Journal of Computational Mathematics
Vol.xx, No.x, 2026, 1–31.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2509-m2025-0064

HIGH ORDER PROBABILISTIC NUMERICAL METHODS FOR FORWARD BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS*

Qiang Han and Yurong Liu[1)]
School of Mathematical Science, Yangzhou University, Yangzhou 225002, China
Emails: hanqiang@yzu.edu.cn, yrliu@yzu.edu.cn

Abstract

In this paper, we design novel high order probabilistic numerical algorithms for forward backward stochastic differential equations. Moreover, we derive the error estimates and prove the high order convergence rates of the proposed schemes. Because the proposed scheme involves conditional expectations, an estimator based on the multilevel Monte Carlo method is applied to approximate the conditional expectations. Furthermore, we theoretically demonstrate that the computational complexity of our numerical method is proportional to the square of prescribed accuracy. Numerical experiments are given to illustrate the theoretical results.

Mathematics subject classification: 60H35, 60Y20, 65C30, 65C05.
Key words: Forward backward stochastic differential equation, Probabilistic numerical method, Multilevel Monte Carlo, Computational complexity, High order discretization.

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered complete probability space where $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ is the natural filtration of the standard q-dimensional Brownian motion. On the space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, we focus on the probabilistic numerical methods of the solutions to the following forward backward stochastic differential equations (FBSDEs):

$$\begin{cases} X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, & \text{(SDE)} \\ Y_t = \Phi(X_T) + \int_t^T f(Y_s, Z_s)ds - \int_t^T Z_s dW_s, & \text{(BSDE)} \end{cases} \tag{1.1}$$

where $T > 0$ denotes a fixed terminal time, $X_0 \in \mathbb{R}^d$ is a given initial condition of the SDE in (1.1), $\Phi(X_T) \in \mathbb{R}^n$ is a given terminal condition of the BSDE in (1.1), $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is the drift coefficient, $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$ is the diffusion matrix, $f : \mathbb{R}^n \times \mathbb{R}^{n\times q} \to \mathbb{R}^n$ is the real valued generator function and $\Phi : \mathbb{R}^d \to \mathbb{R}^n$ is the real valued function. For convenience, assume that the functions b, σ, f and Φ satisfy the following assumptions:

(i) The functions $b, \sigma \in C_b^1$, C_b^k denote the set of continuous functions with uniformly bounded derivatives up to order k, assume

$$\sup_{0\le t\le T} \{|b(t, 0)| + |\sigma(t, 0)|\} \le L,$$

where the non-negative constant L denotes all the Lipschitz constants.

* Received February 24, 2024 / Revised version received August 4, 2025 / Accepted September 8, 2025 / Published online February 24, 2026 /
1) Corresponding author

(ii) $d = q$ and we assume that σ satisfies

$$\sigma(t,x)\sigma^\top(t,x) \geq \frac{I_d}{L}, \quad \forall\,(t,x) \in [0,T] \times \mathbb{R}^d.$$

(iii) b, σ, f and $\Phi \in C_L$, C_L is the set of uniformly Lipschitz continuous function with respect to the spatial variables, furthermore, assume that

$$|f(0,0)| + |\Phi(0)| \leq L.$$

Note that under the above conditions, it is clear that the FBSDEs (1.1) is well-posed; the resulting integrands obtained by taking the conditional expectation on both sides of the backward component are continuous with respect to time.

The existence and uniqueness of the solutions of the FBSDE (1.1) were deduced in [30]. Papers [31,33] demonstrate that the solution (Y_t, Z_t) of the FBSDE (1.1) can be represented as

$$Y_t = u(t, X_t), \quad Z_t = \nabla u(t, X_t)\sigma(t, X_t), \quad \forall\, t \in [0,T], \tag{1.2}$$

where ∇u is the gradient of $u(t,x)$ with respect to the variable x, $u(t,x) \in C_b^{1,2}([0,T] \times \mathbb{R}^d)$ is the solution of the nonlinear parabolic partial differential equation as below

$$\mathcal{L}^{(0)} u(t,x) + f\big(u(t,x), \nabla u(t,x)\sigma(t,x)\big) = 0 \tag{1.3}$$

with the terminal condition $u(T,x) = \Phi(x)$,

$$\mathcal{L}^{(0)} = \frac{\partial}{\partial t} + \sum_{i=1}^{d} b_i(t,x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j}.$$

After the well-posedness result of nonlinear FBSDEs was rigorously proved by [30, 31, 33], the nonlinear FBSDEs are investigated by more and more researchers because they are widely employed in various areas of social and natural sciences, such as pricing problems, partial differential equations, stochastic optimal control problems, risk measures, image processing problems and so on. However, it is rarely possible to find the analytical solutions of FBSDEs because the vast majority of FBSDEs possess of complex structures and are nonlinear. Thus, an increasing number of numerical algorithms is designed to approximate the solution (Y_t, Z_t) of FBSDEs.

To the best of our knowledge, [9] was the first work which dealt with the numerical methods for FBSDEs by means of the nonlinear Feynman-Kac formulation (see [31, 32]). [3, 17, 34, 42] studied the Euler schemes for FBSDEs and proved that the convergence order of the Euler scheme is 1/2. After that, the Euler schemes also have been tailored to FBSDEs driven by the G-Brownian motion [25], quadratic BSDEs with the reflection [36], reflected BSDEs with two continuous barriers [40] and so on. Furthermore, to achieve a higher order of convergence, the multistep schemes [1, 2, 4, 18, 19, 21, 38, 43] are developed as an extension of the Euler schemes with convergence at an arbitrary order. To avoid solving an algebraic nonlinear equation at each time step, the predictor-corrector schemes [20,22,23] are designed. In this paper, we design a new high order multi-step scheme (see the schemes (2.14)) for the FBSDEs (1.1).

Note that as the parameter $a_0 = -1$, our time-discretization of the Y-part of the BSDE is the same as the time-discretization of the Y-part of the BSDE in [4] and the corrector term of time-discretization of the Y-part of the BSDE in [23], as the parameter $a_0 = -1, k = 2$, our time-discretization of the Y-part of the BSDE is the same as [21], as the parameter $a_0 = -1, a_1 = 1$, $a_2 = a_3 = \cdots = a_k = 0$, our time-discretization of the Y-part of the BSDE degenerates to the Adams schemes; as the parameter $a_0 = -1, a_k = 1, a_1 = a_2 = \cdots = a_{k-1} = 0$, $b_0 = b_1 = \cdots = b_k = 0, k = N$, our time-discretization of the Y-part of the BSDE degenerates to the multi-step forward dynamic programming (MDP) scheme in [19]. That is to say, the time-discretization of the Y-part of the BSDE in [4, 19, 21, 23] is the particular case of our time-discretization of the Y-part of the BSDE. Besides, the time-discretization of the Z-part differs from the existing discretization methods for the Z-part.

The differences between this paper and the reference [22] are that

(i) The discrete-time approximations of BSDE with respect to Y and Z in this paper are designed via the differential method, the undetermined coefficient method and the Itô-Taylor expansion, while [22] adopts the Itô-Taylor expansion and iterative processes. The advantage is that the scheme in this paper avoids instability for $k \leq 4$ with respect to Z, where k denotes the step number in [22]. Moreover, the error estimates are demonstrated under Dahlquist's root condition (see Theorem 3.1). Based on the result of the error estimates, we prove that the convergence order of the proposed new scheme is $k, k \in \mathbb{N}^+$ (see Theorem 4.1).

(ii) [22] devotes to designing an algorithm for conditional expectations via the multilevel Monte Carlo (MLMC) method, while this work aims to analyze the computational complexity of the proposed high order probabilistic numerical algorithm. In other words, we conducted a more in-depth theoretical study on the approximation of conditional mathematical expectations using MLMC methods.

Note that the conditional expectations appear in the numerical schemes no mater what the Euler schemes [3, 17, 42] or the multistep schemes [1, 2, 4, 18–23, 38, 43, 44]. In order to obtain the numerical solutions of FBSDEs, it is necessary to design algorithms for approximating the conditional expectations. Up to now, many algorithms have been designed to calculate the conditional expectations, such as the random walk method [13, 14], the Malliavin calculus and Monte Carlo methods [8], the MLMC method [22,24], the empirical regression method [2,18–20], the cubature method [7], the Fourier transform method [12], the BSDE-COS method [35], the stochastic grid bundling method [6], the wavelet-based SWIFT method [5], the extreme gradient boosting regression method [39], the Fourier cos-cos transform method [29], and the homotopy analysis method [45]. Among these efficient algorithms, we are interested in the MLMC method which is designed in [15,16] and can significantly reduce the computational complexity (see [10]) of Monte Carlo method. In order to better deal with stochastic computing problems, researchers introduce some different MLMC methods, such as the optimal MLMC method [37], the adaptive MLMC [27] method and the multilevel Monte Carlo finite-element method [26] and so on. In this paper, we introduce the MLMC method for approximating conditional expectations and this is an optimization problem with constraints in nature. In addition, we theoretically prove that the computational complexity of our proposed method is proportional to ϵ^{-2}, where ϵ denotes the prescribed accuracy. Compared with Monte Carlo method, in the same time-discretisation numerical scheme, this estimator reduces the computational complexity from $\mathcal{O}(\epsilon^{-2-1/\lambda})$ to $\mathcal{O}(\epsilon^{-2})$, where λ is the convergence rate of the time-discretisation schemes.

The main contributions of this paper are:

(i) We derive a novel high order scheme for FBSDEs. The error estimates and high order property of the proposed scheme are rigorously proved. Note that the scheme is always stable with respect to Y and Z. The numerical schemes [4, 19, 21, 23] are also special cases of our multi-step scheme.

(ii) For the conditional expectations appearing in the proposed scheme, we construct an algorithm via the MLMC method designed by Giles [15,16]. The advantage of this algorithm is to significantly reduce the computational complexity. Specifically, we utilize the MLMC method to approximate the conditional expectations and then analyze the computational complexity of our numerical method (see Theorem 5.1). To the best of our knowledge, this is the first attempt at reducing computational complexity of Monte Carlo simulation for FBSDEs.

The remainder of this paper is organized as follows. We introduce the high order probabilistic numerical method for time-discretization of the FBSDEs (1.1) in Section 2. We derive the error estimates of the proposed numerical schemes in Section 3. Section 4 is devoted to analyzing the convergence order of our numerical schemes. Section 5 presents an algorithm to approximate conditional expectations appearing in our numerical schemes and the corresponding complexity of this algorithm is also demonstrated. We report the numerical results in Section 6. In the end, Section 7 is devoted to the conclusion of this paper.

2. Time-discretization for FBSDEs

In this part, we design the time-discretization scheme for the FBSDEs (1.1). To be specific, we first introduce some notations and two ordinary differential equations (ODEs). Then, the function derivatives and the functions in the derived two ODEs are approximated by some expressions that contain parameters. Finally, we provide the condition of the solvability of equations in the scheme (2.14).

We use π to denote the grid of the time interval $[0, T]$, namely

$$\pi = \{0 = t_0 < t_1 < \cdots < t_N = T\}, \quad t_i = ih, \quad h = \frac{T}{N}, \quad i = 0, 1, \ldots, N, \quad N \in \mathbb{N}^+,$$

and

$$\Delta W_{i,i+j} := W_{t_{i+j}} - W_{t_i}, \quad j \in \mathbb{N}$$

represents the Brownian motion increment.

In what follows, two ODEs are derived (see [43] for detail). Taking the conditional expectation $\mathbb{E}_i[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_{t_i}]$ on both sides of the BSDE in (1.1), we have

$$\mathbb{E}_i[Y_t] = \mathbb{E}_i[\Phi(X_T)] + \int_t^T \mathbb{E}_i[f(Y_s, Z_s)]ds, \quad t \in [t_i, T]. \tag{2.1}$$

From Section 1, the integrand $\mathbb{E}_i[f(Y_s, Z_s)]$ is continuous with respect to s. Thus, taking the derivative with respect to t on both sides of (2.1), we have the ODE

$$\frac{d\mathbb{E}_i[Y_t]}{dt} = -\mathbb{E}_i[f(Y_t, Z_t)], \quad t \in [t_i, T]. \tag{2.2}$$

From the BSDE in (1.1), we obtain

$$Y_{t_i} = Y_t + \int_{t_i}^t f(Y_s, Z_s)ds - \int_{t_i}^t Z_s dW_s, \quad t \in [t_i, T]. \tag{2.3}$$

Multiplying $\Delta W_{t_i,t}^\top = (W_t - W_{t_i})^\top$ on both sides of (2.3), and then taking the conditional expectation on both sides of the resulting equation, we obtain

$$0 = \mathbb{E}_i\big[Y_t \Delta W_{t_i,t}^\top\big] + \int_{t_i}^t \mathbb{E}_i\big[f(Y_s, Z_s)\Delta W_{t_i,s}^\top\big]ds - \int_{t_i}^t \mathbb{E}_i[Z_s]ds, \quad t \in [t_i, T]. \tag{2.4}$$

From Section 1, we know the result that two integrands $\mathbb{E}_i[f(Y_s, Z_s)\Delta W_{t_i,s}^\top]$ and $\mathbb{E}_i[Z_s], s > t_i$ are continuous functions of s. Taking derivative with respect to s, we have the ODE

$$\frac{d\mathbb{E}_i\big[Y_t \Delta W_{t_i,t}^\top\big]}{dt} = -\mathbb{E}_i\big[f(Y_t, Z_t)\Delta W_{t_i,t}^\top\big] + \mathbb{E}_i[Z_t], \quad t \in [t_i, T]. \tag{2.5}$$

Now, we introduce the undetermined coefficient method to approximate the function derivatives (e.g., $d\varphi(t)/dt$) and the functions (e.g., $\varphi(t)$) which play a vital role in designing our high order numerical schemes for the FBSDEs (1.1). Specifically, let $\varphi(t) \in C_b^\infty$. The approximation of $\varphi(t)$ and $d\varphi(t)/dt$ can be described as

$$\begin{aligned} \varphi(t) &\approx \sum_{n=0}^k b_n \varphi(t + nh), \\ \frac{d\varphi(t)}{dt} &\approx \sum_{n=0}^k a_n \frac{\varphi(t + nh)}{h}, \end{aligned} \tag{2.6}$$

where the real numbers $\{a_n\}_{0\le n\le k}$ and $\{b_n\}_{0\le n\le k}$ are independent of h with $k \in \mathbb{N}^+$. From (2.6), we obtain

$$\left.\frac{d\mathbb{E}_i[Y_t]}{dt}\right|_{t=t_i} \approx \sum_{n=0}^k \frac{a_n \mathbb{E}_i[Y_{t_{i+n}}]}{h}, \tag{2.7}$$

$$\left.\frac{d\mathbb{E}_i\big[Y_t \Delta W_{t_i,t}^\top\big]}{dt}\right|_{t=t_i} \approx \sum_{n=0}^k \frac{\alpha_n \mathbb{E}_i\big[Y_{t_{i+n}} \Delta W_{t_i,t_{i+n}}^\top\big]}{h}, \tag{2.8}$$

$$\mathbb{E}_i[f(Y_{t_i}, Z_{t_i})] \approx \sum_{n=0}^k b_n \mathbb{E}_i[f(Y_{t_{i+n}}, Z_{t_{i+n}})], \tag{2.9}$$

$$\mathbb{E}_i\big[f(Y_{t_i}, Z_{t_i})\Delta W_{t_i,t_i}^\top\big] = 0, \tag{2.10}$$

$$\mathbb{E}_i[Z_{t_i}] \approx \sum_{n=0}^k \beta_n \mathbb{E}_i[Z_{t_{i+n}}], \tag{2.11}$$

where the real numbers $\{\alpha_n\}_{0\le n\le k}$ and $\{\beta_n\}_{0\le n\le k}$ are also independent of h. Plugging (2.7) and (2.9) into (2.2), we have

$$\sum_{n=0}^k \frac{a_n \mathbb{E}_i[Y_{t_{i+n}}]}{h} \approx -\sum_{n=0}^k b_n \mathbb{E}_i[f(Y_{t_{i+n}}, Z_{t_{i+n}})]. \tag{2.12}$$

Inserting (2.8), (2.10) and (2.11) into (2.5), we have

$$\sum_{n=0}^k \frac{\alpha_n \mathbb{E}_i\big[Y_{t_{i+n}} \Delta W_{t_i,t_{i+n}}^\top\big]}{h} \approx \sum_{n=0}^k \beta_n \mathbb{E}_i[Z_{t_{i+n}}]. \tag{2.13}$$

Thus, we approximate the analytical solution (Y_t, Z_t) of the FBSDEs (1.1) by the following manner:

1. To initialize the novel multi-step schemes with $k\in\mathbb{N}^+$; k terminal conditions $(Y^\pi_{N-j}, Z^\pi_{N-j})$ are given, where $(Y^\pi_{N-j}, Z^\pi_{N-j})$ is $\mathcal{F}_{t_{N-j}}$-measurable square integrable random variables for $0 \le j \le k-1$.

2. For $0 \le i \le N-k$, the transition from $i+1$ to i is given by

$$\begin{aligned} a_0 Y^\pi_i &= -\sum_{j=1}^{k} a_j \mathbb{E}_i\left[Y^\pi_{i+j}\right] - \sum_{j=0}^{k} b_j h \mathbb{E}_i\left[f^\pi_{i+j}\right], \\ \beta_0 Z^\pi_i &= -\sum_{j=1}^{k} \beta_j \mathbb{E}_i\left[Z^\pi_{i+j}\right] + \sum_{j=1}^{k} \frac{\alpha_j \mathbb{E}_i\left[Y^\pi_{i+j}\Delta W^\top_{i,i+j}\right]}{h}, \end{aligned} \tag{2.14}$$

where Y^π_i and Z^π_i denote the discretization form of Y and Z at t_i, $f^\pi_i = f(Y^\pi_i, Z^\pi_i)$, $\{a_\ell\}_{0\le\ell\le k}$, $\{b_\ell\}_{0\le\ell\le k}$, $\{\alpha_\ell\}_{1\le\ell\le k}$ and $\{\beta_\ell\}_{0\le\ell\le k}$ are real numbers, in particular, $a_0 \ne 0$, $\beta_0 \ne 0$.

We now discuss the solvability of equations in (2.14). If $b_0 = 0$ in (2.14), the scheme is explicit with respect to both Y and Z. Besides, every iteration (Y^π_i, Z^π_i) for $i = N-k, N-k-1, \dots, 0$ can be obtained explicitly. Naturally, the numerical scheme (2.14) has a unique solution. If $b_0 \ne 0$ in (2.14), the scheme is still explicit with respect to Z while implicit with respect to Y. Thus, we have to consider the solvability of the nonlinear equation with respect to Y in (2.14). Moreover, we have to verify that the iterations exist.

Lemma 2.1. *Assume that $b_0 \ne 0$ and $2hb_0L < 1$ hold. Then the time-discretization scheme* (2.14) *has a unique solution.*

Proof. The proof follows the line of proofs used in the analysis of the multi-step schemes for ODEs. Hence, we omit it here (see [11, Theorem 6.1.1] for detail). □

3. Error Estimates

In this part, we measure the error between the numerical solutions (Y^π_i, Z^π_i) obtained by the scheme (2.14) and the analytical solutions (Y_{t_i}, Z_{t_i}) of the FBSDEs (1.1) with the criterion $\mathbb{E}[|Y^\pi_i - Y_{t_i}|^2] + \sum_{\ell=i}^{N-k} \mathbb{E}[h|Z^\pi_\ell - Z_{t_\ell}|^2]$. Before showing the error estimates, we first provide two definitions and an essential lemma.

Definition 3.1. *Let (Y_{t_i}, Z_{t_i}) be the analytical solution of the FBSDEs* (1.1) *at t_i, for $i = N-k, N-k-1, \dots, 0$,*

$$\begin{aligned} a_0 Y_{t_i} &= -\sum_{j=1}^{k} a_j \mathbb{E}_i[Y_{t_{i+j}}] - \sum_{j=0}^{k} b_j h \mathbb{E}_i[f_{t_{i+j}}] + h r^y_i, \\ \beta_0 Z_{t_i} &= -\sum_{j=1}^{k} \beta_j \mathbb{E}_i[Z_{t_{i+j}}] + \sum_{j=1}^{k} \frac{\alpha_j \mathbb{E}_i\left[Y_{t_{i+j}}\Delta W^\top_{i,i+j}\right]}{h} + r^z_i, \end{aligned} \tag{3.1}$$

where $f_{t_i} = f(Y_{t_i}, Z_{t_i})$; if

$$\max_{0\le t_i\le T} \mathbb{E}\left[\left|r^y_i\right|\right] \;\to\; 0 \quad as \quad h \;\to\; 0,$$

we call the numerical schemes with respect to Y *in* (3.1) *approximation the differential equation* (2.2)*; analogously, if*

$$\max_{0\le t_i\le T}\mathbb{E}\big[|r_i^z|\big] \;\to\; 0 \quad as \quad h \;\to\; 0,$$

we call the numerical schemes with respect to Z *in* (3.1) *approximation the differential equation* (2.5).

Definition 3.2. *The characteristic polynomials of the numerical scheme* (2.14) *are given by*

$$P_y(\zeta) = a_0\zeta^k + a_1\zeta^{k-1} + a_2\zeta^{k-2} + \cdots + a_k, \tag{3.2}$$

$$P_z(\zeta) = \beta_0\zeta^k + \beta_1\zeta^{k-1} + \beta_2\zeta^{k-2} + \cdots + \beta_k. \tag{3.3}$$

The numerical scheme (2.14) *is said to fulfil Dahlquist's root condition, if*

(i) The roots of the characteristic polynomials $P_y(\zeta)$ *and* $P_z(\zeta)$ *lie on or within the unit circle;*

(ii) The roots on the unit circle are simple.

Lemma 3.1. *Assume that all the characteristic roots* λ_i *of the matrix* A *lie on or within the unit circle and the dimension of the Jordan block whose modulus of the characteristic roots is equal to* 1 *is* 1*. Then there exists a matrix* $\mathcal{D}$ *such that* $\Lambda = \mathcal{D}^{-1}A\mathcal{D}$ *and the norm of the matrix* Λ *satisfies* $\|\Lambda\|_\infty \le 1$*, where*

$$\|\mathcal{D}\|_\infty = \max_{1\le i\le m}\left(\sum_{j=1}^m |d_{ij}|\right),$$

$\mathcal{D} = (d_{ij})_{m\times m}$ *is a matrix.*

Proof. Let $\eta = 1 - \max_{|\lambda_i|<1}|\lambda_i|$. The eigenvalues of the matrix A/η are λ_i/η. We make the matrix A/η as Jordan form, namely

$$\mathcal{D}^{-1}\frac{A}{\eta}\mathcal{D} = \begin{pmatrix} \frac{\lambda_1}{\eta} & \alpha_{12} & 0 & \cdots \\ 0 & \frac{\lambda_2}{\eta} & \alpha_{23} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix},$$

where $\alpha_{i,i+1} = 0$ or $\alpha_{i,i+1} = 1$. Multiplying η on both sides of the above equation, we have

$$\Lambda = \mathcal{D}^{-1}A\mathcal{D} = \begin{pmatrix} \lambda_1 & \alpha_{12}\eta & 0 & \cdots \\ 0 & \lambda_2 & \alpha_{23}\eta & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix}.$$

Therefore,

$$\|\Lambda\|_\infty = \max_i\{|\lambda_i| + |\alpha_{i,i+1}\eta|\} \le 1.$$

The proof is complete. □

We now state the main result of this section, namely, the error estimates with respect to Y and Z.

Theorem 3.1. *Suppose that the assumptions (i)-(iii) hold. Furthermore, assume that the numerical scheme* (2.14) *fulfils Dahlquist's root condition. Then the following estimates hold:*

$$\begin{aligned}&\mathbb{E}\big[|\Delta Y_i|^2\big]+\sum_{\ell=i}^{N-k}\mathbb{E}\big[h|Z_\ell|^2\big]\\ &\le C\Bigg(\max_{N-k<j\le N}\mathbb{E}\big[|\Delta Y_j|^2\big]+\max_{N-k<j\le N}\mathbb{E}\big[h|\Delta Z_j|^2\big]+\sum_{\ell=i}^{N-k}\mathbb{E}\Big[h^2\big|r_\ell^y\big|^2+h\big|r_\ell^z\big|^2\Big]\Bigg),\end{aligned}$$

where C is a positive constant which may change from line to line and depends $\{a_\ell\}_{0\le\ell\le k}$, $\{b_\ell\}_{0\le\ell\le k}, \{\alpha_\ell\}_{1\le\ell\le k}, \{\beta_\ell\}_{0\le\ell\le k}, T, L, k, q$.

Proof. Let

$$\Delta Y_i = Y_i^\pi - Y_{t_i},\quad \Delta Z_i = Z_i^\pi - Z_{t_i},\quad \Delta f_i = f\big(Y_i^\pi, Z_i^\pi\big) - f(Y_{t_i}, Z_{t_i}).$$

We complete the proof of the theorem in three steps.

Step 1. Estimates for ΔZ_i. Subtracting (3.1) from (2.14) with respect to Z, we obtain

$$\sum_{j=0}^{k}\beta_j\mathbb{E}_i[\Delta Z_{i+j}] = \sum_{j=1}^{k}\frac{\alpha_j\mathbb{E}_i\big[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\big]}{h} - r_i^z. \tag{3.4}$$

We rewrite (3.4) as

$$\Delta Z_i = -\sum_{j=1}^{k}\frac{\beta_j}{\beta_0}\mathbb{E}_i[\Delta Z_{i+j}] + \sum_{j=1}^{k}\frac{\alpha_j\mathbb{E}_i\big[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\big]}{\beta_0 h} - \frac{r_i^z}{\beta_0}. \tag{3.5}$$

We rearrange the k-step recursion to a one-step recursion as follows:

$$\mathbb{E}_i[\mathcal{Z}_i] = A_z\mathbb{E}_i[\mathcal{Z}_{i+1}] + \widetilde{R}_i, \tag{3.6}$$

where

$$\mathcal{Z}_i = \begin{pmatrix}\Delta Z_i\\ \Delta Z_{i+1}\\ \vdots\\ \Delta Z_{i+k-1}\end{pmatrix},$$

$$A_z = \begin{pmatrix}-\frac{\beta_1}{\beta_0} & -\frac{\beta_2}{\beta_0} & \cdots & -\frac{\beta_{k-1}}{\beta_0} & -\frac{\beta_k}{\beta_0}\\ 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \cdots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & 0\end{pmatrix},$$

$$\widetilde{R}_i = \begin{pmatrix}\displaystyle\sum_{j=1}^{k}\frac{\alpha_j\mathbb{E}_i\big[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\big]}{\beta_0 h} - \frac{r_i^z}{\beta_0}\\ 0\\ \vdots\\ 0\end{pmatrix}.$$

The characteristic polynomial of the matrix A_z is

$$\begin{aligned}P_z(\lambda) &= (-1)^{k+1}p_{z,k}(\lambda)\\ &= (-1)^k\left(\lambda^k+\frac{\beta_1}{\beta_0}\lambda^{k-1}+\cdots+\frac{\beta_k}{\beta_0}\right)\\ &= (-1)^k\frac{1}{\beta_0}\big(\beta_0\lambda^k+\beta_1\lambda^{k-1}+\cdots+\beta_k\big),\end{aligned}$$

where

$$\begin{aligned}&p_{z,1}(\lambda) = -\frac{\beta_1}{\beta_0}-\lambda,\\ &p_{z,2}(\lambda) = -\frac{\beta_2}{\beta_0}+\lambda\left(-\frac{\beta_1}{\beta_0}-\lambda\right),\\ &\cdots\cdots\cdots\\ &p_{z,k}(\lambda) = -\frac{\beta_k}{\beta_0}+\lambda\left(-\frac{\beta_{k-1}}{\beta_0}+\lambda\left(-\frac{\beta_{k-2}}{\beta_0}+\cdots+\lambda\left(-\frac{\beta_1}{\beta_0}-\lambda\right)\cdots\right)\right).\end{aligned}$$

Thus, the characteristic equation of the matrix A_z and the characteristic equation of the numerical scheme (2.14) with respect to Z differ by a constant factor. Hence, the characteristic roots of characteristic equation of the matrix A_z satisfy Dahlquist's root condition. From Lemma 3.1, there exists a matrix $\mathcal{D}_z$ such that $\|\mathcal{D}_z^{-1}A_z\mathcal{D}_z\|_\infty \le 1$. Let $\mathcal{Z}_i = \mathcal{D}_z\mathbb{Z}_i$. Multiplying $\mathcal{D}_z^{-1}$ on both sides of (3.6), we have

$$\mathbb{E}_i[\mathbb{Z}_i] = \mathcal{D}_z^{-1}A_z\mathcal{D}_z\mathbb{E}_i[\mathbb{Z}_{i+1}] + \mathcal{D}_z^{-1}\widetilde{R}_i. \tag{3.7}$$

Applying $\|\cdot\|_\infty$ to (3.7), we obtain

$$\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty \le \|\mathbb{E}_i[\mathbb{Z}_{i+1}]\|_\infty + \left\|\mathcal{D}_z^{-1}\right\|_\infty\left|\sum_{j=1}^{k}\frac{\alpha_j\mathbb{E}_i\left[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\right]}{\beta_0 h}-\frac{r_i^z}{\beta_0}\right|. \tag{3.8}$$

Squaring (3.8) and by the inequalities

$$(a+b)^2 \le (1+\gamma)a^2+\left(1+\frac{1}{\gamma}\right)b^2, \quad a,b\in\mathbb{R}, \quad \gamma>0,$$

we have

$$\begin{aligned}\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty^2 &\le (1+\gamma)\|\mathbb{E}_i[\mathbb{Z}_{i+1}]\|_\infty^2+\left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2\left|\sum_{j=1}^{k}\frac{\alpha_j\mathbb{E}_i\left[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\right]}{\beta_0 h}-\frac{r_i^z}{\beta_0}\right|^2\\ &\le (1+\gamma)\|\mathbb{E}_i[\mathbb{Z}_{i+1}]\|_\infty^2+\left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2\frac{2}{\beta_0^2}\left(r_i^z\right)^2\\ &\quad+\left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2\frac{2k\max_{1\le j\le k}\left\{\alpha_j^2\right\}}{\beta_0^2h^2}\sum_{j=1}^{k}\left(\mathbb{E}_i\left[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\right]\right)^2.\end{aligned} \tag{3.9}$$

Since the Brownian increment $\Delta W_{i,i+j}$ is conditionally centered, it follows that

$$\mathbb{E}_i\left[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\right] = \mathbb{E}_i\left[(\Delta Y_{i+j}-\mathbb{E}_i[\Delta Y_{i+j}])\Delta W_{i,i+j}^\top\right].$$

By the Cauchy-Schwarz inequality, we obtain

$$\left|\mathbb{E}_i\left[(\Delta Y_{i+j} - \mathbb{E}_i[\Delta Y_{i+j}])\Delta W_{i,j}^\top\right]\right|^2 \le qjh\Big(\mathbb{E}_i\left[(\Delta Y_{i+j})^2\right] - \left(\mathbb{E}_i[\Delta Y_{i+j}]\right)^2\Big). \tag{3.10}$$

Plugging (3.10) into (3.9), we deduce

$$\begin{aligned}
\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty^2 &\le (1+\gamma)\|\mathbb{E}_i[\mathbb{Z}_{i+1}]\|_\infty^2 + \left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2 \\
&\quad \times \left|\sum_{j=1}^k \frac{\alpha_j \mathbb{E}_i\left[\Delta Y_{i+j}\Delta W_{i,i+j}^\top\right]}{\beta_0 h} - \frac{r_i^z}{\beta_0}\right|^2 \\
&\le (1+\gamma)\|\mathbb{E}_i[\mathbb{Z}_{i+1}]\|_\infty^2 + \left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2 \frac{2}{\beta_0^2}\left(r_i^z\right)^2 \\
&\quad + \left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2 \frac{2k^2 q \max_{1\le j\le k}\left\{\alpha_j^2\right\}}{\beta_0^2 h} \\
&\quad \times \sum_{j=1}^k \Big(\mathbb{E}_i\left[(\Delta Y_{i+j})^2\right] - \left(\mathbb{E}_i[\Delta Y_{i+j}]\right)^2\Big).
\end{aligned} \tag{3.11}$$

Step 2. Estimates for ΔY_i. From the expression with respect to Y in (2.14) and (3.1), one obtains

$$a_0 \Delta Y_i = -\mathbb{E}_i\left[\sum_{j=1}^k a_j \Delta Y_{i+j} + \sum_{j=0}^k b_j h \Delta f_{i+j}\right] - h r_i^y. \tag{3.12}$$

We rearrange the k-step recursion (3.12) to a one-step recursion as follows:

$$\mathbb{E}_i[\mathcal{Y}_i] = \mathbb{E}_i[A_y \mathcal{Y}_{i+1} + hF_i] + R_i, \tag{3.13}$$

where

$$\mathcal{Y}_i = \begin{pmatrix} \Delta Y_i \\ \Delta Y_{i+1} \\ \vdots \\ \Delta Y_{i+k-1} \end{pmatrix}, \qquad A_y = \begin{pmatrix} -\frac{a_1}{a_0} & -\frac{a_2}{a_0} & \cdots & -\frac{a_{k-1}}{a_0} & -\frac{a_k}{a_0} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$F_i = \begin{pmatrix} \sum_{j=0}^k \frac{b_j}{a_0}\Delta f_{i+j} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad R_i = \begin{pmatrix} -\frac{h r_i^y}{a_0} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The characteristic polynomial of the matrix A_y is

$$\begin{aligned}
P_y(\lambda) &= (-1)^{k+1} p_{y,k}(\lambda) \\
&= (-1)^k \left(\lambda^k + \frac{a_1}{a_0}\lambda^{k-1} + \cdots + \frac{a_k}{a_0}\right) \\
&= (-1)^k \frac{1}{a_0}\left(a_0\lambda^k + a_1\lambda^{k-1} + \cdots + a_k\right),
\end{aligned}$$

where

$$
\begin{aligned}
p_{y,1}(\lambda) &= -\frac{a_1}{a_0} - \lambda,\\
p_{y,2}(\lambda) &= -\frac{a_2}{a_0} + \lambda\Big(-\frac{a_1}{a_0} - \lambda\Big),\\
&\cdots\cdots\cdots\\
p_{y,k}(\lambda) &= -\frac{a_k}{a_0} + \lambda\Big(-\frac{a_{k-1}}{a_0} + \lambda\Big(-\frac{a_{k-2}}{a_0} + \cdots + \lambda\Big(-\frac{a_1}{a_0} - \lambda\Big)\cdots\Big)\Big).
\end{aligned}
$$

Thus, the characteristic equation of the matrix A_y and the characteristic equation of the numerical scheme (2.14) with respect to Y differ by a constant factor. Hence, we have that the characteristic roots of characteristic equation of the matrix A_y satisfy Dahlquist's root condition. From Lemma 3.1, we have that there exists a matrix $\mathcal{D}_y$ such that $\|\mathcal{D}_y^{-1}A_y\mathcal{D}_y\|_\infty \le 1$. Let $\mathcal{Y}_i = \mathcal{D}_y\mathbb{Y}_i$. Multiplying $\mathcal{D}_y^{-1}$ on both sides of (3.13), we have

$$
\mathbb{E}_i[\mathbb{Y}_i] = \mathcal{D}_y^{-1}A_y\mathcal{D}_y\mathbb{E}_i[\mathbb{Y}_{i+1}] + h\mathcal{D}_y^{-1}\mathbb{E}_i[F_i] + \mathcal{D}_y^{-1}R_i. \tag{3.14}
$$

Applying $\|\cdot\|_\infty$ to (3.14), we obtain

$$
\begin{aligned}
\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty &\le \|\mathbb{E}_i[\mathbb{Y}_{i+1}]\|_\infty + \big\|\mathcal{D}_y^{-1}\big\|_\infty h\bigg|\sum_{j=0}^{k}\frac{b_j}{a_0}\mathbb{E}_i[\Delta f_{i+j}]\bigg| + \big\|\mathcal{D}_y^{-1}\big\|_\infty h\bigg|\frac{r_i^y}{a_0}\bigg|\\
&\le \|\mathbb{E}_i[\mathbb{Y}_{i+1}]\|_\infty + \big\|\mathcal{D}_y^{-1}\big\|_\infty hL\frac{\max_{0\le j\le k}\{|b_j|\}}{|a_0|}\bigg(\sum_{j=0}^{k}\mathbb{E}_i[|\Delta Y_{i+j}| + |\Delta Z_{i+j}|]\bigg)\\
&\quad + \big\|\mathcal{D}_y^{-1}\big\|_\infty h\bigg|\frac{r_i^y}{a_0}\bigg|.
\end{aligned} \tag{3.15}
$$

Squaring (3.15) and by the inequalities

$$
(a+b)^2 \le (1+\delta h)a^2 + \Big(1+\frac{1}{\delta h}\Big)b^2, \quad a,b\in\mathbb{R}, \quad \delta>0,
$$

we have

$$
\begin{aligned}
\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2 &\le (1+\delta h)\|\mathbb{E}_i[\mathbb{Y}_{i+1}]\|_\infty^2 + 2\Big(1+\frac{1}{\delta h}\Big)\big\|\mathcal{D}_y^{-1}\big\|_\infty^2 h^2\bigg|\frac{r_i^y}{a_0}\bigg|^2\\
&\quad + 4\Big(1+\frac{1}{\delta h}\Big)\big\|\mathcal{D}_y^{-1}\big\|_\infty^2\|\mathcal{D}_y\|_\infty^2 h^2kL^2\frac{\max_{0\le j\le k}\big\{|b_j|^2\big\}}{a_0^2}\\
&\quad \times\sum_{j=0}^{1}\big(\|\mathbb{E}_i[\mathbb{Y}_{i+j}]\|_\infty^2 + \|\mathbb{E}_i[\mathbb{Z}_{i+j}]\|_\infty^2\big).
\end{aligned} \tag{3.16}
$$

Step 3. Estimates for $\mathbb{E}[|\Delta Y_i|^2] + \sum_{\ell=i}^{N-k}\mathbb{E}[h|Z_\ell|^2]$. Multiplying (3.11) by h and then adding the derived inequality to (3.16), we have

$$
\begin{aligned}
&\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2 + h\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty^2\\
&\le \bigg(1 + h\bigg(\delta + 4h\Big(1+\frac{1}{\delta h}\Big)\big\|\mathcal{D}_y^{-1}\big\|_\infty^2\|\mathcal{D}_y\|_\infty^2 kL^2\frac{\max_{0\le j\le k}\big\{|b_j|^2\big\}}{a_0^2}\bigg)\bigg)\mathbb{E}_i\big[\|\mathbb{Y}_{i+1}\|_\infty^2\big]
\end{aligned}
$$

$$
\begin{aligned}
&+4\left(1+\frac{1}{\delta h}\right)\left\|\mathcal{D}_y^{-1}\right\|_\infty^2\|\mathcal{D}_y\|_\infty^2 h^2 kL^2\frac{\max_{0\le j\le k}\left\{|b_j|^2\right\}}{a_0^2}\left(\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2+\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty^2\right)\\
&+h\left(1+\gamma+4\left(1+\frac{1}{\delta h}\right)\left\|\mathcal{D}_y^{-1}\right\|_\infty^2\|\mathcal{D}_y\|_\infty^2 hkL^2\frac{\max_{0\le j\le k}\left\{|b_j|^2\right\}}{a_0^2}\right)\|\mathbb{E}_i[\mathbb{Z}_{i+1}]\|_\infty^2\\
&+2\left(1+\frac{1}{\delta h}\right)\left\|\mathcal{D}_y^{-1}\right\|_\infty^2 h^2\left|\frac{r_i^y}{a_0}\right|^2+\left(1+\frac{1}{\gamma}\right)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2\frac{2h}{\beta_0^2}\left(r_i^z\right)^2. \qquad (3.17)
\end{aligned}
$$

Eq. (3.17) can be simplified to

$$
\begin{aligned}
&\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2+C_2h\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty^2\\
&\le(1+C_3h)\mathbb{E}_i\left[\|\mathbb{Y}_{i+1}\|_\infty^2\right]+C_4h\mathbb{E}_i\left[\|\mathbb{Z}_{i+1}\|_\infty^2\right]+C_5\left(h^2\left|r_i^y\right|^2+h\left|r_i^z\right|^2\right), \qquad (3.18)
\end{aligned}
$$

where

$$
\begin{aligned}
C_1&=4\left(1+\frac{1}{\delta h}\right)\left\|\mathcal{D}_y^{-1}\right\|_\infty^2\|\mathcal{D}_y\|_\infty^2 kL^2\frac{\max_{0\le j\le k}\left\{|b_j|^2\right\}}{a_0^2},\\
C_2&=\frac{1-C_1h}{1-C_1h^2},\\
C_3&=\frac{\delta+2C_1h}{1-C_1h^2},\\
C_4&=\frac{1+\gamma+C_1h}{(1-C_1h^2)(1+C_3h)},\\
C_5&=\max\left(\frac{2(1+1/\delta h)\left\|\mathcal{D}_y^{-1}\right\|_\infty^2/a_0^2}{1-C_1h^2},\frac{(1+1/\gamma)\left\|\mathcal{D}_z^{-1}\right\|_\infty^2 2/\beta_0^2}{1-C_1h^2}\right).
\end{aligned}
$$

Then we choose suitable γ,δ and a small enough h such that

$$
C_3\le C,\quad C_5\le C,\quad C_2-C_4>\widehat{C}>0,
$$

where $C,\widehat{C}$ are two positive constants. By (3.18), we have

$$
\begin{aligned}
&\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2+C_2h\|\mathbb{E}_i[\mathbb{Z}_i]\|_\infty^2\\
&\le(1+Ch)\left(\mathbb{E}_i\left[\|\mathbb{Y}_{i+1}\|_\infty^2\right]+C_4h\mathbb{E}_i\left[\|\mathbb{Z}_{i+1}\|_\infty^2\right]\right)+C\left(h^2\left|r_i^y\right|^2+h\left|r_i^z\right|^2\right). \qquad (3.19)
\end{aligned}
$$

Repeating the above iteration (3.19), we obtain

$$
\begin{aligned}
&\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2+\widehat{C}h\sum_{\ell=i}^{N-k}(1+Ch)^{\ell-i}\|\mathbb{E}_i[\mathbb{Z}_\ell]\|_\infty^2\\
&\le(1+Ch)^{N-k-i+1}\left(\mathbb{E}_i\left[\|\mathbb{Y}_{N-k+1}\|_\infty^2\right]+C_4h\mathbb{E}_i\left[\|\mathbb{Z}_{N-k+1}\|_\infty^2\right]\right)\\
&\quad+\sum_{\ell=i}^{N-k}(1+Ch)^{\ell-i}C\left(h^2\left|r_\ell^y\right|^2+h\left|r_\ell^z\right|^2\right)\\
&\le(1+Ch)^N\left(\mathbb{E}_i\left[\|\mathbb{Y}_{N-k+1}\|_\infty^2\right]+C_4h\mathbb{E}_i\left[\|\mathbb{Z}_{N-k+1}\|_\infty^2\right]\right)\\
&\quad+\sum_{\ell=i}^{N-k}(1+C_3h)^NC\left(h^2\left|r_\ell^y\right|^2+h\left|r_\ell^z\right|^2\right)\\
&\le\exp\{CT\}\left(\mathbb{E}_i\left[\|\mathbb{Y}_{N-k+1}\|_\infty^2\right]+C_4h\mathbb{E}_i\left[\|\mathbb{Z}_{N-k+1}\|_\infty^2\right]+\sum_{\ell=i}^{N-k}C\left(h^2\left|r_\ell^y\right|^2+h\left|r_\ell^z\right|^2\right)\right). \qquad (3.20)
\end{aligned}
$$

Then we have

$$\begin{aligned} &\|\mathbb{E}_i[\mathbb{Y}_i]\|_\infty^2 + \sum_{\ell=i}^{N-k} h\|\mathbb{E}_i[\mathbb{Z}_\ell]\|_\infty^2 \\ &\leq C\left(\mathbb{E}_i\left[\|\mathbb{Y}_{N-k+1}\|_\infty^2\right] + h\mathbb{E}_i\left[\|\mathbb{Z}_{N-k+1}\|_\infty^2\right] + \sum_{\ell=i}^{N-k}\left(h^2\left|r_\ell^y\right|^2 + h\left|r_\ell^z\right|^2\right)\right). \end{aligned} \tag{3.21}$$

From

$$\begin{aligned} &\mathbb{E}_i\left[\|\mathbb{Y}_i\|_\infty^2\right] \leq \left\|\mathcal{D}_y^{-1}\right\|_\infty^2 \mathbb{E}_i\left[\|\mathcal{Y}_i\|_\infty^2\right] = \left\|\mathcal{D}_y^{-1}\right\|_\infty^2 \max_{i\leq j<i+k} \mathbb{E}_i\left[|\Delta Y_j|^2\right], \\ &\mathbb{E}_i\left[|\Delta Y_i|^2\right] \leq \mathbb{E}_i\left[\|\mathcal{Y}_i\|_\infty^2\right] \leq \|\mathcal{D}_y\|_\infty^2 \mathbb{E}_i\left[\|\mathbb{Y}_i\|_\infty^2\right], \\ &\mathbb{E}_i\left[\|\mathbb{Z}_i\|_\infty^2\right] \leq \left\|\mathcal{D}_z^{-1}\right\|_\infty^2 \mathbb{E}_i\left[\|\mathcal{Z}_i\|_\infty^2\right] = \left\|\mathcal{D}_z^{-1}\right\|_\infty^2 \max_{i\leq j<i+k} \mathbb{E}_i\left[|\Delta Z_j|^2\right], \\ &\mathbb{E}_i\left[|\Delta Z_i|^2\right] \leq \mathbb{E}_i\left[\|\mathcal{Z}_i\|_\infty^2\right] \leq \|\mathcal{D}_z\|_\infty^2 \mathbb{E}_i\left[\|\mathbb{Z}_i\|_\infty^2\right], \end{aligned}$$

we have

$$\begin{aligned} &\|\mathcal{D}_z\|_\infty^2 \|\mathbb{E}_i[\mathcal{Y}_i]\|_\infty^2 + \|\mathcal{D}_y\|_\infty^2 \sum_{\ell=i}^{N-k} h\|\mathbb{E}_i[\mathcal{Z}_\ell]\|_\infty^2 \\ &\leq C\|\mathcal{D}_y\|_\infty^2\|\mathcal{D}_z\|_\infty^2\Bigg(\left\|\mathcal{D}_y^{-1}\right\|_\infty^2 \mathbb{E}_i\left[\|\mathcal{Y}_{N-k+1}\|_\infty^2\right] + h\left\|\mathcal{D}_z^{-1}\right\|_\infty^2 \mathbb{E}_i\left[\|\mathcal{Z}_{N-k+1}\|_\infty^2\right] \\ &\qquad + \sum_{\ell=i}^{N-k}\left(h^2\left|r_\ell^y\right|^2 + h\left|r_\ell^z\right|^2\right)\Bigg). \end{aligned} \tag{3.22}$$

Thus, there exists a constant C such that

$$\begin{aligned} &\|\mathbb{E}_i[\mathcal{Y}_i]\|_\infty^2 + \sum_{\ell=i}^{N-k} h\|\mathbb{E}_i[\mathcal{Z}_\ell]\|_\infty^2 \\ &\leq C\left(\mathbb{E}_i\left[\|\mathcal{Y}_{N-k+1}\|_\infty^2\right] + h\mathbb{E}_i\left[\|\mathcal{Z}_{N-k+1}\|_\infty^2\right] + \sum_{\ell=i}^{N-k}\left(h^2\left|r_\ell^y\right|^2 + h\left|r_\ell^z\right|^2\right)\right). \end{aligned} \tag{3.23}$$

Furthermore,

$$\begin{aligned} &\mathbb{E}\left[|\Delta Y_i|^2\right] + \sum_{\ell=i}^{N-k} \mathbb{E}\left[h|Z_\ell|^2\right] \\ &\leq C\left(\max_{N-k<j\leq N} \mathbb{E}\left[|\Delta Y_j|^2\right] + \max_{N-k<j\leq N} \mathbb{E}\left[h|\Delta Z_j|^2\right] + \sum_{\ell=i}^{N-k} \mathbb{E}\left[h^2\left|r_\ell^y\right|^2 + h\left|r_\ell^z\right|^2\right]\right). \end{aligned}$$

The proof is complete. □

4. Convergence Results

We concentrate on demonstrating the convergence rate of the proposed multi-step scheme (2.14) in this section. Before that, we introduce an essential proposition and then demonstrate two propositions.

For readers' convenience, we introduce some symbols before providing Itô-Taylor expansions (see [28, Theorem 5.5.1]). For a multi-index with finite length $\gamma := (\gamma_1, \gamma_2, \cdots, \gamma_p), p \in \mathbb{N}^+$, $\gamma_\ell \in \{0, 1, 2, \dots, d\}$, for $\ell \in \{1, 2, \dots, p\}$, let $\ell(\gamma)$ be the length of a multi-index of γ; $\mathcal{A}^\gamma$ is the set of all functions $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d_1}$ for which $\mathcal{L}^\gamma v$ is well defined and continuous where $\mathcal{L}^\gamma = \mathcal{L}^{(\gamma_1)} \circ \cdots \circ \mathcal{L}^{(\gamma_p)}$, $\mathcal{L}^{(j)} = \sum_{k=1}^d \sigma_{kj} \partial_{x_k}$ for $j \in \{1, 2, \dots, d\}$; $\mathcal{A}_b^\gamma$ denotes the subset of all functions $v \in \mathcal{A}^\gamma$ such that the function $\mathcal{L}^\gamma v$ is bounded; for positive integer m, $\mathcal{A}_b^m$ is the set of functions v such that $v^\gamma \in \mathcal{A}_b^\gamma$ for all $\gamma \in \{\gamma | \ell(\gamma) \le m\} \setminus \{\oslash\}$.

Proposition 4.1 ([4, 41]) *Let $m \ge 0$. Then for a function $v \in \mathcal{A}_b^{m+1}$,*

$$\begin{aligned} &\mathbb{E}_t[v(t+h, X_{t+h})] \\ &= v_t + h v_t^{(0)} + \frac{h^2}{2} v_t^{(0,0)} + \cdots + \frac{h^m}{m!} v_t^{(0)_m} + \mathcal{O}(h^{m+1}), \end{aligned} \tag{4.1}$$

$$\begin{aligned} &\mathbb{E}_t[(W_{t+h} - W_t) v(t+h, X_{t+h})] \\ &= h v_t^{(1)} + h^2 v_t^{(1,0)} + \frac{h^3}{2} v_t^{(1)*(0,0)} + \cdots + \frac{h^{m+1}}{m!} v_t^{(1)*(0)_m} + \mathcal{O}(h^{m+2}), \end{aligned} \tag{4.2}$$

where

$$\begin{aligned} &v_t = v(t, X_t), \\ &v_t^{(0)} = \mathcal{L}^{(0)} v(t, X_t), \\ &v_t^{(0,0)} = \mathcal{L}^{(0)} \circ \mathcal{L}^{(0)} v(t, X_t), \\ &\cdots\cdots\cdots \\ &v_t^{(0)_m} = \underbrace{\mathcal{L}^{(0)} \circ \cdots \circ \mathcal{L}^{(0)}}_{m} v(t, X_t), \\ &v_t^{(1)} = \mathcal{L}^{(1)} v(t, X_t), \\ &v_t^{(1,0)} = \mathcal{L}^{(1)} \circ \mathcal{L}^{(0)} v(t, X_t), \\ &\cdots\cdots\cdots \\ &v_t^{(1)*(0)_m} = \mathcal{L}^{(1)} \circ \underbrace{\mathcal{L}^{(0)} \circ \cdots \circ \mathcal{L}^{(0)}}_{m} v(t, X_t), \\ &\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_t]. \end{aligned}$$

From (2.2), (1.3) and (4.1), we have

$$\frac{d\mathbb{E}_i[Y_{t_{i+j}}]}{dt} = -\mathbb{E}_i[f_{t_{i+j}}] = \mathbb{E}_i\big[u_{t_{i+j}}^{(0)}\big], \quad j = 0, 1, 2, \dots, k.$$

Assume that $u(t, X_t) \in \mathcal{A}_b^r$ i.e. $|u_t^{(0)_p}| \le M_p^y < +\infty, t \in [0, T], p = 0, 1, \dots, r \in \mathbb{N}^+$. By Itô-Taylor expansions (4.1), we obtain the expression of $u_{t_{i+j}}$ and $u_{t_{i+j}}^{(0)}$ as below

$$u_{t_{i+j}} = \sum_{p=0}^{r-1} u_{t_i}^{(0)_p} \frac{(jh)^p}{p!} + \eta_i^j, \tag{4.3}$$

$$u_{t_{i+j}}^{(0)} = \sum_{p=1}^{r-1} u_{t_i}^{(0)_p} \frac{(jh)^{p-1}}{(p-1)!} + \gamma_i^j, \tag{4.4}$$

where
$$\left|\eta_i^j\right| \le \frac{M_r^y (jh)^r}{r!}, \quad \left|\gamma_i^j\right| \le \frac{M_r^y (jh)^{r-1}}{(r-1)!}.$$
Substituting (4.3) and (4.4) into the right-hand side of the expression of
$$r_i^y = \sum_{j=0}^{k} \frac{a_j}{h}\mathbb{E}_i[Y_{t_{i+j}}] + \sum_{j=0}^{k} b_j \mathbb{E}_i[f_{t_{i+j}}],$$
we obtain
$$r_i^y = E_0^y \frac{u_{t_i}}{h} + E_1^y u_{t_i}^{(0)} + \cdots + E_{r-1}^y h^{r-2} u_{t_i}^{(0)_{r-1}} + \varepsilon_i^y, \tag{4.5}$$
where
$$\begin{aligned} E_0^y &= \sum_{j=0}^{k} a_j, \\ E_p^y &= \sum_{j=0}^{k} \frac{a_j j^p}{p!} - \sum_{j=0}^{k} \frac{b_j j^{p-1}}{(p-1)!}, \quad p > 0, \\ \varepsilon_i^y &= \sum_{j=0}^{k} \frac{a_j \eta_i^j}{h} - \sum_{j=0}^{k} b_j \gamma_i^j = \mathcal{O}(h^{r-1}). \end{aligned} \tag{4.6}$$
Furthermore, we have
$$\left|\varepsilon_i^y\right| \le D_r^y M_r^y h^{r-1}, \tag{4.7}$$
where
$$D_r^y = \sum_{j=0}^{k} \frac{|a_j| j^r}{r!} + \sum_{j=0}^{k} \frac{|b_j| j^{r-1}}{(r-1)!}.$$

Note that there are $2k+2$ unknowns in the linear homogeneous algebraic equations $E_0^y = E_1^y = \cdots = E_m^y = 0$. If the number of the unknowns is not less than the number of the equations, i.e. $2k+2 \ge m+1$, this system of equations has a non-zero solution.

In what follows, we analyze the order accuracy of Z. Assume that $Z_t \in \mathcal{A}_b^r$ i.e. $|Z_t^{(0)_p}| \le M_p^z < +\infty$, $t \in [0,T]$, $p = 0,1,\ldots,r \in \mathbb{N}^+$. By the relation (1.2), the solution (Y_t, Z_t) of (1.1), the solution $u(t, X_t)$ of (1.3) and (4.2), we obtain the expression of $\mathbb{E}_i[Z_{t_{i+j}}]$ and $\mathbb{E}_i[Y_{t_{i+j}} \Delta W_{i,i+j}^\top]$ as below
$$\mathbb{E}_i[Z_{t_{i+j}}] = \sum_{p=0}^{r-1} Z_{t_i}^{(0)_p} \frac{(jh)^p}{p!} + \tau_i^j = \sum_{p=0}^{r-1} u_{t_i}^{(1)*(0)_p} \frac{(jh)^p}{p!} + \tau_i^j, \tag{4.8}$$
$$\mathbb{E}_i\left[Y_{t_{i+j}} \Delta W_{i,i+j}^\top\right] = \sum_{p=0}^{r-1} u_{t_i}^{(1)*(0)_p} \frac{(jh)^{p+1}}{p!} + \varsigma_i^j, \tag{4.9}$$
where
$$\left|\tau_i^j\right| \le \frac{M_r^z (jh)^r}{r!}, \quad \left|\varsigma_i^j\right| \le \frac{M_r^z (jh)^{r+1}}{r!}.$$
Substituting (4.8) and (4.9) into the right-hand side of the expression of
$$r_i^z = \sum_{j=0}^{k} \beta_j \mathbb{E}_i[Z_{t_{i+j}}] - \sum_{j=1}^{k} \frac{\alpha_j \mathbb{E}_i\left[Y_{t_{i+j}} \Delta W_{i,i+j}^\top\right]}{h},$$

we obtain

$$r_i^z = E_0^z u_{t_i}^{(1)} + E_1^z h u_{t_i}^{(1,0)} + \cdots + E_{r-1}^z h^{r-1} u_{t_i}^{(1)*(0)_{r-1}} + \varepsilon_i^z, \tag{4.10}$$

where

$$\begin{aligned} E_0^z &= \sum_{j=0}^{k} (\beta_j - \alpha_j j), \\ E_p^z &= \sum_{j=1}^{k} \frac{(\beta_j - \alpha_j j) j^p}{p!}, \quad p > 0, \\ \varepsilon_i^z &= \sum_{j=0}^{k} \beta_j \tau_i^j - \sum_{j=1}^{k} \alpha_j \varsigma_i^j = \mathcal{O}(h^r). \end{aligned} \tag{4.11}$$

Furthermore, we have

$$\left|\varepsilon_i^z\right| \leq D_r^z M_r^z h^r, \tag{4.12}$$

where

$$D_r^z = \sum_{j=0}^{k} \frac{|\beta_j| j^r}{r!} + \sum_{j=1}^{k} \frac{|\alpha_j| j^{r+1}}{r!}.$$

Note that there are $2k+1$ unknowns in the linear homogeneous algebraic equations $E_0^z = E_1^z = \cdots = E_m^z = 0$. If the number of the unknowns is more than or equal to the number of the equations, i.e. $2k+1 \geq m+1$, this system of equations has a non-zero solution.

Proposition 4.2. *Assume that $u_t \in \mathcal{A}_b^2$, we have*

$$\begin{aligned} &\lim_{h \to 0} \sum_{j=0}^{k} \frac{a_j \mathbb{E}_t[u_{t+jh}]}{h} = u_t^{(0)}, \\ &\lim_{h \to 0} \sum_{j=0}^{k} b_j \mathbb{E}_t[f(Y_{t+jh}, Z_{t+jh})] = f(Y_t, Z_t), \end{aligned} \tag{4.13}$$

if and only if

$$E_0^y = E_1^y = 0, \quad \sum_{j=0}^{k} b_j = 1. \tag{4.14}$$

Proof. From Itô-Taylor expansions, we have

$$\begin{aligned} &\mathbb{E}_t[u_{t+jh}] = u_t + jh u_t^{(0)} + \mathcal{O}(h^2), \\ &\mathbb{E}_t[f(Y_{t+jh}, Z_{t+jh})] = f(Y_t, Z_t) + \mathcal{O}(h). \end{aligned} \tag{4.15}$$

Substituting (4.15) into the left-hand side of (4.13), we get

$$\begin{aligned} &\lim_{h \to 0} \left(\left(\sum_{j=0}^{k} \frac{a_j u_t}{h} \right) + \sum_{j=0}^{k} a_j j u_t^{(0)} + \mathcal{O}(h) \right) = u_t^{(0)}, \\ &\lim_{h \to 0} \left(\left(\sum_{j=0}^{k} b_j \right) f(Y_t, Z_t) + \mathcal{O}(h) \right) = f(Y_t, Z_t). \end{aligned} \tag{4.16}$$

The relations in (4.16) hold if and only if

$$\sum_{j=0}^{k} a_j = 0, \quad \sum_{j=0}^{k} a_j j = 1, \quad \sum_{j=0}^{k} b_j = 1, \tag{4.17}$$

where $\sum_{j=0}^{k} a_j$ equals E_0^y, $\sum_{j=0}^{k} a_j j - \sum_{j=0}^{k} b_j$ equals E_1^y. Thus, we demonstrate that the result holds. □

Proposition 4.3. *Assume that $Z_t \in \mathcal{A}_b^2$, we have*

$$\lim_{h\to 0} \sum_{j=0}^{k} \beta_j \mathbb{E}_t[Z_{t+jh}] = Z_t, \tag{4.18}$$

if and only if

$$\sum_{j=0}^{k} \beta_j = 1. \tag{4.19}$$

Proof. From Itô-Taylor expansions, we have

$$\mathbb{E}_t[Z_{t+jh}] = Z_t + \mathcal{O}(jh). \tag{4.20}$$

Substituting (4.20) into the left-hand side of (4.18), we get

$$\lim_{h\to 0} \left(\left(\sum_{j=0}^{k} \beta_j \right) Z_t + \mathcal{O}(h) \right) = Z_t. \tag{4.21}$$

The relation in (4.21) holds if and only if

$$\sum_{j=0}^{k} \beta_j = 1,$$

Thus, we demonstrate that the result holds. □

Theorem 4.1. *Suppose that the assumptions (i)-(iii) and $Y_t, Z_t \in \mathcal{A}_b^k$ hold. Furthermore, assume that the numerical scheme* (2.14) *fulfils Dahlquist's root condition and let the initial errors satisfy*

$$\max_{N-k+1\le i\le N} \mathbb{E}\left[\left|Y_i^\pi - Y_{t_i}\right|^2 + h\left|Z_i^\pi - Z_{t_i}\right|^2 \right] = \mathcal{O}(h^{2k}),$$

and the approximate errors satisfy

$$\max_{0\le i\le N-k} \left\{ \mathbb{E}\left[\left|r_i^y\right|\right], \mathbb{E}\left[\left|r_i^z\right|\right] \right\} = \mathcal{O}(h^k),$$
$$r_i^y = \sum_{j=0}^{k} \frac{a_j}{h} \mathbb{E}_i[Y_{t_{i+j}}] + \sum_{j=0}^{k} b_j \mathbb{E}_i[f_{t_{i+j}}],$$
$$r_i^z = \sum_{j=0}^{k} \beta_j \mathbb{E}_i[Z_{t_{i+j}}] - \sum_{j=1}^{k} \frac{\alpha_j \mathbb{E}_i\left[Y_{t_{i+j}} \Delta W_{i,i+j}^\top\right]}{h}.$$

Then the following estimates hold:

$$\max_{0\le i\le N-k} \mathbb{E}\left[|\Delta Y_i|^2 + h|\Delta Z_i|^2\right] \le C h^{2k}.$$

Proof. The proof follows the line of proofs of Theorem 3.1. Hence, we omit it here. □

5. Multilevel Simulation for $\mathbb{E}_i[\cdot]$

In this section, we discuss how to design the algorithm for approximating the conditional expectations in the scheme (2.14) by utilizing the MLMC method and analyze the complexity of the algorithm. The reason why the MLMC method is adopted is that this way can significantly reduce the computational complexity of Monte Carlo simulation (see [15, 16] for details). The MLMC method is reviewed in Section 5.1, we will analyze the computational complexity of the proposed methods in Section 5.2.

5.1. MLMC

Here, we review the idea of MLMC (see [15, 16]) as below.

An output functional U is approximated by a given sequence $U_l, l = 0, 1, \dots$ with increasing accuracy and cost as l increases. The simple identity is deduced

$$\mathbb{E}[U] = \mathbb{E}[U_0] + \sum_{l=1}^{+\infty} \mathbb{E}[U_l - U_{l-1}] \approx \mathbb{E}[U_0] + \sum_{l=1}^{\widehat{L}} \mathbb{E}[U_l - U_{l-1}], \tag{5.1}$$

where $\widehat{L}$ is chosen large enough to ensure that the weak error $\mathbb{E}[U - U_{\widehat{L}}]$,

$$U_{\widehat{L}} = U_0 + \sum_{l=1}^{\widehat{L}} (U_l - U_{l-1})$$

is acceptably small. Every of the expectations on the right-hand side of (5.1) is estimated independently utilizing N_l samples so that the MLMC estimator is

$$P = \sum_{l=0}^{\widehat{L}} P_l,$$

where

$$P_0 = \frac{1}{N_0} \sum_{n=1}^{N_0} U_0^{(n)}, \quad P_l = \frac{1}{N_l} \sum_{n=1}^{N_l} \big(U_l^{(n)} - U_{l-1}^{(n)}\big), \quad l = 1, 2, \dots, \widehat{L},$$

$(U_l^{(n)}, U_{l-1}^{(n)})$ are independent copies of (U_l, U_{l-1}) for $n = 1, 2, \dots, N_l$. The computational savings come from the fact that $U_l - U_{l-1}$ is smaller and has a smaller variance on the finer levels. Thus, fewer samples N_l are required to accurately estimate its expected value.

5.2. Computational complexity

Throughout this section, we assume that time steps satisfy $h_l = h_{l-1}/M, h_0 = h$, for integer $M \geq 2$, $l = 1, 2, \dots, \widehat{L}$, where $\widehat{L}$ is an integer and no less than 2. And the approximation on level l with respect to (Y_i^π, Z_i^π) is denoted by $(Y_{i,l}^\pi, Z_{i,l}^\pi)$. From the key identity of the MLMC method, we obtain, for $i = N - k, N - k - 1, \dots, 0$,

$$\sum_{j=1}^{k} a_j \mathbb{E}_i\big[Y_{i+j,\widehat{L}}^\pi\big] = \sum_{j=1}^{k} a_j \mathbb{E}_i\big[Y_{i+j,0}^\pi\big] + \sum_{j=1}^{k} a_j \sum_{l=1}^{\widehat{L}} \mathbb{E}_i\big[Y_{i+j,l}^\pi - Y_{i+j,l-1}^\pi\big], \tag{5.2a}$$

$$\sum_{j=0}^{k} b_j \mathbb{E}_i\big[h_{\widehat{L}} f_{i+j,\widehat{L}}^\pi\big] = \sum_{j=0}^{k} b_j \mathbb{E}_i\big[h_0 f_{i+j,0}^\pi\big] + \sum_{j=0}^{k} b_j \sum_{l=1}^{\widehat{L}} \mathbb{E}_i\big[h_l f_{i+j,l}^\pi - h_{l-1} f_{i+j,l-1}^\pi\big], \tag{5.2b}$$

$$\sum_{j=1}^{k}\beta_j\mathbb{E}_i\big[Z^{\pi}_{i+j,\widehat{L}}\big]=\sum_{j=1}^{k}\beta_j\mathbb{E}_i\big[Z^{\pi}_{i+j,0}\big]+\sum_{j=1}^{k}\beta_j\sum_{l=1}^{\widehat{L}}\mathbb{E}_i\big[Z^{\pi}_{i+j,l}-Z^{\pi}_{i+j,l-1}\big], \tag{5.2c}$$

$$\sum_{j=1}^{k}\frac{\alpha_j\mathbb{E}_i\big[Y^{\pi}_{i+j}\Delta_{h_{\widehat{L}}}W^{\top}_{i,i+j}\big]}{h_{\widehat{L}}}=\sum_{j=1}^{k}\alpha_j\mathbb{E}_i\left[Y^{\pi}_{i+j,0}\frac{\Delta_{h_0}W^{\top}_{i,i+j}}{h_0}\right]+\sum_{j=1}^{k}\alpha_j\sum_{l=1}^{\widehat{L}}\mathbb{E}_i\left[Y^{\pi}_{i+j,l}\frac{\Delta_{h_l W^{\top}_{i,i+j}}}{h_l}-Y^{\pi}_{i+j,l-1}\frac{\Delta_{h_{l-1}W^{\top}_{i,i+j}}}{h_{l-1}}\right], \tag{5.2d}$$

where

$$\Delta_{h_l}W_{i,i+j}=W_{(i+j)h_l}-W_{i*h_l},\quad f^{\pi}_{i+j,l}=f\big(Y^{\pi}_{i+1,l},Z^{\pi}_{i+1,l}\big).$$

Note that instead of approximating $\sum_{j=1}^{k}a_j\mathbb{E}_i[Y^{\pi}_{i+j,\widehat{L}}]$, $\sum_{j=0}^{k}b_j\mathbb{E}_i[h_{\widehat{L}}f^{\pi}_{i+j,\widehat{L}}]$, $\sum_{j=1}^{k}\beta_j\mathbb{E}_i[Z^{\pi}_{i+j,\widehat{L}}]$, $\sum_{j=1}^{k}\alpha_j\mathbb{E}_i[Y^{\pi}_{i+j}\Delta_{h_{\widehat{L}}}W^{\top}_{i,i+j}]/h_{\widehat{L}}$ by Monte Carlo running with $\widehat{N}$ simulations, the expectations

$$\sum_{j=1}^{k}a_j\mathbb{E}_i\big[Y^{\pi}_{i+j,0}\big],\quad \sum_{j=1}^{k}a_j\sum_{l=1}^{\widehat{L}}\mathbb{E}_i\big[Y^{\pi}_{i+j,l}-Y^{\pi}_{i+j,l-1}\big],\quad \sum_{j=0}^{k}b_j\mathbb{E}_i\big[f^{\pi}_{i+j,0}\big],$$

$$\sum_{j=0}^{k}b_j\sum_{l=1}^{\widehat{L}}\mathbb{E}_i\big[f^{\pi}_{i+j,l}-f^{\pi}_{i+j,l-1}\big],\quad \sum_{j=1}^{k}\beta_j\mathbb{E}_i\big[Z^{\pi}_{i+j,0}\big],\quad \sum_{j=1}^{k}\alpha_j\mathbb{E}_i\left[Y^{\pi}_{i+j,0}\frac{\Delta_{h_0}W^{\top}_{i,i+j}}{h_0}\right],$$

$$\sum_{j=1}^{k}\beta_j\sum_{l=1}^{\widehat{L}}\mathbb{E}_i\big[Z^{\pi}_{i+j,l}-Z^{\pi}_{i+j,l-1}\big],\quad \sum_{j=1}^{k}\alpha_j\sum_{l=1}^{\widehat{L}}\mathbb{E}_i\left[Y^{\pi}_{i+j,l}\frac{\Delta_{h_l}W^{\top}_{i,i+j}}{h_l}-Y^{\pi}_{i+j,l-1}\frac{\Delta_{h_{l-1}}W^{\top}_{i,i+j}}{h_{l-1}}\right]$$

in the Eq. (5.2) are estimated independently in a way in which the overall variance for a given computational complexity, namely the number of random numbers, arithmetic operations and function evaluations, is minimised. First let

$$P^{y}_{i,0}=-\frac{1}{N_0}\sum_{n=1}^{N_0}\left(\sum_{j=1}^{k}a_jY^{\pi,(n)}_{i+j,0}+h_0\sum_{j=1}^{k}b_jf^{\pi,(n)}_{i+j,0}\right),$$

$$P^{z}_{i,0}=\frac{1}{N_0}\sum_{n=1}^{N_0}\left(-\sum_{j=1}^{k}\beta_jZ^{\pi,(n)}_{i+j,0}+\sum_{j=1}^{k}\alpha_jY^{\pi,(n)}_{i+j,0}\frac{\Delta^{(n)}_{h_0}W^{\top}_{i,i+j}}{h_0}\right),$$

where

$$\Delta^{(n)}_{h_0}W_{i,i+j}=W^{(n)}_{(i+j)h_0}-W^{(n)}_{i*h_0},\quad f^{\pi,(n)}_{i+j,0}=f\big(Y^{\pi,(n)}_{i+j,0},Z^{\pi,(n)}_{i+j,0}\big),$$

$Y^{\pi,(n)}_{i+j,0}$ and $Z^{\pi,(n)}_{i+j,0}$ are independent copies of $Y^{\pi}_{i+j,0}$ and $Z^{\pi}_{i+j,0}$ for $n=1,2,\ldots,N_0, j=1,2,\ldots,k$. Let $Y^{\pi}_{i+j,l}$ and $Y^{\pi}_{i+j,l-1}$ be approximations of $Y_{t_{i+j}}$ with step size h_l and h_{l-1}, which utilize the same Brownian motion(see [15] for details). Correspondingly, $Z^{\pi}_{i+j,l}$ and $Z^{\pi}_{i+j,l-1}$ are approximations of $Z_{t_{i+j}}$ with step size h_l and h_{l-1}, which use the same Brownian motion as $Y_{t_{i+j}}$, define

$$P^{y}_{i,l}=-\frac{1}{N_l}\sum_{n=1}^{N_l}\left[\sum_{j=1}^{k}a_j\big(Y^{\pi,(n)}_{i+j,l}-Y^{\pi,(n)}_{i+j,l-1}\big)+\sum_{j=1}^{k}b_j\big(h_lf^{\pi,(n)}_{i+j,l}-h_{l-1}f^{\pi,(n)}_{i+j,l-1}\big)\right],$$

$$P^{z}_{i,l}=\frac{1}{N_l}\sum_{n=1}^{N_l}\left[-\sum_{j=1}^{k}\beta_j\big(Z^{\pi,(n)}_{i+j,l}-Z^{\pi,(n)}_{i+j,l-1}\big)+\sum_{j=1}^{k}\alpha_j\left(Y^{\pi,(n)}_{i+j,l}\frac{\Delta^{(n)}_{h_l}W^{\top}_{i,i+j}}{h_l}-Y^{\pi,(n)}_{i+j,l-1}\frac{\Delta^{(n)}_{h_{l-1}}W^{\top}_{i,i+j}}{h_{l-1}}\right)\right],$$

where

$$\Delta_{h_l}^{(n)} W_{i,i+j} = W_{(i+j)h_l}^{(n)} - W_{i*h_l}^{(n)}, \quad f_{i+j,l}^{\pi,(n)} = f\big(Y_{i+j,l}^{\pi,(n)}, Z_{i+j,l}^{\pi,(n)}\big),$$

$(Y_{i+j,l}^{\pi,(n)}, Y_{i+j,l-1}^{\pi,(n)})$ are independent copies of $(Y_{i+j,l}^{\pi}, Y_{i+j,l-1}^{\pi})$, $(Z_{i+j,l}^{\pi,(n)}, Z_{i+j,l-1}^{\pi,(n)})$ are independent copies of $(Z_{i+j,l}^{\pi}, Z_{i+j,l-1}^{\pi})$ for $n = 1, 2, \ldots, N_l, j = 1, 2, \ldots, k$. Thus, the MLMC estimators of the quantities

$$\begin{aligned} &-\sum_{j=1}^{k} a_j \mathbb{E}_i\big[Y_{i+j}^{\pi}\big] - \sum_{j=0}^{k} b_j h \mathbb{E}_i\big[f_{i+j}^{\pi}\big], \\ &-\sum_{j=1}^{k} \beta_j \mathbb{E}_i\big[Z_{i+j}^{\pi}\big] + \sum_{j=1}^{k} \frac{\alpha_j \mathbb{E}_i\big[Y_{i+j}^{\pi} \Delta W_{i,i+j}^{\top}\big]}{h} \end{aligned}$$

are given by

$$P_i^y = \sum_{l=0}^{\widehat{L}} P_{i,l}^y, \tag{5.3}$$

$$P_i^z = \sum_{l=0}^{\widehat{L}} P_{i,l}^z, \tag{5.4}$$

where the estimators $(P_{i,l}^y, P_{i,l}^z)$ of the different levels must be independent for $l = 0, 1, \ldots, \widehat{L}$.

In what follows, we show the computational complexity of the proposed fully discrete multi-step method.

Theorem 5.1. *Let assumptions of Theorem* 4.1 *hold. Furthermore, let* (Y_{t_i}, Z_{t_i}) *be the analytical solution of the BSDE in* (1.1) *and denote by* $(Y_{i,l}^{\pi}, Z_{i,l}^{\pi})$ *the corresponding numerical approximation of* (Y_{t_i}, Z_{t_i}) *on level* l *with step size* $h_l = M^{-l} h_0$. *Suppose that there exist positive constants* c_1 *and* κ *such that* $\mathcal{C}_l \leq c_1 N_l h_l^{-\kappa}$, *where* $\mathcal{C}_l$ *denotes the computational complexity at level* l. *Then for every* $\epsilon > 0$, *there are choices* $\widehat{L}$ *and* N_l *for* $l = 0, 1, \ldots, \widehat{L}$ *to be provided in* (5.16) *and* (5.17) *respectively and constants* c_2, c_3, c_4 *and* c_5 *in* (5.18) *such that the estimators* (P_i^y, P_i^z) *satisfy the error bound*

$$\max_{0 \leq i \leq N} \mathbb{E}\Big[\big|Y_{t_i} - P_i^y\big|^2 + h_{\widehat{L}}\big|Z_{t_i} - P_i^z\big|^2\Big] \leq \epsilon^2$$

with the computational complexity bound

$$\mathcal{C} \leq \begin{cases} c_2 \epsilon^{-2}, & 2k > \kappa, \\ (c_2 + c_3 + c_4)\epsilon^{-2}, & 2k = \kappa, \\ c_4 \epsilon^{\frac{-\kappa}{k}}, & 2k < \kappa. \end{cases}$$

Proof. **Step 1.** We deal with the term $\max_{0 \leq i \leq N} \mathbb{E}[|Y_{t_i} - P_i^y|^2 + h_{\widehat{L}}|Z_{t_i} - P_i^z|^2]$. We know

$$\begin{aligned} &\mathbb{E}_i\left[\big(Y_{t_i} - P_i^y\big)^2 + h_{\widehat{L}}\big(Z_{t_i} - P_i^z\big)^2\right] \\ &= \big(Y_{t_i} - \mathbb{E}_i\big[P_i^y\big]\big)^2 + \mathbb{V}_i\big(P_i^y\big) + h_{\widehat{L}}\big(Z_{t_i} - \mathbb{E}_i\big[P_i^z\big]\big)^2 + h_{\widehat{L}} \mathbb{V}_i\big(P_i^z\big), \end{aligned} \tag{5.5}$$

where $\mathbb{V}_i[\cdot] = \mathbb{V}[\cdot|\mathcal{F}_{t_i}]$ denotes the conditional variance. From the definitions of $P^y_{i,l}$ and $P^z_{i,l}$, we have for $l = 1, 2, \ldots, \widehat{L}$,

$$
\begin{aligned}
\mathbb{E}_i\big[P^y_{i,0}\big] &= -\mathbb{E}_i\Bigg[\sum_{j=1}^{k} a_j Y^\pi_{i+j,0} + h_0 \sum_{j=1}^{k} b_j f^\pi_{i+j,0}\Bigg],\\
\mathbb{E}_i\big[P^y_{i,l}\big] &= -\mathbb{E}_i\Bigg[\sum_{j=1}^{k} a_j \big(Y^\pi_{i+j,l} - Y^\pi_{i+j,l-1}\big) + \sum_{j=1}^{k} b_j \big(h_l f^\pi_{i+j,l} - h_{l-1} f^\pi_{i+j,l-1}\big)\Bigg],\\
\mathbb{E}_i\big[P^z_{i,0}\big] &= \mathbb{E}_i\Bigg[-\sum_{j=1}^{k} \beta_j Z^\pi_{i+j,0} + \sum_{j=1}^{k} \alpha_j Y^\pi_{i+j,0} \frac{\Delta_{h_0} W^\top_{i,i+j}}{h_0}\Bigg],\\
\mathbb{E}_i\big[P^z_{i,l}\big] &= \mathbb{E}_i\Bigg[-\sum_{j=1}^{k} \beta_j \big(Z^\pi_{i+j,l} - Z^\pi_{i+j,l-1}\big)\\
&\qquad + \sum_{j=1}^{k} \alpha_j \bigg(Y^\pi_{i+j,l} \frac{\Delta_{h_l} W^\top_{i,i+j}}{h_l} - Y^\pi_{i+j,l-1} \frac{\Delta_{h_l} W^\top_{i,i+j}}{h_{l-1}}\bigg)\Bigg].
\end{aligned}
\tag{5.6}
$$

From (5.6), (5.3) and (5.4), we obtain

$$
\begin{aligned}
\mathbb{E}_i\big[P^y_{i}\big] &= -\mathbb{E}_i\Bigg[\sum_{j=1}^{k} a_j Y^\pi_{i+j,0} + h_0 \sum_{j=1}^{k} b_j f^\pi_{i+j,0}\Bigg]\\
&\quad - \sum_{l=1}^{\widehat{L}} \mathbb{E}_i\Bigg[\sum_{j=1}^{k} a_j \big(Y^\pi_{i+j,l} - Y^\pi_{i+j,l-1}\big) + \sum_{j=1}^{k} b_j \big(h_l f^\pi_{i+j,l} - h_{l-1} f^\pi_{i+j,l-1}\big)\Bigg]\\
&= -\sum_{j=1}^{k} a_j \mathbb{E}_i\big[Y^\pi_{i+j,\widehat{L}}\big] - \sum_{j=1}^{k} b_j h_{\widehat{L}} \mathbb{E}_i\big[f^\pi_{i+j,\widehat{L}}\big],\\
\mathbb{E}_i\big[P^z_{i}\big] &= \mathbb{E}_i\Bigg[-\sum_{j=1}^{k} \beta_j Z^\pi_{i+j,0} + \sum_{j=1}^{k} \alpha_j Y^\pi_{i+j,0} \frac{\Delta_{h_0} W^\top_{i,i+j}}{h_0}\Bigg]\\
&\quad + \sum_{l=1}^{\widehat{L}} \mathbb{E}_i\Bigg[-\sum_{j=1}^{k} \beta_j \big(Z^\pi_{i+j,l} - Z^\pi_{i+j,l-1}\big)\\
&\qquad + \sum_{j=1}^{k} \alpha_j \bigg(Y^\pi_{i+j,l} \frac{\Delta_{h_l} W^\top_{i,i+j}}{h_l} - Y^\pi_{i+j,l-1} \frac{\Delta_{h_{l-1}} W^\top_{i,i+j}}{h_{l-1}}\bigg)\Bigg]\\
&= -\sum_{j=1}^{k} \beta_j \mathbb{E}_i\big[Y^\pi_{i+j,\widehat{L}}\big] + \sum_{j=1}^{k} \alpha_j \mathbb{E}_i\Bigg[Y^\pi_{i+j,\widehat{L}} \frac{\Delta_{h_{\widehat{L}}} W^\top_{i,i+j}}{h_{\widehat{L}}}\Bigg].
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\mathbb{E}\Big[\big(Y_{t_i} - \mathbb{E}_i\big[P^y_i\big]\big)^2 + h_{\widehat{L}}\big(Z_{t_i} - \mathbb{E}_i\big[P^z_i\big]\big)^2\Big]\\
&= \mathbb{E}\Big[\big(Y_{t_i} - Y^\pi_{i,\widehat{L}}\big)^2 + h_{\widehat{L}}\big(Z_{t_i} - Z^\pi_{i,\widehat{L}}\big)^2\Big] \le C h^{2k}_{\widehat{L}},
\end{aligned}
\tag{5.7}
$$

where the inequality can be verified via Theorem 4.1,

$$
Y^\pi_{i,l} = -\sum_{j=1}^{k} a_j \mathbb{E}_i\big[Y^\pi_{i+j,l}\big] - \sum_{j=1}^{k} b_j h_l \mathbb{E}_i\big[f^\pi_{i+j,l}\big],
$$

$$Z_{i,l}^{\pi}=-\sum_{j=1}^{k}\beta_j\mathbb{E}_i\big[Y_{i+j,l}^{\pi}\big]+\sum_{j=1}^{k}\alpha_j\mathbb{E}_i\bigg[Y_{i+j,l}^{\pi}\frac{\Delta_{h_l}W_{i,i+j}^{\top}}{h_l}\bigg]$$

for $l=1,2,\ldots,\widehat{L}$. By the definitions of $P_{i,l}^{y}$ and $P_{i,l}^{z}$, we deduce, for $l=1,2,\ldots,\widehat{L}$,

$$\begin{aligned}\mathbb{V}_i\big[P_{i,l}^{y}\big]&=\frac{1}{N_l^2}\sum_{n=1}^{N_l}\mathbb{V}_i\bigg[-\sum_{j=1}^{k}a_j\big(Y_{i+j,l}^{\pi,(n)}-Y_{i+j,l-1}^{\pi,(n)}\big)-\sum_{j=1}^{k}b_j\big(h_lf_{i+j,l}^{\pi,(n)}-h_{l-1}f_{i+j,l-1}^{\pi,(n)}\big)\bigg]\\&=\frac{1}{N_l}\mathbb{V}_i\bigg[-\sum_{j=1}^{k}a_j\big(Y_{i+j,l}^{\pi}-Y_{i+j,l-1}^{\pi}\big)-\sum_{j=1}^{k}b_j\big(h_lf_{i+j,l}^{\pi}-h_{l-1}f_{i+j,l-1}^{\pi}\big)\bigg],\end{aligned}\tag{5.8}$$

$$\begin{aligned}\mathbb{V}_i\big[P_{i,l}^{z}\big]&=\frac{1}{N_l^2}\sum_{n=1}^{N_l}\mathbb{V}_i\bigg[-\sum_{j=1}^{k}\beta_j\big(Z_{i+j,l}^{\pi,(n)}-Z_{i+j,l-1}^{\pi,(n)}\big)\\&\qquad+\sum_{j=1}^{k}\alpha_j\bigg(Y_{i+j,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+j}^{\top}}{h_l}-Y_{i+j,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+j}^{\top}}{h_{l-1}}\bigg)\bigg]\\&=\frac{1}{N_l}\mathbb{V}_i\bigg[-\sum_{j=1}^{k}\beta_j\big(Z_{i+j,l}^{\pi}-Z_{i+j,l-1}^{\pi}\big)\\&\qquad+\sum_{j=1}^{k}\alpha_j\bigg(Y_{i+j,l}^{\pi}\frac{\Delta_{h_l}W_{i,i+j}^{\top}}{h_l}-Y_{i+j,l-1}^{\pi}\frac{\Delta_{h_{l-1}}W_{i,i+j}^{\top}}{h_{l-1}}\bigg)\bigg].\end{aligned}\tag{5.9}$$

Furthermore,

$$\begin{aligned}&\mathbb{V}_i\bigg[-\sum_{j=1}^{k}a_j\big(Y_{i+j,l}^{\pi}-Y_{i+j,l-1}^{\pi}\big)-\sum_{j=1}^{k}b_j\big(h_lf_{i+j,l}^{\pi}-h_{l-1}f_{i+j,l-1}^{\pi}\big)\bigg]\\&\le 2\mathbb{V}_i\bigg[-\sum_{j=1}^{k}a_jY_{i+j,l}^{\pi}-\sum_{j=1}^{k}b_jh_lf_{i+j,l}^{\pi}-Y_{t_i}\bigg]\\&\quad+2\mathbb{V}_i\bigg[Y_{t_i}+\sum_{j=1}^{k}a_jY_{i+j,l-1}^{\pi}+\sum_{j=1}^{k}b_jh_{l-1}f_{i+j,l-1}^{\pi}\bigg]\\&=2\mathbb{V}_i\big[Y_{i,l}^{\pi}-Y_{t_i}\big]+2\mathbb{V}\big[Y_{t_i}-Y_{i,l-1}^{\pi}\big],\end{aligned}\tag{5.10}$$

$$\begin{aligned}&\mathbb{V}_i\bigg[-\sum_{j=1}^{k}\beta_j\big(Z_{i+j,l}^{\pi}-Z_{i+j,l-1}^{\pi}\big)+\sum_{j=1}^{k}\alpha_j\bigg(Y_{i+j,l}^{\pi}\frac{\Delta_{h_l}W_{i,i+j}^{\top}}{h_l}-Y_{i+j,l-1}^{\pi}\frac{\Delta_{h_{l-1}}W_{i,i+j}^{\top}}{h_{l-1}}\bigg)\bigg]\\&\le 2\mathbb{V}_i\bigg[-\sum_{j=1}^{k}\beta_jZ_{i+j,l}^{\pi}+\sum_{j=1}^{k}\alpha_jY_{i+j,l}^{\pi}\frac{\Delta_{h_l}W_{i,i+j}^{\top}}{h_l}-Z_{t_i}\bigg]\\&\quad+2\mathbb{V}_i\bigg[Z_{t_i}+\sum_{j=1}^{k}\beta_jZ_{i+j,l-1}^{\pi}-\sum_{j=1}^{k}\alpha_jY_{i+j,l-1}^{\pi}\frac{\Delta_{h_{l-1}}W_{i,i+j}^{\top}}{h_{l-1}}\bigg]\\&=2\mathbb{V}_i\big[Z_{i,l}^{\pi}-Z_{t_i}\big]+2\mathbb{V}_i\big[Z_{t_i}-Z_{i,l-1}^{\pi}\big].\end{aligned}\tag{5.11}$$

From (5.8)-(5.11) and Theorem 4.1, we have, for $l=1,2,\ldots,\widehat{L}$,

$$\mathbb{V}_i\big[P_{i,l}^{y}\big]+h_l\mathbb{V}_i\big[P_{i,l}^{z}\big]\le\frac{C}{N_l}\big(h_l^{2k}+h_{l-1}^{2k}\big).$$

Analogously, one has

$$\mathbb{V}_i\big[P^y_{i,0}\big]+\mathbb{V}_i\big[P^z_{i,0}\big]\le\frac{C}{N_0}\big(h_0^{2k}+\mathbb{V}_i(Y_{t_i})+\mathbb{V}_i(Z_{t_i})\big)\le\frac{C}{N_0}h_0^{2k}\left(1+\frac{\mathbb{V}_i(Y_{t_i})+\mathbb{V}_i(Z_{t_i})}{T^{2k}/N^{2k}}\right).$$

Therefore,

$$\mathbb{V}_i\big[P^y_i\big]+h_{\widehat{L}}\mathbb{V}_i\big[P^z_i\big]\le C\sum_{l=0}^{\widehat{L}}\frac{h_l^{2k}}{N_l}. \tag{5.12}$$

Combined with (5.7) and (5.12), (5.5) can be rewritten as

$$\mathbb{E}_i\Big[\big(Y_{t_i}-P^y_i\big)^2+h_{\widehat{L}}\big(Z_{t_i}-P^z_i\big)^2\Big]\le C\left(h_{\widehat{L}}^{2k}+\sum_{l=0}^{\widehat{L}}\frac{h_l^{2k}}{N_l}\right). \tag{5.13}$$

Step 2. Under the condition

$$\max_{0\le i\le N}\mathbb{E}\Big[\big|Y_{t_i}-P^y_i\big|^2+h_{\widehat{L}}\big|Z_{t_i}-P^z_i\big|^2\Big]\le\epsilon^2,$$

we minimise the computational complexity of the multilevel estimator

$$\mathcal{C}\le c_1\sum_{l=0}^{\widehat{L}}N_lh_l^{-\kappa}.$$

Obviously, this is an optimization problem with constraints. Naturally, we consider the Lagrange function

$$g\big(N_0,N_1,\cdots,N_{\widehat{L}},\mu^2\big)\equiv c_1\sum_{l=0}^{\widehat{L}}N_lh_l^{-\kappa}+\mu^2\left(Ch_{\widehat{L}}^{2k}+C\sum_{l=0}^{\widehat{L}}N_l^{-1}h_l^{2k}-\epsilon^2\right).$$

Taking derivatives with respect to N_l, for $0\le l\le\widehat{L}$, we have

$$\frac{\partial g}{\partial N_l}=c_1h_l^{-\kappa}-C\mu^2h_l^{2k}N_l^{-2}.$$

Let $\partial g/\partial N_l=0$ for $0\le l\le\widehat{L}$. We obtain

$$N_l=\mu\sqrt{\frac{C}{c_1}}h_l^{\frac{1}{2}(2k+\kappa)}. \tag{5.14}$$

Furthermore, the Lagrange multiplier is

$$\mu=\sqrt{c_1C}\left(\frac{T}{N}\right)^{\frac{1}{2}(2k-\kappa)}\frac{1-M^{(1+\widehat{L})(\kappa-2k)/2}}{1-M^{(\kappa-2k)/2}}\big(\epsilon^2-Ch_{\widehat{L}}^{2k}\big)^{-1}. \tag{5.15}$$

Note that for any such choice of $N_0,N_1,\ldots,N_{\widehat{L}}$, the error $\max_{0\le i\le N}\mathbb{E}[|Y_{t_i}-P^y_i|^2+|Z_{t_i}-P^z_i|^2]$ is really bounded by ϵ^2. For fixed $\widehat{L}$, the computational complexity of the multilevel estimator is provided as

$$\mathcal{C}(\widehat{L})=\frac{c_1C}{\epsilon^2-Ch_{\widehat{L}}^{2k}}\left(\frac{T}{N}\right)^{2k-\kappa}\left(\frac{1-M^{(1+\widehat{L})(\kappa-2k)/2}}{1-M^{(\kappa-2k)/2}}\right)^2.$$

Now, the optimal integer-valued choice of $\widehat{L}$ would be the arg-min of the computational complexity estimate corresponding to the above choices of N_l for $l = 0, 1, \ldots, \widehat{L}$, namely

$$\widehat{L} = \left\lceil \frac{1}{2k} \log_M \left(C\theta^{-1}\epsilon^{-2} \left(\frac{T}{N}\right)^{2k} \right) \right\rceil, \tag{5.16}$$

where $\theta \in (0,1)$; the notation $\lceil x \rceil$ denotes the unique integer n satisfying the constraint $x \leq n < x+1$. Furthermore,

$$\begin{aligned} N_l = \Bigg\lceil &\frac{1}{1-\theta}\epsilon^{-2}C\left(\frac{T}{N}\right)^{2k} M^{-\frac{l}{2}(2k+\kappa)} \\ &\times \frac{1 - C^{(\kappa-2k)/(4k)}\theta^{-(\kappa-2k)/(4k)}\epsilon^{-(\kappa-2k)/(2k)}(TM/N)^{(\kappa-2k)/2}}{1-M^{(\kappa-2k)/2}} \Bigg\rceil. \end{aligned} \tag{5.17}$$

Thus, we deduce the computational complexity expression

$$\begin{aligned} \mathcal{C} &\leq c_1 \sum_{l=0}^{\widehat{L}} N_l h_l^{-\kappa} \\ &\leq \frac{c_1}{1-\theta}\epsilon^{-2}C\left(\frac{T}{N}\right)^{\frac{2k-\kappa}{2}} \\ &\quad \times \frac{1 - C^{(\kappa-2k)/(4k)}\theta^{-(\kappa-2k)/(4k)}\epsilon^{-(\kappa-2k)/(2k)}(TM/N)^{(\kappa-2k)/2}}{1-M^{(\kappa-2k)/2}} \sum_{l=0}^{\widehat{L}} h_l^{\frac{2k-\kappa}{2}} + c_1 \sum_{l=0}^{\widehat{L}} h_l^{-\kappa} \\ &= \frac{c_1 C}{1-\theta}\epsilon^{-2}\left(\frac{T}{N}\right)^{2k-\kappa} \\ &\quad \times \frac{1 - C^{(\kappa-2k)/(4k)}\theta^{-(\kappa-2k)/(4k)}\epsilon^{-(\kappa-2k)/(2k)}(TM/N)^{(\kappa-2k)/2}}{1-M^{(\kappa-2k)/2}} \frac{1-M^{(1+\widehat{L})(\kappa-2k)/2}}{1-M^{(\kappa-2k)/2}} \\ &\quad + c_1\left(\frac{T}{N}\right)^{-\kappa} \frac{1-M^{(1+\widehat{L})\kappa}}{1-M^{\kappa}} \\ &\leq \frac{c_1 C}{1-\theta}\epsilon^{-2}\left(\frac{T}{N}\right)^{2k-\kappa} \frac{(1 - C^{(\kappa-2k)/(4k)}\theta^{-(\kappa-2k)/(4k)}\epsilon^{-(\kappa-2k)/(2k)}(TM^2/N)^{(\kappa-2k)/2})^2}{(1-M^{(\kappa-2k)/2})^2} \\ &\quad + c_1\left(\frac{T}{N}\right)^{-\kappa} \frac{1 - C^{\kappa/(2k)}\theta^{-\kappa/(2k)}\epsilon^{-\kappa/k}(TM^2/N)^{\kappa}}{1-M^{\kappa}} \\ &= c_2\epsilon^{-2} + c_3\epsilon^{-1-\frac{\kappa}{2k}} + c_4\epsilon^{\frac{-\kappa}{k}} + c_5, \end{aligned} \tag{5.18}$$

where

$$\begin{aligned} c_2 &= \frac{c_1 C T^{2k-\kappa}}{(1-\theta)N^{2k-\kappa}(1-M^{(\kappa-2k)/2})^2}, \\ c_3 &= -\frac{2c_1 T^{k-\kappa/2} C^{(\kappa+2k)/(4k)}\theta^{-(\kappa-2k)/(4k)} M^{\kappa-2k}}{(1-\theta)N^{k-\kappa/2}(1-M^{(\kappa-2k)/2})^2}, \\ c_4 &= \frac{c_1 C^{\kappa/(2k)}\theta^{-(\kappa-2k)/(2k)} M^{2\kappa-4k}}{(1-\theta)(1-M^{(\kappa-2k)/2})^2} - \frac{c_1 C^{\kappa/(2k)}\theta^{-\kappa/(2k)} M^{2\kappa}}{1-M^{\kappa}}, \\ c_5 &= \frac{c_1 T^{-\kappa}}{N^{-\kappa}(1-M^{\kappa})}. \end{aligned}$$

The proof is complete. □

Remark 5.1. From Theorem 4.1, we have that the convergence order of the proposed probabilistic numerical method is k; we obtain that the convergence order of the conditional variance on each level is k by Theorem 5.1. From the condition the computational complexity on each level satisfy $\mathcal{C}_l = \mathcal{O}(N_l/h_l^\kappa)$ in Theorem 5.1. Thus, we deduce that the convergence order of the computational complexity on each level is κ. Obviously, $k \geq \min(k,\kappa)/2$. From the [16, Theorem 2.1], we have the computational complexity

$$\mathcal{C} \leq \begin{cases} c_2\epsilon^{-2}, & 2k > \kappa, \\ (c_2+c_3+c_4)\epsilon^{-2}, & 2k = \kappa, \\ c_4\epsilon^{\frac{-\kappa}{k}}, & 2k < \kappa. \end{cases}$$

In other words, our result (Theorem 5.1) aligns with Giles' standard complexity theorem.

6. Numerical Experiments

In this section, we will test the accuracy and the convergence rate of our scheme with respect to the time step h; we will also test the computational complexity of our scheme with respect to the accuracy ϵ. For these two aims, we provide two numerical examples to show the performance of our methods.

The error between the numerical solution and the analytical solution of the FBSDEs at the time $t=0$ is denoted by $|Y_0 - \widehat{Y}_0^\pi|$ and $|Z_0 - \widehat{Z}_0^\pi|$. In the tables, the notation CR represents the convergence rate with respect to the time step size which is computed by means of the least square method; the notation CC denotes the computational complexity; the notation CCS1, CCS2 and CCS3 respectively denote the computational complexity of Scheme 1, computational complexity of Scheme 2 and computational complexity of Scheme 3. Scheme 1, Scheme 2 and Scheme 3 will be given soon. Note that we just exhibit the cases of the step number $k=1,2,3$ in the following two examples. We also make the step number $k=4$ or bigger if we want to. The accuracy of the numerical solutions becomes higher as the step number increases.

In order to obtain the numerical solutions of the FBSDEs, the numerical schemes need to be provided with known parameters by means of (4.6) and (4.11). For the SDE in (1.1), given the initial value X_0, we solve X_i^π for $i=1,\ldots,N$ by the explicit Itô-Taylor schemes [28]. X_i^π denotes the discretization form of X at t_i based on the time grid π. For the BSDE in (1.1), we discuss the numerical schemes according to the following three cases:

1. If $k=1$, let Scheme 1 denote, for $i=N-1,N-2,\ldots,0$,

$$\begin{aligned}\widehat{Y}_i^\pi &= \frac{1}{N_0}\sum_{n=1}^{N_0}\left(\widehat{Y}_{i+1,0}^{\pi,(n)} + \frac{h_0}{2}\left(\widehat{f}_{i,0}^{\pi,(n)} + \widehat{f}_{i+1,0}^{\pi,(n)}\right)\right) \\ &\quad + \sum_{l=1}^{\widehat{L}}\frac{1}{N_l}\sum_{n=1}^{N_l}\left[\left(\widehat{Y}_{i+1,l}^{\pi,(n)} - \widehat{Y}_{i+1,l-1}^{\pi,(n)}\right) + \sum_{j=0}^{1}\frac{1}{2}\left(h_l\widehat{f}_{i+j,l}^{\pi,(n)} - h_{l-1}\widehat{f}_{i+j,l-1}^{\pi,(n)}\right)\right], \\ \widehat{Z}_i^\pi &= \frac{1}{N_0}\sum_{n=1}^{N_0}\left(\frac{1}{2}\widehat{Y}_{i+1,0}^{\pi,(n)}\frac{\Delta_{h_0}^{(n)}W_{i,i+1}^\top}{h_0} + \frac{1}{2}\widehat{Z}_{i+1,0}^{\pi,(n)}\right) \\ &\quad + \sum_{l=1}^{\widehat{L}}\frac{1}{N_l}\sum_{n=1}^{N_l}\left[\frac{1}{2}\left(\widehat{Y}_{i+1,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+1}^\top}{h_l} - \widehat{Y}_{i+1,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+1}^\top}{h_{l-1}}\right) + \frac{1}{2}\left(\widehat{Z}_{i+1,l}^{\pi,(n)} - \widehat{Z}_{i+1,l-1}^{\pi,(n)}\right)\right].\end{aligned}$$

2. If $k=2$, let Scheme 2 denote, for $i=N-2,N-3,\dots,0$,

$$
\begin{aligned}
\widehat{Y}_i^\pi &= \frac{1}{N_0}\sum_{n=1}^{N_0}\left(\widehat{Y}_{i+2,0}^{\pi,(n)}+\frac{h_0}{3}\widehat{f}_{i,0}^{\pi,(n)}+\frac{4h_0}{3}\widehat{f}_{i+1,0}^{\pi,(n)}+\frac{h_0}{3}\widehat{f}_{i+2,0}^{\pi,(n)}\right)\\
&\quad+\sum_{l=1}^{\widehat{L}}\frac{1}{N_l}\sum_{n=1}^{N_l}\Bigg[\left(\widehat{Y}_{i+2,l}^{\pi,(n)}-\widehat{Y}_{i+2,l-1}^{\pi,(n)}\right)+\frac{1}{3}\left(h_l\widehat{f}_{i,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i,l-1}^{\pi,(n)}\right)\\
&\qquad+\frac{4}{3}\left(h_l\widehat{f}_{i+1,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i+1,l-1}^{\pi,(n)}\right)+\frac{1}{3}\left(h_l\widehat{f}_{i+2,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i+2,l-1}^{\pi,(n)}\right)\Bigg],\\
\widehat{Z}_i^\pi &= \frac{1}{N_0}\sum_{n=1}^{N_0}\left(\widehat{Y}_{i+1,0}^{\pi,(n)}\frac{\Delta_{h_0}^{(n)}W_{i,i+1}^\top}{h_0}-\frac{3}{8}\widehat{Y}_{i+2,0}^{\pi,(n)}\frac{\Delta_{h_0}^{(n)}W_{i,i+2}^\top}{h_0}+\widehat{Z}_{i+1,0}^{\pi,(n)}-\frac{1}{4}\widehat{Z}_{i+2,0}^{\pi,(n)}\right)\\
&\quad+\sum_{l=1}^{\widehat{L}}\frac{1}{N_l}\sum_{n=1}^{N_l}\Bigg[\left(\widehat{Y}_{i+1,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+1}^\top}{h_l}-\widehat{Y}_{i+1,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+1}^\top}{h_{l-1}}\right)\\
&\qquad-\frac{3}{8}\left(\widehat{Y}_{i+2,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+2}^\top}{h_l}-\widehat{Y}_{i+2,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+2}^\top}{h_{l-1}}\right)\\
&\qquad+\left(\widehat{Z}_{i+1,l}^{\pi,(n)}-\widehat{Z}_{i+1,l-1}^{\pi,(n)}\right)-\frac{1}{4}\left(\widehat{Z}_{i+2,l}^{\pi,(n)}-\widehat{Z}_{i+2,l-1}^{\pi,(n)}\right)\Bigg].
\end{aligned}
$$

3. If $k=3$, let Scheme 3 denote, for $i=N-3,N-4,\dots,0$,

$$
\begin{aligned}
\widehat{Y}_i^\pi &= \frac{1}{N_0}\sum_{n=1}^{N_0}\Bigg(\frac{1}{3}\widehat{Y}_{i+1,0}^{\pi,(n)}+\frac{1}{3}\widehat{Y}_{i+2,0}^{\pi,(n)}+\frac{1}{3}\widehat{Y}_{i+3,0}^{\pi,(n)}+\frac{13h_0}{36}\widehat{f}_{i,0}^{\pi,(n)}\\
&\qquad+\frac{13h_0}{12}\widehat{f}_{i+1,0}^{\pi,(n)}+\frac{5h_0}{12}\widehat{f}_{i+2,0}^{\pi,(n)}+\frac{5h_0}{36}\widehat{f}_{i+3,0}^{\pi,(n)}\Bigg)\\
&\quad+\sum_{l=1}^{\widehat{L}}\frac{1}{N_l}\sum_{n=1}^{N_l}\Bigg[\sum_{j=1}^{3}\frac{1}{3}\left(\widehat{Y}_{i+j,l}^{\pi,(n)}-\widehat{Y}_{i+j,l-1}^{\pi,(n)}\right)+\frac{13}{36}\left(h_l\widehat{f}_{i,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i,l-1}^{\pi,(n)}\right)\\
&\qquad+\frac{13}{12}\left(h_l\widehat{f}_{i+1,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i+1,l-1}^{\pi,(n)}\right)+\frac{5}{12}\left(h_l\widehat{f}_{i+2,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i+2,l-1}^{\pi,(n)}\right)\\
&\qquad+\frac{5}{36}\left(h_l\widehat{f}_{i+3,l}^{\pi,(n)}-h_{l-1}\widehat{f}_{i+3,l-1}^{\pi,(n)}\right)\Bigg],\\
\widehat{Z}_i^\pi &= \frac{1}{N_0}\sum_{n=1}^{N_0}\Bigg(\frac{7}{6}\widehat{Y}_{i+1,0}^{\pi,(n)}\frac{\Delta_{h_0}^{(n)}W_{i,i+1}^\top}{h_0}-\widehat{Y}_{i+2,0}^{\pi,(n)}\frac{\Delta_{h_0}^{(n)}W_{i,i+2}^\top}{h_0}+\frac{5}{18}\widehat{Y}_{i+3,0}^{\pi,(n)}\frac{\Delta_{h_0}^{(n)}W_{i,i+3}^\top}{h_0}\\
&\qquad+\frac{11}{6}\widehat{Z}_{i+1,0}^{\pi,(n)}-\widehat{Z}_{i+2,0}^{\pi,(n)}+\frac{1}{6}\widehat{Z}_{i+3,0}^{\pi,(n)}\Bigg)\\
&\quad+\sum_{l=1}^{\widehat{L}}\frac{1}{N_l}\sum_{n=1}^{N_l}\Bigg[\frac{7}{6}\left(\widehat{Y}_{i+1,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+1}^\top}{h_l}-\widehat{Y}_{i+1,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+1}^\top}{h_{l-1}}\right)\\
&\qquad-\left(\widehat{Y}_{i+2,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+2}^\top}{h_l}-\widehat{Y}_{i+2,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+2}^\top}{h_{l-1}}\right)\\
&\qquad+\frac{5}{18}\left(\widehat{Y}_{i+3,l}^{\pi,(n)}\frac{\Delta_{h_l}^{(n)}W_{i,i+3}^\top}{h_l}-\widehat{Y}_{i+3,l-1}^{\pi,(n)}\frac{\Delta_{h_{l-1}}^{(n)}W_{i,i+3}^\top}{h_{l-1}}\right)
\end{aligned}
$$

$$+\frac{11}{6}\left(\widehat{Z}^{\pi,(n)}_{i+1,l}-\widehat{Z}^{\pi,(n)}_{i+1,l-1}\right)-\left(\widehat{Z}^{\pi,(n)}_{i+2,l}-\widehat{Z}^{\pi,(n)}_{i+2,l-1}\right)$$
$$+\frac{1}{6}\left(\widehat{Z}^{\pi,(n)}_{i+3,l}-\widehat{Z}^{\pi,(n)}_{i+3,l-1}\right)\Bigg].$$

In what follows, we apply our algorithms to two FBSDEs with closed-form solutions.

Example 6.1. Consider the BSDE

$$Y_t=\frac{\exp\left(T+\sum_{k=1}^q W_k\right)}{1+\exp\left(T+\sum_{k=1}^q W_k\right)}+\int_t^T\left(\sum_{k=1}^q Z_{k,s}\right)\left(Y_s-\frac{2+q}{q}\right)ds-\int_t^T Z_s dW_s \tag{6.1}$$

with the analytic solutions

$$Y_t=\frac{\exp(t+\sum_{k=1}^q W_k)}{1+\exp(t+\sum_{k=1}^q W_k)},\quad Z_{k,t}=\frac{\exp(t+\sum_{k=1}^q W_k)}{\left(1+\exp(t+\sum_{k=1}^q W_k)\right)^2}, \tag{6.2}$$

where $Z_{k,t}$ denotes the value of the component k-th of the Z at t. Let $T=1, q=9, N=2$. Table 6.1 shows the relationship between the absolute errors of Y, Z at the time $t=0$ and the time step $h_{\widehat{L}}$ (both in log scales) under the Scheme 1, Scheme 2 and Scheme 3. Table 6.1 indicates that

(i) the absolute errors of Y and Z calculated by means of Scheme 3 are the smallest and Scheme 1 are the largest;

(ii) the absolute errors of Y and Z become smaller with the time step $h_{\widehat{L}}$ decreasing whether Scheme 1, Scheme 2 or Scheme 3 is applied;

(iii) if the time step $h_{\widehat{L}}$ is fixed, the absolute errors of Z are bigger than those of Y when we approximate them via the same scheme. This coincides with the previous analyses.

Table 6.2 shows relationship between the computational complexity of Scheme 1, Scheme 2 and Scheme 3 and the accuracy ϵ (both in log scales). The computational complexity of the Scheme 1, Scheme 2 and Scheme 3 coincide with the theory that is proportional to ϵ^2.

Example 6.2. We consider the FBSDE (taken from [44]):

$$\begin{cases} X_t=\int_0^t \sin(s+X_s)ds+\int_0^t \frac{3}{10}\cos(s+X_s)dW_s, \\ Y_t=\sin(T+X_T)+\int_t^T\left(\frac{3}{20}Y_sZ_s-\cos(s+X_s)(1+Y_s)\right)ds-\int_t^T Z_s dW_s. \end{cases} \tag{6.3}$$

Table 6.1: Errors and convergence rates of Scheme 1, Scheme 2 and Scheme 3 for Example 6.1.

Scheme	Errors \ $\log_2(h_{\widehat{L}})$	-2	-3	-4	-5	-6	-7	-8	CR
Scheme 1	$\log_2(\lvert Y_0-\widehat{Y}_0^\pi\rvert)$	-2.1112	-2.8306	-3.9454	-5.5098	-6.0694	-6.9285	-8.3312	1.0367
	$\log_2(\lvert Z_0-\widehat{Z}_0^\pi\rvert)$	-1.3912	-2.2406	-3.0454	-3.8098	-4.4802	-5.2285	-5.9712	0.7467
Scheme 2	$\log_2(\lvert Y_0-\widehat{Y}_0^\pi\rvert)$	-4.4612	-6.8212	-8.9900	-11.3740	-13.4254	-15.8609	-17.8690	2.2376
	$\log_2(\lvert Z_0-\widehat{Z}_0^\pi\rvert)$	-2.7603	-3.9914	-5.7601	-7.0554	-8.6657	-9.7516	-11.5237	1.4106
Scheme 3	$\log_2(\lvert Y_0-\widehat{Y}_0^\pi\rvert)$	-6.2552	-9.1822	-12.5104	-15.5637	-18.7656	-21.8632	-25.0286	3.1276
	$\log_2(\lvert Z_0-\widehat{Z}_0^\pi\rvert)$	-5.0612	-8.1414	-10.9624	-13.5578	-16.6636	-19.0106	-20.9737	2.7156

Table 6.2: Computational complexities of Scheme 1, Scheme 2 and Scheme 3 for Example 6.1.

CC \ $\log_2(\epsilon)$	-2	-3	-4	-5	-6	-7	-8
$\log_2(\text{CCS1})$	4.4452	6.8678	8.9905	11.1131	13.3357	15.4386	18.1617
$\log_2(\text{CCS2})$	4.0976	6.1465	8.6953	10.2442	12.2930	14.0418	16.7907
$\log_2(\text{CCS3})$	3.9699	5.9548	7.9398	9.9248	11.9097	13.8947	15.8798

Table 6.3: Errors and convergence rates of Scheme 1, Scheme 2 and Scheme 3 for Example 6.2.

Scheme	Errors \ $\log_2(h_{\widehat{L}})$	-2	-3	-4	-5	-6	-7	-8	CR
Scheme 1	$\log_2(\lvert Y_0 - \widehat{Y}_0^\pi \rvert)$	-2.4466	-3.3399	-4.7532	-6.2665	-7.3798	-8.4931	-9.8064	1.2133
	$\log_2(\lvert Z_0 - \widehat{Z}_0^\pi \rvert)$	-1.3928	-1.8892	-2.2856	-3.092	-3.6784	-4.1008	-4.8712	0.5964
Scheme 2	$\log_2(\lvert Y_0 - \widehat{Y}_0^\pi \rvert)$	-4.0204	-6.1081	-8.0518	-10.0635	-12.1362	-14.0089	-16.0016	2.0127
	$\log_2(\lvert Z_0 - \widehat{Z}_0^\pi \rvert)$	-2.4288	-3.7432	-5.2006	-6.3720	-7.9864	-9.4218	-10.6152	1.3144
Scheme 3	$\log_2(\lvert Y_0 - \widehat{Y}_0^\pi \rvert)$	-5.7072	-8.6848	-11.7744	-14.4681	-17.3616	-20.0552	-22.7488	2.8936
	$\log_2(\lvert Z_0 - \widehat{Z}_0^\pi \rvert)$	-4.6886	-7.3329	-9.4772	-11.6215	-14.1658	-16.4901	-18.4544	2.3443

Table 6.4: Computational complexities of Scheme 1, Scheme 2 and Scheme 3 for Example 6.2.

CC \ $\log_2(\epsilon)$	-2	-3	-4	-5	-6	-7	-8
$\log_2(\text{CCS1})$	4.4451	6.6676	8.8902	11.1127	13.3353	15.5578	17.7804
$\log_2(\text{CCS2})$	3.8612	5.8012	7.7461	9.6826	11.6191	13.8609	15.4922
$\log_2(\text{CCS3})$	3.7334	5.6001	7.4668	9.3335	11.2002	13.0669	14.9336

From the Itô formula, we verify that the analytic solutions of the FBSDE (6.3) can be represented in the following form:

$$Y_t = \sin(t + X_t), \quad Z_t = \frac{3}{10}\cos^2(t + X_t). \tag{6.4}$$

Let $T = 1, N = 2$. Tables 6.3 and 6.4 show the errors between the numerical solution and the analytical solution of the FBSDE (6.3) and the computational complexity via the Scheme 1, Scheme 2 and Scheme 3. Table 6.3 compares the errors in terms of the absolute errors of Y, Z at the time $t = 0$ and the time step $h_{\widehat{L}}$ (both in log scales) by the Scheme 1, Scheme 2 and Scheme 3. Table 6.4 compares the computational complexities in terms of the Scheme 1, Scheme 2 and Scheme 3 and the accuracy ϵ (both in log scales).

7. Conclusion

In this article, a high order probabilistic numerical scheme is constructed to approximate the solutions of the FBSDEs (1.1) by utilizing the multilevel Monte Carlo method, the numerical differentiation method, the Itô-Taylor expansion and so on. Then, we demonstrate the high order property under the Dahlquist's root condition. Conditional expectations involved in the resulting probabilistic numerical scheme are approximated by the multilevel Monte Carlo method.

Compared with Monte Carlo method, in the same time-discretisation numerical scheme, the multilevel Monte Carlo method reduces the computational complexity of our method to the square of prescribed accuracy. In a word, we design a new high order fast probabilistic numerical algorithm for FBSDEs and our method not only enriches the probabilistic numerical schemes for FBSDEs but also provides valuable insights for developing high order fast method for FBSDEs.

Acknowledgements. The authors would like to thank the editor and anonymous referees for their helpful comments and suggestions which lead to a considerably improved presentation.

This work is supported by the National Natural Science of China (Grant Nos. 12501664, 62173292) , by the Golden Phoenix of the Green City-Yang Zhou (Grant No. 137013391), by the National Natural Science of Yangzhou, China (Grant No. YZ2025145) and by the Yangzhou Innovation Capability Enhancement Fund through grant (Grant No. YZ2024245).

References

[1] C. Bender and R. Denk, A forward scheme for backward SDEs, *Stochastic Process. Appl.*, **117**:12 (2007), 1793–1812.

[2] C. Bender and J. Steiner, Least-squares Monte Carlo for backward SDEs, in: *Numerical Methods in Finance. Springer Proceedings in Mathematics*, Vol. 12, Springer, (2012), 257–289.

[3] B. Bouchard and N. Touzi, Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations, *Stochastic Process. Appl.*, **111**:2 (2004), 175–206.

[4] J.F. Chassagneux, Linear multistep schemes for BSDEs, *SIAM J. Numer. Anal.*, **52**:6 (2014), 2815–2836.

[5] K.W. Chau and C.W. Oosterlee, On the wavelet-based SWIFT method for backward stochastic differential equations, *IMA J. Numer. Anal.*, **38** (2018), 1051–1083.

[6] K.W. Chau and C.W. Oosterlee, Stochastic grid bundling method for backward stochastic differential equations, *Int. J. Comput. Math.*, **96** (2019), 2272–2301.

[7] D. Crisan and K. Manolarakis, Solving backward stochastic differential equations using the cubature method: Application to nonlinear pricing, *SIAM J. Financ. Math.*, **3**:1 (2012), 534–571.

[8] D. Crisan, K. Manolarakis, and N. Touzi, On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, *Stochastic Process. Appl.*, **120:7** (2010), 1133–1158.

[9] J. Douglas Jr., J. Ma, and P. Protter, Numerical methods for forward-backward stochastic differential equations, *Ann. Appl. Probab.*, **6**:3 (1996), 940–968.

[10] D. Duffie and P. Glynn, Efficient Monte Carlo simulation of security prices, *Ann. Appl. Probab.*, **5**:4 (1995), 897–905.

[11] W. Gautschi, *Numerical Analysis: An Introduction*, Birkhäuser, 1997.

[12] Y. Ge, L. Li, and G. Zhang, A Fourier transform method for solving backward stochastic differential equations, *Methodol. Comput. Appl. Probab.*, **24**:1 (2022), 385–412.

[13] C. Geiss, C. Labart, and A. Luoto, Mean square rate of convergence for random walk approximation of forward-backward SDEs, *Adv. Appl. Probab.*, **52**:3 (2020), 735–771.

[14] C. Geiss, C. Labart, and A. Luoto, Random walk approximation of BSDEs with Hölder continuous terminal condition, *Bernoulli*, **26**:1 (2020), 159–190.

[15] M. Giles, Multilevel Monte Carlo path simulation, *Oper. Res.*, **56**:3 (2008), 607–617.

[16] M. Giles, Multilevel Monte Carlo methods, *Acta Numer.*, **24** (2015), 259–328.

[17] E. Gobet and C. Labart, Error expansion for the discretization of backward stochastic differential equations, *Stochastic Process. Appl.*, **117**:7 (2007), 803–829.

[18] E. Gobet, J.G. López-Salas, and C. Vázquez, Quasi-regression Monte-Carlo scheme for semi-linear PDEs and BSDEs with large scale parallelization on GPUs, *Arch. Comput. Methods Eng.*, **27**:3

(2020), 889–921.

[19] E. Gobet and P. Turkedjiev, Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions, *Math. Comput.*, **85**:299 (2016), 1359–1391.

[20] Q. Han and S. Ji, A multi-step algorithm for BSDEs based on a predictor-corrector scheme and least-squares Monte Carlo, *Methodol. Comput. Appl. Probab.*, **24**:4 (2022), 2403–2426.

[21] Q. Han and S. Ji, Two-step scheme for backward stochastic differential equations, *J. Comput. Math.*, **41**:2 (2023), 287–304.

[22] Q. Han and S. Ji, Solving BSDEs based on novel multi-step schemes and multilevel Monte Carlo, *J. Comput. Appl. Math.*, **417** (2023), 114543.

[23] Q. Han and S. Ji, Novel multi-step predictor-corrector schemes for backward stochastic differential equations, *Commun. Nonlinear Sci. Numer. Simul.*, **139** (2024), 108269.

[24] Q. Han and Y. Liu, Tamed Euler interpolation multilevel Monte Carlo method for quadratic forward backward stochastic differential equations, *Math. Comput. Simulation*, **240** (2026), 20–37.

[25] M. Hu and L. Jiang, An efficient numerical method for forward-backward stochastic differential equations driven by G-Brownian motion, *Appl. Numer. Math.*, **165** (2021), 578–597.

[26] A. Khodadadian, M. Parvizi, M. Abbaszadeh, M. Dehghan, and C. Heitzinger, A multilevel Monte Carlo finite element method for the stochastic Cahn-Hilliard-Cook equation, *Comput. Mech.*, **64** (2019), 937–949.

[27] A. Khodadadian, M. Parvizi, and C. Heitzinger, An adaptive multilevel Monte Carlo algorithm for the stochastic drift-diffusion-Poisson system, *Comput. Methods Appl. Mech. Engrg.*, **368** (2020), 113163.

[28] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, in: *Stochastic Modelling and Applied Probability*, Vol. 23, Springer, 1992.

[29] X. Li, Y. Wu, Q. Zhu, S. Hu, and C. Qin, A regression-based Monte Carlo method to solve two-dimensional forward backward stochastic differential equations, *Adv. Differential Equations*, **2021** (2021), 207.

[30] E. Pardoux and S. Peng, Adapted solution of a backward stochastic differential equation, *Syst. Control Lett.*, **14**:1 (1990), 55–61.

[31] E. Pardoux and S. Peng, Backward stochastic differential equations and quasilinear parabolic partial differential equations, in: *Stochastic Partial Differential Equations and Their Applications. Lecture Notes in Control and Information Sciences*, Vol. 176, Springer, (1992), 200–217.

[32] E. Pardoux and S. Tang, Forward-backward stochastic differential equations and quasilinear parabolic PDEs, *Probab. Theory Relat. Field*, **114**:2 (1999), 123–150.

[33] S. Peng, Probabilistic interpretation for systems of quasilinear parabolic partial differential equations, *Stoch. Stoch. Rep.*, **37**:1-2 (1991), 61–74.

[34] S. Peng and M. Xu, Numerical algorithms for backward stochastic differential equations with 1-d Brownian motion: Convergence and simulations, *ESAIM Math. Model. Numer. Anal.*, **45**:2 (2011), 335–360.

[35] M.J. Ruijter and C.W. Oosterlee, A Fourier cosine method for an efficient computation of solutions to BSDEs, *SIAM J. Sci. Comput.*, **37**:2 (2015), A859–A889.

[36] D. Sun, G. Liang, and S. Tang, Quantitative stability and numerical analysis of Markovian quadratic BSDEs with reflection, *Probab. Uncertain. Quant. Risk*, **7**:1 (2022), 13–30.

[37] L. Taghizadeh, A. Khodadadian, and C. Heitzinger, The optimal multilevel Monte-Carlo approximation of the stochastic drift-diffusion-Poisson system, *Comput. Methods Appl. Mech. Engrg.*, **318** (2017), 739–761.

[38] X. Tang and J. Xiong, Stability analysis of general multistep methods for Markovian backward stochastic differential equations, *IMA J. Numer. Anal.*, **42**:2 (2022), 1789–1805.

[39] L. Teng, Gradient boosting-based numerical methods for high-dimensional backward stochastic differential equations, *Appl. Math. Comput.*, **426** (2022), 127119.

[40] M. Xu, Numerical algorithms and simulations for reflected backward stochastic differential equa-

tions with two continuous barriers, *J. Comput. Appl. Math.*, **236**:6 (2011), 1137–1154.

[41] C. Zhang, J. Wu, and W. Zhao, One-step multi-derivative methods for backward stochastic differential equations, *Numer. Math. Theory Methods Appl.*, **12**:4 (2019), 1213–1230.

[42] J. Zhang, A numerical scheme for BSDEs, *Ann. Appl. Probab.*, **14**:1 (2004), 459–488.

[43] W. Zhao, Y. Fu, and T. Zhou, New kinds of high-order multi-step schemes for forward backward stochastic differential equations, *SIAM J. Sci. Comput.*, **36**:4 (2014), A1731–A1751.

[44] W. Zhao, W. Zhang, and L. Ju, A multistep scheme for decoupled forward-backward stochastic differential equations, *Numer. Math. Theory Methods Appl.*, **9**:2 (2016), 262–288.

[45] X. Zhong and S. Liao, On the homotopy analysis method for backward/forward-backward stochastic differential equations, *Numer. Algorithms*, **76**:2 (2017), 487–519.