

# Rough Paths based Numerical Algorithms in Computational Finance

L.G.Gyurkó\* & T.J.Lyons†

July 23, 2008

## Abstract

The paper connects asymptotic estimations of [3] and [7] with the Rough Paths perspective ([13], [14]) to present a general framework for deriving high order, stable and tractable path-wise approximations of stochastic differential equations. The approach, which can be traced back to [17] and probably earlier, is based on locally deriving and solving random ordinary differential equations. A sufficient condition on the accuracy of the numerical ODE solver is given to ensure the global order is  $\gamma - 1/2$  if the local order is  $\gamma$ . We also point out some practical solutions which make the high order schemes tractable.

## 1 Introduction

This paper describes methods for approximating solutions to Stratonovich differential equations in  $\mathbb{R}^n$  of the type

$$\left. \begin{aligned} d\tilde{\zeta}_t &= V_0(\tilde{\zeta}_t)dt + \sum_{i=1}^d V_i(\tilde{\zeta}_t) \circ dB_t^i \\ \tilde{\zeta}_0 &= x_0 \end{aligned} \right\} \quad (1)$$

where  $B_t = (B_t^1, \dots, B_t^d)^T$  denotes a  $d$ -dimensional Brownian motion and the coefficient functions  $V_i$ ,  $i = 0, \dots, d$  are smooth  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  functions with bounded derivatives up to a certain order. Throughout the paper we assume that the coefficient functions  $V_i$  for  $i = 0, \dots, d$  satisfy sufficient regularity conditions implying the existence and uniqueness of solution to the SDE.

In many applications (e.g. in computational finance) the SDEs are given in Itô form. However, if the coefficient functions are smooth enough, we have strong technical reasons to rewrite those equations in the equivalent Stratonovich form.

---

\*gyurko@maths.ox.ac.uk

†tlyons@maths.ox.ac.uk

We will say, that an approximation  $\widehat{X}_T$  of  $\xi_T$  has (global) strong order  $\gamma$  if there exists a constant  $C$ , not depending on the number of time steps such that

$$\mathbb{E} \left[ \left\| \xi_T - \widehat{X}_T \right\|^2 \right]^{1/2} < C \Delta t^\gamma \quad (2)$$

for some norm  $\| \cdot \|$  on  $\mathbb{R}^n$  where  $\Delta t$  denotes the longest step length produced by discretization.

In computational finance, path-wise approximation of solutions for SDEs is required for example when evaluating path-dependent products (Asian options, Bermudan options, knock-out options etc.). However, note that some simple versions of these problems can be transformed to a European-type structure, and strong approximation is no longer necessary.

The paper is structured as follows. In section 2, the notation is introduced and some basic lemmas are reviewed. In section 3, we collect the ingredients of Rough Paths Theory which are applied throughout the paper and then in Theorem 3.1 we rephrase and prove the main result of [3] in terms of Rough Paths terminology. The random ordinary differential equations derived from the log-signature theorem of section 3 is what the numerical approximations are based on locally. In section 4, we give a general sufficient condition (Theorem 4.1) for the global convergence of the numerical schemes as well as showing (Corollaries 4.1 and 4.2) when the Theorem 3.1 based schemes satisfy this condition, i.e. how accurately one has to approximate the solution to the derived ODEs to ensure the high order global convergence. The high order schemes require the simulation of high order Brownian iterated integrals, however one can make small enough error without losing the global convergence's high order. In section 5, we show how one can approximate the Brownian iterated integrals based on piece-wise linear interpolations of the Brownian paths on a fine sub-scale. A sufficient condition on the fineness of the sub-scale preserving the scheme's order of convergence is provided by Theorem 5.1. Finally in section 6, we demonstrate the efficiency of the log-signature theorem based approximations for some numerical examples.

## 2 Problem setting and notation

In the following subsections, we introduce some rather technical but necessary notation and assertions.

### 2.1 Multi-indices

In the present paper, we introduce objects indexed by *multi-indices* of the form

$$J = (j_1, \dots, j_k) \in \{0, 1, \dots, d\}^k, \quad k = 0, 1, 2, \dots$$

**Definition 2.1.** *The set of all finite multi-indices is denoted by  $\mathcal{A}$ . Given a multi-index  $J = (j_1, \dots, j_k)$  we define*

(i) the length  $|J|$  of the multi-index  $J$  by  $|J| = |(j_1, \dots, j_k)| := k$ .

(ii) the function  $\|\cdot\| : \mathcal{A} \rightarrow \mathbb{N}$  referred to as the norm of a multi-index as

$$\|J\| = \|(j_1, \dots, j_k)\| := |J| + \text{card}\{j_i = 0 \mid 1 \leq i \leq k\}$$

(iii) the concatenation of multi-indices by  $\circ : \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}$ , e.g.

$$(j_1, \dots, j_k) \circ (i_1, \dots, i_l) = (j_1, \dots, j_k, i_1, \dots, i_l)$$

(iv) and finally the right and left decrement respectively as

$$\begin{aligned} J_- &= (j_1, \dots, j_k)_- := (j_1, \dots, j_{k-1}) \\ -J &= -(j_1, \dots, j_k) := (j_2, \dots, j_k) \end{aligned}$$

## 2.2 Vector fields and other differential operators

In some cases, the functions  $V_i, i = 0, \dots, d$  are regarded as vector fields (or first order differential operators) on  $\mathbb{R}^n$ , i.e.

$$V_i(f) = \sum_{j=1}^n V_i^j \frac{\partial}{\partial x_j} f$$

where  $V_i^j$  denotes the  $j$ th coordinate function of  $V_i$  and  $f$  is a smooth  $\mathbb{R}^n \rightarrow \mathbb{R}$  function. The identity function on  $\mathbb{R}^n$  is denoted by  $H$  and  $V_i(H)$  is understood coordinate-wise.

By compositions of coefficient functions, we will refer to the operator composition, e.g.  $V_i \circ V_j$  is defined

$$V_i \circ V_k(f) = \sum_{j=1}^n V_i^j \frac{\partial}{\partial x_j} \left( \sum_{l=1}^n V_k^l \frac{\partial}{\partial x_l} (f) \right)$$

for a smooth function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

Given a multi-index  $J = (j_1, \dots, j_k)$ , we introduce the notation

$$V_J = V_{j_1} \circ \dots \circ V_{j_k}.$$

## 2.3 Stratonovich and Itô iterated integrals

In this section we review some properties of the Stratonovich and the Itô integrals.

**Definition 2.2.** Given two continuous semi-martingales  $X_t$  and  $Y_t$ , we define the Stratonovich integral as

$$\int^\cdot X \circ dY = \int^\cdot X dY + \frac{1}{2} \langle X, Y \rangle. \quad (3)$$

where  $\int^\cdot X dY$  denotes the Itô integral and  $\langle X, Y \rangle$  is the cross-variation process.

To simplify the notation, we introduce  $B_t^0 := t$ , i.e. the 0th coordinate of the Brownian motion is the time. One collection of basic objects in the paper is given by the Stratonovich iterated integrals.

**Definition 2.3.** Given a multi-index  $J = (j_1, \dots, j_k)$  we define the Stratonovich iterated integrals as

$$B_{s,t}^J := \int_{s < t_1 < \dots < t_k < t} \circ dB_{t_1}^{j_1} \circ \dots \circ dB_{t_k}^{j_k} \quad (4)$$

$$B_{s,t}^J(Y) := \int_{s < t_1 < \dots < t_k < t} Y_{t_1} \circ dB_{t_1}^{j_1} \circ \dots \circ dB_{t_k}^{j_k} \quad (5)$$

for an integrable process  $Y_t$ .

The analogous Itô iterated integrals are denoted by  $D_{t,s}^J$  and  $D_{t,s}^J(Y)$  respectively.

We will mainly work with Stratonovich iterated integrals, but for some cases we will need the Itô forms. To derive a general Stratonovich-Itô transformation we need the following relation between multi-indices.

**Definition 2.4.** Let's regard a partition  $(J_1, \dots, J_k)$  of a multi-index  $J$ , i.e.  $J = J_1 \circ \dots \circ J_k$  for some  $k$  such that  $\|J_i\| \leq 2$  for  $i = 1, \dots, k$ . and suppose there exists a multi-index  $I$  which can be partitioned into sub-indices  $I_1, \dots, I_k$  such that  $I = I_1 \circ \dots \circ I_k$  and for each  $i = 1, \dots, k$  either  $J_i = I_i$  with both length 1 or  $J_i = (l, l)$  and  $I_i = (0)$  for some  $l \in \{1, \dots, d\}$ . Then we will say that  $I$  is related to  $J$  through the partitions  $(I_1, \dots, I_k)$  and  $(J_1, \dots, J_k)$  of length  $k$  and denote this relationship by

$$J \sim_k I$$

where  $k$  denotes the number of subsets in the related partitions.

If  $J \sim_k I$ , the function  $v : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{N}$  is defined as

$$v(J, I) := \text{card}\{i \mid 1 \leq i \leq k, J_i \neq I_i\}$$

where  $(J_1, \dots, J_k)$  and  $(I_1, \dots, I_k)$  denotes the partitions of  $J$  and  $I$  relating the two multi-indices.

Applying the definition of the Stratonovich integral, one can derive

$$\int_0^t \int_0^s \circ dB_u^i \circ dB_s^j = \int_0^t \int_0^s dB_u^i dB_s^j + \delta_{i,j} \frac{1}{2} \int_0^t dB_s^0$$

where  $\delta_{i,j} = 1$  if  $i = j$  and 0 otherwise. Using this assertion repeatedly, we get the following general formula

**Lemma 2.1.** For any multi-index  $J$

$$B_{s,t}^J = \sum_{\substack{J \sim_k I \\ k \in \mathbb{N}}} \frac{1}{2^{v(J,I)}} D_{s,t}^I \quad (6)$$

Again, by (3), for some integrable process  $\xi_t$  and smooth function  $f$ , we get

$$\int_0^t f(\xi_u) \circ dB_u^i = \int_0^t f(\xi_u) dB_u^i + (1 - \delta_{i,0}) \frac{1}{2} \int_0^t V_i f(\xi_u) dB_u^0$$

This implies

**Lemma 2.2.** *For any multi-index  $J$*

$$\begin{aligned} B_{s,t}^J(f(\xi \cdot)) &= \sum_{\substack{J \sim_k I \\ k \in \mathbb{N}}} \frac{1}{2^{v(J,I)}} D_{s,t}^I(f(\xi \cdot)) \\ &\quad + (1 - \delta_{j_1,0}) \frac{1}{2} \sum_{\substack{-J \sim_k I \\ k \in \mathbb{N}}} \frac{1}{2^{v(J,I)}} D_{s,t}^I \left( D_{s,\cdot}^0(V_{j_1} f(\xi \cdot)) \right) \end{aligned}$$

The equation (3) also implies the reverse Lemma:

**Lemma 2.3.** *For any multi-index  $J$*

$$\begin{aligned} D_{s,t}^J &= \sum_{\substack{J \sim_k I \\ k \in \mathbb{N}}} (-1)^{v(J,I)} \frac{1}{2^{v(J,I)}} B_{s,t}^I \\ D_{s,t}^J(f(\xi \cdot)) &= \sum_{\substack{J \sim_k I \\ k \in \mathbb{N}}} (-1)^{v(J,I)} \frac{1}{2^{v(J,I)}} B_{s,t}^I(f(\xi \cdot)) \\ &\quad + (1 - \delta_{j_1,0}) \frac{1}{2} \sum_{\substack{-J \sim_k I \\ k \in \mathbb{N}}} (-1)^{v(J,I)+1} \frac{1}{2^{v(J,I)}} B_{s,t}^I \left( B_{s,\cdot}^0(V_{j_1} f(\xi \cdot)) \right) \end{aligned}$$

**Remark 2.1.** Note that the Itô form of a Stratonovich iterated integral of the type (4) corresponding to a multi-index  $J$  has a term corresponding to an all-zero multi-index if and only if  $J = J_1 \circ \dots \circ J_k$  such that either  $J_i = (0)$  or  $J_i = (l, l)$  for all  $i = 1, \dots, k$  and for some  $1 \leq l \leq d$ .

The Itô isometry and the Hölder inequality imply the  $L^2$  boundedness of the Itô iterated integrals, i.e.

**Lemma 2.4.** *There exist constants  $C_1$  and  $C_2$  depending on the multi-index  $J$  only, such that*

$$\begin{aligned} \mathbb{E} \left[ |D_{s,t}^J|^2 \right]^{1/2} &\leq C_1 (t-s)^{\|J\|/2} \\ \mathbb{E} \left[ |D_{s,t}^J(f(\xi \cdot))|^2 \right]^{1/2} &\leq C_2 \|f\|_\infty (t-s)^{\|J\|/2} \end{aligned}$$

assuming that  $\|f\|_\infty$  is finite.

The error of the numerical schemes we consider in this paper can be represented in terms of iterated integrals. The following lemma is the key step for determining an upper bound for the global error.

**Lemma 2.5.** For each multi-index  $J$ , there exists a constant  $C_J$  not depending on  $k$  such that for any Borel function  $f$  satisfying  $\sup_{t \in [0, T]} \mathbb{E} [\|f(\xi_t)\|^2] < \infty$ , the following inequality holds

$$\mathbb{E} \left[ \sup_{1 \leq l \leq k} \left( \sum_{i=1}^l D_{t_{i-1}, t_i}^J(f(\xi_{\cdot})) \right)^2 \right]^{1/2} < C_J \Delta t^m \sup_{t \in [0, T]} \mathbb{E} [\|f(\xi_t)\|^2]^{1/2} \quad (7)$$

where

- (i)  $m = (\|J\| - 2)/2$  if  $\|J\| = |J|$  or  $J = (j, j)$  for some  $j, 1 \leq j \leq d$
- (ii)  $m = (\|J\| - 1)/2$  otherwise

and furthermore  $\Delta t$  denotes the longest step length in the partition  $0 = t_0 < t_1 < \dots < t_n \leq T$ .

One can derive the proof for the above Lemma by applying Lemma 2.4 and Doob's inequality, exploiting the fact that in case  $\|J\| \neq |J|$  the terms in the expected value (7) are orthogonal. A detailed proof can be found in Chapter 10 of [8].

The Stratonovich version of Lemma 2.5 is as follows.

**Lemma 2.6.** For each multi-index  $J$  there exists a constant  $C_J$  not depending on  $k$  such that for any smooth function  $f$  as in Lemma 2.5, the following inequality holds

$$\mathbb{E} \left[ \sup_{1 \leq l \leq k} \left( \sum_{i=1}^l B_{t_{i-1}, t_i}^J(f(\xi_{\cdot})) \right)^2 \right]^{1/2} < C_J \Delta t^m \max \left\{ \sup_{t \in [0, T]} \mathbb{E} [\|f(\xi_t)\|^2]^{1/2}, \sup_{t \in [0, T], 1 \leq i \leq d} \mathbb{E} [\|V_i f(\xi_t)\|^2]^{1/2} \right\} \quad (8)$$

where

- (i)  $m = (\|J\| - 2)/2$  if  $J$  or  $-J$  is of the form  $J_1 \circ \dots \circ J_s$  for some  $s$  such that for all  $i = 1, \dots, s$  either  $J_i = (0)$  or  $J_i = (j, j)$  for some  $j, 1 \leq j \leq d$  and
- (ii)  $m = (\|J\| - 1)/2$  otherwise

and furthermore  $\Delta t$  denotes the longest step length in the partition  $0 = t_0 < t_1 < \dots < t_n \leq T$ .

*Proof.* There exist constants  $C_1$  and  $C_2$  such that using Lemma 2.2

$$\mathbb{E} \left[ \left( \sup_{1 \leq l \leq k} \sum_{i=1}^l B_{t_{i-1}, t_i}^J(f(\xi_{\cdot})) \right)^2 \right]$$

$$\begin{aligned}
&\leq C_1 \sum_{\substack{J \sim_l I \\ l \in \mathbb{N}}} \mathbb{E} \left[ \sup_{1 \leq l \leq k} \left( \sum_{i=1}^l \frac{1}{2^{v(J,I)}} D_{t_{i-1}, t_i}^I (f(\xi_{\cdot})) \right)^2 \right] \\
&+ C_2 \sum_{\substack{J \sim_l I \\ l \in \mathbb{N}}} \mathbb{E} \left[ \sup_{1 \leq l \leq k} \left( \sum_{i=1}^l \frac{1}{2^{v(J,I)}} D_{t_{i-1}, t_i}^I \left( D_{t_{i-1}, \cdot}^0 (V_{j_1} f(\xi_{\cdot})) \right) \right)^2 \right]
\end{aligned}$$

Then Lemma 2.5 and Remark 2.1 imply the assertion.  $\square$

The following lemma is a special case of Lemma 2.6.

**Lemma 2.7.** *For each multi-index  $J$  there exists a constant  $C_J$  not depending on  $k$  such that for any Borel function  $f$  as in Lemma 2.5, the following inequality holds*

$$\mathbb{E} \left[ \left( \sup_{1 \leq l \leq k} \sum_{i=1}^l f(\xi_{t_{i-1}}) B_{t_{i-1}, t_i}^J \right)^2 \right]^{1/2} < C_J \Delta t^m \sup_{t \in [0, T]} \mathbb{E} [\|f(\xi_t)\|^2]^{1/2} \quad (9)$$

where

- (i)  $m = (\|J\| - 2)/2$  if  $J$  is of the form  $J_1 \circ \dots \circ J_s$  for some  $s$  such that for all  $i = 1, \dots, s$  either  $J_i = (0)$  or  $J_i = (j, j)$  for some  $j, 1 \leq j \leq d$  and
- (ii)  $m = (\|J\| - 1)/2$  otherwise

and furthermore  $\Delta t$  denotes the longest step length in the partition  $0 = t_0 < t_1 < \dots < t_n \leq T$ .

**Lemma 2.8.** *Given the multi-indices  $J_1, \dots, J_k$  there exists a constant not depending on  $s$  and  $t$  such that*

$$\mathbb{E} \left[ \left| \prod_{i=1}^k B_{s,t}^{J_i} \right|^2 \right]^{1/2} < C(t-s)^{(1/2) \sum_{i=1}^k \|J_i\|}$$

The lemma is a corollary of Lemma 2.2, Lemma 2.4 and the fact that for the multi-indices  $I = (i_1, \dots, i_q)$  and  $J = (j_1, \dots, j_r)$

$$\begin{aligned}
B_{s,t}^I B_{s,t}^J &= \int_{s < u_1 < \dots < u_q < t} \circ dB_{u_1}^{i_1} \dots \circ dB_{u_q}^{i_q} \int_{s < v_1 < \dots < v_r < t} \circ dB_{v_1}^{j_1} \dots \circ dB_{v_r}^{j_r} \\
&= \sum_{\sigma \in \text{Shuffles}(q,r)} \int_{s < w_1 < \dots < w_{q+r} < t} \circ dB_{w_1}^{\sigma^{-1}(1)} \dots \circ dB_{w_{q+r}}^{\sigma^{-1}(q+r)}
\end{aligned}$$

where  $\text{Shuffles}(q, r)$  is the set of all permutations of the set  $\{1, 2, \dots, q+r\}$  such that  $\sigma \in \text{Shuffles}(q, r)$  if and only if

$$\sigma(1) < \sigma(2) < \dots < \sigma(q) \text{ and } \sigma(q+1) < \sigma(q+2) < \dots < \sigma(q+r).$$

### 3 The log-signature theorem

In the log-signature theorem, a random ordinary differential equation is derived which approximates the solution for the SDE locally. In some forms, this theorem was proved in [1], [3] and [7] and in addition a local error formula was given. The log-signature theorem is also an extension of the result in [16] to non-linear and stochastic differential equations. In this section the theorem is reformulated in terms of the log-signature of the driving noise and a more specific local error formula is also derived.

#### 3.1 Stochastic Taylor expansion

The expansion presented in this section is a stochastic extension of the deterministic Taylor expansion using Itô's differentiation rule. Given the SDE

$$\left. \begin{aligned} d\tilde{\xi}_t &= \sum_{i=0}^d V_i(\tilde{\xi}_t) \circ dB_t^i \\ \tilde{\xi}_0 &= x_0 \end{aligned} \right\} \quad (10)$$

the Itô lemma in Stratonovich form is

$$df(\tilde{\xi}_t) = \sum_{i=0}^d V_i(f(\tilde{\xi}_t)) \circ dB_t^i$$

By repeated use of Itô's lemma coordinate-wise on the identity function  $H$  of  $\mathbb{R}^n$ , one can derive the *stochastic Taylor expansion*

**Lemma 3.1.**

$$\tilde{\xi}_t = \tilde{\xi}_0 + \sum_{\|J\| \leq m} V_J(\tilde{\xi}_0) B_{0,t}^J + \sum_{\substack{\|(j_2, j_3, \dots)\| = m \\ \|(j_1, j_2, \dots)\| > m}} B_{0,t}^{(j_1, j_2, \dots)} (V_{(j_1, j_2, \dots)}(\tilde{\xi}_\cdot)) \quad (11)$$

where the last term is referred to as the remainder and denoted by  $R_{0,t}^m$ . Furthermore, there exists a constant  $C$  not depending on  $t$  such that

$$\begin{aligned} \mathbb{E} \left[ \|R_{0,t}^m\|^2 \right]^{1/2} &\leq C \sum_{\substack{\|(j_2, j_3, \dots)\| = m \\ \|(j_1, j_2, \dots)\| = m+1}} t^{(m+1)/2} \|V_I H\|_\infty \\ &\quad + C \sum_{\substack{\|(j_2, j_3, \dots)\| = m \\ \|(j_1, j_2, \dots)\| = m+2}} t^{(m+2)/2} \|V_I H\|_\infty \end{aligned}$$

*Proof.* The  $L^2$  bound on the remainder term is implied by Lemma 2.2 and Lemma 2.4.  $\square$



### 3.2 Algebraic setting

Lemma 3.1 demonstrates that the stochastic Taylor expansion on short time steps gives a good approximation of the solution to the corresponding SDE. It also demonstrates the importance of the information provided by the set of iterated integrals on the time interval  $[0, t]$  corresponding to multi-indices of norm at most  $m$ . As it was pointed out in [13] and [14], these iterated integrals can be formulated into an algebraic object called the Brownian *signature*. The signature is a random element in the tensor algebra  $\mathcal{T}$  defined as

**Definition 3.1.**  $\mathcal{T}$  denotes the non-commutative tensor algebra generated by the letters  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_d$ . For a multi-index  $J = (j_1, \dots, j_k)$  we introduce the notation

$$\varepsilon_J = \varepsilon_{j_1} \otimes \dots \otimes \varepsilon_{j_k}$$

Furthermore, the empty word is denoted by  $\mathbf{1}$ .

A general element in  $\mathcal{T}$  is of the form

$$\mathbf{a} = \sum_{J \in \mathcal{A}} a_J \varepsilon_J$$

with real coefficients  $a_J$ ,  $J \in \mathcal{A}$ . We define the operations in  $\mathcal{T}$  as follows.

**Definition 3.2.** The multiplication of the tensor elements is written as

$$\begin{aligned} \mathbf{a} \otimes \mathbf{b} &= \sum_{J, I \in \mathcal{A}} a_J b_I \varepsilon_{J \circ I} \\ \mathbf{a}^{\otimes k} &:= \underbrace{\mathbf{a} \otimes \dots \otimes \mathbf{a}}_{k \text{ times}} \end{aligned}$$

The exponential of a tensor element is defined by the power series

$$\exp(\mathbf{a}) := \mathbf{1} + \sum_{k \geq 1} \frac{1}{k!} \mathbf{a}^{\otimes k}$$

If  $\mathbf{a} = \mathbf{1} + \sum_{J \in \mathcal{A}, |J| > 0} a_J \varepsilon_J$ , then we define the logarithm of  $\mathbf{a}$  as

$$\log(\mathbf{a}) := \sum_{k \geq 1} \frac{(\mathbf{1} - \mathbf{a})^k}{k} (-1)^k$$

We denote the simplex  $\{(s, t) \mid 0 \leq s \leq t \leq T\}$  by  $\Delta_T$ .

**Definition 3.3. (Signature, log-signature)** The Stratonovich Brownian signature is a  $\Delta_T \rightarrow \mathcal{T}$  function, defined as

$$\mathbf{S}_{s,t}(\circ B) := \mathbf{1} + \sum_{J \in \mathcal{A}} B_{s,t}^J \varepsilon_J$$

the log-signature is defined as

$$L_{s,t} := \log(\mathbf{S}_{s,t}).$$

**Definition 3.4.** We define the Lie bracket of the elements  $\mathbf{a}$  and  $\mathbf{b}$  as  $[\mathbf{a}, \mathbf{b}] := \mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}$ . The vector space spanned by all finite lie brackets formed by the letters  $\varepsilon_0, \dots, \varepsilon_d$  is the Lie algebra  $\mathcal{L}$ . The elements of the closure  $\overline{\mathcal{L}}$  are referred to as a Lie series.

For further algebraic background we refer the reader to [19].

The proof for following properties of the Brownian signature can be found in [14].

**Lemma 3.2. (Properties of the Brownian signature)**

(i) The signature is multiplicative, i.e.

$$\mathbf{S}_{s,t} \otimes \mathbf{S}_{t,u} = \mathbf{S}_{s,u}$$

(ii) The log signature  $L_{0,t}$  is a.s. a Lie series.

(iii)  $\exp(L_{0,t}) = S_{0,t}$

**Definition 3.5.** We also introduce the truncation operator  $\pi_m$  defined as

$$\pi_m \sum_{J \in \mathcal{A}} a_J \varepsilon_J := \sum_{\substack{J \in \mathcal{A} \\ \|J\| \leq m}} a_J \varepsilon_J.$$

**Definition 3.6.** Finally we introduce an algebra homomorphism  $\Gamma$  mapping tensor algebra elements to differential operators generated by  $\Gamma(\varepsilon_i) := V_i, i = 0, 1, \dots, d$ .

Note that the function  $\Gamma$  maps Lie algebra elements to vector fields.

### 3.3 Proof of the log-signature theorem

Given the algebraic setting of the previous section we rephrase the result of [1], [3] and [7] in terms of the log-signature. One possible way to put the main theorem of [3] into the new terminology is to adapt and combine the proof of [3] and the second proof for the main theorem in [21]. However, there is an alternative way as follows. We apply some ideas from [9] and [10].

Given a vector field  $W$ , the solution for the ordinary differential equation

$$\left. \begin{aligned} dy_t &= W(y_t)dt \\ y_0 &= x_0 \end{aligned} \right\} \quad (12)$$

is denoted by

$$y_t = \text{Exp}(tW)(y_0).$$

**Theorem 3.1. (Log-signature theorem)** Given the SDE (10) with smooth coefficient functions  $V_0, \dots, V_d$  with bounded derivatives up to order  $2m$ , and a small enough  $t$ , then

$$\xi_t - y_1 = R_1 \quad (13)$$

and there exists a constant  $C_1$  not depending on  $t$ , such that

$$\mathbb{E} \left[ \|R_1\|^2 \right]^{1/2} \leq C_1 \sum_{m+1 \leq \|I\| \leq 2m} t^{\|I\|/2} \|V_I H\|_\infty \quad (14)$$

where

$$\begin{aligned} y_1 &= \text{Exp}(\Gamma(\pi_m L_{0,t}))(x_0) \\ L_{0,t} &= \log[\mathbf{S}_{0,t}(\circ B)] \end{aligned}$$

Furthermore

$$R_1 = \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}](\xi_0) + R_2 \quad (15)$$

such that there exists a constant  $C_2$  not depending on  $t$  but satisfying

$$\mathbb{E} \left[ \|R_2\|^2 \right]^{1/2} \leq C_2 \sum_{m+2 \leq \|I\| \leq 2m} t^{\|I\|/2} \|V_I H\|_\infty \quad (16)$$

**Remark 3.1.** Note that the differential operator defined as  $W := \Gamma(\pi_m L_{0,t})$  is a.s. the projection of a Lie algebra element and hence a vector field.

*Proof.* Applying the notation introduced in the previous section, the stochastic Taylor expansion can also be represented as the image of the signature under the algebra homomorphism

$$\xi_t = \Gamma[\pi_m \mathbf{S}_{0,t}(\circ B)](\xi_0) + R_{0,t}^m. \quad (17)$$

Since  $\exp(L_{0,t}) = \mathbf{S}_{0,t}(\circ B)$ , the stochastic Taylor expansion (11) can be rewritten as

$$\xi_t = \Gamma \left[ \pi_m \sum_{j=0}^m \frac{1}{j!} (L_{0,t})^{\otimes j} \right] (\xi_0) + R_{0,t}^m = \Gamma \left[ \pi_m \sum_{j=0}^m \frac{1}{j!} (\pi_m L_{0,t})^{\otimes j} \right] (\xi_0) + R_{0,t}^m. \quad (18)$$

On the other hand, we derive a deterministic Taylor-type expansion of  $y_1$  as follows. For any differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$df(y_t) = W(f(y_t))dt. \quad (19)$$

Applying (19) coordinate-wise on the identity function  $H$  of  $\mathbb{R}^n$ , we get

$$\begin{aligned} y_1 &= y_0 + \int_0^1 W \circ H(y_t) dt \\ &= y_0 + W(y_0) + \int_{0 < s_1 < s_2 < 1} W \circ W \circ H(y_{s_1}) ds_1 ds_2. \end{aligned} \quad (20)$$

By Definition 3.2 of the log function, the vector field  $W$  can be represented as a finite linear combination of the functions  $V_J$  of the form

$$W = \sum_{\|J\| \leq m} \left( \sum_{\substack{I_1 \circ \dots \circ I_l = J \\ l \leq |J|}} a_{I_1, \dots, I_l} B_{0,t}^{I_1} \cdots B_{0,t}^{I_l} \right) V_J$$

where the real coefficients  $a_{J_1, \dots, J_l}$  only depend on the multi-indices  $J_1, \dots, J_l$ . We introduce the notation

$$A_{s,t}^J = \sum_{\substack{J_1 \circ \dots \circ J_l = J \\ l \leq |J|}} a_{J_1, \dots, J_l} B_{s,t}^{J_1} \cdots B_{s,t}^{J_l}.$$

The expression under the integral in (20) can be written as a weighted sum of functions where the weights are products of  $A_{0,t}^J$ -type terms. If the weight of a term is  $A_{0,t}^{J_1} \cdots A_{0,t}^{J_k}$  for some multi-indices  $J_1, \dots, J_k$  we will say that the degree of the term is  $\|J_1\| + \dots + \|J_k\|$ . Applying (19) repeatedly on the terms under the integral sign with degree at most  $m$ , one can derive the following formula

$$y_1 = \Gamma \left[ \pi_m \sum_{j=0}^m \frac{1}{j!} (\pi_m L_{0,t})^{\otimes j} \right] (y_0) + \bar{R}_{0,t}^m. \quad (21)$$

Lemma 2.8 implies

$$\mathbb{E} \left[ \left\| \bar{R}_{0,t}^m \right\|^2 \right]^{1/2} < C \sum_{m+1 \leq \|I\| \leq 2m} t^{\|I\|/2} \|V_I H\|_\infty. \quad (22)$$

for some constant not depending on  $t$ .

Comparing the solution for the SDE (18) and the solution for the ODE (21), the difference is given by  $\xi_t - y_1 = R_{0,t}^m - \bar{R}_{0,t}^m$ . Given the stochastic and deterministic Taylor expansions,  $R_{0,t}^m$  contains terms of degree  $m+1$  and  $m+2$ . Hence the inequality (22) and Lemma 2.4 imply (14), (15) and (16).  $\square$

Theorem 3.1 shows the importance of locally derived random ODEs.

**Definition 3.7.** We call a numerical approximation for the solution to the SDE (10) ODE-approach based corresponding to the  $m$ -truncated log-signature if at each step a random ODE is derived from  $\pi_m L$  and solved or numerically approximated.

**Remark 3.2.** Both [5] and [8] define discretization schemes (e.g. the Heun, Milstein and other higher order schemes), which are based on the stochastic Taylor expansion as well as some other similar schemes. One can regard those schemes as particular versions of the ODE approach based schemes.

**Remark 3.3.** There are at least two advantages of the ODE approach based schemes implemented with highly accurate ODE solvers over the purely stochastic Taylor expansion based schemes. The first one is that the error of the former method is given by  $R_{0,t}^m - \bar{R}_{0,t}^m$  whereas the error of the latter scheme is  $R_{0,t}^m$ . They are of the same order, however in absolute terms the former might be more accurate because the terms in  $R_{0,t}^m$  are reduced or completely canceled out by the terms in  $\bar{R}_{0,t}^m$ . The second advantage is the stability, namely the ODE approach approximates  $\xi_t$  with points which are reachable by the diffusion. E.g. if the diffusion a.s. does not leave a subset of  $\mathbb{R}^n$ , the ODE approach results in points from or very close to this subset. The purely stochastic Taylor expansion based schemes do not have this property.

### 3.4 Intermediate error

Theorem 3.1 only gives information on the distance of  $\xi_t$  and  $y_1$ . One might be interested in how accurate an approximation the ODE gives in intermediate points of the time interval  $[0, t]$ . The following lemma gives an answer.

**Lemma 3.3.** *Given the conditions of Theorem 3.1 and  $\lambda \in (0, 1)$ ,*

$$\xi_{\lambda t} - y_\lambda = \Gamma \left[ \pi_m \sum_{j=0}^m \frac{1}{j!} \left\{ \lambda^j (\pi_m L_{0,t})^{\otimes j} - (\pi_m L_{0,\lambda t})^{\otimes j} \right\} \right] (\xi_0) + R \quad (23)$$

where

$$\sqrt{\mathbb{E} [\|R\|^2]} \leq C \sum_{m+1 \leq \|I\| \leq 2m} t^{\|I\|/2} \|V_I H\|_\infty$$

with a constant  $C$  not depending on  $t$  or  $\lambda$ .

*Proof.* Similarly as in (18)

$$\xi_{\lambda t} = \Gamma \left[ \pi_m \sum_{j=0}^m \frac{1}{j!} (\pi_m L_{0,\lambda t})^{\otimes j} \right] (\xi_0) + R_{0,\lambda t}^m. \quad (24)$$

The solution for the ODE at  $\lambda$ , based on (21) can be represented as

$$y_\lambda = \Gamma \left[ \pi_{2m} \sum_{j=0}^m \frac{\lambda^j}{j!} (\pi_m L_{0,t})^{\otimes j} \right] (y_0) + \bar{R}_{0,\lambda t}^m \quad (25)$$

which implies the lemma.  $\square$

### 3.5 The nilpotent case

In this section we assume that there exists a positive integer  $m$  such that the coefficient functions  $V_i$ ,  $i = 0, \dots, d$  satisfy the  $m$ -nilpotent condition, i.e.

**Definition 3.8.** *Let's suppose that there exist a positive integer  $m$  such that any bracket of the letters  $\varepsilon_{i_1}, \dots, \varepsilon_{i_k}$  (e.g.  $[\varepsilon_{i_1}, [\dots, [\varepsilon_{i_{k-1}}, \varepsilon_{i_k}] \dots]]$  etc.),  $i_j \in \{0, \dots, d\}$  for  $j = 1, \dots, k$  satisfying  $\|(i_1, \dots, i_k)\| \geq m$  projected to the zero operator by the function  $\Gamma$ . Then we say that the coefficient functions  $V_i$ ,  $i = 0, \dots, d$  satisfy the  $m$ -nilpotent condition.*

This case  $\Gamma [\pi_k L_{0,t}] = \Gamma [L_{0,t}]$  for all  $k \geq m$ , in particular

$$\Gamma [\pi_k \mathbf{S}_{0,t}(\circ B)] = \Gamma [\pi_k \exp(L_{0,t})] = \Gamma [\pi_k \exp(\pi_m L_{0,t})] = \Gamma \left[ \pi_k \sum_{j=0}^{\infty} \frac{1}{j!} \pi_m L_{0,t}^{\otimes j} \right] \quad (26)$$

Hence, we have the following.

**Lemma 3.4.** *Let's assume that the coefficient functions of the SDE (10) satisfy the  $m$ -nilpotent condition for some positive integer  $m$ . If for a given  $x_0$  the series*

$$\Gamma[\mathbf{S}_{0,t}(\circ B)](x_0) = \Gamma\left[\sum_{j=0}^{\infty} \frac{1}{j!} \pi_m L_{0,t}^{\otimes j}\right](x_0) = \sum_{j=0}^{\infty} \frac{1}{j!} \Gamma[\pi_m L_{0,t}]^j(x_0) \quad (27)$$

*is convergent, then*

$$\xi_t = \text{Exp}(\Gamma(\pi_m L_{0,t}))(x_0) = \text{Exp}(\Gamma(L_{0,t}))(x_0) \quad (28)$$

*is satisfied without any remainder term.*

**Example 3.1.** From the practical point of view, Lemma 3.4 tells us that given the vector field corresponding to the  $m$ -truncated log-signature, the solution for the derived ODE results in an exact solution for the SDE at time  $t$ . Note that it is only exact at time  $t$ . For the intermediate time points see the previous section.

One simple example for the nilpotent case is given by the geometric Brownian motion

$$d\xi_t = \left(\mu - \frac{1}{2}\sigma^2\right) \xi_t dt + \sigma \xi_t \circ dB$$

which is 2-nilpotent. Lemma 3.4 states that

$$\xi_t = \text{Exp}[\Gamma(\varepsilon_0 t + \varepsilon_1 B_{0,t})](\xi_0) = \text{Exp}(W)(\xi_0). \quad (29)$$

where

$$W(x) = ((\mu - \sigma^2/2)t + \sigma B_{0,t})x$$

On the other hand, it is well known that

$$\xi_t = \xi_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_{0,t}}$$

i.e. the solution only depends on the increment and is equivalent to (29).

## 4 From local to global error

In this section, discretization schemes given by a random sequence  $\xi_0 = \widehat{X}_{t_0}, \widehat{X}_{t_1}, \dots, \widehat{X}_{t_k}$  approximating the solution to the SDE (10) are analysed. The local error of the scheme is defined as follows.

**Definition 4.1.** *Given a discretization scheme by a random sequence  $\widehat{X}_{t_0}, \dots, \widehat{X}_{t_k}$  approximating the solution to the SDE (10), for  $i = 1, \dots, k$  we define the local error corresponding to the time interval  $[t_{i-1}, t_i]$  as  $E_i := Y_{t_i}^i - \widehat{X}_{t_i}$  where  $Y_t^i$  is the solution to the SDE*

$$\left. \begin{aligned} dY_t^i &= \sum_{i=0}^d V_i(Y_t^i) \circ dB_t^i \\ Y_{t_{i-1}}^i &= \widehat{X}_{t_{i-1}} \end{aligned} \right\} \quad (30)$$

The global error of the same scheme corresponding to the time interval  $[0, t_k]$  is defined as  $\xi_{t_k} - \widehat{X}_{t_k}$ .

The local error of discretization schemes based on the exact solution of the ODE derived from the log-signature of the driving noise is given by Theorem 3.1. The practical discretization schemes are derived by choosing a (numerical) solver for the ODE. However, from Theorem 3.1 the order of the global convergence is not obvious and also the accuracy required from the numerical ODE solver to ensure the global convergence is unclear.

Firstly, we give a sufficient condition on the local error for the global convergence using similar techniques as in [4] and [8]. Then we show when a numerical ODE solver satisfies this condition. The proof of Theorem 4.1 applies the following version of Gronwall's Lemma.

**Lemma 4.1.** *Let  $s, t \in \mathbb{R}$ ,  $s < t$  and suppose that  $b, c$  and  $r$  are  $\mathbb{R} \rightarrow \mathbb{R}$  functions such that  $b$  is continuous,  $c$  is continuously differentiable and  $r$  is piece-wise continuous on  $[s, t]$ , furthermore*

$$r(u) \leq c(u) + \int_s^u b(v)r(v)dv$$

Then

$$r(u) \leq c(u) + \int_s^u b(v)c(v)e^{[\int_v^u b]}dv$$

**Theorem 4.1. (Bound on the global error)** *Let's suppose that a discretization scheme  $\widehat{X}_{t_0}, \dots, \widehat{X}_{t_k}$  approximating the solution to the SDE (10) and corresponding to the partition  $0 = t_0 < t_1 < \dots < t_k = T$  with longest time step  $\Delta t$  has local error  $E_i$   $i = 1, \dots, k$  as in Definition 4.1, such that*

$$\mathbb{E} \left[ \sup_{1 \leq l \leq k} \left\| \sum_{i=1}^l E_i \right\|^2 \right]^{1/2} < C_1 \Delta t^{m/2} \quad (31)$$

for some positive constant  $C_1$ . Furthermore suppose that the coefficient function  $V_0$  is continuously differentiable with bounded gradient and the functions  $V_1, \dots, V_d$  are twice continuously differentiable with bounded first and second derivatives.

Then there exists a constant  $C_2$  not depending on  $\Delta t$  such that

$$\mathbb{E} \left[ \|\xi_T - \widehat{X}_T\|^2 \right]^{1/2} < C_2 \Delta t^{m/2}.$$

*Proof.* Given the partition  $0 = t_0 < t_1 < \dots < t_k = T$ , for each  $1 \leq i \leq k$  we consider the process  $Y_t^i$  introduced in Definition 4.1. Note that  $Y_{t_i}^i$  can be equally represented as

$$\begin{aligned} Y_{t_i}^i &= \widehat{X}_{t_{i-1}} + \sum_{j=0}^d \int_{t_{i-1}}^{t_i} V_j(Y_t^i) \circ dB_t^j \\ Y_{t_i}^i &= \widehat{X}_{t_i} + E_i \end{aligned} \quad (32)$$

We also define

$$Y_t := Y_t^i \text{ for } t \in [t_{i-1}, t_i).$$

By (32)

$$\begin{aligned} \mathbb{E} \left[ \|\xi_t - Y_t\|^2 \right]^{1/2} &\leq \mathbb{E} \left[ \left\| \sum_{i=0}^d \int_0^t V_i(\xi_s) - V_i(Y_s) \circ dB_s^i \right\|^2 \right]^{1/2} \\ &\quad + \mathbb{E} \left[ \sup_{\{l|t \geq t_{l-1}\}} \left\| \sum_{j=1}^l E_j \right\|^2 \right]^{1/2} \end{aligned} \quad (33)$$

Let's introduce the notation

$$r(t) := \mathbb{E} \left[ \|\xi_t - Y_t\|^2 \right]^{1/2}$$

By Lemma 2.2, Lemma 2.4 and the boundedness assumption on the coefficient functions and their derivatives, there exists a constant  $K$  such that

$$\mathbb{E} \left[ \left\| \sum_{i=0}^d \int_0^t V_i(\xi_s) - V_i(Y_s) \circ dB_s^i \right\|^2 \right]^{1/2} \leq K \int_0^t r(u) du.$$

Then applying Gronwall's lemma on the functions  $r, b \equiv K$  and

$$c(t) := \mathbb{E} \left[ \sup_{\{l|t \geq t_{l-1}\}} \left\| \sum_{j=1}^l E_j \right\|^2 \right]^{1/2}$$

we get

$$\mathbb{E} \left[ \|\xi_T - Y_T\|^2 \right]^{1/2} \leq c(T) e^{TK}$$

Using the inequality (31) and defining  $C_2 := C_1 e^{TK}$  the assertion is proved.  $\square$

**Corollary 4.1.** *Let the sequence of local errors  $E_i, i = 1, \dots, k$  corresponding to the discretisation scheme  $\hat{X}_{t_0}, \dots, \hat{X}_{t_k}$  be given as in Theorem 4.1. Let's suppose that for  $i = 1, \dots, k, E_i = M_i + N_i$  satisfying*

$$(i) \quad \mathbb{E} \left[ \|M_i\|^2 \right]^{1/2} \leq C_1 \Delta t^{(m+1)/2} \text{ and } \mathbb{E} [M_i] = 0$$

$$(ii) \quad \mathbb{E} \left[ \|N_i\|^2 \right]^{1/2} \leq C_1 \Delta t^{(m+2)/2}$$

for some constant  $C_1$ . Furthermore, assume that the coefficient functions satisfy the conditions of Theorem 4.1.

Then there exists a constant  $C_2$  not depending on  $\Delta t$  such that

$$\mathbb{E} \left[ \|\xi_T - \hat{X}_T\|^2 \right]^{1/2} < C_2 \Delta t^{m/2}$$



*Proof.* We only need to prove that  $E_i, i = 1, \dots, k$  satisfy the condition (31) of Theorem 4.1. Note that

$$\left\| \sum_{i=1}^l E_i \right\|^2 \leq 2 \left\| \sum_{i=1}^l M_i \right\|^2 + 2 \left\| \sum_{i=1}^l N_i \right\|^2 \quad (34)$$

Given the assumptions on  $M_i$ , the first term on the right hand side of (34) is a discrete time sub-martingale indexed by  $l$ . Then using the conditions of the Corollary and applying Doob's inequality the assertion is proved.  $\square$

**Corollary 4.2.** *Let's assume that the coefficient functions  $V_0, \dots, V_d$  satisfy the conditions of Theorem 3.1 for a positive even integer  $m \geq 2$ . Let's suppose that a random sequence  $\widehat{X}_{t_0}, \dots, \widehat{X}_{t_k}$  with longest step size  $\Delta t$  is the result of an ODE-approach based scheme derived from the  $m$ -truncated log-signature implemented with a numerical ODE solver satisfying*

$$\begin{aligned} E_i &= \Gamma [(\pi_{m+1} - \pi_m) L_{t_i, t_{i+1}}] (\xi_0) + R_i \\ \mathbb{E} [\|R_i\|^2]^{1/2} &\leq C_1 \Delta t^{(m+2)/2} \end{aligned}$$

for some constant  $C_1$ .

Then there exists a constant  $C_2$  such that

$$\mathbb{E} [\|\xi_T - \widehat{X}_T\|^2]^{1/2} < C_2 \Delta t^{m/2}.$$

*Proof.* Since  $m$  is even, Remark 2.1 and Lemma 2.6 imply that defining

$$M_i := \Gamma [(\pi_{m+1} - \pi_m) L_{t_i, t_{i+1}}] (\xi_0)$$

and  $N_i = R_i$ , the conditions of Corollary 4.1 are satisfied and so the assertion is proved.  $\square$

**Example 4.1.** For any even integer  $m$  the discretization schemes defined by the  $m$ -truncated stochastic Taylor expansion satisfy the conditions of Corollary 4.2 and so they have a global order  $m/2$ .

Recalling Remark 2.1, the Itô form of the Stratonovich iterated integrals corresponding to the multi-index  $J$  has an all-zero multi-index corresponding integral if  $J = J_1 \circ \dots \circ J_k$  such that either  $J_i = (0)$  or  $J_i = (l_i, l_i)$  for all  $i = 1, \dots, k$  and for some  $1 \leq l_i \leq d$ . This implies that for an odd integer  $m$ , there are all-zero multi-index corresponding integrals in the Itô form of  $\Gamma [(\pi_{m+1} - \pi_m) L_{t_i, t_{i+1}}] (\xi_{t_i})$  and hence the conditions of Corollary 4.2 are not necessarily satisfied. To ensure the convergence of the  $m = 1$ -truncated log-signature based scheme we introduce the extended scheme as follows.

**Definition 4.2.** *Given the time interval  $[s, t]$  we introduce the random vector field*

$$W := \Gamma [B_{s,t}^0 \varepsilon_0 + \pi_1 L_{s,t}]$$

*which will be referred to as the vector field corresponding to the extended ODE approach based 1-truncated scheme.*

Note that the Itô form of  $\Gamma [\pi_2 L_{s,t} - B_{s,t}^0 \varepsilon_0 - \pi_1 L_{s,t}] (\zeta_s)$  only includes stochastic integrals and so the conditions of Corollary 4.2 are satisfied.

**Example 4.2.** Let's consider the extended ODE approach based 1-truncated scheme given by the vector field

$$W(x) = B^0 V_0(x) + \sum_{i=1}^d B^i V_i(x)$$

If the ODE is

- (i) solved directly (if the exact solution is known)
- (ii) or approximated with the predictor-corrector scheme
- (iii) or approximated with the two step Taylor expansion based ODE solver

By Corollary 4.2, this scheme has a global order 1 if  $d = 1$  and global order  $1/2$  if  $d \geq 2$ . Note that by [2], no other scheme based on only the increment of the driving noise has higher rate of convergence than the the extended ODE approach based 1-truncated scheme.

**Remark 4.1.** Note that by Corollary 4.2, the convergence of the extended ODE approach based 1-truncated scheme is due to the fact that the terms in  $(\pi_2 - \pi_1) L_{s,t} - B_{s,t}^0 \varepsilon_0$  have zero expectation (in other words the Lévy area of the Brownian motion has zero expectation). For a general driving path  $Y_t \in \mathbb{R}^d$  of finite  $2 + \varepsilon$ -variation ( $0 < \varepsilon < 1$ ) in the sense of [13] and [14], the 1-truncated scheme would not converge to the solution of the SDE. The use of the (at least) 2-truncated scheme would guarantee the convergence, but similar arguments as in Corollary 4.2 show that the scheme adjusted with the expected area, i.e. the ODE approach based on the vector field

$$W(x) := Y^0 V_0(x) + \sum_{i=1}^d Y^i V_i(x) + \frac{1}{2} \sum_{i < j} \mathbb{E} [Y^{(i,j)} - Y^{(j,i)}] [V_i, V_j](x)$$

is sufficient for the convergence.

**Remark 4.2.** Note that the ODE solver schemes in Example 4.2 explicitly or implicitly include the terms

$$\widehat{X}_{t_i} B_{t_i, t_{i+1}}^{(jj)} = \frac{1}{2} \widehat{X}_{t_i} \left( B_{t_i, t_{i+1}}^j \right)^2$$

for  $j = 1, \dots, d$ . On the global level, these terms form a non-vanishing contribution to the error, tending to the cross variation process i.e. the difference between the solution to SDE (10) and the solution to the SDE defined by the same coefficient functions but with Itô integrals. Therefore, the one step Taylor expansion (i.e. the deterministic Euler scheme) based numerical solver is convergent, however it converges to the solution of the SDE given by Itô integrals.

## 5 High order schemes

The ODE approach based on the vector field  $W = \sum_{i=0}^d B^i V_i$  has global order 1 if  $d = 1$  and  $1/2$  otherwise (see Example 4.2). To construct higher order schemes, the simulation of higher order iterated integrals is required. In this section we show how the iterated integrals can be constructed from piece-wise linear Brownian paths.

### 5.1 Piece-wise linear Brownian paths

We start our analysis with the tensor algebraic representation of linear paths.

**Lemma 5.1.** *Let's suppose that a path  $\omega_t : [0, T] \rightarrow \mathbb{R} \oplus \mathbb{R}^d$  is linear, i.e.*

$$\omega_t = (t, z_1 t, \dots, z_d t)^T.$$

*Then the log-signature of  $\omega$  is*

$$\log(\mathbf{S}_{0,T}(\omega)) = T \left( \varepsilon_0 + \sum_{i=1}^d z_i \varepsilon_i \right).$$

The multiplicative property of the signature (see Lemma 3.2) and Lemma 5.1 imply the following lemma.

**Lemma 5.2.** *Given a piece-wise linear interpolation  $B_t^{\mathcal{P}}$  of a Brownian motion corresponding to the partition  $\mathcal{P} = (0, t_1, \dots, t_{k-1}, T)$ , i.e. the concatenation of paths of the form*

$$\begin{aligned} \omega_0^i &= 0, \\ \omega_t^i &= \left( t, \frac{B_{t_{i-1}, t_i}^1}{B_{t_{i-1}, t_i}^0} t, \dots, \frac{B_{t_{i-1}, t_i}^d}{B_{t_{i-1}, t_i}^0} t \right)^T, \quad t \in (0, t_i - t_{i-1}] \end{aligned}$$

*then*

$$L_{0,T}^{\mathcal{P}} := \log(\mathbf{S}_{0,T}(\circ B)) = \log \left[ \exp \left( \sum_{j=0}^d B_{0,t_1}^j \varepsilon_j \right) \otimes \dots \otimes \exp \left( \sum_{j=0}^d B_{t_{k-1}, T}^j \varepsilon_j \right) \right] \quad (35)$$

If the partition  $\mathcal{P}$  contains  $k$  subintervals of equal length, we will use the notation  $\mathcal{P}_k$ .

### 5.2 High order approximations based on piece-wise linear paths

In the previous sections we explained how one can construct numerical schemes based on  $L_{0,t}$ . In general we have a small error with these schemes, and if we could replace  $L_{0,t}$  with another random Lie series which is in some sense close enough to  $L_{0,t}$ , then it is possible that the ODE scheme based on the replacement might also

be a good approximation to the exact solution for the SDE. We prove that  $L_{0,t}^{\mathcal{P}_k}$  for a properly chosen  $k$  is a good choice of replacement.

Let's denote the  $i$ th coordinate of  $B_t^{\mathcal{P}_k}$  by  $B_t^{i,\mathcal{P}_k}$ . The following random ODE is a.s. well defined

$$d\hat{\xi}_t = \sum_{i=1}^d V_i(\hat{\xi}_t) dB_t^{i,\mathcal{P}_k} \quad (36)$$

Since the equation (36) is a.s. driven by a path of finite length, we have a differentiation rule analogous to the Itô lemma

$$df(\hat{\xi}_t) = \sum_{i=1}^d V_i f(\hat{\xi}_t) dB_t^{i,\mathcal{P}_k} \quad (37)$$

for any smooth function  $f$ .

We introduce the notation with analogy to section 2

$$B_{s,t}^{J,\mathcal{P}_l} := \int_{s < t_1 < \dots < t_k < t} \circ dB_{t_1}^{j_1,\mathcal{P}_l} \circ \dots \circ dB_{t_k}^{j_k,\mathcal{P}_l} \quad (38)$$

$$B_{s,t}^{J,\mathcal{P}_l}(Y) := \int_{s < t_1 < \dots < t_k < t} Y_{t_1} \circ dB_{t_1}^{j_1,\mathcal{P}_l} \circ \dots \circ dB_{t_k}^{j_k,\mathcal{P}_l} \quad (39)$$

for a continuous function  $Y$ .

Given the differentiation rule (37) one can derive the analogous version of the log-signature theorem.

**Lemma 5.3.** *Let's suppose that  $\hat{\xi}_t$  is (a.s.) the unique solution for the random ODE (36) and  $\hat{y}_t$  is the solution to the following random ODE*

$$\left. \begin{aligned} d\hat{y}_t &= \widehat{W}(\hat{y}_t) dt \\ \hat{y}_0 &= \hat{\xi}_0 \end{aligned} \right\} \quad (40)$$

where  $\widehat{W} = \Gamma \left[ \pi_m L_{0,t}^{\mathcal{P}_k} \right]$ .

Then

$$\hat{\xi}_t - \hat{y}_1 = \Gamma \left[ (\pi_{m+1} - \pi_m) L_{0,t}^{\mathcal{P}_k} \right] (\hat{\xi}_0) + \bar{R}$$

and for  $0 < t < 1$  there exists a constant  $\bar{C}$  not depending on  $t$  such that

$$\mathbb{E} \left[ \|\bar{R}\|^2 \right]^{1/2} < \bar{C} t^{(m+2)/2}$$

**Remark 5.1.** The above lemma can be regarded as the probabilistic extension of [21]. From [21] one can derive an explicit formula for  $\Gamma \left[ \pi_m L_{0,t}^{\mathcal{P}_k} \right]$ .

By Example 4.2,  $\hat{\xi}_t$  is also the extended ODE approach based  $m = 1$ -truncated approximation of the solution for (10) on  $[0, t]$  corresponding to the partition of  $k$  subintervals of equal length. Hence, we link Lemma 5.3 with Theorems 3.1 and 4.1 as follows.

**Theorem 5.1.** *Let's suppose that the conditions of Theorem 3.1 are satisfied and  $0 < t < 1$ . Then there exists a positive integer  $k$  such that*

$$\xi_t - \hat{y}_1 = \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}](\xi_0) + R \quad (41)$$

where  $\hat{y}_t$  is the solution to the random ODE defined in Lemma 5.3 and corresponding to the partition  $\mathcal{P}_k$ , and furthermore, there exists a constant  $C$  not depending on  $t$ , such that

$$\mathbb{E} \left[ \|R\|^2 \right]^{1/2} \leq Ct^{(m+2)/2}.$$

*Proof.* We split  $\xi_t - \hat{y}_1$  into two terms as follows

$$\xi_t - \hat{y}_1 = (\xi_t - \hat{\xi}_t) + (\hat{\xi}_t - \hat{y}_1) \quad (42)$$

and regard the two terms separately.

Given that  $\hat{\xi}_t$  is an extended ODE approach based  $m = 1$ -truncated approximation of  $\xi_t$  corresponding to the partition  $\mathcal{P}_k$ , by Example 4.2, there exists a constant  $C_1$  such that

$$\mathbb{E} \left[ \|\xi_t - \hat{\xi}_t\|^2 \right]^{1/2} \leq C_1 \left[ \frac{t}{k} \right]^\gamma \quad (43)$$

where  $\gamma = 1$  if  $d = 1$  and  $\gamma = 1/2$  otherwise.

Lemma 5.3, equation (42) imply the following representation

$$\begin{aligned} \xi_t - \hat{y}_1 &= \xi_t - \hat{\xi}_t + \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}](\hat{\xi}_0) \\ &\quad + \left[ \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}^{\mathcal{P}_k}](\xi_0) - \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}](\hat{\xi}_0) \right] + \bar{R} \end{aligned}$$

where  $\bar{R}$  is defined in Lemma 5.3.

Define

$$R := \xi_t - \hat{\xi}_t + \left[ \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}^{\mathcal{P}_k}](\xi_0) - \Gamma[(\pi_{m+1} - \pi_m)L_{0,t}](\hat{\xi}_0) \right] + \bar{R}$$

Let's rewrite  $\Gamma[(\pi_{m+1} - \pi_m)L_{0,t}^{\mathcal{P}_k}](\hat{\xi}_0)$  as follows

$$\begin{aligned} &\Gamma[(\pi_{m+1} - \pi_m)L_{0,t}^{\mathcal{P}_k}](\hat{\xi}_0) = \\ &= \sum_{\substack{J \in \mathcal{A} \\ \|J\|=m+1}} \left( \sum_{\substack{J_1 \circ \dots \circ J_l = J \\ l \leq \|J\|}} a_{J_1, \dots, J_l} B_{0,t}^{J_1, \mathcal{P}_k} \dots B_{0,t}^{J_l, \mathcal{P}_k} \right) \Gamma(\varepsilon_J)(\hat{\xi}_0) \quad (44) \end{aligned}$$

with deterministic real coefficients  $a_{J_1, \dots, J_l}$  depending only on the multi-indices  $J_1, \dots, J_l$ .

Since  $B_{0,t}^{J, \mathcal{P}_k}$  is the lowest order ODE approach based approximation of  $B_{0,t}^J$  for each multi-index  $J$ , therefore there exists a constant  $C$  not depending on  $t$  or  $k$  such that

$$\mathbb{E} \left[ \|R\|^2 \right]^{1/2} \leq \mathbb{E} \left[ \|\bar{R}\|^2 \right]^{1/2} + C \left[ \frac{t}{k} \right]^\gamma \quad (45)$$

The proof is complete by choosing a high enough  $k$  satisfying

$$C \left[ \frac{t}{k} \right]^\gamma \leq \bar{C} t^{(m+2)/2} \quad (46)$$

where  $\bar{C}$  is the constant from Lemma 5.3.  $\square$

Theorem 5.1 shows, that for high enough  $k$ ,  $\hat{y}_1$  and  $y_1$  (corresponding to the  $m$ -truncated log signature) are numerically equivalent (local) approximations of  $\xi_t$ . However, the left-hand side of (46) should not be much smaller than the right-hand side, because high values of  $k$  makes the computation of  $\Gamma \left[ \pi_m L_{0,t}^{\mathcal{P}_k} \right]$  less tractable, i.e.

$$k \approx \left[ \frac{\bar{C}}{C} t^{(m+2)/2-\gamma} \right]^{-1/\gamma} \quad (47)$$

is the recommended order of magnitude for  $k$ .

### 5.3 Practical considerations

The implementation of the high order schemes based on piece-wise linear approximation of the Brownian paths on a fine sub-scale requires the repeated computation of  $\Gamma \left[ \pi_m L_{0,t}^{\mathcal{P}_k} \right]$ , where  $L_{0,t}^{\mathcal{P}_k}$  is given by (35). To work out the referred formula, one could use the Campbell-Baker-Hausdorff formula at each discretization step, but this would not be efficient.

Furthermore, the principle given by (46) and (47) implies that shorter time intervals require higher values of  $k$  to ensure the high order convergence of our scheme. However, for higher values of  $k$  the computation of (35) increases rapidly and one would lose the linear growth of computational expense.

**Lemma 5.4.** *The log-signature (35) written in a Lie basis  $\{\ell_i, i \in \mathbb{N}\}$  is of the form*

$$\pi_m L_{0,t}^{\mathcal{P}_k} = \sum_{\substack{i, \\ \pi_m \ell_i = \ell_i}} p_i \ell_i \quad (48)$$

where each  $p_i$  is a polynomial of the variables

$$\{B_{t_{i-1}, t_i}^j \mid j = 0, \dots, d, i = 1, \dots, k\}.$$

Note that the structure of (48) does not change if  $k$  is increased but the polynomials  $p_i$  have more variables. Hence to preserve efficiency, one can pre-compute either this formula or the polynomial coefficients therein for a number of independent Brownian paths. The pre-computation might take some time, however it only has to be done once. After all, the reuse of the pre-computed values results in a fast and high order numerical algorithm.

Since the formula (48) does not depend on the SDE but the dimension and the choice of  $k$ , one could create a universal database of pre-computed coefficients usable for many SDEs.

## 5.4 A note on the computational expense

The implementation of the high order scheme presented in this section is after all a high order approximation of a lowest order scheme corresponding to a discretization on a finer sub-scale. One might ask if one can ever benefit from working out the high order scheme instead of just solving the lowest order scheme on the fine scale. Let's do a brief cost-analysis to answer the question.

Let's assume that the computational cost of evaluating one step corresponding to the lowest order scheme is  $E_1$  and the cost of evaluating one step of an  $m$ -truncated ODE based scheme is  $E_2$ . In general  $E_1$  is much smaller than  $E_2$ . If we apply the lowest order scheme on an interval with a  $k$ -substep fine scale, then the cost is  $kE_1$ . As long as  $kE_1$  is smaller than  $E_2$ , the lowest order scheme is recommended. However, if one needs a more accurate approximation and wants to choose a finer scale, then  $k$  will increase according to (47) and eventually the implementation of the higher order scheme will become more efficient than the corresponding low order scheme.

## 6 Numerical examples

In this section, some numerical examples for the ODE approach are presented. We demonstrate the efficiency of different ODE solvers and the high order methods. The primary aim is to estimate and compare the order of strong convergence, however we also present some weak approximation results.

### 6.1 Estimating the order of convergence

For the estimation of the convergence order, we apply some ideas from [20]. Let's review some basic results.

**Definition 6.1.** *Let's regard a discretization method generating  $\hat{X}_T$  as an approximation of  $\xi_T$ . If the discretization scheme is based on partitioning  $[0, T]$  into subintervals each of length  $\Delta t$ , then this particular case, the resulting approximation is denoted by  $\hat{X}_T(\Delta t)$ .*

**Lemma 6.1.** *Let's suppose that a discretization method  $\hat{X}_T$  approximates  $\xi_T$ , i.e.*

$$\mathbb{E} \left[ \left\| \xi_T - \hat{X}_T(\Delta t) \right\|^2 \right]^{1/2} \rightarrow 0 \text{ as } \Delta t \rightarrow 0$$

Furthermore, let's suppose the existence of constants  $C_2$  and  $\gamma$  not depending on  $\Delta t$  such that for  $0 < \Delta t < 1$

$$\mathbb{E} \left[ \left\| \hat{X}_T(\Delta t/2) - \hat{X}_T(\Delta t) \right\|^2 \right]^{1/2} \leq C_2 \Delta t^\gamma \quad (49)$$

is satisfied. Then for  $0 < \Delta t < 1$

$$\mathbb{E} \left[ \left\| \xi_T - \hat{X}_T(\Delta t) \right\|^2 \right]^{1/2} \leq C_2 \Delta t^\gamma \frac{1}{1 - (1/2)^\gamma} \quad (50)$$

The proof of the lemma is based on the triangle inequality.

We apply Lemma 6.1 to estimate the order of convergence  $\gamma$  and an upper bound on the strong approximation error as follows. Estimating the expected value on the left hand side of (49) for a sequence of step lengths  $T/2^k$  for  $k = l, l+1, \dots$  where  $T/2^l < 1$  and fitting a log-regression, one can estimate  $C_2$  and  $\gamma$ . If  $C_2$  is estimated, then by (50) we get an estimation for an upper bound of the error at step size  $\Delta t$ . When  $\gamma$  is known, we only estimate  $C_2$ .

As it was demonstrated in [20], one can observe the expected value on the left hand side of (49) very accurately, if the schemes corresponding to the different step sizes  $\Delta t$  and  $\Delta t/2$  are run on the same Brownian paths. In that case, the random variables  $\hat{X}_T(\Delta t)$  and  $\hat{X}_T(\Delta t/2)$  are correlated and a simple Monte Carlo method results in a low variance and unbiased estimation of the expected value of their distance. Running schemes corresponding to different step lengths on the same paths can be achieved by using Lévy's construction of the Brownian paths (see e.g. [4] for details).

The idea of running schemes on the same Brownian paths is also useful when comparing two different discretization methods generating  $\hat{X}_T$  and  $\bar{X}_T$  respectively as approximations of  $\xi_T$ . The high correlation between  $\hat{X}_T$  and  $\bar{X}_T$  results a low variance for the Monte Carlo estimation of

$$\mathbb{E} \left[ \left\| \bar{X}_T - \hat{X}_T \right\|^2 \right]^{1/2} \quad (51)$$

This comparison can be further applied as the triangle inequality implies.

**Lemma 6.2.** *Let two discretization methods producing  $\hat{X}_T$  and  $\bar{X}_T$  respectively as approximations of  $\xi_T$  be given, Let's suppose that there exists a constant  $C$  such that*

$$\mathbb{E} \left[ \left\| \xi_T - \bar{X}_T \right\|^2 \right]^{1/2} < C$$

then

$$\mathbb{E} \left[ \left\| \xi_T - \hat{X}_T \right\|^2 \right]^{1/2} \geq \mathbb{E} \left[ \left\| \hat{X}_T - \bar{X}_T \right\|^2 \right]^{1/2} - C \quad (52)$$

Lemma 6.2 implies, that if one can estimate a small enough upper bound  $C$  on the global error of  $\bar{X}_T$ , then (52) gives an accurate estimation for the lower bound on the error of  $\hat{X}_T$ .

In the following sections, for different SDEs and discretization methods we ran simple Monte Carlo simulation based estimations of the following quantities

- (i) the estimated upper bound of the global error, i.e. (50)
- (ii) the  $L^2$  distance of different schemes, i.e. (51)
- (iii) in some cases the estimated lower bound on the global error, i.e. (52)
- (iv) and  $\mathbb{E} \left[ \hat{X}_T(\Delta T) \right]$



for different step sizes. Each method was run on  $10^6$  paths.

Since each Monte Carlo simulation is based on sampling, the resulted estimation is a random variable with positive variance. Using this variance, when estimating (iv), we fit a 99% confidence interval centered at the resulting realization of the random variable. The length of these confidence intervals is proportional to the square root of the number of runs. In case of (i), (ii) and (iii), the calculated confidence intervals were very small.

## 6.2 First order approximation of the CIR process

Firstly, we regarded simple SDEs, driven by one dimensional Brownian motion and tested the first order ODE approach with different ODE solvers as well as comparing it to the Euler-Maruyama scheme. The test results, presented in Figures 1 and 2, are the weak and strong approximation results respectively for the CIR process, i.e. in the Itô form

$$dr_t = a(b - r)dt + \sigma\sqrt{r}dB \quad (53)$$

where  $a, b$  and  $\sigma$  are positive constants satisfying  $ab/2 > \sigma^2$ , which ensures that  $r_t$  is a.s. positive.

We implemented the ODE approach with three different ODE solvers, namely the predictor-corrector, Runge Kutta order 4 and splitting. By splitting we mean the ODE solver in which  $W$  is written as  $W := W_1 + W_2$  where

$$\begin{aligned} W_1(x) &:= B^0 \left[ a(b - x) - \sigma^2/4 \right] \\ W_2(x) &:= B^1 \sigma \sqrt{x} \end{aligned}$$

and at each step (starting at  $\hat{X}_{t_i}$ ), three ODEs are solved as follows

$$\begin{aligned} x_1 &:= \text{Exp} \left[ \frac{1}{2} W_1 \right] (\hat{X}_{t_i}) \\ x_2 &:= \text{Exp} [W_2] (x_1) \\ \hat{X}_{t_{i+1}} &:= \text{Exp} \left[ \frac{1}{2} W_1 \right] (x_2) \end{aligned}$$

In case of the CIR SDE, the exact solution for the ODEs appearing in the splitting is known, and the computation is very fast. The splitting method is recommended in for example [18] and some nice weak approximation properties of it are presented as well.

The first order ODE approach implemented with the predictor-corrector ODE solver is referred to as the Heun scheme.

**Remark 6.1.** Note, that despite the guaranteed positivity of  $r_t$ , all of the tested schemes except the Splitting version can result in negative interest rates. In case of the ODE approach based method, the exact solution to the derived random

ODEs preserves the positivity, but their numerical approximations might not. One possible way to overcome this difficulty, is to adaptively reduce the ODE numerical solver's step size when the solver results in negative interest rates. In case of the splitting, the chosen combination of parameters guarantees positive solutions and no extra care is required. The Euler-Maruyama method is fixed by taking a positive value of the resulting interest rate at each step. We refer the reader to [12] for a review of simulation schemes approximating the CIR process.

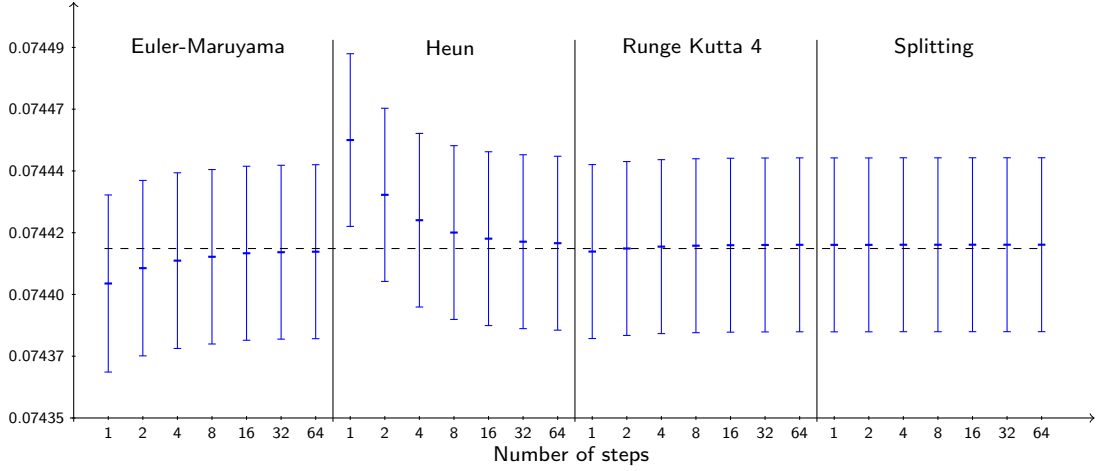


Figure 1: Weak approximation results (CIR)

In Figure 1, the weak approximation results, i.e. the confidence intervals corresponding to Monte Carlo estimations for  $\mathbb{E}[\hat{X}_T]$ , are presented for different schemes implemented with different numbers of steps. The horizontal dashed line is the exact value of  $\mathbb{E}[\xi_T]$ . Given the number of runs ( $10^6$ ) the variance of the evaluated estimation was relatively large, and at the weak level one cannot make a difference between the Runge Kutta 4 and the splitting versions. The weak error of the Euler method at 64 steps seems smaller than the weak error of the Runge Kutta 4, but the calculated values are samples from random variables with positive variance. Note that the Euler-Maruyama method results in a first order weak approximation, i.e. at the weak level there is no difference in the convergence order of the tested methods.

In Figure 2, the accuracy of the path-wise approximations is presented. The graph on the left-hand side presents the estimated upper bound on the error based on (50). In case of the Euler method, we could apply (52) and a lower bound on the error is estimated. This lower bound demonstrates, that despite the nice weak approximation properties, the Euler-Maruyama scheme is less accurate in the path-wise approximations when compared to the other schemes. The estimated order of convergence (i.e. the slope of the curves) are as expected. In case of the Euler-

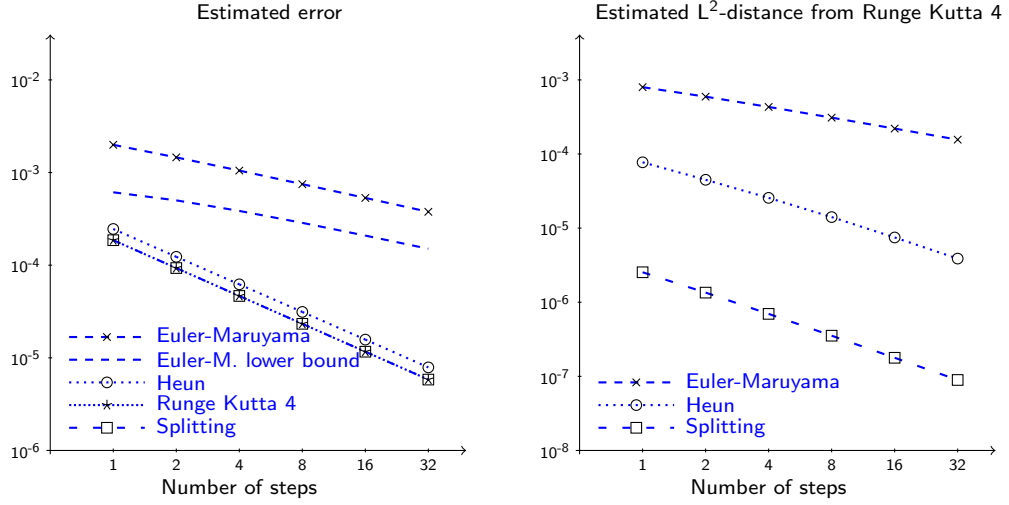


Figure 2: Estimated error, comparison of solvers (CIR)

Maruyama scheme it's close to 1/2 and in case of the ODE approach based methods it is close to 1.

The graph on the right-hand side of Figure 2 compares each method with the Runge Kutta 4 version, estimating the  $L^2$ -distance based on (51). Note that the  $L^2$  distance of the splitting version from the Runge Kutta 4 based method has an order of magnitude  $10^{-6}$ - $10^{-7}$ , whereas the estimated global error of both methods has order  $10^{-5}$  and so numerically the two versions are equivalent. However the splitting is a bit faster than the Heun scheme and more than twice as faster than Runge Kutta 4, so in this particular case the Splitting version is recommended.

Each ODE solver tested here can be easily extended to handle the term  $\int_0^t r_u du$ . The simulation of this term is required when pricing or hedging bonds and derivatives of bonds. In case of the Splitting, the exact solution to the new ODEs appearing in this extension are known.

### 6.3 Second order approximation implemented

In this section we present some numerical results of some tests run with the second order ODE approach based method. The SDE chosen here is less meaningful in finance. However due to its' nice properties it is proved to be a useful test case.

The SDE is as follows

$$\begin{aligned} dx_t^1 &= \sin(x_t^1) \circ dB^1 \\ dx_t^2 &= \sin(x_t^1) dt \end{aligned}$$

The left-hand side of Figure 3 presents the weak approximation results in the form of confidence intervals. According to the rule (46), for the 1-step second order

version we used a linear interpolation of the Brownian motion on a two step sub-scale, i.e.  $k = 2$ . In case of the 2-step version  $k = 4$  was chosen whereas the 4-step version was run with  $k = 8$ .

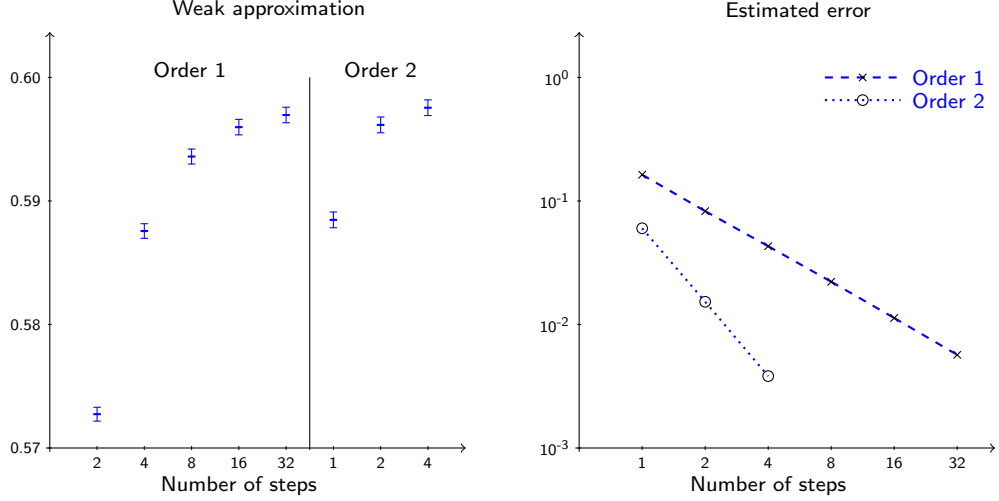


Figure 3: Implementing high order approximations

The estimated upper bound on the strong global error is presented on the right-hand side of Figure 3. The second order method, implemented as described in section 5 has an estimated order close to 2.

We analysed the computational expense for this particular SDE as described in section 5.4. In this case  $E_2/E_1 \approx 3/2$ , so the second order scheme is computationally more efficient than the lowest order at every step. For higher order schemes and in higher dimensions, the lower step high order versions might be relatively less efficient.

## 7 Conclusions

The approach presented by this paper turned out to be very useful in the case of strong approximations as well as in the case of weak approximations. The ideas presented here can be applied to construct weak approximations under reasonably general conditions (see [9], [10] and [11]). The concept of Cubature on Wiener space ([15]) led to the construction of finitely supported measures on the truncated Lie algebra converging to the exact distribution of the Brownian log-signature and resulting in high order weak approximations. The application of these ideas in computational finance as well as in variance reduction is described in [6].

The theorems of this paper ensuring the global accuracy are based on global and in some cases rather strong conditions. The focus of a further work is to ex-

plore how these conditions could be weakened and how one could replace the global conditions with local ones.

## References

- [1] Gérard Ben Arous, *Flots et séries de Taylor stochastiques*, Probab. Theory Related Fields, 81(1):29-77, 1989.
- [2] Clark, J.M.C & Cameron, R.J., *The maximum rate of convergence of discrete approximations for stochastic differential equations*, Lecture Notes in Control and Information Sciences, Springer Berlin/Heidelberg, Volume 25, 162-171, 1980
- [3] Castell, F. *Asymptotic expansion of stochastic flows*, Probability Theory and Related Fields, Springer-Verlag, 1993
- [4] Gaines, J. & Lyons, T. *Variable step size control in the numerical solution of stochastic differential equations*, SIAM J. Appl. Math 57, 1997
- [5] Gard, T. *Introduction to Stochastic Differential Equations*, Marcel Dekker Ltd, 1988
- [6] Gyurkó, L.G, Litterer, C., Lyons, T.J. *Monte Carlo methods based on high order cubature formulae on the Wiener space*, in progress, 2007
- [7] Hu, Y. *Calculs formels sur les e.d.s. de Stratonovich*, Séminaire de Probabilités XXIV 1988-89, 1990
- [8] Kloeden, P.E. & Platen, E. *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag Berlin Heidelberg New York, 1999
- [9] Kusuoka, S. *Approximation of expectation of diffusion process and mathematical finance*, Adv. Stud. Pure Math., 31, 147-165, 2001
- [10] Kusuoka, S. *Approximation of expectation of diffusion process based on Lie algebra and Malliavin calculus*, preprint 2003,  
<http://kyokan.ms.u-tokyo.ac.jp/users/preprint/pdf/2003-34.pdf>
- [11] Litterer, C. *Error bounds for Cubature on Wiener space under relaxed assumption on the drift*, preprint 2007
- [12] Lord, R., Koekkoek, R. & van Dijk, D. *A Comparison of Biased Simulation Schemes for Stochastic Volatility Models*, Tinbergen Institute Discussion Paper, 2006  
<http://www.tinbergen.nl/discussionpapers/06046.pdf>
- [13] Lyons, T.J. 1998 *Differential equations driven by rough signals*, Revista Matematica Iber. Vol 14, Nr. 2, 215-310

- [14] Lyons, T.J. & Qian, Z. 2002 *System Control and Rough Paths*, Oxford mathematical monographs, Clarendon Press, Oxford
- [15] Lyons, T.J. & Victoir, N. 2004 *Cubature on Wiener Space*, Proc. R. Soc. Lond. A 460, 169-198
- [16] Magnus, W. *On the exponential solution of differential equations for a linear operator*, Commun. Pure Appl. Math. 7, 1954
- [17] Newton, N., *An asymptotically efficient difference formula for solving stochastic differential equations*, Stochastics An International Journal of Probability and Stochastic Processes, Volume 19, Issue 3, 175-206, 1986
- [18] Ninomiya, S. & Victoir, N. *Weak approximation of stochastic differential equations and application to derivative pricing*, preprint 2006  
[http://arxiv.org/PS\\_cache/math/pdf/0605/0605361v3.pdf](http://arxiv.org/PS_cache/math/pdf/0605/0605361v3.pdf)
- [19] Reutenauer, C. *Free Lie algebras*, London Mathematical Society Monographs, vol. 7. Oxford Science Publications, 1993
- [20] Schmitz Abe, K. & Shaw, W.T. *Measure order of convergence without an exact solution, Euler vs. Milstein scheme*, International Journal of Pure and Applied Mathematics, Volume 24, No 3, 2005
- [21] Strichartz, R.S. *The Campbell-Baker-Hausdorff-Dynkin Formula and Solutions of Differential Equations*, Journal of Functional Analysis 72, 320-345, 1987