CONVERGENCE RATES FOR DIFFUSIONS ON CONTINUOUS-TIME LATTICES

CLAUDIO ALBANESE AND ALEKSANDAR MIJATOVIĆ
IMPERIAL COLLEGE LONDON

Abstract. In this paper we introduce a discretization scheme based on a continuous-time Markov chain for the Black-Scholes diffusion process. Our principal aim is to find the optimal convergence rate for the probability density function of the discretized process as the distance h between the nodes of the state-space of the Markov chain goes to zero. The main theorem of the paper (theorem 4.1) states that the probability kernel $\mathbf{P}_t^h(x,y)$ of the discretized process converges at the rate $O(h^2)$ to the probability density function $p_t(x,y)$ of the diffusion process. We also show that this convergence is uniform in the state variables x and y and that the proposed discretization scheme converges at a rate which is no faster than $O(h^2)$.

Key Words: continuous-time Markov chains, spectral theory, functional calculus, convergence estimates for probability kernels

1. Introduction

Discretization schemes for stochastic processes are at the very core of modern mathematical finance. Their relevance is both theoretical, as they shed light on the nature of the stochasticity of the underlying process, and practical, since they lend themselves well to numerical methods. Consequently there has been a plethora of publications devoted to various aspects of the topic (e.g. (Kloeden & Platen 1992), (Howison, Dewynne & Wilmott 1995), (Glasserman 2004)).

In the seminal paper (Cox, Ross & Rubinstein 1979) the authors established binomial trees, and numerous generalizations thereof, as the paradigm for the constructive understanding of pricing theory. This led to algorithms for determining the numerical values of derivatives for a wide variety of models. The fundamental principle in this approach is to discretize time as well as space. Because of the theoretical interest in and wide applicability of binomial models, there has been a considerable expenditure of effort to extend and generalize these ideas (see for example (Hull & White 1988), (Madan, Milne & Shefrin 1991), (Derman, Kani & Chriss 1996)).

The fact that the prices obtained from these discretization schemes converge to the prices in continuous-time and -space models has been shown in (He 1990), (Amin & Khanna 1994) and elsewhere. The key issue of the *rate* of convergence of the discrete option price to its continuous limit is studied in (Heston & Zhou 2000). The authors show that the convergence rate of a put option is at least of the order $O(h)$, where $h^2 = \frac{c}{n}$ for some constant c and the number of time steps n in the binomial model. Looking at a wider class of payoff functions, the authors consider twice continuously differentiable functions defined on a bounded interval in $\mathbb{R}$ and show that the prices in that case converge at the rate $O(h^2)$. They propose to improve the convergence rate for non-differentiable payoffs by using smoothing strategies and applying the convergence result for smooth functions. This approach is well-suited to general vanilla payoffs since they are continuous and non-differentiable only at a single point. It is less applicable to functions characterized by a higher degree of irregularity such as the payoffs of European double digitals and butterfly spreads.

The probability density function (PDF) of a stochastic process can be expressed, in terms of pricing theory, as a current value of an option whose payoff equals the Dirac delta function. The question of the convergence rate of the probability density function of a discrete state-space model to the PDF of a continuous state-space model is an interesting problem, the solution of which has been hampered by the singular nature of the corresponding payoff: the Dirac delta function clearly does not possess the smoothness necessary for the application of the methods from (Heston & Zhou 2000). Other numerical discretization methods, such as the algorithms for solving partial differential equations described in (Howison et al. 1995)), are also ill-suited for solving PDEs with such singular boundary conditions.[1] In this paper we put forth a modelling paradigm which enables us to obtain the exact convergence rates for the probability density functions of the models therein.

The discretization scheme proposed in this paper is a continuous-time Markov chain. The idea of employing Markov chains to approximate diffusion processes is not new (see e.g. (Platen 1992)). In our main result, contained in theorem 4.1, we find the optimal convergence rate for the probability density function of the continuous-time lattice model Y_t^h (defined in section 2) to the PDF of the Brownian motion with drift Y_t (see (8) in section 3). It emerges that this convergence rate is $O(h^2)$ and no faster. The method of proof is based on detailed comparison between the spectrum of the Markov generator of Y_t^h and that of Y_t. We apply the same idea to conclude that the delta and the gamma of Arrow-Debreu securities also converge at the rate of $O(h^2)$.

The paper is organized as follows. In section 2 we define our continuous-time lattice model and find the spectral representation for its probability density function. In 2.1 we describe the model by defining its Markov generator in such a way that certain natural moment conditions are satisfied. In 2.2 we apply spectral theory to this Markov generator to get an integral representation of the probability density function. Section 3 recalls the spectral representation for the Black-Scholes probability kernel. Section 4 contains the main convergence result for the probability density functions. In section 5 we prove that the convergence rates for delta and gamma of Arrow-Debreu securities are of the order $O(h^2)$ (see theorem 5.1). Section 6 concludes the paper.

2. The discrete probability kernel

2.1. **Description of the model.** The stochastic process that we are interested in will be a continuous-time Markov chain that will be specified in terms of its Markov generator.[2] Intuitively

[1]The algorithm produces an approximate solution but the convergence rate of the approximation could be arbitrarily slow if we allow for arbitrarily singular payoffs.

[2]See chapter 6 in (Grimmett & Stirzaker 2001) for a general theory of Markov chains and their generators. See appendix A for a brief introduction to Markov generators of time-homogeneous diffusions.

one can think of a Markov generator, of a continuous-time Markov chain taking values in a state-space S, as a first order change of transition probabilities to jump from some state $x \in S$ to some other state $y \in S$ during the infinitesimal time interval dt. It is a well-known fact that a bounded operator $\mathcal{L} : l^2(S) \to l^2(S)$ (for definition of the Hilbert space $l^2(S)$ see appendix B) is a Markov generator for some stochastic process if and only if the following two conditions are satisfied: $\mathcal{L}(x, y) \geq 0$ for all $x \neq y$ and $\sum_{y \in S} \mathcal{L}(x, y) = 0$ for all $x \in S$.[3] The numbers $\mathcal{L}(x, y)$ are simply the coordinate values of the operator $\mathcal{L}$ with respect to the natural basis of $l^2(S)$. The first condition ensures that the transition probabilities are positive and the second implies that they sum to one. For a stochastic process X_t defined by the generator $\mathcal{L}$, the coordinate values of the operator can be interpreted in terms of the conditional probability density of X_t in the following way

$$\mathbb{P}(X_{t+dt} = y | X_t = x) = \mathcal{L}(x, y)dt, \;\; x \neq y, \quad \text{and} \;\; \mathbb{P}(X_{t+dt} = x | X_t = x) = 1 + \mathcal{L}(x, x)dt. \tag{1}$$

In this section we shall build our Markov generator and find a representation for the probability kernel of the underlying stochastic process. This will allow us to obtain explicit estimates on the convergence rates of the discrete kernel to the probability kernel of the Black-Scholes diffusion process. The state-space for our Markov chain will be the set $h\mathbb{Z}$, viewed as a subset of $\mathbb{R}$, where h is a small positive real number. The process will be a discrete version of Brownian motion with drift taking values in $h\mathbb{Z}$. In order to specify its Markov generator we need the following definitions.

Definition. Let $l^2(h\mathbb{Z})$ denote the Hilbert space of sequences indexed by $h\mathbb{Z}$ (i.e. maps from $h\mathbb{Z}$ to $\mathbb{C}$) as defined in appendix B (the measure on $h\mathbb{Z}$ assigns value 1 to each singleton in $h\mathbb{Z}$). A *discrete Laplace operator* $\Delta_h : l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$ is defined to be

$$\Delta_h f(x) := \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}, \quad x \in h\mathbb{Z},$$

for every sequence f in $l^2(h\mathbb{Z})$. A *discrete gradient operator* $\nabla_h : l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$ is defined in the following way

$$\nabla_h f(x) := \frac{f(x+h) - f(x-h)}{2h}, \quad x \in h\mathbb{Z},$$

for any $f \in l^2(h\mathbb{Z})$.

It is clear that these linear operators are well-defined. More to the point, the two operators are bounded (see appendix B for the definition) and therefore continuous. It is not hard to see that their respective operator norms are $\|\Delta_h\| = \frac{4}{h^2}$ and $\|\nabla_h\| = \frac{1}{h}$. These facts, together with the above definition, give a clear indication that in the limit, as h tends to zero, we reobtain the classical Laplace and gradient operators and that these operators must be unbounded.

Let us now choose a real drift μ and a positive log-normal volatility σ. Our aim is to define a generator of a Markov chain that will approximate well the process Y_t (given by (8) in section 3) which is a Brownian motion with drift. It is well-known that the behaviour of any time-homogeneous diffusion is uniquely determined by the first and the second moment of the

[3]For further discussion of the classification of Markov generators see section 6.10 in (Grimmett & Stirzaker 2001) and the references therein.

increment of the process in the infinitesimal time interval dt. These is because these two quantities determine the instantaneous drift and volatility, and hence specify uniquely the stochastic differential equation for Y_t, via[4]

$$\mathbb{E}[(Y_{t+dt} - Y_t)|\mathcal{F}_t] = \mu dt \quad \text{and} \quad \mathbb{E}[(Y_{t+dt} - Y_t)^2|\mathcal{F}_t] = \sigma^2 dt. \tag{2}$$

When defining our continuous-time Markov chain we will stipulate condition (2) as a natural restriction for our discretization scheme. It should be noted however that, since we are modelling in continuous time (i.e. the length of the time interval dt takes its limiting value zero), condition (2) is not, strictly speaking, necessary. In other words, we could choose a different discretization scheme which does not posses property (2) and yet converges to the correct limit. But it emerges that these natural conditions on the first two moments of the discrete process improve the rate of convergence.

The Markov generator of the discrete process we want to define is given as a bounded operator $\mathcal{L}_h : l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$ in the following way

$$\mathcal{L}_h := \mu \nabla_h + \frac{\sigma^2}{2} \Delta_h. \tag{3}$$

Our first task is to make sure that the operator $\mathcal{L}_h$ satisfies the conditions, from the beginning of this section, for being a Markov generator. The operator $\frac{1}{2}\Delta_h$ represents a diffusion with equal probabilities of jumping one state up or one state down from the current state, in the natural ordering of the set $h\mathbb{Z}$. Since no other states are reachable in the infinitesimal time interval dt, we get the following expression in coordinates $x, y \in h\mathbb{Z}$:

$$\Delta_h(x,y) = \frac{1}{h^2} \begin{cases} 1; & \text{if } y = x+h, \quad x-h, \\ -2; & \text{if } y = x, \\ 0; & \text{otherwise.} \end{cases}$$

It is clear that the effect of the volatility parameter σ is to amplify (or reduce, depending on its size) the first order change of the transition probabilities. It follows immediately from the discussion in the beginning of the section that the operator $\frac{1}{2}\Delta_h$ satisfies the conditions for being a Markov generator and that the process it specifies is a discrete state Brownian motion. This is so because, for the discrete Laplace operator, the equations in (1) imply zero drift and instantaneous volatility equal to one.

On the other hand the discrete gradient operator ∇_h, as defined above, is not a Markov generator since it has negative elements off the diagonal. It can be expressed in coordinates as follows:

$$\nabla_h(x,y) = \frac{1}{2h} \begin{cases} 1; & \text{if } y = x+h, \\ -1; & \text{if } y = x-h, \\ 0; & \text{otherwise.} \end{cases}$$

Since the sums of the elements in each row of ∇_h is zero, it is clear that the same will be true for the operator $\mathcal{L}_h$. The condition $\mathcal{L}_h(x,y) \geq 0$, for $x \neq y$, translates into conditions

[4]These formulae, when interpreted in the integral form, follow directly from the definitions of stochastic integrals and stochastic differential equations. See for example chapters 3 and 4 in (Øksendal 2003).

$\mathcal{L}_h(x, x+h) = \frac{\sigma^2}{2h^2} + \frac{\mu}{2h} \geq 0$ and $\mathcal{L}_h(x, x-h) = \frac{\sigma^2}{2h^2} - \frac{\mu}{2h} \geq 0$, for all $x \in h\mathbb{Z}$, which are equivalent to the inequalities

$$-\frac{\sigma^2}{h} \leq \mu \leq \frac{\sigma^2}{h}. \tag{4}$$

This condition however will be satisfied for an arbitrary real μ and positive σ, provided the lattice spacing h is small enough. Since we are interested in the behaviour of the discrete model in the limit as h goes to zero, we can without loss of generality assume that condition (4) holds. This implies that the operator $\mathcal{L}_h$, defined in (3), is a genuine Markov generator which describes a stochastic process Y_t^h. We should also note that the proposed discretization scheme is the unique scheme that yields a "tridiagonal" Markov generator which satisfies the conditions in (2). This can be seen by a straightforward calculation which involves finding a unique solution of a small linear system of equations.

It has by now emerged that our modelling paradigm for discretizing diffusion processes consists of using Markov generators that allow transition from the current state to the neighbouring states only. An instantaneous variance of a diffusion is a result of two contributions: one is the volatility of the process and the other is the drift. It should be noted that condition (4) says that, if we discretize the process as in (3), the lattice with spacing h can only support the variance with a drift component no larger than $\frac{\sigma^2}{h}$. We should note however that this does not imply that a continuous-time lattice with spacing h cannot support a drift of arbitrary size. If we redefined the generator $\mathcal{L}_h$ by specifying the gradient operator as $\nabla_h(f)(x) = \frac{f(x+h)-f(x)}{h}$, we would get a Markov chain which drifts at an arbitrarily large positive rate μ. Such a process is ill-suited in our case because it does not satisfy the second equality in (2) and, more importantly, converges to the continuous limit at a slower rate than Y_t^h.

2.2. **Spectral representation for the discrete probability kernel.** Recall that our general strategy is to use the operator $\mathcal{L}_h$ as a generator of a continuous-time Markov chain Y_t^h. In order to do this we will need to find the transitional probability density function $\mathbf{P}_t^h$ of the underlying stochastic process Y_t^h. To achieve this goal we will use functional calculus on the operator $\mathcal{L}_h$ in the sense of appendix C. But the prerequisite for applying theorem C.1 is to obtain a spectral decomposition of $\mathcal{L}_h$ first, which we will now do.

Consider the family of functions defined on the interval $[-\frac{\pi}{h}, \frac{\pi}{h}]$ in the following way

$$g_n : p \longmapsto \sqrt{\frac{h}{2\pi}} \exp(-inhp), \quad n \in \mathbb{Z}.$$

It is well-known that the functions g_n represent a countable orthonormal basis of the Hilbert space $L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$ with respect to the Lebesgue measure (see theorem II.9 in (Reed & Simon 1980)). This means that $\langle g_n, g_m \rangle = \delta_{nm}$, where

$$\langle \phi, \psi \rangle := \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \phi(p)\overline{\psi(p)}dp$$

is the inner product in $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ and δ_{nm} is the Kronecker delta, and that any function ϕ in the Hilbert space $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ can be represented uniquely by a convergent series

$$\phi = \sum_{n\in\mathbb{Z}} \langle \phi, g_n\rangle g_n. \tag{5}$$

It should be noted however that this series converges to f in the topology induced by the inner product on $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ and not necessarily pointwise in x.

We can now define an isometry $\mathcal{F}_h : l^2(h\mathbb{Z}) \to L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ which will allow us to find the spectral representation of the Markov generator $\mathcal{L}_h$. By viewing an element f of the Hilbert space $l^2(h\mathbb{Z})$ as a function from $h\mathbb{Z}$ to $\mathbb{C}$, we can introduce the following definition

$$\mathcal{F}_h(f)(p) := \sum_{n\in\mathbb{Z}} f(hn) g_n(p),$$

for any point p in the interval $[-\frac{\pi}{h},\frac{\pi}{h}]$. In the literature, the transformation $\mathcal{F}_h$ is sometimes referred to as a *semidiscrete Fourier transform* which maps the sequences from $l^2(h\mathbb{Z})$ to periodic functions on $\mathbb{R}$ with period $\frac{2\pi}{h}$. It follows from the definitions of the respective inner products on Hilbert spaces $l^2(h\mathbb{Z})$ and $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ that $\mathcal{F}_h$ is a unitary transformation (for definition see appendix B). Identity (5) implies that the inverse transform $\mathcal{F}_h^{-1}$ can be expressed as

$$\mathcal{F}_h^{-1}(\phi)(hn) = \langle \phi, g_n\rangle,$$

for any function ϕ in $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ and all points hn in the set $h\mathbb{Z}$.

We are now in the position to find a spectral representation of the generator $\mathcal{L}_h$. Let f be any sequence in $l^2(h\mathbb{Z})$. Then the following calculation holds in the Hilbert space $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ because all infinite sums that feature in it are well-defined elements of $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$:

$$\begin{aligned}
\mathcal{F}_h\Delta_h(f)(p) &= \sum_{n\in\mathbb{Z}} \frac{f(h(n+1)) + f(h(n-1)) - 2f(hn)}{h^2} g_n(p) \\
&= \frac{1}{h^2}\left(e^{ihp}\sum_{n\in\mathbb{Z}} f(hn)g_n(p) + e^{-ihp}\sum_{n\in\mathbb{Z}} f(hn)g_n(p) - 2\sum_{n\in\mathbb{Z}} f(hn)g_n(p)\right) \\
&= \frac{2(\cos(hp)-1)}{h^2}\mathcal{F}_h(f)(p).
\end{aligned}$$

A similar calculation reveals that $\mathcal{F}_h\nabla_h(f)(p) = \frac{e^{ihp}-e^{-ihp}}{2h}\mathcal{F}_h(f)(p)$. We have therefore found that the Markov generator $\mathcal{L}_h$ has a spectral representation of the form

$$\mathcal{F}_h\mathcal{L}_h\mathcal{F}_h^{-1}(\phi)(p) = \left(\mu\frac{e^{ihp}-e^{-ihp}}{2h} + \sigma^2\frac{\cos(hp)-1}{h^2}\right)\phi(p), \tag{6}$$

where ϕ is any element of $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$.

Let T be a time horizon and let us choose a time t before T. We can define a stochastic semigroup $\mathbf{P}_t^h := \exp((T-t)\mathcal{L}_h)$ by applying theorem C.1 to the spectral representation (6) of the bounded linear operator $\mathcal{L}_h$. In other words, for any pair of elements $x, y \in h\mathbb{Z}$, the coordinate value $\mathbf{P}_t^h(x,y)$, also known as the *probability kernel*, can be interpreted as a conditional probability density function $\mathbb{P}(Y_T^h = y | Y_t^h = x)$, where the process Y_t^h is a continuous-time Markov chain corresponding to the generator $\mathcal{L}_h$. This interpretation follows from theorem 10 in section 6.10 of (Grimmett & Stirzaker 2001) and the fact that the operator $\mathcal{L}_h$ is a Markov generator. On the other hand $\mathbf{P}_t^h(x,y)$, viewed as a function of the starting point x, is a solution of the backward Kolmogorov equation (see corollary A.3) for the continuous-time lattice process, where the boundary condition is given by a dirac delta function concentrated at y.

By applying functional calculus we will now establish a spectral representation for the discrete probability kernel which will, together with the representation from section 3, allow us to prove our central convergence result, given in theorem 4.1.

Let $x, y \in h\mathbb{Z}$ be of the form $x = hn$ and $y = hm$ and let δ_y denote the sequence in $l^2(h\mathbb{Z})$ that takes value 1 at the point y and value 0 everywhere else. As mentioned above we can then express $\mathbf{P}_t^h(x,y)$ as $\mathbf{P}_t^h(\delta_y)(x)$. Let $\tilde{\mathcal{L}_h}$ denote the diagonal operator given by the spectral representation (6) of the generator $\mathcal{L}_h$ (see also the first paragraph of appendix C for a definition of such a diagonal operator). Using the definition of functional calculus from appendix C, the probability kernel $\mathbf{P}_t^h(x,y)$ can be obtained directly in the following way:

$$\begin{aligned}\mathbf{P}_t^h(x,y) &= \mathcal{F}_h^{-1} e^{(T-t)\tilde{\mathcal{L}_h}} \mathcal{F}_h(\delta_y)(x) = \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{(T-t)F_{\mathcal{L}_h}(p)} g_m(p)\overline{g_n(p)}dp \\ &= \frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{F_{\mathcal{L}_h}(p)(T-t)} e^{ip(x-y)}dp. \qquad (7)\end{aligned}$$

Function $F_{\mathcal{L}_h}(p) := \sigma^2\frac{\cos(hp)-1}{h^2} + i\mu\frac{\sin(hp)}{h}$ in this expression is the diagonal multiplier that arose in the spectral representation of the Markov generator $\mathcal{L}_h$. Formula (7) for the discrete probability kernel has the crucial property that the time parameter t and the space parameters x, y feature in an independent way. That is, the kernel of integral (7) is a product of two functions, one depending on time t and the other depending on the dislocation $(x - y)$. It is precisely this feature of the spectral representation in (7), and the fact that in section 3 we will establish a similar representation for the probability kernel of the diffusion, that will allow us to find uniform bounds for convergence rates which are independent of the space coordinate.

3. Spectral representation for the Black-Scholes probability kernel

Let X_t be a stochastic process, starting at some positive real value X_0, which is given by the stochastic differential equation

$$dX_t = \nu X_t dt + \sigma X_t dW_t,$$

where W_t is the standard Brownian motion. The real constant ν is the average growth rate of the solution X_t and the positive number σ is the instantaneous log-normal volatility. Such a process is known in the literature as a *geometric Brownian motion* (GBM) and is used to model the stock price in the famous work (Black & Scholes 1973).

It is well-known (see for example chapter 5 in (Øksendal 2003)) that the process X_t can be expressed as an exponential of a Brownian motion with drift, given by

$$dY_t = \mu dt + \sigma dW_t, \qquad (8)$$

if the constant drift μ is set to be $(\nu - \sigma^2/2)$. It is therefore sufficient, for pricing purposes in general, to find the spectral representation of (8). The Markov generator $\mathcal{L}$, acting on any element f in $C_0^2(\mathbb{R})$, can be expressed as $\mathcal{L}f(x) = \mu\nabla f(x) + \frac{1}{2}\sigma^2\Delta f(x)$ by theorem A.1, where ∇

and Δ are the gradient and the Laplace operators respectively (see appendix A for definition). We will now show that the Fourier transform can be used to obtain the spectral representation of the unbounded operator $\mathcal{L}$ in the sense of appendix B.

Recall that the *Fourier transform* is defined, for any function f in the Banach space $L^1(\mathbb{R})$, in the following way

$$\mathcal{F}(f)(p) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ipx} dx.$$

The measure dx in this expression is the usual Lebesgue measure on the real line. Plancherel's theorem tells us that the Fourier transform $\mathcal{F}$ extends uniquely to a unitary map of $L^2(\mathbb{R})$ onto itself and that the inverse Fourier transform, defined by

$$\mathcal{F}^{-1}(g)(x) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(p) e^{ixp} dp,$$

extends uniquely to the adjoint of $\mathcal{F}$. For the proof of these statements see theorem IX.6 in (Reed & Simon 1980).

Let f be a twice differentiable function with compact support, i.e. $f \in C_0^2(\mathbb{R})$. One of the fundamental properties of the Fourier transform tells us that the following equalities hold (see lemma 1 of chapter IX in (Reed & Simon 1980))

$$\mathcal{F}(\nabla f)(p) = ip\mathcal{F}(f)(p) \quad \text{and} \quad \mathcal{F}(\Delta f)(p) = -p^2 \mathcal{F}(f)(p).$$

This yields a spectral representation for the operator $\mathcal{L}$ by substituting f with $\mathcal{F}^{-1}g$ into the identities above. In other words we find that the following holds

$$\mathcal{F}\mathcal{L}\mathcal{F}^{-1}(g)(p) = (i\mu p - \frac{\sigma^2}{2} p^2) g(p) \tag{9}$$

for every function g in $C_0^2(\mathbb{R})$.

Using this decomposition we can find a spectral representation for the probability density function $p_t(x,y)$ of the process Y_t, which can be intuitively described as $\mathbb{P}(Y_T = y | Y_t = x)$. It is a consequence of corollary A.3 and theorem C.1 that the probability kernel must be of the form $p_t(x,y) = e^{(T-t)\mathcal{L}} \delta(x-y)$, where δ is the Dirac delta function and the operator $e^{(T-t)\mathcal{L}}$, which is defined in appendix C, acts on the variable x.

By applying functional calculus and spectral representation (9) we find that the exponential operator can be expressed as $e^{(T-t)\mathcal{L}} = \mathcal{F}^{-1} e^{(T-t)\tilde{\mathcal{L}}} \mathcal{F}$, where $\tilde{\mathcal{L}}$ is an unbounded diagonal operator which acts on the elements of its domain by multiplying them with the function $F_{\mathcal{L}}(p) := i\mu p - \frac{\sigma^2}{2} p^2$ (for more details see appendix C). The spectral representation of the probability kernel therefore takes the form

$$p_t(x,y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{(i\mu p - \frac{\sigma^2}{2} p^2)(T-t)} e^{ip(x-y)} dp. \tag{10}$$

Since Y_t is just a Gaussian process, we know that the transition probability density function is of the form $p_t(x,y) = \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \exp(-\frac{(y-(\mu(T-t)+x))^2}{2\sigma^2(T-t)})$.

The spectral representation formula (10) is much more than another way of calculating a contour integral in the complex plane. It provides a representation of the probability kernel where time and space coordinates feature independently. We will need this property to make the uniform convergence estimates when passing from the discrete to the continuous model.

4. The convergence

In this section we shall examine the convergence of the discrete probability kernel $\mathbf{P}_t^h(x,y)$ in (7) to the conditional probability density function $p_t(x,y)$ (see (10)) of the Brownian motion with drift. We will establish the precise convergence rate of $\mathbf{P}_t^h(x,y)$ to $p_t(x,y)$ and prove that it is uniform in the state variables x and y (see theorem 4.1).

The limit as h goes to zero involves a passage from a discrete state-space $h\mathbb{Z}$ to the continuum of $\mathbb{R}$. Let us start by stating precisely how this limiting procedure should be interpreted. First of all we need to fix a strictly decreasing sequence of positive real numbers $(h_n)_{n\in\mathbb{N}}$ with the following two properties: $\lim_{n\to\infty} h_n = 0$ and $\frac{h_i}{h_j}$ is an integer for every index j which is larger than i. The first property is required because we wish to study the behaviour of transition probabilities as the lattice spacing goes to zero and the second ensures that the lattice $h_j\mathbb{Z}$ contains the lattice $h_i\mathbb{Z}$. A simple example of such a sequence is $h_n = (\frac{1}{2})^n$.

In all that follows we will be assuming that the distance h between two consecutive points in the lattice is equal to one of the elements of the sequence $(h_n)_{n\in\mathbb{N}}$. Similarly when dealing with limits where spacing h goes to zero, we will be assuming that the parameter h visits all elements of the sequence $(h_n)_{n\in\mathbb{N}}$ from some index n onwards. This technical assumption is very easy to satisfy and is essential in ensuring that the limit "as h goes to zero" is well-defined. We may now state our main result.

Theorem 4.1. *Let a positive real number T be a time horizon and let $t \in [0,T)$ denote the current time. Let $\mathbf{P}_t^h$ be the probability kernel (consisting of transition probabilities from time t to time T) of a stochastic process which takes values in $h\mathbb{Z}$, given by expression (7) in section 2. For any two elements x and y in $h\mathbb{Z}$ let $p_t(x,y)$ denote the coordinate expression for the probability kernel of the Brownian motion with drift, as defined in (10) of section 3. Then the following holds*

$$p_t(x,y) = \frac{1}{h}\mathbf{P}_t^h(x,y) + O(h^2)$$

and the error term $O(h^2)$ is independent of x and y (i.e. there exist positive constants C and δ such that the inequality $|p_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y)| \leq Ch^2$ holds for all $h < \delta$ and all $x, y \in h\mathbb{Z}$). Furthermore, this convergence rate of the discrete probability kernel is optimal in the following sense: for any function f with the property $\lim_{h\to 0} f(h) = 0$, the convergence at the rate $O(h^2 f(h))$ cannot be attained.

Before proceeding with the proof of theorem 4.1 we should note that, by definition, a function $f(x)$ is of type $O(x)$ if and only if it is bounded above by Mx, for a positive constant M, on some interval around 0, i.e. $\limsup_{x\to 0}\left|\frac{f(x)}{x}\right| \leq M$.

The statement that the discrete kernel cannot converge to the continuous kernel at a rate faster than $O(h^2)$ should not come as a surprise, because it follows from sections 2 and 3 that on the level of Markov generators the convergence rate is of the order $O(h^2)$ and not faster.

It should also be noted that $\mathbf{P}_t^h(x,y)$, as a function of y, defines a probability density function of a probability measure on the discrete state-space $h\mathbb{Z}$. The value $\mathbf{P}_t^h(x,y)$ for a fixed y must therefore be compared with the probability $\mathbb{P}(y \leq Y_t < y+h) = \int_y^{y+h} p_t(x,z)dz$ that the continuous process Y_t lies in the interval $[y, y+h)$. It is therefore clear that $\frac{1}{h}\mathbf{P}_t^h(x,y)$ should converge to $p_t(x,y)$, since the latter is a limit (as h goes to zero) of the average value of $p_t(x,z)$ on the interval $[y, y+h)$.

Proof. Let x and y be two arbitrary points in the state-space $h\mathbb{Z}$. Recall from sections 2 and 3 that we have the following representations for the probability density functions: $p_t(x,y) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{(i\mu p - \frac{\sigma^2}{2}p^2)(T-t)} e^{ip(x-y)}dp$ and $\frac{1}{h}\mathbf{P}_t^h(x,y) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{F_{\mathcal{L}_h}(p)(T-t)} e^{ip(x-y)}dp$, where $F_{\mathcal{L}_h}(p) = \sigma^2\frac{\cos(hp)-1}{h^2} + i\mu\frac{\sin(ph)}{h}$. It follows from the discussion at the beginning of this section that the points x, y are contained in $h\mathbb{Z}$ throughout the limiting process. Recall also that T is our time horizon and t is the current time. Let us denote time to expiry by $s := T - t$.

Our main aim is to estimate the behaviour of the difference of the two probability kernels which are both expressed as Riemann integrals. Our strategy is as follows. We will start by defining a positive function $K(h)$, which goes to infinity as h approaches 0 but is bounded above by $\frac{\pi}{h}$. This will enable us to express the difference of the kernels as a sum of two contributions: (a) the integral over the interval $[-K(h), K(h)]$ and (b) the contribution of the two kernels over the subset $[-\infty, -K(h)) \cup (K(h), \infty]$.

This is advantageous because the family of functions $p \mapsto \frac{\cos(hp)-1}{h^2}$ converges uniformly to the function $p \mapsto -p^2/2$ on the interval $[-K(h), K(h)]$, as long as $\lim_{h\to 0} K(h)^4 h^2 = 0$ (this follows directly from the Taylor expansion for cosine; see also figure 1). Since we are going to integrate this difference over the interval $[-K(h), K(h)]$, and will therefore obtain an additional factor of $K(h)$ in our final expression, we require that the following holds

$$\lim_{h\to 0} K(h)^5 h^2 = 0. \tag{11}$$

The second summand in the decomposition of the difference of the kernels consists of the integrals over set $\mathbb{R} - [-K(h), K(h)]$, which will be controlled by the rapid decay of the integrands in the representations (7) and (10). The definition of the function $K(h)$ is therefore determined by two opposing requirements: condition (11) means that $K(h)$ should go to infinity as slowly as possible, while the decay condition over the complement $\mathbb{R} - [-K(h), K(h)]$ requires $K(h)$ to go to infinity as quickly as possible. A compromise is reached by the function

$$K(h) := \sqrt{\frac{6}{s\sigma^2}\log(1/h)}, \tag{12}$$

where s is time to expiry as defined above (see figure 1 for an intuitive justification of the definition of $K(h)$).

Before proceeding to the calculation we should introduce another piece of notation:

$$f_h(p) := \frac{\cos(hp)-1}{h^2} + \frac{p^2}{2} \quad \text{and} \quad g_h(p) := i\left(\frac{\sin(hp)}{h} - p\right).$$

Notice that $f_h(p)$ is a non-negative function and that the following inequalities hold:

$$f_h(p) \leq \frac{h^2p^4}{4!} \qquad \text{and} \qquad |g_h(p)| \leq \frac{h^2p^3}{3!}, \quad \text{for} \quad p \in [0, K(h)], \tag{13}$$

when h satisfies $hK(h) < 1$. Both of these inequalities follow from the fundamental fact that the error of a partial sum of an alternating series, whose elements form a monotonically decreasing

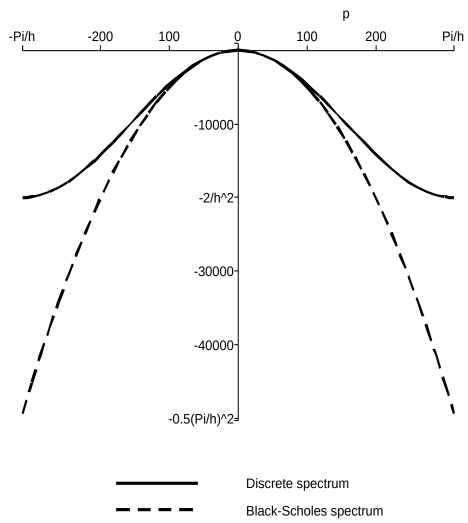

FIGURE 1. In the simple case with no drift ($\mu = 0$), volatility normalized to one ($\sigma = 1$) and time to maturity equal to 1 year, the function in the spectral representation of the Black-Scholes generator is the parabola $p \mapsto -\frac{p^2}{2}$ (see the graph above). Under the same conditions the spectral representation of the Markov generator for the discrete model with spacing h is given by the function $p \mapsto \frac{\cos(hp)-1}{h^2}$. Here, we plot this function for the value of h equal to 100^{-1}. Notice that the value of the function $K(h)$ (defined in (12)) at $h = 100^{-1}$ is less than 5.6. This observation gives intuitive justification to our intended strategy for the proof of theorem 4.1, since it is clear from the figure that the difference of the two spectral representations is negligible on the interval $[-K(h), K(h)]$ and that on its complement the spectra are mapped by the exponential to extremely small positive values.

sequence, is bounded above by the absolute value of the first element in the series, after the partial sum. The condition on h ensures that the elements of the series for $f_h(p)$ and $ig_h(p)$ are monotonically decreasing for all p in the interval $[0, K(h)]$, and therefore implies the inequalities in (13).

We may now proceed to our main task of estimating the difference of the two probability kernels, which we denote by $D(h) := |p_t(x,y) - \frac{1}{h}\mathbf{P}_t^h(x,y)|$. We get the following inequalities:

$$2\pi D(h) \leq \int_{-K(h)}^{K(h)} \left| e^{-\frac{p^2}{2}\sigma^2 s} e^{ip(s\mu+x-y)} \left(1 - e^{s(\mu g_h(p)+\sigma^2 f_h(p))}\right) \right| dp + \tag{14}$$

$$2\int_{K(h)}^{\pi/h} \left| e^{\sigma^2 s \frac{\cos(hp)-1}{h^2}} e^{is\mu\frac{\sin(hp)}{h}+ip(x-y)} \right| dp + \tag{15}$$

$$2\int_{K(h)}^{\infty} \left| e^{-\frac{p^2}{2}\sigma^2 s} e^{ip(s\mu+x-y)} \right| dp \tag{16}$$

$$\leq \int_{-K(h)}^{K(h)} e^{-\frac{p^2}{2}\sigma^2 s} \left| e^{s(\mu g_h(p)+\sigma^2 f_h(p))} - 1 \right| dp + \tag{17}$$

$$2\frac{\pi}{h} e^{\sigma^2 s \frac{\cos(hK(h))-1}{h^2}} + 2\int_{K(h)}^{\infty} e^{-\frac{p^2}{2}\sigma^2 s} dp. \tag{18}$$

The modulus of the exponential factor in the integral of (14) is equal to $e^{-\frac{p^2}{2}\sigma^2 s}$ and the integral is therefore bounded above by (17). The integral in (15) appears with the factor of 2 in front of it because the domain of integration consists of two symmetric regions, one of which is omitted. The first summand of (18) bounds above integral (15) because the function $p \mapsto \frac{\cos(hp)-1}{h^2}$ is monotonically decreasing on the interval $\left[K(h), \frac{\pi}{h}\right]$ (this is obvious since its derivative is negative).

In order to proceed with the estimates, we must recall the following fact which is an easy consequence of l'Hospital's rule:

$$\lim_{x\to\infty} \frac{\int_x^\infty p^\rho e^{-\lambda p^2} dp}{x^\rho e^{-\lambda x^2}} = 0, \quad \text{for any } \lambda \in (0,\infty) \text{ and } \rho \in [0,\infty). \tag{19}$$

In particular this implies that, for large x and ρ equal to zero, the integral in the numerator is less than the exponential $e^{-\lambda x^2}$. Using this fact, the elementary inequality[5]

$$|e^z - 1| \leq e^{|z|} - 1, \tag{20}$$

which is valid for any $z \in \mathbb{C}$, and the observation that functions $|g_h(p)|$ and $|f_h(p)|$ are even, we obtain the following estimate:

$$\pi D(h) \leq \int_0^{K(h)} \left(e^{s(|\mu g_h(p)|+\sigma^2|f_h(p)|)} - 1\right) e^{-\frac{p^2}{2}\sigma^2 s} dp + \frac{\pi}{h} e^{\frac{\cos(hK(h))-1}{h^2}\sigma^2 s} + e^{-\frac{K(h)^2}{2}\sigma^2 s}.$$

Applying the inequalities in (13) to the first two summands in the above expression yields

$$\begin{aligned}
\pi D(h) &\leq \int_0^{K(h)} \left(e^{s(|\mu|h^2p^3+\sigma^2h^2p^4)} - 1\right) e^{-\frac{p^2}{2}\sigma^2 s} dp + \frac{\pi}{h} e^{(-\frac{K(h)^2}{2}+h^2K(h)^4)\sigma^2 s} + e^{-\frac{K(h)^2}{2}\sigma^2 s} \\
&\leq \int_0^{K(h)} \left(e^{sh^2(|\mu|p^3+\sigma^2p^4)} - 1\right) e^{-\frac{p^2}{2}\sigma^2 s} dp + \frac{2\pi}{h} e^{-\frac{K(h)^2}{2}\sigma^2 s} + e^{-\frac{K(h)^2}{2}\sigma^2 s} \\
&\leq \pi C \left(\int_0^{K(h)} sh^2\left(|\mu|p^3 + \sigma^2p^4\right) e^{-\frac{p^2}{2}\sigma^2 s} dp + \frac{1}{h} e^{-\frac{K(h)^2}{2}\sigma^2 s}\right),
\end{aligned}$$

for small h and some positive constant C. The second inequality follows from the fact that $h^2K(h)^4$ is less than $\frac{\log(2)}{s\sigma^2}$, for small h, and the third follows from the elementary relation $e^x - 1 \leq 2x$ which is valid for small non-negative x. Note also that any function of the form $m(p)e^{-p^2\sigma^2 s/2}$, where $m(p)$ is a polynomial of any degree, is integrable over the real line.[6] Therefore the integral in the last inequality is finite. Now, by substituting the definition of $K(h)$ into the final expression, we get the inequality which holds for all sufficiently small h:

$$\frac{D(h)}{h^2} \leq C\left(\int_0^\infty s\left(|\mu|p^3 + \sigma^2p^4\right) e^{-\frac{p^2}{2}\sigma^2 s} dp + 1\right) < \infty.$$

This proves the convergence estimate in the theorem.

In order to prove the second part let us choose a function $f(h)$, such that $\lim_{h\to 0} f(h) = 0$, and let us assume that $D(h)$ converges to zero at the rate $O(h^2 f(h))$. Since we are now looking for a contradiction we can take $\mu = 0$, $\sigma^2 s = 1$ and $y = x$. If we also take $K(h) = \sqrt{8\log(\frac{1}{h})}$,

[5] This inequality follows from taking the absolute value of the Taylor series for the function $z \mapsto e^z - 1$ and applying the triangle inequality to it: $|e^z - 1| = |\sum_{n=1}^\infty z^n| \leq \sum_{n=1}^\infty |z|^n = e^{|z|} - 1$.

[6] This follows from the facts that $|m(p)|$ is bounded above by Ae^p for some constant A and that the function Ae^{-p^2+p} is bounded above by $Ae^{-(p-1)^2}$, which is clearly in $L^1(\mathbb{R})$.

we get, by a similar calculation as above, that the integrals in (15) and (16) converge to zero as fast as $O(h^3)$. Under our assumptions this implies that the integral $\int_0^{K(h)} \left(e^{f_h(p)} - 1\right) e^{-\frac{p^2}{2}} dp$ is of the order $O(h^2 g(h))$, where $g(h) = \max\{f(h), h\}$, since it can be bounded by a linear combination of $D(h)$, (15) and (16). The following calculation yields a contradiction

$$\frac{1}{h^2 g(h)} \int_0^{K(h)} \left(e^{f_h(p)} - 1\right) e^{-\frac{p^2}{2}} dp \geq \frac{1}{g(h)} \int_0^{K(h)} \left(\frac{p^4}{4!} - \frac{h^2 p^6}{6!}\right) e^{-\frac{p^2}{2}} dp,$$

since the last integral clearly converges to a finite positive value, while the limit $\lim_{h\to 0} g(h)$ equals 0. This concludes the proof.

□

5. The sensitivities of the probability kernel

In this section we are going to study the convergence rate of the PDF-delta $\nabla_h \mathbf{P}_t^h(x,y)$ and the PDF-gamma $\Delta_h \mathbf{P}_t^h(x,y)$ of the Markov chain to the corresponding quantities $\nabla p_t(x,y)$ and $\Delta p_t(x,y)$ for the Bronwian motion with drift. In all the cases gradient and Laplace operators act on the x-coordinate of the respective densities. The same approach as the one utilized in section 4 can be used to prove the following theorem.

Theorem 5.1. *Let $\mathbf{P}_t^h$ be the probability kernel given by expression (7) of section 2. For any two elements x and y in $h\mathbb{Z}$, let $p_t(x,y)$ denote the coordinate expression for the probability kernel of the Brownian motion with drift as defined in (10) of section 3. Then the following convergence relations hold for the PDF-delta and PDF-gamma*

$$\begin{aligned} \nabla p_t(x,y) &= \frac{1}{h}\nabla_h \mathbf{P}_t^h(x,y) + O\left(h^2\right), \\ \Delta p_t(x,y) &= \frac{1}{h}\Delta_h \mathbf{P}_t^h(x,y) + O\left(h^2\right) \end{aligned}$$

respectively. Furthermore the error terms $O\left(h^2\right)$ are independent of the coordinates x and y. In other words there exist positive constants C_1, C_2 and δ such that the inequalities $|\nabla p_t(x,y) - \frac{1}{h}\nabla_h \mathbf{P}_t^h(x,y)| \leq C_1 h^2$ and $|\Delta p_t(x,y) - \frac{1}{h}\Delta_h \mathbf{P}_t^h(x,y)| \leq C_2 h^2$ hold for all $h < \delta$ and all $x, y \in h\mathbb{Z}$.

Proof. In order to find the convergence rates for the delta and gamma of the probability density we shall apply the same idea as in the proof of theorem 4.1. More specifically we will use the spectral representations

$$\begin{aligned} \nabla_h \mathbf{P}_t^h(x,y) &= \frac{hi}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \frac{\sin(hp)}{h} e^{F_{\mathcal{L}_h}(p)(T-t)} e^{ip(x-y)} dp, \\ \nabla p_t(x,y) &= \frac{i}{2\pi} \int_{\mathbb{R}} p e^{(i\mu p - \frac{\sigma^2}{2}p^2)(T-t)} e^{ip(x-y)} dp, \\ \Delta_h \mathbf{P}_t^h(x,y) &= \frac{h}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \frac{2(\cos(hp)-1)}{h^2} e^{F_{\mathcal{L}_h}(p)(T-t)} e^{ip(x-y)} dp, \\ \Delta p_t(x,y) &= -\frac{1}{2\pi} \int_{\mathbb{R}} p^2 e^{(i\mu p - \frac{\sigma^2}{2}p^2)(T-t)} e^{ip(x-y)} dp, \end{aligned}$$

which follow from the properties of the gradient and Laplace operators, to determine the speed of convergence in much the same way as it was done for the probability kernel itself.

Let $K(h) := \frac{4}{\sigma\sqrt{s}}\sqrt{\log(1/h)}$ where $s = T - t$ is time to expiry. We will first determine the convergence behaviour of delta by analysing the above spectral representations on the disjoint sets $[-K(h), K(h)]$ and $\mathbb{R} - [-K(h), K(h)]$.

Let $D(h) := |\nabla p_t(x,y) - \frac{1}{h}\nabla_h \mathbf{P}^h_t(x,y)|$ be the difference we are trying to estimate and let us define the following quantities:

$$\begin{aligned} A(h) &:= \int_{-K(h)}^{K(h)} \left| \frac{\sin(hp)}{h} e^{F_{\mathcal{L}_h}(p)s} - pe^{(i\mu p - \frac{\sigma^2}{2}p^2)s} \right| dp, \\ B(h) &:= \int_{K(h)}^{\frac{\pi}{h}} \left| \frac{\sin(hp)}{h} \right| e^{\sigma^2 \frac{\cos(hp)-1}{h^2}s} dp, \\ C(h) &:= \int_{K(h)}^{\infty} pe^{-\frac{\sigma^2}{2}p^2 s} dp. \end{aligned}$$

By decomposing the integration region $\mathbb{R} = [-K(h), K(h)] \cup (\mathbb{R} - [-K(h), K(h)])$ of the above representations we find that the following inequality holds

$$2\pi D(h) \leq A(h) + 2(B(h) + C(h)). \tag{21}$$

It is clear that the bound $pe^{-\frac{\sigma^2}{4}p^2 s} < 1$ holds for all sufficiently large values of p. This fact, combined with (19) from section 4 and the definition of $K(h)$, implies that the following sequence of inequalities

$$C(h) \leq \int_{K(h)}^{\infty} e^{-\frac{\sigma^2}{4}p^2 s} dp \leq e^{-K(h)^2 \frac{\sigma^2 s}{4}} \leq h^4 \tag{22}$$

holds for all sufficiently small values of h. Similarly, since $\sin(hp) \leq 1$ for all real p, we get the following estimates

$$B(h) \leq \frac{1}{h}\int_{K(h)}^{\frac{\pi}{h}} e^{\sigma^2 \frac{\cos(hp)-1}{h^2}s} dp \leq \frac{\pi}{h^2} e^{\sigma^2 \frac{\cos(hK(h))-1}{h^2}s} \leq \frac{\pi e^{\sigma^2 s}}{h^2} e^{-K(h)^2 \frac{\sigma^2 s}{2}}$$

for all small enough values of h. The second inequality follows from the fact that the function $p \mapsto \frac{\cos(hp)-1}{h^2}$ is monotonically decreasing on the interval $\left[K(h), \frac{\pi}{h}\right]$. The third inequality is implied by the bound $\frac{\cos(hK(h))-1}{h^2} \leq -\frac{K(h)^2}{2} + 1$, which follows from the first inequality in (13) from section 4 coupled with the fact that $\lim_{h\to 0} h^2 K(h)^4 = 0$. By substituting the definition of $K(h)$ into the last inequality we obtain a bound on $B(h)$:

$$B(h) \leq \pi e^{\sigma^2 s} h^6. \tag{23}$$

We are now left with the task of finding a bound on $A(h)$. We start by adding and subtracting the summand $\frac{\sin(hp)}{h} e^{(i\mu p - \frac{p^2}{2}\sigma^2)s}$ in the above definition of $A(h)$ and then applying the triangle inequality. This yields the following bounds

$$A(h) \leq \int_{-K(h)}^{K(h)} \left| \frac{\sin(hp)}{h} \right| e^{-\frac{p^2}{2}\sigma^2 s} \left| e^{s(\mu g_h(p) + \sigma^2 f_h(p))} - 1 \right| dp + \int_{-K(h)}^{K(h)} \left| \frac{\sin(hp)}{h} - p \right| e^{-\frac{p^2}{2}\sigma^2 s} dp.$$

Functions $g_h(p)$ and $f_h(p)$ are defined on page 10. Since $\frac{\sin(x)}{x} \leq 1$ for all $x \in \mathbb{R}$, it follows that $\frac{\sin(hp)}{h} \leq p$ for all real numbers p. This fact, together with the second inequality in (13), implies the following:

$$A(h) \quad \leq \quad 2\int_0^{K(h)} pe^{-\frac{p^2}{2}\sigma^2 s}\left|e^{s(\mu g_h(p)+\sigma^2 f_h(p))}-1\right|dp + 2h^2\int_0^{K(h)}\frac{p^3}{3!}e^{-\frac{p^2}{2}\sigma^2 s}dp.$$

Inequalities (20) and (13) imply that the following holds

$$\left|e^{s(\mu g_h(p)+\sigma^2 f_h(p))}-1\right| \leq e^{sh^2(|\mu|p^3+\sigma^2 p^4)}-1 \leq 2sh^2\left(|\mu|p^3+\sigma^2 p^4\right)$$

for all $p \in [0, K(h)]$. The last inequality follows from the elementary relationship $e^x - 1 \leq 2x$ for small positive x and the fact that $\lim_{h\to 0} h^2 K(h)^4 = 0$. In particular this implies the inequality

$$A(h) \quad \leq \quad 2h^2\int_0^{\infty}\left(p\left(|\mu|p^3+\sigma^2 p^4\right)s + \frac{p^3}{3!}\right)e^{-\frac{p^2}{2}\sigma^2 s}dp. \tag{24}$$

Inequalities (21), (22), (23) and (24) together imply the convergence rate of $O(h^2)$ for the PDF-delta, which is clearly independent of the coordinates $x, y \in h\mathbb{Z}$.

In order to show that the same rate of convergence holds for the PDF-gamma one can follow the same scheme as the one above, where the expression $\frac{\sin(hp)}{h}$ is substituted by $\frac{2(\cos(hp)-1)}{h^2}$ and the linear polynomial p is replaced by the quadratic $-p^2$. Using these substitutions we can decompose the difference $|\Delta p_t(x,y) - \frac{1}{h}\Delta_h \mathbf{P}^h_t(x,y)|$ in the same way as in (21). Since the exponential function with a negative exponent decays faster than any polynomial of finite degree, inequality (22) is valid after the substitution. Since $|\cos(hp)|\leq 1$ we find that an upper bound on $B(h)$, analogous to (23), can be obtained as follows:

$$B(h) \leq \frac{4}{h^2}\int_{K(h)}^{\frac{\pi}{h}} e^{\sigma^2\frac{\cos(hp)-1}{h^2}s}dp \leq \frac{4\pi}{h^4}e^{\sigma^2\frac{\cos(hK(h))-1}{h^2}s} \leq \frac{4\pi e^{\sigma^2 s}}{h^3}e^{-K(h)^2\frac{\sigma^2 s}{2}} \leq 4\pi e^{\sigma^2 s}h^5.$$

Our final challenge is to find a bound on $A(h)$. Using the same idea as above we add and subtract the summand $\frac{2(\cos(hp)-1)}{h^2}e^{(i\mu p-\frac{p^2}{2}\sigma^2)s}$ in order to mimic the first step. It follows from the first inequality of (13) that $\frac{2(\cos(hp)-1)}{h^2} \leq -p^2 + p^4$ for all $p \in [0, K(h)]$. Since the integrand $|\frac{2(\cos(hp)-1)}{h^2} + p^2|$ is of the order $h^2 p^4$ and the function $p \mapsto p^4 e^{-\frac{p^2}{2}\sigma^2 s}$ is integrable over $\mathbb{R}$, the rest of the proof of the convergence of PDF-gamma proceeds in exactly the same way as in the case of delta. □

6. CONCLUSION

In this paper we defined a sequence of continuous-time Markov chains parametrized by lattice spacing whose limit is the stochastic process underlying the Black-Scholes model. Our main goal has been to study the convergence rate of the probability kernel of the Markov chains to the probability density function of the stochastic process in the continuous state-space. We found that the probability kernel of the process in the discrete state-space converges to the PDF of the continuous state-space process at the rate of $O(h^2)$ (see theorem 4.1). Furthermore this convergence is uniform in the state variables.

This result is somewhat surprising because the value of a probability density function can be viewed as a non-discounted price of an Arrow-Debreu security which is best described by its payoffs, namely the Dirac delta function concentrated at a given point of the domain of the underlying process. Theorem 4.1 can therefore be viewed as a convergence result for an option with a very singular payoff. This should be contrasted with a well-known fact (see for example (Heston & Zhou 2000)) that the smoothness of the payoff function is crucial for proving the convergence estimates for discretization schemes such as trinomial trees and PDE methods.

We also studied the convergence rates of sensitivities of the probability kernels as the lattice spacing goes to zero. We proved that both PDF-delta and PDF-gamma in our discretization scheme converge at the same rate as the densities (see theorem 5.1).

The reason for studying the convergence properties of this very simple family of continuous-time Markov chains lies in the fact that discretization of this kind, on the level of Markov generators, is extremely flexible and allows a direct generalization to processes of larger financial significance, such as jump-diffusions with stochastic volatility and other local Lévy processes. Recent research has shown (see (Albanese & Mijatović 2006) and (Albanese, Lo & Mijatović 2006)) that general continuous-time Markov chains can be used to define stationary processes which may be calibrated to the entire volatility surface without distorting their forward smiles. The convergence properties of these more complicated Markov chains remain unknown and constitute an attractive topic for future research.

Appendix A. Markov generators for time-homogeneous diffusions

A stochastic process X_t is called a *time-homogeneous diffusion* if it is a solution of a stochastic differential equation (SDE)

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{25}$$

where μ and σ are given deterministic functions of the form $\mu, \sigma : \mathbb{R} \to \mathbb{R}$ and W_t is a Brownian motion on some probability space.

Definition. Let X_t be a time-homogeneous diffusion in $\mathbb{R}$. The *Markov generator* $\mathcal{L}$ of X_t is a differential operator defined by

$$\mathcal{L}f(x) := \lim_{t\to 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t},$$

where the real value x is the starting point of the process X_t, i.e. $X_0 = x$, and the linear operator $\mathbb{E}^x[\cdot]$ is the expectation of any random variable that is a function on the process X_t which started at x. The set of all functions $f : \mathbb{R} \to \mathbb{R}$, such that the limit exists for all $x \in \mathbb{R}$, is the domain of the generator $\mathcal{L}$ and is denoted by $\mathcal{D}(\mathcal{L})$.

It is a well-known fact that the Markov generator $\mathcal{L}$ can be calculated explicitly in terms of the coefficients of SDE (25) if the function f satisfies certain regularity conditions which are specified in theorem A.1.

Theorem A.1. *Let X_t be a time-homogeneous diffusion given by stochastic differential equation (25) where μ and σ satisfy the linear-growth and Lipschitz conditions. Let $C_0^2(\mathbb{R})$ denote the set of all twice differentiable functions on $\mathbb{R}$ with compact support. Then the domain $\mathcal{D}(\mathcal{L})$, of the Markov generator $\mathcal{L}$ of X_t, contains $C_0^2(\mathbb{R})$ and*

$$\mathcal{L}f(x) = \mu(x)\nabla f(x) + \frac{1}{2}\sigma(x)^2\Delta f(x)$$

for every function f in $C_0^2(\mathbb{R})$. Here Δ and ∇ are the one-dimensional Laplace and gradient operators acting as $\Delta f(x) = f''(x)$ and $\nabla f(x) = f'(x)$.

The proof of theorem A.1 is based on Itô's lemma and can be found in section 7.3 of (Øksendal 2003). Since the set $C_0^2(\mathbb{R})$ is a dense subspace of the Hilbert space $L^2(\mathbb{R})$ the Markov generator $\mathcal{L}$ is a densely defined unbounded linear operator. The following theorem is a special case of the famous *Feynman-Kac formula* for diffusions. The proof of theorem A.2 can be found in chapter 8 of (Øksendal 2003).

Theorem A.2. *Let T be some time horizon and let X_t be the solution of SDE (25) for all $t \leq T$ with the Markov generator $\mathcal{L}$. For every element f in $C_0^2(\mathbb{R})$ we can define a function*

$$u : [0,T] \times \mathbb{R} \to \mathbb{R} \quad \text{as} \quad u(t,x) = \mathbb{E}^x[f(X_{T-t})]. \tag{26}$$

Then $u(t,\cdot)$ lies in $\mathcal{D}(\mathcal{L})$ for all $t \in [0,T]$ and the function u is a solution of the partial differential equation

$$\frac{\partial u}{\partial t}(t,x) + \mathcal{L}u(t,x) = 0 \quad \text{for all} \quad (t,x) \in [0,T] \times \mathbb{R}, \tag{27}$$

satisfying the boundary condition $u(T,x) = f(x)$. Moreover any bounded function in $C^{1,2}([0,T] \times \mathbb{R})$ which solves PDE (27) and satisfies the above boundary condition must be of the form (26).

Assume further that there exist a family of probability density functions $p_t(x,y) = \mathbb{P}(X_T = y|X_t = x)$ for the diffusion X_t. Given that the process X_t is time-homogeneous we have $p_t(x,y) = \mathbb{P}(X_T = y|X_t = x) = \mathbb{P}(X_{T-t} = y|X_0 = x)$ for any time t before the time horizon T. It follows from theorem A.2 that the solution $u(t,x)$ of PDE (27) can be expressed as

$$u(t,x) = \int_{\mathbb{R}} f(y) p_t(x,y) dy.$$

The probability kernel $p_t(x,y)$ is known (in the language of PDEs) as *Green's function* for problem (27). The following corollary is an easy consequence of theorem A.2.

Corollary A.3 (Backward Kolmogorov Equation)**.** *Let X_t be as in theorem A.2 and let T be a time horizon. Assume further that there exists a family of probability density functions $p_t(x,y) = \mathbb{P}(X_T = y|X_t = x)$, for all $t \in [0,T]$, which is smooth in t and twice differentiable in x. Then, for very $y \in \mathbb{R}$, the probability kernel $p_t(x,y)$ satisfies the following partial differential equation*

$$\frac{\partial p_t}{\partial t} + \mathcal{L}p_t = 0$$

with the boundary condition $p_T(x,y) = \delta(x-y)$, where δ is a Dirac delta function and the operator $\mathcal{L}$ acts on the coordinate x.

APPENDIX B. SPECTRAL REPRESENTATION FOR OPERATORS ON HILBERT SPACES

The general references for this section are (Reed & Simon 1980) and (Huston & Pym 1980). Let us start with some basic definitions.

Definition. Let $\mathcal{H}$ be a vector space equipped with an inner product, $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$, which gives rise to the norm on $\mathcal{H}$ defined as $\|x\| := \sqrt{\langle x, x \rangle_{\mathcal{H}}}$, for every $x \in \mathcal{H}$. The vector space $\mathcal{H}$ is a *Hilbert space* if every Cauchy sequence in $\mathcal{H}$, with respect to this norm, has a limit in $\mathcal{H}$.

In other words this definition is saying that a vector space is a Hilbert space if and only if it is complete with respect to the norm induced by the inner product. An example of a Hilbert space, denoted by $L^2(\mathbb{R},\mu)$, is the set of equivalence classes of measurable functions $f:\mathbb{R}\to\mathbb{C}$ such that

$$\int_{\mathbb{R}} f(x)\overline{f(x)}d\mu(x) < \infty,$$

where the integral is a Lebesgue integral over the real line with respect to the positive measure μ (for the proof of this fundamental fact see for example section 2.5 in (Huston & Pym 1980)). If the measure μ is concentrated on a discrete set M in $\mathbb{R}$ (i.e. $\mu(\mathbb{R}-M)=0$) then, since the set M must be countable[7], we can represent the above integral condition as

$$\sum_{m\in M} f(m)\overline{f(m)}\mu(m) < \infty.$$

The Hilbert space of all sequences $(f(m))_{m\in M}$, satisfying this condition, is in this case denoted by $l^2(M,\mu)$ (or simply by $l^2(M)$, if it is clear what the measure μ is).

A function $A:\mathcal{N}\to\mathcal{M}$, mapping a Hilbert space $\mathcal{N}$ into a Hilbert space $\mathcal{M}$, is called a *linear operator* if it is additive and invariant under scalar multiplication. If, in addition, the operator A preserves the inner product (i.e. $\langle Ax,Ay\rangle_{\mathcal{M}} = \langle x,y\rangle_{\mathcal{N}}$ for all $x,y\in\mathcal{N}$), then A is called a *unitary operator.*

Definition. Let $\mathcal{N}$ and $\mathcal{M}$ be Hilbert spaces and let $\mathcal{D}(A)$ be a dense[8] linear subspace of $\mathcal{N}$. A linear operator $A:\mathcal{D}(A)\to\mathcal{M}$ is *bounded* if the inequality $\|Ax\|_{\mathcal{M}} \leq C\|x\|_{\mathcal{N}}$ holds for all $x\in\mathcal{D}(A)$ and some positive real constant C. If no such constant exists we say that the operator A is *unbounded.*

Bounded operators are simpler to analyze than unbounded ones because of the plethora of desirable properties they exhibit. For example a bounded operator A can always be extended uniquely to the closure of its domain $\mathcal{D}(A)$ (i.e. the whole Hilbert space $\mathcal{N}$) as a continuous linear function which maps $\mathcal{N}$ to $\mathcal{M}$ (the proof of these basic facts can be found in chapter I of (Reed & Simon 1980)). No such statement is true for unbounded operators. On the other hand we are forced to consider the unbounded case, because the Markov generators of time-homogeneous diffusions are differential operators, which are, by their very nature, unbounded[9].

Bounded operators however come in ample supply. For example it follows directly form the definition that any unitary operator is a bounded linear operator. It should also be noted that, if the Hilbert space $\mathcal{N}$ is finite-dimensional, all linear operators on $\mathcal{N}$ are necessarily bounded and

[7]A discrete subspace of a second countable topological space cannot have cardinality larger than natural numbers $\mathbb{N}$.

[8]A subset Y of a topological space X is said to be *dense* in X, if the only closed set containing Y is X itself.

[9]If we take the operator A to equal the simplest possible differential operator ∇, it can easily be seen that there exists no constant C, such that the L^2 norm of ∇f is bounded above by $C\|f\|$ for all differentiable functions f defined on $\mathbb{R}$.

therefore continuous. From the modelling perspective this implies that the Markov generators for processes defined on finite state-spaces are necessarily continuous linear operators.

It is a fortunate fact that, from the point of view of spectral representation, both bounded and unbounded linear operators on Hilbert spaces behave in much the same way. Before we can elaborate on this point, we need to recall one more fundamental concept.

Definition. Let A be a (possibly unbounded) linear operator on a Hilbert space $\mathcal{H}$. The *resolvent set* of A is the set of all complex numbers λ, such that the inverse $(A - \lambda I_{\mathcal{H}})^{-1}$ exists and is a bounded operator on $\mathcal{H}$ (the operator $I_{\mathcal{H}}$ is the identity on $\mathcal{H}$). The *spectrum* of A, denoted by $\sigma(A)$, is the complement (in $\mathbb{C}$) of the resolvent set.

If the Hilbert space $\mathcal{H}$ is finite-dimensional, then the spectrum of A consists solely of the set of eigenvalues of A. If, on the other hand, $\mathcal{H}$ is infinite-dimensional, then the spectrum $\sigma(A)$ is no longer necessarily discrete but is still a closed subset of the complex plane. In the case of a unitary operator it is not difficult to see directly from the definition that every element of the spectrum must have modulus equal to 1. An effective way of understanding any operator A on a Hilbert space $\mathcal{H}$ is to understand its spectrum $\sigma(A)$. In order to achieve the latter one usually resorts to some sort of spectral representation.

Definition. Let A be a (possibly unbounded) linear operator with a domain $\mathcal{D}(A)$ in a Hilbert space $\mathcal{H}$. The *spectral representation* of A consists of a pair (M, μ), where M is a measure space and μ is a positive measure on M, together with a unitary operator $U : \mathcal{H} \to L^2(M, \mu)$ and a measurable function $F_A : M \to \mathbb{C}$ such that the following holds

$$(UAU^{-1}f)(x) = F_A(x)f(x)$$

for all functions f in the dense linear subspace $U(\mathcal{D}(A))$ of the Hilbert space $L^2(M, \mu)$.

It follows from this definition that the spectral representation of an operator A, defined on a finite-dimensional space, is equivalent to diagonalizing a matrix that represents the operator A in some basis. If we assume that the matrix for A can be diagonalized and that its eigenvalues are all distinct, then the ingredients of the spectral decomposition can be described easily as follows: the space M consists of a discrete set of eigenvalues (i.e. $M = \sigma(A)$), μ is a positive measure which assigns a non-zero weight to every point in M and the function F_A is a natural inclusion of M into $\mathbb{C}$. The eigenvector corresponding to an element λ of M is simply the indicator function of the set $\{\lambda\} \subset M$.

It comes as no surprise that the spectral representation of an operator, as described in the previous definition, might not exist, since many matrices are not equivalent to diagonal transformations. Nevertheless in spectral theory, much like in linear algebra, there are sufficient conditions which guarantee that a bounded or unbounded linear operator possesses a spectral representation (see chapters VII and VIII in (Reed & Simon 1980) and chapters 9 and 10 in (Huston & Pym 1980)). We make no use of these important results here, because we are able to find an explicit spectral representation of the Markov generators defined in sections 3 and 2.

The main reason we are interested in the spectral representation of operators on Hilbert spaces is because it allows us to apply functional calculus in a straightforward way (see appendix C).

Appendix C. Functional calculus

Let $\mathcal{D}(A)$ be a domain of an unbounded operator A on some Hilbert space $\mathcal{H}$, which allows a spectral representation as described in appendix B. In other words the operator $\tilde{A}f := UAU^{-1}f = F_A f$ is a (possibly unbounded) diagonal linear operator defined for all $f \in \mathcal{D}(\tilde{A}) = U(\mathcal{D}(A))$ (i.e. $f = Ug$ for some $g \in \mathcal{D}(A)$), which is a dense subset of the Hilbert space $L^2(M, \mu)$. The mapping U is a unitary linear operator from $\mathcal{H}$ onto $L^2(M, \mu)$.

Let $\phi : \mathbb{C} \to \mathbb{C}$ be an entire function.[10] Our goal is to define an operator $\phi(A)$. It is clear how to proceed in the case of a diagonal operator $\tilde{A}$ where we can define $(\phi(\tilde{A})f)(x) := (\phi \circ F_A)(x)f(x)$ for all f in $\mathcal{D}(\phi(\tilde{A})) = \{h \in L^2(M, \mu) : (\phi \circ F_A)h \in L^2(M, \mu)\}$. Assuming that $\mathcal{D}(\phi(\tilde{A}))$ is a dense subspace of $L^2(M, \mu)$, we can define an operator $\phi(A)$ in the following way

$$\phi(A) := U^{-1}\phi(\tilde{A})U,$$

where $U : \mathcal{H} \to L^2(M, \mu)$ is a unitary operator as above. Since $\mathcal{D}(\phi(\tilde{A}))$ is a dense subspace in $L^2(M, \mu)$ by assumption, the domain $\mathcal{D}(\phi(A)) = U^{-1}\mathcal{D}(\phi(\tilde{A}))$ must be dense in $\mathcal{H}$ which makes $\phi(A)$ a well-defined (possibly unbounded) linear operator.

The reason why we are interested in spectral analysis is contained in the following theorem.

Theorem C.1. *Let $A : \mathcal{D}(A) \to \mathcal{H}$ be a (possibly unbounded) linear operator whose domain $\mathcal{D}(A)$ is a dense subset of the Hilbert space $\mathcal{H}$. Assume also that there exists a spectral representation of A as described above. Then a differential equation*

$$\frac{\partial u}{\partial t} + Au = 0, \qquad u(T) = \psi \tag{28}$$

for a function $u : [0, T] \to \mathcal{D}(A)$ with a boundary value ψ in the space $U^{-1}(\mathcal{D}(\tilde{A}e^{T\tilde{A}}) \cap \mathcal{D}(e^{T\tilde{A}}))$ has a solution of the form

$$u(t) = \exp((T - t)A)\psi, \tag{29}$$

where the linear operator $\exp((T - t)A)$ is defined as in the discussion above for the entire function $\phi(z) = \exp((T - t)z)$. The operator $\tilde{A} = UAU^{-1}$ is a "diagonal version" of A and U is the unitary transformation as described above.

Before proceeding to the proof of theorem C.1 we should note that the derivative $\frac{\partial u}{\partial t}(t_0)$, for any $t_0 \in [0, T]$, is a vector in the Hilbert space $\mathcal{H}$ with the property

$$\lim_{\delta \to 0} \frac{1}{\delta} \|u(t_0 + \delta) - u(t_0) - \delta \frac{\partial u}{\partial t}(t_0)\|_{\mathcal{H}} = 0. \tag{30}$$

We should also note that the condition in theorem C.1 on the function $U\psi$ being in a certain subspace is, in the case of differential operators, just a condition on the decay rate of the function $U\psi$. Since we are interested in the local properties of solution (29), the role of $U\psi$ will be played by smooth functions with compact support which clearly satisfy any condition on the rate of decay. On the other hand if the operator A in theorem C.1 is bounded, then the condition on

[10] A complex function defined on $\mathbb{C}$ is termed *entire* if and only if the convergence radius of its Taylor series around 0 is infinite.

boundary value is satisfied for all $\psi \in \mathcal{H}$, since in that case the spectrum of A is a bounded subset of $\mathbb{C}$.

The importance of theorem C.1 for mathematical finance lies in the fact that it allows us to obtain an explicit form for the transition probabilities of the underlying, which can then be used for pricing of derivatives.

Proof. We shall proceed in two steps as follows. First we will apply the differential operator $\frac{\partial}{\partial t}$ to the candidate solution (29). In the second step we will apply the operator A on (29) and verify that the two procedures yield quantities which are equal in size and opposite in sign. This will prove the theorem.

Let $\psi \in D(A)$ be the chosen boundary value for the solution of equation (28). By our assumption the operator A has a spectral decomposition. In other words there exists a diagonal operator $\tilde{A}$, on a Hilbert space $L^2(M,\mu)$, defined as $\tilde{A}f(x) = F_A(x)f(x)$ for some measurable function F_A which maps the measure space M into $\mathbb{C}$. Furthermore there exists a unitary operator $U : \mathcal{H} \to L^2(M,\mu)$ with the property $A = U^{-1}\tilde{A}U$. We therefore have the following equalities

$$e^{(T-t)A}\psi = U^{-1}e^{(T-t)\tilde{A}}U\psi = U^{-1}\left(e^{(T-t)F_A} \cdot U\psi\right).$$

It should be noted that the vector $e^{(T-t)F_A} \cdot U\psi$ in $L^2(M,\mu)$ is just a product of the functions $e^{(T-t)F_A(x)}$ and $(U\psi)(x)$ for all $x \in M$. Since U is a unitary operator and hence norm preserving, the differential operator $\frac{\partial}{\partial t}$ commutes with U^{-1}. We therefore get $\frac{\partial}{\partial t}\left(e^{(T-t)A}\psi\right) = U^{-1}\left(\frac{\partial}{\partial t}\left(e^{(T-t)F_A} \cdot U\psi\right)\right)$. We now need to show that

$$\frac{\partial}{\partial t}\left(e^{(T-t)F_A} \cdot U\psi\right) = -F_A \cdot \left(e^{(T-t)F_A} \cdot U\psi\right), \tag{31}$$

which is by definition (see (30) above) equivalent to

$$\lim_{\delta\to 0}\frac{1}{\delta^2}\int_M \left|e^{(T-t-\delta)F_A} - e^{(T-t)F_A} - \delta F_A e^{(T-t)F_A}\right|^2 |U\psi|^2 d\mu = 0.$$

Since $U\psi$ lies in $\mathcal{D}(\tilde{A}e^{T\tilde{A}}) \cap \mathcal{D}(e^{T\tilde{A}})$ by assumption, this integral must exist for small positive delta. Since we are dealing with the exponentials of ordinary functions, we can rewrite the integrand as $\left|e^{-\delta F_A} - e^{F_A} - \delta F_A\right|^2 |e^{(T-t)F_A}U\psi|^2$. This expression, divided by δ^2, clearly converges to zero for all x in the space M. From the dominated convergence theorem we can conclude that the integral must also converge to zero. This proves claim (31).

It is now clear that applying the operator $\frac{\partial}{\partial t}$ to $e^{(T-t)A}$ yields the following result

$$\frac{\partial}{\partial t}e^{(T-t)A}\psi = -U\tilde{A}e^{(T-t)\tilde{A}}U\psi.$$

On the other hand note that the equality $AU^{-1} = U\tilde{A}$ implies

$$Ae^{(T-t)A}\psi = AU^{-1}e^{(T-t)\tilde{A}}U\psi = U\tilde{A}e^{(T-t)\tilde{A}}U\psi,$$

which completes the proof of the theorem. □

References

Albanese, C. & A. Mijatović (2006), 'A stochastic volatility model for risk-reversals in foreign exchange', *International Journal of Theoretical and Applied Finance, forthcoming* .

Albanese, C., H. Lo & A. Mijatović (2006), 'Spectral methods for volatility derivatives', *working paper* .

Amin, K. & A. Khanna (1994), 'Convergence of American option values from discrete to continuous-time financial models', *Math. Finance* **4**, 289–304.

Black, F. & M. Scholes (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Cox, J. C., S. A. Ross & M. Rubinstein (1979), 'Option pricing: a simple approach', *Journal of Financial Economics* **7**, 229–263.

Derman, E., G. Kani & N. Chriss (1996), 'Implied trinomial trees of the volatility smile', *Goldman Sachs Quantitative Strategies Research Notes* .

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering*, Applications of Mathematics, 53, Springer-Verlag Berlin Heidelberg New York.

Grimmett, G. & D. Stirzaker (2001), *Probability and Random Processes*, 3rd edn, Oxford University Press.

He, H. (1990), 'Convergence from discrete to continuous time contingent claims prices', *Rev. Financial Stud.* **3**, 523–546.

Heston, S. & G. Zhou (2000), 'On the rate of convergence of discrete-time contingent claims', *Mathematical Finance* **10**(1), 53–75.

Howison, S., J. Dewynne & P. Wilmott (1995), *The Mathematics of Financial Derivatives*, Cambridge University Press.

Hull, J. & A. White (1988), 'The use of the control variate technique in option pricing', *J. Financial Quant. Anal.* **23**, 237–251.

Huston, V. & J. S. Pym (1980), *Applications of Functional Analysis and Operator Theory*, Academic Press.

Kloeden, P. E. & E. Platen (1992), *Numerical Solutions of Stochastic Differential Equations*, Applications of Mathematics, 23, Springer-Verlag Berlin Heidelberg New York.

Madan, D., F. Milne & H. Shefrin (1991), 'The multinomial option pricing model and its Brownian and Poisson limits', *Review of Financial Studies* **2**, 251–265.

Øksendal, B. (2003), *Stochastic Differential Equations*, 6th edn, Springer-Verlag Berlin Heidelberg.

Platen, E. (1992), 'Higher order weak approximation of Itô diffusions by Markov chains', *Probability in the Engineering and Information Sciences* **6**, 391–408.

Reed, M. & B. Simon (1975), *Methods of Modern Mathematical Physics 2: Fourier Analysis, Self-Adjointness*, Academic Press.

Reed, M. & B. Simon (1980), *Methods of Modern Mathematical Physics 1: Functional Analysis*, Academic Press.

CLAUDIO ALBANESE, DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON
E-mail address: `claudio.albanese@imperial.ac.uk`

ALEKSANDAR MIJATOVIĆ, INSTITUTE FOR MATHEMATICAL SCIENCES, IMPERIAL COLLEGE LONDON
E-mail address: `a.mijatovic@imperial.ac.uk`