Convergence of the deep BSDE method for stochastic control problems formulated through the stochastic maximum principle

Huang, Zhipeng; Negyesi, Balint; Oosterlee, Cornelis W.

DOI
10.1016/j.matcom.2024.08.002

Publication date
2025

Document Version
Final published version

Published in
Mathematics and Computers in Simulation

Citation (APA)
Huang, Z., Negyesi, B., & Oosterlee, C. W. (2025). Convergence of the deep BSDE method for stochastic control problems formulated through the stochastic maximum principle. *Mathematics and Computers in Simulation*, *227*, 553-568. https://doi.org/10.1016/j.matcom.2024.08.002

Mathematics and Computers in Simulation 227 (2025) 553–568

Contents lists available at ScienceDirect

Mathematics and Computers in Simulation

journal homepage: www.elsevier.com/locate/matcom

Original articles

Convergence of the deep BSDE method for stochastic control problems formulated through the stochastic maximum principle

Zhipeng Huang [a,*], Balint Negyesi [b], Cornelis W. Oosterlee [a]

[a] Mathematical Institute, Utrecht University, The Netherlands
[b] Delft Institute of Applied Mathematics (DIAM), Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Stochastic control
Deep SMP-BSDE
Stochastic maximum principle
Vector-valued FBSDE

ABSTRACT

It is well-known that decision-making problems from stochastic control can be formulated by means of a forward–backward stochastic differential equation (FBSDE). Recently, the authors of Ji et al. (2022) proposed an efficient deep learning algorithm based on the stochastic maximum principle (SMP). In this paper, we provide a convergence result for this deep SMP-BSDE algorithm and compare its performance with other existing methods. In particular, by adopting a strategy as in Han and Long (2020), we derive *a-posteriori estimate*, and show that the total approximation error can be bounded by the value of the loss functional and the discretization error. We present numerical examples for high-dimensional stochastic control problems, both in the cases of drift- and diffusion control, which showcase superior performance compared to existing algorithms.

1. Introduction

Stochastic control theory is a powerful paradigm for modeling and analyzing decision-making problems that are subject to some random dynamics. Classical approaches for solving these kinds of problems include methods based on the dynamic programming principle (DP) [1], the stochastic maximum principle (SMP) [2,3] and other techniques, see e.g. [4–7]. However, these approaches cannot easily handle high-dimensional problems and suffer from the "curse of dimensionality". A recent candidate solution technique for stochastic control problems is formed by deep learning-based approaches, due to their remarkable performance in high-dimensional settings. The paper [8] developed a deep learning algorithm that directly approximates the optimal control process by a neural network at each step in time, and by training a terminal loss functional for all time steps simultaneously. Similar approaches have been explored in the control theory community [9–11], before the rise of computing power and machine learning.

Inspired by the remarkable performance in [8], the research community has developed several neural network-based algorithms for stochastic control problems, see e.g. [12–14]. Many of these algorithms build upon deriving a forward–backward stochastic differential equation (FBSDE), associated with the control problem. Pioneered by the well-known deep BSDE method, initially proposed by [15] and later extended by [16] to the coupled setting, several solution approaches have been proposed, showing outstanding empirical performance in high-dimensional frameworks. These methods mainly derive the FBSDE through a stochastic representation of the solution of the Hamilton–Jacobi–Bellman (HJB) equation, motivated by the non-linear Feyman–Kac formula. However, such techniques become infeasible when the diffusion of the state process is also controlled, as they do not solve for the value function's second derivative, which is necessary to compute the optimal (diffusion) control. As a remedy, enabling diffusion control, the authors in [17] proposed a deep BSDE algorithm where the associated FBSDE is derived from the stochastic maximum

* Corresponding author.
E-mail address: z.huang1@uu.nl (Z. Huang).

https://doi.org/10.1016/j.matcom.2024.08.002
Received 30 January 2024; Received in revised form 31 July 2024; Accepted 3 August 2024
Available online 10 August 2024

principle (SMP), which we call the deep SMP-BSDE. Other SMP-based algorithms include [18,19] in the context of mean-field control and mean-field games. We refer to [20,21] for a detailed overview of deep learning algorithms for stochastic control problems.

Furthermore, the authors in [22] found that deep BSDE methods may fail to converge for FBSDEs stemming from stochastic control problems via DP, due to local minima. They proposed a robust counterpart by adding a regularization component to the loss function which resolved this issue in the case of drift control. Despite the fact that there are many studies concerning stochastic control, only a few theoretical derivations are available regarding the convergence of machine learning-based approaches for FBSDEs stemming from stochastic control. Theoretical work in this direction includes the study of convergence of the deep BSDE method by [16], which provides *a-posteriori estimate*, and a non-Lipschitz counterpart by [23] that only allows the diffusion coefficient to be non-Lipschitz and independent of the BSDE. The authors of [22] prove the convergence of a robust deep BSDE method by exploiting the special structure of the FBSDE. In more recent research, the authors in [24] proposed an efficient and reliable *a-posteriori estimate* for fully coupled McKean–Vlasov FBSDEs, which naturally extended the error estimator for decoupled FBSDEs studied in [25]. For other works in the decoupled framework, see e.g. [26–28].

In this paper, we provide convergence results for the deep SMP-BSDE algorithm and compare the method with existing algorithms supported by numerical results. Unlike [15,22], we consider FBSDEs that come from the SMP, instead of the HJB equation, and therefore the results and standard estimates for FBSDE stemming from DP cannot be directly applied. Nevertheless, with some extra effort, we are able to adopt a similar strategy as in [16], and derive the *a-posteriori estimate* for the numerical solutions of the deep SMP-BSDE algorithm in a multi-dimensional setting. By this, we are able to tackle diffusion control problems.

This paper is organized as follows: In Section 2, we briefly review the theoretical foundations related to the SMP. In Section 3, we formulate the deep SMP-BSDE algorithm for diffusion control problems. We carry out a convergence analysis in Section 4, and, in particular, *a-posteriori estimate* is derived for the deep SMP-BSDE algorithm. In Section 5, we demonstrate the performance of the algorithm through numerical examples both in the cases of drift- and diffusion control.

2. Background

In this section, we review some basic results from stochastic control theory and show how to reformulate a stochastic control problem into an FBSDE through the SMP.

Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\le t\le T}, \mathbb{P}\right)$ be a filtered probability space, supporting an m-dimensional Brownian motion W and its natural filtration $\mathcal{F} = \{\mathcal{F}_t\}_{0\le t\le T}$, augmented by all $\mathbb{P}$-null sets. Fixing $0 < T < \infty$, we consider the following finite horizon stochastic control problem

$$\begin{cases} \inf\limits_{u\in\mathcal{U}[0,T]} J\left(0, x_0; u(\cdot)\right) := \inf\limits_{u\in\mathcal{U}[0,T]} \mathbb{E}\left(\int_0^T \bar{f}(s, X_s, u_s)\, ds + g(X_T)\right), \\ X_t = x_0 + \int_0^t \bar{b}(s, X_s, u_s)\, ds + \int_0^t \bar{\sigma}(s, X_s, u_s)\, dW_s, \quad t \in [0, T], \end{cases} \tag{1}$$

where $\bar{b} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^\ell \to \mathbb{R}^d$, $\bar{\sigma} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^\ell \to \mathbb{R}^{d\times m}$, $\bar{f} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^\ell \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ are deterministic functions, and X_t, u_t are $\mathbb{R}^d, \mathbb{R}^\ell$-valued stochastic processes, respectively. The set of admissible controls, $\mathcal{U}[0,T]$, is defined as

$$\mathcal{U}[0,T] := \left\{u : [0,T]\times\Omega \to U \mid u \in L^2_{\mathcal{F}}\left(0, T; \mathbb{R}^\ell\right)\right\},$$

with

$$L^2_{\mathcal{F}}\left(0, T; \mathbb{R}^\ell\right) := \left\{x : [0,T]\times\Omega \to \mathbb{R}^\ell \mid x \text{ is } \mathcal{F}\text{-adapted and } E\left[\int_0^T \|x_t\|^2\, dt\right] < \infty\right\},$$

where we shall denote $\|\cdot\|$ for both the usual Euclidean norm and the Frobenius norm for matrices. We assume that the control domain U is a convex body in $\mathbb{R}^\ell$.

Any process $u_t \in \mathcal{U}[0,T]$ is called an admissible control of (1), and the (X_t, u_t) consisting of the corresponding state process is called an admissible pair. Furthermore, (X_t, u_t) is an optimal pair whenever the infimum of (1) is achieved, and accordingly, we define the value function $V(\cdot,\cdot)$ of (1) as follows

$$\begin{cases} V(t,x) := \inf\limits_{u\in\mathcal{U}[t,T]} J(t, x; u(\cdot)), \quad \forall (t,x) \in [0,T)\times\mathbb{R}^d, \\ V(T,y) := g(y), \quad \forall y \in \mathbb{R}^d, \end{cases} \tag{2}$$

and denote its partial derivatives as $V_x := \partial_x V(t,x)$, $V_{xx} := \partial_{xx} V(t,x)$ and $\partial_t V = \partial_t V(t,x)$ without specifying their dependencies in (t,x).

Remark 1. We shall further distinguish two important classes of problem (1). We call (1) a drift control problem if the drift coefficient depends on the control u_t but not the diffusion coefficient, and it is called a diffusion control problem if the diffusion coefficient also includes u_t as an argument.

Associated with the stochastic control problem (1), we introduce the adjoint equation, as follows

$$\begin{cases} P_t = P_0 - \int_0^t \nabla_x \bar{H}(s, X_s, u_s, P_s, Q_s)\, ds + \int_0^t Q_s\, dW_s, \quad t \in [0,T] \\ P_T = -\nabla_x g(X_T), \end{cases} \tag{3}$$

where the $\nabla_x \bar{H}$ is the derivative of Hamiltonian $\bar{H}$, which is defined by

$$\bar{H}(t,x,u,p,q) := p^\top \bar{b}(t,x,u) + \mathrm{Tr}\left(q^\top \bar{\sigma}(t,x,u)\right) - \bar{f}(t,x,u),$$
$$\forall (t,x,u,p,q) \in [0,T] \times \mathbb{R}^d \times U \times \mathbb{R}^d \times \mathbb{R}^{d\times m}.$$

Eq. (3) is a BSDE whose solution is formed by a pair of processes, $(P(\cdot), Q(\cdot)) \in L^2_{\mathcal{F}}\left(0,T;\mathbb{R}^d\right) \times \left(L^2_{\mathcal{F}}\left(0,T;\mathbb{R}^d\right)\right)^m$. Concerning the well-posedness of BSDE (3), we state the following assumption first.

Assumption 1. Let $\bar{\varphi} = \bar{b}, \bar{\sigma}, \bar{f}$ and g. The map $\bar{\varphi}$ is C^2 in x, and $\bar{\varphi}(t,0,u)$ is bounded for any $(t,u) \in [0,T]\times U$. Moreover, $\bar{\varphi}$, $\bar{\varphi}_x$ and $\bar{\varphi}_{xx}$ are uniformly Lipschitz in x and u.

Remark 2. With Assumption 1, the adjoint equation, or BSDE (3), admits a unique solution for every admissible pair (X_t, u_t). However, there is only one of the admissible 4-tuples (X_t, u_t, P_t, Q_t) that also minimizes the objective function in (1). Therefore, to have an FBSDE that admits a unique solution without including the objective function of the control problem, we need extra effort in the reformulation. For this purpose, we recall the SMP below.

Theorem 1 (*Stochastic Maximum Principle*). *Let Assumption 1 hold, and* $\left(X_t^*, u_t^*, P_t^*, Q_t^*\right)$ *be an admissible 4-tuple. Suppose that* $g(\cdot)$ *is convex,* $\bar{H}\left(t,\cdot,\cdot,P_t^*,Q_t^*\right)$ *defined by 1 is concave for all* $t \in [0,T]$ $\mathbb{P}$ *almost surely, and the maximum condition*

$$\bar{H}\left(t, X_t^*, u_t^*, P_t^*, Q_t^*\right) = \max_{u\in U} \bar{H}\left(t, X_t^*, u, P_t^*, Q_t^*\right), \quad \text{a.e. } t \in [0,T], \quad \mathbb{P}\text{-a.s.} \tag{4}$$

holds. Then, $\left(X_t^*, u_t^*\right)$ *is an optimal pair of problem (1).*

Remark 3. The proof of this theorem is found in [29, pp. 149–150], or, in a more general setting, [30, pp. 138–140]. Theorem 1 provides a sufficient condition for the optimal control u_t^*, when certain concavity and convexity conditions hold, which are crucial in general, see for instance, [30, Example 3.1, pp. 138–140]. On the other hand, Theorem 1 itself does not constitute a necessary condition unless there is no diffusion control in problem (1). In the rest of this paper, we shall use superscript $*$ to indicate that a process is associated with the optimal control u_t^* whenever it needs to be further distinguished.

Under sufficient smoothness and concavity assumptions of $\bar{H}$ in Theorem 1, the optimization (4) is uniquely solved by the following first-order conditions,

$$\begin{aligned} \nabla_u \bar{H}(t, X_t^*, u_t^*, P_t^*, Q_t^*) =& (\nabla_u \bar{b}(t, X_t^*, u_t^*))^\top P_t^* \\ &+ \nabla_u \mathrm{Tr}(\bar{\sigma}^\top(t, X_t^*, u_t^*) Q_t^*) - \nabla_u \bar{f}(t, X_t^*, u_t^*) = 0. \end{aligned} \tag{5}$$

Under the setting of Algorithm 3 in [17], without loss of generality we assume that the first-order condition yields an explicit formula for the mapping $\mathcal{M} : (t, X_t^*, P_t^*, Q_t^*) \mapsto u_t^*$,

$$u_t^* = \mathcal{M}(t, X_t^*, P_t^*, Q_t^*), \quad \forall t \in [0,T]. \tag{6}$$

We will call this function the *feedback map*, and remark that for a rather wide range of interesting problems such an expression is available in closed-form.

Let us define the map $\varphi = b, \sigma, f$ and g, which is given by the composition $\varphi := \bar{\varphi}(t,x,\mathcal{M}(t,x,p,q))$ for $\bar{\varphi} = \bar{b}, \bar{\sigma}$ and $\bar{f}$.

Similarly, we define the function $\bar{F}(t,x,u,p,q) := \nabla_x \bar{H}(t,x,u,p,q)$ and write

$$F(t,x,p,q) := \bar{F}(t,x,\mathcal{M}(t,x,p,q),p,q). \tag{7}$$

With this in hand, we reformulate (3) and the controlled SDE of (1) as a fully-coupled FBSDE, for $t \in [0,T]$,

$$\begin{cases} X_t = x_0 + \int_0^t b(s, X_s, P_s, Q_s)\, ds + \int_0^t \sigma(s, X_s, P_s, Q_s)\, dW_s, \\ P_t = P_0 - \int_0^t F(s, X_s, P_s, Q_s)\, ds + \int_0^t Q_s\, dW_s, \\ P_T = -\nabla_x g(X_T). \end{cases} \tag{8}$$

We call the BSDE part in (8), subject to its boundary condition, the SMP-BSDE.

Remark 4. An important feature of (8) is that its solution is equivalent to the solution of (1). Let (X_t^*, u_t^*) be an optimal pair obtained by (1), then there exists a unique pair (P_t^*, Q_t^*) which follows from Remark 2. Therefore, we have the optimal (X_t^*, P_t^*, Q_t^*) that also solves (8). On the other hand, solving (8) gives us $(\tilde{X}_t, \tilde{P}_t, \tilde{Q}_t)$, and we obtain $\tilde{u}_t$ by the feedback map (6). Consequently, we must have $\tilde{u}_t = u_t^*$ and $(\tilde{X}_t, \tilde{P}_t, \tilde{Q}_t) = (X_t^*, P_t^*, Q_t^*)$, due to the uniqueness of feedback map (6) and Theorem 1.

3. Existing algorithms and the deep SMP-BSDE algorithm

In this section, we first give a comparison between the existing deep BSDE algorithm and the deep SMP-BSDE algorithm, and show why the former cannot be used to solve diffusion control problems.

Suppose the DP holds, then the value function $V(t,x)$ of problem (1) solves the following HJB equation

$$\begin{cases} -\partial_t V + \sup_{u\in U} \mathcal{G}\left(t,x,u,-V_x,-V_{xx}\right) = 0, \\ V(T,x) = g(x), \end{cases} \tag{9}$$

where the generalized Hamiltonian $\mathcal{G}$ is defined by

$$\mathcal{G}\left(t,x,u,-V_x,-V_{xx}\right) := -\frac{1}{2}\operatorname{Tr}\left(\bar{\sigma}(t,x,u)^\top V_{xx}\bar{\sigma}(t,x,u)\right) - V_x^\top \bar{b}(t,x,u) - \bar{f}(t,x,u)$$

By the stochastic verification theorem, see [30, pp. 268–269], a given admissible pair (X_t^*, u_t^*) is optimal if and only if

$$\mathcal{G}\left(t,X_t^*,u_t^*,-V_x,-V_{xx}\right) = \max_{u\in U}\mathcal{G}\left(t,X_t^*,u,-V_x,-V_{xx}\right),$$

provided that V is a classical solution to (9) with sufficient smoothness.

Suppose (9) admits a unique classical solution given by the value function $V(t,x)$. Then, it is straightforward to obtain a new feedback map $\tilde{\mathcal{M}}$ by maximizing $\mathcal{G}$ over u, which allows us to write the optimal control u_t^* in terms of X_t^*, V_x and V_{xx}

$$u_t^* = \tilde{\mathcal{M}}(t, X_t^*, -V_x, -V_{xx}), \qquad t\in[0,T].$$

Notice that whenever the diffusion includes the control variable, $\tilde{\mathcal{M}}$ will always depend on V_{xx}. From the stochastic representation of $V(t,x)$, we obtain for $t\in[0,T]$

$$V(t,x) = g(X_T^*) + \int_t^T \bar{f}(s,X_s^*,u_s^*)\,ds - \int_t^T V_x^\top \bar{\sigma}(t,X_t^*,u_t^*)\,dW_s. \tag{10}$$

Eq. (10) defines a BSDE whose unique solution coincides with $(Y_t,Z_t) = \left(V(t,X_t^*), V_x^\top\bar{\sigma}(t,X_t^*,u_t^*)\right)$, whenever the feedback map $\tilde{\mathcal{M}}$ is only a function of V_x but not of V_{xx}. However, this is only the case when the diffusion coefficient does not depend on u, i.e., in the case of drift control. In particular, Eq. (10) coincides with the deep BSDE method of [15,22], applied to stochastic control problems. In numerical settings where one does not have direct access to V_{xx}, this makes the dynamic programming approaches described above infeasible, whenever the diffusion coefficient depends on u.

Moreover, the robust counterpart of the deep BSDE method in [22], which first simulates X_t forward in time and computes the samples of Y_0, denoted by $\mathcal{Y}_0$, backwards in time, according to (10), introduces an objective functional of the form

$$\inf_{\theta^Z} E\left(\mathcal{Y}_0\right) + \lambda \operatorname{Var}\left(\mathcal{Y}_0\right) \tag{11}$$

for some parametric space θ^Z of the neural networks for approximating Z_t and a chosen constant $\lambda>0$. Such formulation builds upon the facts that the corresponding DP approach results in a BSDE whose driver is independent of $Y_t = V(t,X_t^*)$, and $Y_0 = V(t_0,x_0)$ is a deterministic quantity coinciding with the value function of the control problem. Moreover, $\operatorname{Var}\left(\mathcal{Y}_0\right)$ is shown to be equal to the terminal condition of the BSDE, see [22, sec. 3], and therefore (11) should be minimized. However, this robustness technique does not apply to the SMP-BSDE setting, as in general F in (7) depends on P_t, and since $P_0 = -V_x(t_0,x_0)$ its mean is not necessarily minimized at t_0. From this point of view, the deep SMP-BSDE algorithm seems particularly promising for solving general stochastic control problems.

For the reasons above, in order to be able to treat diffusion control problems, in what follows we focus on the SMP-based FBSDE formulation. Let us state the discrete scheme of the deep SMP-BSDE algorithm originating from Algorithm 3 of [17], but instead of using one single neural network for both P_0 and the process Q_t, as in the aforementioned paper, we approximate P_0 and Q_t at each step in time by a separate neural network, respectively. We consider the following classical Euler scheme corresponding to (8)

$$\inf_{\mu_0^\pi\in\theta_0^P,\phi_i^\pi\in\theta_i^Q} E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2, \tag{12a}$$

$$\text{s.t.}\begin{cases} X_0^\pi = x_0, \\ P_0^\pi = \mu_0^\pi(x_0), \\ X_{t_{i+1}}^\pi = X_{t_i}^\pi + b\left(t_i, X_{t_i}^\pi, P_{t_i}^\pi, Q_{t_i}^\pi\right)\Delta t \\ \qquad\qquad + \sigma\left(t_i, X_{t_i}^\pi, P_{t_i}^\pi, Q_{t_i}^\pi\right)\Delta W_i, \\ Q_{t_i}^\pi = \phi_i^\pi\left(X_{t_i}^\pi\right), \\ P_{t_{i+1}}^\pi = P_{t_i}^\pi - F\left(t_i, X_{t_i}^\pi, P_{t_i}^\pi, Q_{t_i}^\pi\right)\Delta t + Q_{t_i}^\pi \Delta W_i, \end{cases} \tag{12b}$$

with a time partition, $\pi: 0=t_0<t_1<\cdots<t_N=T$, $h=T/N$, $t_i=ih$ and $\Delta W_i := W_{t_{i+1}} - W_{t_i}$ for $i=0,1,\ldots,N-1$. We recall the definitions of b, σ and F in Section 2. Moreover, we let θ_0^P and θ_i^Q be the corresponding parametric spaces for the neural networks $\theta_0^P \ni \mu_0^\pi : \mathbb{R}^d \to \mathbb{R}^d$ and $\theta_i^Q \ni \phi_i^\pi : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, respectively. The objective function (12a) serves as the loss function in the machine

learning algorithm, and through the training of the neural networks we wish to find appropriate functions $\mu_0^\pi(x_0)$ and $\phi_i^\pi\left(X_{t_i}^\pi\right)$ that can approximate P_0 and Q_{t_i} sufficiently well. The complete pseudo-code for the deep SMP-BSDE algorithm is given in Algorithm 1, which will be used in the later Section 5 for numerical experiments.

Algorithm 1 deep SMP-BSDE algorithm.

Input: Initial parameters $\left(\theta_0^P, \theta_0^Q, \ldots, \theta_{N-1}^Q\right)$, learning rate η; batch size M; number of iteration K.
Data: Simulated Brownian increments $\left\{\Delta W_{t_i,k}\right\}_{0\le i\le N-1, 1\le k\le K}$
Output: The triple $(X_{t_i}, P_{t_i}, Q_{t_i})$
for $k = 1$ to K **do**
  $X_{t_0,k}^\pi = x_0$, $P_{t_0,k}^\pi = \mu_0^\pi(x_0;\theta_0^P)$
  for $i = 0$ to $N-1$ **do**
    $Q_{t_i,k}^\pi = \phi_i^\pi\left(X_{t_i}^\pi;\theta_i^Q\right)$
    $u_{t_i,k}^\pi = \mathcal{M}\left(t_i, X_{t_i,k}^\pi, P_{t_i,k}^\pi, Q_{t_i,k}^\pi\right)$
    $X_{t_{i+1},k}^\pi = X_{t_i,k}^\pi + \bar{b}\left(t_i, X_{t_i,k}^\pi, u_{t_i,k}^\pi\right)\Delta t_i + \bar{\sigma}\left(t_i, X_{t_i,k}^\pi, u_{t_i,k}^\pi\right)\Delta W_{t_i,k}$
    $P_{t_{i+1},k}^\pi = P_{t_i,k}^\pi - \bar{F}\left(t_i, X_{t_i,k}^\pi, u_{t_i,k}^\pi\right)\Delta t_i + Q_{t_i,k}^\pi \Delta W_{t_i,k}$
  end for
  $\text{Loss} = \frac{1}{M}\sum_{j=1}^M \left\| -\nabla_x g\left(X_{t_N,k}^\pi\right) - P_{t_N,k}^\pi \right\|^2$
  $\left(\theta_0^P, \theta_0^Q, \ldots, \theta_{N-1}^Q\right) \longleftarrow \left(\theta_0^P, \theta_0^Q, \ldots, \theta_{N-1}^Q\right) - \eta\nabla\text{Loss}$
end for

4. Convergence analysis

This section is dedicated to the convergence analysis for the deep SMP-BSDE method, reviewed in Section 2, and the discrete scheme (12b). In particular, we show that the total error of the numerical solution to the FBSDE is bounded by the time discretization error and the simulation error of the objective function, and such error in theory could be made arbitrarily small in a sufficiently fine grid, due to the universal approximation theorem. Our analysis follows a similar strategy as [16]. For the sake of this section, in order to be able to use techniques established therein, we need the following restriction.

Assumption 2. The coefficients $\bar{b}, \bar{\sigma}$ in (1) imply a feedback map such that the compositions $\bar{b}(t,x,\mathcal{M}(t,x,p,q))$ and $\bar{\sigma}(t,x,\mathcal{M}(t,x,p,q))$ do not depend on q.

Even though this condition may seem abstract, note that it in particular covers the important subclass of drift control problems, i.e. whenever $\bar{\sigma}$ does not depend on u. In that case, under the convexity conditions of Theorem 1, the feedback map in (6) is only a function of (t, X_t, P_t) and not of Q_t — see the first-order conditions in (5). Consequently, the solution pair of the backward equation in the FBSDE (8) only couples into the forward component via P_t, similarly as in [16]. Nevertheless, in what follows we follow a more general presentation with Assumption 2, such that special cases of diffusion control problems can also be treated.

Our main contribution is establishing a convergence result for the numerical solution of the stochastic control problem (1) by Algorithm 1, where the corresponding FBSDE is vector-valued and derived from the SMP. In this regard, our results can be viewed as an extension to [16], where the backward equation in (8) was only scalar-valued.

We acknowledge that a recent article [24] has studied the convergence for a more general problem formulation, i.e., a fully coupled McKean–Vlasov FBSDE (MV-FBSDE), where the maps b, σ and F may also depend on the Q_t process and the law of the solution. The main difference between these two works is the set of assumptions and the corresponding methods for studying the well-posedness of the discretization of the equation. To be specific, [24] have studied the well-posedness and stability of an implicit Euler discretization of the MV-FBSDE, and consequently adapted the method of continuation for studying this discretization, which requires a structural monotonicity assumption about the MV-FBSDE. On the contrary, our work utilizes the well-posedness and convergence of the implicit scheme from [31], which is developed through a fixed-point argument and uses a different set of weak coupling and monotonicity conditions formally stated in [16,31]. Indeed, by adopting the structural monotonicity assumptions that are also used in [32,33], the authors of [24] study a more general error estimator compared to the one in [16], and extends the work [25] to a coupled MV-FBSDE setting. On the other hand, our work is an extension of [16] and a convergence study of the algorithm proposed by [17], building on a different set of weak coupling and monotonicity conditions, and therefore these two works do not contain each other.

To conduct our analysis, we require the following technical, standing assumptions to hold.

Assumption 3. Suppose

1. The maps $\bar{b}, \bar{\sigma}, \bar{F}$ are uniformly Hölder-$\frac{1}{2}$ continuous in t.
2. The feedback map $\mathcal{M}(t,x,p,q)$ is uniformly Hölder-$\frac{1}{2}$ continuous in t, and Lipschitz continuous in x, p and q, respectively.

Assumption 4. There exist constants k^b and k^F, that are possibly negative, such that

$$\begin{aligned}\left(b(t,x_1,p)-b(t,x_2,p)\right)^\top \Delta x &\le k^b\|\Delta x\|^2,\\ \left(F(t,x,p_1,q)-F(t,x,p_2,q)\right)^\top \Delta p &\le k^F\|\Delta p\|^2.\end{aligned}$$

Remark 5. With Assumptions 1, 2 and 3, this implies that b,σ,F and g are uniformly Lipschitz continuous in all spatial variables, and therefore we may write

$$\begin{aligned}\left\|b(t,x_1,p_1)-b(t,x_2,p_2)\right\|^2 &\le L_x^b\|\Delta x\|^2+L_p^b\|\Delta p\|^2,\\ \left\|\sigma(t,x_1,p_1)-\sigma(t,x_2,p_2)\right\|^2 &\le L_x^\sigma\|\Delta x\|^2+L_p^\sigma\|\Delta p\|^2,\\ \left\|F\left(t,x_1,p_1,q_1\right)-F\left(t,x_2,p_2,q_2\right)\right\|^2 &\le L_x^F\|\Delta x\|^2+L_p^F\|\Delta p\|^2+L_q^F\|\Delta q\|^2,\\ \left\|\nabla_x g(x_1)-\nabla_x g(x_2)\right\|^2 &\le L_x^{g_x}\|\Delta x\|^2.\end{aligned}$$

Similarly, the Hölder-continuity of b,F and σ follow from the same set of assumptions, which imply that $b(t,0,0),F(t,0,0,0)$ and $\sigma(t,0,0)$ are bounded in t, and the boundedness of $\nabla_x g(x)$ directly follows from Assumption 1. For convenience, we use $\mathscr{L}$ to denote the set of all constants mentioned above and denote its upper bound by L.

Next, we introduce the following system of quasi-linear parabolic PDEs, which is naturally associated with (8),

$$\begin{cases}\partial_t v^i+\frac{1}{2}\operatorname{Tr}\left(\partial_{xx}v^i\sigma\sigma^\top(t,x,v)\right)+\partial_x v^i b\left(t,x,v,\partial_x v\sigma(t,x,v)\right)\\ \qquad +F^i\left(t,x,v,\partial_x v\sigma(t,x,v)\right)=0,\quad \forall i=1,\dots,d;\\ v(T,x)=-\nabla_x g(x),\end{cases}\tag{13}$$

where v^i denotes the ith component of the vector v. Note that this is not the same as HJB Eq. (9), which is associated with a different FBSDE given by (10).

For this system of PDEs, we shall recall the so-called weak and monotonicity conditions studied in [16,31], as well as in earlier literature [34,35]. With these additional conditions, the system of PDEs (13) admits a unique viscosity solution v and there is a unique solution $\left(X_t,P_t,Q_t\right)$ to the FBSDE (8), connected by $v\left(t,X_t\right)=P_t$. For a complete statement of these conditions and results, we refer to [16,31].

Remark 6. It is worth to mention that with additional stronger assumptions such as smoothness and boundedness conditions, the system of PDEs (13) admits a unique classical solution, and enjoys the representations $v(t,X_t)=P_t$ and $\partial_x v(t,X_t)\sigma(t,X_t,v(t,X_t))=Q_t$ via the nonlinear Feynman–Kac formulae. Moreover, a connection to the derivatives of value function $V(t,X_t^*)$ (2) evaluated at the optimal pair (X_t^*,u_t^*) to the control problem can be established by

$$P_t=-V_x(t,X_t^*),\quad Q_t=-V_{xx}(t,X_t^*)\bar\sigma(t,X_t^*,u_t^*),\tag{14}$$

that follows directly from [29, pp. 151–152] or [36, pp. 250–253], see also Remark 4.

Although a classical solution enables us to solve control problems whenever both V_x and V_{xx} are of interest, e.g. delta-gamma hedging problems in finance, we stick with the viscosity solution setting which requires weaker conditions and is sufficient enough for us to derive the results in this paper. With the viscosity solution v, also called decoupling field in BSDE literature, one can decouple the FBSDE (8) and obtain the following.

Theorem 2 (*Convergence of the Implicit Scheme*). *Suppose Assumptions 1–4 hold, and furthermore let the weak and monotonicity conditions in [16] hold. Then for a sufficiently small h, the following discrete-time equation for $0\le i\le N-1$,*

$$\begin{cases}\bar X_0^\pi=x_0,\\ \bar X_{t_{i+1}}^\pi=\bar X_{t_i}^\pi+b\left(t_i,\bar X_{t_i}^\pi,\bar P_{t_i}^\pi\right)h+\sigma\left(t_i,\bar X_{t_i}^\pi,\bar P_{t_i}^\pi\right)\Delta W_i,\\ \bar P_T^\pi=-\nabla_x g\left(\bar X_T^\pi\right),\\ \bar Q_{t_i}^\pi=\frac{1}{h}E\left(\bar P_{t_{i+1}}^\pi\Delta W_i^\top\mid\mathcal{F}_{t_i}\right),\\ \bar P_{t_i}^\pi=E\left(\bar P_{t_{i+1}}^\pi+F\left(t_i,\bar X_{t_i}^\pi,\bar P_{t_i}^\pi,\bar Q_{t_i}^\pi\right)h\mid\mathcal{F}_{t_i}\right),\end{cases}\tag{15}$$

has a solution, $\left(\bar X_{t_i}^\pi,\bar P_{t_i}^\pi,\bar Q_{t_i}^\pi\right)$, such that $\bar X_{t_i}^\pi\in L^2\left(\Omega,\mathcal{F}_{t_i},\mathbb{P}\right)$ and

$$\begin{aligned}&\sup_{t\in[0,T]}\left(E\left\|X_t-\bar X_t^\pi\right\|^2+E\left\|P_t-\bar P_t^\pi\right\|^2\right)\\ &\qquad+\int_0^T E\left\|Q_t-\bar Q_t^\pi\right\|^2dt\le C\left(1+E\|x_0\|^2\right)h,\end{aligned}\tag{16}$$

where $\bar X_t^\pi=\bar X_{t_i}^\pi,\bar P_t^\pi=\bar P_{t_i}^\pi,\bar Q_t^\pi=\bar Q_{t_i}^\pi$, for $t\in\left[t_i,t_{i+1}\right)$, $\left(X_t,P_t,Q_t\right)$ is the solution to (8), and C is a constant depending on $\mathscr{L}$ and T.

Remark 7. We briefly outline how to derive the above theorem using weak and monotonicity conditions and the results of [31]. First, the existence of the solution $\left(\bar{X}^{\pi}_{t_i}, \bar{P}^{\pi}_{t_i}, \bar{Q}^{\pi}_{t_i}\right)$ is proved by the convergence of the approximated decoupling field using a fixed point argument, see Theorem 5.1 (ii) in [31]. Then the error estimates (16) can obtained by Theorem 6.5 in the same literature, which essentially is built upon estimates for the approximated decoupling fields and the true counterpart.

Remark 8. We further remark that we only need the well-posedness and the estimates (16) for the implicit scheme here, and there exist different sets of conditions such that similar results hold since the weak and monotonicity conditions are sufficient only. For instance, [24] derive a similar result for an implicit and forward Euler discretization of a coupled MV-FBSDE, using a different set of monotonicity assumptions and the method of continuation.

Now, recall the classical Euler scheme (12b). For the discrete equation of $P^{\pi}_{t_{i+1}}$ in (12b), we take the conditional expectation, $E\left(\cdot \mid \mathcal{F}_{t_i}\right)$, at both sides and obtain the conditional expectation representation for $P^{\pi}_{t_i}$. For the same discrete equation of $P^{\pi}_{t_{i+1}}$, we multiply by $\left(\Delta W_i\right)^{\top}$ and take $E\left(\cdot \mid \mathcal{F}_{t_i}\right)$ afterwards. Therefore, we obtain a formulation, as follows

$$\begin{cases} X^{\pi}_0 = x_0, \\ X^{\pi}_{t_{i+1}} = X^{\pi}_{t_i} + b\left(t_i, X^{\pi}_{t_i}, P^{\pi}_{t_i}\right) h + \sigma\left(t_i, X^{\pi}_{t_i}, P^{\pi}_{t_i}\right) \Delta W_i, \\ Q^{\pi}_{t_i} = \dfrac{1}{h} E\left(P^{\pi}_{t_{i+1}} \Delta W_i^{\top} \mid \mathcal{F}_{t_i}\right), \\ P^{\pi}_{t_i} = E\left(P^{\pi}_{t_{i+1}} + F\left(t_i, X^{\pi}_{t_i}, P^{\pi}_{t_i}, Q^{\pi}_{t_i}\right) h \mid \mathcal{F}_{t_i}\right). \end{cases} \tag{17}$$

Remark 9. A key feature of this formulation is that it does not include the boundary condition for P^{π}_T, and therefore there are infinitely many solutions. In particular, it is easy to see that both the classic Euler scheme (12b) and the implicit scheme (15) are solutions to this formulation. Such a feature is particularly suitable for the analysis of the algorithm, as we set our loss function (12a) to measure the distance between P^{π}_T and $-\nabla_x g(X^{\pi}_T)$, and one may expect that the closer the two solutions of (17) are, the smaller loss we will have, and vice versa.

In what follows, we shall apply the same techniques in [16] to the current vector-valued BSDE setting in order to prove Lemma 1 and Theorem 3.

Lemma 1. *For $j = 1, 2$, suppose $\left\{(X^{\pi,j}_{t_i}, P^{\pi,j}_{t_i}, Q^{\pi,j}_{t_i})\right\}_{0\le i\le N-1}$ are two solution triples of (17), with $X^{\pi,j}_{t_i}, P^{\pi,j}_{t_i} \in L^2\left(\Omega, \mathcal{F}_{t_i}, \mathbb{P}\right), 0 \le i \le N$. For any $\lambda_1 > 0, \lambda_2 \ge L^F_q$, and sufficiently small h, denote*

$$\begin{aligned} A_1(h) &:= 2k^b + \lambda_1 + L^{\sigma}_x + L^b_x h, \\ A_2(h) &:= \left(\lambda_1^{-1} + h\right) L^b_p + L^{\sigma}_p, \\ A_3(h) &:= -\frac{\ln\left(1 - \left(2k^F + \lambda_2\right) h\right)}{h}, \\ A_4(h) &:= \frac{L^F_x}{\left(1 - \left(2k^F + \lambda_2\right) h\right) \lambda_2}. \end{aligned}$$

Then, we have, for $0 \le n \le N$,

$$E\left\|X^{\pi,1}_{t_n} - X^{\pi,2}_{t_n}\right\|^2 \le A_2 \sum_{i=0}^{n-1} e^{A_1(n-i-1)h} E\left\|P^{\pi,1}_{t_i} - P^{\pi,2}_{t_i}\right\|^2 h, \tag{18}$$

$$\begin{aligned} E\left\|P^{\pi,1}_{t_n} - P^{\pi,2}_{t_n}\right\|^2 \le{}& e^{A_3(N-n)h} E\left\|P^{\pi,1}_{t_N} - P^{\pi,2}_{t_N}\right\|^2 \\ &+ A_4 \sum_{i=n}^{N-1} e^{A_3(i-n)h} E\left\|X^{\pi,1}_{t_i} - X^{\pi,2}_{t_i}\right\|^2 h. \end{aligned} \tag{19}$$

Proof. We remark that even though A_1, A_2, A_3, A_4 are all functions of h, in order to ease the presentation, we do not make this dependence explicit. Let us define

$$\begin{aligned} \delta X_i &:= X^{\pi,1}_{t_i} - X^{\pi,2}_{t_i}, \\ \delta P_i &:= P^{\pi,1}_{t_i} - P^{\pi,2}_{t_i}, \\ \delta b_i &:= b\left(t_i, X^{\pi,1}_{t_i}, P^{\pi,1}_{t_i}\right) - b\left(t_i, X^{\pi,2}_{t_i}, P^{\pi,2}_{t_i}\right), \\ \delta \sigma_i &:= \sigma\left(t_i, X^{\pi,1}_{t_i}, P^{\pi,1}_{t_i}\right) - \sigma\left(t_i, X^{\pi,2}_{t_i}, P^{\pi,2}_{t_i}\right), \\ \delta F_i &:= F\left(t_i, X^{\pi,1}_{t_i}, P^{\pi,1}_{t_i}, Q^{\pi,1}_{t_i}\right) - F\left(t_i, X^{\pi,2}_{t_i}, P^{\pi,2}_{t_i}, Q^{\pi,2}_{t_i}\right). \end{aligned}$$

Then, we have

$$\delta X_{i+1} = \delta X_i + \delta b_i h + \delta \sigma_i \Delta W_i, \tag{20}$$

$$\delta P_i = E\left(\delta P_{i+1} + \delta F_i h \mid \mathcal{F}_{t_i}\right), \tag{21}$$

and motivated by (17), we define

$$\delta Q_i := \frac{1}{h} E\left(\delta P_{i+1} \Delta W_i^\top \mid \mathcal{F}_{t_i}\right). \tag{22}$$

The martingale representation theorem implies the existence of an $\mathcal{F}_t$-adapted square-integrable process $\{\delta Q_t\}_{t_i \le t \le t_{i+1}}$, such that

$$\delta P_{i+1} = E\left(\delta P_{i+1} \mid \mathcal{F}_{t_i}\right) + \int_{t_i}^{t_{i+1}} \delta Q_t \, dW_t,$$

which, together with (21), gives us

$$\delta P_{i+1} = \delta P_i - \delta F_i h + \int_{t_i}^{t_{i+1}} \delta Q_t \, dW_t. \tag{23}$$

From (20) and (23), noting that $\delta X_i, \delta P_i, \delta b_i, \delta \sigma_i$, and δF_i are all $\mathcal{F}_{t_i}$ measurable, and $E\left[\Delta W_i \mid \mathcal{F}_{t_i}\right] = 0$, $E\left(\int_{t_i}^{t_{i+1}} Q_t \, dW_t \mid \mathcal{F}_{t_i}\right) = 0$, we have

$$\begin{aligned} E\left\|\delta X_{i+1}\right\|^2 &= E\left\|\delta X_i + \delta b_i h\right\|^2 + hE\left\|\delta \sigma_i\right\|^2, \\ E\left\|\delta P_{i+1}\right\|^2 &= E\left\|\delta P_i - \delta F_i h\right\|^2 + \int_{t_i}^{t_{i+1}} E\left\|\delta Q_t\right\|^2 \, dt. \end{aligned}$$

We first establish the proclaimed upper bound for the forward diffusion part in (18). Using Assumption 4 and Remark 5, we can apply the root-mean-square and geometric mean inequality (RMS-GM) and derive, for any $\lambda_1 > 0$,

$$\begin{aligned} E\left\|\delta X_{i+1}\right\|^2 \;=\; & \left(1 + \left(2k^b + \lambda_1 + L_x^\sigma + L_x^b h\right) h\right) E\left\|\delta X_i\right\|^2 \\ & + \left(\left(\lambda_1^{-1} + h\right) L_p^b + L_p^\sigma\right) E\left\|\delta P_i\right\|^2 h, \end{aligned}$$

where we recall the definitions of A_1, A_2. Notice that $E\left\|\delta X_0\right\|^2 = 0$, and, thus, by induction, we have that, for $1 \le n \le N$,

$$E\left\|\delta X_n\right\|^2 \le A_2 \sum_{i=0}^{n-1} e^{A_1 (n-i-1)h} E\left\|\delta P_i\right\|^2 h,$$

where in above derivation we have used the inequality $(1 + x) \le e^x$, $\forall x \in \mathbb{R}$.

In order to show , we use a similar approach and apply the RMS-GM inequality for any $\lambda_2 > 0$, which yields

$$\begin{aligned} E\left\|\delta P_{i+1}\right\|^2 \ge\; & E\left\|\delta P_i\right\|^2 + \int_{t_i}^{t_{i+1}} E\left\|\delta Q_t\right\|^2 \, dt - 2k^F h E\left\|\delta P_i\right\|^2 \\ & - \left(\lambda_2 E\left\|\delta P_i\right\|^2 + \lambda_2^{-1}\left(L_x^F E\left\|\delta X_i\right\|^2 + L_q^F E\left\|\delta Q_i\right\|^2\right)\right) h. \end{aligned} \tag{24}$$

To deal with the integral term in the last inequality, we derive the following relation via Ito's isometry, (22) and (23),

$$\delta Q_i = \frac{1}{h} E\left(\int_{t_i}^{t_{i+1}} \delta Q_t \, dt \mid \mathcal{F}_{t_i}\right).$$

Thereafter, the Jensen and the Cauchy–Schwartz inequalities imply the following lower bound for the integral term, which extends the estimate in [16] to a matrix-valued setting,

$$\begin{aligned} E\left\|\delta Q_i\right\|^2 h &= \sum_{j=1}^{d} \sum_{k=1}^{m} \frac{1}{h} E\left(E\left(\int_{t_i}^{t_{i+1}} \left(\delta Q_t\right)_{j,k} dt \mid \mathcal{F}_{t_i}\right)\right)^2 \\ &\le \int_{t_i}^{t_{i+1}} E\left\|\delta Q_t\right\|^2 \, dt, \end{aligned} \tag{25}$$

where $(\cdot)_{j,k}$ denotes the (j,k)-entry of the matrix. Combining (24) with (25), subsequently gives

$$\begin{aligned} E\left\|\delta P_{i+1}\right\|^2 \ge\; & \left(1 - \left(2k^F + \lambda_2\right) h\right) E\left\|\delta P_i\right\|^2 \\ & + \left(1 - L_q^F \lambda_2^{-1}\right) E\left\|\delta Q_i\right\|^2 h - L_x^F \lambda_2^{-1} E\left\|\delta X_i\right\|^2 h. \end{aligned} \tag{26}$$

For any $\lambda_2 \ge L_q^F$ and sufficiently small h satisfying $\left(2k^F + \lambda_2\right) h < 1$, we get the following inequality

$$E\left\|\delta P_i\right\|^2 \le \left(1 - \left(2k^F + \lambda_2\right) h\right)^{-1} \left(E\left\|\delta P_{i+1}\right\|^2 + L_x^F \lambda_2^{-1} E\left\|\delta X_i\right\|^2 h\right).$$

Finally, recalling the definitions of A_3, A_4, by induction, we obtain that, for $0 \le n \le N-1$,

$$E\left\|\delta P_n\right\|^2 \le e^{A_3(N-n)h} E\left\|\delta P_N\right\|^2 + A_4 \sum_{i=n}^{N-1} e^{A_3(i-n)h} E\left\|\delta X_i\right\|^2 h. \quad \square$$

With the help of the a-priori result above, we can now state the main result of this section, which establishes the *a-posteriori estimate* for the convergence of the deep BSDE algorithm for the SMP formulation of the stochastic control problem under the aforementioned assumptions.

Theorem 3. *Suppose the conditions for Theorem 2 hold true, and there exist $\lambda_1 > 0, \lambda_2 \ge L_q^F, \lambda_3 > 0$ and constants in $\mathscr{L}$ such that $\overline{A_0} < 1$, where*

$$\begin{aligned}
\overline{A_1} &:= \lim_{h\to 0} A_1(h) = 2k^b + \lambda_1 + L_x^\sigma,\\
\overline{A_2} &:= \lim_{h\to 0} A_2(h) = L_p^b \lambda_1^{-1} + L_p^\sigma,\\
\overline{A_3} &:= \lim_{h\to 0} A_3(h) = 2k^F + \lambda_2,\\
\overline{A_4} &:= \lim_{h\to 0} A_4(h) = L_x^F \lambda_2^{-1},\\
\overline{A_0} &:= \overline{A_2}\frac{1 - e^{-\left(\overline{A_1}+\overline{A_3}\right)T}}{\overline{A_1}+\overline{A_3}}\left\{L_x^{g_x} e^{\left(\overline{A_1}+\overline{A_3}\right)T} + \overline{A_4}\frac{e^{\left(\overline{A_1}+\overline{A_3}\right)T}-1}{\overline{A_1}+\overline{A_3}}\right\}.
\end{aligned} \tag{27}$$

Then, there exists a constant $C > 0$, depending on $E\|x_0\|^2, \mathscr{L}, T, \lambda_1, \lambda_2$ and λ_3 such that, for sufficiently small h,

$$\begin{aligned}
&\sup_{t\in[0,T]}\left(E\left\|X_t - \hat{X}_t^\pi\right\|^2 + E\left\|P_t - \hat{P}_t^\pi\right\|^2\right)\\
&\qquad + \int_0^T E\left\|Q_t - \hat{Q}_t^\pi\right\|^2 dt \le C\left(h + E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2\right),
\end{aligned} \tag{28}$$

where $\hat{X}_t^\pi = X_{t_i}^\pi, \hat{P}_t^\pi = P_{t_i}^\pi, \hat{Q}_t^\pi = Q_{t_i}^\pi$ for $t \in [t_i, t_{i+1})$, and (X_t, P_t, Q_t) is the solution to (8).

Proof. Let $X_{t_i}^{\pi,1} = X_{t_i}^\pi, P_{t_i}^{\pi,1} = P_{t_i}^\pi, Q_{t_i}^{\pi,1} = Q_{t_i}^\pi$, given by the Euler scheme (12b), and $X_{t_i}^{\pi,2} = \bar{X}_{t_i}^\pi, P_{t_i}^{\pi,2} = \bar{P}_{t_i}^\pi, Q_{t_i}^{\pi,2} = \bar{Q}_{t_i}^\pi$, given by the implicit scheme (15). Since both of these schemes solve (17), we can apply Lemma 1. In what follows, we adopt the notation of Lemma 1.

Through the same reasoning in [16], we first apply the RMS-GM inequality for any $\lambda_3 > 0$,

$$\begin{aligned}
E\left\|\delta P_N\right\|^2 = E\left\|-\nabla_x g\left(\bar{X}_T^\pi\right) - P_T^\pi\right\|^2 \le \left(1+\lambda_3^{-1}\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2\\
+ L_x^{g_x}\left(1+\lambda_3\right) E\left\|\delta X_N\right\|^2,
\end{aligned} \tag{29}$$

where we also used that $\nabla_x g$ is Lipschitz continuous.

Let

$$\mathcal{X} := \max_{0\le n\le N} e^{-A_1 nh} E\left\|\delta X_n\right\|^2, \quad \mathcal{P} := \max_{0\le n\le N} e^{A_3 nh} E\left\|\delta P_n\right\|^2. \tag{30}$$

From the upper bound in (18), it follows that

$$e^{-A_1 nh} E\left\|\delta X_n\right\|^2 \le A_2 \sum_{i=0}^{n-1} e^{-A_1(i+1)h} E\left\|\delta P_i\right\|^2 h \le A_2 \mathcal{P} \sum_{i=0}^{n-1} e^{-A_1(i+1)h - A_3 ih} h. \tag{31}$$

Similarly, imply the following estimate, extending the scalar case in [16] to the vector-valued setting,

$$\begin{aligned}
e^{A_3 nh} E\left\|\delta P_n\right\|^2 &\le e^{A_3 T} E\left\|\delta P_N\right\|^2 + A_4 \sum_{i=n}^{N-1} e^{A_3 ih} E\left\|\delta X_i\right\|^2 h\\
&\le e^{A_3 T}\left(1+\lambda_3^{-1}\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2\\
&\quad + \left(L_x^{g_x}\left(1+\lambda_3\right) e^{(A_1+A_3)T} + A_4 \sum_{i=n}^{N-1} e^{(A_1+A_3)ih} h\right)\mathcal{X},
\end{aligned} \tag{32}$$

where we used (29) to obtain the second inequality. Combining (32), (31) with (30) thus yields

$$\mathcal{X} \le A_2 h e^{-A_1 h} \frac{e^{-(A_1+A_3)T} - 1}{e^{-(A_1+A_3)h} - 1} \mathcal{P}, \tag{33}$$

$$\begin{aligned}
\mathcal{P} &\le e^{A_3 T}\left(1+\lambda_3^{-1}\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2\\
&\quad + \left(L_x^{g_x}\left(1+\lambda_3\right) e^{(A_1+A_3)T} + A_4 h \frac{e^{(A_1+A_3)T}-1}{e^{(A_1+A_3)h}-1}\right)\mathcal{X}.
\end{aligned} \tag{34}$$

Now, we define

$$A(h) := A_2 h e^{-A_1 h} \frac{e^{-(A_1+A_3)T}-1}{e^{-(A_1+A_3)h}-1}\left(L_x^{g_x}\left(1+\lambda_3\right)e^{(A_1+A_3)T} + A_4 h \frac{e^{(A_1+A_3)T}-1}{e^{(A_1+A_3)h}-1}\right), \tag{35}$$

by which (35) can be rewritten, as follows

$$\mathcal{P} \le e^{A_3 T}\left(1+\lambda_3^{-1}\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2 + A(h)\mathcal{P}.$$

Hence, whenever $A(h) < 1$, the estimates in (33) and (35) take the following form

$$\mathcal{X} \le \frac{e^{A_3 T}\left(1+\lambda_3^{-1}\right) A_2 h e^{-A_1 h}\frac{e^{-(A_1+A_3)T}-1}{e^{-(A_1+A_3)h}-1} E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2}{1-A(h)},$$

$$\mathcal{P} \le \frac{e^{A_3 T}\left(1+\lambda_3^{-1}\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2}{1-A(h)}.$$

Recall $\lim_{h\to 0} A_i(h) = \overline{A_i}$. From (35), it then follows that

$$\lim_{h\to 0} A(h) = \overline{A_2}\frac{1-e^{-\left(\overline{A_1}+\overline{A_3}\right)T}}{\overline{A_1}+\overline{A_3}}\left(L_x^{g_x}\left(1+\lambda_3\right)e^{\left(\overline{A_1}+\overline{A_3}\right)T} + \overline{A_4}\frac{e^{\left(\overline{A_1}+\overline{A_3}\right)T}-1}{\overline{A_1}+\overline{A_3}}\right). \tag{36}$$

Recall the definition of $\overline{A_0}$ in (27). Whenever $\overline{A_0} < 1$, comparing (36) with $\overline{A_0}$, there exists a sufficiently small $\lambda_3 > 0$ such that $\lim_{h\to 0} A(h) < 1$ and therefore $A(h) < 1$ holds for sufficiently small h. Then we can write, for sufficiently small h and any $\epsilon > 0$,

$$\mathcal{X} \le (1+\epsilon)\frac{\overline{A_2}e^{\overline{A_3}T}}{1-\lim_{h\to 0} A(h)}\left(1+\lambda_3^{-1}\right)\frac{1-e^{-\left(\overline{A_1}+\overline{A_3}\right)T}}{\overline{A_1}+\overline{A_3}} E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2,$$

$$\mathcal{P} \le (1+\epsilon)\frac{e^{\overline{A_3}T}}{1-\lim_{h\to 0} A(h)}\left(1+\lambda_3^{-1}\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2.$$

Note that we have a slight difference in these two estimates compared to [16], where we keep λ_3 on the right-hand side as it has a significant impact on the term $(1+\lambda_3^{-1})$ and should be chosen appropriately. Consequently, by fixing ϵ and choosing the corresponding small $h > 0$, we obtain the following error estimates

$$\max_{0\le n\le N} E\left\|\delta X_n\right\|^2 \le e^{A_1 T\vee 0}\mathcal{X} \le C\left(\lambda_1,\lambda_2,\lambda_3\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2, \tag{37}$$

$$\max_{0\le n\le N} E\left\|\delta P_n\right\|^2 \le e^{(-A_3 T)\vee 0}\mathcal{P} \le C\left(\lambda_1,\lambda_2,\lambda_3\right) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2. \tag{38}$$

Finally, in order to estimate $E\left\|\delta Q_n\right\|^2$ for $0 \le n \le N-1$, we consider estimate (26) from the proof of Lemma 1, in which λ_2 can take any value such that $\lambda_2 \ge L_q^F$. In particular, when $L_q^F \ne 0$, we choose $\lambda_2 = 2L_q^F$ and obtain

$$\frac{1}{2}E\left\|\delta Q_i\right\|^2 h \le \frac{L_x^F}{2L_q^F} E\left\|\delta X_i\right\|^2 h + E\left\|\delta P_{i+1}\right\|^2 - \left(1-\left(2k^F + 2L_q^F\right)h\right) E\left\|\delta P_i\right\|^2.$$

Summing from 0 to $N-1$ therefore gives

$$\sum_{i=0}^{N-1} E\left\|\delta Q_i\right\|^2 h \le \frac{L_x^F T}{L_q^F}\max_{0\le n\le N} E\left\|\delta X_n\right\|^2 + \left(4\left(k^F + L_q^F\right)T \vee 0 + 2\right)\max_{0\le n\le N} E\left\|\delta P_n\right\|^2.$$

Using the estimates established by (37) and (38), we collect

$$\sum_{i=0}^{N-1} E\left\|\delta Q_i\right\|^2 h \le C(\lambda_1,\lambda_2,\lambda_3) E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2. \tag{39}$$

The case $L_q^F = 0$ can be dealt with similarly by choosing $\lambda_2 = 1$ and the same type of estimate can be derived. Finally, combining estimates (37), (38) and (39) with Theorem 2, we prove our statement. □

Corollary 1. *Let (X_t, u_t, P_t, Q_t) be the optimal 4-tuple that solves problem (1). Under the setting of Theorem 3, there exists a constant, $C > 0$, depending on $E\|x_0\|^2, \mathscr{L}, T, \lambda_1$, and λ_2, such that for sufficiently small h,*

$$\int_0^T E\|u_t^* - \hat{u}_t^\pi\|^2\, dt \le C\left(h + E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2\right), \tag{40}$$

where $\hat{u}_{t_i}^\pi := \mathcal{M}(t, \hat{X}_{t_i}^\pi, \hat{P}_{t_i}^\pi, \hat{Q}_{t_i}^\pi)$, and $\hat{u}_t^\pi = \hat{u}_{t_i}^\pi$ for $t \in [t_i, t_{i+1})$.

Proof. Recall the optimal feedback map obtained by using the SMP,

$$u_t = \mathcal{M}(t, X_t, P_t, Q_t), \quad \forall t \in [0, T],$$

where (X_t, u_t, P_t, Q_t) is the optimal 4-tuple that solves stochastic control problem (1) and the FBSDE (8), see Remark 4. Then, by Theorem 3 and the Lipschitz continuity of $\mathcal{M}$ in Assumption 3, we immediately have

$$\begin{aligned}\int_0^T E\left\|u_t - \hat{u}_t^\pi\right\|^2 dt &= \int_0^T E\left\|\mathcal{M}(t, X_t, P_t, Q_t) - \mathcal{M}(t, \hat{X}_{t_i}^\pi, \hat{P}_{t_i}^\pi, \hat{Q}_{t_i}^\pi)\right\|^2 dt \\ &\le L\int_0^T \Big(E\left\|X_t - \hat{X}_t^\pi\right\|^2 + E\left\|P_t - \hat{P}_t^\pi\right\|^2 \\ &\qquad + E\left\|Q_t - \hat{Q}_t^\pi\right\|^2 \Big)\, dt \\ &\le C\big(h + E\left\|-\nabla_x g\left(X_T^\pi\right) - P_T^\pi\right\|^2\big).\end{aligned}$$

□

5. Numerical results

In this section, we demonstrate the accuracy and robustness of the proposed scheme. We consider stochastic control problems, both in the case of drift and diffusion control. Algorithm 1 has been implemented in TensorFlow 2.15, and the experiments were run on a Dell Alienware Aurora R10 machine, equipped with an AMD Ryzen 9 3950X CPU (16 cores, 64 Mb cache, 4.7 GHz) and an Nvidia GeForce RTX 3090 GPU (24 Gb). The library used in this paper will be publicly accessible under the github repository of the second author. In all numerical experiments presented below, we use standard fully-connected feedforward neural networks to parametrize $\mu_0^\pi : \mathbb{R}^d \to \mathbb{R}^d$ and each $\phi_i^\pi : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, $i = 0, \dots, N-1$. Each neural network has two hidden layers of width 100 and a hyperbolic tangent activation function. All parameters are initialized according to default TensorFlow settings. We employ the Adam optimizer with default settings to minimize the empirical loss function and use a batch size of $2^{12} = 4096$ independent paths of the Brownian motion for each SGD step during training. We apply an adaptive strategy for training the networks. In particular, for $N = 2, 5, 10, 20, 50, 100$, we apply an exponential learning rate schedule starting from an initial learning rate η_0^N, and a decay rate is set to $10^{-6}/\eta_0^N$. An adaptive number of iteration steps I^N also applies to each N.[1] The Monte Carlo errors are computed over an independent validation sample of $2^{14} = 16384$ paths. In order to assess the inherent randomness of the deep BSDE methods and the SGD iterations, we run each experiment 5 times and report on the mean and standard deviation of the resulting independent approximations of Algorithm 1. Computations were carried out with single floating-point precision. We denote the total approximation errors by $\delta\hat{X}_n^\pi$ at time t_n and similarly for other processes. In the comparison with [22], we use $\lambda = 1$ as described in their paper, which is also explained in our Section 3. We denote by Y_n^π the discrete time counterpart of the objective functional in (1), which we compute similarly to [22, eq.(3.1)] via a backward summation, using the relations in Remark 6.

Our numerical examples are given for so-called linear quadratic (LQ) type problems, a special case of (1), where

$$\begin{aligned}&\bar{b}(t, x, u) = Ax + Bu + \beta,\ \bar{\sigma}_j(t, x, u) = C_j x + D_j u + \Sigma_j, \\ &\bar{f}(t, x, u) = \frac{1}{2}(x^\top R_x x + u^\top R_{xu} x + u^\top R_u u),\ g(x) = \frac{1}{2}x^\top G x,\end{aligned} \tag{41}$$

with $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times \ell}$, $\beta \in \mathbb{R}^d$, $R_x \in \mathbb{R}^{d\times d}$, $R_{xu} \in \mathbb{R}^{l\times d}$, $R_u \in \mathbb{R}^{\ell\times\ell}$ and $G \in \mathbb{R}^{d\times d}$, R_x, R_u and G are symmetric, and, for $j = 1, \dots, m$, $C_j \in \mathbb{R}^{d\times d}$, $D_j \in \mathbb{R}^{d\times\ell}$, $\Sigma_j \in \mathbb{R}^d$. For $\bar{\sigma}, \Sigma, q \in \mathbb{R}^{d\times m}$, we denote the jth column by $\bar{\sigma}_j, \Sigma_j, q_j$. Utilizing the SMP, we obtain the feedback map for the optimal control

$$u = \mathcal{M}(t, x, p, q) = -R_u^{-1}\big(R_{xu}x - B^\top p - \sum_{j=1}^m D_j^\top q_j\big), \tag{42}$$

and therein we obtain the corresponding FBSDE (8) for the algorithm, by substituting (42) back into the forward diffusion and adjoint equation.

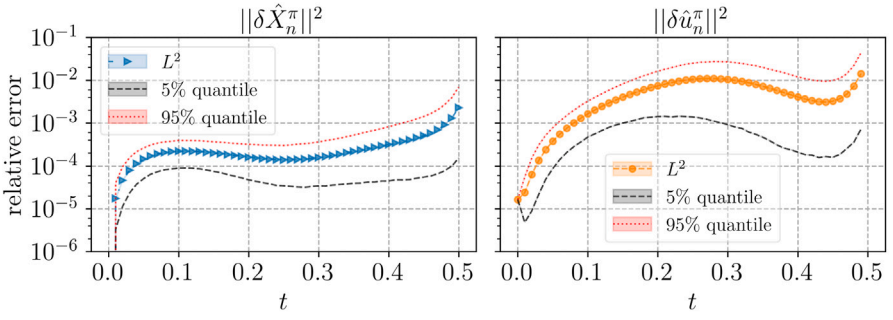
A reference solution to the LQ problem can be derived semi-analytically through numerically solving a system of ODEs, induced by the HJB equation. To obtain this reference solution for a multi-dimensional setting, we extend a well-known result in the literature [30] for the case $m = 1$, and the detailed derivation is given in the appendix. In what follows, we compare our approximations to a reference solution obtained through the numerical integration of the Riccati ODEs (44) in the appendix, using an equidistant time grid with $N_{\text{ode}} = 10^5$, and simulating the coupled FBSDE system in (8) on the same time grid with a discrete Euler–Maruyama approximation.

5.1. Example 1 — drift control

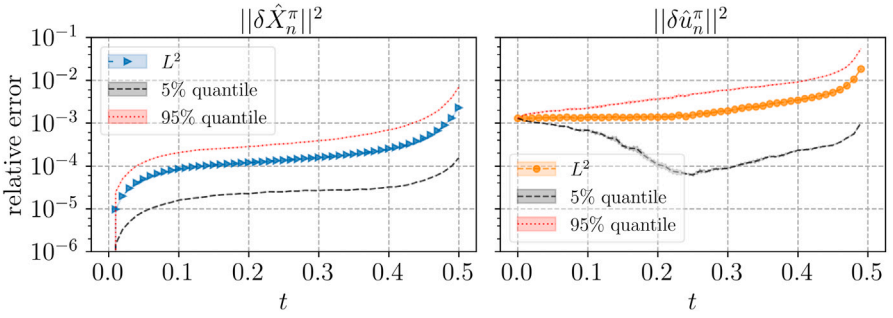
The following stochastic control problem can be found in [22, sec. 5.1.2]. The control space is $\ell = 2$ dimensional, whereas for the state space $d = m = 6$. The coefficients in (1) are of LQ type and the corresponding matrices in (41) are defined, as follows,

$$A = -\text{diag}([1, 2, 3, 1, 2, 3]), \tag{43}$$

[1] We set the initial learning rates as 5e−4, 5e−4, 1e−3, 2e−3, 4e−3, 8e−3, 8e−3, and the number of iterations as 2^{12}, 2^{12}, 2^{13}, 2^{14}, 2^{15}, 2^{16}, 2^{17}, for $N = 2, 5, 10, 20, 50, 100$, respectively.



(a) Algorithm 1.



(b) Andersson et al. (2023) in [22].

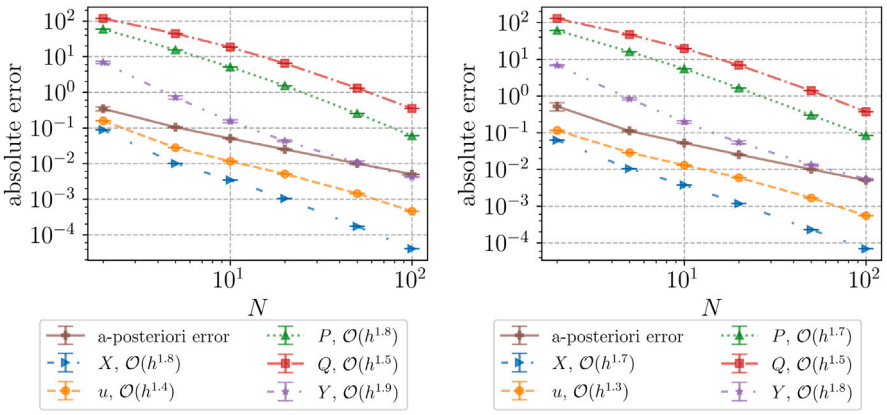
Fig. 1. Example 1, $N = 50$. Errors of approximations of the optimally controlled state space and control strategy. On top: results obtained through Algorithm 1 and the SMP. On the bottom: reference method from [22] using dynamic programming. Lines correspond to the mean of 5 independent runs of the algorithm, shaded areas to the standard deviation. Graphs computed over an independent Monte Carlo sample of size $M = 2^{14}$.

$$B = \begin{pmatrix} 1, & 1, & 0.5, & 1, & 0, & 0 \\ -1, & 1, & 1, & -1, & -1, & 1 \end{pmatrix}^\top,$$
$$\beta = -A([-0.2, -0.1, 0, 0, 0.1, 0.2])^\top, \quad C_j = 0, \quad D_j = 0,$$
$$\Sigma = \mathrm{diag}([0.05, 0.25, 0.05, 0.25, 0.05, 0.25]), \quad R_{xu} = 0, \quad R_u = 2I_l,$$
$$R_x = 2\mathrm{diag}([25, 1, 25, 1, 25, 1]), \quad G = 2\mathrm{diag}([1, 25, 1, 25, 1, 25]),$$

for all $j = 1, \dots, m$. We consider a deterministic initial condition, $x_0 = (0.1, \dots, 0.1)$, and a terminal time of $T = 1/2$. We remark that this problem falls under the class of drift control problems, and, recalling the discussion following Assumption 2, it in particular satisfies the conditions of Theorem 3 and Corollary 1 over any compact subset of the state space.

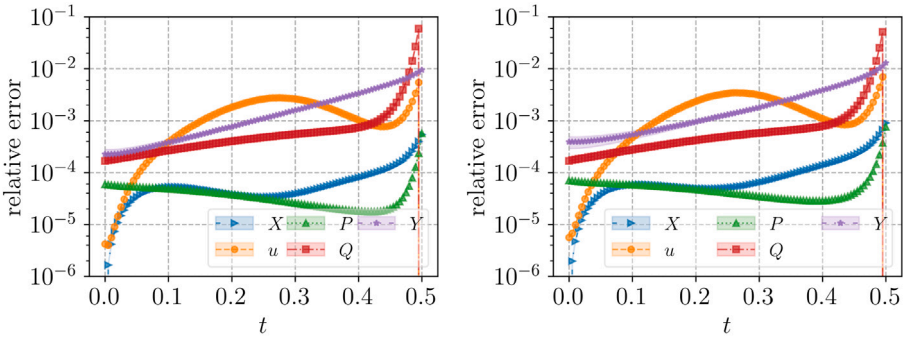
In Fig. 1, the relative approximation errors of the optimal control strategy and the optimally controlled state space are compared to the method proposed in [22], explained in Section 3. We remark that, as found in the aforementioned paper, the deep BSDE method on the dynamic programming equation fails to converge to the true solution. As it can be seen our deep BSDE formulation, via the SMP in Algorithm 1, leads to accurate and robust approximations of both the controlled diffusion and the optimal control. The relative approximation errors of both processes are $\mathcal{O}(10^{-3})$ for over 95% of the sampling paths. Even though, standard to the FBSDE literature, our theory only establishes error bounds in the L^2 sense, see (28) and (40), the tight quantile estimates in Fig. 1 suggest that convergence is achieved in a stronger sense as well. Increasing the number of discretization points N further reduces the errors.

The convergence of the proposed algorithm is collected in Fig. 2(a), where the absolute errors of all corresponding processes are plotted for different values of N. We recall that the insights of Theorem 3 suggest that these errors admit to an upper bound depending on the *a-posteriori estimate* which is defined as the sum of time step size $h = T/N$ and the value of the objective functional $E\| - \nabla_x g(\hat{X}_N^\pi) - \hat{P}_N^\pi\|^2$, see (28). Indeed, Fig. 2(a) confirms our theoretical findings, where the approximation errors of all processes decay with a rate of at least $\mathcal{O}(h)$. Fig. 4(a) depicts the convergence of the components of the *a-posteriori estimate*. As can be seen from the right end of the curves, once N is large, the optimization problem in Algorithm 1 gets more complex, and in the case of $N = 100$ our optimization strategy only manages to yield a slight decrease for terminal loss, leading to a slower convergence. Note that for a given N whenever the loss functional corresponding to (12a) is orders of magnitude smaller than the time step size, the complete *a-posteriori estimate* is dominated by the discretization component. Therefore, we still obtain an empirical convergence rate of the *a-posteriori estimate* of $\mathcal{O}(h^{1.07})$. In order to preserve convergence for very fine time grids, one needs to make sure the terminal loss is sufficiently small. A good practical guideline, as implied by Theorem 3 and the discussion above, is to ensure that the loss functional's value is significantly smaller than h.



(a) Example 1. (b) Example 2.

Fig. 2. Convergence and empirical convergence rates of Algorithm 1 over N. Lines correspond to the mean of 5 independent runs of the algorithm, error bars to the standard deviation. Graphs compute over an independent Monte Carlo sample of size $M = 2^{14}$.



(a) Example 1. (b) Example 2.

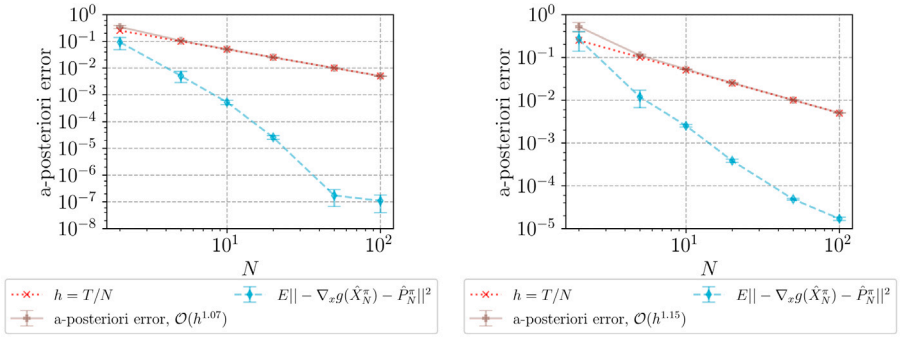
Fig. 3. Relative L^2 approximation errors over time, $N = 100$. Lines correspond to the mean of 5 independent runs of the algorithm, shaded areas to the standard deviation. Graphs computed over an independent Monte Carlo sample of size $M = 2^{14}$.

The relative L^2 approximation errors over time are depicted in Fig. 3(a), for $N = 100$. The method yields highly accurate approximations for all time steps, and of a similar magnitude for all processes. It is worth to mention the accuracy at $t = 0$, where no discretization error from the reference solution is present. All error measures are collected in Table 1. In line with the figures, we see that convergence is achieved in the natural norms of all processes. The resulting approximations are robust regardless of the underlying randomness of the Monte Carlo machinery, standard deviations of the errors are orders of magnitude smaller than their corresponding means. As shown in the last column, Algorithm 1 also offers a competitive runtime for high-dimensional problems, as an SGD iteration step takes approximately 0.102 s on average when $N = 100$. In our experiments, this beats the runtime of [22] with a factor of 5 due to the lack of the extra backward summation step considered therein.

5.2. Example 2 — drift and diffusion control

In the next section, we provide an extension to the previous problem by introducing diffusion control to the LQ problem in Section 5.1. The coefficients read,

$$C_j = \frac{1}{60}\mathrm{diag}([1,2,3,1,2,3]), \quad D_j = \frac{1}{60}\begin{pmatrix} 1, & 0, & 1, & 0, & 1, & 0 \\ 0, & -1, & 0, & -1, & 0, & -1 \end{pmatrix}^T,$$



(a) Example 1. (b) Example 2.

Fig. 4. Convergence of the a-posteriori error estimate defined in (28). Lines correspond to the mean of 5 independent runs of the algorithm, error bars to the standard deviation. Graphs compute over an independent Monte Carlo sample of size $M = 2^{14}$.

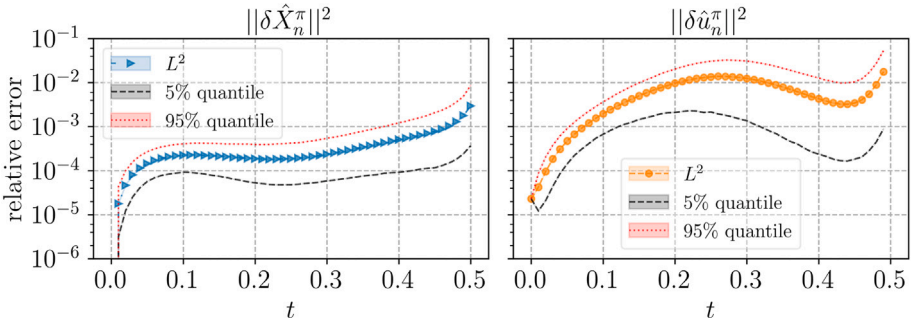


Fig. 5. Example 2, $N = 50$. Errors of the approximations of the optimally controlled state space and control strategy through Algorithm 1 and the SMP. Lines correspond to the mean of 5 independent runs of the algorithm, shaded areas to the standard deviation. Graphs computed over an independent Monte Carlo sample of size $M = 2^{14}$.

Table 1
Example 1. Error measures for different values of N. Figures correspond to the mean (standard deviation) of 5 independent runs of the algorithm. Expectations computed over an independent Monte Carlo sample of size $M = 2^{14}$.

	max. $E\|\|\delta\hat{X}_n^\pi\|\|^2$	max. $E\|\|\delta\hat{P}_n^\pi\|\|^2$	avg. $E\|\|\delta\hat{Q}_n^\pi\|\|^2$	avg. $E\|\|\delta\hat{u}_n^\pi\|\|^2$	$E\|\|-\nabla_X g(\hat{X}_N^\pi)-\hat{P}_N^\pi\|\|^2$	$\|\delta\hat{Y}_0^\pi\|$	iter. time (s)
2	8.869e−2(4e−5)	5.919e+1(5e−2)	1.1746e+2(8e−2)	1.5929e−1(2e−5)	9e−2(4e−2)	9.1(6e−1)	4.8e−3(3e−4)
5	9.9527e−3(6e−7)	1.538 42e+1(6e−4)	4.4459e+1(2e−3)	2.7842e−2(4e−6)	5e−3(2e−3)	7.0e−1(8e−2)	7.1e−3(1e−4)
10	3.442 41e−3(9e−8)	5.1267(2e−4)	1.835 73e+1(3e−4)	1.1609e−2(2e−6)	5.3e−4(1e−4)	1.2e−1(1e−2)	1.18e−2(2e−4)
20	1.045 21e−3(2e−8)	1.528 50(3e−5)	6.499 80(9e−5)	5.092e−3(2e−6)	2.6e−5(4e−6)	2.5e−2(3e−3)	2.5e−2(2e−3)
50	1.725 12e−4(1e−9)	2.525 43e−1(6e−6)	1.312 91(2e−5)	1.4437e−3(3e−7)	2e−7(1e−7)	3e−3(1e−3)	5.72e−2(4e−4)
100	4.064 71e−5(4e−10)	6.100 16e−2(5e−7)	3.495 17e−1(7e−6)	4.5906e−4(8e−8)	1.1e−7(7e−8)	5e−4(1e−4)	1.019e−1(2e−4)

for all $j = 1, \dots, m$, and the rest as in (43). Note that due to the presence of diffusion control, this equation cannot be treated via standard DP, such as [22], therefore no reference method is provided.

In Fig. 5, the relative approximation errors of the controlled state process and the optimal control strategy are depicted. Even with the introduction of diffusion control, we obtain accurate approximations up to $\mathcal{O}(10^{-2})$ relative accuracy in both processes. The tight quantile bounds suggest that our approximations remain accurate even in stronger senses than L^2.

As can easily be seen from (42), Q couples back into the forward diffusion and hence the conditions of Assumption 2 are not satisfied. Therefore, this example does not fall under the setting of Theorem 3. Nevertheless, the convergence drawn in Fig. 2(b), together with a related theoretical result Corollary 4.7 in [24], gives hope that a similar theoretical bound using the weak and monotonicity conditions may also be established in the diffusion control case. Fig. 4(b) collects the convergence of the *a-posteriori estimate*. Similar to the drift control case we find that the *a-posteriori estimate* is dominated by the discretization component over the considered time steps N, and therefore *a-posteriori estimate* converges with a rate of $\mathcal{O}(h^{1.15})$. These results are comparable to Example 1.

The relative L^2 approximation errors are depicted in Fig. 3(b) for $N = 100$. As we see, even in the case of diffusion control, we preserve a similar relative approximation error of $\mathcal{O}(10^{-3})$, as in Example 1. The slight increase in the relative error of u is due to the feedback map (42), also accumulating the errors in Q. Finally, all error measures are collected in Table 2, which shows that all processes converge as expected. Moreover, the lack of difference in total runtime in Table 2, compared to the drift control case in 1, highlights the great potential of such deep BSDE formulations in the framework of high-dimensional diffusion control problems.

Table 2
Example 2. Error measures for different values of N. Figures correspond to the mean (standard deviation) of 5 independent runs of the algorithm. Expectations computed over an independent Monte Carlo sample of size $M = 2^{14}$.

	max. $E\|\|\delta\hat{X}_n^\pi\|\|^2$	max. $E\|\|\delta\hat{P}_n^\pi\|\|^2$	avg. $E\|\|\delta\hat{Q}_n^\pi\|\|^2$	avg. $E\|\|\delta\hat{u}_n^\pi\|\|^2$	$E\|\| - \nabla_x g(\hat{X}_N^\pi) - \hat{P}_N^\pi\|\|^2$	$\|\delta\hat{Y}_0^\pi\|$	iter. time (s)
2	6.242e−2(5e−5)	6.065e+1(7e−2)	1.277e+2(1e−1)	1.1582e−1(8e−5)	3e−1(1e−1)	8.3(3e−1)	4.7e−3(4e−4)
5	1.048 19e−2(2e−7)	1.5949e+1(3e−3)	4.666e+1(1e−2)	2.880e−2(3e−5)	1.2e−2(5e−3)	7.8e−1(6e−2)	7.1e−3(1e−4)
10	3.7835e−3(1e−7)	5.4819(5e−4)	1.945 61e+1(1e−3)	1.3102e−2(2e−6)	2.5e−3(2e−4)	1.5e−1(2e−2)	1.16e−2(1e−4)
20	1.182 83e−3(7e−8)	1.6518(1e−4)	6.8986(4e−4)	5.9651e−3(7e−7)	3.8e−4(3e−5)	3.2e−2(6e−3)	2.5e−2(2e−3)
50	2.2904e−4(2e−8)	2.9989e−1(2e−5)	1.390 10(9e−5)	1.6851e−3(2e−7)	4.9e−5(2e−6)	4.1e−3(7e−4)	5.75e−2(5e−4)
100	6.9501e−5(4e−9)	8.4388e−2(9e−6)	3.7029e−1(5e−5)	5.5453e−4(9e−8)	1.7e−5(1e−6)	9e−4(2e−4)	1.023e−1(3e−4)

CRediT authorship contribution statement

Zhipeng Huang: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Balint Negyesi:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Cornelis W. Oosterlee:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Acknowledgments

The first author would like to thank the China Scholarship Council (CSC) for its financial support. The second author acknowledges financial support from the Peter Paul Peterich Foundation via the TU Delft University Fund. The authors would also like to thank the anonymous referees for their valuable comments and suggestions for improving the paper.

Appendix

In this appendix, we follow the same approach used for $m = 1$ in [30], and derive the reference solution for general dimensions d, ℓ and m. By the DP, the value function $V(t,x)$ for a LQ problem should satisfy the HJB Eq. (9) with a boundary condition $V(T,x) = \frac{1}{2}x^\top Gx$, and, in the LQ case, the generalized Hamiltonian $\mathcal{G}$ is given by

$$\begin{aligned}\mathcal{G}\left(t,x,u,-V_x,-V_{xx}\right) = &-\frac{1}{2}u^\top R_u u - u^\top R_{xu}x - \frac{1}{2}x^\top R_x x - (Ax+Bu+\beta)^\top V_x \\ &- \sum_{j=1}^m \frac{1}{2}(C_j x + D_j u + \Sigma_j)^\top V_{xx}(C_j x + D_j u + \Sigma_j).\end{aligned}$$

We conjecture that the value function takes the form $V(t,x) = \frac{1}{2}x^\top \Gamma(t)x + x^\top\gamma(t) + \kappa(t)$, for some unknowns symmetric $\Gamma(t) \in \mathbb{R}^{d\times d}$, $\gamma(t) \in \mathbb{R}^d$ and $\kappa(t) \in \mathbb{R}$, and for simplicity we omit the argument t in the following. Then

$$\begin{aligned}\mathcal{G}(t,x,u,-V_x,-V_{xx}) = &-\frac{1}{2}(u+\Psi x+\psi)^\top \hat{R}(u+\Psi x+\psi) + \frac{1}{2}x^\top \hat{S}^\top \hat{R}^{-1}\hat{S}x \\ &+ \frac{1}{2}x^\top\left(-R_x - \sum_{j=1}^m C_j^\top \Gamma C_j - \Gamma^\top A - A^\top\Gamma\right)x \\ &- x^\top\left(A^\top\gamma + \sum_{j=1}^m C_j^\top\Gamma\Sigma_j - \Psi^\top\hat{R}\psi + \Gamma\beta\right) \\ &+ \frac{1}{2}\psi^\top\hat{R}\psi - \beta^\top\gamma - \sum_{j=1}^m \frac{1}{2}\Sigma_j^\top\Gamma\Sigma_j,\end{aligned}$$

where we define the following

$$\hat{R} := R_u + \sum_{j=1}^m D_j^\top\Gamma D_j, \quad \hat{S} := B^\top\Gamma + R_{xu} + \sum_{j=1}^m D_j^\top\Gamma C_j,$$
$$\Psi := \hat{R}^{-1}\hat{S}, \quad \psi := \hat{R}^{-1}\left(B^\top\gamma + \sum_{j=1}^m D_j^\top\Gamma\Sigma_j\right),$$

provided that $\hat{R}$ is positive definite. Due to the first quadratic term, we immediately see that the optimal control should take the following feedback form,

$$\begin{aligned}u^* = -\Big(R_u + \sum_{j=1}^m D_j^\top V_{xx}D_j\Big)^{-1}\Big(&B^\top V_x + \big(R_{xu} + \sum_{j=1}^m D_j^\top V_{xx}C_j\big)x \\ &+ \sum_{j=1}^m D_j^\top V_{xx}\Sigma_j\Big).\end{aligned} \tag{44}$$

Substituting $\mathcal{G}(t,x,u^*,-V_x,-V_{xx})$ back to (9), we find that $V(t,x)=\frac{1}{2}x^\top\Gamma(t)x+x^\top\gamma(t)+\kappa(t)$ solves the HJB equation whenever the following relations hold, at each $t\in[0,T]$

$$\begin{cases} 0=\dot\Gamma+\Gamma A+A^\top\Gamma+\sum_{j=1}^m C_j^\top\Gamma C_j+R_x-\widehat S^\top\widehat R^{-1}\widehat S,\\ 0=\dot\gamma+A^\top\gamma+\sum_{j=1}^m C_j^\top\Gamma\Sigma_j-\Psi^\top\widehat R\psi+\Gamma\beta,\\ 0=\dot\kappa-\frac{1}{2}\psi^\top\widehat R\psi+\beta^\top\gamma+\sum_{j=1}^m\frac{1}{2}\Sigma_j^\top\Gamma\Sigma_j,\\ \Gamma(T)=G,\quad \gamma(T)=0,\quad \kappa(T)=0. \end{cases} \tag{45}$$

This system of ODEs can be integrated numerically with practically arbitrary accuracy. With the thereby obtained numerical solution of (45), one can subsequently compute the value function and all of its derivatives using the conjecture, and obtain the optimal control u_t^* through the feedback map (44). The reference solution to the corresponding BSDE (8), derived via the SMP, can similarly be computed using the relations (14) in Remark 6.

References

[1] Richard Bellman, Dynamic programming and stochastic control processes, Inf. Control 1 (3) (1958) 228–239.
[2] Jean-Michel Bismut, An introductory approach to duality in optimal stochastic control, SIAM Rev. 20 (1) (1978) 62–78.
[3] Lev Semenovich Pontryagin, Mathematical Theory of Optimal Processes, Routledge, 2018.
[4] Harold J. Kushner, Numerical methods for stochastic control problems in continuous time, SIAM J. Control Optim. 28 (5) (1990) 999–1048.
[5] Nicolai V. Krylov, The rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients, Appl. Math. Optim. 52 (3) (2005) 365–399.
[6] Hongjie Dong, Nicolai V. Krylov, The rate of convergence of finite-difference approximations for parabolic Bellman equations with Lipschitz coefficients in cylindrical domains, Appl. Math. Optim. 56 (2007) 37–66.
[7] Espen Robstad Jakobsen, On the rate of convergence of approximation schemes for Bellman equations associated with optimal stopping time problems, Math. Models Methods Appl. Sci. 13 (05) (2003) 613–644.
[8] Jiequn Han, Weinan E., Deep learning approximation for stochastic control problems, 2016, arXiv preprint arXiv:1611.07422.
[9] Kenneth J. Hunt, D. Sbarbaro, R. Żbikowski, Peter J. Gawthrop, Neural networks for control systems—a survey, Automatica 28 (6) (1992) 1083–1112.
[10] Charles-Albert Lehalle, Robert Azencott, Piecewise affine neural networks and nonlinear control, in: ICANN 98: Proceedings of the 8th International Conference on Artificial Neural Networks, Skövde, Sweden, 2–4 September 1998 8, Springer, 1998, pp. 633–638.
[11] D. Psaltis, A. Sideris, A.A. Yamamura, A multilayered neural network controller, IEEE Control Syst. Mag. 8 (2) (1988) 17–21.
[12] Achref Bachouch, Côme Huré, Nicolas Langrené, Huyen Pham, Deep neural networks algorithms for stochastic control problems on finite horizon: numerical applications, Methodol. Comput. Appl. Probab. 24 (1) (2022) 143–178.
[13] Côme Huré, Huyên Pham, Achref Bachouch, Nicolas Langrené, Deep neural networks algorithms for stochastic control problems on finite horizon: convergence analysis, SIAM J. Numer. Anal. 59 (1) (2021) 525–557.
[14] Marcus Pereira, Ziyi Wang, Tianrong Chen, Emily Reed, Evangelos Theodorou, Feynman-Kac neural network architectures for stochastic control using second-order FBSDE theory, in: Learning for Dynamics and Control, PMLR, 2020, pp. 728–738.
[15] Jiequn Han, Arnulf Jentzen, Weinan E., Solving high-dimensional partial differential equations using deep learning, Proc. Natl. Acad. Sci. 115 (34) (2018) 8505–8510.
[16] Jiequn Han, Jihao Long, Convergence of the deep BSDE method for coupled FBSDEs, Probab. Uncertain. Quant. Risk 5 (2020) 1–33.
[17] Shaolin Ji, Shige Peng, Ying Peng, Xichuan Zhang, Solving stochastic optimal control problem via stochastic maximum principle with deep learning method, J. Sci. Comput. 93 (1) (2022) 30.
[18] Jean-Pierre Fouque, Zhaoyu Zhang, Deep learning methods for mean field control problems with delay, Front. Appl. Math. Stat. 6 (2020) 11.
[19] René Carmona, Mathieu Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games: II—the finite horizon case, Ann. Appl. Probab. 32 (6) (2022) 4065–4105.
[20] Ruimeng Hu, Mathieu Laurière, Recent developments in machine learning methods for stochastic control and games, 2023, arXiv preprint arXiv:2303.10257.
[21] Maximilien Germain, Huyên Pham, Xavier Warin, et al., Neural networks-based algorithms for stochastic control and PDEs in finance, 2021, arXiv preprint arXiv:2101.08068.
[22] Kristoffer Andersson, Adam Andersson, Cornelis W. Oosterlee, Convergence of a robust deep FBSDE method for stochastic control, SIAM J. Sci. Comput. 45 (1) (2023) A226–A255.
[23] Yifan Jiang, Jinfeng Li, Convergence of the deep BSDE method for FBSDEs with non-Lipschitz coefficients, Probab. Uncertain. Quant. Risk 6 (4) (2021) 391–408.
[24] Christoph Reisinger, Wolfgang Stockinger, Yufei Zhang, A posteriori error estimates for fully coupled McKean–Vlasov forward-backward SDEs, IMA J. Numer. Anal. (2023) drad060.
[25] Christian Bender, Jessica Steiner, A posteriori estimates for backward SDEs, SIAM/ASA J. Uncertain. Quantif. 1 (1) (2013) 139–163.
[26] Côme Huré, Huyên Pham, Xavier Warin, Deep backward schemes for high-dimensional nonlinear PDEs, Math. Comp. 89 (324) (2020) 1547–1579.
[27] Balint Negyesi, Kristoffer Andersson, Cornelis W. Oosterlee, The One Step Malliavin scheme: new discretization of BSDEs implemented with deep learning regressions, IMA J. Numer. Anal. (2024) drad092.
[28] Chengfan Gao, Siping Gao, Ruimeng Hu, Zimu Zhu, Convergence of the backward deep BSDE method with applications to optimal stopping problems, SIAM J. Finan. Math. 14 (4) (2023) 1290–1303.
[29] Huyên Pham, Continuous-time Stochastic Control and Optimization with Financial Applications, vol. 61, Springer Science & Business Media, 2009.
[30] Jiongmin Yong, Xun Yu Zhou, Stochastic Controls: Hamiltonian Systems and HJB Equations, vol. 43, Springer Science & Business Media, 1999.
[31] Christian Bender, Jianfeng Zhang, Time discretization and Markovian iteration for coupled FBSDEs, Ann. Appl. Probab. 18 (1) (2008) 143–177.
[32] A. Bensoussan, S.C.P. Yam, Z. Zhang, Well-posedness of mean-field type forward–backward stochastic differential equations, Stochastic Process. Appl. 125 (9) (2015) 3327–3354.
[33] Shige Peng, Zhen Wu, Fully coupled forward-backward stochastic differential equations and applications to optimal control, SIAM J. Control Optim. 37 (3) (1999) 825–843.
[34] Fabio Antonelli, Backward Forward Stochastic Differential Equations, Purdue University, 1993.
[35] Etienne Pardoux, Shanjian Tang, Forward-backward stochastic differential equations and quasilinear parabolic PDEs, Probab. Theory Related Fields 114 (1999) 123–150.
[36] Jianfeng Zhang, Backward Stochastic Differential Equations, Springer, 2017.