

Novel multi-step predictor-corrector schemes for backward stochastic differential equations

Qiang Han ^{*} Shaolin Ji[†]

February 12, 2021

Abstract. Novel multi-step predictor-corrector numerical schemes have been derived for approximating decoupled forward-backward stochastic differential equations (FBSDEs). The stability and high order rate of convergence of the schemes are rigorously proved. We also present a sufficient and necessary condition for the stability of the schemes. Numerical experiments are given to illustrate the stability and convergence rates of the proposed methods.

Key words. decoupled forward-backward stochastic differential equations, multi-step predictor-corrector schemes, high order discretization, stability

AMS subject classifications. 93E20, 60H10, 35K15

1 Introduction

To the best of our knowledge, the numerical algorithms for decoupled FBSDEs can be divided into two:

One branch explores the connection with partial differential equations (PDEs). To be specific, the solution (Y_t, Z_t) of the BSDE in (2.1) can be represented as $Y_t = u(t, x_t)$, $Z_t = \nabla_x u(t, x_t)\sigma(t, x_t)$, $t \in [0, T]$, $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$, and $u(t, x_t)$ is solution of the parabolic PDE

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

with the terminal condition $u(T, x) = \Phi(x)$. In turn, suppose (Y_t, Z_t) is the solution of the BSDE in (2.1). $u(t, x_t) = Y_t$ is a viscosity solution to the PDE. Thus, the numerical approximation of decoupled FBSDEs is to solve the corresponding parabolic PDEs numerically (see [12, 27, 28]). This algorithm may be limited due to high-dimensionality or lack of smoothness of the coefficients. For this issue, Weinan E et al propose the deep learning algorithm which can deal with 100-dimensional nonlinear PDEs (see [3, 13, 14, 21, 25]). Also,

^{*}Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, Shandong 250100, PR China. 201411272@mail.sdu.edu.cn.

[†]Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, Shandong 250100, PR China. jsl@sdu.edu.cn (Corresponding author). Research supported by National Natural Science of China No. 11971263; 11871458 and National Key R&D Program of China No.2018YFA0703900.

the branching diffusion method does not suffer from the curse of dimensionality (see [22]) and this method is extended to the non-Markovian case and the non-linearities case (see [24] and [23] respectively).

The second branch of algorithms can be implemented via a two-step procedure which consists of a time-discretization of decoupled FBSDEs and an approximation procedure for the conditional expectations. Specifically, if the Euler scheme (explicit, implicit or generalized) is utilized to discretize decoupled FBSDEs, the order of discretization error is $\frac{1}{2}$ and sometimes can reach 1 (see [1, 4, 6, 11, 16–18, 26, 35, 37]). To obtain high order accuracy scheme, authors in [38, 39] develop two kinds of multi-step schemes to solve decoupled FBSDEs. The Runge-Kutta schemes and linear multi-step schemes for approximating decoupled FBSDEs have been investigated in [10] and [8]. In this paper we extend the predictor-corrector type based on Adams schemes (see [9]) to the predictor-corrector type based on general linear multi-step schemes. We also provide an indicator for the local truncation error by utilizing the difference between the predicted and the corrected values at each time step (see Proposition 3.2). Furthermore, we present a sufficient and necessary condition for the stability of the general scheme (see Theorem 3.6). Finally, parameters in multi-step schemes are obtained by different methods. That is to say, the paper [38] adopts derivative approximation; papers [8, 9, 39] use Lagrange interpolating polynomials; and we utilize Itô-Taylor expansion.

From the above review, the time-discretization of decoupled FBSDEs can adopt low order schemes or high order schemes. Notice that there are a large number of documents about low order schemes and this implies that the theory of implementable numerical methods of decoupled FBSDEs is booming. Compared with the development of the numerical methods of ordinary differential equations (ODEs) and stochastic differential equations (SDEs), the investigation of high order accuracy schemes for decoupled FBSDEs is meaningful and necessary. Moreover, the analysis of Section 4.4 of [17] also maintains that development of high order accuracy schemes for decoupled FBSDEs is significant. Hence, for this motivation, we design an available high order accuracy scheme called the general multi-step predictor-corrector schemes (see (3.4)) (see [7] about SDEs which do not have the predictor term). And this kind of schemes possess the advantage of simple type of error estimates for decoupled FBSDEs.

The contributions of this paper are as follows.

First, we derive a novel high order scheme for decoupled FBSDEs. The advantage does not require the solution of an algebraic equation at each step. Therefore, this can reduce the complexity of calculation. Simultaneously, our schemes also inherit the virtues of implicit scheme. Second, the stability and high order property of the scheme (3.4) are rigorously proved. Note that we present a sufficient and necessary condition for the stability of the scheme (3.4). A property of predictor-corrector scheme (see Proposition 3.2) is established in the frame of decoupled FBSDEs. And this property provides an indicator for the local truncation error by utilizing the difference between the predicted and the corrected values at each time step. The high order property of the scheme (3.4) is also established.

The structure of this paper is as follows. In Section 2, we present some fundamental definitions, assumptions and lemmas that can be used in the following sections. Moreover the Adams schemes of decoupled FBSDEs are reviewed. We first construct the predictor-corrector schemes (3.4). Then, the stability and high order properties of scheme (3.4) are also found in Section 3. Section 4 presents numerical experiments to illustrate the stability and convergence rates of algorithms.

2 Preliminaries

In this section, we provide some preliminary results and recall the predictor-corrector scheme of decoupled FBSDEs based on Adams types.

2.1 Decoupled FBSDE

In this subsection, we review the decoupled FBSDE and the corresponding propositions.

Let $T > 0$ be a fixed terminal time and $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered complete probability space where $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ is the natural filtration of the standard d -dimensional Brownian motion. In the space $(\Omega, \mathcal{F}_T, \mathbb{F}, \mathbb{P})$, we consider discretizing the decoupled FBSDEs as below:

$$\begin{cases} X_t = x_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, & (SDE) \\ Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s, & (BSDE) \end{cases} \quad (2.1)$$

where $(X_s)_{t \leq s \leq T}$ is a d -dimensional diffusion process driven by the finite d -dimensional Brownian motion $(W_t)_{0 \leq t \leq T}$ which is defined in a filtered complete probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Set the σ -algebra $\mathcal{F}_{t,s} = \sigma\{W_r - W_t, t \leq r \leq s\}$, $\mathcal{F} = \mathcal{F}_{0,T}$. In addition, functions b, σ, Φ and f satisfy:

Assumption 1. There exists a non-negative constant L satisfying

$$|b(t, x_1) - b(t, x_2)| + |\sigma(t, x_1) - \sigma(t, x_2)| \leq L|x_1 - x_2|, \quad \forall x_1, x_2 \in \mathbb{R}^d.$$

Assumption 2. There exist non-negative constants C_f and L_f such that

- (i) $|f(t_1, x_1, y_1, z_1) - f(t_2, x_2, y_2, z_2)| \leq L_f(\sqrt{|t_1 - t_2|} + |x_1 - x_2| + |y_1 - y_2| + \|z_1 - z_2\|)$ for all $t_1, t_2 \in [0, T]$, $x_1, x_2 \in \mathbb{R}^d$, $y_1, y_2 \in \mathbb{R}$ and $z_1, z_2 \in \mathbb{R}^d$;
- (ii) $|f(t, x, 0, 0)| \leq C_f$ on $[0, T] \times \mathbb{R}^d$;
- (iii) Function Φ is measurable and bounded.

For readers' convenience, here we present two lemmas and adapt them to our context.

Lemma 2.1 (see [33]) *Assume that functions b, σ, f and Φ are uniformly Lipschitz with respect to (w.r.t.) (x, y, z) and $\frac{1}{2}$ -Hölder continuous w.r.t. t . In addition, assume Φ is of class $C_b^{2+\kappa}$ for some $\kappa \in (0, 1)$ and the matrix valued function $a = \sigma\sigma^\top = (a_{ij})$ is uniformly elliptic. Then the solution (Y_t, Z_t) of the BSDE in (2.1) can be represented as*

$$Y_t = u(t, X_t), \quad Z_t = \nabla_x u(t, X_t) \sigma(t, X_t), \quad t \in [0, T],$$

where $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ satisfies the parabolic PDE as below:

$$\mathcal{L}^{(0)} u(t, x) + f(t, x, u(t, x), \nabla_x u(t, x) \sigma(t, x)) = 0, \quad (2.2)$$

with the terminal condition $u(T, x) = \Phi(x)$ where $\mathcal{L}^{(0)} = \frac{\partial}{\partial t} + \sum_{i=1}^d b_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d (\sigma\sigma^\top)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$.

Lemma 2.2 (see Proposition 2.2 in [10]) Let $n \geq 0$. Then for a function $v \in \mathcal{A}_b^{n+1}$,

$$\mathbb{E}_t[v(t+h, X_{t+h})] = v_t + hv_t^{(0)} + \frac{h^2}{2}v_t^{(0,0)} + \cdots + \frac{h^n}{n!}v_t^{(0)^n} + O(h^{n+1}),$$

where $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$; $v_t^\alpha = v^\alpha(t, X_t)$; $\mathcal{A}_b^n, n \geq 1$ is the set of functions $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that $v \in \mathcal{A}_b^\alpha$ for all multi-index with finite length $\alpha \in \{\alpha | \ell(\alpha) \leq n\} \setminus \{\emptyset\}$ is well defined, continuous and bounded; \mathcal{A}_b^α denotes the subset of all functions $v \in \mathcal{A}^\alpha$ such that the function $\mathcal{L}^\alpha v$ is bounded; \mathcal{A}^α is the set of all functions $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ for which $\mathcal{L}^\alpha v$ is well defined and continuous; $\ell(\alpha)$ is the length of a multi-index of α ; let $v^{(0)} = \mathcal{L}^{(0)}v, v^{(0,0)} = \mathcal{L}^{(0)} \circ \mathcal{L}^{(0)}v, \dots, v^{(0)^n} = \underbrace{\mathcal{L}^{(0)} \circ \cdots \circ \mathcal{L}^{(0)}}_n v$.

2.2 Predictor-corrector discretization of the BSDE via Adams types

In this subsection, for readers' convenience to understanding the following text, we review the predictor-corrector discrete-time approximations of BSDE with respect to Y by Adams types (see [9]). As for the time-discretization of Z , we adopt the scheme proposed in [38].

Before approximating solutions of the BSDEs, we first define a uniform partition $\pi = \{t_0 := 0 < t_1 < t_2 \cdots < t_N := T\}$ and the step $h = \frac{T}{N}$, $\Delta W_i = W_{t_{i+1}} - W_{t_i}, W_i = W_{t_i}$. We consider the classical Euler discretization X^π of the SDE

$$\begin{cases} X_{i+1}^\pi = X_i^\pi + hb(t_i, X_i^\pi) + \sigma(t_i, X_i^\pi)\Delta W_i, & i = 0, 1, \dots, N-1, \\ X_0^\pi = x_0. \end{cases}$$

It is known that $\sup_{0 \leq i \leq N} \mathbb{E}[|X_{t_i} - X_i^\pi|^2] \rightarrow 0$, as $h \rightarrow 0$.

For non-stiff problems, Adams type is the most important linear multi-step method. Its solution approximation at t_i is defined either as

$$Y_i^\pi = \mathbb{E}_i \left[Y_{i+1}^\pi + h \sum_{\ell=1}^k \beta_\ell f_{i+\ell}^\pi \right], \quad (2.3)$$

or as

$$Y_i^\pi = \mathbb{E}_i \left[Y_{i+1}^\pi + h\beta_0 f_i^\pi + h \sum_{\ell=1}^k \beta_\ell f_{i+\ell}^\pi \right], \quad (2.4)$$

where Y_i^π and Z_i^π denote the discretization form of Y and Z at t_i and $f_i^\pi = f(t_i, X_i^\pi, Y_i^\pi, Z_i^\pi), i = 0, 1, \dots, N$; $\mathbb{E}_i[\cdot] = \mathbb{E}_{t_i}[\cdot]$; $\beta_0 \neq 0$ and $\{\beta_\ell\}_{1 \leq \ell \leq k}$ are real numbers and $k \in \mathbb{N}^+$.

If we utilize the equation (2.4) as the time-discretization of Y , we are required the solution of an algebraic equation at each step because the equation (2.4) is implicit. To solve Y in an explicit way, we can first approximate Y by the equation (2.3). Now, the obtained value of Y is denoted as \tilde{Y}_i^π , namely

$$\tilde{Y}_i^\pi = \mathbb{E}_i [Y_{i+1}^\pi + h \sum_{j=1}^{\tilde{k}} \tilde{\beta}_j f_{i+j}^\pi], \quad (2.5)$$

where $\tilde{k} \in \mathbb{N}^+$; $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_{\tilde{k}}$ are constants and would be given in the following. Next, we use the improved equation (2.4) to approximate Y , namely

$$Y_i^\pi = \mathbb{E}_i [Y_{i+1}^\pi + h\beta_0 \tilde{f}_i^\pi + h \sum_{j=1}^k \beta_j f_{i+j}^\pi], \quad (2.6)$$

where $\tilde{f}_i^\pi = f(t_i, X_i^\pi, \tilde{Y}_i^\pi, Z_i^\pi)$, for $i = N - 1, N - 2, \dots, 0$.

Next, we review the time-discretization of Z . From the BSDE in (2.1), we know

$$Y_{t_i} = Y_r + \int_{t_i}^r f(s, X_s, Y_s, Z_s) ds - \int_{t_i}^r Z_s dW_s, \quad r \in [t_i, T]. \quad (2.7)$$

Multiplying the above equation by $(W_r - W_i)^\top$, and taking conditional expectation, we obtain

$$0 = \mathbb{E}_i \left[Y_r (W_r - W_i)^\top \right] + \int_{t_i}^r \mathbb{E}_i \left[f(s, X_s, Y_s, Z_s) (W_r - W_i)^\top \right] ds - \int_{t_i}^r \mathbb{E}_i [Z_s] ds. \quad (2.8)$$

Differentiating the equation (2.8) w.r.t. r , we have

$$\frac{d\mathbb{E}_i \left[Y_r (W_r - W_i)^\top \right]}{dr} = -\mathbb{E}_i \left[f(r, X_r, Y_r, Z_r) (W_r - W_i)^\top \right] + \mathbb{E}_i [Z_r]. \quad (2.9)$$

Let $u \in C_b^{m+1}$, we apply Taylor's expansion at t_i for function $u(t)$, that is, for $n = 0, 1, 2, \dots, m$

$$u(t_i + nh) = u(t_i) + nh u'(t_i) + \frac{(nh)^2}{2!} u''(t_i) + \frac{(nh)^3}{3!} u'''(t_i) + \dots. \quad (2.10)$$

Moreover,

$$\sum_{n=0}^m \lambda_{m,n} u(t_i + nh) = \sum_{j=0}^2 \frac{\sum_{n=0}^m \lambda_{m,n} (nh)^j}{j!} \frac{d^j u}{dt^j}(t_i) + \mathcal{O} \left(\sum_{n=0}^m \lambda_{m,n} (nh)^{m+1} \right), \quad (2.11)$$

where $\lambda_{m,0}, \lambda_{m,1}, \dots, \lambda_{m,n}$ are real numbers. Let $\lambda_{m,n}, n = 0, 1, 2, \dots, m$ such that

$$\frac{1}{j!} \sum_{n=0}^m \lambda_{m,n} (nh)^j = \begin{cases} 1, & j = 1, \\ 0, & j \neq 1. \end{cases} \quad (2.12)$$

Hence, we deduce

$$\frac{du}{dt}(t_i) = \sum_{n=0}^m \lambda_{m,n} u(t_i + nh) + \mathcal{O} \left(\sum_{n=0}^m \lambda_{m,n} (nh)^{m+1} \right). \quad (2.13)$$

From the above equation, we have

$$\frac{d\mathbb{E}_i \left[Y_r (W_r - W_i)^\top \right]}{dr} \Bigg|_{r=t_i} = \sum_{n=0}^m \lambda_{m,n} \mathbb{E}_i \left[Y_{t_i+n} (W_{i+n} - W_i)^\top \right] + R_{Z,i}, \quad (2.14)$$

where $R_{Z,i} = \frac{d\mathbb{E}_i \left[Y_r (W_r - W_i)^\top \right]}{dr} \Bigg|_{r=t_i} - \sum_{n=0}^m \lambda_{m,n} \mathbb{E}_i \left[Y_{t_i+n} (W_{i+n} - W_i)^\top \right]$. Combining (2.9) with (2.14), we obtain

$$Z_{t_i} = \sum_{n=1}^m \lambda_{m,n} \mathbb{E}_i \left[Y_{t_i+n} (W_{i+n} - W_i)^\top \right] + R_{Z,i}.$$

Hence, the time-discretization of Z is, for $i = N - m, N - m - 1, \dots, 0$

$$Z_i^\pi = \sum_{n=1}^m \lambda_{m,n} \mathbb{E}_i \left[Y_{i+n}^\pi (W_{i+n} - W_i)^\top \right]. \quad (2.15)$$

Let \tilde{Y}_i^π denote the approximation to $u(t_i, X_i^\pi)$ via the predictor part. Set the improved approximation Y_i^π found in the corrector part. $\tilde{\beta}_j$ replaces the value of β_j in the Adams-Bashforth formula while β_j denotes the Adams-Moulton coefficients. Correspondingly, the parameter k denotes in the Adams-Moulton formula and the Adams-Bashforth formula can be denoted by \tilde{k} . Hence, the predictor-corrector scheme based on the Adams types could be expressed as below, for $i = N - \max(\tilde{k}, k), \dots, 1, 0$:

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi + h \sum_{j=1}^{\tilde{k}} \tilde{\beta}_j f_{i+j}^\pi], \\ Y_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi + h\beta_0 \tilde{f}_i^\pi + h \sum_{j=1}^k \beta_j f_{i+j}^\pi], \\ Z_i^\pi = \sum_{n=1}^{\max(\tilde{k}, k)} \lambda_{\max(\tilde{k}, k), n} \mathbb{E}_i[Y_{i+n}^\pi (W_{i+n} - W_i)^\top], \end{cases} \quad (2.16)$$

where $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_{\tilde{k}}$ and $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ are constants and would be given in the following. This scheme is implemented by means of Adams types i.e. Adams-Bashforth method is adopted by a preliminary computation. Subsequently, this numerical solution is used in the Adams-Moulton formula to yield the derivative value at the new point. The original idea of this scheme is extending the Euler method via allowing the numerical solution to depend on several previous step values of solutions and derivatives (see [2, 29–32] for detail about ODEs and [34] w.r.t. SDEs). The scheme (2.16) is referred to as the predictor-corrector method because the total calculation in a step is made up of a preliminary prediction of the numerical solution and followed by a correction of this predicted answer.

Usually, the coefficients k and \tilde{k} can take different values. To obtain the same order of local truncation error, the coefficients k and \tilde{k} have the relation $\tilde{k} = k + 1$. In addition, the scheme (2.16) can be rewritten as, for $i = N - k - 1, \dots, 1, 0$:

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi + h \sum_{j=1}^{k+1} \tilde{\beta}_j f_{i+j}^\pi], \\ Y_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi + h\beta_0 \tilde{f}_i^\pi + h \sum_{j=1}^k \beta_j f_{i+j}^\pi], \\ Z_i^\pi = \sum_{n=1}^{k+1} \lambda_{k+1, n} \mathbb{E}_i[Y_{i+n}^\pi (W_{i+n} - W_i)^\top]. \end{cases} \quad (2.17)$$

The scheme (2.16) provides an algorithm for calculating $(Y_{N-k-1}^\pi, Z_{N-k-1}^\pi)$ in terms of $(Y_N^\pi, Z_N^\pi), (Y_{N-1}^\pi, Z_{N-1}^\pi), \dots, (Y_{N-k}^\pi, Z_{N-k}^\pi)$. The subsequent approximation solutions can be found via the same manner. However, one has to consider how to obtain the value of $(Y_{N-1}^\pi, Z_{N-1}^\pi), (Y_{N-2}^\pi, Z_{N-2}^\pi), \dots, (Y_{N-k}^\pi, Z_{N-k}^\pi)$. Of course, it is possible to evaluate $(Y_{N-1}^\pi, Z_{N-1}^\pi), (Y_{N-2}^\pi, Z_{N-2}^\pi), \dots, (Y_{N-k}^\pi, Z_{N-k}^\pi)$ via a low order method, such as Euler scheme. Nevertheless, this maybe introduce much bigger errors and lead to nullification of the advantages of the subsequent use of the high order scheme. For this difficulty, we can utilize the Runge-Kutta scheme which is presented by J.-F. Chassagneux and D. Crisan [10] or the scheme (2.16) with $\tilde{k} = 1, k = 0$ with a smaller time step (see [38] for details).

In what follows, before providing the parameters in scheme (2.17), we first give the following definition.

Definition 2.3 *Suppose that $(u(t, X_t), \nabla_x u(t, X_t)\sigma(t, X_t))$ is the exact solution of the BSDE in (2.1). Let the local truncation error with respect to Y be*

$$T_i = u(t_i, X_i^\pi) - Y_i^\pi,$$

where Y_i^π denotes the numerical solution of the BSDE in (2.1). Furthermore, the multi-step scheme (2.17) with respect to Y is said to have n -order accuracy ($n \in \mathbb{N}^+$) if the local truncation error T_i satisfies $T_i = O(h^{n+1})$.

From Lemma 2.1, the integrand $\mathbb{E}_t[f(s, X_s, Y_s, Z_s)], s > t$ is a continuous function w.r.t. s . Then, by taking derivative w.r.t. s on

$$\mathbb{E}_t[Y_s] = \mathbb{E}_t[\Phi(X_T)] + \int_s^T \mathbb{E}_t[f(\bar{s}, X_{\bar{s}}, Y_{\bar{s}}, Z_{\bar{s}})]d\bar{s}, \quad \forall s \in [t, T],$$

we obtain the following reference ordinary differential equation

$$\frac{d\mathbb{E}_t[Y_s]}{ds} = -\mathbb{E}_t[f(s, X_s, Y_s, Z_s)], \quad s \in [t, T]. \quad (2.18)$$

Assume that no errors have yet been introduced when the approximation at (t_i, X_i) is about to be calculated. By (2.18), we get $\frac{d\mathbb{E}_i[Y_{i+j}^\pi]}{dt} = -\mathbb{E}_i[f_{i+j}^\pi] = \mathbb{E}_i[u^{(0)}(t_{i+j}, X_{i+j}^\pi)], j = 0, 1, 2, \dots$. Thus,

$$\begin{aligned} T_i &= \mathbb{E}_i \left[u(t_i, X_i^\pi) - u(t_{i+1}, X_{i+1}^\pi) - h \sum_{\ell=0}^k \beta_\ell f_{i+\ell}^\pi \right] \\ &= \mathbb{E}_i \left[u(t_i, X_i^\pi) - u(t_{i+1}, X_{i+1}^\pi) + h \sum_{\ell=0}^k \beta_\ell u^{(0)}(t_{i+\ell}, X_{i+\ell}^\pi) \right] \\ &= \mathbb{E}_i \left[hu^{(0)}(t_i, X_i^\pi)(-1 + \beta_0 + \beta_1 + \beta_2 + \dots + \beta_k) \right. \\ &\quad + h^2 u^{(0,0)}(t_i, X_i^\pi)(-\frac{1}{2} + \beta_1 + 2\beta_2 + \dots + k\beta_k) \\ &\quad + h^3 u^{(0,0,0)}(t_i, X_i^\pi)(-\frac{1}{6} + \frac{1}{2}(\beta_1 + 2^2\beta_2 + \dots + k^2\beta_k)) \\ &\quad + \dots \\ &\quad \left. + h^k u^{(0)k}(t_i, X_i^\pi)(-\frac{1}{k!} + \frac{1}{(k-1)!}(\beta_1 + 2^{k-1}\beta_2 + \dots + k^{k-1}\beta_k)) \right]. \end{aligned} \quad (2.19)$$

Then T_i has an expression as below via the equation (2.19)

$$C_0 u(t_i, X_i^\pi) + C_1 h u^{(0)}(t_i, X_i^\pi) + C_2 h^2 u^{(0,0)}(t_i, X_i^\pi) + \dots + C_k h^k u^{(0)k}(t_i, X_i^\pi) + O(h^{k+1}). \quad (2.20)$$

If $C_0 = C_1 = \dots = C_k = 0, C_{k+1} \neq 0$, then the local truncation error can be estimated as $O(h^{k+1})$. Now, the method has order k . In Table 1, we provides the value of parameters for $k = 1, 2, 3, 4, 5, 6$ (for $k = 1, 2, 3, 4$ see the Table in page 16 of [9]).

3 Main results

In this part, we introduce the predictor-corrector type general linear multi-step schemes of decoupled FBSDEs in detail and investigate the corresponding stability and convergence.

Table 1: coefficients for predictor-corrector scheme based on Adams type

order	term	β_0	β_1	β_2	β_3	β_4	β_5	β_6	error constant
1	predictor	0	1						$\frac{1}{2}$
	corrector	1	0						$-\frac{1}{2}$
2	predictor	0	$\frac{3}{2}$	$-\frac{1}{2}$					$-\frac{5}{12}$
	corrector	$\frac{1}{2}$	$\frac{1}{2}$						$\frac{1}{12}$
3	predictor	0	$\frac{23}{12}$	$-\frac{4}{3}$	$\frac{5}{12}$				$\frac{3}{8}$
	corrector	$\frac{5}{12}$	$\frac{2}{3}$	$-\frac{1}{12}$					$-\frac{1}{24}$
4	predictor	0	$\frac{55}{24}$	$-\frac{59}{24}$	$\frac{37}{24}$	$-\frac{3}{8}$			$-\frac{251}{720}$
	corrector	$\frac{3}{8}$	$\frac{19}{24}$	$-\frac{5}{24}$	$\frac{1}{24}$				$\frac{19}{720}$
5	predictor	0	$\frac{1901}{720}$	$-\frac{1387}{360}$	$\frac{109}{30}$	$-\frac{637}{360}$	$\frac{251}{720}$		$\frac{95}{288}$
	corrector	$\frac{251}{720}$	$\frac{323}{360}$	$-\frac{11}{30}$	$\frac{53}{360}$	$-\frac{19}{720}$			$-\frac{3}{160}$
6	predictor	0	$\frac{4277}{1440}$	$-\frac{2641}{480}$	$\frac{4991}{720}$	$-\frac{3649}{720}$	$\frac{959}{480}$	$-\frac{95}{288}$	$-\frac{19087}{60480}$
	corrector	$\frac{95}{288}$	$\frac{1427}{1440}$	$-\frac{133}{240}$	$\frac{241}{720}$	$-\frac{173}{1440}$	$\frac{3}{160}$		$\frac{863}{60480}$

3.1 Predictor-corrector discretization via the general linear multi-step scheme

In this subsection, we extend linear multi-step schemes ([8, 9]) to the predictor-corrector type general linear multi-step schemes.

Our aim is to deduce the discretization of BSDE backward in time based on the general linear multi-step scheme if $\{Y_l^\pi\}_{N-m+1 \leq l \leq N}$ and $\{Z_l^\pi\}_{N-m+1 \leq l \leq N}$ are available. Namely, for $i = N - m, N - m - 1, \dots, 0$

$$Y_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{i+j}^\pi + \sum_{j=1}^m \gamma_j h f_{i+j}^\pi \right], \quad (3.1)$$

or as

$$Y_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{i+j}^\pi + \gamma_0 h f_i^\pi + \sum_{j=1}^m \gamma_j h f_{i+j}^\pi \right], \quad (3.2)$$

where $\{\alpha_l\}_{1 \leq l \leq m}$ and $\{\gamma_l\}_{1 \leq l \leq m}$ are real numbers. In particular, let $\gamma_0 \neq 0$ be a real number. Now, (3.1) is an explicit scheme with respect to Y , while (3.2) is an implicit scheme.

As for the time-discretization of the term Z , we adopt the scheme presented in the subsection 2.2. Thus, the equations (3.2) and (2.15) consist of a discrete-time approximation (Y_i^π, Z_i^π) for (Y, Z) at t_i : for $i = N$

$$Y_N^\pi = \Phi(X_N^\pi), \quad Z_N^\pi = \sigma(t_N, X_N^\pi) D_x \Phi(X_N^\pi).$$

For $i = N - 1, N - 2, \dots, N - m + 1$, an appropriate one-step scheme can be utilized to solve the BSDE. For example, we can adjust the parameters of the scheme (2.16) such that it becomes one-step scheme and satisfies the required accuracy by using a smaller time step. For $i = N - m, N - m - 1, \dots, 1, 0$

$$\begin{cases} Y_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{i+j}^\pi + \gamma_0 h f_i^\pi + \sum_{j=1}^m \gamma_j h f_{i+j}^\pi \right], \\ Z_i^\pi = \sum_{n=1}^m \lambda_{m,n} \mathbb{E}_i [Y_{i+n}^\pi (W_{i+n} - W_i)^\top]. \end{cases} \quad (3.3)$$

This scheme is explicit w.r.t. Z and implicit w.r.t. Y . Of course, we can calculate the numerical solutions of BSDE via (3.3). But in general, the implicit scheme requires an algebraic equation to be solved at each time step. This imposes an additional computational burden. For this difficulty, we introduce the predictor-corrector method. The general linear multi-step predictor-corrector method is constructed as below:

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \tilde{\alpha}_j Y_{i+j}^\pi + \sum_{j=1}^m \tilde{\gamma}_j h f_{i+j}^\pi \right], \\ Y_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{i+j}^\pi + \gamma_0 h \tilde{f}_i^\pi + \sum_{j=1}^m \gamma_j h f_{i+j}^\pi \right], \\ Z_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \lambda_{m,j} Y_{i+j}^\pi (W_{i+j} - W_i)^\top \right], \end{cases} \quad (3.4)$$

where $\{\tilde{\alpha}_l\}_{1 \leq l \leq m}$ and $\{\tilde{\gamma}_l\}_{1 \leq l \leq m}$ are real numbers. At the i -th time step, the predictor is constructed by using an explicit general linear multi-step scheme which predicts a value of Y denoted by \tilde{Y}_i^π . Then the corrector whose structure is similar to an implicit general linear multi-step scheme is applied to correct the predicted value. We emphasize that not only the predictor step is explicit, but also the corrector step is explicit.

Next, we provide two schemes which are the variant forms of the scheme (3.4). In other words, these schemes are the special cases of (3.4). If the predictor term \tilde{Y} is calculated via the Adams-Bashforth method, the scheme (3.4) can be restated as below:

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i [Y_{i+1}^\pi + h \sum_{j=1}^{\tilde{m}} \tilde{\beta}_j f_{i+j}^\pi], \\ Y_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{i+j}^\pi + \gamma_0 h \tilde{f}_i^\pi + \sum_{j=1}^m \gamma_j h f_{i+j}^\pi \right], \\ Z_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \lambda_{m,j} Y_{i+j}^\pi (W_{i+j} - W_i)^\top \right]. \end{cases} \quad (3.5)$$

We can also naturally derive the following linear multi-step scheme by changing the calculation expression of Z (see [17]).

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \tilde{\alpha}_j Y_{i+j}^\pi + \sum_{j=1}^m \tilde{\gamma}_j h f_{i+j}^\pi \right], \\ Y_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{i+j}^\pi + \gamma_0 h \tilde{f}_i^\pi + \sum_{j=1}^m \gamma_j h f_{i+j}^\pi \right], \\ Z_i^\pi = \mathbb{E}_i \left[\left(\sum_{j=1}^m \alpha_j Y_{i+1+j}^\pi + \gamma_0 h \tilde{f}_{i+1}^\pi + \sum_{j=1}^m \gamma_j h f_{i+1+j}^\pi \right) \frac{\Delta W_i^\top}{h} \right]. \end{cases} \quad (3.6)$$

In what follows, our goal is to investigate the relation of the parameters α_j and γ_j under the conditions of stability and high order rate of convergence. This is necessary for the reason that we cannot implement the scheme (3.4) to calculate BSDEs if the parameters α_j and γ_j are not known. Combined (2.18), (2.20) with Itô-Taylor expansion, the local truncation error \tilde{T}_i of scheme (3.4) w.r.t. Y is, for $k = m$

$$\begin{aligned} \tilde{T}_i &= \mathbb{E}_i \left[u(t_i, X_i^\pi) - \sum_{\ell=1}^m \alpha_\ell u(t_{i+\ell}, X_{i+\ell}^\pi) - h \sum_{\ell=0}^m \gamma_\ell f_{i+\ell}^\pi \right] \\ &= \mathbb{E}_i \left[u(t_i, X_i^\pi) - \sum_{\ell=1}^m \alpha_\ell u(t_{i+\ell}, X_{i+\ell}^\pi) + h \sum_{\ell=0}^m \gamma_\ell u^{(0)}(t_{i+\ell}, X_{i+\ell}^\pi) \right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}_i \left[u(t_i, X_i^\pi) \left(1 - \sum_{\ell=1}^m \alpha_\ell \right) + hu^{(0)}(t_i, X_i^\pi) \left(- \sum_{\ell=1}^m \ell \alpha_\ell + \sum_{\ell=0}^m \gamma_\ell \right) \right. \\
&\quad + h^2 u^{(0,0)}(t_i, X_i^\pi) \left(- \frac{1}{2} \sum_{\ell=1}^m \ell^2 \alpha_\ell + \sum_{\ell=1}^m \ell \gamma_\ell \right) + h^3 u^{(0,0,0)}(t_i, X_i^\pi) \left(- \frac{1}{6} \sum_{\ell=1}^m \ell^3 \alpha_\ell + \frac{1}{2} \sum_{\ell=1}^m \ell^2 \gamma_\ell \right) \\
&\quad + \dots \\
&\quad \left. + h^m u^{(0)^m}(t_i, X_i^\pi) \left(- \frac{1}{m!} \sum_{\ell=1}^m \ell^m \alpha_\ell + \frac{1}{(m-1)!} \sum_{\ell=1}^m \ell^{m-1} \gamma_\ell \right) \right] + O(h^{m+1}). \tag{3.7}
\end{aligned}$$

Set

$$\left\{ \begin{array}{l} C_0 = 1 - \sum_{\ell=1}^m \alpha_\ell, \\ C_1 = - \sum_{\ell=1}^m \ell \alpha_\ell + \sum_{\ell=0}^m \gamma_\ell, \\ C_2 = - \frac{1}{2} \sum_{\ell=1}^m \ell^2 \alpha_\ell + \sum_{\ell=1}^m \ell \gamma_\ell, \\ \dots \\ C_m = - \frac{1}{m!} \sum_{\ell=1}^m \ell^m \alpha_\ell + \frac{1}{(m-1)!} \sum_{\ell=1}^m \ell^{m-1} \gamma_\ell. \end{array} \right. \tag{3.8}$$

If $C_0 = C_1 = C_2 = \dots = C_m = 0$ and $C_{m+1} \neq 0$, then the local truncation error accuracy of scheme (3.4) reaches m -order.

Remark 3.1 The (3.8) implies that we could obtain a family of schemes reaching m -order because the number of unknowns are greater than those of equations. This is the main difference from the scheme (2.16). Moreover, it indicates that the scheme (2.16) is a special form of the scheme of (3.4).

3.2 Error estimates of the scheme (3.4)

In this subsection, we concentrate on exploring the stability and high order accuracy of the scheme (3.4). Before demonstrating them, we first present a necessary property, a lemma and two definitions.

Proposition 3.2 Assume that f_i^π is smooth enough and $\tilde{k} = k + 1$ in scheme (2.16). For $i < N - k, k \in [0, N)$, it follows that

$$|u(t_i, X_i^\pi) - Y_i^\pi| = \left| \frac{C_{k+2}}{C_{k+2} - \tilde{C}_{k+2}} \right| |\tilde{Y}_i^\pi - Y_i^\pi|, \tag{3.9}$$

where \tilde{C}_{k+2} denotes the error constant for the predictor $(k + 1)$ -order term and C_{k+2} denotes the error constant for $(k + 1)$ -order corrector term.

Proof. It is straightforward that there exist two approximations to the exact solution $u(t_i, X_i^\pi)$ in every step in scheme (2.16). Moreover, the predictor term and the corrector term possess different error constants even though both of them have the same order. Thus, the error in the predictor term is equal to

$$u(t_i, X_i^\pi) = \tilde{Y}_i^\pi + h^{k+2} \tilde{C}_{k+2} u^{(0)k+2}(t_i, X_i^\pi) + o(h^{k+2}). \tag{3.10}$$

Similarly, we can obtain the error of the corrector term at the time step i

$$u(t_i, X_i^\pi) = Y_i^\pi + h^{k+2} C_{k+2} u^{(0)k+2}(t_i, X_i^\pi) + o(h^{k+2}). \quad (3.11)$$

Subtracting (3.10) from (3.11) and ignoring higher order term, one has

$$u^{(0)k+2}(t_i, X_i^\pi) = \frac{1}{h^{k+2}(C_{k+2} - \tilde{C}_{k+2})} (\tilde{Y}_i^\pi - Y_i^\pi). \quad (3.12)$$

Plugging (3.12) into (3.11) and neglecting higher order term, we obtain

$$u(t_i, X_i^\pi) - Y_i^\pi = \frac{C_{k+2}}{C_{k+2} - \tilde{C}_{k+2}} (\tilde{Y}_i^\pi - Y_i^\pi).$$

The proof is completed. ■

Next, we provide a lemma and two definitions which will be used to deduce the stability and high order accuracy of the scheme (3.4),

Lemma 3.3 (see Lemma 3 in [39]) Suppose that N and K are two nonnegative integers with $N \geq K$ and h any positive number. Let $\{\eta_i\}$ be a series satisfying

$$|\eta_i| \leq \beta + \alpha h \sum_{j=i+1}^N |\eta_j|, \quad i = N - K, N - K - 1, \dots, 0,$$

where α and β are two positive constants. Let $M_0 = \max_{N-K \leq j \leq N} |\eta_j|$ and $T = Nh$; then

$$|\eta_i| \leq \exp(\alpha T)(\beta + \alpha K h M_0), \quad i = N - K, N - K - 1, \dots, 0.$$

Definition 3.4 The characteristic polynomials of (3.2) are given by

$$P(\zeta) = \zeta^m - \alpha_1 \zeta^{m-1} - \alpha_2 \zeta^{m-2} - \dots - \alpha_m. \quad (3.13)$$

The equation (3.2) is said to fulfil Dahlquist's root condition, if

- i) The roots of $P(\zeta)$ lie on or within the unit circle;
- ii) The roots on the unit circle are simple.

Definition 3.5 Let $(Y_i^\pi, Z_i^\pi), i = 0, 1, \dots, N - m$ be the time-discretization approximate solution given by the scheme (3.4) and $(\tilde{Y}_i^\pi, \tilde{Z}_i^\pi)$ is the solution of its perturbed form (see (3.15)). Then the scheme (3.4) is said to be \mathbb{L}_2 -stable if

$$\max_{0 \leq i \leq N-m} \mathbb{E}[|Y_i^\pi - \tilde{Y}_i^\pi|^2] + \sum_{i=0}^{N-m} h \mathbb{E}[|Z_i^\pi - \tilde{Z}_i^\pi|^2] \leq C \left(\max_{N-m+1 \leq k \leq N} |Y_k^\pi - \tilde{Y}_k^\pi|^2 + \sum_{i=0}^{N-m} \mathbb{E}[h|\varepsilon_i^Z|^2 + \frac{1}{h}|\varepsilon_i^Y|^2] \right), \quad (3.14)$$

where C is a constant; $(\tilde{Y}_i^\pi, \tilde{Z}_i^\pi)$ satisfies a perturbed form of (3.4) for $i = N - m, N - m - 1, \dots, 0$

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \tilde{\alpha}_j \tilde{Y}_{i+j}^\pi + \sum_{j=1}^m \tilde{\gamma}_j h f(t_{i+j}, X_{i+j}^\pi, \tilde{Y}_{i+j}^\pi, \tilde{Z}_{i+j}^\pi) \right], \\ \tilde{Y}_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j \tilde{Y}_{i+j}^\pi + \gamma_0 h f(t_{i+j}, X_{i+j}^\pi, \tilde{Y}_i^\pi, \tilde{Z}_i^\pi) + \sum_{j=1}^m \gamma_j h f(t_{i+j}, X_{i+j}^\pi, \tilde{Y}_{i+j}^\pi, \tilde{Z}_{i+j}^\pi) \right] + \varepsilon_i^Y, \\ \tilde{Z}_i^\pi = \mathbb{E}_i \left[\sum_{j=1}^m \lambda_{m,j} \tilde{Y}_{i+j}^\pi (W_{i+j} - W_i)^\top \right] + \varepsilon_i^Z. \end{cases} \quad (3.15)$$

Sequences ε_i^Y and ε_i^Z which belong to $\mathbb{L}_2(\mathcal{F}_i)$ are random variables.

Note that we are merely interested in the solution of the BSDE in (2.1). Therefore, we assume that the solution of SDE in (2.1) can be obtained perfectly. Thus, we do not consider the error caused by X_t (see [38]).

Theorem 3.6 *Suppose **Assumption 2 (i)** and **Assumption 2 (iii)** hold. Then the stochastic multi-step method is numerically stable if and only if its characteristic polynomial (3.13) satisfies Dahlquist's root condition.*

Proof. Sufficiency: Let $\Delta Y_i = Y_i^\pi - \bar{Y}_i^\pi$, $\Delta Z_i = Z_i^\pi - \bar{Z}_i^\pi$, $\Delta f_i = f(t_i, X_i^\pi, Y_i^\pi, Z_i^\pi) - f(t_i, X_i^\pi, \bar{Y}_i^\pi, \bar{Z}_i^\pi)$ for $i = N - m, N - m - 1, \dots, 0$. We complete the proof of the theorem in three steps.

step 1. From (3.4) and (3.15) w.r.t. Y , one obtains

$$\Delta Y_i = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j \Delta Y_{i+j} + \gamma_0 h \Delta \hat{f}_i + \sum_{j=1}^m \gamma_j h \Delta f_{i+j} \right] - \varepsilon_i^Y,$$

where $\Delta \hat{f}_i = f(t_i, X_i^\pi, \bar{Y}_i^\pi, Z_i^\pi) - f(t_i, X_i^\pi, \bar{Y}_i^\pi, \bar{Z}_i^\pi)$. We rearrange the m -step recursion to a one-step recursion as follow

$$\mathbb{E}_i[\mathcal{Y}_i] = \mathbb{E}_i[A\mathcal{Y}_{i+1} + F_i + R_i], \quad (3.16)$$

where

$$\mathcal{Y}_i = \begin{pmatrix} \Delta Y_i \\ \Delta Y_{i+1} \\ \vdots \\ \Delta Y_{i+m-1} \end{pmatrix}, A = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}, F_i = \begin{pmatrix} \gamma_0 h \Delta \hat{f}_i + \sum_{j=1}^m \gamma_j h \Delta f_{i+j} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, R_i = \begin{pmatrix} -\varepsilon_i^Y \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

To ensure the stability of the m -step scheme, the norm of the matrix A in the equation (3.16) is no more than 1 (see [20], Chapter III.4, Lemma 4.4). This can be satisfied if the eigenvalues $\text{eig}(A)$ of the matrix A make $|\text{eig}(A)| \leq 1$ and in which the eigenvalues are simple if $|\text{eig}(A)| = 1$. In addition, the eigenvalues of A satisfy the root condition by Definition 3.4. By the Dahlquist's root condition, it is possible that there exists a non-singular matrix \mathcal{D} such that $\|\mathcal{D}^{-1}A\mathcal{D}\|_2 \leq 1$ where $\|\cdot\|_2$ denotes the spectral matrix norm induced by Euclidian vector norm in $\mathbb{R}^{m \times n}$. Hence, we can choose a scalar product for $\bar{A}, \tilde{A} \in \mathbb{R}^{m \times n}$ as $\langle \bar{A}, \tilde{A} \rangle_* := \langle \mathcal{D}^{-1}\bar{A}, \mathcal{D}^{-1}\tilde{A} \rangle = \bar{A}^\top (\mathcal{D}^{-1})^\top \mathcal{D}^{-1}\tilde{A}$. And we have $|\bar{A}|_*^2 := \langle \bar{A}, \bar{A} \rangle_*$ with the induced vector norm on $\mathbb{R}^{m \times n}$. Let $\|A\|_* = \|\mathcal{D}^{-1}A\mathcal{D}\|_2$ be the induced matrix norm. Owing to the norm equivalence, we know that there exist positive constants c_1, c_2 such that

$$c_1 |\bar{A}|_2^2 \leq |\bar{A}|_*^2 \leq c_2 |\bar{A}|_2^2, \quad \forall \bar{A} \in \mathbb{R}^{m \times n} \quad (3.17)$$

where $|\bar{A}|_2^2 = \sum_{j=1,2,\dots,m} |a_j|^2$ for $\bar{A} = (a_1^\top, \dots, a_m^\top)^\top$. Applying $|\cdot|_*$ to the equation (3.16), we have

$$\begin{aligned} |\mathbb{E}_i[\mathcal{Y}_i]|_* &= \left| \mathbb{E}_i[A\mathcal{Y}_{i+1} + F_i + R_i] \right|_* \\ &= \|A\|_* |\mathbb{E}_i[\mathcal{Y}_{i+1}]|_* + |\mathbb{E}_i[F_i]|_* + |\mathbb{E}_i[R_i]|_* \\ &\leq |\mathbb{E}_i[\mathcal{Y}_{i+1}]|_* + |\mathbb{E}_i[F_i]|_* + |\mathbb{E}_i[R_i]|_*. \end{aligned} \quad (3.18)$$

Squaring the above (3.18), then from the inequality $(\sum_{i=1}^n a_i)^2 \leq n \sum_{i=1}^n a_i^2$ and (3.17), one deduces

$$\begin{aligned}
|\mathbb{E}_i[\mathcal{Y}_i]_*|^2 &\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 + 3|\mathbb{E}_i[F_i]_*|^2 + 3|\mathbb{E}_i[R_i]_*|^2 \\
&\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 + 3c_2|\mathbb{E}_i[\gamma_0 h \Delta \widehat{f}_i + \sum_{j=1}^m \gamma_j h \Delta f_{i+j}]|^2 + 3c_2|\mathbb{E}_i[\varepsilon_i^Y]|^2 \\
&\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 + 3(m+1)h^2 c_2 \left(|\mathbb{E}_i[\gamma_0 \Delta \widehat{f}_i]|^2 + \sum_{j=1}^m |\mathbb{E}_i[\gamma_j \Delta f_{i+j}]|^2 \right) + 3c_2|\mathbb{E}_i[\varepsilon_i^Y]|^2.
\end{aligned} \tag{3.19}$$

By the Lipschitz condition of f with respect to (y, z) and

$$\widetilde{Y}_k^\pi - \widetilde{Y}_k^\pi = \mathbb{E}_k \left[\sum_{j=1}^m \widetilde{\alpha}_j \Delta Y_{k+j} + \sum_{j=1}^m \widetilde{\gamma}_j h \Delta f_{k+j} \right],$$

(3.19) can be restated as

$$\begin{aligned}
|\mathbb{E}_i[\mathcal{Y}_i]_*|^2 &\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 + 6(m+1)h^2 c_2 L_f^2 \left(|\mathbb{E}_i[\gamma_0 (\widetilde{Y}_i^\pi - \widetilde{Y}_i^\pi)]|^2 + \sum_{j=1}^m |\mathbb{E}_i[\gamma_j \Delta Y_{i+j}]|^2 \right) \\
&\quad + \sum_{j=0}^m |\mathbb{E}_i[\gamma_j \Delta Z_{i+j}]|^2 + 3c_2|\mathbb{E}_i[\varepsilon_i^Y]|^2 \\
&\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 + 6(m+1)h^2 c_2 L_f^2 \left(\sum_{j=1}^m (4m^2 \gamma_0^2 \widetilde{\alpha}_j^2 + 8m^2 \gamma_0^2 L_f^2 h^2 \widetilde{\gamma}_j^2 + \gamma_j^2) |\mathbb{E}_i[\Delta Y_{i+j}]|^2 \right) \\
&\quad + \sum_{j=0}^m (8m^2 \gamma_0^2 L_f^2 h^2 \widetilde{\gamma}_j^2 + \gamma_j^2) |\mathbb{E}_i[\Delta Z_{i+j}]|^2 + 3c_2|\mathbb{E}_i[\varepsilon_i^Y]|^2 \\
&\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 + 6(m+1)h^2 \frac{c_2}{c_1} L_f^2 \max_{1 \leq j \leq m} \{4m^2 \gamma_0^2 \widetilde{\alpha}_j^2 + 8m^2 \gamma_0^2 L_f^2 h^2 \widetilde{\gamma}_j^2 + \gamma_j^2\} |\mathbb{E}_i[\mathcal{Y}_{i+1}]_*|^2 \\
&\quad + 6(m+1)h^2 c_2 L_f^2 \sum_{j=0}^m (8m^2 \gamma_0^2 L_f^2 h^2 \widetilde{\gamma}_j^2 + \gamma_j^2) |\mathbb{E}_i[\Delta Z_{i+j}]|^2 + 3c_2|\mathbb{E}_i[\varepsilon_i^Y]|^2.
\end{aligned} \tag{3.20}$$

step 2. Subtracting (3.15) from (3.4) with respect to Z , we obtain

$$\Delta Z_i = \mathbb{E}_i \left[\sum_{j=1}^m \lambda_{m,j} \Delta Y_{i+j} (W_{i+j} - W_i)^\top \right] - \varepsilon_i^Z. \tag{3.21}$$

Moreover, we get

$$\begin{aligned}
|\Delta Z_i| &= \left| \mathbb{E}_i \left[\sum_{j=1}^m \lambda_{m,j} \Delta Y_{i+j} (W_{i+j} - W_i)^\top \right] - \varepsilon_i^Z \right| \\
&\leq \sum_{j=1}^m \left| \lambda_{m,j} \mathbb{E}_i [\Delta Y_{i+j} (W_{i+j} - W_i)^\top] \right| + |\varepsilon_i^Z|.
\end{aligned} \tag{3.22}$$

Squaring the above equation (3.22) and then by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|\Delta Z_i|^2 &\leq (m+1) \sum_{j=1}^m \max_{1 \leq j \leq m} \{\lambda_{m,j}^2\} \left| \mathbb{E}_i [\Delta Y_{i+j} (W_{i+j} - W_i)^\top] \right|^2 + (m+1) |\varepsilon_i^Z|^2 \\
&\leq (m+1) m d h \sum_{j=1}^m \max_{1 \leq j \leq m} \{\lambda_{m,j}^2\} |\mathbb{E}_i [\Delta Y_{i+j}]|^2 + (m+1) |\varepsilon_i^Z|^2.
\end{aligned} \tag{3.23}$$

Summing over the above inequality from i to $N - m$ and taking expectation, we have

$$\begin{aligned} \sum_{k=i}^{N-m} h \mathbb{E}_i [|\Delta Z_k|^2] &\leq (m+1)mdh^2 \sum_{k=i}^{N-m} \sum_{j=1}^m \max_{1 \leq j \leq m} \{\lambda_{m,j}^2\} |\mathbb{E}_i [\Delta Y_{i+j}]|^2 + (m+1)h \sum_{k=i}^{N-m} \mathbb{E}_i [|\varepsilon_k^Z|^2] \\ &\leq (m+1)mdh^2 \max_{1 \leq j \leq m} \{\lambda_{m,j}^2\} \sum_{k=i+1}^{N-m+1} \mathbb{E}_i [|\mathcal{Y}_k|^2] + (m+1)h \sum_{k=i}^{N-m} \mathbb{E}_i [|\varepsilon_k^Z|^2]. \end{aligned} \quad (3.24)$$

step 3. Inserting (3.23) into (3.20), we obtain

$$\begin{aligned} |\mathbb{E}_i[\mathcal{Y}_i]|_*^2 &\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]|_*^2 + 6(m+1)h^2 \frac{c_2}{c_1} L_f^2 \max_{1 \leq j \leq m} \{4m^2 \gamma_0^2 \tilde{\alpha}_j^2 + 8m^2 \gamma_0^2 L_f^2 h^2 \tilde{\gamma}_j^2 + \gamma_j^2\} |\mathbb{E}_i[\mathcal{Y}_{i+1}]|_*^2 \\ &\quad + 6(m+1)^2 mdh^3 \frac{c_2}{c_1} L_f^2 \max_{1 \leq j \leq m} \{\lambda_{m,j}^2\} \max_{1 \leq j \leq m} \{8m^2 \gamma_0^2 L_f^2 h^2 \tilde{\gamma}_j^2 + \gamma_j^2\} \sum_{j=0}^m \mathbb{E}_i [|\mathcal{Y}_{i+1+j}|_*^2] \\ &\quad + 6(m+1)^2 h^2 c_2 L_f^2 \sum_{j=0}^m (8m^2 \gamma_0^2 L_f^2 h^2 \tilde{\gamma}_j^2 + \gamma_j^2) \mathbb{E}_i [|\varepsilon_{i+j}^Z|^2] + 3c_2 |\mathbb{E}_i[\varepsilon_i^Y]|^2 \\ &\leq 3|\mathbb{E}_i[\mathcal{Y}_{i+1}]|_*^2 + 6(m+1)h^2 \frac{c_2}{c_1} L_f^2 \left(\max_{1 \leq j \leq m} \{4m^2 \gamma_0^2 \tilde{\alpha}_j^2 + 8m^2 \gamma_0^2 L_f^2 h^2 \tilde{\gamma}_j^2 + \gamma_j^2\} \right. \\ &\quad \left. + (m+1)^2 mdh \max_{1 \leq j \leq m} \{\lambda_{m,j}^2\} \max_{1 \leq j \leq m} \{8m^2 \gamma_0^2 L_f^2 h^2 \tilde{\gamma}_j^2 + \gamma_j^2\} \right) \sum_{k=i+1}^{N-m+1} |\mathbb{E}_i[\mathcal{Y}_k]|_*^2 \\ &\quad + 6(m+1)^2 h^2 c_2 L_f^2 \max_{1 \leq j \leq m} \{8m^2 \gamma_0^2 L_f^2 h^2 \tilde{\gamma}_j^2 + \gamma_j^2\} \sum_{j=0}^m \mathbb{E}_i [|\varepsilon_{i+j}^Z|^2] + 3c_2 |\mathbb{E}_i[\varepsilon_i^Y]|^2. \end{aligned} \quad (3.25)$$

There exists a constant C which changes from line to line such that

$$|\mathbb{E}_i[\mathcal{Y}_i]|_*^2 \leq C \left((h+h^2) \sum_{k=i+1}^{N-m+1} |\mathbb{E}_i[\mathcal{Y}_k]|_*^2 + \sum_{k=0}^m \mathbb{E}_i [|\varepsilon_k^Y|^2 + h^2 |\varepsilon_{i+k}^Z|^2] \right). \quad (3.26)$$

From Lemma 3.3, we have

$$|\mathbb{E}_i[\mathcal{Y}_i]|_*^2 \leq C \left(\max_{i+1 \leq k \leq N-m+1} mh |\mathbb{E}_i[\mathcal{Y}_k]|_*^2 + \sum_{k=0}^m \mathbb{E}_i [|\varepsilon_k^Y|^2 + h^2 |\varepsilon_{i+k}^Z|^2] \right). \quad (3.27)$$

Inserting (3.27) into (3.24), we get, for h small enough

$$\sum_{k=i}^{N-m} h \mathbb{E}_i [|\Delta Z_k|^2] \leq C \left(\max_{i+1 \leq k \leq N} |\mathbb{E}[Y_k]|^2 + \sum_{k=i}^{N-m} \mathbb{E}_i [h |\varepsilon_k^Z|^2 + \frac{1}{h} |\varepsilon_k^Y|^2] \right). \quad (3.28)$$

Adding (3.27) to the above (3.28), we derive that there exists a constant C such that

$$\max_{0 \leq i \leq N-m} |\mathbb{E}[\mathcal{Y}_i]|_*^2 + \sum_{i=0}^{N-m} h \mathbb{E} [|\Delta Z_i|^2] \leq C \left(\max_{N-m+1 \leq k \leq N} |Y_k^\pi - \bar{Y}_k^\pi|^2 + \sum_{i=0}^{N-m} \mathbb{E}_i [h |\varepsilon_i^Z|^2 + \frac{1}{h} |\varepsilon_i^Y|^2] \right).$$

Necessity: The proof is analogous to ordinary differential equations (see Theorem 6.3.3 of [15]). So we omit it. ■

Theorem 3.7 Suppose that Assumption 2 holds. Furthermore, $f(t, x, y, z)$ and $\Phi(x_T)$ are smooth enough functions. Let (Y_{t_i}, Z_{t_i}) and (Y_i^π, Z_i^π) be solutions of the BSDE in (2.1) and solutions of the scheme (3.4)

respectively. The terminal values satisfy $\mathbb{E}[\sup_{N-m < i \leq N} |Y_i^\pi - Y_{t_i}|^2 + h|Z_i^\pi - Z_{t_i}|^2]^{\frac{1}{2}} \leq Ch^{m+1}$. Then, as h is small enough

$$\mathbb{E}[\sup_{0 \leq i \leq N-m} |Y_i^\pi - Y_{t_i}|^2 + h|Z_i^\pi - Z_{t_i}|^2]^{\frac{1}{2}} \leq Ch^{m+1},$$

where C is a constant changing from line to line.

Proof. The BSDE in (2.1) is discretized by the scheme as below:

$$\begin{cases} \tilde{Y}_i = \mathbb{E}_i \left[\sum_{j=1}^m \tilde{\alpha}_j Y_{t_{i+j}} + \sum_{j=1}^m \tilde{\gamma}_j h f_{i+j} \right], \\ Y_{t_i} = \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j Y_{t_{i+j}} + \gamma_0 h \tilde{f}_i + \sum_{j=1}^m \gamma_j h f_{i+j} \right] + R_{Y,i}, \\ Z_{t_i} = \mathbb{E}_i \left[\sum_{j=1}^m \lambda_{m,j} Y_{t_{i+j}} (W_{i+j} - W_i)^\top \right] + R_{Z,i}, \end{cases} \quad (3.29)$$

where $R_{Y,i}$ and $R_{Z,i}$ denote the error of the exact solutions and the approximation solutions w.r.t. Y and Z ; $f_i = f(t_i, X_{t_i}, Y_{t_i}, Z_{t_i})$, $\tilde{f}_i = f(t_i, X_{t_i}, \tilde{Y}_{t_i}, Z_{t_i})$, $i = 0, 1, \dots, N$. Set $\Delta \mathcal{Y}_i = Y_i^\pi - Y_{t_i}$, $\Delta \tilde{\mathcal{Y}}_i = \tilde{Y}_i^\pi - \tilde{Y}_{t_i}$, $\Delta \mathcal{Z}_i = Z_i^\pi - Z_{t_i}$. From (3.4) and (3.29), we have

$$\begin{aligned} |\Delta \mathcal{Y}_i| &= \left| \mathbb{E}_i \left[\sum_{j=1}^m \alpha_j \Delta \mathcal{Y}_{i+j} + h \gamma_0 (\tilde{f}_i^\pi - \tilde{f}_i) + h \sum_{j=1}^m \gamma_j (f_{i+j}^\pi - f_{i+j}) \right] - R_{Y,i} \right| \\ &\leq \mathbb{E}_i \left[\sum_{j=1}^m |\alpha_j| |\Delta \mathcal{Y}_{i+j}| + h |\gamma_0| |\tilde{f}_i^\pi - \tilde{f}_i| + h \sum_{j=1}^m |\gamma_j| |f_{i+j}^\pi - f_{i+j}| \right] + |R_{Y,i}| \\ &\leq \mathbb{E}_i \left[\sum_{j=1}^m |\alpha_j| |\Delta \mathcal{Y}_{i+j}| + h |\gamma_0| L_f |\Delta \tilde{\mathcal{Y}}_i + \Delta \mathcal{Z}_i| + h \sum_{j=1}^m |\gamma_j| L_f |\Delta \mathcal{Y}_{i+j} + \Delta \mathcal{Z}_{i+j}| \right] + |R_{Y,i}|. \end{aligned} \quad (3.30)$$

$$\begin{aligned} |\Delta \tilde{\mathcal{Y}}_i| &= |\tilde{Y}_i^\pi - \tilde{Y}_{t_i}| = \left| \mathbb{E}_i \left[\sum_{j=1}^m \tilde{\alpha}_j \Delta \mathcal{Y}_{i+j} + h \sum_{j=1}^m \tilde{\gamma}_j (f_{i+j}^\pi - f_{i+j}) \right] \right| \\ &\leq \mathbb{E}_i \left[\sum_{j=1}^m |\tilde{\alpha}_j| |\Delta \mathcal{Y}_{i+j}| + h \sum_{j=1}^m |\tilde{\gamma}_j| L_f |\Delta \mathcal{Y}_{i+j} + \Delta \mathcal{Z}_{i+j}| \right]. \end{aligned} \quad (3.31)$$

$$\Delta \mathcal{Z}_i = \sum_{j=1}^m \lambda_{m,j} \mathbb{E}_i \left[\Delta \mathcal{Y}_{i+j} (W_{i+j} - W_i)^\top \right] - R_{Z,i}. \quad (3.32)$$

Inserting (3.31) into (3.30), we obtain

$$\begin{aligned} |\Delta \mathcal{Y}_i| &\leq \mathbb{E}_i \left[\sum_{j=1}^m |\alpha_j| |\Delta \mathcal{Y}_{i+j}| + h |\gamma_0| L_f \left| \mathbb{E}_i \left[\sum_{j=1}^m |\tilde{\alpha}_j| |\Delta \mathcal{Y}_{i+j}| + h \sum_{j=1}^m |\tilde{\gamma}_j| L_f |\Delta \mathcal{Y}_{i+j} + \Delta \mathcal{Z}_{i+j}| \right] \right. \right. \\ &\quad \left. \left. + \Delta \mathcal{Z}_i + h \sum_{j=1}^m |\gamma_j| L_f |\Delta \mathcal{Y}_{i+j} + \Delta \mathcal{Z}_{i+j}| \right] + |R_{Y,i}| \\ &\leq \mathbb{E}_i \left[\sum_{j=1}^m (|\alpha_j| + h L_f |\gamma_0 \tilde{\alpha}_j|) |\Delta \mathcal{Y}_{i+j}| + h |\gamma_0| L_f |\Delta \mathcal{Z}_i| \right. \\ &\quad \left. + \sum_{j=1}^m \left(h^2 L_f^2 |\gamma_0 \tilde{\gamma}_j| + h L_f |\gamma_j| \right) |\Delta \mathcal{Y}_{i+j} + \Delta \mathcal{Z}_{i+j}| \right] + |R_{Y,i}|. \end{aligned} \quad (3.33)$$

Squaring the inequality (3.33) and inserting (3.32) into the derived equation yield, for $i = N - m, N - m - 1, \dots, 0$

$$|\Delta \mathcal{Y}_i|^2 \leq 4 \left(\mathbb{E}_i \left[\sum_{j=1}^m m (|\alpha_j| + h L_f |\gamma_0 \tilde{\alpha}_j|)^2 |\Delta \mathcal{Y}_{i+j}|^2 + h^2 L_f^2 \gamma_0^2 \left| \sum_{j=1}^m \lambda_{m,j} \mathbb{E}_i \left[\Delta \mathcal{Y}_{i+j} (W_{i+j} - W_i)^\top \right] - R_{Z,i} \right|^2 \right. \right.$$

$$\begin{aligned}
& + \sum_{j=1}^m 2m (h^2 L_f^2 |\gamma_0 \tilde{\gamma}_j| + h L_f |\gamma_j|)^2 (|\Delta \mathcal{Y}_{i+j}|^2 \\
& + |\sum_{n=1}^m \lambda_{m,n} \mathbb{E}_{i+j} [\Delta \mathcal{Y}_{i+j+n} (W_{i+j+n} - W_{i+j})^\top] - R_{Z,i+j}|^2) + |R_{Y,i}|^2 \\
& \leq 4 \left(\mathbb{E}_i \left[(m(|\alpha_j| + h L_f |\gamma_0 \tilde{\alpha}_j|)^2 + h m(m+1) L_f^2 d \gamma_0^2 (\max_{1 \leq j \leq m} \lambda_{m,j} h)^2 + 4 m h^2 L_f^2 (h^2 L_f^2 |\gamma_0 \tilde{\gamma}_j|^2 + |\gamma_j|^2) \right. \right. \\
& \quad \left. \left. + 4 m^2 (m+1) d h L_f^2 (\max_{1 \leq j \leq m} \lambda_{m,j} h)^2 (h^2 L_f^2 |\gamma_0 \tilde{\gamma}_j|^2 + |\gamma_j|^2) \right) \sum_{j=1}^m |\Delta \mathcal{Y}_{i+j}|^2 \right] + R_{Y,i}^2 \\
& \quad \left. + \left(h^2 (m+1) L_f^2 \gamma_0^2 + 4 m (m+1) h^2 L_f^2 (h^2 L_f^2 |\gamma_0 \tilde{\gamma}_j|^2 + |\gamma_j|^2) \right) \max_{0 \leq j \leq m} R_{Z,i+j}^2 \right) \\
& \leq C \left(\mathbb{E}_i \left[(h + h^2) \sum_{j=i+1}^N |\Delta \mathcal{Y}_j|^2 \right] + |R_{Y,i}|^2 + h^2 \max_{i+1 \leq j \leq N} R_{Z,j}^2 + O(h^{2m+2}) \right). \tag{3.34}
\end{aligned}$$

From Lemma 3.3, the inequality (3.34) can be rewritten as

$$|\Delta \mathcal{Y}_i|^2 \leq C \left(m h \max_{N-m \leq j \leq N} |\Delta \mathcal{Y}_j|^2 + |R_{Y,i}|^2 + h^2 \max_{i+1 \leq j \leq N} R_{Z,j}^2 + O(h^{2m+2}) \right). \tag{3.35}$$

From Proposition 3.2, we derive

$$|R_{Y,i}| \leq C h^{m+1}. \tag{3.36}$$

By Lemma 2.1 in [36], we have

$$|R_{Z,i}| \leq C h^m. \tag{3.37}$$

Combining (3.35), (3.36) with (3.37), we deduce that

$$|\Delta \mathcal{Y}_i|^2 \leq C (h^{2m+3} + h^{2m+2}), \tag{3.38}$$

for $i = N - m - 1, N - m - 2, \dots, 0$ recursively. Hence, $\sup_{0 \leq i < N-m} |\Delta \mathcal{Y}_i| \leq C h^{m+1}$.

Squaring (3.32) and multiplying h and then with the help of Cauchy-Schwarz inequality, we have

$$\begin{aligned}
h |\Delta \mathcal{Z}_i|^2 & = h \left| \sum_{j=1}^m \lambda_{m,j} \mathbb{E}_i [\Delta \mathcal{Y}_{i+j} (W_{i+j} - W_i)^\top] - R_{Z,i} \right|^2 \\
& \leq h(m+1) \left(\sum_{j=1}^m \lambda_{m,j}^2 m h d \mathbb{E}_i [|\Delta \mathcal{Y}_{i+j}|^2] + |R_{Z,i}|^2 \right) \\
& \leq C \left(\max_{i+1 \leq j \leq N} \mathbb{E}_i [|\Delta \mathcal{Y}_j|^2] + h |R_{Z,i}|^2 \right) \\
& \leq C h^{2m+2}. \tag{3.39}
\end{aligned}$$

Hence, we deduce the conclusion with the help of (3.38) and (3.39). The proof is completed. ■

4 Numerical Experiments

In this section, we provide two numerical examples to show the performance of the scheme (3.4). Specifically, in the **Example 1**, we provide stable numerical schemes for the step number $m = 1, 2, 3, 4$ to show their

convergence rates w.r.t. the time step sizes, absolute errors and running times. And the comparisons with explicit Adams method in [8] are also given. In the **Example 2**, we also present unstable numerical schemes for the step number $m = 2, 3$ to illustrate the previous theory analysis.

To assess the performance of our algorithms, we had better to find a BSDE with closed-form solutions and establish criterions. Let $\epsilon = \mathbb{E}[|Y_0 - Y_0^\pi|]$ denote the error between closed-form solutions and numerical solutions. From the Central Limit Theorem, one gets the error $\bar{\epsilon} := \frac{1}{M} \sum_{k=1}^M |Y_0 - Y_{0,k}^\pi|$ that converges in distribution to ϵ as $M \rightarrow \infty$.

In implementation, one can calculate the variance $\hat{\sigma}_\epsilon^2$ of $\hat{\epsilon}$ and then utilize it to construct a confidence interval (CI) for the absolute error ϵ . To realize this idea, one arranges the simulations into \widetilde{M} batches of M simulations each and estimates the variance $\hat{\sigma}_\epsilon^2$. To be precise, define the average errors $\hat{\epsilon}_j = \frac{1}{M} \sum_{k=1}^M |Y_0 - Y_{0,k,j}^\pi|, j = 1, 2, \dots, \widetilde{M}$, where $Y_{0,k,j}^\pi$ is k -th trajectory generated by our schemes in the j th batch at time 0. These average errors are independent and approximately Gaussian when M is large enough. Thus, the mean of the batch averages is $\hat{\epsilon} = \frac{1}{M} \sum_{j=1}^{\widetilde{M}} \hat{\epsilon}_j = \frac{1}{M\widetilde{M}} \sum_{j=1}^{\widetilde{M}} \sum_{k=1}^M |Y_0 - Y_0^\pi|$ and the variance of the batch averages is $\hat{\sigma}_\epsilon^2 = \frac{1}{M-1} \sum_{j=1}^{\widetilde{M}} (\hat{\epsilon}_j - \hat{\epsilon})^2$. Experience has shown that the batch averages can be interpreted as being Gaussian for batch sizes $\widetilde{M} \geq 15$. A $1 - \alpha$ confidence interval for ϵ has the form $(\hat{\epsilon} - t_{1-\alpha, \widetilde{M}-1} \sqrt{\frac{\hat{\sigma}_\epsilon^2}{M}}, \hat{\epsilon} + t_{1-\alpha, \widetilde{M}-1} \sqrt{\frac{\hat{\sigma}_\epsilon^2}{M}})$ where $t_{1-\alpha, \widetilde{M}-1}$ is determined from t -distribution with $\widetilde{M} - 1$ degrees of freedom.

Next, algorithms are founded via our schemes, and the emerged conditional expectations in our schemes are simulated by means of least squares Monte Carlo method (see [4, 16–18]). Let *OLS* denote the ordinary least squares. Define the empirical probability measure $\nu_{i,M} = \frac{1}{M} \sum_{\hat{m}=1}^M \delta_{(\Delta W_i^{(i,\hat{m})}, X_i^{(i,\hat{m})}, \dots, X_N^{(i,\hat{m})})}$ where δ_x is the Dirac measure and $\{(\Delta W_i^{(i,\hat{m})}, X_i^{(i,\hat{m})}) : \hat{m} = 1, 2, \dots, M\}$ is the independent copies of $(\Delta W_i, X)$; the finite functional linear space $\mathcal{K}_{Y,i} := \{p_{Y,i}^{(1)}(\cdot), p_{Y,i}^{(2)}(\cdot), \dots, p_{Y,i}^{(K_{Y,i})}(\cdot)\}$, the basis function $p_{Y,i}^{(k)} : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\mathbb{E}[|p_{Y,i}^{(k)}(X_i)|^2] < +\infty$ and the finite functional linear space $\mathcal{K}_{Z,i} := \{p_{Z,i}^{(1)}(\cdot), p_{Z,i}^{(2)}(\cdot), \dots, p_{Z,i}^{(K_{Z,i})}(\cdot)\}$, the basis function $p_{Z,i}^{(k)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\mathbb{E}[|p_{Z,i}^{(k)}(X_i)|^2] < +\infty$ where $K_{Y,i}$ and $K_{Z,i}$ denote the dimension of the finite functional linear spaces $\mathcal{K}_{Y,i}$ and $\mathcal{K}_{Z,i}$. Suppose that $\mathcal{T}_L(x)$ is the truncation operator and it is defined as $\mathcal{T}_L(x) = (-L \vee x_1 \wedge L, \dots, -L \vee x_n \wedge L)$ for any finite $L > 0, x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Note that there are measurable, deterministic (but unknown) functions $y_i(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ and $z_i(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ for $i = 0, 1, \dots, N-1$ such that the solution (Y_i^π, Z_i^π) of the discrete BSDE (3.4) is given by $(Y_i^\pi, Z_i^\pi) := (y_i(X_i^\pi), z_i(X_i^\pi))$ (see Theorem 3.1 in [5]).

In what follows, we apply our **Algorithms** to two BSDEs with closed-form solutions.

Example 1. Consider the BSDE as below:

$$Y_t = 1 + \eta + \sin(\tau \mathbf{1}_d^\top W_T) + \int_t^T \min \left\{ 1, (Y_s - \eta - 1 - \frac{\sin(\tau \mathbf{1}_d^\top W_s)}{\exp(\tau^2 d(T-t)/2)})^2 \right\} ds - \int_t^T Z_s dW_s, \quad (4.1)$$

which appears in [19] and is used to illustrate the variance reduction problem with closed-form solutions. Here $\eta > 0; \tau > 0; \mathbf{1}_d$ is a d -dimensional vector with components all 1. Now, the solution to the above BSDE is

$$Y_t = 1 + \eta + \frac{\sin(\tau \mathbf{1}_d^\top W_t)}{\exp(\tau^2 d(T-t)/2)}, \quad (Z_t)_\lambda = \frac{\tau \cos(\tau \mathbf{1}_d^\top W_t)}{\exp(\tau^2 d(T-t)/2)},$$

Algorithm the stable high order predictor-corrector scheme based on (3.4)

1. Initialization $X_0 := x_0, y_N^{(M)}(\cdot) := \Phi(X_N^\pi)$
 2. sample $(t_i, x_i^{\widehat{m}})_{1 \leq i \leq N}$ by $X_{i+1}^{\pi, \widehat{m}} = X_i^{\pi, \widehat{m}} + b(t_i, X_i^{\pi, \widehat{m}})h + \sigma(t_i, X_i^{\pi, \widehat{m}})\Delta W_{i+1}$
 3. for $i = N - 1$ until 1
 4. for $\widehat{m} = 1$ until M
 5. set $S_{Z,i}^{(M)}(\mathbf{W}, \mathbf{X}) = \sum_{j=1}^m \lambda_{m,j} y_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}})(W_{i+j} - W_i)^\top$, where $\mathbf{W} = (W_i, \dots, W_{i+m+1}) \in (\mathbb{R}^d)^{m+2}$, $\mathbf{X} = (X_i^\pi, \dots, X_{i+m+1}^\pi) \in (\mathbb{R}^d)^{m+2}$;
compute $z_i^{(M)}(X_i^\pi) = \mathcal{T}_{C_z}(\text{OLS}(S_{Z,i}^{(M)}, \mathcal{K}_{Z,i}, \nu_{i,M}))$, where C_z denotes the upper bound of Z
 6. set $\widetilde{S}_{Y,i}^{(M)}(\mathbf{X}) = \sum_{j=1}^m y_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}}) + h \sum_{j=1}^m \widetilde{\gamma}_j f(t_{i+j}, X_{i+j}^{\pi, \widehat{m}}, y_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}}), z_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}}))$, compute $\widetilde{y}_i^{(M)}(X_i^\pi) = \mathcal{T}_{C_y}(\text{OLS}(\widetilde{S}_{Y,i}^{(M)}, \mathcal{K}_{Y,i}, \nu_{i,M}))$
where C_y denotes the upper bound of Y
 7. set $S_{Y,i}^{(M)}(\mathbf{X}) = \sum_{j=1}^m \alpha_j y_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}}) + h \gamma_0 f(t_i, X_i^{\pi, \widehat{m}}, \widetilde{y}_i^{(M)}(X_i^{\pi, \widehat{m}}), z_i^{(M)}(X_i^{\pi, \widehat{m}})) + h \sum_{j=1}^m \gamma_j f(t_{i+j}, X_{i+j}^{\pi, \widehat{m}}, y_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}}), z_{i+j}^{(M)}(X_{i+j}^{\pi, \widehat{m}}))$,
compute $y_i^{(M)}(X_i^\pi) = \mathcal{T}_{C_y}(\text{OLS}(S_{Y,i}^{(M)}, \mathcal{K}_{Y,i}, \nu_{i,M}))$
 8. end for
 9. end for
-

where $(Z_t)_\lambda$ is the λ th component of the d -dimensional function $Z_t \in \mathbb{R}^d$. Take $T = 1, \eta = 0.6, \tau = \frac{1}{\sqrt{d}}, d = 2, \widetilde{M} = 21, h = \frac{T}{N}$. The basis functions which are spanned by polynomials whose degree is 2 are applied to compute the value of $Y_i^{(M)}$ and $Z_i^{(M)}$. In the Tables, the notations CR and RT represent the convergence rate w.r.t. the time step sizes and the running time respectively. The unit of RT is the second. In the Figures, the notations GPC scheme and EAM scheme represent the scheme (3.4) and the usual explicit Adams methods from [8] respectively.

If we want to implement the **Algorithm**, we have to determine parameters. Specifically, the following equations should be satisfied for $m = 1$

$$\begin{cases} 0 = 1 - \widetilde{\alpha}_1, \\ 0 = -\widetilde{\alpha}_1 + \widetilde{\gamma}_1, \\ |\widetilde{\alpha}_1| \leq 1, \end{cases} \quad \begin{cases} 0 = 1 - \alpha_1, \\ 0 = -\alpha_1 + \gamma_0 + \gamma_1, \\ |\alpha_1| \leq 1, \end{cases} \quad \text{and} \quad \begin{cases} 0 = \lambda_{1,0}h + \lambda_{1,1}h, \\ 1 = \lambda_{1,1}h. \end{cases}$$

Thus, $\widetilde{\alpha}_1 = 1, \widetilde{\gamma}_1 = 1, \alpha_1 = 1$. Let $\gamma_0 = \frac{1}{2}$, then $\gamma_1 = \frac{1}{2}$. $\lambda_{1,0}h = -1, \lambda_{1,1}h = 1$. Now, the characteristic polynomial becomes $P(\zeta) = \zeta - 1$. Its root 1 fulfils Dahlquist's root condition. That is to say, this one-step scheme is stable and given as

$$\begin{cases} \widetilde{Y}_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi + h f_{i+1}^\pi], \\ Y_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi + \frac{1}{2} h \widetilde{f}_i^\pi + \frac{1}{2} h f_{i+1}^\pi], \\ Z_i^\pi = \mathbb{E}_i[Y_{i+1}^\pi \frac{(W_{i+1} - W_i)^\top}{h}]. \end{cases}$$

Analogously, we present the stable predictor-corrector type general linear multi-step scheme for $m = 2, 3, 4$.

Table 2: Errors and convergence rates based on the **Algorithm**

Step	N	M	$ Y_0 - Y_0^{(M)} $	95%CI of Y	$ Z_0 - Z_0^{(M)} $	95%CI of Z	RT
1	5	2778	1.257e-02	(8.461e-03, 1.668e-02)	1.279e-02	(8.685e-03, 1.690e-02)	0.1275
	10	5996	7.969e-03	(5.449e-03, 1.049e-02)	8.621e-03	(5.526e-03, 1.172e-02)	0.4401
	15	8809	6.877e-03	(4.757e-03, 8.998e-03)	7.393e-03	(4.432e-03, 1.035e-02)	1.548
	20	12018	5.276e-03	(3.966e-03, 6.586e-03)	6.158e-03	(3.734e-03, 8.582e-03)	2.992
	CR		1.021		1.004		
2	5	2778	7.960e-03	(5.440e-03, 1.048e-02)	8.776e-03	(6.257e-03, 1.130e-02)	0.1779
	10	5996	7.012e-04	(4.617e-04, 1.081e-03)	7.193e-03	(4.672e-03, 9.713e-03)	0.9788
	15	8809	6.128e-04	(4.006e-04, 8.250e-04)	6.328e-04	(3.887e-04, 8.769e-04)	2.254
	20	12018	4.276e-04	(2.967e-04, 5.586e-04)	4.450e-04	(2.150e-04, 6.749e-04)	3.795
	CR		1.998		2.001		
3	5	2778	6.604e-04	(4.163e-04, 9.045e-04)	6.860e-04	(4.737e-04, 8.982e-04)	0.2553
	10	5996	6.177e-04	(4.141e-04, 8.213e-04)	6.397e-04	(4.277e-04, 8.518e-04)	1.434
	15	8809	5.857e-05	(3.097e-05, 9.018e-05)	5.596e-05	(3.560e-05, 7.633e-05)	3.484
	20	12018	3.512e-05	(1.763e-05, 4.661e-05)	4.164e-05	(2.227e-05, 6.101e-05)	4.996
	CR		3.109		3.014		
4	5	2778	6.026e-05	(3.726e-05, 8.325e-05)	6.456e-05	(5.147e-05, 7.766e-05)	0.3046
	10	5996	5.645e-06	(3.708e-06, 7.582e-06)	5.717e-05	(4.407e-05, 7.026e-05)	1.6809
	15	8809	5.163e-06	(2.738e-06, 7.587e-06)	5.069e-06	(3.620e-06, 6.518e-06)	3.741
	20	12018	3.001e-06	(1.602e-06, 4.399e-06)	3.689e-06	(2.291e-06, 5.088e-06)	6.966
	CR		4.227		3.894		

For example, if $m = 3$, we provide the following three-step scheme

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i \left[\frac{1}{3}Y_{i+1}^\pi + \frac{1}{3}Y_{i+2}^\pi + \frac{1}{3}Y_{i+3}^\pi + \frac{39}{18}hf_{i+1}^\pi - \frac{2}{3}hf_{i+2}^\pi + \frac{1}{2}hf_{i+3}^\pi \right], \\ Y_i^\pi = \mathbb{E}_i \left[\frac{1}{3}Y_{i+1}^\pi + \frac{1}{3}Y_{i+2}^\pi + \frac{1}{3}Y_{i+3}^\pi + \frac{5}{6}h\tilde{f}_i^\pi - \frac{1}{3}hf_{i+1}^\pi + \frac{11}{6}hf_{i+2}^\pi - \frac{1}{3}hf_{i+3}^\pi \right], \\ Z_i^\pi = \mathbb{E}_i \left[3Y_{i+1}^\pi \frac{(W_{i+1}-W_i)^\top}{h} - \frac{3}{2}Y_{i+2}^\pi \frac{(W_{i+2}-W_i)^\top}{h} + \frac{1}{3}Y_{i+3}^\pi \frac{(W_{i+3}-W_i)^\top}{h} \right]. \end{cases}$$

Now, the characteristic polynomial becomes $P(\zeta) = \zeta^3 - \frac{1}{3}\zeta^2 - \frac{1}{3}\zeta - \frac{1}{3}$. Its roots $1, -\frac{1}{3} + \frac{1121}{2378}i, -\frac{1}{3} - \frac{1121}{2378}i$ fulfil Dahlquist's root condition. That is to say, this three-step scheme is stable.

Table 2 indicates: (i) The larger time points and simulations, the smaller error of closed-form solutions and numerical solutions no matter which-step scheme we utilize. (ii) If the number of time points and the number of simulations are fixed, the errors of closed-form solutions and numerical solutions become smaller as steps become bigger. (iii) If one's aim for the error of closed-form solutions and numerical solutions to reach given accuracy, one cannot only increase time points and simulations but also adopt multi-step methods, such as the scheme (3.4). In other words, this paper presents a stable high order method to

calculate numerical solutions of BSDEs.

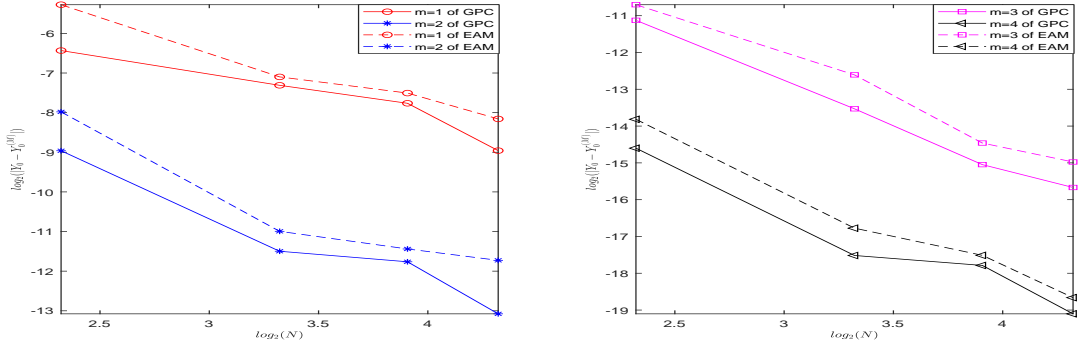


Figure 1: The plots of $\log_2(|Y_0 - Y_0^{(M)}|)$ versus $\log_2(N)$ with GPC scheme and EAM scheme, $M = 3000$

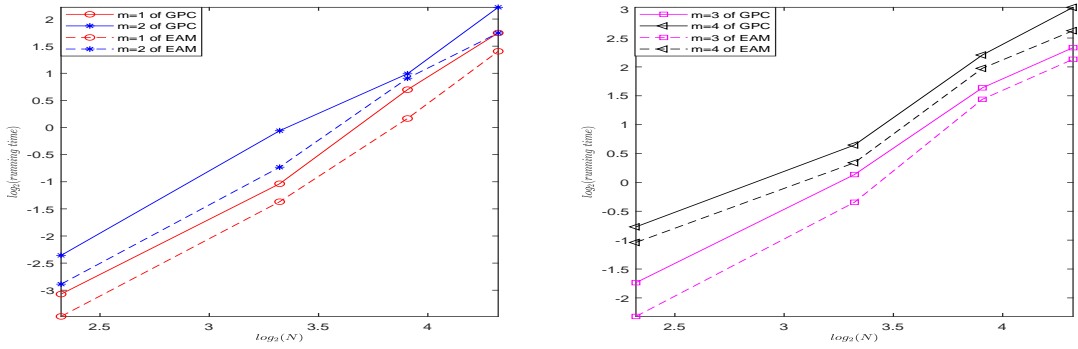


Figure 2: The plots of $\log_2(\text{running time})$ versus $\log_2(N)$ with GPC scheme and EAM scheme, $M = 3000$

Figure 1 compares the GPC scheme with the EAM scheme in terms of the accuracy. The left plot in Figure 1 displays the error of $|Y_0 - Y_0^{(M)}|$ for the one-step scheme and two-step scheme. The right plot describes the error of $|Y_0 - Y_0^{(M)}|$ for the three-step scheme and four-step scheme. Obviously, the accuracy of Y obtained by the GPC scheme is higher than that of the EAM scheme no matter the number of step is 1, 2, 3 or 4. Figure 2 compares the GPC scheme with the EAM scheme in terms of the computational cost. The left plot in Figure 1 displays the running time of these two methods for the one-step scheme and two-step scheme. The right plot describes the running time of these two methods for the three-step scheme and four-step scheme. It is straightforward that the running time of the EAM scheme is smaller than that of the GPC scheme no matter the number of step is 1, 2, 3 or 4.

Example 2. Consider the decoupled FBSDEs (taken from [38])

$$\begin{cases} dX_t = \frac{1}{1+2\exp(t+X_t)} dt + \frac{\exp(t+X_t)}{1+\exp(t+X_t)} dW_t, \\ X_0 = x, \\ -dY_t = \left(-\frac{2Y_t}{1+\exp(t+X_t)} - \frac{1}{2} \left(\frac{Y_t Z_t}{1+\exp(t+X_t)} - Y_t^2 Z_t \right) \right) dt - Z_t dW_t, \\ Y_T = \frac{\exp(T+X_T)}{1+\exp(T+X_T)}, \end{cases} \quad (4.2)$$

with the analytic solutions

$$\begin{cases} Y_t = \frac{\exp(t+X_t)}{1+\exp(t+X_t)}, \\ Z_t = \frac{(\exp(t+X_t))^2}{(1+\exp(t+X_t))^3}. \end{cases}$$

Take $T = 1, x = 1, d = 2, \widetilde{M} = 21, h = \frac{T}{N}$. The basis functions which are spanned by polynomials whose degree is 2 are applied to compute the value of $Y_i^{(M)}$ and $Z_i^{(M)}$.

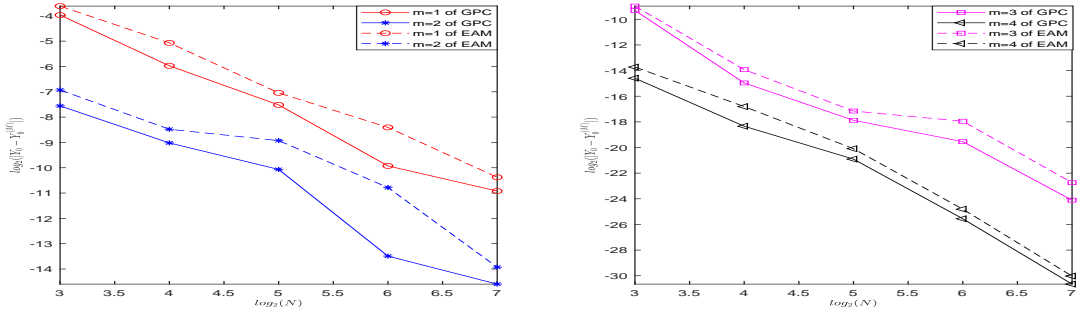


Figure 3: The plots of $\log_2(|Y_0 - Y_0^{(M)}|)$ versus $\log_2(N)$ with GPC scheme and EAM scheme, $M = 10000$

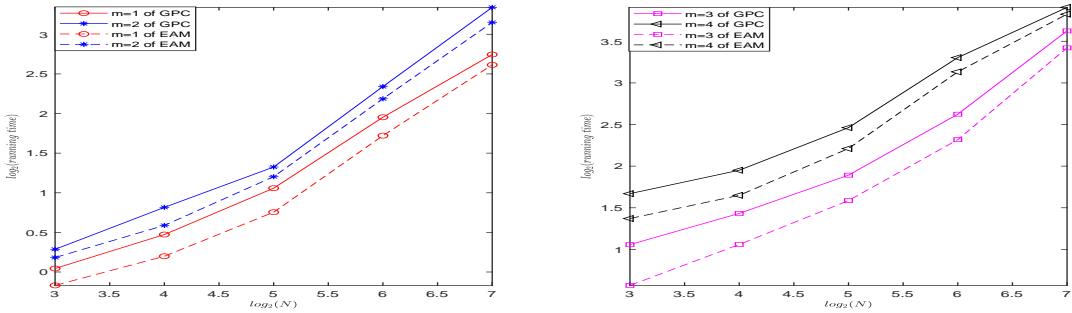


Figure 4: The plots of $\log_2(\text{running time})$ versus $\log_2(N)$ with GPC scheme and EAM scheme, $M = 10000$

Figure 3 compares the GPC scheme with the EAM scheme in terms of the error of $|Y_0 - Y_0^{(M)}|$. Figure 4 compares the GPC scheme with the EAM scheme in terms of the computational cost. These two figures imply that the GPC scheme possesses higher accuracy than the EAM scheme while the running time of the GPC scheme is bigger than that of the EAM scheme.

In what follows, we illustrate the case in which the condition (3.8) is satisfied and Dahlquist's root condition does not hold. In other words, we provide unstable numerical scheme for decoupled FBSDE (2.1). For $m = 2$, we introduce a two-step scheme as below

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i \left[3Y_{i+1}^\pi - 2Y_{i+2}^\pi + \frac{1}{2}hf_{i+1}^\pi - \frac{3}{2}hf_{i+2}^\pi \right], \\ Y_i^\pi = \mathbb{E}_i \left[3Y_{i+1}^\pi - 2Y_{i+2}^\pi + h\tilde{f}_i^\pi - \frac{3}{2}hf_{i+1}^\pi - \frac{1}{2}hf_{i+2}^\pi \right], \\ Z_i^\pi = \mathbb{E}_i \left[2Y_{i+1}^\pi \frac{(W_{i+1} - W_i)^\top}{h} - \frac{1}{2}Y_{i+2}^\pi \frac{(W_{i+2} - W_i)^\top}{h} \right]. \end{cases} \quad (4.3)$$

The characteristic polynomial of this two-step scheme is $P(\zeta) = \zeta^2 - 3\zeta + 2$. Its roots 1, 2 do not fulfil Dahlquist's root condition. That is to say, this two-step scheme is not stable.

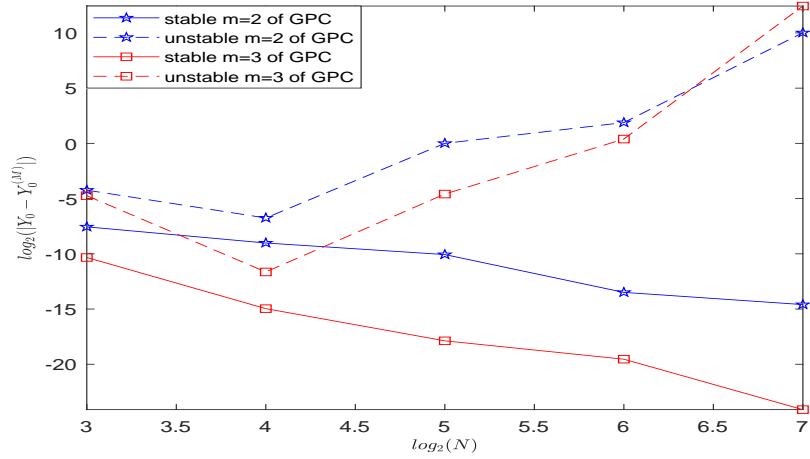


Figure 5: The plots of $\log_2(|Y_0 - Y_0^{(M)}|)$ versus $\log_2(N)$, $M = 10000$

For $m = 3$, we provide the following three-step scheme

$$\begin{cases} \tilde{Y}_i^\pi = \mathbb{E}_i \left[2Y_{i+1}^\pi + 5Y_{i+2}^\pi - 6Y_{i+3}^\pi + 2hf_{i+1}^\pi - 6hf_{i+2}^\pi - 2hf_{i+3}^\pi \right], \\ Y_i^\pi = \mathbb{E}_i \left[2Y_{i+1}^\pi + 5Y_{i+2}^\pi - 6Y_{i+3}^\pi - 3h\tilde{f}_i^\pi + 11hf_{i+1}^\pi - 15hf_{i+2}^\pi + hf_{i+3}^\pi \right], \\ Z_i^\pi = \mathbb{E}_i \left[3Y_{i+1}^\pi \frac{(W_{i+1} - W_i)^\top}{h} - \frac{3}{2}Y_{i+2}^\pi \frac{(W_{i+2} - W_i)^\top}{h} + \frac{1}{3}Y_{i+3}^\pi \frac{(W_{i+3} - W_i)^\top}{h} \right]. \end{cases} \quad (4.4)$$

The characteristic polynomial of the above scheme is $P(\zeta) = \zeta^3 - 2\zeta^2 - 5\zeta + 6$. Its roots $-2, 1, 3$ do not fulfil Dahlquist's root condition. That is to say, this three-step scheme is not stable.

Figure 5 provides the predictor-corrector method (3.4) in terms of the error of $|Y_0 - Y_0^{(M)}|$. Figure 5 indicates that the variation of errors is irregular for the unstable two-scheme (the scheme (4.3)) and the unstable three-scheme (the scheme (4.4)). That is to say, both the scheme (4.3) and the scheme (4.4) are not stable. Meanwhile, Figure 5 shows that the errors of $|Y_0 - Y_0^{(M)}|$ become smaller with the time step sizes N increasing for the stable two-scheme and the stable three-scheme (These two schemes come from **Example 1**). In other words, we verify that the given stable two-scheme and stable three-scheme are indeed stable by means of a numerical example.

References

- [1] V. Bally and G. Pagès, *Error analysis of the optimal quantization algorithm for obstacle problems*. Stochastic processes and their applications, 2003, 106(1): 1-40.
- [2] F. Bashforth and J. C. Adams, *An Attempt to Test the Theories of Capillary Action by Comparing the Theoretical and Measured Forms of Drops of Fluid*. University Press, 1883.
- [3] C. Beck, W. E and A. Jentzen, *Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations*. submitted. arXiv:1709.05963.
- [4] C. Bender and R. Denk, *A forward scheme for backward SDEs*. Stochastic processes and their applications, 2007, 117(12): 1793-1812.
- [5] C. Bender and T. Moseler, *Importance sampling for backward SDEs*. Stochastic Analysis and Applications, 2010, 28(2): 226-253.
- [6] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*. Stochastic Processes and their applications, 2004, 111(2): 175-206.
- [7] E. Buckwar, and R. Winkler, *Multi-step methods for SDES and their application to problems with small noise*. SIAM journal on numerical analysis, 2006, 44(2): 779-803.
- [8] J.-F. Chassagneux, *Linear multistep schemes for BSDEs*. SIAM Journal on Numerical Analysis, 2014, 52(6): 2815-2836.
- [9] J.-F. Chassagneux, *Linear multistep schemes for BSDEs*. arXiv preprint arXiv:1306.5548, 2013.
- [10] J.-F. Chassagneux and D. Crisan, *Runge-Kutta schemes for backward stochastic differential equations*. The Annals of Applied Probability, 2014, 24(2): 679-720.
- [11] F. Delarue and S. Menozzi, *A forward-backward stochastic algorithm for quasi-linear PDEs*. The Annals of Applied Probability, 2006, 16(1): 140-184.
- [12] J. Douglas, J. Ma and P. Protter, *Numerical methods for forward-backward stochastic differential equations*. The Annals of Applied Probability, 1996, 6(3): 940-968.

- [13] W. E, J. Q. Han and A. Jentzen, *Deep learning-based numerical methods for high dimensional parabolic partial differential equations and backward stochastic differential equations*. Communications in Mathematics and Statistics, 2017, 5(4): 349-380.
- [14] W. E, M. Hutzenthaler, A. Jentzen and T. Kruse, *On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations*. Journal of Scientific Computing, 2019, 79(3): 1534-1571.
- [15] W. Gautschi, *Numerical analysis. An introduction*, Birkhäuser, Boston, 1997.
- [16] E. Gobet, J-P. Lemor and X. Warin, *A regression-based Monte Carlo method to solve backward stochastic differential equations*. The Annals of Applied Probability, 2005, 15(3): 2172-2202.
- [17] E. Gobet and P. Turkedjiev, *Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions*. Mathematics of Computation, 2016, 85(299): 1359-1391.
- [18] E. Gobet and P. Turkedjiev, *Approximation of backward stochastic differential equations using Malliavin weights and least-squares regression*. Bernoulli, 2016, 22(1): 530-562.
- [19] E. Gobet and P. Turkedjiev, *Adaptive importance sampling in least-squares Monte Carlo algorithms for backward stochastic differential equations*. Stochastic Processes and their Applications, 2017, 127(4): 1171-1203.
- [20] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations. I, volume 8 of Springer Series in Computational Mathematics*. 1993.
- [21] J. Q. Han, A. Jentzen and W. E, *Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning*. arXiv preprint arXiv:1707.02568, 2017.
- [22] P. Henry-Labordere, *Counterparty risk valuation: a marked branching diffusion approach*. arXiv:1203.2369, 2012.
- [23] P. Henry-Labordere, N. Oudjane, X. Tan, N. Touzi and X. Warin, *Branching diffusion representation of semilinear PDEs and Monte Carlo approximation*. arXiv preprint arXiv:1603.01727, 2016.
- [24] P. Henry-Labordere, X. Tan and N. Touzi, *A numerical algorithm for a class of BSDEs via the branching process*. Stochastic Processes and their Applications, 2014, 124(2): 1112-1140.
- [25] M. Hutzenthaler, A. Jentzen and T. Kruse, *Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities*. arXiv preprint arXiv:1912.02571, 2019.
- [26] J. Ma, P. Protter, J. San Martín and S. Torres, *Numerical method for backward stochastic differential equations*. The Annals of Applied Probability, 2002, 12(1): 302-316.
- [27] J. Ma, J. Shen and Y. Zhao, *On numerical approximations of forward-backward stochastic differential equations*. SIAM Journal on Numerical Analysis, 2008, 46(5): 2636-2661.

- [28] C. N. Milstein and M. V. Tretyakov, *Numerical algorithms for forward-backward stochastic differential equations*. SIAM Journal on Scientific Computing, 2006, 28(2): 561-582.
- [29] W. E. Milne, *Numerical integration of ordinary differential equations*. The American Mathematical Monthly, 1926, 33(9): 455-460.
- [30] W. E. Milne, *Numerical Solution of Differential Equations*, John Wiley and Sons Inc, New York, 1953.
- [31] F. R. Moulton, *New Methods in Exterior Ballistics*. Chicago: University of Chicago Press, 1926.
- [32] E. J. Nyström, *Über die numerische Integration von Differentialgleichungen*. Societas scientiarum Fennica, 1925.
- [33] S. Peng, *Probability interpretation for systems of quasilinear parabolic partial differential equations*. Stochastics and Stochastic Reports. 1991, 37: 61-74.
- [34] E. Platen and N. Bruti-Liberati, *Numerical solution of stochastic differential equations with jumps in finance*. <https://doi.org/10.1080/14697688.2013.828240>.
- [35] M. J. Ruijter and C. W. Oosterlee, *A fourier cosine method for an efficient computation of solutions to BSDEs*, SIAM Journal on Scientific Computing, 2015, 37(2): A859-A889.
- [36] J. Yang and W. Zhao, *Convergence of recent multistep schemes for a forward-backward stochastic differential equation*. East Asian Journal on Applied Mathematics, 2015, 5(4): 387-404.
- [37] J. Zhang, *A numerical scheme for BSDEs*. The annals of applied probability, 2004, 14(1), 459-488.
- [38] W. Zhao, Y. Fu and T. Zhou, *New kinds of high-order multi-step schemes for forward backward stochastic differential equations*. SIAM Journal on Scientific Computing, 2014, 36(4): A1731-A1751.
- [39] W. Zhao, G. Zhang and L. Ju, *A stable multistep scheme for solving backward stochastic differential equations*. SIAM Journal on Numerical Analysis, 2010, 48(4): 1369-1394.