A Runge-Kutta Type Scheme to Solve $dX_t = \sigma(X_t) \circ dW_t$ Under Commutative Noise *

Luis Jose Roman†

November, 1999

Abstract

First, we give a finite-difference scheme of global order h^2 in the mean-square sense to solve numerically certain stochastic differential equations (SDEs), which is inspired on Runge-Kutta methods of order 4 in the case of deterministic ordinary differential equations (ODEs). This scheme depends only on increments of the Wiener processes evaluated at the partition points. We found a class of SDEs for which the scheme works and give its order of convergence. Finally some computational results are given. Second, we study the efficiency of this scheme when computing weak approximations of the solutions,in other words, we are interested in quantities of the form $Ef(X_T)$ where $f(x)$ is a sufficiently smooth function and X_T is the solution of a SDE at time T. In this case, we proved that our scheme has global order of convergence h^2 whereas other known schemes given in [19],[24],[25] and [28] while having order h^2 for the weak convergence have global order h in the mean-square sense. Therefore, for this class of SDEs our scheme gives order of convergence h^2 in both situations, mean-square as well as for weak convergence and does not involve neither derivatives nor stochastic integrals.

*1991 *Mathematics Subject Classification.* 34F05, 60H10, 65L06, 65L70.

Key words and phrases. Stochastic differential equations, Stratonovich integrals, Runge-Kutta methods, Commutative noise, rate of convergence.

†School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA. e-mail: roman@math.umn.edu.

PART I

1 Introduction

We first provide a motivation to this paper. Stochastic differential equations have been used to model systems subject to random influences and have become increasingly important in the analysis of a broad range of phenomena in natural sciences and economics. Many systems are described by differential equations where some of the parameters and/or the initial data are not known with complete certainty due to lack of information, uncertainty in the measurements or incomplete knowledge of the mechanisms themselves.
In most cases we are concerned with temporal variation, and the state of the variable system at any particular instant t is described by an equation of the form:

$$dX_t = a(X_t)dt + \sum_{r=1}^{m} \sigma_r(X_t)dW_r(t) \tag{1}$$

where $a(x)$ and $\sigma_r(x)$ functions defined on $\mathbf{R}^d$ with values in $\mathbf{R}^d$ and the $W_{r'_s}(t)$ are one-dimensional independent Wiener processes It has been of much interest to find ways of finding numerical approximations of (1) see [4], [5], [6], [10], [14], [20], [21] and [18] for instance, but, the literature on that subject is much more extensive. The main idea of most of the methods currently being used are based on the stochastic Taylor expansion of (1), which involves derivatives of the functions $a(x)$,$\sigma_{r'_s}(x)$ and stochastic integrals. These present some complications in simulation and approximations so it would be very useful to come up with other method in which we do not have to deal neither with derivatives (or at least not many) nor stochastic integrals.
On the other hand, one of the main concerns of numerical analysts of deterministic differential equations have been to produce effective and accurate schemes where derivatives are avoided as much as possible. In pursuing that goal many results have been obtained and for further details refer to [3], [23]. In particular, Runge-Kutta methods are well known and have been used for so many years already in solving numerically ordinary differential equations. The idea of the method is to replace derivatives of the solution of an ODE in its Taylor expansion by suitable iterative substitutions of the vector field function $F(t, x)$. These "suitable" substitutions are found using the mean value theorem, the Taylor expansion of the

solution and equating terms of the same order. Thus, it is reasonable to expect that similar methods can be translated into the stochastic case to solve stochastic differential equations and this is our main inspiration for the current paper. Runge-Kutta schemes for SDEs are, for instance, introduced in [4], and [5] for a rather general class of SDEs. However, to achieve better order of accuracy they include additional random variables representing higher order stochastic integrals. In the best scenario the best order of convergence they get is $3/2$ and it is only attained in one dimensional cases. Furthermore, they analyze the L_1-error whereas in most of the literature it seems to be of interest the mean square error (L_2 sense) and the error of weak approximations. Here, we found that the classical Runge-Kutta scheme of order 4 for ODEs works perfectly as a scheme of order 2 in both senses, mean square and weak approximations for certain type of SDEs, those of the form $dX_t = \sum_{r=1}^{m} \sigma_r(X_t) \circ dW_r(t)$ where "$\circ$" stands for stochastic integral in the Stratonovich sense, and satisfying the "commutativity noise condition", in other words $\Lambda_j \sigma_k = \Lambda_k \sigma_j, \quad \forall 1 \leq j \leq k \leq m$. Moreover, this scheme does not use stochastic integrals.

We will concentrate on approximations to (1) that depend only on samples of W_t taken at the points of the partition. Under this assumption, it has been shown (see [7]) that in general no numerical method can guarantee accuracy along the trajectory, in the mean-square sense, of higher order than $O(h)$ for a Wiener process of dimension one, and in dimension greater than one when the commutativity condition is satisfied. However, we prove in this paper that under commutativity noise and one more condition: no drift (or drift "zero") when written in the Stratonovich sense, any order of accuracy can be achieved. In [20] is mentioned that when there is no drift, or equivalently, when $a(x) = \frac{1}{2} \sum_{r=1}^{m} \Lambda_r \sigma_r(x)$ in (1) and commutativity holds, arbitrarily high orders of convergence can be achieved, but, no proof or details of this fact are given. In this paper, by using the ODEs approach to analyze SDEs given by Doss in [9], we are able to prove it in a very simple and clear way. Moreover, it lead us to conclude that for this very restrictive class of equations, any scheme of order $2n$ to solve deterministic ODEs becomes a scheme of order n when solving SDEs. This is basically the main value of this paper at this point. It is worth to emphasize that only commutativity is not enough to achieve any order of accuracy using this kind of approximations as it is wrongly said in [6].

The author is not aware at this time of any work where this observation has been

proved and in fact only mentioned in [6] and [20] despite all the numerous studies and reports made on this subject.
In particular, we proved that the classical Runge-Kutta method of order 4 (RK4 in short) to solve ODEs, it's a scheme of order 2 in the mean-square sense when solving $dX_t = \sum_{r=1}^{m} \sigma_r(X_t) \circ dW_r(t)$ under the commutativity condition. Its implementation is a lot easier and shorter than those approximating schemes given in [20] and [6] among others, and achieve better order of accuracy. Besides, because of its simple structure it is not difficult to show than it is also a scheme of order 2 when used to approximate functionals of solutions of (1), in other words, weak approximations.
It is worth mentioning that the study of numerical solutions of SDEs is a very active ongoing research area, among other reasons, because they come up when solving numerically stochastic partial differential equations (SPDEs in short) once a discretization on the space variable has been made by the method of lines, so that a system of SDEs is obtained. This method of "semi-discretization" of stochastic partial differential equations has been already explored by others, we only mention for instance [12]. Another method which has been used successfully in solving numerically SPDEs is the Splitting-Up method described in [1] and [2], where nevertheless we will have to deal with solving a SDE.

2 Runge-Kutta methods for ODEs

We give a very brief summary of how to derive some Runge-Kutta schemes. Consider the following ordinary differential equation:

$$\frac{dx}{dt} = F(t, x(t)) \tag{2}$$

where $F(t, x)$ is a real valued function defined on $[0, T] \times [a, b]$ and is sufficiently smooth so that solution of (2) exists and has as many derivatives as we need in its Taylor expansion. For example, the Taylor polynomial of order 2 of $x(t)$ at $t + h$ for small h is:

$$x(t + h) = x(t) + x'(t)h + x''(t)\frac{h^2}{2} \tag{3}$$

which by using (2) can be written as:

$$x(t+h) = x(t)+F(t,x(t))h+\left(\frac{\partial F}{\partial t}+\frac{\partial F}{\partial x}F\left(t,x(t)\right)\right)\frac{h^2}{2}+\textit{higher order terms} \tag{4}$$

On the other hand, a Taylor expansion gives us

$$a_1F(t+\alpha,x+\beta) = a_1F(t,x)+a_1\alpha\frac{\partial F}{\partial t}+a_1\beta\frac{\partial F}{\partial x}+\textit{higher order terms} \tag{5}$$

where all the derivatives above are evaluated at the point (t,x). Thus, if we want to write (4) in the form

$$x(t+h) = x(t)+hF(t+\alpha,x(t)+\beta) \tag{6}$$

for some choice of a_1,α and β, then by equating (4) and (5) we see that $a_1 = 1$, $\alpha = \frac{h}{2}$ and $\beta = F(t,x)\frac{h}{2}$, therefore the resulting scheme can be written as

$$x(t+h) = x(t)+hF\left(t+\frac{h}{2},x(t)+F\left(t,x(t)\right)\frac{h}{2}\right) \tag{7}$$

The scheme above is known as the "explicit midpoint method". It's been proved that this method has order accuracy equal to 2 (see [3]), which is not hard to believe since it contains all terms of the Taylor expansion of $x(t)$ up to order 2. In a similar way, higher order schemes are derived, and in particular the one that we are interested in this paper is Runge-Kutta of order 4 which is as follows:

$$\begin{aligned}
k_1 &= hF(t,x(t)),\\
k_2 &= hF(t+\frac{h}{2},x(t)+\frac{1}{2}k_1),\\
k_3 &= hF(t+\frac{h}{2},x(t)+\frac{1}{2}k_2),\\
k_4 &= hF(t+h,x(t)+k_3),\\
x(t+h) &= x(t)+\frac{1}{6}(k_1+2k_2+2k_3+k_4).
\end{aligned} \tag{8}$$

Now, it is reasonable to expect that these methods may have similar versions for the stochastic case. It turns out that, indeed, at least, when the the drift "$a(x)$" and diffusion term "$\sigma(x)$" satisfy certain differential equation, these methods work

too, and numerical experiments show much better results than other schemes referred in the literature as those obtained by taking terms from the stochastic Taylor expansion. One of the main advantages of our method is that we do not have to compute derivatives and still obtain a scheme of better order of accuracy than others involving derivatives. Another fact worth mentioning, it is that the scheme obtained in this way is very simple and easier to implement in computers.

3 The Stochastic Case

In order to study the stochastic case we need a sort of "Taylor representation" for the solution of (1), and some results which grant us the convergence of of a given scheme as well as its order of accuracy. Indeed, we have all those tools and they are presented in the subsection below.

3.1 Facts from Numerical Approximation of SDEs

• **The One-step Approximation.**

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let $\mathcal{F}_t$, $t_0 \leq t \leq t_0 + T$, be a nondecreasing family of σ-algebras of $\mathcal{F}$, and let $(W_r(t), \mathcal{F}_t)$, $r = 1, \ldots, m$, be independent Wiener processes. Consider the *system of stochastic differential equations in the sense of Itô*:

$$dX_t = a(t, X_t)dt + \sum_{r=1}^{m} \sigma_r(t, X_t)dW_r(t) \tag{9}$$

where X, a, σ_r are vectors of dimension d. Assume that the functions $a(t, x)$ and $\sigma_r(t, x)$ are defined and continuous for $t \in [t_0, t_0 + T]$, $x \in \mathbf{R}^d$ and satisfy a Lipschitz condition: for all $t \in [t_0, t_0 + T]$, $x \in \mathbf{R}^d$, $y \in \mathbf{R}^d$ there is an inequality

$$|a(t, x) - a(t, y)| + \sum_{r=1}^{m} |\sigma_r(t, x) - \sigma_r(t, y)| \leq \mathbf{K}|x - y|. \tag{10}$$

Here and below $|x|$ denotes the Euclidean norm of the vector x, and we denote by xy the scalar inner product of two vectors x and y. Let $(X(t), \mathcal{F}_t)$, $t_0 \leq t \leq t_0 + T$, be a solution of the system (9) with $\mathbf{E}|X(t_0)|^2 < \infty$. The one-step

approximation $\overline{X}_{t,x}(t+h), t_0 \le t \le t+h \le t_0+T$, is defined as follows, and depends on x, t, h, and $\{W_1(\theta)-W_1(t), \ldots, W_m(\theta)-W_m(t) : t \le \theta \le t+h\}$:

$$\overline{X}_{t,x}(t+h) = x + \mathbf{A}(t, x, h; Ws, i = 1, \ldots, m, t \le \theta \le t+h) \qquad (11)$$

where $\mathbf{A}$ is a function of t, x, h and $W_i(\theta) - W_i(t)$, examples of such $\mathbf{A}$ will be given later.
Using the one-step approximation we recurrently construct the approximations $(\overline{X}_k, \mathcal{F}_{t_k}), k = 0, \ldots, N, t_{k+1} - t_k = h_{k+1}, t_N = t_0 + T$:

$$\begin{aligned} \overline{X}_0 &= X_0 = X(t_0) \\ \overline{X}_{k+1} &= \overline{X}_{t_k, \overline{X}_k}(t_{k+1}) \qquad (12) \\ &= \overline{X}_k + \mathbf{A}(t_k, \overline{X}_k, h_{k+1}; W_i(\theta) - W_i(t_k), i = 1, \ldots, m, t_k \le \theta \le t_{k+1}) \end{aligned}$$

In order to get a better understanding of the *One-step Approximation process* let's give some examples:
Euler's Method:

$$\overline{X}_{t,x}(t,h) = x + a(t,x)h + \sum_{r=1}^{m} \sigma_r(t,x)\Delta_t W_r(h)$$

where $\Delta_t W_r(h) = W_r(t+h) - W_r(t)$. By (12), this approximation generates the Euler's method:

$$X_{k+1} = X_k + a_k h + \sum_{r=1}^{m} \sigma_{r_k} \Delta_k W_r(h) \qquad (13)$$

where a_k, σ_{r_k} are the values of the coefficients a and σ_r at the point (t_k, X_k), and $\Delta_k W_r(h) = W_r(t_k + h) - W_r(t_k)$.
Other examples of *one-step approximation processes* are:

$$\overline{X}^1_{t,x}(t+h) = x + \sum_{r=1}^{m} \sigma_r(t,x)\Delta_t W_r(h) \qquad (14)$$

$$\overline{X}^{mil}_{t,x}(t+h) = \qquad (15)$$

$$\overline{X}^1_{t,x}(t+h) + ah + \sum_{r=1}^{m}\sum_{i=1}^{m} \Lambda_i \sigma_r \int_t^{t+h} (W_i(\theta) - W_i(t)) dW_r(\theta)$$

where Λ_i is an operator which will be defined later and the coefficients a and $\Lambda_i\sigma_r$ are evaluated at (t, x). Notice that the last one is just the Euler step plus an extra term, this one well known as the Milshtein's scheme.
It is clear that for each *one-step approximation* there is an **error** committed, later in Theorem 2 the *smallness* of this error is discussed.
However, at this point, it seems appropriate to ask the following question:
Given a finite interval, without losing generality, let's say $[0, T]$.
Suppose we want to approximate the solution of (9) at time $t = T$ by successive *one-step approximations*. Since at each step we make an error, the question is: How large is the error carried up to time $t = T$?
The answer to this question is given by the next theorem below, but, let's first get familiar with notation.
As usual, $X_{t_k,X}(t)$ denotes the solution of (9) for $t_k \leq t \leq t_0 + T$ satisfying the following initial condition at $t = t_k$: $X(t_k) = X$. By $\overline{X}_{t_k,X}(t_i), t_i \geq t_k$, we denote the approximation of the solution at step i and such that $\overline{X}_k = X$,(where $\overline{X}_k = \overline{X}(t_k)$). For example,

$$\overline{X}_{t_0,X_0}(t_{k+1}) = \overline{X}_{t_k,\overline{X}_k}(t_{k+1}) = \overline{X}_{k+1} \tag{16}$$

For simplicity reasons we assume that $t_{k+1} - t_k = h = \frac{T}{N}$.

Theorem 1 *([18])*

Suppose the one-step approximation $\overline{X}_{t,x}(t + h)$ has order of accuracy p_1 for the mathematical expectation of the deviation and order of accuracy p_2 for the mean-square deviation; more precisely, for arbitrary $t_0 \leq t \leq t_0 + T - h, x \in \mathbf{R}$ the following inequalities hold:

$$|\mathbf{E}(X_{t,x}(t + h) - \overline{X}_{t,x}(t + h))| \leq \mathbf{K_1}(1 + |x|^2)^{\frac{1}{2}} h^{p_1} \tag{17}$$

$$[\mathbf{E}|X_{t,x}(t + h) - \overline{X}_{t,x}(t + h)|^2]^{\frac{1}{2}} \leq \mathbf{K_2}(1 + |x|^2)^{\frac{1}{2}} h^{p_2} \tag{18}$$

Also, let

$$p_2 \geq \frac{1}{2}, \qquad p_1 \geq p_2 + \frac{1}{2} \tag{19}$$

Then for any N and $k = 0, 1, \ldots, N$ the following inequality holds:

$$[\mathbf{E}|X_{t_0,X_0}(t_k) - \overline{X}_{t_0,X_0}(t_k)|^2]^{\frac{1}{2}} \leq \mathbf{K_3}(1 + \mathbf{E}|X_0|^2)^{\frac{1}{2}} h^{p_2 - \frac{1}{2}} \tag{20}$$

i.e. the global order of accuracy of the method constructed using the one-step approximations $\overline{X}_{t,x}(t+h)$ is $p = p_2 - \frac{1}{2}$.
From now, whenever we talk about global order we mean $p = p_2 - \frac{1}{2}$.
• **Expansion of the solution of a system of stochastic differential equations (Wagner-Platen expansion, see [29]).**

Let $X_{t,x}(s) = X(s)$ be the solution of the system (9), and let $f(t,x)$ be a sufficiently smooth (scalar or vector) function. By Itô's formula we have for $t_0 \leq t \leq \theta \leq t_0 + T$:

$$f(\theta, X(\theta)) = f(t,x) + \sum_{r=1}^{m} \int_t^\theta \Lambda_r f(\theta_1, X(\theta_1)) dW_r(\theta_1) + \int_t^\theta Lf(\theta_1, X(\theta_1)) d\theta_1 \tag{21}$$

where the operators $\Lambda_r, r = 1, \ldots, m$, and L are given by:

$$\begin{aligned} \Lambda_r &= (\sigma_r, \frac{\partial}{\partial x}), \\ L &= \frac{\partial}{\partial t} + (a, \frac{\partial}{\partial x}) + \frac{1}{2} \sum_{r=1}^{m} \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_r^i \sigma_r^j \frac{\partial^2}{\partial x^i \partial x^j}. \end{aligned}$$

Formula (21) is the analog of the Taylor Expansion of an ODE in the deterministic case.

Apply (21) to the functions $\Lambda_r f$ and Lf, and subsequently insert the expressions obtained for $\Lambda_r f(\theta, X(\theta))$ and $Lf(\theta, X(\theta))$ into (21). We find

$$\begin{aligned} f(s, X(s)) &= f + \sum_{r=1}^{m} \Lambda_r f \int_t^s dW_r(\theta) + Lf \int_t^s d\theta \\ &\quad + \sum_{r=1}^{m} \int_t^s \left(\sum_{s}^{m} \int_t^\theta \Lambda_s \Lambda_r f(\theta_1, X(\theta_1)) dW_s(\theta_1) \right) dW_r(\theta) \\ &\quad + \sum_{r=1}^{m} \int_t^s \left(\int_t^\theta L \Lambda_r f(\theta_1, X(\theta_1)) d\theta_1 \right) dW_r(\theta) \\ &\quad + \sum_{r=1}^{m} \int_t^s \left(\int_t^\theta \Lambda_r L f(\theta_1, X(\theta_1)) dW_r(\theta_1) \right) d\theta \\ &\quad + \int_t^s \left(\int_t^\theta L^2 f(\theta_1, X(\theta_1)) d\theta_1 \right) d\theta \end{aligned} \tag{22}$$

where, e.g., $\Lambda_r f$ and Lf are computed at (t, x).
Continuing this way we obtain an expansion for $f(t+h, X(t+h))$. As proved before, in the deterministic situation this expansion is the Taylor expansion in powers of h with remainder of integral type. In the stochastic situation the role of powers is played by random variables of the form (they are independent of $\mathcal{F}_t$)

$$I_{i_1,\dots,i_j}(h) = \int_t^{t+h} dW_{i_j}(\theta) \int_t^{\theta} dW_{i_{j-1}}(\theta_1) \int_t^{\theta_1} \dots \int_t^{\theta_{j-2}} dW_{i_1}(\theta_{j-1}) \qquad (23)$$

where $i_1, \dots, i_j$ take values in the set $\{0, 1, \dots, m\}$, and where $dW_0(\theta_r)$ is understood to mean $d\theta_r$.

•Order of Accuracy.

The next theorem we state is the one we used to determine the order of accuracy of a given scheme. In what follows $\Lambda_0 = L$. Let

$$\bar{i}_k = \begin{cases} 0, & i_k = 0 \\ 1, & i_k \neq 0 \end{cases}$$

Then we have the following:

Theorem 2 *([14],[18])*

Suppose that $\overline{X}_{t,x}(t+h)$ includes all terms of the form $\Lambda_{i_1} \dots \Lambda_{i_j} f I_{i_1,\dots,i_j}$, where $f \equiv x$, up to order m inclusive.
Suppose that all functions $\Lambda_{i_1} \dots \Lambda_{i_j} f(t,x)$, where $f \equiv x$, $\sum_{k=1}^{j}(2-\bar{i}_k)/2 \leq m+1$, satisfy $\left|\Lambda_{i_1} \dots \Lambda_{i_j} f(t,x)\right| \leq \mathbf{K}(1+|x|^2)^{1/2}$. Then the mean-square order of accuracy of the method based on this approximation is equal to m.
Suppose that $\overline{X}_{t,x}(t+h)$ includes all terms of the form $\Lambda_{i_1} \dots \Lambda_{i_j} f I_{i_1,\dots,i_j}$, where $f \equiv x$, up to order $m+1/2$ inclusive, as well as the term $L^m a \int_t^{t+h} d\theta \int_t^{\theta} d\theta_1 \dots \int_t^{\theta_{m-1}} d\theta_m = L^m a h^{m+1}/(m+1)!$. Suppose that all functions $\Lambda_{i_1} \dots \Lambda_{i_j} f(t,x)$, where $f \equiv x$, $\sum_{k=1}^{j}(2-\bar{i}_k)/2 \leq m+2$, satisfy $\left|\Lambda_{i_1} \dots \Lambda_{i_j} f(t,x)\right| \leq \mathbf{K}(1+|x|^2)^{1/2}$. Then the mean-square order of accuracy of the method based on this approximation is equal to $m+1/2$.

3.2 Runge-Kutta type schemes for certain type of SDEs (1-dimensional case).

To begin, let us consider a simpler version of (9), $m = 1$, the case for arbitrary m follows easily from this. Thus, we study a stochastic differential equation of the

form:

$$\begin{cases} dX_t = a(X_t)dt + \sigma(X_t)dW_t \\ X(0) = X_0 \end{cases} \tag{24}$$

where $t \in [0, T]$, $a(x)$ and $\sigma(x)$ are real valued functions defined on $\mathbf{R}$ and satisfy the following Lipschitz condition:

$$|a(x) - a(y)| + |\sigma(x) - \sigma(y)| \leq K|x - y| \tag{25}$$

and W_t is a standard Brownian motion.
Under condition (25) it is proved that equation (24) with initial condition X_0 given has unique solution (see [8]). We can think of (24) "informally" speaking as:

$$\begin{cases} \dfrac{dX_t}{dt} = a(X_t) + \sigma(X_t)\dot{W}_t \\ X(0) = X_0 \end{cases} \tag{26}$$

where $\dot{W}_t$ has no meaning in the strict sense since trajectories of Brownian motion are nowhere differentiable, but, just for intuition purposes in our paper let us write it in that way. If this is really bothersome, the reader could substitute $\dot{W}_t$ by $f'(t)$ where $f(t)$ is any function differentiable in time. Now, it is well known (see [3]) that the Euler scheme for deterministic ODEs has order of convergence equal to 1, so a natural start could be trying to mimic this method to solve (26). By doing this we obtain the following scheme:

$$X_{t+h} = X_t + a(X_t)h + \sigma(X_t)\Delta_h W_t \tag{27}$$

where $X_t = x(t)$, and $\Delta_h W_t = W(t + h) - W(t)$. Notice that we have replaced $\dot{W}_t h$ by $\Delta_h W_t$ which is very natural. On the other hand, if we look at the stochastic Taylor expansion of order 1, for the solution of (1) we have the Milshtein's scheme:

$$X_{t+h} = X_t + a(X_t)h + \sigma(X_t)\Delta_h W_t + \Lambda\sigma\left(\frac{(\Delta_h W_t)^2 - h}{2}\right) \tag{28}$$

where $\Lambda := \sigma\frac{\partial}{\partial x}$ and $\Lambda\sigma$ is also evaluated at X_t. Therefore, we see that the Euler scheme (27) cannot be of order 1 since it does not include all terms of order 1 from its Stochastic Taylor Expansion (STE, in short). Indeed, it does converge also, but,

its order of accuracy is $1/2$ as follows easily from Theorem 2. Thus, we can say that accuracy of Euler's method is reduced by half from deterministic equations to stochastic ones. Next, a natural question that comes to mind is: What happens with the "explicit midpoint method"?, well, in that case the implementation of it will be as follows:

$$\begin{aligned} k &= a(X_t)h + \sigma(X_t)\Delta_h W_t \\ X_{t+h} &= X_t + a(X_t + \tfrac{k}{2})h + \sigma(X_t + \tfrac{k}{2})\Delta_h W_t \end{aligned} \tag{29}$$

Now, using the Taylor expansions of $a(x)$ and $\sigma(x)$ up to order 1 in h and 2 in $\Delta_h W_T$, around X_t, in (29) we have:

$$X_{t+h} = X_t + a(X_t)h + \sigma(X_t)\Delta_h W_t + \frac{1}{2}\sigma\sigma'(X_t)(\Delta_h W_t)^2 \tag{30}$$

Still, we observe, by comparison, that in (30) we are missing one term of order 1, namely "$-\frac{1}{2}\sigma\sigma' h$", and it is not hard to see that it would be impossible to get that term in (30) in a natural way, in other words, it will never come out from the Taylor expansion of it. So, we need to fix it, and we do that by simply subtracting that term since the beginning, so that (29) becomes:

$$\begin{aligned} k &= a(X_t)h + \sigma(X_t)\Delta_h W_t - \tfrac{1}{2}\sigma\sigma'(X_t)h \\ X_{t+h} &= X_t + a(X_t + \tfrac{k}{2})h + \sigma(X_t + \tfrac{k}{2})\Delta_h W_t - \tfrac{1}{2}(\sigma\sigma')(X_t + \tfrac{1}{2}k)h \end{aligned}$$

Usually, we will omit the evaluation point in some of the terms of the expansion, understanding that all are evaluated at the same point unless something else is mentioned.

So, by adding the "*corrector term*" $-\frac{1}{2}\sigma\sigma' h$, to (30) we get a scheme of order of accuracy at least 1, again by Theorem 2. Moreover, it cannot be of higher order since it will have to include cubic powers of $\Delta_h W_t$ and it will not happen with just one "iteration" in the arguments of $a(\cdot)$, $\sigma(\cdot)$ and $\frac{1}{2}\sigma\sigma'(\cdot)$ respectively. Therefore, (31) has order of accuracy equal to 1, whereas the same method (midpoint) has order of accuracy equal to 2 for deterministic equations. Once again, the order of accuracy drops by half from deterministic to stochastic equations with respect mean square calculus.

3.2.1 RK4 is a scheme of order 2 for SDEs of the form: "$dX_t = \sigma(X_t) \circ dW_t$".

Our next goal is to prove that the classical Runge-Kutta method of order 4,(RK4 in short), used in deterministic ODEs, works as a scheme to solve numerically certain

SDEs, with order of accuracy equal to 2. Unfortunately, we do not have yet an answer in the general case yet, but, we show that when $a(x) = \frac{1}{2}\sigma(x)\sigma'(x)$, i.e., for equations of the form $dX_t = \sigma(X_t) \circ dW_t$, where "$\circ$" means stochastic integral in the Stratonovich sense, we actually have a scheme with the desired accuracy, which does not involves derivatives of $\sigma(x)$ and it is very easy to implement.
Next, we give the STE of order 2, of the solution of (9):

$$\begin{aligned} X_{t+h} &= X_t + a(X_t)h + \sigma(X_t)\Delta_h W_t + \frac{1}{2}\sigma(X_t)\sigma'(X_t)\left((\Delta_h W_t)^2 - h\right) \\ &+ L(\sigma)\int_t^{t+h}(\theta - t)dW(\theta) + \Lambda(a)\int_t^{t+h}(W(\theta) - W(t))d\theta \\ &+ \Lambda^2(\sigma)\int_t^{t+h}\int_t^{\theta}\int_t^{\theta_1} dW(\theta_2)dW(\theta_1)dW(\theta) \\ &+ L(a)\frac{h^2}{2} + \Lambda^3(\sigma)\int_t^{t+h}\int_t^{\theta}\int_t^{\theta_1}\int_t^{\theta_2} dW(\theta_3)dW(\theta_2)dW(\theta_1)dW(\theta) \\ &+ L\Lambda(\sigma)\int_t^{t+h}\int_t^{\theta}\int_t^{\theta_1} d\theta_2 dW(\theta_1)dW(\theta) \\ &+ \Lambda L(\sigma)\int_t^{t+h}\int_t^{\theta}\int_t^{\theta_1} dW(\theta_2)d\theta_1 dW(\theta) \\ &+ \Lambda^2(a)\int_t^{t+h}\int_t^{\theta}\int_t^{\theta_1} dW(\theta_2)dW(\theta_1)d\theta \end{aligned} \tag{31}$$

where $L := a\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}$ and $\Lambda := \sigma\frac{\partial}{\partial x}$ and all coefficients are evaluated at the point X_t.
Now, two of the terms of order $3/2$, namely the factors of $L(\sigma)$ and $\Lambda(a)$ respectively, can be combined using the identity $\int_t^{t+h}(\theta - t)dW(\theta) = h\Delta_h W_t - \int_t^{t+h}(W(\theta) - W(t))d\theta$ (see [18]), to obtain instead:

$$L(\sigma)h\Delta_h W_t + (\Lambda(a) - L(\sigma))\int_t^{t+h}(W(\theta) - W(t))d\theta \tag{32}$$

Notice that it is impossible to obtain the factor "$\int_t^{t+h}(W(\theta) - W(t))d\theta$" by just iterating in the arguments of $a(\cdot)$ and $\sigma(\cdot)$ like in Runge-Kutta of order 4 method (see (8) in section 2), in fact, in (8) we get an expansion of the form $p(h, \Delta_h W_t)$,

where $p(x,y)$ is a polynomial of two variables of degree≤ 4. Therefore, if we expect this scheme to give a solution to (24) we better assume that:

$$\Lambda(a)(x) - L(\sigma)(x) = 0 \tag{33}$$

which in our case becomes:

$$\sigma(x)a'(x) - a(x)\sigma'(x) = \frac{1}{2}\sigma^2(x)\sigma''(x) \tag{34}$$

whose general solution is given by:

$$a(x) = \frac{1}{2}\sigma(x)\sigma'(x) + c\sigma(x) \tag{35}$$

where c is any arbitrary constant. In particular, for $c = 0$, we have $a(x) = \frac{1}{2}\sigma(x)\sigma'(x)$ so (24) turns out to be of the form:

$$dX_t = \frac{1}{2}\sigma(X_t)\sigma'(X_t)dt + \sigma(X_t)dW_t = \sigma(X_t) \circ dW_t \tag{36}$$

Now, it is a straightforward computation to proof that if $\Lambda(a)(x) - L(\sigma)(x) = 0$, then the following identities hold:

$$\begin{aligned} L\Lambda(\sigma) = \Lambda L(\sigma) &= \Lambda^2(a) = \frac{1}{2}\Lambda^3(\sigma) \\ \Lambda(a) &= \frac{1}{2}\Lambda^2(\sigma) \\ L(a) &= \frac{1}{4}\Lambda^3(\sigma) \end{aligned} \tag{37}$$

Define

$$\mathcal{I}_{i_1,\ldots,i_j}(h) = \int_t^{t+h} dW_{i_j}(\theta) \int_t^{\theta} dW_{i_{j-1}}(\theta_1) \int_t^{\theta_1} \cdots \int_t^{\theta_{j-2}} dW_{1_1}(\theta_{j-1}) \tag{38}$$

where the indices $i_1, \ldots, i_j$ can take either the value 0 (then $dW_0(\theta) = d\theta$) or 1 (then $dW_1(\theta) = dW(\theta)$).
Then (31) can be written as:

$$\begin{aligned} X_{t+h} &= X_t + \sigma(X_t)\Delta_h W_t + \frac{1}{2}\sigma(X_t)\sigma_x(X_t)(\Delta_h W_t)^2 \\ &\quad + \frac{1}{2}\Lambda^2(\sigma)h\Delta_h W_t + \frac{1}{2}\Lambda^3(\sigma)\left(\mathcal{I}_{0,1,1}(h) + \mathcal{I}_{1,1,0}(h) + \mathcal{I}_{1,0,1}(h)\right) \\ &\quad + \Lambda^2(\sigma)\mathcal{I}_{1,1,1}(h) + \frac{1}{4}\Lambda^3(\sigma)\frac{h^2}{2} \\ &\quad + \Lambda^3(\sigma)\mathcal{I}_{1,1,1,1}(h) \end{aligned} \tag{39}$$

by (37).
Moreover, it is shown in the appendix that

$$\mathcal{I}_{0,1,1}(h) + \mathcal{I}_{1,1,0}(h) + \mathcal{I}_{1,0,1}(h) = h\left(\frac{(\Delta_h W_t)^2 - h}{2}\right) \tag{40}$$

Also, after many mistakes, corrections and so much time spent in computing some of the stochastic integrals that appear in (39), it can be proved that:

$$\begin{aligned}\mathcal{I}_{1,1,1}(h) &= \frac{1}{6}\left(W^3(t+h) - W^3(t)\right) \\ &- \frac{1}{2}\Delta_h W_t\left(W^2(t) + h + W\Delta_h W_t\right)\end{aligned}$$

and

$$\begin{aligned}\mathcal{I}_{1,1,1,1}(h) &= \frac{1}{24}\left(W^4(t+h) + W^4(t)\right) \\ &- \frac{1}{4}\left(W^2(t+h)h - W^2(t)W^2(t+h) + W^2(t)h\right) \\ &+ \frac{1}{2}W(t)W(t+h)h - \\ &+ \frac{1}{2}\left(\frac{W(t)W^3(t+h)}{3} - \frac{W(t+h)W^3(t)}{3} + \frac{h^2}{4}\right)\end{aligned}$$

In these previous calculations, identities obtained by using the Hermite polynomials evaluated at $(t, \Delta_h W(t))$ are used, as well as construction of stochastic integrals in a recursive way (see [18]).
Thus, substituting the last two expressions above into (39) we have:

$$\begin{aligned}X_{t+h} &= X_t + \sigma(X_t)\Delta_h W_t + \frac{1}{2}\sigma(X_t)\sigma_x(X_t)(\Delta_h W_t)^2 \\ &+ \frac{1}{2}\Lambda^2(\sigma)h\Delta_h W_t + \frac{1}{2}\Lambda^3(\sigma)h\left(\frac{(\Delta_h W_t)^2 - h}{2}\right) \\ &+ \Lambda^2(\sigma)\left(\frac{1}{6}\left(W^3(t+h) - W^3(t)\right) - \frac{1}{2}\Delta_h W_t\left(W^2(t) + h + W\Delta_h W_t\right)\right) \\ &+ \frac{1}{4}\Lambda^3(\sigma)\frac{h^2}{2} + \Lambda^3(\sigma)\left(\frac{1}{24}(W^4(t+h) + W^4(t))\right) \\ &- \frac{\Lambda^3(\sigma)}{4}\left(W^2(t+h)h - W^2(t)W^2(t+h) + W^2(t)h\right)\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{\Lambda^3(\sigma)}{2}\,(W(t)W(t+h)h) \\
&+ \quad \Lambda^3(\sigma)\left(-\frac{W(t)W^3(t+h)}{6} - \frac{W(t+h)W^3(t)}{6} + \frac{h^2}{8}\right) \qquad (41)
\end{aligned}
$$

which after some straightforward but tedious computations can be written as:

$$
\begin{aligned}
X_{t+h} &= X_t + \sigma(X_t)\Delta_h W_t + \Lambda\sigma(X_t)\frac{(\Delta_h W_t)^2}{2!} + \Lambda^2\sigma(X_t)\frac{(\Delta_h W_t)^3}{3!} \\
&+ \Lambda^3\sigma(X_t)\frac{(\Delta_h W_t)^4}{4!} \qquad (42)
\end{aligned}
$$

On the other hand, we perform our RK4 scheme as follow:

$$
\begin{aligned}
k_1 &= \sigma(x(t))\Delta_h W_t, \\
k_2 &= \sigma(x(t) + \frac{1}{2}k_1)\Delta_h W_t, \\
k_3 &= \sigma(x(t) + \frac{1}{2}k_2)\Delta_h W_t, \\
k_4 &= \sigma(x(t) + k_3)\Delta_h W_t, \\
x(t+h) &= x(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \qquad (43)
\end{aligned}
$$

Then, we compute the expansion of (43) of order 4 in $\Delta_h W_t$ using Maple V (see appendix for Maple routines). If the reader feels insecure of Maple, the author has to say that doing the expansion by hand could be one of the most challenging and longest computations from which any learning or useful information is gained other than just applications of the mean value theorem over and over. Besides, the chances of making mistakes are enormous and it is a very tedious task.
After getting the expansion of (43), we subtract it from (41) and Eureka !!, in fact, our scheme includes all terms of the stochastic Taylor expansion of $x(t)$ up to order 2, so by theorem 2 we conclude that RK4 has mean square order of accuracy equal to 2. Once again, we notice that the order of accuracy is reduced by half compared with Runge-Kutta of order 4 in the case of deterministic ordinary differential equations. Therefore, our result can be briefly described by:
Runge-Kutta of order 4 for ODEs becomes a scheme of order 2 for SDEs of the form: $dX_t = \sigma(X_t) \circ dW_t$, $X_t \in \mathbf{R}$.

3.3 Multidimensional Case

So far we have work only in the case when X_t and W_t are one dimensional stochastic processes, and $a(x)$,$\sigma(x)$ real valued functions. In this section we consider the following equation:

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t \tag{44}$$

where X_t is a d-dimensional stochastic process, $a : \mathbf{R}^d \mapsto \mathbf{R}^d$, $\sigma : \mathbf{R}^d \mapsto \mathcal{L}(\mathbf{R}^m, \mathbf{R}^d)$ (the space of linear functionals from $\mathbf{R}^m$ to $\mathbf{R}^d$), and W_t is a m-dimensional Wiener processes.
We also observe that (44) can be written as:

$$dX_t = a(X_t)dt + \sum_{r=1}^{m} \sigma_r(X_t)dW_r(t) \tag{45}$$

where

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_{dm} \end{pmatrix}$$

and

$$\sigma_r = \begin{pmatrix} \sigma_{1r} \\ \sigma_{2r} \\ \vdots \\ \sigma_{dr} \end{pmatrix} \qquad W_t = \begin{pmatrix} W_1(t) \\ W_2(t) \\ \vdots \\ W_m(t) \end{pmatrix}$$

The $W_{i's}$ are m-independent real valued Wiener processes.

3.3.1 Motivation

In [9] Doss shows that the integration of a stochastic differential equation in $\mathbf{R}$ can be reduced to the integration of an ordinary differential equation parameterized by a variable in the base probability space. They study some properties of the solution and showed that the result remains true, under some conditions, in a finite or infinite dimensional case. Later, in an article by Talay (see [27]) he shows the same results for a more specific case which applies to our problem considered in this paper.

We first give some of their results obtained in the real case ($n = m = 1$).
Let X_t be a solution of the following SDE:

$$X_t = x_0 + \int_0^t a(x_s)ds + \int_0^t \sigma(x_s)dW_s \tag{46}$$

where W_s is a one dimensional Wiener process and $a(x)$,$\sigma(x)$ satisfies:
i) $a(x)$ is Lipschitz.
ii) $\sigma(x)$ is in $\mathcal{C}^1$ and its derivative is bounded.
Under conditions i) and ii), it is proved in [9], [27] that there exists a function h differentiable, and a real valued process D_t such that

$$X_t = h(D_t, W_t) \quad \text{almost surely} \tag{47}$$

Furthermore, h is solution of the following ODE:

$$\frac{\partial h}{\partial y}(x, y) = \sigma(h(x, y)); \quad h(x, 0) = x \tag{48}$$

and D_t is solution of:

$$\begin{cases} D_0 = x_0 \\ D'_t = \left(\exp[-\int_0^{W_t} \sigma'(h(D_t, s))ds]\right) \left(a - \frac{1}{2}\sigma\sigma'\right)(h(D_t, W_t)) \end{cases} \tag{49}$$

where equation (49) is understood also pathwise. The function h satisfies many properties which can be found in [27]. Our main observation at this point is that when "$a = \frac{1}{2}\sigma\sigma'$", (49) gives us $D_t = X_0$ for all t and thus $X_t = h(X_0, W_t)$, in other words X_t is function of the initial data X_0 and the Wiener process W_t.
This result gives a hint on when the traditional Runge-Kutta method of order 4 for systems of ODEs may also work for systems of SDEs, so we state the following claim:
If (45) can be written as a SDE in the sense of Stratonovich with no drift, then Runge-Kutta of order 4 for ODEs gives a scheme of order 2 for such SDE.
Notice that in the real case ($d = m = 1$) the conjecture has been proved in the previous sections of this paper. In order to show that the result remains true in the multi-dimensional our proof will be based on this representation of the solution given by Doss. However, this representation given by Doss holds in finite or infinite dimensions under other restrictive conditions on σ_r called "*commutativity*".

3.3.2 Main Result

We begin by giving the notation used in this part. For $u \in \mathbf{R}^d$ we denote by $u^{(i)}$, or u^{Δ}, any of its components. In particular, if $h(x,y) = (h^1(x,y), \ldots, h^d(x,y))$ then $h^{\Delta}(x,y)$ represents an arbitrary $h^i(x,y)$ for $1 \leq i \leq d$. We consider equation (44) as in the previous subsection, but now we require two more conditions. Thus, let us assume the following:
H1) $a(X)$ is Lipschitz.
H2) σ is in $\mathcal{C}^2$.
H3)Commutative Noise:

$$\sum_{i=1}^{d} \sigma_{ik} \frac{\partial \sigma_{lj}}{\partial x_i} = \sum_{i=1}^{d} \sigma_{ij} \frac{\partial \sigma_{lk}}{\partial x_i} \tag{50}$$

$\forall j,k,l: \quad 1 \leq j \leq m; 1 \leq k \leq m; 1 \leq l \leq d$, and where σ_{ij} denotes the (i,j) component of σ.
Also, consider the system of ODEs:

$$\begin{cases} \dfrac{\partial h^{(i)}}{\partial y_j}(x,y) = \sigma_{ij}(h(x,y)), \quad 1 \leq i \leq d, 1 \leq j \leq m \\ h(x,0) = x \end{cases} \tag{51}$$

Further, let $D_t \in \mathcal{R}^d$ be the solution of the following system of ODEs:

$$\begin{cases} D_0 = X_0 \\ \dfrac{d}{dt} D_t^{\Delta} = \displaystyle\sum_{i=1}^{d} \left\{ \frac{\partial h^{\Delta}}{\partial x_i}(h(D_t, W_t), -W_t) \right\} \\ \qquad \times \left\{ a^i(h(D_t,W_t)) - \frac{1}{2} \displaystyle\sum_{r=1}^{d} \sum_{j=1}^{m} \frac{\partial \sigma_{ij}}{\partial x_r}(h(D_t,W_t)) \sigma_{rj}(h(D_t,W_t)) \right\} \end{cases} \tag{52}$$

Then, under conditions **H1**, **H2** and **H3**, it is proved in [27] the following:
• (51) has unique solution $h(\cdot,\cdot) : \mathbf{R}^d \times \mathbf{R}^m \to \mathbf{R}^d$ and satisfies several identities involving its first derivatives and some of higher order. Among them we mention the following:

$$\frac{\partial h}{\partial \alpha}(\alpha, t\beta) = I_{d\times d} + \int_0^t D\sigma(h(\alpha, s\beta)).\frac{\partial h}{\partial \alpha}(\alpha, s\beta).\beta ds \ \ \forall t \in \mathbf{R} \tag{53}$$

$$\frac{\partial h}{\partial \beta}(\alpha, \beta) = \sigma\left(h(\alpha,\beta)\right) \tag{54}$$

where $I_{d\times d}$ is the identity matrix in $\mathbf{R}^d \times \mathbf{R}^d$.
• $\forall T > 0$, D_t exists on $[0, T]$, and

$$\exists M_1 > 0 / \forall t \in [0, T], \quad \|D_t\| \leq M_1 \tag{55}$$

where $\|(u_1, \ldots, u_d)\| = \sup_{1\leq i\leq d} |u_i|$.
• $X_t = h(D_t, W_t)$.
They also showed in [9] and [27] that if $\overline{D}_k$ is a scheme approximating the solution of (52) then $\overline{X}_k = h(\overline{D}_k, W_t)$ is an approximation of X_t at time t_k.

Proposition 1 *Assume* **H1**, **H2** *and* **H3**. *Besides, suppose that*

$$a(x) = \frac{1}{2}\sum_{r=1}^{m} \Lambda_r \sigma_r(x), \quad \forall x \tag{56}$$

where Λ_r is the differential operator given by $\Lambda_r = \left(\sigma_r, \frac{\partial}{\partial x}\right)$ and finally as in Theorem 2 assume that $|\Lambda_{i_1} \ldots \Lambda_{i_j} f(t,x)| \leq \mathbf{K}\left(1 + |x|^2\right)^{1/2}$ where $f \equiv x$ and $\sum_{k=1}^{j} i_k \leq 2n+1$. Then any numerical scheme of order $2n$ to approximate solutions of ODEs becomes a scheme of order n in the mean-square sense when approximating SDEs satisfying all the assumptions above.

Remark: By a numerical scheme of order $2n$ we mean there exists an operator $\Phi_h^{2n} : \mathcal{C}^{2n+1}\left(\mathbf{R}^d\right) \to \mathcal{C}\left(\mathbf{R}^d\right)$ such that $\Phi_h^{2n}(f)(x)$ gives the multidimensional Taylor expansion of f at x of order $2n$ in powers of h and $\overline{X}_{k+1} = \Phi_h^{2n}(X)(\overline{X}_k)$.
Proof:

In [9] is proved that under these conditions there is a solution $h(x, y)$ to equation (51). Further, (56) implies that:

$$a^i(x) = \frac{1}{2}\sum_{r=1}^{d}\sum_{j=1}^{m} \frac{\partial \sigma_{ij}}{\partial x_r}\sigma_{rj}(x), \quad \forall i = 1, \ldots, d. \tag{57}$$

Hence the solution of the ODE (52) is given by $D_t = X_0$ for all t and moreover,

$$X_t = h(D_0, W_t) \tag{58}$$

Assume that $\sigma \in \mathcal{C}^{2n}\left(\mathbf{R}^d\right)$ in order to make sense of the scheme of order $2n$ then:

$$\begin{aligned} X_{t+h} &= h(D_0, W_{t+h}) \\ &= h(D_0, W_t) + \sum_{k=1}^{2n}\left(\sum_{\gamma=k, \gamma\in\Sigma} a_{i_{\phi(1)}\ldots i_{\phi(m)}} \mathcal{I}_1^{i_{\phi(1)}}(h)\ldots\mathcal{I}_m^{i_{\phi(m)}}(h)\right) \\ &\quad + \rho_{2n} \end{aligned} \tag{59}$$

where $\Sigma \doteq \left\{ i_{\phi(1)} + \ldots + i_{\phi(m)} : \phi \in \mathcal{S}_m \right\}$, and $\mathcal{S}_m$ is the set of permutations of m elements, the $a_{i_1 i_2 \ldots i_{m's}}$ are the coefficients of the d-dimensional Taylor expansion of $h(D_0, \cdot)$ evaluated at (D_0, W_t), and

$$\rho_{2n} = \sum_{s_1=1}^{m} \ldots \sum_{s_{2n}=1}^{m} \left(\int_t^{t+h} \ldots \left(\int_t^{\theta_{2n-2}} \Lambda_{s_1} \ldots \Lambda_{s_{2n}} h(\theta_{2n-1}, X(\theta_{2n-1})) dW_{s_{2n}}(\theta_{2n-1}) \right) \ldots dW_{s_1}(\theta) \right)$$

is the remainder of the expansion. Moreover, by the growth condition assumed in the statement it follows that:

$$|\rho_{2n}| \leq \mathbf{K}_\rho \left(1 + |X(t)|^2\right) \sum_{\gamma=2n+1, \gamma \in \Sigma} \mathcal{I}_1^{i_{\phi(1)}}(h) \ldots \mathcal{I}_m^{i_{\phi(m)}}(h) \tag{60}$$

where $\mathcal{I}_s(h)$ is defined as in section 3.2.1.
Since $\{W_1(t+h) - W_1(t), W_2(t+h) - W_2(t), \ldots, W_m(t+h) - W_m(t)\}$ are independent and each is a Gaussian process with mean zero and variance h it is easy to verify that:

$$\begin{aligned} \mathbf{E}|\rho_{2n}| &\leq \mathbf{K_1} \left(1 + |X(t)|^2\right) h^{n+1} \\ \left(\mathbf{E}(\rho_{2n})^2\right)^{1/2} &\leq \mathbf{K_2} \left(1 + |X(t)|^2\right) h^{n+1/2} \end{aligned} \tag{61}$$

as in the notation of theorems 1 and 2 of section 3.1.
Notice that (59) shows that the only information we would need in order to approximate X_{t+h} are the increments of the Wiener processes at the partition points, i.e, $\Delta_h W_i(t_k)$. Therefore, any scheme which produces all the terms of the Taylor expansion of h up to order $2n$ will give the expansion of X_{t+h} in powers of $(W_{t+h} - W_t)$ up to the $2nth$ power.
Define $\overline{X}_{t_{k+1}} = \Phi^{2n}_{\Delta_h W_{t_k}}(X)(\overline{X}_{t_k})$ where $\Phi^{2n}_h(x,s)$ is defined as in the remark above. Then by taking expectation in (59) and (61) this schemes satisfies:

$$\begin{aligned} \mathbf{E}|X_{t+h} - \overline{X}_{t+h}| &\leq K_1 \left(1 + |X_t|^2\right) h^{n+1} \\ \left(\mathbf{E}(X_{t+h} - \overline{X}_{t+h})^2\right)^{1/2} &\leq K_2 \left(1 + |X_t|^2\right) h^{n+1/2} \end{aligned}$$

Hence by theorem 1 the scheme generated by using the one-step approximation process given above is of mean-square order n. This ends the proof.

We make the remark that Doss' representation is what makes possible this proof in such a trivial terms.
Now, in particular if we want to implement RK4 (used in the 1-dimensional case) for multidimensional equations, we need to know whether this method works for systems of first order equations. Indeed, it does work and some details can be found in [11] where it is proved that the method converges and still it is of order 4 under sufficiently smooth conditions on the vector field. Therefore, we can expect in the d-dimensional case that the classical Runge-Kutta method of order 4 also works for systems of SDEs, in other words it gives the Taylor expansions in terms of $\Delta W_t(h)$ up to the fourth power so that the method itself will be of order 2 as in the 1-dimensional case. In fact, let's explore in more detail some of the implications of these assumptions. For instance, straightforward computations show that **(H3)** implies:

$$\Lambda_j \sigma_k = \Lambda_k \sigma_j, \quad \forall \ \ 1 \leq k, j \leq m \tag{62}$$

Also, a significant amount of work but requiring only multidimensional calculus shows that under the assumption

$$a(x) = \frac{1}{2} \sum_{r=1}^{m} \Lambda_r \sigma_r(x)$$

we have

$$\Lambda_r a = L\sigma_r, \quad \forall \ \ 1 \leq r \leq m \tag{63}$$

where as before,

$$L = \left(a, \frac{\partial}{\partial x} \right) + \frac{1}{2} \sum_{r=1}^{m} \sum_{k=1}^{d} \sum_{j=1}^{d} \sigma_{jr} \sigma_{kr} \frac{\partial^2}{\partial x_k \partial x_j}$$

Now, let us consider the stochastic Taylor expansion of X_t up to order 2:

$$\begin{aligned} X_{t,x}(t+h) &= x + \sum_{r=1}^{m} \sigma_r \Delta_h W_r(t) + ah + \sum_{r=1}^{m} \sum_{i=1}^{m} (\Lambda_i \sigma_r) \mathcal{I}_{i,r}(h) \\ &\quad + \sum_{r=1}^{m} (L\sigma_r) \int_t^{t+h} (\theta - t) dW_r(\theta) + \sum_{r=1}^{m} (\Lambda_r a) \mathcal{I}_{r,0}(h) \\ &\quad + \sum_{r=1}^{m} \sum_{i=1}^{m} \sum_{s=1}^{m} (\Lambda_s \Lambda_i \sigma_r) \mathcal{I}_{s,i,r}(h) \end{aligned}$$

$$+ La\frac{h^2}{2} \tag{64}$$

which can be written as

$$\begin{aligned} X_{t,x}(t+h) = x &+ \sum_{r=1}^{m} \sigma_r \Delta_h W_r(t) + ah + \sum_{i=1}^{m} \Lambda_i \sigma_i \left(\frac{(\Delta_h W_i)^2 - h}{2} \right) \\ &+ \sum_{1 \le r < i \le m} \Lambda_i \sigma_r \left[\mathcal{I}_{i,r}(h) + \mathcal{I}_{r,i}(h) \right] \\ &+ \sum_{r=1}^{m} L\sigma_r h \Delta_h W_r + \sum_{r=1}^{m} (\Lambda_r a - L\sigma_r) \int_t^{t+h} (W_r(\theta) - W_r(t)) d\theta \\ &+ \sum_{r=1}^{m} \sum_{i=1}^{m} \sum_{s=1}^{m} (\Lambda_s \Lambda_i \sigma_r) \mathcal{I}_{s,i,r}(h) + La\frac{h^2}{2} \end{aligned} \tag{65}$$

Notice that in the second line of (65) we used the commutativity condition of $\Lambda_i \sigma_r$ given by (62). Also, it is not difficult to see that

$$\int_t^{t+h} (W_i(\theta) - W_i(t)) dW_r(\theta) + \int_t^{t+h} (W_r(\theta) - W_r(t)) dW_i(\theta) = \Delta_h W_i \Delta_h W_r \tag{66}$$

Therefore substituting in (65) and using (63) we obtain

$$\begin{aligned} X_{t,x}(t+h) &= x + \sum_{r=1}^{m} \sigma_r \Delta_h W_r(t) + \sum_{r=1}^{m} (\Lambda_r \sigma_r) \frac{(\Delta_h W_r)^2}{2} \\ &\quad + \sum_{1 \le r < i \le m} \Delta_i \sigma_r \Delta_h W_i \Delta_h W_r + \sum_{r=1}^{m} L\sigma_r h \Delta_h W_r \\ &\quad + \text{higher order terms} \end{aligned} \tag{67}$$

Hence we see that there is no need to simulate the stochastic integrals $\int_t^{t+h} (W_i(\theta) - W_i(t)) dW_r(\theta)$ and $\int_t^{t+h} (W_r(\theta) - W_r(t)) d\theta$. Moreover, by computing the double stochastic integral which appears in (65) we should get terms which allow to cancel all factors of $''h\Delta_h W_r(t)''$ and so that the stochastic expansion is given in powers of the increments $\Delta_h W_r$ as it was done in the one dimensional case. Indeed, the author was able to prove it in 2-dimensions ($d = 2$) being once again lots of calculations from which there are not many things to say, except that it seems interesting to know the analogous of the identities (37) here in the multidimensional case.

4 Numerical Examples

In this section we give three different examples where the Milhstein scheme of order 2 and the Runge-Kutta scheme developed in this paper are compared against each other. The programs were elaborated in Matlab and they are accessible upon request. For each experiment it is necessary to simulate many trajectories of the Wiener process and for simplicity we take always $M = 100$, where M stands for the number of different realizations of the Wiener process. Of course, the programs are made in such a way that users could choose M as they wish. Two different errors are computed at each trial. The first one, named e_1 is just the L_2 error on $[0, T] \times \Omega$ between the exact and the approximate solution, we refer to it as the pathwise error. The second, named e_2, is the mean-square deviation error (see [18]), between the exact solution X_t and the approximate solution $\overline{x}_t$ at $t = T$, where T is the end point of the interval $[0, T]$. We estimate e_2 by taking the average of the errors committed for each path of the Wiener process. In other words, $e_2 = \left(\frac{1}{M} \sum_{i=1}^{M} |X_t^i - \overline{x}_T^i|^2 \right)^{1/2}$, where the superscript i indicates the error corresponding to the ith path of the Wiener process. Finally, h stands always for the size-step of the discretization.

Example 1.

The first of the examples we show is the following equation:

$$\begin{cases} dX_t = a^2 X_t(1 + X_t^2)dt + a(1 + X_t^2)dW_t = a(1 + X_t^2) \circ dW_t \\ X(0) = 0 \end{cases}$$

where a is any real number.

The exact solution is $X_t = \tan(aW_t + \arctan X_0)$ and the results are the following:

Schemes	errors	a=.5, T=.2 h=.1	a=.5, T=.2 h=.01	a=.2, T=1 h=.1	a=.2, T=1 h=.01
STE	e_1	3.4×10^{-9}	$O(10^{-10})$	1.7×10^{-9}	$O(10^{-10})$
of order 2	e_2	1.2×10^{-4}	9.8×10^{-6}	6.9×10^{-5}	7×10^{-6}
Runge-Kutta	e_1	$O(10^{-10})$	$O(10^{-10})$	$O(10^{-10})$	$O(10^{-10})$
of order 2	e_2	6.6×10^{-7}	1.2×10^{-8}	5.8×10^{-7}	8.8×10^{-9}

There we observe some small difference in the pathwise error, but, Runge-Kutta shows considerably better results when estimating e_2.

Example 2.
The next example we will consider is the equation:

$$\begin{cases} dX_t = -\frac{1}{2}a^2 X_t dt - a\sqrt{1-X_t^2}dW_t = -a\sqrt{1-X_t^2}\circ dW_t \\ X(0) = 0 \end{cases}$$

where again a is any real number.
The exact solution is given by $X_t = \cos\left(aW_t + \arccos X_0\right)$ and we obtain:

Schemes	errors	a=.5, T=.2 h=.1	a=.5, T=.2 h=.01	a=.2, T=1 h=.1	a=.2, T=1 h=.01
STE	e_1	$O(10^{-10})$	$O(10^{-10})$	$O(10^{-10})$	$O(10^{-10})$
of order2	e_2	7.2×10^{-7}	1.4×10^{-8}	6.8×10^{-7}	1×10^{-8}
Runge-Kutta	e_1	$O(10^{-10})$	$O(10^{-10})$	$O(10^{-10})$	$O(10^{-10})$
of order 2	e_2	3.7×10^{-7}	6.2×10^{-9}	3.4×10^{-7}	4.9×10^{-9}

In this case, we do not see much difference in the accuracy of both schemes, in fact, they are of the same order for $h = .1$, but Runge-Kutta seems to be of one degree better for e_2 when $h = .01$. However, in order to implement the Milhstein scheme many different derivatives were needed, thus, its implementation is more tedious and longer than using Runge-Kutta.

Example 3.
Finally, our third example is the general linear SDE:

$$\begin{cases} dX_t = aX_t dt + bX_t dW_t = (a - \frac{1}{2}b^2)X_t dt + bX_t \circ dW_t \\ X(0) = 1 \end{cases}$$

where a and b are any real numbers.
The exact solution is given by $X_t = X_0 \exp\left((a - \frac{1}{2}b^2)W_t + bW_t\right)$.
Of course, for most choices of a and b, drift and diffusion terms respectively, this equation will not satisfy the condition $a = \frac{1}{2}\sigma\sigma'$, however, we observe that the Runge-Kutta scheme performs much better than Milhstein's scheme (i.e. STE of order 1).

Schemes	errors	a=.2, b=1 h=.1	a=.2, b=1 h=.01	a=1, b=1 h=.1	a=1, b=1 h=.01
STE	e_1	3.3×10^{-5}	7.4×10^{-8}	1.2×10^{-3}	1.8×10^{-6}
of order 1	e_2	6.7×10^{-3}	4.4×10^{-4}	5.9×10^{-2}	2.1×10^{-3}
Runge-Kutta	e_1	4.1×10^{-6}	8×10^{-10}	2.1×10^{-5}	3.3×10^{-9}
of order 2	e_2	2.2×10^{-3}	4.4×10^{-5}	8.4×10^{-3}	1.2×10^{-4}

In the opinion of the author, this is a very interesting example because even though the equation is not of the type studied in this paper, still we see how Runge-Kutta performs way better that the STE scheme of order 1. In fact, empirically, it seems to behave as a scheme of order 2 even though there is not theory that supports this fact. Below, we show a plot of the errors obtained when performing the STE scheme of order 1, Runge-Kutta scheme, and Runge-Kutta with Splitting-Up (see [1] and [2]), to the general linear equation above when $a = 1$ and $b = 1$. We plot $-\ln(error)$ versus $-\ln(h)$, where h is the discretization step, to see, at least in a empirical way, what happens with the order of convergence of these schemes.

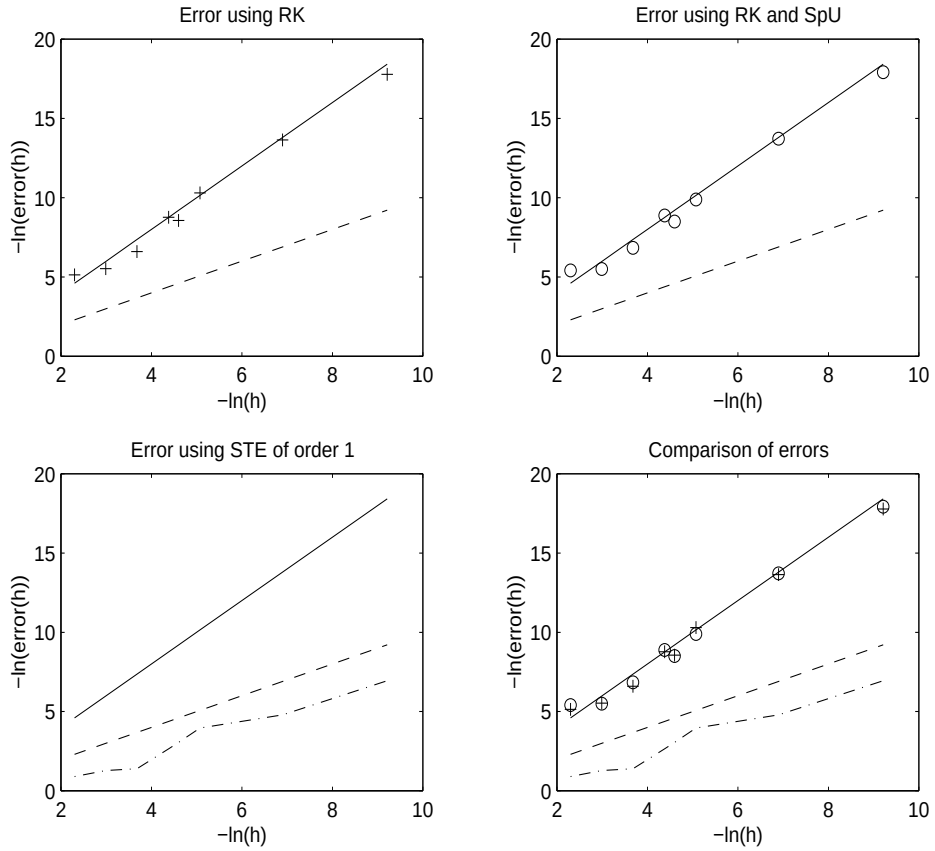

Figure 1: Errors using RK, RKSpU, and STE1. 'RK +', 'RKSpU o', and 'STE of order 1 -.' .

PART II

5 On The Approximation of Expected Values Of Functionals Of X_T.

Once again we consider the solution of:

$$\begin{cases} dX_t = a(X_t)dt + \sigma(X_t)dW_t \\ X(0) = X_0 \end{cases} \tag{68}$$

where $t \in [0,T]$, $a(x)$ and $\sigma(x)$ are as defined in (24). But, this time we are interested in computing $\mathbf{E}f(X_T)$ where $f(x)$ is a sufficiently smooth function.
This problem has been studied in [19],[24],[25],and [28] among others. They state the problem as follows:
"*Given (X_t) the solution of a SDE, and f any "regular" function of growth at most polynomial, develop a good method which can be performed efficiently on computers to approximate:* $\mathbf{E}f(X_T)$."
In view of solving this problem we would like to simulate M realizations of X_t. Unfortunately, in most cases, we do not know the law of X_t. Hence we discretize (68) to get an approximate solution of it, $\overline{X}_t$, of which we simulate M independent realizations, $\overline{X}_t^{(i)}, 1 \leq i \leq M$,and we do compute:

$$\frac{1}{M}\sum_{i=1}^{M} f(\overline{X}_T^{(i)}) \tag{69}$$

The error of the method satisfies the following inequality:

$$|\mathbf{E}f(X_T) - \frac{1}{M}\sum_{i=1}^{M} f(\overline{X}_T^{(i)})| \leq$$

$$|\mathbf{E}f(X_T) - \mathbf{E}f(\overline{X}_T)| + |\mathbf{E}f(\overline{X}_T) - \frac{1}{M}\sum_{i=1}^{M} f(\overline{X}_T^{(i)})| = \varepsilon_1 + \varepsilon_2$$

To minimize ε_2 one only needs to choose M large because of the law of Large Numbers.
The error ε_1 depends on the step of the discretization of $[0,T]$ which is used and denoted by h.
So the goals are to produce an algorithm of discretization (numerical scheme) which has the following properties:
1) the error ε_1 satisfies the inequality:

$$\varepsilon_1 \leq Ch^r \tag{70}$$

where C is a constant and r is large enough.
2) in view to minimize the computation time, (remember that we must get M independent trajectories of $\overline{X}_t$), the complexity of the algorithm is reduced.
3) in the same view, the laws of the random variables occurring in the algorithm are simple to simulate.

6 Previous results

Let $h = \frac{T}{N}$, $M \in \mathbf{N}$, be the step of the discretization of $[0, T]$ given by: $\{0 = t_0, h = t_1, \ldots, T = t_N\}$.

Let us consider the following schemes:

$$X_{k+1} = X_k + a_k h + \sigma_k \Delta_k W(h) \tag{71}$$

$$\overline{X}^2_{k+1} = \overline{X}^2_k + a_k h + \sigma_k \Delta_k W(h) + \Lambda\sigma_k \left(\frac{(\Delta_k W(h))^2 - h}{2} \right) \tag{72}$$

$$\begin{aligned} \overline{X}^{mc}_{k+1} &= \overline{X}^{mc}_k + \sigma_k R_{k+1}\sqrt{h} \\ &\quad + \{a_k - \frac{1}{2}\Lambda\sigma_k\}h + \frac{1}{2}\Lambda\sigma_k (R_{k+1})^2 h \\ &\quad + \frac{1}{2}\{L\sigma_k + \Lambda a_k\} R_{k+1} h^{\frac{3}{2}} \\ &\quad + \frac{1}{2}\{L a_k\} h^2 \end{aligned} \tag{73}$$

where a_k,σ_k are the values of the coefficients $b(x)$ and $\sigma(x)$ at $\overline{X}^{(.)}_k$,$\Delta_k W(h) = W(t_{k+1}) - W(t_k)$, L, Λ as defined before in section 3.3 and $\{R_k\}$ is as follows:
a) $R_o = 0$;
b) R_{k+1} is a random variable independent of the σ-field generated by $\{R_j, 0 \leq j \leq k\}$.
c) R_{k+1} satisfies:

$$\begin{aligned} \mathbf{E}(R_{k+1}) &= \mathbf{E}(R_{k+1})^3 = \mathbf{E}(R_{k=1})^5 = 0 \\ \mathbf{E}(R_{k+1})^2 &= 1, \ \ \mathbf{E}(R_{k+1})^4 = 3 \\ \mathbf{E}(R_{k+1})^6 &< \infty \end{aligned}$$

Remark: It is clear that $R_{k+1}\sqrt{h}$ plays the part of $W(t_{k+1}) - W(t_k)$.

The first scheme (71) is the usual Euler scheme, (72) is the Milshtein's scheme

and (73) is known as "The Monte-Carlo Scheme" (see [24]). It is known that the Euler scheme has mean-square order of convergence equal to $h^{\frac{1}{2}}$ whereas (72) has mean-square order equal to h (see [18]). Also, it is not difficult to see that (73) has mean-square order to h too. On the other hand, it is proved in [19],[24],and [28] that (71) and (72) satisfy:

$$|\mathbf{E}f(X_T) - \mathbf{E}f(\overline{X}_T)| \leq Ch^r \tag{74}$$

for $r = 1$ whereas (73) satisfies the same inequality but with $r = 2$. Actually, this Monte-Carlo scheme was constructed because having order of weak convergence equal to h is too low, so a better scheme was needed. However, this "better" scheme is of order h in the mean square error, thus in that sense is not any better than (72) which is much easier to implement on computers than (73) itself.
In the next section we show that for equations of the form: $dX_t = \sigma(X_t) \circ dW_t$ the RK4 scheme derived in Part I has order of convergence equal to h^2 in both senses, i.e., in the mean-square error and weak approximation too. Moreover, its implementation on computers is easier than (72) and (73), and even though the Euler scheme would be the easiest to implement it has the disadvantage that pathwise is not a very good approximation.

7 Accuracy of Weak Approximations for the Runge-Kutta scheme. Case: $\mathbf{a} = \frac{1}{2}\sigma\sigma'$.

We wish to prove the following:
For any given function f in a large subset of the space of continuous functions, it exists a real C not depending on h such that, if $(\overline{X}_t^h)$ is the approximate solution of (68) given by the Runge-Kutta scheme, piecewise constant on each interval $[t_k = kh, t_{k+1} = (k+1)h]$, then we have:

$$|\mathbf{E}f(X_T) - \mathbf{E}f(\overline{X}_T^{(h)})| \leq Ch^2 \tag{75}$$

The main tool of this proof is the Kolmogorov's backward equation (KBE, in short) associated to the diffusion process (X_t) and the function f:

$$\begin{cases} \frac{\partial v}{\partial t}(t,\cdot) + Lv(t,\cdot) = 0, & ,0 \leq t \leq T \\ v(T,\cdot) = f(\cdot) \end{cases} \tag{76}$$

where $L := a(\cdot)\frac{\partial}{\partial x}+\frac{1}{2}\sigma^2(\cdot)\frac{\partial^2}{\partial x^2}$. This tool was suggested to D.Talay by A.Bensoussan. The idea is use (76) to rewrite (75) as:

$$|\mathbf{E}v(T,X_T) - \mathbf{E}v(T,\overline{X}^h_T)| \leq Ch^2 \tag{77}$$

Before giving the main lines of the proof, we state 3 lemmas.

Lemma 1 *Let us suppose:*
*(**H1**) b, σ are continuous functions and their derivatives up to order 6 are continuous bounded functions.*
*(**H2**) f and its derivatives up to order 6 are continuous functions such that:*

$$\exists M_1 \in \mathbf{R}, \exists r \in \mathbf{N}/\forall x \in \mathbf{R} : |f^{(i)}(x)| \leq M_1(|x|^r + 1), \quad , 0 \leq i \leq 6 \tag{78}$$

Then the PDE (KBE) has a solution; the solution we exhibit has the following properties:
(i) v is 6 times continuously differentiable with respect to x; moreover: $\exists s \in \mathbf{N}, \forall T, \exists M_2 \in \mathbf{R}, \forall t \in [0,T]$,

$$|\frac{\partial^i v(t,x)}{\partial x^i}| \leq M_2(|x|^s + 1), \quad 0 \leq i \leq 6 \tag{79}$$

(ii) v and its derivatives with respect to x are continuous in (t,x).

Also,

Lemma 2 *Under the assumption (H1) of Lemma 1 one gets:*
$\forall T, \forall n \in \mathbf{N}, \exists M_4 \in \mathbf{R}/\forall h = \frac{T}{N}, N \in \mathbf{N}, \forall k = 0,1,\ldots,\frac{T}{h}$

$$\mathbf{E}|\overline{X}^h_k|^n \leq M_4 \tag{80}$$

and finally, we restrict ourselves to the case $a = \frac{1}{2}\sigma\sigma'$, then we have:

Lemma 3 *The process $(\overline{X}^h_k)_k$ defined by the RK4 scheme given in (43) satisfies the following:*

$$\begin{aligned}
(i) \quad \mathbf{E}(\overline{X}^h_{k+1} - \overline{X}^h_k) &= \sigma\sigma'(\overline{X}^h_k)\frac{h}{2} + (\Lambda^3\sigma)(\overline{X}^h_k)\frac{h^2}{8} \\
(ii) \quad \mathbf{E}(\overline{X}^h_{k+1} - \overline{X}^h_k)^2 &= \sigma^2(\overline{X}^h_k)h + (\Lambda\sigma(\overline{X}_k))^2\frac{3h^2}{4} + (\Lambda^2\sigma)(\overline{X}^h_k)h^2 + O(h^3) \\
(iii) \quad \mathbf{E}(\overline{X}^h_{k+1} - \overline{X}^h_k)^i &= O(h^3), \forall i \geq 3
\end{aligned}$$

For a prove of the first two lemmas see [24], a straightforward computation leads to lemma 3.
Moreover, it is shown in [13] that a solution of the KBE is given by: $v(t,x) = \mathbf{E}f(X_T^{t,x})$ where the process $X_\theta^{t,x}$ is defined by:

$$X_\theta^{t,x} = x + \int_t^\theta a(X_u^{t,x})du + \int_t^\theta \sigma(X_u^{t,x})dW(u) \tag{81}$$

Now we can state the main result of this paragraph:

Theorem 3 *Let us suppose* **(H1)**, **(H2)** *of lemma 1 and that* $a = \frac{1}{2}\sigma\sigma'$. *Then the Runge-Kutta scheme satisfies the estimation:*
$\forall T, \exists C \in \mathbf{R}/\forall h = \frac{T}{N}, N \in \mathbf{N},$

$$|\mathbf{E}f(X_T) - \mathbf{E}f(\overline{X}_T^h)| \leq Ch^2 \tag{82}$$

Proof:
The KBE equation can be written as:

$$\left(\frac{\partial}{\partial t} + L\right) v = 0 \tag{83}$$

therefore,

$$\left(\frac{\partial}{\partial t} + L\right)^2 v = \left(\frac{\partial^2}{\partial t^2} + 2\frac{\partial}{\partial t}L + L^2\right) v = 0 \tag{84}$$

Moreover, by the mean value theorem applied to $v(t,x)$ we have:

$$\begin{aligned} v(t_{k+1}, \overline{X}_{k+1}^h) &= v(t_k, \overline{X}_k^h) + \frac{\partial v}{\partial t}h + \frac{\partial v}{\partial x}(\overline{X}_{k+1}^h - \overline{X}_k^h) \\ &+ \frac{1}{2!}\left[\frac{\partial^2 v}{\partial t^2}h^2 + \frac{\partial^2 v}{\partial x^2}(\overline{X}_{k+1}^h - \overline{X}_k^h)^2 + 2\frac{\partial^2 v}{\partial x \partial t}h(\overline{X}_{k+1}^h - \overline{X}_k^h)\right] \\ &+ \frac{1}{3!}\left[3\frac{\partial^3 v}{\partial t \partial x^2}h(\overline{X}_{k+1}^h - \overline{X}_k^h)^2 + \text{other derivatives of order 3}\right] \\ &+ \text{higher order derivatives} \end{aligned}$$

where all derivatives in the right-hand side are evaluated at the point $(t_k, \overline{X}_k^h)$. We also remark that the other derivatives of order 3 omitted in the above expression

are such that after expectation gives terms of order h^3.
We introduce the following notation: for two integrable random variables X and Y: we write $X \doteq Y$ instead of $\mathbf{E}X = \mathbf{E}Y$.
Now, taking expectations in the above equality and using lemma 3 we obtain:

$$\begin{aligned} v(t_{k+1}, \overline{X}^h_{k+1}) \doteq v(t_k, \overline{X}^h_k) &+ \left(\frac{\partial v}{\partial t} + \frac{1}{2}\sigma\sigma' \frac{\partial v}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 v}{\partial x^2} \right) h \\ &+ \frac{1}{2} \left(\frac{\partial^2 v}{\partial t^2} + \frac{1}{4}\Lambda^3(\sigma)\frac{\partial v}{\partial x} + \left(\frac{3}{4}(\Lambda\sigma)^2 + \sigma(\Lambda^2\sigma) \right) \frac{\partial^2 v}{\partial x^2} + (\Lambda\sigma)\frac{\partial^2 v}{\partial x \partial t} \right) h^2 \\ &+ \left(\sigma^2 \frac{\partial^3 v}{\partial t \partial x^2} \right) h^2 + O(h^3) \end{aligned} \tag{85}$$

On the other hand, we have that in the general case (83) becomes:

$$\frac{\partial v}{\partial t} + a(x)\frac{\partial v}{\partial x} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} = 0 \tag{86}$$

and direct computations shows that (84) is:

$$\begin{aligned} \frac{\partial^2 v}{\partial t^2} + 2\frac{\partial (Lv)}{\partial t} + (L^2)v &= \frac{\partial^2 v}{\partial t^2} + \left(a(x)a'(x) + \frac{1}{2}\sigma^2 a'' \right) \frac{\partial v}{\partial x} \\ &+ \left(a^2(x) + a(x)\sigma(x)\sigma'(x) + \sigma^2(x)a'(x) + \frac{1}{2}\sigma^2(x)(\sigma'(x))^2 \right) \frac{\partial^2 v}{\partial x^2} \\ &+ \left(\sigma^3(x)\frac{(\sigma'')}{2} \right) \frac{\partial^2 v}{\partial x^2} \\ &+ 2a(x)\frac{\partial^2 v}{\partial t \partial x} + \sigma^2(x)\frac{\partial^3 v}{\partial t \partial x^2} \\ &= 0 \end{aligned} \tag{87}$$

Now, under the assumption that $a(x) = \frac{1}{2}\sigma\sigma'(x)$ and using the identities given in (37) from section 3.3 these equations become:

$$\frac{\partial v}{\partial t} + \frac{1}{2}\sigma\sigma'(x)\frac{\partial v}{\partial x} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} = 0 \tag{88}$$

and

$$\frac{\partial^2 v}{\partial t^2} + 2\frac{\partial (Lv)}{\partial t} + (L^2)v \;\; = \;\; \frac{\partial^2 v}{\partial t^2} + \frac{1}{4}\Lambda^3(\sigma)\frac{\partial v}{\partial x}$$

$$
\begin{aligned}
&+ \quad \left(\frac{3}{4}(\Lambda\sigma)^2 + \sigma(\Lambda^2(\sigma)\right)\frac{\partial^2 v}{\partial x^2} \\
&+ \quad (\Lambda\sigma)\frac{\partial^2 v}{\partial t \partial x} + \sigma^2(x)\frac{\partial^3 v}{\partial t \partial x^2} \\
&= \quad 0 \qquad (89)
\end{aligned}
$$

Then, substituting (88) and (89) into equation (85) we get:

$$
\begin{aligned}
v(t_{k+1}, \overline{X}^h_{k+1}) \quad &\doteq \quad v(t_k, \overline{X}^h_k) + \left(\frac{\partial}{\partial t} + L\right) v(t_k, \overline{X}^h_k)h \\
&\quad + \left(\frac{\partial}{\partial t} + L\right)^2 v(t_k, \overline{X}^h_k)\frac{h^2}{2} + O(h^3)
\end{aligned}
$$

so we have:

$$
\begin{aligned}
v(t_{k+1}, \overline{X}^h_{k+1}) \quad &\doteq v(t_k, \overline{X}^h_k) + O(h^3), \quad \text{since } v \text{ is a solution of KBE} \\
&\doteq v(0, \overline{X}^h_0) + O(h^2), \quad \text{by recurrence on } k \\
&\doteq v(0, X_0) + O(h^2), \quad \text{because } \overline{X}_0 = X_0 \\
&\doteq v(T, X_T) + O(h^2), \quad \text{by definition of } v
\end{aligned}
$$

Therefore it follows that:

$$
\mathbf{E}f(X_T) - \mathbf{E}f(\overline{X}^h_T) = e_2(T)h^2 + O(h^3) \qquad (90)
$$

where $e_2(T)$ is a constant that does not depend on h. Moreover, now that we know (90) we could use the *Romberg extrapolation* (see [26] and [28]) between values corresponding to two different step-sizes. More precisely, consider the following approximation:

$$
Z^h_T = \frac{4}{3}\mathbf{E}f(\overline{X}^{h/2}_t) - \frac{1}{3}\mathbf{E}f(\overline{X}^h_T) \qquad (91)
$$

then

$$
\mathbf{E}f(\overline{X}^h_T) - Z^h_T = O(h^3) \qquad (92)
$$

The above expressions are easy to check from the error expansion (90).
Thus, we get a scheme of precision of order h^3 from a result given by a second-order scheme.
Remark: This extrapolation technique has been generalized by Hoffman, Kloeden and Platen.

8 Conclusions

We found a scheme whose order of precision is h^2 in both senses: mean-square convergence and weak approximation. This actually improves some results, for certain class of SDEs (no drift in the Stratonovich sense and commutative noise), given in [19], [24], [28] among others, where they give also a scheme of order of precision h^2 for approximations of $\mathbf{E}f(X_T)$ but they are only of order h in the mean-square error.

The author is currently working in the general case. Two different approaches have been attempted. The first one is to write $dX_t = a(X_t)dt + \sigma(X_t)dW_t$ in the Stratonovich sense, in other words, $dX_t = (a - \frac{1}{2}\sigma\sigma')(X_t)dt + \sigma(X_t) \circ dW_t$, then use the "Splitting-Up" method (see [1],[2]). So far, the author has proved in the one dimensional case that in this way we obtain a scheme of order h instead, so it seems like splitting-up takes away one degree of precision. However, as mentioned in the introduction, the result given by Cameron-Clark in [7] tell us that order 1 is the best we can achieve in the general case for approximations taking only increments of the Wiener process at the partition points thus we are actually giving another scheme of order 1 for general SDEs.

Another way could be to use a change of measure and then apply Girsanov's theorem, however, the new Wiener process after the change of measure depends on the paths of X_t, therefore, it is not clear if it really helps, this also is within the future research of the author.

APPENDIX

A Calculations Of Some Stochastic Integrals.

Here, we present some of the calculations involving stochastic integrals used through this paper, specifically when writing down the scheme obtained by using Runge-Kutta in the stochastic case. Now, in section 3.3 of Part I we claimed that:

$$\mathcal{I}_{0,1,1}(h) + \mathcal{I}_{1,1,0}(h) + \mathcal{I}_{1,0,1}(h) = h\left(\frac{(\Delta_h W_t)^2 - h}{2}\right) \tag{93}$$

where

$$\mathcal{I}_{i_1,\ldots,i_j}(h) = \int_t^{t+h} dW_{i_j}(\theta) \int_t^{\theta} dW_{i_{j-1}}(\theta_1) \int_t^{\theta_1} \ldots \int_t^{\theta_{j-2}} dW_{i_1}(\theta_{j-1}) \tag{94}$$

where the indices $i_1, \ldots, i_j$ can take either the value 0 or 1 in which cases $dW_0(\theta) = d\theta$ and $dW_1(\theta) = dW(\theta)$ respectively.
In order to prove (93) we need to introduce the *Hermite polynomials*:

$$H_n(t,x) = \frac{(-t)^n}{n!} \exp\left(\frac{x^2}{2t}\right) \frac{\partial^n}{\partial x^n} \exp\left(-\frac{x^2}{2t}\right), \qquad n \geq 0 \tag{95}$$

Consider $H_n(t, W(t))$ and compute, using Itô's formula and some properties of the Hermite polynomials we get:

$$dH_n = \left(\frac{\partial H_n}{\partial t} + \frac{1}{2}\frac{\partial^2 H_n}{\partial x^2}\right) dt + \frac{\partial H_n}{\partial x} dW = H_{n-1} dW \tag{96}$$

whence

$$\begin{aligned} H_n(\alpha s, \beta W(s)) - H_n(\alpha t, \beta W(t) &= \frac{1}{2}(\beta^2 - \alpha) \int_t^s H_{n-2}(\alpha\theta, \beta W(\theta)) d\theta \\ &\quad + \beta \int_t^s H_{n-1}(\alpha\theta, \beta W(\theta)) dW(\theta) \end{aligned} \tag{97}$$

(for more details on these computations see [18]). Writing down (97) for $n = 2, \ldots, 7$ in succession and equating identical powers of α and β, we obtain the

required relations between the $n/2$-integrals. For $n = 2$ and powers of β^2 (97) gives:

$$\int_t^{t+h} W(\theta)dW(\theta) = \frac{(W^2(t+h) - W^2(t)) - h}{2} \tag{98}$$

Similarly, for $n = 3$, powers of β^3 and $\alpha\beta$ we obtain respectively:

$$\begin{aligned} \int_t^{t+h} W^2(\theta)dW(\theta) &= \frac{W^3(t+h) - W^3(t)}{3} - \int_t^{t+h} W(\theta)d\theta \\ \int_t^{t+h} \theta dW(\theta) &= (t+h)W(t+h) - tW(t) - \int_t^{t+h} W(\theta)\theta \end{aligned} \tag{99}$$

and for $n = 4$, powers of β^4 and $\beta^2\alpha$ we get:

$$\begin{aligned} \int_t^{t+h} W^3(\theta)dW(\theta) &= \frac{W^4(t+h) - W^4(t)}{4} - \frac{3}{2}\int_t^{t+h} W^2(\theta)d\theta \\ \int_t^{t+h} \theta W(\theta)dW(\theta) &= -\frac{(t+h)^2 - t^2}{4} + \frac{W^2(t+h) - W^2(t)t}{2} \\ &\quad -\frac{1}{2}\int_t^{t+h} W^2(\theta)d\theta \end{aligned} \tag{100}$$

Now, using these relations and straightforward computations (very tedious though!), leads to:

$$\begin{aligned} \mathcal{I}_{0,1,1}(h) &= -\frac{h^2}{4} + \frac{W^2(t+h)h}{2} - W(t+h)\int_t^{t+h} W(\theta)d\theta + \frac{1}{2}\int_t^{t+h} W^2(\theta)d\theta \\ \mathcal{I}_{1,1,0}(h) &= \frac{1}{2}\int_t^{t+h} W^2(\theta)d\theta - \frac{h^2}{4} + \frac{W^2(t)h}{2} - W(t)\int_t^{t+h} W(\theta)d\theta \\ \mathcal{I}_{1,0,1}(h) &= W(t+h)\int_t^{t+h} W(\theta)d\theta - W(t+h)W(t)h + W(t)\int_t^{t+h} W(\theta)d\theta - \int_t^{t+h} W^2(\theta)d\theta \end{aligned}$$

Therefore (93) follows easily.

It is worth mentioning that to get the above results we had to change the order integration of certain integrals, for instance, when computing $\mathcal{I}_{0,1,1}(h)$ we faced the following integral: $\int_t^{t+h}\int_t^{\theta} W(\theta_1)d\theta_1 dW(\theta)$ which by changing the order of integration becomes: $\int_t^{t+h}\int_{\theta_1}^{t+h} W(\theta_1)dW(\theta)d\theta_1$ and we do know how to solve the last one.

Similar situation is met when dealing with $\mathcal{I}_{1,0,1}(h)$.

References

[1] Allan Bensoussan, R.Glowinski, and A.Rascanu. *Approximation of the Zakai Equation By the Splitting Up Method.* SIAM journal, Control and Optimization, vol.28, No 6, 1990.

[2] Allan Bensoussan, R.Glowinski, and A.Rascanu.it Approximation Of Some Stochastic Differential Equations by The Splitting Up Method. Appl. Math. Optim, 25, 1992.

[3] Burden, Richard and Faires J.Douglas. *Numerical Analysis.* 5th edition, PWS-Kent Publishing Company, 1993.

[4] Burrage, K. and Burrage, P.M.. *High Strong Order Explicit Runge-Kutta Methods For Stochastic Ordinary Differential Equations.* Appl.Num.Math., 22, 1996, p.81-101.

[5] Burrage, K. and Burrage, P.M.. *General Order Conditions For Stochastic Runge-Kutta Methods For Both Commuting And Non-Commuting Stochastic Ordinary Equations.* Appl.Num.Math., 28, 1998, p.161-177..

[6] Castell, Fabienne and Jessica Gaines. *The Ordinary Differential Equation Approach To Asymptotically Efficient Schemes For Solution Of Stochastic Differential Equations*, Ann. Inst. Henri Poincare, Vol.32, 1996, p.231-250.

[7] J.M.C Clark, R.J. Cameron *The Maximum Rate of Convergence Of Discrete Approximations For Stochastic Differential Equations*, Lectures Notes in Control and Information Sciences, 25, 1978.

[8] N, Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*, Second edition, North-Holland Publishing Company, 1989.

[9] Doss, Halim. *Liens entre équations différentielles stochastiques et ordinaires*, Ann. Inst. Henri Poincaré,Vol. XIII, no 2, 1977, p.99-125.

[10] J.G.Gaines. *Numerical Experiments with S(P)DE's*,

[11] Gear, C.William. *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall Inc, NJ, 1971.

[12] A.Germany, M.Piccioni. *Semi-discretization of Stochastic Partial Differential Equations in R^d by a finite-element technique*, Stochastics, Vol.23, pp.131-148, 1988.

[13] Gikhman, I. and Skorokhod, A. *Introduction à la théorie des processus aléatoires*, Ed. Mir. 1980.

[14] Kloeden, P.E. and Platen, E. *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1992.

[15] Kloeden, P.E, Platen, E. and Schurz, H. *Numerical Solution of SDE through Computer Experiments*,1997.

[16] Krylov N.V. *Introduction to the Theory of Diffusion Processes*, American Mathematical Society, 1995.

[17] Francois LeGland. *Splitting-Up Approximation For SPDE's and SDE's with Application To Nonlinear Filtering*.

[18] Milshtein G.N., *Numerical Integration of Stochastic Differential Equations*, Kluwer Academic Publishers, 1995.

[19] Milshtein G.N., it A Method Of Second-Order Accuracy Integration Of Stochastic Differential Equations 1976.

[20] Nigel.J.Newton. *Asymptotically Efficient Runge-Kutta Methods For A Class Of Itô And Stratonovich Equations*. SIAM J. Appl. Math, Vol.51, No.2, pp. 542-567.

[21] Nigel.J.Newton. *An Asymptotically Efficient Difference Formula For Solving Stochastic Differential Equations*. Stochastics, Vol.19, pp. 175-206, 1986.

[22] Schurz, Henri. *Stability, Stationarity, And Boundedness Of Some Implicit Numerical Methods For Stochastic Differential Equations And Applications*. Logos-Verlag, Berlin, 1997.

[23] Strikwerda, John C. *Finite Difference Schemes and Partial Differential Equations*. Wadsworth and Brooks, 1989.

[24] Talay, Denis. *Efficient Numerical Schemes For The Approximation Of Expectations Of Functionals Of The Solution Of A S.D.E, And Applications*,Lecture Notes in Control and Information Sciences, Filtering and Control of Random Processes, 61, Proceedings of the E.N.S.T.-C.N.E.T. Colloquium Paris, France, February 23-24, 1983.

[25] Talay, Denis. *Discrétisation D'une Équation Différentielle Stochastique Et Calcul Approché D'Epérances De Fonctionnelles De La Slution*, Mathematical Modeling and Numerical Analysis, Vol. 20, n^o 1, 1986, p. 141 à 179.

[26] Talay, Denis. *Simulation of Stochastic Differential Systems*,Probabilistic Methods in Applied Physics, P. Krée, W. Wedig (Eds.), Springer, 1995.

[27] Talay, Denis. *Resolution Trajectorielle et Analyse Numerique des Equations Differentielles Stochastiques*, Stochastics, 1983, Vol.9, pp.275-306.

[28] Talay, Denis and Tubaro, Luciani.*Expansion of the global error for numerical solving Stochastic Differential Equations*, Stochastic Analysis and Applications, 8(4), 483-509, 1990.

[29] Wolfgang Wagner and Eckhard Platen. *Approximation Of Itô Integral Equations*. Als Preprint gedruckt, Berlin, 1978.