AN OPTIMAL POLYNOMIAL APPROXIMATION OF
BROWNIAN MOTION *

JAMES FOSTER†, TERRY LYONS†, AND HARALD OBERHAUSER†

Abstract. In this paper, we will present a strong (or pathwise) approximation of standard Brownian motion by a class of orthogonal polynomials. Remarkably the coefficients obtained from the expansion of Brownian motion in this polynomial basis are independent Gaussian random variables. Therefore it is practical (requires $N$ independent Gaussian coefficients) to generate an approximate sample path of Brownian motion that respects integration of polynomials with degree less than $N$. Moreover, since these orthogonal polynomials appear naturally as eigenfunctions of the Brownian bridge covariance function, the proposed approximation is optimal in a certain weighted $L^2(P)$ sense. In addition, discretizing Brownian paths as piecewise parabolas gives a locally higher order numerical method for stochastic differential equations (SDEs) when compared to the piecewise linear approach. We shall demonstrate these ideas by simulating Inhomogeneous Geometric Brownian Motion (IGBM). This numerical example will also illustrate the deficiencies of the piecewise parabola approximation when compared to a new version of the asymptotically efficient log-ODE (or Castell-Gaines) method.

Key words. Brownian motion, polynomial approximation, numerical methods for SDEs

AMS subject classifications. 41A10, 60H35, 60J65

1. Introduction. Brownian motion is a central object for modelling real-world systems that evolve under the influence of random perturbations [1]. In applications where methods discretize Brownian motion, usually only increments of the path are generated [2]. In this setting, the best $L^2(P)$ approximation of Brownian motion that is measurable with respect to these increments is given by the piecewise linear path that agrees on discretization points [3]. This motivates the following natural question: Are there better discrete approximations of Brownian motion than piecewise linear?

The next simplest approximant would be a piecewise polynomial, though it is not clear whether this would be advantageous for tackling problems such as SDE simulation. This paper can be viewed as a logical continuation of [4], where a polynomial wavelet representation of Brownian motion was proposed. These wavelets were constructed to capture certain “geometrical features” of the path, namely the integrals of the Brownian motion against monomials. The goal of this paper is to convince the reader of the practical implications of these polynomials in the numerical analysis of SDEs.

!!![Fig. 1.1. Sample paths of Brownian motion with corresponding polynomial approximations.]

*Research supported by the Engineering and Physical Sciences Research Council [EP/N509711/1].
†Mathematical Institute, University of Oxford, Woodstock Road, Oxford, OX2 6GG. (james.foster@maths.ox.ac.uk, terry.lyons@maths.ox.ac.uk, harald.oberhauser@maths.ox.ac.uk)
The paper is organised as follows. In Section 2, we shall state and prove the main result of the paper. This will be a Karhunen-Loève theorem for the Brownian bridge, where the orthogonal functions used in the approximation turn out to be polynomials. In Section 3, we will investigate some significant consequences of the main theorem. In particular, the following theorem can be proved immediately from the main result.

**Theorem 1.1.** Let $W$ denote a standard real-valued Brownian motion on $[0, 1]$. Let $W^n$ be the unique $n$-th degree polynomial with a root at 0 and satisfying

\[ \int_0^1 u^k dW^n_u = \int_0^1 u^k dW_u, \quad \text{for} \quad k = 0, 1, \cdots, n - 1. \]

Then $W = W^n + Z^n$, where $Z^n$ is a centred Gaussian process independent of $W^n$.

The above theorem has a simple yet striking conclusion, namely that polynomials can be unbiased approximants of Brownian motion. Moreover, the first non-trivial case ($n = 2$) already has interesting applications within the numerical analysis of SDEs. The reason is that parabolas can capture the “space-time area” of Brownian motion.

**Fig. 1.2.** Brownian motion can be expressed as a (random) parabola plus independent noise. Moreover, the approximating parabola has the same increment and time integral as the original path.

Therefore discretizing Brownian motion using a piecewise parabola gives a locally high order methodology for numerically solving one-dimensional SDEs. However, since certain triple iterated integrals of Brownian motion and time are partially matched by these parabolas, we expect this method to have only an $O(h)$ rate of convergence (where $h$ denotes the step size used). This gives motivation for the following theorem:

**Theorem 1.2.** Let $\tilde{W}$ be the unique quadratic polynomial with a root at 0 and

\[ \tilde{W}_1 = W_1, \quad \int_0^1 \tilde{W}_u du = \int_0^1 W_u du. \]

Then the following third order iterated integral of Brownian motion can be estimated:

\[ \mathbb{E} \left[ \int_0^1 W_u^2 du \mid W_1, \int_0^1 W_u du \right] = \int_0^1 \tilde{W}_u^2 du + \frac{1}{15}. \]

The above theorem can be directly incorporated into the stochastic Taylor method as well as the log-ODE or Castell-Gaines method (see [5], [6]). The hope is that by estimating this non-trivial iterated integral with its conditional expectation, we can design numerical methods that enjoy high orders of both strong and weak convergence. Specifically, for a general one-dimensional SDE with sufficiently regular vector fields, numerical methods that correctly utilize the above conditional expectation will have a strong convergence rate of $O(h^{\frac{3}{2}})$ as well as a weak convergence rate of $O(h^2)$. Since the piecewise parabola approach doesn’t entirely match the expectation of this iterated integral, it will have a lower order of convergence than the log-ODE method.
In Section 4, we shall demonstrate the above ideas through various discretizations of Inhomogeneous Geometric Brownian Motion (IGBM)

\[ dy_t = a(b - y_t) dt + \sigma y_t dW_t, \]

where \(a \geq 0\) and \(b \in \mathbb{R}\) are the mean reversion parameters and \(\sigma \geq 0\) is the volatility.

In mathematical finance, IGBM is an example of a short rate model that can be both mean-reverting and non-negative. As a result it is suitable for modelling interest rates, stochastic volatilities and default intensities [7]. From a mathematical viewpoint, IGBM is one of the simplest SDEs that has no known method of exact simulation [8]. By incorporating the ideas provided by the main theorem into the log-ODE method, we will produce a state-of-the-art numerical approximation of IGBM. Although the vector fields for IGBM are not bounded, our numerical evidence indicates that the method has a strong convergence rate of \(O(h^{3/2})\) and a weak convergence rate of \(O(h^2)\).

1.1. Notation. Below is some of the notation that is used throughout the paper.

- \(W\) denotes a standard real-valued Brownian motion. It is often convenient to write \(W\) with an additional coordinate corresponding to time, i.e. \(W_t(0) := t\).
- Any Stratonovich SDE on the interval \([0, T]\) in this paper will be of the form
  \[ dy_t = f_0(y_t) dt + f_1(y_t) \circ dW_t, \]
  \[ y_0 = \xi, \]
  where \(y, \xi \in \mathbb{R}^e\), and \(f_i : \mathbb{R}^e \to \mathbb{R}^e\) denote the vector fields.
- (Similarly, Itô SDEs will be defined on fixed intervals and have the same form)
  - \(B\) denotes a standard real-valued Brownian bridge on \([0, 1]\).
  - \([s, t]\) is a general closed subinterval of \([0, T]\).
  - \(h\) is the step size at which a numerical solution is propagated, usually \(h = t - s\).
  - \(\{e_k\}_{k \geq 1}\) are a family of Jacobi-like polynomials with \(\deg(e_k) = k + 1\) that are orthogonal with respect to the weight function \(w(x) := \frac{1}{x(1-x)}\) for \(x \in (0, 1)\).
  - \(P^{(\alpha, \beta)}_k\) is the \(k\)-th order \((\alpha, \beta)\)-Jacobi polynomial on \([-1, 1]\) where \(\alpha, \beta > -1\).
  - \(I_k\) denotes the time integral of \(B\) over \([0, 1]\) against the polynomial \(e_k(t)w(t)\),
  \[ I_k := \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)} dt. \]
  - \(H_{s,t}\) denotes the rescaled space-time Lévy area of Brownian motion over \([s, t]\),
  \[ H_{s,t} := \frac{1}{h} \int_s^t W_{s,u} - \frac{u - s}{h} W_{s,t} du. \]
  - \(L_{s,t}\) denotes the space-space-time Lévy area of Brownian motion over \([s, t]\),
  \[ L_{s,t} := \frac{1}{6} \left( \int_s^t \int_s^u \int_s^v dW_r \circ dW_u \circ dW_v - 2 \int_s^t \int_s^u \int_s^v dW_r \circ dW_v \circ dW_u \right. \]
  \[ \left. + \int_s^t \int_s^u \int_s^v dW_r \circ dW_v \circ dW_u \right). \]
  - \(\bar{W}\) denotes the Brownian parabola corresponding to \(W\) over a fixed interval.
  \[ \bar{W}_u = W_s + \frac{u - s}{h} W_{s,t} + \frac{6(u - s)(t - u)}{h^2} H_{s,t}, \quad \forall u \in [s, t]. \]
  - \(Z\) is the Brownian arch corresponding to \(W\) over an interval (\(Z := W - \bar{W}\)).
2. Main result. It was shown in [4] that Brownian motion can be generated using Alpert-Rokhlin multiwavelets (see [9]). The mother functions that generate this wavelet basis are supported on $[0, 1]$ and are defined using polynomials as follows:

**Definition 2.1 (Alpert-Rokhlin wavelets).** For $q \geq 1$, define the $q$ functions $\phi^{q, 1}, \ldots, \phi^{q, q} : [0, 1] \to \mathbb{R}$ as piecewise polynomials of degree $q - 1$ with pieces on $[0, \frac{1}{2}), \left(\frac{1}{2}, 1\right]$ and which satisfies the following conditions for $1 \leq p \leq q$ and $t \in [0, \frac{1}{2})$:

\[
\phi^{q, p}(t) = (-1)^{q+p-1} \phi^{q, p}(1 - t), \tag{2.1}
\]

\[
\int_0^1 \phi^{q, p}(t) \phi^{q, r}(t) \, dt = \delta_{qr}, \quad \text{for} \ 1 \leq r \leq q, \tag{2.2}
\]

\[
\int_0^1 t^k \phi^{q, p}(t) \, dt = 0, \quad \text{for} \ 0 \leq k \leq q - 1. \tag{2.3}
\]

The Alpert-Rokhlin multiwavelets of order $q$ can now be generated by translating and scaling the mother functions $\phi^{q, p}$:

\[
\phi^{q, p}_{nk}(t) := \frac{1}{\sqrt{2^n}} \phi^{q, p}(2^n t - k),
\]

for $n \geq 0$ and $k \in \{0, \ldots, 2^n - 1\}$.

Whilst our results will not be presented in terms of the above wavelets, we shall see that the polynomials of interest are directly related to the conditions (2.1), (2.2) and (2.3). The main result of this paper gives an effective method for approximately Brownian sample paths using a class of Jacobi-like polynomials. The proof is based on the new discovery that these polynomials can be viewed as eigenfunctions of an integral operator defined by the Brownian bridge covariance function. These polynomials, which lie at the heart of this paper, will also help us interpret the geometrical features that certain normally distributed iterated integrals encode about the Brownian path.

**Theorem 2.2 (A polynomial Karhunen-Loève theorem for the Brownian bridge).** Let $B$ denote a Brownian bridge on $[0, 1]$ and consider the Borel measure $\mu$ given by

\[
\mu(a, b) := \int_a^b \frac{1}{x(1-x)} \, dx \quad \text{for all open intervals} \ (a, b) \subset [0, 1].
\]

Then there exists a family of orthogonal polynomials $\{e_k\}_{k \geq 1}$ with $\deg(e_k) = k + 1$ and

\[
\int_0^1 e_i e_j \, d\mu = \delta_{ij},
\]

with $\delta_{ij}$ denoting the Kronecker delta, such that $B$ admits the following representation

\[
B = \sum_{k=1}^{\infty} I_k e_k, \tag{2.4}
\]

where $\{I_k\}$ is the collection of independent centered Gaussian random variables with

\[
I_k := \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)} \, dt, \tag{2.5}
\]

and

\[
\text{Var}(I_k) = \frac{1}{k(k+1)}.
\]
Moreover, \( \{e_k\} \) is an optimal orthonormal basis of \( L^2([0,1], \mu) \) for approximating \( B \) by truncated series expansions with respect to the following weighted \( L^2(\mathcal{P}) \) norm

\[
\|X\|_{L^2(\mathcal{P})} := \sqrt{E \left[ \int_0^1 (X_s)^2 d\mu(s) \right]},
\]

where \( X \) is a square \( \mu \)-integrable process.

\textbf{Proof.} Our argument is that of the Karhunen-Loève theorem in general \( L^2 \) spaces. Note that \( B \) is a square \( \mu \)-integrable process as

\[
E \left[ \int_0^1 (B_s)^2 d\mu(s) \right] = \int_0^1 (B_s)^2 d\mu(s) = \int_0^1 s(1-s) \cdot \frac{1}{1-s} du = 1 < \infty.
\]

Let \( K_B \) denote the covariance function for the standard Brownian bridge on \([0,1]\). One can show by direct calculation that

\[
\|K_B\|_{L^2([0,1]^2), \mu}^2 = \int_0^1 \int_0^1 (\min(s,t) - st)^2 d\mu(s) d\mu(t) = \frac{1}{3} \pi^2 - 3 < \infty.
\]

It now follows that the linear operator \( T_K : L^2([0,1]^2, \mu) \to L^2([0,1]^2, \mu) \) given by

\[
(T_K f)(t) := \int_0^1 K_B(s,t)f(s) \, d\mu(s),
\]

is well-defined and continuous. Furthermore the function \( k_B \) on \([0,1]\) defined as \( k_B(x) := K_B(x,x) \) satisfies

\[
\int_0^1 |k_B(x)| \, d\mu(x) = \int_0^1 x(1-x) \cdot \frac{1}{x(1-x)} \, dx = 1 < \infty.
\]

It is now possible to apply Mercer’s theorem for kernels on general \( L^2 \) spaces (see \cite{Mercer}). It then follows from Mercer’s theorem that there exists an orthonormal set \( \{e_k\}_{k \geq 1} \) of \( L^2([0,1], \mu) \) consisting of eigenfunctions of \( T_K \) such that the corresponding sequence of eigenvalues \( \{\lambda_n\}_{n \geq 1} \) is non-negative. Moreover, the eigenfunctions corresponding to non-zero eigenvalues are continuous on \([0,1]\) and the kernel \( K_B \) has the representation

\[
K_B(s,t) = \sum_{k=1}^{\infty} \lambda_k e_k(s)e_k(t),
\]

(2.6)

where the series convergences absolutely and uniformly on compact subsets of \([0,1]\).

In the next part of the proof, we’ll show that each \( e_k \) is a polynomial of degree \( k+1 \). As each \( e_k \) is an eigenfunction of \( T_K \), we have

\[
\int_0^1 \frac{\min(s,t) - st}{s(1-s)} e_k(s) \, ds = \lambda_k e_k(t).
\]

(2.7)

Since \( e_k \in L^2([0,1], \mu) \), it follows that \( e_k(0) = 0 \) and \( e_k(1) = 0 \) for each \( k \geq 1 \). Therefore by using the Leibniz integral rule to twice differentiate both sides of (2.7), we see that \( e_k \) must satisfy the following differential equation

\[
\lambda_n e''_k(t) = -\frac{1}{t(1-t)} e_k(t).
\]

(2.8)
Since \( e_k \neq 0 \), we have that \( \lambda_k \neq 0 \). Differentiating both sides of the ODE (2.8) gives
\[
t(1-t)\frac{d^2}{dt^2}(e_k'(t)) + (1-2t)\frac{d}{dt}(e_k'(t)) + \frac{1}{\lambda_k}e_k'(t) = 0.
\]
For \( x \in [0,1] \), we define the function
\[
y_k(x) := e_k'\left(\frac{1}{2}(1+x)\right).
\]
Thus \( y_k \) satisfies the following differential equation
\[
(1-x^2)y_k''(x) - 2xy_k'(x) + \frac{1}{\lambda_k}y_k(x) = 0. \tag{2.9}
\]
Remarkably, this is the Legendre differential equation \([11]\). It follows from classical theory that \( \frac{1}{\lambda_k} = k(k+1) \), and \( y_k \) is proportional to the \( k \)-th Legendre polynomial. Therefore \( e_k' \) is a constant multiple of the \( k \)-th shifted Legendre polynomial and hence each \( e_k \) is a polynomial of degree \( k+1 \).

We can now define the following integrals for \( k \geq 1 \),
\[
I_k := \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)} \, dt.
\]
It follows from Fubini's theorem that
\[
E[I_k] = 0,
\]
\[
E[I_i I_j] = E\left[ \int_0^1 \int_0^1 B_s B_t e_i(s) e_j(t) \, d\mu(s) \, d\mu(t) \right]
= \int_0^1 \int_0^1 E[B_s B_t] e_i(s) e_j(t) \, d\mu(s) \, d\mu(t)
= \int_0^1 e_j(t) \left( \int_0^1 K_B(s,t) e_i(s) \, d\mu(s) \right) \, d\mu(t)
= \lambda_i \delta_{ij}.
\]
Since each \( I_k \) is defined by a linear functional on the same Gaussian process \( B \), we see from the above that \( \{I_k\} \) is a collection of uncorrelated (and therefore independent) Gaussian random variables with
\[
E[I_k] = 0, \quad \text{Var}(I_k) = \frac{1}{k(k+1)}.
\]
Finally, the \( L^2(\mathbb{P}) \) convergence we require follows as
\[
E \left[ \left( B_t - \sum_{k=1}^N I_k e_k(t) \right)^2 \right] = k_B(t) + E \left[ \sum_{i,j=1}^N I_i I_j e_i(t) e_j(t) \right] - 2E \left[ B_t \sum_{k=1}^N I_k e_k(t) \right]
= k_B(t) + \sum_{i=1}^N \lambda_i e_i^2(t) - 2E \left[ \int_0^1 B_s B_t e_i(s) e_i(t) \, d\mu(s) \right]
= k_B(t) - \sum_{i=1}^N \lambda_i e_i^2(t),
\]
which converges to 0 by Mercer’s theorem (2.6).
All that remains is to prove optimality for the truncated series expansions of $[24]$. Let $\{f_k\}_{k \geq 1}$ denote an orthonormal basis of $L^2([0, 1], \mu)$ such that

$$B = \sum_{k=1}^{\infty} J_k f_k, \quad \text{where} \quad J_k := \int_0^1 B_t f_k(t) \, d\mu(t), \quad \forall k \geq 1.$$ 

For $n \geq 1$, define the error process $r_n := \sum_{k=n+1}^{\infty} J_k f_k$.

Then the square $L^2(\mathbb{P})$ norm of the error process admits the following expansion,

$$\|r_n(t)\|_{L^2(\mathbb{P})}^2 = \mathbb{E} \left[ \sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} J_i J_j f_i(t) f_j(t) \right]$$

$$= \sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} \mathbb{E} \left[ \int_0^1 \int_0^1 B_s B_t f_i(s) f_j(t) \, d\mu(s) \, d\mu(t) \right] f_i(t) f_j(t)$$

$$= \sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} \left( \int_0^1 \int_0^1 K_B(s, t) f_i(s) f_j(t) \, d\mu(s) \, d\mu(t) \right) f_i(t) f_j(t).$$

Integrating the above with respect to $\mu$ and using the orthogonality of $\{f_k\}$ gives

$$\|r_n\|_{L^2(\mathbb{P})}^2 = \int_0^1 \|r(t)\|_{L^2(\mathbb{P})}^2 \, d\mu(t) = \sum_{k=n+1}^{\infty} \int_0^1 \int_0^1 K_B(s, t) f_k(s) f_k(t) \, d\mu(s) \, d\mu(t).$$

Note that an optimal orthonormal basis of $L^2([0, 1], \mu)$ solves the following problem

$$\min_{f_k} \|r_n\|_{L^2(\mathbb{P})}^2 \quad \text{subject to} \quad \|f_k\|_{L^2([0, 1], \mu)} = 1.$$ 

By introducing Lagrange multipliers $\nu_k$, we wish to find functions $\{f_k\}$ that minimize

$$E_n[\{f_k\}] := \sum_{k=n+1}^{\infty} \int_0^1 \int_0^1 K_B(s, t) f_k(s) f_k(t) \, d\mu(s) \, d\mu(t) - \nu_k \left( \int_0^1 (f_k(s))^2 \, d\mu(s) - 1 \right).$$

We will now consider the following square integrable functions, defined for $s, t \in (0, 1)$:

$$\tilde{f}_k(t) := f_k(t) \cdot \frac{1}{\sqrt{t(1-t)}}, \quad \tilde{K}_B(s, t) := K_B(s, t) \cdot \frac{1}{\sqrt{s(1-s)}} \cdot \frac{1}{\sqrt{t(1-t)}}.$$ 

Therefore it is enough to find a family of functions $\{\tilde{f}_k\}$ in $L^2([0, 1])$ that minimizes

$$\tilde{E}_n[\{\tilde{f}_k\}] := \sum_{k=n+1}^{\infty} \int_0^1 \int_0^1 \tilde{K}_B(s, t) \tilde{f}_k(s) \tilde{f}_k(t) \, ds \, dt - \nu_k \left( \int_0^1 (\tilde{f}_k(s))^2 \, ds - 1 \right).$$

To find a minimizer, we set the functional derivative of $\tilde{E}_n$ with respect to $\tilde{f}_k$ to zero.

$$\frac{\partial \tilde{E}_n}{\partial \tilde{f}_k(t)} = 2 \int_0^1 \tilde{K}_B(s, t) \tilde{f}_k(s) \, ds - 2\nu_k \tilde{f}_k(t) = 0.$$ 

By using the definitions of $\tilde{f}_k$ and $\tilde{K}_B$, it is trivial to show the above is equivalent to

$$\int_0^1 K_B(s, t) f_k(s) \, d\mu(s) = \nu_k f_k(t),$$

which is satisfied if and only if $f_k$ are eigenfunctions of $T_K$. $\square$
Corollary 2.3. The above result can naturally be extended to Brownian motion. If $W$ is a standard Brownian motion on $[0,1]$ and $B$ is the associated bridge process then by Theorem 2.2 we have the following representation of $W$:

$$W = W_1 e_0 + \sum_{k=1}^{\infty} I_k e_k,$$

(2.10)

where $e_0(t) := t$ for $t \in [0,1]$, and the random variables $\{I_k\}$ are independent of $W_1$.

In the rest of this section, we shall study the key objects introduced in Theorem 2.2. Since each orthogonal polynomial lies in $L^2([0,1], \mu)$, it must have roots at 0 and 1. Therefore $e_k \cdot \frac{t}{(1-t)}$ is itself a polynomial but with degree $k-1$, and one can repeatedly apply the integration by parts formula to the stochastic integrals $\{I_k\}$ defined by (2.5). This enables us to express each $I_k$ in terms of iterated integrals of Brownian motion. Moreover, as $e_k \cdot \frac{t}{(1-t)}$ has precisely $k-2$ non-zero derivatives, the highest order iterated integral that is required to fully describe $I_k$ is $\int_{0<s_1<\cdots<s_k<1} B_{s_1} ds_1 \cdots ds_k$.

So by applying the integration by parts formula as above, we can construct a lower triangular $n \times n$ matrix $M_n$ with non-zero diagonal entries that characterizes the relationship between $\{I_k\}_{1 \leq k \leq n}$ and a set of $n$ iterated integrals of Brownian motion.

$$\begin{pmatrix} I_1 \\ \vdots \\ I_n \end{pmatrix} = M_n \begin{pmatrix} \int_{0<s_1<1} B_{s_1} ds_1 \\ \vdots \\ \int_{0<s_1<\cdots<s_n<1} B_{s_1} ds_1 \cdots ds_n \end{pmatrix}.$$  

(2.11)

Since $M_n$ is an invertible matrix, it follows that the column vectors appearing in (2.11) both encode the same information about the Brownian bridge. This will now enable us to establish a fundamental connection between Brownian motion and polynomials.

Theorem 2.4. Consider the following unbiased estimator of a one-dimensional Brownian motion over the interval $[0,1]$,

$$W^n_t := \mathbb{E} \left[ W_1, \int_{0<s_1<1} W_{s_1} ds_1, \cdots, \int_{0<s_1<\cdots<s_{n-1}<1} W_{s_1} ds_1 \cdots ds_{n-1} \right].$$

(2.12)

Then $W^n_t$ is the unique polynomial of degree $n$ with a root at 0 that matches the increment and $n-1$ iterated time integrals of the Brownian path appearing in (2.12).
Proof. It is an immediate consequence of (2.11) that $W^n_t = E[W_t | W_1, I_1, \ldots, I_{n-1}]$. Hence by (2.10) and independence of the random variables $\{W_1, I_1, \ldots\}$, we have that

$$W^n_t = W_1 e_0 + \sum_{k=1}^{n-1} I_k e_k.$$ 

Thus $W^n$ is indeed a polynomial of degree $n$ with a root at 0 and that matches the increment of the Brownian path. Without loss of generality we can now assume $n \geq 2$.

All that remains is to argue $W^n$ matches the $n-1$ iterated integrals given in (2.12). Using the orthgonality of $\{e_k\}$, it follows that for $1 \leq k \leq n-1$:

$$I_k = \int_0^1 (W_t - W_1 e_0) \cdot \frac{e_k(t)}{t(1-t)} dt = \int_0^1 (W^n_t + \sum_{m=n}^{\infty} I_m e_m - W_1 e_0) \cdot \frac{e_k(t)}{t(1-t)} dt = \int_0^1 (W^n_t - W_1 e_0) \cdot \frac{e_k(t)}{t(1-t)} dt + \sum_{m=n}^{\infty} I_m \int_0^1 \frac{e_k(t) e_m(t)}{t(1-t)} dt = \int_0^1 (W^n_t - W_1 e_0) \cdot \frac{e_k(t)}{t(1-t)} dt.$$ 

Thus $W^n$ matches the integrals of Brownian motion against polynomials with degree at most $n-1$. By the same argument used to derive (2.11), the result then follows. 

![Fig. 2.2. Sample paths of Brownian motion with corresponding polynomial approximations.](image)

Although Theorem 2.2 and Corollary 2.3 are interesting results from a theoretical point of view, both lack an explicit construction of the orthogonal polynomials $\{e_k\}$. On the other hand, it was shown that the defining eigenfunction property of each $e_k$ implies that its derivative $e'_k$ is proportional to the $k$-th shifted Legendre polynomial. Hence the family $\{e_k\}$ are the (normalized) shifted $(\alpha, \beta)$-Jacobi polynomials but with $\alpha = \beta = -1$. Since Jacobi polynomials are typically studied with $\alpha, \beta > -1$, it is necessary to check that there are no complications when the parameters approach $-1$.

**Definition 2.5.** For $k \geq 2$, the $k$-th order $(-1, -1)$-Jacobi polynomial, $P_k^{(-1, -1)}$, is defined as

$$P_k^{(-1, -1)} := \lim_{\alpha, \beta \to -1+} P_k^{(\alpha, \beta)}. \quad (2.13)$$ 

Naturally, for the above definition to make sense we require the following lemma.
Lemma 2.6. Let \( P_k^{(\alpha,\beta)} \) denote the \( k \)-th order \((\alpha,\beta)\)-Jacobi polynomial on \([-1,1]\). Then for \( k \geq 2 \), there exists a real-valued polynomial \( P_k \) such that \( \| P_k - P_k^{(\alpha,\beta)} \|_\infty \to 0 \) as \( \alpha,\beta \to -1^+ \).

Proof. Consider the following integral relationship for \((\alpha,\beta)\)-Jacobi polynomials with \( \alpha,\beta > -1 \) (see [12]).

\[
P_k^{(\alpha,\beta)}(x) = \frac{k + \alpha + \beta + 1}{2} \int_{-1}^{x} P_{k-1}^{(\alpha+1,\beta+1)}(u) \, du, \quad \text{for all } k \geq 2. \tag{2.14}
\]

Define a real-valued polynomial \( P_k \) on \([-1,1]\) by

\[
P_k(x) := \frac{k - 1}{2} \int_{-1}^{x} P_{k-1}^{(0,0)}(u) \, du, \quad \text{for all } k \geq 2. \tag{2.15}
\]

Since the terms of \((\alpha,\beta)\)-Jacobi polynomials depend continuously on \((\alpha,\beta)\), we have \( P_n^{(0,0)} = \lim_{\alpha,\beta \to 0} P_n^{(\alpha,\beta)} \) for \( n \geq 1 \). The result then follows from (2.14) and (2.15).

Using the above definition for \((-1,-1)\)-Jacobi polynomials, we can give an explicit formula for the polynomials \( \{e_k\}_{k \geq 1} \) which appear in Theorem 2.2 and Corollary 2.3.

Theorem 2.7. Suppose that each \( e_k \) has a positive leading coefficient. Then

\[
e_k(t) = \frac{1}{k} \sqrt{k(k+1)(2k+1)} P_{k+1}^{(-1,-1)}(2t - 1), \quad \forall t \in [0,1], \forall k \geq 1.
\]

Proof. The following result is stated in [12]:

\[
\int_{-1}^{1} (1-x)^{\alpha} (1+x)^{\beta} (P_n^{(\alpha,\beta)}(x))^2 \, dx = \frac{2^{\alpha+\beta+1} \Gamma(k+\alpha+1) \Gamma(n+\beta+1)}{2n+\alpha+\beta+1 \Gamma(n+\alpha+\beta+1) n!},
\]

for \( n \geq 1 \) and \( \alpha,\beta > -1 \). Applying the change of variables, \( t := \frac{1}{2}(x + 1) \), we have

\[
\int_{0}^{1} t^{\beta}(1-t)^{\alpha} (P_n^{(\alpha,\beta)}(2t - 1))^2 \, dt = \frac{1}{2n+\alpha+\beta+1} \frac{\Gamma(n+\alpha+1) \Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1) n!},
\]

for \( n \geq 1 \) and \( \alpha,\beta > -1 \). Taking the limit \( \alpha,\beta \to -1^+ \) gives

\[
\int_{0}^{1} \frac{1}{t(1-t)} (P_n^{(-1,-1)}(2t - 1))^2 \, dt = \frac{1}{2n-1} \frac{(n-1)! (n-1)!}{n! (n-2)!} = \frac{1}{2n-1} \frac{n-1}{n}, \quad \text{for all } n \geq 2.
\]

Therefore by setting \( k := n - 1 \), we have

\[
\int_{0}^{1} \frac{1}{t(1-t)} \left( \frac{1}{k} \sqrt{k(k+1)(2k+1)} P_{k+1}^{(-1,-1)}(2t - 1) \right)^2 \, dt = 1, \quad \text{for all } k \geq 1.
\]

Note that the construction of \( P_k \) in the proof of Lemma 2.6 implies that each \( e_k(t) \) is proportional to \( P_{k+1}^{(-1,-1)}(2t - 1) \). The result now follows from the above calculations.

To conclude this section, we will address the relationship between the orthogonal Jacobi-like polynomials \( \{e_k\} \) and the Alpert-Rokhlin wavelets given in Definition 2.1. Since each \( e_k \) is proportional to the \( k \)-th shifted Legendre polynomial, the family of polynomials \( \{e_k\} \) is orthogonal with respect to the standard \( L^2([0,1]) \) inner product. This orthogonality is exactly what is needed to satisfy the conditions (2.2) and (2.3). So for any \( q \geq 1 \) there exists an Alpert-Rokhlin mother function of order \( q \) that is a piecewise polynomial where both pieces can be rescaled and translated to give \( e_{q-1} \).
3. Applications to SDEs. Consider the Stratonovich SDE on the interval \([0, T]\)
\[
dy_t = f_0(y_t) \, dt + f_1(y_t) \circ dW_t,
\]
where \(\xi \in \mathbb{R}^p\) and \(f_i\) denote bounded \(C^\infty\) vector fields on \(\mathbb{R}^p\) with bounded derivatives. It then follows from the standard Picard iteration argument that there exists a unique strong solution \(y\) to (3.1). An important tool in the numerical analysis of this solution is the stochastic Taylor expansion (see chapter 5 of [13] for a comprehensive review). For the purposes of this paper, we only require the following specific Taylor expansion.

**Theorem 3.1 (High order Stratonovich-Taylor expansion).** Let \(y\) denote the unique strong solution to (3.1) and let \(0 \leq s \leq t\). Then \(y_t\) can be expanded as follows:
\[
y_t = y_s + f_0(y_s) \, h + f_1(y_s) \, W_{s,t} + \frac{1}{2} f'_1(y_s) f_1(y_s) W_{s,t}^2 + \frac{1}{2} f''_0(y_s) f_0(y_s) h^2 \tag{3.2}
\]
\[
+ f'_0(y_s) f_1(y_s) \int_s^t \int_s^u \circ dW_v \, du + f'_1(y_s) f_0(y_s) \int_s^t \int_s^u \circ dW_v \, du + \frac{1}{6} \left( f'_1(y_s) f'_1(y_s) f_1(y_s) + f''_1(y_s) (f_1(y_s), f_1(y_s)) \right) W_{s,t}^3
\]
\[
+ \int_s^t \int_s^u \circ dW_v \, dv \circ dW_u \tag{3.3}
\]
\[
+ \left( f'_1(y_s) f'_1(y_s) f_0(y_s) f_1(y_s) f_1(y_s) \right) W_{s,t}^3
\]
where \(h := t - s\) and the remainder term has the following uniform estimate for \(h < 1\),
\[
\sup_{y_s \in \mathbb{R}^p} \| R_4(h, y_s) \|_{L^2(\mathbb{P})} \leq C h^2,
\]
where the constant \(C\) depends only on the vector fields of the differential equation.

From a numerical perspective, the most challenging terms presented in (3.2) are those that involve non-trivial third order iterated integrals of Brownian motion and time. Moreover, the most significant source of discretization error that high order numerical methods will experience is generally due to approximating these stochastic integrals. By representing Brownian motion as a (random) polynomial plus independent noise, we shall derive a new optimal and unbiased estimator for these third order integrals.

**Theorem 3.2.** Let \(W\) denote a standard real-valued Brownian motion on \([0, 1]\). Let \(W^n\) be the unique \(n\)-th degree polynomial with a root at 0 and satisfying
\[
\int_0^1 u^k \, dW^n_u = \int_0^1 u^k \, dW_u, \quad \text{for} \ k = 0, 1, \ldots, n - 1.
\]
Then \(W = W^n + Z^n\), where \(Z^n\) is a centred Gaussian process independent of \(W^n\).
Furthermore, $Z^n$ has the following covariance function:

$$
\text{cov}(Z^n_s, Z^n_t) = K_B(s, t) - \sum_{k=1}^{n-1} \lambda_k e_k(s)e_k(t), \quad \text{for } s, t \in [0, 1].
$$

where $K_B$ denotes the standard Brownian bridge covariance function and $\{\lambda_k\}, \{e_k\}$ are the same eigenvalues and eigenfunctions defined by Theorem 2.2.

Proof. It follows from the integration by parts formula that $W^n$ matches the increment and $n-1$ iterated time integrals of Brownian motion that appear in (2.12). Hence $W^n$ is also the polynomial defined in Theorem 2.4 and $W = W^n + Z^n$ where

$$
W^n = W_1 e_0 + \sum_{k=1}^{n-1} I_k e_k,
$$

$$
Z^n = \sum_{k=n}^{\infty} I_k e_k.
$$

Then by Theorem 2.2 $Z^n$ is a centered Gaussian process that is independent of $W^n$. In addition, the covariance function defining $Z^n$ can be directly computed as follows:

$$
\text{cov}(Z^n_s, Z^n_t) = \text{cov}
\left(
\sum_{i=n}^{\infty} I_i e_i(s), \sum_{j=n}^{\infty} I_j e_j(t)
\right)
= \sum_{k=n}^{\infty} \lambda_k e_k(s)e_k(t)
= K_B(s, t) - \sum_{k=1}^{n-1} \lambda_k e_k(s)e_k(t), \quad \text{for } s, t \in [0, 1].
$$

Note that the final line is achieved using the representation of $K_B$ given by (2.6).

The above theorem has a significant conclusion, namely that there exist unbiased polynomial approximants of Brownian motion for which the error process can be uniformly estimated in an $L^2(\mathbb{P})$ sense. In particular, this theorem already gives interesting applications in the case when $n = 2$ and motivates the following definitions:

**Definition 3.3.** The standard Brownian parabola $\widetilde{W}$ is the unique quadratic polynomial on $[0, 1]$ with a root at 0 and satisfying

$$
\widetilde{W}_1 = W_1, \quad \int_0^1 \widetilde{W}_u du = \int_0^1 W_u du.
$$

**Definition 3.4.** The standard Brownian arch $Z$ is the process $Z := W - \widetilde{W}$. By Theorem 3.2 $Z$ is the centred Gaussian process on $[0, 1]$ with covariance function

$$
K_Z(s, t) = \min(s, t) - st - 3st(1-s)(1-t), \quad \text{for } s, t \in [0, 1].
$$

**Definition 3.5.** The rescaled space-time Lévy area of Brownian motion over the interval $[s, t]$ is the average excursion experienced by the associated bridge process,

$$
H_{s,t} := \frac{1}{h} \int_s^t \frac{u-s}{h} W_{s,t} du,
$$

where $h = t - s$. Since $e_1(t) = \sqrt{6}t(1-t)$, one can see that $H_{0,1}$ corresponds to $\frac{\sqrt{6}}{12}I_1$ as defined in Theorem 2.2. Therefore $H_{s,t} \sim \mathcal{N}(0, \frac{1}{12}h)$ and is independent of $W_{s,t}$. 


By applying the natural scaling of Brownian motion, one can define the Brownian parabola and Brownian arch processes over any interval $[s, t]$ with finite size $h = t - s$. Whilst the Brownian arch can be viewed in a similar light to the Brownian bridge, there are clear qualitative and quantitative differences in their covariance functions. In particular, the Brownian arch has less variance at its midpoint compared to most points in $[s, t]$ (by which we mean that $|\{u \in [s, t] : \text{Var}(Z_u) \leq \text{Var}(Z_{\frac{1}{2}(s+t)})\}| < \frac{1}{2} h$). This is in contrast to the Brownian bridge, which has most variance at its midpoint. In fact, the Brownian parabola gives a relatively uniform estimate of the original path.

**Fig. 3.1.** Brownian motion can be expressed as a (random) parabola plus independent noise. The approximating parabola has the same increment and space-time Lévy area as the original path.

**Fig. 3.2.** Variance profile of the standard Brownian arch.

Using these new definitions, we can study the high order integrals appearing in (3.2).

**Theorem 3.6** (Conditional expectation of non-trivial Brownian time integral).

$$
\mathbb{E} \left[ \int_s^t W_{s,u}^2 \, du \, \bigg| \, W_{s,t}, H_{s,t} \right] = \frac{1}{3} h W_{s,t}^2 + h W_{s,t} H_{s,t} + \frac{6}{5} h H_{s,t}^2 + \frac{1}{15} h^2. \tag{3.4}
$$

**Proof.** By the natural Brownian scaling it is enough to prove the result on $[0, 1]$. Note that $W = \tilde{W} + Z$ where the parabola $\tilde{W}$ is completely determined by $(W_1, H_1)$ and $Z$ is independent of $(W_1, H_1)$. This gives a decomposition for the LHS of (3.4).

$$
\mathbb{E} \left[ \int_0^1 W_u^2 \, du \, \bigg| \, W_1, H_1 \right] = \mathbb{E} \left[ \int_0^1 (\tilde{W}_u + Z_u)^2 \, du \, \bigg| \, W_1, H_1 \right] = \mathbb{E} \left[ \int_0^1 \tilde{W}_u^2 \, du + 2 \int_0^1 \tilde{W}_u Z_u \, du + \int_0^1 Z_u^2 \, du \, \bigg| \, W_1, H_1 \right] = \int_0^1 \tilde{W}_u^2 \, du + 2 \int_0^1 \tilde{W}_u \mathbb{E}[Z_u] \, du + \int_0^1 \mathbb{E}[Z_u^2] \, du = \int_0^1 (u W_1 + 6 u(1 - u) H_1)^2 \, du + \int_0^1 u - u^2 - 3 u^2 (1 - u)^2 \, du.
$$

The result now follows by evaluating the above integrals. ∎
The above theorem has practical applications for SDE simulation as $W$ and $H$ are independent Gaussian random variables and can be easily generated or approximated. That said, we must first recall how the various iterated integrals in (3.2) are connected.

**Definition 3.7.** The **space-space-time Lévy area** of Brownian motion over an interval $[s, t]$ is defined as

$$L_{s,t} := \frac{1}{6} \left( \int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_v \, du - 2 \int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_u \right) + \int_s^t \int_s^u \int_s^v \, dr \circ dW_r \circ dW_u \right).$$

**Theorem 3.8.** Let $H_{s,t}$ and $L_{s,t}$ denote the Lévy areas of Brownian motion given by definitions 3.3 and 3.7 respectively. Then the following integral relationships hold,

$$\int_s^t \int_s^u \circ dW_v \, du = \frac{1}{2} hW_{s,t} + hH_{s,t},$$

$$\int_s^t \int_s^u \circ dW_u \, du = \frac{1}{2} hW_{s,t} - hH_{s,t},$$

$$\int_s^t \int_s^u \circ dW_r \circ dW_v \, du = \frac{1}{6} hW_{s,t}^2 + \frac{1}{2} hW_{s,t} H_{s,t} + L_{s,t},$$

$$\int_s^t \int_s^u \circ dW_r \circ dW_u \, du = \frac{1}{6} hW_{s,t}^2 - 2L_{s,t},$$

$$\int_s^t \int_s^u \circ dW_r \circ dW_v \circ dW_u = \frac{1}{6} hW_{s,t}^2 - \frac{1}{2} hW_{s,t} H_{s,t} + L_{s,t}.$$

**Proof.** The result follows from numerous applications of integration by parts. □

We can now present the new unbiased estimator for third order iterated integrals of Brownian motion and time. The proposed estimator is fast to compute and the best $L^2(\mathbb{P})$ approximation of these integrals that is measurable with respect to $(W_{s,t}, H_{s,t})$.

**Theorem 3.9** (Conditional expectation of Brownian space-space-time Lévy area).

$$\mathbb{E} \left[ L_{s,t} \mid W_{s,t}, H_{s,t} \right] = \frac{3}{5} hH_{s,t}^2 + \frac{1}{30} h^2. \quad (3.5)$$

**Proof.** The result follows immediately from Theorem 3.6 and Theorem 3.8 □

Therefore in order to propagate a numerical solution of (3.1) over an interval $[s, t]$, one can generate $(W_{s,t}, H_{s,t})$ exactly and then approximate $L_{s,t}$ using Theorem 3.9. However, there are many numerical methods we can choose for solving a given SDE. For the purposes of this paper, we will consider the following two numerical methods:

**Definition 3.10** (High order log-ODE method). For a fixed number of steps $N$ we can construct a numerical solution $(Y_k)_{0 \leq k \leq N}$ of (3.1) by setting $Y_0 := \xi$ and for each $k \in \{0, \ldots, N-1\}$, defining $Y_{k+1}$ to be the solution at $u = 1$ of the following ODE:

$$\frac{d}{du} \xi = f_0(\xi) h + f_1(\xi) W_{t_k, t_{k+1}} + [f_1, f_0](\xi) \cdot hH_{t_k, t_{k+1}} + [f_1, [f_1, f_0]](\xi) \cdot \mathbb{E} \left[ L_{t_k, t_{k+1}} \mid W_{t_k, t_{k+1}}, H_{t_k, t_{k+1}} \right],$$

$$z_0 = Y_k.$$
In both numerical methods the true solution \( y \)

\[ y_k \]

Moreover, the only terms of degree four that are not \( O(h^2) \)

\[ F \]

Then

\[ \text{Let} \]

Note that \( O(h^2) \) the resulting terms, we are left with a Taylor expansion for

\[ Y \]

Whilst there are different ways of interpolating between the successive approximations

\[ Y \]

\[ h \]

\[ k \]

we can construct a numerical solution of the following ODE:

\[ \frac{dY}{dt} = f_0(z)h + f_1(z) \left( W_{t_k,t_{k+1}} + (6 - 12u) H_{t_k,t_{k+1}} \right), \]

\[ z_0 = Y_k, \]

where \( h := \frac{t}{N} \) and \( t_k := kh \).

In both numerical methods the true solution \( y \) at time \( t_k \) can be approximated by \( Y_k \). Whilst there are different ways of interpolating between the successive approximations \( Y_k \) and \( Y_{k+1} \), for this paper we shall simply interpolate between such points linearly. To analyse the above methods, we shall first note the key differences between them. The first important distinction between the two methods is a purely practical one. Although these methods both involve computing a numerical solution of an ODE, the parabola method does not require one to explicitly resolve the vector field derivatives. The second significant difference can be seen in the Taylor expansions of the methods.

**Theorem 3.12.** Let \( Y_{\log} \) be the one-step approximation defined by the log-ODE method on the interval \([s, t]\) with initial value \( Y_0^{\log} = y_s \). Then for sufficiently small \( h \)

\[ Y_1^{\log} = y_s - \left[ f_1, [f_1, f_0] \right] (y_s) \left( L_{s,t} - E \left[ L_{s,t} | W_{s,t}, H_{s,t} \right] \right) + O(h^{3/2}). \]  

Similarly, let \( Y_{\text{para}} \) denote the one-step approximation given by the parabola-ODE method on the interval \([s, t]\) with the same initial value. Then for sufficiently small \( h \)

\[ Y_1^{\text{para}} = y_s - \left[ f_1, [f_1, f_0] \right] (y_s) \left( L_{s,t} - \frac{3}{5} H_{s,t}^2 \right) + O(h^{3/2}). \]

Note that \( O(h^{3/2}) \) denotes terms which can be estimated in an \( L^2(\mathbb{P}) \) sense as in (3.3).

**Proof.** In order to derive (3.8), we must compute the Taylor expansion of (3.6). Let \( F \) denote the vector field defined in (3.6) that was constructed from \( f_0 \) and \( f_1 \). Then \( F \) is smooth, and it follows from the classical Taylor’s theorem for ODEs that

\[ Y_1^{\log} = y_s + F'(y_s)F(y_s) + \frac{1}{2} F''(y_s)(F'(y_s), F(y_s)) + \frac{1}{6} F'''(y_s)F'(y_s)F(y_s) + \frac{1}{24} F''(y_s)F'(y_s)F(y_s) + \frac{1}{24} F'(y_s)F''(y_s)(F'(y_s), F(y_s)) + \frac{1}{8} F''(y_s)(F'(y_s), F(y_s)) + \frac{1}{24} F''(y_s)(F'(y_s), F(y_s)), \]

\[ + \text{higher order terms}. \]

One can define the degree of each term in the above Taylor expansion by counting the number of times functions from \{\( F, F', F'', \ldots \)\} appear. Therefore, we can see that the higher order terms are precisely the terms with degree strictly greater than four. Since the largest component of \( F \) is \( f_1(\cdot)W_{s,t} \), both \( F \) and its derivatives are \( O(h^{3/2}) \).

Hence the “higher order terms” in the above Taylor expansion are \( O(h^{3/2}) \) as in (3.3). Moreover, the only terms of degree four that are not \( O(h^{3/2}) \) are those involving \( W_{s,t}^4 \). It is now enough to analyse just the terms appearing in the first line of the expansion. By substituting the formula for \( F \) given by (3.6) into the first line and then rearranging the resulting terms, we are left with a Taylor expansion for \( Y_1^{\log} \) that resembles (3.2).

The largest difference are the \( O(h^{3/2}) \) terms that correspond to third order integrals. Hence (3.8) follows from Theorem (3.8) with the definitions of \([f_1, f_0]\) and \([f_1, [f_1, f_0]]\).
Arguing (3.9) is fairly straightforward and does not require extensive computations. Using the substitution $Z_u := z^u_{k(u-s)}$ for $u \in [s, t]$, the ODE (3.7) can be rewritten as

$$dZ_u = f_0(Z_u) \, du + f_1(Z_u) \, d\tilde{W}_u,$$

(3.10)

where $\tilde{W}$ denotes the Brownian parabola defined by $(W_{s,t}, H_{s,t})$ on the interval $[s, t]$. By emulating the derivation of the Stratonovich-Taylor expansion (3.2), it is possible to Taylor expand (3.10) in the same fashion. The only difference is that Stratonovich integrals with respect to $W$ are replaced by Riemann-Stieltjes integrals against $\tilde{W}$. Therefore each term in the expansion of (3.10) can be estimated in $L^2(P)$ by applying the natural Brownian scaling to the corresponding iterated integral of $\tilde{W}$ with time.

The result (3.9) is now a direct consequence of Theorem 3.8 along with the above. \[ \square \]

Theorem 3.12 shows that both methods give a one-step approximation error of $O(h^2)$. This means that the log-ODE and parabola-ODE methods are both locally high order; however there is a significant difference in how these methods propagate local errors. The reason is that the $O(h^2)$ components of the log-ODE local errors give a martingale, whilst the $O(h^2)$ part for each parabola-ODE local error has non-zero expectation. Thus the log-ODE method is globally high order whilst the parabola method is not.

In the numerical example, we shall see that the parabola method has the same order of convergence as the piecewise linear approach but gives significantly smaller errors.

To conclude this section, we will present the orders of convergence for both methods.

**Definition 3.13 (Strong convergence).** A numerical solution $Y$ for (3.1) is said to converge in a strong sense with order $\alpha$ if there exists a constant $C$ such that

$$\|Y_N - y_T\|_{L^2(P)} \leq Ch^\alpha,$$

for all sufficiently small step sizes $h = \frac{T}{N}$.

**Definition 3.14 (Weak convergence).** A numerical solution $Y$ for (3.1) is said to converge in a weak sense with order $\beta$ if for any polynomial $p$ there exists $C_p > 0$ such that

$$|E[p(Y_N)] - E[p(y_T)]| \leq C_p h^\beta,$$

for all sufficiently small step sizes $h = \frac{T}{N}$.

**Theorem 3.15 (Orders of convergence).** For a general SDE (3.1), the log-ODE method converges in a strong sense with order 1.5 and a weak sense with order 2. The parabola-ODE method converges in both strong and weak senses with order 1.

**Proof.** Note that Theorem 3.12 establishes the Taylor expansions of both methods. The strong convergence can then be shown as in the proof of Theorem 11.5.1 in [13]. Moreover, the proof of Theorem 11.5.1 also provides the orders of strong convergence. Similarly weak convergence follows directly from the Taylor expansions (3.8) and (3.9), and the rate of convergence can be shown as in the proof of Theorem 14.5.2 in [13]. \[ \square \]
4. **A numerical example.** We shall demonstrate the ideas presented so far through various discretizations of Inhomogeneous Geometric Brownian Motion (IGBM)

\[ dy_t = a(b - y_t) \, dt + \sigma y_t \, dW_t, \]

(4.1)

where \( a \geq 0 \) and \( b \in \mathbb{R} \) are the mean reversion parameters and \( \sigma \geq 0 \) is the volatility. As the vector fields are smooth, the SDE (4.1) can be expressed in Stratonovich form:

\[ dy_t = \tilde{a} (\tilde{b} - y_t) \, dt + \sigma y_t \circ dW_t, \]

(4.2)

where \( \tilde{a} := a + \frac{1}{2} \sigma^2 \) and \( \tilde{b} := \frac{2ab}{2a + \sigma^2} \) denote the “adjusted” mean reversion parameters.

IGBM is an example of a one-factor short-rate model and has seen recent attention in the mathematical finance literature as an alternative to popular models (see \[7\], \[8\]). IGBM is also one of the simplest SDEs that has no known method of exact simulation. We will investigate the strong and weak convergence rates of the following methods:

1. **Log-ODE method** (see definition 3.10)

   Due to the analytically tractable vector fields of (4.2), this method becomes

   \[ Y_{\log}^{\text{log}}_{k+1} := Y_{\log}^{\text{log}}_{k} e^{-\tilde{a}h + \sigma W_{tk,tk+1}} + abh \left( 1 - \sigma H_{tk,tk+1} + \sigma^2 \left( \frac{3}{5} H_{tk,tk+1}^2 + \frac{1}{30} h \right) \right) e^{-\tilde{a}h + \sigma W_{tk,tk+1}} - 1 \]

   \[ -\tilde{a}h + \sigma W_{tk,tk+1}, \]

   \[ Y_{\log}^{\text{log}}_0 := y_0. \]

2. **Parabola-ODE method** (see definition 3.11)

   Due to the analytically tractable vector fields of (4.2), this method becomes

   \[ Y_{\text{para}}_{k+1} := e^{-\tilde{a}h + \sigma W_{tk,tk+1}} \left( Y_{\text{para}}_{k} + ab \int_{tk}^{tk+1} e^{\tilde{a}(s - tk) - \sigma W_{tk,s}} \, ds \right), \]

   \[ Y_{\text{para}}^0 := y_0. \]

   The integral above will be computed by 3-point Gauss-Legendre quadrature.

3. **Piecewise linear method** (see \[14\] for definition and proof of convergence)

   Due to the analytically tractable vector fields of (4.2), this method becomes

   \[ Y_{\text{lin}}_{k+1} := Y_{\text{lin}}_{k} e^{-\tilde{a}h + \sigma W_{tk,tk+1}} + abh \frac{e^{-\tilde{a}h + \sigma W_{tk,tk+1}} - 1}{-\tilde{a}h + \sigma W_{tk,tk+1}}, \]

   \[ Y_{\text{lin}}^0 := y_0. \]

4. **Milstein method** (see section 6 of \[2\] and section 10.3 of \[13\] for overviews)

   \[ Y_{\text{mil}}_{k+1} := \left( Y_{\text{mil}}_{k} + \tilde{a}(\tilde{b} - Y_{\text{mil}}_{k}) h + \sigma Y_{\text{mil}}_{k} W_{tk,tk+1} + \frac{1}{2} \sigma^2 Y_{\text{mil}}_{k} W_{tk,tk+1}^2 \right)^+, \]

   \[ Y_{\text{mil}}^0 := y_0. \]

As before, we shall be discretizing the SDE on a uniform partition with mesh size \( h \). For IGBM, the log-ODE method has the advantage that it leads to a closed formula since the vector field lie brackets that appear in the ODE (3.6) turn out to be constant. Similarly, the parabola-ODE and piecewise linear methods are both straightforward to implement numerically due to analytical tractability of the SDE governing IGBM. We have included the (non-negative) Milstein method into the numerical experiment as a benchmark to test how the proposed methods compare to a well-known method. That said, the Euler-Maruyama method will not be displayed since it has a similar computational cost as Milstein’s method but has a lower order of strong convergence.
In this numerical example, we shall use the same parameter values as in [7], namely $a = 0.1$, $b = 0.04$, $\sigma = 0.6$ and $y_0 = 0.06$. We will also fix the time horizon at $T = 5$. Below is the definition of the error estimators used to analyse the numerical methods.

**Definition 4.1 (Strong and weak error estimators).** For each $N \geq 1$, let $Y_N$ denote a numerical solution of (4.1) computed at time $T$ using a fixed step size $h = \frac{T}{N}$. We can define the following estimators for quantifying strong and weak convergence:

$$S_N := \sqrt{\mathbb{E} \left[ (Y_N - Y_{T}^{\text{fine}})^2 \right]},$$  \hspace{1cm} (4.3)

$$E_N := \mathbb{E} \left[ (Y_N - b)^+ \right] - \mathbb{E} \left[ (Y_{T}^{\text{fine}} - b)^+ \right],$$  \hspace{1cm} (4.4)

where the above expectations are approximated by standard Monte-Carlo simulation and $Y_{T}^{\text{fine}}$ is the numerical solution of (4.1) obtained at time $T$ using the log-ODE method with a “fine” step size of $\min \left( \frac{T}{1000}, \frac{T}{10} \right)$. The fine step size is chosen so that the $L^2(\mathbb{P})$ error between $Y_{T}^{\text{fine}}$ and the true solution $y$ is negligible compared to $S_N$. Note that $Y_N$ and $Y_{T}^{\text{fine}}$ are both computed with respect to the same Brownian paths.

We will now present our results for the numerical experiment that is described above.

**Table 4.1**

| Method         | Computation time (s) |
|----------------|----------------------|
| Log-ODE        | 2.44                 |
| Parabola-ODE   | 2.95                 |
| Piecewise Linear | 1.48               |
| Milstein       | 1.18                 |

From the above graph we see that the log-ODE method gives the best performance. In addition, whilst the other numerical methods share the same order of convergence it is evident there are magnitudes of difference between their respective accuracies. For example, the parabola method is seven times more accurate than piecewise linear.
The above graph demonstrates that the log-ODE method is especially well-suited for weak approximation as it achieves a second order convergence rate in this example. Surprisingly, the other three methods all perform with a comparable level of accuracy. We expect the log-ODE and parabola methods to have about twice the computational cost as the other methods because each step requires generating two random variables. Table 4.1 confirms this and thus sampling is the main bottleneck for the methods. So overall, the numerical evidence supports our claim that the high order log-ODE method is currently a state-of-the-art method for the pathwise discretization of IGBM.

5. Conclusion. There are primarily three significant results given in this paper:

- **An efficient strong polynomial approximation of Brownian motion**
  The main result allows one to construct a “smoother” Brownian motion as a finite sum of (-1, -1)-Jacobi polynomials with independent Gaussian weights. Moreover, it was shown that the approximation is optimal in a weighted $L^2(\mathbb{P})$ sense and the surrounding noise is an independent centered Gaussian process.

- **Unbiased approximation of third order Brownian iterated integrals**
  Iterated integrals of Brownian motion and time are important objects in the study of SDEs as they appear naturally within stochastic Taylor expansions. We have derived the $L^2(\mathbb{P})$-optimal estimator for a class of such integrals that is measurable with respect to the path’s increment and space-time Lévy area.

- **Simulation of Inhomogeneous Geometric Brownian Motion (IGBM)**
  IGBM is a mean-reverting short rate model used in mathematical finance and also one of the simplest SDEs that has no known method of exact simulation. By incorporating the new iterated integral estimator into the log-ODE method we have developed a high order state-of-the-art numerical method for IGBM. Moreover, the results of this paper immediately produce the following open problems:

  - Is it possible to generalize the main theorem to fractional Brownian motion?
  - What are the most efficient Runge-Kutta methods for general one-dimensional SDEs that correctly use the new estimator for third order iterated integrals?
REFERENCES

[1] R. M. Mazo, *Brownian Motion: Fluctuations, Dynamics and Applications*, Clarendon Press, Oxford, 2002.
[2] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, Volume 43, 2001.
[3] J. M. C. Clark and R. J. Cameron, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, Stochastic Differential Systems Filtering and Control, Volume 25 in Lecture Notes in Control and Information Sciences, Springer, 1988.
[4] D. S. Grebenkov, D. Belyaev and P. W. Jones, *A multiscale guide to Brownian motion*, Journal of Physics A, Volume 49, 2015.
[5] F. Castell and J. G. Gaines, *The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations*, Annales de l’Institut Henri Poincaré, Volume 32, 1996.
[6] S. J. A. Malham and A. Wiese, *Stochastic Lie group integrators*, SIAM Journal on Scientific Computing, Volume 30, 2008.
[7] L. Capriotti, Y. Jiang and G. Shaimerdenova, *Approximation Methods for Inhomogeneous Geometric Brownian Motion*, International Journal of Theoretical and Applied Finance, Volume 22, 2019.
[8] B. Zhao, C. Yan and S. Hodges, *Three One-Factor Processes for Option Pricing with a Mean-Reverting Underlying: The Case of VIX*, Financial Review, Volume 54, 2019.
[9] G. Beylkin, R. Coifman, and V. Rokhlin, *Wavelets in Numerical Analysis*, Wavelets and Their Applications, Jones and Bartlett, 1992.
[10] C. Carmeli, A. De Vito, and E. Toigo, *Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem*, Analysis and Applications, Volume 4, 2006.
[11] B. G. S. Doman, *The Classical Orthogonal Polynomials*, World Scientific, 2016.
[12] E. W. Weisstein, *Jacobi Polynomial*, From MathWorld - A Wolfram Web Resource, [http://mathworld.wolfram.com/JacobiPolynomial.html](http://mathworld.wolfram.com/JacobiPolynomial.html)
[13] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
[14] E. Wong, M. Zakai, *On the Convergence of Ordinary Integrals to Stochastic Integrals*, Annals of Mathematical Statistics, Volume 36, 1965.