formula to derive a recursion formula for the Fourier coefficients c_α.

Chapter 4

Markov Cubature

In this chapter we want to present the concept of Markov cubature rules as done by Filipoviv et al. in [10]. Cubature, also known as multi dimensional quadrature is by itself an active research area. In particular high dimensional problems are subject to the curse of dimensions. The methodology introduced in [10] is however different from classical cubature. The idea is to take a continuous time polynomial preserving process (we only present the pure diffusion case here) and replace it with a continious time Markov process living on a finite state space, where this process is supposed to match the moments of the initial process exactly up to a certain degree. What is a very strong requirement here, is that this moment matching has to happen for a certain set of times, usealle $t \geq 0$. Being able to replace the original process with a finitely supported process can increase flexibility and tractability, leading to applications such as filtering.

Cubature on the Wiener space has already been studied, see i.e. [21] and [2], and applied to filtering, see [18]. There however, the concept is different and the approach we will present is an alternative to those.

4.1. Continous time Markov cubature rules

We start by presenting the basic setup before and define Markov cubature rules. We do not consider the more general class of polynomial preserving processes[1] that would allow for jumps as done in [10]. Instead we only focus on the case where we have a pure diffusion processes with the property of its generator being polynomial

[1]The difference would only be in the assumptions on the process in order to be able to compute moments.

preserving. As demonstrated in Theorem 1.5, these processes allow the computation of moments by means of computing a matrix exponential. Recall that we are considering operators of the form

$$\mathcal{G}f(x) = b(x)^\top \nabla f(x) + \frac{1}{2} Tr(a(x)\nabla^2 f(x)).$$

In the sequel we will assume $\mathbb{E}\left[||X_0||^{2n}\right] < \infty$ so we can compute moments up to degree n using Theorem 1.5. As in the first part, let $E \subset \mathbb{R}^d$ any subset and with $Pol_n(E)$ we denote the set of multivariate polynomials on E. Let N_n be the dimension of $Pol_n(E)$ and fix any arbitrary basis $\beta = (h_1(x), \ldots, h_{N_n}(x))^\top$ of $Pol_n(E)$.

Since $\mathcal{G}$ is polynomial preserving, there exists a representing matrix on $Pol_n(E)$ whose real coefficients are given by

$$\mathcal{G}h_j(x) = \sum_{i=1}^{N_n} G_{ij} h_i(x). \tag{4.1}$$

As it follows from Theorem 1.5, moments can be computed by the formula

$$\mathbb{E}_x\left[h_j(X_t)\right] = \sum_{i=1}^{N_n} (\exp[tG])_{ij} h_i(x), \tag{4.2}$$

where the expectation is taken under $\mathbb{P}_x$, the solution to the martingale problem for $(\mathcal{G}, E)$ with initial law $\mu = \delta_x$. We now extend the definition of the moment vector H in the first part to several input arguments. Let $x_1, \ldots, x_{N_n} \in E$ be points in the state space E, we will call them cubature points, and define

$$H = H(x_1, \ldots, x_{N_n}) := (h_j(x_i))_{1 \le i \le M, 1 \le j \le N_n}.$$

In particular, this definition coincides with the previous definition for H in the first part for $M = 1$ and after transposition. Using the above notation, we can restate (4.1) and (4.2) and get

$$\mathcal{G}h_j(x_i) = (HG)_{ij}, \tag{4.3}$$

$$\mathbb{E}_x\left[h_j(X_t)\right] = (H\exp(tG))_{ij}. \tag{4.4}$$

Since (4.3) and (4.4) have to hold for all i and j, they establish a relation by posing a system of linear equations between generator and semigroup on the polynomial space $Pol_n(E)$ with a set of points in E, evaluated by the basis polynomials. Recalling that for polynomial preserving diffusions conditional moments are polynomials in the initial state space, the motivation for the above relationship will become clear with the following definition.

Definition 4.1. *Using the above notation, we call a finite state time-homogeneous Markov process Y with state space $E^Y := \{x_1, \dots, x_M\} \subset E$ a n-Markov cubature rule for X on $\mathcal{J} \subset [0, \infty)$, if for all $t \in \mathcal{J}, 1 \leq i \leq M$ and $1 \leq j \leq N_n$ we have*

$$\mathbb{E}_{x_i}\left[h_j(X_t)\right] = \mathbb{E}^Y_{x_i}\left[h_j(Y_t)\right] = \sum_{k=1}^{M} \mathbb{P}^Y_{x_i}(Y_t = x_k) h_j(x_k).$$

The above concept is similar, but different to classical cubature. Given a domain $\Omega \subset \mathbb{R}^d$, recall that one can see cubature as a finitely supported measure on Ω, that integrates a set of test functions exactly with respect to the original measure q on Ω. Usually these test functions are polynomials up to a certain degree.

In the above setting however, the measure one wants to integrate exact on $Pol_n(E)$, namely the law of X_t, depends on the initial value. Further we want the exact integration for all $t \in \mathcal{J}$, unlike classical cubature where the measure is just one. Another way to say this is that if above, $\mathcal{J}$ is a singelton and we only want the above equality to hold for one specific i, then we would have the situation of classical cubature where the transition rates $\mathbb{P}^Y_{x_i}$ are the weights and E^Y are the cubature points.

Another important point is that in the above definition, since we want equality for all i, the weights $\mathbb{P}^Y_{x_i}$ come from a finite-state Markov process and hence they must have a semigroup structure. The following theorem gives a characterization for continious time n-Markov cubature rules.

Theorem 4.2 (Theorem 3.1 in [10])**.** *Given a set of points $E^Y = \{x_1, \dots, x_M\} \subset E$ the following statements are equivalent.*

1. *There exists a continuous-time n-Markov cubature rule, Y, with state space E^Y.*
2. *For H as defined above, $HG = LH$, where the matrix $L \in \mathbb{R}^{M \times M}$ is some transition matrix, i.e*

$$\sum_{k=1}^{M} L_{ik} = 0 \ \forall i \text{ and } L_{ij} \geq 0 \text{ for } i \neq j.$$

3. *For each $x \in E^Y$ the vector*

$$v(x) := (\mathcal{G}h_1(x), \dots, \mathcal{G}h_{N_n}(x))$$

points inside the convex hull of the rows of $H = H(x_1, \dots, x_M)$ at the points $(h_1(x), \dots, h_{N_n}(x))$.

Additionally, if $N_n = M$ and H invertible (in one dimension this follows if the points are distinct), there exists a Lagrange basis of $Pol_n(E)$, denoted by $\tilde{\beta} = (\tilde{h}_1, \dots, \tilde{h}_{N_n})$, i.e. a basis with $\tilde{h}_j(x_i) = \delta_{ij}$. Further the following statement is equivalent to the previous ones:

4. $\mathcal{G}\tilde{h}_j(x_i) \geq 0$ for $i \neq j$.

In particular, the matrix L in 2. can be taken as the transition rate (or Q-) matrix for the n-Markov cubature rule Y.

We stop the discussions about continuous-time Markov cubature rules at this point, since the question about their existence, in general, has to be answered to the negative. In fact, Filipovic et al. show that one can not construct non trivial time continuous Markov cubature rules of arbitrary order for polynomial preserving diffusions. One way to approach this problem is by lifting the state space E^Y to signed measures. We will focus on another approach proposed in [10] and present it in the following.

4.2. Discrete time Markov cubature rules

The idea is to have a setting as close to classical cubature as possible. Existence and construction of cubature formulas are a well studied and still active field. Therefore we start with defining classical cubature.

In the same spirit as in one dimension, the name cubature is the higher dimensional analogy to find a cube with the same volume as a given ball. Cubature rules have already been studied by Gauss, and up to this day, they are of interest in current research. In many cases the results are for very specific geometries or weight functions. For a theoretical introduction about existence of cubature formulas on general compact domains using a functional analytical approach, we refer to the book by Sobolev and Vaskevic [30]. For explicit constructions of cubature formulas using optimization methods see [23]. We further want to refer to the paper by Ryu and Boyd, see [27] where the construction of cubature formulas is described as a certain linear program in infinite dimensions and a link to the exactness of these formulas (motivated by Gauss quadrature rules in one dimension) is established.

We start by defining a cubature rule. Let $E \subset \mathbb{R}^d$. In the sequel, let μ be a Borel probability measure supported on E. We define a cubature rule as follows.

Definition 4.3. *Let $\xi = \{\xi_1, \dots, \xi_N\} \subset E$ and $w = \{w_1, \dots, w_N\} \subset \mathbb{R}$. Denote by $p = \{p_0, \dots, p_n\}$ a set of real valued test functions on E, usually polynomials. We call (ξ, w) a cubature rule for (E, μ, p) of exactness n if*

$$\int_E p_i(x) d\mu(x) = \sum_{i=1}^{N} p_k(\xi_i) w_i, \quad \forall k = 1, \dots, n$$

and we call the cubature rule positive if all $w_i > 0$. We will call the points ξ_i cubature nodes and the numbers w_i weights.

When E is compact, the existence of cubature formulas has been proved, see e.g. [1] and the references given within. It is desirable to have N as small as possible for given n. In one dimension this translates to the fact that Gauss quadrature is optimal with respect to this point. In our situation this translates into the wish, that the state space of the cubature rule Y to be of as small as possible cardinality.

We will now assume to be able to find cubature points for a given measure on arbitrary E. We will use in the following to construct discrete time cubature rules.

The first step is to state the following theorem.

Theorem 4.4 (Theorem A.1 in [10])**.** *The following are equivalent:*

1. $\mathbb{E}\left[||X_0||^{2n}\right] < \infty$
2. *The sequence of matrices* $(\exp(tG))$ *converges as* $t \to \infty$.
3. *For all* $x \in E$ *and all* $j = 1, \dots, N_n$, $\mathbb{E}_x\left[h_j(X_t)\right]$ *converges as* $t \to \infty$.
4. *For all* $x \in E$ *and all* $f \in Pol_n(E)$, $\mathbb{E}_x\left[f(X_t)\right]$ *converges as* $t \to \infty$.

For the proof see [10]. In the sequel we will assume that the above theorem holds. We continue by defining

$$\mu_j(x) := \lim_{t \to \infty} \mathbb{E}_x\left[h_j(X_t)\right]$$

for $j = 1, \dots, N_n$ and $x \in E$. The next lemma is the foundation of our construction approach.

Lemma 4.5. *Assume $\mathbb{E}\left[||X_0||^{2n}\right] < \infty$. Suppose there are points*

$$x_1, \dots, x_M \in E$$

and

$$W = (w_{ij})_{i,j=1}^{M} \in \mathbb{R}^{M \times M}$$

with positive entries such that

$$\mu_j(x_k) = \sum_{i=1}^{M} w_{ki} H_j(x_i).$$

If the rank of $H(x_1, \ldots, x_M)$ has rank N_n, then for t big enough it holds

$$H \exp(tG) = Q(t)H,$$

where $Q(t)$ is a probability matrix with positive entries.

In practice we will have that the asymptotic means $\mu_j(x)$ do not depend on x, which will be important for our construction later. Sufficient conditions for this stationary character are Proposition A.2 and Corollary A.3 in [10]. In particular, under this assumption, we can simply write $\mu_j = \mu_j(x)$.

The construction is now as follows:

Consider the one dimensional case. Compute the time until one has (almost) the asymptotic moments for all j. Use these moments to construct the corresponding orthogonal polynomial of degree n using the Gram Schmidt process. Compute the roots of this polynomial to get the nodes of the Gauss quadrature for the asymptotic measure and use these points as Markov cubature points. Fix an initial time t. Solve the equation $H \exp(tG) = Q(t)H$ in Lemma 4.5 and check if Q is a probability matrix. If not, augment t and repeat the last step.

Remark 4.6. *In the last step above, the solving of $H \exp(tG) = Q(t)H$, the solution does not need to be unique. Therefore one should look at solutions that give probability matrices. But then this is nothing but a feasibility problem. Further, by introducing an auxiliary variable, one can formulate the task of finding a solution as on optimization problem where the target function forces the entries of the solution to be true positive numbers.*

Remark 4.7. *In the multi dimensional case, the task of finding the cubature points becomes finding a cubature rule for the asymptotic measure.*

We have now constructed a cubature rule for $\mathcal{J} = \Delta$ where Δ is the time for whcich $Q(t)$ is a solution. Using this matrix as transition matrix, the cubature rule can now be extended ro a cubature rule on $\mathcal{J} = \{l\Delta | l \in \mathbb{N}\}$.

Bibliography

[1] Christian Bayer and Josef Teichmann. The proof of tchakaloffs theorem. *Proceedings of the American Mathematical Society*, 134(10):3035–3040, 2006.

[2] Christian Bayer and Josef Teichmann. Cubature on wiener space in infinite dimension. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 464(2097):2493–2516, 2008.

[3] Christian Berg and Marco Thill. Rotation invariant moment problems. *Acta Mathematica*, 167(1):207–227, 1991.

[4] Patrick Billingsley. Probability and measure. 1979.

[5] Christa Cuchiero. *Affine and polynomial processes*. PhD thesis, ETH ZURICH, 2011.

[6] Christa Cuchiero, Martin Keller-Ressel, and Josef Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740, 2012.

[7] Darrell Duffie, Damir Filipović, and Walter Schachermayer. Affine processes and applications in finance. *Annals of applied probability*, pages 984–1053, 2003.

[8] Charles F Dunkl and Yuan Xu. *Orthogonal polynomials of several variables*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2014.

[9] D. Filipovic. *Term-Structure Models: A Graduate Course*. Springer Finance. Springer, 2009.

[10] D. Filipovic, M. Larsson, and S. Pulido. Markov cubature rules for polynomial preserving processes. Working paper, 2014.

[11] D. Filipovic and S. Pulido. The power Heston model. Working paper, 2014.

[12] Damir Filipovic and Martin Larsson. Polynomial preserving diffusions and applications in finance. *Swiss Finance Institute Research Paper*, 2014.

[13] Damir Filipović, Eberhard Mayerhofer, and Paul Schneider. Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111, 2013.

[14] G.B. Folland. *Fourier Analysis and Its Applications*. Pure and applied undergraduate texts. American Mathematical Society, 1992.

[15] Julie Lyng Forman and Michael Sørensen. The pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35(3):438–465, 2008.

[16] W. Gautschi, G.H. Golub, and G. Opfer. *Applications and Computation of Orthogonal Polynomials: Conference at the Mathematical Research Institute Oberwolfach, Germany, March 22-28, 1998*. International Series of Numerical Mathematics. Springer, 1999.

[17] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[18] Wonjung Lee and Terry Lyons. The adaptive patched particle filter and its implementation. *arXiv preprint arXiv:1311.6755*, 2013.

[19] E. Lukacs. *Characteristic Functions*. Griffin books of cognate interest. Hafner Publishing Company, 1970.

[20] Eugene Lukacs and Otto Szsz. On analytic characteristic functions. *Pacific Journal of Mathematics*, 2(4):615–625, 1952.

[21] Terry Lyons and Nicolas Victoir. Cubature on wiener space. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004.

[22] Olivier Mazet. Classification des semi-groupes de diffusion sur ir associés à une famille de polynômes orthogonaux. In *Séminaire de Probabilités XXXI*, pages 40–53. Springer, 1997.

[23] Sanjay Mehrotra and Dávid Papp. Generating moment matching scenarios using optimization techniques. *SIAM Journal on Optimization*, 23(2):963–999, 2013.

[24] Bernard P. *Lecture Notes in Physics*, volume 451/1995, chapter Some Remarks Concerning Convergence of Orthogonal Polynomial Expansions, pages 327–334. Springer, 1995.

[25] LC Petersen. On the relation between the multidimensional moment problem and the one-dimensional moment problem. *Mathematica Scandinavica*, 51:361–366, 1982.

[26] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*. Cambridge Mathematical Library. Cambridge University Press, 2000.

[27] Ernest K Ryu and Stephen P Boyd. Extensions of gauss quadrature via linear programming. *Foundations of Computational Mathematics*, pages 1–19, 2013.

[28] S. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance / Springer Finance Textbooks. Springer New York, 2010.

[29] Alicja Smoktunowicz. Backward stability of clenshaw's algorithm. *BIT Numerical Mathematics*, 42(3):600–610, 2002.

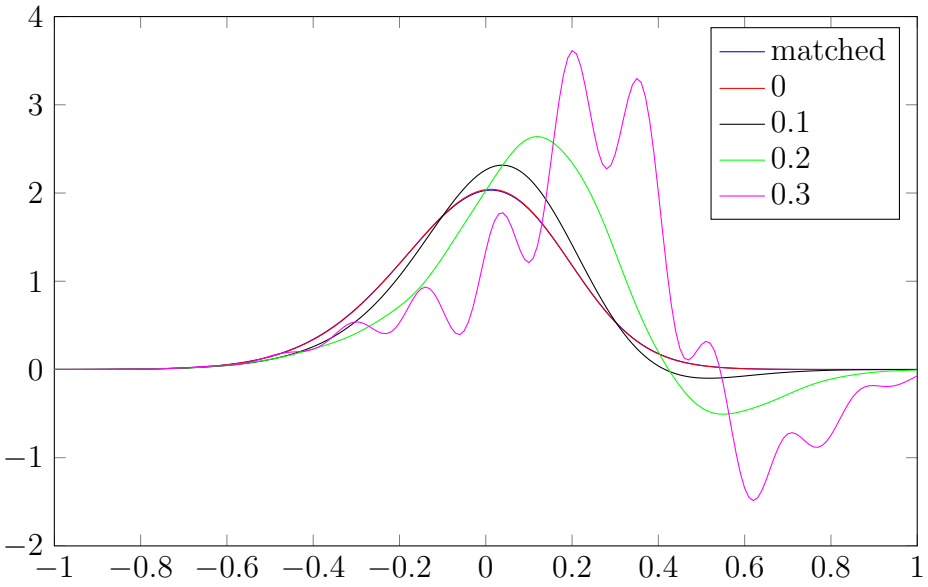
[30] S.L. Sobolev and V.L. Vaskevich. *The Theory of Cubature Formulas*. Mathematics and Its Applications. Springer, 1997.

[31] Nikolai G Ushakov. *Selected topics in characteristic functions*. Walter de Gruyter, 1999.

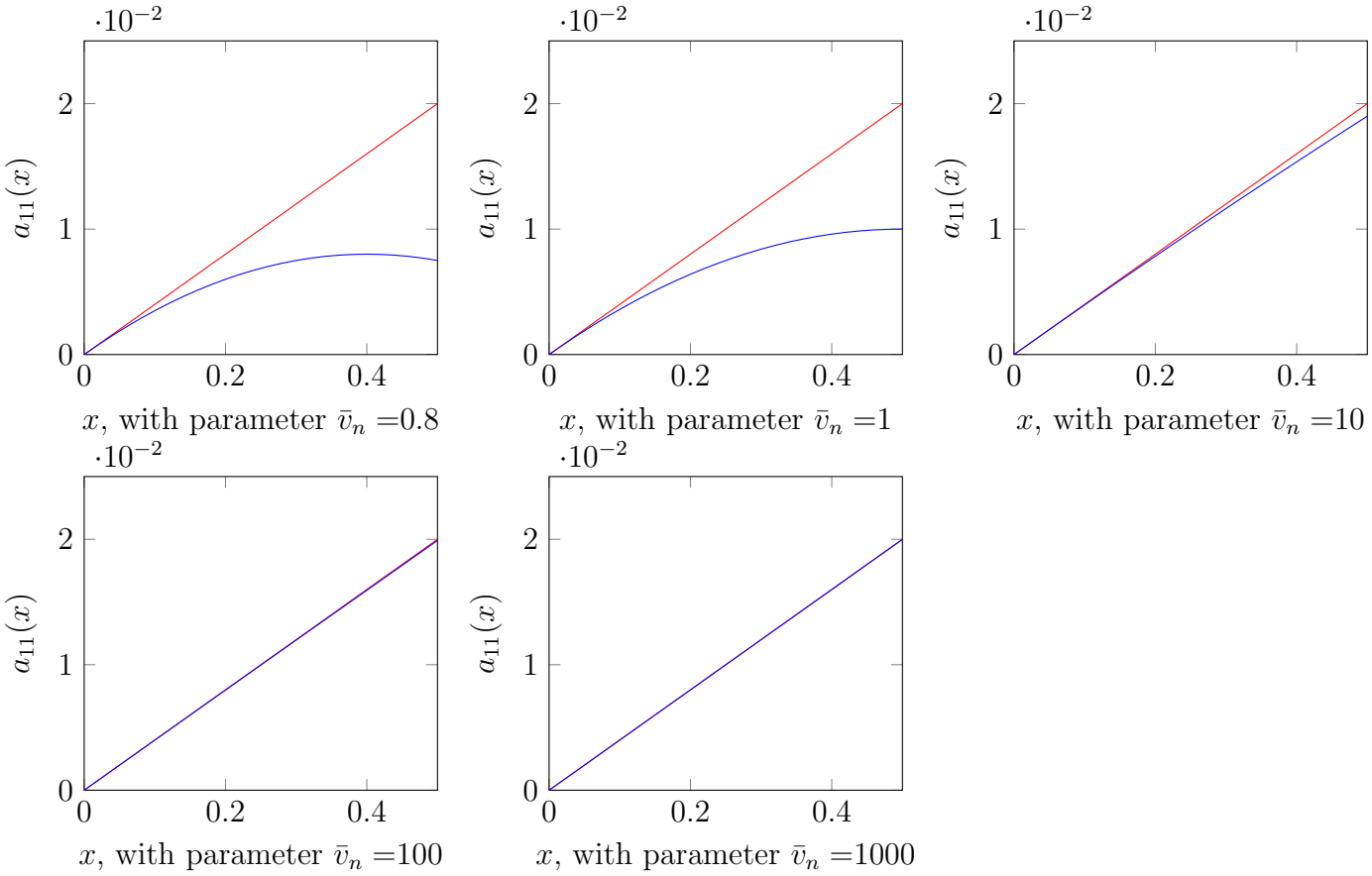
[32] E. Wong. The construction of a class of stationary markoff processes. *Stochastic Processes in Mathematical Physics and Engineering*, pages 264–276, 1964.

[33] Hao Zhou. Itô conditional moment generator and the estimation of short-rate processes. *Journal of Financial Econometrics*, 1(2):250–271, 2003.

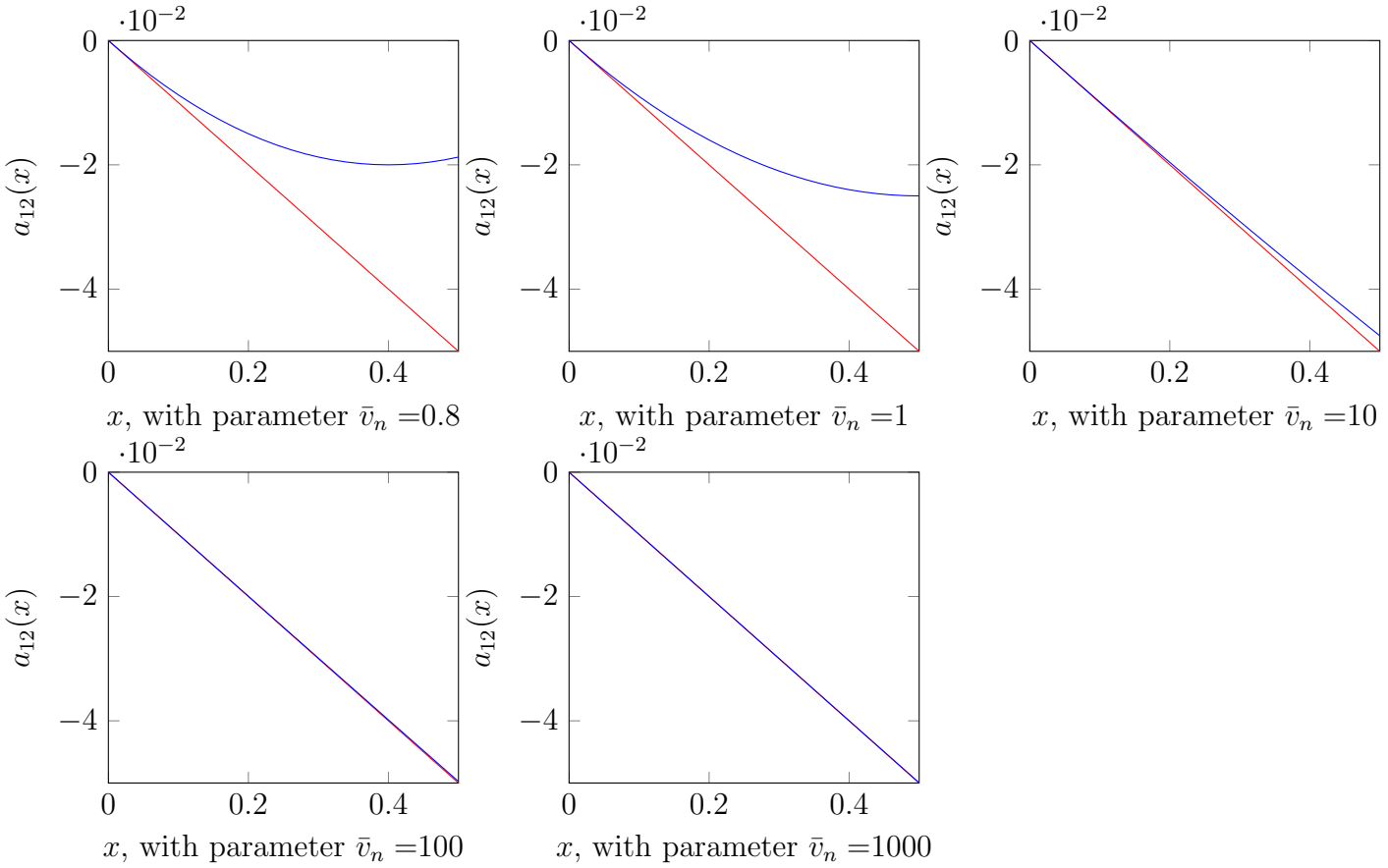
List of Figures



Appendix Figure 1: Plots of pseudo density functions with parameters given in Table 3.1. Variance is matched for the auxiliary density. Mean varies. True mean is -0.02.



Appendix Figure 2: Plots of diffusion matrix component $a(x)_{11}$ for Heston (red) and power Heston model (blue) as a function if $x = x_1$. Plots for different $\bar{v}$.



Appendix Figure 3: Plots of diffusion matrix component $a(x)_{12}$ for Heston (red) and power Heston model (blue) as a function if $x = x_1$. Plots for different $\bar{v}$.

Appendix A

Hermite Polynomials

In the chapter about the density approximation, we have seen that a right choice for the auxiliary density $w(x)$ is the normal density with general mean and variance. The associated orthogonal polynomials are (generalized) Hermite polynomials, which have the advantage of having a known closed form recurrence relation. Therefore one can constructing these polynomials (or avoid having to construct them at all) without moment based characterizations. This in return increases stability being crucial when dealing with high order approximations. Another advantage is that one can evaluate linear combinations using the Clenshaw algorithm without the need to know the coefficients in the monomial base.

Hermite polynomials are maybe the most popular example of orthogonal polynomials in one variable associated to a weight function with unbounded support. There are different orthogonal polynomials referred to as Hermite polynomials in the literature. The one kind, often referred to as physicists' Hermite polynomials, is associated to the weight function $w(x) = e^{-x^2}$, see e.g. [8]. The other kind, the probabilists' Hermite polynomials, is associated to the standard normal density (usually without the normalization constant). For our purposes we are interested in the latter with normalization and their generalizations to general mean and variance.

Orthogonal polynomials are not necessarily uniquely determined, even if one restricts to the normalized ones. An additional constraint could be to fix the leading coefficient. For our purposes it does not matter which unique definition we take, so we will only present one specific definition for Hermite polynomials.

There are several ways to define Hermite polynomials, e.g. using a generating function, or as a solution to an ordinary differential equation. We will take a purely constructive approach since we are interested in numerical applications and refer the reader for further reading and a more comprehensive study of Hermite polynomials and other orthogonal polynomials, to the existing literature, such as [16]. For

the case of orthogonal polynomials in several variables, we want to refer to the great book by Dunkl and Xu [8].

We will prove the recurrence relation in the most general case, i.e. $w(x)$ has general mean and variance, by starting from the standard normal case. We conclude with presenting a recurrence method to compute the coefficients c_α.

A.0.1. The recurrence relation for Generalized Hermite Polynomials

The first step is to define which polynomials we refer to as standard Hermite polynomials (i.e. the weight function will be the standard normal density) and show certain properties. Based on that we will make the transition to the general normal density, in particular we will prove the recurrence relation for the general case.

Definition A.1 (Standard Hermite polynomials)**.** *Let $H = (H_n)_{n\in\mathbb{N}_o}$ be a family of polynomials defined by $H_0 = 1$, $H_1 = x$ and for $n \geq 2$*

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x).$$

By definition, H is a family of monic polynomials with $deg(H_n) = n$. Another useful property of these polynomials is presented in the next Lemma.

Lemma A.2. *Let $H_n(x)$ be the sequence of standard Hermite polynomials defined above. These polynomials form an Appell sequence, i.e. $\forall n \in \mathbb{N}$ it holds*

$$H_n'(x) = nH_{n-1}(x).$$

Proof. TO DO: Follows directly by induction □

Our goal is to prove the next theorem.

Theorem A.3. *Define the family of polynomials $\tilde{H}^{\mu,\sigma}=(\tilde{H}_n^{\mu,\sigma})_{n\in\mathbb{N}_0}$ by $\tilde{H}_0^{\mu,\sigma} = 1$, $\tilde{H}_1^{\mu,\sigma} = (x-\mu)/\sigma$ and for $n \geq 2$ by the recurrence*

$$\tilde{H}_{n+1}^{\mu,\sigma}(x) = \frac{x-\mu}{\sigma\sqrt{n+1}}\tilde{H}_n^{\mu,\sigma}(x) - \sqrt{\frac{n}{n+1}}\tilde{H}_{n-1}^{\mu,\sigma}(x).$$

Denote by $\phi^{\mu,\sigma}$ the normal density function with mean μ and variance σ^2. Then $\tilde{H}^{\mu,\sigma}$ is a family of normalized orthogonal polynomials in $\mathcal{L}^2_{\phi^{\mu,\sigma}}$ with $deg(\tilde{H}_n^{\mu,\sigma}(x)) = n$ forming an orthonormal basis of $\mathcal{L}^2_{\phi^{\mu,\sigma}}$.

We will prove Theorem A.3 by constructing them from the standard Hermite Polynomials. The property of being an orthonormal basis follows from the existence of Lebesgue densities in the normal case and that the characteristic function of a normal distribution is analytical around zero, ensuring density of polynomials in $\mathcal{L}^2_{\phi^{\mu,\sigma}}$, see [24].

Recall that the inner product $\langle \cdot, \cdot \rangle_{\mathcal{L}^2_{\phi^{\mu,\sigma}}}$ of $\mathcal{L}^2_{\phi^{\mu,\sigma}}$ is given by

$$\langle f, g \rangle_{\mathcal{L}^2_{\phi^{\mu,\sigma}}} := \int_{\mathbb{R}} f(\xi) g(\xi) \phi^{\mu,\sigma}(\xi) d\xi.$$

For the case where $\phi^{\mu,\sigma}$ is the standard normal density, we will just write ϕ to simplify notation. The next step is to derive the recurrence relation for the normalized standard Hermite polynomials.

Lemma A.4. *Let $\phi(\xi)$ be the standard normal density function. Then*

$$\langle H_n, H_m \rangle_{\mathcal{L}^2_\phi} = \int_{\mathbb{R}} H_n(\xi) H_m(\xi) \phi(\xi) d\xi = \delta_{n,m} n!$$

Consequently, these polynomials are orthogonal polynomials in $\mathcal{L}^2_w$ and the polynomials $\tilde{H}_n := H_n / ||H_n||_{\mathcal{L}^2_\phi}$ form an orthonormal basis of $\mathcal{L}^2_\phi$ with $deg(\tilde{H}_n) = n$. Further they satisfy the recurrence relation $\tilde{H}_0 = 1, \tilde{H}_1 = x$ and for $n \geq 2$

$$\tilde{H}_{n+1}(x) = \frac{x}{\sqrt{n+1}} \tilde{H}_n(x) - \sqrt{\frac{n}{n+1}} \tilde{H}_{n-1}(x)$$

Proof. By induction, recurrence and integration by parts and using $\phi'(\xi) = -\xi\phi(\xi)$. □

A simple change of variable and noticing that the polynomial $(x - \mu)/\sigma$ is of degree one gives the following lemma, which proves Theorem A.3.

Lemma A.5. *The polynomials defined by $\tilde{H}^{\mu,\sigma}_n(x) := \tilde{H}_n(\frac{x-\mu}{\sigma})$ are orthonormal polynomials in $\mathcal{L}^2_{\phi^{\mu,\sigma}}$ with $deg(\tilde{H}^{\mu,\sigma}_n) = n$.*

Since we have all recurrence coefficients of the generalized Hermite polynomials $\tilde{H}^{\mu,\sigma}_n$ in closed form, we can efficiently evaluate a linear combination of them using the Clenshaw algorithm. Alternatively we apply the Clenshaw algorithm to (A.4) after transforming the variable.

However we still need to construct the $\tilde{H}^{\mu,\sigma}_n$ to compute the linear combination coefficients c_k. Using Theorem A.3, we can circumvent this as described in the following subsection.

A.0.2. Computation of Generalized Hermite Moments

We continue by presenting a method to compute the moments of orthogonal polynomials (such as the generalized Hermite polynomials) without having to compute the coefficients of those polynomials. These polynomials satisfy a certain three term recursion which we will make use of. The moments up to a previously specified order will be computed recursively (for stability reasons) by a triangular scheme and our method works for all probability measures with Lebesgue density g with finite and known moments. We start by motivating our method with the task of computing the coefficients c_k for the pseudo densities, see chapter 3. Our method is for the one dimensional case, but we believe that it is also applicable for multivariate settings after modification, in particular since there is a three term recursion principle for multivariate orthogonal polynomials as well. Further one can generalize to a family of polynomials satisfying a general multi term recursion. This however is not of interest in this thesis, which is why we will only consider the specific setting of univariate orthogonal polynomials.

Denote by $w(x)$ the auxiliary density under which the sequence $(H_n)_{n\in\mathbb{N}_0}$ is orthonormal. Recall that the coefficients c_k are given by

$$c_k = \left\langle \frac{g}{w}, H_k \right\rangle_{\mathcal{L}^2_w} = \int_{\mathbb{R}} H_k(\xi) g(\xi) d\xi$$

In case of the Hermite polynomials, see Definition A.1, for general normal density $w(x)$, we call c_k the k'th (g, μ, σ) Hermite moment, or in short k'th Hermite moment if there is no ambiguity. Using the three term recursion (A.3), we get

$$\begin{aligned} c_{n+1} &= \frac{1}{\sigma\sqrt{n+1}} \int_{\mathbb{R}} (x-\mu) H_n(\xi) g(\xi) d\xi - \sqrt{\frac{n}{n+1}} c_{n-1} \\ &= \frac{1}{\sigma\sqrt{n+1}} \left(\int_{\mathbb{R}} x H_n(\xi) g(\xi) d\xi - \mu c_n \right) - \sqrt{\frac{n}{n+1}} c_{n-1}. \end{aligned}$$

The problem is now that even if we know the two preceding values c_k and c_{k-1}, we can not compute c_{k+1} since $\langle g/w, xH_k \rangle_{\mathcal{L}^2_w}$ is unknown. For given moments of g, we would need to know the coefficients of H_k in the monomial basis, but constructing H_k is what we want to avoid for stability reasons.

Our method is now based on the following inside. By repeating the above substitution iteratively for $H_m, m \in \{k, k-1, \cdots, 0\}$ we can express all Hermite moments $c_k, k \leq n$ if we know $\langle g/w, x^l H_0 \rangle_{\mathcal{L}^2_w}$ for all $l \leq n$ using the known three term recursion for $(H_n)_{n\in\mathbb{N}_0}$. But since $H_0 \equiv 1$, we have $\langle g/w, x^l H_0 \rangle_{\mathcal{L}^2_w} = \mu_g(l)$, the l'th moment of g, which we can compute for polynomial preserving processes.

We want to formalize this. Define the matrix $M = (M_{i,j})_{i,j=0}^{n}$ by

$$M_{i,j} := \int_{\mathbb{R}} \xi^j H_i(\xi) g(\xi) d\xi.$$

In particular, the first row of M contains the known moments of g. We are interested in the Hermite moments up to order n, which form exactly the first column of M. We will now proceed by computing the next row of M with the exception of the element $M_{1,n}$ using only the first row and the recurrence relation. In the next step we will compute all elements of the third row except the last two. Repeating this n times, we will have computed all values of M in the upper triangle (including the diagonal entries), hence the expression triangle recursion.

We present this algorithm in a slightly more general framework. In particular it can be applied to any family of orthogonal polynomials. Given $H_0 \equiv 1$, let H_n be a polynomial sequence satisfying

$$H_{n+1} = A_{n+1} x H_n + B_{n+1} H_n + C_{n+1} H_{n-1}$$

for constants A_n, B_n and C_n only depending on n where we set $H_{-1} = 0 = C_1$ so that (A.0.2) is well defined for $n \geq 1$. In particular that implies $M_{-1,l} = 0$ for all l.

Given $\mu_g(l)$ for $l = 0, \cdots n$, the following algorithm computes the values $c_l = \langle g/w, H_l \rangle_{\mathcal{L}^2_w}$ for $l = 0, \cdots, n$ by computing the matrix M (we use Matlab inspired pseudo code):

Listing A.1: Algorithm for Polynomial Moments

```
Input:  [mom_0,...,mom_n, A_1,...,A_n, B_1,...,B_n, C_2,...,
   C_n]
Output: c=[c_0,...,c_n]

C_1=0;
% We deviate from Matlab and start indexing at zero
M = zeros(n); % M is now a (n+1)*(n+1) zero Matrix
M(0,:) = [mom_0,...,mom_n];

for k=1:n
        for l=0:n-k
                M(k,l) = A_k * M(k-1,l+1) + B_k * M(k-1,l)...
                 +C_k * M(k-2,l); % for k=1, M(k-2,l) out of
                    % bounds. Interpret as zero.
        end
end
```

```
c = M(:,0);
return;
```

Remark A.6. *If one is only interested in a specific c_k and not the whole range, the computational effort can be reduced.*

Overall we have constructed a method to efficiently evaluate the pseudo densities in a stable way without having to construct the orthonormal polynomials.