

Article

Convergence Analysis for an Online Data-Driven Feedback Control Algorithm

Siming Liang ^{1,*}, Hui Sun ², Richard Archibald ³ and Feng Bao ¹

¹ Department of Mathematics, Florida State University, Tallahassee, FL 32304, USA; fbao@fsu.edu

² Citigroup Inc., Wilmington, DE 19801, USA; hui1.sun@citi.com

³ Devision of Computational Science and Mathematics, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA; archibaldrk@ornl.gov

* Correspondence: sliang@fsu.edu

Abstract: This paper presents convergence analysis of a novel data-driven feedback control algorithm designed for generating online controls based on partial noisy observational data. The algorithm comprises a particle filter-enabled state estimation component, estimating the controlled system's state via indirect observations, alongside an efficient stochastic maximum principle-type optimal control solver. By integrating weak convergence techniques for the particle filter with convergence analysis for the stochastic maximum principle control solver, we derive a weak convergence result for the optimization procedure in search of optimal data-driven feedback control. Numerical experiments are performed to validate the theoretical findings.

Keywords: stochastic optimal control; nonlinear filtering; data driven; maximum principle; stochastic optimization

MSC: 49N35



Citation: Liang, S.; Sun, H.; Archibald, R.; Bao, F. Convergence Analysis for an Online Data-Driven Feedback Control Algorithm. *Mathematics* **2024**, *12*, 2584. <https://doi.org/10.3390/math12162584>

Academic Editors: Mudassir M. Rashid and Zukui Li

Received: 10 July 2024

Revised: 9 August 2024

Accepted: 19 August 2024

Published: 21 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this paper, we carry out numerical analysis demonstrating the convergence of a data-driven feedback control algorithm designed for generating online control based on partial noisy observational data.

Our focus lies on the stochastic feedback control problem, which aims to determine optimal controls. These control actions are used to guide a controlled state dynamical system towards meeting certain optimality conditions, leveraging feedback from the system's current state. There are two practical challenges in solving the feedback control problem. First, when the control problem's dimension is high, the computational cost for searching the optimal control escalates exponentially. This is known as the "curse of dimensionality". Second, in numerous scenarios, the state of the controlled system is not directly observable and must be inferred through detectors or observation facilities. These sensors are typically subject to noise that originates from the device itself or the surrounding environment. For instance, radar receives noisy data and processes them through the arctangent function. Therefore, state estimation techniques become necessary to estimate the current state for designing optimal control, with observations gathered to aid in estimating the hidden state.

To address the aforementioned challenges, a novel online data-driven feedback control algorithm has been developed [1]. This algorithm introduces a stochastic gradient descent optimal control solver within the stochastic maximum principle framework to combat the high dimensionality issue in optimal control problems. Traditionally, stochastic optimal control problems are solved using dynamical programming or the stochastic maximum principle, both requiring numerical simulations for large differential systems [2–4]. However, the stochastic maximum principle stands out for its capability to handle random coefficients in the state model and finite-dimensional terminal state constraints [5]. In the

stochastic maximum principle approach, a system of backward stochastic differential equations (BSDEs) is derived as the adjoint equation of the controlled state process. Then, the solution of the adjoint BSDE is utilized to formulate the gradient of the cost function with respect to the control process [6,7]. However, solving BSDEs numerically entails significant computational costs, especially in high-dimensional problems, demanding a large number of random samples [8,9]. To bolster efficiency, a sample-wise optimal control solver method has been devised [10], where the solution of the adjoint BSDE is represented using only one realization or a small batch of samples. This approach justifies the application of stochastic approximation in the optimization procedure [11,12], and it shifts the computational cost from solving BSDEs to searching for the optimal control, thereby enhancing overall efficiency [13].

In data-driven feedback control, optimal filtering methods also play a pivotal role in dynamically estimating the state of the controlled system. Two prominent approaches for *nonlinear* optimal filtering are the Zakai filter and the particle filter. While the Zakai filter aims to compute the conditional probability density function (pdf) for the target dynamical system using a parabolic-type stochastic partial differential equation known as the Zakai equation [14], the particle filter, also known as a sequential Monte Carlo method, approximates the desired conditional pdf using the empirical distribution of a set of random samples (particles) [15]. Although the Zakai filter theoretically offers more accurate approximations for conditional distributions, the particle filter is favored in more practical applications due to the high efficiency of the Monte Carlo method in approximating high-dimensional distributions [16].

The aim of this study is to examine the convergence of the data-driven feedback control algorithm proposed in [1], providing mathematical validation for its performance. While convergence in particle filter methods has been well studied [17–19], this work adopts the analysis technique outlined in [18] to establish weak convergence results for the particle filter regarding the number of particles. Analysis techniques for BSDEs alongside classical convergence results for stochastic gradient descent [13,20] are crucial for achieving convergence in the stochastic gradient descent optimal control solver. The theoretical framework of this analysis merges the examination of particle filters with the analysis of optimal control, and the overarching objective of this paper is to derive a comprehensive weak convergence result for the optimal data-driven feedback control.

In this paper, we present two numerical examples to demonstrate the baseline performance and convergence trend of our algorithm. The first example involves a classic linear quadratic optimal control problem, comparing the analytical control with the estimated control. The second example addresses a nonlinear scenario, specifically a Dubins vehicle maneuvering problem, where both the system and observations exhibit significant nonlinearity.

The rest of this paper is organized as follows. In Section 2, we introduce the data-driven feedback control algorithm. Convergence analysis will be presented in Section 3, and in Section 4, we conduct two numerical experiments to validate our theoretical findings.

2. An Efficient Algorithm for Data-Driven Feedback Control

We first briefly introduce the data-driven feedback control problem that we consider in this work. Then, we describe our efficient algorithm for solving the data-driven feedback control problem by using a stochastic gradient descent-type optimization procedure for the optimal control.

2.1. Problem Setting for the Data-Driven Optimal Control Problem

In probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider the following augmented system on time interval $[0, T]$

$$d \begin{pmatrix} X_t \\ M_t \end{pmatrix} = \begin{pmatrix} b(t, X_t, u_t) \\ g(X_t) \end{pmatrix} dt + \begin{pmatrix} \sigma(t, X_t, u_t) & 0 \\ 0 & I \end{pmatrix} d \begin{pmatrix} W_t \\ B_t \end{pmatrix}, \quad \begin{pmatrix} X_0 = \xi \\ M_0 = 0 \end{pmatrix}, \quad (1)$$

where $X := \{X_t\}_{t=0}^T$ is the \mathbb{R}^d -dimensional controlled state process with dynamics $b : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^{d \times q}$ is the diffusion coefficient for a d -dimensional Brownian motion W that perturbs the state X , and u is an m -dimensional control process valued in some set U that controls the state process X . In the case that the state X is not directly observable, we have an observation process M that collects partial noisy observations on X with observation function $g : \mathbb{R}^d \rightarrow \mathbb{R}^p$, and B is a p -dimensional Brownian motion that is independent from W .

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), Equation (1) 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. The \mathbb{F}^M progressively measurable control processes, denoted by u^M , are control actions driven by the information contained in observational data.

We introduce the set of data-driven admissible controls as

$$\mathcal{U}_{ad}[0, T] = \left\{ u^M : [0, T] \times \Omega \rightarrow U \subset \mathbb{R}^m \mid u^M \text{ is } \mathbb{F}^M\text{-progressively measurable} \right\},$$

and the cost functional that measures the performance of data-driven control u^M is defined as

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

where f is the running cost, and h is the terminal cost.

The goal of the data-driven feedback control problem is to find the optimal data-driven control $u^* \in \mathcal{U}_{ad}[0, T]$ such that

$$J(u^*) = \inf_{u^M \in \mathcal{U}_{ad}[0, T]} J(u^M). \quad (3)$$

2.2. The Algorithm for Solving the Data-Driven Optimal Control Problem

To solve the data-driven feedback control problem, we will use the algorithm from [1], which is derived from the stochastic maximum principle.

2.2.1. The Optimization Procedure for Optimal Control

When the optimal control u^* is in the interior of \mathcal{U}_{ad} , the gradient process of the cost functional J^* with respect to the control process on time interval $t \in [0, T]$ can be derived using the Gâteaux derivative of u^* and the stochastic maximum principle in the following form:

$$(J^*)'_u(u_t^*) = E \left[b_u(t, X_t^*, u_t^*)^\top Y_t + \sigma_u(t, X_t^*, u_t^*)^\top Z_t + f_u(t, X_t^*, u_t^*)^\top \mid \mathcal{F}_t^M \right], \quad (4)$$

where stochastic processes Y and Z are solutions of the following forward–backward stochastic differential equations (FBSDEs) system:

$$\begin{cases} dX_t^* = b(t, X_t^*, u_t^*) dt + \sigma(t, X_t^*, u_t^*) dW_t, & X_0 = \xi \\ dM_t^* = \phi(X_t) dt + dB_t, & M_0 = 0 \\ dY_t = (-b_x(t, X_t^*, u_t^*)^\top Y_t - \sigma_x(t, X_t^*, u_t^*)^\top Z_t - f_x(t, X_t^*, u_t^*)^\top) dt \\ \quad + Z_t dW_t + \zeta_t dB_t, & Y_T = (g_x(X_T))^\top \end{cases} \quad (5)$$

where Z is the martingale representation of Y with respect to W and ζ is the martingale representation of Y with respect to B .

To solve the data-driven feedback optimal control problem, we also use gradient descent-type optimization, and the gradient process $(J^*)'_u$ is defined in (4). Then, we can use the following gradient descent iteration to find the optimal control u_t^* at any time instant $t \in [0, T]$

$$u_t^{l+1,M} = u_t^{l,M} - r(J^*)'_u(u_t^{l,M}), \quad l = 0, 1, 2, \dots, \quad (6)$$

where r is the step size for the gradient. We know that the observational information \mathcal{F}_t^M is gradually increased as we collect more and more data over time. Therefore, at a certain time instant t , we target finding the optimal control u_t^* with accessible information \mathcal{F}_t^M . Since the evaluation for $(J^*)'_u(u_t^{l,M})$ requires trajectories $(Y_s, Z_s)_{t \leq s \leq T}$ as Y_t and Z_t are solved backwards from T to t , we take conditional expectation $E[\cdot | \mathcal{F}_t^M]$ to the gradient process $\{(J^*)'_u(u_t^{l,M})\}_{t \leq s \leq T}$, i.e.,

$$\begin{aligned} E[(J^*)'_u(u_s^{l,M}) | \mathcal{F}_t^M] &= E \left[b_u(s, X_s, u_s^{l,M})^\top Y_s + \sigma_u(s, X_s, u_s^{l,M})^\top Z_s \right. \\ &\quad \left. + f_u(s, X_s, u_s^{l,M})^\top | \mathcal{F}_t^M \right], \quad s \in [t, T], \end{aligned} \quad (7)$$

where X_s , Y_s and Z_s correspond to the estimated control $u_s^{l,M}$. For the gradient descent iteration (6) on the time interval $[t, T]$, by taking conditional expectation $E[\cdot | \mathcal{F}_t^M]$, we obtain

$$E[u_s^{l+1,M} | \mathcal{F}_t^M] = E[u_s^{l,M} | \mathcal{F}_t^M] - rE[(J^*)'_u(u_s^{l,M}) | \mathcal{F}_t^M], \quad l = 0, 1, 2, \dots \quad s \in [t, T]. \quad (8)$$

When $s > t$, the observational information $\{\mathcal{F}_s^M\}_{t \leq s \leq T}$ is not available at time t . We use conditional expectation $E[u_s^{l+1,M} | \mathcal{F}_t^M]$ to replace $u_s^{l,M}$ since it provides the best approximation for $u_s^{l,M}$ given the current observational information \mathcal{F}_t^M . We denote

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

and then the gradient descent iteration is

$$u_s^{l+1,M}|_t = u_s^{l,M}|_t - rE[(J^*)'_u(u_s^{l,M}|_t) | \mathcal{F}_t^M], \quad l = 0, 1, 2, \dots \quad s \in [t, T], \quad (9)$$

where $E[(J^*)'_u(u_s^{l,M}|_t) | \mathcal{F}_t^M]$ can be obtained by solving the following FBSDEs

$$\begin{aligned} dX_t &= b(s, X_s, u_s^{l,M}|_t)ds + \sigma(s, X_s, u_s^{l,M}|_t)dW_s, & s \in [t, T] \\ dY_t &= (-b_x(s, X_s, u_s^{l,M}|_t)^\top Y_s - \sigma_x(s, X_s, u_s^{l,M}|_t)^\top Z_s - f_x(s, X_s, u_s^{l,M}|_t)^\top)ds \\ &\quad + Z_s dW_s + \zeta_s dB_s, & Y_T = (g_x(X_T))^\top \end{aligned} \quad (10)$$

and evaluated effectively using the numerical algorithm, which will be introduced later.

When the controlled dynamics and the observation function ϕ are nonlinear, we will use optimal filtering techniques to obtain the conditional expectation $E[\Psi(s) | \mathcal{F}_t^M]$. Before applying the particle filter method, which is one of the most important particle-based optimal filtering methods, we define

$$\Psi(s, X_s, u_s^{l,M}|_t) := b_u^\top(s, X_s, u_s^{l,M}|_t)Y_s + \sigma_u^\top(s, X_s, u_s^{l,M}|_t)Z_s + f_u^\top(s, X_s, u_s^{l,M}|_t) \quad (11)$$

for $s \in [t, T]$. With the conditional probability density function (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 $E[(J^*)'_u(u_t^{l,M}|_t) | \mathcal{F}_t^M]$ in (7) can be obtained by the following integral

$$E[(J^*)'_u(u_s^{l,M}|_t) | \mathcal{F}_t^M] = \int_{\mathbb{R}^d} E[\Psi(s, X_s, u_s^{l,M}|_t) | X_t = x] \cdot p(x | \mathcal{F}_t^M) dx, \quad s \in [t, T]. \quad (12)$$

2.2.2. Numerical Approach for Data-Driven Feedback Control by PF-SGD

For the numerical framework, we need the temporal partition Π_{N_T}

$$\Pi_{N_T} = \{t_n, 0 = t_0 < t_1 < \dots < t_{N_T} = T\},$$

and we use the control sequence $\{u_{t_n}^*\}_{n=1}^{N_T}$ to represent the control process u^* over the time interval $[0, T]$.

Numerical Schemes for FBSDEs

For the FBSDEs system, we adopt the following schemes:

$$\begin{aligned} X_{i+1} &= X_i + b(t_i, X_i, u_{t_i}^{l,M}|_{t_n}) \Delta t_i + \sigma(t_i, X_i, u_{t_i}^{l,M}|_{t_n}) \Delta W_{t_i}, \\ Y_i &= E_i[Y_{i+1}] + E_i[b_x(t_{i+1}, X_{i+1}, u_{t_{i+1}}^{l,M}|_{t_n})^\top Y_{i+1} \\ &\quad + \sigma_x(t_{i+1}, X_{i+1}, u_{t_{i+1}}^{l,M}|_{t_n})^\top Z_{i+1} + f_x(t_{i+1}, X_{i+1}, u_{t_{i+1}}^{l,M}|_{t_n})^\top] \Delta t_i, \\ Z_i &= \frac{1}{\Delta t_i} E_i[Y_{i+1} \Delta W_{t_i}], \end{aligned} \quad (13)$$

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.

Then, the standard Monte Carlo method can approximate expectations with K random samples:

$$\begin{aligned} X_{i+1}^k &= X_i + b(t_i, X_i, u_{t_i}^{l,M}|_{t_n}) \Delta t_i + \sigma(t_i, X_i, u_{t_i}^{l,M}|_{t_n}) \sqrt{\Delta t_i} \omega_i^k, k = 1, 2, \dots, K, \\ Y_i &= \sum_{k=1}^K \frac{Y_{i+1}^k}{K} + \frac{\Delta t_i}{K} \sum_{k=1}^K [b_x(t_{i+1}, X_{i+1}^k, u_{t_{i+1}}^{l,M}|_{t_n})^\top Y_{i+1}^k \\ &\quad + \sigma_x(t_{i+1}, X_{i+1}^k, u_{t_{i+1}}^{l,M}|_{t_n})^\top Z_{i+1}^k + f_x(t_{i+1}, X_{i+1}^k, u_{t_{i+1}}^{l,M}|_{t_n})^\top], \\ Z_i &= \frac{1}{\Delta t_i} \sum_{k=1}^K \frac{Y_{i+1}^k \sqrt{\Delta t_i} \omega_i^k}{K}, \end{aligned} \quad (14)$$

where $\{\omega_i^k\}_{k=1}^K$ is a set of random samples following the standard Gaussian distribution that we use to describe the randomness of ΔW_{t_i} .

The above schemes solve the FBSDE system (5) as a recursive algorithm, and the convergence of the schemes is well studied—cf. ([20,21]).

Particle Filter Method for Conditional Distribution

To apply the particle filter method, we consider the controlled process on time interval $[t_{n-1}, t_n]$

$$X_{t_n} = X_{t_{n-1}} + \int_{t_{n-1}}^{t_n} b(s, X_s, u_s) ds + \int_{t_{n-1}}^{t_n} \sigma(s, X_s, u_s) dW_s \quad (15)$$

Assume that at time instant t_{n-1} , we have S particles, denoted by $\{x_{n-1}^{(s)}\}_{s=1}^S$, that follow an empirical distribution $\pi(X_{t_{n-1}} | \mathcal{F}_{t_{n-1}}^M) := \frac{1}{S} \sum_{s=1}^S \delta_{x_{n-1}^{(s)}}(X_{t_{n-1}})$ as an approximation for $p(X_{t_{n-1}} | \mathcal{F}_{t_{n-1}}^M)$. The prior pdf that we want to find in the prediction stage is approximated as

$$\tilde{\pi}(X_{t_n} | \mathcal{F}_{t_{n-1}}^M) := \frac{1}{S} \sum_{s=1}^S \delta_{\tilde{x}_n^{(s)}}(X_{t_n}) \quad (16)$$

where $\tilde{x}_n^{(s)}$ is sampled from $\pi(X_{t_{n-1}} | \mathcal{F}_{t_{n-1}}^M) p(X_{t_n} | X_{t_{n-1}})$ and $p(X_{t_n} | X_{t_{n-1}})$ is the transition probability derived from the state dynamics (15). As a result, the sample cloud $\{\tilde{x}_n^{(s)}\}_{s=1}^S$ provides an approximate distribution for the prior $p(X_{t_n} | \mathcal{F}_{t_{n-1}}^M)$. Then, in the update stage, we have

$$\tilde{\pi}(X_{t_n} | \mathcal{F}_{t_n}^M) := \frac{\sum_{s=1}^S \delta_{\tilde{x}_n^{(s)}}(X_{t_n}) p(M_{t_n} | \tilde{x}_n^{(s)})}{\sum_{s=1}^S p(M_{t_n} | \tilde{x}_n^{(s)})} = \sum_{s=1}^S w_n^{(s)} \delta_{\tilde{x}_n^{(s)}}(X_{t_n}) \quad (17)$$

In this way, we obtain a weighted empirical distribution $\tilde{\pi}(X_{t_n} | \mathcal{F}_{t_n}^M)$ that approximates the posterior pdf $p(X_{t_n} | \mathcal{F}_{t_n}^M)$ with the importance density weight $w_n^{(s)} \propto p(M_{t_n} | \tilde{x}_n^{(s)})$. Then, to avoid the degeneracy problem, we need the resampling step. Thus, we have

$$\pi(X_{t_n} | \mathcal{F}_{t_n}^M) = \frac{1}{S} \sum_{s=1}^S \delta_{x_n^{(s)}}(X_{t_n}) \quad (18)$$

Then, we combine the numerical schemes for the adjoint FBSDEs system (10) and the particle filter algorithm to formulate an efficient stochastic optimization algorithm to solve the optimal control process u^* .

Stochastic Optimization for Control Process

In this subsection, we combine the numerical schemes for the adjoint FBSDEs system (10) and the particle filter algorithm to formulate an efficient stochastic optimization algorithm to solve the optimal control process u^* .

On a time instant $t_n \in \Pi_{N_T}$, we have

$$E[(J^*)'_u(u_{t_i}^{l,M} |_{t_n}) | \mathcal{F}_{t_n}^M] = \int_{\mathbb{R}^d} E[\Psi(t_i, X_{t_i}, u_{t_i}^{l,M} |_{t_n}) | X_{t_n} = x] \cdot p(x | \mathcal{F}_{t_n}^M) dx, \quad (19)$$

where $t_i \geq t_n$ is a time instant after t_n .

Then, we use the approximate solutions (Y_i, Z_i) of FBSDEs from schemes (14) to replace (Y_{t_i}, Z_{t_i}) and the conditional distribution $p(X_{t_n} | \mathcal{F}_{t_n}^M)$ is approximated by the empirical distribution $\pi(X_{t_n} | \mathcal{F}_{t_n}^M)$ obtained from the particle filter algorithm (16)–(18). Then, we can solve the optimal control $u_{t_n}^*$ through the following gradient descent optimization iteration

$$\begin{aligned} u_{t_i}^{l+1,M} |_{t_n} &= u_{t_i}^{l,M} |_{t_n} - r \frac{1}{S} \sum_{s=1}^S E \left[b_u^\top(t_i, X_{t_i}, u_{t_i}^{l,M} |_{t_n}) Y_i + \sigma_u^\top(t_i, X_{t_i}, u_{t_i}^{l,M} |_{t_n}) Z_i \right. \\ &\quad \left. + f_u^\top(t_i, X_{t_i}, u_{t_i}^{l,M} |_{t_n}) | X_{t_n} = x_n^{(s)} \right] \end{aligned} \quad (20)$$

Then, the standard Monte Carlo method can approximate expectation $E[\cdot | X_{t_n} = x_n^{(s)}]$ by Λ samples:

$$\begin{aligned} u_{t_i}^{l+1,M} |_{t_n} &\approx u_{t_i}^{l,M} |_{t_n} - r \frac{1}{S} \frac{1}{\Lambda} \sum_{s=1}^S \sum_{\lambda=1}^{\Lambda} \left[b_u^\top(t_i, X_{t_i}^{(\lambda,s)}, u_{t_i}^{l,M} |_{t_n}) Y_i + \sigma_u^\top(t_i, X_{t_i}^{(\lambda,s)}, u_{t_i}^{l,M} |_{t_n}) Z_i \right. \\ &\quad \left. + f_u^\top(t_i, X_{t_i}^{(\lambda,s)}, u_{t_i}^{l,M} |_{t_n}) | X_{t_n} = x_n^{(s)} \right] \end{aligned} \quad (21)$$

We can see from the above Monte Carlo approximation that in order to approximate the expectation in one gradient descent iteration step, we need to generate $S \times \Lambda$ samples. This is even more computationally expensive when the controlled system is a high-dimensional process.

Thus, we want to apply the idea of stochastic gradient descent (SGD) to improve the efficiency of classic gradient descent optimization and combine it with the particle filter method. Instead of using the fully calculated Monte Carlo simulation to approximate the conditional expectation, we use only one realization of X_{t_i} to represent the expectation. For the conditional distribution of the controlled process, we can use the particles to describe. Thus, we have

$$E[(J^*)'_u(u_{t_i}^{l,M}|_{t_n})|\mathcal{F}_{t_n}^M] \approx b_u^\top(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})Y_i^{(\hat{l},\hat{s})} + \sigma_u^\top(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})Z_i^{(\hat{l},\hat{s})} + f_u^\top(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n}) \quad (22)$$

where l is the iteration index, and the index \hat{l} indicates that the random generation of the controlled process varies among the gradient descent iteration steps. $X_{t_i}^{(\hat{l},\hat{s})}$ 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 $\{x_n^{(s)}\}_{s=1}^S$.

Then, we have the following SGD schemes:

$$u_{t_i}^{l+1,M}|_{t_n} = u_{t_i}^{l,M}|_{t_n} - r \left(b_u^\top(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})Y_i^{(\hat{l},\hat{s})} + \sigma_u^\top(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})Z_i^{(\hat{l},\hat{s})} + f_u^\top(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n}) \right) \quad (23)$$

where $Y_i^{(\hat{l},\hat{s})}$ is the approximate solution Y_i corresponding to the random sample $X_{t_i}^{(\hat{l},\hat{s})}$, and the path of $X_{t_i}^{(\hat{l},\hat{s})}$ is generated as follows

$$X_{t_{i+1}}^{(\hat{l},\hat{s})} = X_{t_i}^{(\hat{l},\hat{s})} + b(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})\Delta t_i + \sigma(t_i, X_{t_i}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})\sqrt{\Delta t_i}\omega_i^{(\hat{l},\hat{s})} \quad (24)$$

where $\omega_i^{(\hat{l},\hat{s})} \sim N(0, 1)$. Then, an estimate for our desired data-driven optimal control at time instant t_n is

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

The scheme for FBSDEs is

$$Y_i^{(\hat{l},\hat{s})} = Y_{i+1}^{(\hat{l},\hat{s})} + \left[b_x(t_{i+1}, X_{t_{i+1}}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top Y_{i+1}^{(\hat{l},\hat{s})} + \sigma_x(t_{i+1}, X_{t_{i+1}}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top Z_{i+1}^{(\hat{l},\hat{s})} + f_x(t_{i+1}, X_{t_{i+1}}^{(\hat{l},\hat{s})}, u_{t_i}^{l,M}|_{t_n})^\top \right] \Delta t_i \quad (25)$$

$$Z_i^{(\hat{l},\hat{s})} = \frac{Y_{i+1}^{(\hat{l},\hat{s})}\omega_i^{(\hat{l},\hat{s})}}{\sqrt{\Delta t_i}}$$

Then, we have the following Algorithm 1:

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 $n = 0, 1, 2, \dots, N_T$, **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 $l = 0, 1, 2, \dots, L$, **do**

 Simulate one realization of controlled process $\{X_{t_{i+1}}^{(\hat{l},\hat{s})}|_{t_n}\}_{i=n}^{N_T-1}$ through scheme (24)

 with

$$\{X_{t_n}^{(\hat{l},\hat{s})}\} = x_n^{(\hat{s})} \in \{x_n^{(s)}\}_{s=1}^S;$$

 Calculate solution $\{Y_{t_i}^{(\hat{l},\hat{s})}|_{t_n}\}_{i=N_T}^n$ of the FBSDEs system (14) corresponding to $\{X_{t_{i+1}}^{(\hat{l},\hat{s})}|_{t_n}\}_{i=n}^{N_T-1}$ through schemes (25);

 Update the control process to obtain $\{u_{t_i}^{(l+1,M)}|_{t_n}\}_{i=n}^{N_T}$ through scheme (23);

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 (16)–(18) to obtain $\{x_{n+1}^{(s)}\}_{s=1}^S$ by using the estimated optimal control $\hat{u}_{t_n}^*$.

end while

3. Convergence Analysis

Our convergence analysis aims to show the convergence of the distribution of the state to the "true state" under the temporal model discretization N . We also show the convergence of the estimated control to the "true control" under the expectation restricting it to a compact set. To proceed, we first introduce our notations and the assumptions required in the proof in Section 3.1. Then, in Section 3.2, we shall provide the main convergence theorems.

3.1. Notations and Assumptions

Notations

1. We use $U_n : \{t_n, \dots, T\} \rightarrow \mathbb{R}^d$ to denote the control process starts from time t_n and ends at time T . We use

$$\mathcal{U}_n := \{U_n | U_n : \{t_n, \dots, T\} \rightarrow \mathbb{R}^d, U_n \text{ is } \mathcal{F}_{t_n}^M\text{-adapted}\}$$

to denote the collection of the admissible controls starting at time t_n .

2. We define the control at time t_n to be $u_n := U_n|_{t_n}$, the conditional distribution coming from a particle filter algorithm.
3. We define $\mu_n^N := \pi_{t_n|t_n}^N$ where the superscript means that the measure is obtained through the particle filter method, and so it is random.
4. We use \mathcal{S}^N to denote the sampling operator: $\pi_{t_n|t_n}^N \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{x_t^{(i)}}$ and L_n to denote the updating step in the particle filter. We use P_n^N to denote the transition operator (the prediction step) under the SGD–particle filter framework. P_n is the deterministic transition operator for the exact case (the control is exact in SDG). We mention "deterministic" here to distinguish the case where the control u_n may be random due to the SGD optimization algorithm.
5. We use $\langle \cdot, \cdot \rangle$ to denote the deterministic L_2 inner product, i.e., if $f, g \in L^2([0, T]; \mathbb{R}^d)$, then

$$\langle f, g \rangle := \int_0^T f \cdot g \, dt \quad (26)$$

6. We define $J_N'(U_n) := \mathbb{E}[J'_N(U_n) | X_n = x]$. We then have $\mathbb{E}[J_N'^{X_n}(U_n)] := \int \mathbb{E}[J'_N(U_n) | X_n = s] \, d\mu_n^N(s)$. We remark that U_n is a process that starts from time t_n , and so X_n is essentially the initial condition of the diffusion process.
7. We define the distance between two random measures to be the following:

$$d(\mu, \nu) := \sup_{\|f\|_\infty \leq 1} \sqrt{\mathbb{E}^\omega[|\mu^\omega f - \nu^\omega f|^2]} \quad (27)$$

where the expectation is taken over by the randomness of the measure.

8. We use the total variation distance between two deterministic probability measures μ, ν :

$$d_{TV}(\mu, \nu) := \sup_{\|f\|_\infty \leq 1} |\mu f - \nu f| \quad (28)$$

9. We use K_n to denote the total number of iterations taken in the SGD algorithm at time t_n ; we use \mathcal{N} to denote the total number of particles in the system. We use C to denote a generic constant which may vary from line to line.
10. Abusing the notation, we will denote $J_N'(U_n)$ in the following way where the argument U_n can be a vector of any length $1 \leq n \leq N$:

$$J_N'(U_n)|_{t_i} := \mathbb{E}^{X_{t_n}=x}[b_u(X_{t_i}, U_n|_{t_i})^\top Y_{t_i} + \sigma_u(X_{t_i}, U_n)^\top Z_{t_i} + f_u(X_{t_i}, U_n|_{t_i})] \quad (29)$$

Assumptions

1. We assume that J'_N satisfy the following strong condition: for any $x \in X$, there exist a constant $\lambda > 0$ such that for all $U, V \in \mathcal{U}_0$:

$$\lambda ||U - V||^2 \leq \langle J'_N(x)(U) - J'_N(x)(V), U - V \rangle \quad (30)$$

Notice that (30) implies that such inequality is true for any $U_n, V_n \in \mathcal{U}_n$, and it can be seen from simply fixing all the $U_n|_{t_i}, V_n|_{t_i}, 0 \leq i \leq n - 1$ to be 0.

This is a very strong assumption, and one should consider relaxing it to

$$\lambda ||U - V||^2 \leq \mathbb{E}^\omega [\mathbb{E}^{\mu_{n'}} [\langle J'_N(x)(U) - J'_N(x)(V), U - V \rangle]] \quad (31)$$

That is, this relation holds in expectation instead of point-wise.

2. Both b and σ are deterministic and in $C_b^{2,2}(\mathbb{R}^d \times \mathbb{R}^m; \mathbb{R}^d)$ in space variable x and control u .
3. $b, b_x, b_u, \sigma, \sigma_x, f_x, f_u$ are all uniformly Lipschitz in x, u and uniformly bounded.
4. σ satisfies the uniform elliptic condition.
5. The initial condition $X_0 := \xi \in L^2(\mathcal{F}_0)$.
6. The terminal (Loss) function Φ is C^1 and positive, and Φ_x has the most linear growth at infinity.
7. We assume that the function g_n (related to the Bayesian step) has the following bound: there exists $0 < \kappa < 1$ such that

$$\kappa \leq g_n \leq \kappa^{-1}$$

3.2. The Convergence Theorem for the Data-Driven Feedback Control Algorithm

Our algorithm combines the particle filter method and the stochastic gradient descent method. Lemma 1 (combine Lemma 4.7–4.9 from the book [22]) provides the convergence result for the particle filter method alone. It shows that each prediction and updating step is guaranteed to be convergent.

Recall that S^N is the sampling operator where we sample N particles. P_n^N denotes the transition operator (the prediction step) under the SGD–particle filter framework. P_n denotes the deterministic transition operator assuming that SGD gives the exact control. L_n denotes the updating step in the particle filter method.

Lemma 1. *We assume that there exists $\kappa \in (0, 1]$. The following is true:*

$$\sup_{\mu \in \mathcal{P}(\mu)} d(S^N \mu, \mu) \leq \frac{1}{\sqrt{N}} \quad (32)$$

$$d(P_n^N \mu, P_n^N \nu) \leq d(\mu, \nu) \quad (32)$$

$$d(P_n \mu, P_n \nu) \leq d(\mu, \nu)$$

$$d(L_n \nu, L_n \mu) \leq 2\kappa^{-2} d(\nu, \mu) \quad (33)$$

Given Lemma 1, Theorem 4.5 in [22] tells us the particle filter framework is convergent. Then, following Lemma 1, we can have the distance between the true distribution of the state and the estimated distribution through the SGD–particle filter framework.

$$\begin{aligned}
d(\mu_{n+1}^{N,\cdot}, \mu_{n+1}) &\equiv d(L_n S^N P_n^N \mu_n^{N,\cdot}, L_n P_n \mu_n) \\
&\leq d(L_n S^N P_n^N \mu_n^{N,\cdot}, L_n S^N P_n \mu_n) + d(L_n S^N P_n \mu_n, L_n P_n \mu_n) \\
&\leq 2\kappa^{-2} \left(\frac{2}{\sqrt{N}} + d(P_n^N \mu_n^{N,\cdot}, P_n \mu_n^{N,\cdot}) + d(S^N P_n \mu_n^{N,\cdot}, P_n \mu_n) \right) \\
&\leq 2\kappa^{-2} \left(\frac{3}{\sqrt{N}} + d(P_n^N \mu_n^{N,\cdot}, P_n \mu_n^{N,\cdot}) + d(\mu_n^{N,\cdot}, \mu_n) \right)
\end{aligned} \tag{34}$$

where in the above inequalities, we have used triangle inequalities and Lemma 1.

Hence, if we can show that the inequality of the following form holds

$$d(P_n^N \mu_n^{N,\cdot}, P_n \mu_n^{N,\cdot}) \leq C_n d(\mu_n^{N,\cdot}, \mu_n) + \epsilon_n \tag{35}$$

for some constant C_n and ϵ_n that we can tune, then by recursion, we can show that by using (34), the convergence holds true.

Remark. We point out that the difficulty lies in showing (35). Recall that the distance between two random measures is defined in (27) and involves testing the overall measurable function bounded by 1. However, we will see later that it is more desirable to test against Lipschitz functions. Hence, since the underlying measure is a finite Borel probability measure, we want to identify the function first with a continuous function on a compact set (Lusin). Then, we approximate this continuous function uniformly by a Lipschitz function since now the domain is compact. This way, we can roughly show that a form close to (35) is true.

Remark. Notice that the first measure in $d(P_n^N \mu_n^{N,\cdot}, P_n \mu_n^{N,\cdot})$ has two sources of randomness: the randomness in P_n which comes from the SGD algorithm used to find the control, and the randomness in the measure μ_n^N . However, we do not distinguish the two when we take the expectation.

To prove the convergence, we need to create a subspace where all particles X_i (obtained from the particle filter method) at any time n are within this bounded subspace. Or we can relax it to the statement that the probability of any particles X_i escaping from a very large region is very small. Lemma 2 shows we can restrict the particles to a compact subspace with the radius M by starting from any particles and any admissible control U_0 .

Lemma 2. *There exists M and constant C , such that under any admissible control U_0*

$$\mathbb{P}\left(\sup_{\substack{i \in \{1, \dots, N\} \\ U_0 \in \mathcal{U}}} |X_i| \geq M\right) \leq \frac{C}{M^2}, \quad X_i \sim \pi_{t_i|t_{i-1}}^N \text{ or } X_i \sim \pi_{t_i|t_i}^N \tag{36}$$

Proof. See Appendix A.1. \square

Remark. Lemma 2 tells that starting from any random selection of particles and any admissible control U_0 , at any time t , all particles are restricted by a compact set \mathcal{M} with diameter $\text{diam}(\mathcal{M}) \leq M$, such that

$$\mathbb{P}\left(\sup_{\substack{i \in \{1, \dots, N\} \\ U_0 \in \mathcal{U}}} |X_i| \geq \text{diam}(\mathcal{M})\right) \leq \frac{C}{M^2}, \quad X_i \sim \pi_{t_i|t_{i-1}}^N \text{ or } X_i \sim \pi_{t_i|t_i}^N \tag{37}$$

We will use the following result extensively later

$$\mathbb{E}[\mathbf{1}_{\{|X_n| \geq M\}}] \leq \frac{C}{M^2}, \quad \forall 1 \leq n \leq N \tag{38}$$

The following Lemma 3 describes the difference between the estimated optimal control u_n and the true control u_n^* . Let $\mathcal{G}_k := \{\Delta W_n^i, x^i\}_{i=0}^{k-1}$. We can see that knowing \mathcal{G}_k essentially means that we know the control u_k in the SGD framework at time t_n , since according to our scheme, the control is \mathcal{G}_k measurable.

Lemma 3. *Under a fixed temporal discretization number N , with the particle cloud $\mu_n^{N,\omega}$, a deterministic u_n^* and a compact domain \mathcal{K}'_n , (such that $\mathbb{E}^\omega \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n}] \leq \frac{C}{M_n^2}$ and $\text{diam}(\mathcal{K}'_n) \leq M_n$), we have for any iteration number K that the following holds*

$$\mathbb{E}^\omega \left[\mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{K}'_n} | u_n^\omega - u_n^* |^2 \mid X_n = x \right] \right] \leq CM_n^2 \sup_{\|q\|_\infty \leq 1} \mathbb{E}^\omega [\| \mu_n^{N,\omega} q - \mu_n q \|^2] + \frac{C}{M_n} + \frac{CM_n^2}{K} \quad (39)$$

Remark. The value of $\sup_{\|q\|_\infty \leq 1} \mathbb{E}^\omega [\| \mu_n^{N,\omega} q - \mu_n q \|^2]$ depends on $\mu_n^{N,\omega}$, which is obtained from the previous step, and it does not depend on the current M_n . As a result, we can see that, as long as

$$\sup_{\|q\|_\infty \leq 1} \mathbb{E}^\omega [\| \mu_n^{N,\omega} q - \mu_n q \|^2] \rightarrow 0$$

(39) can be made arbitrarily small on any compact domain \mathcal{K}'_n , and this indicates the point-wise convergence for u at any time t_n .

Proof. For simplicity in the proof, we denote control as U_n^{K+1} where K is the iteration in SGD, and n is the current time t_n . $j_n'^{x^k}(U_n^K)$ denote the SGD process using estimated control U_n^K , and $J_N'(U_n^*)$ denote the process using true control U_n^* .

$$U_n^{K+1} = U_n^K - \eta_k j_n'^{x^k}(U_n^K) \quad (40)$$

$$U_n^* = U_n^* - \eta_k \mathbb{E}^{\mu_n} [J_N'(U_n^*)] \quad (41)$$

where x^k is drawn from the current distribution $\mu_n^{N,\omega}$ and $\mathbb{E}^{\mu_n} [J_N'(U_n^*)] = 0$ by the optimality condition. Take the difference between (40) and (41), square both sides, and take conditional expectation $\mathbb{E}[\cdot | \mathcal{G}_K]$, and this conditional expectation is taken with respect to the following randomness:

1. The randomness comes from the selection on the initial point x_n^k .
2. The randomness comes from the pathwise approximated Brownian motion used for FBSDEs.
3. The randomness comes from the accumulation of the past particle sampling.

We can write $\mathbb{E}[j'^x(U^K) | \mathcal{G}_K] = \mathbb{E}^{\mu_n^{N,\omega}} [j'^x(U^K) | \mathcal{G}_K]$, which can be seen from the following

$$\begin{aligned} \mathbb{E}[j'^x(U^K) | \mathcal{G}_K] &= \mathbb{E}^{\mu_n^{N,\omega}} [\mathbb{E}^{X_{t_n}=x} [j'^x(U^K) | \mathcal{G}_K]] \\ &= \mathbb{E}^{\mu_n^{N,\omega}} [J_N'(U^K) | \mathcal{G}_K] \end{aligned} \quad (42)$$

Then, take the square norm on both sides, multiply by an indicator function $\mathbf{1}_{\mathcal{K}'_n}$ and take conditional expectation $\mathbb{E}[\cdot | \mathcal{G}_K]$. Noticing that U_n^K is \mathcal{G}_K is measurable and U^* is deterministic. In this case, we obtain the following

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||U_n^{K+1} - U_n^*||^2 | \mathcal{G}_K] &= \mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||U_n^K - U_n^*||^2 | \mathcal{G}_K] - 2\eta_k \langle \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U_n^K)] | \mathcal{G}_K \rangle \\
&\quad - \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U_n^*), U_n^K - U_n^*] + \eta_k^2 \mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||J'_N(x)(U_n^K) - \mathbb{E}^{\mu_n} [J'_N(x)(U_n^*)]|| | \mathcal{G}_K] \\
&= \mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||U_n^K - U_n^*||^2 | \mathcal{G}_K] - 2\eta_k \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} \langle J'_N(x)(U_n^K) - J'_N(x)(U_n^*) \\
&\quad + J'_N(x)(U_n^*) - \mathbb{E}^{\mu_n} [J'_N(x)(U_n^*)], U_n^K - U_n^* | \mathcal{G}_K] \\
&\quad + \eta_k^2 \mathbb{E}^{\mu_n^N} [\mathbf{1}_{\mathcal{K}'_n} ||J'_N(x)(U_n^K) - \mathbb{E}^{\mu_n} [J'_N(x)(U_n^*)]||] \\
&= \mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||U_n^K - U_n^*||^2 | \mathcal{G}_K] - 2\eta_k \mathbb{E}^{\mu_n^{N,\omega}} [\langle \mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U_n^K) \\
&\quad - \mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*), U_n^K - U_n^* | \mathcal{G}_K] \\
&\quad - 2\eta_k \langle \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U_n^*)] - \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} \mathbb{E}^{\mu_n} [J'_N(x)(U_n^*)]], \mathbf{1}_{\mathcal{K}'_n} (U_n^K - U_n^*) \rangle \\
&\quad + \eta_k^2 \mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||J'_N(x)(U_n^K) - \mathbb{E}^{\mu_n} [J'_N(x)(U_n^*)]|| | \mathcal{G}_K] \\
&\leq (1 - \lambda\eta_k) \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} ||U_n^K - U_n^*||^2 | \mathcal{G}_K] \\
&\quad + \underbrace{\frac{\eta_k}{\lambda} ||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U_n^*) - \mathbf{1}_{\mathcal{K}'_n} \mathbb{E}^{\mu_n} [J'_N(x)(U_n^*)]]||^2 + \eta_k^2 \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} C|x_i|^2 + C]}_{**}
\end{aligned} \tag{43}$$

where in the last line, we used the following Lemma from [13] that states that there exists C such that

$$\mathbb{E}[||J'_N(x)||^2] \leq C|x|^2 + C \tag{44}$$

Recall that $\mathbb{E}^{\mu_n} [J'_N(x)] = 0$, we then have

$$\begin{aligned}
** &\leq ||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] - \mathbb{E}^{\mu_n^{N,\omega}} [\mathbb{E}^{\mu_n} [J'_N(x)(U^*)]] + \mathbb{E}^{\mu_n^{N,\omega}} [\mathbb{E}^{\mu_n} [J'_N(x)(U^*)]] \\
&\quad - \mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} \mathbb{E}^{\mu_n} [J'_N(x)(U^*)]]||^2 \\
&\leq ||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] - \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] + \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] - \mathbb{E}^{\mu_n^{N,\omega}} [\mathbb{E}^{\mu_n} [J'_N(x)(U^*)]]||^2 \\
&\leq (1 + \epsilon) ||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] - \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)]||^2 + (1 + \frac{1}{\epsilon}) ||\mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] \\
&\quad - \mathbb{E}^{\mu_n} [J'_N(x)(U^*)]||^2 \\
&\leq C ||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] - \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)]||^2 + \frac{C}{M_n}
\end{aligned} \tag{45}$$

Then, we take the expectation on both sides over the randomness and we have

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_{\mathcal{K}'_n} ||U^{K+1} - U^*||^2] &\leq (1 - \lambda\eta_k) \mathbb{E}^{\mu_n^N} [||U^K - U^*||^2] + \frac{\eta_k}{\lambda} \mathbb{E}^{\omega} [C ||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] \\
&\quad - \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)]||^2 + \frac{C}{M_n}] + \eta_k^2 C M_n^2 \\
&\leq \frac{||U^0 - U^*||^2}{K} + C \mathbb{E}^{\omega} [||\mathbb{E}^{\mu_n^{N,\omega}} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)] - \mathbb{E}^{\mu_n} [\mathbf{1}_{\mathcal{K}'_n} J'_N(x)(U^*)]||_2^2] \\
&\quad + \frac{C}{M_n} + \frac{C M_n^2}{K}
\end{aligned} \tag{46}$$

Notice that for the control U^* , we know that for a fixed x , $J'_N(x)(U^*)$ is uniformly bounded:

$$\begin{aligned}
& \|\mathbb{E}^{\mu_n^{N,\omega}}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)] - \mathbb{E}^{\mu_n}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)]\|_2^2 \\
&= \sum_{i=n}^N \Delta t \left| \mathbb{E}^{\mu_n^{N,\omega}}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)]_i - \mathbb{E}^{\mu_n}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)]_i \right|^2 \\
&\leq \sum_{i=n}^N \Delta t \sup_{j \in \{n, \dots, N\}} \left| \mathbb{E}^{\mu_n^{N,\omega}}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)(x)]_j - \mathbb{E}^{\mu_n}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)(x)]_j \right|^2 \\
&\leq \sup_{j \in \{n, \dots, N\}} \left| \mathbb{E}^{\mu_n^{N,\omega}}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)]_j - \mathbb{E}^{\mu_n}[\mathbf{1}_{\mathcal{K}'_n} J'_N(U^*)]_j \right|^2
\end{aligned} \tag{47}$$

However, since from the result in (44):

$$\sup_{j \in \{n, \dots, N\}} \left| J'_N(U^*) \right|_j^2 \leq C|x|^2 + C \tag{48}$$

we have that

$$\mathbf{1}_{\mathcal{K}'_n} \sup_{j \in \{n, \dots, N\}} \left| J'_N(U^*) \right|_j^2 \leq C|M_n|^2 \mathbf{1}_{\mathcal{K}'_n} |q(x)| \tag{49}$$

for some $q(x)$, where $\|q(x)\|_\infty \leq 1$. As a result, we see that

$$\begin{aligned}
(46) &\leq CM_n^2 \mathbb{E}^\omega \left[\left| \mathbb{E}^{\mu_n^{N,\omega}}[q(x)] - \mathbb{E}^{\mu_n}[q(x)] \right|^2 \right] + \frac{CM_n^2}{K} \\
&\leq CM_n^2 \sup_{\|q\|_\infty \leq 1} \mathbb{E}^\omega [|\mu_n^{N,\omega} q - \mu_n q|^2] + \frac{CM_n^2}{K}
\end{aligned} \tag{50}$$

Thus, we have

$$\mathbb{E}^\omega \left[\mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{K}'_n} \sup_n |u^{K+1} - u^*|^2 \right] \right] \leq CM_n^2 \sup_{\|q\|_\infty \leq 1} \mathbb{E}^\omega [|\mu_n^{N,\omega} q - \mu_n q|^2] + \frac{C}{M_n} + \frac{CM_n^2}{K} \tag{51}$$

where we have absorbed the constant term N in C . \square

Lemma 3 shows that when the empirical distribution μ_n^N is close enough to the true distribution μ_n , the difference between u_n and u_n^* under the expectation restricted to a compact set is bounded by the difference between the true distribution of the state and the estimated distribution. Thus, suppose we can show the convergence of the distribution. In that case, we have the convergence result of the estimated control to the "true control" under the expectation restricted to a compact set.

Next, we want to show that by moving forward in one step, the distance between the true distribution of the state and the estimated distribution through the SGD-particle filter framework is bounded by the distance of the previous step with some constants in Lemma 4.

Lemma 4. *For each $n = 0, 1, \dots, N-1$, there exist M_n, L_n, δ_n, K_n such that the following inequality holds*

$$d(\mu_{n+1}^{N,\cdot}, \mu_{n+1}) \leq 2\kappa^{-2} \left((1 + C\Delta t L_n M_n) d(\mu_n^{N,\cdot}, \mu_n) + \frac{C}{M_n} + \frac{CM_n}{K_n} + 2\delta_n + \frac{3}{\sqrt{N}} \right) \tag{52}$$

Proof. The key step is to estimate the quantity $d^2(P_n^N \mu_n^{N,\cdot}, P_n \mu_n^{N,\cdot})$ in (34). WLOG, we assume that the sup is realized by the function f with $\|f\|_\infty \leq 1$; then, we have

$$d^2(P_n^N \mu_n^{N,\cdot}, P_n \mu_n^{N,\cdot}) = \mathbb{E}^\omega [|P_n^N \mu_n^{N,\cdot} f - P_n \mu_n^{N,\cdot} f|^2] \quad (53)$$

Notice that P_n^N is the prediction operator that uses the control u_n which carries the randomness from SGD, and P_n uses the control u_n^* . Then $P_n^N \mu_n^{N,\omega}$ is a random measure, and we comment that both u_n^* and μ_n are deterministic.

Without loss of generality, we use u_n^ω and $\mu_n^{N,\omega}$ to denote the random control and the random measure. (Even though the randomness can be different, we can concatenate $(\omega_1, \omega_2) := \omega$ to define them as ω in general.)

We have for the fixed randomness ω , and by Fubini's theorem

$$\begin{aligned} |P_n^N \mu_n^{N,\omega} f - P_n \mu_n^{N,\omega} f|^2 &= \left| \mathbb{E}^{\mu_n^{N,\omega}} \left[\underbrace{\mathbb{E}[f(X_n + b(X_n, u_n^\omega) \Delta t + \sigma(X_n) \Delta W_n) | X_n = x]}_{f_1^\omega} \right. \right. \\ &\quad \left. \left. - \mathbb{E}^{\mu_n^{N,\omega}} \left[\underbrace{\mathbb{E}[f(X_n + b(X_n, u_n^*) \Delta t + \sigma(X_n) \Delta W_n) | X_n = x]}_{f_2} \right] \right] \right|^2 \\ &= \left| \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbb{E}[f_1^\omega - f_2 | X_n = x] \right] \right|^2 \\ &= \underbrace{\left| \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{M}_n} \mathbb{E}[f_1^\omega - f_2 | X_n = x] \right] \right|^2}_{A_1} \\ &\quad + \underbrace{\left| \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{M}_n^c} \mathbb{E}[f_1^\omega - f_2 | X_n = x] \right] \right|^2}_{A_2} \end{aligned} \quad (54)$$

where the inner conditional expectation is taken with respect to ΔW_n .

Now, since we can pick \mathcal{M}_n to be a large compact set containing the origin, then

$$\mathbb{P}(\sup_{n, U_0} |X_n| \geq \text{diam}(\mathcal{M}_n)) \leq \frac{C}{M_n^2} \quad (55)$$

To deal with A_1, A_2 , we see that it is desirable that the function f has the *Lipschitz property*. However, it is only measurable in general. The strategy to overcome this difficulty is to first use *Lusin's Theorem* to find a continuous identification \tilde{f} with f on a large compact set; then, on this compact set, we can approximate \tilde{f} uniformly by a Lipschitz function.

We see that

$$A_1 \leq \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{M}_n} \mathbb{E}[|f_1^\omega - f_2|^2 | X_n = x] \right] \quad (56)$$

Then, by taking expectation on both sides over all the randomness in this quantity, we have

$$\mathbb{E}^\omega [A_1] \leq \mathbb{E}^\omega \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{M}_n} \mathbb{E}[|f_1^\omega - f_2|^2] \right] \quad (57)$$

We know that there exists a big compact \mathcal{K}_n (so a large M_n) containing the origin such that

$$\mathbb{P}(\sup_{n, U_0} |X_n| \geq \text{diam}(\mathcal{K}_n)) \leq \frac{C}{M_n^2} \quad (58)$$

and a continuous \tilde{f}^n with $\tilde{f}^n|_{\mathcal{K}_n} = f|_{\mathcal{K}_n}$ by *Lusin's theorem*.

Thus, we know that $\tilde{f}^n|_{\mathcal{K}_n \cap \mathcal{M}_n} = f|_{\mathcal{K}_n \cap \mathcal{M}_n}$, and we also have the following inequality:

$$\begin{aligned} \mathbb{E}^\omega \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{M}_n} \mathbb{E}[|f_1^\omega - f_2|^2] \right] &= \mathbb{E}^\omega \mathbb{E}^{\mu_n^{N,\omega}} \left[(\mathbf{1}_{\mathcal{M}_n \cap \mathcal{K}_n} + \mathbf{1}_{\mathcal{M}_n \cap \mathcal{K}_n^c}) \mathbb{E}[|f_1^\omega - f_2|^2] \right] \\ &\leq \mathbb{E}^\omega \mathbb{E}^{\mu_n^{N,\omega}} \left[(\mathbf{1}_{\mathcal{M}_n \cap \mathcal{K}_n} + \mathbf{1}_{\mathcal{K}_n^c}) \mathbb{E}[|f_1^\omega - f_2|^2] \right] \end{aligned} \quad (59)$$

Moreover, since both \mathcal{K}_n and \mathcal{M}_n are compact, $\mathcal{K}'_n := \mathcal{K}_n \cap \mathcal{M}_n$ is also compact with $\text{diam}(\mathcal{K}'_n) \leq M_n$. From Lemma 2, we know that there exists some constant C such that for any $\pi_{t_n|t_{n-1}}^N, \pi_{t_n|t_n}^N$ that one obtains from or particle filter–SGD algorithm, $X \sim \pi_{t_n|t_{n-1}}^N$ or $\pi_{t_n|t_n}^N$:

$$\mathbb{E}^\omega \left[\mathbb{E}[\mathbf{1}_{\{X \in \mathcal{K}_n^c\}}] \right] \leq \frac{C}{M_n^2} \quad (60)$$

Hence, we have that

$$(59) \leq \mathbb{E}^\omega \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{K}'_n} \mathbb{E}[|f_1^\omega - f_2|^2] \right] + \frac{C}{M_n^2} \quad (61)$$

To deal with A_2 , notice that $|f_1^\omega - f_2| \leq 2$ by the choice of f , we have the following.

$$\begin{aligned} A_2 &\leq \mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{M}_n^c} \mathbb{E}[|f_1^\omega - f_2|^2 | X_n = x] \right] \\ &\Rightarrow \mathbb{E}^\omega[A_2] \leq 4 \mathbb{E}^\omega \left[\mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{K}_n^c} \right] \right] \\ &\leq \frac{C}{M_n^2} \end{aligned} \quad (62)$$

by Lemma 2.

To deal with A_1 , we have by the density of the Lipschitz function that there exists $\|f^n - \tilde{f}^n\|_{\mathcal{K}'_n, \infty} \leq \delta_n$ with Lipschitz constant L_n . We point out that L_n may depend on \mathcal{K}'_n, δ_n and the function $\tilde{f}|_{\mathcal{K}'_n}$. Now, by taking the expectation on both sides and using the Lipschitz property, we have

$$\mathbb{E}^\omega[A_1] \leq (C \Delta t L_n)^2 \underbrace{\mathbb{E}^\omega \left[\mathbb{E}^{\mu_n^{N,\omega}} \left[\mathbf{1}_{\mathcal{K}'_n} |u_n^\omega - u_n^*|^2 | X_n = x \right] \right]}_* + \frac{C}{M_n^2} + 4\delta_n^2 \quad (63)$$

We realize that $*$ is the SGD optimization part of the algorithm in expectation, and we note that we have dropped the inner expectation. The expectation $\mathbb{E}^{\mu_n^{N,\omega}}[\cdot]$ means that given the initial condition $X_n = x \in \mathbf{1}_{\mathcal{K}'_n}$, with $X_n \sim \mu_n^{N,\omega}$, one wants to find the difference in expectation between u_n and u_n^* . The outer expectation $\mathbb{E}^\omega[\cdot]$ means averaging overall the randomness in both the measure and the SGD.

Now, by using (50) in Lemma 3, absorbing N in the constant C, we obtain the following

$$\mathbb{E}^\omega[A_1] \leq (C \Delta t L_n)^2 N M_n^2 \sup_{\|q\| \leq 1} \mathbb{E}^\omega[|\mu_n^{N,\omega} q - \mu_n q|^2] + \frac{C M_n^2}{K_n} + \frac{C}{M_n^2} + 4\delta_n^2 \quad (64)$$

By definition of the distance between two random measures, we have that:

$$\begin{aligned}\mathbb{E}[A_1] &\leq (C\Delta t L_n)^2 N M_n^2 d^2(\mu_n^{N,\cdot}, \mu_n) + \frac{CM_n^2}{K_n} + \frac{C}{M_n^2} + 4\delta_n^2 \\ \Rightarrow \sqrt{\mathbb{E}[A_1]} &\leq C\Delta t L_n M_n d(\mu_n^{N,\cdot}, \mu_n) + \frac{CM_n}{\sqrt{K_n}} + \frac{C}{M_n} + 2\delta_n\end{aligned}\quad (65)$$

Since $\sqrt{\mathbb{E}[A_2]} \leq \frac{C}{M_n}$, we have that

$$\begin{aligned}(34) &\leq 2\kappa^{-2} \left(\frac{3}{\sqrt{N}} + C\Delta t L_n M_n d(\mu_n^{N,\cdot}, \mu_n) + \frac{CM_n}{\sqrt{K_n}} + \frac{C}{M_n} + 2\delta_n + \frac{2}{M_n} + d(\mu_n^{N,\cdot}, \mu_n) \right) \\ \Rightarrow d(\mu_{n+1}^{N,\cdot}, \mu_{n+1}) &\leq 2\kappa^{-2} \left((1 + C\Delta t L_n M_n) d(\mu_n^{N,\cdot}, \mu_n) + \frac{C}{M_n} + \frac{CM_n}{\sqrt{K_n}} + 2\delta_n + \frac{3}{\sqrt{N}} \right)\end{aligned}\quad (66)$$

where in (66), we have merged \sqrt{N} into C . \square

Remark. Lusin's theorem requires the underlying measure to be finite Borel regular, and in this case, we are looking at the measure $\tilde{\mu}$ defined as follows: for $A \subset \mathbb{R}^n$, $\tilde{\mu}(A) = \mathbb{P}(\{\omega \mid \text{there exist } n, U_0 \text{ such that } X_n(\omega) \in A\})$. $\tilde{\mu}$ is clearly a probability measure induced on the Polish space \mathbb{R}^n , and so it is tight by the inverse implication of Prokhorov's theorem (or we can use the fact that all finite Borel measures defined on a complete metric space are tight). Thus, it is inner regular; since now $\tilde{\mu}$ is also clearly locally finite, it also implies the outer regularity.

Finally, we can use Lemma 4 repeatedly to show the convergence result:

Theorem 5. By taking $\mu_0^N = \mu^0$, there exist $\{M_n \mid M_n \in \mathbb{R}, n = 0, 1, \dots, N-1\}$, $\{L_n \mid L_n \in \mathbb{R}, n = 0, 1, \dots, N-1\}$ and $\{\delta_n \mid \delta_n \in \mathbb{R}, n = 0, 1, \dots, N-1\}$ such that

$$d(\mu_N^{N,\cdot}, \mu_N) \leq \sum_{i=0}^{N-1} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \left(\frac{C}{M_{N-i}} + 2\delta_{N-i} + \frac{CM_{N-i}}{\sqrt{K_{N-i}}} + \frac{3}{\sqrt{N}} \right) \quad (67)$$

where $C_j := 1 + C\Delta t L_j M_j$. Then, for any $M > 0$, we have by picking $\{M_n\}$, $\{K_n\}$, N large enough and $\{\delta_n\}$ small enough, and then the following holds

$$d(\mu_N^{N,\cdot}, \mu_N) \leq \frac{C}{M} \quad (68)$$

for some fixed constant C which depends only on κ .

Proof. With C_n defined as

$$C_n := 1 + C\Delta t L_n M_n \quad (69)$$

and by using (66) repeatedly, we obtain the following result:

$$\begin{aligned}d(\mu_N^{N,\cdot}, \mu_N) &\leq \sum_{i=0}^{N-1} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \left(\frac{C}{M_{N-i}} + \frac{CM_{N-i}}{K_{N-i}} + 2\delta_{N-i} + \frac{3}{\sqrt{N}} \right) \\ &\quad + (2\kappa^{-2})^N \prod_{j=0}^{N-1} C_{N-j} d(\mu_0^N, \mu_0) \\ &\leq \sum_{i=0}^{N-1} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \left(\frac{C}{M_{N-i}} + \frac{CM_{N-i}}{K_{N-i}} + 2\delta_{N-i} + \frac{3}{\sqrt{N}} \right)\end{aligned}\quad (70)$$

Since we know that $d(\mu_0^N, \mu_0) = 0$, we now just need to show that (70) vanishes when K_l, N gets large and δ_i gets small, $i \in \{0, 1, \dots, N\}$. Notice that M_l comes from the domain truncation for each time step and δ_l comes from the uniform approximation which is free to choose. The choice of δ_l will potentially determine the value of L_n .

We fix $M_N := NM, \delta_N := \frac{1}{NM}$ where N is the number of time discretization and M is potentially a large number.

Then, we define δ_l, M_l through the following:

$$(2\kappa^{-2})^{i+1} \prod_{j=0}^i C_{N-j} 2\delta_{N-i-1} = (2\kappa^{-2})^i \prod_{j=0}^{i-1} 2\delta_{N-i} \quad (71)$$

$$(2\kappa^{-2})^{i+1} \prod_{j=0}^i C_{N-j} \frac{C}{M_{N-i-1}} = (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \frac{C}{M_{N-i}} \quad (72)$$

Here, we define $C_{N+1} \equiv 1$.

Notice that one should iterate (71) and (72) iteratively, since defining δ_i will lead to the Lipschitz constant L_i at stage i , which is needed for the definition for C_i .

Then, we have that

$$\begin{aligned} \sum_{i=0}^{N-1} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \frac{C}{M_{N-i}} &\leq N \frac{C}{NM} \\ &\leq \frac{C}{M} \end{aligned} \quad (73)$$

and we also have

$$\begin{aligned} \sum_{i=0}^{N-1} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} 2\delta_{N-i} &\leq N \frac{1}{NM} \\ &\leq \frac{1}{M} \end{aligned} \quad (74)$$

By picking K_{N-i} to be large, we then can have

$$\begin{aligned} \sum_{i=0}^{N-1} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \frac{CM_{N-i}}{K_{N-i}} &\leq N \frac{1}{NM} \\ &\leq \frac{1}{M} \end{aligned} \quad (75)$$

Last but not least, by taking \mathcal{N} so large such that

$$\left(\sum_{i=0}^{N-i} (2\kappa^{-2})^i \prod_{j=0}^{i-1} C_{N-j} \right) \frac{3}{\sqrt{\mathcal{N}}} \leq \frac{1}{M} \quad (76)$$

we can see that (70) converges to 0 by taking M to be very large. \square

Remark. Notice that in Theorem 5, it is natural to have terms that depend on $\frac{1}{K_n}$ and $\frac{1}{\mathcal{N}}$. The presence of M_n and δ_n are due to technical difficulties. M_n basically gives the growth of the particles in the worst-case scenario (we want our domain to be compact), while L_n and δ_n come from the Lipschitz approximation for the test function f .

4. Numerical Example

In this section, we carry out two numerical examples. In the first example, we consider a classic linear quadratic optimal control problem, in which the optimal control can be derived analytically. We use this example as a benchmark example to show the baseline

performance and the convergence trend of our algorithm. In the second example, we solve a more practical Dubins vehicle maneuvering problem, and we design control actions based on bearing angles to let the target vehicle follow a pre-designed path.

4.1. Example 1. Linear Quadratic Control Problem with Nonlinear Observations

Assume B, K are symmetric, positive definite. The forward process Y and the observation process M are given by

$$\begin{aligned} dY(t) &= A(u(t) - r(t))dt + \sigma Bu(t)dW_t \\ dM(t) &= \sin(Y(t))dt + dB_t \end{aligned} \quad (77)$$

The cost functional is given by

$$J[u] = \frac{1}{2}E \left[\int_0^T \langle R(Y_t - Y_t^*), (Y_t - Y_t^*) \rangle dt + \frac{1}{2} \int_0^T \langle Ku_t, u_t \rangle dt + \frac{1}{2} \langle QY_T, Y_T \rangle \right] \quad (78)$$

and we want to find $J(u^*) = \inf_{u \in \mathcal{U}_{ad}[0,T]} J(u)$.

4.1.1. Experimental Design

An interesting fact of such an example is that one can construct a time deterministic exact solution which depends only on x_0 .

By simplifying (78), we have

$$J[u] = \frac{1}{2} \int_0^T (RE[Y_t^T Y_t] - 2RY_t^T E[Y_t] + RY^{*T} Y^* + \langle Ku_t, u_t \rangle dt) + \frac{1}{2}E[\langle QY_T, Y_T \rangle] \quad (79)$$

Then, we define:

$$\begin{aligned} X_t &= E[Y_t] = E[Y_0 + \int_0^t A(u(s) - r(s))ds + \int_0^t \sigma Bu(s)dW_s] \\ &= E[Y_0 + \int_0^t A(u(s) - r(s))ds] \end{aligned} \quad (80)$$

Hence, we see that

$$\begin{aligned} E[Y_t^T Y_t] &= E[(Y_0 + \int_0^t A(u(s) - r(s))ds + \int_0^t \sigma Bu(s)dW_s)^2] \\ &= E[Y_0^T Y_0 + \int_0^t (u(s) - r(s))^T A^T A(u(s) - r(s))ds + Y_0^T \int_0^t A(u(s) - r(s))ds \\ &\quad + \int_0^t (u(s) - r(s))^T A^T Y_0 ds] + E[\int_0^t \sigma^2 u(s)^T B^T B u(s)ds] \\ &= X_t^T X_t + \sigma^2 \int_0^t u(s)^T B^T B u(s)ds \end{aligned} \quad (81)$$

and (81) is true because all the terms are deterministic in time given x_0 . Moreover, we observe that

$$\begin{aligned} E[Y_T^T Y_T] &= E[(Y_0 + \int_0^T A(u(s) - r(s))ds + \int_0^T \sigma Bu(s)dW_s)^2] \\ &= X_T^T X_T + \sigma^2 \int_0^T u(s)^T B^T B u(s)ds \end{aligned} \quad (82)$$

As a result, we see that (79) now takes the form:

$$\begin{aligned} J[u] = & \frac{1}{2} \int_0^T (X_s^T R X_s - 2 X_s^T R X_s^* + X_s^{*T} R X_s^* + u_s^T (\sigma^2 B^T Q B + K) u_s) ds \\ & + \frac{1}{2} \sigma^2 \int_0^T \int_0^t u_s^T B^T R B u_s ds dt + \frac{1}{2} X_T^T Q X_T \end{aligned} \quad (83)$$

and by performing a simple integration by part, we have

$$\begin{aligned} J[u] = & \frac{1}{2} \int_0^T (X_s^T R X_s - 2 X_s^T R X_s^* + X_s^{*T} R X_s^* + u_s^T (\sigma^2 B^T Q B + K) u_s) ds \\ & + \frac{1}{2} \sigma^2 \int_0^T (T - t) u_s^T B^T R B u_s ds + \frac{1}{2} X_T^T Q X_T \end{aligned} \quad (84)$$

As a result, we have the following standard deterministic control problem:

$$\begin{aligned} J[u] = & \frac{1}{2} \int_0^T \underbrace{(X_s^T R X_s - 2 X_s^T R X_s^* + X_s^{*T} R X_s^* + u_s^T (\sigma^2 B^T Q B + K) u_s + \sigma^2 (T - t) u_s^T B^T R B u_s)}_{2f} ds \\ & + \frac{1}{2} X_T^T Q X_T \end{aligned} \quad (85)$$

$$\frac{dX_t}{dt} = \underbrace{A(u(t) - r(t))}_b, X_{t_0} = X_0 \quad (86)$$

Then, one can form the following Hamiltonian

$$H(x, p, u) = bp + f \quad (87)$$

where p is $\nabla_x v$, and v is the value function.

Then, to find the optimal control, we have

$$\frac{\partial}{\partial u} H = 0 \quad (88)$$

which is

$$Ap + u(\sigma^2 B^T R B(T - t) + (K + \sigma^2 B^T Q B)) = 0 \quad (89)$$

Thus, we obtain

$$u_t = \frac{-Ap(t)}{\sigma^2 B^T R B(T - t) + (K + \sigma^2 B^T Q B)} \quad (90)$$

Additionally, notice that

$$\frac{d}{dt} p(t) = -R(X_t - X_t^*), \quad p(T) = Q X_T \quad (91)$$

and then,

$$p(t) = Q X_T + \int_0^T R(X_s - X_s^*) ds \quad (92)$$

Combining (86), (90) and (92) together, we can solve the control of the system.

Then, set

$$\mathbf{A} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix} \quad (93)$$

Set \mathbf{B} , \mathbf{R} , \mathbf{K} and \mathbf{Q} as identity matrices. With $\mathbf{X}_0 = \mathbf{0}$, we have the following solution according to this setup.

To solve (92), let

$$\begin{aligned} X_t^1 - X_t^{*1} &:= t \\ X_t^2 - X_t^{*2} &:= \cos(t) \\ X_t^3 - X_t^{*3} &:= t^2 \\ X_t^4 - X_t^{*4} &:= -2\pi \sin(2\pi t) \end{aligned}$$

Then, we have

$$p(t) = \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} - \frac{t^2}{2} \\ \sin(T) - \sin(t) \\ \frac{T^3}{3} - \frac{t^3}{3} \\ \cos(2\pi T) - \cos(2\pi t) \end{bmatrix} \quad (94)$$

Let $\hat{r}(t)$ be

$$\hat{r}_t^1 = \frac{-t^2/2}{\beta_t}, \quad \hat{r}_t^2 = \frac{-\sin(t)}{\beta_t}, \quad \hat{r}_t^3 = \frac{-t^3}{3\beta_t}, \quad \hat{r}_t^4 = \frac{-\cos(2\pi t)}{\beta_t}$$

where $\beta_t = (1 + \sigma^2) + \sigma^2(T - t)$. Then,

$$r(t) = A\hat{r}(t)$$

Then, plug (94) into (90), and we solve (86)

$$\frac{dX_t}{dt} = A(u(t) - r(t)) = A\left(-\frac{Ap(t)}{\beta_t} - A\hat{r}(t)\right) \quad (95)$$

and obtain

$$X_t = \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \end{bmatrix} = \frac{\alpha_t A^2}{\sigma^2} \left(\begin{bmatrix} \frac{T^2}{2} \\ \sin(T) \\ \frac{T^3}{3} \\ \cos(2\pi T) \end{bmatrix} - \begin{bmatrix} X_T^1 \\ X_T^2 \\ X_T^3 \\ X_T^4 \end{bmatrix} \right) \quad (96)$$

Thus, replacing X_t with X_t^* , we obtain

$$X_t^* = \begin{bmatrix} X_t^{1,*} \\ X_t^{2,*} \\ X_t^{3,*} \\ X_t^{4,*} \end{bmatrix} = \begin{bmatrix} t \\ \cos(t) \\ t^2 \\ -2\pi \sin(2\pi t) \end{bmatrix} + \frac{\alpha_t A^2}{\sigma^2} \left(\begin{bmatrix} \frac{T^2}{2} \\ \sin(T) \\ \frac{T^3}{3} \\ \cos(2\pi T) \end{bmatrix} - \begin{bmatrix} X_T^1 \\ X_T^2 \\ X_T^3 \\ X_T^4 \end{bmatrix} \right) \quad (97)$$

where X_T^i can be obtained from the system (96) by letting $t = T$, and

$$\alpha_t = \ln \frac{1 + \sigma^2 + \sigma^2 T}{(1 + \sigma^2) + \sigma^2(T - t)}$$

Then, to find the exact form by following the trajectory of y_t in this setup, one will have to solve the following coupled forward–backward ODE.

$$\frac{dX_t}{dt} = A(u_t - r_t) \quad x_n = y_{t_n} \quad (98)$$

$$\frac{dp(t)}{dt} = X_t - X_t^*, \quad p_T = x_T^{t_n, y_{t_n}} \quad (99)$$

with $u_t = -Ap_t / (\sigma^2(T-t) + (1+\sigma^2))$. As a result, we have

$$\frac{dX_t}{dt} = A(-Ap_t / (\sigma^2(T-t) + (1+\sigma^2)) - r_t) \quad x_n = y_{t_n} \quad (100)$$

$$\frac{dp(t)}{dt} = X_t - X_t^*, \quad p_T = x_T^{t_n, y_{t_n}} \quad (101)$$

That is, we need to solve the above coupled FBODE. Then, seeing that $p_t = x_T^{t_n, y_{t_n}} + \int_{t_n}^T X_s - X_s^* ds$, and writing $a_t := 1 / (\sigma^2(T-t) + (1+\sigma^2))$, we have

$$\frac{dX_t}{dt} = A(-a_t A X_T - a_t A \int_{t_n}^T (X_s - X_s^*) ds - r_t), \quad x_{t_n} = y_{t_n} \quad (102)$$

To solve (102) numerically, we conduct a numerical discretization:

$$\begin{aligned} x_{t_{n+1}} - x_{t_n} &= -a_{t_n} A^2 X_T \Delta t - a_{t_n} (\Delta t)^2 A^2 \sum_{i=n}^{N-1} (X_{t_i} - X_{t_i}^*) - A r_t \\ \Rightarrow A r_t - a_{t_n} (\Delta t)^2 A^2 \sum_{i=n}^{N-1} X_{t_i}^* &= X_{t_n} - X_{t_{n+1}} - a_{t_n} (\Delta t)^2 A^2 \sum_{i=n}^{N-1} X_{t_i} - a_{t_n} A^2 X_T, \quad x_{t_n} = y_{t_n} \end{aligned} \quad (103)$$

We can put (103) into a large linear system and solve it numerically.

4.1.2. Performance Experiment

We set the total number of discretizations to be $N = 50$. Set iteration $L = 10^4$, $\sigma = 0.1$, the number of particles in each dimension is 128, $T = 1$, and $X_0 = 0$.

In Figure 1, we present the estimated data-driven control and the true optimal control.

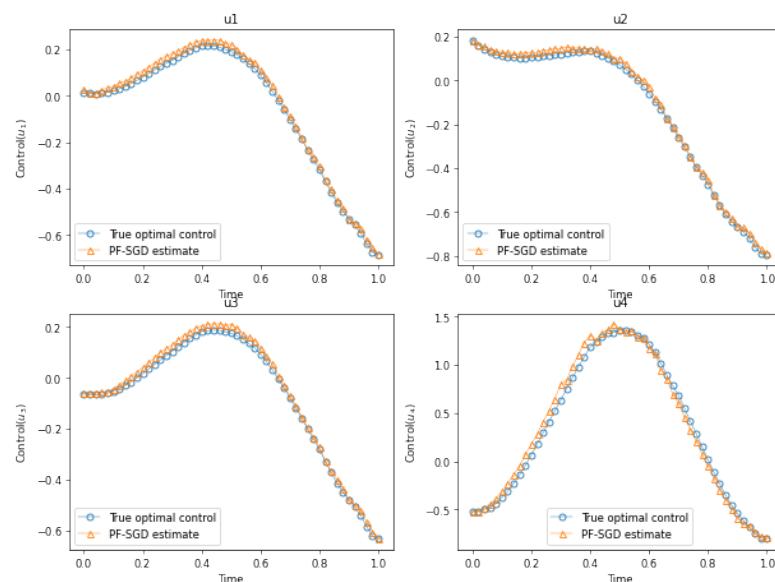


Figure 1. Estimated control vs. true optimal control.

In Figure 2, we show the estimated state trajectories with respect to true state trajectories in each dimension.

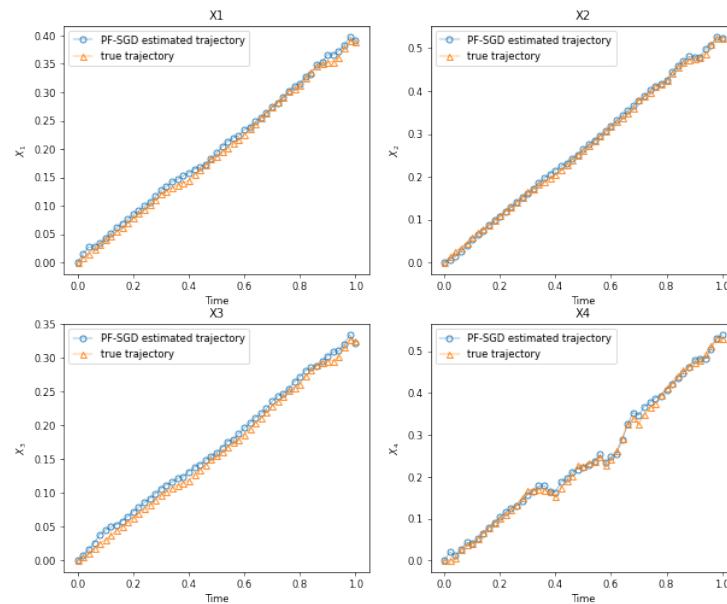


Figure 2. Estimated state vs. true state.

We can see from these Figures that our data-driven feedback control algorithm works very well for this 4-D linear quadratic control problem despite there being nonlinear observations.

4.1.3. Convergence Experiment

In this experiment, we demonstrate the convergence performance of our algorithm, and we study the error decay of the algorithm in the L_2 norm with respect to the number of particles used. Each result is an average of $\|u^{est} - u^*\|_2$ of 50 independent tests.

Specifically, we set $L = 10^4$ and we just increase the number of particles $S = \{2, 8, 32, 128, 512, 2048, 4096, 8192, 16384, 32768\}$, and we obtained the result in Figure 3.

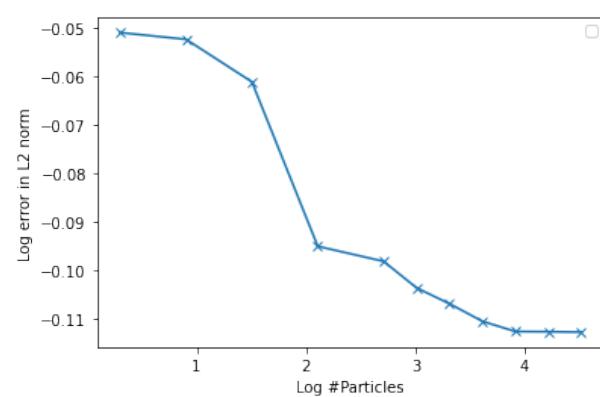


Figure 3. Error vs. number of particles.

Set the number of particles $S = \{2, 8, 32, 128, 512, 1024, 2048, 4096\}$, and $L = S^2$. We obtained the result in Figure 4.

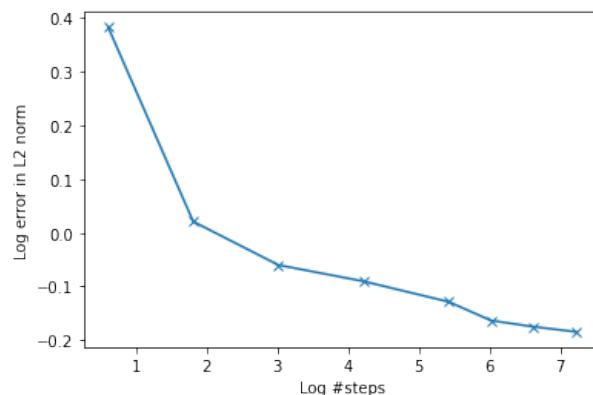


Figure 4. Error vs. number of steps.

From the results above, we can see that the error will decrease and converge as we increase the number of particles and the number of iterations.

4.2. Example 2. Two-Dimensional Dubins Vehicle Maneuvering Problem

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

$$dS_t = \begin{bmatrix} dX_t \\ dY_t \end{bmatrix} = \begin{bmatrix} \sin(\theta_t)dt \\ \cos(\theta_t)dt \end{bmatrix} + \sigma dW_t \quad (104)$$

$$d\theta_t = u_t dt + \sigma^2 dW_t \quad (105)$$

$$dM_t = [\arctan\left(\frac{X_t + 1}{Y_t - 1}\right), \arctan\left(\frac{X_t - 2}{Y_t - 1}\right)]^T + \eta_t \quad (106)$$

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 , and σ is the noise that perturbs the motion and control actions. Assume that we do not have direct observations on the robot. Instead, we use two detectors located on different observation platforms at $(-1, 1)$ and $(2, 1)$ to collect bearing angles of the target robot as indirect observations. Thus, we have the observation process M_t . Given the expected path S^* , the car should follow it and arrive at the terminal position on time. The performance cost functional based on observational data that we aim to minimize is defined as:

$$J[u] = E \left[\frac{1}{2} \int_0^T \langle R(S_t - S_t^*), (S_t - S_t^*) \rangle dt + \frac{1}{2} \int_0^T \langle Ku_t, u_t \rangle dt + \langle Q(S_T - S_T^*), (S_T - S_T^*) \rangle \right] \quad (107)$$

In our numerical experiments, we let the car start from $(X_0, Y_0) = (0, 0)$ to $(X_T, Y_T) = (1, 1)$. The expected path S_t^* is $X_t^2 + Y_t^2 = 1$. Other settings are $T = 1$, $\Delta t = 0.02$, i.e., $N_T = 50$, $\sigma = 0.1$, $\eta_t \sim N(0, 0.1)$, $L = 1000$, $K = 1$, and the initial heading direction is $\pi/2$. To emphasize the importance of following the expected path and arriving at the target location at the terminal time, let $R = Q = 20$.

In Figure 5, we plot our algorithm's designed trajectory and the estimated trajectory. We can see from this figure that the car moves towards the target along the designed path and is "on target" at the final time with a very small error.

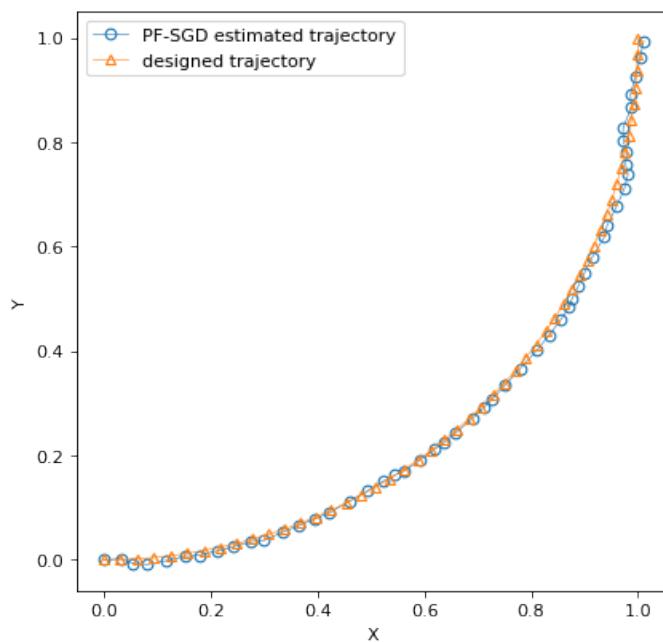


Figure 5. Controlled trajectory from $(0,0)$ to $(1,1)$.

We set $L = 10^3$ and we just increase the number of particles $S = \{2, 8, 32, 128, 512, 1024, 2048, 4096, 8192, 16384, 32768\}$. To provide the convergence of our algorithm in solving this Dubins vehicle maneuvering problem, we repeated the above experiment 50 times and we obtained the error $= \frac{1}{M_{rept.}} \sum_{m=1}^{M_{rept.}} \langle (S_t - S_t^*), (S_t - S_t^*) \rangle$ in Figure 6 where $M_{rept.} = 50$.

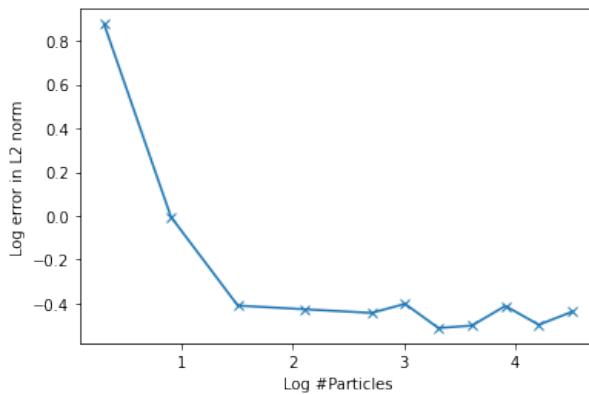


Figure 6. Error vs. number of particles.

Set the number of particles $S = \{8, 16, 32, 64, 128, 256, 512, 1024\}$, and $L = S^2$. We obtained the error $= \frac{1}{M_{rept.}} \sum_{m=1}^{M_{rept.}} \langle (S_t - S_t^*), (S_t - S_t^*) \rangle$ in Figure 7, where the error is the average of $\|S_t - S_t^*\|_2$.

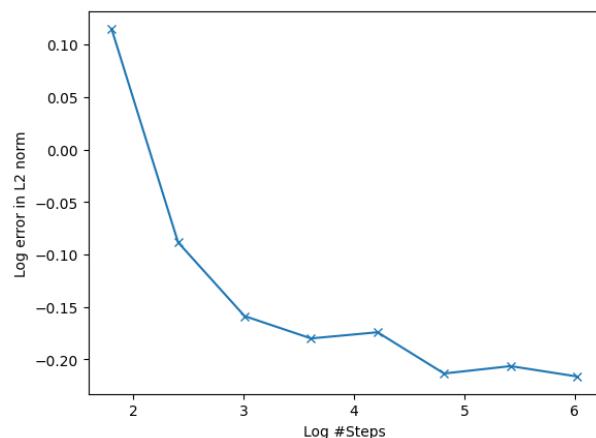


Figure 7. Error vs. number of steps.

From the results above, we can see that the error will decrease and converge as we increase the number of particles and the number of iterations.

5. Conclusions

In this paper, we present the weak convergence of the data-driven feedback control algorithm proposed in [1]. We do not discuss the convergence rate due to the challenge of determining the radius M of the compact subspace that bounds all particles X_i . However, in practice, given a terminal time T , one can use Monte Carlo simulations to find an M that satisfies a certain probability in Lemma 2. Our numerical experiments indicate that both the estimated control and estimated distribution converge at a rate related to the number of particles and iterations.

Future work can focus on analyzing the convergence rate and error bounds for a given state system. This will provide clarity on the number of particles and iterations required to achieve the desired estimation accuracy when applying the algorithm from [1].

Author Contributions: Conceptualization, S.L., H.S. and F.B.; methodology, F.B. and R.A.; software, S.L.; validation, S.L. and F.B.; formal analysis, S.L. and H.S.; investigation, S.L.; resources, S.L.; data curation, S.L.; writing—original draft preparation, S.L. and H.S.; writing—review and editing, F.B. and R.A.; visualization, S.L.; supervision, F.B.; project administration, F.B.; funding acquisition, F.B. and R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by U.S. Department of Energy through FASTMath Institute and Office of Science, Advanced Scientific Computing Research program under the grant DE-SC0022297. FB would also like to acknowledge the support from U.S. National Science Foundation through project DMS-2142672.

Data Availability Statement: All code written as part of this study will be made available on GitHub upon completing the peer review process for this article.

Conflicts of Interest: Author Hui Sun was employed by the company Citigroup Inc. The research work performed by the author does not represent any corporate opinion. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BSDEs	Backward stochastic differential equations
FBSDEs	Forward–backward stochastic differential equations
PF	Particle filter
SGD	Stochastic gradient descent

Appendix A

Appendix A.1

Proof of Lemma 2.

Proof. We start with time t_0 .

Step 1. Starting from $X_0 \sim \xi$ with $\mathbb{E}[\xi^2] \leq C_0$, and by fixing an arbitrary control u_0 , we have for the prediction step:

$$\begin{aligned}\mathbb{E}[|X_1^-|^2] &= \mathbb{E}[|X_0 + b(X_0, u_0)\Delta t + \sigma(X_0)\Delta W_0|^2] \\ &\leq \mathbb{E}[(1 + \Delta t)X_0^2 + (1 + \frac{1}{\Delta t})b^2(\Delta t)^2 + C_\sigma^2\Delta t] + C_\sigma^2\Delta t \\ &\leq (1 + \Delta t)C_0^2 + (C_b^2(\Delta t + 1) + C_\sigma^2)\Delta t \\ &:= C_0^-\end{aligned}\tag{A1}$$

Step 2. We denote the distribution $\mathcal{L}(X_1^-) \sim \pi_{t_1|t_0}$, and then the particle method will perform a random resampling from such a distribution and obtain a random distribution

$$\frac{1}{N} \sum_{i=1}^N \delta_{x_i(\omega)} := \pi_{t_1|t_0}^N\tag{A2}$$

Hence, we have for $X \sim \pi_{t_1|t_0}^N$, taking the expectation where the expectation is taken over all randomness in the measure

$$\mathbb{E}[X^2] = \mathbb{E}[\mathbb{E}[X^2 | \mathcal{G}_1^-]] = \frac{N}{N} \mathbb{E}[\mathbb{E}[x_1^2 | \mathcal{G}_1^-]] = \mathbb{E}[\tilde{X}^2]\tag{A3}$$

where $x_i \sim \pi_{t_1|t_0}$ are i.i.d random samples, \mathcal{G}_0 contains the sampling randomness and $\tilde{X} \sim \pi_{t_1|t_0}$. The conditional expectation is meant to show that all the particles ξ are conditionally independent (since there is other randomness that has accumulated in the history if we want to apply this argument recursively.) Thus, by (A1)

$$\mathbb{E}[\tilde{X}^2] \leq C_0^-\tag{A4}$$

Step 3. We now have the random measure $\pi_{t_1|t_0}^N$, and we proceed to the analysis step. We have by definition

$$X_1 \sim \frac{g(x)d\pi_{t_1|t_0}^N(x)}{\int g(x)d\pi_{t_1|t_0}^N(x)} := \tilde{\pi}_{t_1|t_0}^N\tag{A5}$$

where $\pi_{t_1|t_0}^N(x)$ is the distribution of the terminal state X_1^- from the previous step. Recalling assumption 7, we give an estimate over $\mathbb{E}[|X_1|^2]$:

$$\mathbb{E}[|X_1|^2] \leq (\frac{1}{\kappa})^2 \mathbb{E}[\int x^2 d\pi_{t_1|t_0}^N(x)] \leq (\frac{1}{\kappa})^2 C_0^- := C_1\tag{A6}$$

Step 4. Now, we again apply the random sampling step

$$\frac{1}{N} \sum_{i=1}^N \delta_{x_i(\omega)} := \pi_{t_1|t_1}^N\tag{A7}$$

where $x_i(\omega) \sim \tilde{\pi}_{t_1|t_1}^N$. Then, for $X \sim \pi_{t_1|t_1}^N$, we have

$$\begin{aligned}\mathbb{E}[X^2] &= \mathbb{E}[\mathbb{E}[X^2|\mathcal{G}_1]] \\ &= \frac{\mathcal{N}}{\mathcal{N}}\mathbb{E}[\mathbb{E}[x_i^2|\mathcal{G}_1]] \\ &= \mathbb{E}[\tilde{X}^2]\end{aligned}\quad (\text{A8})$$

where $\tilde{X} \sim \tilde{\pi}_{t_1|t_1}^N$ and \mathcal{G}_1 is the filtration that builds on \mathcal{G}_1^- and the randomness of the current sampling. Then, by (A6), we have

$$\mathbb{E}[X^2] \leq C_1, \quad X \sim \tilde{\pi}_{t_1|t_1}^N \quad (\text{A9})$$

and this completes all the estimates for the first time-stepping. Hence, after one time step, we have

$$C_1 = \kappa^{-2}(1 + \Delta t)C_0 + C\Delta t \quad (\text{A10})$$

which means that by applying the same argument, we will have the following recursion in general:

$$C_{n+1} = \kappa^{-2}(1 + \Delta t)C_n + C\Delta t \quad (\text{A11})$$

As a result, by picking arbitrary u_n , using this same argument repeatedly until N , we have that for all $n = 1, \dots, N$:

$$C_n = (\kappa^{-2}(1 + \Delta t))^n C_0 + \sum_{i=0}^{n-1} (\kappa^{-2}(1 + \Delta t))^i C \Delta t \quad (\text{A12})$$

and we notice that C_n is increasing in n . As a result, we know that for any $X_n \sim \pi_{t_n|t_{n-1}}^N, \pi_{t_n|t_n}^N$, we have that

$$\mathbb{E}[|X_n|^2] \leq C_N \quad (\text{A13})$$

Hence, by Chebyshev's inequality, we have

$$\mathbb{P}(|X_n| \geq M) \leq \frac{C_N}{M^2}, \quad \forall n \in \{1, 2, \dots, N\} \quad (\text{A14})$$

and then we have that

$$\mathbb{P}(\sup_n |X_n| \geq M) \leq \frac{C_N}{M^2} \quad (\text{A15})$$

By noticing that the control values are arbitrarily picked, we have that

$$\mathbb{P}(\sup_{n, U_0} |X_n| \geq M) \leq \frac{C}{M^2}, \quad X_n \sim \pi_{t_n|t_{n-1}}^N \text{ or } \pi_{t_n|t_n}^N \quad (\text{A16})$$

□

References

1. Archibald, R.; Bao, F.; Yong, J.; Zhou, T. An efficient numerical algorithm for solving data driven feedback control problems. *J. Sci. Comput.* **2020**, *85*, 51. [[CrossRef](#)]
2. Bellman, R. Dynamic programming. *Science* **1966**, *153*, 34–37. [[CrossRef](#)] [[PubMed](#)]
3. Feng, X.; Glowinski, R.; Neilan, M. Recent developments in numerical methods for fully nonlinear second order partial differential equations. *SIAM Rev.* **2013**, *55*, 205–267. [[CrossRef](#)]
4. Peng, S. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.* **1990**, *28*, 966–979. [[CrossRef](#)]

5. Yong, J.; Zhou, X.Y. *Stochastic Controls: Hamiltonian Systems and HJB Equations*; Springer Science & Business Media: Cham, Switzerland, 2012.
6. Gong, B.; Liu, W.; Tang, T.; Zhao, W.; Zhou, T. An efficient gradient projection method for stochastic optimal control problems. *SIAM J. Numer. Anal.* **2017**, *55*, 2982–3005. [\[CrossRef\]](#)
7. Tang, S. The maximum principle for partially observed optimal control of stochastic differential equations. *SIAM J. Control Optim.* **1998**, *36*, 1596–1617. [\[CrossRef\]](#)
8. Zhang, J. A numerical scheme for BSDEs. *Ann. Appl. Probab.* **2004**, *14*, 459–488. [\[CrossRef\]](#)
9. Zhao, W.; Fu, Y.; Zhou, T. New kinds of high-order multistep schemes for coupled forward backward stochastic differential equations. *SIAM J. Sci. Comput.* **2014**, *36*, A1731–A1751. [\[CrossRef\]](#)
10. Archibald, R.; Bao, F.; Yong, J. A stochastic gradient descent approach for stochastic optimal control. *East Asian J. Appl. Math.* **2020**, *10*, 635–658. [\[CrossRef\]](#)
11. Sato, I.; Nakagawa, H. Approximation analysis of stochastic gradient Langevin dynamics by using Fokker-Planck equation and Ito process. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; Volume 32, pp. 982–990.
12. Shapiro, A.; Wardi, Y. Convergence analysis of gradient descent stochastic algorithms. *J. Optim. Theory Appl.* **1996**, *91*, 439–454. [\[CrossRef\]](#)
13. Archibald, R.; Bao, F.; Cao, Y.; Sun, H. Numerical analysis for convergence of a sample-wise backpropagation method for training stochastic neural networks. *SIAM J. Numer. Anal.* **2024**, *62*, 593–621. [\[CrossRef\]](#)
14. Zakai, M. On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **1969**, *11*, 230–243. [\[CrossRef\]](#)
15. Gordon, N.J.; Salmond, D.J.; Smith, A.F. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F (Radar Signal Process.)* **1993**, *140*, 107–113. [\[CrossRef\]](#)
16. Morzfeld, M.; Tu, X.; Atkins, E.; Chorin, A.J. A random map implementation of implicit filters. *J. Comput. Phys.* **2012**, *231*, 2049–2066. [\[CrossRef\]](#)
17. Andrieu, C.; Doucet, A.; Holenstein, R. Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* **2010**, *72*, 269–342. [\[CrossRef\]](#)
18. Crisan, D.; Doucet, A. A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **2002**, *50*, 736–746. [\[CrossRef\]](#)
19. Künsch, H.R. Particle filters. *Bernoulli* **2013**, *19*, 1391–1403. [\[CrossRef\]](#)
20. Bao, F.; Cao, Y.; Meir, A.; Zhao, W. A first order scheme for backward doubly stochastic differential equations. *SIAM/ASA J. Uncertain. Quantif.* **2016**, *4*, 413–445. [\[CrossRef\]](#)
21. Zhao, W.; Zhou, T.; Kong, T. High order numerical schemes for second-order FBSDEs with applications to stochastic optimal control. *Commun. Comput. Phys.* **2017**, *21*, 808–834. [\[CrossRef\]](#)
22. Law, K.; Stuart, A.; Zygalakis, K. *Data Assimilation*; Springer: Cham, Switzerland, 2015.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.