A WEAK SIMPSON METHOD FOR A CLASS OF STOCHASTIC DIFFERENTIAL EQUATIONS

RAM SHARAN ADHIKARI

Faculty of Mathematics, Rogers State University

Claremore, Oklahoma

Email:radhikari@rsu.edu

Abstract.

This paper proposes a novel weak Simpson method for numerical solution for a class of stochastic differential equations. We show that such a method has weak convergence of order one in general and weak convergence of order three under certain additional assumptions, which improves the weak convergence order of two for the weak trapezoid method developed in (1). The main results are illustrated with numerical examples.

Keywords: Stochastic differential equation, higher order method, numerical method, Simpson rule, Euler-Maruyama method, and weak convergence

1. Introduction

We consider the problem of constructing accurate approximations on fixed time intervals to solutions of the following system of stochastic differential equations (SDEs)

$$X(s) = x + \int_0^s b(X(r))\, dr + \sum_{k=1}^{M} \int_0^S \sigma_k (X(r))\, v_k\, dW_k(r), \quad s \geq 0, \tag{1.1}$$

where $x \in \mathbb{R}^d$, $M \in \mathbb{N}$ is a positive integer, $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma_k : \mathbb{R}^d \to \mathbb{R}$, and for $k = 1.......M$, $v_k \in \mathbb{R}^d$, and $W_k(t)$ are independent one-dimensional Brownian motions. Here for each k, v_k represents the direction along which the random noise W_k enters the system (1.1). Suppose the coefficients b and σ are measurable and are such that a weak solution to (1.1) exists and is unique in probability law. Typically the coefficients b and σ are assumed to satisfy the Lipschitz continuity and the linear growth condition; see, for example, (9) or (5). SDEs have a wide range of applications in areas such as ecosystem modeling, mathematical finance, and risk management. In most of the practical applications, we cannot find explicit solutions for the underlying SDEs, as with most ordinary differential equations (ODEs). In such situations, numerical approximation then becomes the one viable approach.

There are generally two numerical approximation methods for SDEs, namely strong and weak methods. The strong methods focus on the sample path properties while the weak methods deal with the distributions of the underlying SDEs. Strong approximations are necessary to explore characteristics of the systems that depend on path wise properties. The order of convergence of strong approximation is sometimes less in the stochastic case than in the corresponding deterministic case (6). It is also observed that strong explicit methods, particularly, the widely used Euler-Maruyama method, sometimes works unreliably and generate large errors for certain step-sizes, (2).

Motivated by financial applications, we are interested in evaluating functional of the form $\mathbb{E}[f(X(T))]$ for some $T > 0$ and some function f. Therefore, we will focus on *weak methods* in this work. For instance, the arbitrage-free price of a European call option is given by $\mathbb{E}_{\mathbb{Q}}[e^{-r(T-t)}(S(T)-K)^{+} | \mathcal{F}_t]$, in which $\mathbb{Q}$ is a risk-neutral measure, $r > 0$ is the discounting factor, K is the strike price, and $S(T)$ is the price of the underlying asset at time T. In approximating such functionals, an efficient approximation to the probability distribution of the solutions is sufficient.

This work is motivated by (1) and improves their weak trapezoidal method. In (1), the weak trapezoidal method has weak convergence order of two and seems to require a large number sample paths (10 million). Motivated by the fact that Simpson's rule is usually an improvement of trapezoidal rule, we use the Simpson-like rule to approximate the area under the diffusion term in (1.1). In other words, we use a weak Simpson method to approximate the stochastic integral of (1.1) in our algorithm. We show that our method has weak convergence order one in general and weak convergence of order three under certain additional assumptions (Assumption (A4)). However, we note that our examples in Section 2.4 all demonstrate weak order three convergence, even though they do not necessarily satisfy the Assumption (A4). Moreover, in these examples, our method requires fewer number (in the order of 50,000) of sample paths compared to (1). Unfortunately, at this point, we are not able to prove that the weak Simpson method enjoys weak order three convergence without Assumption (A4).

It is worth noting that there are many other higher order weak Taylor schemes to solve stochastic differential equations, see, for example, (6). Generally, these higher order methods are much more complicated than ours, and contain a large number of terms, such as all of the multiple Itô integrals of higher multiplicity from Itô-Taylor expansion (6). For instance, we need to include all of the third order multiple Itô integrals from the Itô-Taylor expansion to construct order 3.0 weak Taylor scheme. This makes these methods hard to implement in practice. Compared to those higher order weak Taylor schemes, our algorithm is simple and derivative free, yet still enjoys weak convergence order 3.0 under Assumption (A4).

The algorithm of our method consists of two steps, in the first an explicit Euler-Maruyama type step is used and in the second the resulting fractional point is used in combination with initial point to obtain higher order. We use variable steps to obtain better approximation. The use of different paths for each time step-size make sense in our setting because we are only concerned with the mean of the solution. We can choose any sample $\sqrt{h}\,N(0, 1)$ for the increment $W(t_k) - W(t_{k-1})$. This paper develops the method which produce weak approximation rather than strong approximation. Hence, we produce an approximating sample path without giving proper attention to underlying Wiener process. Our algorithm does not require simulation of the Itô integral.

The rest of the paper is organized as follows. In Section 2 we propose our method and we describe why and how our method works by considering the simple case. We give a proof that our method converge with weak order three under the suitable conditions. We also discuss about the numerical performance of our method with some examples. Finally we give the proof of Lemma 2.6 in Appendix which we need to prove our local approximation theorem. To facilitate the presentation of the paper, we introduce some notations here. The notation a T denotes the transpose of a vector or matrix a. For a smooth function $f: \mathbb{R}^d \mapsto \mathbb{R}^d$, $Df(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_d}\right)^T$ denotes the gradient of f and $D^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)$ *is the Heissian of* f *at* x. For $\eta, x \in \mathbb{R}^d$, $f'[\eta](x)$ is the derivative of f in the direction of η evaluated at the point x. And for $\eta, v, x \in \mathbb{R}^d$, $f''[\eta, v](x) := f'(f'[\eta])[v](x)$. In a similar fashion, we define $f'''[\eta, v, \xi](x)$ etc. Note that $f''[\eta, v][x] = \mathrm{tr}(\eta v^T D^2 f(x)) := f''[v, \eta](x)$.

A vector $\alpha = (\alpha_1, \ldots\ldots, \alpha_d)$ with each component α_i taking values from the set of nonnegative integers is called a multi-index. Moreover, we denote $|\alpha| := \alpha_1 + \ldots + \alpha_d$ and $D^\alpha f(x) = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} f(x)$.

2. The weak Simpson Method

2.1. Preliminaries. To begin with, we state the standing assumptions throughout the paper: (A1) The coefficients of (1.1) satisfy the Lipschitz and linear growth condition:

$$|b(x) - b(y)| + \sum_{k=1}^{M} |\sigma_k(x) - \sigma_k(y)| \le L\,|x - y|,$$

$$|b(x)| + \sum_{k=1}^{M} |\sigma_k(x)| \le K(1 + |x|), \tag{2.1}$$

for all $x, y \in \mathbb{R}^d$, where L, K are positive constants.

(A2) for each $k = 1, \ldots, M$, we have $\inf_{x \in \mathbb{R}^d}\{\sigma_k(x)\} > 0$. In addition, there exists a positive constant $\lambda \in (0,1]$ so that for any $x, \xi \in \mathbb{R}^d$ we have

$$\lambda |\xi|^2 \le \xi^T a(x) \xi \le \lambda^{-1} |\xi|^2, \tag{2.2}$$

where $a(x) := \sum_{k=1}^{M} \sigma^2(x)\, v_k v_K^T$.

` (A3) For all multi-index α with $|\alpha| \le 8$, we have

$$|D^\alpha b(x)| + \sum_{k=1}^{M} |D^\alpha \sigma_k(x)| \le K(1 + |x|^P), \text{ for all } x \in \mathbb{R}^d, \tag{2.3}$$

where K and p are positive numbers.

It is well-known that under Assumption (A1), the stochastic differential equation (1.1) has a unique strong solution; see, for example, (5),(9) or (12). Moreover, we have the following moment estimate:

Lemma 2.1 (12). Assume (A1). Let $T > 0$ be fixed. Then for any positive constant p, we have

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X^x(t)|^p\right] \leq C < \infty,\ x \in \mathbb{R}^d \times M, \tag{2.4}$$

Where the constant C satisfies $C = C(x,T,p) > 0$ and X^x denotes the solution to (1.1) with initial condition $x \in \mathbb{R}^d$.

Remark 2.2. Assumption (A1) is a standard condition in the literature and it guarantees the existence and uniqueness of a strong solution to the stochastic differential equation (1.1). In addition, we note that Assumptions (A1)–(A3) are weaker than those in (1), where it is assumed that b, σ_k, $k = 1, \ldots, M$ are bounded with bounded and continuous partial derivatives up to the sixth order and that $\inf_{x\in\mathbb{R}^d} \{\sigma_k(x)\} > 0$ for each k. Also; (2.2) plays an important role in a certain Gaussian tail estimate in the proof of Lemma 2.6.

We denote the space of continuous and bounded functions whose first through kth order partial derivatives are continuous and bounded by $C^k(\mathbb{R}^d)$, that is,

$C^k(\mathbb{R}^d) = \{f : \mathbb{R}^d \rightarrow \mathbb{R}$ s.t. $D^\alpha f(x)$ exists and is bounded and continuous for all $x \in \mathbb{R}^d\}$, where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index such that $|\alpha| := \alpha_1 + \cdots + \alpha_d \leq k$. In addition, we define the norm of $C^k(\mathbb{R}^d)$ in following way

$$\| f \|_k = \sup \{|D^\alpha f(x)| : x \in \mathbb{R}^d, \alpha = (\alpha_1, \ldots, \alpha_d) \text{ is a multi-index with } |\alpha| \leq k\} \tag{2.5}$$

Note that when $k = 0$, C^0 is the family of bounded and continuous functions.

Next we define the Markov semigroup ρ_t: $C^k \rightarrow C^k$ related to (1.1) by

$$(\rho_t f)(x) \stackrel{\text{def}}{=} \mathbb{E}_x[f(X(t))], \qquad [2.6]$$

where $X(0) = x$ and Markov semigroup P_h: $C^k \rightarrow C^k$ associated with single full step size h of the numerical method (2.9)–(2.10) by

$$(P_h f)(y) \stackrel{\text{def}}{=} \mathbb{E}_y[f(Y_1)], \qquad [2.7]$$

where $Y_0 = y$. Since $\| f \|_0 = \sup\{|f(x)|: x \in \mathbb{R}^d\}$, it follows that $\| \rho_t f \|_0 \leq \| f \|_0$ and similarly $\| P_h f\|_0 \leq \| f \|_0$. The following Proposition is from (1).

Proposition 2.3. Assume (A1)–(A3). *Then for any $0 < t \leq T$ and $k \in \mathbb{N}$, there exists a $C = C(T, k, b, \sigma) > 0$ such that $\| \rho_t f\|_k \leq C\|f\|_k$.*

As in (1), we define the induced operator norm for any linear operator $L : C^k \rightarrow C^l$ by

$$\|L\|_{k\rightarrow l} = \sup_{f\in C^k,\ f\neq 0} \frac{\|Lf\|_l}{\|f\|_k}$$

Then it follows from Proposition 2.3 that $\| \rho_t \|_{k\rightarrow k} \leq C$. In particular, we have

$$\|\rho_t\|_{0\to 0} = \sup_{f\in C^0,\, f\neq 0} \frac{\|\rho_t f\|_0}{\|f\|_0} \leq \sup_{f\in C^0,\, f\neq 0} \frac{\|f\|_0}{\|f\|_0} = 1$$

Similarly, $\|P_h\|_{0\to 0} \leq 1$.

2.2. The Algorithm. The weak Simpson method can be summarized as follows. Let $T > 0$ and $\Pi := \{0 = t_0 < t_1 < \ldots < t_N = T\}$ be a subdivision of $[0, T]$. Let $\{\eta^{(i)}_{1k}, \eta^{(i)}_{2k} : i \in \mathbb{N}, k \in \{1, 2, \ldots, M\}\}$ be a collection of mutually independent normal random variables with mean zero and variance 1. Fix $\theta \in (0, 1)$ and define

$$\alpha_1 = \frac{5}{12\theta(1-\theta)} \text{ and } \alpha_2 = \alpha_1 - 1 = \frac{5 - 12\theta + 12\theta^2}{12\theta(1-\theta)} \tag{2.8}$$

In this work, we take constant discretization stepsize $h = T/N$ and so $t_i = ih$ for i = 0, 1, . . . , N. Let $Y_0 = X(0) = x_0$ and for $i = 1, 2, \ldots, N$, we repeat the following steps:

Step 1.

$$Y_i^* = Y_{i-1} + b(Y_{i-1})\theta h + \sum_{k=1}^{M} \sigma_k(Y_{i-1})v_k\eta^{(i)}_k\sqrt{\theta h} \tag{2.9}$$

Step 2.

$$\begin{aligned} Y_i &= Y_i^* + (\alpha_1 b(Y_i^*) - \alpha_2 b(Y_{i-1}))(1-\theta)h \\ &+ \sum_{k=1}^{M} \sqrt{[\alpha_1\sigma_k^2(Y_i^*) - \alpha_2\sigma_k^2(Y_{i-1})^+}\, v_k\,\eta^{(i)}_{2k}\sqrt{(1-\theta)h} \end{aligned} \tag{2.10}$$

We call such an algorithm the *weak Simpson method*. To motivate such a name, let's temporarily ignore the diffusion terms in (1.1). In addition, if we take $\theta = \frac{1}{2}$, then $\alpha_1 = \frac{5}{3}$, $\alpha_2 = \frac{2}{3}$, and (2.9) reduces to $Y_i^* = Y_{i-1} + \frac{h}{2}b(Y_{i-1})$. Next we insert it into (2.10) to obtain

$$Y_i = Y_{i-1} + \frac{h}{6}[b(Y_{i-1}) + 4b(Y_i^*) + b(Y_i^*)] \tag{2.11}$$

On the other hand, the Simpson rule approximates the deterministic integral $\int_{t_{i-1}}^{t_i} b(Y(s))\,ds$ by

$$\int_{t_{i-1}}^{t_i} b(Y(s))\,ds \approx \frac{t_i - t_{i-1}}{6}\left[b(Y(t_{i-1})) + 4b\left(Y\left(\frac{t_{i-1}+t_i}{2}\right)\right) + b(Y(t_i))\right]$$

$$= \frac{h}{6}\left[b\,(Y(t_{i-1})) + 4b\left(Y\left(t_{i-1} + \frac{h}{2}\right)\right) + b\,(Y\,(t_i))\right]$$

Compare this with the second term on the right-hand side of (2.11), and notice that

$$Y_i^* = Y_{i-1} + \frac{h}{2}\, b\,(Y\,(t_{i-1})) \approx Y\left(t_{i-1} + \frac{h}{2}\right)$$

Therefore our algorithm (2.9)–(2.10) is similar to the deterministic Simpson rule, though we use the θ-midpoint value $b(Y_i^*)$ instead of the terminal value $b(Y(t_i))$ in (2.11).

To further illustrate the idea behind the algorithm (2.9)–(2.10), we note that (1) the solution of (1.1) is equivalent in distribution to

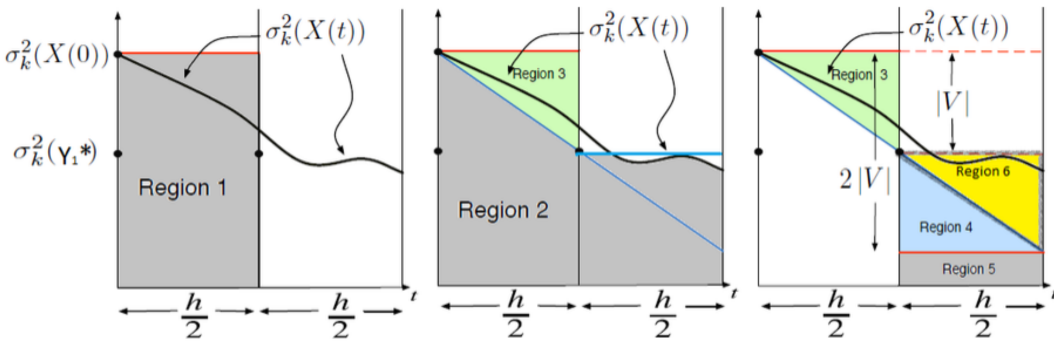
$$X(t) = x + \int_0^t b(X(s))\, ds + \sum_{k=1}^{M} \nu_k \int_0^{\infty}\int_0^t I_{[0,\sigma_k^2(X(s))]}(u)\,\xi_k\,(\mathrm{d}u \times \mathrm{d}s), \tag{2.12}$$

Where ξ_k is a time-space white noise for each k = 1, ..., M. Recall that $\{\xi_k(t, x)\}_{t\geq 0,\, x\in |^+}$ is a time-space white noise if it is an $\mathbb{F}$-adapted centered Gaussian random field with

$$\mathbb{E}[\xi\,(s,y)\;\xi\,(t,x)] = \delta\,(t-s)\;\delta\,(x-y),\ \text{for all } t, s \geq \text{ and } x, y \in |^+,$$

where, $\delta(.)$ is the delta function. In particular, it follows that is A,B are disjoint subsets of $[\,0, \infty)^2$, then $\xi_k(A)$ and $\xi_k(B)$ are independent normal random variables with means 0 and variances $|A|$ and $|B|$, respectively, where | . | denotes the Lebesgue measure on $[0,\infty)^2$. We refer to (11) for introduction to space-time white noise and stochastic partial differential equations.

To illustrate the idea, let's fix $\theta = \frac{1}{2}$ and modify a figure from (1) as in Figure 1. For simplicity, we take i = 1 in (2.9) and (2.10). In order to approximate the diffusion term in (2.12) over the interval $[0, h]$, we must approximate $\xi_k\,(A_{[0,h]}\left(\sigma_k^2\right))$, where $A_{[0,h]}\left(\sigma_k^2\right)$ is the region under the curve $\sigma_k^2\,(X\,(t))$ for $0 \leq t \leq h$. Since ξ_k is a space-time white noise, $\xi_k\,(A_{[0,h]}\left(\sigma_k^2\right)$ is normally distributed with mean zero and variance equals the area of the region $A_{[0,h]}\left(\sigma_k^2\right)$. Thus it is enough to find an accurate approximation to the area of the region $A_{[0,h]}\left(\sigma_k^2\right)$.



(a) First step (b) Desired second step (c) Used second step

Figure 1. A graphical illustration of weak Simpson scheme for $\theta = \frac{1}{2}$, where $V = \sigma_k^2\,(X\,(0)) - \sigma_k^2\,(Y_1^*)$

We would like to approximate the area under the curve $\sigma_k^2\,(X(t))$ using the Simpson rule. Recall for a positive function *f(x)*, the Simpson rule to approximate $\int_a^b f(x)\, dx$ which gives the area of the region under the curve $y = f(x)$ between *a* and *b* is given by

$$
\begin{aligned}
I\,(f) &\approx \frac{h}{6}\left[f\,(\mathrm{a}) + 4f\left(\frac{\mathrm{a}+\mathrm{b}}{2}\right) + f\,(\mathrm{b})\right] \\
&= \frac{1}{3} \cdot \frac{\mathrm{h}}{2}\,(f\,(\mathrm{a}) + f\,(\mathrm{b})) + \frac{2}{3} \,.\, hf\left(\frac{\mathrm{a}+\mathrm{b}}{2}\right) \qquad (2.13) \\
&= \frac{1}{3}\,(\text{Area of } BCDF) + \frac{2}{3}\,(\text{Area of } ACDE),
\end{aligned}
$$

Where we put $h = b - a$, *BCDF* is the trapezoid with base [*a*, *b*] and heights *f* (a) and *f* (b), and *ACDE* is the rectangle with base [*a*, *b*] and height $f\left(\frac{\mathrm{a}+\mathrm{b}}{2}\right)$; see the illustration in Figure 2.

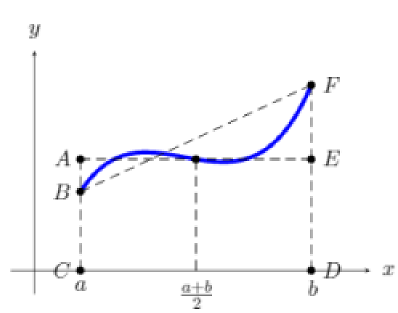


Figure2. The Simpson Rule

To approximate the area of the region $A_{[0,h]}\left(\sigma_k^2\right)$, we first note that

$$\xi_k \text{ (Region 1) } \stackrel{d}{=} N\left(0, \sigma_k^2 (X(0)) \frac{h}{2}\right) \stackrel{d}{=} \sigma_k (X(0)) \sqrt{\frac{h}{2}}\, N(0,1),$$

Which is equivalent in distribution to the summand of the right-hand side of (2.9) in Step 1 of our algorithm. Here and throughout the paper, $N(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2.

Now using the estimated θ-midpoint Y_1^* from step 1, we approximate the area under the curve $\sigma_k^2 (X(t))$. Since we used the area of the green shaded region (Region 3 in Figure 1 (b)) in (2.9) of Step 1, we need to discard this area in Step 2. Observe that Region 3 and the blue shaded region (Region 4 in Figure 1 (c)) have equal areas. Thus in Step 2, we do not add the area of Region 4.

Since Region 5 in Figure 1(c) is the part of both the rectangle and the trapezoid, we take the whole region under consideration.

$$\xi_k \text{ (Region 5) } \stackrel{d}{=} N\left(0, \left[\sigma_k^2 (X(0)) - 2(\sigma_k^2(X(0)) - \sigma_k^2 (Y_1^*))\right]\frac{h}{2}\right)$$

$$\stackrel{d}{=} N\left(0, \left[2\sigma_k^2(Y_1^*) - \sigma_k^2 (X(0))\right]\frac{h}{2}\right)$$

On the other hand, Region 6 in Figure 1(c) is part of the rectangle only, we take $\frac{2}{3}$ of the area of Region 6 only:

$$\xi_k\left(\frac{2}{3}\text{ Region 6}\right) \stackrel{d}{=} \mathrm{N}\left(0, \frac{2}{3}\cdot\frac{1}{2}\cdot\left(\sigma_k^2(X(0)) - \sigma_k^2(Y_1^*)\right)\cdot\frac{h}{2}\right)$$

$$\stackrel{d}{=} \mathrm{N}\left(0, \frac{1}{3}\left(\sigma_k^2(X(0)) - \sigma_k^2 (Y_1^*)\right)\frac{h}{2}\right).$$

Note that Region 5 and Region 6 are disjoint. Thus we have

$$\xi_k \text{ (Region 5) } + \xi_k\left(\frac{2}{3}\text{ Region 6}\right)$$

$$\stackrel{d}{=} N\left(0, \left[2\sigma_k^2 (Y_1^*) - \sigma_k^2 (X(0))\right]\frac{h}{2}\right) + N\left(0, \frac{1}{3}\left(\sigma_k^2 (X(0)) - \sigma_k^2 (Y_1^*)\right)\frac{h}{2}\right)$$

$$\stackrel{d}{=} N\left(0, \left(\frac{5}{3}\sigma_k^2(Y_1^*) - \frac{2}{3}\sigma_k^2 (X(0))\right)\frac{h}{2}\right)$$

$$\stackrel{d}{=} \sqrt{\frac{5}{3}\sigma_k^2 (Y_1^*) - \frac{2}{3}\sigma_k^2 (X(0))}\ \sqrt{\frac{h}{2}}\, N(0,1).$$

Note that for $\theta = \frac{1}{2}$ we have $\alpha_1 = \frac{5}{3}$ and $\alpha_2 = \frac{2}{3}$. Again, we find that ξ_k (Region 5) + $\xi_k \left(\frac{2}{3}\text{ Region } 6\right)$ is equivalent in distribution to the last term on the right-hand side of (2.10) of Step 2. Therefore, it is reasonable to anticipate that the weak Simpson algorithm shall work well. We note that the weak trapezoid method in (1) does not consider Region 6.

2.3 The Weak Convergence Rage. Similar to the notion in (6), we give the following definition:

Definition 2.4. We say that the approximating sequence $Y = Y^h$ converges weakly to X at time T with order $\beta > 0$ if

$$|\mathbb{E}[f(X(T))] - \mathbb{E}[f(Y(N))]| \leq C \, || f||_k \, h^{\beta} \text{ for all } f \in C^k(\mathbb{R}^d) \tag{2.14}$$

where $k = 2(\beta + 1)$, $Nh = T$, and $C = C(T) > 0$ is a constant. In terms of the induced operator norm. (2.14) can be rewritten equivalently as

$$|| \rho_t - P_h^N ||_{k \to 0} \leq Ch^{\beta} \tag{2.15}$$

Next we state the following assumption, which is necessary to show that the weak Simpson method (2.9) – (2.10) has weak convergence order three.

(A4) For any sufficiently smooth functions $f\colon \mathbb{R}^d \to \mathbb{R}$, we have

$$(A^2 f)(x_0) = (B_1^2 f)(x_0), \text{ and} \tag{2.16}$$

$$(A^3 f)(x_0) = (B_1^2(Af))(x_0) = (B_1 g)(x_0) = (B_1(A(B_1 f)))(x_o) = (B_1^3 f)(x_0), \tag{2.17}$$

for all $x_0 \in \mathbb{R}$, where

$$g(x) = f''[b(x), b(x)](x) + \sum_{k=1}^{M} \sigma_k^2(x) f'''[v_k, v_k, b(x)](x)$$

$$+ \frac{1}{4} \sum_{k,j=1}^{M} \sigma_j^2(x)\, \sigma_k^2(x)\, f^{(4)}[v_k, v_k, v_j, v_j](x), \tag{2.18}$$

$$(Af)(x) = f'[b(x)](x) + \frac{1}{2} \sum_{k=1}^{M} \sigma_k^2(x) f''[v_k, v_k](x), \tag{2.19}$$

$$(B_1 f)(x) = f'[b(x_0)](x) + \frac{1}{2} \sum_{k=1}^{M} \sigma_k(x_0)^2 f''[v_k, v_k](x). \tag{2.20}$$

and we define $(A^n f)(x) = (A(A^{n-1} f))(x)$ for any integer $n \geq 2$, and similarly for B_1 and B.

The following theorem gives the weak local convergence rate of the weak Simpson method (2.9) – (2.10).

Theorem 2.5 (Local Approximation). *Assume* (A1) – (A4). *Then there exist a constant k so that*

$\|\rho_h - P_h\|_{\infty \to 0} \leq kh^4$ for all $h > 0$ *sufficiently small.*

The proof of Theorem 2.5 depends on the following lemma, whose proof is relegated to Appendix A.

Lemma 2.6. *Assume* (A1) – (A4). *Then for all* $h > 0$ *sufficiently small and* $f \in C^\infty$ *we have*

$$\mathbb{E}\left[f(Y_1^*) + (Bf)(Y_1^*)(1-\theta)h + (B^2 f)(Y_1^*)\frac{(1-\theta)^2h^2}{2} + (B^3 f)(Y_1^*)\frac{(1-\theta)^3h^3}{6}\right] \tag{2.21}$$

$$= f(x_0) + (Af)(x_0)h + (A^2 f)(x_0)\frac{h^2}{2} + (A^3 f)(x_0)\frac{h^3}{6} + O(h^4),$$

where

$$(Bf)(x) = f'[\alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](x) + \frac{1}{2}\sum_{k=1}^{M}[\alpha_1\sigma_k^2(Y_1^*) - \alpha_2\sigma_k^2(x_0)]^+ f''[v_k, v_k](x). \tag{2.22}$$

Proof of Theorem 2.5. We need to show that for any $f \in C^\infty$ and $h > 0$ sufficiently small, there exists a constant $K > 0$ so that

$$|\mathbb{E}_{x_0}[f(Y_1)] - \mathbb{E}_{x_0}[f(X(h))]| \leq K\|f\|_\infty h^4, \text{ for all } x_0 \in \mathbb{R}^d \tag{2.23}$$

To this end, we consider the stochastic differential equation.

$$dy(t) = b(x_0)dt + \sum_{k=1}^{M}\sigma_k(x_0)v_k\, dW_k(t),\ t \geq 0,\ y(0) = x_0. \tag{2.24}$$

Then we have,

$$y(\theta h) = x_0 + b(x_0)\,\theta h + \sum_{k=1}^{M}\sigma_k(x_0)v_k(W_k(\theta h) - W_k(0)). \tag{2.25}$$

Since $W_k(\theta h) - W_k(0) \overset{d}{=} N(0, \theta h) \overset{d}{=} N(0, 1)\sqrt{\theta h}$, we see that (2.9) in Step 1 of the weak Simpson algorithm produces a value Y_1^* which is equal to $y(\theta h)$ in distribution. Similarly, (2.10) in Step 2 produces a value Y_1 that is equivalent to $z(h)$ in distribution, where $z(t)$ solves the stochastic differential equation

$$\begin{cases} dz(t) = (\alpha_1 b(Y_1^*) - \alpha_2 b(x_0))\,dt + \sum_{k=1}^{M}\sqrt{[\alpha_1\sigma_k^2(Y_1^*) - \alpha_2\sigma_k^2(x_0)]^+}\, v_k\, dW_k(t),\ t \geq \theta h, \\ z(\theta h) = Y_1^*. \end{cases} \tag{2.26}$$

Recall the definitions of the operators A and B in (2.19) and (2.22), respectively. Note that A is the infinitesimal generator of (1.1) and B the infinitesimal generator of (2.26).

Let F_t denote the filtration generated by the Brownian motion processes $W_k(t)$ in (2.24) and (2.26). Then

$$\mathbb{E}\,[f(z(h))] = \mathbb{E}[\mathbb{E}[f(z(h))\,|F_{\theta h}]] \overset{def}{=} \mathbb{E}\,[\mathbb{E}_{\theta h}[f(z(h))]], \tag{2.27}$$

where we have defined $\mathbb{E}_{\theta h}[.] \overset{def}{=} \mathbb{E}[.|F_{\theta h}]$. Let $z(h)$ be the solution to (2.26). Since $f \in C^8$, we can use Dynkin's formula repeatedly to obtain

$$\begin{aligned}
\mathbb{E}_{\theta h}[f(z(h))] &= f(Y_1^*) + \int_{\theta h}^{h} \mathbb{E}_{\theta h}[(Bf)(z(s))]\,ds \\
&= f(Y_1^*) + (Bf)(Y_1^*)(1-\theta)h + \int_{\theta h}^{h}\int_{\theta h}^{s} \mathbb{E}_{\theta h}[(B^2 f)(z(r))]\,dr\,ds \\
&= f(Y_1^*) + (Bf)(Y_1^*)(1-\theta)h + (B^2 f)(Y_1^*)\frac{(1-\theta)^2h^2}{2} \\
&\quad + \int_{\theta h}^{h}\int_{\theta h}^{s}\int_{\theta h}^{r} \mathbb{E}_{\theta h}[B^3 f)(z(u))]\,du\,dr\,ds \\
&= f(Y_1^*) + (Bf)(Y_1^*)(1-\theta)h + (B^2 f)(Y_1^*)\frac{(1-\theta)^2h^2}{2} + (B^3 f)(Y_1^*)\frac{(1-\theta)^3h^3}{6} \\
&\quad + \int_{\theta h}^{h}\int_{\theta h}^{s}\int_{\theta h}^{r}\int_{\theta h}^{v} \mathbb{E}_{\theta h}[(B^4 f)(z(w))]\,dw\,du\,dr\,ds
\end{aligned} \tag{2.28}$$

The term $(B^4 f)(y(w))$ in the last integral above depends only on the first eight derivatives of f. Therefore, using the fact that $f \in C^8$ and Lemma 2.1, we obtain

$$\left| \int_{\theta h}^{h}\int_{\theta h}^{s}\int_{\theta h}^{r}\int_{\theta h}^{v} \mathbb{E}_{\theta h}[(B^4 f)(z(w))]\,dw\,du\,dr\,ds \right| \le C_1 \,\|f\|_8\, h^4 \tag{2.29}$$

for some constant C_1 independent of h.

Recall that Y_1 of (2.10) and $z(h)$ of (2.26) have the same distribution, and in particular, we have $\mathbb{E}_{X_0}[f(Y_1)] = \mathbb{E}_{X_0}[f(z(h))]$. Then it follows from (2.27), (2.28), and (2.29) that

$$\begin{aligned}
\mathbb{E}_{X_0}[f(Y_1)] &= \mathbb{E}_{X_0}[f(z(h))] \\
&= \mathbb{E}_{X_0}\Big[f(Y_1^*) + (Bf)(Y_1^*)(1-\theta)h + (B^2 f)(Y_1^*)\frac{(1-\theta)^2h^2}{2}
\end{aligned}$$

$$+ (B^3 f)(Y_1^*) \frac{(1-\theta)^3 h^3}{6} + O(h^4) \Big]$$

On the other hand, proceeding as above and applying Dynkin's formula to (1.1) repeatedly gives

$$\mathbb{E}_{x_0} [f(X(h))] = f(x_0) + (Af(x_0)h + (A^2 f)(x_0) \frac{h^2}{2} + (A^3 f)(x_0) \frac{h^3}{6} + O(h^4).$$

Then (2.23) follows from Lemma 2.6 and the above two displayed equations. This completes the proof of the theorem.

With Theorem 2.5 at our hands, we can proceed to derive the global weak convergence rate for the weak Simpson method (2.9) – (2.10).

Theorem 2.7. *Assume* (A1) – (A4). *Then for any $T > 0$ there exists a constant $C(T) > 0$* such that

$$\sup_{0 \le nh \le T} \left\| \rho_{nh} - P_{\text{h}}^{\text{n}} \right\|_{8 \to 0} \le C(T) h^3 \tag{2.30}$$

Proof. This theorem follows directly from Theorem 2.5; the argument is similar to that of (1, Theorem 2.4). We shall omit the details here.

2.4. Examples. To illustrate the main results in Section 2.3, we present three examples in this subsection. We start with the one-dimensional geometric Brownian motion (2.31) in Example 2.8. so that we can compare the numerical computations using the weak Simpson method (2.9) – (2.10) with the theoretical values. Next we consider two nonlinear two dimensional SDEs which are investigated in (1). In particular, we want to compare the performance of the weak Simpson method with that of the weak trapezoidal method proposed in (1).

Example 2.8. We consider the one-dimensional geometric Brownian motion in the following form

$$dX(t) = \lambda X(t)\, dt + \mu X(t) dW(t),$$

$$X(0) = X(0) \in |, \tag{2.31}$$

where λ and μ are real constants. The solution to (2.31) is $X(t) = X(0) e^{\left(\lambda - \frac{1}{2}\mu^2\right)t + \mu W(t)}$ and we have

$$\mathbb{E}[X(t)] = \mathbb{E}[X(0)]\, e^{\lambda t}, \quad \mathbb{E}[(X(t)^2] = \mathbb{E}[X(0)^2]\, e^{(2\lambda + \mu^2)t}$$

We test the performance of the weak Simpson method (2.9) – (2.10) for (2.31). We take $T = 1$, $\lambda = 3$, $\mu = 0.3$, and $X(0) = 1$ in (2.31). Also, we use step sizes $h_p = \frac{1}{100p}$, for $p = 1, \ldots\ldots, 5$ to generate $N = 48, 200$ sample paths of (2.31) using the weak Simpson algorithm (2.9) – (2.10), We then compute

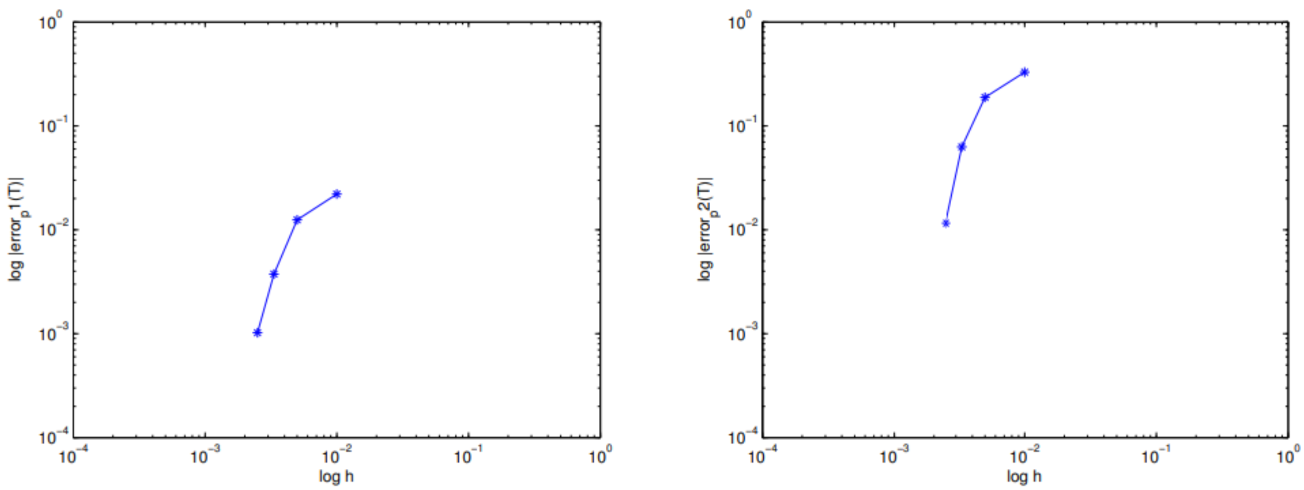
$$\text{Error}\,{}^{1}_{\text{p}}(1) = \mathbb{E}[X(1)] - \frac{1}{N} \sum_{k=1}^{N} Y_{N_p}^{k, h_p}, \tag{2.32}$$

$$\text{Error}_1^2(t) = \mathbb{E}\,|X(1)^2] - \frac{1}{N}\sum_{i=1}^{N}\left(Y_{N_p}^{k,h_p}\right)^2. \tag{2.33}$$

where $N_p = \frac{1}{h_p}$, and $Y_{N_p}^{k,h_p}$ is the value obtained using the weak Simpson method (2.9) – (2.10) in the k^{th} simulation with step size h_p. The results are plotted in Figure 3 in the log-log scale. Note that in both plots of Figure 3, the best fit lines have slope 3, as expected.

Example 2.9. Here we investigate the same example considered in (1).

$$\begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} X_1(t) \\ 0 \end{bmatrix} dt + X_1(t)\begin{bmatrix} 0 \\ 1 \end{bmatrix} dW_1(t) + \frac{1}{10}\begin{bmatrix} 1 \\ 1 \end{bmatrix} dW_2(t), \tag{2.34}$$



(a) $\text{Error}_p^1(1)$ of (2.32) (b) $\text{Error}_p^2(1)$ of (2.33)

***Figure* 3.** Log-Log plots of error versus step-size for example 2.8

where $W_1(t)$ and $W_2(t)$ are independent standard one-dimensional Brownian motion processes. It is straight forward to show that

$$\mathbb{E}[X_2(t)^2] = \mathbb{E}\,[X_2(0)^2] - \frac{1}{2}\,\mathbb{E}\,[X_1(0)^2] + \frac{1}{400}\,e^{2t}\,(200\,\mathbb{E}\,[X_1(0)^2] + 1) + \frac{t}{200} - \frac{1}{400}\,. \tag{2.35}$$

Next we use (2.35) to compute the error

$$\text{Error}_p(1) = \mathbb{E}\,[X_2(1)^2] - \frac{1}{N}\sum_{k=1}^{N}\left(Y_{2,N_p}^{k,h_p}\right)^2, \tag{2.36}$$

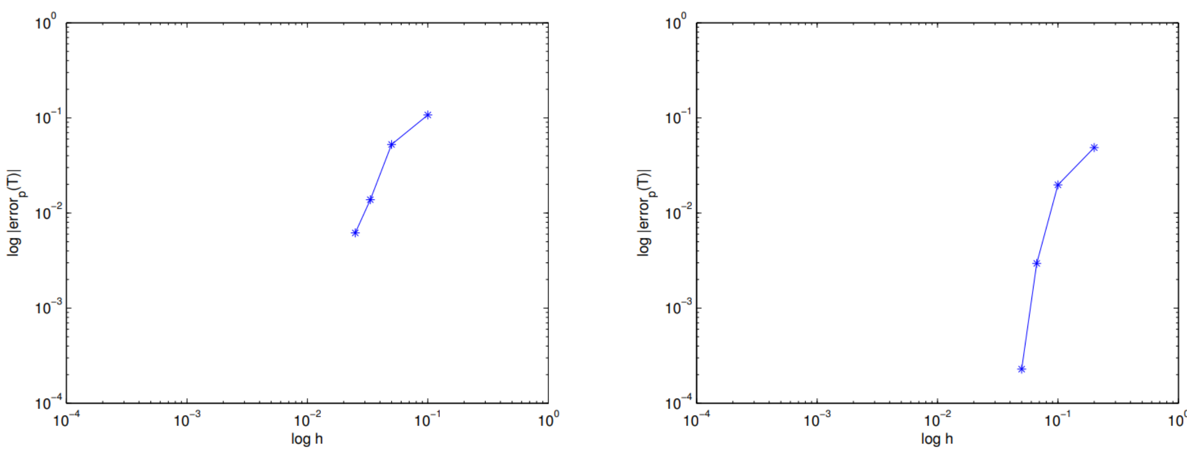
by generating N=46,300 sample paths of (2.34), where $h_p = \dfrac{1}{10p}$ for $p = 1,\ldots,4$. $N_p = \dfrac{1}{h_p}$, and $\left(Y^{k,h_p}_{1,N_p}, Y^{k,h_p}_{2,N_p}\right)' \in \mathbb{R}^2$ is the approximated value of (2.34) obtained using the weak Simpson method (2.9) – (2.10) in the k^{th} simulation, for $1 \le k \le N$. The resulting error is displayed in Figure 4(a), where we have plotted the weak error against h_p on log-log scale. We observe that the slope of the best fit line in Figure 4(a) is three as the step size h tends to zero.

Example 2.10. Here we consider the system

$$\begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} -X_2(t) \\ X_1(t) \end{bmatrix} dt + \sqrt{\frac{\sin^2(X_1(t) + X_2(t))}{t+1}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} dW_1(t) + \sqrt{\frac{\cos^2(X_1(t) + X_2(t))}{t+1}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} dW_2(t). \tag{2.37}$$

where $W_1(t)$ and $W_2(t)$ are independent standard one-dimensional Brownian motions. The system (2.37) is similar to the one considered in (1). Then $\mathbb{E}\,[|X(t)|^2]$ can be calculated as

$$\mathbb{E}\,[|X(t)|^2] = \mathbb{E}[X(0)^2] + \log(1+t). \tag{2.38}$$



(a) $\text{Error}_p(1)$ of (2.36) in Example 2.9 (b) $\text{Error}_p(1)$ of (2.39) in Example 2.10

***Figure* 4.** Log-Log plots of error versus step-size for Examples 2.9 and 2.10.

We generate N=51,000 different sample paths of (2.37) using the weak Simpson algorithm (2.9) – (2.10) with step sizes $h = \dfrac{1}{5p}$ for $1 \le p \le 4$ and initial condition $X(0) = (1, 1)'$. We then compute

$$\text{Error}_p(1) = [|X(1)|^2] - \frac{1}{N}\sum_{k=1}^{N} \left| Y^{k,h_p}_{N_p} \right|^2, \tag{2.39}$$

where $N_p = \frac{1}{h_p}$, for $1 \leq k \leq 51{,}000$, $Y_{N_p}^{k,h_p} \in |^2$ is the k^{th} numerical value of (2.37) obtained from the weak Simpson method and $\mathbb{E}[|X(1)|^2]$ is from (2.38). The result of numerical experiment is shown in Figure 4 (b), where we have plotted the error against h on log-log scale. It is observed that the slope of the best fit line is three.

Remark 2.11. We note that the weak trapezoidal method in (1) gives a weak convergence order two; and requires 10 million sample paths to obtain such a convergence order for Examples 2.9 and 2.10. In contrast, the weak Simpson method (2.9)–(2.10) gives a weak convergence order three and requires only approximately 5×10^4 sample paths. Therefore, it is a substantial improvement over the method in (1).

Remark 2.12. The computation of the error may be sometimes influenced by different kind of errors like sampling error, random number bias and rounding error. In our algorithm we generate more than one sample of random numbers. So, for large number of sample path there is a greater chance of dependency in the samples and might degrade the order of convergence of our algorithm.

Appendix A. The Proof of Lemma 2.6

The proof of Lemma 2.6 depends on the following Lemmas.

Lemma A.1. *Let X, Z, W, and Y be real valued random variables on some probability space* $(\Omega, F, \mathbb{P})$. *Let* $p, q > 1$ *with* $\frac{1}{p} + \frac{1}{q} = 1$. *Then the following assertions are true:*

(i) *if* $\|XY\|_{L^p(\Omega)} < \infty$, *then*

$$\mathbb{E}\,[|YX^+ - YX|] \leq \|XY\|_{L^p(\Omega)}(\mathbb{P}\{X < 0\})^{\frac{1}{q}},$$

(ii) *if* $\|ZXY\|_{L^p(\Omega)} < \infty$, *then*

$$\mathbb{E}\,[|ZY^+X^+ - ZYX|] \leq \|XYZ\|_{L^p(\Omega)}(\mathbb{P}\{X < 0 \text{ or } Z < 0\})^{\frac{1}{q}},$$

(iii) *if* $\|WXYZ\|_{L^p(\Omega)} < \infty$, *then*

$$[|ZW^+Z^+Y^+ - ZWXY|] \leq \|WYXZ\|_{L^p(\Omega)}(\mathbb{P}\{W < 0 \text{ or } X < 0 \text{ or } Y < 0\})^{\frac{1}{q}}.$$

Proof: We only prove assertion (iii) here; the proofs of (i) and (ii) are similar and can be found in (1). Let $A := \{W < 0 \text{ or } X < 0 \text{ or } Y < 0\}$. Then we have $A^c = \{W \geq 0 \text{ and } X \geq 0 \text{ and } Y \geq 0\}$. Moreover, on the set A^c, $ZW^+X^+Y^+ = ZWXY$. Thus it follows from the Holders inequality that

$$\mathbb{E}\,[|ZW^+X^+Y^+ - ZWXY|] = \mathbb{E}\,[| - ZWXYI_A|] \leq \|ZWXY\|_{L^p(\Omega)}\, \mathbb{P}(A)^{1/q}.$$

Lemma A.2. *Suppose Assumptions (A1) and (A2). Then for any $p > 0$, there exists and $h_0 > 0$ so that*

$$\mathbb{P}\,\{\alpha_1\, \sigma_k^2\,(Y_1^*) - \alpha_2\, \sigma_k^2\,(x_0) < 0\} = O(h^p) \text{ for all } 0 < h < h_0, \tag{A.1}$$

where Y_1^* is given in (2.9).

Proof: Denote $E_k := \{\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0) < 0\}$. As we noted before, Y_1^* of (2.9) and $y(\theta h)$ of (2.25) have the same distribution. Thus

$$\mathbb{P}(E_K) = \mathbb{P}\{\alpha_1 \sigma_k^2 (y(\theta h)) - \alpha_2 \sigma_k^2 (x_0) < 0\} = \mathbb{P}\left\{|\sigma_k(y(\theta h))| < \sqrt{\alpha_2/\alpha_1}\, |\sigma_k(x_0)|\right\} \tag{A.2}$$

$$= \mathbb{P}\left\{|\sigma_k(y(\theta h))| - |\sigma_k(x_0)| < \left(\sqrt{\alpha_2/\alpha_1} - 1\right) |\sigma_k(x_0)|\right\}$$

On the other hand, using the triangle inequality and (2.1), we have

$$|\sigma_k(y(\theta h))| - |\sigma_k(x_0)| \geq -|\sigma_k(y(\theta h)) - \sigma_k(x_0)| \geq -L\,|y(\theta h) - x_0|.$$

Putting this into (A.2), we have

$$\mathbb{P}(E_k) \leq \mathbb{P}\left\{-L\,|y(\theta h) - x_0| \leq \left(\sqrt{\alpha_2/\alpha_1} - 1\right) |\sigma_k(x_0)|\right\} = \mathbb{P}\{|y(\theta h) - x_0| \geq C\}$$

$$= \mathbb{P}\left\{\left| b(x_0)\theta h + \sum_{k=1}^{M} \sigma_k(x_0)\, v_k\, (W_k(\theta h) - W_k(0))\right| \geq C\right\}$$

$$\leq \mathbb{P}\left\{\left| \sum_{k=1}^{M} \sigma_k(x_0)\, v_k\, (W_k(\theta h) - W_k(0))\right| \geq C - |b(x_0)|\,\theta h\right\}$$

$$= \mathbb{P}\left\{\left| \sum_{k=1}^{M} \sigma_k(x_0)\, v_k \frac{W_k(\theta h) - W_k(0)}{\sqrt{\theta h}}\right| \geq \frac{C - |b(x_0)|\,\theta h}{\sqrt{\theta h}}\right\}$$

where $C := \frac{1}{L}\left(1 - \sqrt{\alpha_2/\alpha_1}\right) |\sigma_k(x_0)|$. Recall from (2.8) that $0 < \alpha_2 = \alpha - 1 < \alpha_1$. Consequently $C > 0$. Note that for each $k = 1, \ldots, M$ and $h > 0$, $Z_k := \frac{W_k(\theta h) - W_k(0)}{\sqrt{\theta h}}$ has standard normal distribution. This, together with the assumption that $W_1, \ldots, W_M$ are independent Brownian motions, implies that

$\sum_{k=1}^{M} \sigma_k(x_0) v_k \frac{W_k(\theta h) - W_k(0)}{\sqrt{\theta h}}$ has multivariate normal distributions with mean zero and covariance matrix $\sum_{k=1}^{M} \sigma_k^2(x_0)\, v_k v_k^T = a(x_0)$. Now (A.1) follows from (2.2) and the usual Gaussian tail estimation (see, for instance, Theorem 1 of Hüsler et al. (2002)).

Corollary A.3. *Assume the conditions of Lemma A.2. Suppose* $f \in C^8(\mathbb{R}^d)$ *and that for all multi-index* α *with* $|\alpha| \le 8$, *we have*

$$|D^{\alpha} f(x)| \le K(1 + |x|^q) \tag{A.3}$$

for some positive constants K and $p \ge 1$. *Then for any* $k, j, l = 1, \ldots, M$ *and* $p \ge 1$, *there exists an* $h_0 > 0$ *so that for all* $h \in (0, h_0]$, *we have*

$$\mathbb{E}\,[\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0))^+ f''[v_k, v_k](Y_1^*)] \tag{A.4}$$

$$= \mathbb{E}\left[(\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0)) f''[v_k, v_k](Y_1^*)\right] + O(h^p),$$

$$\mathbb{E}\left[[\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0)]^+ \; [\alpha_1 \sigma_j^2(Y_1^*) - \alpha_2 \sigma_j^2(x_0)]^+ f^{(4)}[v_k, v_k, v_j, v_j](Y_1^*)\right] \tag{A.5}$$

$$= \mathbb{E}\left[[\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0)]\,[\alpha_1 \sigma_j^2(Y_1^*) - \alpha_2 \sigma_j^2(x_0)]\, f^{(4)}[v_k, v_k, v_j, v_j](Y_1^*)\right] + O(h^p),$$

and

$$\mathbb{E}\left[[\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0)]^+ [\alpha_1 \sigma_j^2(Y_1^*) - \alpha_2 \sigma_j^2(x_0)]^+ [\alpha_1 \sigma_l^2(Y_1^*) - \alpha_2 \sigma_l^2(x_0)\right]^+ \tag{A.6}$$

$$\times f^{(6)}[[v_k, v_k, v_j, v_j, v_l, v_l](Y_1^*)]$$

$$= \mathbb{E}\left[[\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0)\right]\left[[\alpha_1 \sigma_j^2(Y_1^*) - \alpha_2 \sigma_j^2(x_0)\right]\left[[\alpha_1 \sigma_l^2(Y_1^*) - \alpha_2 \sigma_l^2(x_0)\right]$$

$$\times f^{(6)}[[v_k, v_k, v_j, v_j, v_l, v_l](Y_1^*)] + O(h^P),$$

Proof: As observed in the proof of Lemma A.2, Y_1^* is equal to $y(\theta h)$ in distribution, where $y(\theta h)$ is given by (2.25). Therefore, in view of (A.3), the standard arguments as those in (9) or (12) yield

$$\|(\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0)) f''[v_k, v_k](Y_1^*)\|_{L^p} = \|(\alpha_1 \sigma_k^2(y(\theta h)) - \alpha_2 \sigma_k^2(x_0)) f''[v_k, v_k](y(\theta h))\|_{L^p}$$

$$\le K < \infty.$$

Then (A.4) follows from Lemmas A.1 and A.2.

Observe that

$$\mathbb{P}\left\{\left[\alpha_1 \sigma_k^2(Y_1^*) - \alpha_2 \sigma_k^2(x_0) < 0\right] \text{or} \left[\alpha_1 \sigma_j^2(Y_1^*) - \alpha_2 \sigma_j^2(x_0) < 0\right]\right\}$$

$$\leq \mathbb{P}\left\{ \alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0) < 0 \right\} + \mathbb{P}\left\{ \alpha_1 \sigma_j^2 (Y_1^*) - \alpha_2 \sigma_j^2 (x_0) < 0 \right\}.$$

Then (A.5) follows from a similar argument as above using Lemmas A.1 and A.2.

In a similar fashion, we can establish (A.6).

Lemma A.4. *For any sufficiently smooth function* $f: \mathbb{R}^d \to \mathbb{R}$, *we have*

$$B_1^2 f(x) = f''[b(x_0), b(x_0)](x) + \sum_{k=1}^{M} \sigma_k^2 (x_0) f'''[v_k, v_k, b(x_0)](x)$$

$$+ \frac{1}{4} \sum_{k,j=1}^{M} \sigma_k^2 (x_0)\, \sigma_j^2 (x_0) f^{(4)} [v_k, v_k, v_j, v_j](x),$$

$$A(B_1 f)(x) = f''[b(x), b(x_0)](x) + \frac{1}{2} \sum_{k=1}^{M} \sigma_k^2 (x_0) f'''[v_k, v_k, b(x)](x)$$

$$+ \frac{1}{2} \sum_{k=1}^{M} \sigma_k^2 (x) f'''[v_k, v_k, b(x_0)](x) + \frac{1}{4} \sum_{k,j=1}^{M} \sigma_k^2 (x)\, \sigma_j^2 (x_0) f^{(4)} [v_k, v_k, v_j, v_j](x),$$

$$B^2 f(x) = f''[\alpha_1 b(Y_1^*) - \alpha_2 b(x_0), \alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](x)$$

$$+ \sum_{k=1}^{M} (\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0))^+ f'''[v_k, v_k, \alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](x)$$

$$+ \frac{1}{4} \sum_{k,j=1}^{M} (\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0))^+ (\alpha_1 \sigma_j^2 (Y_1^*) - \alpha_2 \sigma_j^2 (x_0))^+ f^{(4)} [v_k, v_k, v_j, v_j](x),$$

$$B_1^3 f(x) = f'''[b(x_0), b(x_0), b(x_0)](x) + \frac{3}{2} \sum_{k=1}^{M} \sigma_k^2 (x_0) f^{(4)} [v_k, v_k, b(x_0), b(x_0)](x)$$

$$+ \frac{3}{4} \sum_{k,j=1}^{M} \sigma_k^2 (x_0)\, \sigma_j^2 (x_0) f^{(5)} [v_k, v_k, v_j, v_j, b(x_0)](x)$$

$$+ \frac{1}{8} \sum_{k,j,l=1}^{M} \sigma_k^2 (x_0)\, \sigma_j^2 (x_0)\, \sigma_l^2 (x_0) f^{(6)} [v_k, v_k, v_j, v_j, v_l, v_l](x)$$

and

$$B^3 f(x) = f''' [\alpha_1 b (Y_1^*) - \alpha_2 b (x_0), \alpha_1 b(Y_1^*) - \alpha_2 b (x_0), \alpha_1 b (Y_1^*) - \alpha_2 b (x_0)] (x)$$

$$+ \frac{3}{2} \sum_{k=1}^{M} (\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0))^+ \; f^{(4)} [v_k, v_k, \alpha_1 b (Y_1^*) - \alpha_2 b (x_0) , \alpha_1 b (Y_1^*) - \alpha_2 b (x_0)] (x)$$

$$+ \frac{3}{4} \sum_{k,j=1}^{M} (\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0))^+ (\alpha_1 \sigma_j^2 (Y_1^*) - \alpha_2 \sigma_j^2 (x_0))^+$$

$$\times f^{(5)} [v_k, v_k, v_j, v_j, \; \alpha_l \, b(Y_l^*) - \alpha_2 b (x_0)] (x)$$

$$\frac{1}{8} \sum_{k,j=1}^{M} (\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0))^+ (\alpha_1 \sigma_j^2 (Y_1^*) - \alpha_2 \sigma_j^2 (x_0))^+ \; \alpha_1 \sigma_l^2 \; (Y_1^*) - \alpha_2 \sigma_l^2 \; (x_0))^+$$

$$\times f^{(6)} [v_k, v_k, v_j, v_j, v_1, v_1] (x).$$

***Proof**:* This lemma follows from straight forward and tedious calculations. We shall omit the details here.

Now we prove Lemma 2.6.

Proof of Lemma 2.6. We analyze every term on the left hand side of (2.21).

Step 1. Since Y_1^* is equal to $y(\theta h)$ of (2.25) in distribution, and noting that the infinitesimal generator of (2.24) is given by B_1 in (2.20), we can apply the Dynkin formula repeatedly to obtain

$$\mathbb{E} [f(Y_1^*) \; = f(x_0) + (B_1 f) (x_0) \theta h + (B_1^2 f) (x_0) \frac{\theta^2 h^2}{2} + (B_1^3 f) (x_0) \frac{\theta^3 h^3}{6} + O(h^4) \tag{A.7}$$

$$= f(x_0) + (Af) (x_0) \; \theta h + (B_1^2 f) (x_0) \frac{\theta^2 h^2}{2} + (B_1^3 f) (x_0) \frac{\theta^3 h^3}{6} + O(h^4),$$

where the second equality follows from the observation that $(B_1 f) (x_0) = (Af) (x_0)$.

Step 2. Next we deal with the term $\mathbb{E} [(Bf) (Y_1^*)]$. It follows from the definition of B_2 and (A.4) that

$$\mathbb{E} [(Bf) (Y_1^*)] = \mathbb{E} \left[\left(f' [\alpha_1 \, b (Y_1^*) - \alpha_2 b (x_0)] + \frac{1}{2} \sum_{k=1}^{M} [\alpha_1 \, \sigma_k^2 (Y_1^*) - \alpha_2 \, \sigma_k^2 \; (x_0)]^+ f'' \, |v_k , v_k| \right) (Y_1^*) \right]$$

$$= \mathbb{E} \left[f' [\alpha_1 \, b(Y_1^*) - \alpha_2 b (x_0)] (Y_1^*) \right]$$

$$+ \frac{1}{2} \sum_{k=1}^{M} \mathbb{E} \left[[\alpha_1 \sigma_k^2 (Y_1^*) - \alpha_2 \sigma_k^2 (x_0)] \, f'' \, [v_k , v_k] (Y_1^*) \right] + O(h^3)$$

$$= \mathbb{E}\,[f'\,[\alpha_1 b\,(y(\theta h)) - \alpha_2 b\,(x_0)]\,(y(\theta h))]$$

$$+\frac{1}{2}\sum_{k=1}^{M}\mathbb{E}\Big[[\alpha_1\,\sigma_k^2\,(y(\theta h)) - \alpha_2\sigma_k^2\,(x_0)]\,f''\,[v_k\,,v_k]\,(y(\theta h))\Big] + O(h^3).$$

In the above, we again used the fact that Y_1^* and $y(\theta h)$ of (2.25) have the same distribution to obtain the third equality. Moreover, since

$$f'\,[\alpha_1 b(y(\theta h)) - \alpha_2 b\,(x_0)]\,(y(\theta h)) = \alpha_1 b\,(y(\theta h)).\,Df\,(y(\theta h)) - \alpha_2 b(x_0).\,Df\,(y(\theta h))$$

$$= \alpha_1 f'\,[b(y(\theta h))]\,(y(\theta h)) - \alpha_2 f'\,[b(x_0)]\,(y\,(\theta h)),$$

and for each $k = 1, \ldots..,M$.

$$[\alpha_1\sigma_k^2\,(y\,(\theta h)) - \alpha_2\sigma_k^2\,(x_0)]\,f''\,[v_k,\,v_k]\,(y\,(\theta h))$$

$$= \alpha_1\sigma_k^2\,(y(\theta h))\,f''\,[v_k\,,v_k]\,(y\,(\theta h)) - \alpha_2\sigma_k^2\,(x_0)\,f''\,[v_k\,,v_k]\,(y\,(\theta h)),$$

We have

$$\mathbb{E}\,[(Bf\,)\,(Y_1^*)] = \alpha_1\mathbb{E}\left[f'\,[\,b\,(y\,(\theta h))]\,(y\,(\theta h)) + \frac{1}{2}\sum_{k=1}^{M}\sigma_k\,(y\,(\theta h))^2 f''\,\,[\,v_k\,,v_k]\,(y\,(\theta h))\right]$$

$$-\,\alpha_2\,\mathbb{E}\left[f'\,[b\,(x_0)]\,(y\,(\theta h)) + \frac{1}{2}\sum_{k=1}^{M}\sigma_k\,\,(x_0)^2 f''\,[v_k,\,v_k]\,(y\,(\theta h))\right] \quad \text{(A.8)}$$

$$= \alpha_1\,\mathbb{E}\,[Af\,(y\,(\theta h))] - \alpha_2\,\mathbb{E}\,[B_1 f\,(y\,(\theta h))],$$

where A and B_1 are defined in (2.19) and (2.20); they are the infinitesimal generators for the stochastic differential equations (1.1) and (2.24), respectively. Now we apply Dynkin's formula repeatedly to obtain

$$\mathbb{E}\,[Af\,(y\,(\theta h))] = Af\,(x_0) + \int_0^{\theta h}\mathbb{E}\,[B_1\,(Af)\,(y\,(s))]\,\mathrm{d}s$$

$$= Af\,(x_0) + B_1\,(Af)\,(x_0)\,\,\theta h + \int_0^{\theta h}\int_0^{s}[B_1^2\,(Af\,)(y(r))]\,\mathrm{d}r\mathrm{d}s$$

$$= Af\,(x_0) + B_1\,(Af)\,(x_0)\,\,\theta h + B_1^2\,\,(Af)\,(x_0)\,\frac{\theta^2 h^2}{2} \quad \text{(A.9)}$$

$$+\int_0^{\theta h}\int_0^{s}\int_0^{r}\mathbb{E}\,[B_1^3\,(Af\,(y\,(u))]\,\mathrm{d}u\,\mathrm{d}r\,\mathrm{d}s$$

$$= Af(x_0) + B_1(Af)(x_0)\theta h + B_1^2(Af)(x_0)\frac{\theta^2 h^2}{2} + O(h^3)$$

Similarly, we have

$$\mathbb{E}[B_1 f(y(\theta h))] = B_1 f(x_0) + B_1^2 f(x_0)\,\theta h + B_1^3 f(x_0)\frac{\theta^2 h^2}{2} + O(h^3). \qquad \text{(A. 10)}$$

Notice that $B_1 f(x_0) = Af(x_0)$ and hence $B_1(Af)(x_0) = A(Af)(x_0) = A^2 f(x_0)$. Using these observations in (A.9) and (A.10) and plugging them into (A.8) and noting $\alpha_1 - \alpha_2 = 1$, we have

$$\mathbb{E}[(Bf)(Y_1^*)] = (Af)(x_0) + \alpha_1(A^2 f)(x_0)\,\theta h - \alpha_2(B_1^2 f)(x_0)\theta h$$

$$+ \alpha_1 B_1^2(Af)(x_0)\frac{\theta^2 h^2}{2} - \alpha_2(B_1^3 f)(x_0)\frac{\theta^2 h^2}{2} + O(h^3). \qquad \text{(A. 11)}$$

Step 3. Next we evaluate $\mathbb{E}(B^2 f)(Y_1^*)$. Thanks to Lemma A.4 and Corollary A.3, we have

$$\mathbb{E}[B^2 f(Y_1^*)] = \mathbb{E}[f''[\alpha_1 b(Y_1^*) - \alpha_2 b(x_0), \alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](Y_1^*)]$$

$$+ \sum_{k=1}^{M} \mathbb{E}[(\alpha_1\sigma_k^2(Y_1^*) - \alpha_2\sigma_k^2(x_0))^+ f'''[v_k, v_k, \alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](Y_1^*)]$$

$$+ \frac{1}{4}\sum_{k,j=1}^{M} \mathbb{E}[(\alpha_1\sigma_k^2(Y_1^*) - \alpha_2\sigma_k^2(x_0))^+ (\alpha_1\sigma_j^2(Y_1^*) - \alpha_2\sigma_j^2(x_0))^+ f^{(4)}[v_k, v_k, v_j, v_j](Y_1^*)]$$

$$= \mathbb{E}[f''[\alpha_1 b(Y_1^*) - \alpha_2 b(x_0), \alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](Y_1^*)]$$

$$+ \sum_{k=1}^{M} \mathbb{E}[(\alpha_1\sigma_k^2(Y_1^*) - \alpha_2\sigma_k^2(x_0)) f'''[v_k, v_k, \alpha_1 b(Y_1^*) - \alpha_2 b(x_0)](Y_1^*)]$$

$$+ \frac{1}{4}\sum_{k,j=1}^{M} \mathbb{E}[(\alpha_1\sigma_k^2(Y_1^*) - \alpha_2\sigma_k^2(x_0))(\alpha_1\sigma_j^2(Y_1^*) - \alpha_2\sigma_j^2(x_0)) f^{(4)}[v_k, v_k, v_j, v_j](Y_1^*)] + O(h^2).$$

Moreover, detailed calculations using Lemma A.4 reveal that

$$\mathbb{E}[B^2 f(Y_1^*)] = \mathbb{E}\Bigg[\alpha_2^2 B_1^2 f(Y_1^*) - 2\alpha_1\alpha_2 A(B_1 f)(Y_1^*) + \alpha_1^2\Bigg[\sum_{k=1}^{M}\sigma_k^2(Y_1^*) f'''[v_k, v_k, b(Y_1^*)](Y_1^*)$$

$$+ f''[b(Y_1^*), b(Y_1^*)](Y_1^*) + \frac{1}{4}\sum_{k,j=1}^{M}\sigma_k^2(Y_1^*)\,\sigma_j^2(Y_1^*) f^{(4)}[v_k, v_k, v_j, v_j](Y_1^*)\Bigg]\Bigg] + O(h^2).$$

Next we apply Dynkin's formula repeatedly to obtain

$$\mathbb{E}[B^2 f(Y_1^*)] = \mathbb{E}[B^2 f(y(\theta h))]$$

$$= \alpha_2^2 \left[B_1^2 f(x0) + B_1^3 f(x0)\, h + \int_0^h \int_0^s \square\, [B_1^4 f(y(r))]\, dr\, ds \right]$$

$$- 2\alpha_1\alpha_2 \left[A(B_1 f)(x_0) + B_1 (A(B_1 f))(x_0)\, \theta h + \int_0^{\theta h}\int_0^s \mathbb{E}\, [B_1^2 (A(B_1 f))(y(r))\, dr\, ds] \right]$$

$$+ \alpha_1^2 \left[(B_1^2 f)(x_0) + (B_1^3 f)(x_0)\, \theta h + \int_0^{\theta h}\int_0^s \mathbb{E}\, [B_1^4 f(y(r)) \right] dr\, ds] + O(h^2).$$

The detailed calculation shows that,

$$\mathbb{E}\, [B^2 f(Y_1^*)] = (B_1^2 f)(x_0) + (B_1^3 f)(x_0)\theta h + O(h^2). \tag{A.12}$$

Step 4. Proceeding in the same way as above, we have

$$\mathbb{E}\, (B^3 f)(Y_1^*) = (B_1^3 f)(x_0) + O(h) \tag{A.13}$$

Step 5. Combining (A.7), (A.11), (A.12), (A.13) and observing

$$\alpha_1 \theta (1 - \theta) h^2 - \alpha_2 \theta (1 - \theta)\, h^2 + \frac{(1-\theta)^2 h^2}{2} + \frac{\theta^2 h^2}{2} = \frac{h^2}{2}$$

and

$$\frac{\theta^3 h^3}{6} + \alpha_1 \frac{\theta^2 (1-\theta)}{2} h^3 - \alpha_2 \frac{\theta^2 (1-\theta)}{2} h^3 + \alpha_1 \frac{\theta (1-\theta)^2}{2} h^3 - \alpha_2 \frac{\theta (1-\theta)^2}{2} h^3 + \frac{(1-\theta)^3}{6} h^3 = \frac{h^3}{6}$$

gives the desired result.

References

[1] Anderson-M-11 Anderson, D.F. and Mattingly, J.C. (2011). A weak trapezoidal method for a class of stochastic differential equations. *Common. Math. Sci.*, 9 (1): 301 – 318

[2] Bruti-Liberati, N. and Platen. E. (2008). Strong predictor-corrector Euler methods for stochastic differential equations. *Stoch. Dyn.*, 8(3): 561 – 581.

[3] Higham, D.J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, 43 (3): 525 – 546 (electronic).

[4] Hisler, J., Liu, R. Y., and Singh, K. (2002). A formula for the tail probability of a multivariate normal distribution and its applications. *Journal of multivariate analysis*, 82 (2): 422 – 430.

[5] Karatzas, I. and Shreve, S.E. (1991). *Brownian motion and stochastic calculus,* volume 113 of *Graduate Texts in Mathematics.* Springer- Verlag. New York, second edition.

[6] Kloeden, P.E. and Platen, E. (1992). *Numerical solution of stochastic differential equations,* volume 23 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin.

[7] Mao, X. and Yuan, C. (2006). *Stochastic differential equations with Markovian switching.* Imperial College Press, London.

[8] Meyn, S.P. and Tweedie, R.L. (2009). *Markov chains and stochastic stability*. Cambridge University Press, Cambridge.

[9] Øksendal, B. (2003). *Stochastic differential equations*, *An introduction with applications.* Universitext. Springer- Verlag. Berlin, sixth edition.

[10] Saito. Y. and Mitsui, T. (1996). Stablity analysis of numerical schemes for stochastic differential equations. *SIAM J. Numer, Anal.*, 33(6): 2254 – 2267.

[11] Walsh. J.B. (1986). *An introduction to stochastic partial differential equations.* Springer, Berlin.

[12] Yin, G.G. and Zhu. C. (2010). *Hybrid Switching Diffusions: Properties and Applications*, volume 63 of *Stochastic Modelling and Applied Probability.* Springer, New York.