A Stochastic Gradient Descent Approach for Stochastic Optimal Control

Richard Archibald, * Feng Bao, † Jiongmin Yong ‡

February 5, 2020

1 Introduction

Stochastic optimal control is an important and active research topic in applied mathematics, and it has extensive applications in numerous areas, including engineering, finance and economics, biology, public health, communication networks, to mention a few [8, 18]. In the past half century, fundamental results of stochastic optimal control theory have been established: Pontryagin type maximum principle (MP, for short) ([6, 16, 15]), Bellman dynamic programming principle (DPP, for short) ([2, 3]) and Hamilton-Jacobi-Bellman (HJB, for short) equation theory ([7]), and linear-quadratic (LQ, for short) optimal control and Riccati equation theory ([11, 17]). These are three well-recognized mile stones of stochastic optimal control theory.

It is known that except for some limited special cases, such as LQ problems, one-dimensional linear state equation with convex/concave performance index (such as Merton type problem in mathematical finance), most stochastic optimal control problems are not explicitly solvable and therefore numerical algorithms to generate approximate solutions are needed. One of the most widely used approaches to solve the stochastic optimal control problem is the above-mentioned DPP, mainly due to Bellman ([2, 3]). The main idea of the DPP approach is to consider a family of optimal control problems with different initial states and times, and establish relationships among these problems through the HJB equation, which is a fully nonlinear PDE. Taking the advantage of well-established numerical schemes for solving PDEs, many computational methods for stochastic optimal control are developed under the DPP approach [23, 24, 26, 30]. Although all of these methods solve the control problem successfully, due to the complexity of numerical approximations for solutions of PDEs and the nonlinearity of the HJB equation, methods that follow the DPP approach are computationally expensive, and even infeasible when the dimension exceeds 3, although, in recent years, some efforts have been made to pursue the relaxation of the dimensional restriction ([1]). Another disadvantage of DPP is that the optimal control problem considered could not have any state constraints. The presence of state constraints will lead to the discontinuity of the value function, for which the suitable HJB equation theory is not available as of today.

*Computational Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

†Department of Mathematics, Florida State University, Tallahassee, Florida, (bao@math.fsu.edu).

‡Department of Mathematics, University of Central Florida, Orlando, Florida.

Another important approach to solve the stochastic optimal control problem is the *stochastic maximum principle*. The classic deterministic maximum principle was first introduced by Pontryagin and his students ([6, 16]). Stochastic version was developed by several researchers since 1960s ([12, 10, 5, 4, 15]). The central idea of maximum principle is that any optimal control problem must satisfy an optimization condition of a function called the Hamiltonian, and it is much easier to optimize a Hamiltonian than solving the original optimal control problem, which is infinite-dimensional. The MP approach for stochastic optimal control problems as three major advantages over the DPP approach: First, there is no dimension restriction; Second, it allows to have some state constraints, especially some finite dimensional terminal state constraints; Third, it allows to have random coefficients in the state equation and/or in the performance functional (to be optimized).

The goal of this paper is to introduce a stochastic gradient descent approach to solve stochastic optimal control problems. Let us elaborate a little more on the MP approach. When optimal control problems become stochastic, the corresponding stochastic maximum principle leads to a stochastic Hamiltonian system that consists of two forward backward stochastic differential equations (FBSDEs) ([13]). In this way, solving stochastic optimal control problems through stochastic maximum principle involves solving FBSDEs that meet certain optimization condition, which is typically achieved by gradient descent based approaches. It can be shown (under appropriate assumptions) that the gradient process of the optimization condition can also be expressed by a FBSDE system. Therefore, numerical implementation of stochastic maximum principle requires solving FBSDEs repeatedly to reach the optimization condition. However, since computational methods for solving FBSDEs are not as well developed as those for solving PDEs, numerical studies for stochastic optimal control through stochastic maximum principle are only beginning.

The numerical method for solving the stochastic optimal control problem that we shall introduce in this paper lies in stochastic maximum principle, and we adopt gradient descent as our framework to accomplish the optimization task in the Hamiltonian. The motivation of our approach is based on the fact that the optimization condition is under expectation due to the stochastic nature of the problem, and obtaining solutions of FBSDEs contained in the optimization condition involves expensive calculations. In the deterministic gradient descent approach, one needs accurate evaluations for expectations, which is computational expensive, in order to derive corresponding expected gradients in the optimization procedure. One of the most successful ways to improve the efficiency of deterministic gradient descent when the optimization condition is under expectation is the stochastic gradient descent method. The methodology of stochastic gradient descent is to represent the gradient under expectation by its single-sample simulation (or simulations from small batches of samples) to avoid complete calculation for expectations in each gradient descent iteration step. As a result, the stochastic gradient descent reduces the computational cost in evaluating expectations and achieves higher computational efficiency in trade of larger number of iterations in the optimization procedure [28, 29]. In this connection, we propose in this work to apply stochastic gradient descent to improve the efficiency of gradient descent optimization procedure when solving the stochastic optimal control problem through stochastic maximum principle.

On the other hand, in the stochastic maximum principle approach for stochastic optimal control, solutions of FBSDEs in the Hamiltonian system are needed. In most practical applications, obtaining general solutions for FBSDEs requires numerical ap-

proximation schemes for FBSDEs, which typically involve extensive calculations – either through numerical solutions for PDEs [19, 34] or numerical approximations for conditional expectations [20, 21, 35, 33]. Since stochastic gradient descent is an effective method to replace an expectation by its single-sample representation in the gradient descent optimization procedure, in this work we utilize single-sample simulations to approximate conditional expectations when solving FBSDEs. Although the concept of single-sample representation for conditional expectation does not provide accurate approximation for solutions (compared to classic numerical methods for FBSDEs), FBSDEs in the maximum principle are essentially used to represent the gradient of optimization condition. Therefore, similar justification for the effectiveness of stochastic gradient descent would also apply to sample-wise approximation for FBSDEs when optimizing the Hamiltonian in the maximum principle approach. It's worthy to mention that FBSDEs are systems of stochastic ordinary differential equations (SDEs), hence the computational cost of implementing single-sample simulation for FBSDEs would be comparable to simulating sample-wise SDEs. We also want to point out that, similar to the methodology of classic stochastic gradient descent, our proposed stochastic gradient descent approach for stochastic optimal control would transfer the extensive computational cost of approximating expectations when solving FBSDEs to larger number of iterations in the gradient descent optimization procedure, and we could save unnecessary computational cost in deriving fully calculated approximation for solutions of FBSDEs that is not needed in the optimization task. This is even more advantageous when the system of the control problem is in higher dimensions since it requires much more effort to carry out complete numerical approximation for solutions of FBSDEs.

The rest of this paper is organized as following. In Section 2, we provide a brief introduction to the stochastic optimal control problem and the gradient projection framework, which can be considered as a generalization of gradient descent approach to solve the optimal control problem. In Section 3, we introduce numerical optimization for solving the control problem. We shall start our discussion from conventional methods and then introduce our stochastic gradient descent approach that improves the efficiency of the existing algorithms. In Section 4, we carry out numerical experiments to demonstrate the effectiveness and efficiency of our method.

2 Problem Statement

2.1 Stochastic optimal control

Given a filtered probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\geq 0},P)$ on which a d-dimensional standard Brownian motion W_t is defined, we consider the following dynamical system in the form of a stochastic differential equation (SDE) over a deterministic time interval $[0,T]$

$$X_t = X_0 + \int_0^t b(s,X_s,u_s)ds + \int_0^t \sigma(s,X_s,u_s)dW_s, \quad 0<t\leq T, \tag{2.1}$$

where X_t is the state of some dynamical system which is usually called the controlled process; the function $b:[0,T]\times\mathbb{R}^d\times\mathcal{U}\to\mathbb{R}^d$ describes the dynamics of some model that is governed by a control process u_t representing the controlling action of the controllers, which is chosen in a separable metric space $\mathcal{U}$ with values in $\mathbb{R}$; W_t is a standard d-dimensional Brownian motion that drives the Itô type stochastic integral $\int_0^t \cdot dW_s$; the stochastic process $\sigma:[0,T]\times\mathbb{R}^d\times\mathcal{U}\to\mathbb{R}^d$ is the diffusion coefficient that represents

the size of noises added to the system X_t, and $X_0 \in \mathbb{R}^d$ is the initial condition for the controlled state.

In an optimal control problem, we have a cost functional defined as

$$J(u) := E\Big[\int_0^T f(t, X_t, u_t)dt + h(X_T)\Big] \tag{2.2}$$

for all $u \in \mathcal{U}$. The control process u_t is called an admissible control, and the pair (X_t, u_t) is called an admissible pair if u_t is a feasible control process (see [] for details), X_t is a unique solution of the SDE (2.1), and

$$f(\cdot, X, u) \in L_{\mathcal{F}^1(0,T,\mathbb{R})}, \qquad h(X_T) \in L_{\mathcal{F}_T^1(0,T,\mathbb{R})}.$$

The goal of the optimal control problem is to minimize the cost functional (2.2) and find an optimal control process u^*, such that

$$J(u^*) = \inf_{u \in \mathcal{U}} J(u). \tag{2.3}$$

The main efforts to solve the above stochastic optimal control problem can be summarized as two types of methods: (i) *dynamic programming*, in which we solve a partial differential equation system named Hamilton-Jacobi-Bellman (HJB) equation, and (ii) *maximum principle*, which aims to optimize the corresponding Hamiltonian system. Our approach in this work lies in the general framework of maximum principle, which provides the adjoint process of the cost functional with respect to the control process u.

To proceed, let $\mathcal{U}$ be a nonempty convex set and all the controls in $\mathcal{U}$ be square integrable. Suppose (X_t^*, u_t^*) be an optimal pair for the control problem (2.1) - (2.3), we have that for any $u \in \mathcal{U}$,

$$\begin{aligned} 0 \le & \lim_{\epsilon \to 0} \frac{J(u^\epsilon) - J(u^*)}{\epsilon} \\ = & E\Big[\int_0^T f_x(t, X_t, u_t)\mathcal{D}X_t + f_u(t, X_t, u_t)[u_t - u^*(t)]dt + h_x \mathcal{D}X_T\Big], \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} \mathcal{D}X_t = & \int_0^t \Big(b_x(s, X_s, u_s)\mathcal{D}X_s + b_u(s, X_s, u_s)[u_s - u_s^*]\Big)ds \\ & + \int_0^t \Big(\sigma_x(s, X_s, u_s)\mathcal{D}X_s + \sigma_u(s, X_s, u_s)[u_s - u_s^*]\Big)dW_s \end{aligned} \tag{2.5}$$

is the adjoint equation with initial condition $\mathcal{D}X_0 = 0$, and for a given function ϕ we let ϕ_x be the partial derivative with respect to the state variable x and ϕ_u be the partial derivative of the control variable u.

Although (2.4) provides a formulation for the differential of J with respect to u, it's typically very difficult to carry out optimization procedures for (2.4) given an SDE system in the form of (2.5). In this approach, we introduce an alternative formulation to cancel out the $\mathcal{D}X_t$ term. To this end, we consider the following forward backward stochastic differential equations (FBSDEs)

$$\begin{aligned} dX_s = & b(s, X_s, u_s)ds + \sigma(s, X_s, u_s)dW_s, \\ dY_s = & \Big(-b_x(s, X_s, u_s)^\top Y_s - \sigma_x(s, X_s, u_s)^\top Z_s - f_x(s, X_s, u_s)^\top\Big)ds \\ & + Z_s dW_s, \quad Y_T = h_x^\top, \end{aligned} \tag{2.6}$$

where X_t is the same controlled process in the stochastic optimal control problem, f and h are defined in the cost functional (2.2). The first equation in the FBSDE system is usually called the forward SDE and the second equation is a backward SDE (or BSDE). The solution of the above FBSDE system is a triple (X,Y,Z), which is adapted with the $\mathcal{F}_t$ filtration. The third element in the solution triple, i.e. Z, is the martingale representation of the solution Y such that

$$Z_t = \sigma_t \nabla Y_t, \tag{2.7}$$

where ∇Y_t is the gradient of Y_t with respect to X_t. The existence and uniqueness of solution in (2.6) are guaranteed under mild assumptions and it can be shown that Y and Z are functions of X (see [14, 15]). For the convenience of presentation, in the rest of this paper we abuse the notation and write $\phi(s) := \phi(s, X_s, u_s)$ for a function $\phi : [0,T] \times \mathbb{R}^d \times \mathcal{U} \to \mathbb{R}^d$. With the equation for $\mathcal{D}X_t$ (defined in (2.5)) and the solution Y_t for the FBSDE system (2.6), we have

$$\begin{aligned}
&h_x \mathcal{D}X_T = \langle Y_T, \mathcal{D}X_T \rangle - \langle Y_0, \mathcal{D}X_0 \rangle \\
=&\int_0^T \Big(\big\langle -b_x(s)^\top Y_s - \sigma_x(s)^\top Z_s - f_x(s)^\top, \mathcal{D}X_s \big\rangle \\
&\quad + \langle Y_s, b_x(s)\mathcal{D}X_s + b_u(s)[u_s - u_s^*] \rangle + \langle Z_s, \sigma_x(s)\mathcal{D}X_s + \sigma_u(s)[u_s - u_s^*] \rangle \Big) ds \\
&\quad + \int_0^T \Big(\langle Z_s, \mathcal{D}X_s \rangle + \langle Y_s, \sigma_x(s)\mathcal{D}X_s + \sigma_u(s)[u_s - u_s^*] \rangle \Big) dW_s \\
=&\int_0^T \Big(\big\langle -f_x(s)^\top, \mathcal{D}X_s \big\rangle + \big\langle b_u^\top(s) Y_s + \sigma_u(s)^\top Z_s, u_s - u_s^* \big\rangle \Big) ds \\
&\quad + \int_0^T \Big(\langle Z_s, \mathcal{D}X_s \rangle + \langle Y_s, \sigma_x(s)\mathcal{D}X_s + \sigma_u(s)[u_s - u_s^*] \rangle \Big) dW_s.
\end{aligned}$$

Therefore, (2.4) becomes

$$\begin{aligned}
0 \le& \lim_{\epsilon \to 0} \frac{J(u^\epsilon) - J(u^*)}{\epsilon} \\
=& E\Big[\int_0^T \big(f_x(t)\mathcal{D}X_t + f_u(t)[u_t - u^*(t)] \big) dt + h_x \mathcal{D}X_T \Big] \\
=& E\Big[\int_0^T \Big(\big\langle b_u^\top(t) Y_t + \sigma_u(t)^\top Z_t + f_u(t)^\top, u_t - u_t^* \big\rangle \Big) dt \Big],
\end{aligned}$$

which leads to the representation for the differential expression of J with respect to u, i.e.

$$J'(u)|_t = E\big[b_u^\top(t) Y_t + \sigma_u(t)^\top Z_t + f_u(t)^\top \big] \tag{2.8}$$

for the time instant t.

Based on the above discussions and the differential J' provided in (2.8), in what follows we shall introduce a gradient projection method to determine u^*.

2.2 Gradient projection method for Optimal control

The main theme of our approach in this work is to solve the optimal control problem through an optimization procedure that targets on the minimization task (2.3), and the gradient of the cost functional is derived by the maximum principle which is described

in (2.8). There are many successful optimization methods to achieve the minimization (2.3), such like Newton's method, sequential quadratic programming, gradient descent method and its variations, etc. The framework of our optimization procedure in this work adopts the gradient projection method introduced in [22, 25], which will be discussed in the following. We want to point out that the gradient projection method can be considered as a generalization of gradient descent methods for optimization and therefore the stochastic gradient descent concept that we propose in this work can be generally implemented for any gradient based methods with appropriate formulation of the differential expression J' in (2.8).

For the stochastic optimal control problem (2.1) - (2.3), we observe that

$$\langle J'(u^*), u-u^*\rangle \geq 0, \quad u\in\mathcal{U},$$

where u^* is the optimal control in (2.3). Let r be any given positive constant, the above variational inequality can be rewritten as

$$\big\langle u^* - \big(u^* - rJ'(u^*)\big), u-u^*\big\rangle \geq 0, \quad u\in\mathcal{U}, \tag{2.9}$$

In addition, we define a projection operator $P(\cdot)$ to $\mathcal{U}$ as

$$\|P(w)-w\| = \min_{u\in\mathcal{U}}\|u-w\|, \tag{2.10}$$

which leads to

$$\langle P(w)-w, u-P(w)\rangle \geq 0, \quad \forall u\in\mathcal{U}.$$

If we choose $w = u^* - rJ'(u^*)$, the above inequality becomes

$$\big\langle P\big(u^* - rJ'(u^*)\big) - \big(u^* - rJ'(u^*)\big), u - P\big(u^* - rJ'(u^*)\big)\big\rangle \geq 0, \quad \forall u\in\mathcal{U}. \tag{2.11}$$

By comparing (2.9) with (2.11) and due to the fact of convex optimization, we can see that the optimal control u^* satifsfies

$$u^* = P\big(u^* - rJ'(u^*)\big),$$

which indicates that u^* is the fixed point of the expression $P\big(u - rJ'(u)\big)$. Therefore, the optimal control u^* can be determined through fixed point iteration. Specifically, we choose an initial guess, denoted by u^0, for the control process u. Assume that we have an estimate u^i for u^* at the iteration step i, then the estimated optimal control u^{i+1} is calculated by

$$u^{i+1} = P(u^i - rJ'(u^i)). \tag{2.12}$$

The convergence of the above iteration is guaranteed by the well-posedness of the control problem and we refer to [25] for the proof. In this work, we use the differential expression J' derived in (2.8) to carry out the fixed point iteration. However, there are two specific challenges to implement (2.8). First of all, the expression of J' is an expectation and effective evaluation for the expected value is needed. Secondly, the solution of BSDE, i.e. Y and Z in (2.8), is usually not explicitly solvable, and we need approximate solutions for BSDEs. Therefore, it is necessary to derive effective numerical methods to address both challenges mentioned above. In what follows, we shall introduce our numerical schemes to implement the iteration (2.12).

3 Numerical Optimization for Stochastic Optimal Control

From the formulation of the gradient projection method for the stochastic optimal control problem, we can see that the primary computational challenge is the calculation of J', which is derived in Section 2.1. In this work, we use the expression described in (2.8) to calculate J', which contains solutions of the FBSDE system (2.6). Therefore, the structure of our computational framework is composed by two steps: (i) numerical solution for FBSDEs and (ii) numerical implementation of gradient projection iteration (2.12). The numerical solution that we derive for the FBSDE system (2.6) will be used to construct the gradient J', which will be applied to the iteration of the gradient projection. The main contribution of this work is to introduce a stochastic gradient descent algorithm to replace the deterministic implementation of the aforementioned gradient projection optimization. To proceed, we first introduce a temporal partition Π_N defined by

$$\Pi_N = \{t_n, 0 = t_0 < t_1 < t_2 < \cdots < t_N = T\},$$

where N is a positive integer represents the number of partition points in time and T is the terminal time in the control problem. For convenience of presentation, we assume that Π_N is a uniform partition with $\Delta t = \frac{T}{N}$.

The goal of our numerical approach is to obtain a discrete approximation for the optimal control u^* as a sequence in time, and we aim to find an approximation for the piece-wise constant representation of u^*. Specifically, we define $u^{*,N}$ as

$$u_t^{*,N}\big|_{t\in[t_n t_{n+1})} = u_{t_n}^*,$$

and our efforts focus on finding a sequence that approaches $u^{*,N}$. In this way, the space in which we explore the optimal control becomes $\mathcal{U}_N$, where

$$\mathcal{U}_N := \Big\{ u \in \mathcal{U} \Big| u = \sum_{n=1}^{N} \alpha_n I_{[t_n, t_{n+1})}, \alpha_n \in \mathbb{R} \Big\},$$

and the original optimal control problem becomes its approximate version

$$J(u^{*,N}) = \inf_{u\in\mathcal{U}_N} J(u). \tag{3.13}$$

As a result, the projection operator that we defined in (2.10) takes projection to $\mathcal{U}_N$ and we denoted this operator by $P_N(\cdot)$. Through similar derivation, we get

$$u^{*,N} = P(u^{*,N} - rJ'(u^{*,N})),$$

which leads the gradient projection iteration (2.12) to its discrete counterpart

$$u^{i+1,N} = P_N(u^{i,N} - rJ'_N(u^{i,N})), \tag{3.14}$$

where J'_N is a discrete approximation for J'. The following theorem shows the first order convergence of the above iteration procedure.

Theorem 3.1 *Assume that J' is Lipschitz continuous (with Lipschitz constant C) and uniformly monotone (with the rate of change c) around u^* and $u^{*,N}$, J'_N approaches J' as $N \to \infty$, and ρ is a properly selected constant such that $0 < 1 - 2cr + (1+2C)r^2 < \delta^2$, where $\delta \in (0,1)$. Then, the discretized approximation $u^{i,N}$ obtained in the iteration scheme* (3.14) *gives a first order approximation for u^*, i.e.*

$$\|u^* - u^{i,N}\| \sim O(\Delta t).$$

The proof of Theorem 3.1 can be found in [25] from Theorem 3.1 and Corollary 3.2.

With the above convergence theorem that justifies the temporal discretization of the stochastic optimal control problem, we shall discuss our numerical approach under the discretization Π_N in this section. Since in many practical problems, it's hard to take effective control actions to influence the size of diffusion noise, we assume that the diffusion coefficient σ does not contain the control term, i.e. $\sigma(t)=\sigma(t,X_t)$ in the rest of this paper.

3.1 Numerical solution for FBSDEs

For the temporal partition Π_N, we consider the FBSDE system on time interval $[t_n,t_{n+1}]$, $n=0,1,2,\ldots,N-1$, i.e.

$$\begin{aligned} X_{t_{n+1}} &= X_{t_n}+\int_{t_n}^{t_{n+1}} b(s,X_s,u_s)ds+\int_{t_n}^{t_{n+1}} \sigma(s,X_s)dW_s,\\ Y_{t_n} &= Y_{t_{n+1}}+\int_{t_n}^{t_{n+1}} \Big(b_x(s,X_s,u_s)^\top Y_s+\sigma_x(s,X_s)^\top Z_s+f_x(s,X_s,u_s)^\top\Big)ds \\ &\quad -\int_{t_n}^{t_{n+1}} Z_s dW_s. \end{aligned} \tag{3.15}$$

In order to derive approximation schemes for solutions of the FBSDE system, we first provide numerical integration methods to approximate the integrals in (3.15).

For the forward SDE, we use Euler scheme to approximate the deterministic integral and the so-called *Euler-Maruyama* scheme to approximate the stochastic Itô integral. As a result, we have the following approximation equation for the solution X_t:

$$X_{t_{n+1}} = X_{t_n}+b(t_n,X_{t_n},u_{t_n})\Delta t+\sigma(t_n,X_{t_n})\Delta W_n+R_X^n, \tag{3.16}$$

where $\Delta W_n := W_{t_{n+1}}-W_{t_n}$, and R_X^n is the approximation error of numerical integrations such that

$$R_X^n=\int_{t_n}^{t_{n+1}} b(s,X_s,u_s)ds-b(t_n,X_{t_n},u_{t_n})\Delta t+\int_{t_n}^{t_{n+1}} \sigma(s,X_s)dW_s-\sigma(t_n,X_{t_n})\Delta W_n.$$

For the BSDE, due to the discrepancy between the integration direction of the Itô integral (forward) and the propagation direction of the solution pair (Y,Z) (backward), the Euler-Maruyama scheme for the Itô integral typically results the implicity of the algorithm in solving Z, which requires additional iterations in the scheme. In our approach, we take conditional expectation $E_n[\cdot]$, which is defined by

$$E_n[\cdot] := E[\cdot|\mathcal{F}_{t_n}],$$

on both sides of the BSDE. Then, it follows from the identity

$$E_n[\int_{t_n}^{t_{n+1}} Z_s dW_s]=0$$

that the BSDE in (3.15) becomes

$$Y_{t_n}=E_n[Y_{t_{n+1}}]+\int_{t_n}^{t_{n+1}} E_n\Big[b_x(s,X_s,u_s)^\top Y_s+\sigma_x(s,X_s)^\top Z_s+f_x(s,X_s,u_s)^\top\Big]ds,$$

where we have also used the fact that Y_t is $\mathcal{F}_t$ adapted, which results $E_n[Y_{t_n}] = Y_{t_n}$. Since the propagation direction of Y and Z is from t_{n+1} to t_n, we use the right point formula to approximate the deterministic integral on the right hand side of the above equation. It's worthy to mention that both the right-point formula and left-point formula enjoy the same level of accuracy for the deterministic integral. In this way, the above equation can be written as the following approximation equation

$$\begin{aligned} Y_{t_n} =& E_n[Y_{t_{n+1}}] + E_n\Big[b_x(t_{n+1}, X_{t_{n+1}}, u_{t_{n+1}})^\top Y_{t_{n+1}} \\ &+ \sigma_x(t_{n+1}, X_{t_{n+1}})^\top Z_{t_{n+1}} + f_x(t_{n+1}, X_{t_{n+1}}, u_{t_{n+1}})^\top\Big]\Delta t + R_Y^n, \end{aligned} \tag{3.17}$$

where R_Y^n is the approximation error for the numerical integration with

$$\begin{aligned} R_Y^n =& \int_{t_n}^{t_{n+1}} E_n\Big[b_x(s, X_s, u_s)^\top Y_s + \sigma_x(s, X_s)^\top Z_s + f_x(s, X_s, u_s)^\top\Big] ds \\ &- E_n\Big[b_x(t_{n+1}, X_{t_{n+1}}, u_{t_{n+1}})^\top Y_{t_{n+1}} + \sigma_x(t_{n+1}, X_{t_{n+1}})^\top Z_{t_{n+1}} + f_x(t_{n+1}, X_{t_{n+1}}, u_{t_{n+1}})^\top\Big]\Delta t. \end{aligned}$$

In order to get a numerical scheme for Z, we reconsider the BSDE in (3.15) and multiply ΔW_n to both sides of the equation to get

$$\begin{aligned} Y_{t_n}\Delta W_n =& Y_{t_{n+1}}\Delta W_n + \int_{t_n}^{t_{n+1}} \Big(b_x(s, X_s, u_s)^\top Y_s + \sigma_x(s, X_s)^\top Z_s \\ &+ f_x(s, X_s, u_s)^\top \Delta W_n\Big) ds - \int_{t_n}^{t_{n+1}} Z_s \Delta W_n dW_s. \end{aligned}$$

Then, we take conditional expectation $E_n[\cdot]$ on both sides of the above equation. Since Y_{t_n} is independent from ΔW_n, we have $E_n[Y_{t_n}\Delta W_n] = 0$, and it follows from Itô isometry that

$$\begin{aligned} 0 =& E_n[Y_{t_{n+1}}\Delta W_n] + \int_{t_n}^{t_{n+1}} E_n\Big[b_x(s, X_s, u_s)^\top Y_s + \sigma_x(s, X_s)^\top Z_s \\ &+ f_x(s, X_s, u_s)^\top \Delta W_n\Big] ds - \int_{t_n}^{t_{n+1}} E_n[Z_s] ds. \end{aligned}$$

Approximating the deterministic integrals with the left point formula, the above equation becomes

$$\begin{aligned} 0 =& E_n[Y_{t_{n+1}}\Delta W_n] + E_n\Big[b_x(t_n, X_{t_n}, u_{t_n})^\top Y_{t_n} + \sigma_x(t_n, X_{t_n})^\top Z_{t_n} \\ &+ f_x(t_n, X_{t_n}, u_{t_n})^\top \Delta W_n\Big]\Delta t - E_n[Z_{t_n}]\Delta t + R_Z^n. \end{aligned} \tag{3.18}$$

where R_Z^n is the approximation error. Since X_{t_n}, Y_{t_n}, Z_{t_n} are $\mathcal{F}_{t_n}$ adapted, we have $E_n[Z_{t_n}] = Z_{t_n}$ and

$$E_n\Big[b_x(t_n, X_{t_n}, u_{t_n})^\top Y_{t_n} + \sigma_x(t_n, X_{t_n})^\top Z_{t_n} + f_x(t_n, X_{t_n}, u_{t_n})^\top \Delta W_n\Big]\Delta t = 0.$$

Therefore, the approximation error R_Z^n can be written by

$$\begin{aligned} R_Z^n =& \int_{t_n}^{t_{n+1}} E_n\Big[b_x(s, X_s, u_s)^\top Y_s + \sigma_x(s, X_s)^\top Z_s + f_x(s, X_s, u_s)^\top \Delta W_n\Big] ds \\ &+ E_n[Z_{t_n}]\Delta t - \int_{t_n}^{t_{n+1}} E_n[Z_s] ds, \end{aligned}$$

and we derive the approximation equation for Z from equation (3.18) as following

$$Z_{t_n}\Delta t = E_n[Y_{t_{n+1}}\Delta W_n] + R_Z^n. \tag{3.19}$$

By dropping the error terms R_X^n, R_Y^n and R_Z^n in approximation equations (3.16), (3.17) and (3.19) respectively, we obtain numerical schemes for the solution triple (X,Y,Z) for a given control process u_t. Specifically, for the time instant t_n with the inverse order $n = N-1, N-2, \ldots, 1, 0$ and the given variables X_n, Y_{n+1} and Z_{n+1}, we define approximate solutions (X_{n+1}, Y_n, Z_n) for $(X_{t_{n+1}}, Y_{t_n}, Z_{t_n})$ as following

$$\begin{aligned} X_{n+1} &= X_n + b(t_n, X_n, u_{t_n})\Delta t + \sigma(t_n, X_n)\Delta W_n \\ Y_n &= E_n[Y_{n+1}] + E_n\Big[b_x(t_{n+1}, X_{n+1}, u_{t_{n+1}})^\top Y_{n+1} \\ &\qquad + \sigma_x(t_{n+1}, X_{n+1})^\top Z_{n+1} + f_x(t_{n+1}, X_{n+1}, u_{t_{n+1}})^\top\Big]\Delta t \\ Z_n &= \frac{1}{\Delta t} E_n[Y_{n+1}\Delta W_n]. \end{aligned} \tag{3.20}$$

It's worthy to mention that the propagation direction of X is from 0 to T and the propagation direction of Y, Z is from T to 0. Therefore, our numerical solution X_{n+1} is at the time instant t_{n+1} and the numerical solutions Y_n and Z_n are at the time instant t_n. Given the initial condition $Y_N = Y_T$ and $Z_N = Z_T$, where Z_T is derived through (2.7), the above schemes solve the FBSDE system (2.6) as a recursive algorithm, and the convergence of the schemes is well studied (see [36, 32]).

We also observe that to implement the numerical schemes (3.20), we need to obtain X_n in addition to Y_{n+1} and Z_{n+1}, which can be calculated directly from the schemes. A straight forward option to obtain X_n is to simulate random trajectories of X_t from time 0 to t_n and use the samples at t_n to represent X_{t_n}. However, since X_t is a diffusion process, it's very expensive to calculate an accurate representation for X_{t_n} through path-wise simulations. In [35], the authors use a pre-determined grid mesh to represent X_n, which successfully describes the solutions Y and Z as functions of X. On the other hand, although the grid mesh description of X_n avoids long-term simulations for X_t, it suffers the so-called "curse of dimensionality". It is well-known that the size of grid mesh needs to grow exponentially as the dimension increases.

We want to point out that the above numerical schemes (3.20) provide semi-discretization in time. To derive a full-discretization scheme, we need to handle another challenge in the implementation of the schemes (3.20) – the approximation of conditional expectation. This would influence the accuracy of the algorithm since the numerical approximations for Y and Z are essentially under the conditional expectation E_n. The most widely accepted method to approximate the mathematical expectation is Monte Carlo type methods – although numerical integration methods are also successful alternatives [33], they typically work well only in low dimensional spaces. The central concept of Monte Carlo method is to use the mean values of empirical samples to approximate expectations. In our applications, we calculate the mean value of X_{n+1} samples (given X_n) to approximate the conditional expectation E_n in the schemes for Y and Z. Specifically, for any given initial state $X_n = x$ and a function $\phi(X_{n+1}, Y_{n+1}, Z_{n+1})$, we approximate the conditional expectation $E_n[\phi]$ by

$$\hat{E}_n[\phi(X_{n+1}, Y_{n+1}, Z_{n+1})] = \sum_{m=1}^{M} \frac{\phi(\tilde{X}_{n+1}^{m,x}, \tilde{Y}_{n+1}^{m,x}, \tilde{Z}_{n+1}^{m,x})}{M},$$

where

$$\tilde{X}_{n+1}^{m,x} := x + b(t_n, x, u_{t_n})\Delta t + \sigma(t_n, x)\sqrt{\Delta t}\ \xi^m$$

is the m-th path of X_{n+1}, $\xi^m \sim N(0,1)$ is a Gaussian random sample. We define

$$\tilde{Y}_{n+1}^{m,x} := \tilde{Y}_{n+1}(\tilde{X}_{n+1}^{m,x}), \quad \tilde{Z}_{n+1}^{m,x} = \tilde{Z}_{n+1}(\tilde{X}_{n+1}^{m,x}),$$

where $\tilde{Y}_{n+1}$ and $\tilde{Z}_{n+1}$ are interpolatory approximations for Y_{n+1} and Z_{n+1} over the $\mathbb{R}^d$ space. Since Y_{n+1} and Z_{n+1} are calculated on a pre-determined set of spatial points, denoted by Π_X, $\tilde{Y}_{n+1}(\tilde{X}_{n+1}^{m,x})$ and $\tilde{Z}_{n+1}(\tilde{X}_{n+1}^{m,x})$ are obtained by interpolating functions Y_{n+1} and Z_{n+1} with their values on Π_X.

With the above approximation scheme for conditional expectation and appropriate interpolatory approximation methods, we derive the full-discretization schemes for solving stochastic optimal control related FBSDE system as following:

$$\begin{aligned}
\tilde{X}_{n+1}^{m,x} &= X_n + b(t_n, x, u_{t_n})\Delta t + \sigma(t_n, x)\sqrt{\Delta t}\ \xi^m \\
\tilde{Y}_n(x) &= \sum_{m=1}^{M} \frac{\tilde{Y}_{n+1}^{m,x}}{M} + \sum_{m=1}^{M} \frac{b_x(t_{n+1}, \tilde{X}_{n+1}^{m,x}, u_{t_{n+1}})^\top \tilde{Y}_{n+1}^{m,x}}{M}\Delta t \\
&\quad + \sum_{m=1}^{M} \frac{\sigma_x(t_{n+1}, \tilde{X}_{n+1}^{m,x})^\top \tilde{Z}_{n+1}^{m,x}}{M}\Delta t + \sum_{m=1}^{M} \frac{f_x(t_{n+1}, \tilde{X}_{n+1}^{m,x}, u_{t_{n+1}})^\top}{M}\Delta t \\
\tilde{Z}_n(x) &= \frac{1}{\Delta t}\sum_{m=1}^{M} \frac{\tilde{Y}_{n+1}^{m,x}\sqrt{\Delta t}\ \xi^m}{M},
\end{aligned} \tag{3.21}$$

where $n = N-1, N-2, \ldots, 1, 0$, and $x \in \mathbb{R}^d$ is selected from a pre-determined grid mesh denoted by Π_X. With the above computable schemes (3.21), we can calculate J' by using the expression (2.8) and use gradient projection iteration (3.14) to determine an optimal control u^*.

3.2 Optimization through fully calculated gradient descent iteration

In this subsection, we discuss numerical optimization of the gradient projection approach for stochastic optimal control problems. To proceed, we recall the gradient projection iteration as introduced in (3.14)

$$u^{i+1,N} = P_N(u^{i,N} - rJ'_N(u^{i,N})),$$

where N is the number of temporal partition points in partition Π_N and J'_N (introduced in (2.8)) at time instant t_n is given by

$$J'_N(u)|_{t_n} = E\big[b_u^\top(t_n)Y_{t_n} + \sigma_u(t_n)^\top Z_{t_n} + f_u(t_n)^\top\big],$$

where Y_t and Z_t are solutions of FBSDE system (2.6). Combining the above equations, for a given initial guess for the control process $u^{0,N}$ and a chosen projection operator P_N, the optimization procedure to determine u^* is

$$u^{i+1,N} = P_N\Big(u^{i,N} - rE\big[b_u^\top(t_n)Y_{t_n} + \sigma_u(t_n)^\top Z_{t_n} + f_u(t_n)^\top\big]\big|_{u^{i,N}}\Big), \quad i = 0,1,2,\cdots \tag{3.22}$$

where the expectation $E[\cdot]_{u^{i,N}}$ indicates that the random variables in the expectation are derived under the control term $u^{i,N}$. In order to implement the above optimization

procedure numerically, we can use $\tilde{Y}_n$, $\tilde{Z}_n$ calculated in (3.21) as approximations for Y_{t_n}, Z_{t_n} and have the approximate optimization procedure as

$$u^{i+1,N} = P_N\Big(u^{i,N} - rE\big[b_u^\top(t_n)\tilde{Y}_n + \sigma_u(t_n)^\top \tilde{Z}_n + f_u(t_n)^\top\big]\big|_{u^{i,N}}\Big), \quad i=0,1,2,\cdots \tag{3.23}$$

Notice that there's another expectation in (3.23), so we need to apply another Monte Carlo approximation to simulate $E[\cdot]\big|_{u^{i,N}}$, i,e. for each iteration step i, the above scheme becomes

$$u^{i+1,N} = P_N\Big(u^{i,N} - \frac{r}{Q}\sum_{q=1}^{Q}\big(b_u^\top(t_n)\tilde{Y}_n^q + \sigma_u(t_n)^\top \tilde{Z}_n^q + f_u(t_n)^\top\big)\big|_{u^{i,N}}\Big), \tag{3.24}$$

where Q is the number Monte Carlo samples to approximate $E[\cdot]\big|_{u^{i,N}}$, $\tilde{Y}_n^q$ and $\tilde{Z}_n^q$ are defined by

$$\tilde{Y}_n^q := \tilde{Y}_n(\tilde{X}_n^q), \quad \tilde{Z}_n^q = \tilde{Z}_n(\tilde{X}_n^q), \tag{3.25}$$

and $\tilde{X}_n^q$ is the q-th realization of the path-wise simulation of the SDE with the Euler-Maruyama scheme from time $t=0$ to time $t=t_n$. It's worthy to recall that $\tilde{Y}_n^q$, $\tilde{Z}_n^q$ are obtained by using interpolatory approximation with values of $\tilde{Y}_n$, $\tilde{Z}_n$ on Π_X as introduced in (3.21). Also, numerical approximations $\tilde{Y}_n$ and $\tilde{Z}_n$ are based on the current estimation of the optimal control, i.e. $u^{i,N}$, and at each iteration step we need to calculate $\tilde{Y}_n$ and $\tilde{Z}_n$ to update the new estimate for u in the schemes. Although the Monte Carlo simulation for $E[\cdot]_{u^{i,N}}$ does not require extra approximations for Y and Z (with only interpolation cost at the time instant t_n), there's significant computational effort to calculate $\tilde{Y}_n$ and $\tilde{Z}_n$.

Before we move forward to stochastic gradient descent, in the rest of this subsection we shall have a brief discussion on the computational cost of the numerical implementation of the iteration scheme (3.24) for solving the stochastic optimal control problem.

We can see from the fully discretized schemes (3.21) that for each pre-selected spatial point $X_n = x$ and a given random sample ξ^m, an important computational cost is the interpolatory approximation for Y and Z, i.e. to obtain $\tilde{Y}_{n+1}^{m,x}$ and $\tilde{Z}_{n+1}^{m,x}$. Besides interpolation, another component of computational cost in the scheme (3.21) is to simulate $\{\tilde{X}_{n+1}^m\}_{m=1}^M$ and calculate the corresponding functions, which is relatively small compare to interpolation. If we let the number of spatial points at time level n be L_n, the primary computational cost to implement the numerical schemes (3.21) can be described by

$$C_{YZ}^n \approx (C_{interp.} + C_{sampl.}) \times L_n \times M \tag{3.26}$$

where M is the number of Monte Carlo samples to approximate the conditional expectation E_n, C_{YZ}^n is the rough computational cost for solving $\tilde{Y}_n$ and $\tilde{Z}_n$ at certain time level n, $C_{interp.}$ denotes the computational cost of interpolatory approximation to obtain $\tilde{Y}_{n+1}^{m,x}$ and $\tilde{Z}_{n+1}^{m,x}$, and $C_{sampl.}$ denotes the computational cost for sampling one realization of $\tilde{X}_{n+1}^m$ together with evaluating functions corresponding to the sample $\tilde{X}_{n+1}^m$ in the schemes. Here, we have ignored marginal computational costs such like simple summation and multiplication in (3.21).

In addition to solving the FBSDE system with the total cost of $\sum_{n=0}^{N-1} C_{YZ}^n$ (calculating from time t_0 to t_{N-1}), since the representation for J' with Y and Z is also under expectation, the calculation of $J_N'(u)|_{t_n}$ requires another Monte Carlo sampling for the forward SDE X_t from time 0 to t_n as we described in (3.24). As a result, the total cost

of maximum principle based gradient projection approach for stochastic optimal control problems with the above classic numerical implementation is

$$\text{total cost} \approx K\Big(\sum_{n=0}^{N-1} C_{YZ}^n + \sum_{n=0}^{N-1} QC_{J'}^n\Big), \tag{3.27}$$

where K is the number of iteration to determine u^* in the projection gradient descent, and $C_{J'}^n$ is the cost of generating one realization of X_t sample, i.e. $\tilde{X}_n^q$ in (3.25), together with the cost of interpolation for $\tilde{Y}_n^q$ and $\tilde{Z}_n^q$ in (3.24). Similar to the argument in (3.26), we have ignored marginal computational costs in (3.27).

From the above discussion for the computational cost, we can see that the pricey term in (3.27) is the accurate approximation for solutions of FBSDEs, i.e. $K\Big(\sum_{n=0}^{N-1} C_{YZ}^n\Big)$. At every time level t_n, the computational cost C_{YZ}^n involves interpolation and function evaluations and the number of calculations depends on both the number of Monte Carlo samples, i.e. M, and the number of spatial points, i.e. L_n, on which we evaluate solutions Y and Z. Especially, when the dimension of the problem increases, the number of spatial points typically increases exponentially, which results significant growth of the computational cost. In what follows, we introduce our methodology of applying stochastic gradient descent for solving the stochastic optimal control problem under the maximum principle framework.

3.3 Stochastic gradient descent for stochastic optimal control

In this subsection, we first give a brief discussion on the classic application of stochastic gradient descent (SGD) for optimization problems. Then, we shall discuss how we apply SGD to solve the optimal control problem through gradient projection.

Consider the following optimization problem

$$F(v^*) = \min_{v\in\mathcal{V}} E\big[F(v,\Gamma)\big], \tag{3.28}$$

where v is the optimization parameter with the optimum v^* selected from the set $\mathcal{V}$, F is the objective function governed by the parameter v and Γ is a random variable with a given distribution. The standard gradient descent method takes gradient of the objective function as a direction to improve the estimation for the target parameter through iteration, i.e.

$$v^{i+1} = v^i - rE\big[\nabla F(v,\Gamma)\big], \qquad i = 0,1,2,\cdots, \tag{3.29}$$

where v^0 is the initial guess for v^* and r is the step-size (sometimes called the learning rate). Under certain restrictions and assumptions, the above gradient descent iteration for v^i converges to the optimal parameter v^*. In practice, the expectation in (3.28) is usually approximated by Monte Carlo methods, i.e. for a given sample size M, the iteration scheme (3.29) becomes

$$v^{i+1} = v^i - r\frac{\sum_{m=1}^{M}\nabla F(v,\gamma_m)}{M}, \qquad i = 0,1,2,\cdots, \tag{3.30}$$

where γ_m is a sample of the random variable Γ. One challenge in the above gradient descent (3.30) is that repeating sampling the random variable Γ in ∇F requires M times calculation of ∇F to get one update for the estimation of v^*. When the calculation for

∇F is time comsuming, the optimization procedure (3.30) is computationally expensive. The main theme of stochastic gradient descent is that, in stead of carrying out complete evaluation for the expectation in the optimization problem (3.28), we can use one realization of sample in Γ to be a rough approximation for the expectation. Although the expectation is not well approximated, it only requires one-time calculation for ∇F, compared to M times calculation for ∇F in (3.30), to get an updated estimate v^{i+1}. Specifically, in the stochastic gradient descent method, the gradient iteration is given by

$$v^{i+1} = v^i - r\nabla F(v, \gamma_i), \qquad i = 0, 1, 2, \cdots. \tag{3.31}$$

It's worthy to mention that the low accuracy of the single-sample approximation for expectation results more iteration steps to achieve an accurate estimate for the target parameter v^*. However, practical applications of stochastic gradient descent indicate that it's a more efficient approach to solve the optimization problem – especially when the objective function F is a complicated high dimensional model [].

From the above discussion, we can see that the main contribution of stochastic gradient descent is to transfer the expensive computational cost in estimating expectation to more iteration steps in the gradient descent optimization. Analysis for the efficiency and effectiveness of the stochastic gradient descent are extensively studied due to its successful applications in machine learning [27, 31].

In this work we adopt the methodology of stochastic gradient descent as presented in (3.31) to solve the stochastic optimal control problem. By comparing the optimization procedure (3.22) with (3.29), we can see that the expectation term in (3.22) is an alternative expression for J'_N, which plays the role of $E\big[\nabla F(v, \Gamma)\big]$ in (3.29), and (3.24) is an analogue of (3.30) with Monte Carlo approximation for expectation. Then, adopting the structure of stochastic gradient descent introduced in (3.31) in the gradient projection optimization as described by (3.22), we obtain the following stochastic gradient descent scheme

$$u^{i+1,N} = P_N\Big(u^{i,N} - r\big[b_u^\top(t_n)Y_{t_n}^i + \sigma_u(t_n)^\top Z_{t_n}^i + f_u(t_n)^\top\big]\big|_{u^{i,N}}\Big), \quad i = 0, 1, 2, \cdots \tag{3.32}$$

where $Y_{t_n}^i$ and $Z_{t_n}^i$ are i-th samples of solutions Y_{t_n} and Z_{t_n}, and $\big[b_u^\top(t_n)Y_{t_n}^i + \sigma_u(t_n)^\top Z_{t_n}^i + f_u(t_n)^\top\big]\big|_{u^{i,N}}$ can be considered as a single-sample approximation for the expectation expression of the gradient $J'_N(u^{i,N})$. From the above iteration scheme and the discussion on the computational cost of gradient descent approach for solving the stochastic optimal control problem, we can see that the primary computational cost in (3.32) is to generate samples $Y_{t_n}^i$ and $Z_{t_n}^i$. Apparently, obtaining accurate approximations for Y_{t_n} and Z_{t_n} through (3.21) and then generating samples $Y_{t_n}^i$ and $Z_{t_n}^i$ requires extensive computation with cost C_{YZ}^n as we discussed in (3.26). From the semi-discretization scheme (3.20), we notice that the semi-discretization schemes for Y_{t_n} and Z_{t_n} also involve expectations, i.e. $E_n[\cdot]$. Since we are using a single-sample to represent Y_{t_n} and Z_{t_n} in (3.32), we pass the *"single-sample approximation for expectation"* concept in the stochastic gradient descent to the numerical solution for FBSDEs. Although this would cause low-accuracy in approximating solutions of FBSDEs, since we are using solutions of FBSDEs to represent the gradient of J' as indicated in (3.22), the solutions Y and Z essentially play the role of gradient that is under expectation. In this way, the single-sample approximation for expectation is applicable for representing solutions of FBSDEs in the stochastic gradient descent method.

More specifically, we let $M = 1$ in the full-discretized scheme (3.21) and use $\tilde{X}_{t_n}^i$, $\tilde{Y}_{t_n}^i$ and $\tilde{Z}_{t_n}^i$ to denote the single-sample approximation for X_{t_n}, Y_{t_n} and Z_{t_n} respectively.

With initial condition $\tilde{X}_0^i = X_0$, we first sample the entire path of $\tilde{X}_t^i$, i.e. $\{\tilde{X}_{t_n}^i\}_{n=1}^N$ by

$$\tilde{X}_{t_{n+1}}^i = \tilde{X}_{t_n}^i + b(t_n, \tilde{X}_{t_n}^i, u_{t_n})\Delta t + \sigma(t_n, \tilde{X}_{t_n}^i)\sqrt{\Delta t}\ \xi_{t_n}^i, \quad n = 0, 1, \cdots, N-1, \tag{3.33}$$

where $\xi_{t_n}^i$ is a sample drew from a standard Gaussian distribution. Then, we derive $\tilde{Y}_{t_n}^i$ and $\tilde{Z}_{t_n}^i$ with initial condition $\tilde{Y}_T^i = Y_T(\tilde{X}_T^i)$ and $\tilde{Z}_T^i = Z_T(\tilde{X}_T^i)$ as following

$$\begin{aligned} \tilde{Y}_{t_n}^i =& \tilde{Y}_{t_{n+1}}^i + b_x(t_{n+1}, \tilde{X}_{t_{n+1}}^i, u_{t_{n+1}})^\top \tilde{Y}_{t_{n+1}}^i \Delta t, \\ &+ \sigma_x(t_{n+1}, \tilde{X}_{t_{n+1}}^i)^\top \tilde{Z}_{t_{n+1}}^i \Delta t + f_x(t_{n+1}, \tilde{X}_{t_{n+1}}^i, u_{t_{n+1}})^\top \Delta t \\ \tilde{Z}_{t_n}^i =& \frac{\tilde{Y}_{t_{n+1}}^i \xi_{t_n}^i}{\sqrt{\Delta t}}. \end{aligned} \tag{3.34}$$

Recall that $b_u^\top$, $f_u^\top$ are functions of t, X_t and u_t, and σ_u is a function of t and X_t, we now have the following computable version of scheme (3.32) to solve for the optimal control u^*

$$u^{i+1,N} = P_N\Big(u^{i,N} - r\big[b_u^\top(t_n, \tilde{X}_{t_n}^i, u^{i,N})\tilde{Y}_{t_n}^i + \sigma_u^\top(t_n, \tilde{X}_{t_n}^i)\tilde{Z}_{t_n}^i + f_u^\top(t_n, \tilde{X}_{t_n}^i, u^{i,N})\big]\Big), \tag{3.35}$$

where $\tilde{X}_{t_n}^i$, $\tilde{Y}_{t_n}^i$ and $\tilde{Z}_{t_n}^i$ are sample-wise approximation obtained in schemes (3.33)-(3.34).

The schemes (3.33)-(3.34) and (3.35) compose the computational framework of our stochastic gradient descent approach for solving the stochastic optimal control problem, and we can tell that the schemes (3.33)-(3.34) carry out the computational task of "Numerical solution for FBSDEs" as discussed in Section 3.1 and the scheme (3.35) carries out the computational task of "Optimization through gradient based iteration" as discussed in Section 3.2. It's important to point out that the numerical schemes (3.33)-(3.34) for solving the FBSDEs is essentially different from the fully discretized schemes (3.21) and we avoid computational cost in both Monte Carlo approximation for the conditional expectation $E_n[\cdot]$ and spatial dimension approximation for Y_{t_n} and Z_{t_n} on all the spatial points, i.e. Π_X, that we use to describe X_{t_n}. As a result, the computational effort for implementing the numerical simulation for obtaining $\tilde{Y}_{t_n}^i$ and $\tilde{Z}_{t_n}^i$ only consists the cost of generating samples governed by the schemes (3.33)-(3.34), multiplied by the total number of optimization iterations K. In this way, the computational cost C_{SGD} of our stochastic gradient descent approach can be simply described by

$$C_{SGD} \approx K C_{path}, \tag{3.36}$$

where K is the total number of iterations required in (3.35) and C_{path} is the cost of generating one realization of $\tilde{X}_{t_n}^i$, $\tilde{Y}_{t_n}^i$ and $\tilde{Z}_{t_n}^i$ from t_0 to T through schemes (3.33)-(3.34).

4 Numerical experiments

In this section, we present three numerical experiments to demonstrate the performance of our stochastic gradient descent approach for solving the stochastic optimal control problem – we denote it "SGD-SOC" for convenience of presentation in this section. In the first example, we solve a classic 1-dimensional stochastic optimal control problem and compare effectiveness and efficiency of our approach with the gradient projection

method as a benchmark method for deterministic gradient descent approach. Then, in Example, 2 we solve a 3-dimensional stochastic optimal control problem, which would challenge most existing grid-based methods due to the exponential increase of spatial grid points. In Example 3, we are going to present the performance of our SGD-SOC in solving a feedback control problem which we have full information on the state of controlled process.

Example 1

In this example, we consider the following controlled process

$$dX_t = u_t X_t dt + \sigma X_t dW_t, \quad X_0 = x_0 \in \mathbb{R}^2,$$

where both the drift and diffusion terms are linear functions of X_t, and σ in the diffusion coefficient is a constant. The cost functional is given by

$$J(u) = 0.5 \int_0^T E[(X_t - X_t^*)^2] dt + 0.5 \int_0^T u_t^2 dt.$$

From [22], we know that if we let X_t^* in the cost functional be

$$X_t^* = \frac{e^{\sigma^2 t} - (T-t)^2}{1 - Tt + \frac{t^2}{2}} + 1,$$

the control u_t^* for the above stochastic optimal control problem has the following explicit expression

$$u_t^* = \frac{T-t}{\frac{1}{x_0} - Tt + \frac{t^2}{2}}. \tag{4.37}$$

In our numerical experiments, we choose $T = 1$ and discretize the time interval $[0,1]$ with 20 time steps, i.e. $N = 20$ and $\Delta t = 0.05$. We also let $x_0 = -0.8$ and $\sigma = 0.01$ in the controlled process. For the SGD-SOC, the learning rate in the gradient optimization is set to be $r = 10^{-3}$ so that our stochastic optimization procedure would have more stable performance. In figure 1, we compare the SGD-SOC estimate u_t^* with the estimated optimal control obtained by using the standard gradient descent under the gradient projection framework introduced in [25]. The red curve marked by crosses is the analytic optimal control u_t^* given by (4.37); the black curve marked by crosses is the estimated optimal control obtained by the gradient projection method with deterministic optimization procedure; and the blue curve marked by circles gives the estimated optimal control obtained by using SGD-SOC. We can see from the figure that both gradient projection method and SGD-SOC work well and their estimates are very close to the true optimal control. In what follows, we give a brief discussion on the computational cost of implementing the optimization procedure.

When solving the FBSDEs (2.6) with conventional methods (on full spatial grid mesh), we use 50 spatial points to describe the state of controlled process X_{t_n} in this experiment, i.e. $L_n = 50$. To simulate expectations E_n and E, we take 1000 samples in Monte Carlo approximation, i.e. $M = Q = 1000$, and the total number of iteration we use is $K = 1000$. On the other hand, we use the total number of $K = 2 \times 10^7$ iterations in the SGD-SOC to obtain the result as presented in the figure. From the cost estimate (3.27), we know that the computational cost through deterministic gradient descent

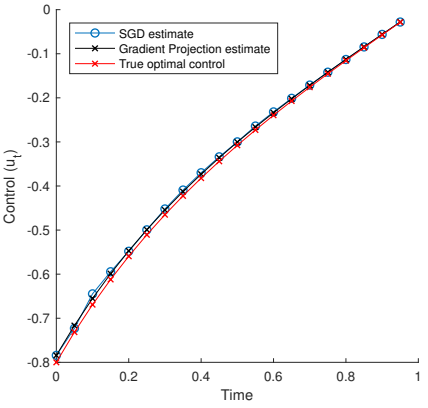


Figure 1: Example 1. Comparison

type approach is approximately

$$\begin{aligned}\text{total cost} \approx & K\Big(\sum_{n=0}^{N-1} C_{YZ}^n + \sum_{n=0}^{N-1} QC_{J'}^n\Big)\\ \approx & KM\sum_{n=0}^{N-1}(C_{interp.}+C_{sampl.})\times L_n + KQ\sum_{n=0}^{N-1} C_{J'}^n \\ = & 5\times 10^7 \times \sum_{n=0}^{N-1}(C_{interp.}+C_{sampl.}) + 10^6 \sum_{n=0}^{N-1} C_{J'}^n,\end{aligned} \tag{4.38}$$

given that $K=1000$, $M=Q=1000$, $L_n=50$ as we choose in this numerical experiment, and the first part of the above cost analysis is the total cost of solving FBSDEs numerically. However, the total computational cost of the SGD-SOC is roughly $2\times 10^7 \times C_{path}$, where C_{path} is the cost of path-wise simulation of single-sample solutions as we introduced in (3.36). Since the cost of sampling and evaluating an entire stochastic process, i.e. C_{path}, is similar to summing the cost of one temporal step sampling $C_{sampl.}$ (defined in (3.26)), i.e. $\sum_{n=0}^{N-1} C_{sampl.}$, the computational cost of SGD-SOC is close to $2\times 10^7 \times \big(\sum_{n=0}^{N-1} C_{sampl.}\big)$, which is only a portion of the computational cost of solving FBSDEs numerically and is much lower than implementing the fully calculated deterministic gradient descent approach with the total cost given by (4.38).

Example 2

In this example, we consider the following three dimensional controlled process

$$dX_t = (u_t\mathbf{1} - \mathbf{r}_t)dt + \sigma dW_t, \quad X_0 = x_0 \in \mathbb{R}^3, \tag{4.39}$$

where $\mathbf{1}=(1,1,1)^T$, u_t is a scaler control process, $\mathbf{r}_t$ is a given process in $\mathbb{R}^3$ chosen as $\mathbf{r}_t=(\frac{t}{2},\frac{t}{2},\frac{t}{2})^T$, and σ is an identity matrix $\sigma := \mathbf{I}_3$. The cost function in the control problem is defined by

$$J(u) = 0.5\Big[\int_0^T \mathbf{c}_1 E[(X_t - X_t^*)^2]dt + c_2\int_0^T u_t^2 dt\Big], \tag{4.40}$$

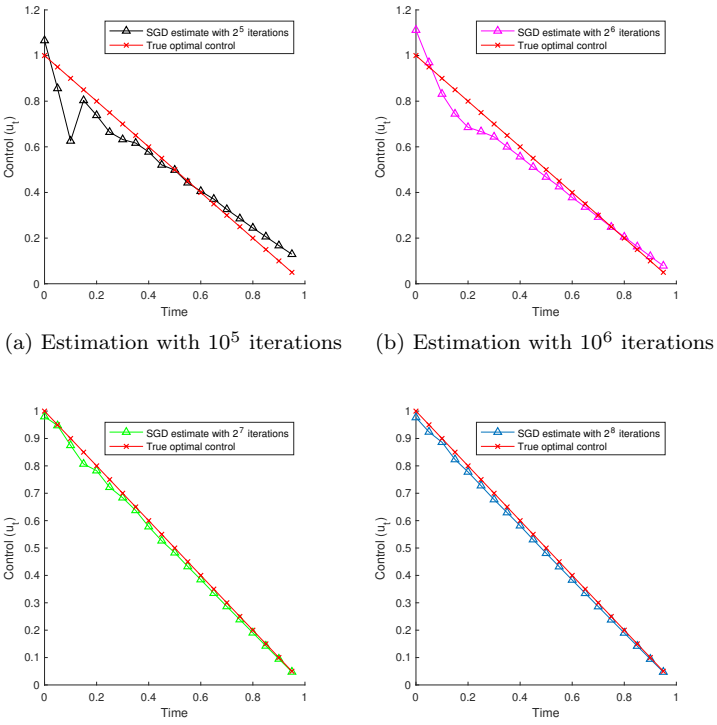
where we choose $\mathbf{c}_1 = (3,1,2)$ and $c_2 = 1$. If we let

$$X_t^* = E[X_t] + \frac{dp_t}{dt} = (3Tt - \frac{t^2}{2}, 3Tt - \frac{t^2}{2}, 3Tt - \frac{t^2}{2})^T + (-\frac{1}{2}, 0, 1)^T,$$

one can derive that

$$u_t^* = 3T - \frac{t}{2},$$

is the analytic optimal control for the above stochastic optimal control problem (4.39)-(4.40). Since this is a 3-dimensional control problem, the number of spatial points needed to describe X_t increases exponentially, which would significant increase the computational cost of fully calculated deterministic gradient descent approach as we discussed in (4.38). Therefore, the deterministic gradient descent type optimization is prohibitive in this example and we only demonstrate the performance of our SGD-SOC approach.



(a) Estimation with 10^5 iterations (b) Estimation with 10^6 iterations

(c) Estimation with 10^7 iterations (d) Estimation with 10^8 iterations

Figure 2: Example 2. Performance of estimation for control with 2^5, 2^6, 2^7 and 2^8 iteration steps in the optimization procedure.

In the numerical experiments, we let $T = 1$ and discretize the time interval $[0,1]$ with 20 partition points, i.e. $\Delta t = 0.05$. In Figure 2, we demonstrate the performance of SGD-SOC estimation for the optimal control u_t^* with different iteration steps in the stochastic gradient descent procedure (3.35), i.e. $K = 10^5, 10^6, 10^7, 10^8$. Specifically, in Figures 2 (a), we plot the estimation performance with 10^5 iterations, where the red curve marked by crosses is the real optimal control and the black curve marked by triangles is the estimate of SGD-SOC; in Figures 2 (b), the red curve marked by crosses

is the real optimal control and the magenta curve marked by triangles is the estimate of SGD-SOC with 10^6 iterations; in Figures 2 (c), the red curve marked by crosses is the real optimal control and the green curve marked by triangles is the estimate of SGD-SOC with 10^7 iterations; and in Figures 2 (d), the red curve marked by crosses is the real optimal control and the blue curve marked by triangles is the estimate of SGD-SOC with 10^8 iterations. From these subplots, we can see that the accuracy of SGD-SOC estimation improves as the number of iteration increases and both the 10^7-iteration estimate and 10^8-iteration estimate are very close to the true optimal control u_t^*.

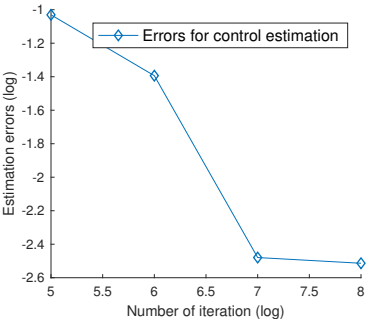


Figure 3: Example 2. Convergence of errors.

To further demonstrate the convergence of accuracy with increasing iteration steps, we calculate the approximation errors and present the overall errors with respect to the number of iterations in Figure 3. The x-axis in the figure shows the number of iterations ($\log(K)$) and the y-axis is the overall logarithmic estimate error for u_t^* under L_2 norm. From this figure, we can see a clear convergence trend with more iteration steps.

Example 3

In this example, we aim to examine the performance of SGD-SOC in solving a feedback control problem. Consider the following controlled process

$$dX_t = A(t)X_t dt + BU_t dt + CdW_t, \quad X_0 = x_0,$$

where $A(t)$ is a given function, B C are given constants and U_t is the control process. The cost functional $J(U)$ is defined by

$$J(U) = E\Big[0.5\int_0^T \Big(< Q(s)X_s, X_s > + R(s) < U_s, U_s >^2 \Big) ds + 0.5M < X_T, X_T >^2 \Big],$$

where $Q(t)$ is a function, and R, M are also given constants.

It is well known that the above linear-quadratic stochastic optimal control is a feedback control, in which the control U_t depends on the state of the controlled process. One can derive that the optimal control U_t^* for the above stochastic optimal control problem is

$$U_t^* = -R^{-1}(t)B^T P(t)X_t, \tag{4.41}$$

where $P(t)$ is the unique, symmetric, positive definite solution of the following Riccati equation

$$\frac{dP(t)}{dt} = -P(t)A(t) - A^T(t)P(t) + P(t)BR^{-1}(t)B^TP(t) - Q(t), \quad P(T) = M.$$

For the 1-dimensional case, which we will solve in this example, the above equation becomes

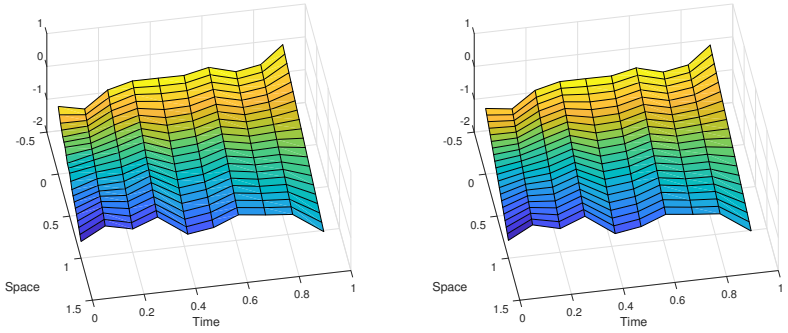
$$dP(t) = (-2P(t)A(t) + \frac{P^2(t)B^2}{R(t)} - Q(t))dt, \quad P(T) = M.$$

Since the optimal control depends on the state of controlled process, the approximation for the control process needs to consider the spatial dimension in addition to the temporal dimension. Therefore, in the feedback control problem that we solve in this example, we approximate the optimal control on a time-space domain and the control U_t^* is described by a "control surface" – instead of a control trajectory as we presented in the first two examples. In this way, the expectation in the stochastic maximum principle, which requires Monte Carlo type simulations for X_t from time 0 to t_n, is no longer needed and the expression for J' (as defined in (2.8)) becomes

$$J'_N(u)|_{t_n} = b_u(t_n, x, u_{t_n})Y_{t_n}(x) + \sigma_u(t_n, x)Z_{t_n}(x) + f_u(t_n, x, u_{t_n}),$$

where $x \in \mathbb{R}$ is a given spatial point on which we estimate its corresponding control. As a result, the simulation for $\tilde{X}^i_{t_n}$ from time 0 to t_n is not needed either in the SGD-SOC approach since we can use a pre-selected "point of interest" x to replace $\tilde{X}^i_{t_n}$ in schemes (3.33)-(3.35). At the same time, the stochastic gradient descent approach is still necessary since approximations for the conditional expectation E_n when solving FBSDEs can not be avoided and stochastic gradient descent could reduce computational effort by using the single-sample $\xi^i_{t_n}$ (introduced in (3.33)) to evaluate conditional expectations.

In the first case that we examine in this example, we let $A(t) = 0.5$, $B = 0.5$, and $C = 0.1$ in the controlled process with the initial state of the controlled process as $X_0 = 0.5$. For the cost function, we let $Q(t) = 1$, $R = M = 1$. In the numerical experiment, we let $T = 1$ and discretize the time interval $[0,1]$ with 10 steps, i.e. $\Delta t = 0.1$. In Figure



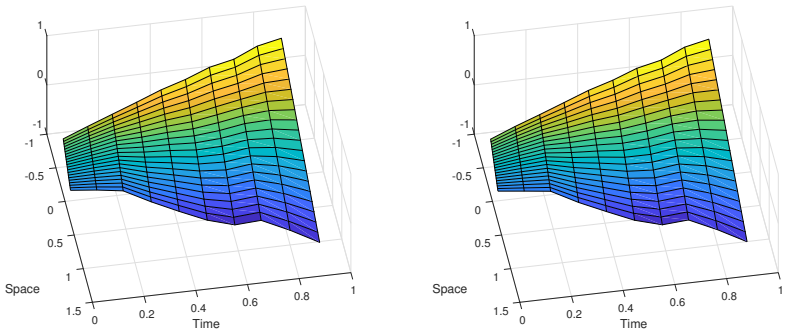
(a) Estimate control surface (b) Real control surface

Figure 4: Example 3. Performance of estimation for control – case 1.

4 we compare our estimated optimal control (in subplot (a)) with the real optimal

control surface U_t^* (in subplot (b)) over the time-space domain, where the axes represent "time", "space" and values of control actions. Since the feedback control framework only requires estimation for the optimal control on pre-selected spatial points, long-term simulations for the control process is not required. In this experiment, we choose the number of iteration steps to be $K = 10^5$. From this figure, we can see that our estimate control surface is very close to the real control as derived in (4.41).

In the second case that we examine in this example, we let $A(t) = \sin(t)$, $B = 0.5$, and $C = 0.01$ in the controlled process with the initial state of the controlled process as $X_0 = 0.1$. For the cost function, we let $Q(t) = \exp(-t)$, $R = M = 1$. In Figure 5 we



(a) Estimate control surface (b) Real control surface

Figure 5: Example 3. Performance of estimation for control – case 2.

compare our estimated optimal control (in subplot (a)) with the real control surface U_t^* (in subplot (b)) over the time-space domain, where the axes also represent "time", "space" and values of control actions. In this experiment, we also choose the number of iteration steps to be $K = 10^5$, and we can see from this figure that our estimate control surface is very close to the real control as derived in (4.41).

5 Acknowledgement

This work is partially supported by the Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research through FASTMath Institute and CompFUSE project. The second author also acknowledges support by U.S. National Science Foundation under Contract DMS-1720222. The third author acknowledges the partial support by NSF grant DMS-1812921.

References

[1] 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, J. Nonlinear Sci.*, 29 (2019), 1563–1619.

[2] R. Bellman, *An introduction to the theory of dynamic programming,* RAND Corp. Report 153, 1949.

[3] R. Bellman, *Dynamic Programming*, Princeton Univ. Press, 1957.

[4] A. Bensoussan, *Stochastic maximum principle for distributd parameter systems, J. Franklin Inst.,* 315 (1983), 387–406.

[5] J. M. Bismut, *An introductory approach to duality in optimal stochastic control, SIAM Rev.,* 20 (1978), 62–78.

[6] V. G. Boltyanskii, R. V. Gamkrelidze, L. S. Pontryagin, *On the theory of optimal processes, Dokl. Akad. Nauk SSSR (N.S.),* 110 (1956), 7–10 (in Russian).

[7] M. G. Crandall, H. Ishii, P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc. (N.S.),* 27 (1992), 1–67.

[8] W. H. Fleming, *Future Directions in Control Theory: A Mathematical Perspective,* SIAM, 1989.

[9] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions, 2nd Ed.,* Springer, 2006.

[10] U. G. Haussmann, *General necessary conditions for optimal control of stochastic systemsn Math. Prog. Study,* 6 (1976), 34–48.

[11] R. E. Kalman, *Contributions to the theory of optimal control, Bol. Soc. Mat. Mexicana,* 5 (1960), 102–119.

[12] H. J. Kushner, *On the stochastic maximum principle: Fixed time of control, J. Math. Anal. Appl.,* 12 (1965), 78–92.

[13] J. Ma and J. Yong, *Forward-Backward Stochastic Differential Equations and Their Applications,* Lecture Notes in Math. vol. 1702, Springer-Verlag, 1999.

[14] É. Pardoux and S. Peng, *Adapted solution of a backward stochastic differential equation, Systems Control Lett.,* 14 (1990), 55–61.

[15] S. Peng, *A general stochastic maximum principle for optimal control problems, SIAM J. Control Optim.,* 28 (1990), 966–979.

[16] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, E. F. Mishchenko, *The Mathematical Theory of Optimal Processes,* John Wiley & Sons, Inc., New York, 1962.

[17] J. Sun and J. Yong, *Stochastic Linear-Quadratic Optimal Control Theory: Open-Loop and Closed-Loop Solutions,* Springer, to appear.

[18] J. Yong, and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations,* Springer-Verlag, New York, 1999.

[19] J. Ma, P. Protter, and J. Yong. Solving forward-backward stochastic differential equations explicitly — a four-step scheme. *Probab. Theory & Rel. Fields*, 98:339–359, 1994.

[20] Feng Bao, Yanzhao Cao, Amnon Meir, and Weidong Zhao. A first order scheme for backward doubly stochastic differential equations. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):413–445, 2016.

[21] Feng Bao, Yanzhao Cao, and Weidong Zhao. A backward doubly stochastic differential equation approach for nonlinear filtering problems. *Commun. Comput. Phys.*, 23(5):1573–1601, 2018.

[22] Ning Du, Jingtao Shi, and Wenbin Liu. An effective gradient projection method for stochastic optimal control. *Int. J. Numer. Anal. Model.*, 10(4):757–774, 2013.

[23] Xiaobing Feng, Roland Glowinski, and Michael Neilan. Recent developments in numerical methods for fully nonlinear second order partial differential equations. *SIAM Rev.*, 55(2):205–267, 2013.

[24] Xiaobing Feng and Max Jensen. Convergent semi-Lagrangian methods for the Monge-Ampère equation on unstructured grids. *SIAM J. Numer. Anal.*, 55(2):691–712, 2017.

[25] Bo Gong, Wenbin Liu, Tao Tang, Weidong Zhao, and Tao Zhou. An efficient gradient projection method for stochastic optimal control problems. *SIAM J. Numer. Anal.*, 55(6):2982–3005, 2017.

[26] Behzad Kafash, Ali Delavarkhalafi, and Seyed Mehdi Karbassi. A computational method for stochastic optimal control problems in financial mathematics. *Asian J. Control*, 18(4):1501–1512, 2016.

[27] Q. Li, C. Tai, and W. E. Stochastic modified equations and dynamics of stochastic gradient algorithms i: mathematical foundations. *Journal of Machine Learning Research*, pages 40–47, 2019.

[28] I. Sato and H. Nakagawa. Convergence analysis of gradient descent stochastic algorithms. *Proceedings of the 31st International Conference on Machine Learning*, pages 982–990, 2014.

[29] A. Shapiro and Y. Wardi. Convergence analysis of gradient descent stochastic algorithms. *Journal of Optimization Theory and Applications*, pages 439–454, 1996.

[30] Iain Smears and Endre Süli. Discontinuous Galerkin finite element methods for time-dependent Hamilton-Jacobi-Bellman equations with Cordes coefficients. *Numer. Math.*, 133(1):141–176, 2016.

[31] L. Wu, C Ma, and W. E. How sgd selects the global minima in over-parameterized learning: A dynamical stability perspective. *NeurIPS 2018*, pages 8289–8298, 2018.

[32] Jie Yang, Guannan Zhang, and Weidong Zhao. A first-order numerical scheme for forward-backward stochastic differential equations in bounded domains. *J. Comput. Math.*, 36(2):237–258, 2018.

[33] Guannan Zhang, Max Gunzburger, and Weidong Zhao. A sparse-grid method for multi-dimensional backward stochastic differential equations. *J. Comput. Math.*, 31(3):221–248, 2013.

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

[35] Weidong Zhao, Lifeng Chen, and Shige Peng. A new kind of accurate numerical method for backward stochastic differential equations. *SIAM J. Sci. Comput.*, 28(4):1563–1581, 2006.

[36] Weidong Zhao, Yu Fu, and Tao Zhou. New kinds of high-order multistep schemes for coupled forward backward stochastic differential equations. *SIAM J. Sci. Comput.*, 36(4):A1731–A1751, 2014.