BIT Numerical Mathematics (2020) 60:917–957
https://doi.org/10.1007/s10543-020-00803-6

BIT

Check for updates

Modified averaged vector field methods preserving multiple invariants for conservative stochastic differential equations

Chuchu Chen[1,2] · **Jialin Hong**[1,2] · **Diancong Jin**[1,2]

Received: 27 October 2018 / Accepted: 11 February 2020 / Published online: 24 February 2020

Abstract

A novel class of conservative numerical methods for general conservative Stratonovich stochastic differential equations with multiple invariants is proposed and analyzed. These methods, which are called modified averaged vector field methods, are constructed by modifying the averaged vector field method to preserve multiple invariants simultaneously. Based on the a prior estimate for high-order moments of the modification coefficient, the mean square convergence order 1 of the proposed methods is proved in the case of commutative noises. In addition, the effect of the quadrature formula on the mean square convergence order and the preservation of invariants for modified averaged vector field methods is considered. Numerical experiments are performed to verify the theoretical analyses and to show the superiority of the proposed methods in the long time simulation.

Keywords Stochastic differential equations · Invariants · Conservative methods · Mean square convergence order · Quadrature formula

---

Communicated by Antonella Zanna Munthe-Kaas.

This work is supported by National Natural Science Foundation of China (Nos. 91630312, 11971470, 11871068, 11926417, 11711530017).

✉ Diancong Jin
diancongjin@lsec.cc.ac.cn

Chuchu Chen
chenchuchu@lsec.cc.ac.cn

Jialin Hong
hjl@lsec.cc.ac.cn

1 Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

2 School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

Mathematics Subject Classification 60H10 · 60H35 · 65D30

1 Introduction

Numerical methods for stochastic differential equations (SDEs) have attracted extensive attention over the past decades, in view of the difficulty of obtaining explicit solutions of original systems (see e.g. [2,7,10]). It is important to construct numerical methods which preserve properties of original systems as much as possible. For conservative SDEs with one invariant, there have been many works related to numerical methods in recent years. On the one hand, aiming at the SDEs with single noise, [12] proposes an energy-preserving difference method for stochastic Hamiltonian systems and analyzes the local error. Based on the equivalent "skew gradient" (SG) form for conservative SDEs with one invariant, [6] proposes the direct discrete gradient methods and the indirect discrete gradient methods, and proves that these two kinds of methods are of mean square order 1. Authors in [4] construct energy-preserving methods for stochastic Poisson systems, and prove that those methods are of mean square order 1 and preserve quadratic Casimir functions. On the other hand, in the case of SDEs with multiple noises, [3] proposes the averaged vector field (AVF) method for conservative SDEs (see [13] for the original derivation of the AVF method in the deterministic setting). It is shown that the mean square order of the AVF method is 1 if noises are commutative and that the weak order is 1 in the general case. For the case of quadratic invariants, [5] constructs stochastic Runge–Kutta (SRK) methods for SDEs with quadratic invariants and [14] gives the order conditions for SRK methods preserving quadratic invariants.

For conservative SDEs with multiple invariants, one difficulty is to preserve multiple invariants simultaneously. One approach is via a projection technique, which combines an arbitrary one-step approximation together with a projection onto the invariant submanifold in each step. [15] shows that this approach is feasible in stochastic settings, and the proposed methods could reach high strong order as supporting methods do. In this paper, we focus on constructing a new class of multi-invariant-preserving methods, which are called modified averaged vector field (MAVF) methods. More precisely, we add modification terms to the AVF method to preserve multiple invariants simultaneously, motivated by the ideas of line integral methods (LIMs) for deterministic conservative ordinary differential equations (ODEs) in [1].

As is seen in (3.3), the modification terms contain a vector-valued random variable $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ which is called the modification coefficient, hence a prerequisite to acquire the convergence order of MAVF methods is the boundedness of the high-order moments of $\boldsymbol{\alpha}$. To this end, one technique is to truncate the Brownian increments, which not only ensures the solvability of MAVF methods, but also makes sure that for sufficiently small stepsize, $\boldsymbol{\alpha}$ is uniformly small with respect to the sample point ω . Another technique is the use of the orthogonality of Legendre polynomials, which makes us get rid of the effect of low-order terms and then acquire the estimate for high-order moments of $\boldsymbol{\alpha}$. We compare MAVF methods with Milstein method to prove that MAVF methods are of mean square order 1.

When the integrals contained in MAVF methods can not be obtained directly, numerical integration is an option to approximate these integrals. Thus it is necessary to investigate the effect of the numerical integration on the mean square convergence order and the preservation of invariants for the proposed methods. It is proved that the induced MAVF methods are still of mean square order 1 provided that the order of the quadrature formula is not less than 2. Generally, the invariant-preserving order in mean square sense of the MAVF method using the numerical integration only depends on the order of the quadrature formula.

The rest of this paper is organized as follows. In Sect. 2, we give some concepts about conservative SDEs with invariants and preliminary theorems and lemmas for numerical analyses. Section 3 proposes MAVF methods for conservative SDEs with single or multiple noises and shows the properties of these methods. Section 4 investigates MAVF methods using the numerical integration, and analyzes their convergence orders and preservation of invariants. Numerical experiments are performed in Sect. 5 to verify the theoretical analyses and to show the advantages of MAVF methods in the long time simulation.

In the sequel, for convenience, we will use the following notations:

- $|x|$: The trace norm of a vector or a matrix x, i.e., $|x| = \sqrt{\mathrm{Tr}(x^\top x)}$.
- $\mathbf{C}^k(\mathbb{R}^m, \mathbb{R}^n)$: The space of k times continuously differentiable functions $f : \mathbb{R}^m \to \mathbb{R}^n$.
- $\mathbf{C}_b^k(\mathbb{R}^m, \mathbb{R}^n)$: The space of k times continuously differentiable functions $f : \mathbb{R}^m \to \mathbb{R}^n$ with uniformly bounded jth order derivatives, $j = 1, \ldots, k$.
- ∇f: The gradient of a scalar function $f \in \mathbf{C}^1(\mathbb{R}^m, \mathbb{R})$, i.e., $\nabla f = \big(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_m}\big)$; or the Jacobian matrix of a vector function $f \in \mathbf{C}^1(\mathbb{R}^m, \mathbb{R}^n)$, i.e., $\nabla f = (\nabla f_1^\top, \ldots, \nabla f_n^\top)^\top$.
- Gâteaux derivative $f^{(k)}(x)(\xi_1, \ldots, \xi_k)$: If $f(x) \in \mathbf{C}^k(\mathbb{R}^m, \mathbb{R})$ and $\xi_1, \ldots, \xi_k \in \mathbb{R}^m$, then $f^{(k)}(x)(\xi_1, \ldots, \xi_k) = \sum_{i_1,\ldots,i_k=1}^m \frac{\partial^k f(x)}{\partial x^{i_1} \cdots \partial x^{i_k}} \xi_1^{i_1} \cdots \xi_k^{i_k}$.
- The closed ball $\bar{B}(y, r)$ with center $y \in \mathbb{R}^m$ and radius $r > 0$: $\bar{B}(y, r) = \big\{x \in \mathbb{R}^m \big| \, |x - y| \le r\big\}$.

2 Preliminary

In this section, we give the definition of an invariant for conservative SDEs and introduce some lemmas and theorems for the proof of convergence.

Consider the general m-dimensional autonomous SDE in the sense of Stratonovich

$$\mathrm{d}Y(t) = f\big(Y(t)\big)\mathrm{d}t + \sum_{r=1}^{D} g_r\big(Y(t)\big) \circ \mathrm{d}W_r(t), \quad 0 \le t \le T, \quad Y(0) = Y_0, \tag{2.1}$$

where $W_r(t)$, $r = 1, \ldots, D$, are D independent one-dimensional Brownian motions defined on a complete filtered probability space $\big(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\ge 0}, \mathbb{P}\big)$ with $\{\mathscr{F}_t\}_{t\ge 0}$ satisfying the usual conditions. Assume that Y_0 is a deterministic initial value, and that $f : \mathbb{R}^m \to \mathbb{R}^m$, $g_r : \mathbb{R}^m \to \mathbb{R}^m$, $r = 1, \ldots, D$, are such that (2.1) has a unique global solution. Next we give the definition of an invariant.

Definition 2.1 (*see* [15]) SDE (2.1) is said to have ν invariants $L^i(y) \in \mathbf{C}^1(\mathbb{R}^m, \mathbb{R})$, $i = 1, \ldots, \nu$, if

$$\nabla L^i(y) f(y) = 0,\ \nabla L^i(y) g_r(y) = 0,\ r = 1, \ldots, D,\ i = 1, \ldots, \nu, \quad \forall\, y \in \mathbb{R}^m. \tag{2.2}$$

If we define the vector-valued function $L(y) := \left(L^1(y), \ldots, L^\nu(y)\right)^\top$, then (2.2) can be written compactly as

$$\nabla L(y) f(y) = \nabla L(y) g_r(y) = \mathbf{0},\ r = 1, \ldots, D, \quad \forall\, y \in \mathbb{R}^m. \tag{2.3}$$

Hereafter, we also say that the vector-valued function $L(y)$, which satisfies (2.3), is the invariant of (2.1). According to the definition of the invariant, it follows from the stochastic chain rule that $\mathrm{d}L(Y(t)) = \mathbf{0}$, where $Y(t)$ is the exact solution of (2.1). This implies that $L(Y(t)) = L(Y_0)$, a.s. This is to say, $L(y)$, along the exact solution $Y(t)$, is invariant almost surely.

The following two theorems give the relationship between local errors and global errors of numerical methods for general SDEs. In the sequel, we always assume that the assumptions of these two theorems hold unless we make an additional statement.

Theorem 2.2 (see [8]) *Suppose that the one-step approximation $\bar{Y}_{t,y}(t+h)$ has order of accuracy p_1 for the expectation of the deviation and order of accuracy p_2 for the mean square deviation; more precisely, for arbitrary $t_0 \le t \le t_0 + T - h$, $y \in \mathbb{R}^m$ the following inequalities hold:*

$$|E\left(Y_{t,y}(t+h) - \bar{Y}_{t,y}(t+h)\right)| \le K \cdot \left(1 + |y|^2\right)^{1/2} h^{p_1}, \tag{2.4}$$

$$\left[E|Y_{t,y}(t+h) - \bar{Y}_{t,y}(t+h)|^2\right]^{1/2} \le K \cdot \left(1 + |y|^2\right)^{1/2} h^{p_2}. \tag{2.5}$$

Also, let

$$p_2 \ge \frac{1}{2},\ p_1 \ge p_2 + \frac{1}{2}.$$

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

$$\left[E|Y_{t_0,Y_0}(t_k) - \bar{Y}_{t_0,Y_0}(t_k)|^2\right]^{1/2} \le K \cdot \left(1 + E|Y_0|^2\right)^{1/2} h^{p_2 - 1/2}, \tag{2.6}$$

i.e., the mean-square order of accuracy of the method constructed using the one-step approximation $\bar{Y}_{t,y}(t+h)$ is $p = p_2 - 1/2$.

Theorem 2.3 (see [8]) *Let the one-step approximation $\bar{Y}_{t,y}(t+h)$ satisfy the conditions of Theorem 2.2. Suppose that $\tilde{Y}_{t,y}(t+h)$ is such that*

$$|E\left(\tilde{Y}_{t,y}(t+h)-\bar{Y}_{t,y}(t+h)\right)| = \mathscr{O}(h^{p_1}), \tag{2.7}$$

$$\left[E|\tilde{Y}_{t,y}(t+h)-\bar{Y}_{t,y}(t+h)|^2\right]^{1/2} = \mathscr{O}(h^{p_2}). \tag{2.8}$$

with the same p_1 and p_2. Then the method based on the one-step approximation $\tilde{Y}_{t,y}(t+h)$ has the same mean square order of accuracy as the method based on $\bar{Y}_{t,y}(t+h)$, i.e., its order is equal to $p = p_2 - 1/2$.

Generally speaking, when implementing implicit numerical methods, the truncated random variables $\Delta\widehat{W}_r(h)$ for the Brownian increments $\Delta W_r(h) = W_r(t+h) - W_r(t)$, $r = 1,\ldots,D$, need to be introduced (see [8]). For this end, one can represent $\Delta W_r(h) = \sqrt{h}\xi_r$, $r = 1,\ldots,D$, where ξ_r, $r = 1,\ldots,D$, are independent $N(0,1)$-distributed random variables. Then define $\Delta\widehat{W}_r(h) = \sqrt{h}\zeta_{rh}$ as follows:

$$\zeta_{rh} = \begin{cases} \xi_r, & \text{if } \ |\xi_r| \le A_h, \\ A_h, & \text{if } \ \xi_r > A_h, \\ -A_h, & \text{if } \ \xi_r < -A_h, \end{cases} \tag{2.9}$$

with $A_h := \sqrt{2k|\ln h|}$, where k is an arbitrary positive integer. The following properties hold for the truncated Brownian increments.

Lemma 2.4 (see [8]) *Let $A_h := \sqrt{2k|\ln h|}$, $k \ge 1$, and ζ_{rh} be defined by (2.9). Then it holds that*

$$E(\zeta_{rh}-\xi_r)^2 \le h^k, \tag{2.10}$$

$$0 \le E(\xi_r^2-\zeta_{rh}^2) = 1 - E\zeta_{rh}^2 \le (1+2\sqrt{2k|\ln h|})h^k. \tag{2.11}$$

Moreover, it is not difficult to obtain the following properties:

$$\begin{aligned} &E\left(|\Delta\widehat{W}_r(h)|^{2p}\right)^{\frac{1}{2p}} \le E\left(|\Delta W_r(h)|^{2p}\right)^{\frac{1}{2p}} \le c_p h^{1/2}, \quad \forall\, p \in \mathbb{N}^+, \\ &E(\Delta\widehat{W}_r(h))^{2p-1} = E\,(\Delta W_r(h))^{2p-1} = 0, \quad \forall\, p \in \mathbb{N}^+, \\ &E\left(\Delta\widehat{W}_i\Delta\widehat{W}_j\Delta\widehat{W}_k\right) = E\left(\Delta W_i\Delta W_j\Delta W_k\right) = 0, \quad \forall\, i,j,k \in \{1,2,\ldots,D\}, \end{aligned} \tag{2.12}$$

where c_p is a constant independent of h.

3 MAVF methods for stochastic SDEs

Consider the following autonomous m-dimensional SDE with single noise

$$\mathrm{d}Y(t) = f\big(Y(t)\big)\mathrm{d}t + g\big(Y(t)\big)\circ \mathrm{d}W(t), \quad 0 \le t \le T, \quad Y(0) = Y_0, \tag{3.1}$$

where f and g satisfy the global Lipschitz condition. Let $L(y) : \mathbb{R}^m \to \mathbb{R}^\nu$ be the invariant of (3.1), i.e., $\nabla L(y) f(y) = \nabla L(y) g(y) = \mathbf{0}$, for all $y \in \mathbb{R}^m$. Hereafter, we alway assume that ∇L is continuous on $\mathbb{R}^m$.

We consider the numerical approximation for (3.1) in the interval $[0, T]$. Let $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$ be a uniform partition of interval $[0, T]$, where $t_n = nh,\ n = 0, 1, \ldots, N$. Let $\{y_n\}_{n=0}^N$ be some numerical approximation.

3.1 Introduction on AVF method

In this part, we recall the AVF method for conservative SDEs. The AVF method, as a special discrete gradient method, is originally proposed by [13] in dealing with conservative ODEs. For the SDE (3.1), the AVF method is proposed similarly as in deterministic settings:

$$\bar{Y} = y + h \int_0^1 f(y + \tau(\bar{Y} - y))\,\mathrm{d}\tau + \Delta\widehat{W} \int_0^1 g(y + \tau(\bar{Y} - y))\,\mathrm{d}\tau. \tag{3.2}$$

If $I \in \mathbf{C}(\mathbb{R}^m, \mathbb{R})$ is an invariant of (3.1), then there are two skew-symmetric matrices $S(x)$ and $T(x)$ such that $f(y) = S(y)\nabla I(y)$ and $g(y) = T(y)\nabla I(y)$ (see e.g. [6]). Especially, if S and T are constant matrices, then (3.2) preserves the invariant I. In order to preserve the invariant I for general f and g, authors in [3] combine the skew gradient form of (3.1) and the AVF method (3.2) to give a variant of the AVF method. Generally, the method (3.2) and the one in [3] could not preserve multiple invariants. This inspires us to seek for a new class of conservative methods to preserve multiple invariants simultaneously.

3.2 MAVF methods for conservative SDEs with single noise

In this part, we propose MAVF methods preserving multiple invariants for conservative SDEs with single noise and prove that these methods are of mean square order 1.

We denote $y_{n+1} = y_{t_n, y_n}(t_{n+1}),\ n = 0, 1, \ldots, N-1$. For convenience, we write the one-step approximation as $\bar{Y} = \bar{Y}_{t,y}(t+h)$. Next, we give the MAVF method for (3.1). It is generated by the following one-step approximation $\bar{Y}$:

$$\begin{aligned}\bar{Y} = y &+ h\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\,\alpha_0\right]\\ &+ \Delta\widehat{W}\left[\int_0^1 g(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\,\alpha_1\right],\end{aligned} \tag{3.3a}$$

$$\left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\right]\alpha_0 = \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau, \tag{3.3b}$$

$$\left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\right]\alpha_1 = \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 g(\sigma(\tau))\,\mathrm{d}\tau, \tag{3.3c}$$

where $\sigma(\tau) = y + \tau(\bar{Y} - y)$, $\Delta\widehat{W} = \sqrt{h}\zeta_h$ and ζ_h is defined by (2.9) with $A_h = \sqrt{4|\ln h|}$, i.e., $k = 2$, and α_0, α_1 are $\mathbb{R}^\nu$-valued random variables.

As is seen above, the MAVF method (3.3) can be regarded as the modification of the AVF method in [3]. Here, $\int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\, \alpha_0$ and $\int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\, \alpha_1$ are called modification terms. Let $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ and we call $\boldsymbol{\alpha}$ the modification coefficient of the method (3.3). The modification coefficient $\boldsymbol{\alpha}$ satisfies (3.3b) and (3.3c) in (3.3) to make the MAVF method conservative.

More precisely, in order to preserve multiple invariants, we perturb (3.2) as follows:

$$\begin{aligned}\bar{Y} = \ & y + h\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\,\alpha_0\right]\\ &+\Delta\widehat{W}\left[\int_0^1 g(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\,\alpha_1\right],\end{aligned} \tag{3.4}$$

where α_0 and α_1 are undetermined coefficients to make this method preserve multiple invariants. To this end, by the Taylor expansion and (3.4), one has

$$\begin{aligned}&L(\bar{Y}) - L(y)\\ &= \int_0^1 \nabla L(y + \tau(\bar{Y} - y))\,\mathrm{d}\tau(\bar{Y} - y)\\ &= h\left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\,\alpha_0\right]\\ &= +\Delta\widehat{W}\left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 g(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\,\alpha_1\right].\end{aligned}$$

It is observed that choosing α_0 and α_1 from (3.3b) and (3.3c) in (3.3) ensures the preservation of multiple invariants, i.e., $L(\bar{Y}) = L(y)$.

3.2.1 The a prior estimate of the modification coefficient

Let $\{y_n\}_{n=0}^N$, with $y_0 = Y_0$, be the numerical scheme generated by (3.3). We introduce the level set

$$\mathscr{M} = \left\{y \in \mathbb{R}^m \middle| L(y) = L(y_0)\right\} \tag{3.5}$$

and the δ-neighbourhood of $\mathscr{M}$

$$\mathscr{M}_\delta = \left\{y \in \mathbb{R}^m \,|dist(y, \mathscr{M}) \le \delta\right\}, \qquad \text{for some } \delta > 0, \tag{3.6}$$

where $dist(y, \mathscr{M}) = \inf_{x\in\mathscr{M}}|x - y|$. Moreover, for some subset $U \subset \mathbb{R}^m$ and $k \in \mathbb{N}$, we denote by $\mathbf{C}^k(U, \mathbb{R}^n)$ the set of k times continuously differentiable functions from U to $\mathbb{R}^n$, and by $\mathbf{C}_b^k(U, \mathbb{R}^n)$ the subset of $\mathbf{C}^k(U, \mathbb{R}^n)$ with uniformly bounded jth order derivatives, $j = 1, \ldots, k$.

The following lemma gives the solvability and conservation of the MAVF method (3.3).

Lemma 3.1 *Suppose that for some bounded $\mathscr{M}_\delta$, f, g, $\nabla L \in \mathbf{C}^1(\mathscr{M}_\delta)$ and $\nabla L(y)\nabla L(y)^\top$ is invertible on $\mathscr{M}$. If $y \in \mathscr{M}$, a.s., then there is some $h_0 > 0$ such that for all $h \le h_0$ and almost sure $\omega \in \Omega$, the method (3.3) is uniquely solvable, that is, there is a unique $\bar{Y} = \bar{Y}_{t,y}(t+h)$, a.s., corresponding to (3.3). Moreover, it holds that*

(I) *The MAVF method (3.3) preserves exactly multiple invariants L^i, $i = 1, \ldots, \nu$ of SDE (3.1), i.e., $L(\bar{Y}) = L(y)$, a.s.;*

(II) *For every $\epsilon > 0$, there exists some $h_1 \in (0, h_0)$ such that for all $h \le h_1$,*

$$|\boldsymbol{\alpha}| \le \epsilon, \quad a.s. \tag{3.7}$$

Proof We rewrite $\bar{Y} = \bar{Y}_{t,y}(t+h)$ as

$$\begin{aligned}\bar{Y} = y + h\Bigg[&\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\left(\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau\int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\right)^{-1}\\ &\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau\Bigg] + \Delta\widehat{W}\Bigg[\int_0^1 g(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\\ &\left(\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau\int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\right)^{-1}\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau\int_0^1 g(\sigma(\tau))\,\mathrm{d}\tau\Bigg],\ a.s.\end{aligned} \tag{3.8}$$

Noting that $y \in \mathscr{M}$, one has that there is a closed ball $\bar{B}(y,\delta) \subseteq \mathscr{M}_\delta$. Let $\sigma_{y,z}(\tau) = y + \tau(z-y)$ and $P_y(z) = \int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma_{y,z}(\tau))^\top\,\mathrm{d}\tau$. Next, we prove that there is a constant $r \in (0,\delta)$ such that for every fixed $y \in \mathscr{M}$, $P_y(\cdot)$ is invertible on $\bar{B}(y,r)$.

Since $\mathscr{M}$ is bounded, there is a bounded convex closed set $F \supseteq \mathscr{M}$. For any $y, z \in F$, we define $Q(y,z) = \det(P_y(z))$, where $\det(P_y(z))$ denotes the determinant of $P_y(z)$. Due to the continuity of ∇L and the fact that $F \times F$ is a bounded closed set in $\mathbb{R}^m \times \mathbb{R}^m$, $Q(\cdot,\cdot)$ is uniformly continuous on $F \times F$. Noting that $\nabla L \nabla L^\top$ is invertible on $\mathscr{M}$, we have that $Q(y,y) \ne 0$ for any $y \in \mathscr{M}$. The continuity of $Q(y,y)$ with respect to y yields $Q(y,y) > 0$ for any $y \in \mathscr{M}$ or $Q(y,y) < 0$ for any $y \in \mathscr{M}$. Without loss of generality, we assume that $Q(y,y) > 0$ for each $y \in \mathscr{M}$. It follows from the continuity of $Q(y,y)$ on the bounded closed set $\mathscr{M}$ that there is a constant $c_0 > 0$ such that $Q(y,y) \ge c_0$ for any $y \in \mathscr{M}$. By the uniform continuity of Q, there is some $r \in (0,\delta)$ such that for every $(y,y), (y,z) \in F \times F$ with $|y-z| \le r$, $|Q(y,z) - Q(y,y)| < \frac{c_0}{2}$. In this way, for every $y \in \mathscr{M}$, $z \in \bar{B}(y,r)$, $|Q(y,z) - Q(y,y)| < \frac{c_0}{2}$. This implies that $Q(y,z) > Q(y,y) - \frac{c_0}{2} \ge c_0 - \frac{c_0}{2} = \frac{c_0}{2} > 0$ for $y \in \mathscr{M}$, $z \in \bar{B}(y,r)$. That is to say, there is a constant $r \in (0,\delta)$ such that for every fixed $y \in \mathscr{M}$, $P_y(\cdot)$ is invertible on $\bar{B}(y,r)$.

Let $y \in \mathscr{M}$ be fixed. For each $z \in \bar{B}(y, r)$, we define

$$\begin{aligned}\varphi(z) = y + h \Bigg[&\int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma_{y,z}(\tau))^\top \mathrm{d}\tau\, P_y(z)^{-1} \\ &\int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \Bigg] \\ &+ \Delta\widehat{W} \Bigg[\int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma_{y,z}(\tau))^\top \mathrm{d}\tau\, P_y(z)^{-1} \\ &\int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \Bigg], \\ = y + h &\left[\int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau - M(z) \right] \\ &+ \Delta\widehat{W} \left[\int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau - N(z) \right], \quad a.s., \end{aligned} \tag{3.9}$$

where $M(z) = \int_0^1 \nabla L(\sigma_{y,z}(\tau))^\top \mathrm{d}\tau\, P_y(z)^{-1} \int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau$ and $N(z) = \int_0^1 \nabla L(\sigma_{y,z}(\tau))^\top \mathrm{d}\tau\, P_y(z)^{-1} \int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau$. Then (3.8) can be written as $\bar{Y} = \varphi(\bar{Y})$, a.s.

We claim that φ is a contraction mapping from $\bar{B}(y, r)$ into itself. In fact, since $\mathscr{M}_\delta$ is a bounded closed set, it follows from assumptions on f, g and ∇L that, there exists a constant $K_\delta > 0$ such that

$$\max_{y \in \mathscr{M}_\delta} (|f(y)| + |g(y)| + |\nabla L(y)| + |f'(y)| + |g'(y)| + |\nabla L'(y)|) \le K_\delta. \tag{3.10}$$

In what follows, we use K_δ to denote a generic constant dependent on $\mathscr{M}_\delta$, but independent of y and h, and it may vary from one line to another. For an invertible and continuously differentiable matrix-valued function G on $\mathbb{R}^m$, we notice that $[G(y)^{-1}]' = -G(y)^{-1}G'(y)G(y)^{-1}$. By the assumption $\nabla L \in \mathbf{C}^1(\mathscr{M}_\delta)$, we have that for each $y \in \mathscr{M}$, $P_y(\cdot)^{-1} \in \mathbf{C}^1(\bar{B}(y, r))$. In addition, one has that $\int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau, \int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau, \int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \in \mathbf{C}^1(\bar{B}(y, r))$. Hence, it follows from (3.10) that

$$\begin{aligned} &\max_{z \in \bar{B}(y,r)} \left| \frac{\mathrm{d}}{\mathrm{d}z} \int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \right| \\ &\le \max_{z \in \bar{B}(y,r)} \int_0^1 \left| f'(\sigma_{y,z}(\tau))\tau \right| \mathrm{d}\tau \le \int_0^1 \max_{z \in \bar{B}(y,r)} \left| f'(\sigma_{y,z}(\tau)) \right| \tau\,\mathrm{d}\tau \\ &\le \int_0^1 \max_{y \in \bar{B}(y,r)} \left| f'(y) \right| \tau\,\mathrm{d}\tau \le \int_0^1 \max_{y \in \mathscr{M}_\delta} \left| f'(y) \right| \tau\,\mathrm{d}\tau \le \frac{1}{2} K_\delta. \end{aligned}$$

This means that $\int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau \in \mathbf{C}_b^1(\bar{B}(y,r))$. Clearly, $\max_{z\in\bar{B}(y,r)}\left|\int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau\right| \le K_\delta$. Same arguments also hold for $\int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau$, $\int_0^1 \nabla L(\sigma_{y,z}(\tau))\,\mathrm{d}\tau$, and $P_y(z)$. Since $P_y(\cdot)^{-1} \in \mathbf{C}^1(\bar{B}(y,r))$, it holds that $\max_{z\in\bar{B}(y,r)}\left|P_y(z)^{-1}\right| \le K_\delta$. According to the fact $[P_y(z)^{-1}]' = -P_y(z)^{-1}P_y'(z)P_y(z)^{-1}$, we obtain that $\max_{z\in\bar{B}(y,r)}\left|\left[P_y(z)^{-1}\right]'\right| \le K_\delta$.

Using the chain rule, one has that $\max_{z\in\bar{B}(y,r)}\left(|M(z)|+|N(z)|+|M'(z)|+|N'(z)|\right) \le K_\delta$. Thus, there is some constant K_δ such that $\int_0^1 f(\sigma_{y,z}(\tau))\,\mathrm{d}\tau$, $\int_0^1 g(\sigma_{y,z}(\tau))\,\mathrm{d}\tau$, $M(z)$ and $N(z)$ are Lipschitz continuous on $\bar{B}(y,r)$ with uniform Lipschitz constant K_δ. Then we deduce that for any $z, z_1, z_2 \in \bar{B}(y,r)$,

$$|\varphi(z)-y| \le K_\delta(h+|\Delta\widehat{W}|), \qquad a.s, \tag{3.11}$$

$$|\varphi(z_1)-\varphi(z_2)| \le K_\delta(h+|\Delta\widehat{W}|)|z_1-z_2|, \qquad a.s. \tag{3.12}$$

Since for almost sure $\omega \in \Omega$, $|\Delta\widehat{W}| \le A_h\sqrt{h} = \sqrt{4h|\ln h|} \to 0$ as $h \to 0$, there exists a positive h_0 independent of ω, such that for every $h < h_0$, $K_\delta(h+|\Delta\widehat{W}|) < \min\{1, r\}$, a.s. It follows from (3.11) and (3.12) that for every $h < h_0$, φ is a contraction mapping from $\bar{B}(y,r)$ into itself. Consequently, by the contraction mapping principle, there is a unique $\bar{Y}$ such that $\bar{Y} = \varphi(\bar{Y})$, a.s., for all sufficiently small stepsizes h, and

$$|\bar{Y}-y| \le K_\delta(h+|\Delta\widehat{W}|), \qquad a.s. \tag{3.13}$$

Further, the Taylor expansion leads to

$$L(\bar{Y})-L(y) = \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau(\bar{Y}-y), \qquad a.s.$$

Substituting (3.8) into the above formula yields $L(\bar{Y}) = L(y)$, a.s.

It follows from the definition of α_0 and $\max_{z\in\bar{B}(y,r)}\left|P_y(z)^{-1}\right| \le K_\delta$ that

$$|\alpha_0| \le K_\delta\left|\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau\right|, \qquad a.s. \tag{3.14}$$

By the Taylor expansion, it holds that

$$\int_0^1 f(\sigma(\tau))\mathrm{d}\tau = f(y) + \int_0^1 \tau \int_0^1 f'(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y)\,\mathrm{d}\theta\,\mathrm{d}\tau, \qquad a.s. \tag{3.15}$$

$$\int_0^1 \nabla L(\sigma(\tau))\mathrm{d}\tau = \nabla L(y) + \int_0^1 \tau \int_0^1 \nabla L'(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y)\,\mathrm{d}\theta\,\mathrm{d}\tau, \qquad a.s. \tag{3.16}$$

Substituting (3.15) and (3.16) into (3.14), and using the fact $\nabla Lf = 0$, we have

$$|\alpha_0| \leq K_\delta(|\bar{Y} - y| + |\bar{Y} - y|^2) \leq K_\delta|\bar{Y} - y| \leq K_\delta(h + |\Delta\widehat{W}|), \qquad a.s.$$

Similarly, one has that $|\alpha_1| \leq K_\delta(h+|\Delta\widehat{W}|)$, a.s. Finally, using the fact $(h+|\Delta\widehat{W}|) \to 0$ as h tends to 0, we complete the proof. □

As is seen in the above proof, although the randomness of y leads to the randomness of $\bar{B}(y, r)$, the constants K_δ and r are independent of ω. Thus, h_0 is also independent of ω. Using the fact $y_0 \in \mathscr{M}$, we obtain a unique sequence $\{y_n\}_{n=0}^N$ satisfying $y_{n+1} = \bar{Y}_{t_n,y_n}(y_n)$, a.s., for sufficiently small stepsize h. Furthermore, $L(y_n) = L(y_0)$, a.s., $n = 1, 2, \ldots, N$.

Remark 3.2 As is seen in (3.7), we actually have that for sufficiently small h independent of ω, $|\boldsymbol{\alpha}| \leq \varepsilon$, a.s., which is essential to give the high-order-moment estimates of the modification coefficient $\boldsymbol{\alpha}$. The key to the proof of the boundedness of $\boldsymbol{\alpha}$ is the use of the truncated Brownian increment $|\Delta\widehat{W}|$. Otherwise, one can only obtain that for every $\varepsilon > 0$ and every $\omega \in \Omega$, there exists an $h_0(\omega)$ such that for all $h(\omega) \leq h_0(\omega)$, $|\boldsymbol{\alpha}| \leq \varepsilon$.

Remark 3.3 If the initial value Y_0 is a random variable, we can also establish the solvability of the method (3.3). In this case, stronger assumptions on $f, g, \nabla L$ are required. For example, if $Y_0 \in U_0$, a.s., for some $U_0 \subseteq \mathbb{R}^m$, we define $\mathscr{N} = \left\{y \in \mathbb{R}^m \middle| L(y) = L(x) \text{ for some } x \in U_0\right\}$, $\mathscr{N}_\delta = \{y \in \mathbb{R}^m \,|dist(y, \mathscr{N}) \leq \delta\}$. If we replace $\mathscr{M}$ by $\mathscr{N}$, and $\mathscr{M}_\delta$ by $\mathscr{N}_\delta$ in Lemma 3.1, then the solvability of (3.3) can be given similarly.

Next we introduce Legendre polynomials $\{P_j(t)\}_{j\geq 0}$, which is a family of orthogonal polynomials on the interval [0, 1]. For Legendre polynomials $\{P_j(t)\}_{j\geq 0}$, it holds that

$$\int_0^1 P_j(t)\,\mathrm{d}t = 0,\ \forall\, j \geq 1, \qquad \int_0^1 P_j(t)t^k\,\mathrm{d}t = 0,\ \forall\, k < j. \tag{3.17}$$

In the following, we will use the properties of Legendre polynomials to derive some important lemmas. The authors in [1] give Lemma 3.1 and some facts in Chapter 6 by means of Legendre polynomials, when dealing with numerical methods for conservative ODEs. Likewise, we obtain some useful lemmas in stochastic cases.

Lemma 3.4 *Assume that f, g and ∇L are continuous, then*

$$\begin{aligned} &\sum_{j\geq 0}\int_0^1 P_j(\tau)\nabla L(\sigma(\tau))\,\mathrm{d}\tau \cdot \int_0^1 P_j(\tau)f(\sigma(\tau))\,\mathrm{d}\tau = \mathbf{0}, \\ &\sum_{j\geq 0}\int_0^1 P_j(\tau)\nabla L(\sigma(\tau))\,\mathrm{d}\tau \cdot \int_0^1 P_j(\tau)g(\sigma(\tau))\,\mathrm{d}\tau = \mathbf{0}. \end{aligned} \tag{3.18}$$

Proof Since Legendre polynomials $\{P_j(t)\}_{j\geq 0}$ form an orthonormal basis of the Hilbert space $L^2([0,1])$, it follows that $f(\sigma(\tau)) = \sum_{j\geq 0} P_j(\tau)\big[\int_0^1 P_j(\tau)f(\sigma(\tau))\,\mathrm{d}\tau\big]$. Noting that $\nabla L(y)f(y) = 0$, for all $y \in \mathbb{R}^m$, one has

$$\begin{aligned}\mathbf{0} &= \int_0^1 \nabla L(\sigma(\tau))f(\sigma(\tau))\,\mathrm{d}\tau \\ &= \int_0^1 \nabla L(\sigma(\tau)) \cdot \sum_{j\geq 0} P_j(\tau)\left[\int_0^1 P_j(\tau)f(\sigma(\tau))\,\mathrm{d}\tau\right]\mathrm{d}\tau \\ &= \sum_{j\geq 0}\int_0^1 P_j(\tau)\nabla L(\sigma(\tau))\,\mathrm{d}\tau \cdot \int_0^1 P_j(\tau)f(\sigma(\tau))\,\mathrm{d}\tau.\end{aligned}$$

Likewise, we obtain the second equality of (3.18). □

In the sequel, we will use a generic constant K, dependent on y but independent of h, which may vary from one line to another.

Lemma 3.5 *Let the assumptions of Lemma 3.1 hold. Let G be a scalar or vector-valued function defined on $\mathscr{M}_\delta$ and $G(y) \in \mathbf{C}^{(j+1)}(\mathscr{M}_\delta)$. If $y \in \mathscr{M}$, a.s., there is a representation*

$$\int_0^1 P_j(t)G(\sigma(\tau))\,\mathrm{d}\tau = c_j G^{(j)}(y)(\underbrace{\bar{Y}-y,\ldots,\bar{Y}-y}_{j}) + M_{j,G}, \quad a.s., \quad \forall\, j \geq 0, \tag{3.19}$$

where $c_j = \frac{1}{j!}\int_0^1 P_j(\tau)\tau^j\,\mathrm{d}\tau$, and $[E|M_{j,G}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{(j+1)/2})$ for all $p = 1, 2, 3, \ldots$

Proof Denote $F := G\circ\sigma$, then F is $(j+1)$ times continuously differentiable. So the Taylor expansion gives

$$F(\tau) = \sum_{k=0}^{j}\frac{F^{(k)}(0)}{k!}\tau^k + \int_0^1 \frac{(1-\theta)^j}{j!}F^{(j+1)}(\theta\tau)\tau^{j+1}\,\mathrm{d}\theta.$$

Noting that $\int_0^1 P_j(\tau)\tau^k\,\mathrm{d}\tau = 0$, for all $k < j$, we obtain

$$\begin{aligned}&\int_0^1 P_j(t)G(\sigma(\tau))\,\mathrm{d}\tau = \int_0^1 P_j(t)F(\tau)\,\mathrm{d}\tau \\ &= \frac{F^{(j)}(0)}{j!}\int_0^1 P_j(\tau)\tau^j\,\mathrm{d}\tau + \int_0^1 P_j(\tau)\left[\int_0^1 \frac{(1-\theta)^j}{j!}F^{(j+1)}(\theta\tau)\tau^{j+1}\,\mathrm{d}\theta\right]\mathrm{d}\tau \\ &=: c_j F^{(j)}(0) + M_{j,G},\end{aligned}$$

where $c_j = \frac{1}{j!}\int_0^1 P_j(\tau)\tau^j\,\mathrm{d}\tau$ and $M_{j,G} = \int_0^1 P_j(\tau)\left[\int_0^1 \frac{(1-\theta)^j}{j!}F^{(j+1)}(\theta\tau)\tau^{j+1}\,\mathrm{d}\theta\right]\mathrm{d}\tau$.

Further, $F^{(k)}(\tau) = G^{(k)}(y+\tau(\bar{Y}-y))(\underbrace{\bar{Y}-y,\ldots,\bar{Y}-y}_{k})$, $k=0,1,\ldots$, leads to (3.19).

It remains to estimate the moments of $M_{j,G}$. It follows from the boundedness of $G^{(j+1)}$ that $|M_{j,G}| \le K_j|\bar{Y}-y|^{j+1}$. By (3.13), the Hölder inequality and (2.12), we have

$$[E|M_{j,G}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}\left(h^{(j+1)/2}\right), \quad \forall\, p \ge 1.$$

This completes the proof. □

Next, we give the a prior estimate of the modification coefficient $\boldsymbol{\alpha}$.

Lemma 3.6 *Under the assumptions of Lemma 3.1, if $y \in \mathscr{M}$ a.s., then $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ determined by (3.3) satisfies*

$$[E|\boldsymbol{\alpha}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h), \qquad |E(\Delta\widehat{W}\boldsymbol{\alpha})| = \mathscr{O}(h^2). \tag{3.20}$$

Proof It follows from (3.13) that

$$\bar{Y} = y + R_{1,0} \qquad \text{with} \qquad [E|R_{1,0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2}). \tag{3.21}$$

By the Taylor expansion of $f(\sigma(\tau))$ at y,

$$f(\sigma(\tau)) = f(y) + \tau\int_0^1 f'(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y)\,\mathrm{d}\theta.$$

Integrating the above formula on both sides yields

$$\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau = f(y) + R_{1,f} \quad \text{with}$$
$$R_{1,f} = \int_0^1 \tau\int_0^1 f'(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y)\,\mathrm{d}\theta\mathrm{d}\tau. \tag{3.22}$$

For sufficiently small h, it follows from (3.13) that for some $r>0$, $\sigma(\tau) \in \bar{B}(y,r) \subseteq \mathscr{M}_\delta$. Thus the boundedness of f' on $\mathscr{M}_\delta$ and (3.21) lead to $[E|R_{1,f}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2})$. Similarly, we obtain

$$\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau = \nabla L(y) + R_{1,L} \quad \text{with} \quad [E|R_{1,L}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2}). \tag{3.23}$$

It follows from (3.23) that (3.3b) can be written as

$$(\nabla L + R_{1,L})(\nabla L^\top + R_{1,L}^\top)\alpha_0 = \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau. \tag{3.24}$$

By Lemmas 3.4, 3.5 and $P_0(t) \equiv 1$, it holds that

$$
\begin{aligned}
&\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau \\
&\quad = -\sum_{j\geq 1}\int_0^1 P_j(\tau)\nabla L(\sigma(\tau))\,\mathrm{d}\tau \cdot \int_0^1 P_j(\tau) f(\sigma(\tau))\,\mathrm{d}\tau \\
&\quad = -\sum_{j\geq 1}\Bigg[c_j \nabla L^{(j)}(y)(\underbrace{\bar{Y}-y,\ldots,\bar{Y}-y}_{j}) + M_{j,\nabla L}\Bigg] \\
&\qquad \times \Bigg[c_j f^{(j)}(y)(\underbrace{\bar{Y}-y,\ldots,\bar{Y}-y}_{j}) + M_{j,f}\Bigg] \\
&\quad = K_1 \nabla L'(y)(\bar{Y}-y)(f'(y)(\bar{Y}-y)) + R_{1,\alpha_0},
\end{aligned}
\tag{3.25}
$$

where $K_1 = -c_1^2$ with $c_1 = \int_0^1 P_1(\tau)\tau\,\mathrm{d}\tau$, and $[E|R_{1,\alpha_0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5})$.

In what follows, our idea is to establish a rough estimate of α_0 such that α_0 can reach $\frac{1}{2}$ order in the mean square sense, then by the iteration argument, the order is improved to be 1. Combining (3.24) and (3.25), we have

$$
\begin{aligned}
\alpha_0 = &-[\nabla L \nabla L^\top]^{-1}(\nabla L R_{1,L}^\top + R_{1,L}\nabla L^\top + R_{1,L}R_{1,L}^\top)\alpha_0 \\
&+ K_1[\nabla L \nabla L^\top]^{-1}\nabla L'(y)(\bar{Y}-y)(f'(y)(\bar{Y}-y)) + [\nabla L \nabla L^\top]^{-1}R_{1,\alpha_0}.
\end{aligned}
\tag{3.26}
$$

By (3.13), for sufficiently small h, $|\alpha_0| \leq 1$, a.s. From assumptions on f and ∇L, it follows that

$$
|\alpha_0| \leq K|R_{1,L}| + K|R_{1,L}|^2 + K|\bar{Y}-y|^2 + K|R_{1,\alpha_0}|. \tag{3.27}
$$

This implies that

$$
[E|\alpha_0|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2}). \tag{3.28}
$$

Similarly, one has

$$
[E|\alpha_1|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2}). \tag{3.29}
$$

According to (3.21), (3.26) and (3.28), we obtain

$$
\begin{aligned}
E|\alpha_0|^2 &\leq KE(|R_{1,L}|^2|\alpha_0|^2) + KE|R_{1,L}|^4 + KE|\bar{Y}-y|^4 + KE|R_{1,\alpha_0}|^2 \\
&\leq K[E(|R_{1,L}|^4]^{1/2}[E|\alpha_0|^4]^{1/2} + Kh^2 + Kh^3 \\
&\leq Kh^2.
\end{aligned}
$$

That is to say, $[E|\alpha_0|^2]^{1/2} = \mathscr{O}(h)$. Similarly, we have

$$[E|\alpha_0|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h). \tag{3.30}$$

Thus (3.26) can be rewritten as

$$\alpha_0 = K_1[\nabla L \nabla L^\top]^{-1} \nabla L'(y)(\bar{Y} - y)(f'(y)(\bar{Y} - y)) + \widehat{R}_{1,\alpha_0}, \tag{3.31}$$

with $[E|\widehat{R}_{1,\alpha_0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5})$.

We claim that $\bar{Y} = y + \Delta\widehat{W} g + R_{1,0}$ with $[E|R_{1,0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h)$. Applying the Taylor expansion to $g(\sigma(\tau))$ gives

$$g(\sigma(\tau)) = g(y) + \tau \int_0^1 g'\left(y + \theta\tau(\bar{Y} - y)\right)(\bar{Y} - y)\mathrm{d}\theta.$$

Since g' is bounded, integrating the above formula on [0, 1] leads to

$$\int_0^1 g(\sigma(\tau))\mathrm{d}\tau = g(y) + \int_0^1 \tau \int_0^1 g'\left(y + \theta\tau(\bar{Y} - y)\right)(\bar{Y} - y)\mathrm{d}\theta\mathrm{d}\tau =: g(y) + R_{1,g} \tag{3.32}$$

with $|R_{1,g}| \le K|\bar{Y} - y|$. Thus, it follows from (3.21) that

$$[E|R_{1,g}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2}). \tag{3.33}$$

Substituting (3.32) into (3.3a) produces

$$\bar{Y} = y + \Delta\widehat{W} g + R_{1,0}, \tag{3.34}$$

where

$$R_{1,0} = h\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \,\mathrm{d}\tau\, \alpha_0\right] + \Delta\widehat{W}\left[R_{1,g} - \int_0^1 \nabla L(\sigma(\tau))^\top \,\mathrm{d}\tau\, \alpha_1\right].$$

Finally, using (3.28), (3.29), (3.33) and the Hölder inequality, we have $[E|R_{1,0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h)$.

Substituting (3.34) into (3.31) gives

$$\Delta\widehat{W}\alpha_0 = K_1 \Delta\widehat{W}^3 [\nabla L \nabla L^\top]^{-1} \nabla L' g(f'g) + \widetilde{R}_{1,\alpha_0} \quad \text{with} \quad [E|\widetilde{R}_{1,\alpha_0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^2).$$

Hence, $|E(\Delta\widehat{W}\alpha_0)| = \mathscr{O}(h^2)$ due to (2.12).

As for α_1, analogous to the estimate on α_0, one can show that

$$\begin{aligned}\alpha_1 &= K_1[\nabla L \nabla L^\top]^{-1}\nabla L'(y)(\bar{Y}-y)(g'(y)(\bar{Y}-y)) + \widehat{R}_{1,\alpha_1}\\ &\quad \text{with} \quad [E|\widehat{R}_{1,\alpha_1}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}),\\ \Delta\widehat{W}\alpha_1 &= K_1\Delta\widehat{W}^3[\nabla L \nabla L^\top]^{-1}\nabla L' g(g'g) + \widetilde{R}_{1,\alpha_1}\\ &\quad \text{with} \quad [E|\widetilde{R}_{1,\alpha_1}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^2).\end{aligned}$$

Thus, $[E|\alpha_1|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h)$ and $|E(\Delta\widehat{W}\alpha_1)| = \mathscr{O}(h^2)$. □

In the proof of Lemma 3.6, we obtain the low-oder expansion $\bar{Y} - y = \Delta\widehat{W}g + R_{1,0}$ (see (3.34)). In next subsection concerning the convergence order of the MAVF method, we will make a high-order expansion on $\bar{Y} - y$.

3.2.2 Convergence of MAVF methods for SDEs with single noise

Theorem 3.7 *Let the assumptions of Lemma 3.1 hold. If $f \in \mathbf{C}_b^2(\mathbb{R}^m, \mathbb{R}^m)$, $g \in \mathbf{C}_b^3(\mathbb{R}^m, \mathbb{R}^m)$ then the numerical method generated by (3.3) for SDE (3.1) has first order convergence in mean square sense.*

Proof Our idea is to use the Taylor expansion repeatedly to represent (3.3a) so that we can use the Theorems 2.2 and 2.3 to prove the convergence. For this end, we divide the proof into three steps:

Step 1: We claim that $\bar{Y} = y + \Delta\widehat{W}g + hf + \frac{1}{2}\Delta\widehat{W}^2 g'g + R_{2,0}$ with $[E|R_{2,0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5})$.
Applying the Taylor expansion to g produces

$$g(\sigma(\tau)) = g(y) + \tau g'(y)(\bar{Y}-y) + \tau^2\int_0^1 (1-\theta)g''(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y, \bar{Y}-y)\mathrm{d}\theta.$$

Noting the boundedness of g'', we obtain

$$\int_0^1 g(\sigma(\tau))\mathrm{d}\tau = g + \frac{1}{2}g'(y)(\bar{Y}-y) + R_{2,g}^{(1)} \tag{3.35}$$

with

$$\begin{aligned}R_{2,g}^{(1)} &= \int_0^1 \tau^2 \int_0^1 (1-\theta)g''(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y, \bar{Y}-y)\mathrm{d}\theta\mathrm{d}\tau \quad \text{and}\\ &\quad \times [E|R_{2,g}^{(1)}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h).\end{aligned}$$

Substituting (3.34) into (3.35), we have

$$\int_0^1 g(\sigma(\tau))\mathrm{d}\tau = g + \frac{1}{2}\Delta\widehat{W}g'g + \frac{1}{2}g'R_{1,0} + R_{2,g}^{(1)}$$
$$=: g + \frac{1}{2}\Delta\widehat{W}g'g + R_{2,g} \text{ with } [E|R_{2,g}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h). \quad (3.36)$$

Substituting (3.22) and (3.36) into (3.3a) leads to

$$\bar{Y} = y + \Delta\widehat{W}g + hf + \frac{1}{2}\Delta\widehat{W}^2 g'g + R_{2,0}, \quad (3.37)$$

where

$$R_{2,0} = hR_{1,f} - h\int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\alpha_0 + \Delta\widehat{W}R_{2,g} - \Delta\widehat{W}\int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\alpha_1, \quad (3.38)$$

and $[E|R_{2,0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5})$.

Step 2: Estimate of the expectation of $R_{2,0}$.
Recall that $R_{1,f} = \int_0^1 \tau \int_0^1 f'(y+\theta\tau(\bar{Y}-y))(\bar{Y}-y)\,\mathrm{d}\theta\mathrm{d}\tau$ (see (3.22)). By the Taylor expansion and the boundedness of f'',

$$R_{1,f} = \frac{1}{2}f'(y)(\bar{Y}-y) + \int_0^1 \tau^2 \int_0^1 \left[\int_0^1 f''(y+\lambda\theta\tau)(\bar{Y}-y, \bar{Y}-y)\mathrm{d}\lambda\right]\theta\mathrm{d}\theta\mathrm{d}\tau$$
$$= \frac{1}{2}\Delta\widehat{W}f'g + \widehat{R}_{1,f}$$

with $[E|\widehat{R}_{1,f}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h)$. Thus, we have

$$|E(hR_{1,f})| = \mathscr{O}(h^2). \quad (3.39)$$

Next we estimate the expectation of $R_{2,g} = \frac{1}{2}g'R_{1,0} + R_{2,g}^{(1)}$. Comparing (3.34) and (3.37), we obtain

$$R_{1,0} = hf + \frac{1}{2}\Delta\widehat{W}^2 g'g + R_{2,0}. \quad (3.40)$$

Further, it follows from the Taylor expansion, the boundedness of g''' and (3.34) that

$$R_{2,g}^{(1)} = \frac{1}{6}\Delta\widehat{W}^2 g''(g,g) + \widehat{R}_{2,g}^{(1)} \quad \text{with} \quad [E|\widehat{R}_{2,g}^{(1)}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}). \quad (3.41)$$

By (3.40), (3.41) and the definition of $R_{2,g}$, one has

$$R_{2,g} = \frac{1}{2}hg'f + \frac{1}{4}\Delta\widehat{W}^2 g'(g'g) + \frac{1}{6}\Delta\widehat{W}^2 g''(g,g) + \tilde{R}_{2,g} \quad \text{with}$$
$$[E|\tilde{R}_{2,g}^{(1)}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}).$$

The above formula and (2.12) yield

$$\left|E(\Delta\widehat{W}R_{2,g})\right| = \mathscr{O}(h^2). \tag{3.42}$$

Due to (3.23) and Lemma 3.6,

$$\begin{aligned}&\left|E\left(h\int_0^1 \nabla L(\sigma(\tau))\mathrm{d}\tau\alpha_0\right)\right| = \mathscr{O}(h^2) \quad \text{and}\\ &\left|E\left(\Delta\widehat{W}\int_0^1 \nabla L(\sigma(\tau))\mathrm{d}\tau\alpha_1\right)\right| = \mathscr{O}(h^2).\end{aligned} \tag{3.43}$$

Combining (3.38), (3.39), (3.42) and (3.43), we have

$$\left|ER_{2,0}\right| = \mathscr{O}(h^2). \tag{3.44}$$

Step 3: Comparison between (3.37) *and Milstein method.*
Consider the one-step Milstein approximation $\bar{Y}^{[M]}$ for (3.1):

$$\bar{Y}^{[M]} = y + \Delta W g + hf + \frac{1}{2}\Delta W^2 g' g. \tag{3.45}$$

As is well known, Milstein method (3.45) satisfies Theorem 2.2 with $p_1 = 1.5,\ p_2 = 2$. Comparing the expansion (3.37) of our method (3.3) with Milstein method, we get

$$\begin{aligned}\bar{Y} - \bar{Y}^{[M]} &= (\Delta\widehat{W} - \Delta W)g + \frac{1}{2}(\Delta\widehat{W}^2 - \Delta W^2)g'g + R_{2,0}\\ &= \sqrt{h}(\zeta_h - \xi)g + \frac{1}{2}h(\zeta_h^2 - \xi^2)g'g + R_{2,0}.\end{aligned} \tag{3.46}$$

Owing to Lemma 2.4, $|E(\zeta_h^2-\xi^2)| = \mathscr{O}(h^{2-\epsilon})$, for every $\epsilon \in (0,1)$, and $E(\zeta_h-\xi)^2 = \mathscr{O}(h^2)$. Moreover, one can prove $E(\zeta_h - \xi)^4 = \mathscr{O}(h^2)$. The Hölder inequality and the above evaluations lead to

$$E|\bar{Y} - \bar{Y}^{[M]}|^2 = \mathscr{O}(h^3), \quad |E(\bar{Y} - \bar{Y}^{[M]})| = \mathscr{O}(h^2). \tag{3.47}$$

Thus the proof of (II) is completed by Theorem 2.3. □

3.3 MAVF methods for conservative SDEs with multiple noises

In this section, we propose MAVF methods for conservative SDEs with multiple noises and prove that these methods are of mean square order 1 if noises are commutative.

We still suppose that $L(y) : \mathbb{R}^m \to \mathbb{R}^\nu$ is the invariant of (2.1). Based on the ideas of dealing with conservative SDEs with single noise, we construct the MAVF method for (2.1) as follows:

$$\begin{cases} \bar{Y} = y + h\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \,\mathrm{d}\tau\, \alpha_0\right] \\ \quad + \sum_{r=1}^{D} \Delta \widehat{W}_r \left[\int_0^1 g_r(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \,\mathrm{d}\tau\, \alpha_r\right], \\ \left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top \,\mathrm{d}\tau\right] \alpha_0 = \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau, \\ \left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top \,\mathrm{d}\tau\right] \alpha_r = \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 g_r(\sigma(\tau))\,\mathrm{d}\tau, \\ \quad r = 1, \ldots, D, \end{cases} \tag{3.48}$$

where $\sigma(\tau) = y + \tau(\bar{Y} - y)$ and $\Delta \widehat{W}_r = \sqrt{h}\zeta_{rh}$ is defined by (2.9) with $k = 2$. In addition, α_r, $r = 0, 1, \ldots, D$, are $\mathbb{R}^\nu$-valued random variables. And $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_D)$ is called the modification coefficient.

Remark 3.8 If the invariant L of (2.1) is quadratic, and f, g_r, $r = 1, \ldots, D$, are linear functions, then $\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau = f(\frac{\bar{Y}+y}{2})$, $\int_0^1 g_r(\sigma(\tau))\,\mathrm{d}\tau = g_r(\frac{\bar{Y}+y}{2})$, $\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau = \nabla L(\frac{\bar{Y}+y}{2})$. Noting that $\nabla L f = \nabla L g_r = \mathbf{0}$, we have that $\boldsymbol{\alpha} = \mathbf{0}$. In this case, method (3.48) becomes the stochastic midpoint method.

Theorem 3.9 *Let $f \in \mathbf{C}_b^2(\mathbb{R}^m, \mathbb{R}^m)$ and $g_r \in \mathbf{C}_b^3(\mathbb{R}^m, \mathbb{R}^m)$, $r = 1, \ldots, D$. Suppose that $\nabla L(y) \nabla L(y)^\top$ is invertible on $\mathscr{M}$, and that for some bounded $\mathscr{M}_\delta$, $\nabla L \in \mathbf{C}^1(\mathscr{M}_\delta)$. Then the numerical method generated by (3.48) for (2.1) possesses the following properties:*

(I) *It preserves multiple invariants L^i, $i = 1, \ldots, \nu$, of (2.1), i.e., $L(\bar{Y}) = L(y)$.*

(II) *If the noises of (2.1) satisfy the commutative conditions, i.e., $g_r' g_i = g_i' g_r$, $i, r = 1, \ldots, D$, it is of mean square order* 1.

Proof Given that the proof is similar to that of Theorem 3.7, we only give the sketch. The first property (I) easily comes out as in the proof of Lemma 3.1.

Let us proceed to the proof of (II). First, as is done in Lemma 3.1, we acquire the solvability of (3.48) and have that for every $\epsilon \geq 0$, there exists an h_0 such that for all $h \leq h_0$,

$$|\boldsymbol{\alpha}| \leq \epsilon \quad a.s.$$

Then, analogous to Lemma 3.6, we have that

$$[E|\boldsymbol{\alpha}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h), \qquad |E(\Delta \widehat{W}_r \boldsymbol{\alpha})| = \mathscr{O}(h^2),\ r = 1, \ldots, D.$$

Further, using the Taylor expansion repeatedly, we acquire

$$\bar{Y} = y + \sum_{r=1}^{D} \Delta \widehat{W}_r g_r + \frac{1}{2} \sum_{r=1}^{D} \sum_{i=1}^{D} \Delta \widehat{W}_r \Delta \widehat{W}_i g_r' g_i + hf + R, \tag{3.49}$$

where

$$R = \frac{1}{2}h\sum_{r=1}^{D}\Delta\widehat{W}_r f' g_r + \frac{1}{2}h\sum_{r=1}^{D}\Delta\widehat{W}_r g_r' f - \sum_{r=1}^{D}\Delta\widehat{W}_r \nabla L^\top \alpha_r$$
$$+\frac{1}{6}\sum_{r,i,j=1}^{D}\Delta\widehat{W}_r\Delta\widehat{W}_i\Delta\widehat{W}_j g_r''(g_i,g_j) + \frac{1}{4}\sum_{r,i,j=1}^{D}\Delta\widehat{W}_r\Delta\widehat{W}_i\Delta\widehat{W}_j g_r'(g_i' g_j) + \tilde{R},$$

with $[E|\tilde{R}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^2)$.

Thus, it follows that

$$[E|R|^2]^{1/2} = \mathscr{O}(h^{1.5}), \qquad |ER| = \mathscr{O}(h^2).$$

Further, Milstein method for (2.1) is

$$\bar{Y}^{[M]} = y + hf + \sum_{r=1}^{D} g_r \Delta W_r + \frac{1}{2}\sum_{r=1}^{D} g_r' g_r \Delta W_r^2$$
$$+\sum_{i\neq r} g_r' g_i \int_t^{t+h} (w_i(\theta) - w_i(t))\,\mathrm{d}W_r(\theta), \tag{3.50}$$

which has first order convergence in mean square sense under the assumptions on f, g_r. Note the fact

$$\int_t^{t+h}(w_i(\theta) - w_i(t))\,\mathrm{d}W_r(\theta) + \int_t^{t+h}(w_r(\theta) - w_r(t))\,\mathrm{d}W_i(\theta) = \Delta w_i \Delta w_r.$$

This means, in the case of commutative noises,

$$\sum_{i\neq r} g_r' g_i \int_t^{t+h}(w_i(\theta) - w_i(t))\,\mathrm{d}W_r(\theta) = \sum_{i<r} g_r' g_i \Delta w_i \Delta w_r = \frac{1}{2}\sum_{i\neq r} g_r' g_i \Delta w_i \Delta w_r. \tag{3.51}$$

With (3.51), Milstein method (3.50) becomes

$$\bar{Y}^{[M]} = y + \sum_{r=1}^{D}\Delta W_r g_r + \frac{1}{2}\sum_{r=1}^{D}\sum_{i=1}^{D}\Delta W_i \Delta W_r g_r' g_i + hf. \tag{3.52}$$

Comparing (3.49) and (3.52), we have

$$[E|\bar{Y} - \bar{Y}^{[M]}|^2]^{1/2} = \mathscr{O}(h^{1.5}), \qquad |E(\bar{Y} - \bar{Y}^{[M]})| = \mathscr{O}(h^2), \tag{3.53}$$

which means that the method (3.48) is of mean square order 1 by Theorem 2.3. □

Remark 3.10 It is noted that if the noises are not commutative, the mean square order of the method (3.48) is only $\frac{1}{2}$. In this case, (3.51) could not be obtained. Thus when comparing (3.48) and (3.50), one obtains $[E|\bar{Y}-\bar{Y}^{[M]}|^2]^{1/2}=\mathscr{O}(h)$ and $|E(\bar{Y}-\bar{Y}^{[M]})|=\mathscr{O}(h^2)$, which leads to $\frac{1}{2}$ convergence order in the mean square sense.

4 Numerical integration

When the integrals contained in MAVF methods can not be obtained directly, we need to use the numerical integration to approximate the integrals. In this section, we investigate the effect of the numerical integration on MAVF methods, including the mean square convergence order and the preservation of invariants.

Here, we recall some concepts of numerical integration. Consider the quadrature formula $(c_i, b_i)_{i=1}^M$ on the interval [0, 1]:

$$\int_0^1 f(x)\,\mathrm{d}x \approx \sum_{i=1}^M b_i f(c_i). \tag{4.1}$$

The quadrature formula (4.1) is said to have order q if it is exact for polynomials of degree no larger than $q-1$, i.e.,

$$\int_0^1 x^k\,\mathrm{d}x = \sum_{i=1}^M b_i c_i^k, \quad k=0,1,\ldots,q-1.$$

Here are some examples of the quadrature formulas:

$$\int_0^1 f(x)\,\mathrm{d}x \approx \frac{1}{2}[f(0)+f(1)], \tag{4.2}$$

$$\int_0^1 f(x)\,\mathrm{d}x \approx \frac{1}{4}\left[3f\left(\frac{1}{3}\right)+f(1)\right], \tag{4.3}$$

$$\int_0^1 f(x)\,\mathrm{d}x \approx \frac{1}{2}\left[f\left(\frac{3-\sqrt{3}}{6}\right)+f\left(\frac{3+\sqrt{3}}{6}\right)\right], \tag{4.4}$$

$$\int_0^1 f(x)\,\mathrm{d}x \approx \frac{1}{18}\left[5f\left(\frac{5-\sqrt{15}}{10}\right)+8f\left(\frac{1}{2}\right)+5f\left(\frac{5+\sqrt{15}}{10}\right)\right], \tag{4.5}$$

and their orders are 2, 3, 4, 6, respectively.

As is well known, if $f\in\mathbf{C}^q$ with q being the order of the quadrature formula (4.1), then it holds that

$$\int_0^1 f(x)\,\mathrm{d}x = \sum_{i=1}^M b_i f(c_i) + \rho_q f^{(q)}(\eta), \tag{4.6}$$

where $\eta \in (0, 1)$ and ρ_q is independent of f. Next, we use the numerical integration to approximate the integrals in (3.48). The induced numerical method using the quadrature formula (4.1) is

$$\widetilde{Y} = y + h\left[\sum_{i=1}^{M} b_i f(\sigma(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))^\top \alpha_0\right] \\ + \sum_{r=1}^{D} \Delta\widehat{W}_r \left[\sum_{i=1}^{M} b_i g_r(\sigma(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))^\top \alpha_r\right], \tag{4.7a}$$

$$\left[\sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))\right]\left[\sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))^\top\right]\alpha_0 \\ = \left[\sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))\right]\left[\sum_{i=1}^{M} b_i f(\sigma(c_i))\right], \tag{4.7b}$$

$$\left[\sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))\right]\left[\sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))^\top\right]\alpha_r \\ = \left[\sum_{i=1}^{M} b_i \nabla L(\sigma(c_i))\right]\left[\sum_{i=1}^{M} b_i g_r(\sigma(c_i))\right], \quad r = 1, \ldots, D, \tag{4.7c}$$

where $\sigma(\tau) = y + \tau(\widetilde{Y} - y)$.

4.1 Mean square convergence order

In this part, we study the mean square convergence order of (4.7). Following the procedure in Sect. 3, we firstly present the boundedness of $\boldsymbol{\alpha}$, and the expansion formula of $\bar{Y}$. Hereafter, we denote $P(z, y) := \left[\sum_{i=1}^{M} b_i \nabla L(y + c_i(z - y))\right]\left[\sum_{i=1}^{M} b_i \nabla L(y + c_i(z - y))^\top\right]$.

Lemma 4.1 *Assume that $f, g_r \in \mathbf{C}_b^1(\mathbb{R}^m, \mathbb{R}^m) \cap C_b(\mathbb{R}^m, \mathbb{R}^m)$, $r = 1, \ldots, D$, $L \in \mathbf{C}_b^2(\mathbb{R}^m)$, and that $P(y, z)$ is invertible for each $y, z \in \mathbb{R}^m$ with $P^{-1} \in C_b^1(\mathbb{R}^m \times \mathbb{R}^m) \cap C_b(\mathbb{R}^m \times \mathbb{R}^m)$. Then for any given $y \in \mathbb{R}^m$, the method (4.7) is uniquely solvable with respect to $\widetilde{Y}$ and $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_D)$, a.s., for sufficiently small stepsize h. In addition, there is a representation*

$$\widetilde{Y} = y + \sum_{r=1}^{D} \Delta\widehat{W}_r g_r + R_{\widetilde{Y}}, \tag{4.8}$$

with $\left[E|R_{\widetilde{Y}}|^{2p}\right]^{\frac{1}{2p}} = \mathscr{O}(h)$, $p = 1, 2, \ldots$. Moreover, for every $\epsilon > 0$, there exists some $h_0 > 0$ such that for all $h \le h_0$,

$$|\boldsymbol{\alpha}| \le \epsilon, \quad a.s. \tag{4.9}$$

Proof Considering that this proof is similar to that of Lemma 3.1, we give the sketch of the proof. Under the assumptions of this lemma, one can use the contraction mapping principle to prove that there is a unique $\widetilde{Y}$ satisfying (4.7) for sufficiently small stepsize h independent of ω. It follows from the assumptions on f, g_r and L that (4.8) holds. Then using similar arguments as in the proof of Lemma 3.1, one can derive (4.9). □

Remark 4.2 In Lemma 4.1, in order to acquire the solvability of the MAVF method (4.7) using numerical integration, we make stronger assumptions than those in Lemma 3.1. In practice, even though the numerical solution of this method may leave the manifold $\mathscr{M}$, the invariant-preserving error can still be controlled (see Theorem 4.9), and the numerical solutions stay in some neighbourhood of $\mathscr{M}$ (see for instance, the numerical experiments in Sect. 5.2). It means that (4.7) could be applied to some systems provided that $\nabla L \nabla L^T$ is invertible on some neighbourhood of $\mathscr{M}$. It would be interesting to investigate the mild assumptions to ensure the solvability of the MAVF method (4.7) for a uniform stepsize independent of ω.

The following lemma is used to estimate the accuracy of the numerical integration in (4.7).

Lemma 4.3 *Let q be the order of the quadrature formula $(c_i, b_i)_{i=1}^M$ and G be an arbitrary scalar or vector-valued function. Let the assumptions of Lemma 4.1 hold. If $q \geq 2$ and $G^{(q)} \in \mathbf{C}_b(\mathbb{R}^m)$, then we have*

$$\int_0^1 G(\sigma(\tau))\,\mathrm{d}\tau = \sum_{i=1}^{M} b_i G(\sigma(c_i)) + \Psi_{G,q}, \tag{4.10}$$

where $\Psi_{G,q}$ satisfies

$$\left[E|\Psi_{G,q}|^{2p}\right]^{\frac{1}{2p}} = \mathscr{O}\left(h^{\frac{q}{2}}\right), \quad \forall\ p = 1, 2, \ldots \tag{4.11}$$

In addition, it holds that

(I) *If q is odd, then*

$$\left|E\left(\Delta\widehat{W}_r \Psi_{G,q}\right)\right| = \mathscr{O}\left(h^{\frac{q+1}{2}}\right), \quad r = 1, \ldots, D. \tag{4.12}$$

(II) *If q is even and $G^{(q+1)} \in \mathbf{C}_b(\mathbb{R}^m)$, then*

$$\left|E\left(\Delta\widehat{W}_r \Psi_{G,q}\right)\right| = \mathscr{O}\left(h^{\frac{q+2}{2}}\right), \quad r = 1, \ldots, D. \tag{4.13}$$

Proof We only prove the case that G is a scalar function, the case of vector-valued function is analogous. According to (4.6), we have

$$\int_0^1 G(\sigma(\tau))\,\mathrm{d}\tau = \sum_{i=1}^{M} b_i G(\sigma(c_i)) + \Psi_{G,q},$$

where $\Psi_{G,q} = \rho_q \frac{\mathrm{d}^q}{\mathrm{d}\tau^q} G(\sigma(\tau))\Big|_{\tau=\eta}$ with $\eta \in (0, 1)$.

Besides, it holds that

$$\frac{\mathrm{d}^q}{\mathrm{d}\tau^q} G(\sigma(\tau))\Big|_{\tau=\eta} = G^{(q)}(\sigma(\eta))(\underbrace{\widetilde{Y} - y, \ldots, \widetilde{Y} - y}_{q}). \tag{4.14}$$

Due to (4.8) and boundedness of $G^{(q)}$, we have

$$[E|\Psi_{G,q}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}\left(h^{\frac{q}{2}}\right).$$

(I) If q is odd, the Hölder inequality yields

$$|E\Delta\widehat{W}_r\Psi_{G,q}| = \mathscr{O}\left(h^{\frac{q+1}{2}}\right), \quad r = 1, \ldots, D.$$

(II) If q is even, it holds that

$$E\left(\Delta\widehat{W}_{i_1}\Delta\widehat{W}_{i_2}\cdots\Delta\widehat{W}_{i_{q+1}}\right) = 0, \quad \forall\, i_1, i_2, \ldots, i_{q+1} \in \{1, 2, \ldots, D\}.$$

Since $G^{(q+1)} \in \mathbf{C}_b$, we have $G^{(q)}(\sigma(\eta)) = G^{(q)}(y) + R_G$, with $[E|R_G|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1/2})$. Thus, using the Hölder inequality and (4.8) gives

$$|E\Delta\widehat{W}_r\Psi_{G,q}| = \mathscr{O}(h^{\frac{q+2}{2}}), \quad r = 1, \ldots, D.$$

□

Next, we give the a prior estimate of the modification coefficient $\boldsymbol{\alpha}$ in (4.7).

Lemma 4.4 *Let q be the order of the quadrature formula $(c_i, b_i)_{i=1}^M$ in (4.7). Let the assumptions of Lemma 4.1 hold. Assume that $f,\ g_r,\ \nabla L \in \mathbf{C}_b^3(\mathbb{R}^m)$, $r = 1, \ldots, D$. If $q \geq 2$, then we have*

$$\left[E|\boldsymbol{\alpha}|^{2p}\right]^{\frac{1}{2p}} = \mathscr{O}(h),\ \ p = 1, 2, \ldots \qquad \textit{and} \qquad |E(\Delta\widehat{W}_r\boldsymbol{\alpha})| = \mathscr{O}(h^2), \quad r = 1, \ldots, D. \tag{4.15}$$

Proof By (4.10), (4.7b) in (4.7) can be rewritten as

$$\begin{aligned}&\left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau - \Psi_{\nabla L,q}\right]\left[\int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau - \Psi_{\nabla L,q}^\top\right]\alpha_0\\ &\quad= \left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau - \Psi_{\nabla L,q}\right]\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \Psi_{f,q}\right].\end{aligned} \tag{4.16}$$

By arranging the above formula, we have

$$\begin{aligned}&\left[\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau\right]\alpha_0\\&\quad=\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau + T_{\alpha_0},\end{aligned} \tag{4.17}$$

where

$$\begin{aligned}T_{\alpha_0} =& -\int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau \Psi_{f,q} - \Psi_{\nabla L,q}\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau + \Psi_{\nabla L,q}\Psi_{f,q}\\&+\left[\Psi_{\nabla L,q}\int_0^1 \nabla L(\sigma(\tau))^\top\,\mathrm{d}\tau + \int_0^1 \nabla L(\sigma(\tau))\,\mathrm{d}\tau\Psi_{\nabla L,q}^\top - \Psi_{\nabla L,q}\Psi_{\nabla L,q}^\top\right]\alpha_0.\end{aligned}$$

Using (3.23) and (3.25), we write (4.17) as

$$(\nabla L + R_{1,L})(\nabla L^\top + R_{1,L}^\top)\alpha_0 = K_1\nabla L'(y)(\widetilde{Y}-y)(f'(y)(\widetilde{Y}-y)) + R_{1,\alpha_0} + T_{\alpha_0},$$

where $[E|R_{1,\alpha_0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5})$.

Further, since $[\nabla L\nabla L^\top]^{-1}$ is invertible, we have

$$\begin{aligned}\alpha_0 =& -[\nabla L\nabla L^\top]^{-1}(\nabla L R_{1,L}^\top + R_{1,L}\nabla L^\top + R_{1,L}R_{1,L}^\top)\alpha_0 + [\nabla L\nabla L^\top]^{-1}T_{\alpha_0}\\&+K_1[\nabla L\nabla L^\top]^{-1}\nabla L'(y)(\widetilde{Y}-y)(f'(y)(\widetilde{Y}-y)) + [\nabla L\nabla L^\top]^{-1}R_{1,\alpha_0}.\end{aligned} \tag{4.18}$$

Next we prove the conclusion for different values of q.

Case 1 Since $q=2$ is even and $f, g_r, \nabla L \in \mathbf{C}_b^3(\mathbb{R}^m)$, it follows from Lemma 4.3 that

$$\left[E|\Psi_{f,q}|^{2p}\right]^{\frac{1}{2p}} = \mathscr{O}(h),\quad \left[E|\Psi_{\nabla L,q}|^{2p}\right]^{\frac{1}{2p}} = \mathscr{O}(h),$$

and

$$\left|E\left(\Delta\widehat{W}_r\Psi_{f,q}\right)\right| = \mathscr{O}(h^2),\quad \left|E\left(\Delta\widehat{W}_r\Psi_{\nabla L,q}\right)\right| = \mathscr{O}(h^2).$$

Hence, $[E|T_{\alpha_0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h)$, and $|E\Delta\widehat{W}_rT_{\alpha_0}| = \mathscr{O}(h^2)$. Analogous to the proof of Lemma 3.6, we have

$$[E|\boldsymbol{\alpha}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h),\qquad |E(\Delta\widehat{W}_r\boldsymbol{\alpha})| = \mathscr{O}(h^2),\quad r=1,\ldots,D.$$

Case 2 If $q=3$, according to Lemma 4.3, we obtain

$$[E|\Psi_{f,q}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}),\quad [E|\Psi_{\nabla L,q}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}).$$

Then the Hölder inequality yields

$$[E|T_{\alpha_0}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}), \qquad |E\Delta\widehat{W}_r T_{\alpha_0}| = \mathscr{O}(h^2).$$

Analogous to the proof of Lemma 3.6, we have

$$[E|\boldsymbol{\alpha}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h), \qquad |E(\Delta\widehat{W}_r\boldsymbol{\alpha})| = \mathscr{O}(h^2), \quad r = 1, \ldots, D.$$

Case 3 For the case of $q \geq 4$, since $f, \nabla L \in \mathbf{C}_b^3(\mathbb{R}^m)$, the proof is the same as that of *Case* 2.

Combining the above conclusions, we complete the proof. □

Next we give the result of convergence of the method (4.7).

Theorem 4.5 *Let the assumptions of Lemma 4.4 hold. If the noises satisfy the commutative conditions, i.e., $g_r' g_i = g_i' g_r$, $i, r = 1, \ldots, D$, then the method generated by (4.7) is of mean square order* 1.

Proof This proof is analogous to that of Theorem 3.9, we only give the sketch. It follows from Lemma 4.3 that

$$\begin{aligned}
\widetilde{Y} = & \, y + h\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \Psi_{f,q} - \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\alpha_0 + \Psi_{\nabla L,q}^\top \alpha_0\right] \\
& + \sum_{r=1}^{D} \Delta\widehat{W}_r \left[\int_0^1 g_r(\sigma(\tau))\,\mathrm{d}\tau - \Psi_{g_r,q} - \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\alpha_r + \Psi_{\nabla L,q}^\top \alpha_r\right] \\
= & \, y + h\left[\int_0^1 f(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\alpha_0\right] \\
& + \sum_{r=1}^{D} \Delta\widehat{W}_r \left[\int_0^1 g_r(\sigma(\tau))\,\mathrm{d}\tau - \int_0^1 \nabla L(\sigma(\tau))^\top \mathrm{d}\tau\alpha_r\right] + R^{[0]},
\end{aligned} \tag{4.19}$$

where

$$R^{[0]} = -h\Psi_{f,q} + h\Psi_{\nabla L,q}^\top \alpha_0 + \sum_{r=1}^{D} \Delta\widehat{W}_r \left[-\Psi_{g_r,q} + \Psi_{\nabla L,q}^\top \alpha_r\right].$$

According to Lemmas 4.3 and 4.4, we have

$$[E|R^{[0]}|^{2p}]^{\frac{1}{2p}} = \mathscr{O}(h^{1.5}), \quad |ER^{[0]}| = \mathscr{O}(h^2).$$

Using the Taylor expansion, analogous to the proof of Theorem 3.9, we obtain

$$\widetilde{Y} = y + \sum_{r=1}^{D} \Delta\widehat{W}_r g_r + \frac{1}{2}\sum_{r=1}^{D} \Delta\widehat{W}_r^2 g_r' g_r + \sum_{r=1}^{D-1}\sum_{i=r+1}^{D} \Delta\widehat{W}_i \Delta\widehat{W}_r g_r' g_i + hf + T^{[0]}. \tag{4.20}$$

where

$$T^{[0]} = \frac{1}{2}h\sum_{r=1}^{D}\Delta\widehat{W}_r f' g_r + \frac{1}{2}h\sum_{r=1}^{D}\Delta\widehat{W}_r g_r' f - \sum_{r=1}^{D}\Delta\widehat{W}_r \nabla L^{\top}\alpha_r + \frac{1}{6}\sum_{r,i,j=1}^{D}\Delta\widehat{W}_r\Delta\widehat{W}_i\Delta\widehat{W}_j g_r''(g_i, g_j) + \frac{1}{4}\sum_{r,i,j=1}^{D}\Delta\widehat{W}_r\Delta\widehat{W}_i\Delta\widehat{W}_j g_r'(g_i' g_j) + R^{[0]} + \widehat{R}^{[0]}, \tag{4.21}$$

with $\left[E|\widehat{R}^{[0]}|^2\right]^{1/2} = \mathscr{O}(h^2)$. Thus, it holds that

$$[E|T^{[0]}|^2]^{1/2} = \mathscr{O}(h^{1.5}), \quad |ET^{[0]}| = \mathscr{O}(h^2).$$

Comparing (4.20) with Milstein method for (2.1), we obtain that the method (4.7) is of mean square order 1. □

Remark 4.6 As is explained in Remark 3.10, if the noises are not commutative, the mean square convergence order of (4.7) is $\frac{1}{2}$.

4.2 Invariant-preserving order in mean square sense

It is worth noting that the method (4.7) does not generally preserve exactly the invariant of the original system, due to the use of the quadrature formula, which makes it necessary to study the preservation of invariant of (4.7). In the following, we give the definition of invariant-preserving order in mean square sense.

Definition 4.7 A numerical discretization $\{y_n\}_{n=0}^N$ is said to have invariant-preserving order p in mean square sense, if the invariant $L(y)$ of (2.1) satisfies

$$[E|L(y_N) - L(y_0)|^2]^{1/2} = \mathscr{O}(h^p). \tag{4.22}$$

Let $\{y_n\}_{n=0}^N$ be the numerical discretization corresponding to the one-step approximation (4.7) with $y_0 = Y_0$, and denote $\widetilde{Y}_{t_n,y_n}(t_{n+1}) = y_{n+1}$, $n = 0, 1, \ldots, N-1$. The numerical method generated from the one-step approximation (4.7) reads

$$\widetilde{Y}_{t_n,y_n}(t_{n+1}) = y_n + h\left[\sum_{i=1}^{M} b_i f(\sigma_n(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^{\top}\alpha_0\right] + \sum_{r=1}^{D}\Delta\widehat{W}_{r,n}\left[\sum_{i=1}^{M} b_i g_r(\sigma_n(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^{\top}\alpha_r\right], \tag{4.23a}$$

$$\left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))\right]\left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^{\top}\right]\alpha_0 = \left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))\right]\left[\sum_{i=1}^{M} b_i f(\sigma_n(c_i))\right], \tag{4.23b}$$

$$\left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))\right]\left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^\top\right]\alpha_r = \left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))\right] \times \left[\sum_{i=1}^{M} b_i g_r(\sigma_n(c_i))\right], \quad r = 1, \ldots, D, \tag{4.23c}$$

where $\sigma_n(\tau) = y_n + \tau(\widetilde{Y}_{t_n,y_n}(t_{n+1}) - y_n))$, and $\Delta\widehat{W}_{r,n} = \Delta\widehat{W}_r(t_{n+1}) - \Delta\widehat{W}_r(t_n)$, $r = 1, \ldots, D$, $n = 0, 1, \ldots, N-1$, are mutually independent truncated Brownian increments.
The following lemma gives the one-step error estimate on the conservation of the invariant of (4.23).

Lemma 4.8 *Let the assumptions of Lemma 3.1 hold. Assume that* $(\nabla L)^{(q+1)} \in \mathbf{C}_b(\mathbb{R}^m)$. *Let q be the order of the quadrature formula* $(c_i, b_i)_{i=1}^M$ *in (4.7). If* $q \geq 2$, *then it holds that*

$$\left[E|L(\widetilde{Y}_{t_n,y_n}(t_{n+1})) - L(y_n)|^2\right]^{1/2} = \mathscr{O}\left(h^{\frac{q+1}{2}}\right), \quad n = 0, 1, \ldots, N-1, \tag{4.24}$$

and

$$\left|E\left[L(\widetilde{Y}_{t_n,y_n}(t_{n+1})) - L(y_n)\middle| \mathscr{F}_{t_n}\right]\right| = \begin{cases} \mathscr{O}\left(h^{\frac{q+2}{2}}\right), & \text{if } q \text{ is even}, \\ \mathscr{O}\left(h^{\frac{q+1}{2}}\right), & \text{if } q \text{ is odd}. \end{cases} \tag{4.25}$$

Proof Let y_n denote a random variable and $y \in \mathscr{M}$ denote a deterministic variable in this proof. It follows from the Taylor expansion that

$$L(\widetilde{Y}_{t_n,y_n}(t_{n+1})) - L(y_n) = \int_0^1 \nabla L(\sigma_n(\tau))\,\mathrm{d}\tau(\widetilde{Y}_{t_n,y_n}(t_{n+1}) - y_n). \tag{4.26}$$

By (4.10), we have

$$\int_0^1 \nabla L(\sigma_n(\tau))\,\mathrm{d}\tau = \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i)) + \Psi_{\nabla L,q}, \tag{4.27}$$

with $\left[E|\Psi_{\nabla L,q}|^{2p}\right]^{\frac{1}{2p}} = \mathscr{O}\left(h^{\frac{q}{2}}\right)$.

Substituting (4.27) and (4.23a) into (4.26) gives

$$\begin{aligned} &L(\widetilde{Y}_{t_n,y_n}(t_{n+1})) - L(y_n) \\ &= h\left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i)) + \Psi_{\nabla L,q}\right]\left[\sum_{i=1}^{M} b_i f(\sigma_n(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^\top \alpha_0\right] \\ &\quad + \sum_{r=1}^{D} \Delta\widehat{W}_{r,n}\left[\sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i)) + \Psi_{\nabla L,q}\right] \end{aligned}$$

$$\times \left[\sum_{i=1}^{M} b_i g_r(\sigma_n(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^\top \alpha_r \right].$$

Utilizing (4.23b) and (4.23c), we obtain

$$\begin{aligned} & L(\widetilde{Y}_{t_n,y_n}(t_{n+1})) - L(y_n) \\ & = h\Psi_{\nabla L,q} \left[\sum_{i=1}^{M} b_i f(\sigma_n(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^\top \alpha_0 \right] \\ & \quad + \sum_{r=1}^{D} \Delta \widehat{W}_{r,n} \Psi_{\nabla L,q} \left[\sum_{i=1}^{M} b_i g_r(\sigma_n(c_i)) - \sum_{i=1}^{M} b_i \nabla L(\sigma_n(c_i))^\top \alpha_r \right]. \end{aligned} \tag{4.28}$$

In order to acquire (4.24), it suffices to estimate the lowest-order term

$$\sum_{r=1}^{D} \Delta \widehat{W}_{r,n} \Psi_{\nabla L,q} \left[\sum_{i=1}^{M} b_i g_r(\sigma_n(c_i)) \right].$$

According to assumptions on f, g_r, ∇L and the Hölder inequality, it follows from (4.11) that

$$\begin{aligned} E|L(\widetilde{Y}_{t_n,y_n}(t_{n+1})) - L(y_n)|^2 & \le K \sum_{r=1}^{D} E\left[|\Delta \widehat{W}_{r,n}|^2 |\Psi_{\nabla L,q}|^2 \right] + Kh^{q+2} \\ & \le Kh \left[E|\Psi_{\nabla L,q}|^4 \right]^{1/2} + Kh^{q+2} \\ & \le Kh^{q+1}, \end{aligned} \tag{4.29}$$

which proves (4.24).

If y_n is replaced by the deterministic variable y, we are able to use Lemma 4.3 and acquire that

$$|E\Delta \widehat{W}_r \Psi_{\nabla L,q}| = \begin{cases} \mathscr{O}\left(h^{\frac{q+2}{2}}\right), & \text{if } q \text{ is even}, \\ \mathscr{O}\left(h^{\frac{q+1}{2}}\right), & \text{if } q \text{ is odd}. \end{cases} \tag{4.30}$$

Thus, we have

$$\left| E\left[L(\widetilde{Y}_{t_n,y}(t_{n+1})) - L(y) \right] \right| = \begin{cases} \mathscr{O}\left(h^{\frac{q+2}{2}}\right), & \text{if } q \text{ is even}, \\ \mathscr{O}\left(h^{\frac{q+1}{2}}\right), & \text{if } q \text{ is odd}. \end{cases}$$

Notice that y_n is $\mathscr{F}_{t_n}$-measurable and that $\widetilde{Y}_{t_n,y}(t_{n+1}) - L(y)$ is $\mathscr{F}_{t_n}$-independent. According to the property of conditional expectations (see [10, Chapter 1]), we have

$$E\left[L(\widetilde{Y}_{t_n,y_n}(t_{n+1}))-L(y_n)\Big|\,\mathscr{F}_{t_n}\right]=\left(E\left[L(\widetilde{Y}_{t_n,y}(t_{n+1}))-L(y)\right]\right)\Big|_{y=y_n}. \quad (4.31)$$

In this way, we obtain (4.25). □

We now give the result about the invariant-preserving order in mean square sense for (4.23).

Theorem 4.9 *Under the assumptions of Lemma 4.8, it holds that*

$$\left[E|L(y_N)-L(y_0)|^2\right]^{1/2}=\begin{cases}\mathscr{O}\left(h^{\frac{q}{2}}\right), & \text{if } q \text{ is even},\\ \mathscr{O}\left(h^{\frac{q-1}{2}}\right), & \text{if } q \text{ is odd}.\end{cases} \quad (4.32)$$

Proof Denote $e_n=E|L(y_n)-L(y_0)|^2$, $n=0,1,\ldots,N$, and we have

$$\begin{aligned} e_{n+1} &= E|L(y_{n+1})-L(y_0)|^2\\ &= E|L(y_{n+1})-L(y_n)+L(y_n)-L(y_0))|^2\\ &= E|L(y_{n+1})-L(y_n)|^2+E|L(y_n)-L(y_0)|^2\\ &\quad +2E\left[(L(y_n)-L(y_0))^\top(L(y_{n+1})-L(y_n))\right]\\ &= e_n+E|L(y_{n+1})-L(y_n)|^2+2S, \quad n=0,1,\ldots,N-1, \end{aligned} \quad (4.33)$$

where $S=E\left[(L(y_n)-L(y_0))^\top(L(y_{n+1})-L(y_n))\right]$. Since y_n and y_0 are $\mathscr{F}_{t_n}$-measurable, it follows that

$$\begin{aligned} S &= E\left(E\left[(L(y_n)-L(y_0))^\top(L(y_{n+1})-L(y_n))\Big|\,\mathscr{F}_{t_n}\right]\right)\\ &= E\left[(L(y_n)-L(y_0))^\top E\left(L(y_{n+1})-L(y_n)|\,\mathscr{F}_{t_n}\right)\right]\\ &\le \left[E|L(y_n)-L(y_0)|^2\right]^{1/2}\left\{E\left|E\left[L(y_{n+1})-L(y_n)|\,\mathscr{F}_{t_n}\right]\right|^2\right\}^{1/2}\\ &= e_n^{1/2}\left\{E\left|E\left[L(y_{n+1})-L(y_n)|\,\mathscr{F}_{t_n}\right]\right|^2\right\}^{1/2}. \end{aligned} \quad (4.34)$$

Substituting (4.34) into (4.33) and using Young's inequality $ab\le\frac{1}{2}(a^2+b^2)$, we obtain

$$e_{n+1}\le e_n+E|L(y_{n+1})-L(y_n)|^2+he_n+h^{-1}\left\{E\left|E\left[L(y_{n+1})-L(y_n)|\,\mathscr{F}_{t_n}\right]\right|^2\right\}. \quad (4.35)$$

Note that $y_{n+1}=\widetilde{Y}_{t_n,y_n}(t_{n+1})$. Utilizing (4.24) and (4.25) in Lemma 4.8, we have

(1) If q is odd, then

$$\begin{aligned} e_{n+1} &\le e_n(1+h)+Kh^{q+1}+h^{-1}Kh^{q+2}\\ &\le e_n(1+h)+Kh^{q+1}. \end{aligned} \quad (4.36)$$

It follows from Gronwall's inequality (see [10, Lemma 1.6]) that $e_n \leq Kh^q$, i.e., $e_n^{1/2} \leq Kh^{\frac{q}{2}}, n = 0, 1, \ldots, N$.

(2) If q is odd, we similarly have

$$e_{n+1} \leq e_n(1+h) + Kh^q. \tag{4.37}$$

Thus, Gronwall's inequality leads to $e_n^{1/2} \leq Kh^{\frac{q-1}{2}}, n = 0, 1, \ldots, N$. This completes the proof. □

Remark 4.10 In fact, if $f, g_r,\ r = 1, \ldots, D$, and $\nabla L(y)$ are polynomials with degree no larger than $q-1$, then the quadrature formulas in (4.23) are exactly equal to the integrals in (3.48). In this case, (4.23) exactly preserves the invariant $L(y)$. For the general case, Theorem 4.9 implies that the invariant-preserving order in mean square sense of MAVF methods using numerical integration only depends on the order of the quadrature formulas.

5 Numerical experiments

In this section, we implement numerical experiments to verify our theoretical analyses. And we show the superiority of MAVF methods when applied to conservative SDEs.

5.1 MAVF methods

In this part, we take the method (3.48) as an example to explain how MAVF methods are applied to concrete problems. For a conservative SDE with coefficients satisfying the conditions of Theorem 3.9, then the numerical method generated by (3.48) can be written as

$$\begin{cases} y_{n+1} = y_n + h\left[\int_0^1 f(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau - \int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))^\top \mathrm{d}\tau\, \alpha_{0,n}\right] \\ \quad + \sum_{r=1}^{D} \Delta\widehat{W}_{r,n}\left[\int_0^1 g_r(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau - \int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))^\top \mathrm{d}\tau\, \alpha_{r,n}\right], \\ \alpha_{0,n} = \left[\int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau \int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))^\top \mathrm{d}\tau\right]^{-1} \\ \quad \times \left[\int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau \int_0^1 f(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau\right], \\ \alpha_{r,n} = \left[\int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau \int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))^\top \mathrm{d}\tau\right]^{-1} \\ \quad \times \left[\int_0^1 \nabla L(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau \int_0^1 g_r(y_n+\tau(y_{n+1}-y_n))\,\mathrm{d}\tau\right], \quad r = 1, \ldots, D, \end{cases} \tag{5.1}$$

where $\Delta\widehat{W}_{r,n} = \sqrt{h}\zeta_{rh,n}$ is the truncation of $W_r(t_{n+1}) - W_r(t_n) = \sqrt{h}\xi_{r,n}$ with $A_h = \sqrt{4|\ln h|}$ (see (2.9)). Note that (5.1) is an implicit method. We denote the right hand side of (5.1) by $F(y_n, y_{n+1}, \alpha_{0,n}, \ldots, \alpha_{D,n}, h, \Delta\widehat{W}_{1,n}, \ldots, \Delta\widehat{W}_{D,n})$, and then (5.1) is rewritten as $(y_{n+1}, \alpha_{0,n}, \ldots, \alpha_{D,n}) = F(y_n, y_{n+1}, \alpha_{0,n}, \ldots, \alpha_{D,n}, h, \Delta\widehat{W}_{1,n}, \ldots, \Delta\widehat{W}_{D,n})$. In order to approximately solve y_{n+1} and $\alpha_{i,n},\ i = 0, 1, \ldots, D$, the fixed point iteration is applied, i.e.,

$$\begin{aligned}&\left(y_{n+1}^{(P)}, \alpha_{0,n}^{(P)}, \ldots, \alpha_{D,n}^{(P)}\right)\\ &\quad= F\left(y_n, y_{n+1}^{(P-1)}, \alpha_{0,n}^{(P-1)}, \ldots, \alpha_{D,n}^{(P-1)}, h, \Delta\widehat{W}_{1,n}, \ldots, \Delta\widehat{W}_{D,n}\right),\end{aligned}$$

where $P = 1, 2, \ldots$ The initial value of iteration at each step can be taken as $y_{n+1}^{(0)} = y_n$ and $\alpha_{i,n}^{(0)} = 0,\ i = 0, 1, \ldots, D$. The iteration stops if $\max\Big\{|y_{n+1}^{(P)} - y_{n+1}^{(P-1)}|, |\alpha_{0,n}^{(P)} - \alpha_{0,n}^{(P-1)}|, \ldots, |\alpha_{D,n}^{(P)} - \alpha_{D,n}^{(P-1)}|\Big\} < \epsilon$, where ϵ is the given error tolerance.

In the following, we present three examples to compare MAVF methods with stochastic midpoint method (see e.g. [5]) and the projection method based on Milstein, i.e., the MilsteinP method (see [15]). These methods have first order convergence in mean square sense, if noises are commutative.

5.1.1 Example 1: Kubo oscillator

Consider the following stochastic harmonic oscillator in [9]

$$\begin{cases} \mathrm{d}X_1(t) = -aX_2(t)\mathrm{d}t - \sigma X_2(t) \circ \mathrm{d}W(t), \\ \mathrm{d}X_2(t) = aX_1(t)\mathrm{d}t + \sigma X_1(t) \circ \mathrm{d}W(t), \end{cases} \tag{5.2}$$

where a and σ are constants, and $W(t)$ is a one-dimensional Brownian motion. The quadratic function $I(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$ is the invariant of (5.2). In the numerical test, we take $a = \sigma = 1$ and the initial value $(X_1(0), X_2(0)) = (1, 0)$. For this system, the level set $\mathscr{M} = \left\{(y_1, y_2) \in \mathbb{R}^2 \middle| y_1^2 + y_2^2 = 1\right\}$. Then one could take $\mathscr{M}_\delta = \left\{(y_1, y_2) \in \mathbb{R}^2 \middle| 0.5 \le y_1^2 + y_2^2 \le 1.5\right\}$ and verify that the coefficients and the invariant I satisfy the conditions of Theorem 3.7. In fact, from Remark 3.8, we know that the MAVF method (3.3) for this system is reduced to stochastic midpoint method.

Figure 1 displays the convergence order of the MAVF method (3.3) and the MilsteinP method in mean square sense. Here, the reference solution is obtained by Milstein method with stepsize $h_{ref} = 2^{-14}$. The mean square errors are computed at the endpoint $T = 1$ by adopting five different stepsizes $h = 2^{-5},\ 2^{-6},\ 2^{-7}, 2^{-8}, 2^{-9}$. The expectation is realized by using the average of 1000 independent sample paths. The convergence of order one, as is shown in this figure, is observed for the MAVF method, which is consistent with theoretical analyses of Theorem 3.7.

Figures 2 and 3 show the long time behavior, including the evolution of errors of invariant and the evolution of the global mean square errors, of the MAVF method and the MilsteinP method when numerically simulating Kubo oscillator. The exact solution is simulated by using Milstein method with $h_{ref} = 10^{-5}$, and the expectation is approximated by using the average of 1000 independent sample paths. Figure 2

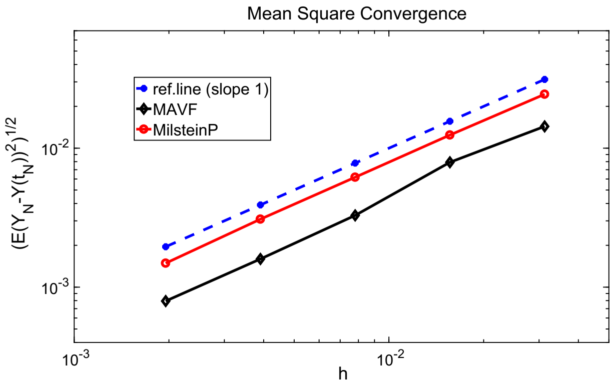

Fig. 1 Mean square errors of the MAVF method and the MilsteinP method at T = 1 for Kubo oscillator. The dashed reference line has slope 1

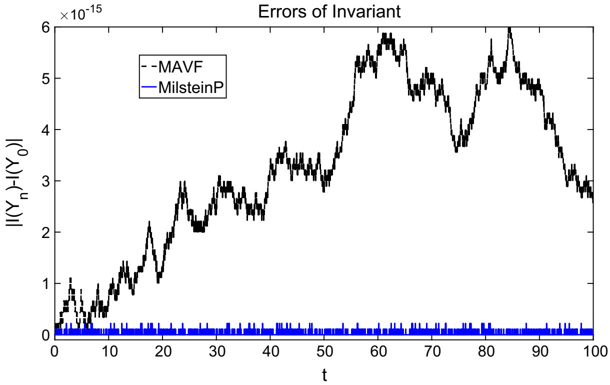

Fig. 2 Errors of invariant $I(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$ of the MAVF method and the MilsteinP method for Kubo oscillator with $T = 100$, $h = 0.01$

illustrates that both methods preserve the invariant if ignoring the round-off error. In terms of the evolution of the errors of invariant, the MilsteinP method behaves slightly better. Figure 3 displays the evolution of the global mean square errors in two different kinds of time intervals; see Fig. 3a for the time interval [0, 1], and Fig. 3b for the time interval [0, 100]. From Fig. 3, we observe that the mean square errors of the MAVF method are always smaller than that of the MilsteinP method, and that both of the errors evolve linearly.

5.1.2 Example 2: Stochastic cyclic Lotka–Volterra system

Consider the following stochastic dynamical system

$$\mathrm{d}\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} = \begin{pmatrix} X(t)(Z(t)-Y(t)) \\ Y(t)(X(t)-Z(t)) \\ Z(t)(Y(t)-X(t)) \end{pmatrix} \mathrm{d}t + c \begin{pmatrix} X(t)(Z(t)-Y(t)) \\ Y(t)(X(t)-Z(t)) \\ Z(t)(Y(t)-X(t)) \end{pmatrix} \circ \mathrm{d}W(t), \quad (5.3)$$

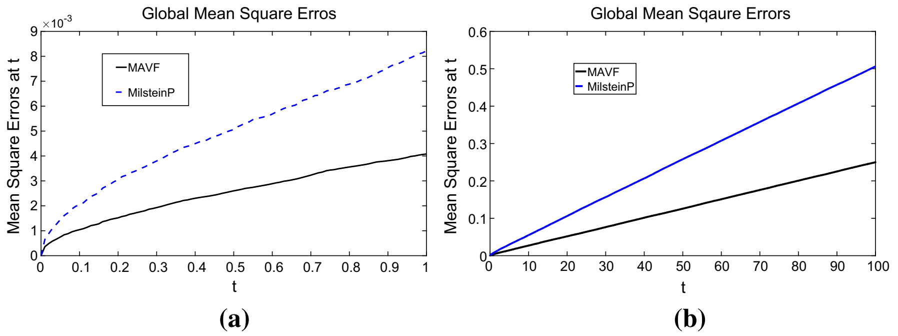

Fig. 3 Global mean square errors of the MAVF method and the MilsteinP method for Kubo oscillator with $h = 0.01$. The left one is plotted with $T = 1$ and the right one is with $T = 100$

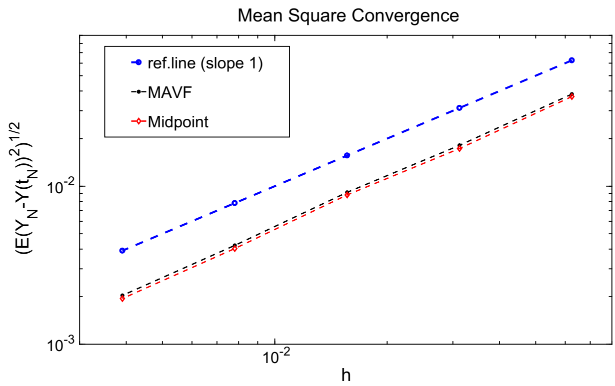

Fig. 4 Mean square errors of the MAVF method and midpoint method at T = 1 for stochastic cyclic Lotka–Volterra system with $y_0 = [1, 2, 1]^\top$. The dashed reference line has slope 1

where c is a real-valued constant and $W(t)$ is a one-dimensional Brownian motion. It can be regarded as a cyclic Lotka–Volterra system of competing 3-species in a chaotic environment [15]. It is verified that (5.3) has two conservative quantities

$$I_1(x, y, z) = x + y + z, \qquad I_2(x, y, z) = x \cdot y \cdot z. \tag{5.4}$$

In this experiment, we set $c = 0.5$. Then, the exact solution of system (5.3) remains on the one-dimensional manifold

$$\mathscr{M} = \left\{(x, y, z) \in \mathbb{R}^3 | I_1(x, y, z) = X_0 + Y_0 + Z_0, \quad I_2(x, y, z) = X_0 \cdot Y_0 \cdot Z_0\right\},$$

which is a closed curve in three-dimensional Euclid space. We compare the MAVF method with midpoint method to demonstrate the strengths of the proposed method.

Figure 4 shows the convergence orders of the MAVF method and midpoint method. The reference solution is obtained by Milstein method with step size $h_{ref} = 2^{-14}$. The mean square errors are computed at the endpoint $T = 1$ by adopting five different

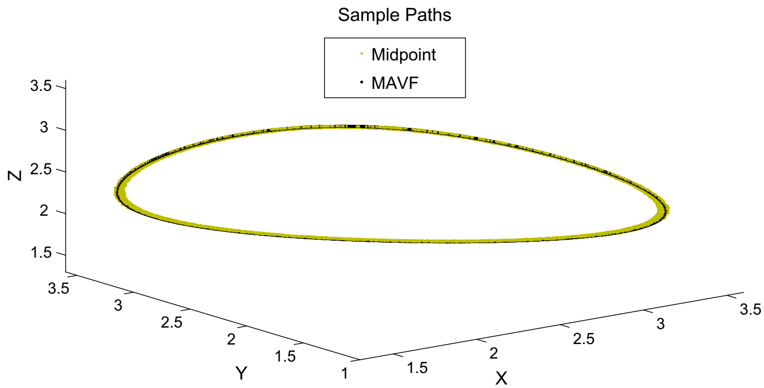

Fig. 5 Numerical sample paths of the MAVF method and midpoint method for stochastic cyclic Lotka–Volterra system with $T = 100$, $h = 0.01$ and $y_0 = [1.8, 3.6, 1.8]^\top$

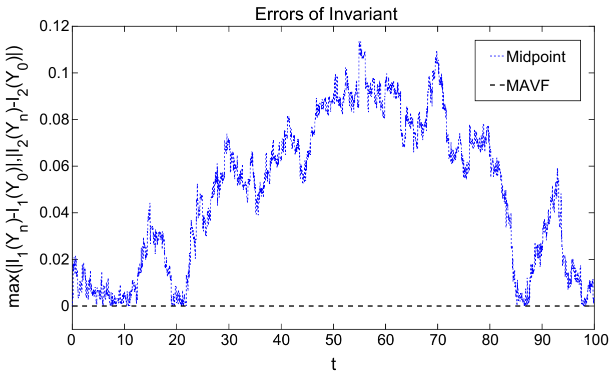

Fig. 6 Errors of invariants of the MAVF method and midpoint method for stochastic cyclic Lotka–Volterra system with $T = 100$, $h = 0.01$ and $y_0 = [1.8, 3.6, 1.8]^\top$

stepsizes $h = 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}$. The expectation is approximated using the average of 1000 independent sample paths. It is observed that the MAVF method for this system is of mean square order 1.

The numerical sample paths of the MAVF method and midpoint method are shown in Fig. 5 with time interval length $T = 100$ and step size $h = 0.01$. We observe that numerical solutions of the MAVF method, along one sample, lie in the manifold $\mathscr{M}$, but those of midpoint method do not. Figure 6 displays the errors of invariants of these two methods. Here the error is denoted by $\max\{|I_1(Y_n) - I_1(Y_0)|, |I_2(Y_n) - I_2(Y_0)|\}$. As is seen in this figure, the MAVF method exactly preserves the two invariants. Although the coefficients of system (5.3) do not satisfy the globally Lipschitz conditions as required in Theorem 3.9, the MAVF method for original system still works well, which indicates that MAVF methods can be applied to more general systems.

Figure 7 presents the evolution of the global mean square errors, and Fig. 8 presents their errors at three different time intervals. The expectation is approximated by using the average of 1000 independent sample paths. As is seen in Fig. 7, the mean square

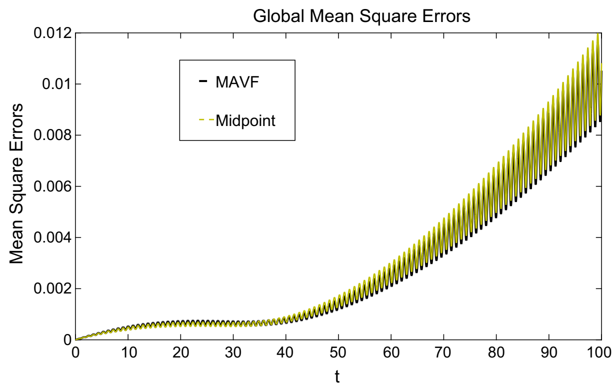

Fig. 7 Global mean square errors of the MAVF method and midpoint method for stochastic cyclic Lotka–Volterra system with $y_0 = [1, 2, 1]^\top$ and $h = 0.01$

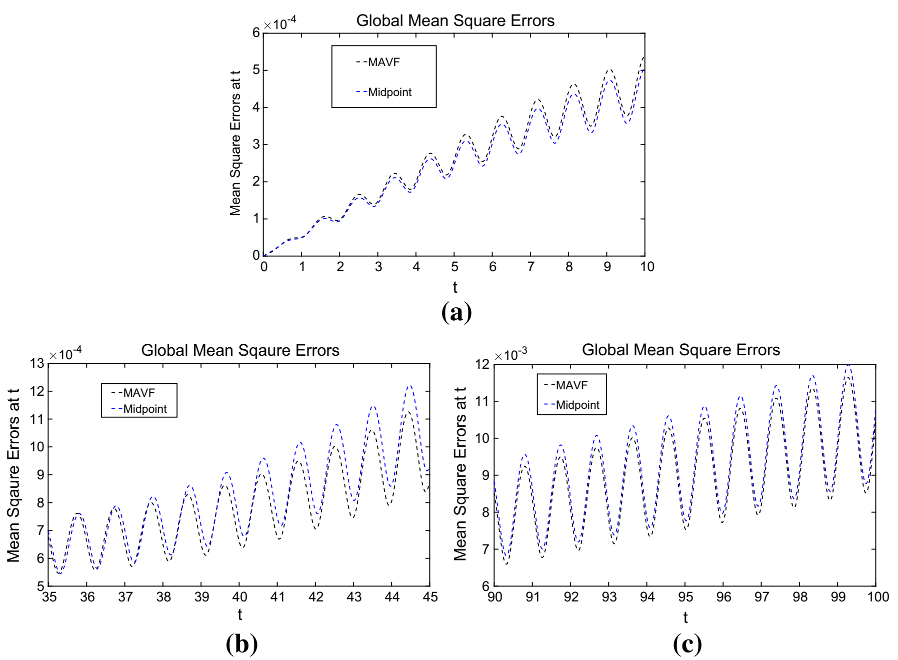

Fig. 8 Global mean square errors, in different time intervals, of the MAVF method and midpoint method for stochastic cyclic Lotka–Volterra system with $y_0 = [1, 2, 1]$ and $h = 0.01$

errors of the MAVF method are very close to those of midpoint method. One observes from Fig. 8a that midpoint method has smaller global errors in the very beginning. Figure 8b, c show that the mean square errors of midpoint method gradually exceed that of the MAVF method.

5.1.3 Example 3: Stochastic Hamiltonian system with multiple invariants

In this experiment, we consider the following Stochastic Hamiltonian system with commutative noises

$$\mathrm{d}\begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \\ Y_4(t) \end{pmatrix} = \begin{pmatrix} Y_3(t) \\ Y_4(t) \\ -Y_1(t) \\ -Y_2(t) \end{pmatrix} \left(\mathrm{d}t + c_1 \circ \mathrm{d}W_1(t) + c_2 \circ \mathrm{d}W_2(t)\right), \tag{5.5}$$

where c_1 and c_2 are constants, and $W_1(t)$ and $W_2(t)$ are two independent Brownian motions. The system (5.5) can be regarded as the extension of Example 3.1 in [11]. One can verify that this system has three invariants

$$\begin{aligned} L_1(y_1, y_2, y_3, y_4) &= y_1 y_4 - y_2 y_3, \\ L_2(y_1, y_2, y_3, y_4) &= \frac{1}{2}\left(y_1^2 - y_2^2 + y_3^2 - y_4^2\right), \\ L_3(y_1, y_2, y_3, y_4) &= y_1 y_2 + y_3 y_4. \end{aligned} \tag{5.6}$$

In this experiment, we take parameters $c_1 = 1$, $c_2 = 0.5$ and the initial value $Y_0 = (-0.5, 0, 0.5, 1)^\top$. We compare the MAVF method with the MilsteinP method for (5.5). It follows from Remark 3.8 that the MAVF method (3.48) becomes midpoint method when applied to this stochastic Hamiltonian system.

We can observe from Fig. 9 that the mean square convergence order is one when applying the MAVF method to (5.5). The mean square errors are computed at the endpoint $T = 1$ by adopting five different stepsizes $h = 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}$. The reference solution is obtained by Milstein method with step size $h_{ref} = 2^{-14}$. The expectation is evaluated by the average of 1000 independent sample paths. This verifies the conclusion about convergence in Theorem 3.9 under the case of commutative noises.

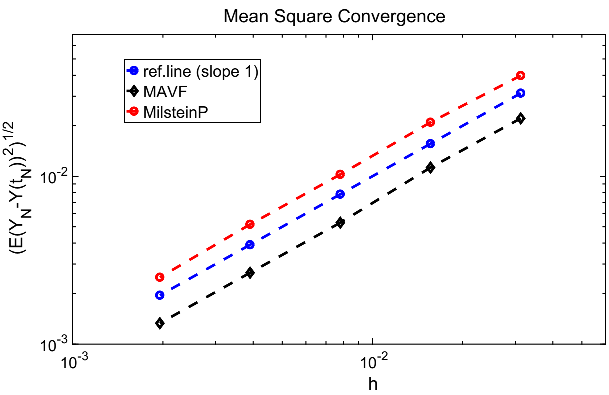

Fig. 9 Mean square errors of the MAVF method and the MilsteinP method at T = 1 for stochastic Hamiltonian system. The dashed reference line has slope 1

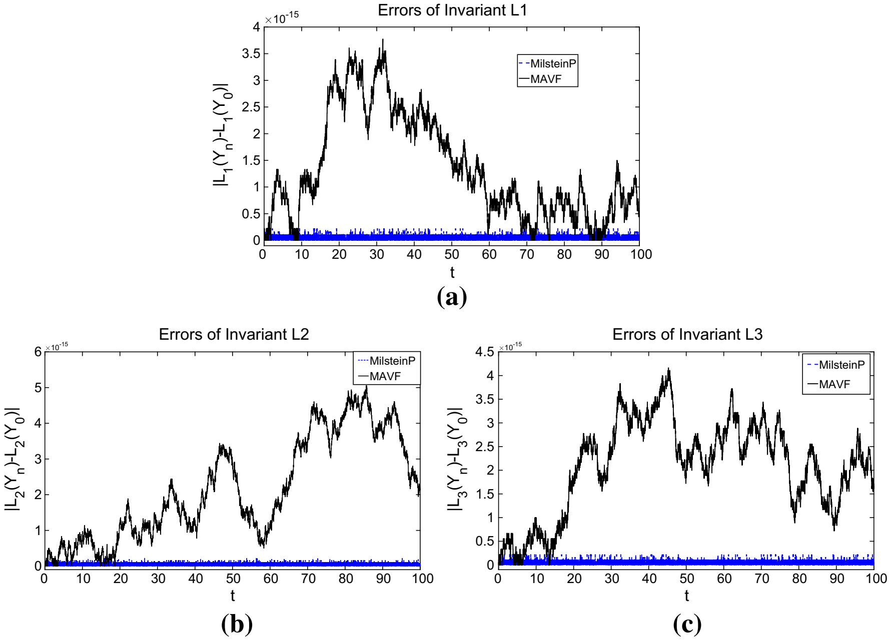

Fig. 10 Errors of invariants of the MAVF method and the MilsteinP method for stochastic Hamiltonian system with $T = 100$ and $h = 0.01$

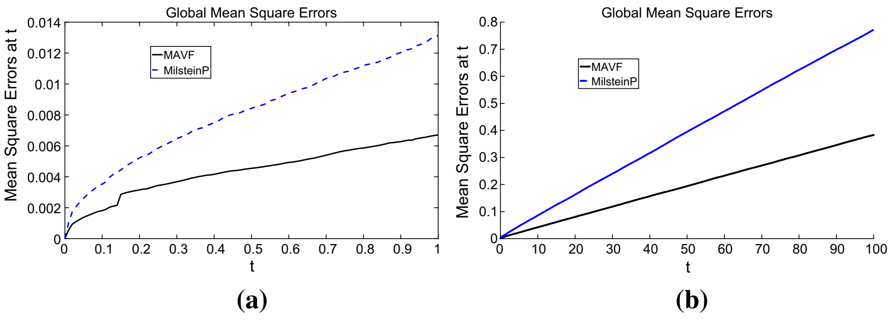

Fig. 11 Global mean square errors of the MAVF method and the MilsteinP method for stochastic Hamiltonian system with $h = 0.01$. The left one is plotted with $T = 1$ and the right one with $T = 100$

Figure 10 displays the errors of invariants L_1, L_2 and L_3, respectively, when applying the MAVF method and the MilsteinP method to (5.5). We observe that both methods preserve these three invariants up to the round-off error, but the MilsteinP method preserves better. In the aspect of the global mean square error, as shown in Fig. 11, the MAVF method behaves better than the MilsteinP method and both of the errors evolve linearly.

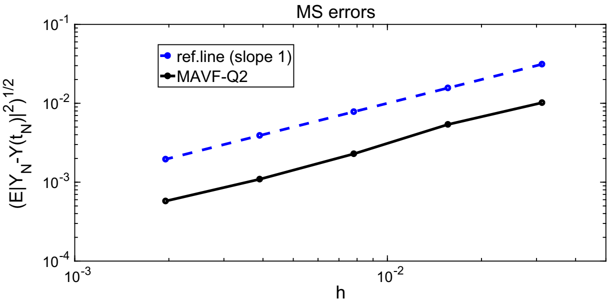

Fig. 12 Mean square errors of MAVF-Q2 method at T = 1 for stochastic pendulum problem. The dashed reference line has slope 1

Table 1 Errors of invariant of MAVF-Q2 method, MAVF-Q4 method, and MAVF-Q6 method at different times along single sample path for stochastic pendulum problem with T = 10000 and h = 0.01

t	100	500	1000	5000	10000	CPU time
MAVF-Q2	0.0009	0.0036	0.0189	0.0193	0.0101	78 s
MAVF-Q4	0.0145E−05	0.2108E−05	0.4764E−05	0.8180E−05	0.9179E−05	61 s
MAVF-Q6	0.0007E−05	0.0095E−05	0.0170E−05	0.0738E−05	0.1478E−05	59 s

5.2 MAVF methods using numerical integration

In this section, we perform numerical experiments to present the effect of numerical integration on MAVF methods. The MAVF methods (4.7) using the quadrature formulas (4.2), (4.3), (4.4) and (4.5) are called MAVF-Q2 method, MAVF-Q3 method, MAVF-Q4 method and MAVF-Q6 method respectively.

Consider the following SDE with commutative noises (see [3])

$$\mathrm{d}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} p \\ \sin(q) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} 0 & -\cos(q) \\ \cos(q) & 0 \end{pmatrix} \times \begin{pmatrix} p \\ \sin(q) \end{pmatrix}(c_1 \circ \mathrm{d}W_1(t) + c_2 \circ \mathrm{d}W_2(t)), \tag{5.7}$$

where c_1, c_2 are constants, and $W_1(t)$, $W_2(t)$ are two independent Brownian motions. This system has $I(p,q) = \frac{1}{2}p^2 - \cos(q)$ as its invariant. We take $c_1 = 1$, $c_2 = 0.5$, and the initial value $(p_0, q_0) = (0.2, 1)$ in this experiment.

Figure 12 shows the convergence order of MAVF-Q2 method. The reference solution is obtained by Milstein method with step size $h_{ref} = 2^{-14}$. The mean square errors are computed at the endpoint $T = 1$ by adopting five different stepsizes $h = 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}$. The expectation is approximated by using the average of 1000 independent sample paths. It is observed that the conservative method for this system is of mean square order 1, which is consistent with the conclusion of Theorem 4.5.

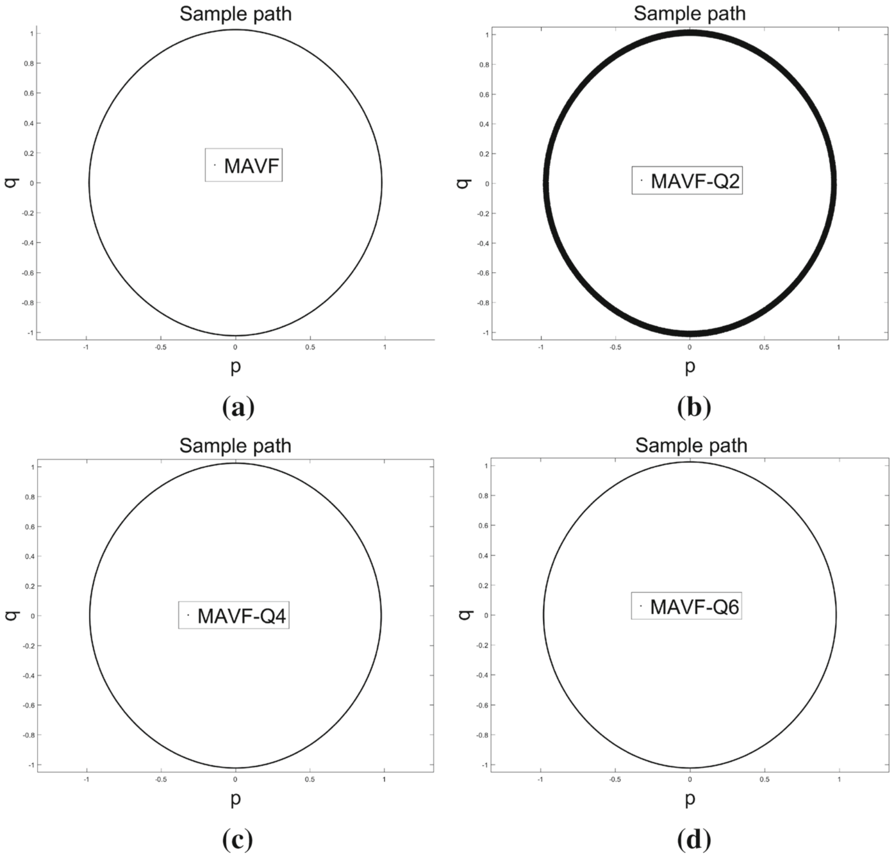

Fig. 13 Numerical sample paths of MAVF method, MAVF-Q2 method, MAVF-Q4 method and MAVF-Q6 method for stochastic pendulum problem with $T = 10000$ and $h = 0.01$

Figure 13 presents the sample paths of the MAVF method, MAVF-Q2 method, MAVF-Q4 method and MAVF-Q6 method. The interval length is $T = 10000$ and the stepsize $h = 0.01$. Table 1 shows the errors of invariant of these three methods and their computation time along a single sample path. As is seen in Fig. 13 and Table 1, as the order of the quadrature formula enlarges, the invariant is preserved better.

Figure 14 shows invariant-preserving orders in mean square sense of the MAVF methods using numerical integration. Here, we use MAVF-Q2 method, MAVF-Q3 method, MAVF-Q4 method to perform numerical experiments. The reference solution is obtained by Milstein method with step size $h_{ref} = 2^{-14}$. The invariant-preserving orders in mean square sense are computed at the endpoint $T = 1$ by adopting five different stepsizes $h = 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}$. The expectation is approximated by using the average of 1000 independent sample paths. It is shown that MAVF-Q2 method and MAVF-Q3 method have mean square order 1 in the preservation of invariants, while MAVF-Q4 method has mean square order 2. These results coincide with those of Theorem 4.9.

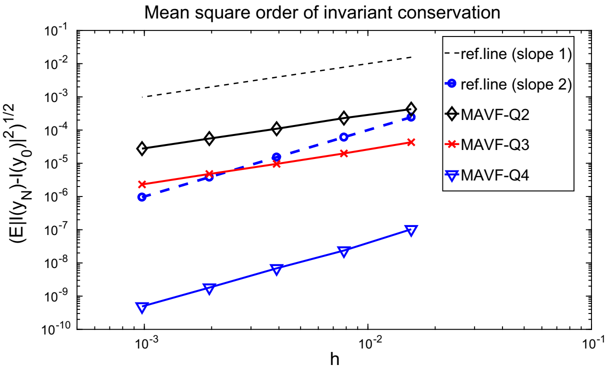

Fig. 14 Invariant-preserving orders in mean square sense of MAVF-Q2 method, MAVF-Q3 method and MAVF-Q4 method at T = 1 for stochastic pendulum problem. The two dashed reference lines have slope 1 and 2 respectively

References

1. Brugnano, L., Iavernaro, F.: Line Integral Methods for Conservative Problems. CRC Press, Boca Raton, FL (2016)
2. Burrage, K., Burrage, P.M., Tian, T.: Numerical methods for strong solutions of stochastic differential equations: an overview. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **460**(2041), 373–402 (2004)
3. Chen, C., Cohen, D., Hong, J.: Conservative methods for stochastic differential equations with a conserved quantity. Int. J. Numer. Anal. Model. **13**(3), 435–456 (2016)
4. Cohen, D., Dujardin, G.: Energy-preserving integrators for stochastic Poisson systems. Commun. Math. Sci. **12**(8), 1523–1539 (2014)
5. Hong, J., Xu, D., Wang, P.: Preservation of quadratic invariants of stochastic differential equations via Runge–Kutta methods. Appl. Numer. Math. **87**, 38–52 (2015)
6. Hong, J., Zhai, S., Zhang, J.: Discrete gradient approach to stochastic differential equations with a conserved quantity. SIAM J. Numer. Anal. **49**(5), 2017–2038 (2011)
7. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)
8. Milstein, G.N., Repin, Y.M., Tretyakov, M.V.: Mean-square symplectic methods for Hamiltonian systems with multiplicative noise. WIAS preprint **670** (2001)
9. Milstein, G.N., Repin, Y.M., Tretyakov, M.V.: Numerical methods for stochastic systems preserving symplectic structure. SIAM J. Numer. Anal. **40**(4), 1583–1604 (2002)
10. Milstein, G.N., Tretyakov, M.V.: Stochastic Numerics for Mathematical Physics. Springer, Berlin (2004)
11. Misawa, T.: Conserved quantities and symmetry for stochastic dynamical systems. Phys. Lett. A **195**(3–4), 185–189 (1994)
12. Misawa, T.: Energy conservative stochastic difference scheme for stochastic Hamilton dynamical systems. Jpn J. Indust. Appl. Math. **17**(1), 119–128 (2000)
13. Quispel, G.R.W., McLaren, D.I.: A new class of energy-preserving numerical integration methods. J. Phys. A **41**(4), 045206 (2008)
14. Sverre, A., Anne, K.: Order conditions for stochastic Runge-Kutta methods preserving quadratic invariants of Stratonovich SDEs. J. Comput. Appl. Math. **316**, 40–46 (2017)
15. Zhou, W., Zhang, L., Hong, J., Song, S.: Projection methods for stochastic differential equations with conserved quantities. BIT **56**(4), 1497–1518 (2016)