

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. The coefficients that are obtained from the expansion of Brownian motion in this polynomial basis are independent Gaussian random variables. Therefore, it is practical (i.e., 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 $L^2(\mathbb{P})$ 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, 60J65, 60L90, 65C30

DOI. 10.1137/19M1261912

1. Introduction. Brownian motion is a central object for modeling 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 $L^2(\mathbb{P})$ 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 stochastic differential equation (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. We shall investigate the practical applications of these polynomials and their geometrical features in the numerical analysis of SDEs. Figure 1.1 illustrates these polynomial approximations on two sample paths of Brownian motion.

The paper is organized as follows. In section 2, we shall state and prove the main result of the paper (Theorem 2.2). This will be a Karhunen–Loève theorem for the Brownian bridge, where the orthogonal functions used in the approximation are polynomials. Furthermore, we shall explicitly show that each basis func-

*Received by the editors May 15, 2019; accepted for publication (in revised form) February 14, 2020; published electronically May 4, 2020.

<https://doi.org/10.1137/19M1261912>

Funding: The work of the first author was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/N509711/1. The work of the second and third authors was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant DATASIG, the Alan Turing Institute, and the Oxford-Man Institute.

[†]Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK (james.foster@maths.ox.ac.uk, terry.lyons@maths.ox.ac.uk, harald.oberhauser@maths.ox.ac.uk).

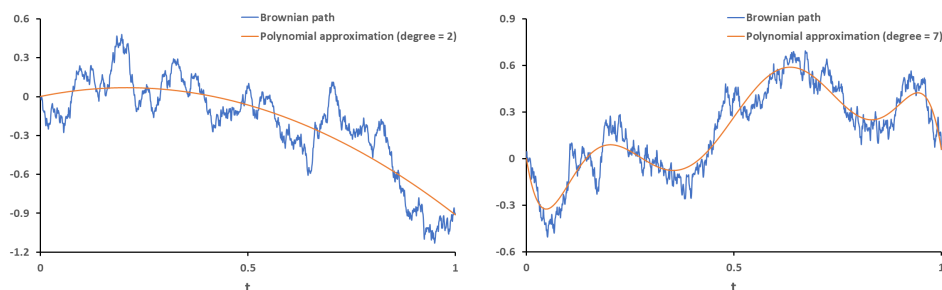


FIG. 1.1. Sample paths of Brownian motion with corresponding polynomial approximations.

tion is proportional to a shifted (α, β) -Jacobi polynomial but with the nonstandard exponents $\alpha = \beta = -1$. This enables us to construct these orthogonal polynomials using recurrence relations, or as the difference of two shifted Legendre polynomials whose degrees differ by two.

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 random polynomial with a root at 0 and satisfying*

$$(1.1) \quad \int_0^1 u^k dW_u^n = \int_0^1 u^k dW_u \quad \text{for } k = 0, 1, \dots, n-1.$$

Then $W = W^n + Z^n$, where Z^n is a centered 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 nontrivial case ($n = 2$) already has interesting applications within the numerical analysis of SDEs. One reason is that parabolas can capture the “space-time area” of Brownian motion (as illustrated by Figure 1.2).

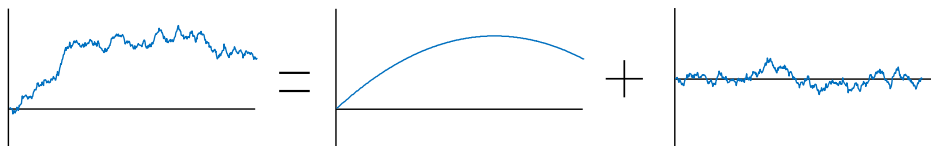


FIG. 1.2. Brownian motion can be expressed as a (random) parabola plus independent noise. Moreover, the approximating parabola has the same increment and time integral as the original path.

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 \widehat{W} be the (unique) quadratic polynomial with a root at 0 and*

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

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

$$(1.2) \quad \mathbb{E} \left[\int_0^1 W_u^2 du \mid W_1, \int_0^1 W_u du \right] = \int_0^1 \widehat{W}_u^2 du + \frac{1}{15}.$$

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]). We will show that by estimating this nontrivial 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 SDE that is driven by a one-dimensional Brownian motion and governed by sufficiently regular vector fields (smooth with bounded derivatives), the 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 $O(h^2)$.

These high orders of convergence can also be achieved in the multidimensional setting provided the vector fields governing the SDE satisfy certain commutativity conditions. For example, the estimator (1.2) has applications for simulating SDEs with additive noise:

$$dy_t = f(y_t) dt + \sigma dW_t,$$

where f is a smooth vector field on \mathbb{R}^d , $\sigma > 0$ is constant, and W is now d -dimensional. By considering Theorem 1.2, we expect that y_t is well approximated (for small t) by

$$\widehat{y}_t + \frac{1}{30} t^2 \sigma^2 \Delta f(y_0),$$

where \widehat{y} is the solution of the following ODE driven by a “Brownian parabola” \widehat{W} :

$$\begin{aligned} d\widehat{y}_t &= f(\widehat{y}_t) dt + \sigma d\widehat{W}_t, \\ \widehat{y}_0 &= y_0. \end{aligned}$$

This parabola-driven ODE can then be discretized using a three-stage Runge–Kutta method, and the resulting SDE approximation shall be investigated in future work.

Since these methods are based on the conditional expectation given by Theorem 1.2, they are designed to minimize the leading error term within local Taylor expansions. This sense of optimality is conceptually similar to that of the asymptotically efficient SDE approximations developed by Clark [7], Newton [8, 9], and Castell and Gaines [5]. The key difference is that we are employing additional integral information about W . Hence this line of research could provide further insight into the approximation of Itô integrals using linear path information, where there already are a number of results concerning the computational complexity of methods (see [10, 11, 12]).

Most notably, Tang and Xiao [10] consider the same triple iterated integral as in (1.2) and present an asymptotically optimal approximation that performs well when a limited number of random variables are used (see Table 2 in [10] for these numerical results). While there are other senses of optimality (such as those discussed in [13, 14]) that could be used when analyzing the proposed approximations of Brownian motion and SDE solutions, we shall estimate errors in an $L^2(\mathbb{P})$ sense throughout the paper. In particular, we can apply the main result to quantify the error of the new estimator.

THEOREM 1.3. *Using the same notation as before, we have the following variance:*

$$\text{Var} \left(\int_0^1 W_u^2 du \mid W_1, \int_0^1 W_u du \right) = \frac{11}{6300} + \frac{1}{180} W_1^2 + \frac{1}{175} \left(\int_0^1 W_u du - \frac{1}{2} W_1 \right)^2.$$

In section 4, we demonstrate the applicability of these ideas to SDE simulation 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 nonnegative. It is therefore suitable for modeling interest rates, stochastic volatilities, and default intensities [15]. From a mathematical viewpoint, IGBM is one of the simplest SDEs that has no known method of exact simulation [16]. 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 $O(h^{\frac{3}{2}})$ and a weak convergence rate of $O(h^2)$.

1.1. Notation. Below we list some of the notation used throughout the paper.

Symbol	Meaning
W	a standard real-valued Brownian motion.
B	a standard real-valued Brownian bridge on $[0, 1]$.
μ	a Borel measure on $[0, 1]$ defined by a singular weight function,
	$\mu(a, b) = \int_a^b \frac{1}{x(1-x)} dx,$
	for all open intervals $(a, b) \subset [0, 1]$.
$\{e_k\}_{k \geq 1}$	a family of Jacobi-like polynomials with $\deg(e_k) = k + 1$ that are orthogonal with respect to weight function $w(x) := \frac{1}{x(1-x)}$.
I_k	a time integral of B times the polynomial $e_k(t) w(t)$ over $[0, 1]$,
	$I_k = \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)} dt.$
K_B	the covariance function of B , that is, $K_B(s, t) = \min(s, t) - st$.
$P_k^{(\alpha, \beta)}$	the k th order (α, β) -Jacobi polynomial on $[-1, 1]$ ($\alpha, \beta > -1$).
Q_k	the k th order Legendre polynomial on $[-1, 1]$, i.e., $Q_k = P_k^{(0, 0)}$.
y	a solution of the Stratonovich SDE on the finite interval $[0, T]$,
	$dy_t = f_0(y_t) dt + f_1(y_t) \circ dW_t,$
	$y_0 = \xi,$
	where $y, \xi \in \mathbb{R}^e$, and $f_i : \mathbb{R}^e \rightarrow \mathbb{R}^e$ denote smooth vector fields. (Itô SDEs will be defined on fixed intervals with the same form.)
$[s, t]$	a general closed subinterval of $[0, T]$, usually considered small.
h	the step size that a numerical method uses, typically $h = t - s$.
$W_{s,t}$	the increment of Brownian motion over $[s, t]$, $W_{s,t} := W_t - W_s$.
\widehat{W}	the Brownian parabola corresponding to W over some interval,
	$\widehat{W}_u = W_s + \frac{u-s}{h} W_{s,t} + \frac{6(u-s)(t-u)}{h^2} H_{s,t} \quad \forall u \in [s, t].$

- Z the Brownian arch corresponding to W defined as $Z := W - \widehat{W}$.
- $H_{s,t}$ the rescaled space-time Lévy area of Brownian motion on $[s, t]$,

$$H_{s,t} = \frac{1}{h} \int_s^t W_{s,u} - \frac{u-s}{h} W_{s,t} du.$$
- $L_{s,t}$ the space-space-time Lévy area of Brownian motion over $[s, t]$,

$$L_{s,t} = \frac{1}{6} \left(\int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_v du - 2 \int_s^t \int_s^u \int_s^v \circ dW_r dv \circ dW_u + \int_s^t \int_s^u \int_s^v dr \circ dW_v \circ dW_u \right),$$
- Y an approximation for the true solution y of a Stratonovich SDE.
- $[\cdot, \cdot]$ the standard Lie bracket of vector fields, $[f_0, f_1] = f'_1 f_0 - f'_0 f_1$.

2. Main result. It was shown in [4] that Brownian motion can be generated using Alpert–Rokhlin multiwavelets (see [17]). 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 $\phi^{q,1}, \dots, \phi^{q,q} : [0, 1] \rightarrow \mathbb{R}$ as piecewise polynomials of degree $q-1$ with pieces on $[0, \frac{1}{2}]$, $[\frac{1}{2}, 1]$ that satisfy the following conditions for all $p \in \{1, \dots, q\}$ and $t \in [0, \frac{1}{2}]$:*

$$(2.1) \quad \phi^{q,p}(t) = (-1)^{q+p-1} \phi^{q,p}(1-t),$$

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

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

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

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

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

While our results will not be presented in terms of the above wavelets, we shall see that the polynomials of interest are directly related to conditions (2.1), (2.2), and (2.3). The main result of this paper gives an effective method for approximating sample paths of Brownian motion by a class of Jacobi-like polynomials. The proof is based on the interpretation of these polynomials as eigenfunctions of an integral operator defined by the Brownian bridge covariance function.¹ These orthogonal 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 μ given by*

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

¹The Brownian bridge is the centered Gaussian process with covariance $K_B(s, t) = \min(s, t) - st$.

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 δ_{ij} denoting the Kronecker delta, such that B admits the representation

$$(2.4) \quad B = \sum_{k=1}^{\infty} I_k e_k,$$

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

$$(2.5) \quad I_k := \int_0^1 B_t \cdot \frac{e_k(t)}{t(1-t)} dt$$

and variance

$$\text{Var}(I_k) = \frac{1}{k(k+1)}.$$

Furthermore, $\{e_k\}$ is an optimal orthonormal basis of $L^2([0, 1], \mu)$ for approximating B by truncated series expansions with respect to the following weighted $L^2(\mathbb{P})$ norm:

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

where X is a square μ -integrable process.

Proof. Our argument is that of the Karhunen–Loève theorem in general L^2 spaces. Note that B is a square μ -integrable process as

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

Let K_B denote the covariance function for the standard Brownian bridge on $[0, 1]$. Since $K_B(s, t) = \min(s, t) - st$, it can be shown by direct calculation that K_B satisfies

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

Hence, it follows that the integral operator $T_K : L^2([0, 1], \mu) \rightarrow L^2([0, 1], \mu)$, given by

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

is well defined and continuous. In addition, the variance function $k_B(x) := K_B(x, x)$ for $x \in [0, 1]$ is μ -integrable as

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

Therefore, we can apply Mercer's theorem for kernels on general L^2 spaces (see [18]). 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 $\{\lambda_k\}_{k \geq 1}$ is nonnegative. Moreover, the eigenfunctions corresponding to nonzero eigenvalues are continuous on $[0, 1]$, and the kernel K_B has the representation

$$(2.6) \quad K_B(s, t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t),$$

where the series (2.6) converges absolutely and uniformly on compact subsets of $[0, 1]$.

In the next part of the proof, we will see that each e_k is a polynomial of degree $k + 1$. As each e_k is an eigenfunction of T_K , we have

$$(2.7) \quad \int_0^1 \frac{\min(s, t) - st}{s(1-s)} e_k(s) ds = \lambda_k e_k(t).$$

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) and then multiplying by $t(1-t)$, we observe that e_k satisfies the differential equation

$$(2.8) \quad t(1-t)\lambda_k e_k''(t) + e_k(t) = 0.$$

Since $e_k \neq 0$, we have that $\lambda_k \neq 0$. Differentiating the left-hand side of the ODE (2.8) produces

$$t(1-t) \frac{d^2}{dt^2}(e_k') + (1-2t) \frac{d}{dt}(e_k') + \frac{1}{\lambda_k} e_k' = 0.$$

For $x \in [-1, 1]$, we define the function

$$y_k(x) := e_k' \left(\frac{1}{2}(1+x) \right).$$

Thus, y_k satisfies the differential equation

$$(2.9) \quad (1-x^2)y_k''(x) - 2xy_k'(x) + \frac{1}{\lambda_k} y_k(x) = 0.$$

Remarkably, this is the Legendre differential equation [19]. It then follows using classical Sturm–Liouville theory that y_k is proportional to the k th Legendre polynomial and $\frac{1}{\lambda_k} = k(k+1)$. Therefore, the derivative e_k' will be 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

$$\begin{aligned} \mathbb{E}[I_k] &= 0, \\ \mathbb{E}[I_i I_j] &= \mathbb{E} \left[\int_0^1 \int_0^1 B_s B_t e_i(s) e_j(t) d\mu(s) d\mu(t) \right] \\ &= \int_0^1 \int_0^1 \mathbb{E}[B_s B_t] e_i(s) e_j(t) d\mu(s) d\mu(t) \\ &= \int_0^1 e_j(t) \left(\int_0^1 K_B(s, t) e_i(s) d\mu(s) \right) d\mu(t) \\ &= \lambda_i \delta_{ij}. \end{aligned}$$

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

$$\begin{aligned}\mathbb{E}[I_k] &= 0, \\ \text{Var}(I_k) &= \frac{1}{k(k+1)}.\end{aligned}$$

Finally, the $L^2(\mathbb{P})$ convergence we require follows as

$$\begin{aligned}\mathbb{E} \left[\left(B_t - \sum_{k=1}^N I_k e_k(t) \right)^2 \right] &= k_B(t) + \mathbb{E} \left[\sum_{i,j=1}^N I_i I_j e_i(t) e_j(t) \right] - 2 \mathbb{E} \left[B_t \sum_{k=1}^N I_k e_k(t) \right] \\ &= k_B(t) + \sum_{i=1}^N \lambda_i e_i^2(t) - 2 \mathbb{E} \left[\sum_{i=1}^N \int_0^1 B_s B_t e_i(s) e_i(t) d\mu(s) \right] \\ &= k_B(t) - \sum_{i=1}^N \lambda_i e_i^2(t),\end{aligned}$$

which converges to 0 by Mercer's theorem (2.6).

All that remains is to prove optimality for the truncated series expansions of (2.4). Let $\{f_k\}_{k \geq 1}$ denote an orthonormal basis of $L^2([0, 1], \mu)$ such that

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

For $n \geq 1$, we consider the following error process associated with the above:

$$r_n := \sum_{k=n+1}^{\infty} J_k f_k.$$

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

$$\begin{aligned}\|r_n(t)\|_{L^2(\mathbb{P})}^2 &= \mathbb{E} \left[\sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} J_i J_j f_i(t) f_j(t) \right] \\ &= \sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} \mathbb{E} \left[\int_0^1 \int_0^1 B_s B_t f_i(s) f_j(t) d\mu(s) d\mu(t) \right] f_i(t) f_j(t) \\ &= \sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} \left(\int_0^1 \int_0^1 K_B(s, t) f_i(s) f_j(t) d\mu(s) d\mu(t) \right) f_i(t) f_j(t).\end{aligned}$$

Integrating the above with respect to μ and using the orthogonality of $\{f_k\}_{k \geq 1}$ gives

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

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

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

By introducing Lagrange multipliers ν_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 $L^2([0, 1])$ which 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 that 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 . □

This result can naturally be extended to express Brownian motion using polynomials.

THEOREM 2.3. *If W is a standard Brownian motion and B is the associated bridge process on $[0, 1]$, then by Theorem 2.2, we have the following representation of W :*

$$(2.10) \quad W = W_1 e_0 + \sum_{k=1}^\infty I_k e_k,$$

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

We shall illustrate the polynomial representation given by the above theorem in Figure 2.1.

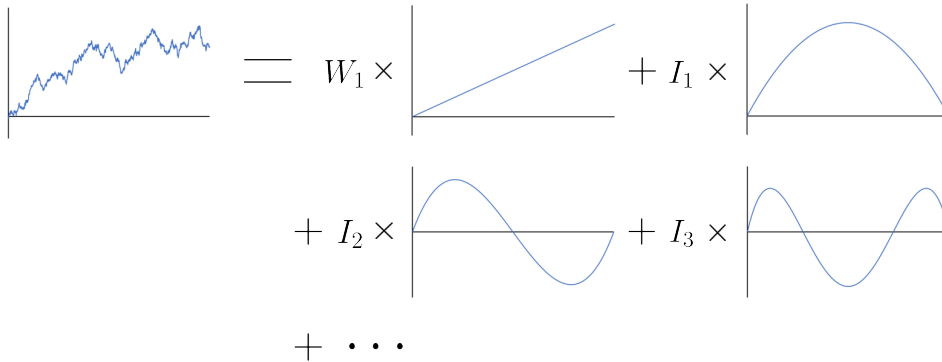


FIG. 2.1. Brownian motion can be expressed as a sum of polynomials with independent weights. Moreover, these polynomials are orthogonal and capture different time integrals of the original path.

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, $e_k \cdot \frac{1}{t(1-t)}$ 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 $e_k \cdot \frac{1}{t(1-t)}$ has precisely $k-2$ nonzero derivatives, the highest order iterated integral that is required to fully describe I_k is $\int_{0 < s_1 < \dots < s_k < 1} B_{s_1} ds_1 \dots ds_k$.

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

$$(2.11) \quad \begin{pmatrix} I_1 \\ \vdots \\ I_n \end{pmatrix} = M_n \begin{pmatrix} \int_{0 < s_1 < 1} B_{s_1} ds_1 \\ \vdots \\ \int_{0 < s_1 < \dots < s_n < 1} B_{s_1} ds_1 \dots ds_n \end{pmatrix}.$$

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

THEOREM 2.4. *Consider the conditional expectation of Brownian motion,*

$$(2.12) \quad W_t^n := \mathbb{E} \left[W_t \mid W_1, \int_{0 < s_1 < 1} W_{s_1} ds_1, \dots, \int_{0 < s_1 < \dots < s_{n-1} < 1} W_{s_1} ds_1 \dots ds_{n-1} \right].$$

where $t \in [0, 1]$. Then W_t^n is the unique polynomial of degree n with a root at 0 that matches the increment W_1 and $n-1$ iterated time integrals of the path W given by

$$(2.13) \quad \int_{0 < s_1 < 1} W_{s_1} ds_1, \dots, \int_{0 < s_1 < \dots < s_{n-1} < 1} W_{s_1} ds_1 \dots ds_{n-1}.$$

Proof. It is a direct consequence of (2.11) that $W_t^n = \mathbb{E}[W_t \mid W_1, I_1, \dots, I_{n-1}]$. Hence by (2.10) and the independence of the random variables $\{W_1, I_1, \dots\}$, we have

$$(2.14) \quad W^n = W_1 e_0 + \sum_{k=1}^{n-1} I_k e_k.$$

Thus, W^n is indeed a polynomial of degree n with a root at 0 and matches the increment of the Brownian path. Without loss of generality we can now assume $n \geq 2$. All that remains is to argue that W^n matches the $n - 1$ iterated integrals given in (2.13). Using the orthogonality of $\{e_k\}$, it follows directly from (2.14) that for $1 \leq k \leq n - 1$,

$$\begin{aligned} I_k &= \int_0^1 (W_t - W_1 e_0(t)) \cdot \frac{e_k(t)}{t(1-t)} dt \\ &= \int_0^1 \left(W_t^n + \sum_{m=n}^{\infty} I_m e_m(t) - W_1 e_0(t) \right) \cdot \frac{e_k(t)}{t(1-t)} dt \\ &= \int_0^1 (W_t^n - W_1 e_0(t)) \cdot \frac{e_k(t)}{t(1-t)} dt + \underbrace{\sum_{m=n}^{\infty} I_m \int_0^1 \frac{e_m(t) e_k(t)}{t(1-t)} dt}_{=0} \end{aligned}$$

Hence W^n matches the integrals of Brownian motion against polynomials with degree at most $n - 1$. By the same argument used in the derivation of (2.11), it follows that W^n matches the various iterated time integrals given in the statement of the theorem. The uniqueness of W^n is now a consequence of having $n + 1$ different constraints. \square

2.1. Properties of orthogonal polynomials. Although Theorems 2.2 and 2.3 are interesting results from a theoretical point of view, both lack an explicit construction of the polynomials $\{e_k\}$ that could be implemented in practice. 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 $\{e_k\}$ is the (normalized) shifted (α, β) -Jacobi polynomials but with $\alpha = \beta = -1$. Since Jacobi polynomials are typically studied with $\alpha, \beta > -1$, it is necessary to show that there exists a well-defined limit when the parameters approach -1 .

DEFINITION 2.5. For $k \geq 2$, the k th degree $(-1, -1)$ -Jacobi polynomial $P_k^{(-1, -1)}$ is

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

Naturally, for this definition to be unambiguous, we will require the following lemma.

LEMMA 2.6. Let $P_k^{(\alpha, \beta)}$ denote the k th degree (α, β) -Jacobi polynomial on $[-1, 1]$. Then for $k \geq 2$, there exists a real-valued polynomial P_k such that $\|P_k - P_k^{(\alpha, \beta)}\|_{\infty} \rightarrow 0$ as $\alpha, \beta \rightarrow -1^+$.

Proof. The following is an identity for Jacobi polynomials, with $\alpha, \beta > -1$, given in [20]:

$$(2.15) \quad P_k^{(\alpha, \beta)}(x) = \frac{k + \alpha + \beta + 1}{2} \int_{-1}^x P_{k-1}^{(\alpha+1, \beta+1)}(u) du \quad \text{for all } k \geq 2.$$

Therefore, we shall define the k th degree polynomial P_k over the interval $[-1, 1]$ by

$$(2.16) \quad P_k(x) := \frac{k - 1}{2} \int_{-1}^x P_{k-1}^{(0, 0)}(u) du \quad \text{for all } k \geq 2.$$

It is straightforward to verify that $\lim_{\alpha, \beta \rightarrow 0} \|P_n^{(\alpha, \beta)} - P_n^{(0, 0)}\|_\infty = 0$ for $n \in \{0, 1\}$. So by induction and the recurrence relation for Jacobi polynomials (see [20]), we have

$$\|P_n^{(\alpha, \beta)} - P_n^{(0, 0)}\|_\infty \rightarrow 0 \quad \text{as } \alpha, \beta \rightarrow 0$$

for all $n \geq 0$. Hence by the dominated convergence theorem with (2.15) and (2.16), it follows that $P_k^{(\alpha, \beta)}$ converges pointwise to P_k as $\alpha, \beta \rightarrow -1$ for each $k \geq 1$. Finally, the result follows as $P_k^{(\alpha, \beta)}$ and P_k are always polynomials with degree k . \square

Using the above definition for $(-1, -1)$ -Jacobi polynomials, we can give an explicit formula for the orthonormal polynomials $\{e_k\}_{k \geq 1}$ appearing in Theorems 2.2 and 2.3.

THEOREM 2.7. *Suppose each e_k has a positive leading coefficient. Then for $k \geq 1$,*

$$(2.17) \quad e_k(t) = \frac{1}{k} \sqrt{k(k+1)(2k+1)} P_{k+1}^{(-1, -1)}(2t-1) \quad \text{for all } t \in [0, 1].$$

Proof. The following identity for (α, β) -Jacobi polynomials is stated in [20]:

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

for $n \geq 1$ and $\alpha, \beta > -1$. 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 $\alpha, \beta > -1$. By Definition 2.5, taking the limit $\alpha, \beta \rightarrow -1^+$ will yield

$$\begin{aligned} \int_0^1 \frac{1}{t(1-t)} (P_n^{(-1, -1)}(2t-1))^2 dt &= \frac{1}{2n-1} \frac{(n-1)!(n-1)!}{n!(n-2)!} \\ &= \frac{1}{2n-1} \frac{n-1}{n} \quad \text{for all } n \geq 2. \end{aligned}$$

Therefore, by setting $k := n-1$ we have

$$\int_0^1 \frac{1}{t(1-t)} \left(\frac{1}{k} \sqrt{k(k+1)(2k+1)} P_n^{(-1, -1)}(2t-1) \right)^2 dt = 1 \quad \text{for all } k \geq 1.$$

Recall that e'_k is proportional to the k th shifted Legendre polynomial $P_k^{(0, 0)}(2t-1)$. Similarly, we saw in the proof of Lemma 2.6 that the derivative of $P_{k+1}^{(-1, -1)}$ is $\frac{k}{2} P_k^{(0, 0)}$. As e_k and $P_{k+1}^{(-1, -1)}$ are zero at their respective endpoints, we have that each e_k must be proportional to $P_{k+1}^{(-1, -1)}(2t-1)$. The result now follows from the above calculations. \square

Having identified an explicit formula for the eigenfunctions $\{e_k\}$ in (2.17), we shall now describe two methodologies for computing the Jacobi-like polynomials $\{P_k^{(-1, -1)}\}$.

Given these constructions, it is then straightforward to produce illustrations such as Figures 1.1 and 2.2. The first approach is to use the three-term recurrence relation in the theorem below.

THEOREM 2.8 (recurrence relation for $(-1, -1)$ -Jacobi polynomials). *For $n \geq 2$,*

$$(2.18) \quad n(n+2)P_{n+2}^{(-1, -1)}(x) = (n+1)(2n+1)xP_{n+1}^{(-1, -1)}(x) - n(n+1)P_n^{(-1, -1)}(x),$$

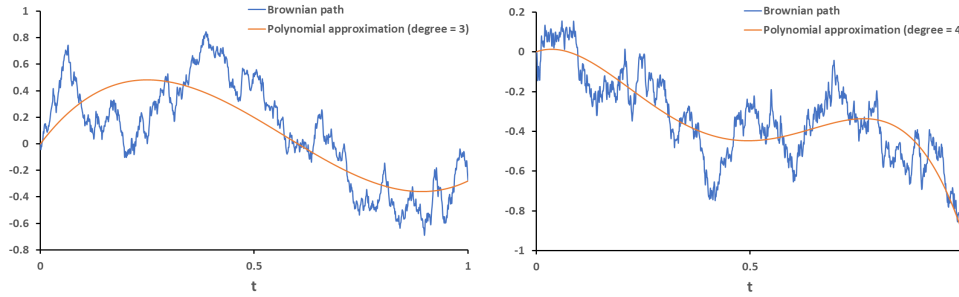


FIG. 2.2. Sample paths of Brownian motion with corresponding polynomial approximations.

where the initial polynomials are given by

$$P_2^{(-1,-1)}(x) = \frac{1}{4}(x - 1)(x + 1),$$

$$P_3^{(-1,-1)}(x) = \frac{1}{2}x(x - 1)(x + 1).$$

Proof. The following recurrence relation for Jacobi polynomials is presented in [20]:

$$\begin{aligned} &2(k + 1)(k + \alpha + \beta + 1)(2k + \alpha + \beta)P_{k+1}^{(\alpha,\beta)}(x) \\ &= (2k + \alpha + \beta + 1)((2k + \alpha + \beta)(2k + \alpha + \beta + 2)x + \alpha^2 - \beta^2) P_k^{(\alpha,\beta)}(x) \\ &\quad - 2(k + \alpha)(k + \beta)(2k + \alpha + \beta + 2) P_{k-1}^{(\alpha,\beta)}(x) \end{aligned}$$

for $k \geq 1$ and $\alpha, \beta > -1$. By Definition 2.5, it is possible to take the limit $\alpha, \beta \rightarrow 1^+$ provided that $k \geq 3$. Therefore, taking this limit and setting $n = k - 1 \geq 2$ produces

$$4n^2(n + 2)P_{n+2}^{(-1,-1)}(x) = 4n(n + 1)(2n + 1)xP_{n+1}^{(-1,-1)}(x) - 4n^2(n + 1)P_n^{(-1,-1)}(x)$$

for $n \geq 2$. Dividing the above by $4n$ gives the required recurrence relation (2.18). Finally, the following formula, stated in [20], can be used to compute $P_2^{(-1,-1)}$ and $P_3^{(-1,-1)}$:

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{d^n}{dx^n} \left((1 - x)^{\alpha+n} (1 + x)^{\beta+n} \right) \quad \text{for } n \geq 2.$$

As before, we take $\alpha, \beta \rightarrow 1^+$ in the above to obtain an explicit formula for $P_n^{(-1,-1)}$. □

In addition to computing these polynomials via a recurrence relation, it is also possible to represent each $P_n^{(-1,-1)}$ as the difference of two (rescaled) Legendre polynomials. Since the Legendre polynomials already have efficient implementations in the majority of high-level programming languages, this second approach is particularly appealing.

THEOREM 2.9 (relationship between the Jacobi-like and Legendre polynomials). *For $n \geq 1$, we have*

$$P_{n+1}^{(-1,-1)}(x) = \frac{n}{4n + 2} (Q_{n+1}(x) - Q_{n-1}(x)),$$

where Q_k denotes the k th degree Legendre polynomial defined on $[-1, 1]$.

Proof. Recall that $\frac{d}{dx}(P_{n+1}^{(-1,-1)}) = \frac{n}{2}P_n^{(0,0)}$ for $n \geq 1$, where $P_n^{(0,0)} (= Q_n)$ is the n th degree Legendre polynomial. Therefore, differentiating both sides of (2.18) yields

$$\frac{1}{2}n(n+1)(n+2)Q_{n+1}(x) = (n+1)(2n+1)P_{n+1}^{(-1,-1)}(x) + \frac{1}{2}n(n+1)(2n+1)xQ_n(x) - \frac{1}{2}n(n-1)(n+1)Q_{n-1}(x).$$

Hence by simplifying and rearranging the above, we have that for $n \geq 1$,

$$(2n+1)P_{n+1}^{(-1,-1)}(x) = \frac{1}{2}n(Q_{n+1}(x) - Q_{n-1}(x)) + \frac{1}{2}n((n+1)Q_{n+1}(x) - (2n+1)xQ_n(x) + nQ_{n-1}(x)).$$

We see that the last term is zero by a recurrence relation for Legendre polynomials [19]. \square

In addition to viewing the polynomials $\{e_k\}$ as orthogonal with respect to the weight function $w(x) := \frac{1}{x(1-x)}$, we can characterize them via their iterated time integrals. In particular, for $1 \leq k \leq n-1$, it follows from the integration by parts formula that

$$\begin{aligned} \int_{0 < s_1 < \dots < s_k < 1} e_n(s_1) ds_1 \cdots ds_k &= \int_0^1 e_n(s) d\left(\frac{1}{k!}s^k\right) \\ &= \frac{1}{(k-1)!} \int_0^1 s^{k-1} e_n(s) ds \\ &= -\frac{1}{k!} \int_0^1 s^k e'_n(s) ds \\ &= 0 \quad (\text{by the orthogonality of } e'_n). \end{aligned}$$

Hence for $k \geq 1$, e_k is a polynomial with degree $k+1$ that has roots at 0 and 1 as well as $k-1$ trivial iterated integrals against time. By additionally specifying the k th iterated time integral, it is then possible to characterize the k th polynomial e_k .

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). Hence 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 $[0, T]$,

$$(3.1) \quad \begin{aligned} dy_t &= f_0(y_t) dt + f_1(y_t) \circ dW_t, \\ y_0 &= \xi, \end{aligned}$$

where $\xi \in \mathbb{R}^e$ and f_i denote bounded C^∞ vector fields on \mathbb{R}^e 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 [21] 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:*

$$\begin{aligned}
 (3.2) \quad y_t &= y_s + f_0(y_s)h + f_1(y_s)W_{s,t} + \frac{1}{2}f_1'(y_s)f_1(y_s)W_{s,t}^2 + \frac{1}{2}f_0'(y_s)f_0(y_s)h^2 \\
 &+ f_0'(y_s)f_1(y_s) \int_s^t \int_s^u \circ dW_v du + f_1'(y_s)f_0(y_s) \int_s^t \int_s^u dv \circ dW_u \\
 &+ \frac{1}{6}(f_1'(y_s)f_1'(y_s)f_1(y_s) + f_1''(y_s)(f_1(y_s), f_1(y_s)))W_{s,t}^3 \\
 &+ (f_0'(y_s)f_1'(y_s)f_1(y_s) + f_0''(y_s)(f_1(y_s), f_1(y_s))) \int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_v du \\
 &+ (f_1'(y_s)f_0'(y_s)f_1(y_s) + f_1''(y_s)(f_0(y_s), f_1(y_s))) \int_s^t \int_s^u \int_s^v \circ dW_r dv \circ dW_u \\
 &+ (f_1'(y_s)f_1'(y_s)f_0(y_s) + f_1''(y_s)(f_1(y_s), f_0(y_s))) \int_s^t \int_s^u \int_s^v dr \circ dW_v \circ dW_u \\
 &+ \frac{1}{24}(f_1'(y_s)f_1'(y_s)f_1'(y_s)f_1(y_s) + 3f_1''(y_s)(f_1'(y_s)f_1(y_s), f_1(y_s)) \\
 &\quad + f_1'(y_s)f_1''(y_s)(f_1(y_s), f_1(y_s)) + f_1'''(y_s)(f_1(y_s), f_1(y_s), f_1(y_s)))W_{s,t}^4 \\
 &+ R(h, y_s),
 \end{aligned}$$

where $h := t - s$ and the remainder term has the following uniform estimate for $h < 1$:

$$(3.3) \quad \sup_{y_s \in \mathbb{R}^e} \|R(h, y_s)\|_{L^2(\mathbb{P})} \leq C h^{\frac{5}{2}},$$

where the constant $C > 0$ 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 nontrivial 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 random polynomial with a root at 0 and satisfying*

$$\int_0^1 u^k dW_u^n = \int_0^1 u^k dW_u \quad \text{for } k = 0, 1, \dots, n - 1.$$

Then $W = W^n + Z^n$, where Z^n is a centered Gaussian process independent of W^n . Furthermore, Z^n has the following covariance function:

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

where K_B denotes the standard Brownian bridge covariance function and $\{\lambda_k\}, \{e_k\}$ are the eigenvalues and eigenfunctions that were defined in the proof of 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 (2.13). 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:

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

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

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

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

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

DEFINITION 3.4. *The standard Brownian arch Z is the process $Z := W - \widehat{W}$. By Theorem 3.2, Z is the centered 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 an interval $[s, t]$ with length h encodes the signed area of the associated bridge process,*

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

Remark 3.6. Since $e_1(t) = \sqrt{6}t(1-t)$, we have that $H_{0,1}$ corresponds to $\frac{\sqrt{6}}{6}I_1$ as defined in Theorem 2.2. Thus, $H_{s,t} \sim \mathcal{N}(0, \frac{1}{12}h)$, and $H_{s,t}$ 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 $h = t - s$. While the Brownian arch can be viewed similarly 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 the most variance at its midpoint. In fact, the Brownian parabola gives a relatively uniform estimate of the original path (as illustrated by Figure 3.1).

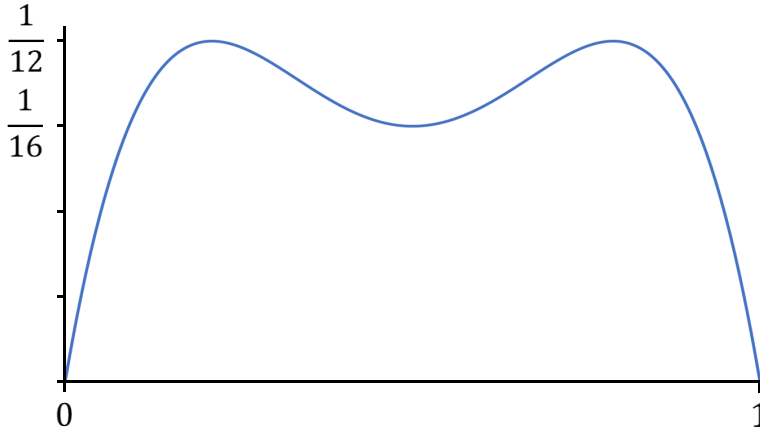


FIG. 3.1. Variance profile of the standard Brownian arch.

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

THEOREM 3.7 (conditional expectation of a nontrivial Brownian time integral).

$$(3.4) \quad \mathbb{E} \left[\int_s^t W_{s,u}^2 du \mid W_{s,t}, H_{s,t} \right] = \frac{1}{3}hW_{s,t}^2 + hW_{s,t}H_{s,t} + \frac{6}{5}hH_{s,t}^2 + \frac{1}{15}h^2.$$

Proof. By the natural Brownian scaling it is enough to prove the result on $[0, 1]$. Recall that $W = \widehat{W} + Z$ where the parabola \widehat{W} is completely determined by (W_1, H_1) and Z is independent of (W_1, H_1) . This leads to the following decomposition for the left-hand side of (3.4):

$$\begin{aligned} \mathbb{E} \left[\int_0^1 W_u^2 du \mid W_1, H_1 \right] &= \mathbb{E} \left[\int_0^1 (\widehat{W}_u + Z_u)^2 du \mid W_1, H_1 \right] \\ &= \mathbb{E} \left[\int_0^1 \widehat{W}_u^2 du + 2 \int_0^1 \widehat{W}_u Z_u du + \int_0^1 Z_u^2 du \mid W_1, H_1 \right] \\ &= \int_0^1 \widehat{W}_u^2 du + 2 \int_0^1 \widehat{W}_u \mathbb{E}[Z_u] du + \int_0^1 \mathbb{E}[Z_u^2] du \\ &= \int_0^1 (uW_1 + 6u(1-u)H_1)^2 du + \int_0^1 u - u^2 - 3u^2(1-u)^2 du. \end{aligned}$$

The result now follows by evaluating the above integrals. □

The above theorem has practical applications for SDE simulation, as $W_{s,t}$ and $H_{s,t}$ are independent Gaussian random variables and can be easily generated or approximated. That said, we should first discuss how the iterated integrals within (3.2) are connected.

DEFINITION 3.8. *The space-space-time Lévy area of Brownian motion over an interval $[s, t]$ is defined as*

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

We can interpret $L_{s,t}$ as an area between processes $\{W_{s,u}\}_{u \in [s,t]}$ and $\{H_{s,u}\}_{u \in [s,t]}$. Moreover, rough path theory provides an algebraic structure (called the log-signature) that relates $(W_{s,t}, H_{s,t}, L_{s,t})$ to the iterated integrals of space-time Brownian motion and ultimately to SDE solutions via the log-ODE method (see [22] for an overview). For our purposes, it is enough to give formulae relating these Lévy areas to integrals.

THEOREM 3.9. *Let $H_{s,t}$ and $L_{s,t}$ denote the Lévy areas of Brownian motion given by Definitions 3.5 and 3.8, respectively. Then the following integral relationships hold:*

$$\begin{aligned} \int_s^t \int_s^u \circ dW_v \circ dW_u &= \frac{1}{2} h W_{s,t} + h H_{s,t}, \\ \int_s^t \int_s^u dv \circ dW_u &= \frac{1}{2} h W_{s,t} - h H_{s,t}, \\ \int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_v \circ dW_u &= \frac{1}{6} h W_{s,t}^2 + \frac{1}{2} h W_{s,t} H_{s,t} + L_{s,t}, \\ \int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_u \circ dW_v &= \frac{1}{6} h W_{s,t}^2 - 2L_{s,t}, \\ \int_s^t \int_s^u \int_s^v dr \circ dW_v \circ dW_u &= \frac{1}{6} h W_{s,t}^2 - \frac{1}{2} h W_{s,t} H_{s,t} + L_{s,t}. \end{aligned}$$

Proof. The result follows from numerous applications of integration by parts. \square

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

THEOREM 3.10 (conditional moments of Brownian space-space-time Lévy area).

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

$$(3.6) \quad \text{Var} (L_{s,t} \mid W_{s,t}, H_{s,t}) = \frac{11}{25200} h^4 + h^3 \left(\frac{1}{720} W_{s,t}^2 + \frac{1}{700} H_{s,t}^2 \right).$$

Proof. The expectation (3.5) is simply a consequence of Theorems 3.7 and 3.9. Without loss of generality, we will consider the above conditional variance on $[0, 1]$.

Since \widehat{W} is determined using the increment W_1 and space-time Lévy area H_1 , we have

$$\begin{aligned} \text{Var} \left(\int_0^1 W_u^2 du \mid W_1, H_1 \right) &= \text{Var} \left(\int_0^1 \widehat{W}_u^2 du + 2 \int_0^1 \widehat{W}_u Z_u du + \int_0^1 Z_u^2 du \mid W_1, H_1 \right) \\ &= 4 \text{Var} \left(\int_0^1 \widehat{W}_u Z_u du \mid W_1, H_1 \right) + \text{Var} \left(\int_0^1 Z_u^2 du \right) \\ &\quad + 4 \text{cov} \left(\int_0^1 \widehat{W}_u Z_u du, \int_0^1 Z_u^2 du \mid W_1, H_1 \right). \end{aligned}$$

Recall $Z = \sum_{k=2}^\infty I_k e_k$ where $\{I_k\}$ are independent centered Gaussian random variables. In particular, this means that Z and $-Z$ have the same law. Therefore, we have that

$$\begin{aligned} \mathbb{E} \left[\int_0^1 \widehat{W}_u Z_u du \mid W_1, H_1 \right] &= 0, \\ \text{cov} \left(\int_0^1 \widehat{W}_u Z_u du, \int_0^1 Z_u^2 du \mid W_1, H_1 \right) &= 0. \end{aligned}$$

The remaining two terms were resolved with the assistance of Mathematica as follows:

$$\begin{aligned} \text{Var} \left(\int_0^1 \widehat{W}_u Z_u du \mid W_1, H_1 \right) &= \int_0^1 \int_0^1 \widehat{W}_u \widehat{W}_v \mathbb{E} [Z_u Z_v \mid W_1, H_1] du dv \\ &= \int_0^1 \int_0^1 \widehat{W}_u \widehat{W}_v (\min(u, v) - uv - 3uv(1-u)(1-v)) du dv \\ &= 2 \int_0^1 \widehat{W}_v \int_0^v u \widehat{W}_u du dv - \left(\int_0^1 u \widehat{W}_u du \right)^2 - 3 \left(\int_0^1 u(1-u) \widehat{W}_u du \right)^2 \\ &= 2 \left(\frac{1}{15} W_1^2 + \frac{13}{60} W_1 H_1 + \frac{13}{70} H_1^2 \right) - \left(\frac{1}{3} W_1 + \frac{1}{2} H_1 \right)^2 - 3 \left(\frac{1}{12} W_1 + \frac{1}{5} H_1 \right)^2 \\ &= \frac{1}{720} W_1^2 + \frac{1}{700} H_1^2. \end{aligned}$$

$$\begin{aligned} \text{Var} \left(\int_0^1 Z_u^2 du \right) &= \mathbb{E} \left[\left(\int_0^1 Z_u^2 du \right)^2 \right] - \left(\mathbb{E} \left[\int_0^1 Z_u^2 du \right] \right)^2 \\ &= \int_0^1 \int_0^1 \mathbb{E} [Z_u^2 Z_v^2] du dv - \left(\mathbb{E} \left[\int_0^1 Z_u^2 du \right] \right)^2 \\ &= \int_0^1 \int_0^1 \mathbb{E} [Z_u^2] \mathbb{E} [Z_v^2] + 2 (\mathbb{E} [Z_u Z_v])^2 du dv - \left(\mathbb{E} \left[\int_0^1 Z_u^2 du \right] \right)^2 \\ &= 4 \int_0^1 \int_0^v (u - uv - 3uv(1-u)(1-v))^2 du dv \\ &= \frac{11}{6300}. \end{aligned}$$

By Theorem 3.9, the above gives the following explicit formula for the conditional variance (3.6):

$$\text{Var}(L_1 | W_1, H_1) = \text{Var}\left(\frac{1}{2} \int_0^1 W_u^2 du \mid W_1, H_1\right) = \frac{11}{25200} + \frac{1}{720} W_1^2 + \frac{1}{700} H_1^2.$$

By the natural Brownian scaling, the result on the interval $[s, t]$ directly follows. \square

Remark 3.11. The conditional variance (3.6) allows one to estimate local $L^2(\mathbb{P})$ errors for certain numerical methods and thus may be useful when choosing step sizes.

Thus, 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 3.10. However, there are many numerical methods that could be used to solve a given SDE.

3.1. Examples of ODE methods. We will consider the following two methods.

DEFINITION 3.12 (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 $k \in [0 \dots N - 1]$, defining Y_{k+1} to be the solution at $u = 1$ of the following ODE:*

$$(3.7) \quad \begin{aligned} \frac{dz}{du} &= f_0(z)h + f_1(z)W_{t_k, t_{k+1}} + [f_1, f_0](z) \cdot hH_{t_k, t_{k+1}} \\ &\quad + [f_1, [f_1, f_0]](z) \cdot \mathbb{E}[L_{t_k, t_{k+1}} \mid W_{t_k, t_{k+1}}, H_{t_k, t_{k+1}}], \\ z_0 &= Y_k, \end{aligned}$$

where $h := \frac{T}{N}$, $t_k := kh$, and $[\cdot, \cdot]$ denotes the standard Lie bracket of vector fields.

DEFINITION 3.13 (the parabola-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 $k \in [0 \dots N - 1]$, defining Y_{k+1} to be the solution at $u = 1$ of the following ODE:*

$$(3.8) \quad \begin{aligned} \frac{dz}{du} &= f_0(z)h + f_1(z)(W_{t_k, t_{k+1}} + (6 - 12u)H_{t_k, t_{k+1}}), \\ z_0 &= Y_k, \end{aligned}$$

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 . While there are different ways of interpolating between the successive approximations Y_k and Y_{k+1} , for this paper we will simply interpolate between such points linearly. To analyze 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 both of these methods involve computing a numerical solution of an ODE, the parabola method does not require explicitly resolving vector field derivatives. The second significant difference can be seen in the Taylor expansions of the methods.

THEOREM 3.14. *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 ,*

$$(3.9) \quad Y_1^{\log} = y_t - [f_1, [f_1, f_0]](y_s) (L_{s,t} - \mathbb{E}[L_{s,t} \mid W_{s,t}, H_{s,t}]) + O(h^{\frac{5}{2}}).$$

Similarly, let Y^{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 ,

$$(3.10) \quad Y_1^{\text{para}} = y_t - [f_1, [f_1, f_0]](y_s) \left(L_{s,t} - \frac{3}{5} H_{s,t}^2 \right) + O(h^{\frac{5}{2}}).$$

Note that $O(h^{\frac{5}{2}})$ denotes terms which can be estimated in an $L^2(\mathbb{P})$ sense as in (3.3).

Proof. In order to derive (3.9), we must compute the Taylor expansion of (3.7). Let F denote the vector field defined in (3.7) 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

$$\begin{aligned} Y_1^{\text{log}} &= y_s + F(y_s) + \frac{1}{2} F'(y_s) F(y_s) + \frac{1}{6} F''(y_s)(F(y_s), F(y_s)) + \frac{1}{6} F'(y_s) F'(y_s) F(y_s) \\ &\quad + \frac{1}{24} F'(y_s) F'(y_s) F'(y_s) F(y_s) + \frac{1}{24} F'(y_s) F''(y_s)(F(y_s), F(y_s)) \\ &\quad + \frac{1}{8} F''(y_s)(F'(y_s) F(y_s), F(y_s)) + \frac{1}{24} F'''(y_s)(F(y_s), F(y_s), F(y_s)) \\ &\quad + \frac{1}{24} \int_0^1 (1-u)^4 \frac{d^5}{du^5} (Y^{\text{log}})(u) du. \end{aligned}$$

We shall first consider the remainder term, which can be directly estimated as follows:

$$\left\| \int_0^1 (1-u)^4 \frac{d^5}{du^5} (Y^{\text{log}})(u) du \right\|_{L^2(\mathbb{P})} \leq \sup_{u \in [0,1]} \left\| \frac{d^5}{du^5} (Y^{\text{log}}(u)) \right\|_{L^2(\mathbb{P})}.$$

One can define the degree of each term in the above Taylor expansion by counting the number of times functions from $\{F, F', F'', \dots\}$ appear. Therefore, after expanding the fifth derivative of Y^{log} we can see that the remainder term has a degree of five. Since the largest component of F is $f_1(\cdot)W_{s,t}$, both F and its derivatives are $O(h^{\frac{1}{2}})$. Hence the remainder term in the above Taylor expansion will be $O(h^{\frac{5}{2}})$ as in (3.3). Moreover, the only terms of degree four that are not $O(h^{\frac{5}{2}})$ are those involving $W_{s,t}^4$.

It is now enough to analyze just the terms appearing in the first line of the expansion. By substituting the formula for F given by (3.7) into the first line and then rearranging the resulting terms, we can obtain a Taylor expansion for Y_1^{log} that resembles (3.2) as

$$\begin{aligned} F(y_s) &= f_0(y_s)h + f_1(y_s)W_{s,t} + (f'_0(y_s)f_1(y_s) - f'_1(y_s)f_0(y_s))hH_{s,t} \\ &\quad + (f'_0(y_s)f'_1(y_s)f_1(y_s) + f''_0(y_s)(f_1(y_s), f_1(y_s)) \\ &\quad \quad - 2f'_1(y_s)f'_0(y_s)f_1(y_s) - 2f''_1(y_s)(f_0(y_s), f_1(y_s))) \\ &\quad + f'_1(y_s)f'_1(y_s)f_0(y_s) + f''_1(y_s)(f_1(y_s), f_0(y_s))) \left(\frac{1}{30}h^2 + \frac{3}{5}hH_{s,t}^2 \right), \\ F'(y_s)F(y_s) &= f'_1(y_s)f_1(y_s)W_{s,t}^2 + (f'_0(y_s)f_1(y_s) + f'_1(y_s)f_0(y_s))hW_{s,t} \\ &\quad + (f'_0(y_s)f'_1(y_s)f_1(y_s) - f'_1(y_s)f'_1(y_s)f_0(y_s) \\ &\quad \quad + f''_0(y_s)(f_1(y_s), f_1(y_s)) - f''_1(y_s)(f_0(y_s), f_1(y_s)))hW_{s,t}H_{s,t} \\ &\quad + f'_0(y_s)f_0(y_s)h^2 + O(h^{\frac{5}{2}}), \end{aligned}$$

$$F'(y_s)F'(y_s)F(y_s) = f'_1(y_s)f'_1(y_s)f_1(y_s)W_{s,t}^3 + f'_0(y_s)f'_1(y_s)f_1(y_s)hW_{s,t}^2 \\ + (f'_1(y_s)f'_0(y_s)f_1(y_s) + f'_1(y_s)f'_1(y_s)f_0(y_s))hW_{s,t}^2 + O(h^{\frac{5}{2}}),$$

$$F''(y_s)(F(y_s), F(y_s)) = f''_1(y_s)(f_1(y_s), f_1(y_s))W_{s,t}^3 + 2f''_1(y_s)(f_0(y_s), f_1(y_s))hW_{s,t}^2 \\ + f''_0(y_s)(f_1(y_s), f_1(y_s))hW_{s,t}^2 + O(h^{\frac{5}{2}}).$$

Therefore, by summing the above formulae for F (and its derivatives) we can derive an expansion of Y_1^{\log} in terms of f_0, f_1 and $(h, W_{s,t}, H_{s,t})$ that has an $O(h^{\frac{5}{2}})$ remainder. By comparing this with the stochastic Taylor expansion (3.2), we see that the result (3.9) follows.

Arguing (3.10) is fairly straightforward and does not require extensive computations. Using the substitution $\widehat{Y}_u = z_{\frac{1}{h}(u-s)}$ for $u \in [s, t]$, the ODE (3.8) can be rewritten as

$$(3.11) \quad d\widehat{Y}_u = f_0(\widehat{Y}_u) du + f_1(\widehat{Y}_u) d\widehat{W}_u, \\ \widehat{Y}_s = y_s,$$

where \widehat{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 (3.11) in the same fashion. The only difference is that Stratonovich integrals with respect to W are replaced with Riemann–Stieltjes integrals against \widehat{W} .

In particular, by the change-of-variable formula for ODEs (exercise 3.17 in [23]) given in [23], we see that the remainder term of such a Taylor expansion will have the following form:

$$\widehat{R} = \sum_{\substack{i_1, \dots, i_n \in \{0,1\} \\ i_1 + \dots + i_n = 2n-4}} \int_{s < r_1 < \dots < r_n < t} f_{i_1, \dots, i_n}(\widehat{Y}_{r_1}) - f_{i_1, \dots, i_n}(\widehat{Y}_s) d\widehat{W}_{r_1}^{i_1} \dots d\widehat{W}_{r_n}^{i_n},$$

where we have identified an additional “zero” coordinate of \widehat{W} with time, $\widehat{W}_t^0 := t$, and for each index (i_1, \dots, i_n) , the function $f_{i_1, \dots, i_n} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ consists of finitely many compositions of f_0, f_1 along with their derivatives (and thus is Lipschitz continuous).

Therefore, each term in the expansion of (3.11) can be estimated in $L^2(\mathbb{P})$ by applying the natural Brownian scaling to the corresponding iterated integral of \widehat{W} with time. As before, the largest differences are the $O(h^2)$ terms involving third order integrals. Fortunately, these iterated integrals of the Brownian parabola can be computed explicitly as

$$\int_s^t \int_s^u \int_s^v \circ dW_r \circ dW_v du - \int_s^t \int_s^u \int_s^v d\widehat{W}_r d\widehat{W}_v du = L_{s,t} - \frac{3}{5}H_{s,t}^2.$$

The result (3.10) is a direct consequence of Theorem 3.9 along with the above. \square

Theorem 3.14 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, while the $O(h^2)$ part for each parabola-ODE local error has

nonzero expectation. Thus, the log-ODE method is globally high order, while the parabola method is not. However, since the parabola-ODE method is straightforward to implement and locally high order, one could expect it to perform well compared to other low order methods. 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 3.15 (strong convergence). *A numerical solution Y for (3.1) is said to converge in a strong sense with order α if there exists a constant $C > 0$ such that*

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

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

DEFINITION 3.16 (weak convergence). *A numerical solution Y for (3.1) is said to converge in a weak sense with order β if for any polynomial p there exists $C_p > 0$ such that*

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

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

THEOREM 3.17 (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 2.0. The parabola-ODE method converges in both strong and weak senses with order 1.0.*

Proof. Recall that Theorem 3.14 establishes the Taylor expansions of both methods. The strong convergence can then be shown as in the proof of Theorem 11.5.1 in [21]. Moreover, the proof of Theorem 11.5.1 also provides the orders of strong convergence. Similarly, weak convergence follows directly from the Taylor expansions (3.9) and (3.10), and the rate of convergence can be shown as in the proof of Theorem 14.5.2 in [21]. \square

4. A numerical example. We shall demonstrate the ideas presented so far using various discretizations of inhomogeneous geometric Brownian motion (IGBM)

$$(4.1) \quad dy_t = a(b - y_t) dt + \sigma y_t dW_t,$$

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

$$(4.2) \quad dy_t = \tilde{a}(\tilde{b} - y_t) dt + \sigma y_t \circ dW_t,$$

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 [15, 16]. 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 3.12). Since the vector fields of (4.2) give constant Lie brackets, this method becomes

$$Y_{k+1}^{\log} := Y_k^{\log} e^{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}} + abh \left(1 - \sigma H_{t_k, t_{k+1}} + \sigma^2 \left(\frac{3}{5} H_{t_k, t_{k+1}}^2 + \frac{1}{30} h \right) \right) \frac{e^{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}} - 1}{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}},$$

$$Y_0^{\log} := y_0.$$

2. **Parabola-ODE method** (see Definition 3.13). As the SDE (4.2) is quite analytically tractable, this method is expressible as

$$Y_{k+1}^{\text{para}} := e^{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}} \left(Y_k^{\text{para}} + ab \int_{t_k}^{t_{k+1}} e^{\tilde{a}(s-t_k) - \sigma \widehat{W}_{t_k, s}} ds \right),$$

$$Y_0^{\text{para}} := y_0.$$

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

3. **Piecewise linear method** (see [24] for definition and proof of convergence). Just as above, this method can be simplified to give a straightforward formula:

$$Y_{k+1}^{\text{lin}} := Y_k^{\text{lin}} e^{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}} + abh \frac{e^{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}} - 1}{-\tilde{a}h + \sigma W_{t_k, t_{k+1}}},$$

$$Y_0^{\text{lin}} := y_0.$$

4. **Milstein method** (see section 6 of [2] and section 10.3 of [21] for overviews). For this method, we shall take the positive part to guarantee nonnegativity:

$$Y_{k+1}^{\text{mil}} := \left(Y_k^{\text{mil}} + \tilde{a}(\tilde{b} - Y_k^{\text{mil}})h + \sigma Y_k^{\text{mil}} W_{t_k, t_{k+1}} + \frac{1}{2} \sigma^2 Y_k^{\text{mil}} W_{t_k, t_{k+1}}^2 \right)^+,$$

$$Y_0^{\text{mil}} := y_0.$$

5. **Euler–Maruyama method** (see sections 4 and 5 of [2] and section 10.2 of [21]). Just as above, we take the positive part of each step to ensure nonnegativity:

$$Y_{k+1}^{\text{eul}} := \left(Y_k^{\text{eul}} + a(b - Y_k^{\text{eul}})h + \sigma Y_k^{\text{eul}} W_{t_k, t_{k+1}} \right)^+,$$

$$Y_0^{\text{eul}} := y_0.$$

Note that the explicit formula for the log-ODE method comes from the Lie brackets:

$$\begin{aligned} [f_1, f_0](y) &= f_0'(y) f_1(y) - f_1'(y) f_0(y) \\ &= -\tilde{a}\sigma y - \tilde{a}\sigma(\tilde{b} - y) \\ &= -ab\sigma, \end{aligned}$$

$$\begin{aligned} [f_1, [f_1, f_0]](y) &= f_0'(y) f_1'(y) f_1(y) - 2f_1'(y) f_0'(y) f_1(y) + f_1'(y) f_1'(y) f_0(y) \\ &\quad + f_0''(y)(f_1(y), f_1(y)) - 2f_1''(y)(f_0(y), f_1(y)) + f_1''(y)(f_1(y), f_0(y)) \\ &= -\tilde{a}\sigma^2 y + 2\tilde{a}\sigma^2 y + \tilde{a}\sigma^2(\tilde{b} - y) \\ &= ab\sigma^2, \end{aligned}$$

and the formula for the parabola-ODE method was derived using a change of variable.

The Euler–Maruyama and Milstein methods are included in the numerical experiment as benchmarks to test how the proposed methods compare to well-known methods. As before, we will be discretizing the SDE over a uniform partition with mesh size h . Sample paths of IGBM computed using the log-ODE method are shown in Figure 4.1.

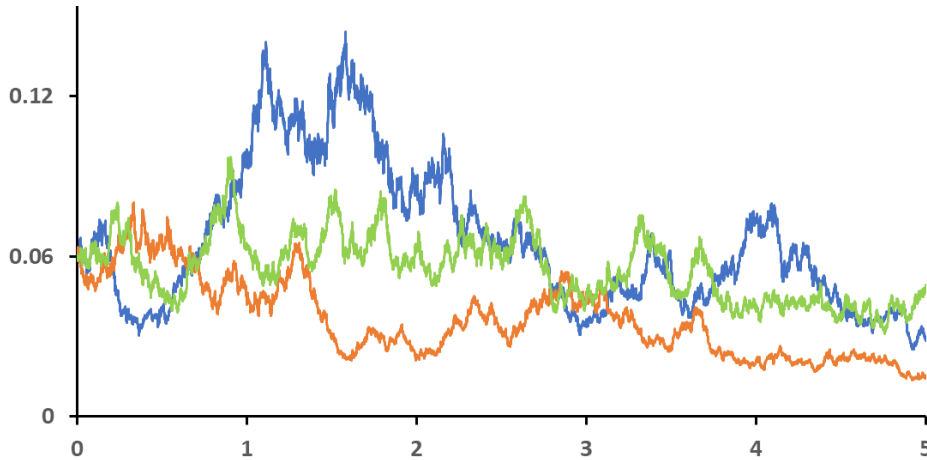


FIG. 4.1. Log-ODE sample paths of IGBM where $a = 0.1$, $b = 0.04$, $\sigma = 0.6$, and $y_0 = 0.06$. The above sample paths experience larger fluctuations the further away from zero they are.

Below is the definition of the error estimators used to analyze 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 define the following estimators for quantifying strong and weak convergence rates:

$$(4.3) \quad S_N := \sqrt{\mathbb{E}[(Y_N - Y_T^{fine})^2]},$$

$$(4.4) \quad E_N := \left| \mathbb{E}[(Y_N - b)^+] - \mathbb{E}[(Y_T^{fine} - b)^+] \right|,$$

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

In this numerical example, we shall use the same parameter values as in [15], namely $a = 0.1$, $b = 0.04$, $\sigma = 0.6$, and $y_0 = 0.06$. We will also fix the time horizon at $T = 5$.

We will now present our results for the numerical experiment that is described above. (Code for this example can be found at <http://github.com/james-m-foster/igbm-simulation>.)

From Figure 4.2 we see that the log-ODE method is by far the most accurate. This is epitomized by the fact that the numerical error produced by 100 steps of the

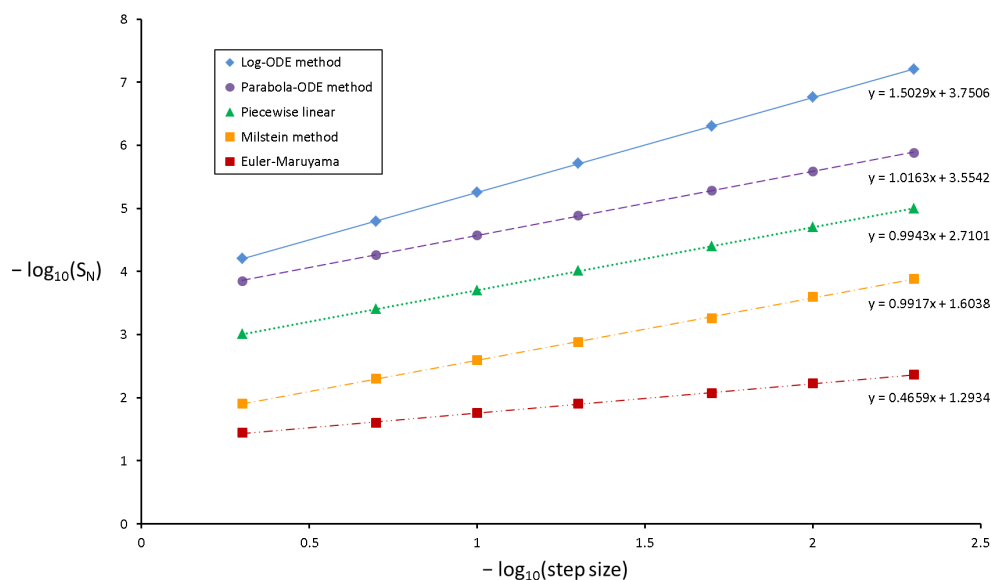


FIG. 4.2. S_N computed with 100,000 sample paths as a function of step size $h = \frac{T}{N}$.

log-ODE method is comparable to the error of the parabola method with 1000 steps. In addition, while there are three methods that 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. As one might expect, the Euler–Maruyama and Milstein schemes both perform poorly. Nevertheless, in order to truly measure the performance of these numerical methods, we should consider the computational costs required for achieving a specified accuracy.

TABLE 4.1

Estimated simulation times for computing 100,000 sample paths that achieve a given accuracy using a single-threaded C++ program on a desktop computer.

	Log-ODE	Parabola	Linear	Milstein	Euler
Estimated time to achieve an accuracy of $S_N = 10^{-4}$	0.179 (s)	0.405 (s)	1.47 (s)	15.4 (s)	0.437 (days)
Estimated time to achieve an accuracy of $S_N = 10^{-5}$	0.827 (s)	3.90 (s)	14.9 (s)	157 (s)	61.2 (days)

TABLE 4.2

Simulation times for computing 100,000 sample paths with 100 steps per path using a single-threaded C++ program on a desktop computer.

	Log-ODE	Parabola	Linear	Milstein	Euler
Computation time (s)	2.44	2.95	1.48	1.18	1.17

The simulation times given by Table 4.1 are estimated from Figure 4.2 along with Table 4.2, which is given below.

Finally, we will investigate the rates of weak convergence for these numerical methods.

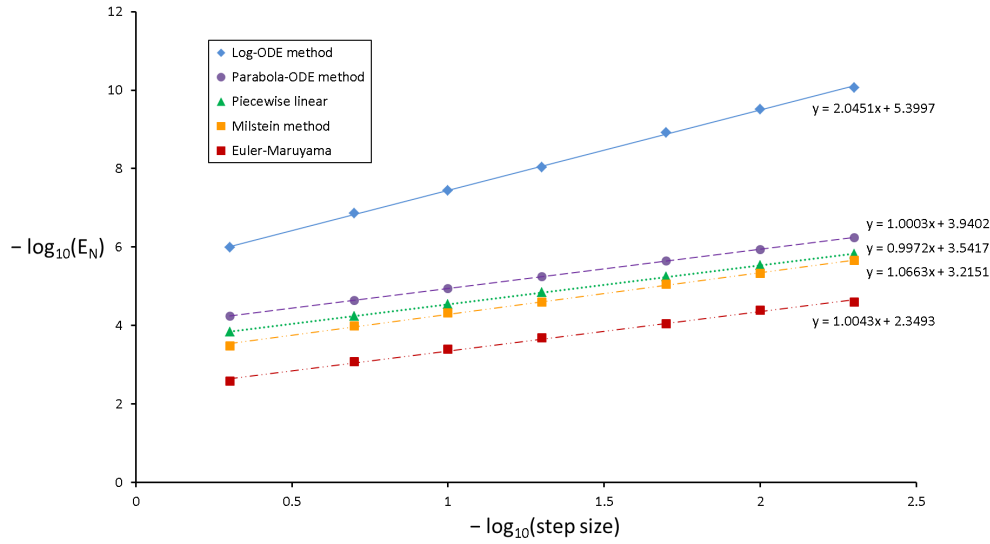


FIG. 4.3. E_N computed with 500,000 sample paths as a function of step size $h = \frac{T}{N}$.

Figure 4.3 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 middle three methods exhibit almost identical convergence profiles. As before, we estimate the simulation times needed to achieve given accuracies (Table 4.3).

TABLE 4.3

Estimated simulation times for computing 100,000 sample paths that achieve a given accuracy using a single-threaded C++ program on a desktop computer.

	Log-ODE	Parabola	Linear	Milstein	Euler
Estimated time to achieve an accuracy of $E_N = 10^{-5}$	< 0.240 (s)	1.69 (s)	2.15 (s)	2.78 (s)	25.5 (s)
Estimated time to achieve an accuracy of $E_N = 10^{-6}$	0.240 (s)	16.9 (s)	21.6 (s)	24.1 (s)	252 (s)

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 4.2 confirms this, and thus sampling may be a bottleneck for these 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 new results established 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.

Furthermore, the results of this paper naturally lead to the following open questions:

- Which weight functions give “explicit eigenfunctions” for Brownian motion? (For example, we could try $w(x) = x$ or $w(x) = \frac{1}{x}$ with $K_W(s, t) = \min(s, t)$.)
- 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?
- Is this polynomial expansion optimal for approximating Lévy area? (See [12].)
- Which conditional moments can be computed for a given stochastic integral?
- How might we construct a piecewise linear path γ with the following properties?

1. $\gamma_s = W_s, \quad \gamma_t = W_t.$
2. $\int_s^t \gamma_{s,u} du = \int_s^t W_{s,u} du.$
3. $\int_s^t \gamma_{s,u}^2 du = \mathbb{E} \left[\int_s^t W_{s,u}^2 du \mid \dots \right].$

- Would this method of construction lead to effective cubature paths? (See [25].)

Given such a path, we can approximate (3.1) with the following “piecewise linear” ODE:

$$(5.1) \quad \frac{dY}{du} = f_0(Y) + f_1(Y) \frac{d\gamma}{du}.$$

(Along each piece of γ , we would discretize (5.1) using an appropriate solver.)

- How effective is the above piecewise linear ODE method for simulating SDEs?
- Can we extend the approximations given in this paper to the SPDE setting?

REFERENCES

- [1] R. M. MAZO, *Brownian Motion: Fluctuations, Dynamics and Applications*, Clarendon Press, 2002.
- [2] D. J. HIGHAM, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Rev., 43 (2001), pp. 525–546, <https://doi.org/10.1137/S0036144500378302>.
- [3] J. M. C. CLARK AND R. J. CAMERON, *The maximum rate of convergence of discrete approximations for stochastic differential equations*, in Stochastic Differential Systems Filtering and Control, Lect. Notes Control Inf. Sci. 25, Springer, 1980, pp. 162–171.
- [4] D. S. GREBENKOV, D. BELYAEV, AND P. W. JONES, *A multiscale guide to Brownian motion*, J. Phys. A, 49 (2016), 043001.
- [5] F. CASTELL AND J. G. GAINES, *The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations*, Ann. Inst. H. Poincaré Probab. Statist., 32 (1996), pp. 231–250.
- [6] S. J. A. MALHAM AND A. WIESE, *Stochastic Lie group integrators*, SIAM J. Sci. Comput., 30 (2008), pp. 597–617, <https://doi.org/10.1137/060666743>.
- [7] J. M. C. CLARK, *An efficient approximation scheme for a class of stochastic differential equations*, in Advances in Filtering and Optimal Stochastic Control, Proceedings of the IFIP-WG 7/1 Working Conference (Cocoyoc, Mexico, 1982), Lect. Notes Control Inf. Sci. 42, Springer-Verlag, 1982, pp. 69–78.
- [8] N. J. NEWTON, *An asymptotically efficient difference formula for solving stochastic differential equations*, Stochastics, 19 (1986), pp. 175–206.
- [9] N. J. NEWTON, *Asymptotically efficient Runge–Kutta methods for a class of Itô and Stratonovich equations*, SIAM J. Appl. Math., 51 (1991), pp. 542–567, <https://doi.org/10.1137/0151028>.
- [10] X. TANG AND A. XIAO, *Asymptotically optimal approximation of some stochastic integrals and its applications to the strong second-order methods*, Adv. Comput. Math., 45 (2019), pp. 813–846.
- [11] P. PRZYBYŁOWICZ, *Linear information for approximation of the Itô integrals*, Numer. Algorithms, 52 (2009), pp. 677–699.
- [12] A. S. DICKINSON, *Optimal approximation of the second iterated integral of Brownian motion*, Stoch. Anal. Appl., 25 (2007), pp. 1109–1128.
- [13] D. LEE AND G. W. WASILKOWSKI, *Approximation of linear functionals on a Banach space with a Gaussian measure*, J. Complexity, 2 (1986), pp. 12–43.
- [14] J. CREUTZIG, T. MÜLLER-GRONBACH, AND K. RITTER, *Free-knot spline approximation of stochastic processes*, J. Complexity, 23 (2007), pp. 867–889.
- [15] L. CAPRIOTTI, Y. JIANG, AND G. SHAIMERDENOVA, *Approximation methods for inhomogeneous geometric Brownian motion*, Int. J. Theor. Appl. Finance, 22 (2019), 1850055.
- [16] B. ZHAO, C. YAN, AND S. HODGES, *Three one-factor processes for option pricing with a mean-reverting underlying: The case of VIX*, Financial Rev., 54 (2019), pp. 165–199.
- [17] G. BEYLKIN, R. COIFMAN, AND V. ROKHLIN, *Wavelets in numerical analysis*, in Wavelets and Their Applications, Jones and Bartlett, 1992, pp. 165–199.
- [18] C. CARMELI, A. DE VITO, AND E. TOIGO, *Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem*, Anal. Appl. (Singap.), 4 (2006), pp. 377–408.
- [19] B. G. S. DOMAN, *The Classical Orthogonal Polynomials*, World Scientific, 2016.
- [20] E. W. WEISSTEIN, *Jacobi Polynomial*, from MathWorld—A Wolfram Web Resource, <http://mathworld.wolfram.com/JacobiPolynomial.html>.
- [21] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
- [22] T. LYONS, *Rough paths, signatures and the modelling of functions on streams*, in Proceedings of the International Congress of Mathematicians—Seoul 2014, Vol. IV, Kyung Moon Sa, 2014, pp. 163–184.
- [23] P. K. FRIZ AND N. B. VICTOIR, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge University Press, 2010.
- [24] E. WONG AND M. ZAKAI, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist., 36 (1965), pp. 1560–1564.
- [25] T. LYONS AND N. VICTOIR, *Cubature on Wiener space: Stochastic analysis with applications to mathematical finance*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460 (2004), pp. 169–198.