Journal of Scientific Computing (2020) 85:51
https://doi.org/10.1007/s10915-020-01358-y

An Efficient Numerical Algorithm for Solving Data Driven Feedback Control Problems

Richard Archibald[1] · Feng Bao[2] · Jiongmin Yong[3] · Tao Zhou[4]

Received: 1 August 2020 / Revised: 25 October 2020 / Accepted: 29 October 2020 /
Published online: 10 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
The goal of this paper is to solve a class of stochastic optimal control problems numerically, in which the state process is governed by an Itô type stochastic differential equation with control process entering both in the drift and the diffusion, and is observed partially. The optimal control of feedback form is determined based on the available observational data. We call this type of control problems the data driven feedback control. The computational framework that we introduce to solve such type of problems aims to find the best estimate for the optimal control as a conditional expectation given the observational information. To make our method feasible in providing timely feedback to the controlled system from data, we develop an efficient stochastic optimization algorithm to implement our computational framework.

Keywords Stochastic optimal control · Nonlinear filtering · Data driven · Maximum principle · Stochastic optimization

Mathematics Subject Classification 93E11 · 60G35 · 65K10

1 Introduction

Stochastic optimal control is an important research subject that attracts scientists and engineers in various fields from theoretical scientific research to practical industrial production. The *control process* (also called *control policy*), which controls a stochastic dynamical system whose solution called *state process*, is designed to meet some optimality conditions. For the classic stochastic optimal control problem with full observation of the state, both theo-

✉ Feng Bao
bao@math.fsu.edu

1 Computational Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

2 Department of Mathematics, Florida State University, Tallahassee, FL, USA

3 Department of Mathematics, University of Central Florida, Orlando, FL, USA

4 Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing, China

retical results and numerical methods are extensively studied. However, in practice the full observation of the state process is often not available. Instead, we have detectors/observation facilities to collect partial observational data, which provide indirect information about the state process. The theoretical formulation for partially observable stochastic optimal control is derived analytically [14,20,26,30,40,42], and the corresponding optimal control is a stochastic process adapted to the observational information. Since the controlled state needs to be inferred from observations, the procedure of finding the optimal control requires data analysis for observational data, and the control actions are driven by information contained in data. To highlight the influence of data in designing control policies, we call the stochastic optimal control with partially observed controlled processes the *data driven feedback control*.

The key to solving a data driven feedback control problem is the effective combination of data analysis techniques and stochastic optimal control methods. One of the most widely used methods to combine data with control is the "separation principle" [31,43]. The main theme of separation principle is that the data driven feedback control problem can be separated into a state estimation problem, and a standard stochastic optimal control problem, in which the state of controlled process is provided by the state estimation. Despite of its significant success in practical applications, the separation principle only works for time-invariant linear controlled systems with linear observations. Although several efforts have been made to extend the applicability of the separation principle by linearizing the nonlinear models [14, 33,37], computational methods for general *nonlinear* problems are still needed.

The goal of this paper is to introduce an effective computational framework that incorporates observational information into stochastic optimal control problems and we shall also design an efficient algorithm to implement our computational framework through stochastic optimization. Since the data driven feedback control is adapted to observational information, we use the conditional expectation for the control process as its "best" approximation given observations. In this way, our computational framework is composed by two components: (I) numerical schemes to obtain conditional distribution of the controlled process given observations; (II) computational methods for stochastic optimal control problems. Component (II) provides the overarching algorithmic structure to find the optimal control, and the conditional distribution obtained in component (I) is used to calculate conditional expectations for the optimal control solver. In this paper, we apply optimal filtering methods, which could find the conditional distribution of a dynamical system based on indirect observations [9], to accomplish component (I) in our computational framework. When solving nonlinear optimal filtering problems, Zakai filter and particle filter are two well-known approaches. The methodology of the Zakai filter is to solve for the conditional probability density function (pdf) for the target dynamical system through a parabolic type stochastic partial differential equation (called Zakai equation) [45]; and the particle filter, also known as a sequential Monte Carlo method, describes the desired conditional pdf by using empirical distribution of a set of random samples (particles) [22,27]. Although the Zakai filter could provide more accurate approximation for conditional distributions theoretcally, the particle filter is more popular in solving practical problems due to the high efficiency of Monte Carlo method in approximating high dimensional distributions [35]. Therefore, in our approach we use the particle filter to approximate the conditional distribution of the controlled process. On the other hand, computational methods for solving the stochastic optimal control problem in component (II) are developed under two general frameworks—dynamic programming principle and maximum principle [18,21,39,44]. In this work, we adopt the maximum principle framework to solve the optimal control problem due to its multiple advantages over dynamic programming principle. For example, it has less restrictions on dimension of the problem and it could solve problems with state constraints—especially with finite dimensional terminal

state constraints, and it also allows to have random coefficients in the controlled state equation and/or in the performance cost functional. In maximum principle, we have a stochastic Hamiltonian system that consists of a system of forward backward stochastic differential equations (FBSDEs), which meet certain optimality condition with respect to the optimal control [36,41]. Therefore, solving the stochastic optimal control problem through maximum principle involves solving the FBSDEs system (usually carried out by numerical methods) and an optimization procedure, which is typically achieved by gradient descent iterations. There are several successful numerical methods for solving FBSDEs, which can be categorized by two types of approaches: the first type of approach solves an FBSDEs system through numerical schemes for its equivalent parabolic partial differential equation [32,34]; the second approach solves FBSDEs as a system of stochastic differential equations and solutions of FBSDEs are approximated by their conditional expectations [2,3,5–7,13,16,17,46]. While both approaches effectively solve FBSDEs, implementing numerical algorithms to solve FBSDEs is still a challenging task—even with some advanced high efficiency methods [24], and the computational cost for solving FBSDEs increases dramatically when the dimension of the problem becomes higher. In our approach, we choose the second approach to solve FBSDEs due to the efficiency and flexibility of stochastic computing methods.

From the above discussion, we can see that the implementation of our data driven feedback control framework includes particle filtering for conditional pdf of the controlled precess and numerical optimization for control process which involves numerical solutions for FBSDEs. Since the purpose of data driven feedback control is to give timely feedback to the controlled system based on observational data, efficiency of a numerical method is essential. To provide a highly efficient algorithm to implement our computational framework, we introduce a stochastic optimization algorithm that combines the particle filter with stochastic gradient descent in the optimization procedure for the data driven optimal control. The motivation of applying stochastic optimization is the fact that our computational framework consists multiple conditional expectations and stochastic gradient descent is a very efficient method to treat expectations in the gradient descent type optimization. In this connection, we consider all the random samples, including particles obtained from the particle filter, to be the "data" which we use to calculate conditional expectations. Then, in each gradient descent iteration step, we pick one random sample in the sample space, as well as one particle in the particle cloud, and represent conditional expectations by the realization of stochastic process generated by our picked sample/particle. In this way, the gradient decent becomes its stochastic counterpart and we transfer the computational cost of fully calculating conditional expectations (by averaging all samples/particles) to more iteration steps with one-sample simulation of stochastic process in each step. The justification of stochastic gradient descent is well studied theoretically [29,38] and the efficiency is verified in many application scenarios. The only potential issue for our application of stochastic optimization is that the single-realization representation of conditional expectations could not solve FBSDEs accurately. Indeed, the usage of only one realization of random sample is not sufficient to provide full characterization for solutions of FBSDEs. However, the purpose of solving the FBSDEs system in maximum principle is to formulate a stochastic Hamiltonian system, which is used to describe the gradient process with respect to the control process. Therefore solving FBSDEs is not the goal of our computational framework. Actually, the conditional expectations that we use to approximate solutions of FBSDEs are eventually used to describe the gradient process, and similar justifications for stochastic gradient descent would also apply to FBSDEs in our approach. Moreover, by using the single-realization representation of expectation, we can avoid unnecessary calculations for obtaining fully calculated solutions of FBSDEs, which is the primary computational cost of maximum principle approach for stochastic optimal control problems.

The rest of this paper is organized as following. In Sect. 2, we briefly discuss the theoretical background for the data driven feedback control problem and introduce an optimization framework for finding the optimal control, which will be used to design our numerical algorithms. The computational framework with our efficient stochastic optimization implementation will be introduced in Sect. 3. In Sect. 4, we demonstrate the effectiveness and efficiency of our numerical algorithms by solving both classic benchmark problems and practical feedback control problems. Some concluding remarks and the plan of our future work are given in Sect. 5

2 Data Driven Feedback Control Problem

2.1 Problem Statement

Let $(\Omega, \mathcal{F}, \mathbb{F}^W, \mathbb{P})$ be a complete filtered probability space on which a d-dimensional standard Brownian motion $W := \{W_s\}_{s\geq 0}$ is defined with $\mathbb{F}^W = \{\mathcal{F}_s^W\}_{s\geq 0}$ being its natural filtration augmented by all the $\mathbb{P}$-null sets in $\mathcal{F}$. Consider the following stochastic differential equation (SDE) over a deterministic time interval $[0, T]$

$$dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dW_t, \qquad t \in [0, T], \qquad X_0 = \xi, \tag{1}$$

where X is the (controlled) state process valued in $\mathbb{R}^d$, u is the control process valued in some set $U \subseteq \mathbb{R}^m$, $b : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}^{d\times d}$ are suitable maps, and ξ is a random variable independent of W following a distribution p_0. We call the terms b and σ the *drift* and the *diffusion*, respectively. Let

$$\mathcal{U}[0, T] = \left\{u : [0, T] \times \Omega \to U \;\middle|\; u \text{ is } \mathbb{F}^W\text{-progressively measurable}\right\}.$$

Under some mild conditions, for every square integrable initial variable ξ and control $u \in \mathcal{U}[0, T]$, Eq. (1) admits a unique solution X. To measure the performance of the control u, we introduce the following cost functional

$$J(u) = \mathbb{E}\left[\int_0^T f(t, X_t, u_t)dt + h(X_T)\right]. \tag{2}$$

Then classical optimal control problem can be stated as

Problem (C). For any given initial condition $X_0 = \xi$, find a $\bar{u} \in \mathcal{U}[0, T]$ such that

$$J(\bar{u}) = \inf_{u\in\mathcal{U}[0,T]} J(u). \tag{3}$$

Any $\bar{u} \in \mathcal{U}[0, T]$ satisfying the above is called an *optimal control*. Under proper conditions, if the optimal control $\bar{u}$ exists, it should be a function of the corresponding state process $\bar{X}$, in some sense. However, in reality, we may not be able to observe the true state X. Instead, we introduce the following equation

$$dM_t = g(X_t)dt + dB_t, \qquad M_0 = 0, \tag{4}$$

where $g : \mathbb{R}^d \to \mathbb{R}^\ell$ and B is an ℓ-dimensional standard Brownian motion independent of W. We call the above an observation equation and M is called the *observation process*, which provides partial (noisy) observations for X. We may put (1) and (4) together to get the following augmented system on time interval $[0, T]$

$$d\begin{pmatrix} X_t \\ M_t \end{pmatrix} = \begin{pmatrix} b(t, X_t, u_t) \\ g(X_t) \end{pmatrix} dt + \begin{pmatrix} \sigma(t, X_t, u_t) & 0 \\ 0 & I \end{pmatrix} d\begin{pmatrix} W_t \\ B_t \end{pmatrix}, \qquad \begin{pmatrix} X_0 = \xi \\ M_0 = 0 \end{pmatrix}. \tag{5}$$

Let $\mathbb{F}^B = \{\mathcal{F}_t^B\}_{t\geq 0}$ be the filtration of B augmented by all the $\mathbb{P}$-null sets in $\mathcal{F}$, and $\mathbb{F}^{W,B} \equiv \{\mathcal{F}_t^{W,B}\}_{t\geq 0}$ be the filtration generated by W and B (augmented by $\mathbb{P}$-null sets in $\mathcal{F}$). Under mild conditions, for any square integrable random variable ξ independent of W and B, and any $\mathbb{F}^{W,B}$-progressively measurable process u (valued in U), (5) admits a unique solution (X, M) which is $\mathbb{F}^{W,B}$-adapted. Next, we let $\mathbb{F}^M = \{\mathcal{F}_t^M\}_{t\geq 0}$ be the filtration generated by M (augmented by all the $\mathbb{P}$-null sets in $\mathcal{F}$). Clearly, $\mathbb{F}^M \subset \mathbb{F}^{W,B}$, and $\mathbb{F}^M \neq \mathbb{F}^W$, $\mathbb{F}^M \neq \mathbb{F}^B$, in general. We introduce the set of admissible controls as

$$\mathcal{U}_{ad}[0,T] = \big\{u : [0,T]\times\Omega \to U \;\big|\; u \text{ is } \mathbb{F}^M\text{-progressively measurable}\big\}.$$

To emphasis the $\mathbb{F}^M$-adaptedness, for a control in the admissible control set $\mathcal{U}_{ad}[0,T]$, we denote it by u^M. Replacing u in (5) by u^M , we still have the unique solution (X, M), which is $\mathbb{F}^{W,B}$-adapted. The cost functional should now be modified as

$$J^*(u^M) = \mathbb{E}\Big[\int_0^T f(t, X_t, u_t^M)dr + h(X_T)\Big]. \tag{6}$$

Then we pose the following optimal control problem.

Problem (**C***). For any given square integrable random variable ξ, find a $u^* \in \mathcal{U}_{ad}[0,T]$ such that

$$J^*(u^*) = \inf_{u^M\in\mathcal{U}_{ad}[0,T]} J^*(u^M) \equiv V^*(t,\xi). \tag{7}$$

It is worthy of mentioning that under the induced probability measure P^M with Girsanov transformation $dP^M = \Theta_t^T dP$, where $\Theta_t^T := \exp\big[-\int_t^T g(X_s)dB_s - \int_t^T \frac{1}{2}|g(X_s)|^2 ds\big]$, the observation process M is a standard Brownian motion. Discussions on the above indirect observations based stochastic optimal control can be traced back to the work of Fleming in 1960s [19], followed by Kwakernaak [28], Fleming–Pardoux [20], Haussmann [25,26], Bensoussan [11,12], Baras–Elliott–Kohlmann [10], Li–Tang [30], Tang [40], and Wang–Wu–Xiong [42], and so on. Since the optimal control u^* defined in (7) implements feedback control actions based on the observational data, we name the optimal control problem (C*) with its corresponding stochastic system (5) *"data driven feedback control problem"*.

2.2 Optimization for Optimal Control

In this work, we solve the data driven feedback control problem through an optimization procedure, which is derived by stochastic maximum principle. In the case that the optimal control u^* is in the interior of $\mathcal{U}_{ad}$, one can deduce by using the Gâteaux derivative of u^* and maximum principle that the gradient process of the cost functional J^* with respect to the control process on time interval $t \in [0,T]$ has the following form (see "Appendix" for details)

$$(J^*)'_u(u_t^*) = \mathbb{E}\big[b_u(t, X_t^*, u_t^*)^\top Y_t + \sigma_u(t, X_t^*, u_t^*)^\top Z_t + f_u(t, X_t^*, u_t^*)^\top \;\big|\; \mathcal{F}_t^M\big], \tag{8}$$

where we use subscripts to denote partial derivatives of functions throughout of this paper, and stochastic processes Y and Z are solutions of the following forward backward stochastic differential equations (FBSDEs) system

$$\begin{cases} dX_t^* = b(t, X_t^*, u_t^*)dt + \sigma(t, X_t^*, u_t^*)dW_t, \qquad X_0 = \xi & \text{(SDE)} \\ dM_t^* = g(X_t^*)dt + dB_t, \qquad M_0 = 0 & \text{(SDE)} \\ dY_t = \Big(-b_x(t, X_t^*, u_t^*)^\top Y_t - \sigma_x(t, X_t^*, u_t^*)^\top Z_t - f_x(t, X_t^*, u_t^*)^\top\Big)dt & \\ \qquad\qquad + Z_t dW_t + \zeta_t dB_t, \qquad Y_T = h_x(X_T^*)^\top, & \text{(BSDE)} \end{cases} \tag{9}$$

where Z is the martingale representation of Y with respect to W and ζ is the martingale representation of Y with respect to B. Due to the backward propagation direction, we call the third equation in (9) the backward stochastic differential equation (BSDE), which is derived as the *adjoint equation* of the controlled state equation.

The main theme of our approach is to solve for the data driven feedback optimal control u^* through gradient descent type optimization and the gradient process $(J^*)'_u$ is defined in (8). Specifically, for a pre-chosen $\mathbb{F}^M$-adapted process $u^{0,M}$ (as our initial guess), we introduce the following gradient descent iteration to find the optimal control u_t^* at any time instant $t \in [0, T]$

$$u_t^{l+1,M} = u_t^{l,M} - \rho (J^*)'_u(u_t^{l,M}), \quad l = 0, 1, \ldots, \tag{10}$$

where ρ is the step-size for the gradient, and we use $(J^*)'_u(u)$ to denote the gradient process corresponding to a control process u. The primary challenge in applying the gradient descent optimization (10) to solve for the data driven optimal control u^* is to obtain the gradient process $\{(J^*)'_u(u_t^{l,M})\}_{0\le t\le T}$, which is described by solutions Y_t and Z_t of the FBSDEs system (9) corresponding to the $\mathcal{F}_t^M$ adapted estimated optimal control $u_t^{l,M}$. However, the observational information $\mathcal{F}_t^M$ is gradually increased as we collect more and more data over time. At a certain time instant t, the information $\mathcal{F}_s^M$ for $s \in [t, T]$ is not available to solve for $\{(Y_s, Z_s)\}_{t\le s\le T}$, and the estimated feedback control $\{u_s^{l,M}\}_{t\le s\le T}$ should not be driven by information $\{\mathcal{F}_s^M\}_{t\le s\le T}$. Therefore, we target on finding the optimal control u_t^* at the current time t with accessible information $\mathcal{F}_t^M$. Since the evaluation for $(J^*)'_u(u_t^{l,M})$ requires trajectories $\{(Y_s, Z_s)\}_{t\le s\le T}$ as Y_t and Z_t are solved backwards from T to t, we take conditional expectation $\mathbb{E}[\,\cdot\,|\mathcal{F}_t^M]$ to the gradient process $\{(J^*)'_u(u_s^{l,M})\}_{t\le s\le T}$, i.e.

$$\begin{aligned}\mathbb{E}[(J^*)'_u(u_s^{l,M})|\mathcal{F}_t^M] = \mathbb{E}\Big[&b_u(s, X_s, u_s^{l,M})^\top Y_s \\ &+ \sigma_u(s, X_s, u_s^{l,M})^\top Z_s + f_u(s, X_s, u_s^{l,M})^\top \;\Big|\; \mathcal{F}_t^M\Big], \quad s \in [t, T],\end{aligned} \tag{11}$$

where X_s, Y_s and Z_s are corresponding to the estimated control $u_s^{l,M}$. Similarly, for the gradient descent iteration (10) on the time interval $[t, T]$, by taking conditional expectation $\mathbb{E}[\,\cdot\,|\mathcal{F}_t^M]$, we obtain

$$\mathbb{E}[u_s^{l+1,M}|\mathcal{F}_t^M] = \mathbb{E}[u_s^{l,M}|\mathcal{F}_t^M] - \rho\mathbb{E}\big[(J^*)'_u(u_s^{l,M})|\mathcal{F}_t^M\big], \quad l = 0, 1, \ldots, \quad s \in [t, T]. \tag{12}$$

When $s = t$, the conditional expectation $\mathbb{E}[u_s^{l+1,M}|\mathcal{F}_t^M]$ on the left hand side of (12) gives us $u_t^{l+1,M} = \mathbb{E}[u_t^{l+1,M}|\mathcal{F}_t^M]$ due to the $\mathcal{F}_t^M$ adaptedness of $u_t^{l+1,M}$. In this work, we shall derive numerical methods to calculate the conditional estimated control process $\mathbb{E}[u_s^{l+1,M}|\mathcal{F}_t^M]$ for $s \in [t, T]$, where t gradually increases according to data reception. The procedure of calculating $\{\mathbb{E}[u_s^{l+1,M}|\mathcal{F}_t^M]\}_{t\le s\le T}$ would give us an estimate for $u_t^{l+1,M}$, which is what we need in the gradient descent scheme (10) to obtain u_t^*. However, in the conditional gradient descent iteration (12), the gradient $(J^*)'_u$ on the right hand side is corresponding to $u_s^{l,M}$, which is $\mathcal{F}_s^M$-adapted. Since the observational information $\{\mathcal{F}_s^M\}_{t\le s\le T}$ is not available at time t, to make the iteration (12) implementable, we use conditional expectation $\mathbb{E}[u_s^{l+1,M}|\mathcal{F}_t^M]$ to replace $u_s^{l,M}$ since it provides the *best approximation* for $u_s^{l+1,M}$ given $\mathcal{F}_t^M$. In this way, we denote

$$u_s^{l,M}|_t := \mathbb{E}[u_s^{l+1,M}|\mathcal{F}_t^M]$$

and carry out the following gradient descent iteration

$$u_s^{l+1,M}|_t = u_s^{l,M}|_t - \rho\mathbb{E}\big[(J^*)'_u(u_s^{l,M}|_t)|\mathcal{F}_t^M\big], \; l = 0, 1, \ldots, \; s \in [t, T], \tag{13}$$

to derive $u_t^{l+1,M}$ as desired. From Eq. (13), we can see that the conditional gradient process $\mathbb{E}\big[(J^*)'_u(u_s^{l,M}|_t)|\mathcal{F}_t^M\big]$ is the key component and the effort to obtain such a term is composed by the following two tasks: (i) obtaining solutions Y and Z of the following FBSDEs system

$$\begin{aligned} dX_s &= b(s, X_s, u_s^{l,M}|_t)ds + \sigma(s, X_s, u_s^{l,M}|_t)dW_s, \qquad s \in [t, T] \\ dY_s &= \big(-b_x(s, X_s, u_s^{l,M}|_t)^\top Y_s - \sigma_x(s, X_s, u_s^{l,M}|_t)^\top Z_s \\ &\qquad - f_x(s, X_s, u_s^{l,M}|_t)^\top\big)ds + Z_s dW_s + \zeta_s dB_s, \quad Y_T = h_x^\top(X_T) \end{aligned} \tag{14}$$

where the controlled process X is corresponding to the estimated control $u_s^{l,M}|_t$; and (ii) obtaining effective evaluations for the conditional expectation $E[\psi(s)|\mathcal{F}_t^M]$ for a $\mathcal{F}_s^X \vee \mathcal{F}_t^M$ adapted stochastic process $\psi(s)$. In most situations, the FBSDEs system (14) is not explicitly solvable. Therefore, we use numerical algorithms to approximate solutions X, Y and Z in this work. The reason that we did not include the observation equation $dM_s = g(X_s)ds + dB_s$ in the FBSDEs system (14) is that M is not contained in the expression (11) for the gradient process. Moreover, when the control process is chosen to be $u_s^{l,M}|_t$, the adjoint BSDE in (14) is no longer driven by observational information beyond time t. In fact, M is the data collected by observation equipments and the observational data $\{M_s\}_{s>t}$ is not available at time t, and therefore the observation dynamics should not be incorporated into the FBSDEs system.

On the other hand, when the controlled dynamics and the observation function g are nonlinear, obtaining the conditional expectation $\mathbb{E}[\psi(s)|\mathcal{F}_t^M]$ (in task (ii)) is not trivial either. In this paper, we introduce an optimal filtering framework to achieve this goal. The standard formulation for an optimal filtering problem is given by the following state-space model,

$$\begin{aligned} dS_t &= \beta(t, S_t)dt + \gamma(t, S_t)dW_t \\ dM_t &= g(S_t)dt + dB_t, \end{aligned} \tag{15}$$

where the drift term $\beta : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ describes some dynamical model, and $\gamma : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^{d\times l}$ is the diffusion coefficient that drives the Brownian motion $W \in \mathbb{R}^l$ perturbing the model. The stochastic process S_t is usually called the *state process*, and M_t provides partial noisy observations for S_t in the same manner of the "measurement process" introduced in (4). The goal of the optimal filtering problem is to obtain the best estimate for $\Psi(S_t)$ given the observational information $\mathcal{F}_t^M$, where Ψ is a test function. Mathematically, we aim to find the optimal filter $\tilde{\Psi}_t$ defined by the conditional expectation $\tilde{\Psi}_t := \mathbb{E}[\Psi(S_t)|\mathcal{F}_t^M]$. When all the functions β, γ and g in the optimal filtering problem are linear, the desired conditional expectation can be derived analytically by the well-known Kalman filter method. However, in most practical applications where the optimal filtering problems are nonlinear, numerical methods are used to approximate the conditional probability density function (pdf) for the state, i.e. $p(S_t|\mathcal{F}_t^M)$, which can be used to calculate the optimal filter through integral $\tilde{\Psi}_t = \int_{\mathbb{R}^d} \Psi(x) \cdot p(x\big|\mathcal{F}_t^M)dx$. Well known nonlinear filtering methods include the Zakai filter, in which the desired filtering density is formulated as solution of a parabolic type stochastic PDE (Zakai equation), and the particle filter, which provides an empirical distribution to describe $p(S_t|\mathcal{F}_t^M)$.

In our data driven feedback control problem, we let the controlled process X_t be the target state process S_t in the optimal filtering problem and generalize the test function Ψ as a stochastic process defined by

$$\Psi(s, X_s, u_s^{l,M}|_t) := b_u(s, X_s, u_s^{l,M}|_t)^\top Y_s + \sigma_u(s, X_s, u_s^{l,M}|_t)^\top Z_s + f_u(s, X_s, u_s^{l,M}|_t)^\top \tag{16}$$

for $s \in [t, T]$. With the conditional pdf $p(X_t|\mathcal{F}_t^M)$ that we obtain through optimal filtering methods and the fact that $\Psi(s, X_s, u_s^{l,M}|_t)$ is a stochastic process depending on the state of random variable X_t, the conditional gradient process $\mathbb{E}\big[(J^*)'_u(u_s^{l,M}|_t)|\mathcal{F}_t^M\big]$ in (13) can be obtained by the following integral

$$\mathbb{E}\big[(J^*)'_u(u_s^{l,M}|_t)|\mathcal{F}_t^M\big] = \int_{\mathbb{R}^d} \mathbb{E}[\Psi(s, X_s, u_s^{l,M}|_t)\big|X_t = x] \cdot p(x\big|\mathcal{F}_t^M)dx, \quad s \in [t, T]. \tag{17}$$

In the following section, we derive a computational framework that carries out the gradient descent optimization (13) to evaluate the data driven feedback control u_t^*, and introduce an efficient stochastic optimization algorithm to implement our computational framework. In many practical applications, it is difficult to control the diffusion. Therefore, in this work we assume that the diffusion coefficient σ does not contain the control term.

3 Numerical Approach for Data Driven Feedback Control

The general framework of our data driven feedback control system is designed on a temporal partition Π_{N_T} that reflects the capability of collecting measurement data and implementing control actions, where Π_{N_T} is defined by

$$\Pi_{N_T} := \{t_n : 0 = t_0 < t_1 < t_2 < \cdots < t_n < \cdots < t_{N_T} = T\}, \quad N_T \in \mathbb{N},$$

and we use the control sequence $\{u_{t_n}^*\}_{n=1}^{N_T}$ to represent the control process u^*. In this way, the gradient descent (10) is restricted to discrete time instants Π_{N_T}, and the conditional gradient descent scheme that we use to obtain $u_{t_n}^*$ becomes

$$u_{t_i}^{l+1,M}|_{t_n} = u_{t_i}^{l,M}|_{t_n} - \rho\mathbb{E}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})|\mathcal{F}_{t_n}^M\big], \quad n = 0, 1, \ldots, N_T, \quad n \le i \le N_T, \tag{18}$$

where the conditional gradient process on the right hand side of (18) is given by

$$\mathbb{E}[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})|\mathcal{F}_{t_n}^M] = \mathbb{E}\Big[b_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top Y_{t_i} + f_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top \;\big|\; \mathcal{F}_{t_n}^M\Big], \tag{19}$$

and X_{t_i} is $\mathcal{F}_{t_i}^W \vee \mathcal{F}_{t_n}^M$ measurable since its corresponding control is $u_{t_i}^{l,M}|_{t_n}$. In what follows, we introduce numerical schemes to calculate X_{t_i}, Y_{t_i} and Z_{t_i} corresponding to the estimated conditional optimal control $u_{t_i}^{l,M}|_{t_n}$ in Sect. 3.1, and introduce a particle filter method, which is one of the most widely accepted methods for solving nonlinear filtering problems, to approximate the conditional distribution $p(X_{t_n}|\mathcal{F}_{t_n}^M)$ in Sect. 3.2. Then, in Sect. 3.3 we combine schemes in Sects. 3.1–3.2 to formulate a computational framework for the data driven feedback control problem with an efficient stochastic optimization implementation, and we summarize our numerical approach in Sect. 3.4.

3.1 Numerical Schemes for FBSDEs

We consider the FBSDEs system (14) on time interval $[t_i, t_{i+1}]$, i.e.

$$\begin{aligned} X_{t_{i+1}} =& X_{t_i} + \int_{t_i}^{t_{i+1}} b(s, X_s, u_s^{l,M}|_{t_n})ds + \int_{t_i}^{t_{i+1}} \sigma(s, X_s)dW_s, \qquad \text{(SDE)} \\ Y_{t_i} =& Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} \Big(b_x(s, X_s, u_s^{l,M}|_{t_n})^\top Y_s + \sigma_x(s, X_s)^\top Z_s + f_x(s, X_s, u_s^{l,M}|_{t_n})^\top\Big)ds \\ & - \int_{t_i}^{t_{i+1}} Z_s dW_s + \int_{t_i}^{t_{i+1}} \zeta_s dB_s. \qquad \text{(BSDE)} \end{aligned} \tag{20}$$

To solve the first equation in (20), which is a standard forward SDE, we use the left-point formula to approximate both the deterministic and stochastic integrals, and obtain the following Euler-Maruyama discretized approximation equation for X as

$$X_{t_{i+1}} = X_{t_i} + b(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})\Delta t_i + \sigma(t_i, X_{t_i})\Delta W_{t_i} + R_{X,n}^i, \tag{21}$$

where $R_{X,n}^i$ is the approximation error for integrals, and we denote $\Delta t_i := t_{i+1} - t_i$ and $\Delta W_{t_i} := W_{t_{i+1}} - W_{t_i}$.

To derive a discretization scheme for solution Y, we take conditional expectation $\mathbb{E}_i[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_{t_i}^{X,B}]$ on both sides of the BSDE in (20) to get

$$Y_{t_i} = \mathbb{E}_i[Y_{t_{i+1}}] + \int_{t_i}^{t_{i+1}} \mathbb{E}_i\big[b_x(s, X_s, u_s^{l,M}|_{t_n})^\top Y_s + \sigma_x(s, X_s)^\top Z_s + f_x(s, X_s, u_s^{l,M}|_{t_n})^\top\big]ds,$$

where the stochastic integrals are eliminated and we have used the fact that $Y_{t_i} = \mathbb{E}_i[Y_{t_i}]$ due to the adaptedness of Y. Then, we use the right-point formula to approximate the deterministic integral in the above equation and obtain

$$\begin{aligned} Y_{t_i} = \mathbb{E}_i[Y_{t_{i+1}}] + \mathbb{E}_i\Big[& b_x(t_{i+1}, X_{t_{i+1}}, u_{t_{i+1}}^{l,M}|_{t_n})^\top Y_{t_{i+1}} + \sigma_x(t_{i+1}, X_{t_{i+1}})^\top Z_{t_{i+1}} \\ & + f_x(t_{i+1}, X_{t_{i+1}}, u_{t_{i+1}}^{l,M}|_{t_n})^\top\Big]\Delta t_i + R_{Y,n}^i, \end{aligned} \tag{22}$$

where $R_{Y,n}^i$ is the approximation error for the integral.

For solution Z, we multiply ΔW_{t_i} and take conditional expectation $\mathbb{E}_i[\cdot]$ on both sides of the BSDE in (20) to get

$$\begin{aligned} \mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} Z_s \Delta W_{t_i} dW_s\right] =& \mathbb{E}_i[Y_{t_{i+1}}\Delta W_{t_i}] + \int_{t_i}^{t_{i+1}} \mathbb{E}_i\Big[\Big(b_x(s, X_s, u_s^{l,M}|_{t_n})^\top Y_s \\ & + \sigma_x(s, X_s)^\top Z_s + f_x(s, X_s, u_s)^\top\Big)\Delta W_{t_i}\Big]ds, \end{aligned}$$

where we have used the fact $\mathbb{E}_i[Y_{t_i}\Delta W_{t_i}] = 0$ and the independence of Brownian motions W and B. Then, we use the left-point formula to approximate both deterministic and stochastic integrals in the above equation and obtain

$$Z_{t_i}\Delta t_i = \mathbb{E}_i[Y_{t_{i+1}}\Delta W_{t_i}] + R_{Z,n}^i, \tag{23}$$

where the deterministic integral is eliminated due to the $\mathbb{F}^W$ adaptedness and the fact that we picked the left-point formula approximation, and $R_{Z,n}^i$ is the term containing approximation errors for integrals.

By dropping approximation error terms $R^i_{X,n}$, $R^i_{Y,n}$ and $R^i_{Z,n}$ in (21), (22) and (23), respectively, we could provide a numerical algorithm for solving X, Y and Z corresponding to the gradient process $\mathbb{E}[(J^*)'_u(u^{l,M}_{t_i}|_{t_n})|\mathcal{F}^M_{t_n}]$ as introduced in (19):
For a fixed $n \leq N_T - 1$ representing the current time t_n, and for $i = N_T - 1, \ldots, n+1, n$ with initial conditions Y_{N_T} and Z_{N_T}, we solve the FBSDEs system through the following schemes

$$\begin{aligned} X_{i+1} =& X_i + b(t_i, X_i, u^{l,M}_{t_i}|_{t_n})\Delta t_i + \sigma(t_i, X_i)\Delta W_{t_i} \\ Y_i =& \mathbb{E}_i[Y_{i+1}] + \mathbb{E}_i\Big[b_x(t_{i+1}, X_{i+1}, u^{l,M}_{t_{i+1}}|_{t_n})^\top Y_{i+1} + \sigma_x(t_{i+1}, X_{i+1})^\top Z_{i+1} \\ & + f_x(t_{i+1}, X_{i+1}, u^{l,M}_{t_{i+1}}|_{t_n})^\top\Big]\Delta t_i \\ Z_i =& \mathbb{E}_i[Y_{i+1}\Delta W_{t_i}] \cdot (\Delta t_i)^{-1} \end{aligned} \tag{24}$$

where X_{i+1}, Y_i and Z_i are numerical approximations for $X_{t_{i+1}}$, Y_{t_i} and Z_{t_i}, respectively. The numerical analysis for schemes (24) and their extended versions are discussed in [4,47].

Remark 3.1 We notice that the random variable ζ is not involved in numerical schemes (24) to calculate approximate solutions X_{i+1}, Y_i and Z_i, and it's not contained in the expression for $(J^*)'_u$. Therefore, we do not need an extra scheme to approximate ζ in our approach.

In order to carry out numerical schemes (24), one needs to evaluate conditional expectations. There are many well-known methods to approximate expectations, such like (quasi) Monte Carlo methods and numerical integration methods, etc. In this work, we adopt the standard Monte Carlo method due to its high effectiveness and efficiency in approximating expectations—especially in high dimensional spaces. Specifically, we use a set of $K \in \mathbb{N}$ random samples to describe ΔW_{t_i} and the expectations are approximated by sample means. As a result, schemes (24) now become

$$\begin{aligned} X^k_{i+1} =& X_i + b(t_i, X_i, u^{l,M}_{t_i}|_{t_n})\Delta t_i + \sigma(t_i, X_i)\sqrt{\Delta t_i}\omega^k_i, \qquad k = 1, 2 \ldots, K, \\ Y_i =& \sum_{k=1}^{K} \frac{Y^k_{i+1}}{K} + \frac{\Delta t_i}{K}\sum_{k=1}^{K}\Big[b_x(t_{i+1}, X^k_{i+1}, u^{l,M}_{t_{i+1}}|_{t_n})^\top Y^k_{i+1} \\ & + \sigma_x(t_{i+1}, X^k_{i+1})^\top Z^k_{i+1} + f_x(t_{i+1}, X^k_{i+1}, u^{l,M}_{t_{i+1}}|_{t_n})^\top\Big], \\ Z_{t_i} =& \frac{1}{\Delta t_i}\sum_{k=1}^{K}\frac{Y^k_{i+1}\sqrt{\Delta t_i}\omega^k_i}{K}, \end{aligned} \tag{25}$$

where $\{\omega^k_i\}_{k=1}^K$ is a set of random samples following the standard Gaussian distribution that we use to describe the randomness of ΔW_{t_i}; Y^k_{i+1} and Z^k_{i+1} denote approximate solutions Y_{i+1} and Z_{i+1} corresponding to the k-th random sample in the Monte Carlo approximation.

3.2 Particle Filter Method for Conditional Distribution

The purpose of our optimal filtering procedure is to obtain the conditional distribution for the controlled process, which will be used to generate the conditional gradient process $\mathbb{E}\big[(J^*)'_u(u^{l,M}_{t_i}|_{t_n})|\mathcal{F}^M_{t_n}\big]$ with respect to the control process. To proceed, we consider the controlled process on time interval $[t_{n-1}, t_n]$, i.e.

$$X_{t_n} = X_{t_{n-1}} + \int_{t_{n-1}}^{t_n} b(s, X_s, u_s)ds + \int_{t_{n-1}}^{t_n} \sigma(s, X_s)dW_s, \tag{26}$$

and the observational data at time instant t_n is given by $M_{t_n} = g(X_{t_n}) + \eta_n$, where η_n is the observational noise. Well known pioneer work by Zakai showed that the desired filtering density is equivalent to the solution of a stochastic PDE (the Zakai equation). Although the Zakai's approach theoretically solves the nonlinear filtering problem, due to the "curse of dimensionality" of solving a stochastic PDE in high dimensional space, the Zakai filter is prohibitive in practically applications. In this work, we adopt the Bayesian filter framework and introduce a widely accepted Bayesian filter—the particle filter method—to solve for the filtering density $p(X_{t_n}|\mathcal{F}^M_{t_n})$. The Bayesian filter framework is composed by two stages: *prediction stage* and *update stage*. In the prediction stage, assuming that we have the distribution $p(X_{t_{n-1}}|\mathcal{F}^M_{t_{n-1}})$ at time instant t_{n-1}, the prior pdf that predicts the state of controlled process at time t_n is given by the following Chapman-Kolmogorov formula

$$p(X_{t_n}|\mathcal{F}^M_{t_{n-1}}) = \int p(X_{t_{n-1}}|\mathcal{F}^M_{t_{n-1}})p(X_{t_n}|X_{t_{n-1}})dX_{t_{n-1}},$$

where $p(X_{t_n}|X_{t_{n-1}})$ is the transition probability derived from the state dynamics (26). With the new observational data M_{t_n}, the update stage uses the Bayesian inference to update the prior pdf and get the posterior pdf $p(X_{t_n}|\mathcal{F}^M_{t_n})$ as

$$p(X_{t_n}|\mathcal{F}^M_{t_n}) = \frac{p(X_{t_n}|\mathcal{F}^M_{t_{n-1}})p(M_{t_n}|X_{t_n})}{p(M_{t_n}|\mathcal{F}^M_{t_{n-1}})}, \tag{27}$$

where $p(M_{t_n}|X_{t_n})$ is the likelihood function that describes the discrepancy between the predicted state and the observations [15]. In what follows, we introduce the benchmark particle filter method (bootstrap filter algorithm [22]) to the implement the aforementioned Bayesian filter framework due to the high effectiveness and efficiency of the particle filter in solving nonlinear filtering problems. To proceed, assume that at time instant t_{n-1} we have S particles, denoted by $\{x^{(s)}_{n-1}\}^S_{s=1}$, that follow an empirical distribution $\pi(X_{t_{n-1}}|\mathcal{F}^M_{t_{n-1}}) := \frac{1}{S}\sum_{s=1}^S \delta_{x^{(s)}_{n-1}}(X_{t_{n-1}})$ as an approximation for $p(X_{t_{n-1}}|\mathcal{F}^M_{t_{n-1}})$, where δ_x is the Dirac delta function at x. The prior pdf that we want to find in the prediction stage is approximated as

$$\tilde{\pi}(X_{t_n}|\mathcal{F}^M_{t_{n-1}}) := \frac{1}{S}\sum_{s=1}^S \delta_{\tilde{x}^{(s)}_n}(X_{t_n}), \tag{28}$$

where $\tilde{x}^{(s)}_n$ is sampled from $\pi(X_{t_{n-1}}|\mathcal{F}^M_{t_{n-1}})p(X_{t_n}|X_{t_{n-1}})$. As a result, the sample cloud $\{\tilde{x}^{(s)}_n\}^S_{s=1}$ provides an approximate distribution for the prior $p(X_{t_n}|\mathcal{F}^M_{t_{n-1}})$. Then, in the update stage, we update the approximated prior pdf to get the posterior pdf by replacing $p(X_{t_n}|\mathcal{F}^M_{t_{n-1}})$ with $\tilde{\pi}(X_{t_n}|\mathcal{F}^M_{t_{n-1}})$ in the Bayesian inference (27) as following

$$\tilde{\pi}(X_{t_n}|\mathcal{F}^M_{t_n}) := \frac{\sum_{s=1}^S \delta_{\tilde{x}^{(s)}_n}(X_{t_n})p(M_{t_n}|\tilde{x}^{(s)}_n)}{\sum_{s=1}^S p(M_{t_n}|\tilde{x}^{(s)}_n)} = \sum_{s=1}^S w^{(s)}_n \delta_{\tilde{x}^{(s)}_n}(X_{t_n}). \tag{29}$$

In this way, we obtain a weighted empirical distribution $\tilde{\pi}(X_{t_n}|\mathcal{F}^M_{t_n})$ that approximates the posterior pdf $p(X_{t_n}|\mathcal{F}^M_{t_n})$ with the importance density weight $w^{(s)}_n \propto p(M_{t_n}|\tilde{x}^{(s)}_n)$. In practice, importance weights $\{w^{(s)}_n\}^S_{s=1}$ tend to concentrate on a few samples after several time steps, which dramatically reduces the effective particle size in the algorithm. To avoid such degeneracy problem, we resample the particles $\{\tilde{x}^{(s)}_n\}^S_{s=1}$ by replacing particles with low density weights with copies of particles with high weights. In the bootstrap particle filter,

we use importance sampling method to generate equally weighted samples $\{x_n^{(s)}\}_{s=1}^S$ from $\tilde{\pi}(X_{t_n}|\mathcal{F}_{t_n}^M)$ to formulate our empirical distribution $\pi(X_{t_n}|\mathcal{F}_{t_n}^M)$ that describes the conditional pdf for the controlled process $p(X_{t_n}|\mathcal{F}_{t_n}^M)$ [22], i.e.

$$\pi(X_{t_n}|\mathcal{F}_{t_n}^M) = \frac{1}{S}\sum_{s=1}^{S}\delta_{x_n^{(s)}}(X_{t_n}). \tag{30}$$

3.3 Stochastic Optimization for Control Process

The procedure that aims to determine the data driven optimal control is the gradient descent iteration. In this subsection, we combine the numerical schemes for the adjoint FBSDEs system (14) (introduced in Sect. 3.1) and the particle filter algorithm (introduced in Sect. 3.2) to formulate an efficient stochastic optimization algorithm to solve for the optimal control process u^*.

The framework that we adopt to design our algorithm is the conditional gradient descent iteration (18) and the conditional gradient process with respect to the control process, i.e. $\mathbb{E}[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})|\mathcal{F}_{t_n}^M]$, is calculated by using the integral expression (17). Specifically, on a time instant $t_n \in \Pi_{N_T}$, we have

$$\mathbb{E}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})|\mathcal{F}_{t_n}^M\big] = \int_{\mathbb{R}^d}\mathbb{E}[\Psi(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})\big|X_{t_n} = x]\cdot p(x\big|\mathcal{F}_{t_n}^M)dx,$$

where $t_i \geq t_n$ is a time instant after t_n and the test function Ψ in (16) is chosen as

$$\Psi(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n}) = b_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top Y_{t_i} + f_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top.$$

In our numerical approach, we use approximate solutions (Y_i, Z_i) of FBSDEs from schemes (25) to replace (Y_{t_i}, Z_{t_i}) and the conditional distribution $p(X_{t_n}\big|\mathcal{F}_{t_n}^M)$ is approximated by the empirical distribution $\pi(X_{t_n}|\mathcal{F}_{t_n}^M)$ obtained from the particle filter algorithm (28)–(30). Specifically, for increasing time instants $t_n \in \Pi_{N_T}$, $n = 0, 1, 2, \ldots, N_T$, we solve for the optimal control $u_{t_n}^*$ through the following gradient descent optimization iteration

$$u_{t_i}^{l+1,M}|_{t_n} = u_{t_i}^{l,M}|_{t_n} - \rho\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big], \quad l = 0, 1, 2, \ldots, L, \quad t_i \geq t_n, \tag{31}$$

where $\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big]$ is defined by

$$\begin{aligned}\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big] := \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}\Big[&b_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top Y_i \\ &+ f_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top \mid X_{t_n} = x_n^{(s)}\Big],\end{aligned} \tag{32}$$

and samples $\{x_n^{(s)}\}_{s=1}^S$ follow distribution $\pi(X_{t_n}|\mathcal{F}_{t_n}^M)$ which describes the conditional pdf of the controlled process X_{t_n} given the observational information $\mathcal{F}_{t_n}^M$. Here, we want to re-emphasize that the purpose of carrying out an optimization procedure for the conditional control process $\{u_{t_i}^{\cdot,M}|_{t_n}\}_{i=n}^{N_T}$ is to find an estimate for the optimal control at time t_n. Since the approximate conditional gradient process on the right hand side of (32) is still under expectation, Monte Carlo simulation for $\mathbb{E}[\cdot|X_{t_n} = x_n^{(s)}]$ is needed in general, and

$\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big]$ is approximated as following

$$\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big] \approx \frac{1}{S}\frac{1}{\Lambda}\sum_{s=1}^{S}\sum_{\lambda=1}^{\Lambda}\Big[b_u(t_i, X_{t_i}^{(\lambda,s)}, u_{t_i}^{l,M}|_{t_n})^\top Y_i^{(\lambda,s)} + f_u(t_i, X_{t_i}^{(\lambda,s)}, u_{t_i}^{l,M}|_{t_n})^\top \mid X_{t_n} = x_n^{(s)}\Big], \tag{33}$$

where Λ is the number of samples to approximate the expectation $\mathbb{E}[\cdot|X_{t_n} = x_n^{(s)}]$, $X_{t_i}^{(\lambda,s)}$ is the λ-th realization of controlled process with initial state $X_{t_n} = x_n^{(s)}$, and $Y_i^{(\lambda,s)}$ is the approximate solution Y_i corresponding to $X_{t_i}^{(\lambda,s)}$. We can see from the above Monte Carlo approximation that in order to approximate the expectation $\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big]$ in one gradient descent iteration step, we need to generate altogether $S \times \Lambda$ samples of controlled process and evaluate the values of solution Y_i corresponding to $\{X_{t_i}^{(\lambda,s)}\}_{\lambda,s}$ as well. This is even more computationally expensive when the controlled system is a high dimensional process.

Inspired by the application of stochastic gradient descent (SGD) in improving the efficiency of classic gradient descent optimization, in this work we introduce a stochastic optimization algorithm that combines the particle filter method with the stochastic gradient descent algorithm to carry out the gradient descent iteration procedure (31) efficiently. When implementing the stochastic gradient descent in the optimization scheme (31), instead of using the fully calculated Monte Carlo simulation to approximate the conditional expectation on the right hand side of (32) [as introduced in (33)], we use only one realization of X_{t_i} to represent the expectation [1]. Specifically, the expectation $\mathbb{E}\big[b_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top Y_i + f_u(t_i, X_{t_i}, u_{t_i}^{l,M}|_{t_n})^\top|X_{t_n} = x_n^{(s)}\big]$ is represented by the random variable $b_u(t_i, X_{t_i}^{(\hat{l},s)}, u_{t_i}^{l,M}|_{t_n})^\top Y_i^{(\hat{l},s)} + f_u(t_i, X_{t_i}^{(\hat{l},s)}, u_{t_i}^{l,M}|_{t_n})^\top$, where $X_{t_i}^{(\hat{l},s)}$ is a randomly generated realization of the controlled process with initial state $X_{t_n}^{(\hat{l},s)} = x_n^{(s)}$, and we use the index $\hat{l}$ to indicate that the random generation of the controlled process varies among the gradient descent iteration steps in (31). In addition, one should notice that the particles $\{x_n^{(s)}\}_{s=1}^S$ are used to describe the conditional distribution of the controlled process and the Monte Carlo average in the the approximation scheme (32) also aims to approximate a conditional expectation. In light of the concept of single-realization representation for expectation, we extend this concept to the Monte Carlo averaging of the particles in the approximation scheme (32). Therefore, we use the following expresssion to represent the conditional expectation $\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big]$ in (31)

$$\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big] \approx b_u(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top Y_i^{(\hat{l},\hat{s})} + f_u(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top, \tag{34}$$

where $X_{t_i}^{(\hat{l},\hat{s})}$ indicates a *randomly generated* realization of the controlled process with a *randomly selected* initial state $X_{t_n}^{(\hat{l},\hat{s})} = x_n^{(\hat{s})}$ from the particle cloud $\{x_n^{(s)}\}_{s=1}^S$. In this way, the gradient descent optimization scheme (31) becomes the following SGD optimization iteration scheme

$$u_{t_i}^{l+1,M}|_{t_n} = u_{t_i}^{l,M}|_{t_n} - \rho\Big(b_u(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top Y_i^{(\hat{l},\hat{s})} + f_u(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top\Big), \quad l = 0, 1, 2, \ldots, L, \quad t_i \geq t_n, \tag{35}$$

where $Y_i^{(\hat{l},\hat{s})}$ is the approximate solution Y_i corresponding to the random sample $X_{t_i}^{(\hat{l},\hat{s})}$ indexed by the sample pair $(\hat{l},\hat{s})$, and $X_{t_i}^{(\hat{l},\hat{s})}$ is generated as following

$$X_{t_{i+1}}^{(\hat{l},\hat{s})} = X_{t_i}^{(\hat{l},\hat{s})} + b(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})\Delta t_i + \sigma(s, X_s)\sqrt{\Delta t_i}\,\omega_i^{(\hat{l},\hat{s})},\ i = n, \ldots, N_T - 1 \quad (36)$$

where $X_{t_n}^{(\hat{l},\hat{s})} = x_n^{(\hat{s})} \in \{x_n^{(s)}\}_{s=1}^{S}$, $\omega_i^{(\hat{l},\hat{s})} \sim N(0,1)$, and $\{\omega_i^{(\hat{l},\hat{s})}\}_{i=1}^{N_T-1}$ form a sequence of random variables corresponding to the sample indices $\hat{l}$ and $x_n^{(\hat{s})}$. Then, the resulted conditional optimal control process $\{u_{t_i}^{L,M}|_{t_n}\}_{i=n}^{N_T}$ leads to an estimate for our desired data driven optimal control at time instant t_n, i.e. $\hat{u}_{t_n}^*$, which is given by

$$\hat{u}_{t_n}^* := u_{t_n}^{L,M}|_{t_n}.$$

As a result, particles in the particle filter method (PF) and random samples in the stochastic gradient descent method (SGD) are combined into one stochastic optimization iteration (35), and we use only one-realization to represent two layers of expectations instead of altogether $S \times \Lambda$ calculations of $X_{t_i}^{(\lambda,s)}$ and $Y_i^{(\lambda,s)}$. For convenience of presentation, we name this methodology the "PF-SGD".

It is important to point out that in order to characterize the dependence of numerical solution Y_i on $X_{t_i}^{(\hat{l},\hat{s})}$ as we indicated in (35), one needs to provide approximations for Y and Z corresponding to the random variable X. In practice, people typically derive approximate solutions Y_i and Z_i on a set of spatial points that represent X_{t_i}, and then construct interpolatory approximations based on discrete solution values on those selected spatial points. This may also be considered as spatial dimension approximation for numerical solutions of FBSDEs. Recently, a machine learning based method has been developed, which could solve FBSDEs in high dimensional space on a set of pre-selected spatial points [24]. However, even with accurately calculated solutions on a large number of spatial points, characterizing Y and Z with respect to the random variable X is still challenging due to the sparsity of spatial points in high dimensional space. To address the challenge in spatial dimension approximation, in this work we take the advantage of SDE nature of FBSDEs and incorporate numerical schemes for FBSDEs into our PF-SGD framework. Specifically, we further extend the concept of singel-realization representation of expectation to the numerical schemes (25) for FBSDEs and use the solution path $\{X_{t_i}^{(\hat{l},\hat{s})}\}_{i=n}^{N_T}$ to drive our numerical solution for Y and Z as following

$$\begin{aligned} Y_i^{(\hat{l},\hat{s})} =& Y_{i+1}^{(\hat{l},\hat{s})} + \Big[b_x(t_{i+1}, X_{t_{i+1}}^{(\hat{l},\hat{s})}, u_{t_{i+1}}^{l,M}|_{t_n})^\top Y_{i+1}^{(\hat{l},\hat{s})} \\ &+ \sigma_x(t_{i+1}, X_{t_{i+1}}^{(\hat{l},\hat{s})})^\top Z_{i+1}^{(\hat{l},\hat{s})} + f_x(t_{i+1}, X_{i+1}^{(\hat{l},\hat{s})}, u_{t_{i+1}}^{l,M}|_{t_n})^\top \Big]\Delta t_i, \\ Z_i^{(\hat{l},\hat{s})} =& Y_{i+1}^{(\hat{l},\hat{s})}\omega_i^{(\hat{l},\hat{s})}(\Delta t_i)^{-1/2}, \end{aligned} \quad (37)$$

where the Monte Carlo type approximations for expectations are replaced by a single sample path corresponding to $X_{t_i}^{(\hat{l},\hat{s})}$. Although the schemes (37) does not provide accurate numerical approximations for solutions of FBSDEs—compared with the conventional fully calculated schemes, such like the schemes (25) and schemes in [13,32,47], we want to point out that the conditional expectations in numerical schemes (24) for Y_i and Z_i appear in the expression for the gradient process $\tilde{\mathbb{E}}_{t_n}\big[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})\big]$ in (31). Therefore, the justification for the application of SGD in the gradient descent method also applies to (37). In other words, the purpose of obtaining numerical solutions Y_i and Z_i for the adjoint FBSDEs system is to carry out the gradient descent optimization—not to derive accurate approximations for FBSDEs.

Table 1 PF-SGD algorithm

Algorithm 1: *PF-SGD algorithm for data driven feedback control problem*

Initialize the particle cloud $\{x_0^{(s)}\}_{s=1}^S \sim \xi$ and the number of iteration $L \in \mathbb{N}$
while $n = 0, 1, 2, \cdots, N_T$, **do**

Initialize an estimated control process $\{u_{t_i}^{0,M}|_{t_n}\}_{i=n}^{N_T}$ and a step-size ρ;

for SGD iteration steps $l = 0, 1, 2, \cdots, L$,

Simulate one realization of controlled process $\{X_{t_{i+1}}^{(\hat{l},\hat{s})}\}_{i=n}^{N_T-1}$ through scheme (36) with $X_{t_n}^{(\hat{l},\hat{s})} = x_n^{(\hat{s})} \in \{x_n^{(s)}\}_{s=1}^S$;

Calculate solution $\{Y_i^{(\hat{l},\hat{s})}\}_{i=N_T}^n$ of the FBSDEs system (14) corresponding to $\{X_{t_{i+1}}^{(\hat{l},\hat{s})}\}_{i=n}^{N_T-1}$ through schemes (37);

Update the control process to obtain $\{u_{t_i}^{l+1,M}|_{t_n}\}_{i=n}^{N_T}$ through scheme (35);

end for

The estimated optimal control is given by $\hat{u}_{t_n}^* = u_{t_n}^{L,M}|_{t_n}$;

Propagate particles through the particle filter algorithm (28) - (30) to obtain $\{x_{n+1}^{(s)}\}_{s=1}^S$ by using the estimated optimal control $\hat{u}_{t_n}^*$.

end while

In this way, the PF-SGD schemes (35)–(37) combined together transfer our computational efforts in solving FBSDEs to optimization iterations of finding the optimal control.

3.4 Summary of the Numerical Approach

In *Algorithm 1* (Table 1), we summarize our PF-SGD algorithm for the data driven feedback control problem (C*).

The methodology of our approach is based on the formulation of a conditional optimization (13), which provides the "best approximation" for the data driven feedback control u^* given the current observational information. The general computational framework combines (a) numerical schemes for FBSDEs, which provide numerical approximation for the gradient process with respect to the optimal control, and (b) the optimal filtering algorithm for the conditional distribution of the controlled process, which is used to computing the conditional expectation of the gradient process. Then, we adopt the concept of SGD and use a single-realization of simulated controlled process to represent expectations in our computational framework. The contribution of the PF-SGD algorithm is that every realization of particle based simulated controlled process $X_{t_i}^{(\hat{l},\hat{s})}$, as well as its corresponding FBSDEs solution pair $(Y_i^{(\hat{l},\hat{s})}, Z_i^{(\hat{l},\hat{s})})$, is effectively used to search for the optimal control, which significantly saves unnecessary computational cost on solving the optimal control problem in the entire controlled state space.

4 Numerical Experiments

In this section, we present three numerical examples to demonstrate the effectiveness and efficiency of our data driven feedback control algorithm. The processor that we use to run the numerical experiments is an Intel Core i7 with 2.5 GHz Dual-Core. In the first example, we demonstrate the effectiveness of our computational framework by solving a classic linear-quadratic (LQ) optimal control problem with *nonlinear* observations. If the real state of controlled process is given, its corresponding optimal control process can be derived analytically. Therefore, this example could be a benchmark test that compares the data driven optimal control calculated from our algorithm with the analytical optimal control process, and we want to use this example to show that our algorithm could accurately provide the optimal control process based on indirect observations of the controlled process. In the second example, we demonstrate the efficiency of the PF-SGD implementation of our data driven feedback control framework by solving a nonlinear optimal control problem (with non-quadratic cost functional), which can not be solved analytically. The benchmark method that we use in this example implements our data driven feedback control framework with the classic Zakai filter to calculate the conditional pdf for the controlled state [8] and the gradient decent optimization with fully calculated FBSDEs to solve the corresponding optimal control problem [21]. Both the Zakai filter and the gradient descent method are well studied with rigorous numerical analysis. In example 3, we solve a practical application problem—the Dubins vehicle problem, in which we control a car-like robot to reach a given target platform in the designated arrival time [23].

Example 1 Consider the following controlled process

$$dX_t = A(t)X_t dt + BU_t dt + CdW_t, \quad X_0 = x_0, \tag{38}$$

where $A(t)$, B and C are given coefficients, X_t is the multi-dimensional controlled process with the given initial position x_0 and U_t is the multi-dimensional control process and the cost function J is defined by

$$J(U) = E\Big[\frac{1}{2}\int_0^T \Big(\langle QX_t, X_t\rangle + \langle RU_t, U_t\rangle\Big)dt + \frac{1}{2}\langle FX_T, X_T\rangle\Big]$$

The optimal control U_t for the above stochastic optimal control problem can be derived analytically as

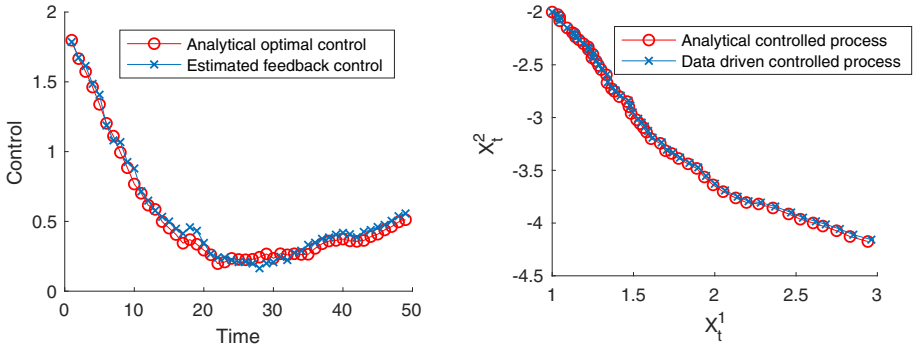
$$\bar{U}_t = -R^{-1}B^T P(t)X_t, \tag{39}$$

where $P(t)$ is the unique solution of the following Riccati equation

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

We can see from (39) that the optimal control $\bar{U}$ depends on the state of the controlled process, which forms a feedback control problem. In our experiments, we assume that the exact state of the controlled process X_t is not known and we use nonlinear indirect observations $M_t = \sin(X_t) + \eta_t$ to collect measurement data about X_t, where η_t is a Gaussian random noise with covariance Γ. Therefore, the data driven feedback control has the cost functional $J^*(U)$

$$J^*(U^M) = E\Big[\frac{1}{2}\int_0^T \Big(\langle QX_t, X_t\rangle + \langle RU_t^M, U_t^M\rangle\Big)dt + \frac{1}{2}\langle FX_T, X_T\rangle\Big], \tag{40}$$



(a) Estimation for the control process **(b)** Estimation for the controlled process

Fig. 1 Comparison between empirical representation for PDF u and Shepard's approximation for PDF u

where U^M is a data driven feedback control based on observations M and we try to find U^* such that $J^*(U^*) := \min_{U^M \in \mathcal{U}_{ad}} J^*(U^M)$.

In this example, we first solve the problem by choosing the controlled process as a 2D system with 1D control process. Specifically, let $A(t) = (2\sin(t), \cos(t)) \cdot I_2$ $B = (0.5, 0.5) \cdot I_2$ and $C = 0.1 I_2$ in (38). For the cost functional, we let $Q = I_2$, $R = 1$ and $F = I_2$. We solve the data driven feedback control problem over the time interval $[0, 1]$ by choosing the temporal step-size $\Delta t = 0.02$, i.e. $N_T = 50$, and we let Γ in the observational noise be $\Gamma = (0.1)^2 I_2$. The initial state of the controlled process is chosen as $x_0 = (1, -2)^T$.

In Fig. 1, we demonstrate the effectiveness of our data driven feedback control algorithm by comparing with the analytical solution, where the analytical optimal control is calculated from (39) with the *exact controlled state* and we let the analytical optimal control drive its corresponding controlled system through Eq. (38) with synthetically generated Brownian motion W to get the analytical controlled process. The estimated optimal control and its corresponding controlled process are calculated by our data driven feedback control *Algorithm 1* based on observations M with 500 particles in the particle filter to describe the conditional pdf and 1000 iteration steps for the SGD optimization for the optimal control, i.e. $S = 500$, $L = 1000$. In Fig. 1a, we compare our estimated data driven feedback control (blue curve marked by crosses) with the analytical optimal control (red curve marked by circles); and In Fig. 1b, we compare the controlled process driven by our estimated feedback control (blue curve marked by crosses) with the analytical optimal controlled process (red curve marked by circles). We can see from this figure that our algorithm accurately recovers the analytical optimal control process based on partial noisy observations for the controlled process, and the controlled process governed by our data driven feedback control well aligns with the real state of analytical controlled process. To further examine the performance of our algorithm, we solve this 2D feedback control problem repeatedly and compare our estimated control with the analytical control to get the accumulated root mean square errors (RMSEs), i.e.

$$Err_{RMSE} := \sqrt{\frac{1}{M_{rept}} \sum_{m=1}^{M_{rept.}} \sum_{n=1}^{N_T} \| \hat{U}_{t_n}^{(m)} - U_{t_n}^{(m)} \|_{L^2}^2}, \tag{41}$$

where $M_{rept.}$ is the total number of repeated experiments, $\hat{U}_{t_n}^{(m)}$ is our estimated data driven feedback control at the time instant t_n for the m-th repeated experiment and $U_{t_n}^{(m)}$ is the

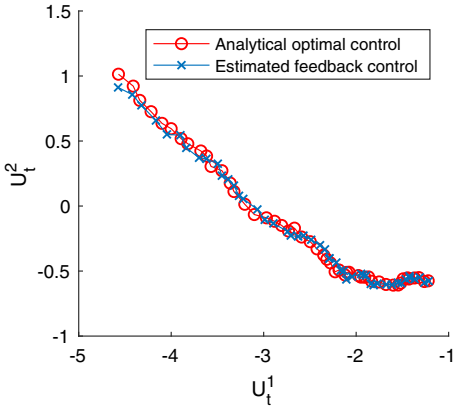


Fig. 2 Estimation for the control process

analytical optimal control corresponding to the m-th realization of the analytical controlled process at time instant t_n. After repeating the above numerical experiments 100 times with different random samples, i.e. $M_{rept.} = 100$, we obtain that $\boldsymbol{Err_{RMSE} = 0.0793}$, which is very low noting that Err_{RMSE} is the accumulated RMSE combining errors of all time instants over the time interval [0, 1], and this confirms the accuracy of our PF-SGD algorithm in solving data driven feedback control problem.

One advantage of our algorithm is that it transfers the computational efforts in solving dimension-dependent stochastic differential systems to the iterations of SGD. This makes our method less sensitive to the dimension of the problem, which would address the curse of dimensionality in solving the data driven optimal control problem. In what follows, we solve a 4D version of the optimal control problem (38)–(40). Specifically, we let $A(t) = (2\sin(t), \cos(t), 1, 0.5) \cdot I_4$ $B = (0.5, 0.5, 1, 1) \cdot I_4$ and $C = 0.1 I_4$ in the controlled process. For the cost functional, we let $Q = I_4$, $R = I_2$ and $F = I_4$. As a result, the controlled process is a 4D dynamical system and the control is a 2D process. We also solve the data driven feedback control problem over the time interval [0, 1] by choosing the temporal step-size $\Delta t = 0.02$, and we let the initial state of the controlled process be $x_0 = (1, 2, -1, 2)^T$.

Similar to the previous experiment, we compare our estimated data driven feedback control (blue curve marked by crosses) with the analytical optimal control (red curve marked by circles) in Fig. 2. To better describe a 4D distribution for the controlled state, we use 1000 particles in the filtering stage of our algorithm and still use 1000 iterations in the SGD optimization. In Fig. 3, we compare the controlled process driven by our estimated control process (blue curve marked by crosses) with the analytical optimal controlled process (red curve marked by circles) in each dimension.

To further examine the performance of our algorithm, we repeat this 4D feedback control problem 100 times with different random samples, i.e. $M_{rept.} = 100$. Then, we calculate the accumulated RMSEs as defined in (41) and get $\boldsymbol{Err_{RMSE} = 0.0649}$, which combines errors on all time instants $\{t_n\}_{n=1}^{50}$.

From the above numerical experiments, we can see that our method could accurately obtain the desired feedback control in both 2D experiments and 4D experiments. On the other hand, the average CPU time for solving the 2D problem is 9.24 s and the average CPU time for solving the 4D problem is 13.51 s, which is not a significant increase although both the dimension of the controlled process and the control process is doubled. Since the only difference between the 2D and 4D problems is the dimension—while they maintain the

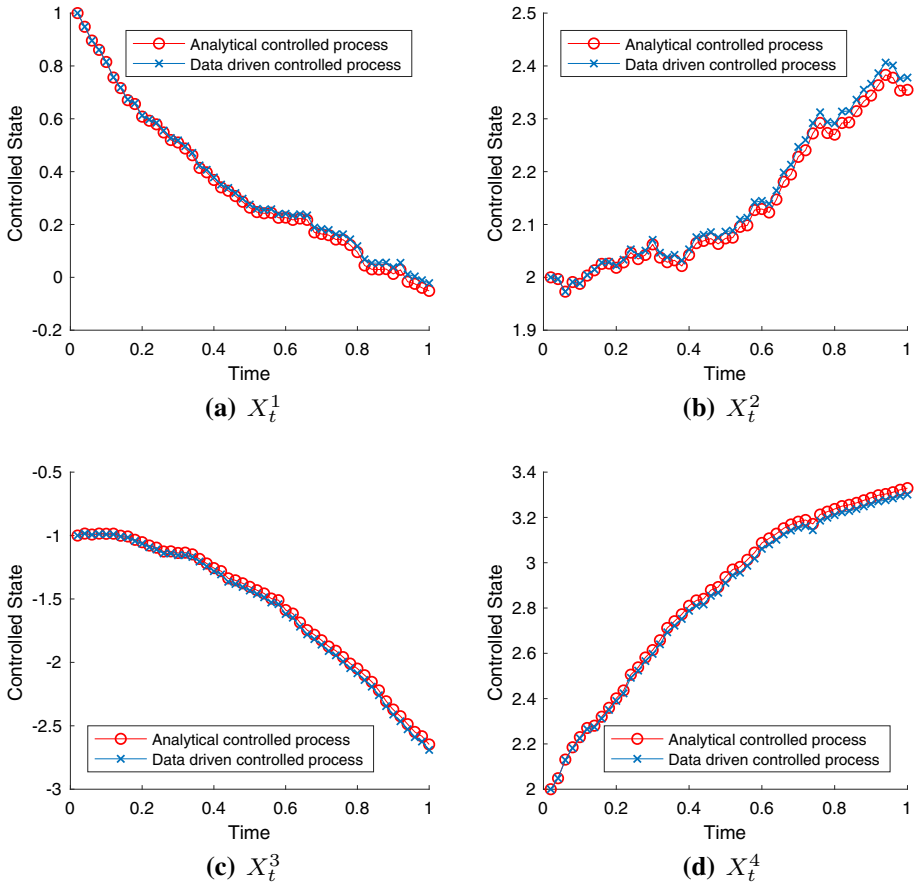


Fig. 3 Estimation for the controlled process

same matrices structure, this CPU time comparison indicates that our algorithm is not very sensitive to the number of dimensions.

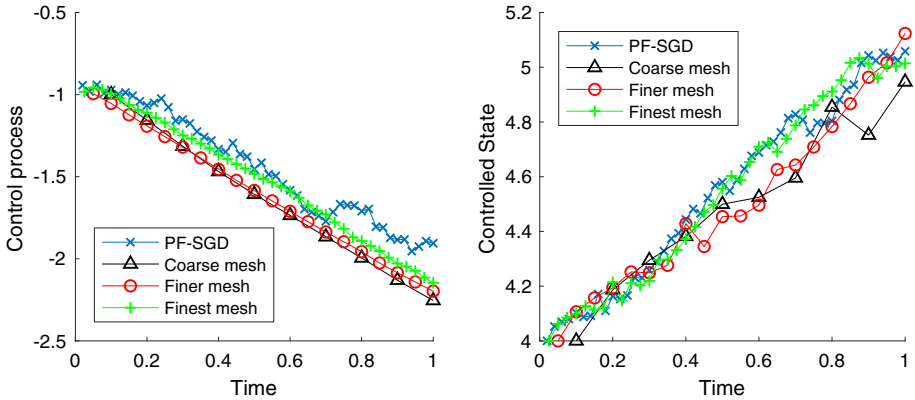
Example 2 With verified effectiveness of our computational framework through numerical experiments in Example 1, in this example we focus on the demonstration of the efficiency. Consider the following 1D nonlinear controlled process

$$dX_t = \arctan(X_t + u_t)dt + \sigma X_t dW_t, \tag{42}$$

where u_t is the control process, σX_t is the coefficient that determines the size of noise perturbing the controlled system and we let $\sigma = 0.05$ in this example. The cost functional we try to minimize is

$$J^*(u^M) = E\Big[\frac{1}{2}\int_0^T \sin^2\big(X_t + u_t^M\big)dt\Big], \tag{43}$$

where the observation process is defined by $M_t = X_t + \eta_t$ with a Gaussian type observational noise η_t and we choose the standard deviation for η_t to be 0.05. We can see that the



(a) Estimated control **(b)** Controlled process

Fig. 4 Estimation for the control process

optimal control problem (42)–(43) is a nonlinear/non-quadratic control problem, which is not explicitly solvable.

In this example, we demonstrate the efficiency of our PF-SGD algorithm by comparing with the conventional benchmark numerical implementation of the data driven feedback control framework (18)–(19), where the conditional pdf for the controlled process, i.e. $p(X_t|\mathcal{F}_t^M)$, is obtained by the Zakai filter through numerical solutions of the Zakai equation, and the gradient process $(J^*)_u$ is calculated by approximate solutions of the adjoint FBSDEs system. Both the Zakai equation and the FBSDEs are fully solved in the controlled state space, hence we call this conventional benchmark approach the *full solution method*. We want to use this one-dimensional problem comparison to show that by avoiding to obtain full solutions of equations in each gradient descent iteration step, the PF-SGD algorithm could solve the data driven feedback control problem (C*) efficiently. Since this is a one-dimensional example, advanced high dimensional numerical techniques are not needed and we use the standard grid mesh based linear interpolation to handle the spatial dimension approximation. In our numerical experiments, we let $T = 1$ and implement the full solution method with a coarse mesh calculation by choosing $\Delta t = 0.1$, i.e. $N_T = 10$, with spatial partition step-size $\Delta x = 0.1$ over the pre-determined spatial interval [3, 6]; we also carry out a finer mesh calculation by choosing $\Delta t = 0.05$, i.e. $N_T = 20$, with spatial partition step-size $\Delta x = \frac{\sqrt{2}}{2} \cdot 0.1$ over [3, 6]; in the finest mesh calculation, we let $\Delta t = 0.025$, i.e. $N_T = 40$, with spatial partition step-size $\Delta x = 0.05$. In this way, for the fixed control period [0, 1], we carry out more frequent control actions when choosing finer grid mesh in the full solution method. On the other hand, the temporal step-size in our PF-SGD algorithm is $\Delta t = 0.02$, , i.e. $N_T = 50$, and we use 500 particles to describe the conditional pdf of X_t and choose the number of iteration to be 1000 in the SGD optimization.

In Fig. 4, we plot the estimated controls (Fig. 4a) and their corresponding controlled processes (Fig. 4b). Specifically, we use the blue curve marked by "crosses" to represent the numerical results obtained by our PF-SGD algorithm. The black curve marked by "triangles" are numerical results obtained by the full solution method with coarse mesh and 10 control actions. The red curve marked by "circles" are numerical results obtained by the full solution method with finer mesh and 20 control actions. The green curve marked by "plus signs" are numerical results obtained by the full solution method with the finest mesh and 40

Table 2 Efficiency comparison

Method	Coarse mesh	Finer mesh	Finest mesh	PF-SGD
Overall cost	0.0481	0.0318	0.0076	0.00095
CPU time (s)	29.78	220.47	1560.15	0.93

control actions. From this figure, we can see that the PF-SGD method could provide more detailed features for the control process which leads to more flexible actions in the controlled process; and when we refine both temporal and spatial step-sizes, the full solution method's evaluation gets closer and closer to the PF-SGD evaluation. To indicate the performance of each algorithm, we compare the overall cost J^* (as defined in (43)) on time interval [0, 1] as well as computing time in Table 2.

We can see from the table that with finer temporal and spatial step-size, the full solution method gets lower cost, which can be explained by the convergence results of the Zakai filter and numerical solutions for FBSDEs, and the PF-SGD has the lowest cost. However, the full solution method spends much more CPU time when carried out with finer partition, and the PF-SGD spends very short time to reach the lowest overall cost in this experiment. The reason that the full solution method is expensive is that we have to solve both the Zakai equation and the FBSDEs system on a large domain in the controlled state space to give *one update* for the estimated optimal control, and thus the computational cost for each gradient descent iteration is high. Moreover, the corresponding computational cost would increase significantly as the dimension of the problem increases, which makes the full solution method prohibitive. On the other hand, as indicated in Example 1, the computational cost of our PF-SGD method is not very sensitive to the number of dimension for the state space, which makes our method even more advantageous in solving high dimensional problems.

To further demonstrate the comparison results, we repeat the above experiment and calculate the average cost $\hat{J}_t^*(\hat{u})$ (with respect to time) defined by

$$\hat{J}_t^*(\hat{u}) := \frac{1}{M_{rept.}} \sum_{m=1}^{M_{rept.}} \frac{1}{2} \int_0^t \sin^2\left(\hat{X}_s^{(m)} + \hat{u}_s^{(m)}\right) ds,$$

where $\hat{u}^{(m)}$ is the calculated control action for the m-th repeated experiment and $\hat{X}^{(m)}$ is the controlled process corresponding to $\hat{u}^{(m)}$.

In Fig. 5, we present the average cost trajectories $\hat{J}_t^*(\hat{u})$ obtained from different numerical methods through $M_{rept.} = 50$ repeated experiments in solving the feedback control problem (42)–(43). The blue curve marked by "crosses" is the cost trajectory of our PF-SGD algorithm over time, the black curve marked by "triangles" is the cost trajectory of the coarse mesh implementation of the full solution method, the red curve marked by "circles" is the cost trajectory of finer mesh implementation of the full solution method, and the green curve marked by "plus signs" is the cost trajectory of the finest mesh implementation of the full solution method. We can see that when solving the problem with finer mesh and more control actions, the conventional full solution method would perform better by achieving lower cost. On the other hand, although the PF-SGD spends much less computing time, the cost trajectory of PF-SGD is much lower than full solution methods, which outperforms the conventional methods in both time and cost.

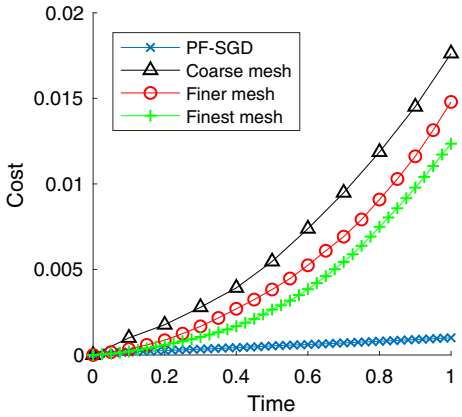


Fig. 5 Comparison of performance cost over time

Example 3 In this example, we solve a Dubins vehicle maneuvering problem. The controlled process is described by the following nonlinear controlled dynamics

$$\begin{aligned} dX_t &= \sin(\theta_t)dt + \sigma dW_t, \\ dY_t &= \cos(\theta_t)dt + \sigma dW_t, \\ d\theta_t &= u_t dt + \sigma^2 dW_t, \end{aligned} \tag{44}$$

where the pair (X, Y) gives the position of a car-like robot moving in the 2D plane, θ is the steering angle that controls the moving direction of the robot, which is governed by the control action u_t, σ is the noise that perturbs the motion and control actions. In our numerical experiments, we choose $\sigma = 0.2$. The performance cost functional based on observational data that we aim to minimize is defined as

$$J^*(u^M) = E\Big[\int_0^T \frac{1}{2}(u_t^M)^2 dt + \delta\Big((X_T - X_P)^2 + (Y_T - Y_P)^2\Big)\Big], \tag{45}$$

where (X_P, Y_P) is a target location for the robot. The goal of this optimal control problem is to let the robot reach the target location at the given terminal time T while trying to minimize the control actions, and we let $\delta = 10$ to give a strong enforcement to emphasize the importance of arriving at the target location at the terminal time. We also want to mention that the speed of the robot is an invariant constant, and the robot has to move appropriately so that it arrives at the target location at the desired time T. In order to make our experiments more practical, we assume that we do not have direct observations on the robot. Instead, we use two detectors located on different observation platforms at $(6, 1)$ and $(-1, 4)$ to collect bearing angles of the target robot as indirect observations. Specifically, the observation process is defined by

$$M_t = \left[\arctan\left(\frac{X_t - 6}{Y_t - 1}\right), \arctan\left(\frac{X_t + 1}{Y_t - 4}\right)\right]^T + \eta_t,$$

where the observational noise η_t is a two dimensional Gaussian random variable with standard deviation $0.01 I_2$.

In our numerical experiments, we let $T = 1$, $\Delta t = 0.02$, i.e. $N_T = 50$, and we choose the target terminal location of the robot to be $(X_P, Y_P) = (5, 3)$. In this way, we process observational data and implement control actions 50 times during the feedback control time period $[0, 1]$. In the PF-SGD algorithm, we use 1000 particles to describe the conditional distribution for the controlled process and choose 1000 iteration steps in SGD optimization.

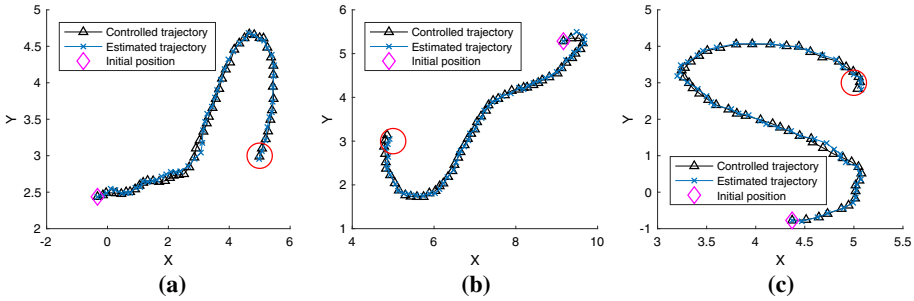


Fig. 6 Controlled trajectories with different initial points

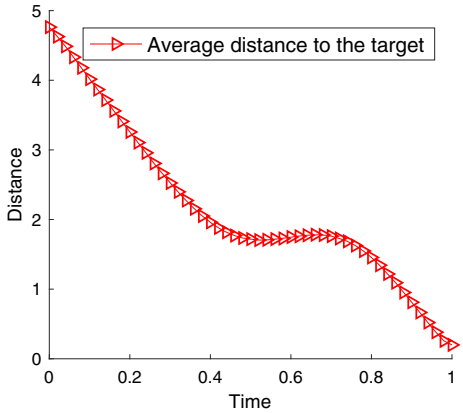


Fig. 7 Estimation for the control process

In Fig. 6, we present trajectories of the controlled car-like robot driven by observational data in three repeated experiments with three different random initial positions/steering angles. The red circle around point (5, 3) is the target at the terminal time; the magenta diamonds mark different initial positions of the robot; the black curves marked by triangles give the real positions of the robot; and the blue curves marked by crosses are estimated positions of the controlled robot from the optimal filter. Due to the high effectiveness of the particle filter method in solving nonlinear filtering problems, we can see from this figure that our estimation for the position of the robot is very accurate. As the main measurement for the performance of data driven optimal control, the maneuvering robot always reaches the target location at the terminal time and when the initial position is too close to the target, the robot lingers around to wait for the *designated* time T.

To provide the robustness of performance of our algorithm in solving this Dubins vehicle maneuvering problem, we repeat the above experiment 50 times with initial position (1, 1) for the robot, which is also perturbed by a standard Gaussian noise; and the initial heading direction is $\pi/2$, which is also perturbed by a Gaussian noise with standard deviation 0.3. The average CPU time to carry out our PF-SGD implementation of data driven feedback control over 50 time steps is 7.47 s, which shows the potential feasibility of our algorithm in real-time feedback control tasks.

In Fig. 7, we plot the average Euclidean distance between the robot and the target point (5, 3) by the red curve marked by triangles. We can see from this figure that the robot moves towards the target platform and is "on target" at the final time with a very small error.

5 Conclusions and Future Work

In this paper, we have introduced an efficient numerical algorithm to solve the data driven feedback control problem, in which the optimal control is driven by indirect observations on the (controlled) state process. The general computational framework is designed by using the particle filter method to estimate conditional expectations in the maximum principle type optimal control solver, and the optimization procedure for the control process is carried out by stochastic optimization. Numerical experiments are presented to demonstrate both the effectiveness and efficiency of our algorithm. In the future, we plan to carry out numerical analysis to derive convergence and efficiency theorems of our algorithm. We shall also study optimization properties, such like the necessity of convexity assumption in optimization, for the application of SGD in our computational framework. In many practical scenarios, due to the transportation of data and the computational implementation of algorithms, we also expect latency between the data generation and control actions. Therefore, we plan to incorporate the latency into the optimal control and extend the current computational framework to address delay issues in data driven feedback control problems in the future.

Acknowledgements 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.

Appendix: Derivation for the Gradient Process

In this "Appendix", we give a detailed discussion on the derivation for the gradient process (8). To proceed, let U be convex and (X^*, u^*) be any state-control pair which could be an optimal pair. For any $u^M \in \mathcal{U}_{ad}[0, T]$, we let $u^{\varepsilon,M} = u^* + \varepsilon(u^M - u^*)$. Then the Gâteaux derivative of $u^M \mapsto J^*(u^M)$ at u^* is given by the following:

$$\lim_{\varepsilon\to 0} \frac{J^*(u^{\varepsilon,M}) - J^*(u^*)}{\varepsilon} = \mathbb{E}\Big[\int_0^T \big(f_x^*(t)\mathcal{D}X_t + f_u^*(t)[u_t^M - u_t^*]\big)dt + h_x\mathcal{D}X_T\Big], \quad (46)$$

where

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

with $\mathcal{D}X_0 = 0$. For convenience of presentation, we denote $\psi^*(t) := \psi(t, X_t^*, u_t^*)$ and use subscript to denote partial derivative of a function. Let (Y, Z, ζ) be the adapted solution to the following *adjoint* backward stochastic differential equation (BSDE)

$$dY_s = \Big(-b_x^*(s)^\top Y_s - \sigma_x^*(s)^\top Z_s - g_x^*(s)^\top\Big)ds + Z_s dW_s + \zeta_s dB_s, \qquad Y_T = (h_x^*)^\top$$

with $h_x^* := h_x(X_T^*)$, where Z is the martingale representation of Y with respect to W and ζ is the martingale representation of Y with respect to B. Then

$$\begin{aligned}
h_x^* \mathcal{D}X_T &= \langle Y_T, \mathcal{D}X_T\rangle - \langle Y_0, \mathcal{D}X_0\rangle \\
&= \int_0^T \Big(\langle -b_x^*(t)^\top Y_t - \sigma_x^*(t)^\top Z_t - f_x^*(t)^\top, \mathcal{D}X_t\rangle \\
&\qquad + \langle Y_t, b_x^*(t)\mathcal{D}X_t + b_u^*(t)[u_t^M - u_t^*]\rangle \\
&\qquad + \langle Z_t, \sigma_x^*(t)\mathcal{D}X_t + \sigma_u^*(t)[u_t^M - u_t^*]\rangle\Big)ds \\
&\qquad + \int_0^T \Big(\langle Z_t, \mathcal{D}X_t\rangle + \langle Y_t, \sigma_x^*(t)\mathcal{D}X_t + \sigma_u^*(t)[u_t^M - u_t^*]\Big)dW_t \\
&\qquad + \int_0^T \langle \zeta_t, \mathcal{D}X_t + \sigma_u^*(t)[u_t^M - u_t^*]\rangle dB_t \\
&= \int_0^T \Big(- f_x^*(t)\mathcal{D}X_t + \langle b_u^*(t)^\top Y_t + \sigma_u^*(t)^\top Z_t, u_t^M - u_t^*\rangle\Big)dt \\
&\qquad + \int_0^T \Big(\langle Z_t, \mathcal{D}X_t\rangle + \langle Y_t, \sigma_x^*(t)\mathcal{D}X_t\rangle + \sigma_u^*(t)[u_t^M - u_t^*]\Big)dW_t \\
&\qquad + \int_0^T \langle \zeta_t, \mathcal{D}X_t + \sigma_u^*(t)[u_t^M - u_t^*]\rangle dB_t.
\end{aligned}$$

Substituting the above equation into the right hand side of (46), we obtain

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

Here, we have used that $u_t^M - u_t^*$ is $\mathcal{F}_t^M$-measurable. Hence, in the case that u^* is in the interior of U, one has

$$(J^*)'_u(u_t^*) = \mathbb{E}\big[b_u^*(t)^\top Y_t + \sigma_u^*(t)^\top Z_t + f_u^*(t)^\top \mid \mathcal{F}_t^M\big], \quad t \in [0, T],$$

as required in (8), where Y and Z are solutions of the FBSDE system (9).

References

1. Archibald, R., Bao, F., Yong, J.: Stochastic gradient descent approach for stochastic optimal control. East Asian J. Appl. Math. **10**(4), s635–658 (2020)
2. Bally, V.: Approximation scheme for solutions of BSDE. In: Backward Stochastic Differential Equations (Paris, 1995–1996). Pitman Research Notes in Mathematics Series, vol. 64, pp. 177–191. Longman, Harlow (1997)
3. Bao, F., Archibald, R., Maksymovych, P.: Backward SDE filter for jump diffusion processes and its applications in material sciences. Commun. Comput. Phys. **27**, 589–618 (2020)

4. Bao, F., Cao, Y., Meir, A., Zhao, W.: A first order scheme for backward doubly stochastic differential equations. SIAM/ASA J. Uncertain. Quantif. **4**(1), 413–445 (2016)
5. Bao, F., Cao, Y., Zhao, W.: Numerical solutions for forward backward doubly stochastic differential equations and Zakai equations. Int. J. Uncertain. Quantif. **1**(4), 351–367 (2011)
6. Bao, F., Cao, Y., Zhao, W.: A first order semi-discrete algorithm for backward doubly stochastic differential equations. Discrete Contin. Dyn. Syst. Ser. B **5**, 1297–1313 (2015)
7. Bao, F., Cao, Y., Zhao, W.: A backward doubly stochastic differential equation approach for nonlinear filtering problems. Commun. Comput. Phys. **23**(5), 1573–1601 (2018)
8. Bao, F., Cao, Y., Webster, C., Zhang, G.: A hybrid sparse-grid approach for nonlinear filtering problems based on adaptive-domain of the Zakai equation approximations. SIAM/ASA J. Uncertain. Quantif. **2**(1), 784–804 (2014)
9. Bao, F., Maroulas, V.: Adaptive meshfree backward SDE filter. SIAM J. Sci. Comput. **39**(6), A2664–A2683 (2017)
10. Baras, J.S., Elliott, R.J., Kohlmann, M.: The partially observed stochastic minimum principle. SIAM J. Control Optim. **27**(6), 1279–1292 (1989)
11. Bensoussan, A.: Maximum principle and dynamic programming approaches of the optimal control of partially observed diffusions. Stochastics **9**(3), 169–222 (1983)
12. Bensoussan, A.: Stochastic Control of Partially Observable Systems. Cambridge University Press, Cambridge (1992)
13. Bouchard, B., Touzi, N.: Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. Stoch. Process. Appl. **111**(2), 175–206 (2004)
14. Charalambous, C.D., Elliott, R.J.: Classes of nonlinear partially observable stochastic optimal control problems with explicit optimal control laws. SIAM J. Control Optim. **36**(2), 542–578 (1998)
15. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, New York (2001)
16. Exarchos, I., Theodorou, E.A.: Stochastic optimal control via forward and backward stochastic differential equations and importance sampling. Automatica **87**, 159–165 (2018)
17. Exarchos, I., Theodorou, E.A., Tsiotras, P.: Stochastic l1 optimal control via forward and backward sampling. Syst. Control Lett. **118**, 101–108 (2018)
18. Feng, X., Glowinski, R., Neilan, M.: Recent developments in numerical methods for fully nonlinear second order partial differential equations. SIAM Rev. **55**(2), 205–267 (2013)
19. Fleming, W.H.: Optimal control of partially observable diffusions. SIAM J. Control **6**, 194–214 (1968)
20. Fleming, W.H., Pardoux, É.: Optimal control for partially observed diffusions. SIAM J. Control Optim. **20**(2), 261–285 (1982)
21. Gong, B., Liu, W., Tang, T., Zhao, W., Zhou, T.: An efficient gradient projection method for stochastic optimal control problems. SIAM J. Numer. Anal. **55**(6), 2982–3005 (2017)
22. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F **140**(2), 107–113 (1993)
23. Gorodetsky, A., Karaman, S., Marzouk, Y.: High-dimensional stochastic optimal control using continuous tensor decompositions. Int. J. Robot. Res. **37**(2–3), 340–377 (2018)
24. Han, J., Jentzen, A., Weinan, E.: Solving high-dimensional partial differential equations using deep learning. Proc. Natl. Acad. Sci. USA **115**(34), 8505–8510 (2018)
25. Haussmann, U.G.: On the existence of optimal controls for partially observed diffusions. SIAM J. Control Optim. **20**(3), 385–407 (1982)
26. Haussmann, U.G.: The maximum principle for optimal control of diffusions with partial information. SIAM J. Control Optim. **25**(2), 341–361 (1987)
27. Kang, K., Maroulas, V., Schizas, I., Bao, F.: Improved distributed particle filters for tracking in a wireless sensor network. Comput. Stat. Data Anal. **117**, 90–108 (2018)
28. Kwakernaak, H.: A minimum principle for stochastic control problems with output feedback. Syst. Control Lett. **1**(1), 74–77 (1981/1982)
29. Li, Q., Tai, C., Weinan, E.: Stochastic modified equations and dynamics of stochastic gradient algorithms i: mathematical foundations. J. Mach. Learn. Res. **20**, 40–47 (2019)
30. Li, X., Tang, S.: General necessary conditions for partially observed optimal stochastic controls. J. Appl. Probab. **32**(4), 1118–1137 (1995)
31. Lindquist, A.: On feedback control of linear stochastic systems. SIAM J. Control **11**, 323–343 (1973)
32. Ma, J., Protter, P., Yong, J.: Solving forward–backward stochastic differential equations explicitly–a four step scheme. Probab. Theory Relat. Fields **98**(3), 339–359 (1994)
33. Medjghou, A., Ghanai, M., Chafaa, K.: Improved feedback linearization control based on PSO optimization of an extended Kalman filter. Optim. Control Appl. Methods **39**(6), 1871–1886 (2018)

34. Milstein, G.N., Tretyakov, M.V.: Numerical algorithms for forward–backward stochastic differential equations. SIAM J. Sci. Comput. **28**(2), 561–582 (2006)
35. Morzfeld, M., Tu, X., Atkins, E., Chorin, A.J.: A random map implementation of implicit filters. J. Comput. Phys. **231**(4), 2049–2066 (2012)
36. Peng, S.: A general stochastic maximum principle for optimal control problems. SIAM J. Control Optim. **28**(4), 966–979 (1990)
37. Sassano, M., Astolfi, A.: A local separation principle via dynamic approximate feedback and observer linearization for a class of nonlinear systems. IEEE Trans. Autom. Control **64**(1), 111–126 (2019)
38. Shapiro, A., Wardi, Y.: Convergence analysis of gradient descent stochastic algorithms. J. Optim. Theory Appl. **91**, 439–454 (1996)
39. Smears, I., Süli, E.: Discontinuous Galerkin finite element methods for time-dependent Hamilton-Jacobi-Bellman equations with Cordes coefficients. Numer. Math. **133**(1), 141–176 (2016)
40. Tang, S.: The maximum principle for partially observed optimal control of stochastic differential equations. SIAM J. Control Optim. **36**(5), 1596–1617 (1998)
41. Touzi, N.: Optimal stochastic control, stochastic target problems, and backward SDE. In: Fields Institute Monographs, vol. 29. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON. With Chapter 13 by Angès Tourin (2013)
42. Wang, G., Wu, Z., Xiong, J.: An Introduction to Optimal Control of FBSDE with Incomplete Information. Springer Briefs in Mathematics. Springer, Cham (2018)
43. Wonham, W.M.: On the separation theorem of stochastic control. SIAM J. Control **6**, 312–326 (1968)
44. Yong, J., Zhou, X.Y.: Stochastic controls. Applications of Mathematics (New York), vol. 43. Springer, New York. Hamiltonian systems and HJB equations (1999)
45. Zakai, M.: On the optimal filtering of diffusion processes. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **11**, 230–243 (1969)
46. Zhao, W., Chen, L., Peng, S.: A new kind of accurate numerical method for backward stochastic differential equations. SIAM J. Sci. Comput. **28**(4), 1563–1581 (2006)
47. Zhao, W., Fu, Y., Zhou, T.: New kinds of high-order multistep schemes for coupled forward backward stochastic differential equations. SIAM J. Sci. Comput. **36**(4), A1731–A1751 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.