DENSITY APPROXIMATIONS FOR MULTIVARIATE AFFINE JUMP-DIFFUSION PROCESSES

DAMIR FILIPOVIĆ1, EBERHARD MAYERHOFER2, AND PAUL SCHNEIDER3

Abstract. We introduce closed-form transition density expansions for multivariate affine jump-diffusion processes. The expansions rely on a general approximation theory which we develop in weighted Hilbert spaces for random variables which possess all polynomial moments. We establish parametric conditions which guarantee existence and differentiability of transition densities of affine models and show how they naturally fit into the approximation framework. Empirical applications in option pricing, credit risk, and likelihood inference highlight the usefulness of our expansions. The approximations are extremely fast to evaluate, and they perform very accurately and numerically stable.

1. Introduction

Most observed phenomena in financial markets are inherently multivariate: stochastic trends, stochastic volatility, and the leverage effect in equity markets are well-known examples. The theory of affine processes provides multivariate stochastic models with a well established theoretical basis and sufficient degree of tractability to model such empirical attributes. They enjoy much attention and are widely used in practice and academia. Among their best-known proponents are Vasicek’s interest rate model (Vasicek, 1977), the square-root model Cox et al. (1985), Heston’s model (cf. Heston, 1993), and affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000; Collin-Dufresne et al., 2008). Affine models owe their popularity and their

1École Polytechnique Fédérale de Lausanne and Swiss Finance Institute, Quartier UNIL-Dorigny, Extranef 218, CH - 1015 Lausanne, Switzerland
2School of Mathematical Sciences, Dublin City University, Ireland
3Institute of Finance, University of Lugano, Via Buffi 13, CH-6900 Lugano.
E-mail addresses: damir.filipovic@epfl.ch, eberhard.mayerhofer@gmail.com, paul.schneider@usi.ch.
Date: 15 March 2013, Journal of Econometrics forthcoming.

Key words and phrases. Affine Processes, Asymptotic Expansion, Density Approximation, Orthogonal Polynomials.

JEL Classification: G12, C13, C32.
name to their key defining property: their characteristic function is of exponential affine form and can be computed by solving a system of generalized Riccati differential equations (cf. Duffie et al. (2003)). This allows for computing transition densities and transition probabilities by means of Fourier inversion (Duffie et al., 2000). Transition densities constitute the likelihood which is an ingredient for both frequentist and Bayesian econometric methodologies. Also, they appear in the pricing of financial derivatives. However, Fourier inversion is a very delicate task. Complexity and numerical difficulties increase with the dimensionality of the process. Efficient density approximations avoiding the need for Fourier inversion are therefore desirable.

This paper is concerned with directly approximating the transition density without resorting to Fourier inversion techniques. We pursue a polynomial expansion approach, an idea that has been proposed by Wong and Thomas (1962), Wong (1964), Schoutens (2000), Schauburg (2001), Aït-Sahalia (2002), and Hurn et al. (2008) among others for univariate diffusion processes. Extensions for multivariate (jump-)diffusions do exist in Aït-Sahalia (2008) (with applications in Aït-Sahalia and Kimmel, 2007, 2010), and Yu (2007), but they follow a different route by approximating the Kolmogorov forward-, and backward partial differential equations. Our approach exploits a crucial property of affine processes. Under some technical conditions, conditional moments of all orders exist and can be explicitly computed in closed form as the solution of a matrix exponential. This ensures that the coefficients of the polynomial expansions can be computed without approximation error.

We present a general theory of density approximations with several traits of the affine model class in mind. The assumptions made for the general theory are then justified by proving existence and differentiability of the true, unknown transition densities of affine models. These theoretical results, contrary to the density approximations themselves, do rely on Fourier theory. Specifically, we investigate the asymptotic behavior of the characteristic function with novel ODE techniques.

Specializing to affine models we improve earlier work along several lines. Our method (i) is applicable to multivariate models; (ii) works equally well for reducible and irreducible processes in the sense of Aït-Sahalia (2008); in particular stochastic volatility models; (iii) produces density approximations the quality of which is independent of the time interval between observations; (iv) allows for expansions on the "correct" state space. That is, the support of the density approximation agrees with the support of the true, unknown transition density as in Hurn et al. (2008) and Schoutens (2000); (v) produces density approximations that integrate to unity by construction, hence are much more amenable to applications that demand the constant of proportionality than the purely polynomial expansions from Aït-Sahalia (2008). This includes Wishart processes Bru (1991) and even general affine matrix-valued processes (Cuchiero

---

1 Various other approaches for parameter estimation for discretely observed Markov processes can be found in the literature (excellent comprehensive surveys are for example in Hurn et al., 2007; Sørensen, 2004; Aït-Sahalia, 2007). The approaches range from likelihood approximation using Bayesian data augmentation (Roberts and Stramer, 2001; Elerian et al., 2001; Eraker, 2001; Jones, 1998), estimating functions (Bibby et al., 2004), up to the efficient method of moment Gallant and Tauchen (2009). Only few of them make use of the properties of affine models, however (e.g. Singleton, 2001; Bates, 2006).

2 A model is said to be reducible in the sense of Aït-Sahalia (2008) if its diffusion function can be transformed one-to-one into a constant. Otherwise it is termed irreducible.

3 The Markov chain Monte Carlo sampling schemes from Stramer et al. (2009) accommodate Bayesian likelihood-based inference using expansions from Aït-Sahalia (2008) even in absence of the normalizing constant, but at a high computational cost.
et al., 2010a). This paper therefore provides a unified framework for econometric inference for financial models, because in applications one typically needs to evaluate, both, the transition densities themselves, as well as integrals of payoff functions against the transition densities for model-based asset pricing. This complements the methods recently developed in Chen and Joslin (2011) and Kristensen and Mele (2011), which are aimed at asset pricing.\footnote{It is of course conceivable to mix the mentioned methods. For example, one could use transition densities developed in this paper, while approximating asset prices using the generalized Fourier transform in Chen and Joslin (2011), whenever the payoff function allows it, or the error expansion method from Kristensen and Mele (2011).}

The paper proceeds as follows: Section 2 develops a general theory of orthonormal polynomial density approximations in certain weighted $L^2$ spaces under suitable integrability and regularity assumptions. These may be validated by the sufficient criteria presented subsequently in Section 3. The density approximations are then specialized within the context of affine processes: Section 4 reviews the affine transform formula and the polynomial moment formula for affine processes, which in turn allows the aforementioned polynomial approximations. The main theoretical contribution – general results on existence and differentiability of transition densities of affine processes – is elaborated in Section 4.3. In Section 5 we introduce candidate weight functions and the Gram-Schmidt algorithm to compute orthonormal polynomial bases corresponding to these weights, along with important examples. Section 6 details computation of the ingredients to the density expansions using an explicit example. Section 7 relates existing techniques for density approximations to ours. An empirical study is presented in Section 8: Applications in stochastic volatility (Section 8.1), option pricing (Section 8.2), credit risk (Section 8.3), and likelihood inference (Section 8.4), support the tractability and usefulness of the likelihood expansions. Section 9 concludes. The proofs of our main results are given in an Internet Appendix.

In the paper we will use the following notational conventions. The nonnegative integers are denoted by $\mathbb{N}_0$. The length of a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ is defined by $|\alpha| = \alpha_1 + \cdots + \alpha_d$, and we write $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$ for any $\xi \in \mathbb{R}^d$. The degree of a polynomial $p(x) = \sum_{|\alpha| \geq 0} p_\alpha x^\alpha$ in $x \in \mathbb{R}^d$ is defined as $\deg p(x) = \max\{|\alpha| \mid p_\alpha \neq 0\}$. For the likelihood ratio functions below we define $0/0 = 0$. The class of $p$-times continuously differentiable (or continuous, if $p = 0$) functions on $\mathbb{R}^d$ is denoted by $C^p$.

2. Density Approximations

Let $g$ denote a probability density on $\mathbb{R}^d$ whose polynomial moments

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

of every order $\alpha \in \mathbb{N}_0^d$ exist and are known in closed form. For example, $g$ may denote the pricing density in a financial market model. Typically, $g$ is not known in explicit form, and needs to be approximated. Let $w$ be an auxiliary probability density function on $\mathbb{R}^d$. The aim is to expand the likelihood ratio $g/w$ in terms of orthonormal polynomials of $w$ in order to get an explicit approximation of the unknown density function $g$. This can be formalized as follows. Define the weighted Hilbert space $L^2_w$ as the set of (equivalence classes of) measurable functions $f$ on $\mathbb{R}^d$...
with finite $L^2_w$-norm defined by

$$\|f\|_{L^2_w}^2 = \int_{\mathbb{R}^d} |f(\xi)|^2 w(\xi) \; d\xi < \infty.$$  

Accordingly, the scalar product on $L^2_w$ is denoted by

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

We will now proceed under the following assumptions. Sufficient conditions for the assumptions to hold are provided in Section 3 below.

**Assumption 1.** There exists an orthonormal basis of polynomials $\{H_\alpha | \alpha \in \mathbb{N}_0^d \}$ of $L^2_w$ with $\deg H_\alpha = |\alpha|$. This implies $H_0 = 1$ in particular.

**Assumption 2.** The likelihood ratio function $g/w$ lies in $L^2_w$. This is equivalent to

$$\int_{\mathbb{R}^d} g(\xi)^2 w(\xi) \; d\xi < \infty.$$  

Consequently, the coefficients

$$c_\alpha = \langle \frac{g}{w}, H_\alpha \rangle_{L^2_w} = \int_{\mathbb{R}^d} H_\alpha(\xi) g(\xi) \; d\xi \quad (= 1 \text{ for } \alpha = 0)$$

are well-defined and given explicitly\(^5\) in terms of the coefficients of $H_\alpha$ and the polynomial moments $\mu_\alpha$ of $g$. Moreover, according to standard $L^2_w$-theory, the sequence of pseudo-likelihood ratios\(^6\) $1 + \sum_{|\alpha|=1}^J c_\alpha H_\alpha$ approximates the likelihood ratio $g/w$ in $L^2_w$ for $J \to \infty$. In fact, defining the pseudo-density functions\(^7\)

$$g^{(J)}(x) = w(x) \left( 1 + \sum_{|\alpha|=1}^J c_\alpha H_\alpha(x) \right)$$

the following properties can be established.

**Theorem 2.1.** The pseudo-density functions $g^{(J)}$ satisfy

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

$$\lim_{J \to \infty} \int_{\mathbb{R}^d} \left| \frac{g^{(J)}}{w} - \frac{g}{w} \right|^2 \frac{d\xi}{w(\xi)} = 0.$$  

**Proof.** A calculation shows that $\int_{\mathbb{R}^d} H_\alpha(\xi) w(\xi) \; d\xi = \langle H_\alpha, 1 \rangle_{L^2_w} = \langle H_\alpha, H_0 \rangle_{L^2_w} = 0$, by the orthogonality of $H_\alpha$ and $H_0 = 1$. Hence $\int_{\mathbb{R}^d} g^{(J)}(\xi) \; d\xi = \int_{\mathbb{R}^d} w(\xi) \; d\xi = 1$, which proves (2.2). Properties (2.3) and (2.4) are formal restatements of the discussion preceding the theorem.  

\(^5\)This is an advantage over the method in A"ıt-Sahalia (2002) which also relies on series expansions, where the coefficients are functions of expectations of nonlinear moments, and therefore have to be approximated in general.

\(^6\)See Footnote 7 below for an explanation of this terminology.

\(^7\)Theorem 2.1 below states that $g^{(J)}$ integrates to one, but $g^{(J)}$ may take negative values. Whence we shall call $g^{(J)}$ a pseudo-density function, and $g^{(J)}/w$ a pseudo-likelihood ratio.
Property (2.2) proves to be very useful for applications where the constant of proportionality is needed, for example option pricing and the computation of Bayes factors.

The idea of expanding the likelihood ratio function \( g/w \) in terms of orthonormal polynomials of \( w \) is simple and powerful. An overview and discussion of related literature can be found e.g. in Bernard (1995). In particular, for the case where \( w \) is the standard Gaussian density, (2.1) is actually the Gram–Charlier expansion of \( g \). But note that Assumption 2 is very restrictive in this case. This is why the Gram–Charlier series diverges in most cases of interest, which is sometimes given as an argument against the use of it. However, the blame is on the choice of the Gaussian as auxiliary density. The efficiency of the approximation (2.3), or equivalently (2.4), lies in the appropriate choice of the auxiliary density function \( w \) and the corresponding orthonormal polynomials \( H_\alpha \).

Here is a first result towards a good choice of \( w \). The intuition is to choose \( w \) as close as possible to the unknown density function \( g \), in the sense that the pseudo-likelihood ratio \( g/w \) is close to one. This should be achieved if many of the coefficients \( c_\alpha \), other than \( c_0 = 1 \), are equal to zero. This will also improve the numerical efficiency of the approximation as the respective orthonormal polynomials \( H_\alpha \) need not be computed. Denote the polynomial moments of \( w \) by \( \lambda_\alpha = \int_{\mathbb{R}^d} \xi^\alpha w(\xi) \, d\xi \).

**Lemma 2.2 (Moment Matching Principle).** Suppose for some \( n \geq 1 \), we have \( \mu_\alpha = \lambda_\alpha \) for all \( |\alpha| \leq n \). Then \( c_\alpha = 0 \) for \( 1 \leq |\alpha| \leq n \).

**Proof.** The assumption implies that, for \( 1 \leq |\alpha| \leq n \), \( c_\alpha = \int_{\mathbb{R}^d} H_\alpha(\xi) g(\xi) \, d\xi = \int_{\mathbb{R}^d} H_\alpha(\xi) w(\xi) \, d\xi = \langle H_\alpha, 1 \rangle_{L^2_w} = \langle H_\alpha, H_0 \rangle_{L^2_w} = 0 \), by the orthogonality of \( H_\alpha \) and \( H_0 = 1 \). \( \square \)

### 3. Sufficient Conditions for Assumptions 1 and 2

In this section we provide sufficient conditions for Assumptions 1 and 2 to hold. The proofs of the following lemmas are deferred to the Internet Appendix B. We first provide sufficient conditions on \( w \) that guarantee that Assumption 1 is satisfied.

**Lemma 3.1.** Suppose that the density function \( w \) has a finite exponential moment

\[
\int_{\mathbb{R}^d} e^{\epsilon_0 \|\xi\|} w(\xi) \, d\xi < \infty
\]

for some \( \epsilon_0 > 0 \). Then the set of polynomials is dense in \( L^2_w \). Moreover, Assumption 1 is satisfied.

In applications, the auxiliary density function \( w \) on \( \mathbb{R}^d \) will often be given as product of marginal densities \( w_i \) on \( \mathbb{R} \). Hence the following modification of Lemma 3.1 will be useful.

**Lemma 3.2.** Let \( w_1, \ldots, w_d \) be density functions on \( \mathbb{R} \) having finite exponential moments

\[
\int_{\mathbb{R}} e^{\epsilon |\xi_i|} w_i(\xi_i) \, d\xi_i < \infty
\]

\( \text{For multivariate applications using polynomial expansions as in this paper, or Ait-Sahalia (2008) and Yu (2007), the number of polynomials increases exponentially with the truncation lag } J. \text{ To overcome this curse of dimensionality (at an efficiency loss) one could pursue a marginalization approach along the lines of the LML-CCF estimator from Singleton (2001).} \)
for some $\epsilon_i > 0$, $i = 1, \ldots, d$. Then the product density $w(\xi) = w_1(\xi_1) \cdots w_d(\xi_d)$ on $\mathbb{R}^d$ admits a finite exponential moment (3.1) for $\epsilon_0 = \min_i \epsilon_i$. Moreover, let $\{H^i_j \mid j \in \mathbb{N}_0\}$ denote the corresponding orthonormal basis of polynomials of $L^2_{w_i}(\mathbb{R})$ by $\deg H^i_j = j$, for $i = 1, \ldots, d$, asserted by Lemma 3.1. Then $H^i_0(\xi) = \alpha_0^i(\xi) \cdots H^d_{\alpha_d}(\xi)$ defines an orthonormal basis of polynomials of $L^2_w$ with $\deg H^i_0 = |\alpha^i|$, and Assumption 1 is satisfied.

Assumption 2 is opposite to Assumption 1 in the sense that there we have to bound the auxiliary density function $w$ from below. The following lemmas provide sufficient conditions for Assumption 2 to hold.

**Lemma 3.3.** Assume that $g$ is bounded and has a finite exponential moment $\int_{\mathbb{R}^d} e^{\epsilon_0 \|\xi\|} g(\xi) \, d\xi < \infty$ for some $\epsilon_0 > 0$. If $w$ decays at most exponentially such that

$$\sup_{x \in \mathbb{R}^d} \frac{e^{-\epsilon_0 \|x\|}}{w(x)} < \infty$$

then Assumption 2 is satisfied.

If the support of $w$ and $g$ is contained in a subset $D$ of $\mathbb{R}^d$, the situation becomes more difficult as one has to control the rate at which $w$ converges to zero at the boundary of the support set. We provide sufficient conditions for the set $D = \mathbb{R}^m_+ \times \mathbb{R}^n$, starting with the scalar case $D = \mathbb{R}_+^+$.\n
**Lemma 3.4.** Let $d = 1$ and $p \in \mathbb{N}$. Assume that $g$ is a bounded density with support in $\mathbb{R}_+$ and has a finite exponential moment $\int_0^\infty e^{\epsilon_0 \xi} g(\xi) \, d\xi < \infty$ for some $\epsilon_0 > 0$. Assume further that $g$ is of class $C^p$. If $w$ has support in $\mathbb{R}_+$, and decays at most polynomially at zero and exponentially at infinity such that

$$\sup_{x \in [0,1]} \frac{x^{2p}}{w(x)} < \infty$$

then Assumption 2 is satisfied.

The case where $D = \mathbb{R}^m_+ \times \mathbb{R}^n$ is similar, but requires stronger conditions on $g$ and $w$. We respect the product structure of the domain by writing $g = g(x,y)$ for $x \in \mathbb{R}^m_+$ and $y \in \mathbb{R}^n$. The following tubular neighborhood of the boundary of $D$

$$\mathcal{I} = \mathbb{R}^m_+ \setminus (1, \infty)^m = \{x \in \mathbb{R}^m_+ \mid \min_i x_i \leq 1\}$$

is the convenient multivariate generalization of the unit interval from the above scalar case.

**Lemma 3.5.** Let $d = m + n$ and $p \in \mathbb{N}$. Assume that $g(x,y)$ is a bounded density with support in $\mathbb{R}^m_+ \times \mathbb{R}^n$ and has a finite exponential moment

$$\int_{\mathbb{R}^m_+} \int_{\mathbb{R}^n} e^{\epsilon_1 \|\xi\| + \epsilon_2 \|\eta\|} g(\xi, \eta) \, d\xi \, d\eta < \infty$$

for some $\epsilon_1, \epsilon_2 > 0$. Assume further that $g(x,y)$ is of class $C^p$ in $x$ and the $p$-th partial derivative $\partial^p_i g(x,y)$ is bounded on $\mathcal{I} \times \mathbb{R}^n$, for all $i = 1, \ldots, m$. If $w$ has support in $\mathbb{R}^m_+ \times \mathbb{R}^n$, and decays at most polynomially around the boundary and exponentially at infinity such that

$$\sup_{(x,y) \in \mathcal{I} \times \mathbb{R}^n} \frac{\min_i x_i \, e^{-\epsilon_2 \|y\|}}{w(x,y)} < \infty$$

and

$$\sup_{(x,y) \in (1, \infty)^m \times \mathbb{R}^n} \frac{e^{-\epsilon_1 \|x\| - \epsilon_2 \|y\|}}{w(x,y)} < \infty$$

then Assumption 2 is satisfied.
We note that the conditions in Lemmas 3.3 - 3.5 can be explicitly verified for transition densities of affine processes, see Corollary 4.4 below.

4. Affine Models

The main application of the polynomial density approximation are affine factor models. In this section, we follow the setup of Duffie et al. (2003), which we now briefly recap. Let \( d = m + n \geq 1 \). We define the index set \( J = \{m+1, \ldots, d\} \), and write \( v_J = (v_{m+1}, \ldots, v_d) \) and \( m_{JJ} = (m_{kl})_{k,l \in J} \), for any vector \( v \) and matrix \( m \). We consider an affine process \( X \) on the canonical state space \( \mathcal{D} = \mathbb{R}^m_+ \times \mathbb{R}^n \) with generator

\[
Af(x) = \sum_{k,l=1}^d \left( \text{diag}(0, a) + \sum_{i=1}^m x_i \alpha_i \right) \frac{\partial^2 f(x)}{\partial x_k \partial x_l} + (b + \beta x)^\top \nabla f(x) + \int_D (f(x + \xi) - f(x) - \chi_J(\xi)^\top \nabla_J f(x)) m(d\xi) + \int_D (f(x + \xi) - f(x)) \left( \sum_{i=1}^m x_i \mu_i(d\xi) \right)
\]

for some appropriate positive semidefinite \( n \times n \)- and \( d \times d \)-matrices \( a \) and \( \alpha_i \), respectively. Here, with \( \text{diag}(0, a) \) we denote the block-diagonal \( d \times d \)-matrix with blocks given by the \( m \times m \)-zero matrix and \( a \). Moreover, \( \chi_J(\xi) \) denotes an \( \mathbb{R}^n \)-valued continuous and bounded truncation function with \( \chi_J(\xi) = \xi_J \) in a neighborhood of the origin \( \xi = 0 \). For detailed parametric restrictions on \( (a, \alpha_i, b, \beta, m, \mu_i) \) we refer the reader to Duffie et al. (2003, Definition 2.6). We assume for simplicity\(^9\) that the jump measures \( \mu_i \) are of finite variation type with integrable large jumps \( \int_D \|\xi\| \mu_i(d\xi) < \infty, \quad i = 1, \ldots, m \).

4.1. Affine Transform Formula. The analytical tractability of affine models stems from the fact that the characteristic function of \( X_t | X_0 = x \) is explicitly given by the affine transform formula

\[
\mathbb{E}\left[e^{iu^\top X_t} \mid X_0 = x\right] = e^{\phi(t, iu) + \psi(t, iu)^\top x}, \quad u \in \mathbb{R}^d, \quad x \in \mathcal{D}
\]

where the \( \mathbb{C}_- \) - and \( \mathbb{C}_+ \times i\mathbb{R}^n \)-valued functions \( \phi = \phi(t, iu) \) and \( \psi = \psi(t, iu) \) solve the generalized Riccati equations, for \( i = 1, \ldots, m \),

\[
\begin{align*}
\partial_t \phi &= \psi_J^\top a \psi_J + b^\top \psi + \int_D \left( e^{\psi^\top \xi} - 1 - \psi_J^\top \chi_J(\xi) \right) m(d\xi), \quad \phi(0) = 0, \\
\partial_t \psi_i &= \psi_i^\top \alpha_i + \psi + \int_D \left( e^{\psi^\top \xi} - 1 \right) \mu_i(d\xi), \quad \psi_i(0) = iu_i, \\
\partial_t \psi_J &= \mathcal{B}_J \psi_J, \quad \psi_J(0) = iu_J,
\end{align*}
\]

where we define \( \mathcal{B} = \beta^\top \) and write \( \mathcal{B}_i \) for the \( i \)th row vector of \( \mathcal{B} \). Obviously, we have \( \psi_J(t, iu) = i e^{B_J t} u_J \), and \( \phi(t, iu) \) is given by simple integration of the right hand side of its equation.

\(^9\)At the cost of more technical analysis, the following results could also be proved for the general case of infinite variation jumps \( \mu_i \) with infinite tail mean.
4.2. Polynomial Moments. It is well known that if \( X_t|X_0 = x \) has finite \( k \)-th moment,

\[
\mathbb{E} \left[ \|X_t\|^k \mid X_0 = x \right] < \infty \quad \text{for all } x \in D,
\]

then \( \phi(t, u) \) and \( \psi(t, u) \) are of class \( C^k \) in \( u \). Moreover, polynomial moments are explicitly given in terms of the respective mixed derivatives of the characteristic function for \( |\alpha| \leq k \), see e.g. Duffie et al. (2003, Lemma A.1). Consequently they are a real polynomial in \( x \) of degree less than or equal to \( |\alpha| \). Recently, generalizing the recursive method used in Forman and Sørensen (2008) for Pearson-type diffusions, Cuchiero et al. (2010b) proposed an alternative method to compute the coefficients of this polynomial. The idea rests on the insight that the affine generator \( A \) formally maps \( P_k \) into \( P_k \), where \( P_k \) denotes the finite-dimensional linear space of all polynomials in \( x \in \mathbb{R}^d \) of degree less than or equal to \( k \).

The generator \( A \) thus restricts to a linear operator \( A_k \) on \( P_k \). Consequently, we obtain the formal representation

\[
\mathbb{E} \left[ X_t^\alpha \mid X_0 = x \right] = e^{A_k t \cdot x^\alpha}
\]

where \( e^{A_k t} = \sum_{j \geq 0} \frac{(A_k t)^j}{j!} \) is the exponential of \( A_k t \). This can be expressed as a matrix. We shall illustrate this for \( d = 1 \). The dimension of \( P_k \) then equals \( k + 1 \), and we can pick as canonical basis of \( P_k \) the set \( Q = \{1, x, \ldots, x^k\} \). For every \( j = 0, \ldots, k \) we then calculate symbolically the coefficients \( q_{ij} \) in

\[
(4.3) \quad A_k x^j = Ax^j = \sum_{i=0}^{k} q_{ij} x^i.
\]

Hence \( A_k \) can be represented by the upper-triangular matrix \( Q = (q_{ij}) \) with respect to the basis \( Q \). In other words, if we identify a generic polynomial \( p(x) = \sum_{i=0}^{k} p_i x^i \) in \( P_k \) with the vector of its coefficients \( p = (p_0, \ldots, p_k)^\top \), then \( A_k p(x) \in P_k \) equals the polynomial with coefficient vector \( Qp \). Moreover,

\[
(4.4) \quad \mathbb{E} \left[ p(X_t) \mid X_0 = x \right] = e^{Qt} p.
\]

See Sections 7 and 8 below for some concrete applications for \( d = 1 \) and \( d = 2 \).

4.3. Existence and Properties of Affine Transition Densities. In this section we present our main theoretical results, which establish existence and smoothness of the density of the conditional distribution of the affine process \( X \). Moreover, we provide explicitly verifiable conditions asserting that Lemmas 3.3–3.5 apply.

Our first result provides sufficient conditions for the existence and smoothness of a density of \( X_t|X_0 = x \), for some \( t > 0 \) and \( x \in \mathbb{R}^m \times \mathbb{R}^n \). These easy to check conditions apply in particular to multi-factor affine term-structure models on \( \mathbb{R}_+^m \times \mathbb{R}^n \) from Duffie and Kan (1996) and Dai and Singleton (2000), and Heston’s stochastic volatility model. The proof is given in the Internet Appendix C.

\(^{10}\)This method is not restricted to affine processes, but can be defined for any Markov process with finite \( k \)-th moments, and whose infinitesimal generator maps \( P_k \) into itself.
Lemma 5.3). From Duffie et al. (2003, Theorem 2.16 and Lemma 6.5) we then infer that

\[ \mathcal{K} = \left[ \sum_{i=1}^{m} \alpha_i, \text{diag}(0, a), \text{diag}(0, B_{fJ}^{-1}a), \ldots, \text{diag}(0, (B_{fJ}^{-1})^{T}a) \right] \]

has full rank. Further, let \( p \) be a nonnegative integer with

\[ p < \min_{i \in \{1, \ldots, m\}} \frac{b_i}{\alpha_{i,ii}} - 1. \]

Then \( X_t|X_0 = x \) admits a density \( g(\xi) \) of class \( C^p \) with support in \( \mathbb{R}^m_+ \times \mathbb{R}^n \) and the partial derivatives of \( g(\xi) \) of orders \( 0, \ldots, p \) tend to 0 as \( \|\xi\| \to \infty \).

We note that condition (4.6) is sharp and cannot be relaxed in general. Consider for instance the scalar square-root diffusion \( X \) on \( \mathbb{R}_+ \) with generator \( A_f(x) = \alpha x f''(x) + b f'(x) \). It is well known that for any parameter values \( \alpha > 0 \) and \( b \geq 0 \), the distribution of \( \frac{2X_t}{\alpha t} | X_0 = x \) is noncentral \( \chi^2 \) with \( \frac{2b}{\alpha} \) degrees of freedom and noncentrality parameter \( \frac{2x}{\alpha} \), (see, for instance Filipović, 2009, Exercise 10.9). The corresponding density function \( g(\xi) \) satisfies \( \lim_{\xi \to 0} g(\xi) = 0 \), and is therefore of class \( C^0 \), if and only if the degrees of freedom \( \frac{2b}{\alpha} \geq 2 \), see Johnson et al. (1995, Chap. 29). This is exactly what condition (4.6) states for \( p = 0 \). As regards exponential moments of \( X_t|X_0 = x \), we combine and rephrase some results from Duffie et al. (2003):

**Theorem 4.2.** Assume that the jump measures admit exponential moments

\[ \int_{\|\xi\| > 1} e^{q^T \xi} m(d\xi) < \infty \quad \text{and} \quad \int_{\|\xi\| > 1} e^{q^T \xi} \mu_i(d\xi) < \infty, \quad i = 1, \ldots, m \]

for all \( q \) in some open neighborhood \( V \) of 0 in \( \mathbb{R}^d \). Then the right hand side of (4.2) is analytic in \( \psi \in V \). Suppose further that (4.2) admits a \( V \)-valued solution \( \psi(t, u) \) with \( \psi(0, u) = u \) for all \( t \in [0, T] \) and for all \( u \) in \([-\epsilon_1, \epsilon_1]^m \times [-\epsilon_2, \epsilon_2]^n \), for some \( \epsilon_1, \epsilon_2 > 0 \). Then \( X_t|X_0 = x \) has a finite exponential moment

\[ E \left[ e^{q^T Y_t + \sum_{i=1}^{m} \epsilon_i |a_{ii}|} Z_t \bigg| X_0 = x \right] < \infty \]

for all \( t \in [0, T] \), where we denote \( Y_t = (X_{1,t}, \ldots, X_{m,t})^T \) and \( Z_t = (X_{m+1,t}, \ldots, X_{d,t})^T \).

**Proof.** That the right hand side of (4.2) is analytic in \( \psi \in V \) follows from Duffie et al. (2003, Lemma 5.3). From Duffie et al. (2003, Theorem 2.16 and Lemma 6.5) we then infer that

\[ E \left[ e^{q^T X_t} \bigg| X_0 = x \right] < \infty \]

for all \( t \in [0, T] \) and for all \( q \in [-\epsilon_1, \epsilon_1]^m \times [-\epsilon_2, \epsilon_2]^n \). Combining this with the elementary inequality

\[ e^{q_1 Y_t + \sum_{i=1}^{m} |a_{ii}| + \sum_{i=m+1}^{d} |a_{i+1}|} \leq \sum_{|\alpha| = 0}^{1} e^{q_0^T X_t}, \]

where we denote \( q_0 = ((-1)^{\alpha_1} \epsilon_1, \ldots, (-1)^{\alpha_m} \epsilon_1, (-1)^{\alpha_{m+1}} \epsilon_2, \ldots, (-1)^{\alpha_d} \epsilon_2)^T \in [-\epsilon_1, \epsilon_1]^m \times [-\epsilon_2, \epsilon_2]^n \), proves (4.7). \( \square \)

\( ^{11} \)Here, for given \( d \times d \)-matrices \( B_1, B_2, \ldots, B_n \) the expression \( [B_1, B_2, \ldots, B_n] \) denotes the \( d \times nd \)-block matrix we obtain by putting the matrices next to each other.
Note that if \( m(d\xi) \) and \( \mu_i(d\xi) \) have light tails of the order \( e^{-r\|\xi\|^2} d\xi \) for some \( r > 0 \), or have compact support in particular, then the first assumption of Theorem 4.2 is satisfied for \( V = \mathbb{R}^d \). Even then, however, the solution \( \psi(t,u) \) exists only on a finite time horizon \( t < T < \infty \) for any nonzero \( u \in \mathbb{R}^d \) in general. We refer to the discussion of the diffusion case in Filipović and Mayerhofer (2009), see also Filipović (2009, Chapter 10).

We further present an additional result which concerns the existence of the marginal transition density of integrated affine jump-diffusions, which are not covered by Theorem 4.1. If \( X \) is a one-dimensional affine process on \( \mathbb{R}_+ \), then the two-dimensional process \( (dX, X dt)^\top \) is affine again, with state space \( \mathbb{R}_+^2 \). However, its diffusion matrix is degenerate and thus violates the conditions of Theorem 4.1. Nevertheless, a slight adaption of its proof yields the existence of the marginal transition density of the integrated process \( \int X dt \) under some more stringent conditions, completing earlier results on the marginal, unconditional density in Wong (1964). The proof of the following theorem is given in the Internet Appendix D.

**Theorem 4.3.** Let \( X \) be an \( \mathbb{R}_+ \)-valued affine process with parameters \((a = 0, \alpha, b, \beta, m, \mu)\). Further, let \( p \) be a nonnegative integer with

\[
p < \frac{b}{2\alpha} - 1.
\]

Then \( \int_0^t X_s ds \mid X_0 = x \) admits a density \( g(\xi) \) of class \( C^p \) with support in \( \mathbb{R}_+ \) and the partial derivatives of \( g(\xi) \) of orders \( 0, \ldots, p \) tend to 0 as \( \xi \to \infty \).

From an application point of view, we can rephrase the statements of the preceding theorems as follows:

**Corollary 4.4.** Theorems 4.1–4.3 provide conditions in terms of the parameters of the affine process \( X \) such that the assumptions in Lemmas 3.3–3.5, and thus eventually the validity of Assumptions 1 and 2, can explicitly be verified for the density of the (marginal) transition distributions of \( X \).

### 5. Examples of Auxiliary Density Functions

For applications of the polynomial density approximations to affine models we are free to choose any auxiliary density function \( w \) as \( L^2 \) weight, limited only by the requirements of Assumptions 1 and 2. For Lévy-driven stochastic differential equations, Schaumburg (2001) uses the densities of the driving Lévy processes. For affine processes there is no such obvious candidate in general. We therefore pursue an approach with auxiliary densities with exponential tails, which are reasonably easy to evaluate, for which the moment-generating function is available in closed-form, and with sufficiently many parameters to render the \( L^2 \) weight as close as possible to the true, unknown transition density. For positive coordinates we use the Gamma density

\[
\gamma(\xi; D) = \frac{e^{-\xi \xi^D}}{\Gamma[1 + D]}
\]

of a \( \Gamma(1 + D, 1) \)-distributed random variable. Here, \( \Gamma[\cdot] \) denotes the Gamma function. It is easily seen that conditions (3.1)–(3.4) are satisfied for the appropriate parameters \( p, \epsilon_0 \) and \( \epsilon_1 \).

For real-valued coordinates we employ the bilateral Gamma density from Küchler and Tappe (2008a). The corresponding family of distributions nests, for example, the Variance Gamma distribution as a special case. It has very flexible shapes (Küchler and Tappe, 2008b). For the
purposes of this paper we make use of a constrained, standardized version with mean centered at zero, unit variance, zero skewness, and excess kurtosis $C > 0$. We denote this standardized bilateral Gamma distribution by $\Gamma_b(C)$. Its characteristic function is given by

$$\Phi_{\Gamma_b}(u; C) = 216 \frac{1}{C} \left( \frac{1}{6 - Cu^2} \right)^{3/C}, \quad u \in \mathbb{R}, \quad C \in \mathbb{R}_+.$$ 

The corresponding density is

$$\gamma_b(\xi; C) = \frac{2^{\frac{3(C-2)}{4C}} 3^{\frac{C+6}{4C}} C^{-\frac{C+6}{4C}} |\xi|^{\frac{3}{2} - \frac{1}{2} K} \frac{1}{\sqrt{\pi}} \Gamma \left( \frac{3}{C} \right)}{\Gamma \left( \frac{3}{C} \right)},$$

where $K_n(\xi)$ denotes the modified Bessel function of the second kind (see Küchler and Tappe, 2008b, eq. (3.6)). It follows from Küchler and Tappe (2008b, Section 6) that conditions (3.1)–(3.4) are satisfied for the appropriate parameters $\epsilon_0$ and $\epsilon_2$. The special case with excess kurtosis $C = 1/3$ leads to the following simple expression for the density $\gamma_b(\xi) = \gamma_b(\xi; C = 1/3)$, since for half-integer indices the modified Bessel functions evaluate to elementary functions

$$\gamma_b(\xi) = \frac{27 e^{-3\sqrt{2}|\xi|}}{1146880 \sqrt{2}} \left( 7776 \sqrt{2} |\xi|^7 + 83160 \sqrt{2} |\xi|^5 + 180180 \sqrt{2} |\xi|^3 ight.$$

$$\left. + 75075 \sqrt{2} |\xi| + 1296 \xi^8 + 45360 \xi^6 + 207900 \xi^4 + 210210 \xi^2 + 25025 \right).$$

Orthonormal polynomial bases can be constructed for any auxiliary density function $w$ which has finite exponential moment (3.1) by the Gram-Schmidt process, which is also used in the proof of Lemma 3.1.

**Algorithm 5.1** (Gram-Schmidt Process).

\begin{align*}
H_0 &= 1, \\
\tilde{H}_\alpha &= \xi^\alpha - \sum_{0 \leq |\beta| \leq |\alpha|, \beta \neq \alpha} \langle \xi^\alpha, H_\beta(\xi) \rangle_{L^2_w} H_\beta(\xi), \\
H_\alpha &= \frac{\tilde{H}_\alpha}{\| \tilde{H}_\alpha(\xi) \|_{L^2_w}} \quad \text{(normalization)}.
\end{align*}

Note that $\deg H_\alpha = \deg \xi^\alpha = |\alpha|$, which is due to the linear independence of the set of monomials $\{ \xi^\alpha \mid \alpha \in \mathbb{N}_0^d \}$. Below are the first five orthonormal polynomials for the Gamma and the bilateral Gamma densities $\gamma$ and $\gamma_b$ introduced above.
Example 5.1. The non-normalized orthogonal polynomials for the Gamma density $\gamma$ are the generalized Laguerre polynomials, the first five of which are
\[\bar{H}_{0}^{\gamma}(\xi) = H_{0}^{\gamma}(\xi) = 1,\]
\[\bar{H}_{1}^{\gamma}(\xi) = -\xi + D + 1,\]
\[\bar{H}_{2}^{\gamma}(\xi) = \frac{1}{2}(\xi^2 - 2\xi(D + 2) + D^2 + 3D + 2),\]
(5.4)
\[\bar{H}_{3}^{\gamma}(\xi) = \frac{1}{6}(-\xi^3 + 3\xi^2(D + 3) - 3\xi(D^2 + 5D + 6) + D^3 + 6D^2 + 11D + 6),\]
\[\bar{H}_{4}^{\gamma}(\xi) = \frac{1}{24}(\xi^4 - 4(D + 4)\xi^3 + 6(D + 3)(D + 4)\xi^2 - 4(D + 2)(D + 3)(D + 4)\xi + (D + 1)(D + 2)(D + 3)(D + 4)).\]
The normalization constants are given by
(5.5)
\[HO_{n}^{\gamma} = \left\| \bar{H}_{n}^{\gamma}(\xi) \right\|_{L_{\gamma}^2} = \sqrt{\prod_{i=1}^{n}(i + D)/n!}.\]
Example 5.2. For the standardized bilateral Gamma density $\gamma_b$ in (5.2), the first five non-normalized orthogonal polynomials are
\[\bar{H}_{0}^{\gamma_b}(\xi) = H_{0}^{\gamma_b}(\xi) = 1,\]
\[\bar{H}_{1}^{\gamma_b}(\xi) = \xi,\]
\[\bar{H}_{2}^{\gamma_b}(\xi) = \xi^2 - 1,\]
(5.6)
\[\bar{H}_{3}^{\gamma_b}(\xi) = (-C - 3)\xi + \xi^3,\]
\[\bar{H}_{4}^{\gamma_b}(\xi) = \frac{-2(5C^2 + 21C + 18)(\xi^2 - 1)}{3(C + 2)} - C + \xi^4 - 3,\]
and the corresponding normalization constants $HO_{n}^{\gamma_b} = \left\| \bar{H}_{n}^{\gamma_b}(\xi) \right\|_{L_{\gamma_b}^2}$ are given by
\[HO_{0}^{\gamma_b} = 1,\]
\[HO_{1}^{\gamma_b} = 1,\]
\[HO_{2}^{\gamma_b} = \sqrt{C + 2},\]
(5.7)
\[HO_{3}^{\gamma_b} = \sqrt{\frac{7C^2}{3} + 9C + 6},\]
\[HO_{4}^{\gamma_b} = \sqrt{\frac{2(55C^4 + 363C^3 + 822C^2 + 756C + 216)}{9(C + 2)}}.\]
Example 5.3 (Product Measure). Define the product density $w_{\gamma_{\gamma}}$ with support on $\mathbb{R}_{+} \times \mathbb{R}$ by
\[w_{\gamma_{\gamma}}(\xi_{1}, \xi_{2}; C, D) = \gamma(\xi_{1}, D)\gamma_b(\xi_{2}, C)\]
with Gamma density $\gamma$ defined in (5.1) and bilateral Gamma density $\gamma_b$ defined in (5.2). Combining Lemma 3.2 and Examples 5.1 and 5.2, we obtain the corresponding orthonormal basis of polynomials $\{H_{n_{1}}^{\gamma_{\gamma}} \cdot H_{n_{2}}^{\gamma_{b}} \mid (n_{1}, n_{2}) \in \mathbb{N}_{0}^2\}$. 
6. Explicit Example

In the present section we illustrate the necessary steps to engineer a density expansion to order \( J = 4 \). We adopt notation used conventionally in finance and econometrics. In particular we deviate from Duffie et al. (2003) notation. From here onward the time interval between observations is denoted by \( \Delta \).

Consider as an example the Basic Affine Jump-Diffusion process (BAJD) \( Y \) with domain \( D = \mathbb{R}_+ \), solving the SDE

\[
dY_t = (\kappa \theta - \kappa Y_t) \, dt + \sigma \sqrt{Y_t} \, dW_t + dL_t.
\]

Here, \( L \) is a compound Poisson process with jump intensity \( l \geq 0 \), and the expected jump size of the exponentially distributed jumps is \( \nu \geq 0 \). We collect the set of parameters in \( \varrho_Y = \{ \kappa \theta, \kappa, \sigma, \nu, l \} \). The explicit form of the transition density \( g_Y(y | y_0, \Delta, \varrho_Y) \) of \( Y \, | \, Y_0 = y_0 \) is unknown, but Theorem 4.1 ensures its existence if \( 2\kappa \theta > \sigma^2 \).

6.1. Transformation. As a first step we linearly transform the random variable \( Y \) for an approximation with the Gamma density \( w = \gamma(\xi, D) \) from (5.1) as \( L \) weight. In view of Lemma 2.2, we aim at matching the first two moments of the scaled \( Y \) and \( \gamma(\xi, D) \). Note that the mean and variance of \( \gamma(\xi, D) \) equal \( D + 1 \). The unique linear transform of \( Y \) which preserves the state space \( D = \mathbb{R}_+ \), and renders mean and variance equal to some value \( D + 1 \) turns out to be

\[
\bar{Y}_\Delta := Y_\Delta \mu_1 / \mu_2 - \mu_1 - 1, \quad \mu_1 = \mathbb{E}[Y_\Delta | Y_0 = y_0, \varrho_Y].
\]

We write \( g_{\bar{Y}} \) for the density of \( \bar{Y}_\Delta \).

6.2. Computing the Coefficients of the Expansion. For the computation of the coefficients \( c_j^\gamma \) from (2.1) corresponding to the \( \gamma(\xi, D) \) distribution, no approximations are necessary. Using the polynomial property of affine processes described in Section 4 we can obtain conditional moments in closed form. Taking recourse to formula (4.4) and acknowledging that the infinitesimal generator of the BAJD is

\[
A f(x) = (\kappa \theta - \kappa x) \frac{\partial f(x)}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 f(x)}{\partial x^2} + l \int_{\mathbb{R}_+} (f(x + \xi) - f(x)) e^{-\xi} \nu d\xi,
\]

we first populate the matrix \( Q = (q_{ij}) \) according to (4.3) relative to the canonical basis \( \{1, x, x^2, x^3, x^4\} \) for the first four conditional moments (for \( J = 4 \))

\[
Q = \begin{pmatrix}
0 & \kappa \theta + l\nu & 2l\nu^2 & 6\nu^3 & 24\nu^4 \\
0 & -\kappa & \sigma^2 + 2\kappa \theta + 2l\nu & 6\nu^2 & 24\nu^3 \\
0 & 0 & -2\kappa & 3\sigma^2 + 3\kappa \theta + 3l\nu & 12\nu^2 \\
0 & 0 & 0 & -3\kappa & 6\sigma^2 + 4\kappa \theta + 4l\nu \\
0 & 0 & 0 & 0 & -4\kappa
\end{pmatrix}.
\]

Note the upper-triangular form. A symbolic mathematics software package such as Mathematica or Maple will be able to compute the matrix exponential \( e^{Qt} \) in closed form. The conditional moments \( \mu_n(y_0, \Delta, g_Y) = \mathbb{E}[Y^n_\Delta | Y_0 = y_0, g_Y] \) may then be obtained by plugging into formula (4.4). Below we will suppress dependence on \( y_0, \Delta, g_Y \) to lighten notation.
Using the corresponding orthogonal polynomials \( \tilde{H}_n^\gamma \) from (5.4) we obtain the coefficients of the density approximation. For each \( j \geq 0 \) we have
\[
c_j^\gamma = \frac{\langle g^\gamma \tilde{H}_j^\gamma \rangle_{L^2_\gamma}}{\|H_j^\gamma\|_{L^2_\gamma}} = \frac{\int_D \tilde{H}_j^\gamma \left( \frac{y_\mu_1}{\mu_2^2-\mu_1^2} \right) g_Y(y \mid y_0; \Delta) dy}{\tilde{H}_j^\gamma},
\]
where \( HO_j^\gamma \) is given in (5.5). In particular, the first five coefficients are of the explicit form
\[
c_0^\gamma = 1, \quad c_1^\gamma = 0, \quad c_2^\gamma = 0, \\
c_3^\gamma = \frac{(D + 1) (D + 2)(D + 3) - (D + 1)^2 \mu_1^2}{\sqrt{6} \sqrt{(D + 1)(D + 2)(D + 3)}}, \\
c_4^\gamma = \frac{(D + 1)^4 \mu_4 + (D + 4)(D + 1) \mu_1 (3(D + 2)(D + 3) \mu_1^2 - 4(D + 1)^2 \mu_3)}{2 \sqrt{6} \sqrt{(D + 1)(D + 2)(D + 3)(D + 4) \mu_1^2}}.
\]
Note that due to the chosen scaling, the first and second order terms of the deforming polynomial (the pseudo likelihood ratio) vanish as stated by Lemma 2.2.

### 6.3. Verification of Assumption 2

By Theorem 4.1, \( g_Y \) and therefore \( \bar{g}_Y \) are of class \( C^p \) for the greatest nonnegative integer \( p \) satisfying
\[
p < \frac{2\kappa \theta}{\sigma^2} - 1.
\]
On the other hand, using Theorem 4.2 one can verify numerically, by solving the corresponding Riccati differential equations, that
\[
\mathbb{E}[e^{\bar{Y}_{\Delta}}] = \mathbb{E}[e^{(D+1) Y_{\Delta}}] < \infty.
\]
This implies finite polynomial moments of \( g_Y \) and \( g_{\bar{Y}} \), and therefore justifies the calculations in Steps 6.1 and 6.2. Note that the Gamma density \( w(\xi) = \gamma(\xi; D) \) satisfies \( \sup_{x \in [0,1]} x^D / w(x) < \infty \) and \( \sup_{x \geq 1} e^{-x} / w(x) < \infty \). In view of (6.5), Lemma 3.4 implies validity of Assumption 2, that is \( g_{\bar{Y}} / w \in L^2_w \), once \( D \leq 2p \). By (6.4), the latter holds if and only if\(^{12}\)
\[
[D/2] < \frac{2\kappa \theta}{\sigma^2} - 1,
\]
which again can easily be checked numerically.

### 6.4. Putting Everything Together

Accounting for the change of variable \( \bar{y}(y) = \frac{y \mu_1}{\mu_2 - \mu_1^2} \), the density proxy equals
\[
g_Y^{(4)}(y \mid y_0, \Delta, g_Y) = \gamma(\bar{y}) \left( 1 + \sum_{j=1}^4 c_j^\gamma H_j^\gamma(\bar{y}) \right) \cdot \frac{\mu_1}{\mu_2 - \mu_1^2}.
\]
\(^{12}[x] \) denotes the smallest integer which is greater than or equal to \( x \).
The coefficients $c_j^\varphi$ and the orthonormal polynomials depend on $y_0$, $\Delta$, and $\varphi_Y$, but they are available in closed form, and the expansion above therefore lends itself to rapid numeric evaluation.

7. Relation to Existing Approximations

In this section we recall facts about closed-form density approximations from previous literature and relate them to the density expansions of the present paper. A short summary of the capabilities and limitations of the different methods is reported in Table 1 below. The closest methodology to the one introduced in Section 2 is Aït-Sahalia (2002) (AS02) and Chapter 1 of Schaumburg (2001). To illustrate the differences between AS02 and our approach we devote the next section to a direct comparison reusing the BAJD example developed in the previous Section 6.

7.1. Direct Comparison to Aït-Sahalia (2002). Since the AS02 approach is designed for diffusions we will assume that $l = 0$ in this section.

7.1.1. Transformation. AS02 applies two transformations such that a Gaussian-weighted $L^2$ expansion converges. The first change of variables yields a unit diffusion process through the Lamperti transform. This step introduces nonlinearities into the drift. The resulting process is then centered and scaled in time in a second step. The two transformations give

$$Y_0^\diamond := \frac{g(Y_\Delta) - g(Y_0)}{\sqrt{\Delta}},$$

where we use $y(x) = \int^x \frac{du}{\sigma \sqrt{\pi}}$. The range of $Y_0^\diamond$ is $\mathbb{R}$, while the scaling transformation in (6.2) preserves that $\tilde{Y}_\Delta$ lives on $\mathbb{R}_+$. Nevertheless we will see below that the resulting density expansions are of a similar functional form, since the approximation of the true, unknown transition density $g_{y_0}(z \mid y_0, \Delta)$ is engineered in terms of a Gaussian-weighted $L^2$ expansion. We denote the density of the standard normal distribution $\varphi$, and the corresponding Hermite polynomials by $H_{\varphi}^\varphi$.

7.1.2. Computing the Coefficients of the Expansion. Computing the coefficients, analogously to (6.3), of the Hermite expansion we obtain

$$(7.1) \quad c_j^\varphi = \frac{1}{j!} \mathbb{E} \left[ H_j^\varphi(Y_0^\diamond) \mid Y_0 = y_0 \right].$$

It follows that the Hermite polynomials have to be evaluated in conjunction with a nonlinear transform of $Y$. In practice this is handled via Monte Carlo integration, or via a Taylor expansion of the form

$$(7.2) \quad \mathbb{E} \left[ f(Y_\Delta) \mid Y_0 = y_0 \right] = \sum_{k=0}^K A_k f(y_0) \frac{\Delta^k}{k!} + \mathbb{E} \left[ A^{K+1} f(Y_\Delta) \mid Y_0 = y_0 \right] \frac{\Delta^{K+1}}{(K + 1)!},$$

which is valid for any smooth test function $f$, provided that $Y$ is stationary (see eqn. (4.3) and Proposition 4 in AS02). This step represents a key difference to our approach, where the coefficients $c_j^\gamma$ of the density expansion are explicit and in closed form in terms of the moments of $Y$.

$^{13}$This is also suggested by (Schaumburg, 2001, Chapter 1.5).
The AS expansion is finally obtained by changing variables as follows

\[
g_{Y}(y | y_0, \Delta) = \frac{\Delta^{-1/2}}{\sigma \sqrt{y}} \varphi \left( \frac{y(y) - y(y_0)}{\sqrt{\Delta}} \right) \left( 1 + \sum_{j=1}^{J} c_{j,K} H_j^{J}(y(y)) \right).
\]  

Figure 1 shows the percentage deviation of the AS02 expansion and the expansions developed in this paper from the true non-central \( \chi^2 \) density of the square-root process. The comparison is engineered for the orders \( J = 2 \) and \( J = 4 \), and \( K = 3 \). It can be seen that the AS02 expansion converges much faster than the approximations developed here, owing to the AS02 standardization which appears to be working better in particular in the tails. This suggests that for univariate applications with diffusion processes the AS02 expansion is the preferred
choice. However, we note that the expansions developed in this paper can be applied to affine multivariate applications with jumps, which the AS02 expansion is not capable of.

7.2. Relation to Other Density Approximations. In the multivariate case a Lamperti transform is rarely possible, since most applications call for stochastic volatility models which are irreducible in the sense of Footnote 2 (Aıt-Sahalia, 2008, Proposition 1). An entirely different strategy is therefore pursued in Aıt-Sahalia (2008) for the irreducible multivariate case, where the log likelihood is expanded in, both, time, and space, so that the coefficients of the expansion may be computed from the Kolmogorov forward and backward equations. This approach is adopted by Yu (2007) with the difference that he also considers jump-diffusions and approximates the transition density itself, rather than the log transition density. The saddlepoint approach in Aıt-Sahalia and Yu (2006) (approximately) solves the Fourier inversion problem by expanding the cumulant generating function about the saddlepoint\textsuperscript{14}, rather than making use of the Kolmogorov forward and backward equations. The maintained assumption here is that the cumulant generating function is available, even though for diffusions Aıt-Sahalia and Yu (2006, Section 4) circumvent this problem by using a Taylor series expansion for small times along the lines of Aıt-Sahalia (2002) for nonlinear moments. Though the saddlepoint approach and this paper both facilitate expansion techniques, the objects of the expansion are different and the formulae are unrelated. Saddlepoint approximations are extremely accurate even for low orders (Aıt-Sahalia and Yu, 2006, Fig. 2). The price to be paid for this precision is the computational burden of having to solve numerically for the saddlepoint for every pair of forward and backward variables.

Schaumburg (2001) considers scalar Markov processes given as solutions of Lévy driven stochastic differential equations with constant volatility function, and general drift function. For a linear drift specification, the Schaumburg (2001) approach is fully nested in our framework. In particular, he uses the same $\mathcal{L}^2$ approximation method as in our paper, with auxiliary density function given by the transition density of the driving Lévy process. However, as the focus in Schaumburg (2001) is mostly on the scalar and constant volatility case, it is not obvious how his results would carry over to the general multivariate affine case that we consider in our paper.

\textsuperscript{14}For a stochastic process $X$ denote by $K(t, u \mid x_0) = \log \mathbb{E}[e^{u^T X_t} \mid X_0 = x_0]$ the cumulant generating function. Suppose $X_t \mid X_0 = x_0$ has an absolutely continuous law. For any state $x$ the saddlepoint is defined as the solution $\hat{u} = \hat{u}(t, x, x_0)$ in $u$ to the implicit equation $\partial_u K(t, u \mid x_0) = x$. 

| Approximations | S01 | AS02 | ASY05 | Y07 | AS08 | this paper |
|----------------|-----|------|-------|-----|------|------------|
| multivariate   | No  | No   | Yes   | Yes | Yes  | Yes        |
| everywhere positive | No  | No   | No    | Yes | No   | No         |
| integrates to one      | Yes | Yes  | No    | No  | No   | Yes        |
| jumps              | Yes | No   | Yes   | No  | No   | Yes        |

Table 1. Comparison of Closed-Form Transition Density Approximations: S01 refers to Schaumburg (2001), AS02 Aıt-Sahalia (2002), ASY05 to Aıt-Sahalia and Yu (2006), Y07 to Yu (2007), and AS08 to Aıt-Sahalia (2008).
In the following we present applications which highlight the usefulness of the transition density approximations developed in this paper. For the empirical investigations considered below we find that there is a trade-off between numerical accuracy and the order of the expansion. Higher-order expansions may perform worse than low-order expansions due to numerical errors that are induced by the limited numeric precision of the computer environment in representing very large or very small numbers. As a general guideline we suggest matching as many moments (cumulants) as possible when choosing $L^2$ weights, and stopping the expansion at a relatively low order such as $J = 4$. We strongly recommend checking the above theoretical foundations for the validation of Assumptions 1 and 2 in numerical applications, as outlined in Section 6 at the example for the BAJD.

8.1. **Heston’s Model.** The Heston (1993) stochastic variance model has been particularly used for the pricing of equity (index) options. The model for the log stock price $X$ and its stochastic variance $V$ can be realized as solution of the following SDE

$$
\begin{align*}
\frac{dV_t}{V_t} &= \left(\kappa \theta - \kappa V_t\right) dt + \sigma \sqrt{V_t} dW^V_t, \\
\frac{dX_t}{V_t} &= \left(\kappa \theta - \frac{1}{2} \kappa V_t\right) dt + \sqrt{V_t} \left(\rho dW^V_t + \sqrt{1 - \rho^2} dW^X_t\right),
\end{align*}
$$

with $(W^V, W^X)$ being a two-dimensional standard Brownian motion. The domain $\mathcal{D}$ of the process equals $\mathbb{R}_+ \times \mathbb{R}$. With $2\kappa \theta > \sigma^2$ and $|\rho| < 1$ Theorem 4.1 guarantees existence of transition densities. Note that it would be perfectly possible to enrich Heston’s model above with jumps in both factors (this has been done for example in Duffie et al. (2000), Eraker et al. (2003), and Eraker (2004)), to multiple variance factors, or even a matrix-valued variance process as in (Da Fonseca et al., 2008).

For the bivariate Heston model, to compute conditional moments up to order two using formula (4.4), the canonical basis is given by $\{1, v, x, v^2, vx, x^2\}$ and the corresponding $Q$ matrix from (4.3) is

$$
Q = \begin{pmatrix}
0 & \kappa \theta & \kappa \theta_x & 0 & 0 & 0 \\
0 & -\kappa & -\frac{1}{2} & \sigma^2 + 2\kappa \theta & \kappa \theta_x + \rho \sigma & 1 \\
0 & 0 & 0 & 0 & \kappa \theta & 2\kappa \theta_x \\
0 & 0 & 0 & -2\kappa & -\frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & -\kappa & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.
$$

**Example 8.1 (Standardizing and Scaling the Heston Model).** Our goal is to work within a $L^2$ space weighted with a product measure $w : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}_+$,

$$
w(\xi, \eta)d(\xi, \eta) = \gamma(\xi)d\xi \cdot \gamma_b(\eta)d\eta,
$$

composed of $\Gamma(D + 1, 1)$ and $\Gamma_b(C)$ densities, from definitions (5.1) and (5.2), respectively. In this space, we may use the polynomials from eqs. (5.4) and (5.6). Lemma 3.2 ensures that the product of the polynomials forms an ONB in the $L^2_w$ space. Acknowledging Lemma 2.2 we want to make sure that we match as many moments as possible to optimize the quality of the approximation. Below we show how this transformation $\varsigma$ is composed.
We introduce lighter notation by defining $U_t = (V_t, X_t)$ and for the first two moments of $(V_t, X_t)$
\[
\mathbb{E}[U_t \mid U_0 = u, \varphi_{VX}] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \text{and} \quad \mathbb{V}[U_t \mid U_0 = u, \varphi_{VX}] = \begin{pmatrix} a_1 & b \\ b & a_2 \end{pmatrix}.
\]
For demeaning and block-diagonalizing define $\varsigma_1(u) = Y_1 u + v_1$ where
\[
Y_1 = \begin{pmatrix} 1 \\ -b/a_1 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 0 \\ b\mu_1/a_1 - \mu_2 \end{pmatrix}.
\]
Then $\varsigma_1(U_t)$ has first two moments of the form
\[
\mathbb{E}[\varsigma_1(U_t) \mid U_0 = u, \varphi_{VX}] = \begin{pmatrix} \mu_1 \\ 0 \end{pmatrix}, \quad \mathbb{V}[\varsigma_1(U_t) \mid U_0 = u, \varphi_{VX}] = \begin{pmatrix} a_1 & 0 \\ 0 & -b^2/a_1 + a_2 \end{pmatrix}.
\]
The next transformation scales the process into the optimal form (according to Lemma 2.2) $\varsigma_2(u) = Y_2 \cdot u$, where
\[
Y_2 = \begin{pmatrix} \mu_1/a_1 & 0 \\ 0 & \sqrt{-b^2/a_1 + a_2} \end{pmatrix}.
\]
Then $\varsigma_2 \circ \varsigma_1(U_t)$ has first two moments of the form
\[
\mathbb{E}[\varsigma_2 \circ \varsigma_1(U_t) \mid U_0 = u, \varphi_{VX}] = \begin{pmatrix} \mu_1^2/a_1 \\ 0 \end{pmatrix}, \quad \mathbb{V}[\varsigma_2 \circ \varsigma_1(U_t) \mid U_0 = u, \varphi_{VX}] = \begin{pmatrix} \mu_1^2/a_1 & 0 \\ 0 & 1 \end{pmatrix},
\]
and choosing $D = \mu_1^2/a_1 - 1$ the bivariate orthogonal expansion of the density of $\varsigma_2 \circ \varsigma_2(U_t)$ may be performed in terms of the polynomials introduced in (5.4) and (5.6). By the transformation the polynomial moments up to second order induced by $w$ agree with the moments of $\varsigma_2 \circ \varsigma_1(U_t)$ and the moment-matching Lemma 2.2 applies up to order 2. We have used $\varsigma : u \mapsto Y_2 \circ (Y_1 u - v_1)$, and its inverse is $\varsigma^{-1} : u \mapsto Y_1^{-1} Y_2^{-1} u + Y_1^{-1} v_1$. The parameter $C$ in (5.2) is set to the exact excess kurtosis of the transformed log stock process and the expansion may be performed analogously to Section 7.

8.2. Option Pricing. Heston’s model (8.1) is used for option pricing because it may be consistent with the implied volatility skew that can be inferred from market prices. As such it is much more compatible with real data than for instance the Black-Scholes model. In stock (index) option pricing the quantity of interest is the marginal transition probability distribution function of the log stock price $X$. We therefore engineer an approximation directly around the marginal density of $X_\Delta \mid X_0, V_0$ by expanding $g_X$ in $L^2_{\mu_0}$. We set the constant $C$ from (5.2) to the excess kurtosis of $X$.

Recall that the price of a European call option with maturity $\Delta$ and strike price $K$ is given by
\[
C(\Delta, K) = e^{-r\Delta} \mathbb{E}\left[\left(e^{X_\Delta} - K\right)^+ \mid X_0 = x, V_0 = v, \varphi_{VX}\right]
\]
(8.2)
\[
e^{-r\Delta} \left( \int_{\log K}^{\infty} e^\xi g_X(\xi \mid x, v, \varphi_{VX}, \Delta) d\xi - K \int_{\log K}^{\infty} g_X(\xi \mid x, v, \varphi_{VX}, \Delta) d\xi \right).
\]
We will denote by $C^{(J)}(\Delta, K)$, and similarly $HA^{(J)}(\Delta, K)$ and $HB^{(J)}(\Delta, K)$, the option price computed with $g^{(J)}_X$ instead of $g_X$. Denoting by $Q(X \leq \xi)$ the marginal distribution function (and accordingly $Q^{(J)}(X \leq \xi)$ the $J$-order approximation thereof) we have that $HB^{(J)}(\Delta, K) = 1 - Q^{(J)}(X \leq \log K)$. Using the standardization from Example 8.1 and the change of variables formula

$$HB^{(J)}(\Delta, K) = 1 - \int_{-\infty}^{\log K} g^{(J)}_X(\xi | x, v, \vartheta V |) d\xi$$

$$= 1 - \frac{1}{\sqrt{a_2}} \int_{-\infty}^{\log K} \gamma_b \left( \frac{\xi - \mu_2}{\sqrt{a_2}} \right) \left( 1 + \sum_{i=1}^{J} c_i^b H_i^{(b)} \left( \frac{\xi - \mu_2}{\sqrt{a_2}} \right) \right) d\xi$$

$$= 1 - \sum_{i=0}^{J} \Gamma_b \left( \frac{\log K - \mu_2}{\sqrt{a_2}}, i \right) \vartheta_i.$$

Here, $H_i^{(b)}$ are from eqs. (5.6) and (5.7) and $\vartheta_i$ are implicitly defined as

$$1 + \sum_{i=1}^{J} c_i^b H_i^{(b)}(\xi) = \sum_{i=0}^{J} \vartheta_i \xi^i.$$

The function $\Gamma_b(K, n) = \int_{-\infty}^{K} \xi^n \gamma_b(\xi) d\xi$ is explicit in terms of the (incomplete) Gamma function $\int_0^x \exp(-\xi) \xi^k d\xi$ $(x > 0)$. The constituent $HA^{(J)}$ of the approximate call price from eq. (8.2) can be similarly computed, since exponential tilting $\gamma_b$ still yields integrals of the same kind.\(^{15}\)

Figure 2 shows that option pricing performance is very good.

\(^{15}\)We thank a referee for this suggestion.
Remark 8.1. Collecting coefficients to compute $\vartheta_i$ from Section 8.2 eq. (8.3) by hand is very error-prone. Instead we recommend using a symbolic mathematics software package such as Mathematica or Maple.

8.3. CDO Pricing. In the reduced-form credit risk framework (Lando, 1998), we model the stochastic default intensity $\lambda$ of a corporation with a positive process such as (6.1). Under the pricing measure $\mathbb{Q}$ the default time $\tau$ of a corporation is then taken to be the first jump of an inhomogeneous Poisson process with intensity $\lambda$. More formally we write the survival probability of a corporation (using the short-hand notation $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$)

$$\mathbb{Q}[\tau > T \mid \mathcal{F}_t] = \mathbb{I}_{\{\tau > t\}} \mathbb{E}_t \left[ e^{-\int_t^T \lambda_u du} \right].$$

All expectations are with respect to the risk-neutral pricing measure $\mathbb{Q}$. For the pricing of portfolio credit derivatives, to introduce dependence between different obligors, Duffie and Garleanu (2001) (and subsequently Mortensen (2006), Eckner (2009), and Feldhütter (2008)) introduce a factor intensity model

$$\lambda_{it} = X_{it} + a_i Y_t,$$

where $X_{it}$ is a firm-specific (idiosyncratic) intensity factor, and $Y_t$ is a (systemic) factor common to all obligors $i = 1, \ldots, n$. We model both $X$ and $Y$ with independent jump-diffusion processes from eq. (6.1). For $n$ obligors we must impose $\sum_{i=1}^n a_i = 1$ to ensure identifiability (see Eckner (2009)).

The survival probability of obligor $i$ according to model (8.4) is then due to independence of the factors

$$\mathbb{Q}[\tau_i > T \mid \mathcal{F}_t] = \mathbb{I}_{\{\tau_i > t\}} \mathbb{E}_t \left[ e^{-\int_t^T X_{iu} du} \right] \mathbb{E}_t \left[ e^{-a_i \int_t^T Y_u du} \right].$$

Defining $Z_{t,T} = \int_t^T Y_s ds$ we may write the default probability of the $i$th obligor conditional on $Z_{t,T}$ as

$$q_i(Z_{t,T}) = \mathbb{Q}_t[ t < \tau_i \leq T \mid Z_{t,T}] = \mathbb{I}_{\{\tau_i > t\}} \left( 1 - \mathbb{E}_t \left[ e^{-\int_t^T X_{iu} du} \right] e^{-a_i Z_{t,T}} \right),$$

and denoting by $P_{t,T}^{(m)}(k \mid Z_{t,T})$ the conditional probability that $k$ of the first $n$ credits in the portfolio default between $t$ and $T$ the recursive algorithm of Andersen et al. (2003) then develops $P_{t,T}^{(0)}(k \mid Z_{t,T})$ as

$$P_{t,T}^{(0)}(k \mid Z_{t,T}) = \mathbb{I}_{\{k=0\}},$$

$$P_{t,T}^{(m+1)}(k \mid Z_{t,T}) = q_{m+1}(Z_{t,T}) P_{t,T}^{(m)}(k-1 \mid Z_{t,T}) + (1-q_{m+1}(Z_{t,T})) P_{t,T}^{(m)}(k \mid Z_{t,T}),$$

for $0 \leq k \leq n$ and $0 \leq m < n$. The expressions $\mathbb{E}_t \left[ e^{-\int_t^T X_{iu} du} \right], i = 1, \ldots, n$ are unproblematic, but computing the unconditional default probability

$$P_{t,T}^{(n)}(k) = \int P_{t,T}^{(n)}(k \mid Z_{t,T}) d\mathbb{Q}(Z_{t,T})$$

involves an integration against the density of $Z_{t,T}$. We can get hold of the distribution of $Z_{t,T}$ by investigating the joint evolution of $Y$ from eq. (6.1) and the integral over $Y$. We therefore
embed \( Y \) into the two-dimensional affine process \((Y, Z)\) described by

\[
\begin{align*}
    dY &= (\kappa\theta - \kappa Y) \, dt + \sigma \sqrt{Y} \, dW_t + dL_t \\
    dZ_t &= Y_t \, dt.
\end{align*}
\] (8.8)

Note that even though the instantaneous covariance matrix of the process (8.8) above is only of rank one, this process is a well-defined affine process in the sense of Duffie et al. (2003) as pointed out also in Section 4.3. Existence of the marginal transition density of \(Z_{t,T} \mid Y_t\) is shown in Theorem 4.3 for \(\kappa\theta > \sigma^2\).

In principle the conditional default probabilities from eq. (8.7) may be computed using the moment generating function of \(Z_{t,T}\). In real-world applications \(n\) is typically larger than 100, however, and the expressions become intractably large, even for small \(k\). In practice, recursion (8.7) is therefore computed through numerical integration. A test of our density expansion in this setting may therefore be reduced to the question of how well we can approximate the true moment generating function. Below we outline how this approximation can be done in closed form using a \(\gamma L^2\) weight.

Denote by \(E_t^{(J)}[f(Z_{t,T})]\) the expectation of \(f(Z_{t,T})\) with respect to a \(J\)-order expansion instead of the true density. Considering the functional form of the expansion (2.1), to approximate the expressions \(E_t \left[ e^{-a Z_{t,T}} \right], i = 1, \ldots, n\) we note that we need to perform the computation

\[
\begin{align*}
    E_t^{(J)} \left[ e^{a Z_{t,T}} \right] &= \int_{\mathbb{R}^+} e^{a\xi} \gamma(\xi) \sum_{j=0}^{J} c_j^\gamma H_j^\gamma(\xi) \, d\xi \\
    &= \sum_{j=0}^{J} c_H(j) \int_{\mathbb{R}^+} e^{a\xi} \xi^j \gamma(\xi) \, d\xi, \quad (8.9)
\end{align*}
\] (8.10)

where \(c_H(j)\) is implicitly defined as

\[
\sum_{j=0}^{J} c_j^\gamma H_j^\gamma(\xi) = \sum_{j=0}^{J} c_H(j) \xi^j.
\]

To compute (8.10) we note that for a random variable \(Z\) that is Gamma distributed \(Z \sim \Gamma(\alpha, \theta)\)

\[
E \left[ e^{a Z} Z^n \right] = \frac{\theta^{-n} \Gamma(\alpha - n)(1 - a\theta)^{a-n}}{\Gamma(\alpha)}, \quad n \in \mathbb{N}, a \in \mathbb{R},
\]

where \(\Gamma\) denotes the Gamma function.

Figure 3 shows that for the order 10 expansion the approximation error is numerically zero. The order 2 expansion also works well, with negligible numeric error.

8.4. **Likelihood-based Inference.** In this section we investigate the performance of the polynomial density expansions in likelihood-based inference, both frequentist and Bayesian. For discrete, equally spaced (with time interval \(\Delta\)) observations \((X_0, X_1, \ldots, X_N) = \tilde{X}\) of a Markov

\(^{16}\)In practice the coefficients may be collected using a symbolic mathematics package such as Mathematica or Maple.
Figure 3. True vs. Approximated Moment Generating Function: The figure shows the log difference between the true moment generating function $E_e^{aZ}$ and the approximated moment generating function $E^{(J)}_{e^{aZ}}$ computed for an order 2 and an order 10 expansion of the integrated BAJD from (8.8). The parameters that generated the picture were $a = 1, T - t = 5, \kappa \theta = 0.00150602, \kappa = 0.4648, \sigma = 0.01, l = 1, \nu = 0.0002, y_0 = (\kappa \theta + \nu)/\kappa$. Results are computed using Mathematica and the picture is generated with a numeric precision of 20 digits.

For a process $(X_t)_{t \geq 0, X_0 = x_0}$ with domain $D$ and parameters $\varrho_X$ we may write the likelihood function $l_X : D^N \times g_X \to \mathbb{R}_+$ as

$$l_X(X | \varrho_X, X_0) = \prod_{i=1}^N g_X(X_i | X_{i-1}, \varrho_X, \Delta).$$

(8.11)

The approximate likelihood function $l^{(J)}_X$ is defined analogously with $g^{(J)}_X$ replacing $g$. The maximum likelihood estimator $\hat{\varrho}_X$ is obtained as the global maximizer of the log likelihood (8.11)

$$\hat{\varrho}_X = \arg \max_{\varrho_X \in \Theta} \sum_{i=1}^N \log g_X(X_i | X_{i-1}, \varrho_X, \Delta).$$

(8.12)

The Bayesian framework (cf. Robert, 1994, for reference and comparison to other methodologies) is aimed at the posterior density

$$p(\varrho_X | X) = \frac{l_X(X | \varrho_X)}{\int l(X | \varrho) \pi(\varrho) d\varrho} \pi(\varrho_X) \propto l_X(X | \varrho_X) \pi(\varrho_X).$$

(8.13)

Subsequently we will denote posteriors where the likelihood is approximated using $l^{(J)}_X$ by $p^{(J)}(\varrho_X | X)$. To test both methodologies we generate realizations from models (6.1) and (8.1) through exact simulation methods. We then perform both frequentist and Bayesian inference using our density approximations and the true density (obtained through Fourier inversion of the characteristic function). Frequentist inference is performed on 1,000 data sets generated by model (8.1), to acquire information about the sampling distribution of the (approximate) maximum likelihood estimators. Bayesian inference is performed on one data set, for the BAJD (eq.
Table 2. Kolmogorov-Smirnov test statistics: The table displays p-values for a two-sided Kolmogorov-Smirnov test applied to posterior density $g_{VX}$ to $g_{VX}^{(2)}$ and $g_{VX}^{(4)}$ using prior (8.15) for the Heston model in the left panel. The right panel displays p-values for the test applied to $g_Y$ to $g_Y^{(2)}$ and $g_Y^{(4)}$, the BAJD model. The prior distribution for this model is defined in eq. (8.14). The true posterior is defined in eq. (8.13) and the approximate posterior densities are defined in terms of density approximations $g_Y^{(2)}$ and $g_Y^{(4)}$.

(6.1)) and Heston’s model (eq. (8.1)), respectively. We then compare the posterior distribution originating from the true density to the posterior distribution from the density approximations from this paper.

The simulation for each data set is started from the unconditional mean and then propagated forward 600 data points. We discard the first 100 observations to eliminate impact of the initial condition. To investigate the behavior of our density expansions for different time horizons we choose a monthly observation frequency for the square-root jump-diffusion (6.1) and weekly observation frequency for the Heston model (8.1).

To obtain exact draws from the BAJD we generate exact draws from $Y_i \mid Y_{i-1}$ using Robert and Casella (2004, Lemma 2.4). For a uniform random variable $U \sim U(0,1)$ we exploit that $G_Y^{-1}(U \mid Y_{i-1}, \varrho_Y) \sim G_Y$ for any distribution function $G_Y$. We simulate from (6.1) using the parameters $\kappa \theta = 0.04, \kappa = 1, \sigma = 0.2, l = 3, \nu = 0.01$ in Algorithm A.1.

For Bayesian inference we specify an uninformative prior

$$
\pi((\varrho_Y)) = \mathbb{1}_{\{2\kappa \theta > \sigma^2, \sigma > 0, l > 0, \nu > 0\}} \frac{1}{\sigma \cdot \kappa \theta \cdot l \cdot \nu}.
$$

The Heston parameters are $\kappa_V = 1, \kappa \theta_V = 0.04, \sigma = 0.2, \kappa \theta_X = 0.03, \rho = -0.8$. To obtain exact draws from this model we refer the reader to the algorithm in Broadie and Kaya (2006). For Bayesian inference we specify the prior distribution as

$$
\pi((\varrho_{VX})) = \mathbb{1}_{\{2\kappa \theta_V > \sigma^2, \sigma > 0, l > 0, \nu > 0\}} \frac{1}{\sigma \cdot \kappa \theta_V}.
$$

To evaluate the true transition density we employ the formulation from Lamoureux and Paseka (2005) through a single numerical integral, instead of the two-dimensional Fourier integral.

With 1,000 datasets of weekly realizations from the Heston model, for each dataset we obtain parameters $\widehat{\varrho}_{VX}$ by maximizing the log likelihood (8.11), respectively the approximate log likelihood. We use the optimizer donlp2 to achieve this task. To relate the density expansions of this paper to existing approximations we perform the estimation experiment with

- the true density (obtained through Fourier inversion) denoted by $MLE$
Table 3. Heston Estimation Success: The table reports the number of estimation successes on 1,000 datasets generated as exact draws from the Heston model using the technology from Broadie and Kaya (2006). Estimation success is defined by the optimizer meeting the termination criterion, which is a function of the norm of the gradient of the log likelihood function. The density approximations used are BG(4), a fourth order expansion using a Bilateral Gamma weight for the log stock variable and a Gamma weight for the variance variable, G(4), a fourth order expansion using a Gaussian weight for the log stock variable and a Gamma weight for the variance variable, QML denotes a Gaussian approximation using the true conditional moments up to order 2, and CF(2) denotes the second-order likelihood expansions from Aït-Sahalia (2008). The optimizer used in the likelihood search is donlp2.

|                | MLE | BG(4) | QML | G(4) | CF(2) |
|----------------|-----|-------|-----|------|-------|
| #Success       | 688 | 841   | 982 | 949  | 981   |

- order 4 likelihood expansions developed in this paper using a product measure with a Gamma weight for the variance process and for the log stock variable a
  - bilateral Gamma weight. Specifically we employ formulation (5.3). Estimates are denoted by $BG(4)$
  - Gaussian weight. Estimates are denoted by $G(4)$
- order 2 closed-form likelihood expansions from Aït-Sahalia (2008) denoted by $CF(2)$
- Gaussian approximation using true conditional moments up to order 2 denoted by $QML$

Table 3 reveals that the true likelihood function exhibits problematic behavior for some parameterizations. Only 688 out of 1,000 estimates turned out to be successful. This is due to numerical integration problems that occur in particular for low values of $\sigma$ that arise in the likelihood search. Density expansions developed in this paper are also not entirely unproblematic. Numerical errors from evaluating the pseudo likelihood ratio accumulate and induce spikes that irritate the optimizer’s numerical differentiation routines. The Hermite polynomials used for $G(4)$ appear better behaved than the polynomials associated with the bilateral Gamma density used in $BG(4)$. 
### Heston Asymptotic Assessment

Panel (a) displays bias and RMSE of the MLE estimator and the approximated MLE estimators. Panel (b) displays mean and standard deviation of ML estimation bias as well as the first two moments of the difference between the MLE estimator and the approximated MLE estimators. Computed over a sample of 1,000 datasets, all of which generated as exact draws from the Heston model using the technology from Broadie and Kaya (2006). For a given dataset only parameter estimates were taken into consideration where all five estimators converged. Out of 1,000 this left 578 samples. The number of estimation successes is reported in Table 3 above. Approximate estimators are obtained through BG(4), a fourth order expansion using a Bilateral Gamma weight for the log stock variable and a Gamma weight for the variance variable, G(4), a fourth order expansion using a Gaussian weight for the log stock variable and a Gamma weight for the variance variable, QML denotes a Gaussian approximation using the true conditional moments up to order 2, and CF(2) denotes the second-order likelihood expansions from Aït-Sahalia (2008).
Table 3a reports bias and RMSE of the estimators. The large bias of 0.2255 for the \( \kappa \) parameter is a well-established phenomenon that has also been reported in Aıt-Sahalia and Kimmel (2007). As an overall impression the results suggest that the density approximations developed in this paper exhibit parameter estimates with properties similar to the true ML estimates, while Aıt-Sahalia (2008) expansions interestingly exhibit lower bias, with the exception of the \( \sigma \) parameter, but higher RMSE. In Table 3b the first column (Mean) in \( \hat{\rho}_{VX}^{MLE} - \rho_{VX}^{TRUE} \) indicates mean deviation from the true ML estimator and the second column (SD) captures statistical noise in the estimation. Estimation bias around the MLE for all estimators appears very small. Except for the \( CF(2) \) estimator, the noise induced through the density approximations is smaller than the estimation noise of the true \( MLE \). Surprisingly, the QML estimator, a special case of the approximations developed in this paper since it is an order two expansion around a Gaussian, performs remarkably well. All around the \( BG(4) \) expansions appear to be the preferable choice. In particular \( \hat{\sigma}_{BG(4)}^{MLE} - \hat{\sigma}_{BG(4)}^{MLE} \) and \( \hat{\rho}_{BG(4)}^{MLE} - \hat{\rho}_{BG(4)}^{MLE} \) point to the right direction, the estimators are closer to the true parameters than \( MLE \).

The results of the Bayesian inference study also appear promising. We see that an order 2 expansion already delivers reasonable results, while the order 4 expansion brings the approximate posterior even closer to the posterior density obtained from the true density function. To assess how close the posteriors \( p_{VX}^{(2)} \) and \( p_{VX}^{(4)} \) densities are to the posterior obtained through the true transition density \( p_{VX} \) we compute Kolmogorov-Smirnov tests. The results can be seen in Table 2. They suggest that while \( p_{VX}^{(2)} \) appears to be quite different from \( p_{VX} \), \( p_{VX}^{(4)} \) is statistically almost indistinguishable from the true posterior \( p_{VX} \) for the majority of the parameters.

9. Discussion

This paper develops a general framework for density approximations for affine processes using orthonormal polynomial expansions in well-chosen weighted \( L^2 \) spaces. We also provide novel existence and smoothness results for their true, unknown transition densities.

The approximations are designed to exploit the explicit polynomial moments of affine processes to compute the coefficients of the expansion without approximation error and in closed form; the computational burden is concentrated only in the initial calculation of the coefficients of the expansions. Once they are implemented, evaluation is rapid, avoiding the heavy computational cost of Fourier methods to obtain transition densities. Empirical applications in option pricing, credit risk, and likelihood-based parameter inference suggest that the density expansions are very accurate.

The paper leaves a number of open points for future research. The first question concerns approximations in higher-order weighted Sobolev spaces. One might suspect that approximation of (sufficiently smooth) densities in weighted higher-order Sobolev spaces are superior to \( L^2 \) expansions. In particular, it could be expected that (i) the quality of approximation might be better (ii) Sobolev embedding theorems could be applied to infer global uniform convergence. However, quite contrary to the \( L^2 \) case, it is unknown whether the space of polynomials is dense in weighted Sobolev spaces. Higher-order Sobolev spaces also impose heavy restrictions on the functional form of the approximation weights, which in turn lead to very slow convergence rates. Indeed, preliminary numerical experiments suggest that the price for global, uniform convergence which potentially comes with higher-order Sobolev spaces is a very slow convergence rate.
Another route worth pursuing is a compact truncation of the state space, such that approximations could be performed in non-weighted Sobolev spaces, for which there is more theory available in the literature.

Suitable approximation weights (such as the bilateral gamma weight of this paper) are a research topic of its own, and they lead to non-trivial problems in the theory of special functions. Also, density expansions for processes on state spaces different from the canonical ones would be highly desirable. As an example we mention the class of matrix-valued processes used in co-volatility modeling (Leippold and Trojani, 2008; Da Fonseca et al., 2008; Buraschi et al., 2008).

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica*, 70:223–262.

Aït-Sahalia, Y. (2007). *Estimating Continuous-Time Models Using Discretely Sampled Data.* in Richard Blundell, Torsten Persson, Whitney K. Newey: *Advances in Economics and Econometrics, Theory and Applications*, chapter 9, Cambridge University Press.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36(4):906–937.

Aït-Sahalia, Y. and Hansen, L. P., editors (2009). *Handbook of Financial Econometrics*. Elsevier.

Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83:413–452.

Aït-Sahalia, Y. and Kimmel, R. L. (2010). Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions. *Journal of Financial Economics*, 98:113–144.

Aït-Sahalia, Y. and Yu, J. (2006). Saddlepoint approximations for continuous-time Markov processes. *Journal of Econometrics*, 134:507–551.

Andersen, L., Sidenius, J., and Basu, S. (2003). All your hedges in one basket. *Risk*.

Bates, D. (2006). Maximum likelihood estimation of latent affine processes. *Review of Financial Studies*, 19(3):909–965.

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

Bibby, B., Jacobsen, M., and Sørensen, M. (2004). Estimating functions for discretely sampled diffusion-type models. Working paper, CAF Centre for Analytical Finance, University of Aarhus.

Broadie, M. and Kaya, Özgür. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231.

Bru, M.-F. (1991). Wishart processes. *Journal of Theoretical Probability*, 4(4):725–751.

Buraschi, A., Cieslak, A., and Trojani, F. (2008). Correlation risk and the term structure of interest rates. Working paper, Imperial College and University of St. Gallen.

Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2(4):61–73.

Chen, H. and Joslin, S. (2011). Generalized transform analysis of affine processes and applications in finance. Working paper, Sloan School of Management and Marshall School of Business.
Collin-Dufresne, P., Goldstein, R. S., and Jones, C. S. (2008). Identification of maximal affine term structure models. *Journal of Finance*, 63(2):743–795.

Cox, J., Ingersoll, J., and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–407.

Cuchiero, C., Filipović, D., Mayerhofer, E., and Teichmann, J. (2010a). Affine processes on positive semidefinite matrices. *The Annals of Applied Probability*, 21:397–463.

Cuchiero, C., Teichmann, J., and Keller-Ressel, M. (2010b). Polynomial processes and their application to mathematical finance. *Finance & Stochastics*, forthcoming.

Da Fonseca, J., Grasselli, M., and Tebaldi, C. (2008). A multifactor volatility heston model. *Quantitative Finance*, 8(6):591 – 604.

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *Journal of Finance*, 55(5):1943–1978.

Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability*, 13:984–1053.

Duffie, D. and Garleanu, N. (2001). Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57(1):41–59.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Eckner, A. (2009). Computational techniques for basic affine models of portfolio credit risk. *Journal of Computational Finance*, 100:1 – 35.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, 19(2):177–191.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59(3):1367–1404.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, 58(3):1269–1300.

Feldhütter, P. (2008). An empirical investigation of an intensity-based model for pricing CDO tranches. Working paper, Copenhagen Business School.

Filipović, D. (2009). *Term-Structure Models: A Graduate Course*. Springer, Berlin.

Filipović, D. and Mayerhofer, E. (2009). Affine diffusion processes: theory and applications.

Forman, J. L. and Sørensen, M. (2008). The pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35:438–465.

Gallant, A. R. and Tauchen, G. (2009). *Simulated Score Methods and Indirect Inference for Continuous-time Models*. In Aït-Sahalia and Hansen (2009).

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343.

Horn, R. A. and Johnson, C. R. (1990). *Matrix analysis*. Cambridge University Press, Cambridge. Corrected reprint of the 1985 original.

Hurn, A., Jeisian, J., and Lindsay, K. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics*, 5:390–455.
Hurn, A., Jeisman, J., and Lindsay, K. (2008). Horses for courses: Polynomial-based approximations of transitional density. Working paper, Queensland University of Technology, University of Glasgow.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). Continuous univariate distributions. Vol. 2. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Jones, C. S. (1998). Bayesian estimation of continuous-time finance models. Working paper, University of Rochester.

Kristensen, D. and Mele, A. (2011). Adding and subtracting black-scholes: A new approach to approximating derivativeprices in continuous-time models. *Journal of Financial Economics*. forthcoming.

Küchler, U. and Tappe, S. (2008a). Bilateral gamma distributions and processes in financial mathematics. *Stochastic Processes and Their Applications*, 118(2):261 – 283.

Küchler, U. and Tappe, S. (2008b). On the shapes of bilateral gamma densities. *Statistics and Probability Letters*, 78(15):2478 – 2484.

Lamoureux, C. G. and Paseka, A. (2005). Information in options and underlying asset dynamics. working paper, University of Arizona.

Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120.

Leippold, M. and Trojani, F. (2008). Asset pricing with matrix jump diffusions. Technical report, Swiss Finance Institute University of Zürich and University of Lugano.

Mortensen, A. (2006). Semi-analytical valuation of basket credit derivatives in intensity-based models. *Journal of Derivatives*, 13(4):8–28.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.

Robert, C. P. (1994). *The Bayesian Choice*. Springer, New York.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621.

Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author.

Schaumburg, E. (2001). *Maximum Likelihood Estimation of Jump Processes*. PhD thesis, Princeton University. Chapter 1 of thesis.

Schoutens, W. (2000). *Stochastic Processes and Orthogonal Polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer, New York.

Singleton, K. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102(6):111–141.

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points: a survey. *International Statistical Review*, 72(3):337–354.

Stramer, O., Bognar, M., and Schneider, P. (2009). Bayesian inference for discretely sampled markov processes with closed–form likelihood expansions. *Journal of Financial Econometrics*. forthcoming.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

Volkmann, P. (1972). Gewöhnliche Differentialungleichungen mit quasinonmonoton wachsenden Funktionen in topologischen Vektorräumen. *Math. Z.*, 127:157–164.
Wong, E. (1964). The construction of a class of stationary markoff processes. *Stochastic Processes in Mathematical Physics and Engineering*, pages 264–276.

Wong, E. and Thomas, J. B. (1962). On polynomial expansions of second-order distributions. *Journal of the Society for Industrial and Applied Mathematics*, 10:507–516.

Yu, J. (2007). Closed-form likelihood estimation of jump-diffusions with an application to the realignment risk of the chinese yuan. *Journal of Econometrics*, 141:1245–1280.
Internet Appendix to

“Density Approximations for Multivariate Affine Jump-Diffusion Processes”
by Damir Filipović, Eberhard Mayerhofer, and Paul Schneider

This separate Internet Appendix collects additional proofs referenced in the paper.
Appendix A. Exact Sampling from the Basic Affine Jump-Diffusion

Algorithm A.1 (Exact draws from BAJD process (6.1)). We perform the following procedure starting from $Y_0 = \mathbb{E}[Y_t]$, the unconditional mean, for a realization $Y_i \mid Y_{i-1}$

(i) Draw $U \sim \mathcal{U}(0, 1)$. Call the realization $u_i$.

(ii) Use the Newton-Raphson algorithm to compute $y : G_Y(Y \leq y \mid Y_{i-1}, \varrho_Y) = u_i$. In this step we substitute $y = e^{w_j} + c$ to keep $y$ on the positive domain. The floor parameter $c$ we set to $10^{-6}$ to avoid numerical difficulties. The iteration is then

$$w_{j+1} = w_j - \frac{e^{w_j} (e^{w_j} + 1)^2 G_Y(Y \leq c + e^{w_j} \mid Y_{i-1}, \varrho_Y) - u_i}{gy(c + e^{w_j} \mid Y_{i-1}, \varrho_Y)}$$

starting from $w_0 = \log \left(\frac{y_{i-1} - c}{c - y_{i-1} + 1}\right)$. Stop the iteration at

$$w^* : \left|G_Y(Y \leq c + e^{w^*} \mid Y_{i-1}, \varrho_Y) - u_i\right| < \varepsilon.$$

Both, $g_Y$, and, $G_Y$ are obtained through Fourier inversion. In our implementation the algorithm terminates after 5 to 6 iterations for $\varepsilon = 10^{-6}$.

(iii) Set $Y_i = c + e^{w^*} \mid Y_{i-1}$ increment $i$ and go back to step (1).

Appendix B. Proofs for Section 3

This appendix gathers the proofs of the lemmas in Section 3.

Proof of Lemma 3.1. That the set of polynomials is dense in $\mathcal{L}^2_w$ is shown in Bernard (1995, Lemma 1). The assumption made in Bernard (1995) that $w$ is strictly positive can easily be omitted by replacing point-wise equality “= 0” by “= 0 $w(\xi)$ $d\xi$-a.s.” at the end of the proof of Bernard (1995, Lemma 1). An orthonormal basis of polynomials $\{H_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ of $\mathcal{L}^2_w$ with $\text{deg} H_\alpha = |\alpha|$ is obtained by applying the Gram–Schmidt process to the linearly independent set of monomials $\{\xi^\alpha \mid \alpha \in \mathbb{N}_0^d\}$, see Algorithm 5.1.

Proof of Lemma 3.2. That the product density $w$ on $\mathbb{R}^d$ has finite exponential moment (3.1) follows from the elementary inequality $\|\xi\| \leq \sum_{i=1}^d |\xi_i|$ for all $\xi \in \mathbb{R}^d$. The orthonormality of $H_\alpha$ follows from the easily verifiable relationship

$$\langle H_\alpha, H_\beta \rangle_{\mathcal{L}^2_w} = \prod_{i=1}^d \langle H^i_\alpha, H^i_\beta \rangle_{\mathcal{L}^2_{w^i}(\mathbb{R})}, \quad \alpha, \beta \in \mathbb{N}_0^d.$$

Moreover, every monomial $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$ can be written as a product of linear combinations of the respective orthonormal polynomials

$$\xi_1^{\alpha_1} = \sum_{j=0}^{\alpha_1} c_{ij} H^1_j(\xi_1).$$

It follows that the set $\{H_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ is dense in $\mathcal{L}^2_w$, and hence forms an orthonormal basis of $\mathcal{L}^2_w$. This proves the lemma.
Proof of Lemma 3.3. The lemma follows from the estimate
\[ \int_{\mathbb{R}^d} \frac{g(\xi)^2}{w(\xi)} d\xi = \int_{\mathbb{R}^d} g(\xi) \frac{e^{-c_0\|\xi\|}}{w(\xi)} e^{c_0\|\xi\|} g(\xi) d\xi \leq \sup_{x \in \mathbb{R}^d} \left( g(x) \frac{e^{-c_0\|x\|}}{w(x)} \right) \int_{\mathbb{R}^d} e^{c_0\|\xi\|} g(\xi) d\xi < \infty. \]

\[ \square \]

Proof of Lemma 3.4. A Taylor expansion of \( g(x) \) around 0 gives \( g(x) = \frac{\partial_p^p g(\xi)}{p!} x^p \) for some \( \xi = \xi(x) \in [0, x] \). Since \( \partial_p^p g \) is continuous we conclude that there exists some finite constant \( K \) such that \( g(x) \leq K x^p \) for all \( x \in [0, 1] \). We then obtain
\[ \int_0^\infty \frac{g(\xi)^2}{w(\xi)} d\xi = \int_0^1 \frac{g(\xi)^2}{w(\xi)} d\xi + \int_1^\infty g(\xi) \frac{e^{-c_0\xi}}{w(\xi)} e^{c_0\xi} g(\xi) d\xi \leq K^2 \int_0^1 \frac{\xi^2p}{w(\xi)} d\xi + \sup_{x \geq 1} \left( g(x) \frac{e^{-c_0x}}{w(x)} \right) \int_1^\infty e^{c_0\xi} g(\xi) d\xi < \infty. \]

This proves the lemma. \( \square \)

Proof of Lemma 3.5. Let \( x \in \mathcal{I} \), and let \( i^\star \) be such that \( x_{i^\star} = \min_i x_i \). A Taylor expansion of \( g(x, y) \in x_{i^\star} \) around \( x_{i^\star} = 0 \) gives \( g(x, y) = \frac{\partial_p^p g(\xi, y)}{p!} x_{i^\star}^p \) for some \( \xi = \xi(x, y) \in \mathcal{I} \). Since \( \partial_p^p g(x, y) \) is bounded on \( \mathcal{I} \times \mathbb{R}^n \) we conclude that there exists some finite constant \( K \) such that \( g(x, y) \leq K \min_i x_i^p \) for all \( (x, y) \in \mathcal{I} \times \mathbb{R}^n \). We then decompose
\[ \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} \frac{g(\xi, \eta)^2}{w(\xi, \eta)} d\xi d\eta = I_1 + I_2 \]
with
\[ I_1 = \int_{\mathcal{I}} \int_{\mathbb{R}^n} \frac{g(\xi, \eta)^2}{w(\xi, \eta)} d\xi d\eta \leq K \int_{\mathcal{I}} \int_{\mathbb{R}^n} \min_i \frac{\xi_i^p}{w(\xi, \eta)} e^{-c_2\|\eta\|} g(\xi, \eta) d\xi d\eta \]
\[ \leq K \sup_{(x, y) \in \mathcal{I} \times \mathbb{R}^n} \left( \frac{\min_i x_i^p e^{-c_2\|y\|}}{w(x, y)} \right) \int_{\mathcal{I}} \int_{\mathbb{R}^n} e^{c_2\|\eta\|} g(\xi, \eta) d\xi d\eta < \infty \]
and
\[ I_2 = \int_{(1, \infty)^m} \int_{\mathbb{R}^n} \frac{g(\xi, \eta)^2}{w(\xi, \eta)} d\xi d\eta \]
\[ = \int_{(1, \infty)^m} \int_{\mathbb{R}^n} g(\xi, \eta) \frac{e^{-c_1\|\xi\| - c_2\|\eta\|}}{w(\xi, \eta)} e^{c_1\|\xi\| + c_2\|\eta\|} g(\xi, \eta) d\xi d\eta \]
\[ \leq \sup_{(x, y) \in (1, \infty)^m \times \mathbb{R}^n} \left( g(x, y) \frac{e^{-c_1\|x\| - c_2\|y\|}}{w(x, y)} \right) \int_{(1, \infty)^m} \int_{\mathbb{R}^n} e^{c_1\|\xi\| + c_2\|\eta\|} g(\xi, \eta) d\xi d\eta < \infty. \]
This proves the lemma.

**Appendix C. Proof of Theorem 4.1**

First we note that the existence and smoothness properties of a density on $\mathbb{R}^d$ for $X_t|X_0 = x$ is invariant with respect to non-singular linear transformations of the state vector $X_t$. In view of Filipović (2009, Theorem 10.7) there exists a non-singular linear transformation of the state vector $X_t$ mapping $\mathcal{D} = \mathbb{R}_+^m \times \mathbb{R}^n$ onto itself, and which renders block diagonal matrices $\alpha_i$ in the form

\[(C.1) \quad \alpha_i = \begin{pmatrix} \text{diag}(0, \ldots, 0, \alpha_{i,ii}, 0, \ldots, 0) & 0 \\ 0 & \alpha_{i,JJ} \end{pmatrix}, \]

so that $x^\top \alpha_i x = \alpha_{i,ii} x_i^2 + x_j^\top \alpha_{i,JJ} x_J$ for all $x \in \mathbb{R}^d$. Moreover, this transformation does not affect the upper diagonal element $\alpha_{i,ii}$ (see the proof of Filipović (2009, Lemma 10.5)), which is important in view of the criterion (4.6). Hence without loss of generality we shall from now on assume that the matrices $\alpha_i$ are of the block diagonal form (C.1).

We now recall a classical result on characteristic functions $\hat{\nu}$ of probability measures $\nu$, see Sato (1999, Proposition 28.1):

**Lemma C.1.** Let $\nu$ be a probability measure on $\mathbb{R}^d$. Assume its characteristic function $\hat{\nu}(iu) = \int_{\mathbb{R}^d} e^{iu^\top \xi} \nu(d\xi)$ satisfies

$$\int_{\mathbb{R}^d} |\hat{\nu}(iu)| \|u\|^k du < \infty$$

for some nonnegative integer $k$. Then $\nu$ has a density $h(x)$ of class $C^k$ and the partial derivatives of $h(x)$ of orders $0, \ldots, k$ tend to $0$ as $\|x\| \to \infty$.

It thus remains to prove the appropriate integrability of the affine characteristic function (4.1), that is, the appropriate tail behavior in $u \in \mathbb{R}^d$ of the functions $\phi(t, iu)$ and $\psi(t, iu)$. The following lemma is our core result, which together with Lemma C.1 completes the proof of Theorem 4.1.

**Lemma C.2.** The following properties are equivalent:

(i) The $d \times (n+1)d$-matrix $\mathcal{K}$ given in (4.5) has full rank.

(ii) For any $t > 0$, the $d \times d$-matrix

$$A(t) = \int_0^t \text{diag}(0, e^{B_{J,J} s} a e^{B_{J,J} s}) ds + t \sum_{i \in L} \alpha_i$$

is nonsingular.

(iii) For any $t > 0$ there exists an $\epsilon > 0$ such that the cones

$$C_0 = \left\{ u \in \mathbb{R}^d \mid u^\top \left( \int_0^t e^{B_{J,J} s} a e^{B_{J,J} s} ds \right) u, J \geq \epsilon \|u\|^2 \right\}$$

$$C_i = \left\{ u \in \mathbb{R}^d \mid u^\top \alpha_i u \geq \epsilon \|u\|^2 \right\}, \quad i = 1, \ldots, m$$

cover $\mathbb{R}^d$. That is, $\bigcup_{i=0}^m C_i = \mathbb{R}^d$. 

Moreover, any of the above properties, (i), (ii), or (iii), implies that
\[
\int_{\mathbb{R}^d} \left| e^{\phi(t,iu)+\psi(t,iu)^T x} \right| \|u\|^p \, du < \infty
\]
for all nonnegative numbers \( p < \min_{i \in \{1, \ldots, m\}} \frac{b_i}{\alpha_{i,ii}} - 1 \).

The remainder of this section is devoted to the proof of Lemma C.2. Let \( t > 0 \) and \( u \in \mathbb{R}^d \setminus \{0\} \). We first claim that \( u^T K = 0 \) if and only if \( u^T A(t) = 0 \), which proves equivalence of (i) and (ii). To prove the claim, note that since \( A(t) \) is positive semidefinite, \( u^T A(t) = 0 \) is equivalent to \( u^T A(t)^{1/2} (= u) \). Since each of the summands in \( A(t) \) is positive semidefinite, this again is equivalent to \( u^T \sum_{i=1}^{m} \alpha_i = 0 \) and \( u^T (B_{ij}^k)^{1/2} a = 0 \) for all \( k \in \mathbb{N}_0 \).

The Cayley–Hamilton theorem (see (Horn and Johnson, 1990, Theorem 2.4.2)) implies that, for all \( k \geq n \), \( B_{ij}^k \) is a linear combination of \( I_d, B_{ij}, \ldots, B_{ij}^{n-1} \). Whence the above property is equivalent to \( u^T K = 0 \), which proves the claim.

The equivalence of (ii) and (iii) follows from the identity
\[
\int_{\mathbb{R}^d} \left| e^{\phi(t,iu)+\psi(t,iu)^T x} \right| \|u\|^p \, du < \infty
\]
and the fact that each of the summands is nonnegative. This establishes the first part of Lemma C.2.

As for the second part of Lemma C.2, we note that as a consequence of (C.1) the real and imaginary parts
\[
f(t,iu) = \Re \psi(t,iu) \quad \text{and} \quad g(t,iu) = \Im \psi(t,iu)
\]
of \( \psi \) satisfy the following system of Riccati equations, for \( i = 1, \ldots, m \):
\[
\begin{align*}
\partial_t f_i & = \alpha_{i,ii} f_i^2 - g^T \alpha_i g + B_i f + \int_D \left( \left( e^{f_j^T \xi} \cos \left( g^T e \xi \right) - 1 \right) \mu_i(d\xi) \right) \\
\partial_t g_i & = \alpha_{i,ii} g_i + B_i g + \int_D e^{f_j^T \xi} \sin \left( g^T e \xi \right) \mu_i(d\xi) \\
\partial_t f_i(0) & = 0 \\
f_i(t) & = 0 \\
\partial_t g_i(0) & = u_i \\
g_i(t) & = e^{B_{ij} t} u_j \\
\end{align*}
\]
In the sequel we will make use, without further notice, of the fact that \( f_i \) is \( \mathbb{R}_- \)-valued for all \( i = 1, \ldots, m \), and that \( f_j = 0 \) for all \( j \in J \). In particular, it follows from above that \( f_i \) satisfies the following system of differential inequalities
\[
\partial_t f_i \leq \alpha_{i,ii} f_i^2 - g^T \alpha_i g + B_{ii} f_i, \quad i = 1, \ldots, m.
\]
For any \( u \neq 0 \), we now define the scaled functions

\[
F(t, u) = \frac{1}{\|u\|} f\left( \frac{t}{\|u\|}, iu \right)
\]

\[
G(t, u) = \frac{1}{\|u\|} g\left( \frac{t}{\|u\|}, iu \right).
\]

Then \( F \) and \( G \) satisfy, for \( i = 1, \ldots, m \):

\[
\partial_t F_i \leq \alpha_{i,ii} F_i^2 - G^\top \alpha_i G + \frac{1}{\|u\|} B_{ii} F_i
\]

\[
F_i(0) = 0
\]

\[
F_J \equiv 0
\]

\[
\partial_t G_i = 2 F_i \alpha_{i,ii} G_i + \frac{1}{\|u\|} (B_i + K_i) G
\]

\[
G_i(0) = \frac{u_i}{\|u\|}
\]

\[
G_J = e^{B_{JJ} t} \frac{u_J}{\|u\|}
\]

where we define the \( d \times d \)-matrix \( K = K(t, u) \) by its \( 1 \times d \)-row vectors

\[
K_i = \begin{cases} 
\int_D \left( \int_0^1 \cos \left( s \|u\| G^\top \xi \right) ds \right) e^{\|u\| F^\top \xi \xi^\top \mu_i(d\xi)}, & i = 1, \ldots, m \\
0, & i = m + 1, \ldots, d,
\end{cases}
\]

and we have used the simple fact that

\[
e^{\|u\| F^\top \xi} \sin \left( \|u\| G^\top \xi \right) = e^{\|u\| F^\top \xi} \int_0^1 \frac{d}{ds} \sin \left( s \|u\| G^\top \xi \right) ds
\]

\[
= \|u\| G^\top \xi e^{\|u\| F^\top \xi} \int_0^1 \cos \left( s \|u\| G^\top \xi \right) ds.
\]

It follows by the assumptions on \( \mu_i \) that the matrix \( K \) is uniformly bounded

\[
\sup_{t,u} \|K(t, u)\| = K < \infty
\]

where the constant \( K \) only depends on the measures \( \mu_i \). The squared norm of \( G \) thus satisfies

\[
\partial_t \|G\|^2 = 2 G^\top \partial_t G = 4 G^\top \text{diag}(F) G + \frac{2}{\|u\|} G^\top (B + K) G
\]

\[
\leq \frac{2}{\|u\|} (\|B\| + K) \|G\|^2
\]

\[
\|G(0)\|^2 = 1.
\]

We shall now and in the sequel make use of the following comparison result, which is a special case of a more general theorem proved by Volkmann (1972):
Lemma C.3. Let $R(t,v)$ be a continuous real map on $\mathbb{R}_+ \times \mathbb{R}$ and locally Lipschitz continuous in $v$. Let $p(t)$ and $q(t)$ be differentiable functions satisfying
\[
\frac{d}{dt} p(t) \leq R(t, p(t)) \\
\frac{d}{dt} q(t) = R(t, q(t)) \\
p(0) \leq q(0).
\]
Then we have $p(t) \leq q(t)$ for all $t \geq 0$.

Applying Lemma C.3 to the above differential inequality for $\|G\|^2$ we obtain
\[
\|G\|^2 \leq e^{\frac{2}{\|\mathcal{B}\|}(\|\mathcal{B}\|+K)t}.
\]
For any $i \in \{1, \ldots, m\}$ we then obtain the differential inequality
\[
\partial_t \left( G^\top \alpha_i G \right) = 2G^\top \alpha_i \partial_t G \\
= 4G^\top \alpha_i \text{diag}(F_i) \alpha_i G + \frac{2}{\|u\|} G^\top \alpha_i (\mathcal{B} + K) G \\
\geq 4\alpha_{i,ii} F_i \alpha_{i,ii} G_i^2 - \frac{2\|\alpha_i\|}{\|u\|} (\|\mathcal{B}\| + K) \|G\|^2 \\
\geq 4\alpha_{i,ii} F_i G^\top \alpha_i G - \frac{2\|\alpha_i\|}{\|u\|} (\|\mathcal{B}\| + K) e^{\frac{2}{\|\mathcal{B}\|}(\|\mathcal{B}\|+K)t} \\
G(0)^\top \alpha_i G(0) = \frac{u^\top \alpha_i u}{\|u\|^2}
\]
where we have used the fact that $\alpha_{i,ii} G_i^2 \leq G^\top \alpha_i G$ and (C.5) for the last inequality. Lemma C.3 again yields the lower bound
\[
G^\top \alpha_i G \geq e^{4\alpha_{i,ii} \int_0^t F_i(s) ds} \frac{u^\top \alpha_i u}{\|u\|^2} \\
\quad - \frac{2\|\alpha_i\|}{\|u\|} (\|\mathcal{B}\| + K) \int_0^t e^{4\alpha_{i,ii} \int_0^r F_i(s) ds} e^{\frac{2}{\|\mathcal{B}\|}(\|\mathcal{B}\|+K)(t-s)} ds \\
\quad \geq e^{4\alpha_{i,ii} \int_0^t F_i(s) ds} \frac{u^\top \alpha_i u}{\|u\|^2} - \|\alpha_i\| \left( e^{\frac{2}{\|\mathcal{B}\|}(\|\mathcal{B}\|+K)t} - 1 \right)
\]
Combining this with the differential inequality (C.4) for $F_i$ we obtain
\[
\partial_t F_i \leq \alpha_{i,ii} F_i^2 + \frac{1}{\|u\|} \mathcal{B}_{ii} F_i - e^{4\alpha_{i,ii} \int_0^t F_i(s) ds} \frac{u^\top \alpha_i u}{\|u\|^2} \\
\quad + \|\alpha_i\| \left( e^{\frac{2}{\|\mathcal{B}\|}(\|\mathcal{B}\|+K)t} - 1 \right) \\
F_i(0) = 0.
\]
We arrive at the following intermediate result.
Lemma C.4. For every $\epsilon > 0$ and $t_0 > 0$ there exists some $\rho > 0$ and $R > 0$ such that

$$F_i(t_0, iu) \leq -\rho$$

for all $u \in \mathbb{R}^d$ with $\|u\| \geq R$ and $u^T \alpha_i u \geq \epsilon \|u\|^2$, for all $i \in \{1, \ldots, m\}$.

Proof. The differential inequality (C.6) is autonomous and smooth in $F_i$. Moreover, the initial slope satisfies

$$\partial_t F_i(t_0, iu) |_{t=0} \leq -\frac{u^T \alpha_i u}{\|u\|^2} \leq -\epsilon$$

uniformly in $i$ and $u$ in the designated set. Also notice the estimate

$$\frac{1}{\|u\|} B_{ii} F_i \leq -\frac{1}{R} |B_{ii}| F_i$$

and the uniform bound on the last summand on the right hand side of (C.6) for $t \leq t_0$ and $\|u\| \geq R$. The claim now follows from Lemma C.3. \hfill \Box

Below we shall make use of the following is easy-to-check auxiliary result on Riccati equations:

Lemma C.5. Let $A > 0, B \in \mathbb{R} \setminus \{0\}$, and $t_0 \geq 0$, and $G_0 < 0$. For $t \geq t_0$, the solution of

$$\partial_t G(t) = AG(t)^2 + BG(t), \quad G(t_0) = G_0$$

is of the form

$$G(t) = \frac{BG_0 e^{B(t-t_0)}}{(AG_0 + B) - AG_0 e^{B(t-t_0)}}.$$

If $B = 0$, then

$$G(t) = \frac{G_0}{1 - AG_0 (t - t_0)}.$$

From (C.4) we deduce the trivial differential inequality

$$\partial_t F_i \leq \alpha_{i,ii} F_i^2 + \frac{B_{ii}}{\|u\|} F_i.$$

By Lemma C.5, the solution of

$$\partial_t h = \alpha_{i,ii} h^2 + \frac{B_{ii}}{\|u\|} h$$

with $h(t_0) < 0$ is explicitly given by

$$h(t) = -\frac{G_{ii} e^{\frac{G_{ii}}{\|u\|}(t-t_0)}}{\|u\| \alpha_{i,ii} \left( e^{\frac{G_{ii}}{\|u\|}(t-t_0)} - 1 \right) - \frac{1}{h(t_0)}}, \quad t \geq t_0.$$

Together with Lemmas C.3 and C.4 and we thus obtain that

$$F_i(t, iu) \leq -\frac{G_{ii} e^{\frac{G_{ii}}{\|u\|}(t-t_0)}}{\|u\| \alpha_{i,ii} \left( e^{\frac{G_{ii}}{\|u\|}(t-t_0)} - 1 \right) + \frac{1}{\rho}}, \quad t \geq t_0$$

(C.7)
for all \( \|u\| \geq R \) with \( u^\top \alpha_i u \geq \epsilon \|u\|^2 \). By rescaling we infer
\[
f_i(t, iu) = \|u\|F_i(t\|u\|, iu)
\leq -\frac{\|u\|e^{B_{ii}(t-t_0)\|u\|^2}}{\|u\|\alpha_{i,ii}\frac{\alpha_{i,ii}}{B_{ii}} e^{B_{ii}(t-t_0)\|u\|^2} - 1 + \frac{1}{\rho}}
\]
for all \( \|u\| \geq R \) with \( u^\top \alpha_i u \geq \epsilon \|u\|^2 \). Integrating this inequality yields
\[
\int_0^t f_i(s, iu) \, ds \leq \int_0^t f_i(s, iu) \, ds
\]
(C.8)
\[
= -\frac{1}{\alpha_{i,ii}} \log \left( \|u\| \alpha_{i,ii}\frac{\alpha_{i,ii}}{B_{ii}} e^{B_{ii}(t-t_0)\|u\|^2} - 1 + \frac{1}{\rho} \right) - \log \left( \frac{1}{\rho} \right)
\]
for all \( \|u\| \geq R \) with \( u^\top \alpha_i u \geq \epsilon \|u\|^2 \). We arrive at the following key result, which completes the proof of Lemma C.2.

**Lemma C.6.** Let \( i \in \{1, \ldots, m\} \). For every \( \epsilon > 0 \) and \( t > 0 \) there exists some \( R > 0 \) and \( C > 0 \) such that
\[
\left| e^{\phi(t, iu) + \psi(t, iu)^\top x} \right| \leq C \left( 1 + \|u\| \right)^{-\frac{b_i}{\alpha_{i,ii}}}
\]
for all \( u \in \mathbb{R}^d \) with \( \|u\| \geq R \) and \( u^\top \alpha_i u \geq \epsilon \|u\|^2 \). Moreover,
\[
\Re \phi(t, iu) \leq -u_J^\top \left( \int_0^t e^{B_{JJ}s} a e^{B_{JJ}s} \, ds \right) u_J
\]
for all \( u \in \mathbb{R}^d \).

**Proof.** Integration of (4.2) implies
\[
\Re \left( \phi(t, iu) + \psi(t, iu)^\top x \right) \leq \Re \phi(t, iu)
\]
\[
\leq -u_J^\top \left( \int_0^t e^{B_{JJ}s} a e^{B_{JJ}s} \, ds \right) u_J + b_i \int_0^t f_i(s, iu) \, ds.
\]
Together with (C.8), this proves the lemma.

**Appendix D. Proof of Theorem 4.3**

Fix some \( \gamma > 0 \). We consider the two-dimensional process \( Z = (X, Y) \), where \( Y_t = y + \gamma \int_0^t X_s^2 \, ds \), with \( y \in \mathbb{R}_+ \). It is easy to see that \((X, Y)\) is an affine process with state space \( \mathbb{R}_+^2 \). In particular, if \( y = 0 \) we have that \( Y_t = \gamma \int_0^t X_s^2 \) has an exponentially affine characteristic function of the form
\[
\mathbb{E} \left[ e^{i v Y_t} \mid X_0 = x \right] = e^{i \phi(t, iv) + \psi(t, iv) x}, \quad v \in \mathbb{R},
\]
where the characteristic exponents $\phi$ and $\psi$ satisfy the generalized Riccati differential equations

\[
\partial_t \phi = b\psi + \int_0^\infty \left( e^{\psi \xi} - 1 \right) m(d\xi)
\]
\[
\phi(0) = 0
\]
\[
\partial_t \psi = \alpha \psi^2 + \beta \psi + i\gamma v + \int_0^\infty \left( e^{\psi \xi} - 1 \right) \mu(d\xi)
\]
\[
\psi(0) = 0.
\]

For any $v \neq 0$, we define the scaled functions\(^{17}\)

\[
F(t, iv) = \frac{1}{|v|} \Re \psi \left( \frac{t}{\sqrt{|v|}}, iv \right)
\]
\[
G(t, iv) = \frac{1}{|v|} \Im \psi \left( \frac{t}{\sqrt{|v|}}, iv \right).
\]

Then $F$ and $G$ satisfy

\[
\partial_t F = \alpha (F^2 - G^2) + \frac{\beta}{\sqrt{|v|}} F + \frac{1}{|v|} \int_0^\infty e^{\sqrt{|v|} F \xi} \left( \cos \left( \sqrt{|v|} F \xi \right) - 1 \right) \mu(d\xi)
\]
\[
F(0) = 0
\]
\[
\partial_t G = 2\alpha FG + \frac{\beta}{\sqrt{|v|}} G + \frac{\gamma}{|v|} v + \frac{1}{|v|} \int_0^\infty e^{\sqrt{|v|} F \xi} \sin \left( \sqrt{|v|} F \xi \right) \mu(d\xi)
\]
\[
G(0) = 0
\]

We now prove a first intermediary result, which is analogous to Lemma C.4:

**Lemma D.1.** There exists some $t_0 > 0$, $\rho > 0$ and $R > 0$ such that

\[
F(t_0, iv) \leq -\rho
\]

for all $v$ with $|v| \geq R$.

**Proof.** For $v \to \pm \infty$, the solutions $F(t, iv)$ and $G(t, iv)$ of (D.1) converge locally uniformly in $t$ to the solutions $F_\infty(t)$ and $G_\infty(t)$ of the system

\[
\partial_t F_\infty = \alpha (F_\infty^2 - G_\infty^2)
\]
\[
F_\infty(0) = 0
\]
\[
\partial_t G_\infty = 2\alpha F_\infty G_\infty \pm 1
\]
\[
G_\infty(0) = 0.
\]

Since $\partial_t G_\infty(t)|_{t=0} = \pm 1$ it follows that there exists some $t_1 > 0$ such that $G_\infty(t) \neq 0$ for all $t \in (0, t_1)$. This again implies that $F_\infty(t_0) < 0$ for some $t_0 \in (0, t_1)$, and the lemma follows. \(\square\)

From (D.1) we deduce the trivial differential inequality

\[
\partial F \leq \alpha F^2 + \frac{\beta}{\sqrt{|v|}} F.
\]

\(^{17}\)Note that here we have to scale by $\sqrt{|v|}$, which is in contrast to the proof of Theorem 4.1, see (C.3).
Arguing as for the derivation of (C.7), we then obtain together with Lemmas C.3 and D.1 that

$$F(t, iv) \leq -\frac{e^{\beta \sqrt{|v|} (t-t_0)}}{\sqrt{|v|} \left( e^{\beta \sqrt{|v|} (t-t_0)} - 1 \right) + \frac{1}{\rho}}, \quad t \geq t_0$$

for all $v \in \mathbb{R}$ with $|v| \geq R$. By rescaling and integrating we infer, arguing as for the derivation of (C.8), that

$$\int_0^t \Re \psi(s, iv) \, ds \leq \int_{t_0}^t \Re \psi(s, iv) \, ds$$

$$= -\frac{1}{\alpha} \left[ \log \left( \sqrt{|v|} \left( e^{\beta \left( t - \frac{t_0}{\sqrt{|v|}} \right)} - 1 \right) + \frac{1}{\rho} \right) - \log \left( \frac{1}{\rho} \right) \right]$$

$$= -\frac{1}{\alpha} \log \left( \rho \sqrt{|v|} \left( e^{\beta \left( t - \frac{t_0}{\sqrt{|v|}} \right)} - 1 \right) + 1 \right), \quad t \geq \frac{t_0}{\sqrt{|v|}}$$

for all $v \in \mathbb{R}$ with $|v| \geq R$. Similarly as in Lemma C.6 we now infer that

$$\left| e^{\phi(t, iv) + \psi(t, iv)x} \right| \leq C \left( 1 + |v| \right)^{-\frac{b}{2\alpha}}.$$ 

Combining this with Lemma C.1 completes the proof of Theorem 4.3.