Solving Backward Doubly Stochastic Differential Equations through Splitting Schemes

Feng Bao [*] Yanzhao Cao [†] He Zhang [‡]

Abstract

A splitting scheme for backward doubly stochastic differential equations is proposed. The main idea is to decompose a backward doubly stochastic differential equation into a backward stochastic differential equation and a stochastic differential equation. The backward stochastic differential equation and the stochastic differential equation are then approximated by first order finite difference schemes, which results in a first order scheme for the backward doubly stochastic differential equation. Numerical experiments are conducted to illustrate the convergence rate of the proposed scheme.

Key words Backward doubly stochastic differential equations, Splitting up scheme, Stochastic partial differential equations, Zakai equations, Nonlinear filtering problems

AMS classification 60H15, 65H35, 65C20, 93E11

1 Introduction

The aim of this paper is to introduce a splitting algorithm for the following backward doubly stochastic differential equation (BDSDE):

$$Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)\, \mathrm{d}s - \int_t^T Z_s\, \mathrm{d}W_s + \int_t^T g(s, X_s, Y_s)\, \mathrm{d}\overleftarrow{B_s}, \quad (1.1)$$

where $0 \le t \le T$, $W := \{W_t\}_{t\ge0}$, $B := \{B_t\}_{t\ge0}$ are two independent Brownian motions and the stochastic process X_t is defined by $X_t = X_0 + W_t$, where X_0 is an initial random variable independent of W and B. The notation $\mathrm{d}\overleftarrow{B}$ stands

[*]Department of Mathematics, Florida State University, 1017 Academic Way, Tallahassee, FL 32304 (`fbao@fsu.edu`).

[†]Department of Mathematics, Auburn University, Auburn, AL 36849 (`yzc0009@auburn.edu`).

[‡]Department of Mathematics, Auburn University, Auburn, AL 36849 (`hzz0077@auburn.edu`).

for the backward Itô integral (see [32]), which is an Itô integral with backward propagation direction. The solution of the BDSDE (1.1) is a pair of stochastic processes (Y_t, Z_t). Here "doubly" refers to the fact that the equation is driven by two independent Brownian motions. Without the $d\overleftarrow{B}_t$ integral, the BDSDE is reduced to a standard backward stochastic differential equation (BSDE), which has been extensively studied [28, 29, 33, 39, 40].

The theory of BDSDEs was first studied in [34] to give a probabilistic interpretation for the solutions of the following class of semilinear stochastic partial differential equations (SPDEs)

$$\begin{aligned} u(t,x) =& \Phi(x) + \int_t^T \big(\mathcal{L}u(s,x) + f(s,x,u(s,x),(\nabla u\sigma)(s,x))\big)\,\mathrm{d}s \\ &+ \int_t^T g(s,x,u(s,x))\,\mathrm{d}\overleftarrow{B_s}, \quad (t,x) \in [0,T] \times \mathbb{R}^d \end{aligned} \tag{1.2}$$

through the relation

$$Y_t = u(t, X_t), \quad Z_t = \nabla u(t, X_t)\sigma(X_t). \tag{1.3}$$

The SPDE system (1.2) provides a stochastic version of parabolic type PDEs which could decribe uncertainties in modeling physical and engineering problems. For example, in the case that f is a linear function, the above SPDE solves the optimal filtering problem which aims to obtain the best estimate for the state of some stochastic dynamical system based on noisy partial observational data [5]. The optimal filtering problem is the key mission in data assimilation and it has been widely used in target tracking, weather forecasting, image processing, parameter estimation, etc.. In an optimal filtering problem, we need to obtain the conditional expectation for the target dynamical system given the observational information. It was proved ([37]) that the solution of the SPDE system (1.2) (in the linear case) is the conditional probability density for the dynamical system in the optimal filtering problem, which is used to calculate the desired conditional expectation. In the connection of the equivalence relation (1.3), the BDSDE (1.1) also provides solution for the optimal filtering problem. In a recent study ([2, 3, 4, 6]), we established a direct link between BDSDEs and optimal filtering problems. The main advantage of solving application problems via BDSDEs instead of SPDEs is twofold. First, solving BDSDEs is mesh free, thus unstructured methods such as Monte Carlo methods and stochastic meshfree approximations can be applied [10]. Moreover, scalable parallel numerical algorithms for BSDEs and BDSDEs enable us to benefit from recent advances in high performance parallel computing and even the deep learning techniques ([16, 18, 24]). Second, while it is very difficult to construct higher order methods to solve SPDEs, high order schemes for BDSDEs are relatively easy to construct ([7, 8, 9]).

In this paper we introduce a numerical scheme for the BDSDE eq. (1.1) using the splitting up method. Our work is inspired by the studies of splitting up method for linear SPDEs. The application of the splitting up methods to linear SPDEs was initiated by A. Bensoussan et al [11] where the SPDE is decomposed into a PDE and an SDE. Bensoussan's method was further developed in [12, 13, 25]. In particular, Gyöngy and Krylov [20], proved the convergence in the maximum norm.

To obtain a splitting up approximation for the BDSDE eq. (1.1), we decompose it into two equations, a BSDE which serves as a predictor or a pre-solving procedure, and an SDE which serves as an update procedure. Both can be solved using highly efficient numerical schemes ([19, 23, 38, 40]). In this paper, we construct a first order scheme by using the Milstein scheme on the SDE and a simple first order scheme on the BSDE. One of the advantages of our splitting up schemes, in comparison with the existing numerical schemes for BDSDEs ([1, 9]), is that it avoids the solve of Z_t in eq. (1.1), which significantly reduces the computing cost. It's also worthy to point out that the conventional splitting up methods under the SPDEs framework are focused on the case that both f and g in eq. (1.2) are linear functions while our methodology applies to more general nonlinear equations. In addition, the significance our splitting up method is boosted by some recent work of E, Han and Jentzen ([16, 21]), where a deep learning technique is used to solve fairly high dimensional BSDEs. Such a method can be applied to solve the BSDE, which is the most computational expensive component in our splitting up algorithm, thus can help solve high dimensional BDSDEs through our splitting up process.

The rest of this paper is organized as follows. In Section 2, we introduce some notations, assumptions and concepts as well as some known theoretical results of BDSDEs. In Section 3, we first present the splitting up method where the BDSDE is split into a BSDE and an SDE, and then prove the first order convergence. The numerical schemes with the corresponding numerical analysis are presented in Section 4, followed by three numerical examples in Section 5.

2 Preliminaries

Let $T > 0$ be a fixed terminal time, $(\Omega, \mathcal{F}, P)$ a probability space, and W and B two mutually independent Brownian motions on this space, with values in $\mathbb{R}^d$ and $\mathbb{R}^l$, respectively. For each $t \in [0, T]$, define two collections $\{\mathcal{F}_t\}_{0\leq t\leq T}$ and $\{\mathcal{G}_t\}_{0\leq t\leq T}$ by

$$\mathcal{F}_t := \mathcal{F}^W_{0,t} \vee \mathcal{F}^B_{t,T}, \text{ and } \mathcal{G}_t := \mathcal{F}^W_{0,t} \vee \mathcal{F}^B_{0,T},$$

where $\mathcal{F}^W_{s,t}$ and $\mathcal{F}^B_{s,t}$ are the completion of $\sigma\{W_r - W_s; s \leq r \leq t\}$ and $\sigma\{B_r - B_s; s \leq r \leq t\}$, respectively. Here $\{\mathcal{F}_t\}_{0\leq t\leq T}$ is neither increasing nor decreasing, while $\{\mathcal{G}_t\}_{0\leq t\leq T}$ is an increasing filtration. To simplify the presentation and make our analysis more readable, we assume throughout the paper that $d = l = 1$. The results obtained in this paper can be extended to multi-dimensional cases through similar procedures.

Denote by $\mathcal{M}^2([0,T];\mathbb{R})$ the set of all $\mathbb{R}$-valued, $\mathcal{F}_t$-measurable processes $\{\varphi(t)\}_{0\le t\le T}$ such that $E\int_0^T |\varphi(t)|^2\,\mathrm{d}t < \infty$, by $\mathcal{S}^2([0,T];\mathbb{R})$ the set of all $\mathbb{R}$-valued, $\mathcal{F}_t$-measurable processes $\{\varphi(t)\}_{0\le t\le T}$ such that $E\Big[\sup\limits_{0\le t\le T}|\varphi(t)|^2\Big] < \infty$, and by $L^2(\Omega,\mathcal{F}_T,P;\mathbb{R})$ the set of all $\mathcal{F}_T$-measurable random variable ξ such that $E|\xi|^2 < \infty$.

We assume that Φ, f and g satisfy the following regularity assumptions:

(H1) $\Phi \in C^3(\mathbb{R},\mathbb{R})$, $f \in C^3([0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R},\mathbb{R})$, and $g \in C^3([0,T]\times\mathbb{R}\times\mathbb{R},\mathbb{R})$. Here $C^k(A,B)$ denotes the set of functions of class C^k from A to B whose partial derivatives of order less than or equal to k are bounded.

(H2) $f\colon \Omega\times[0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ and $g\colon \Omega\times[0,T]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ are jointly measurable. For any $x,y,z\in\mathbb{R}$,

$$f(\cdot,x,y,z)\in\mathcal{M}^2([0,T];\mathbb{R}), \text{ and } g(\cdot,x,y)\in\mathcal{M}^2([0,T];\mathbb{R}). \tag{2.1}$$

(H3) f and g satisfy the Lipschitz conditions. For all $\omega\in\Omega$, $t,s\in[0,T]$, $x,\bar{x}\in\mathbb{R}$, $y,\bar{y}\in\mathbb{R}$, $z,\bar{z}\in\mathbb{R}$, there exists a constant $L>0$ such that

$$\begin{aligned}&|f(t,x,y,z)-f(s,\bar{x},\bar{y},\bar{z})|^2 \le L(|t-s|+|x-\bar{x}|^2+|y-\bar{y}|^2+|z-\bar{z}|^2),\\ &|g(t,x,y)-g(s,\bar{x},\bar{y})|^2 \le L(|t-s|+|x-\bar{x}|^2+|y-\bar{y}|^2).\end{aligned} \tag{2.2}$$

Moreover,

$$\sup_{0\le t\le T}\{|f(t,0,0,0)|^2+|g(t,0,0)|^2\} < L. \tag{2.3}$$

The following theorem is a collection of well posedness and regularity results on BDSDEs which will be used throughout the rest of the paper.

Theorem 2.1 *Let **(H1)-(H3)** hold.*

(1) (Theorem 1.1 in [34]) For any $\Phi(X_T)\in L^2(\Omega,\mathcal{F}_T,P;\mathbb{R})$, BDSDE (1.1) *has a unique solution $(Y,Z)\in\mathcal{S}^2([0,T];\mathbb{R})\times\mathcal{M}^2([0,T];\mathbb{R})$.*

(2) (Theorem 1.4 in [34]) There exists a positive constant M, independent of t, such that

$$E\Big[\sup_{0\le t\le T}|Y_t|^2+\int_0^T |Z_t|^2\,\mathrm{d}t\Big] \le M.$$

(2) (Lemma 4.2 in [9]) For $0\le s\le t\le T$, there exists some positive constant C, independent of t, such that

$$E[(Y_t-Y_s)^2] \le C(t-s), \quad |E[Y_t-Y_s]| \le C(t-s).$$

(3) (Lemma 2.3 in [34]) For any $t \leq s \leq T$, $(\nabla Y_s, \nabla Z_s)$ is the unique solution of the following variational equation

$$\nabla Y_s = \Phi'(X_T)\nabla X_T + \int_s^T \nabla f(r, X_r, Y_r, Z_r)\,\mathrm{d}r - \int_s^T \nabla Z_s\,\mathrm{d}W_r + \int_s^T \nabla g(r, X_r, Y_r)\,\mathrm{d}\overleftarrow{B_r}, \tag{2.4}$$

where ∇ is the gradient operator with respect to X_0 (X_0 denoting the initial condition for X_t),

$$\nabla f(s, X_s, Y_s, Z_s) := f_x(s, X_s, Y_s, Z_s)\nabla X_s + f_y(s, X_s, Y_s, Z_s)\nabla Y_s + f_z(s, X_s, Y_s, Z_s)\nabla Z_s,$$
$$\nabla g(s, X_s, Y_s) := g_x(s, X_s, Y_s)\nabla X_s + g_y(s, X_s, Y_s)\nabla Y_s.$$

Here we use subscripts to indicate partial differentiations.

(4) (Lemma 4.4 in [1]) $\{Z_t\}_{0\leq t\leq T}$ has an a.s. continuous version which is given by

$$Z_t = \nabla Y_t.$$

Furthermore, with the assumptions of the theorem and through similar estimation techniques for the variation equation for Y_t, we have

$$E[(Z_t - Z_s)^2] \leq C(t-s), \quad |E[Z_t - Z_s]| \leq C(t-s), \tag{2.5}$$

for some positive constant C, independent of t.

3 Splitting up method and convergence analysis

In this section, we introduce the splitting up framework for BDSDE (1.1) and show that our splitting up system provides a first order approximation for the original BDSDE.

3.1 Splitting up method

Let $0 = t_0 < t_1 < \cdots < t_N = T$ be an uniform partition of $[0, T]$ with partition size $\Delta t := \frac{T}{N}$, where N is a positive integer. Denote $\Delta W_i := W_{t_{i+1}} - W_{t_i}$ and $\Delta B_i := B_{t_{i+1}} - B_{t_i}$. The approximation $Y_i(t)$ to the solution Y_t of BDSDE (1.1) is defined recursively on each time interval $[t_i, t_{i+1}), i = 1, \cdots, N-1$ as follows. Set $Y_N(T) = \Phi(X_T)$. First define $\tilde{Y}_i(t), t_i \leq t < t_{i+1}$, to be the solution of the BSDE:

$$\tilde{Y}_i(t) = Y_{i+1}(t_{i+1}) + \int_t^{t_{i+1}} f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s))\,\mathrm{d}s - \int_t^{t_{i+1}} \tilde{Z}_i(s)\,\mathrm{d}W_s, \ (BSDE) \tag{3.1}$$

Then $Y_i(t)$ is defined as the solution of the SDE:

$$Y_i(t) = \tilde{Y}_i(t) + \int_t^{t_{i+1}} g(s, X_s, Y_i(s))\, \mathrm{d}\overleftarrow{B}_s. \; (SDE) \tag{3.2}$$

In this way, the approximation of BDSDE (1.1) on subinterval $[t_i, t_{i+1})$ is split into two steps. In the first step, we solve the BSDE (3.1). In the second step, we use the solution $\tilde{Y}_i(t_i)$ of the BSDE at time t_i as the terminal value at time t_{i+1} and solve the SDE (3.2) on $[t_i, t_{i+1})$. These implicit equations are solved using iterative techniques. Here, the solution $(\tilde{Y}_i, \tilde{Z}_i)$ of the BSDE (3.1) plays the role of the intermediate solution before we incorporate the $d\overleftarrow{B}$ integral. Hence $(\tilde{Y}_i(t), \tilde{Z}_i(t))$ is $\mathcal{F}_{0,t}^W \vee \mathcal{F}_{t_{i+1},T}^B$ measurable for any $t \in [t_i, t_{i+1})$. On the other hand, the solution $Y_i(t)$ of the SDE is $\mathcal{F}_{0,t_{i+1}}^W \vee \mathcal{F}_{t,T}^B$ measurable. We let $\tilde{Z}_i$ be our approximation for the solution Z_t for $t \in [t_i, t_{i+1})$. It's worthy noting that Y_i incorporates the $d\overleftarrow{B}$ integral as a solution for SDE, and $\tilde{Z}_i$ incorporates the $d\overleftarrow{B}$ integral only through the variation relationship with $Y_{i+1}(t_{i+1})$ at temporal grid points. Moreover, letting $t \to t_{i+1} - 0$, we have

$$\lim_{t \to t_{i+1} - 0} Y_i(t) = Y_{i+1}(t_{i+1}).$$

Therefore the approximate process $\bar{Y}_t := \sum_{i=0}^{N-1} Y_i(t)\, 1_{[t_i, t_{i+1})}(t) + \Phi(X_T)\, 1_T(t)$ has continuous trajectories.

3.2 Convergence analysis

We now turn to the convergence analysis for the proposed splitting up system (3.1)-(3.2) in approximating the BDSDE eq. (1.1). We first state the main result of our analysis which shows that our splitting up system provides a first order mean square approximation for solution Y_t and half order mean square approximation for solution Z_t.

Theorem 3.1 *Assume that **(H1)-(H3)** hold. Then for sufficiently large N, there exists a positive constant C, independent of Δt and X_0, such that*

$$\max_{1 \le i \le N} \left(E\big[E_{t_i}[Y_i(t_i)] - Y_{t_i}\big]^2 \right) \le C\Delta t^2, \quad \max_{1 \le i \le N} \left(E\big[\tilde{Z}_i(t_i) - Z_{t_i}\big]^2 \right) \le C\Delta t, \tag{3.3}$$

where $E_{t_i}[\cdot]$ denotes the conditional expectation over the σ-algebra $\mathcal{G}_{0,t_i} = \mathcal{F}_{0,t_i}^W \vee \mathcal{F}_{0,T}^B$.

To prove the theorem, we need several estimations concerning the intermediate approximation $\tilde{Y}_i$ and $\tilde{Z}_i$ given by eq. (3.1).

Lemma 3.2 *Under the assumptions **(H1)-(H3)**, for any given interval* $[t_i, t_{i+1})$*, there is a constant* C*, independent of* Δt *and* X_0*, such that*

$$\sup_{t\in[t_i,t_{i+1})} E[(\tilde{Y}_i(t) - Y_{i+1}(t_{i+1}))^2] \le C\Delta t, \quad |E_{t_i}[\tilde{Y}_i(t) - Y_{i+1}(t_{i+1})]| \le C\Delta t,$$

$$\sup_{t\in[t_i,t_{i+1})} E[(\tilde{Z}_i(t) - \tilde{Z}_i(t_i))^2] \le C\Delta t, \quad |E_{t_i}[\tilde{Z}_i(t) - \tilde{Z}_i(t_i)]| \le C\Delta t.$$

Proof: The estimations in the lemma follow directly from Theorem 2.1 with the special case $g \equiv 0$. □

Lemma 3.3 *Under the assumptions **(H1)-(H3)**, for any given interval* $[t_i, t_{i+1})$*, there is a constant* C*, independent of* Δt *and* X_0*, such that*

$$\sup_{t\in[t_i,t_{i+1})} \left(E[(\tilde{Y}_i(t) - Y_t)^2]\right) \le CE[(E_{t_{i+1}}[Y_{i+1}(t_{i+1})] - Y_{t_{i+1}})^2] + C\Delta t.$$

Proof: Subtracting eq. (1.1) for $t \in [t_i, t_{i+1})$ from eq. (3.1), and taking the conditional expectation $E[\cdot|\mathcal{G}_{0,t_{i+1}}]$ gives

$$\begin{aligned}\tilde{Y}_i(t) - Y_t &= E_{t_{i+1}}[Y_{i+1}(t_{i+1})] - Y_{t_{i+1}} - \int_t^{t_{i+1}} [\tilde{Z}_i(s) - Z_s]\,\mathrm{d}W_s \\ &+ \int_t^{t_{i+1}} [f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s)) - f(s, X_s, Y_s, Z_s)]\,\mathrm{d}s - \int_t^{t_{i+1}} g(s, X_s, Y_s)\,\mathrm{d}\overleftarrow{B}_s.\end{aligned} \tag{3.4}$$

Note that $\tilde{Y}_i(t)$ is $\mathcal{F}^W_{0,t} \vee \mathcal{F}^B_{t_{i+1},T}$ measurable for $t < t_{i+1}$, thus it is $\mathcal{F}^W_{0,t} \vee \mathcal{F}^B_{t,T}$ measurable, i.e. $\mathcal{G}_t$ measurable. Applying the generalized Itô's Lemma (see Lemma 1.3 in [34]) to $|\tilde{Y}_i(t) - Y_t|^2$ and taking the expectation, we have, using Young's inequality with $\varepsilon = \frac{1}{2L}$, and assumption **(H3)**,

$$\begin{aligned}&E|\tilde{Y}_i(t) - Y_t|^2 + \int_t^{t_{i+1}} E|\tilde{Z}_i(s) - Z_s|^2\,\mathrm{d}s \\ &\le E|E_{t_{i+1}}[Y_{i+1}(t_{i+1})] - Y_{t_{i+1}}|^2 + 2\int_t^{t_{i+1}} L(E|X_s|^2 + E|Y_s|^2)\,\mathrm{d}s \\ &+ 2\int_t^{t_{i+1}} E|g(s,0,0)|^2\,\mathrm{d}s + (2L + \frac{1}{2})\int_t^{t_{i+1}} E|\tilde{Y}_i(s) - Y_s|^2\,\mathrm{d}s \\ &+ \frac{1}{2}\int_t^{t_{i+1}} E|\tilde{Z}_i(s) - Z_s|^2\,\mathrm{d}s.\end{aligned}$$

The desired result follows from Gronwall's inequality, assumption **(H3)**, and Theorem 2.1. □

Lemma 3.4 *Under assumptions **(H1)-(H3)**, for any given interval* $[t_i, t_{i+1})$*, there exists a constant* C*, independent of* Δt *and* X_0*, such that*

$$\sup_{t\in[t_i,t_{i+1})} \left(E[(Y_i(t) - Y_t)^2]\right) \le CE[(E_{t_{i+1}}[Y_{i+1}(t_{i+1})] - Y_{t_{i+1}})^2] + C\Delta t. \tag{3.5}$$

Proof: Note that

$$Y_i(t) - Y_t = Y_i(t) - \tilde{Y}_i(t) + \tilde{Y}_i(t) - Y_t.$$

This result is then a direct consequence of Lemma 3.3, Itô's isometry, and the assumption **(H3)**. $\square$

Combining Theorem 2.1 and Lemma 3.2, using Young's inequality, we arrive an estimate on $E[(\tilde{Z}_i(t) - Z_t)^2]$.

Lemma 3.5 *Under the assumptions **(H1)-(H3)**, for any given interval $[t_i, t_{i+1})$, there exits constant a C, independent of Δt and X_0, such that*

$$\sup_{t\in[t_i,t_{i+1})} \big(E[(\tilde{Z}_i(t) - Z_t)^2]\big) \le (1+\epsilon_0)E[(\tilde{Z}_i(t_{i+1}-0) - Z_{t_{i+1}})^2] + C\Delta t,$$

for some suitable $\epsilon_0 > 0$.

Proof of Theorem 3.1: The main ingredients of the proof are the estimations for the errors $Y_i(t_i) - Y_{t_i}$ and $\tilde{Z}_i(t_i) - Z_{t_i}$. Once these estimations are obtained, the desired result of the theorem is the consequence of application of the discrete Gronwall inequality.

Estimation for the error $Y_i(t_i) - Y_{t_i}$.
Subtracting eq. (1.1) for $t \in [t_i, t_{i+1})$ from eq. (3.2) and substituting eq. (3.4) with result we have that for $t = t_i$

$$\begin{aligned} Y_i(t_i) - Y_{t_i} &= E_{t_{i+1}}[Y_{i+1}(t_{i+1})] - Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} [g(s, X_s, Y_i(s)) - g(s, X_s, Y_s)]\,\mathrm{d}\overleftarrow{B_s} \\ &+ \int_{t_i}^{t_{i+1}} [f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s)) - f(s, X_s, Y_s, Z_s)]\,\mathrm{d}s - \int_{t_i}^{t_{i+1}} [\tilde{Z}_i(s) - Z_s]\,\mathrm{d}W_s. \end{aligned} \tag{3.6}$$

To simplify notation in subsequent derivations, we shall use the following shorthand notation:

$$\begin{aligned} &e_y^i := E_{t_i}[Y_i(t_i)] - Y_{t_i},\ e_z^{i+1} := \tilde{Z}_i(t_{i+1}-0) - Z_{t_{i+1}}, \\ &\delta f^i(s) := f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s)) - f(s, X_s, Y_s, Z_s), \\ &\delta g^i(s) := g(s, X_s, Y_i(s)) - g(s, X_s, Y_s). \end{aligned}$$

Taking the conditional expectation $E_{t_i}[\cdot]$ on both sides of the above yields

$$\begin{aligned} E_{t_i}[Y_i(t_i)] - Y_{t_i} &= E_{t_i}[E_{t_{i+1}}[Y_{i+1}(t_{i+1})] - Y_{t_{i+1}}] \\ &\quad + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)]\,\mathrm{d}s + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta g^i(s)]\,\mathrm{d}\overleftarrow{B_s}. \end{aligned} \tag{3.7}$$

Here we have used Fubini's theorem and the fact that Y_{t_i} is $\mathcal{F}_{0,t_i}^W \vee \mathcal{F}_{0,T}^B$ measurable, i.e. $\mathcal{G}_{0,t_i}$ measurable.

Next we consider the mean square estimation for e_y^i. Square and then take the expectation on both sides of eq. (3.7) to obtain

$$E[(e_y^i)^2] = E[(E_{t_i}[e_y^{i+1}])^2] + E\Big[\Big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)]\,\mathrm{d}s + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta g^i(s)]\,\mathrm{d}\overleftarrow{B_s}\Big)^2\Big]$$
$$+ 2E\Big[\big(E_{t_i}[e_y^{i+1}]\big)\cdot\Big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)]\,\mathrm{d}s + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta g^i(s)]\,\mathrm{d}\overleftarrow{B_s}\Big)\Big]. \tag{3.8}$$

Using the elementary inequality $(a+b)^2 \le 2(a^2+b^2)$ on eq. (3.8), we have

$$E[(e_y^i)^2] \le I_1 + I_2 + I_3 + I_4 + I_5, \tag{3.9}$$

where

$$\begin{aligned}
I_1 &:= E[(E_{t_i}[e_y^{i+1}])^2],\\
I_2 &:= 2E\Big[\big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)]\,\mathrm{d}s\big)^2\Big],\\
I_3 &:= 2E\Big[\big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta g^i(s)]\,\mathrm{d}\overleftarrow{B_s}\big)^2\Big],\\
I_4 &:= 2E\Big[\big(E_{t_i}[e_y^{i+1}]\big)\cdot\big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)]\,\mathrm{d}s\big)\Big],\\
I_5 &:= 2E\Big[\big(E_{t_i}[e_y^{i+1}]\big)\cdot\big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta g^i(s)]\,\mathrm{d}\overleftarrow{B_s}\big)\Big].
\end{aligned}$$

By Cauchy's inequality, Jensen's inequality, and the assumptions **(H1)-(H3)**, we have

$$\begin{aligned}
I_2 &\le 2\Delta t E\Big[\int_{t_i}^{t_{i+1}} (E_{t_i}[\delta f^i(s)])^2\,\mathrm{d}s\Big]\\
&\le 2\Delta t \int_{t_i}^{t_{i+1}} L\Big(E[(\tilde{Y}_i(s) - Y_s)^2] + E\big[(\tilde{Z}_i(s) - Z_s)^2\big]\Big)\,\mathrm{d}s.
\end{aligned}$$

Then, from Lemma 3.3 and Lemma 3.5, we get

$$I_2 \le C(\Delta t)^2 E[(e_y^{i+1})^2] + 2L(1+\epsilon_0)(\Delta t)^2 E\big[(e_z^{i+1})^2\big] + O(\Delta t^3). \tag{3.10}$$

Next we estimate I_3. To simplify the presentation, we use abbreviated notations $\Theta_r = (r, X_r, Y_r)$ and $\tilde{\Theta}_r = (r, X_r, Y_i(r))$, and use subscripts of function g to indicate partial differentiations. We also use $\mathrm{d}[X]_r$ to denote the quadratic variation of X_r, and $\mathrm{d}[X,Y]_r$ the quadratic covariation of X_r and Y_r. In order to derive an estimation for I_3, we first apply the Itô-Taylor expansions for $g(\Theta_s)$

and $g(\tilde{\Theta}_s)$ on interval $[s, t_{i+1}]$ to obtain

$$g(\Theta_{t_{i+1}}) = g(\Theta_s) + \int_s^{t_{i+1}} g_t(\Theta_r)\,\mathrm{d}r + \int_s^{t_{i+1}} g_x(\Theta_r)\,\mathrm{d}X_r + \int_s^{t_{i+1}} g_y(\Theta_r)\,\mathrm{d}Y_r$$
$$+ \frac{1}{2}\int_s^{t_{i+1}} g_{xx}(\Theta_r)\,\mathrm{d}[X]_r + \frac{1}{2}\int_s^{t_{i+1}} g_{yy}(\Theta_r)\,\mathrm{d}[Y]_r + \int_s^{t_{i+1}} g_{xy}(\Theta_r)\,\mathrm{d}[X,Y]_r, \tag{3.11}$$

and

$$g(\tilde{\Theta}_{t_{i+1}}) = g(\tilde{\Theta}_s) + \int_s^{t_{i+1}} g_t(\tilde{\Theta}_r)\,\mathrm{d}r + \int_s^{t_{i+1}} g_x(\tilde{\Theta}_r)\,\mathrm{d}X_r + \int_s^{t_{i+1}} g_y(\tilde{\Theta}_r)\,\mathrm{d}Y_i(r)$$
$$+ \frac{1}{2}\int_s^{t_{i+1}} g_{xx}(\tilde{\Theta}_r)\,\mathrm{d}[X]_r + \frac{1}{2}\int_s^{t_{i+1}} g_{yy}(\tilde{\Theta}_r)\,\mathrm{d}[Y_i]_r + \int_s^{t_{i+1}} g_{xy}(\tilde{\Theta}_r)\,\mathrm{d}[X,Y_i]_r. \tag{3.12}$$

Note that $dX_r = dW_r$, it then follows from the generalized Itô's Lemma (see Lemma 1.3 in [34]) that

$$\mathrm{d}[X]_r = \mathrm{d}r,\;\; \mathrm{d}[Y]_r = -g^2(\Theta_r)\,\mathrm{d}r + (Z_r)^2\,\mathrm{d}r,\;\; \mathrm{d}[X,Y]_r = Z_r\,\mathrm{d}r$$
$$\mathrm{d}[Y_i]_r = -g^2(\tilde{\Theta}_r)\,\mathrm{d}r + (\tilde{Z}_i(r))^2\,\mathrm{d}r,\;\; \mathrm{d}[X,Y_i]_r = \tilde{Z}_i(r)\,\mathrm{d}r.$$

Subtracting eq. (3.11) from eq. (3.12), we have

$$g(\tilde{\Theta}_s) - g(\Theta_s) = g(\tilde{\Theta}_{t_{i+1}}) - g(\Theta_{t_{i+1}}) + \int_s^{t_{i+1}} [(g_y\cdot g)(\tilde{\Theta}_r) - (g_y\cdot g)(\Theta_r)]\,\mathrm{d}\overleftarrow{B_r}$$
$$- \int_s^{t_{i+1}} [(g_x(\tilde{\Theta}_r) + g_y(\tilde{\Theta}_r)\tilde{Z}_i(r)) - (g_x(\Theta_r) + g_y(\Theta_r)Z_r)]\,\mathrm{d}W_r + R^i_{g,Y}(s), \tag{3.13}$$

where $R^i_{g,Y}$ contains all the $\int_s^{t_{i+1}} \cdot\,\mathrm{d}r$ integrals:

$$\begin{aligned}
R^i_{g,Y}(s) :=& - \int_s^{t_{i+1}} [g_t(\tilde{\Theta}_r) - g_t(\Theta_r)]\,\mathrm{d}r \\
&+ \int_s^{t_{i+1}} [g_y(\tilde{\Theta}_r) f(r, X_r, \tilde{Y}_i(r), \tilde{Z}_i(r)) - g_y(\Theta_r) f(r, X_r, Y_r, Z_r)]\,\mathrm{d}r \\
&- \frac{1}{2}\int_s^{t_{i+1}} [g_{xx}(\tilde{\Theta}_r) + g_{yy}(\tilde{\Theta}_r)(\tilde{Z}_i(r))^2 - g_{xx}(\Theta_r) - g_{yy}(\Theta_r)(Z_r)^2]\,\mathrm{d}r \\
&+ \frac{1}{2}\int_s^{t_{i+1}} [g_{yy}(\tilde{\Theta}_r) g^2(\tilde{\Theta}_r) - g_{yy}(\Theta_r) g^2(\Theta_r)]\,\mathrm{d}r \\
&- \int_s^{t_{i+1}} [g_{xy}(\tilde{\Theta}_r)\tilde{Z}_i(r) - g_{xy}(\Theta_r) Z_r]\,\mathrm{d}r,
\end{aligned}$$

and it's easy to see that $\sup\limits_{s\in[t_i, t_{i+1})} E[(R^i_{g,Y}(s))^2] = O(\Delta t^2)$.

Taking the conditional expectation $E_{t_i}[\cdot]$ on both sides of eq. (3.13), we have

$$E_{t_i}[\delta g^i(s)] = E_{t_i}[\delta g^i(t_{i+1})] + E_{t_i}\left[\int_s^{t_{i+1}} \delta(g_y g)^i(r)\,\mathrm{d}\overleftarrow{B_r}\right] + E_{t_i}[R^i_{g,Y}(s)],$$

where $\delta(g_y g)^i(r) := (g_y \cdot g)(\tilde{\Theta}_r) - (g_y \cdot g)(\Theta_r)$. By the above estimation, Itô's isometry, the elementary inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, and Jensen's inequality, we obtain

$$\begin{aligned} I_3 &= 2E\bigg(\int_{t_i}^{t_{i+1}} (E_{t_i}[\delta g^i(s)])^2\,\mathrm{d}s\bigg) \\ &\le 6\int_{t_i}^{t_{i+1}} \bigg(E\big[\big(E_{t_i}[\delta g^i(t_{i+1})]\big)^2\big] + \int_s^{t_{i+1}} E[(E_{t_i}[\delta(g_y g)^i(r)])^2]\,\mathrm{d}r + E[(R^i_{g,Y}(s))^2]\bigg)\,\mathrm{d}s. \end{aligned} \tag{3.14}$$

Using Lemma 3.4, the assumptions **(H1)-(H3)** and Jensen's inequality, we have

$$I_3 \le 6|g_y|^2_\infty \Delta t E[(e_y^{i+1})^2] + C\Delta t^2 E[(e_y^{i+1})^2] + O(\Delta t^3). \tag{3.15}$$

We now turn to the estimation of I_4. First we decompose $\delta f^i(s)$, which is the abbreviation for $f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s)) - f(s, X_s, Y_s, Z_s)$, into three parts to write I_4 as

$$I_4 = 2E\bigg[E_{t_i}[e_y^{i+1}] \cdot \int_{t_i}^{t_{i+1}} \big(E_{t_i}[\delta f^{i,a}] + E_{t_i}[\delta f^{i,b}] + E_{t_i}[\delta f^{i,c}]\big)\,\mathrm{d}s\bigg],$$

where

$$\begin{aligned} \delta f^{i,a} &:= f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s)) - f(t_{i+1}, X_{t_{i+1}}, Y_{i+1}(t_{i+1}), \tilde{Z}_i(t_{i+1}-0)), \\ \delta f^{i,b} &:= f(t_{i+1}, X_{t_{i+1}}, Y_{i+1}(t_{i+1}), \tilde{Z}_i(t_{i+1}-0)) - f(t_{i+1}, X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}}), \\ \delta f^{i,c} &:= f(t_{i+1}, X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}}) - f(s, X_s, Y_s, Z_s). \end{aligned}$$

By Itô-Taylor expansion and Lemma 3.2, we see that $E_{t_i}[\delta f^{i,a}] = O(\Delta t)$. Hence

$$2E\bigg[E_{t_i}[e_y^{i+1}] \cdot \int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^{i,a}]\,\mathrm{d}s\bigg] \le \frac{\Delta t}{3} E\big[(e_y^{i+1})^2\big] + O(\Delta t^3).$$

On the other hand, by Theorem 2.1, we have $E[\delta f^{i,c}] = O(\Delta t)$. Therefore, it follows from the properties of conditional expectations and Youngs inequality that

$$\begin{aligned} &2E\bigg[E_{t_i}[e_y^{i+1}] \cdot \int_{t_i}^{t_{i+1}} \big(E_{t_i}[\delta f^{i,c}]\big)\,\mathrm{d}s\bigg] = 2E\bigg[E\big[E_{t_i}[e_y^{i+1}] \cdot \int_{t_i}^{t_{i+1}} \big(E_{t_i}[\delta f^{i,c}]\big)\,\mathrm{d}s|\mathcal{F}^B_{t_i,t_{i+1}}\big]\bigg] \\ &= 2E[e_y^{i+1}] \int_{t_i}^{t_{i+1}} E[\delta f^{i,c}]\,\mathrm{d}s \le \frac{\Delta t}{3} E\big[(e_y^{i+1})^2\big] + O(\Delta t^3). \end{aligned} \tag{3.16}$$

Putting the above estimates together, then using the assumptions **(H1)-(H3)** and Youngs inequality, we obtain

$$
\begin{aligned}
I_4 &\leq \Delta t E\big[(E_{t_i}[e_y^{i+1}])^2\big] + 6|f_y|_\infty^2 \Delta t E[(e_y^{i+1})^2] + 6|f_z|_\infty^2 \Delta t E[(E_{t_i}[e_z^{i+1}])^2] + O(\Delta t^3) \\
&= (1+6|f_y|_\infty^2)\Delta t E[(e_y^{i+1})^2] + 6|f_z|_\infty^2 \Delta t E\big[(E_{t_i}[e_z^{i+1}])^2\big] + O(\Delta t^3).
\end{aligned} \tag{3.17}
$$

To estimate the last term, we use an argument similar to (3.16).

$$
\begin{aligned}
I_5 =& 2E\Big[E[E_{t_i}[e_y^{i+1}] \cdot \int_{t_i}^{t_{i+1}} E_{t_i}[\delta g^i(t_{i+1})] \,\mathrm{d}\overleftarrow{B_s} | \mathcal{F}^B_{t_i,t_{i+1}}]\Big] \\
&+ 2E\Big[E[E_{t_i}[e_y^{i+1}] \cdot \int_{t_i}^{t_{i+1}} (E_{t_i}[\delta g^i(s) - \delta g^i(t_{i+1})]) \,\mathrm{d}\overleftarrow{B_s} | \mathcal{F}^B_{t_i,t_{i+1}}]\Big] \\
&= 2E\Big[E_{t_i}[e_y^{i+1}] E_{t_i}[\delta g^i(t_{i+1})]\Big] E[\Delta B_i] + 2E[e_y^{i+1}] \cdot E\Big[\int_{t_i}^{t_{i+1}} (E_{t_i}[\delta g^i(s) - \delta g^i(t_{i+1})]) \,\mathrm{d}\overleftarrow{B_s}\Big] = 0.
\end{aligned}
$$

Substituting the above estimations eqs. (3.10), (3.15) and (3.17) into eq. (3.9), we have

$$
\begin{aligned}
E[(e_y^i)^2] \leq& E[(E_{t_i}[e_y^{i+1}])^2] + C_y^1 \Delta t E\big[(e_y^{i+1})^2\big] + 6|f_z|_\infty^2 \Delta t E\big[(E_{t_i}[e_z^{i+1}])^2\big] \\
&+ C\Delta t^2 E[(e_y^{i+1})^2] + 2L(1+\epsilon_0)\Delta t^2 E\big[(E_{t_i}[e_z^{i+1}])^2\big] + O(\Delta t^3),
\end{aligned} \tag{3.18}
$$

where $C_y^1 := 1 + 6|f_y|_\infty^2 + 6|g_y|_\infty^2$ is a constant.

Estimation for the error $\tilde{Z}_i(t_i) - Z_{t_i}$.
In order to derive an estimation for $\tilde{Z}_i(t_i) - Z_{t_i}$, we subtract eq. (1.1) for $t \in [t_i, t_{i+1})$ from eq. (3.4), let $t = t_i$, multiply both sides of the resulting equation by ΔW_i, and then take the conditional expectation $E_{t_i}[\cdot]$ to obtain

$$
\begin{aligned}
\int_{t_i}^{t_{i+1}} E_{t_i}[\tilde{Z}^i(s) - Z_s] \,\mathrm{d}s =& E_{t_i}[e_y^{i+1}\Delta W_i] + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i] \,\mathrm{d}s \\
&- \int_{t_i}^{t_{i+1}} E_{t_i}\Big[g(s, X_s, Y_s)\Delta W_i\Big] \,\mathrm{d}\overleftarrow{B_s},
\end{aligned} \tag{3.19}
$$

where we have used the Fubini's theorem and the fact that $E_{t_i}[(\tilde{Y}_i(t_i) - Y_{t_i})\Delta W_i] = 0$. Rewrite $E_{t_i}[\tilde{Z}^i(s) - Z_s]$ as

$$
E_{t_i}[\tilde{Z}_i(s) - \tilde{Z}_i(t_{i+1} - 0)] + E_{t_i}[\tilde{Z}_i(t_{i+1} - 0) - Z_{t_{i+1}}] + E_{t_i}[Z_{t_{i+1}} - Z_s].
$$

Using Theorem 2.1 (4), we have

$$
\begin{aligned}
\Delta t E_{t_i}[e_z^{i+1}] =& \ E_{t_i}[e_y^{i+1}\Delta W_i] + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i] \,\mathrm{d}s \\
&- \int_{t_i}^{t_{i+1}} E_{t_i}[g(\Theta_s)\Delta W_i] \,\mathrm{d}\overleftarrow{B_s} + \int_{t_i}^{t_{i+1}} \int_s^{t_{i+1}} E_{t_i}[\nabla g(\Theta_r)] \,\mathrm{d}\overleftarrow{B_r} \,\mathrm{d}s + O(\Delta t^2),
\end{aligned} \tag{3.20}
$$

where $\Theta_r = (r, X_r, Y_r)$. Similar to the argument in eq. (3.11) (notice that here we apply the Itô formula on interval $[t_i, s]$ instead of $[s, t_{i+1}]$ in eq. (3.11)), we obtain

$$g(\Theta_s) = g(\Theta_{t_i}) + \int_{t_i}^{s} \big(g_x(\Theta_r) + g_y(\Theta) Z_r\big) dW_r - \int_{t_i}^{s} (g_y \cdot g)(\Theta_r)\, \mathrm{d}\overleftarrow{B_r} + R_g^i,$$

where all the $\int_{t_i}^{s} \cdot dr$ terms are included in R_g^i. The above equation leads to

$$\int_{t_i}^{t_{i+1}} E_{t_i}\Big[g(\Theta_s)\Delta W_i\Big]\, \mathrm{d}\overleftarrow{B_s} = \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}\big[g_x(\Theta_r) + g_y(\Theta_r) Z_r\big]\, \mathrm{d}r\, \mathrm{d}\overleftarrow{B_s}$$
$$- \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}[(g_y \cdot g)(\Theta_r)\Delta W_i]\, \mathrm{d}\overleftarrow{B_r}\, \mathrm{d}\overleftarrow{B_s} + \int_{t_i}^{t_{i+1}} E_{t_i}\Big[R_g^i \Delta W_i\Big]\, \mathrm{d}\overleftarrow{B_s}.$$

Here we have used the fact $E_{t_i}[g(\Theta_{t_i})\Delta W_i] = 0$. Decompose the second term on the right hand side of the above equation into two terms

$$\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}[(g_y \cdot g)(\Theta_{t_i})\Delta W_i]\, \mathrm{d}\overleftarrow{B_r}\, \mathrm{d}\overleftarrow{B_s}$$
$$+ \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}[\big((g_y \cdot g)(\Theta_r) - (g_y \cdot g)(\Theta_{t_i})\big)\Delta W_i]\, \mathrm{d}\overleftarrow{B_r}\, \mathrm{d}\overleftarrow{B_s}.$$

By properties of conditional expectations, the first part is 0. Therefore

$$\int_{t_i}^{t_{i+1}} E_{t_i}\Big[g(\Theta_s)\Delta W_i\Big]\, \mathrm{d}\overleftarrow{B_s} = \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}\big[g_x(\Theta_r) + g_y(\Theta_r) Z_r\big]\, \mathrm{d}r\, \mathrm{d}\overleftarrow{B_s}$$
$$- \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}[\big((g_y \cdot g)(\Theta_r) - (g_y \cdot g)(\Theta_{t_i})\big)\Delta W_i]\, \mathrm{d}\overleftarrow{B_r}\, \mathrm{d}\overleftarrow{B_s} + \int_{t_i}^{t_{i+1}} E_{t_i}\Big[R_g^i \Delta W_i\Big]\, \mathrm{d}\overleftarrow{B_s}.$$

Putting this into eq. (3.20), and noting that $\nabla g(\Theta_r) = g_x(\Theta_r) + g_y(\Theta_r) Z_r$ (as defined in Theorem 2.1), we obtain

$$\Delta t E_{t_i}[e_z^{i+1}] = E_{t_i}[e_y^{i+1}\Delta W_i] + \int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i]\, \mathrm{d}s + R_z^{i+1}, \qquad (3.21)$$

where

$$R_z^{i+1} = -\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}\big[\nabla g(\Theta_r) - \nabla g(\Theta_{t_i})\big]\, \mathrm{d}r\, \mathrm{d}\overleftarrow{B_s}$$
$$+ \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} E_{t_i}[\big((g_y \cdot g)(\Theta_r) - (g_y \cdot g)(\Theta_{t_i})\big)\Delta W_i]\, \mathrm{d}\overleftarrow{B_r}\, \mathrm{d}\overleftarrow{B_s} - \int_{t_i}^{t_{i+1}} E_{t_i}\Big[R_g^i \Delta W_i\Big]\, \mathrm{d}\overleftarrow{B_s}$$
$$+ \int_{t_i}^{t_{i+1}} \int_{s}^{t_{i+1}} E_{t_i}[\nabla g(\Theta_r) - \nabla g(\Theta_{t_i})]\, \mathrm{d}\overleftarrow{B_r}\, \mathrm{d}s + O(\Delta t^2).$$

It is easy to check that $E[(R_z^{i+1})^2] = O(\Delta t^4)$. Similar to the discussions for the e_y^i error, we square eq. (3.21), and take expectation to get

$$\Delta t^2 E\big[(E_{t_i}[e_z^{i+1}])^2\big] \le E\big[\big(E_{t_i}[e_y^{i+1}\Delta W_i]\big)^2\big] + 2E\Bigg[\bigg(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i]\,\mathrm{d}s\bigg)^2\Bigg]$$
$$+ 2E\Big[E_{t_i}[e_y^{i+1}\Delta W_i]\cdot\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i]\,\mathrm{d}s\Big] + E\Big[E_{t_i}[e_y^{i+1}\Delta W_i]R_z^{i+1}\Big] + O(\Delta t^4). \tag{3.22}$$

For the cross product terms in the above equation, we use Young's inequality to get

$$\begin{aligned}
&2E\Big[E_{t_i}[e_y^{i+1}\Delta W_i]\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i]\,\mathrm{d}s\Big]\\
&\le \frac{L\Delta t}{\epsilon_1}E\big[\big(E_{t_i}[e_y^{i+1}\Delta W_i]\big)^2\big] + \frac{\epsilon_1}{L\Delta t}E\Big[\Big(\int_{t_i}^{t_{i+1}} E_{t_i}[\delta f^i(s)\Delta W_i]\,\mathrm{d}s\Big)^2\Big]\\
&\le \frac{L\Delta t^2}{\epsilon_1}E\big[\big(e_y^{i+1}\big)^2\big] + \epsilon_1\Delta t^2 E\Big[(CE_{t_i}[e_y^{i+1}])^2 + (1+\epsilon_0)(E_{t_i}[e_z^{i+1}])^2\Big] + O(\Delta t^4)
\end{aligned}$$

and

$$E\Big[E_{t_i}[e_y^{i+1}\Delta W_i]R_z^{i+1}\Big] \le \epsilon_2 E\Big[\big(E_{t_i}[e_y^{i+1}\Delta W_i]\big)^2\Big] + O(\Delta t^4).$$

From the assumption **(H3)**, Lemma 3.3 and Lemma 3.5, and the above estimations, we have

$$\begin{aligned}
\Delta t^2 E\big[(E_{t_i}[e_z^{i+1}])^2\big] \le{}& E\big[\big(E_{t_i}[e_y^{i+1}\Delta W_i]\big)^2\big] + C\Delta t^3 E[(e_y^{i+1})^2] + 2L(1+\epsilon_0)\Delta t^3 E\big[\big(E_{t_i}[e_z^{i+1}]\big)^2\big]\\
&+ \frac{L\Delta t^2}{\epsilon_1}E\big[\big(e_y^{i+1}\big)^2\big] + C\epsilon_1\Delta t^2 E[(e_y^{i+1})^2] + \epsilon_1(1+\epsilon_0)\Delta t^2 E[(E_{t_i}[e_z^{i+1}])^2]\\
&+ \epsilon_2 E\Big[\big(E_{t_i}[e_y^{i+1}\Delta W_i]\big)^2\Big] + O(\Delta t^4).
\end{aligned}$$

Set $\frac{1}{1+\epsilon_2} < 1$. Dividing both sides of the above estimate by $\Delta t(1+\epsilon_2)$, and noting that $\big(E_{t_i}[e_y^{i+1}\Delta W_i]\big)^2 \le \Delta t\Big(E_{t_i}[(e_y^{i+1})^2] - (E_{t_i}[e_y^{i+1}])^2\Big)$, we obtain

$$\begin{aligned}
\frac{\Delta t}{1+\epsilon_2}E\big[(E_{t_i}[e_z^{i+1}])^2\big] \le{}& E\big[E_{t_i}[(e_y^{i+1})^2] - (E_{t_i}[e_y^{i+1}])^2\big] + \big(\frac{L}{\epsilon_1} + C\epsilon_1\big)\Delta t E\big[\big(e_y^{i+1}\big)^2\big]\\
&+ \epsilon_1(1+\epsilon_0)\Delta t E\Big[E_{t_i}[(e_z^{i+1})^2]\Big] + C\Delta t^2 E[(e_y^{i+1})^2]\\
&+ 2L(1+\epsilon_0)\Delta t^2 E\big[\big(E_{t_i}[e_z^{i+1}]\big)^2\big] + O(\Delta t^3).
\end{aligned} \tag{3.23}$$

which is the desired estimate for e_z^{i+1}.

Now we use the above estimates, eq. (3.18) for $Y_i(t_i) - Y_{t_i}$, and eq. (3.23) for $\tilde{Z}_i(t_i) - Z_{t_i}$, to derive the error estimate of the theorem. First we combine eq. (3.18) and eq. (3.23) to obtain

$$
\begin{aligned}
&E[(e_y^i)^2] + \frac{\Delta t}{1+\epsilon_2} E\big[(E_{t_i}[e_z^{i+1}])^2\big] \\
&\leq E[(e_y^{i+1})^2] + \big(C_y^1 + \frac{L}{\epsilon_1} + C\epsilon_1\big)\Delta t E[(e_y^{i+1})^2] + \big(6|f_z|_\infty^2 + \epsilon_1(1+\epsilon_0)\big)\Delta t E\big[(E_{t_i}[e_z^{i+1}])^2\big] \\
&\quad + C\Delta t^2 E[(e_y^{i+1})^2] + 4L(1+\epsilon_0)\Delta t^2 E\big[(E_{t_i}[e_z^{i+1}])^2\big] + O(\Delta t^3).
\end{aligned} \tag{3.24}
$$

Next we properly choose ϵ_0, ϵ_1 and ϵ_2 to control the $E\big[(E_{t_i}[e_z^{i+1}])^2\big]$ terms on the right hand side of eq. (3.24). Specifically, we move all the $E\big[(E_{t_i}[e_z^{i+1}])^2\big]$ terms to the left hand side and get

$$
E[(e_y^i)^2] + C_z \Delta t E\big[(E_{t_i}[e_z^{i+1}])^2\big] \leq E[(e_y^{i+1})^2] + C_y \Delta t E[(e_y^{i+1})^2] + O(\Delta t^3) \tag{3.25}
$$

where $C_y = C_y^1 + \frac{L}{\epsilon_1} + C\epsilon_1 + C\Delta t$, $C_z = \frac{1}{1+\epsilon_2} - 6|f_z|_\infty^2 - \epsilon_1(1+\epsilon_0) - 4L(1+\epsilon_0)\Delta t$. Now we choose constants ϵ_0, ϵ_1 and ϵ_2 sufficiently small such that $\frac{1}{1+\epsilon_2} - 6|f_z|_\infty^2 - \epsilon_1(1+\epsilon_0) > 0$. When the temporal step size Δt is chosen such that $\Delta t < (\frac{1}{1+\epsilon_2} - 6|f_z|_\infty^2 - \epsilon_1(1+\epsilon_0))/4L(1+\epsilon_0)$, we have $C_z > 0$. As a result, the estimate eq. (3.25) becomes

$$
E[(e_y^i)^2] \leq E[(e_y^{i+1})^2] + C_y \Delta t E[(e_y^{i+1})^2] + O(\Delta t^3),
$$

which gives

$$
\max_{1\leq i\leq N} \big(E[(e_y^i)^2]\big) \leq C\Delta t^2 \tag{3.26}
$$

according to the discrete Gronwall 's inequality. This is the first part of eq. (3.3).

For the second part of eq. (3.3), we substitute eq. (3.26) into eq. (3.25) to obtain

$$
\max_{1\leq i\leq N-1} E\big[(E_{t_i}[e_z^{i+1}])^2\big] \leq C\Delta t. \tag{3.27}
$$

Since $\tilde{Z}_i(t_i) - Z_{t_i}$ is $\mathcal{F}_{0,t_i}^W \vee \mathcal{F}_{0,T}^B$ measurable, i.e. $\mathcal{G}_{0,t_i}$ measurable, we have

$$
\begin{aligned}
&\max_{1\leq i\leq N-1} E\big[(\tilde{Z}_i(t_i) - Z_{t_i})^2\big] = \max_{1\leq i\leq N-1} E\big[(E_{t_i}[\tilde{Z}_i(t_i) - Z_{t_i}])^2\big] \\
&\leq \max_{1\leq i\leq N-1} 3E\Big[(E_{t_i}[\tilde{Z}_i(t_i) - \tilde{Z}_i(t_{i+1}-0)])^2 + (E_{t_i}[e_z^{i+1}])^2 + (E_{t_i}[Z_{t_{i+1}} - Z_{t_i}])^2\Big].
\end{aligned}
$$

Then the second part of the theorem follows from Theorem 2.1, Lemma 3.2, and eq. (3.27). □

4 A first order splitting up scheme

In this section, we discretize the BSDE (3.1) and SDE (3.2) in the splitting up system to obtain a first order splitting up numerical scheme.

First we define an approximation for X by $X^0 = X_0$, and

$$X^{i+1} = X^i + \Delta W_i, \quad i = 1, \cdots, N. \tag{4.1}$$

It is easy to see that for any $t \in [t_i, t_{i+1})$ and $i = 1 \cdots, N$, there exists a positive constant C, independent of X_0 and Δt, such that

$$E\Big[|X^{i+1} - X_t|^2 + |X_t - X^i|^2\Big] \leq C\Delta t. \tag{4.2}$$

To obtain a first order splitting up scheme, we use the explicit Euler scheme to approximate the BSDE eq. (3.1) and the Milstein scheme to approximate the SDE eq. (3.2). The resulting algorithm is given as follows.

$$\begin{aligned}
&H^{i+1} = Y^{i+1} + \Delta t f(t_{i+1}, X^{i+1}, Y^{i+1}, Z^{i+1}), && (a)\\
&\tilde{Y}^i = E_{t_i}[H^{i+1}], && (b)\\
&Z^i = \frac{1}{\Delta t} E_{t_i}[H^{i+1}\Delta W_i], && (c)\\
&Y^i = \tilde{Y}^i + E_{t_i}[G^{i+1}(\tilde{Y}^i)], && (d)
\end{aligned} \tag{4.3}$$

where for any $\mathcal{F}^B_{t_{i+1},T}$ measurable random variable, $G^{i+1}(\xi)$ is defined by

$$G^{i+1}(\xi) := g(t_{i+1}, X^{i+1}, \xi)\Delta B_i + (g_y \cdot g)(t_{i+1}, X^{i+1}, \xi)\frac{1}{2}(\Delta B_i^2 - \Delta t). \tag{4.4}$$

Note that in eq. (4.3), $\tilde{Y}^i$ is an approximation for $\tilde{Y}_i(t_i)$, Y^i is an approximation for $E_{t_i}[Y_i(t_i)]$, and Z^i is an approximation for $\tilde{Z}_i(t_i)$. Apparently, Y^i is also an approximation for Y_{t_i} and Z^i is also an approximation for Z_{t_i}.

In order to show that Y^i is a first order numerical approximation for Y_t and Z^i is a half order numerical approximation for Z_t, we first show that Y^i is a first order approximation for $E_{t_i}[Y_i(t_i)]$ and Z^i is a half order approximation for $\tilde{Z}_i(t_i)$. Then, the first order convergence rate and half order convergence rate of our numerical schemes in approximating Y_t and Z_t, respectively, is arrived as a direct consequence of Theorem 3.1.

Theorem 4.1 *Assume that assumptions **(H1)-(H3)** hold. Then there exists a positive constant C, independent of Δt and X_0, such that*

$$\max_{1\leq i\leq N} E\big[(Y^i - E_{t_i}[Y_i(t_i)])^2 + \Delta t(Z^i - \tilde{Z}_i(t_i))^2\big] \leq C\Delta t^2. \tag{4.5}$$

Proof: Set $t = t_i$, take conditional expectation on both sides of eq. (3.2), and then subtract the result from eq. (4.3) (d) to get

$$Y^i - E_{t_i}[Y_i(t_i)] = \tilde{Y}^i - \tilde{Y}_i(t_i) + E_{t_i}[G^{i+1}(\tilde{Y}^i) - G^{i+1}(\tilde{Y}_i(t_i))] + R^i_g, \tag{4.6}$$

where

$$R_g^i = \int_{t_i}^{t_{i+1}} E_{t_i}[g(\tilde{\Theta}_s)]\,\mathrm{d}\overleftarrow{B_s} - E_{t_i}[G^{i+1}(\tilde{Y}_i(t_i))].$$

It is easy to verify that $E[(R_g^i)^2] = O(\Delta t^3)$. As in Section 3, denote

$$\hat{e}_y^i = Y^i - E_{t_i}[Y_i(t_i)], \text{ and } \hat{e}_z^i = Z^i - \tilde{Z}_i(t_i).$$

Squaring eq. (4.6) and taking the expectation, we obtain

$$\begin{aligned} E[(\hat{e}_y^i)^2] = {} & E\big[(\tilde{Y}^i - \tilde{Y}_i(t_i))^2\big] + E\Big[\Big(E_{t_i}[G^{i+1}(\tilde{Y}^i) - G^{i+1}(\tilde{Y}_i(t_i))] + R_g^i\Big)^2\Big] \\ & + 2E\Big[\big(\tilde{Y}^i - \tilde{Y}_i(t_i)\big)\Big(E_{t_i}[G^{i+1}(\tilde{Y}^i) - G^{i+1}(\tilde{Y}_i(t_i))] + R_g^i\Big)\Big]. \end{aligned} \tag{4.7}$$

The last term above is 0, which can be proved by an argument similar to (3.16). Next we estimate $\tilde{Y}^i - \tilde{Y}_i(t_i)$ and $E_{t_i}[G^{i+1}(\tilde{Y}^i) - G^{i+1}(\tilde{Y}_i(t_i))]$ in eq. (4.7). By the definition of $\hat{G}$ in eq. (4.4), we have

$$E\Big[\big(G^{i+1}(\tilde{Y}^i) - G^{i+1}(\tilde{Y}_i(t_i))\big)^2\Big] \le C_G \Delta t E[(\tilde{Y}^i - \tilde{Y}_i(t_i))^2], \tag{4.8}$$

where C_G is a constant independent of X_0 and Δt. Therefore, it suffices to estimate $E[(\tilde{Y}^i - \tilde{Y}_i(t_i))^2]$. To this end, we take the conditional expectation $E_{t_i}[\cdot]$ on both sides of eq. (3.4) and subtract it from eq. (4.3) (b) to obtain

$$\tilde{Y}^i - \tilde{Y}_i(t_i) = E_{t_i}[\hat{e}_y^{i+1}] + \Delta t E_{t_i}\Big[f(\Pi^{i+1}) - f(\tilde{\Pi}_{t_{i+1}})\Big] + R_f^i, \tag{4.9}$$

where $R_f^i = \int_{t_i}^{t_{i+1}} E_{t_i}\big[f(\tilde{\Pi}_{i+1}) - f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s))\big]\,\mathrm{d}s$ is the truncation error, and $R_f^i = O(\Delta t^2)$ (for notational simplicity, we denote $\Pi^{i+1} := (t_{i+1}, X^{i+1}, Y^{i+1}, Z^{i+1})$ and $\tilde{\Pi}_{i+1} := (t_{i+1}, X_{t_{i+1}}, Y_{i+1}(t_{i+1}), \tilde{Z}_i(t_{i+1} - 0)))$. Squaring both sides of the above, and then taking the expectation, we have

$$\begin{aligned} E\Big[\big(\tilde{Y}^i - \tilde{Y}_i(t_i)\big)^2\Big] = {} & E\Big[\big(E_{t_i}[\hat{e}_y^{i+1}]\big)^2\Big] + E\Big[\Big(\Delta t E_{t_i}\Big[f(\Pi^{i+1}) - f(\tilde{\Pi}_{t_{i+1}})\Big] + R_f^i\Big)^2\Big] \\ & + 2E\Big[E_{t_i}[\hat{e}_y^{i+1}]\Big(\Delta t E_{t_i}\Big[f(\Pi^{i+1}) - f(\tilde{\Pi}_{t_{i+1}})\Big] + R_f^i\Big)\Big]. \end{aligned} \tag{4.10}$$

Using similar arguments as eqs. (3.10), (3.15) and (3.17), we obtain

$$\begin{aligned} E[(\hat{e}_y^i)^2] \le {} & E\big[\big(E_{t_i}[\hat{e}_y^{i+1}]\big)^2\big] + \Delta t \hat{C}_y^1 E\big[(\hat{e}_y^{i+1})^2\big] \\ & + \epsilon_1 \Delta t E\big[(\hat{e}_z^{i+1})^2\big] + \hat{C}_y^2 \Delta t^2 E\big[(\hat{e}_z^{i+1})^2\big] + O(\Delta t^3), \end{aligned} \tag{4.11}$$

where $\hat{C}_y^1$ and $\hat{C}_y^2$ are constants independent of X_0 and Δt, and ϵ_1 is a constant to be specified later.

To estimate $\hat{e}_z^i$, we multiply ΔW_i on both sides of eq. (3.1), take conditional expectation $E_{t_i}[\cdot]$, and subtract it from eq. (4.3) (c) to obtain

$$\Delta t\big(\hat{e}_z^i\big) = E_{t_i}[\big(\hat{e}_y^{i+1}\big)\Delta W_i] + \Delta t E_{t_i}\Big[\Big(f(\Pi^{i+1}) - f(\tilde{\Pi}_{t_{i+1}})\Big)\Delta W_i\Big] + R_{fW}^i, \tag{4.12}$$

where

$$\begin{aligned} R_{fW}^i = &\int_{t_i}^{t_{i+1}} E_{t_i}\Big[\big(f(\tilde{\Pi}_{t_{i+1}}) - f(s, X_s, \tilde{Y}_i(s), \tilde{Z}_i(s))\big)\Delta W_i\Big]\,\mathrm{d}s \\ &+ \int_{t_i}^{t_{i+1}} E_{t_i}[\tilde{Z}_i(s) - \tilde{Z}_i(t_i)]\,\mathrm{d}s. \end{aligned}$$

It follows from Lemma 3.2 that $R_{fW}^i = O(\Delta t^2)$. Squaring both sides of eq. (4.12), taking the expectation, and using similar analysis techniques as in the proof of Theorem 3.1, we derive that

$$\begin{aligned} \frac{\Delta t}{1+\epsilon} E[(\hat{e}_z^i)^2] \le\; & E\Big[E_{t_i}[\big(\hat{e}_y^{i+1}\big)^2] - (E_{t_i}[\hat{e}_y^{i+1}])^2\Big] + \Delta t \hat{C}_z^1 E\big[(\hat{e}_y^{i+1})^2\big] \\ &+ \epsilon_2 \Delta t E\big[(\hat{e}_z^{i+1})^2\big] + \hat{C}_z^2 \Delta t^2 E\big[(\hat{e}_z^{i+1})^2\big] + O(\Delta t^3), \end{aligned} \tag{4.13}$$

where $\hat{C}_z^1$, $\hat{C}_z^2$ are constants independent of Δ and X_0, and ϵ_2 is a constant that will be determined later.

Finally, we add eq. (4.11) and eq. (4.13) to obtain

$$\begin{aligned} E[(\hat{e}_y^i)^2] + \frac{\Delta t}{1+\epsilon} E[(\hat{e}_z^i)^2] \le\; & E\big[(\hat{e}_y^{i+1})^2\big] + (\epsilon_1 + \epsilon_2)\Delta t E\big[(\hat{e}_z^{i+1})^2\big] \\ &+ \Delta t(\hat{C}_y^1 + \hat{C}_z^1) E\big[(\hat{e}_y^{i+1})^2\big] + (\hat{C}_y^2 + \hat{C}_z^2)\Delta t^2 E\big[(\hat{e}_z^{i+1})^2\big] + O(\Delta t^3). \end{aligned} \tag{4.14}$$

Choosing ϵ, ϵ_1 and ϵ_2 sufficiently small so that $\epsilon_1 + \epsilon_2 < \frac{1}{1+\epsilon}$, and using the discrete Gronwall inequality, we obtain the desired result of the theorem. □

As a direct consequence of Theorem 3.1 and Theorem 4.1, we have the following first order error estimate for our numerical scheme (4.3).

Theorem 4.2 *Assume that assumptions **(H1)-(H3)** hold. Then there exists a positive constant C independent of Δt, such that*

$$\max_{1\le i\le N} E\big[(Y^i - Y_{t_i})^2 + \Delta t(Z^i - Z_{t_i})^2\big] \le C\Delta t^2. \tag{4.15}$$

5 Numerical experiments

In this section, we use three numerical experiments to validate our splitting up scheme and verify the error estimations. In order to implement the numerical schemes (4.3), we need to approximate the conditional expectation E_{t_i} in eq. (4.3). Since the conditional expectation E_{t_i} is essentially an integral with the Gaussian kernel, we use Gauss - Hermite quadrature formula as a numerical integral method to calculate conditional expectations (see [9] for more details). To calculate the general expectation E, we use Monte Carlo method with 300 samples and compute the root mean square error in each example.

Example 1

In the first example, we consider the BDSDE

$$\mathrm{d}Y_t = -f(t, Y_t, Z_t)\,\mathrm{d}t + Z_t\,\mathrm{d}W_t - g(t, Y_t)\,\mathrm{d}\overleftarrow{B}_t,$$

where $f(t, Y_t, Z_t) = \frac{Y_t}{2} - Z_t + \frac{B_t - B_T}{8}$ and $g(t, Y_t) = \frac{1}{4}\big(\cos(t + W_t)^2 + Y_t - (\frac{B_t - B_T}{8})^2\big)$.The exact solution to the above equation is $Y_t = \sin(t+W_t) - \frac{B_t}{4} + \frac{B_T}{4}$ and $Z_t = \cos(t + W_t)$. To demonstrate the performance of our numerical

Table 1: Example 1

Partition	$E[\|\tilde{Y}^0 - Y_0\|_{L^2}]$	$E[\|Y^0 - Y_0\|_{L^2}]$	$E[\|Z^0 - Z_0\|_{L^2}]$
$N = 2^3$	$2.4000e-01$	$2.3489e-01$	$9.8106e-02$
$N = 2^4$	$1.3788e-01$	$1.2745e-01$	$4.2258e-02$
$N = 2^5$	$7.7604e-02$	$6.4257e-02$	$2.0880e-02$
$N = 2^6$	$4.3455e-02$	$3.1834e-02$	$1.2762e-02$
$N = 2^7$	$2.7816e-02$	$1.5770e-02$	$6.0392e-03$
CR	0.79	0.98	0.98

schemes, we compute the root mean square errors (RMSEs) between our approximate solutions and the exact solution. Specifically, we calculate the expectation of the L^2 norm errors $\|\tilde{Y}^0 - Y_0\|_{L^2}$, $\|Y^0 - Y_0\|_{L^2}$ and $\|Z^0 - Z_0\|_{L^2}$ at time $t = 0$ with 300 Monte Carlo samples and discretize the equations with time step sizes $\Delta t = 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}$. The corresponding errors are presented in Table 1. Here CR in the table stands for "convergence rate". We can see from the table that Y^i indeed provides a first order numerical approximation for the solution Y_t, and $\tilde{Y}^i$ provides reasonably accurate approximation for Y_t. However, since the scheme for $\tilde{Y}^i$ does not include the $d\overleftarrow{B}$ integral, it does not provide a first order approximation for the solution. On the other hand, we can see that our numerical solution Z^i converges with first order in approximating Z_t in this example although in our proof we only obtain half order convergence analysis for Z^i. Further investigation is needed to determine if this a super convergence for Z_t on the nodal points.

Example 2

In the second example, we consider the BDSDE

$$\mathrm{d}Y_t = -f(t, Y_t, Z_t)\,\mathrm{d}t + Z_t\,\mathrm{d}W_t - g(t, Y_t)\,\mathrm{d}\overleftarrow{B}_t,$$

with $f(t, Y_t, Z_t) = (Y_t - t - B_t)^2 + \big(\cos(W_t)\big)^2 - \frac{1}{2}\sin(W_t)$, $g(t, Y_t) = (Y_t - t - B_t)^2 + \big(\cos(W_t)\big)^2$. The exact solution is given by $Y_t = \sin(W_t) + t + B_t$ and $Z_t = \cos(W_t)$. We can see that in this example, both f and g are nonlinear

Table 2: Example 2

Partition	$E[\|\tilde{Y}^0 - Y_0\|_{L^2}]$	$E[\|Y^0 - Y_0\|_{L^2}]$	$E[\|Z^0 - Z_0\|_{L^2}]$
$N = 2^3$	$4.1305e-01$	$2.4342e-01$	$1.3810e-01$
$N = 2^4$	$2.9031e-01$	$1.4541e-01$	$9.0722e-02$
$N = 2^5$	$1.9627e-01$	$7.3160e-02$	$5.5327e-02$
$N = 2^6$	$1.3704e-01$	$3.4548e-02$	$2.7033e-02$
$N = 2^7$	$8.9505e-02$	$1.5549e-02$	$1.3429e-02$
CR	0.55	1.00	0.85

function for Y_t. Therefore, this example demonstrates the performance of our schemes in solving nonlinear BDSDE systems. As in the first example, we evaluate he RMSEs $E[\|\tilde{Y}^0 - Y_0\|_{L^2}]$, $E[\|Y^0 - Y_0\|_{L^2}]$ and $E[\|Z^0 - Z_0\|_{L^2}]$, at time $t = 0$. In Table 2, we can see that the convergence order for $\|Y^0 - Y_0\|_{L^2}$ is 1 and the convergence order for $\|\tilde{Y}^0 - Y_0\|_{L^2}$ is roughly 0.55, For the numerical solution Z^i, we can see from the table that the convergence for $\|Z^0 - Z_0\|_{L^2}$ is 0.847, which is less than 1. From this example we can see that Z^i does not always produce first order approximation for Z_{t_i}.

Example 3

In the third example, we consider the BDSDE

$$\mathrm{d}Y_t = -f(t, Y_t, Z_t)\,\mathrm{d}t + Z_t\,\mathrm{d}W_t - g(t, Y_t)\,\mathrm{d}\overleftarrow{B}_t,$$

with $f(t, Y_t, Z_t) = -\frac{1}{2}(\sin(Y_t))^2 - \frac{1}{2}(\cos(t + W_t + \frac{1}{2}B_t))^2 - \frac{1}{2}(Z_t)^2$ and $g(t, Y_t) = -\frac{1}{2}(\sin(Y_t))^2 - \frac{1}{2}(\cos(t + W_t + \frac{1}{2}B_t))^2$. The exact solution for the above equation is $Y_t = t + W_t + \frac{1}{2}B_t$ and $Z_t = 1$. In the last example, f is a nonlinear functions for Y_t, Z_t, g is a nonlinear function for Y_t, and Y_t is in a trigonometric function in both f and g. The purpose of this example to demonstrate the performance of our method in solving a more general BDSDE system. In Table 3, we present the RMSEs between our approximate solutions and the exact solution at time $t = 0$. From this table, we can see that for this example $\tilde{Y}^0$ converges to Y_0 with half order, and $(\hat{Y}_0, \hat{Z}_0)$ converges to $(Y_0, Z_0)$with first order.

Table 3: Example 3

Partition	$E[\|\tilde{Y}^0 - Y_0\|_{L^2}]$	$E[\|Y^0 - Y_0\|_{L^2}]$	$E[\|Z^0 - Z_0\|_{L^2}]$
$N = 2^3$	$1.7759e-01$	$9.2446e-03$	$1.6664e-02$
$N = 2^4$	$1.1907e-01$	$4.3131e-03$	$8.1184e-03$
$N = 2^5$	$9.0338e-02$	$2.2815e-03$	$4.2914e-03$
$N = 2^6$	$5.7300e-02$	$1.1198e-03$	$2.1936e-03$
$N = 2^7$	$4.6539e-02$	$5.5550e-04$	$1.0452e-03$
CR	0.49	1.01	0.99

References

[1] Achref Bachouch, Mohamed Anis Ben Lasmar, Anis Matoussi, and Mohamed Mnif. Euler time discretization of backward doubly SDEs and application to semilinear SPDEs. *Stochastics and Partial Differential Equations: Analysis and Computations*, 4(3):592–634, Sep 2016.

[2] Feng Bao, Richard Archibald, and Peter Maksymovych Backward SDE Filter for Jump Diffusion Processes and Its Applications in Material Sciences *Communications in Computational Physics*, 27:589-618, 2020.

[3] Feng Bao, Yanzhao Cao and Hongmei Chi. Adjoint Forward Backward Stochastic Differential Equations Driven by Jump Processes and Its Application to Nonlinear Filtering Problems *International Journal for Uncertainty Quantification*, 9(2):143-159, 2019.

[4] Feng Bao, Yanzhao Cao and Xiaoying Han. Forward backward doubly stochastic differential equations and the optimal filtering of diffusion processes. *Communications in Mathematical Sciences*, 18(3):635-661, 2020.

[5] Feng Bao, Yanzhao Cao, Clayton Webster, and Guannan Zhang. A Hybrid Sparse-Grid Approach for Nonlinear Filtering Problems Based on Adaptive-Domain of the Zakai Equation Approximations, *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):784-804, 2014.

[6] Feng Bao, Yanzhao Cao, and Weidong Zhao. Numerical Solutions for Forward Backward Doubly Stochastic Differential Equations and Zakai Equations *International Journal for Uncertainty Quantification*, 4(1):351-367, 2011.

[7] Feng Bao, Yanzhao Cao, and Weidong Zhao. A Backward Doubly Stochastic Differential Equation Approach for Nonlinear Filtering Problems *Communications in Computational Physics*, 23(5):1573-1601, 2018.

[8] Feng Bao, Yanzhao Cao, and Weidong Zhao. A First Order Semi-discrete Algorithm for Backward Doubly Stochastic Differential Equations, *Discrete and Continuous Dynamical Systems - Series B*, 5(2):1297-1313, 2015.

[9] Feng Bao, Yanzhao Cao, Amnon Meir, and Weidong Zhao. A first order scheme for backward doubly stochastic differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):413-445, 2016.

[10] Feng Bao and Vasileios Maroulas. Adaptive meshfree backward sde filter. *SIAM Journal on Scientific Computing*, 39(6):A2664–A2683, 2017.

[11] Allan Bensoussan and R Glowinski. Approximation of Zakai equation by the splitting up method. In *Stochastic Systems and Optimization*, pages 255–265. Springer, 1989.

[12] Allan Bensoussan, R Glowinski, and A Rascanu. Approximation of the Zakai equation by the splitting up method. *SIAM Journal on Control and Optimization*, 28(6):1420–1431, 1990.

[13] Allan Bensoussan, R Glowinski, and A Raşcanu. Approximation of some stochastic differential equations by the splitting up method. *Applied Mathematics and Optimization*, 25(1):81–106, 1992.

[14] Bruno Bouchard and Nizar Touzi. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their applications*, 111(2):175–206, 2004.

[15] Jim Douglas, Jin Ma, Philip Protter, et al. Numerical methods for forward-backward stochastic differential equations. *The Annals of Applied Probability*, 6(3):940–968, 1996.

[16] Weinan E, Jiequn Han, and Arnulf Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat.*, 5(4):349–380, 2017.

[17] Aurora Germani and M Piccioni. Semi-discretization of stochastic partial differential equations on R^d by a finite-element technique a. germani. *Stochastics: An International Journal of Probability and Stochastic Processes*, 23(2):131–148, 1988.

[18] E. Gobet, J. Lpez-Salas, P. Turkedjiev, and C. Vzquez. Stratified regression Monte-Carlo scheme for semilinear PDEs and BSDEs with large scale parallelization on GPUs. *SIAM Journal on Scientific Computing*, 38(6):C652–C677, 2016.

[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*, 127(4):1171 – 1203, 2017.

[20] István Gyöngy and Nicolai Krylov. On the splitting-up method and stochastic partial differential equations. *The Annals of Probability*, 31(2):564–591, 2003.

[21] Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA*, 115(34):8505–8510, 2018.

[22] Arnulf Jentzen and Peter Kloeden. Taylor expansions of solutions of stochastic partial differential equations with additive noise. *Ann. Probab.*, 38(2):532–569, 03 2010.

[23] Peter Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.

[24] Céline Labart and Jérôme Lelong. A parallel algorithm for solving BSDEs. *Monte Carlo Methods and Applications*, 19(1):11–39, 2013.

[25] François LeGland. Splitting-up approximation for SPDE's and SDE's with application to nonlinear filtering. In *Stochastic partial differential equations and their applications*, pages 177–187. Springer, 1992.

[26] Pierre-Louis Lions and Panagiotis E Souganidis. Fully nonlinear stochastic partial differential equations. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 326(9):1085–1092, 1998.

[27] S. Lototsky, R. Mikulevicius, and B. Rozovskii. Nonlinear Filtering Revisited: A Spectral Approach. *SIAM Journal on Control and Optimization*, 35(2):435–461, 1997.

[28] Jin Ma, Philip Protter, Jaime San Martin, Soledad Torres, et al. Numberical method for backward stochastic differential equations. *The Annals of Applied Probability*, 12(1):302–316, 2002.

[29] Jin Ma, Philip Protter, and Jiongmin Yong. Solving forward-backward stochastic differential equations explicitlya four step scheme. *Probability theory and related fields*, 98(3):339–359, 1994.

[30] G Milstein and M Tretyakov. Solving parabolic stochastic partial differential equations via averaging over characteristics. *Mathematics of computation*, 78(268):2075–2106, 2009.

[31] Noriaki Nagase. Remarks on nonlinear stochastic partial differential equations: an application of the splitting-up method. *SIAM journal on control and optimization*, 33(6):1716–1730, 1995.

[32] E. Pardoux and P. Protter. A two-sided stochastic integral and its calculus. *Probability Theory and Related Fields*, 76(1):15–49, Sep 1987.

[33] Etienne Pardoux and Shige Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic partial differential equations and their applications*, pages 200–217. Springer, 1992.

[34] Etienne Pardoux and Shige Peng. Backward doubly stochastic differential equations and systems of quasilinear SPDEs. *Probability Theory and Related Fields*, 98(2):209–227, 1994.

[35] E Pardouxt. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics*, 3(1-4):127–167, 1980.

[36] Hyek Yoo. Semi-discretization of stochastic partial differential equations on R^1 by a finite-difference method. *Mathematics of computation*, 69(230):653–666, 2000.

[37] Moshe Zakai. On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(3):230–243, 1969.

[38] Guannan Zhang, Max Gunzburger, and Weidong Zhao. A sparse-grid method for multi-dimensional backward stochastic differential equations. *Journal of Computational Mathematics*, 31(3):221–248, 2013.

[39] Jianfeng Zhang. A numerical scheme for BSDEs. *Ann. Appl. Probab.*, 14(1):459–488, 02 2004.

[40] Jianfeng Zhang. Backward stochastic differential equations. In *Backward Stochastic Differential Equations*, pages 79–99. Springer, 2017.