Discrete and Continuous Dynamical Systems
Vol. 48, 2026, pp. 447-468
doi:10.3934/dcds.2025148

ASYMPTOTIC ERROR DISTRIBUTIONS OF SYMPLECTIC AND NON-SYMPLECTIC METHODS FOR STOCHASTIC HAMILTONIAN SYSTEM WITH ADDITIVE NOISE

Jialin Hong[1,2], Ge Liang[1,2] and Derui Sheng[*3]

[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

[3]Department of Applied Mathematics,
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

(Communicated by Kening Lu)

Abstract. This paper studies the asymptotic error distributions of several symplectic and non-symplectic methods for stochastic Hamiltonian systems. Focusing on stochastic Hamiltonian systems driven by additive noise, we obtain the asymptotic limit of the normalized error distribution of the θ-method ($\theta \in [0,1]$) that is symplectic if and only if $\theta = \frac{1}{2}$. The upper bound for the second moment of the asymptotic error distribution suggests that the midpoint method may minimize the error constant of the θ-method over a large time horizon T. Furthermore, we take the linear stochastic oscillator as a test equation and investigate exact asymptotic error constants of several symplectic and non-symplectic methods. Our result implies that in the long-time computation, the probability that the error deviates from zero decays exponentially faster for the symplectic methods than for the non-symplectic ones.

1. **Introduction.** Consider the following $2d$-dimensional stochastic Hamiltonian system:

$$\mathrm{d}\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix} DH(X_t^1, X_t^2)\mathrm{d}t + \sum_{k=1}^{m} \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix} DH_k(X_t^1, X_t^2) \circ \mathrm{d}W_t^k \tag{1.1}$$

for $t \in [0,T]$ with the initial value $(X_0^1, X_0^2) \in \mathbb{R}^d \times \mathbb{R}^d$. Here, $H, H_1, \ldots, H_m : \mathbb{R}^{2d} \to \mathbb{R}$ are the Hamiltonians and $W = (W^1, \ldots, W^m)^\top$ is an m-dimensional Brownian motion defined a complete filtered probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}\big)$ with the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ satisfying the usual conditions. One of the most intrinsic properties of (1.1) is that its phase flow preserves the symplectic structure

2020 *Mathematics Subject Classification.* Primary: 60H35; Secondary: 65C30, 65P10.

Key words and phrases. Asymptotic error distributions, central limit type theorem, stochastic Hamiltonian system, symplectic method.

The first author is supported by National key R&D Program of China (No. 2020YFA0713701), and by the National Natural Science Foundation of China (Nos. 11971470, 12031020, 12171047).

*Corresponding author: Derui Sheng.

in phase space, i.e., $\mathrm{d}X_t^1 \wedge \mathrm{d}X_t^2 = \mathrm{d}X_0^1 \wedge \mathrm{d}X_0^2$ for $t \in [0,T]$ and almost surely $\omega \in \Omega$ (see, e.g., [1]). Such a property is called symplecticity, which implies that the sum of the oriented areas of the projections of phase flow onto each coordinate plane (x_i, y_i), $i = 1, 2, \dots, d$, is invariant.

Stochastic Hamiltonian systems have numerous applications in various fields, including chemistry, physics, and celestial mechanics. A basic principle in designing efficient numerical methods for (1.1) is that the numerical method should preserve the symplecticity of the phase flow of (1.1). Such a numerical method is called a symplectic method, originating from the pioneering work of Milstein et al. (see, e.g., [18]). Extensive numerical simulations (see, e.g., [8, 12, 13, 26]) show that when applied to stochastic Hamiltonian systems, symplectic methods exhibit long-time stability compared to non-symplectic methods. The underlying mechanism behind the superiority of symplectic methods for stochastic Hamiltonian systems has attracted considerable attention recently. Inspired by deterministic systems, some studies have explained the long-term stability of symplectic methods through modified equations and backward error analysis techniques (see, e.g., [26, 27]). From a probabilistic standpoint, [6, 7] investigated this issue by proving that symplectic methods can asymptotically preserve the large deviation principles of key physical observables associated with stochastic Hamiltonian systems, while [5] addressed this issue from the perspective of the law of iterated logarithm. Following this research line, we study the asymptotic error distributions of numerical methods to reveal the long-time superiority of symplectic methods over non-symplectic ones for stochastic Hamiltonian systems.

When the normalized error process of a numerical method converges in distribution to a limiting process as the stepsize vanishes, the limit distribution is referred to as the asymptotic error distribution. This kind of weak convergence result can be viewed as a central limit type theorem (see, e.g., [14]), which reduces to the classical central limit theorem when the asymptotic error distribution is Gaussian. Fruitful results have been established for various stochastic systems. For instance, [17] proved that for stochastic differential equations with Lipschitz nonlinearity and multiplicative noise, the normalized error process $\{\sqrt{N}U_t^N\}_{t\in[0,T]}$ of the Euler–Maruyama method converges in distribution to some process $U = \{U_t\}_{t\in[0,T]}$. This result was later extended in [22] to equations with locally Lipschitz nonlinearities. The exact rate of convergence of numerical methods for differential equations driven by fractional Brownian motion was investigated in, e.g., [15, 19, 28, 25]. For more related works, we also refer to [10, 21] for the Euler method of stochastic Volterra equations and to [11] for the accelerated exponential Euler method of stochastic partial differential equations. Beyond numerical accuracy, asymptotic error distributions provide deeper insights into the error structure of numerical methods [2]. Prior work established that the limiting error process for the Euler–Maruyama method forms a gradient in the Dirichlet form sense, enabling error analysis via local Dirichlet forms [3, 4]. The error structure plays a crucial role in error propagation in Monte Carlo simulations, particularly in financial modeling (see, e.g., [3]).

In this work, we focus on the stochastic Hamiltonian system with additive noise (i.e., (1.1) with affine H_k, $k = 1, \dots, m$) and study the asymptotic error distribution of the θ-method ($\theta \in [0,1]$). The θ-method is symplectic for (1.1) if and only if $\theta = \frac{1}{2}$, corresponding to the midpoint method. Since this method exhibits first-order strong convergence for the additive noise case, the normalized error is defined using a normalization constant N, rather than $\sqrt{N}$ that is commonly used for the multiplicative noise case (see, e.g., [17, 22]). The normalized error process can be decomposed into a negligible part that vanishes identically in probability and a dominant part that converges in distribution to the solution of a stochastic differential equation. This suggests that the sharp strong convergence order of the θ-method is 1 for (1.1) with additive noise, regardless of the value of $\theta \in [0, 1]$. We further provide in Theorem 2.7 an upper bound estimate for the second moment of the asymptotic error distribution, which depends on θ and T. This upper bound is minimized when $\theta = \frac{1}{2}$ for large T, which implies that the midpoint method has the smallest asymptotic error constant among all θ-methods.

Inspired by [6], we take the linear stochastic oscillator as a test equation to further investigate the exact asymptotic error constants of its numerical methods. In detail, we derive the error constant K_T for several concrete symplectic and non-symplectic methods for the test equation, and find that the growth of K_T is approximately proportional to T for the symplectic methods and to T^3 for the non-symplectic methods. Consequently, at the scale ϵ, the probability that the error deviates from zero decays exponentially faster for symplectic methods than for non-symplectic methods. This comparison suggests that the considered symplectic methods exhibit superior long-term performance over their non-symplectic counterparts for the test equation, particularly in terms of their asymptotic error distribution. Based on these findings, we plan to extend our investigation to the error structure of symplectic methods for stochastic Hamiltonian systems in future work.

The rest of this paper is organized as follows. In section 2, we establish the asymptotic error distribution of the θ-method for (1.1) with additive noise. By taking the linear stochastic oscillator as a test equation, we further study the exact asymptotic error constants of several symplectic and non-symplectic methods in section 3. Numerical experiments are finally performed in section 4 to verify the theoretical results. Throughout the paper, we denote by D the derivative operator and by C a generic constant which may change from occurrence to occurrence. We will explicity write $C(T, p, \sigma, \cdots)$ to emphasize the dependence of the constant C upon the parameters $T, p, \sigma, \cdots$

2. Asymptotic error distribution of θ-method.

In this section, we investigate the asymptotic error distribution of the normalized error for the θ-method applied to the stochastic Hamiltonian system (1.1) with additive noise. Specifically, we consider the following model

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad t \in [0, T], \tag{2.1}$$

where $\sigma \in \mathbb{R}^{2d\times m}$ is a constant matrix and

$$b := J \cdot DH \quad \text{with} \quad J := \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix}.$$

The following assumption ensures that $b : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ has bounded derivatives up to order 3.

Assumption 2.1. *Assume that $H : \mathbb{R}^{2d} \to \mathbb{R}$ is four times differentiable and that there exists some constant $C > 0$ such that for any $x \in \mathbb{R}^{2d}$,*

$$\|D^2H(x)\| + \|D^3H(x)\| + \|D^4H(x)\| \leq C.$$

Examples of H fulfilling Assumption 2.1 include

$$H(x) = \|x\|^2, \quad H(x) = \frac{1}{2}\|x\|^2 + \sin(x_1), \quad x = (x_1, \ldots, x_{2d}) \in \mathbb{R}^{2d}.$$

By introducing a uniform partition of $[0,T]$ with stepsize $h=\frac{T}{N}$, where $N\in\mathbb{N}_+$, the θ-method applied to (2.1) reads

$$\begin{cases}\widehat{X}^{N,\theta}_{k+1}=\widehat{X}^{N,\theta}_{k}+hb\Big(\theta\widehat{X}^{N,\theta}_{k+1}+(1-\theta)\widehat{X}^{N,\theta}_{k}\Big)+\sigma\Delta W_k,\\ \widehat{X}^{N,\theta}_{0}=X_0,\end{cases}\tag{2.2}$$

where $\Delta W_k:=W_{(k+1)h}-W_{kh},\ k=0,1,\dots,N-1$. We define the continuous version of the θ-method (2.2) as

$$X^{N,\theta}_t=X_0+\int_0^t b(Z^{N,\theta}_s)\mathrm{d}s+\int_0^t\sigma\mathrm{d}W_s,\quad t\in[0,T],$$

where $Z^{N,\theta}_s:=\theta X^{N,\theta}_{R(s)}+(1-\theta)X^{N,\theta}_{L(s)}$, $L(s):=\lfloor s/h\rfloor h$, and $R(s):=\lceil s/h\rceil h$. Here, $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ represent the floor and ceiling functions, respectively. It is clear that $X^{N,\theta}_{kh}=\widehat{X}^{N,\theta}_{k}$ for $k\in\{0,1,\dots,N\}$.

Remark 2.2. In view of the fundamental convergence theorem (see [18, Theorem 1.1.1]), for any $p\in\mathbb{N}_+$,

$$\mathbb{E}\big[\|X_t-X^{N,\theta}_t\|^{2p}\big]\le C(p,T)h^{2p}\quad\forall\, t\in[0,T].\tag{2.3}$$

Remark 2.2 reveals that the error of the θ-method for (2.1) has first-order convergence of accuracy, which motivates us to define the normalized error process

$$U^{N,\theta}_t:=N(X_t-X^{N,\theta}_t),\quad t\in[0,T].\tag{2.4}$$

To obtain the limit distribution of (2.4), we introduce an auxiliary process $\widetilde{U}^{N,\theta}=\{\widetilde{U}^{N,\theta}_t,t\in[0,T]\}$ via

$$\begin{aligned}\widetilde{U}^{N,\theta}_t=&\int_0^t Db(X_s)\widetilde{U}^{N,\theta}_s\mathrm{d}s+N\int_0^t T^{N,\theta}_s Db(X^{N,\theta}_s)b(X^{N,\theta}_s)\mathrm{d}s\\&+N\int_0^t Db(X^{N,\theta}_{L(s)})\sigma S^{N,\theta}_s\mathrm{d}s\\&+N\int_0^t D^2b(X^{N,\theta}_{L(s)})\big(\sigma(W_s-W_{L(s)}),\sigma S^{N,\theta}_s\big)\mathrm{d}s\\&-\frac{N}{2}\int_0^t D^2b\big(X^{N,\theta}_{L(s)}\big)\big(\sigma S^{N,\theta}_s,\sigma S^{N,\theta}_s\big)\mathrm{d}s=:\sum_{i=0}^{4}I^{N,\theta}_i(t),\end{aligned}\tag{2.5}$$

where

$$\begin{aligned}T^{N,\theta}_s&:=(1-\theta)(s-L(s))-\theta(R(s)-s),\\ S^{N,\theta}_s&:=(1-\theta)\big(W_s-W_{L(s)}\big)-\theta\big(W_{R(s)}-W_s\big).\end{aligned}\tag{2.6}$$

The following lemma shows that the normalized error process $U^{N,\theta}_{\cdot}$ has the same limit distribution as the auxiliary process $\widetilde{U}^{N,\theta}_{\cdot}$ if either of them converges in distribution.

Lemma 2.3. *Let Assumption 2.1 hold. Then for any $\theta\in[0,1]$, we have*

$$\lim_{N\to\infty}\mathbb{E}\Big[\sup_{0\le t\le T}\|U^{N,\theta}_t-\widetilde{U}^{N,\theta}_t\|^2\Big]=0.$$

Proof. By the mean value theorem,

$$\begin{aligned}
U_t^{N,\theta} &= N\int_0^t b(X_s) - b(X_s^{N,\theta})\mathrm{d}s + N\int_0^t b(X_s^{N,\theta}) - b(Z_s^{N,\theta})\mathrm{d}s \\
&= N\int_0^t Db(X_s)(X_s - X_s^{N,\theta})\mathrm{d}s \\
&\quad - N\int_0^t\int_0^1 (1-\xi)D^2b_\xi(X_s, X_s^{N,\theta})(X_s^{N,\theta} - X_s, X_s^{N,\theta} - X_s)\mathrm{d}\xi\mathrm{d}s \\
&\quad + N\int_0^t Db(X_s^{N,\theta})(X_s^{N,\theta} - Z_s^{N,\theta})\mathrm{d}s \\
&\quad - N\int_0^t\int_0^1 (1-\xi)D^2b_\xi(X_s^{N,\theta}, Z_s^{N,\theta})(Z_s^{N,\theta} - X_s^{N,\theta}, Z_s^{N,\theta} - X_s^{N,\theta})\mathrm{d}\xi\mathrm{d}s \\
&=: \sum_{i=0}^3 R_i^{N,\theta}(t),
\end{aligned} \tag{2.7}$$

where $D^2b_\xi(Y,\widetilde{Y}) := D^2b((1-\xi)Y + \xi\widetilde{Y})$ for $\xi \in [0,1]$ and any two $2d$-dimensional vectors Y and $\widetilde{Y}$. In view of $X_s^{N,\theta} - Z_s^{N,\theta} = b(Z_s^{N,\theta})T_s^{N,\theta} + \sigma S_s^{N,\theta}$, we can further obtain that

$$\begin{aligned}
&R_2^{N,\theta}(t) \\
&= N\int_0^t T_s^{N,\theta}Db(X_s^{N,\theta})b(X_s^{N,\theta})\mathrm{d}s + N\int_0^t T_s^{N,\theta}Db(X_s^{N,\theta})\big(b(Z_s^{N,\theta}) - b(X_s^{N,\theta})\big)\mathrm{d}s \\
&\quad + N\int_0^t Db(X_{L(s)}^{N,\theta})\sigma S_s^{N,\theta}\mathrm{d}s + N\int_0^t D^2b(X_{L(s)}^{N,\theta})\big(\sigma(W_s - W_{L(s)}), \sigma S_s^{N,\theta}\big)\mathrm{d}s \\
&\quad + N\int_0^t D^2b(X_{L(s)}^{N,\theta})\Big(\int_{L(s)}^s b(Z_s^{N,\theta})\mathrm{d}r, \sigma S_s^{N,\theta}\Big)\mathrm{d}s \\
&\quad + N\int_0^t\int_0^1 \big(D^2b_\xi(X_{L(s)}^{N,\theta}, X_s^{N,\theta}) - D^2b(X_{L(s)}^{N,\theta})\big)\big(X_s^{N,\theta} - X_{L(s)}^{N,\theta}, \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s,
\end{aligned}$$

and

$$\begin{aligned}
R_3^{N,\theta}(t) = &- N\int_0^t (T_s^{N,\theta})^2\int_0^1 (1-\xi)D^2b_\xi(X_s^{N,\theta}, Z_s^{N,\theta})\big(b(Z_s^{N,\theta}), b(Z_s^{N,\theta})\big)\mathrm{d}\xi\mathrm{d}s \\
&- 2N\int_0^t T_s^{N,\theta}\int_0^1 (1-\xi)D^2b_\xi(X_s^{N,\theta}, Z_s^{N,\theta})\big(b(Z_s^{N,\theta}), \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s \\
&+ N\int_0^t\int_0^1 (1-\xi)\big(D^2b\big(X_{L(s)}^{N,\theta}\big) - D^2b_\xi\big(X_s^{N,\theta}, Z_s^{N,\theta}\big)\big)\big(\sigma S_s^{N,\theta}, \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s \\
&- N\int_0^t\int_0^1 (1-\xi)D^2b\big(X_{L(s)}^{N,\theta}\big)\big(\sigma S_s^{N,\theta}, \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s.
\end{aligned}$$

It follows from (2.5) and (2.7) that

$$\begin{aligned}
&U_t^{N,\theta} - \widetilde{U}_t^{N,\theta} \\
&= \int_0^t Db(X_s)\big(U_s^{N,\theta} - \widetilde{U}_s^{N,\theta}\big)\mathrm{d}s + R_1^{N,\theta}(t)
\end{aligned}$$

$$
\begin{aligned}
&+ N\int_0^t T_s^{N,\theta} Db(X_s^{N,\theta})\big(b(Z_s^{N,\theta}) - b(X_s^{N,\theta})\big)\mathrm{d}s \\
&+ N\int_0^t D^2b(X_{L(s)}^{N,\theta})\Big(\int_{L(s)}^s b(Z_s^{N,\theta})\mathrm{d}r, \sigma S_s^{N,\theta}\Big)\mathrm{d}s \\
&+ N\int_0^t\int_0^1 \big(D^2b_\xi(X_{L(s)}^{N,\theta}, X_s^{N,\theta}) - D^2b(X_{L(s)}^{N,\theta})\big)\big(X_s^{N,\theta} - X_{L(s)}^{N,\theta}, \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s \\
&- N\int_0^t (T_s^{N,\theta})^2\int_0^1 (1-\xi)D^2b_\xi(X_s^{N,\theta}, Z_s^{N,\theta})\Big(b(Z_s^{N,\theta}), b(Z_s^{N,\theta})\Big)\mathrm{d}\xi\mathrm{d}s \\
&- 2N\int_0^t T_s^{N,\theta}\int_0^1 (1-\xi)D^2b_\xi(X_s^{N,\theta}, Z_s^{N,\theta})\big(b(Z_s^{N,\theta}), \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s \\
&+ N\int_0^t\int_0^1 (1-\xi)\big(D^2b\big(X_{L(s)}^{N,\theta}\big) - D^2b_\xi\big(X_s^{N,\theta}, Z_s^{N,\theta}\big)\big)\big(\sigma S_s^{N,\theta}, \sigma S_s^{N,\theta}\big)\mathrm{d}\xi\mathrm{d}s \\
=:& \int_0^t Db(X_s)\big(U_s^{N,\theta} - \widetilde{U}_s^{N,\theta}\big)\mathrm{d}s + R_1^{N,\theta}(t) + \sum_{j=1}^6 J_j^{N,\theta}(t). \qquad (2.8)
\end{aligned}
$$

In view of (2.3), we have

$$
\mathbb{E}\Big[\sup_{0\le t\le T}\|R_1^{N,\theta}(t)\|^2\Big] \le CN^2\mathbb{E}\Big[\Big(\int_0^T \|X_s^{N,\theta} - X_s\|^2\mathrm{d}s\Big)^2\Big] \le Ch^2.
$$

Note that each of the terms $\{J_j^{N,\theta}(t), j = 1, 2, \dots, 6\}$ in (2.8) has strong convergence order at least $1/2$, i.e., there exists some consant $C > 0$ independent of h such that for any $j = 1, 2, \dots, 6$,

$$
\mathbb{E}\Big[\sup_{0\le t\le T}\|J_j^{N,\theta}(t)\|^2\Big] \le Ch. \qquad (2.9)
$$

We illustrate (2.9) by taking $j = 6$ as an instance. By the mean value theorem and the boundedness of $D^3b = JD^4H$ under Assumption 2.1, one has

$$
\begin{aligned}
J_6^{N,\theta}(t) = N\int_0^t\int_0^1\int_0^1 &(1-\xi)D^3b\big(\zeta X_{L(s)}^{N,\theta} + (1-\zeta)((1-\xi)X_s^{N,\theta} + \xi Z_s^{N,\theta})\big) \\
&\Big(X_{L(s)}^{N,\theta} - (1-\xi)X_s^{N,\theta} - \xi Z_s^{N,\theta}, \sigma S_s^{N,\theta}, \sigma S_s^{N,\theta}\Big)\mathrm{d}\zeta\mathrm{d}\xi\mathrm{d}s.
\end{aligned}
$$

The bound (2.9) for $j = 6$ then comes from the relation $Z_s^{N,\theta} = \theta X_{R(s)}^{N,\theta} + (1-\theta)X_{L(s)}^{N,\theta}$ and the fact that $\{X_t^{N,\theta}, t \in [0,T]\}$ is $1/2$-Hölder continuous in time in $\mathbb{L}^p(\Omega;\mathbb{R}^{2d})$ for any $p \ge 1$. Hence, it holds that

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le s\le t}\big\|U_s^{N,\theta} - \widetilde{U}_s^{N,\theta}\big\|^2\Big] \\
\le\,& C\mathbb{E}\Big[\sup_{0\le s\le t}\Big\|\int_0^s Db(X_r)\big(\widetilde{U}_r^{N,\theta} - U_r^{N,\theta}\big)\mathrm{d}r\Big\|^2\Big] + Ch \\
\le\,& C\int_0^t \mathbb{E}\Big[\sup_{0\le r\le s}\|\widetilde{U}_r^{N,\theta} - U_r^{N,\theta}\|^2\Big]\mathrm{d}s + Ch,
\end{aligned}
$$

which, together with the Gronwall inequality, finishes the proof. □

In the sequel, we denote by " $\stackrel{d}{\Rightarrow}$ " the convergence in distribution for random variables. Let $\{Y_n\}_{n=1}^\infty$ be a sequence of random variables with values in a metric space (E, d_E), all defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ be an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ and let Y be an E-valued random variable on the extension. We say that Y_n $\mathcal{F}$-stably converges in law to Y if

$$\mathbb{E}[Zf(Y_n)] \to \tilde{\mathbb{E}}[Zf(Y)]$$

for all bounded and continuous $f : E \to \mathbb{R}$ and all bounded random variables Z on $(\Omega, \mathcal{F})$, where $\tilde{\mathbb{E}}$ denotes the expectation with respect to $\tilde{\mathbb{P}}$.

The following lemma is a variant of Slutsky's theorem.

Lemma 2.4. *Let $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ be two sequence of E-valued random variables. Assume that X_n converges $\mathcal{F}$-stably in law to X, and the distance $d_E(Y_n, X_n)$ converges in probability to 0. Then Y_n also converges $\mathcal{F}$-stably in law to X.*

Proof. Let $f : E \to \mathbb{R}$ be a bounded and Lipschitz continuous function with Lipschitz constant K, i.e.,

$$|f(x) - f(y)| \leq \min\{K d_E(x, y), 2\|f\|_{L^\infty(E)}\} \quad \forall\, x, y \in E.$$

Then, by the Vitali convergence theorem, we have that for any bounded random variable Z,

$$\lim_{n\to\infty} |\mathbb{E}[Zf(Y_n)] - \mathbb{E}[Zf(X_n)]| = 0.$$

This result, combined with the fact that X_n converges $\mathcal{F}$-stably in law to X, implies

$$\lim_{n\to\infty} |\mathbb{E}[Zf(Y_n)] - \tilde{\mathbb{E}}[Zf(X)]| = 0. \tag{2.10}$$

Now, let $f : E \to \mathbb{R}$ be a bounded and continuous function. We can find two sequences of bounded and Lipschitz continuous functions, $\{f_k\}_{k=1}^{\infty}$ and $\{g_k\}_{k=1}^{\infty}$, such that f_k increases monotonically to f ($f_k \uparrow f$) and g_k decreases monotonically to f ($g_k \downarrow f$) as $k \to \infty$. Then, from (2.10), for any nonnegative bounded random variable Z,

$$\liminf_{n\to\infty} \mathbb{E}[Zf(Y_n)] \geq \liminf_{n\to\infty} \mathbb{E}[Zf_k(Y_n)] = \tilde{\mathbb{E}}[Zf_k(X)] \quad \forall\, k = 1, 2, \ldots$$

Passing to the limit as $k \to \infty$ in the above inequality yields

$$\liminf_{n\to\infty} \mathbb{E}[Zf(Y_n)] \geq \tilde{\mathbb{E}}[Zf(X)].$$

Similarly, by utilizing the sequence g_k, we can show that for any nonnegative bounded random variable Z,

$$\limsup_{n\to\infty} \mathbb{E}[Zf(Y_n)] \leq \tilde{\mathbb{E}}[Zf(X)].$$

The above two inequalities together establish (2.10) for any nonnegative bounded random variable Z and any bounded, continuous function $f : E \to \mathbb{R}$. For a general bounded random variable Z, we can express it as $Z = Z^+ - Z^-$, where $Z^+ = \max\{Z, 0\}$ and $Z^- = -\min\{Z, 0\}$, and apply the previous result to Z^+ and Z^- separately. The proof is completed. $\square$

To identify the limit distribution of the normalized error process $U^{N,\theta}_{\cdot}$ or equivalently that of the auxiliary process $\widetilde{U}^{N,\theta}_{\cdot}$, we next prepare for the convergence of the terms $I_i^{N,\theta}$, $i = 1, 2, 3, 4$, as the discretization parameter N goes to infinity.

Lemma 2.5. *Let Assumption 2.1 hold. Then we have the following results.*

(1) For any $t \in [0, T]$, $I_i^{N,\theta}(t)$ converges to $I_i^{\theta}(t)$ in $\mathbb{L}^2(\Omega; \mathbb{R}^{2d})$ as $N \to \infty$ for $i = 1, 3, 4$. Here, $I_i^{\theta}, i = 1, 3, 4$ are defined as

$$I_1^{\theta}(t) := \frac{(1 - 2\theta)T}{2} \int_0^t Db(X_s)b(X_s)\mathrm{d}s,$$

$$I_3^\theta(t) := \frac{(1-\theta)T}{2}\sum_{k=1}^m \int_0^t D^2 b(X_s)(\sigma_{\cdot,k},\sigma_{\cdot,k})\mathrm{d}s,$$
$$I_4^\theta(t) := -\frac{(1-2\theta+2\theta^2)T}{4}\sum_{k=1}^m \int_0^t D^2 b(X_s)(\sigma_{\cdot,k},\sigma_{\cdot,k})\mathrm{d}s.$$

(2) $I_2^{N,\theta}$ $\mathcal{F}$-stably converges in law to I_2^θ in $\mathbb{C}([0,T];\mathbb{R}^{2d})$ with

$$I_2^\theta(t) := \frac{(1-2\theta)T}{2}\int_0^t Db(X_s)\sigma \mathrm{d}W_s + \frac{\sqrt{3}}{6}T\int_0^t Db(X_s)\sigma \mathrm{d}\widetilde{W}_s, \quad t\in[0,T] \quad (2.11)$$

where $\{\widetilde{W}_t, t\in[0,T]\}$ is an m-dimensional standard Brownian motion independent of $\{W_t, t\in[0,T]\}$.

Proof. We estimate the terms $\{I_i^{N,\theta}(t), i=1,2,3,4\}$ separately.

Estimate of $I_1^{N,\theta}(t)$. Since $NT_s^{N,\theta} = T\big(\frac{Ns}{T} - \lfloor\frac{Ns}{T}\rfloor\big) - \theta T$, we have

$$I_1^{N,\theta}(t) = T\int_0^t b(X_s^{N,\theta})b(X_s^{N,\theta})\Big\{\frac{Ns}{T} - \lfloor\frac{Ns}{T}\rfloor\Big\}\mathrm{d}s - \theta T\int_0^t Db(X_s^{N,\theta})b(X_s^{N,\theta})\mathrm{d}s.$$

By Remark 2.2 and the uniform boundedness of Db and D^2b, for any $t\in[0,T]$,

$$\lim_{N\to\infty}\mathbb{E}\int_0^t \|Db(X_s^{N,\theta})b(X_s^{N,\theta}) - Db(X_s)b(X_s)\|^2\mathrm{d}s = 0.$$

Then applying [11, Proposition 4.2] yields that $\lim_{N\to\infty} I_1^{N,\theta}(t) = I_1^\theta(t)$ in $\mathbb{L}^2(\Omega;\mathbb{R}^{2d})$.

Estimate of $I_2^{N,\theta}(t)$. By (2.6) and the stochastic Fubini theorem, it holds that

$$\begin{aligned}I_2^{N,\theta}(t) &= (1-\theta)N\int_0^t\int_{L(s)}^s Db(X_{L(s)}^{N,\theta})\sigma\mathrm{d}W_r\mathrm{d}s - \theta N\int_0^t\int_s^{R(s)} Db(X_{L(s)}^{N,\theta})\sigma\mathrm{d}W_r\mathrm{d}s\\ &= \widehat{I}_2^{N,\theta}(t) + \widetilde{I}_2^{N,\theta}(t)\end{aligned}$$

for any $t\in[0,T]$, where

$$\begin{aligned}\widehat{I}_2^{N,\theta}(t) &:= (1-\theta)N\int_0^t Db(X_{L(r)}^{N,\theta})\sigma(R(r)-r)\mathrm{d}W_r\\ &\quad - \theta N\int_0^t Db(X_{L(r)}^{N,\theta})\sigma(r-L(r))\mathrm{d}W_r,\\ \widetilde{I}_2^{N,\theta}(t) &:= -(1-\theta)N\int_{L(t)}^t Db(X_{L(r)}^{N,\theta})\sigma(R(t)-t)\mathrm{d}W_r\\ &\quad - \theta N\int_t^{R(t)} Db(X_{L(r)}^{N,\theta})\sigma(t-L(t))\mathrm{d}W_r.\end{aligned}$$

Denoting by $\widehat{I}_2^{N,\theta,k}(t)$ the kth entry of $\widehat{I}_2^{N,\theta}(t)$, we have

$$\begin{aligned}&\widehat{I}_2^{N,\theta,k}(t)\\ &= T\sum_{i=1}^{2d}\sum_{j=1}^m \sigma_{i,j}\int_0^t \partial_i b^k(X_{L(s)}^{N,\theta})\Big\{\lfloor\frac{Ns}{T}\rfloor - \frac{Ns}{T} + 1 - \theta\Big\}\mathrm{d}W_s^j, \quad k=1,\dots,2d.\end{aligned}$$

For $\nu \in \{1, \dots, m\}$ and $\mu \in \{1, \dots, 2d\}$, the cross variation process between $\widehat{I}_2^{N,\theta,\mu}$ and W^ν is

$$\langle \widehat{I}_2^{N,\theta,\mu}, W^\nu \rangle_t = \sum_{i=1}^{2d} \int_0^t \partial_i b^\mu (X_{L(s)}^{N,\theta}) \sigma_{i,\nu} T \Big\{ \lfloor \frac{Ns}{T} \rfloor - \frac{Ns}{T} + 1 - \theta \Big\} \mathrm{d}s,$$

which combined with [11, Proposition 4.2] and Remark 2.2 leads to

$$\lim_{N \to \infty} \langle \widehat{I}_2^{N,\theta,\mu}, W^\nu \rangle_t = \frac{(1-2\theta)T}{2} \int_0^t \sum_{i=1}^{2d} \partial_i b^\mu (X_s) \sigma_{i,\nu} \mathrm{d}s \quad \text{in } \mathbb{L}^2(\Omega; \mathbb{R}). \tag{2.12}$$

Similarly, it holds that for any $\mu, \widetilde{\mu} \in \{1, \dots, 2d\}$,

$$\begin{aligned} &\lim_{N \to \infty} \langle \widehat{I}_2^{N,\theta,\mu}, \widehat{I}_2^{N,\theta,\widetilde{\mu}} \rangle_t \\ &= \frac{\theta^3 + (1-\theta)^3}{3} T^2 \sum_{i_1,i_2=1}^{2d} \sum_{j=1}^{m} \int_0^t \partial_{i_1} b^\mu (X_s) \partial_{i_2} b^{\widetilde{\mu}} (X_s) \sigma_{i_1,j} \sigma_{i_2,j} \mathrm{d}s. \end{aligned} \tag{2.13}$$

Based on (2.12) and (2.13), and by applying [16, Theorem 4-1], we deduce that there is an M-biased $\mathcal{F}$-conditional Gaussian martingale $I_2^\theta = (I_2^{\theta,1}, I_2^{\theta,2}, \dots, I_2^{\theta,2d})$ on some extension of $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,T]}, \mathbb{P})$ such that $\widehat{I}_2^{N,\theta}$ converges $\mathcal{F}$-stably in law to I_2^θ, and satisfies

$$\langle I^{\theta,\mu}, I^{\theta,\widetilde{\mu}} \rangle_t = \frac{\theta^3 + (1-\theta)^3}{3} T^2 \sum_{i_1,i_2=1}^{2d} \sum_{j=1}^{m} \int_0^t \partial_{i_1} b^\mu (X_s) \partial_{i_2} b^{\widetilde{\mu}} (X_s) \sigma_{i_1,j} \sigma_{i_2,j} \mathrm{d}s,$$

$$\langle I_2^{\theta,\mu}, W^\nu \rangle_t = \frac{(1-2\theta)T}{2} \int_0^t \sum_{i=1}^{2d} \partial_i b^\mu (X_s) \sigma_{i,\nu} \mathrm{d}s,$$

where $\mu, \widetilde{\mu} \in \{1, \dots, 2d\}$ and $\nu \in \{1, \dots, m\}$. Furthermore, [16, Proposition 1-4] indicates that $I_2^\theta(t)$ can be expressed in the following form:

$$\frac{(1-2\theta)T}{2} \int_0^t Db(X_s) \sigma \mathrm{d}W_s + \frac{\sqrt{3}}{6} T \int_0^t Db(X_s) \sigma \mathrm{d}\widetilde{W}_s,$$

where $\widetilde{W}$ is an m-dimensional Brownian motion independent of W. This result, coupled with the convergence of $\widetilde{I}_2^{N,\theta}$ to 0 in $\mathbb{L}^2(\Omega; \mathbb{C}([0,T]; \mathbb{R}^{2d}))$, establishes that $I_2^{N,\theta}$ converges $\mathcal{F}$-stably in law to I_2^θ in $\mathbb{C}([0,T]; \mathbb{R}^{2d})$ (see Lemma 2.4).

Estimate of $I_3^{N,\theta}(t)$. By (2.6), we divide $I_3^{N,\theta}(t) = I_{3,1}^{N,\theta}(t) + I_{3,2}^{N,\theta}(t)$, where

$$I_{3,1}^{N,\theta}(t) := (1-\theta) N \int_0^t D^2 b(X_{L(s)}^{N,\theta}) \big(\sigma(W_s - W_{L(s)}), \sigma(W_s - W_{L(s)}) \big) \mathrm{d}s,$$

$$I_{3,2}^{N,\theta}(t) := -\theta N \int_0^t D^2 b(X_{L(s)}^{N,\theta}) \big(\sigma(W_s - W_{L(s)}), \sigma(W_{R(s)} - W_s) \big) \mathrm{d}s.$$

Furthermore, we have

$$\begin{aligned} I_{3,1}^{N,\theta}(t) = (1-\theta) \sum_{i_1,j_1=1}^{2d} \sum_{k=1}^{m} N \int_0^t \partial_{i_1,j_1} b(X_{L(s)}^{N,\theta}) \sigma_{i_1,k} \sigma_{j_1,k} (W_s^k - W_{L(s)}^k)^2 \mathrm{d}s + (1-\theta) \\ \sum_{i_1,j_1=1}^{2d} \sum_{\substack{i_2,j_2=1 \\ i_2 \neq j_2}}^{m} N \int_0^t \partial_{i_1,j_1} b(X_{L(s)}^{N,\theta}) \sigma_{i_1,i_2} \sigma_{j_1,j_2} (W_s^{i_2} - W_{L(s)}^{i_2})(W_s^{j_2} - W_{L(s)}^{j_2}) \mathrm{d}s \end{aligned}$$

$$=: \widetilde{I}_{3,1}^{N,\theta}(t) + \widehat{I}_{3,1}^{N,\theta}(t).$$

By the relation $\mathrm{d}(W_t^i W_t^j) = W_t^i \mathrm{d}W_t^j + W_t^j \mathrm{d}W_t^i + \delta_{i,j}\mathrm{d}t$ for $i, j \in \{1, \dots, m\}$ and the stochastic Fubini theorem, we obtain

$$\begin{aligned}
&\widetilde{I}_{3,1}^{N,\theta}(t) \\
&= 2(1-\theta)N \sum_{i_1,j_1=1}^{2d} \sum_{k=1}^{m} \int_0^t \partial_{i_1,j_1} b(X_{L(r)}^{N,\theta}) \sigma_{i_1,k} \sigma_{j_1,k} (W_r^k - W_{L(r)}^k)(R(r) \wedge t - r) \mathrm{d}W_r^k \\
&\quad + (1-\theta)T \sum_{i_1,j_1=1}^{2d} \sum_{k=1}^{m} \int_0^t \partial_{i_1,j_1} b(X_{L(s)}^{N,\theta}) \sigma_{i_1,k} \sigma_{j_1,k} \Big(\frac{Ns}{T} - \lfloor \frac{Ns}{T} \rfloor \Big) \mathrm{d}s \\
&=: I_{3,1,1}^{N,\theta}(t) + I_{3,1,2}^{N,\theta}(t),
\end{aligned}$$

and

$$\begin{aligned}
&\widehat{I}_{3,1}^{N,\theta}(t) \\
&= (1-\theta)N \sum_{i_1,j_1=1}^{2d} \sum_{\substack{i_2,j_2=1 \\ i_2 \neq j_2}}^{m} \int_0^t \partial_{i_1,j_1} b(X_{L(r)}^{N,\theta}) \sigma_{i_1,i_2} \sigma_{j_1,j_2} (R(r) \wedge t - r)(W_r^{i_2} - W_{L(r)}^{i_2}) \mathrm{d}W_r^{j_2} \\
&\quad + (1-\theta)N \sum_{i_1,j_1=1}^{2d} \sum_{\substack{i_2,j_2=1 \\ i_2 \neq j_2}}^{m} \int_0^t \partial_{i_1,j_1} b(X_{L(r)}^{N,\theta}) \sigma_{i_1,i_2} \sigma_{j_1,j_2} (R(r) \wedge t - r)(W_r^{j_2} - W_{L(r)}^{j_2}) \mathrm{d}W_r^{i_2} \\
&=: I_{3,1,3}^{N,\theta}(t) + I_{3,1,4}^{N,\theta}(t).
\end{aligned}$$

Analogous to the estimate of $I_1^{N,\theta}(t)$, it can be shown that

$$\lim_{N\to\infty} I_{3,1,2}^{N,\theta}(t) = \frac{1-\theta}{2} T \sum_{k=1}^{m} \int_0^t D^2 b(X_s)(\sigma_{\cdot,k}, \sigma_{\cdot,k}) \mathrm{d}s = I_3^\theta(t) \quad \text{in } \mathbb{L}^2(\Omega; \mathbb{R}^{2d}),$$

since $D^3 b = JD^4 H$ is uniformly bounded due to Assumption 2.1. Moreover, $\lim_{N\to\infty} I_{3,1,i}^{N,\theta} = 0$ in $\mathbb{L}^2(\Omega; \mathbb{C}([0,T]; \mathbb{R}^{2d}))$ for $i \in \{1, 3, 4\}$. Hence, for any fixed $t \in [0, T]$,

$$\lim_{N\to\infty} I_{3,1}^{N,\theta}(t) = I_3^\theta(t) \quad \text{in } \mathbb{L}^2(\Omega; \mathbb{R}^{2d}). \tag{2.14}$$

By the stochastic Fubini theorem,

$$\begin{aligned}
I_{3,2}^{N,\theta}(t) = &-\theta N \sum_{i_1,i_2=1}^{2d} \sum_{i_2,j_2=1}^{m} \int_0^t \int_{L(r)}^r \partial_{i_1,j_1} b(X_{L(r)}^{N,\theta}) \sigma_{i_1,i_2} \sigma_{j_1,j_2} \big(W_s^{i_2} - W_{L(r)}^{i_2}\big) \mathrm{d}s \mathrm{d}W_r^{j_2} \\
&- \theta N \sum_{i_1,i_2=1}^{2d} \sum_{i_2,j_2=1}^{m} \int_t^{R(t)} \int_{L(t)}^t \partial_{i_1,j_1} b(X_{L(t)}^{N,\theta}) \sigma_{i_1,i_2} \sigma_{j_1,j_2} \big(W_s^{i_2} - W_{L(t)}^{i_2}\big) \mathrm{d}s \mathrm{d}W_r^{j_2},
\end{aligned}$$

from which we obtain that $\lim_{N\to\infty} I_{3,2}^{N,\theta} = 0$ in $\mathbb{L}^2(\Omega; \mathbb{C}([0,T]; \mathbb{R}^{2d}))$. This along with (2.14) proves the convergence in $\mathbb{L}^2(\Omega; \mathbb{R}^{2d})$ of $I_3^{N,\theta}(t)$ to $I_3^\theta(t)$ for any fixed $t \in [0, T]$.

Estimate of $I_4^{N,\theta}(t)$. For the term $I_4^{N,\theta}(t)$, we have

$$\begin{aligned}
I_4^{N,\theta}(t) = &-\frac{N}{2}(1-\theta)^2 \int_0^t D^2 b(X_{L(s)}^{N,\theta}) \big(\sigma(W_s - W_{L(s)}), \sigma(W_s - W_{L(s)})\big) \mathrm{d}s \\
&+ N(1-\theta)\theta \int_0^t D^2 b(X_{L(s)}^{N,\theta}) \big(\sigma(W_s - W_{L(s)}), \sigma(W_{R(s)} - W_s)\big) \mathrm{d}s
\end{aligned}$$

$$-\frac{N}{2}\theta^2\int_0^t D^2b(X^{N,\theta}_{L(s)})\big(\sigma(W_{R(s)}-W_s),\sigma(W_{R(s)}-W_s)\big)\mathrm{d}s.$$

The first term and third term on the right hand side can be estimated similarly to $I^{N,\theta}_{3,1}(t)$; while the second term on the right hand side converges to 0 (see the estimate of $I^{N,\theta}_{3,2}(t)$). As a result, we derive that

$$\lim_{N\to\infty} I^{N,\theta}_4(t) = -\frac{(1-2\theta+2\theta^2)T}{4}\sum_{k=1}^m\int_0^t D^2b(X_s)(\sigma_{\cdot,k},\sigma_{\cdot,k})\mathrm{d}s \quad \text{in } \mathbb{L}^2(\Omega;\mathbb{R}^{2d}).$$

The proof is completed. □

Combining Lemmas 2.3 and 2.5, we can conclude the asymptotic error distribution of the θ-method (2.2) for (2.1).

Proposition 2.6. *Let Assumption 2.1 hold and $\theta\in[0,1]$. Then for any $t\in[0,T]$, $U^{N,\theta}_t$ converges to U^θ_t in distribution. Here, U^θ_t satisfies the following equation*

$$\begin{aligned} U^\theta_t = &\int_0^t Db(X_s)U^\theta_s\mathrm{d}s + \frac{(1-2\theta)T}{2}\int_0^t Db(X_s)b(X_s)\mathrm{d}s \\ &+\frac{(1-2\theta)T}{2}\int_0^t Db(X_s)\sigma\mathrm{d}W_s + \frac{\sqrt{3}}{6}T\int_0^t Db(X_s)\sigma\mathrm{d}\widetilde{W}_s \\ &+\frac{(1-2\theta^2)T}{4}\sum_{k=1}^m\int_0^t D^2b(X_s)(\sigma_{\cdot,k},\sigma_{\cdot,k})\mathrm{d}s, \end{aligned}$$

where $\{\widetilde{W}_t, t\in[0,T]\}$ is an m-dimensional standard Brownian motion independent of $\{W_t,t\in[0,T]\}$.

Proof. We introduce a process $\mathcal{U}^{N,\theta}=\{\mathcal{U}^{N,\theta}_t, t\in[0,T]\}$ by

$$\mathcal{U}^{N,\theta}_t = \int_0^t Db(X_s)\mathcal{U}^{N,\theta}_s\mathrm{d}s + I^\theta_1(t)+I^\theta_3(t)+I^\theta_4(t)+I^{N,\theta}_2(t). \tag{2.15}$$

Here, $I^\theta_1(t)$, $I^\theta_3(t)$, and $I^\theta_4(t)$ are precisely as defined in Lemma 2.5(1), while $I^{N,\theta}_2(t)$ is given by (2.5). Comparing (2.5) and (2.15), we can infer that

$$\mathbb{E}\left[\|\widetilde{U}^{N,\theta}_t-\mathcal{U}^{N,\theta}_t\|\right] \le C\int_0^t\mathbb{E}\left[\|\widetilde{U}^{N,\theta}_s-\mathcal{U}^{N,\theta}_s\|\right]\mathrm{d}s + \sum_{i=1,3,4}\mathbb{E}\left[\|I^{N,\theta}_i(t)-I^\theta_i(t)\|\right].$$

Applying Gronwall's inequality yields

$$\begin{aligned}\mathbb{E}\left[\|\widetilde{U}^{N,\theta}_t-\mathcal{U}^{N,\theta}_t\|\right] \le &\sum_{i=1,3,4}\mathbb{E}\left[\|I^{N,\theta}_i(t)-I^\theta_i(t)\|\right] \\ &+C\sum_{i=1,3,4}\int_0^t\mathbb{E}\left[\|I^{N,\theta}_i(s)-I^\theta_i(s)\|\right]\mathrm{d}s.\end{aligned}$$

Note that for each fixed $i\in\{1,3,4\}$,

$$\sup_{t\in[0,T]}\mathbb{E}\left[\|I^{N,\theta}_i(s)\|+\|I^\theta_i(s)\|\right]\le C$$

for some constant $C>0$ independent of N. This estimate, coupled with the $\mathbb{L}^2(\Omega;\mathbb{R}^d)$ convergence of $I^{N,\theta}_i(t)$ to $I^\theta_i(t)$ for $i=1,3,4$ (see Lemma 2.5(1)) and the bounded convergence theorem, implies that for any fixed $t\in[0,T]$,

$$\lim_{N\to\infty}\mathbb{E}\left[\|\widetilde{U}^{N,\theta}_t-\mathcal{U}^{N,\theta}_t\|\right]=0.$$

By Slutsky's theorem, to complete the proof, it remains to show that $\mathcal{U}_t^{N,\theta}$ converges in distribution to U_t^θ. Recall from Lemma 2.5(2) that $I_2^{N,\theta}$ converges $\mathcal{F}$-stably in law to I_2^θ in $\mathbb{C}([0,T];\mathbb{R}^{2d})$. Since $X, I_1^\theta, I_3^\theta$, and I_4^θ are $\mathcal{F}$-measurable, this stable convergence further implies that the joint process $(X, I_1^\theta, I_3^\theta, I_4^\theta, I_2^{N,\theta})$ converges in distribution to $(X, I_1^\theta, I_3^\theta, I_4^\theta, I_2^\theta)$ (see [17, Lemma 2.1]). Given that $\mathcal{U}^{N,\theta}$ is a continuous functional of $(X, I_1^\theta, I_3^\theta, I_4^\theta, I_2^{N,\theta})$, it directly follows from the continuous mapping theorem that $\mathcal{U}^{N,\theta}$ converges in distribution to U^θ in $\mathbb{C}([0,T];\mathbb{R}^{2d})$. The proof is completed. □

Proposition 2.6 also holds for general stochastic differential equations with constant diffusion term σ. It suggests that the strong convergence order 1 is sharp for the θ-method applied to nonlinear stochastic differential equations with additive noise.

Theorem 2.7. *Let Assumption 2.1 hold, $\theta \in [0,1]$, and $T \geq 1$. Assume that there exists $R > 0$ such that the Hessian matrix D^2H is uniformly positive definite for all $x \in \mathbb{R}^{2d}$ with $\|x\| > R$. Then there exist some constants C_i, $i = 0,1,2,3$ dependent on H and σ but independent of θ and T such that for any $t \in [0,T]$,*

$$\begin{aligned}\mathbb{E}\left[\|U_t^\theta\|^2\right] &\leq f_\theta(T)\\ &:= e^{C_0 T}\left(C_1(1-2\theta)^2T^4 + C_2(\theta^2-\theta+\frac{1}{3})T^3 + C_3(1-2\theta^2)^2T^2\right).\end{aligned} \tag{2.16}$$

Proof. In this proof, we denote by $\mathrm{K}_1, \mathrm{K}_2, \dots,$ the generic constants that may dependent on H and σ, but independent of θ and T. The process $\widehat{W} := (\theta^2 - \theta + \frac{1}{3})^{-\frac{1}{2}}(\frac{(1-2\theta)}{2}W + \frac{\sqrt{3}}{6}\widetilde{W})$ is a standard Brownian motion with respect to the filtration generated by W and $\widetilde{W}$, and

$$\begin{aligned}\frac{(1-2\theta)T}{2}\int_0^t Db(X_s)\sigma \mathrm{d}W_s + \frac{\sqrt{3}}{6}T\int_0^t Db(X_s)\sigma \mathrm{d}\widetilde{W}_s\\ = \sqrt{\theta^2-\theta+\frac{1}{3}}T\int_0^t Db(X_s)\sigma \mathrm{d}\widehat{W}_s.\end{aligned}$$

Then by the Itô formula and Proposition 2.6, we have

$$\begin{aligned}\frac{1}{2}\mathbb{E}\left[\|U_t^\theta\|^2\right] &= \mathbb{E}\int_0^t (U_s^\theta)^\top Db(X_s)U_s^\theta \mathrm{d}s + \frac{(1-2\theta)T}{2}\mathbb{E}\int_0^t (U_s^\theta)^\top Db(X_s)b(X_s)\mathrm{d}s\\ &+ \frac{(1-2\theta^2)T}{4}\sum_{k=1}^m \mathbb{E}\int_0^t (U_s^\theta)^\top D^2b(X_s)(\sigma_{\cdot,k},\sigma_{\cdot,k})\mathrm{d}s\\ &+ (\theta^2-\theta+\frac{1}{3})T^2\mathbb{E}\int_0^t \mathrm{tr}(Db(X_s)\sigma\sigma^\top Db(X_s)^\top)\mathrm{d}s.\end{aligned}$$

By the Young inequality and Assumption 2.1, for any $t \in [0,T]$,

$$\begin{aligned}\mathbb{E}\left[\|U_t^\theta\|^2\right] \leq \mathrm{K}_0\int_0^t \mathbb{E}\left[\|U_s^\theta\|^2\right]\mathrm{d}s + \mathrm{K}_1(1-2\theta)^2T^2\mathbb{E}\int_0^t \|b(X_s)\|^2\mathrm{d}s\\ + \mathrm{K}_2(1-2\theta^2)^2T^2 + \mathrm{K}_3(\theta^2-\theta+\frac{1}{3})T^3.\end{aligned} \tag{2.17}$$

Utilizing the Itô formula again as well as the anti-symmetry of J, for any $t \in [0, T]$,

$$\mathbb{E}\left[H(X_t)\right] = \frac{1}{2}\mathbb{E}\int_0^t \operatorname{tr}(D^2 H(X_s)\sigma\sigma^\top)\mathrm{d}s \leq \mathrm{K}_4 t.$$

Recall that there exists some constant $R > 0$ such that the Hessian matrix D^2H is uniformly positive definite for all $x \in \mathbb{R}^{2d}$ with $\|x\| > R$, i.e., there exists some constant $\lambda_0 > 0$ such that $x \in \mathbb{R}^{2d}$ with $\|x\| > R$ and $a \in \mathbb{R}^{2d}$,

$$a^\top D^2 H(x) a \geq \lambda_0 \|a\|^2.$$

Moreover, the continuity of H implies that $\min_{x \in \mathbb{R}^{2d}, \|x\| \leq 2R} H(x) =: H_{2R} > -\infty$. Hence for any $x \in \mathbb{R}^{2d}$ with $\|x\| \geq 2R$ and $\hat{x} := \frac{x}{\|x\|}R$,

$$\begin{aligned} H(x) &= H(\hat{x}) + DH(\hat{x}) \cdot (x - \hat{x}) + \int_0^1 (1-\eta)(x-\hat{x})^\top D^2 H(\hat{x} + t(x - \hat{x}))(x - \hat{x})\mathrm{d}\eta \\ &\geq H(\hat{x}) - \lambda_0^{-1}\|DH(\hat{x})\|^2 - \frac{1}{4}\lambda_0\|x - \hat{x}\|^2 + \frac{1}{2}\lambda_0\|x - \hat{x}\|^2 \\ &\geq H_{2R} - \lambda_0^{-1}|H'_R|^2 + \frac{1}{16}\lambda_0\|x\|^2, \end{aligned}$$

where $H'_R := \max\{\|DH(y)\| : y \in \mathbb{R}^{2d}, \|y\| = R\}$ and the last line is due to $\|x - \hat{x}\| = (1 - R/\|x\|)\|x\| \geq \frac{1}{2}\|x\|$. For any $x \in \mathbb{R}^{2d}$ with $\|x\| \leq 2R$,

$$H(x) \geq H_{2R} \geq H_{2R} - \frac{1}{4}\lambda_0 R^2 + \frac{1}{16}\lambda_0\|x\|^2.$$

Denoting $a_1 := \frac{1}{16}\lambda_0 > 0$ and

$$a_2 := \max\{-H_{2R} + \lambda_0^{-1}|H'_R|^2, -H_{2R} + \frac{1}{4}\lambda_0 R^2, 0\} \geq 0,$$

it holds that for any $x \in \mathbb{R}^{2d}$,

$$H(x) \geq a_1\|x\|^2 - a_2.$$

The linear growth of b implies that for any $t \in [0, T]$ and $T \geq 1$,

$$\mathbb{E}\left[\|b(X_t)\|^2\right] \leq \mathrm{K}_5(1 + \mathbb{E}\left[\|X_t\|^2\right]) \leq \mathrm{K}_5\Big(1 + \frac{\mathrm{K}_4 t + a_2}{a_1}\Big) \leq \mathrm{K}_6 T.$$

Plugging the above inequality into (2.17), we obtain (2.16) from the Gronwall inequality, which completes the proof. □

According to (2.16), for any fixed sufficiently large $T > 0$,

$$f_{\frac{1}{2}}(T) = \min_{\theta \in [0,1]} f_\theta(T).$$

This implies that in the long-time computation, the error constant of the midpoint method may be smaller than that for the θ-method with $\theta \neq \frac{1}{2}$.

3. **Test equation.** In this section, by taking the linear stochastic oscillator as the test equation, we study the asymptotic error distribution of several numerical methods. In subsections 3.1 and 3.2, we study the variance of the asymptotic error distributions for several concrete symplectic and non-symplectic methods for the test equation. A comparison between symplectic and non-symplectic methods will be discussed in subsection 3.3. In the sequel, let $R = \mathcal{O}(h^p), p \geq 0$, stand for $|R| \leq Ch^p$ for all sufficiently small $h > 0$, where C is independent of h. Let $K \sim T^p, p \geq 0$, stand for $\lim_{T\to\infty} K/T^p = C$, where C is independent of T. We denote by $\mathcal{N}(\mathbf{m}, \Sigma)$ the normal distribution with mean $\mathbf{m} \in \mathbb{R}$ and variance $\Sigma > 0$.

By introducing $X_t^2 := \dot{X}_t^1$ and $X_t := (X_t^1, X_t^2)^\top$, the linear stochastic oscillator $\ddot{X}_t^1 + X_t^1 = \alpha \dot{W}_t$ ($\alpha > 0$) can be rewritten as

$$\mathrm{d}X_t = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} X_t \mathrm{d}t + \alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrm{d}W_t, \quad t \in [0, T] \tag{3.1}$$

with initial value $X_0 := (X_0^1, X_0^2)^\top$. Here, $W = \{W_t\}_{t\in[0,T]}$ denotes a 1-dimensional standard Brownian motion defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. The exact solution of (3.1) is given by (see e.g., [24])

$$\begin{cases} X_t^1 = X_0^1 \cos t + X_0^2 \sin t + \alpha \displaystyle\int_0^t \sin(t-s)\mathrm{d}W_s, \\ X_t^2 = -X_0^1 \sin t + X_0^2 \cos t + \alpha \displaystyle\int_0^t \cos(t-s)\mathrm{d}W_s. \end{cases} \tag{3.2}$$

The symplectic structure of its phase flow is preserved (see, e.g., [23]), i.e.,

$$\mathrm{d}X_t^1 \wedge \mathrm{d}X_t^2 = \mathrm{d}X_0^1 \wedge \mathrm{d}X_0^2 \quad \forall\, t \geq 0.$$

We consider a general convergent numerical method for (3.1) of the form

$$\begin{pmatrix} \widehat{X}_{k+1,1}^N \\ \widehat{X}_{k+1,2}^N \end{pmatrix} = A \begin{pmatrix} \widehat{X}_{k,1}^N \\ \widehat{X}_{k,2}^N \end{pmatrix} + \alpha b \Delta W_k := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \widehat{X}_{k,1}^N \\ \widehat{X}_{k,2}^N \end{pmatrix} + \alpha \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \Delta W_k, \tag{3.3}$$

with the initial value $(\widehat{X}_{0,1}^N, \widehat{X}_{0,2}^N) = (X_0^1, X_0^2)$, where $\Delta W_k := W_{(k+1)h} - W_{kh}$, for $k = 0, 1, \ldots, N-1$. The numerical method (3.3) has first-order convergence of accuracy for (3.1) if (see [6, Theorem 4.1])

$$|a_{11} - 1| + |a_{22} - 1| + |a_{12} - h| + |a_{21} + h| = \mathcal{O}(h^2), \quad |b_1| + |b_2 - 1| = \mathcal{O}(h). \tag{3.4}$$

A numerical method $\{(\widehat{X}_{k,1}^N, \widehat{X}_{k,2}^N)\}_{k=0}^N$ for (3.1) is called symplectic if

$$\mathrm{d}\widehat{X}_{k+1,1}^N \wedge \mathrm{d}\widehat{X}_{k+1,2}^N = \mathrm{d}\widehat{X}_{k,1}^N \wedge \mathrm{d}\widehat{X}_{k,2}^N \quad \forall\, k \in \{0, 1, \ldots, N-1\}.$$

Recall that (3.3) is a symplectic method if and only if $\det(A) = 1$ (see e.g., [23]). Since $\{X_t^2\}_{t\in[0,T]}$ is the derivative of $\{X_t^1\}_{t\in[0,T]}$ and many physical observations (e.g., the mean position $\frac{1}{T}\int_0^T X_t^1 \mathrm{d}t$ and the mean velocity $\frac{X_T^1}{T}$) of (3.1) depends on $\{X_t^1\}_{t\in[0,T]}$, we mainly consider the error $e_N := \widehat{X}_{N,1}^N - X_T^1$. In terms of the numerical method (3.3), it follows from [23] that

$$\begin{aligned} \widehat{X}_{N,1}^N &= (a_{11}\widehat{\alpha}_{N-1} + \widehat{\beta}_{N-1}) X_0^1 + a_{12}\widehat{\alpha}_{N-1} X_0^2 \\ &\quad + \alpha \sum_{k=0}^{N-1} \left(b_1 \widehat{\alpha}_{N-1-k} + \gamma \widehat{\alpha}_{N-2-k}\right) \Delta W_k, \end{aligned} \tag{3.5}$$

where $\gamma = a_{12} b_2 - a_{22} b_1$, $\widehat{\alpha}_k = (\det(A))^{\frac{k}{2}} \frac{\sin((k+1)\xi)}{\sin(\xi)}$, and $\widehat{\beta}_k = -\left(\det(A)\right)^{\frac{k+1}{2}} \frac{\sin(k\xi)}{\sin(\xi)}$ for any integer k, with $\xi \in (0, \pi)$ satisfying

$$\cos(\xi) = \frac{\mathrm{tr}(A)}{2\sqrt{\det(A)}}, \qquad \sin(\xi) = \frac{\sqrt{4\det(A) - (\mathrm{tr}(A))^2}}{2\sqrt{\det(A)}}. \tag{3.6}$$

Utilizing (3.2) and (3.5), it holds that

$$\begin{aligned} e_N &= [a_{11}\widehat{\alpha}_{N-1} + \widehat{\beta}_{N-1} - \cos(T)] X_0^1 + [a_{12}\widehat{\alpha}_{N-1} - \sin(T)] X_0^2 \\ &\quad + \alpha \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \left(b_1 \widehat{\alpha}_{N-1-j} + \gamma \widehat{\alpha}_{N-2-j} - \sin(T-s)\right) \mathrm{d}W_s. \end{aligned} \tag{3.7}$$

For the linear problem (3.1), the error e_N is a Gaussian random variable whose variance plays a crucial role in determining its distribution. By the Itô isometry,

$$\frac{\mathrm{Var}(e_N)}{\alpha^2} = \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \left| b_1 \widehat{\alpha}_{N-1-j} + \gamma \widehat{\alpha}_{N-2-j} - \sin(T-s) \right|^2 \mathrm{d}s, \tag{3.8}$$

which enables the asymptotic expansion of $\mathrm{Var}(e_N)$ with respect to h for a concrete numerical method. In the following, we discuss several concrete first-order convergent numerical methods for the test equation (3.1). However, this discussion is not intended to cover all types of symplectic and non-symplectic methods. A comparison of higher-order symplectic and non-symplectic methods, as well as the investigation of pseudo-symplectic methods (see, e.g., [20]), is also meaningful and warrants further investigation in the future.

3.1. **Error of symplectic methods.** In this subsection, we focus on the asymptotic error distributions of the errors $\{e_N\}_{N\in\mathbb{N}_+}$ for several symplectic methods, including symplectic β method and general symplectic methods with $\xi = h$.

3.1.1. *Symplectic β method* ($\beta \in [0,1]$). Applying the symplectic β method to (3.1), the coefficients A_β and b_β are given by

$$A_\beta = \frac{1}{1+\beta(1-\beta)h^2} \begin{pmatrix} 1-(1-\beta)^2h^2 & h \\ -h & 1-\beta^2h^2 \end{pmatrix},$$
$$b_\beta = \frac{1}{1+\beta(1-\beta)h^2} \begin{pmatrix} (1-\beta)h \\ 1 \end{pmatrix}.$$

The symplectic β method reduces to the midpoint method when $\beta = \frac{1}{2}$. Note that the entries of A_β and b_β satisfy (3.4). It is clear that

$$\det(A_\beta) = 1, \quad \mathrm{tr}(A_\beta) = \frac{2 - \left(2\beta^2 - 2\beta + 1\right)h^2}{1+\beta\,(1-\beta)\,h^2}, \quad \gamma = \frac{\beta h}{1+\beta(1-\beta)h^2},$$
$$\sin(\xi) = \frac{h\sqrt{4-(1-2\beta)^2h^2}}{2+2\beta(1-\beta)h^2}, \qquad \cos(\xi) = \frac{2-(2\beta^2-2\beta+1)h^2}{2+2\beta(1-\beta)h^2}.$$

By the Taylor expansion $\arcsin(h) = h + \frac{1}{6}h^3 + \frac{3}{40}h^5 + \mathcal{O}(h^7)$, we have $\xi = h + (\frac{\beta^2}{2} - \frac{\beta}{2} + \frac{1}{24})h^3 + (\frac{3\beta^4}{8} - \frac{3\beta^3}{4} + \frac{7\beta^2}{16} - \frac{\beta}{16} + \frac{3}{640})h^5 + \mathcal{O}(h^7)$. Plugging the expressions of ξ, A_β and b_β into (3.8), a lengthy computation leads to

$$\mathrm{Var}(e_N) = (3\beta^2 - 3\beta + 1)\Big(\frac{\alpha^2}{6}T + \frac{\alpha^2}{12}\sin(2T)\Big)h^2 + \mathcal{O}(h^3). \tag{3.9}$$

Hence we derive the central limit theorem of the error of the stochastic β method

$$Ne_N - \mathbb{E}[Ne_N] \overset{d}{\Rightarrow} \mathcal{N}\Big(0, \alpha^2T^2(3\beta^2-3\beta+1)\Big(\frac{T}{6} + \frac{\sin(2T)}{12}\Big)\Big).$$

Remark 3.1. For any fixed $T > 0$, the error constant of the midpoint method ($\beta = \frac{1}{2}$) is minimal among symplectic β methods for the test equation (3.1).

3.1.2. *General symplectic methods with $\xi = h$.* By (3.6), for the symplectic method, the condition $\xi = h$ is equivalent to $\operatorname{tr}(A) = 2\cos(h)$. By assuming further that the coefficients $a_{ij}(\cdot), b_i(\cdot), i, j = 1, 2$, are smooth functions satisfying (3.4), we have

$$\begin{aligned}
a_{11} &= 1 + a_{11}^{(1)}h^2 + a_{11}^{(2)}h^3 + a_{11}^{(3)}h^4 + \mathcal{O}(h^5),\\
a_{22} &= 1 - (1 + a_{11}^{(1)})h^2 - a_{11}^{(2)}h^3 + (\frac{1}{12} - a_{11}^{(3)})h^4 + \mathcal{O}(h^5),\\
a_{12} &= h + a_{12}^{(1)}h^2 + a_{12}^{(2)}h^3 + a_{12}^{(3)}h^4 + \mathcal{O}(h^5),\\
a_{21} &= -h + a_{21}^{(1)}h^2 + a_{21}^{(2)}h^3 + a_{21}^{(3)}h^4 + \mathcal{O}(h^5),\\
b_1 &= b_1^{(1)}h + b_1^{(2)}h^2 + b_1^{(3)}h^3 + b_1^{(4)}h^4 + \mathcal{O}(h^5),\\
b_2 &= 1 + b_2^{(1)}h + b_2^{(2)}h^2 + b_2^{(3)}h^3 + b_2^{(4)}h^4 + \mathcal{O}(h^5).
\end{aligned} \tag{3.10}$$

In view of (3.10), it holds that $\operatorname{Var}(e_N) = K_T^e h^2 + \mathcal{O}(h^3)$, where the error constant K_T^e is given by

$$\begin{aligned}
K_T^e &= \alpha^2 \frac{1 + 3b_1^{(1)}(b_1^{(1)} - 1) + 3(a_{12}^{(1)} + b_2^{(1)})^2}{6} T\\
&\quad + \alpha^2 \frac{(1 - 2b_1^{(1)})(a_{12}^{(1)} + b_2^{(1)})(\cos(2T) - 1)}{2}\\
&\quad + \alpha^2 \frac{1 + 3b_1^{(1)}(b_1^{(1)} - 1) - 3(a_{12}^{(1)} + b_2^{(1)})^2}{12} \sin(2T).
\end{aligned} \tag{3.11}$$

Furthermore, the corresponding central limit theorem of the error is

$$Ne_N - \mathbb{E}[Ne_N] \overset{d}{\Rightarrow} \mathcal{N}\left(0, T^2 K_T^e\right).$$

We present three existing symplectic methods satisfying $\xi = h$ in Table 1.

Symplectic method	A	b	K_T^e
Exponential method	$\begin{pmatrix} \cos(h) & \sin(h) \\ -\sin(h) & \cos(h) \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\frac{\alpha^2}{6}T + \frac{\alpha^2 \sin(2T)}{12}$
Integral method	$\begin{pmatrix} \cos(h) & \sin(h) \\ -\sin(h) & \cos(h) \end{pmatrix}$	$\begin{pmatrix} \sin(h) \\ \cos(h) \end{pmatrix}$	$\frac{\alpha^2}{6}T + \frac{\alpha^2 \sin(2T)}{12}$
Optimal method	$\begin{pmatrix} \cos(h) & \sin(h) \\ -\sin(h) & \cos(h) \end{pmatrix}$	$\frac{1}{h}\begin{pmatrix} 2\sin^2(\frac{h}{2}) \\ \sin(h) \end{pmatrix}$	$\frac{\alpha^2}{24}T + \frac{\alpha^2 \sin(2T)}{48}$

TABLE 1. Symplectic methods with $\xi = h$ for (3.1).

Remark 3.2. When $T \gg 1$, among the symplectic methods with $\xi = h$, the numerical method satisfying

$$b_1^{(1)} = \frac{1}{2} \quad \text{and} \quad a_{12}^{(1)} + b_2^{(1)} = 0 \tag{3.12}$$

has the minimal error constant. Obviously, the optimal method fulfills (3.12). Besides, we construct a symplectic method satisfying (3.12):

$$A = \begin{pmatrix} \cos(h) & \sin(h) \\ -\sin(h) & \cos(h) \end{pmatrix}, \qquad b = \begin{pmatrix} \frac{h}{2} \\ 1 \end{pmatrix}. \tag{3.13}$$

3.2. **Error of non-symplectic methods.** This subsection is devoted to studying the asymptotic error distributions of errors $\{e_N\}_{N\in\mathbb{N}_+}$ for several non-symplectic methods, including the θ-method ($\theta \neq \frac{1}{2}$) and the PC(EM-BEM) method (see e.g., [23]).

3.2.1. *θ-method* ($\theta \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$). For $\theta \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$, the coefficients of the θ-method for (3.1) are given by

$$A^\theta = \frac{1}{1+\theta^2h^2}\begin{pmatrix} 1-(1-\theta)\theta h^2 & h \\ -h & 1-(1-\theta)\theta h^2 \end{pmatrix}, \qquad b^\theta = \frac{1}{1+\theta^2h^2}\begin{pmatrix} \theta h \\ 1 \end{pmatrix},$$

for which (3.4) is satisfied. It can readily be shown that

$$\begin{aligned}
\cos(\xi) &= \frac{1-(1-\theta)\theta h^2}{\sqrt{1+(1-\theta)^2h^2}\sqrt{1+\theta^2h^2}}, \\
\sin(\xi) &= \frac{h}{\sqrt{1+(1-\theta)^2h^2}\sqrt{1+\theta^2h^2}}, \\
\det(A^\theta) &= \frac{1+(1-\theta)^2h^2}{1+\theta^2h^2}, \quad \gamma = \frac{(1-\theta)h}{1+\theta^2h^2}.
\end{aligned}$$

On this basis, we can further formulate ξ in terms of h as follows $\xi = h + (\theta-\theta^2-\frac{1}{3})h^3 + (\theta^4-2\theta^3+2\theta^2-\theta+\frac{1}{5})h^5 + (3\theta^5-\theta^6-5\theta^4+5\theta^3-3\theta^2+\theta-\frac{1}{7})h^7 + \mathcal{O}(h^9)$. Plugging the above relations into (3.8), we can obtain

$$\mathrm{Var}(e_N) = K_T^\theta h^2 + \mathcal{O}(h^3), \tag{3.14}$$

where the error constant K_T^θ is given by

$$\begin{aligned}
K_T^\theta = &\frac{\alpha^2(2\theta-1)^2}{24}T^3 - \frac{\alpha^2\sin(2T)(2\theta-1)^2}{16}T^2 \\
&+ \alpha^2\left(\frac{(1-\theta)^3-5\theta^3}{6} + \frac{(2\theta-1)^2\cos(2T)}{16}\right)T \\
&+ \alpha^2\left(\frac{1}{48} + \frac{(2\theta-1)^2}{32}\right)\sin(2T).
\end{aligned}$$

Thus the error $\{e_N\}_{N\in\mathbb{N}_+}$ of the θ-method with $\theta \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$ satisfies the following central limit theorem

$$Ne_N - \mathbb{E}[Ne_N] \overset{d}{\Rightarrow} \mathcal{N}(0, T^2K_T^\theta).$$

3.2.2. *PC(EM-BEM) method.* The PC(EM-BEM) method is a predictor-corrector method using the Euler–Maruyama method as the predictor and the backward Euler–Maruyama method as the corrector, whose coefficients are given by

$$A = \begin{pmatrix} 1-h^2 & h \\ -h & 1-h^2 \end{pmatrix}, \qquad b = \begin{pmatrix} h \\ 1 \end{pmatrix}.$$

By a straightforward calculation, we have

$$\begin{aligned}
&\det(A) = 1-h^2+h^4, \quad \mathrm{tr}(A) = 2-2h^2, \quad \gamma = h^3, \\
&\sin(\xi) = \frac{h}{\sqrt{1-h^2+h^4}}, \quad \cos(\xi) = \frac{1-h^2}{\sqrt{1-h^2+h^4}},
\end{aligned}$$

which leads to $\xi = h + \frac{2h^3}{3} + \frac{h^5}{5} + \frac{h^7}{7} + \mathcal{O}(h^9)$. Further, we can simplify (3.8) into

$$\operatorname{Var}(e_N) = \alpha^2\Big(\frac{T^3}{24} - \frac{T^2\sin(2T)}{16} + \frac{6\cos^2(T)+5}{48}T + \frac{5\sin(2T)}{96}\Big)h^2 + \mathcal{O}(h^3). \quad (3.15)$$

Then for the PC(EM-BEM) method, $Ne_N - \mathbb{E}[Ne_N]$ converges in distribution to

$$\mathcal{N}\Big(0, \alpha^2T^2\Big(\frac{T^3}{24} - \frac{T^2\sin(2T)}{16} + \frac{6\cos^2(T)+5}{48}T + \frac{5\sin(2T)}{96}\Big)\Big).$$

3.3. **Comparison between symplectic and non-symplectic methods.** Let $e_N^{(s)}$ and $e_N^{(ns)}$ be the errors of the considered symplectic and non-symplectic methods for (3.1), respectively. It has been shown in subsections 3.1 and 3.2 that the error constant $K_T^{(s)} := K_T \sim T$ for symplectic methods and that the error constant $K_T^{(ns)} := K_T \sim T^3$ for non-symplectic methods. Since $K_T^{(s)}/K_T^{(ns)} \to 0$ as $T \to \infty$, one can choose a sufficiently large time $T_0 > 0$ so that $K_T^{(s)} < K_T^{(ns)}$ for all $T \geq T_0$.

Fix $\epsilon > 0$ and $T \geq T_0$. It follows from $K_T^{(s)} < K_T^{(ns)}$ that

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}\big(|Ne_N^{(s)} - \mathbb{E}[Ne_N^{(s)}]| > \epsilon\big) = \mathcal{N}(0, T^2K_T^{(s)})(\{|x| > \epsilon\})\\
&= \mathcal{N}(0, T^2K_T^{(ns)})\big(\{|x| > \epsilon\sqrt{K_T^{(ns)}/K_T^{(s)}}\}\big) < \mathcal{N}(0, T^2K_T^{(ns)})\big(\{|x| > \epsilon\}\big)\\
&= \lim_{N\to\infty} \mathbb{P}\big(|Ne_N^{(ns)} - \mathbb{E}[Ne_N^{(ns)}]| > \epsilon\big).
\end{aligned}$$

Therefore there exists $N_0 \in \mathbb{N}_+$ such that for any $N \geq N_0$,

$$\mathbb{P}(|Ne_N^{(s)} - \mathbb{E}[Ne_N^{(s)}]| > \epsilon) < \mathbb{P}(|Ne_N^{(ns)} - \mathbb{E}[Ne_N^{(ns)}]| > \epsilon). \quad (3.16)$$

The above inequality (3.16) compares the error's deviation of symplectic and non-symplectic methods for the test equation (3.1). For the test equation and from the perspective of their asymptotic error distributions, the considered symplectic methods demonstrate superior performance compared to their non-symplectic counterparts, although they may have the same mean square convergence order.

We can draw a further comparison between symplectic and non-symplectic methods of the test equation, via the theory of the large deviation principle. In fact, by the Gärtner–Ellis theorem (see, e.g., [9, Theorem 2.3.6]), and the fact that $e_N^{(s)}$ and $e_N^{(ns)}$ are Gaussian random variables, a direct computation gives

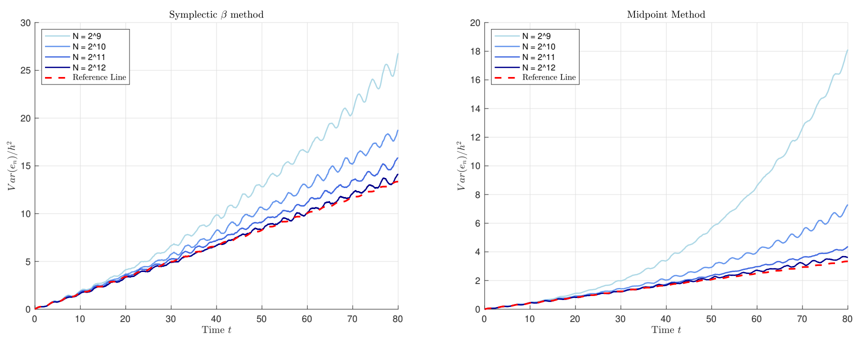
$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N^2}\log\Big(\frac{\mathbb{P}(|e_N^{(s)}| > \epsilon)}{\mathbb{P}(|e_N^{(ns)}| > \epsilon)}\Big)\\
&= \lim_{N\to\infty} \frac{1}{N^2}\log\mathbb{P}(|e_N^{(s)}| > \epsilon) - \lim_{N\to\infty}\frac{1}{N^2}\log\mathbb{P}(|e_N^{(ns)}| > \epsilon) = -\mathcal{R}_\epsilon(T),
\end{aligned}$$

where $\mathcal{R}_\epsilon(T) := \frac{\epsilon^2(K_T^{(ns)} - K_T^{(s)})}{2T^2K_T^{(ns)}K_T^{(s)}} > 0$. Hence there exists some $N_1 \in \mathbb{N}_+$ such that

$$\frac{\mathbb{P}(|e_N^{(s)}| > \epsilon)}{\mathbb{P}(|e_N^{(ns)}| > \epsilon)} \leq \exp\Big(-\frac{1}{2}N^2\mathcal{R}_\epsilon(T)\Big) \quad \forall\ N > N_1. \quad (3.17)$$

The relation (3.17) reveals that at the scale ϵ, the probability of the error's deviation from zero decays exponentially faster for the symplectic method than for the non-symplectic method.

4. **Numerical experiments.** In this section, we present numerical experiments to demonstrate that the normalized errors associated with symplectic methods are smaller than those of non-symplectic methods.



(a) Sympletic β method ($\theta = 0$) (b) Sympletic β method ($\theta = 1/2$)

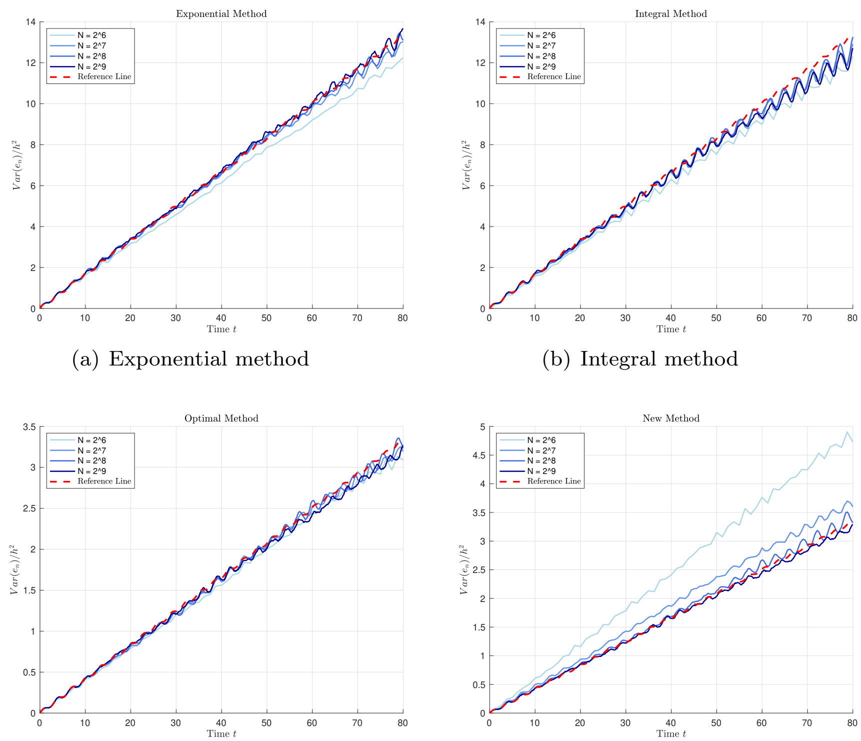
FIGURE 1. The variance $\mathrm{Var}(e_n)/h^2$ of normalized error against time t_n of sympletic β method for the test equation (3.1).

4.1. **Linear stochastic oscillator.** In Figs. 1-3, we plot the variance of the normalized error e_n/h against time $t_n = nh$ based on 2000 sample paths for several numerical methods, applied to the test equation (3.1) with $T = 80$, the initial value $(1, 0)$, and $\alpha = 1$. The corresponding exact solution is computed via numerical integration of (3.2) with partition parameter $N_{\text{ref}} = 2^{17}$. For the symplectic β method (resp. symplectic methods with $\xi = h$), we set the step-size $h = T/N$ with discretization parameters $N \in \{2^9, 2^{10}, 2^{11}, 2^{12}\}$ (resp. $N \in \{2^6, 2^7, 2^8, 2^9\}$). As shown in Fig. 1 (resp. Fig. 2), $\mathrm{Var}(e_n)/h^2$ for these symplectic methods grows nearly linearly with respect to time t_n, in agreement with the reference line $g(t) = 3\beta^2 - 3\beta + 1)(\frac{\alpha^2}{6}t + \frac{\alpha^2}{12}\sin(2t))$ in (3.9) (resp. $g(t) = K_t^e$ in Table 1). For the θ-method and the PC (EM–BEM) method, we set the step-size $h = T/N$ with $N \in \{2^{12}, 2^{13}, 2^{14}, 2^{15}\}$. The reference lines in Fig. 3(a) and Fig. 3(b) are given by the function $g(t) = K_t^\theta$ with $\theta = 1/4$ (see (3.14)) and $g(t) = \frac{t^3}{24} - \frac{t^2\sin(2t)}{16} + \frac{6\cos^2(t)+5}{48}t + \frac{5\sin(2t)}{96}$ (see (3.15)), respectively. It can be observed from Figs. 1-3 that the variance of the normalized error approaches the corresponding reference line as h decreases. Moreover, a comparison between these figures shows that $\mathrm{Var}(e_n)/h^2 \sim t_n^3$ for non-symplectic methods, whereas $\mathrm{Var}(e_n)/h^2 \sim t_n$ for symplectic methods. This demonstrates that symplectic methods exhibit smaller errors compared to non-symplectic ones.

4.2. **Pendulum problem.** The motion of a simple pendulum with unit length and mass can be described by $\ddot{\vartheta}(t) = -g\sin(\vartheta(t)) + \dot{W}(t)$, where ϑ is the angle, $g = 9.8$ is the gravitational acceleration and W is a 1-dimensional standard Brownian motion. By introducing $X_t^1 = \vartheta(t)$ and $X_t^2 = \dot{\vartheta}(t)$, we have

$$\mathrm{d}\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} g\sin(X_t^1) \\ X_t^2 \end{pmatrix} \mathrm{d}t + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrm{d}W_t, \quad t \in [0, T]. \tag{4.1}$$

In Fig. 4, we plot the evolution of $\log(\mathbb{E}[|e_n|^2]/h^2)$ over time t_n for the θ-method and symplectic β method (see [18, (3.6) in pp. 225]) for (4.1) with $T = 10$ and



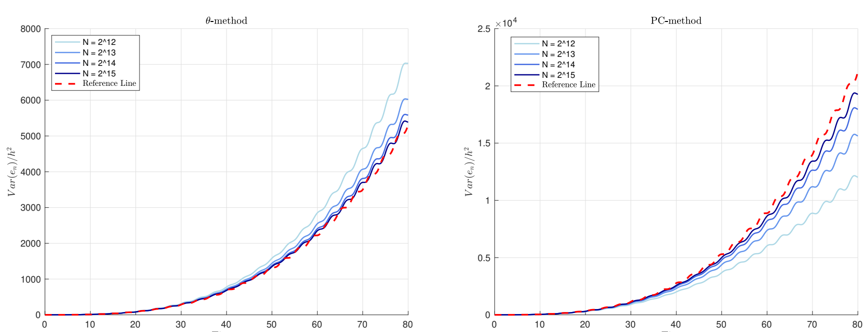
(a) Exponential method

(b) Integral method

(c) Optimal method

(d) New method (3.13)

FIGURE 2. The variance $\mathrm{Var}(e_n)/h^2$ of normalized error against time t_n of symplectic methods with $\xi = h$ for the test equation (3.1).

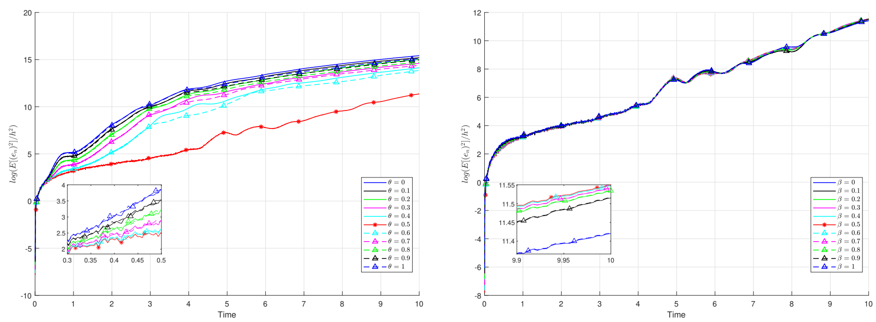


(a) θ-method ($\theta = 1/4$)

(b) PC(EM-BEM) method

FIGURE 3. The variance $\mathrm{Var}(e_n)/h^2$ of normalized error against time t_n of non-symplectic methods for the test equation (3.1).

the initial value $(\frac{\pi}{2}, 0)$, where e_n denotes the error between the exact and numerical solutions at time t_n. The expectation is estimated by averaging over 1000 sample



(a) θ-method (b) Symplectic β method

FIGURE 4. The evolution of the second moment of the normalized error e_n/h over time t_n for the θ-method (left) and the symplectic β method for the pendulum problem (4.1).

paths, and the reference solution is computed using the Euler–Maruyama method with a small stepsize 10^{-5}. The numerical solutions are computed with a stepsize $h = 5/1024$. Among all θ-methods, the midpoint method (i.e., $\theta = \frac{1}{2}$) yields the smallest second moment of the error. Fig. 4(a) also shows that $\log(\mathbb{E}[|e_n|^2]/h^2)$ grows at most linearly with respect to time, indicating that $\mathbb{E}[|e_n|^2]/h^2$ grows at most exponentially in t_n, which is consistent with the upper bound given in Theorem 2.7. Fig. 4(b) suggests that all symplectic β methods exhibit similar second moments of the normalized error for $\beta \in \{0, 0.1, \dots, 1\}$, with the special case $\beta = \frac{1}{2}$ corresponding to the midpoint method. This implies that, for the nonlinear problem, the error constant of the midpoint method ($\beta = \frac{1}{2}$) may not be the smallest among the symplectic β methods, in contrast to the case of the linear test equation (see Remark 3.1).

REFERENCES

[1] J.-M. Bismut, *Mécanique Aléatoire*, volume 866 of *Lecture Notes in Mathematics*, With an English Summary, Springer-Verlag, Berlin-New York, 1981.

[2] N. Bouleau, When and how an error yields a Dirichlet form, *J. Funct. Anal.*, **240** (2006), 445-494.

[3] N. Bouleau, Dirichlet forms methods: An application to the propagation of the error due to the Euler scheme, In *Seminar on Stochastic Analysis, Random Fields and Applications V*, volume 59 of *Progr. Probab.*, Birkhäuser, Basel, 2008, 57-74.

[4] N. Bouleau, *The Mathematics of Errors*, Springer, Cham, French edition, (2021) ©2021.

[5] C. Chen, X. Chen, T. Dang and J. Hong, Superiority of stochastic symplectic methods via the law of iterated logarithm, `arXiv:2404.14842`.

[6] C. Chen, J. Hong, D. Jin and L. Sun, Asymptotically-preserving large deviations principles by stochastic symplectic methods for a linear stochastic oscillator, *SIAM J. Numer. Anal.*, **59** (2021), 32-59.

[7] C. Chen, J. Hong, D. Jin and L. Sun, Large deviations principles for symplectic discretizations of stochastic linear Schrödinger equation, *Potential Anal.*, **59** (2023), 971-1011.

[8] J. Cui, J. Hong, Z. Liu and W. Zhou, Stochastic symplectic and multi-symplectic methods for nonlinear Schrödinger equation with white noise dispersion, *J. Comput. Phys.*, **342** (2017), 267-285.

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*, Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[10] M. Fukasawa and T. Ugai, Limit distributions for the discretization error of stochastic Volterra equations with fractional kernel, *Ann. Appl. Probab.*, **33** (2023), 5071-5110.

[11] J. Hong, D. Jin, X. Wang and G. Yang, Asymptotic error distribution of accelerated exponential Euler method for parabolic SPDEs, `arXiv:2409.13827`.

[12] J. Hong and L. Sun, *Symplectic Integration of Stochastic Hamiltonian Systems*, volume 2314 of *Lecture Notes in Mathematics*, Springer Nature, 2022.

[13] J. Hong and X. Wang, *Invariant Measures for Stochastic Nonlinear Schrödinger Equations. Numerical Approximations and Symplectic Structures*, volume 2251 of *Lecture Notes in Mathematics*, Springer, Singapore, 2019.

[14] W. Hong, S. Li, W. Liu and X. Sun, Central limit type theorem and large deviation principle for multi-scale McKean-Vlasov SDEs, *Probab. Theory Related Fields*, **187** (2023), 133-201.

[15] Y. Hu, Y. Liu and D. Nualart, Rate of convergence and asymptotic error distribution of Euler approximation schemes for fractional diffusions, *Ann. Appl. Probab.*, **26** (2016), 1147-1207.

[16] J. Jacod, On continuous conditional Gaussian martingales and stable convergence in law, In *Séminaire de Probabilités, XXXI*, volume 1655 of *Lecture Notes in Math.*, Springer, Berlin, 1997, 232-246.

[17] J. Jacod and P. Protter, Asymptotic error distributions for the Euler method for stochastic differential equations, *Ann. Probab.*, **26** (1998), 267-307.

[18] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Scientific Computation, Springer-Verlag, Berlin, 2004.

[19] A. Neuenkirch and I. Nourdin, Exact rate of convergence of some approximation schemes associated to SDEs driven by a fractional Brownian motion, *J. Theoret. Probab.*, **20** (2007), 871-899.

[20] X. Niu, J. Cui, J. Hong and Z. Liu, Explicit pseudo-symplectic methods for stochastic Hamiltonian systems, *BIT*, **58** (2018), 163-178.

[21] D. Nualart and B. Saikia, Error distribution of the Euler approximation scheme for stochastic Volterra equations, *J. Theoret. Probab.*, **36** (2023), 1829-1876.

[22] P. Protter, L. Qiu and J. S. Martin, Asymptotic error distribution for the Euler scheme with locally Lipschitz coefficients, *Stochastic Process. Appl.*, **130** (2020), 2296-2311.

[23] M. J. Senosiain and A. Tocino, A review on numerical schemes for solving a linear stochastic oscillator, *BIT*, **55** (2015), 515-529.

[24] A. H. Strømmen Melbø and D. J. Higham, Numerical simulation of a linear stochastic oscillator with additive noise, *Appl. Numer. Math.*, **51** (2004), 89-99.

[25] K. Ueda, Error distribution for one-dimensional stochastic differential equations driven by fractional Brownian motion, *J. Theoret. Probab.*, **38** (2025), Paper No. 20, 61 pp.

[26] L. Wang, J. Hong and L. Sun, Modified equations for weakly convergent stochastic symplectic schemes via their generating functions, *BIT*, **56** (2016), 1131-1162.

[27] Z. Wang, J. Xin and Z. Zhang, Computing effective diffusivity of chaotic and stochastic flows using structure-preserving schemes, *SIAM J. Numer. Anal.*, **56** (2018), 2322-2344.

[28] H. Zhou, Y. Hu and Y. Liu, Backward Euler method for stochastic differential equations with non-Lipschitz coefficients driven by fractional Brownian motion, *BIT*, **63** (2023), Paper No. 40, 37 pp.

Received March 2025; revised July 2025; early access September 2025.