Appl. Math. J. Chinese Univ.
2020, 35(2): 166-183

The implementation of approximate coupling in two-dimensional SDEs with invertible diffusion terms

Yousef Alnafisah

Abstract. We explain and prove some lemmas of the approximate coupling and we give some details of the Matlab implementation of this method. A particular invertible SDEs is used to show the convergence result for this method for general d, which will give an order one error bounds..

§1 Introduction

In this study, an approximate coupling method developed by Davie is investigated [8] that uses coupling and has order-one strong convergence for stochastic differential equations (SDEs). There are several numerical methods for solving SDEs. P.E. Kloeden and E. Platen [14] described a method based on the stochastic Taylor series expansion; however, the major difficulty with this approach is that the double stochastic integrals cannot be 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 in [14], [22], and [24]. However, several random variables should be generated each time, and therefore the computation requires a large amount of time; moreover, this method is difficult to extend to higher order.

In this study, a modified interpretation for the normal random variables generated in the Taylor expansion will be considered. This method has order-one convergence under a non-degeneracy condition for the diffusion term. In standard methods such as Milstein, the approximations for the Taylor expansion terms are separately generated. In the coupling method, the approximation for the Taylor expansion is generated as a combination of random variables. The modification consists in replacing the iterated integrals by different random variables with a good approximation in distribution. Then, a random vector will be obtained from the linear term that is a good approximation in distribution to the original Taylor expansion.

Received: 2018-11-12. Revised: 2020-01-02.
MR Subject Classification: 60H10.
Keywords: stochastic differential equation, milstein method, euler method.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-020-3663-8.
The author gratefully acknowledge Qassim University, represented by the deanship of scientific research, on the material support for this research under the number (3871-cos-2018-1-14-S) during the academic year 2018.

There are several studies that used coupling for the numerical solution of SDEs. Kanagawa [13] investigated the rate of convergence in terms of two probability metrics between approximate solutions with i.i.d. random variables. Rachev and Ruschendorff [19] developed Kanagawa's method by using the Komlós, Major, and Tusnády theorem in [16]. Fournier [10] used the quadratic Vaserstein distance for the approximation of the Euler scheme and the results of Rio [20], which give a very precise rate of convergence for the central limit theorem in the Vaserstein distance. Moreover, Rio [21] provided precise bound estimates. Under uniform ellipticity, Alfonsi, Jourdain, and Kohatsu-Higa [4], [5] studied the Vaserstein bound for the Euler method and proved an $O(h^{(\frac{2}{3}-\epsilon)})$ bound for a one-dimensional diffusion process, where h is the step-size; subsequently, they generalized the result to SDEs of any dimension with an $O(h\sqrt{\log(\frac{1}{h})})$ bound when the coefficients are time-homogeneous. Cruzeiro, Malliavin, and Thalmaier [7] obtained an order-one method, and under non-degeneracy, they constructed a modified Milstein scheme that attains order one for the strong approximation. Charlbonneau, Svyrydov, and Tupper [6] investigated the Vaserstein bound [23] by using weak convergence and the Strassen–Dudley theorem. Convergence of an approximation to a strong solution on a given probability space was established by Gyöngy and Krylov in [11] using coupling. Davie in [9] applied the Vaserstein bound to solutions of vector SDEs and used the Komlós, Major, and Tusnády theorem to obtain order-one approximation under a non-degeneracy assumption.

The remainder of this paper is organized as follows. In Section 2, certain results concerning SDEs are reviewed, and Davies method [8] is introduced. In Section 2, the idea of bounds using two-level coupling is presented and the approximate coupling is described. In the last section, a numerical implementation is provided to demonstrate the convergence behavior for 2-dimensional SDEs using invertible diffusion.

1.1 Stochastic Differential Equations(SDEs)

1.1.1 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).
The stochastic process $X = X(t)$, considered in this study can be described by *SDEs*, namely,

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t), \quad t \in [0, T] \tag{1.1}$$

Let the initial condition $X(0) = x$ be an $\mathcal{F}_0$-measurable random vector in $\mathbb{R}^d$. A $\mathcal{F}_t$-adapted stochastic process $X = (X(t))_{t\geq 0}$ is called a solution of Equation (1.1) if

$$X(t) = X(0) + \int_0^t a(s, X(s))ds + \int_0^t b(s, X(s))dW(s) \tag{1.2}$$

holds almost surely (a.s.)
The conditions that the integral processes

$$\int_0^t a(s, X(s))ds, \qquad \int_0^t b(s, X(s))dW(s),$$

are well-defined are required for(1.2) to hold. Indeed, for the functions $a(s, X(s))$ and $b(s, X(s))$ we have

$$E\int_0^t b^2(s, X(s))ds < \infty, \tag{1.3}$$

and a.s. for all $t \geq 0$,

$$\int_0^t |a(s, X(s))|ds < \infty. \tag{1.4}$$

These conditions imply that the corresponding processes are well defined.
One important property of the stochastic integral is that

$$\int_0^t W(s)dW(s) = \frac{1}{2}\int_0^t d(W^2(s)) - \frac{1}{2}\int_0^t ds = \frac{1}{2}W^2(t) - \frac{t}{2}.$$

For more details on stochastic integrals see [14].

1.2 Strong convergence for SDEs

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where 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. An approximate solution x_h of (1.1) is considered that uses a subdivision of the interval $[0, T]$ into a finite number N of subintervals of length $h = \frac{T}{N}$. Moreover, it is assumed that the approximate solutions x_h are random variables on Ω. Then, the discrete time approximation x_h with step-size h is said to converge strongly of order γ at time $T = Nh$ to the solution $X(t)$ of (1.1) if

$$E|x_h - X(T)|^p \leq Ch^{\gamma p}, \qquad h \in (0, 1),$$

where the strong convergence is in the L^p space. C is a positive constant independent of h. Approximate method will give a strong approximation in the sense of this definition.

1.3 Numerical method for approximating the SDEs

There are several numerical methods for solving SDEs. Here, two important schemes will be mentioned. One is the Euler–Maruyama scheme that has strong order $\frac{1}{2}$, and the other is the Milstein scheme that has strong order one. It is assumed that we have the stochastic differential equation

$$dX_i(t) = a_i(t, X(t))dt + \sum_{k=1}^{d} b_{ik}(t, X(t))dW_k(t), \quad X_i(0) = X_i^{(0)}, \tag{1.5}$$

where i =1,...,d, on an interval $[0, T]$ for a d-dimensional vector $X(t)$ and a d-dimensional Brownian path $W(t)$. To approximate the solution, it is assumed that $[0, T]$ is divided into N equal intervals of length $h = T/N$.

1.3.1 The Milstein scheme

The Milstein scheme is now introduced, which yields an order-one strong Taylor scheme. The Milstein scheme can be obtained by adding the quadratic terms $\sum_{k,l=1}^{d} \rho_{ikl}(jh, x^{(j)})A_{kl}^{(j)}$

to the Euler scheme, namely,

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum_{k=1}^{d} b_{ik}(jh, x^{(j)})\Delta W_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(jh, x^{(j)})A_{kl}^{(j)}, \tag{1.6}$$

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 Euler scheme is easy to implement, as one need only generate the normal distribution for the standard Brownian motion $\Delta W_k^{(j)}$; however, it is not easy to generate the integral $A_{kl}^{(j)}$ for the Milstein scheme for two-dimensional (or higher) SDEs. The two-level approximation will now be described. The increments $\Delta W_k^{(j)}$ should be generated when the solution to (1.1) is approximated by using Euler or other schemes, which will explained later in this section. Therefore, Levy's construction of the Brownian motion will be used to simulate a sequence of approximations that converge to the solution. That is,

$$\Delta W_k^{(r,j)} = \Delta W_k^{(r+1,2j)} + \Delta W_k^{(r+1,2j+1)}, \tag{1.7}$$

where $r \in \mathbb{N}$ and $\Delta W_k^{(r,j)} = W_k((j+1)h^{(r)}) - W_k(jh^{(r)})$ with $h^{(r)} = \frac{T}{2^r}$.
The two-level approximation in (1.7) is called *the trivial coupling.* The normal distribution in (1.7) for the increments for a given level r could be generated by first generating the increments on the left-hand side and then conditionally generating the increments on the right-hand side. The same process is performed for all subsequent levels, and the Brownian path $W(t)$ is thus obtained. The empirical estimation of the error of a numerical method should now be explained.

1.4 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 (1.1). 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

$$E|X(T) - x_h| \leq \sum_{h=0}^{\infty} C_1\big(\frac{h}{2^k}\big)^\gamma = \frac{C_1 h^\gamma}{1 - 2^{-\gamma}}. \tag{1.8}$$

So from (1.8) we could estimate the convergence and the constant.

If the commutativity condition

$$\rho_{ikl}(t,x) = \rho_{ilk}(t,x), \tag{1.9}$$

holds for all $x \in \mathbb{R}^d$, $t \in [0,T]$, and i,k,l, then the Milstein scheme (1.6) is reduced to

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum_{k=1}^{d} b_{ik}(jh, x^{(j)})\Delta W_k^{(j)} + \sum_{k,l=1}^{d} \rho_{ikl}(jh, x^{(j)})B_{kl}^{(j)}, \tag{1.10}$$

which depends only on the generation of the Brownian motion $\Delta W_k^{(j)}$. Scheme (1.10) has order one if $d = 1$, but if $d > 1$ it has order $\frac{1}{2}$. As is described in Davie's study, scheme (1.10) can be modified to obtain order one under a non-degeneracy condition.

1.5 Modification to (1.10) for order-one convergence

As is described in [8], the generation of the normal distribution will be modified in scheme (1.10), leading to order-one convergence under a non-degeneracy condition. In the implementation of the Milstein scheme, the random variables $\Delta W_k^{(j)}$ and $A_{kl}^{(j)}$ are separately generated and are then added to obtain the right-hand side of (1.10). The idea here is to directly generate the following:

$$Y := \sum b_{ik}(jh, x^{(j)})\Delta W_k^{(j)} + \sum \rho_{ikl}(jh, x^{(j)})A_{kl}^{(j)}.$$

If there is a scheme

$$x_i^{(j+1)} = x_i^{(j)} + a_i(jh, x^{(j)})h + \sum b_{ik}(jh, x^{(j)})X_k^{(j)} + \sum \rho_{ikl}(jh, x^{(j)})(X_k^{(j)}X_l^{(j)} - h\delta_{kl}), \tag{1.11}$$

where the increments $X_k^{(j)}$ are independent $N(0,h)$ random variables, then it is the same as scheme (1.10) with $\Delta W_k^{(j)}$ replaced by $X_k^{(j)}$, and $\Delta W_k^{(j)} = X_k^{(j)}$ is not assumed. Furthermore,

$$Z_i := \sum b_{ik}(jh, x^{(j)})X_k^{(j)} + \sum \rho_{ikl}(jh, x^{(j)})(X_k^{(j)}X_l^{(j)} - h\delta_{kl})$$

is assumed to be a good approximation to Y_i, that is, the joint distribution of the random vectors $(\Delta W_k^{(j)}, A_{kl}^{(j)})$ and $(X_k^{(j)})$ should be determined, so that they have the required marginal distribution with bound $E(Y_i - Z_i)^2 = O(h^3)$. In the following section, it will be explained how a coupling can be used to obtain the required marginal distribution, which will give good bounds for the random distributions Y_i and Z_i. Subsequently, an order-one approximation between the two approximate solutions $x(jh)$ and $x^{(j)}$ of the SDE will be obtained, i.e., $E(x(jh) - x^{(j)}) = O(h^2)$.

§2 Bounds using two-level coupling

For simplicity, in (1.11), $b_{ik}(x)$ will be assumed to depend only on x; moreover, the drift term is assumed to be zero. Thus,

$$x_i^{(j+1)} = x_i^{(j)} + \sum b_{ik}(x^{(j)})X_k^{(j)} + \sum \rho_{ikl}(x^{(j)})(X_k^{(j)}X_l^{(j)} - h\delta_{kl}). \tag{2.1}$$

For step-size $h^{(r)} = \frac{T}{2^r}$ there are $2^r d$ independent random variables $X_k^{(r,j)}$. Then, at two consecutive levels, that is, from level r to level $r+1$, $r \in \mathbb{N}$, a coupling between $X_k^{(r,j)}$ should be found that is $N(0, h^{(r)})$, so that $(X_k^{(r+1,2j)}, X_k^{(r+1,2j+1)})$ are independent and $N(0, h^{(r+1)})$. If $\tilde{x}_i^{(r,j)}$ is a solution of 2.1 at level r, then for fixed time j, by comparing $\tilde{x}_k^{(r,j+1)}$ at level r with

$\tilde{x}_k^{(r+1,2j+2)}$ at level $r+1$, we have

$$\tilde{x}_i^{(r,j+1)} = \tilde{x}_i^{(r,j)} + \sum_{k=1}^{d} b_{ik}(\tilde{x}^{(r,j)})X_k^{(r,j)} + \frac{1}{2}\sum_{k,l=1}^{d} \rho_{ikl}(\tilde{x}^{(r,j)})(X_k^{(r,j)}X_l^{(r,j)} - h^{(r)}\delta_{kl}), \tag{2.2}$$

and y is defined as follows:

$$y = \tilde{x}_i^{(r+1,2j)} + \sum_{k=1}^{d} b_{ik}(\tilde{x}^{(r+1,2j)})X_k^{(r,j)} + \frac{1}{2}\sum_{k,l=1}^{d} \rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r,j)}X_l^{(r,j)} - h^{(r)}\delta_{kl}). \tag{2.3}$$

Moreover,

$$\begin{aligned}\tilde{x}_i^{(r+1,2j+1)} &= \tilde{x}_i^{(r+1,2j)} + \sum_{k=1}^{d} b_{ik}(\tilde{x}^{(r+1,2j)})X_k^{(r+1,2j)} \\ &\quad + \frac{1}{2}\sum_{k,l=1}^{d} \rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r+1,2j)}X_l^{(r+1,2j)} - h^{(r+1)}\delta_{kl}).\end{aligned} \tag{2.4}$$

$$\begin{aligned}\tilde{x}_i^{(r+1,2j+2)} &= \tilde{x}_i^{(r+1,2j+1)} + \sum_{k=1}^{d} b_{ik}(\tilde{x}^{(r+1,2j+1)})X_k^{(r+1,2j+1)} \\ &\quad + \frac{1}{2}\sum_{k,l=1}^{d} \rho_{ikl}(\tilde{x}^{(r+1,2j+1)})(X_k^{(r+1,2j+1)}X_l^{(r+1,2j+1)} - h^{(r+1)}\delta_{kl}).\end{aligned} \tag{2.5}$$

It should be mentioned that the notation $X = O(M)$ for the random variable X refers to the L^p bound, i.e., $(E|X|^p)^{1/p} \le CM$. We now have

$$b_{ik}(\tilde{x}^{(r+1,2j+1)}) = b_{ik}(\tilde{x}^{(r+1,2j)}) + \rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r+1,2j)}) + O(h) \tag{2.6}$$

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

Using these relations in (2.5) and combining it with (2.4), we obtain

$$\begin{aligned}\tilde{x_i}^{(r+1,2j+2)} &= \tilde{x_i}^{(r+1,2j)} + \sum_{k=1}^{d} b_{ik}(\tilde{x_i}^{(r+1,2j)})(X_k^{(r+1,2j)} + X_k^{(r+1,2j+1)}) \\ &\quad + \sum_{l,k=1}^{d} \rho_{ikl}(\tilde{x}^{(r+1,2j)})X_k^{(r+1,2j+1)}X_l^{(r+1,2j)} \\ &\quad + \frac{1}{2}\sum_{l,k=1}^{d} \rho_{ikl}(\tilde{x}^{(r+1,2j)})(X_k^{(r+1,2j)}X_l^{(r+1,2j)} + X_k^{(r+1,2j+1)}X_l^{(r+1,2j+1)} - h^{(r)}\delta_{kl}) \\ &\quad + \lambda,\end{aligned} \tag{2.7}$$

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

Let now (c_{ij}) be the inverse matrix of $\big(b_{ik}(\tilde{x}^{(r+1,2j)})\big)$, so that $\sum_j c_{ij}b_{ik}(\tilde{x}^{(r+1,2j)}) = \delta_{ik}$. Then, by Equations (2.3) and (2.7), the local error $y - \tilde{x}_k^{(r+1,2j+2)} = O((h^{(r)})^{3/2})$ requires the coupling to satisfy

$$\begin{aligned}X_i^{(r,j)} &= X_i^{(r+1,2j)} + X_i^{(r+1,2j+1)} + \sum_{k,l=1}^{d} \tau_{ikl}(X_k^{(r+1,2j+1)}X_l^{(r+1,2j)} - X_l^{(r+1,2j+1)}X_k^{(r+1,2j)}) \\ &\quad + O((h^{(r)})^{3/2}),\end{aligned} \tag{2.8}$$

where $\tau_{ikl} = \frac{1}{2}\sum_j c_{ij}\rho_{ikl}$. Equation (2.8) is now reformulated by scaling. r is fixed, and let $\epsilon = (h^{(r)})^{1/2}, X_i^{(r,j)} = \epsilon V_i, X_i^{(r+1,2j)} = \epsilon Y_i$, and $X_i^{(r+1,2j+1)} = \epsilon Z_i$. Then $V_1, \cdots, V_d$ are independent and $N(0,1)$, whereas $(Y_1, \cdots, Y_d, Z_1, \cdots, Z_d)$ are independent and $N(0, 1/2)$. A coupling should now be found between (V_i) and (Y_i, Z_i) so that

$$V_i = Y_i + Z_i + \epsilon \sum_{k,l=1}^{d} \tau_{ikl}(Z_k Y_l - Z_l Y_k) + O(\epsilon^2). \tag{2.9}$$

Let $U_i = Y_i + Z_i$ and $U_i^* = Y_i - Z_i$, so that U_i and U_i^* are independent and $N(0,1)$. We have $U_l^* U_k - U_k^* U_l = 2(Y_l Z_k - Z_l Y_k)$; thus, Equation (2.9) yields

$$V_i = U_i + \epsilon \sum_{k,l=1}^{d} \tau_{ikl}(U_l^* U_k - U_k^* U_l) + O(\epsilon^2). \tag{2.10}$$

Therefore, a coupling between $(V_1, \cdots, V_d)$ and $(U_1, \cdots, U_d, U_1^*, \cdots, U_d^*)$ is required, where all the random variables are $N(0,1)$, $(V_1, \cdots, V_d)$ are mutually independent, $(U_1, \cdots, U_d, U_1^*, \cdots, U_d^*)$ are also mutually independent, and (2.10) holds.

Definition 2.1. Let Σ be a positive definite real $q \times q$ matrix and let f be the density function on $\mathbb{R}^q$ of the $N(0, \Sigma)$ normal distribution. Let $\mathcal{P}$ denote the set of polynomials in d variables with real coefficients and let the projection operator P on $\mathcal{P}$ be defined by $(Pp)(x) = p(x) - \bar{p}$ where $\bar{p} = \int_{\mathbb{R}^q} p(x)f(x)dx$. Then $\bar{P}p = 0$. We have the following

Lemma 2.1. *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 [8]. □

Lemma 2.2. *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 [8]. □

§3 Approximate coupling for general d

In this section we will describe another method for the coupling which satisfies (2.10) 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 3.1.

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

$$f_Y(y) = (2\pi)^{-d/2} e^{-|y|^2/2}\{1 + \epsilon(y^t Ay - trA) + \epsilon^2\Omega\} + O(\epsilon^3) \tag{3.1}$$

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 [8]. □

We need now to apply the lemma to (2.10). The same definition will be used for U and U^* as in (2.10) and define

$$Y_i = U_i + \frac{\epsilon}{2}\sum_{k,l=1}^{d}\tau_{jkl}(U_l^*U_k - U_k^*U_l) \tag{3.2}$$

If we define $\sigma_{ikl} = \frac{1}{2}(\tau_{ikl} - \tau_{ilk})$ then (3.2) could be rewritten in the following way

$$Y_i = U_i + \epsilon\sum_{k,l=1}^{d}\sigma_{ikl}U_kU_l^* \tag{3.3}$$

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 (3.12) from the previous lemma. Now we need to find the unconditional density of Y by substituting for A in (3.12) 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

$$N_1 = \sum_{l=1}^{n} a_lU_l^*, \qquad N_2 = \sum_{k=1}^{n} b_kU_k^*$$

Then

$$\begin{aligned} E(N_1N_2) &= E\sum_{k,l=1}^{n} a_lb_kU_l^*U_k^* \\ &= \sum_{k,l=1}^{n} a_lb_kE(U_l^*U_k^*) \\ &= \sum_{k,l=1}^{n} a_lb_k\delta_{kl} \\ &= \sum_{k=1}^{n} a_kb_k \end{aligned}$$

So from equation (3.12) and taking the expectation w.r.t U^* , we have.

$$\begin{aligned} E((\mathrm{tr}A)y^tAy) &= E(\sum_{i,k=1}^{d}\Big(\sum_{m=1}^{d}a_{mm}\Big)a_{ik}y_ky_i) \\ &= \sum_{i,k=1}^{d}\sum_{m,l,j=1}^{d}\sigma_{mml}\sigma_{ikj}E(U_l^*U_j^*)y_ky_i \\ &= \sum_{i,k=1}^{d}\sum_{m,l=1}^{d}\sigma_{mml}\sigma_{ikl}y_ky_i \end{aligned} \tag{3.4}$$

$$\begin{aligned} E(y^tA^2y) &= E\Big(\sum_{i,k=1}^{d}\Big(\sum_{m=1}^{d}a_{im}a_{mk}y_ky_i\Big)\Big) \\ &= \sum_{i,k=1}^{d}\sum_{m,l,j=1}^{d}\sigma_{iml}\sigma_{mkj}E(U_l^*U_j^*)y_ky_i \\ &= \sum_{i,k=1}^{d}\sum_{m,l=1}^{d}\sigma_{iml}\sigma_{mkl}y_ky_i \end{aligned} \tag{3.5}$$

$$\begin{aligned} E|Ay|^2 = E(y^tA^tAy) &= E\Big(\sum_{i,k=1}^{d}\Big(\sum_{m=1}^{d}a_{mi}a_{mk}y_ky_i\Big)\Big) \\ &= \sum_{i,k=1}^{d}\sum_{m,l,j=1}^{d}\sigma_{mil}\sigma_{mkj}E(U_l^*U_j^*)y_ky_i \\ &= \sum_{i,k=1}^{d}\sum_{m,l=1}^{d}\sigma_{mil}\sigma_{mkl}y_ky_i \end{aligned} \tag{3.6}$$

$$\begin{aligned} E(y^tAy)^2 &= E\Big(\sum_{i,j,k,m=1}^{d}a_{ik}a_{jm}y_ky_iy_my_j\Big) \\ &= \sum_{i,j,k,m=1}^{d}\sum_{l,r=1}^{d}\sigma_{ikl}\sigma_{jmr}E(U_l^*U_r^*)y_ky_iy_my_j \\ &= \sum_{i,j,k,m=1}^{d}\sum_{l=1}^{d}\sigma_{ikl}\sigma_{jml}y_ky_iy_my_j \end{aligned} \tag{3.7}$$

$$\begin{aligned} E((\mathrm{tr}A)^2) &= E\Big(\sum_{i,k=1}^{d}a_{ii}a_{kk}\Big) \\ &= \sum_{i,k=1}^{d}\sum_{l,j=1}^{d}\sigma_{iil}\sigma_{kkj}E(U_l^*U_j^*) \\ &= \sum_{i,k,l=1}^{d}\sigma_{iil}\sigma_{kkl} \end{aligned} \tag{3.8}$$

Finally

$$\begin{aligned} E(\mathrm{tr}A^2) &= E(\mathrm{tr}AA) \\ &= E\Big(\sum_{i,k=1}^{d} a_{ik}a_{ki}\Big) \\ &= \sum_{i,k=1}^{d}\sum_{l,j=1}^{d} \sigma_{ikl}\sigma_{kij}E(U_l^*U_j^*) \\ &= \sum_{i,k,l=1}^{d} \sigma_{ikl}\sigma_{kil} \end{aligned} \tag{3.9}$$

Now from (3.4) to (3.9) the unconditional density of Y satisfies

$$(2\pi)^{-d/2}e^{-|y|^2/2}\Big\{1+\epsilon^2\Big(\mathcal{K}-\sum_{i,k=1}^{d}\theta_{ik}y_iy_k+\sum_{i,j,k,m=1}^{d}\Psi_{ijkm}y_iy_jy_ky_m\Big)\Big\}+O(\epsilon^4) \tag{3.10}$$

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

$$V_i=U_i+\epsilon\sum_{k,l=1}^{d}\sigma_{ikl}U_kU_l^*+\epsilon^2p_i(U) \tag{3.11}$$

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 3.2.

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 $Y=U+\epsilon AU+\epsilon^2p(U)$ where p is an $\mathbb{R}^d$-valued polynomial on $\mathbb{R}^d$. Then the density function of Y satisfies

$$f_Y(y)=(2\pi)^{-d/2}e^{-|y|^2/2}\{1+\epsilon(y^tAy-trA)+\epsilon^2\Omega\}+O(\epsilon^3) \tag{3.12}$$

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. Write $y=(I+\epsilon A)u+\epsilon^2p(u)$.

We have $u=(I+\epsilon A)^{-1}(y-\epsilon^2p(y))+O(\epsilon^3)=(I-\epsilon A+\epsilon^2A^2)(y-\epsilon^2p(y))+O(\epsilon^3)$, which gives that

$$\begin{aligned} |u|^2 &= (y-\epsilon^2p(y))^t(I-\epsilon A+\epsilon^2A^2)^2(y-\epsilon^2p(y))+O(\epsilon^3) \\ &= |y|^2-2\epsilon y^tAy+\epsilon^2(2y^tA^2y+|Ay|^2)-2\epsilon^2y.p(y)+O(\epsilon^3) \end{aligned} \tag{3.13}$$

After we divided equation (3.13) by 2 and taking the exponential, we will obtain

$$\begin{aligned} e^{-|u|^2/2} &= e^{-|y|^2/2+\epsilon y^tAy-\epsilon^2(y^tA^2y+\frac{1}{2}|Ay|^2)+\epsilon^2 y.p(y)} \\ &= e^{-|y|^2/2}e^{\epsilon y^tAy-\epsilon^2(y^tA^2y+\frac{1}{2}|Ay|^2)+\epsilon^2 y.p(y)} \\ &= e^{-|y|^2/2}\Big\{1+\epsilon y^tAy-\epsilon^2\big(y^tA^2y+\frac{1}{2}|Ay|^2\big)+\frac{\epsilon^2}{2}(y^tAy)^2+\epsilon^2 y.p(y)\Big\}+O(\epsilon^3) \quad (3.14) \end{aligned}$$

Here we should mention that there are some small errors in the following process and because of the rapid decay of the normal density, this will not cause a problem. Now we will have the following density function of Y.

$$\begin{aligned} f_Y(y) &= \det(I+\epsilon A+\epsilon^2\nabla p(y))^{-1}f_U(u) \\ &= (2\pi)^{-d/2}\det(I+\epsilon A+\epsilon^2\nabla p(y))^{-1}e^{-|u|^2/2} \\ &= (2\pi)^{-d/2}\det(I+\epsilon A+\epsilon^2\nabla p(y))^{-1}e^{-|y|^2/2}\Big\{1+\epsilon y^tAy-\epsilon^2\big(y^tA^2y+\frac{1}{2}|Ay|^2\big) \\ &+\frac{\epsilon^2}{2}(y^tAy)^2+\epsilon^2 y.p(y)\Big\}+O(\epsilon^3) \quad (3.15) \end{aligned}$$

Now from the expansion of log and the following property $e^{\mathrm{tr}(\log(I+\epsilon A+\epsilon^2\nabla p(y)))}=\det(I+\epsilon A+\epsilon^2\nabla p(y))$ we will have

$$\begin{aligned} \mathrm{tr}\,\log(I+\epsilon A+\epsilon^2\nabla p(y)) &= \epsilon\mathrm{tr}A-\frac{\epsilon^2}{2}\mathrm{tr}(A^2)+\epsilon^2\mathrm{tr}(\nabla p(y))+O(\epsilon^3) \\ &= \epsilon\mathrm{tr}A-\frac{\epsilon^2}{2}\mathrm{tr}(A^2)+\epsilon^2\nabla.p(y)+O(\epsilon^3) \quad (3.16) \end{aligned}$$

therefore

$$\begin{aligned} \det(I+\epsilon A+\epsilon^2\nabla p(y))^{-1} &= \exp\Big(-\epsilon\mathrm{tr}A+\frac{\epsilon^2}{2}\mathrm{tr}(A^2)-\epsilon^2\nabla.p(y)\Big)+O(\epsilon^3) \\ &= 1+\epsilon\mathrm{tr}A+\frac{\epsilon^2}{2}\{(\mathrm{tr}A)^2+\mathrm{tr}(A^2)\}-\epsilon^2\nabla.p(y)+O(\epsilon^3) \quad (3.17) \end{aligned}$$

substituting the result (3.17) in (3.15), then we get the result in (3.12). □

So the density function f_V satisfies

$$f_V(y)=(2\pi)^{-d/2}e^{-|y|^2/2}\big\{1+\epsilon^2\big(\mathcal{K}-\theta(y)+\Psi(y)+y.p(y)-\nabla.p(y)\big)\big\}+O(\epsilon^3) \quad (3.18)$$

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 2.1, 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

$$\nabla.p(y)-y.p(y)=\Psi(y)-\theta(y)+\mu \quad (3.19)$$

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 (3.19). 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

$$\begin{aligned} \nabla(\theta(y)) &= \nabla(\sum_{i,k=1}^{d} \theta_{ik} y_i y_k) \\ &= \sum_{i,k=1}^{d} \theta_{ik} \nabla y_i y_k + \sum_{i,k=1}^{d} \theta_{ik} y_i \nabla y_k \\ &= F'\Theta y + F'\Theta^t y \\ &= I_d \Theta y + I_d \Theta^t y \\ &= \Theta y + \Theta^t y \end{aligned} \tag{3.20}$$

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

$$\begin{aligned}\nabla(\Psi(y)) &= \nabla(\sigma_1(y)^2 + \cdots + \sigma_d(y)^2)\\ &= \nabla\sigma_1(y)^2 + \cdots + \nabla\sigma_d(y)^2\\ &= 2\sigma_1(y)\nabla\sigma_1(y) + \cdots + 2\sigma_d(y)\nabla\sigma_d(y)\\ &= 2\sigma_1(y)(A_1y + A_1^ty) + \cdots + 2\sigma_d(y)(A_dy + A_d^ty)\\ &= 2\sigma_1(y)B_1 + \cdots + 2\sigma_d(y)B_d\end{aligned}$$

Where $B_1 = (A_1y + A_1^ty), \cdots, B_d = (A_dy + A_d^ty)$

$$\begin{aligned}\nabla^2(\Psi(y)) &= \nabla(2\sigma_1(y)B_1 + \cdots + 2\sigma_d(y)B_d)\\ &= 2\nabla\sigma_1(y)B_1 + 2\sigma_1(y)\nabla B_1 + \cdots + 2\nabla\sigma_d(y)B_d + 2\sigma_d(y)\nabla B_d\\ &= 2(A_1y + A_1^ty)^tB_1 + 4\sigma_1(y)(tr(A_1)) + \cdots + 2(A_dy + A_d^ty)^tB_d\\ &+ 4\sigma_d(y)(tr(A_d))\\ &= 2y^t\big[A_1^tA_1 + A_1^2 + 2A_1(tr(A_1)) + (A_1^t)^2 + A_1A_1^t\big]y + \cdots\\ &+ 2y^t\big[A_d^tA_d + A_d^2 + 2A_d(tr(A_d)) + (A_d^t)^2 + A_dA_d^t\big]y\\ &= 2y^tE_1y + \cdots + 2y^tE_dy\\ &= 2y^t(E_1 + \cdots + E_d)y\\ &= 2y^tEy\end{aligned}$$

where $E_1 = A_1^tA_1 + A_1^2 + 2A_1(tr(A_1)) + (A_1^t)^2 + A_1A_1^t$,
$E_d = A_d^tA_d + A_d^2 + 2A_d(tr(A_d)) + (A_d^t)^2 + A_dA_d^t$
and $E = E_1 + \cdots + E_d$

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

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

So from (3.19) we have

$$\begin{aligned}\nabla.p(y) - y.p(y) &= -E + \frac{(\Theta + \Theta^t)}{2} - y^tEy - y^t[-Ey + \frac{(\Theta + \Theta^t)y}{2}\\ &- (\sigma_1(y)B_1 + \cdots + \sigma_d(y)B_d)]\\ &= -E + \frac{(\Theta + \Theta^t)}{2} - y^tEy + y^tEy - \frac{y^t(\Theta + \Theta^t)y}{2}\\ &+ (\sigma_1(y)^2 + \cdots + \sigma_d(y)^2)\\ &= \Psi(y) - \theta(y) + \mu \qquad (3.21)\end{aligned}$$

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 (2.10).

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 2.2 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

$$E|V - \tilde{V}|^2 = O(\epsilon^8) \tag{3.22}$$

The purpose of the following discussion is to show how we could use the empirical estimate as in section 1.4 with the approximate coupling to get an estimate for the error using the exact 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 (2.1). We could get the true implementation of (2.1) 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 (3.22) we obtain the bound $\bar{x}^{(r,j)} - \tilde{x}^{(r,j)} = O(h^2)$.

Now we need to estimate the error as we describe in section 5 in chapter one. 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 (2.1). 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 (1.11) which support the theoretical results. I would like to mention that the following implementation will be for 2-dimensional stochastic differential equation but the codes could be applied to d-dimensional SDEs.

§4 The implementation of an approximate coupling in two-dimensional case with invertible $\left(b_{ik}(\tilde{x}^{(r+1,2j)})\right)$

First of all, we have the 2-dimensional SDE, which is invertible.

$$\begin{aligned} dX_1(t) &= (\sin(X_2(t)))^2 dW_1(t) - \frac{1}{1+X_1^2(t)} dW_2(t), \\ dX_2(t) &= \frac{1}{1+X_2^4(t)} dW_1(t) + (\cos(X_1(t)))^2 dW_2(t), \\ &\text{for } 0 \le t \le 1, \text{ with } X_1(0) = 2 \quad \text{and} \quad X_2(0) = 0 \end{aligned} \tag{4.1}$$

where $W_1(t)$ and $W_2(t)$ 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 will give us the absolute value of the difference between two solutions for the SDEs (4.1) 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 properties of the exact coupling method. We will repeat this with a different step size using (for example, $R = 400$) independent simulations. So the order of convergence of this method between two solution should be 1.

The Matlab code will be run with different step sizes over large number of paths R as described in the table below and see the result of the error ϵ, where each simulation is for the same Brownian path and $\epsilon = \frac{1}{R}\sum_{i=1}^{R} |x_h^{(i)} - x_{h/2}^{(i)}|$ will be our estimator.

Table 1: The error results for the approximate coupling with invertible matrix

step-size	error(ϵ)
0.005	0.0058
0.0025	0.0031
0.00125	0.0016
0.00062	0.00076
0.00031	0.00036
0.00015	0.00018
0.00007	0.000097

The table (1) and the plotting in Figure (1) show the implementation of the approximate solutions of the previous 2-dimensional SDEs for the approximate coupling with different number of steps (200, 400, 800, 1600, 3200, 6400, 12800). Running the code (AproxcouplingInverend)

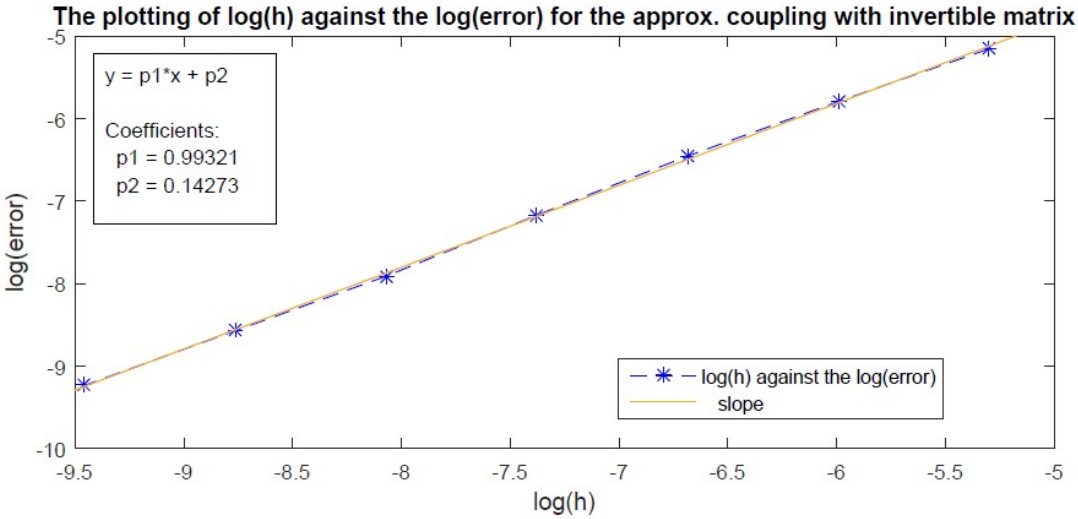

Figure 1: log Error against log of step sizes for approx. coupling with invertible matrix

for 400 simulations gives a value for its estimator ϵ equal to 0.0058 with the step-size 0.005 i.e.

$$\epsilon = \frac{1}{400}\sum_{i=1}^{400} |x_h^{(i)} - x_{h/2}^{(i)}| = 0.0058$$

and 0.0031 with step-size 0.0025 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 the 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 0.99321 which again indicates a strong convergence of $O(h)$ for the stochastic differential equation (4.1).

Therefore from this computational results we could see that we have obtained good agreement between the theoretical bound in [8] with the implementation results.

Acknowledgement

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] Alnafisah Y. *Two-Level Bound for Stochastic Differential Equations Using the Exact Coupling with an Explicit Coefficients*, J Comput Theor Nanosci, 2018, 1954-1964.

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

[5] Alfonsi A, Jourdain B, Kohatsu-Higa A. *Optimal transport bounds between the time-marginals of multidimensional diffusion and its Euler scheme*, arXiv: 1405-7007.

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

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

[8] Davie A. *Pathwise approximation of stochastic differential equations using coupling*, preprint: www.maths.ed.ac.uk/∼ adavie/coum.pdf, 2015.

[9] Davie A. *KMT theory applied to approximations of SDE*, Stochastic Analysis and Applications, 2014, 185-201.

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

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

[12] Hörmander L. *Hypoelliptic second order differential equations*, Acta Math, 1967, 119: 147-171.

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

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

[15] Kloeden P E, Platen E, Wright I. *The approximation of multiple stochastic integrals*, J Stoch Anal Appl, 1992, 10: 431-441.

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

[17] Alhojilan Y. *Using Weak Bound Simulation for Testing Accuracy of Strong Approximation of Order 2 for Stochastic Differential Equations*, J Comput Theor Nanosci, 2018, 15: 3184-3186.

[18] Hairer M. *Malliavin's proof of Hörmander's theorem*, Bull Math Sci, 2011, 135: 650-666.

[19] Rachev S T, Ruschendorff L. *Mass Transportation Problems*, 1998, 1.

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

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

[22] Rydén T, Wiktrosson M. *On the simulation of iteraled Itô integrals*, Stochastic Processes Appl, 2001, 91: 151-168.

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

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

Mathematics Department, College of Science, Qassim University, P.O.Box 6644, Buraydah 51452, Saudi Arabia.

Email: UAN_0261@hotmail.com