Neural, Parallel, and Scientific Computations, **26**, No. 3 (2018), 279-296 ISSN: 1061-5369

THE IMPLEMENTATION RESULT OF APPROXIMATE COUPLING IN THREE-DIMENSIONAL SDE WITH INVERTIBLE DIFFUSION

YOUSEF ALNAFISAH

Mathematics Department
College of Science, Qassim University
P.O. Box 6644, Buraydah 51452, SAUDI ARABIA

ABSTRACT: We study a new method for the strong approximate solution of stochastic differential equations using coupling and we prove by the Matlab implementation an order one error bounds for Davie's scheme [5] in section (8) for the L^p space assuming the invertibility of the diffusion matrix. We apply a particular three-dimensional invertible SDEs to show the convergence result.

Key Words: strong approximate solution, stochastic differential equations, Davie's scheme, Matlab

Received: August 6, 2018 ; **Accepted:** October 11, 2018 ;
Published: November 14, 2018 **doi:** 10.12732/npsc.v26i3.5
Dynamic Publishers, Inc., Acad. Publishers, Ltd. https://acadsol.eu/npsc

1. INTRODUCTION

In this study we investigate the new method developed by Davie [5] which uses coupling and gives order one for the strong convergence for stochastic differential equations (SDEs). There are many numerical methods for solving SDEs. P.E.Kloeden and E.Platen [6] have described a method based on the stochastic Taylor series expansion but the major difficulty with this approach is that the double stochastic integrals cannot be so easily expressed in terms of simpler stochastic integrals when the Wiener process is multi-dimensional. In the multi-dimensional case the Fourier series expansion of Wiener process has been used to represent the double integrals by [6], [11] and [10] but we need to generate many random variables each time therefore it takes a lot of time to compute and also it is hard to extend to higher order. We will see

in this study a modified interpretation for the normal random variables generated in the Taylor expansion. This method will give order one convergence under a non-degeneracy condition for the diffusion term. In standard methods such as Milstein we generate the approximations for the Taylor expansion terms separately. In the coupling method we will generate the approximation for the Taylor expansion as a combination of random variables. The modification is by replacing the iterated integrals by different random variables but with a good approximation in distribution. Then we will obtain a random vector from the linear term which is a good approximation in distribution to the original Taylor expansion. There have been many studies using coupling for the numerical solution of Stochastic differential equations. In [12] Kanagawa investigate the rate of convergence in terms of two probability metrics between approximate solutions with i.i.d random variables. Rachev and Ruschendorff [8] in volume 2 developed Kanagawa's method by using the Komlós, Major and Tusnády theorem in [7]. In [13] Fournier uses the quadratic Vaserstein distance for the approximation of the Euler scheme and the results of Rio [14] which gives a very precise rate of convergence for the central limit theorem in Vaserstein distance. Also Rio in [21] continues his research in [14] for the Vaserstein bound to give precise bound estimates. Under uniform ellipticity, Alfonsi, Jourdain and Kohatsu-Higa [3] and [4] have studied the Vaserstein bound for Euler method and they have proved an $O(h^{(\frac{2}{3}-\epsilon)})$ for one-dimensional diffusion process where h is the step-size and then they generalize the result to SDEs of any dimension with $O(h\sqrt{\log(\frac{1}{h})})$ bound when the coefficients are time-homogeneous. Cruzeiro, Malliavin and Thalmaier [15] get an order one method and under the non-degeneracy they construct a modified Milstein scheme which obtains an order one for the strong approximation. Charlbonneau, Svyrydov and Tupper [16] investigate the Vaserstein bound [9] by using the weak convergence and Strassen- Dudley theorem. Convergence of an approximation to a strong solution on a given probability space was established by Gyöngy and Krylov in [17] using coupling. Davie in [22] applied the Vaserstein bound to solutions of vector SDEs and uses the Komlós, Major and Tusnády theorem to get order one approximation under a non-degeneracy assumption. In this paper we will give the proof of an order one convergence for Davie [5] method by using Matlab implementation for a specific SDEs. The proof will be in L^p space using three-dimensional SDEs. The rest of this paper is organized as follows. Section 2 reviews some results concerning SDE and introduce Davie [5] method. Section 3 presents the idea of bounds using two-level coupling. In section 4 we look at and explain the method of the approximate coupling for general d. In the last section we give numerical example to the show the convergence behavior for 3-dimensional SDE.

2. STOCHASTIC DIFFERENTIAL EQUATIONS(SDES)

Definition: let $\{W(t)\}_{t\geq 0}$ be a d-dimensional standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, $a = a(t,x)$ be a d-dimensional vector function(called *drift* coefficient) and $b = b(t,x)$ a $d \times d$-matrix function(called *diffusion* coefficient). Stochastic processes $X = X(t)$, where $t \in [0,T]$, can be described by *stochastic differential equations*

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t). \tag{2.1}$$

Let the initial condition $X(0) = x$ be an $\mathcal{F}_0$-measurable random vector in $\mathbb{R}^d$. An $\mathcal{F}_t$-adapted stochastic process $X = (X(t))_{t\geq 0}$ is called a solution of equation (2.1) if

$$X(t) = X(0) + \int_0^t a(s, X(s))ds + \int_0^t b(s, X(s))dW(s), \tag{2.2}$$

is satisfied.

2.1. STRONG CONVERGENCE FOR SDES

Suppose that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given. In this probability space Ω is the set of continuous functions with the supremum metric on the interval $[0,T]$, $\mathcal{F}$ is the σ-algebra of Borel sets and $\mathbb{P}$ is the Wiener measure. We consider an approximate solution x_h of (2.1) which uses a subdivision of the interval $[0,T]$ into a finite number N of subintervals which we assume to be of length $h = \frac{T}{N}$. Also we assume the approximate solutions x_h is a random variable on Ω. Now we say that the discrete time approximation x_h with the step-size h converges strongly of order γ at time $T = Nh$ to the solution $X(t)$ if $E|x_h - X(T)|^p \leq Ch^{\gamma p}, h \in (0,1)$ where the strong convergence will be in L^p space and $X(T)$ is the solution to the stochastic differential equation. C is a positive constant and C independent of h.
Our method will give a strong approximation in the sense of this definition.

2.2. NUMERICAL METHOD FOR APPROXIMATING THE SDES

There are many numerical methods for solving stochastic differential equation, here we will mention two important schemes. The first one is the Euler-Maruyama scheme which will give strong order $\frac{1}{2}$ and the second one is the Milstein scheme which has an order one for the strong convergence. Suppose we have the stochastic differential equation.

$$dX_i(t) = a_i(t, X(t))dt + \sum_{k=1}^{d} b_{ik}(t, X(t))dW_k(t), \quad X_i(0) = X_i^{(0)} \tag{2.3}$$

where i =1,...,d on an interval $[0,T]$, for a d-dimensional vector $X(t)$, with a d-dimensional Brownian path $W(t)$. In order to approximate the solution, we assume $[0,T]$ is divided into N equal intervals of length $h = T/N$.

2.2.1. EULER-MARUYAMA SCHEME

The simplest numerical method for approximating the solution of stochastic differential equations is the stochastic Euler scheme (also called Euler Maruyama scheme) which utilizes only the first two terms of the Taylor expansion and it attains the strong convergence $\gamma = \frac{1}{2}$. Firstly, consider the Euler-Maruyama approximation scheme.

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum_{k=1}^{d} b_{ik}(jh, x^{(j)})\Delta W_k^{(j)}, \tag{2.4}$$

where $\Delta W_k^{(j)} = W_k((j+1)h) - W_k(jh)$ and our numerical approximation to $X(jh)$ will be denoted $x^{(j)}$.

2.2.2. THE MILSTEIN SCHEME

We shall now introduce the Milstein scheme which gives an order one strong Taylor scheme. We could obtain the Milstein scheme by adding the quadratic terms $\sum_{k,l=1}^{d} \rho_{ikl}(jh, x^{(j)})A_{kl}^{(j)}$, to Euler scheme which gives the following scheme

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum_{k=1}^{d} b_{ik}(jh, x^{(j)})\Delta W_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(jh, x^{(j)})A_{kl}^{(j)}, \tag{2.5}$$

where $\Delta W_k^{(j)} = W_k((j+1)h) - W_k(jh)$,
$A_{kl}^{(j)} = \int_{jh}^{(j+1)h} \{W_k(t) - W_k(jh)\} dW_l(t)$, and $\rho_{ikl}(t,x) = \sum_{m=1}^{q} b_{mk}(t,x)\frac{\partial b_{il}}{\partial x_m}(t,x)$.

The implementation of the Euler scheme is easy to do as only needs to generate the normal distribution for the standard Brownian motion $\Delta W_k^{(j)}$ but it is not easy to generate the integral $A_{kl}^{(j)}$ for the Milstein scheme when we have two or more dimensional SDEs. We need to mention some facts about the two-level approximation.

2.3. TWO-LEVEL APPROXIMATION

We need to generate the increments $\Delta W_k^{(j)}$ when we approximate the solution to (2.1) by using Euler or other schemes which we will explain later in this section, therefore Levy's construction of the Brownian motion will be used to simulate a sequence of approximations converge to the solution. That is

$$\Delta W_k^{(r,j)} = \Delta W_k^{(r+1,2j)} + \Delta W_k^{(r+1,2j+1)}, \tag{2.6}$$

where $r \in \mathbb{N}$ and $\Delta W_k^{(r,j)} = W_k((j+1)h^{(r)}) - W_k(jh^{(r)})$ with $h^{(r)} = \frac{T}{2^r}$.
We will call the two-level approximation in (2.6) *the trivial coupling*. We could generate the normal distribution in (2.6) for the increments for a given level r by firstly generating the increments in the LHS $\Delta W_k^{(r,j)}$ and then conditionally generating the increments in the RHS. We do the same process for each level $r+2$, $r+3$ and so on. After that we will get the Brownian path $W(t)$.

2.3.1. EMPIRICAL ESTIMATION OF THE ERROR OF A NUMERICAL METHOD

Because usually we do not know the solutions of the stochastic differential equation explicitly therefore we could not directly estimate the mean error $E|X(T) - x_h|$ which is the absolute value of the difference between the approximation solution x_h and the solution $X(T)$ of an SDE (2.1). Assume the approximate solution x_h converges to the solution $X(T)$ as we decrease the step-size and go to zero. Then we can estimate the order of convergence for a particular scheme by repeating R different independent simulations of sample paths. We will use the following estimator $\{\epsilon = \frac{1}{R}E(|x_{(r)} - \hat{x}_{(r)}|)\}$ for different approximation solutions $x_{(r)}$ and $\hat{x}_{(r)}$ for different range value of h. So for any numerical method if we have a bound for the error $E|x_h - x_{h/2}| \leq C_1 h^\gamma$ then $E|x_{h/2} - x_{h/4}| \leq C_1(\frac{h}{2})^\gamma$ and then $E|x_{h/4} - x_{h/8}| \leq C_1(\frac{h}{2^2})^\gamma$ and so on. Therefore we will get a geometric series then we will obtain

$$E|X(T) - x_h| \leq \sum_{h=0}^{\infty} C_1\big(\frac{h}{2^k}\big)^\gamma = \frac{C_1 h^\gamma}{1 - 2^{-\gamma}}. \tag{2.7}$$

So from (2.7) we could estimate the convergence and the constant. If the commutativity condition for $\rho_{ikl}(t,x) = \rho_{ilk}(t,x)$, holds for all $x \in \mathbb{R}^d$, $t \in [0,T]$ and all i,k,l then the Milstein scheme (2.5) reduce to

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum_{k=1}^{d} b_{ik}(jh, x^{(j)})\Delta W_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(jh, x^{(j)})B_{kl}^{(j)}, \tag{2.8}$$

which only depends on the generation of the Brownian motion $\Delta W_k^{(j)}$. Scheme (2.8) will give an order one if $d = 1$, but if $d > 1$ will have order $\frac{1}{2}$. As it is described in Davie's paper we could do a modification to scheme (2.8) which will give an order one under a non-degeneracy condition.

2.4. A MODIFICATION TO (2.8) WHICH GIVES ORDER ONE

As it is described in Davie's paper [5] the interpretation of generating of the normal distribution will be changed in scheme (2.8) which leads to convergence of order one

under a non-degeneracy condition. In the implementation of the Milstein scheme we start by generating the random variables $\Delta W_k^{(j)}$ and $A_{kl}^{(j)}$ separately and then we add these random variables to get the RHS of scheme (2.8). The idea here that we will try to generate the following $Y := \sum b_{ik}(jh, x^{(j)})\Delta W_k^{(j)} + \sum \rho_{ikl}(jh, x^{(j)})A_{kl}^{(j)}$, directly. If we have a scheme

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum b_{ik}(jh, x^{(j)})X_k^{(j)} + \sum \rho_{ikl}(jh, x^{(j)})(X_k^{(j)}X_l^{(j)} - h\delta_{kl}), \quad (2.9)$$

where the increment $X_k^{(j)}$ are independent $N(0, h)$ random variables then it is the same as scheme (2.8) with $\Delta W_k^{(j)}$ replaced by $X_k^{(j)}$ and we do not assume $\Delta W_k^{(j)} = X_k^{(j)}$. Now we need $Z_i := \sum b_{ik}(jh, x^{(j)})X_k^{(j)} + \sum \rho_{ikl}(jh, x^{(j)})(X_k^{(j)}X_l^{(j)} - h\delta_{kl})$, to be a good approximation to Y_i, in other words how we could find a joint distribution of random vectors $(\Delta W_k^{(j)}, A_{kl}^{(j)})$ and $(X_k^{(j)})$ so they have the required marginal distribution, with bound $E(Y_i - Z_i)^2 = O(h^3)$. We will explain in the following section how we can use a coupling to find the required marginal distribution which will give good bound for the random distribution Y_i and Z_i. After that we will get an order one approximation between the two approximate solutions of the SDEs, $x(jh)$ and $x^{(j)}$ i.e. $E(x(jh) - x^{(j)}) = O(h^2)$. We need to mention two facts before we explain the approximate coupling.

(1) In the implementation we will use the explicit versions for the coefficients from the Runge-Kutta scheme coefficients(11.1.7) in Kloeden and Platen's book [6] i.e.

$$\beta_{ikl}(x) = \frac{b_{ik}(\Upsilon_n^l) - b_{ik}(x)}{\sqrt{h}}$$

where $\left(\Upsilon_n^l = x + b^l\sqrt{h}\right)$ for $l = 1, 2, \cdots$ and β_{ikl} we will be used an approximation to ρ_{ikl}.

Lemma 2.1. *Suppose we have the Runge-Kutta scheme coefficients(11.1.7) in Kloeden and Platen's book [6], i.e.*

$$\beta_{ikl}(x) = \frac{b_{ik}(\Upsilon_n^l) - b_{ik}(x)}{\sqrt{h}}$$

with $b_{ik}(x)$ twice differentiable with respect to x and $\Upsilon_n^l = x + b^l\sqrt{h}$ for $l = 1, 2, \cdots$. Moreover the $b_{ik}(x)$ and its second derivative are bounded by constant. Then the difference approximation between $\beta_{ikl}(x)$ and the derivatives term $\rho_{ikl}(x)$ will be $O(h)$.i.e.

$$\left(\left|\beta_{ikl}(x) - \rho_{ikl}(x)\right|^p\right)^{2/p} \leq C_p h \quad (2.10)$$

where C_p is a constant.

Proof. see Alnafisah [1] □

Lemma 2.2. *Let $p \in \mathcal{P}$. Then we can find a vector polynomial $\psi \in \mathcal{P}^q$ such that $\nabla.(f\psi) = fPp$.*

Proof. see Lemma 1 in [5] □

Lemma 2.3. *Let $n \leq N$ and R be positive integers, and for $j = 1, \cdots, N$ let $p_j, r_j \in \mathcal{P}$, all having degree$\leq R$, and such that $p_j = r_j$ for $j \leq n$. Let $\eta > 0$ with $\eta R \leq n$ and let $K > 0$. Then we can find $C > 0$ such that, if $\epsilon > 0$ and we write $\mu_0 = pf\chi_B dx$ and $\nu_0 = rf\chi_B dx$ where $p = 1 + \sum_{j=1}^{N} \epsilon^j p_j$, $r = 1 + \sum_{j=1}^{N} \epsilon^j r_j$ and $B = \{x \in \mathbb{R}^q : |x| \leq \epsilon^{-\eta}\}$, and if μ and ν are probability measures on $\mathbb{R}^q$ with $\int_{\mathbb{R}^q}(1 + |x|^2)d|\mu - \mu_0|(x) < K\epsilon^{2n+2}$ and $\int_{\mathbb{R}^q}(1 + |x|^2)d|\nu - \nu_0|(x) < K\epsilon^{2n+2}$, then $\mathbb{W}_2(\mu, \nu) < C\epsilon^{n+1}$.*

Proof. see Lemma 2 in [5] □

3. BOUNDS USING TWO-LEVEL COUPLING

First we consider scheme (2.9) and for the simplicity we will let $b_{ik}(x)$ depend only on x and also the drift term equal zero, so

$$x_i^{(j+1)} = x_i^{(j)} + \sum b_{ik}(x^{(j)})X_k^{(j)} + \sum \rho_{ikl}(x^{(j)})(X_k^{(j)}X_l^{(j)} - h\delta_{kl}). \tag{3.1}$$

Now for the step-size $h^{(r)} = \frac{T}{2^r}$ we will have $2^r d$ independent random variables $X_k^{(r,j)}$. Then at two consecutive levels, in other words from level r to level $r+1$, $r \in \mathbb{N}$ we need to find a coupling between $X_k^{(r,j)}$ which is $N(0, h^{(r)})$ and $(X_k^{(r+1,2j)}, X_k^{(r+1,2j+1)})$ so they are independent of each other and they are $N(0, h^{(r+1)})$. If we have that $\tilde{x}_i^{(r,j)}$ is a solution of 3.1 at the level r then for a fix time j we compare $\tilde{x}_k^{(r,j+1)}$ at level r with $\tilde{x}_k^{(r+1,2j+2)}$ in the level $r+1$, we have

$$\tilde{x}_i^{(r,j+1)} = \tilde{x}_i^{(r,j)} + \sum_{k=1}^{d} b_{ik}(\tilde{x}^{(r,j)})X_k^{(r,j)} + \frac{1}{2}\sum_{k,l=1}^{d} \rho_{ikl}(\tilde{x}^{(r,j)})(X_k^{(r,j)}X_l^{(r,j)} - h^{(r)}\delta_{kl}), \tag{3.2}$$

and define y as the following

$$y = \tilde{x}_i^{(r+1,2j)} + \sum_{k=1}^{d} b_{ik}(\tilde{x}^{(r+1,2j)})X_k^{(r,j)}$$

$$+\frac{1}{2}\sum_{k,l=1}^{d}\rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r,j)}X_l^{(r,j)}-h^{(r)}\delta_{kl}), \quad (3.3)$$

also we have

$$\tilde{x}_i^{(r+1,2j+1)}=\tilde{x}_i^{(r+1,2j)}+\sum_{k=1}^{d}b_{ik}(\tilde{x}^{(r+1,2j)})X_k^{(r+1,2j)}$$
$$+\frac{1}{2}\sum_{k,l=1}^{d}\rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r+1,2j)}X_l^{(r+1,2j)}-h^{(r+1)}\delta_{kl}). \quad (3.4)$$

$$\tilde{x}_i^{(r+1,2j+2)}=\tilde{x}_i^{(r+1,2j+1)}+\sum_{k=1}^{d}b_{ik}(\tilde{x}^{(r+1,2j+1)})X_k^{(r+1,2j+1)}$$
$$+\frac{1}{2}\sum_{k,l=1}^{d}\rho_{ikl}(\tilde{x}^{(r+1,2j+1)})(X_k^{(r+1,2j+1)}X_l^{(r+1,2j+1)}-h^{(r+1)}\delta_{kl}). \quad (3.5)$$

We should mention that when we write $X=O(M)$ for the random variable X we mean the L^p bound for it i.e. $(E|X|^p)^{1/p}\leq CM$. Now, from lemma 2.1 we have

$$b_{ik}(\tilde{x}^{(r+1,2j+1)})=b_{ik}(\tilde{x}^{(r+1,2j)})+\rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r+1,2j)})+O(h)$$

and $\rho_{ikl}(\tilde{x}^{(r+1,2j+1)})=\rho_{ikl}(\tilde{x}^{(r+1,2j)})+O(h)$.

Using these relations in (3.5) and combining it with (3.4) we get.

$$\tilde{x_i}^{(r+1,2j+2)}=\tilde{x_i}^{(r+1,2j)}+\sum_{k=1}^{d}b_{ik}(\tilde{x_i}^{(r+1,2j)})(X_k^{(r+1,2j)}+X_k^{(r+1,2j+1)})$$
$$+\sum_{l,k=1}^{d}\rho_{ikl}(\tilde{x}^{(r+1,2j)})X_k^{(r+1,2j+1)}X_l^{(r+1,2j)}$$
$$+\frac{1}{2}\sum_{l,k=1}^{d}\rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r+1,2j)}X_l^{(r+1,2j)}$$
$$+X_k^{(r+1,2j+1)}X_l^{(r+1,2j+1)}-h^{(r)}\delta_{kl})$$
$$+\lambda, \quad (3.6)$$

where $\lambda=O((h^{(r)})^{3/2})$.

Now, let (c_{ij}) be the matrix inverse of $\left(b_{ik}(\tilde{x}^{(r+1,2j)})\right)$ so that $\sum_j c_{ij}b_{ik}(\tilde{x}^{(r+1,2j)})=\delta_{ik}$. Then from equation (3.3) and (3.6) if we need the local error $y-\tilde{x}_k^{(r+1,2j+2)}=O((h^{(r)})^{3/2})$, we require the coupling to satisfy

$$X_i^{(r,j)}=X_i^{(r+1,2j)}+X_i^{(r+1,2j+1)}$$

$$+\sum_{k,l=1}^{d}\tau_{ikl}(X_k^{(r+1,2j+1)}X_l^{(r+1,2j)}-X_l^{(r+1,2j+1)}X_k^{(r+1,2j)})+O((h^{(r)})^{3/2}),$$

where $\tau_{ikl}=\frac{1}{2}\sum_j c_{ij}\rho_{ikl}$.

Now we will reformulate (**??**) by a scaling. We fix r write $\epsilon=(h^{(r)})^{1/2}, X_i^{(r,j)}=\epsilon V_i, X_i^{(r+1,2j)}=\epsilon Y_i$ and $X_i^{(r+1,2j+1)}=\epsilon Z_i$. Then $V_1,\cdots,V_d$ are independent and $N(0,1)$, while $(Y_1,\cdots,Y_d,Z_1,\cdots,Z_d)$ are independent and $N(0,1/2)$. Now we need to find a coupling between a vector (V_i) and (Y_i,Z_i) so that

$$V_i=Y_i+Z_i+\epsilon\sum_{k,l=1}^{d}\tau_{ikl}(Z_kY_l-Z_lY_k)+O(\epsilon^2). \tag{3.7}$$

We need to write $U_i=Y_i+Z_i$ and $U_i^*=Y_i-Z_i$ that gives U_i and U_i^* are independent and $N(0,1)$. We have $U_l^*U_k-U_k^*U_l=2(Y_lZ_k-Z_lY_k)$ so that from equation (3.7) we obtain

$$V_i=U_i+\epsilon\sum_{k,l=1}^{d}\tau_{ikl}(U_l^*U_k-U_k^*U_l)+O(\epsilon^2). \tag{3.8}$$

Therefore, we require a coupling between $(V_1,\cdots,V_d)$ and $(U_1,\cdots,U_d,U_1^*,\cdots,U_d^*)$, here all the random variables are $N(0,1)$, and also $(V_1,\cdots,V_d)$ are mutually independent,
$(U_1,\cdots,U_d,U_1^*,\cdots,U_d^*)$ are also mutually independent, and (3.8) holds.

4. APPROXIMATE COUPLING FOR GENERAL *D*

In this section we will describe the approximate coupling which satisfies (3.8) with U, U^* having the required distribution but the random variable V has only approximately a standard normal distribution. Here the error bounds are somewhat less precise but the estimates can easily be made rigorous. First of all, we will start with a lemma.

Lemma 4.1. *Let $U=(U_1,\cdots,U_d)$ be a random vector with $N(0,I)$ distribution and let A be a fixed $d\times d$ matrix. Let $Y=U+\epsilon AU$. Then the density function of Y satisfies*

$$f_Y(y)=(2\pi)^{-d/2}e^{-|y|^2/2}\{1+\epsilon(y^tAy-trA)+\epsilon^2\Omega\}+O(\epsilon^3) \tag{4.1}$$

where $\Omega=-(trA)y^tAy-y^tA^2y-\frac{1}{2}|Ay|^2+\frac{1}{2}(y^tAy)^2+\frac{1}{2}(trA)^2+\frac{1}{2}tr(A^2)$.

Proof. see Davie [5] □

We need now to apply the lemma to (3.8). We will use the same definition for U and U^* as in (3.8) and define

$$Y_i=U_i+\frac{\epsilon}{2}\sum_{k,l=1}^{d}\tau_{jkl}(U_l^*U_k-U_k^*U_l) \tag{4.2}$$

If we define $\sigma_{ikl} = \frac{1}{2}(\tau_{ikl} - \tau_{ilk})$ then we could rewrite (4.2) in the following way

$$Y_i = U_i + \epsilon \sum_{k,l=1}^{d} \sigma_{ikl} U_k U_l^* \tag{4.3}$$

Now we need to find V which is close to $N(0, I)$ such that $V - U = O(\epsilon^2)$. To do this we first apply the previous lemma to approximate the density function of Y. We write $Y = U + \epsilon AU$ where the matrix $A = (a_{ik})$ is given by $a_{ik} = \sum_{l=1}^{d} \sigma_{ikl} U_l^*$. Then the density of Y, conditional on U^*, is given by (4.12) from the previous lemma. Now we need to find the unconditional density of Y by substituting for A in (4.12) and taking the expectation with respect to U^*. We will do this for every term separately. Firstly let

$$\delta_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases}$$

If we have the normal distributions $U_1^*, \cdots, U_n^*$ and let

$$N_1 = \sum_{l=1}^{n} a_l U_l^*, \qquad N_2 = \sum_{k=1}^{n} b_k U_k^*$$

$$E(N_1 N_2) = E \sum_{k,l=1}^{n} a_l b_k U_l^* U_k^* = \sum_{k=1}^{n} a_k b_k$$

So from equation (4.12) and taking the expectation w.r.t U^* , we have.

$$E((\mathrm{tr}A) y^t A y) = E(\sum_{i,k=1}^{d} \Big(\sum_{m=1}^{d} a_{mm} \Big) a_{ik} y_k y_i) = \sum_{i,k=1}^{d} \sum_{m,l=1}^{d} \sigma_{mml} \sigma_{ikl} y_k y_i \tag{4.4}$$

$$E(y^t A^2 y) = E\Big(\sum_{i,k=1}^{d} \Big(\sum_{m=1}^{d} a_{im} a_{mk} y_k y_i \Big) \Big) = \sum_{i,k=1}^{d} \sum_{m,l=1}^{d} \sigma_{iml} \sigma_{mkl} y_k y_i \tag{4.5}$$

$$E|Ay|^2 = E(y^t A^t A y) = E\Big(\sum_{i,k=1}^{d} \Big(\sum_{m=1}^{d} a_{mi} a_{mk} y_k y_i \Big) \Big) = \sum_{i,k=1}^{d} \sum_{m,l=1}^{d} \sigma_{mil} \sigma_{mkl} y_k y_i \tag{4.6}$$

$$E(y^t A y)^2 = E\Big(\sum_{i,j,k,m=1}^{d} a_{ik} a_{jm} y_k y_i y_m y_j \Big) = \sum_{i,j,k,m=1}^{d} \sum_{l=1}^{d} \sigma_{ikl} \sigma_{jml} y_k y_i y_m y_j \tag{4.7}$$

$$E((\mathrm{tr}A)^2) = E\Big(\sum_{i,k=1}^{d} a_{ii} a_{kk} \Big) = \sum_{i,k,l=1}^{d} \sigma_{iil} \sigma_{kkl} \tag{4.8}$$

$$E(\mathrm{tr}A^2) = E(\mathrm{tr}AA) = E\Big(\sum_{i,k=1}^{d} a_{ik}a_{ki}\Big) = \sum_{i,k,l=1}^{d} \sigma_{ikl}\sigma_{kil} \tag{4.9}$$

Now from (4.4) to (4.9) the unconditional density of Y satisfies

$$(2\pi)^{-d/2}e^{-|y|^2/2}\Big\{1+\epsilon^2\Big(\mathcal{K}-\sum_{i,k=1}^{d}\theta_{ik}y_iy_k+\sum_{i,j,k,m=1}^{d}\Psi_{ijkm}y_iy_jy_ky_m\Big)\Big\}+O(\epsilon^4) \tag{4.10}$$

Where $\mathcal{K} = \frac{1}{2}\sum_{i,k,l=1}^{d}(\sigma_{ikl}\sigma_{kil} + \sigma_{iil}\sigma_{kkl})$, $\theta_{ik} = \sum_{l,m=1}^{d}(\sigma_{iml}\sigma_{mkl} + \sigma_{ikl}\sigma_{mml} + \frac{1}{2}\sigma_{mil}\sigma_{mkl})$
and $\Psi_{ijkm} = \frac{1}{2}\sum_{l=1}^{d}\sigma_{ikl}\sigma_{jml}$. Here there is no ϵ^3 term because the density is invariant under $\epsilon \to -\epsilon$. A correction term needs to be added to the distribution Y to make it close to the standard normal distribution. We consider an $\mathbb{R}^d$-valued random variable V given by

$$V_i = U_i + \epsilon\sum_{k,l=1}^{d}\sigma_{ikl}U_kU_l^* + \epsilon^2 p_i(U) \tag{4.11}$$

Here we need to choose the p which is an $\mathbb{R}^d$-valued polynomial on $\mathbb{R}^d$. After we add the correction term to the V then we need to find its density function f_V and we will use the following Lemma.

Lemma 4.2. *Let $U = (U_1, \cdots, U_d)$ be a random vector with $N(0, I)$ distribution and let A be a fixed $d \times d$ matrix. Let $Y = U + \epsilon AU + \epsilon^2 p(U)$ where p is an $\mathbb{R}^d$-valued polynomial on $\mathbb{R}^d$. Then the density function of Y satisfies*

$$f_Y(y) = (2\pi)^{-d/2}e^{-|y|^2/2}\{1 + \epsilon(y^tAy - trA) + \epsilon^2\Omega\} + O(\epsilon^3) \tag{4.12}$$

Where $\Omega = -(trA)y^tAy - y^tA^2y - \frac{1}{2}|Ay|^2 + \frac{1}{2}(y^tAy)^2 + \frac{1}{2}(trA)^2 + \frac{1}{2}tr(A^2) + y.p(y) - \nabla.p(y)$.

Proof. See Alnafisah [1]. □

So the density function f_V satisfies

$$f_V(y) = (2\pi)^{-d/2}e^{-|y|^2/2}\big\{1+\epsilon^2\big(\mathcal{K}-\theta(y)+\Psi(y)+y.p(y)-\nabla.p(y)\big)\big\}+O(\epsilon^3) \tag{4.13}$$

where $\theta(y) = \sum_{i,k=1}^{d}\theta_{ik}y_iy_k$ and $\Psi(y) = \sum_{i,j,k,m=1}^{d}\Psi_{ijkm}y_iy_jy_ky_m$.

From lemma 2.2, the polynomial p could be chosen such that

$$\nabla.(fp(y)) = f(\Psi(y) - \theta(y) + \mu) \Rightarrow f\nabla.p(y) - y.p(y)f = f(\Psi(y) - \theta(y) + \mu)$$

This gives

$$\nabla.p(y) - y.p(y) = \Psi(y) - \theta(y) + \mu \tag{4.14}$$

μ here is a constant. Because f_V is a density and its integral over entire space equals to one then we should have $\mu = \mathcal{K}$. Now we will verify that if we have $p = \frac{1}{2}\nabla(\frac{\nabla^2-\Psi}{4} + \theta - \frac{\Psi}{2})$ then it will satisfy (4.14). We need to find every term separately. If we have that

$$F' = \begin{bmatrix} \frac{\partial y_1}{\partial y_1} & \cdots & \frac{\partial y_d}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial y_d} & \cdots & \frac{\partial y_d}{\partial y_d} \end{bmatrix}$$

$$\Theta = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1d} \\ \vdots & \ddots & \vdots \\ \theta_{d1} & \cdots & \theta_{dd} \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix}$$

then

$$\begin{aligned} \nabla(\theta(y)) &= \nabla(\sum_{i,k=1}^{d} \theta_{ik} y_i y_k) \\ &= \sum_{i,k=1}^{d} \theta_{ik} \nabla y_i y_k + \sum_{i,k=1}^{d} \theta_{ik} y_i \nabla y_k \\ &= F'\Theta y + F'\Theta^t y \\ &= I_d \Theta y + I_d \Theta^t y \\ &= \Theta y + \Theta^t y \end{aligned} \tag{4.15}$$

On the other hand we have

$$\begin{aligned} \Psi(y) &= \sum_{i,j,k,m=1}^{d} \Psi_{ijkm} y_i y_j y_k y_m \\ &= \sum_{i,j,k,m=1}^{d} \sum_{l}^{d} \sigma_{ikl}\sigma_{jml} y_i y_j y_k y_m \\ &= \big(\sum_{i,k=1}^{d} \sigma_{ik1} y_i y_k\big)\big(\sum_{j,m=1}^{d} \sigma_{jm1} y_j y_m\big) + \cdots + \big(\sum_{i,k=1}^{d} \sigma_{ikd} y_i y_k\big)\big(\sum_{j,m=1}^{d} \sigma_{jmd} y_j y_m\big) \\ &= \big|\sum_{i,k=1}^{d} \sigma_{ik1} y_i y_k\big|^2 + \cdots + \big|\sum_{i,k=1}^{d} \sigma_{ikd} y_i y_k\big|^2 \end{aligned}$$

and

$$\nabla(\Psi(y)) = \nabla\big(\big|\sum_{i,k=1}^{d} \sigma_{ik1} y_i y_k\big|^2 + \cdots + \big|\sum_{i,k=1}^{d} \sigma_{ikd} y_i y_k\big|^2\big)$$

Now let $\sigma_1(y) = \sum_{i,k=1}^{d} \sigma_{ik1} y_i y_k$, $\cdots$, $\sigma_d(y) = \sum_{i,k=1}^{d} \sigma_{ikd} y_i y_k$

and $A_1 = \sigma_{ik1}$, $\cdots$,$A_d = \sigma_{ikd}$ Then

$$\begin{aligned}
\nabla(\Psi(y)) &= \nabla(\sigma_1(y)^2 + \cdots + \sigma_d(y)^2) \\
&= \nabla\sigma_1(y)^2 + \cdots + \nabla\sigma_d(y)^2 \\
&= 2\sigma_1(y)\nabla\sigma_1(y) + \cdots + 2\sigma_d(y)\nabla\sigma_d(y) \\
&= 2\sigma_1(y)(A_1 y + A_1^t y) + \cdots + 2\sigma_d(y)(A_d y + A_d^t y) \\
&= 2\sigma_1(y)B_1 + \cdots + 2\sigma_d(y)B_d
\end{aligned}$$

Where $B_1 = (A_1 y + A_1^t y)$, $\cdots$, $B_d = (A_d y + A_d^t y)$

$$\begin{aligned}
\nabla^2(\Psi(y)) &= \nabla(2\sigma_1(y)B_1 + \cdots + 2\sigma_d(y)B_d) \\
&= 2\nabla\sigma_1(y)B_1 + 2\sigma_1(y)\nabla B_1 + \cdots + 2\nabla\sigma_d(y)B_d + 2\sigma_d(y)\nabla B_d \\
&= 2(A_1 y + A_1^t y)^t B_1 + 4\sigma_1(y)(tr(A_1)) + \cdots + 2(A_d y + A_d^t y)^t B_d \\
&+ 4\sigma_d(y)(tr(A_d)) \\
&= 2y^t\big[A_1^t A_1 + A_1^2 + 2A_1(tr(A_1)) + (A_1^t)^2 + A_1 A_1^t\big]y + \cdots \\
&+ 2y^t\big[A_d^t A_d + A_d^2 + 2A_d(tr(A_d)) + (A_d^t)^2 + A_d A_d^t\big]y \\
&= 2y^t E_1 y + \cdots + 2y^t E_d y \\
&= 2y^t(E_1 + \cdots + E_d)y \\
&= 2y^t E y
\end{aligned}$$

where $E_1 = A_1^t A_1 + A_1^2 + 2A_1(tr(A_1)) + (A_1^t)^2 + A_1 A_1^t$,
$E_d = A_d^t A_d + A_d^2 + 2A_d(tr(A_d)) + (A_d^t)^2 + A_d A_d^t$
and $E = E_1 + \cdots + E_d$

Finally we need to find $\nabla[\nabla^2(\Psi(y))]$

$$\begin{aligned}
\nabla[\nabla^2(\Psi(y))] &= \nabla\big(2y^t E y\big) \\
&= 2(Ey + E^t y)
\end{aligned}$$

So from (4.14) we have

$$\begin{aligned}
\nabla.p(y) - y.p(y) &= -E + \frac{(\Theta + \Theta^t)}{2} - y^t E y - y^t[-Ey + \frac{(\Theta + \Theta^t)y}{2} \\
&- (\sigma_1(y)B_1 + \cdots + \sigma_d(y)B_d)] \\
&= -E + \frac{(\Theta + \Theta^t)}{2} - y^t E y + y^t E y - \frac{y^t(\Theta + \Theta^t)y}{2} \\
&+ (\sigma_1(y)^2 + \cdots + \sigma_d(y)^2) \\
&= \Psi(y) - \theta(y) + \mu
\end{aligned} \tag{4.16}$$

After we find the derivation of p we could see the ϵ^2 term will equal zero, also f_V is an even function of ϵ, therefore we will not have an ϵ^3 term in its expansion, so $f_V(y) = (2\pi)^{-d/2}e^{-|y|^2/2} + O(\epsilon^4)$. We also see that this Vsatisfies (3.8).
Now we need to expand the density f_V further as

$$f_V(y) = \phi(y)(1 + q(y)) + O(\epsilon^8)$$

where $\phi(y)$ is the density of standard normal distribution in $\mathbb{R}^d$. i.e.

$$\phi(y) = (2\pi)^{-d/2}e^{-|y|^2/2}$$

and q is a polynomial in ϵ and y and its expansion has ϵ^4 and ϵ^6 terms but the dominant term will be of order ϵ^4, i.e. it has $O(\epsilon^4)$. Then from lemma 2.3 we can deduce that the distance ($\mathbb{W}_2$) between the random variables V and $\tilde{V} = N(0, I)$ will be of $O(\epsilon^4)$. This means $\tilde{V}$ will be coupled to V so that

$$E|V - \tilde{V}|^2 = O(\epsilon^8) \tag{4.17}$$

The purpose of the following discussion is to show how we could use the empirical estimate as in section 2.3.1 with the approximate coupling to get an estimate for the error using the coupling $\tilde{V}$ and hence we get an empirical upper bound for the Vaserstein distance between the approximate solutions at two levels.

So if we need to generate coupled approximate solutions $\tilde{x}^{(r,j)}$ and $\tilde{x}^{(r+1,2j)}$ at two different levels r and $r + 1$ then we could use the above definitions of V, U, U^*. Because V does not have the exact normal distribution $N(0, I)$, therefore we will not get the true implementation for $\tilde{x}^{(r,j)}$ in (3.1). We could get the true implementation of (3.1) which we will call $\bar{x}^{(r,j)}$ by substituting V by $\tilde{V}$, but we do not have a means of generating it jointly with the level $r + 1$ solution, therefore we use $\tilde{x}^{(r,j)}$ as an approximation. As we have done before, we have that the bound between $\tilde{x}^{(r,j)}$ and $\tilde{x}^{(r+1,2j)}$ is $O(h)$, and from (4.17) we obtain the bound $\bar{x}^{(r,j)} - \tilde{x}^{(r,j)} = O(h^2)$.

Now we need to estimate the error as the following. If we have that N is the total number of steps at level r, we need to estimate the following $E|\bar{x}^{(r,N)} - \bar{x}^{(r+1,2N)}|$ where $\bar{x}^{(r,N)}$ is the true implementation of (3.1). But as we mentioned before that we could estimate empirically $E|\tilde{x}^{(r,N)} - \tilde{x}^{(r+1,2N)}|$ by using the approximate coupling method. After that we could get $O(h^2)$ bound between $E|\bar{x}^{(r,N)} - \bar{x}^{(r+1,2N)}|$ and $E|\tilde{x}^{(r,N)} - \tilde{x}^{(r+1,2N)}|$. As we expect $E|\bar{x}^{(r,N)} - \bar{x}^{(r+1,2N)}|$ to be of order h, so the error h^2 between them should be negligible for small h, hence the approximate method is effective for empirical estimation.

In following section we will show the numerical results of the approximate coupling for the scheme (2.9) which support the theoretical results. I would like to mention that the following implementation will be for 3-dimensional stochastic differential equation but the codes could be applied to d-dimensional SDEs.

Theorem 4.1. *Assume the matrix $b_{ik}(x)$ is invertible and twice differentiable with respect to x. Moreover the $b_{ik}(x)$ and its second derivative are bounded by constant. Also we assume the boundedness on the inverse of the matrix $b_{ik}(x)$. Then we obtain the following*

$$(E|\tilde{x}_i^{(r+1,2j)} - \tilde{x}_i^{(r,j)}|^p)^{2/p} \leq k_2 h^2 e^{TL} \tag{4.18}$$

where $\tilde{x}_i^{(r,j)}$ and $\tilde{x}_i^{(r+1,2j)}$ are define the same in (3.2), (3.4) and (3.5) where the explanation of the generation of the random variables X has been shown in the coupling summary.

Proof. See Alnafisah [1]. □

5. THE IMPLEMENTATION OF APPROXIMATE COUPLING IN THREE-DIMENSIONAL SDE WITH INVERTIBLE DIFFUSION

First of all, we have the 3-dimensional SDE, which is invertible.

$$\begin{aligned} dX_1(t) &= -(sin(X_1(t)))^2 dW_1 + \frac{1}{1+(X_2(t))^2} dW_2 + \frac{1}{1+(X_2(t))^2} dW_3, \\ dX_2(t) &= \frac{1}{1+(X_2(t))^2} dW_1 + (cos(X_1(t)))^2 dW_3, \\ dX_3(t) &= (cos(X_3(t)))^2 dW_1 + \frac{1}{1+(X_1(t))^2} dW_2, \\ \text{for } 0 \leq t \leq 1, &\text{ with } X_1(0) = 2\ , X_2(0) = 0 \text{ and } X_3(0) = 1 \end{aligned} \tag{5.1}$$

where $W_1(t)$, $W_2(t)$ and $W_3(t)$ are independent standard Brownian motion.

To apply a numerical method to this SDE we need to simulate solutions (for the same Brownian path) simultaneously using two different step sizes (h and $h/2$).

The Matlab implementation for this SDE using the approximate coupling, which will show us the result of the absolute value of the difference between two solutions with step size h and $h/2$.

To construct this experiment, we will decrease the step size (h) every time when we calculate the error and examine the convergence order of the approximate coupling method. We will repeat this with different step size using (for example, $R = 1000$) independent simulations. Then the order of convergence of this method between two approximate solutions should be 1.

For the SDE we use the Matlab code to estimates the absolute error $\epsilon = \frac{1}{R}\sum_{i=1}^{R} |x_h^{(i)} - x_{h/2}^{(i)}|$, for the approximation solution x_h where each simulation is for the same Brownian path. We will run the Matlab code with different number of steps (50, 100, 200, 400, 800, 1600) over a very large number of paths.

step-size	error(ϵ)
0.02	0.0225
0.01	0.0117
0.005	0.0059
0.00025	0.0028
0.000125	0.001

Table 1: The error results for the Approx. coupling in 3-d.

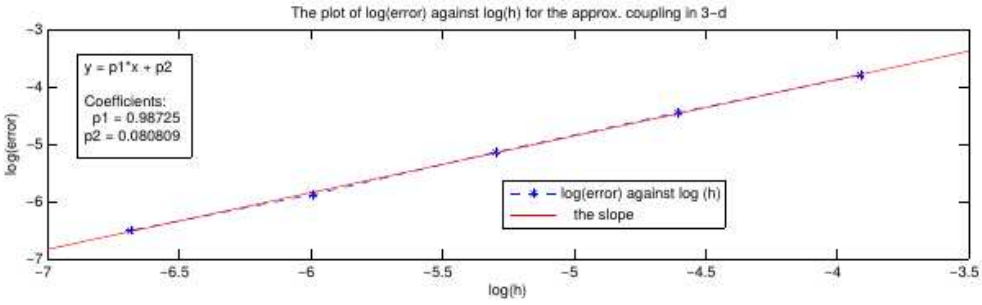

Figure 1: plotting for the convergence of the Apprx. coupling in 3-d

The table (1) and the plotting in Figure (1) show the implementation of the approximation solutions of the previous 3-dimensional SDEs with different number of steps (50, 100, 200, 400, 800, 1600). Running the code for 1000 simulations gives a value for its estimator ϵ equal to 0.0225 with the step-size 0.02 i.e.

$$\epsilon = \frac{1}{1000}\sum_{i=1}^{1000} |x_h^{(i)} - x_{h/2}^{(i)}| = 0.0225$$

and 0.0117 with step-size 0.01 and so on. This means when we increase the number of steps which each time gives a smaller step-size then the estimate error ϵ will give $O(h)$ as it appears in the results in table (1). Also the Figure (1) is a plot of the log of the estimator ϵ i.e. $\log \epsilon$ against the log of step-size h i.e. $\log(h)$ which has a slope of 0.98725 which again indicates a strong convergence of $O(h)$ for the stochastic differential equation (5.1).

Therefore from these computational results we could see that we have obtained good agreement between the theoretical bound in [1] and the implementation results.

ACKNOWLEDGEMENTS

The author would first like to thank Professor Sandy Davie, Professor Istvan Gyongy and Dr Sotirios Sabanis from Edinburgh University for their useful discussions and for their suggestions and improvements to this version of this paper.

REFERENCES

[1] Alnafisah Y. *First-order numerical schemes for stochastic differential equations using coupling*, PhD thesis, University of Edinburgh 2016.

[2] Alnafisah Y., The Implementation of Milstein Scheme in Two-Dimensional SDEs Using the Fourier Method, *Abstract and Applied Analysis,* vol. 2018.

[3] A. Alfonsi, B. Jourdain and A. Kohatsu-Higa, Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme, *Ann. Appl. Probab.* **24** (2014), 1049-1080.

[4] A. Alfonsi, B. Jourdain and A. Kohatsu-Higa, Optimal transport bounds between the time-marginals of multidimensional diffusion and its Euler scheme, *arXiv:* 1405.7007.

[5] A. M. Davie, Pathwise approximation of stochastic differential equations using coupling, preprint: www.maths.ed.ac.uk/∼ adavie/coum.pdf

[6] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag 1995.

[7] J. Komlós, P. Major and G. Tusnády, An approximation of partial sums of independent RV's and the sample DF. I, *Z. Wahr. und Wer. Gebiete* **32** (1975), 111-131.

[8] S. T. Rachev and L. Ruschendorff, *Mass Transportation Problems*, Volume 1, Theory; Volume 2, Applications. Springer-Verlag 1998.

[9] L. N. Vaserstein, Markov processes over denumerable products of spaces describing large system of automata (Russian), *Problemy Peredaci Informacii* **5** (1969), 64-72.

[10] M. Wiktorsson, Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions, *Ann. Appl. Probab.* **11** (2001), 470-487.

[11] T. Rydén and M. Wiktorsson, On the simulation of iteraled Itô integrals, *Stochastic Processes Appl.* **91** (2001), 151-168.

[12] S. Kanagawa, The rate of convergence for the approximate solutions of SDEs, *Tokyo J. Math* **12** (1989), 33-48.

[13] N. Fournier, Simulation and approximation of Lévy-driven SDEs, *ESIAM Probab. Stat.* **15** (2011), 233-248.

[14] E. Rio, Upper bounds for minimal distances in the central limit theorem, *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009), 802-817.

[15] A. B. Cruzeiro, P. Malliavin and A. Thalmaier, Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation, C. R. Math. Acad. Sci. Paris **338** (2004), 481-486.

[16] B. Charbonneau, Y. Svyrydov and P. Tupper, Weak convergence in the Prokhorov metric of methods for stochastic differential equations, *IMA J. Numer. Anal.* **30** (2010), 579-594.

[17] I. Gyöngy and N. Krylov, Existence of strong solutions for Itô's stochastic equations via approximations, *Probab. Theory Related Fields* **105** (1996), 143-158.

[18] Kloeden, P.E; Platen, E. and Wright, I., The approximation of multiple stochastic integrals. J.Stoch. Anal. Appl. **10** (1992) 431-441.

[19] L.Hörmander. Hypoelliptic second order differential equations *Acta. Math.* **119** (1967) 147-171.

[20] Hairer, M. Malliavin's proof of Hörmander's theorem. Bull Math. Sci. **135** (2011), no 6-7, pp. 650-666.

[21] E. Rio, Asymptotic constants for minimal distances in the central limit theorem, Elect. Comm. in Probab. **16** (2011), 96-103.

[22] A. Davie. KMT theory applied to approximations of SDE. In *Stochastic Analysis and Applications 2014*, pages 185-201. Springer, (2014).

[23] Gaines, J., Lyons, T.J., Random generation of stochastic area integrals. In *SIAM J. Appl. Math.* **54**(4), 1132-1146, (2015) .

[24] Simon J.A. Malham, Anke Wiese . Efficient almost-exact Lévy area sampling. In *Statistics and Probability Letters, 2014*, **88**, 50-55.

[25] Guy Flint and Terry Lyons (2015). Pathwise approximation of SDEs by coupling piecewise abelian rough paths, available at https://arxiv.org/abs/1505.0129.