Deep Learning Methods for Mean Field Control Problems With Delay

Jean-Pierre Fouque* and Zhaoyu Zhang

Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, CA, United States

We consider a general class of mean field control problems described by stochastic delayed differential equations of McKean–Vlasov type. Two numerical algorithms are provided based on deep learning techniques, one is to directly parameterize the optimal control using neural networks, the other is based on numerically solving the McKean–Vlasov forward anticipated backward stochastic differential equation (MV-FABSDE) system. In addition, we establish the necessary and sufficient stochastic maximum principle of this class of mean field control problems with delay based on the differential calculus on function of measures, and the existence and uniqueness results are proved for the associated MV-FABSDE system under suitable conditions.

Mathematical Subject Classification (2000): 93E20, 60G99, 68-04

Keywords: deep learning, mean field control, delay, McKean-Vlasov, stochastic maximum principal

1. INTRODUCTION

Stochastic games were introduced to study the optimal behaviors of agents interacting with each other. They are used to study the topic of systemic risk in the context of finance. For example, in Carmona et al. [1], the authors proposed a linear quadratic inter-bank borrowing and lending model, and solved explicitly for the Nash equilibrium with a finite number of players. Later, this model was extended in Carmona et al. [2] by considering delay in the control in the state dynamic to account for the debt repayment. The authors analyzed the problem via a probabilistic approach which relies on stochastic maximum principle, as well as via an analytic approach which is built on top of an infinite dimensional dynamic programming principle.

Both mean field control and mean field games are used to characterize the asymptotic behavior of a stochastic game as the number of players grows to infinity under the assumption that all the agents behave similarly, but with different notion of equilibrium. The mean field games consist of solving a standard control problem, where the flow of measures is fixed, and solving a fixed point problem such that this flow of measures matches the distribution of the dynamic of a representative agent. Whereas, the mean field control problem is a non-standard control problem in the sense that the law of state is present in the McKean–Vlasov dynamic, and optimization is performed while imposing the constraint of distribution of the state. More details can be found in Carmona and Delarue [3] and Bensoussan et al. [4].

In this paper, we considered a general class of mean field control problem with delay effect in the McKean–Vlasov dynamic. We derived the adjoint process associated with the McKean–Vlasov stochastic delayed differential equation, which is an anticipated backward stochastic differential equation of McKean–Vlasov type due to the fact that the conditional expectation of the future of adjoint process as well as the distribution of the state dynamic are involved. This type of anticipated backward stochastic differential equations (BSDE) was introduced in Peng and Yang [5], and for the general theory of BSDE, we refer Zhang [6]. The necessary and sufficient part of stochastic maximum principle for control problem with delay in state and control can be found in Chen and Wu [7]. Here, we also establish a necessary and sufficient stochastic maximum principle based

on differential calculus on functions of measures as we consider the delay in the distribution. In the meantime, we also prove the existence and uniqueness of the system of McKean–Vlasov forward anticipated backward stochastic differential equations (MV-FABSDE) under some suitable conditions using the method of continuation, which can be found in Zhang [6], Peng and Wu [8], Bensoussan et al. [9], and Carmona et al. [2]. For a comprehensive study of FBSDE theory, we refer to Ma and Yong [10].

When there was no delay effect in the dynamic, Ma et al. [11] proved the relation between the solution to the FBSDE and quasi-linear partial differential equation (PDE) via "Four Step Scheme." E et al. [12] and Raissi [13] explored the use of deep learning for solving these PDEs in high dimensions. However, the class of fully coupled MV-FABSDE considered in our paper has no explicit solution. Here, we present one algorithm to tackle the above problem by means of deep learning techniques. Due to the non-Markovian nature of the state dynamic, we apply the long short-term memory (LSTM) network, which is able to capture the arbitrary long-term dependencies in the data sequence. It also partially solves the vanishing gradient problem in vanilla recurrent neural networks (RNNs), as was shown in Hochreiter and Schmidhuber [14]. The idea of our algorithm is to approximate the solution to the adjoint process and the conditional expectation of the adjoint process. The optimal control is readily obtained after the MV-FABSDE being solved. We may also emphasize that the way that we present here for numerically computing conditional expectation may have a wide range of applications, and it is simple to implement. We also present another algorithm solving the mean field control problem by directly parameterizing the optimal control. Similar idea can be found in the policy gradient method in the regime of reinforcement learning [15] as well as in Han and E [16]. Numerically, the two algorithms that we propose in this paper yield the same results. Besides, our approaches are benchmarked to the case with no delay for which we have explicit solutions.

The paper is organized as follows. We start with an N-player game with delay, and let number of players goes to infinity to introduce a mean field control problem in section 2. Next, in section 3, we mathematically formulate the feedforward neural networks and LSTM networks, and we propose two algorithms to numerically solve the mean field control problem with delay using deep learning techniques. This is illustrated on a simple linear-quadratic toy model, however with delay in the control. One algorithm is based on directly parameterizing the control, and the other depends on numerically solving the MV-FABSDE system. In addition, we also provide an example of solving a linear quadratic mean field control problem with no delay both analytically, and numerically. The adjoint process associated with the delayed dynamic is derived, as well as the stochastic maximum principle is proved in section 4. Finally, the uniqueness and existence solution for this class of MV-FABSDE are proved under suitable assumptions via continuation method in section 5.

2. FORMULATION OF THE PROBLEM

We consider an N-player game with delay in both state and control. The dynamic $(X_t^i)_{0 \le t \le T}$ for player $i \in \{1, \ldots, N\}$ is given

by a stochastic delayed differential equation (SDDE),

$$
\begin{aligned}
dX_t^i &= b^i(t, \mathbf{X}_t, \mathbf{X}_{t-\tau}, \alpha_t^i, \alpha_{t-\tau}^i)dt + \sigma^i(t, \mathbf{X}_t, \mathbf{X}_{t-\tau}, \alpha_t^i, \alpha_{t-\tau}^i)dW_t^i, \\
&\quad t \in (0, T], \\
X_0^i &= x_0^i, \\
X_t^i &= \alpha_t^i = 0; \quad t \in [-\tau, 0),
\end{aligned}
\tag{2.1}
$$

for $T > 0$, $\tau > 0$ given constants, where $\mathbf{X}_t = (X_t^1, \cdots, X_t^N)$, and where $((W_t^i)_{t \in [0,T]})_{i=1,\cdots,N}$ are N independent Brownian motions defined on the space $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{0 \le t \le T}$ being the natural filtration of Brownian motions.

$$
(b^i, \sigma^i) : [0, T] \times \times \mathbb{R}^N \times \mathbb{R}^N \times A \times A \to \mathbb{R} \times \mathbb{R},
$$

are progressively measurable functions with values in $\mathbb{R}$. We denote A a closed convex subset of $\mathbb{R}$, the set of actions that player i can take, and denote $\mathbb{A}$ the set of admissible control processes. For each $i \in \{1, \ldots, N\}$, A-valued measurable processes $(\alpha_t^i)_{0 \le t \le T}$ satisfy an integrability condition such that $\mathbb{E}\left[\int_{-\tau}^{T} |\alpha_t^i|^2 dt\right] < +\infty$.

Given an initial condition $\mathbf{x}_0 = (x_0^1, \cdots, x_0^N) \in \mathbb{R}^N$, each player would like to minimize his objective functional:

$$
J^i(\boldsymbol{\alpha}) = \mathbb{E}\left[\int_0^T f^i(t, \mathbf{X}_t, \mathbf{X}_{t-\tau}, \alpha_t^i)dt + g^i(\mathbf{X}_T)\right],
\tag{2.2}
$$

for some Borel measurable functions $f^i : [0, T] \times \mathbb{R}^N \times \mathbb{R}^N \times A \to \mathbb{R}$, and $g^i : \mathbb{R}^N \to \mathbb{R}$.

In order to study the mean-field limit of $(\mathbf{X}_t)_{t \in [0,T]}$, we assume that the system (2.1) satisfy a symmetric property, that is to say, for each player i, the other players are indistinguishable. Therefore, drift b^i and volatility σ^i in (2.1) take the form of

$$
(b^i, \sigma^i)(t, \mathbf{X}_t, \mathbf{X}_{t-\tau}, \alpha_t^i, \alpha_{t-\tau}^i) = (b^i, \sigma^i)(t, X_t^i, \mu_t^N, X_{t-\tau}^i, \mu_{t-\tau}^N, \alpha_t^i, \alpha_{t-\tau}^i),
$$

and the running cost f^i and terminal cost g^i are of the form

$$
\begin{aligned}
f^i(t, \mathbf{X}_t, \mathbf{X}_{t-\tau}, \alpha_t^i) &= f^i(t, X_t^i, \mu_t^N, X_{t-\tau}^i, \mu_{t-\tau}^N, \alpha_t^i) \text{ and } g^i(\mathbf{X}_T) \\
&= g^i(X_T^i, \mu_T^N),
\end{aligned}
$$

where we use the notation μ_t^N for the empirical distribution of $\mathbf{X} = (X^1, \cdots, X^N)$ at time t, which is defined as

$$
\mu_t^N = \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^j}.
$$

Next, we let the number of players N go to $+\infty$ before we perform the optimization. According to symmetry property and the theory of propagation of chaos, the joint distribution of the N dimensional process $(\mathbf{X}_t)_{0 \le t \le T} = (X_t^1, \ldots, X_t^N)_{0 \le t \le T}$ converges to a product distribution, and the distribution of each single marginal process converges to the distribution of $(X_t)_{0 \le t \le T}$ of the following McKean–Vlasov stochastic delayed differential equation (MV-SDDE). For more detail on the

argument without delay, we refer to Carmona and Delarue [3] and Carmona et al. [17].

$$
\begin{aligned}
dX_t &= b(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau})dt \\
&\quad + \sigma(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau})dW_t, \quad t \in (0, T], \\
X_0 &= x_0, \\
X_t &= \alpha_t = 0, \quad t \in [-\tau, 0).
\end{aligned} \tag{2.3}
$$

We then optimize after taking the limit. The objective for each player of (2.2) now becomes

$$
J(\alpha) = \mathbb{E}\left[\int_0^T f(X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t)dt + g(X_T, \mu_T) \right], \tag{2.4}
$$

where we denote $\mu_t := \mathcal{L}(X_t)$ the law of X_t.

3. SOLVING MEAN-FIELD CONTROL PROBLEMS USING DEEP LEARNING TECHNIQUES

Due to the non-Markovian structure, the above mean-field optimal control problem (2.3 and 2.4) is difficult to solve either analytically or numerically. Here we propose two algorithms together with four approaches to tackle the above problem based on deep learning techniques. We would like to use two types of neural networks, one is called the feedforward neural network, and the other one is called Long Short-Term Memory (LSTM) network.

For a feedforward neural network, we first define the set of layers $\mathbb{M}_{d,h}^\rho$, for $x \in \mathbb{R}^d$, as

$$
\mathbb{M}_{d,h}^\rho := \{M : \mathbb{R}^d \to \mathbb{R}^h | M(x) = \rho(Ax + b), A \in \mathbb{R}^{h \times d}, b \in \mathbb{R}^h\}. \tag{3.1}
$$

d is called input dimension, h is known as the number of hidden neurons, $A \in \mathbb{R}^{h \times d}$ is the weight matrix, $b \in \mathbb{R}^h$ is the bias vector, and ρ is called the activation function. The following activation functions will be used in this paper, for some $x \in \mathbb{R}$,

$$
\rho_{ReLU}(x) := x^+ = \max(0, x); \quad \rho_s(x) := \frac{1}{1+e^{-x}};
$$
$$
\rho_{\tanh}(x) := \tanh(x); \quad \rho_{Id}(x) := x.
$$

Then feedforward neural network is defined as a composition of layers, so that the set of feedforward neural networks with l hidden layers we use in this paper is defined as

$$
\begin{aligned}
\mathbb{NN}_{d_1,d_2}^l &= \{\tilde{M} : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2} | \tilde{M} = M_l \circ \cdots \circ M_1 \circ M_0, \\
M_0 &\in \mathbb{M}_{d_1,h_1}^{\rho_{ReLU}}, M_l \in \mathbb{M}_{h_l,d_2}^{\rho_{Id}}, M_i \in \mathbb{M}_{h_i,h_{i+1}}^{\rho_{ReLU}}, h. \in \mathbb{Z}^+, i = 1, \ldots, l-1\}.
\end{aligned} \tag{3.2}
$$

The LSTM network is one of RNN architectures, which are powerful for capturing long-range dependence of the data. It is proposed in Hochreiter and Schmidhuber [14], and it is designed to solve the shrinking gradient effects which basic RNN often suffers from. The LSTM network is a chain of

cells. Each LSTM cell is composed of a cell state, which contains information, and three gates, which regulate the flow of information. Mathematically, the rule inside tth cell follows,

$$
\begin{aligned}
\Gamma_{f_t} &= \rho_s(A_f x_t + U_f a_{t-1} + b_f), \\
\Gamma_{i_t} &= \rho_s(A_i x_t + U_i a_{t-1} + b_i), \\
\Gamma_{o_t} &= \rho_s(A_o x_t + U_o a_{t-1} + b_o), \\
c_t &= \Gamma_{f_t} \odot c_{t-1} + \Gamma_{i_t} \odot \rho_{\tanh}(A_c x_t + U_c a_{t-1} + b_c), \\
a_t &= \Gamma_{o_t} \odot \rho_{\tanh}(c_t),
\end{aligned} \tag{3.3}
$$

where the operator $\odot$ denotes the Hadamard product. $(\Gamma_{f_t}, \Gamma_{i_t}, \Gamma_{o_t}) \in \mathbb{R}^h \times \mathbb{R}^h \times \mathbb{R}^h$ represents forget gate, input gate and output gate, respectively, h refers the number of hidden neurons. $x_t \in \mathbb{R}^d$ is the input vector with d features. $a_t \in \mathbb{R}^h$ is known as the output vector with initial value $a_0 = 0$, and $c_t \in \mathbb{R}^h$ is known as the cell state with initial value $c_0 = 0$. $A. \in \mathbb{R}^{h \times d}$ are the weight matrices connecting input and hidden layers, $U. \in \mathbb{R}^{h \times h}$ are the weight matrix connecting hidden and output layers, and $b \in \mathbb{R}^h$ represents bias vector. The weight matrices and bias vectors are shared through all time steps, and are going to be learned during training process by back-propagation through time (BPTT), which can be implemented in Tensorflow platform. Here we define the set of LSTM network up to time t as

$$
\mathbb{LSTM}_{d,h,t} = \Big\{ M : (\mathbb{R}^d)^t \times \mathbb{R}^h \times \mathbb{R}^h \to \mathbb{R}^h \times \mathbb{R}^h \mid M(x_0, \ldots, x_t,
$$
$$
a_0, c_0) = (a_t, c_t), c_t = \Gamma_{f_t} \odot c_{t-1} + \Gamma_{i_t} \odot \rho_{\tanh}(A_c x_t + U_c a_{t-1} + b_c),
$$
$$
a_t = \Gamma_{o_t} \odot \rho_{\tanh}(c_t), a_0 = c_0 = 0 \Big\}, \tag{3.4}
$$

where $\Gamma_{f.}, \Gamma_{i.}, \Gamma_{o.}$ are defined in (3.3).

In particular, we specify the model in a linear-quadratic form, which is inspired by Carmona et al. [2] and Fouque and Zhang [18]. The objective function is defined as

$$
J(\alpha) = \mathbb{E}\left[\int_0^T \left(\frac{1}{2}\alpha_t^2 + \frac{c_f}{2}(X_t - m_t)^2 \right) dt + \frac{c_t}{2}(X_T - m_T)^2 \right], \tag{3.5}
$$

subject to

$$
\begin{aligned}
dX_t &= (\alpha_t - \alpha_{t-\tau})dt + \sigma dW_t, \quad t \in [0, T], \\
X_0 &= x_0, \\
X_t &= \alpha_t = 0, \quad t \in [-\tau, 0),
\end{aligned} \tag{3.6}
$$

where $\sigma, c_f, c_t > 0$ are given constants, and $m_t := \int_\mathbb{R} x d\mu_t(x)$ denotes the mean of X at time t, and $\mu_t := \mathcal{L}(X_t)$. In the following subsections, we solve the above problem numerically using two algorithms together with four approaches. The first two approaches are to directly approximate the control by either a LSTM network or a feedforward neural network, and minimize the objective (3.5) using stochastic gradient descent algorithm. The third and fourth approaches are to introduce the adjoint process associated with (3.6), and approximate the adjoint process and the conditional expectation of adjoint process using neural networks.

3.1. Approximating the Optimal Control Using Neural Networks

We first set $\Delta t = T/N = \tau/D$ for some positive integer N. The time discretization becomes

$$-\tau = t_{-D} \leq t_{-D+1} \leq \cdots \leq t_0 = 0 = t_0 \leq t_1 \leq \cdots \leq t_N \leq T,$$

for $t_i - t_{i-1} = \Delta t$, $i \in \{-D+1, \cdots, 0, \cdots, N-1, N\}$. The discretized SDDE (3.6) according to Euler-Maruyama scheme now reads

$$X_{t_{i+1}} = X_{t_i} + (\alpha_{t_i} - \alpha_{t_{i-D}})\Delta t + \sigma\sqrt{\Delta t}\Delta W_{t_i}, \text{ for } i \in \{0, \cdots, N-1\}, \quad (3.7)$$

where $(\Delta W_{t_i})_{0 \leq i \leq N-1}$ are independent, normal distributed sequence of random variables with mean 0 and variance 1.

First, from the definition of open loop control, and due to non-Markovian feature of (3.6), the open-loop optimal control is a function of the path of the Brownian motions up to time t, i.e., $\alpha(t, (W_s)_{0 \leq s \leq t})$. We are able to describe this dependency by a LSTM network by parametrizing the control as a function of current time and the discretized increments of Brownian motion path, i.e.,

$$(a_{t_i}, c_{t_i}) = \varphi^1(t_i, (\Delta W_s)_{0 \leq s \leq t_i}|\Phi^1) \text{ for } \varphi^1 \in \mathbb{LSTM}_{2,h_1,t}$$
$$\text{and } \Phi^1 = (A_f, A_i, A_o, A_c, U_f, U_i, U_o, U_c, b_f, b_i, b_o, b_c),$$

$$\alpha(t_i, (W_s)_{0 \leq s \leq t_i}) \approx \psi^1(a_{t_i}|\Psi^1) \text{ for } \psi^1 \in \mathbb{M}_{h_1,1}^{Id} \text{ and } \Psi^1 = (A, b), \quad (3.8)$$

for some $h_1 \in \mathbb{Z}^+$. We remark that the last dense layer is used to match the desired output dimension.

The second approach is again directly approximate the control but with a feedforward neural network. Due to the special structure of our model, where the mean of dynamic in (3.6) is constant, the mean field control problem coincides with the mean field game problem. In Fouque and Zhang [18], authors solved the associated mean field game problem using infinite dimensional PDE approach, and found that the optimal control is a function of current state and the past of control. Therefore, the feedforward neural network with l layers, which we use to approximate the optimal control, is defined as

$$\alpha_{t_i}(X_{t_i}, (\alpha_s)_{t_{i-D} \leq s < t_i}) \approx \psi^2(X_{t_i}, (\alpha_s)_{t_{i-D} \leq s \leq t_i}|\Psi^2)$$
$$\text{for } \psi^2 \in \mathbb{NN}_{D+1,1}^l, \ \Psi^2 = (A_0, b_0, \ldots, A_l, b_l). \quad (3.9)$$

From Monte Carlo algorithm, and trapezoidal rule, the objective function (3.5) now becomes

$$J = \frac{1}{M}\sum_{j=1}^{M}\left[\left(\frac{1}{2}(\alpha_{t_0}^{(j)})^2 + \frac{c_f}{2}(X_{t_0}^{(j)} - \bar{X}_{t_0})^2 + \right.\right.$$
$$\sum_{i=1}^{N-1}\left((\alpha_{t_i}^{(j)})^2 + c_f(X_{t_i}^{(j)} - \bar{X}_{t_i})^2\right)$$
$$\left.+ \frac{1}{2}(\alpha_{t_N}^{(j)})^2 + \frac{c_f}{2}(X_{t_N}^{(j)} - \bar{X}_{t_N})^2\right)\frac{\Delta t}{2} + \left.\frac{c_t}{2}(X_{t_N}^{(j)} - \bar{X}_{t_N})^2\right], \quad (3.10)$$

where M denotes the number of realizations and $\bar{X} := \frac{1}{M}\sum_{j=1}^{M}X^{(j)}$ denotes the sample mean. After plugging in the neural network either given by (3.8) or (3.9), the optimization problem becomes to find the best set of parameters either (Φ^1, Ψ^1) or Ψ^2 such that the objective $J(\Phi^1, \Psi^1)$ or $J(\Psi^2)$ is minimized with respect to those parameters.

The algorithm works as follows:

Algorithm 1: Algorithms for solving mean field control problem with delay by directly approximating the optimal control using neural networks

Initialization of parameters $\Theta_1 = (\Phi^1, \Psi^1)$ for approach 1 (3.8) or $\Theta_1 = (\Psi^2)$ for approach 2 (3.9);

for *each epoch* $e = 1, 2, \ldots$ **do**

- Generate $\Delta W \in \mathbb{R}^{M \times N}$ for $\Delta W_{t_i}^{(j)} := W_{ji} \sim N(0, 1)$, $j \in \{1, \ldots, M\}$ and $i \in \{1, \ldots, N\}$;
- $\alpha_{t_i}^{(j)} = 0$ for $i = \{-D, \ldots, -1\}$, $\forall j$;
- $X_0^{(j)} = \bar{X}_0 = x_0, \alpha_0^{(j)} \approx \varphi_0^{(j)}(\Theta_e), \forall j$, for some network φ given by (3.8) or by (3.9) at t_0 with proper inputs;
- $J = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{2}(\varphi_0^{(j)})^2\frac{\Delta t}{2}$;

for $(i = 0, \ldots, N-1)$ **do**

- $X_{t_{i+1}}^{(j)} = X_{t_i}^{(j)} + (\alpha_{t_i}^{(j)} - \alpha_{t_{i-D}}^{(j)})\Delta t + \sigma\sqrt{\Delta t}\Delta W_{t_i}^{(j)}, \forall j$;
- $\bar{X}_{t_{i+1}} = \frac{1}{M}\sum_{j=1}^{M}X_{t_{i+1}}^{(j)}$;
- $\alpha_{t_{i+1}}^{(j)} \approx \varphi_{t_{i+1}}^{(j)}(\Theta_e), \forall j$, is given by either (3.8) or (3.9) at t_{i+1};

if $(i = N-1)$ **then**

- $J = J + \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{2}(\varphi_{t_{i+1}}^{(j)})^2 + \frac{c_f}{2}(X_{t_{i+1}}^{(j)} - \bar{X}_{t_{i+1}})^2\right)\frac{\Delta t}{2}$;

else

- $J = J + \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{2}(\varphi_{t_{i+1}}^{(j)})^2 + \frac{c_f}{2}(X_{t_{i+1}}^{(j)} - \bar{X}_{t_{i+1}})^2\right)\Delta t$;

end

end

- $J = J + \frac{1}{M}\sum_{j=1}^{M}\frac{c_t}{2}(X_{t_N}^{(j)} - \bar{X}_{t_N})^2, \forall j$;
- Compute the gradient $\nabla J(\Theta_e)$ by backpropagation through time;
- Stop if $J(\Theta_e)$ converges, or $|\nabla J(\Theta_e)| < \delta$ for some threshold δ, and return Θ_e;
- Otherwise, update $\Theta_{e+1} = \Theta_e - \eta\nabla J(\Theta_e)$, according to stochastic gradient descent algorithm, for some learning rate $\eta > 0$ small;

end

In the following graphics, we choose $x_0 = 0, c_f = c_t = 1, \sigma = 1, T = 10, \tau = 4, \Delta t = 0.1, M = 4,000$. For approach 1, the neural network $\varphi \in \mathbb{NN}$, which is defined in (3.2), is composed of 3 hidden layers with $d_1 = 42, h_1 = 64, h_2 = 128, h_3 = 64, d_2 = 1$. For approach 2, the LSTM network $\varphi \in \mathbb{LSTM}$, which is defined in (3.4), consists of 128 hidden neurons. For a specific representative path, the underlying Brownian motion

paths are approximately the same for different approaches. **Figure 1** compares one representative optimal trajectory of the state dynamic and the control, and they coincide. **Figure 2** plots the sample average of optimal trajectory of the state dynamic and the control, which are trajectories of approximately 0, and this is the same as the theoretical mean.

3.2. Approximating the Adjoint Process Using Neural Networks

The third and fourth approaches are based on numerically solving the MV-FABSDE system using LSTM network and feedforward neural networks. From section 4, we derive the adjoint process, and prove the sufficient and necessary parts of stochastic maximum principle. From (4.7), we are able to write the backward stochastic differential equation associated to (3.6) as,

$$dY_t = -c_f(X_t - m_t)dt + Z_t dW_t, \quad t \in [0, T], \quad (3.11)$$

with terminal condition $Y_T = c_t(X_T - m_T)$, and $Y_s = 0$ for $s \in (T, T + \tau]$. The optimal control $(\hat{\alpha}_t)_{0 \le t \le T}$ can be obtained in terms of the adjoint process Y_t from the maximum principle, and it is given by

$$\hat{\alpha}_t = -Y_t + \mathbb{E}[Y_{t+\tau}|\mathcal{F}_t].$$

From the Euler-Maruyama scheme, the discretized version of (3.6) and (3.11) now reads, for $i = \{0, \ldots, N-1\}$,

$$X_{t_{i+1}} = X_{t_i} + (\hat{\alpha}_{t_i} - \hat{\alpha}_{t_{i-D}})\Delta t + \sigma\sqrt{\Delta t}\Delta W_{t_i}, \text{ where } \Delta W_{t_i} \sim N(0, 1).$$

(3.12)

$$Y_{t_{i+1}} = Y_{t_i} - c_f(X_{t_i} - \bar{X}_{t_i})\Delta t + Z_{t_i}\sqrt{\Delta t}\Delta W_{t_i}, \quad (3.13)$$

where we use the sample average $\bar{X}_t = \frac{1}{M}\sum_{j=1}^{M} X_t^{(j)}$ to approximate the expectation of X_t. In order to solve the above MV-FABSDE system, we need to approximate $(Y_{t_i}, \mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}], Z_{t_i})_{0 \le t_i \le t_N}$.

The third approach consists of approximating $(Y_{t_i}, \mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}], Z_{t_i})_{0 \le t_i \le t_N}$ using three LSTM networks as functions of current time and the discretized path of Brownian motions, respectively, i.e.,

$$(a_{t_i}^Y, c_{t_i}^Y) = \varphi^Y(t_i, (\Delta W_s)_{t_0 \le s \le t_i}|\Phi^Y)$$
$$\text{for } \varphi^Y \in \mathbb{LSTM}_{2, h_Y, t_i} \text{ and}$$
$$\Phi^Y = (A_f^Y, A_i^Y, A_o^Y, A_c^Y, U_f^Y, U_i^Y, U_o^Y, U_c^Y, b_f^Y, b_i^Y, b_o^Y, b_c^Y),$$
$$Y_{t_i} \approx \psi^Y(a_{t_i}^Y|\Psi^Y) \text{ for } \psi^Y \in \mathbb{M}_{h_Y, 1}^{Id} \text{ and } \Psi^Y = (A^Y, b^Y),$$

$$(a_{t_i}^{EY}, c_{t_i}^{EY}) = \varphi^{EY}(t_i, (\Delta W_s)_{t_0 \le s \le t_i}|\Phi^{EY})$$
$$\text{for } \varphi^{EY} \in \mathbb{LSTM}_{2, h_{EY}, t_i} \text{ and}$$
$$\Phi^{EY} = (A_f^{EY}, A_i^{EY}, A_o^{EY}, A_c^{EY}, U_f^{EY}, U_i^{EY}, U_o^{EY}, U_c^{EY}, b_f^{EY}, b_i^{EY}, b_o^{EY}, b_c^{EY}),$$
$$\mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}] \approx \psi^{EY}(a_{t_i}^{EY}|\Psi^{EY}) \text{ for } \psi^{EY} \in \mathbb{M}_{h_{EY}, 1}^{Id} \text{ and } \Psi^{EY} = (A^{EY}, b^{EY}),$$

$$(a_{t_i}^Z, c_{t_i}^Z) = \varphi^Z(t_i, (\Delta W_s)_{t_0 \le s \le t_i}|\Phi^Z)$$
$$\text{for } \varphi^Z \in \mathbb{LSTM}_{2, h_Z, t_i}$$
$$\text{and } \Phi^Z = (A_f^Z, A_i^Z, A_o^Z, A_c^Z, U_f^Z, U_i^Z, U_o^Z, U_c^Z, b_f^Z, b_i^Z, b_o^Z, b_c^Z),$$
$$Z_{t_i} \approx \psi^Z(a_{t_i}^Z|\Psi^Z) \text{ for } \psi^Z \in \mathbb{M}_{h_Z, 1}^{Id} \text{ and } \Psi^Z = (A^Z, b^Z), \quad (3.14)$$

for some $h_Y, h_{EY}, h_Z \in \mathbb{Z}^+$. Again, the last dense layers are used to match the desired output dimension.

Since approach 3 consists of three neural networks with large number of parameters, which is hard to train in general, we would like to make the following simplification in approach 4 for approximating $(Y_{t_i}, \mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}], Z_{t_i})_{0 \le t_i \le t_N}$ via combination of one LSTM network and three feedforward neural networks. Specifically,

$$(a_{t_i}, c_{t_i}) = \varphi(t_i, (\Delta W_s)_{t_0 \le s \le t_i}|\Phi)$$
$$\text{for } \varphi \in \mathbb{LSTM}_{2, h, t_i} \text{ and}$$
$$\Phi = (A_f, A_i, A_o, A_c, U_f, U_i, U_o, U_c, b_f, b_i, b_o, b_c),$$
$$Y_{t_i} \approx \psi^Y(a_{t_i}|\Psi^Y) \text{ for } \psi^Y \in \mathbb{NN}_{h, 1}^l \text{ and}$$
$$\Psi^Y = (A_0^Y, b_0^Y, \ldots, A_l^Y, b_l^Y), \mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}] \quad (3.15)$$
$$\approx \psi^{EY}(a_{t_i}|\Psi^{EY}) \text{ for } \psi^{EY} \in \mathbb{NN}_{h, 1}^l \text{ and}$$
$$\Psi^{EY} = (A_0^{EY}, b_0^{EY}, \ldots, A_l^{EY}, b_l^{EY}),$$
$$Z_{t_i} \approx \psi^Z(a_{t_i}|\Psi^Z) \text{ for } \psi^Z \in \mathbb{NN}_{h, 1}^l \text{ and}$$
$$\Psi^Z = (A_0^Z, b_0^Z, \ldots, A_l^Z, b_l^Z).$$

In words, the algorithm works as follows. We first initialize the parameters $(\Theta^Y, \Theta^{EY}, \Theta^Z) = ((\Phi^Y, \Psi^Y), (\Phi^{EY}, \Psi^{EY}), (\Phi^Z, \Psi^Z))$ either in (3.14) or $(\Theta^Y, \Theta^{EY}, \Theta^Z) = ((\Phi, \Psi^Y), \Psi^{EY}), \Psi^Z)$ in (3.15). At time 0, $X_0 = x_0$, $(Y_0, \mathbb{E}[Y_{t_D}|\mathcal{F}_0], Z_0) \approx (\varphi_0^Y(\Theta^Y), \varphi_0^{EY}(\Theta^{EY}), \varphi_0^Z(\Theta^Z))$ for some network $(\varphi^Y, \varphi^{EY}, \varphi^Z)$ given by either (3.14) or (3.15), and $\alpha_0 = -Y_{t_0} + \mathbb{E}[Y_{t_D}|\mathcal{F}_{t_0}]$. Next, we update $X_{t_{i+1}}$ and $Y_{t_{i+1}}$ according to (3.12), and the solution to the backward equation at t_{i+1} is denoted by $\tilde{Y}_{t_{i+1}}$. In the meantime, $Y_{t_{i+1}}$ is also approximated by a neural network. In such case, we refer to $\tilde{Y}.$ as the label, and $Y.$ given by the neural network as the prediction. We would like to minimize the mean square error between these two. At time T, Y_{t_N} is also supposed to match $c_t(X_{t_N} - \bar{X}_{t_N})$, from the terminal condition of (3.11). In addition, the conditional expectation $\mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}]$ given by a neural network should be the best predictor of $\tilde{Y}_{t_{i+D}}$, which implies that we would like to find the set of parameters Θ^{EY} such that $\mathbb{E}[(\tilde{Y}_{t_{i+D}} - \varphi_{t_i}^{EY}(\Theta^{EY}))^2]$ is minimized for all $t_i \in \{t_0, \ldots, t_{N-D}\}$. Therefore, for M samples, we would like to minimize two objective functions L_1 and L_2 defined as

$$L_1(\Theta^Y, \Theta^Z) = \frac{1}{M}\left[\sum_{j=1}^{M}\sum_{i=0}^{N}(\varphi_{t_i}^{Y,(j)} - \tilde{Y}_{t_i}^{(j)})^2 + \sum_{j=1}^{M}\left(\varphi_{t_N}^{Y,(j)} - c_t(X_{t_N}^{(j)} - \bar{X}_{t_N})\right)^2\right], \quad (3.16)$$

$$L_2(\Theta^{EY}) = \frac{1}{M}\sum_{j=1}^{M}\sum_{i=0}^{N-D}\left(\varphi_{t_i}^{EY,(j)} - \tilde{Y}_{t_{i+D}}^{(j)}\right)^2.$$

The algorithm works as follows:

Algorithm 2: Algorithms for solving mean field control problem with delay according to MV-FABSDE

Initialization of parameters $(\Theta_1^Y, \Theta_1^{EY}, \Theta_1^Z)$ for approach 3 as in (3.14) or approach 4 as in (3.15);

for *each epoch* $e = 1, 2, \ldots$ **do**

- Generate $\Delta W \in \mathbb{R}^{M \times N}$ for $\Delta W_{t_i}^{(j)} := \Delta W_{ji} \sim N(0, 1)$, $j \in \{1, \ldots, M\}$ and $i \in \{1, \ldots, N\}$;
- $\alpha_{t_i}^{(j)} = 0$ for $i \in \{-D, \ldots, -1\}, \forall j$;
- $X_0^{(j)} = \bar{X}_0 = x_0, \forall j;$ $\left(Y_0^{(j)}, \mathbb{E}[Y_{t_D}^{(j)}|\mathcal{F}_0], Z_0^{(j)}\right) \approx$ $\left(\varphi_0^{Y,(j)}(\Theta_e^Y), \varphi_0^{EY,(j)}(\Theta_e^{EY}), \varphi_0^{Z,(j)}(\Theta_e^Z)\right)$ given by either (3.14) or by (3.15); $\alpha_0^{(j)} \approx -\varphi_0^{Y,(j)} + \varphi_0^{EY,(j)}$ at t_0;
- $L_1(\Theta_e^Y, \Theta_e^Z) = 0$, $L_2(\Theta_e^{EY}) = 0$;

for $(i = 0, \ldots, N-1)$ **do**

- $X_{t_{i+1}}^{(j)} = X_{t_i}^{(j)} + (\alpha_{t_i}^{(j)} - \alpha_{t_{i-D}}^{(j)})\Delta t + \sigma\sqrt{\Delta t}\Delta W_{t_i}^{(j)}, \forall j;$
- $\bar{X}_{t_{i+1}} = \frac{1}{M}\sum_{j=1}^{M} X_{t_{i+1}}^{(j)};$
- $\tilde{Y}_{t_{i+1}}^{(j)} = Y_{t_i}^{(j)} - c_f(X_{t_i}^{(j)} - \bar{X}_{t_i})\Delta t + Z_{t_i}\sqrt{\Delta t}\Delta W_{t_i}^{(j)}, \forall j;$

  if $(i \leq N - D)$ **then**

  - $\left(Y_{t_{i+1}}^{(j)}, \mathbb{E}[Y_{t_{i+1+D}}^{(j)}|\mathcal{F}_{t_{i+1}}], Z_{t_{i+1}}^{(j)}\right) \approx$ $\left(\varphi_{t_{i+1}}^{Y,(j)}(\Theta_e^Y), \varphi_{t_{i+1}}^{EY,(j)}(\Theta_e^{EY}), \varphi_{t_{i+1}}^{Z,(j)}(\Theta_e^Z)\right), \forall j$ given by (3.14) or by (3.15) at t_{i+1};
  - $L_1 += \frac{1}{M}\sum_{j=1}^{M}\left(\varphi_{t_{i+1}}^{Y,(j)} - \tilde{Y}_{t_{i+1}}^{(j)}\right)^2;$

  else

  - $\left(Y_{t_{i+1}}^{(j)}, \mathbb{E}[Y_{t_{i+1+D}}^{(j)}|\mathcal{F}_{t_{i+1}}], Z_{t_{i+1}}^{(j)}\right) \approx$ $\left(\varphi_{t_{i+1}}^{Y,(j)}(\Theta_e^Y), 0, \varphi_{t_{i+1}}^{Z,(j)}(\Theta_e^Z)\right), \forall j$ given by (3.14) or by (3.15) at t_{i+1};
  - $L_1 += \frac{1}{M}\sum_{j=1}^{M}\left(\varphi_{t_{i+1}}^{Y,(j)} - \tilde{Y}_{t_{i+1}}^{(j)}\right)^2;$

  end

end

for $(i = 0, 1, \ldots, N - D)$ **do**

- $L_2 += \frac{1}{M}\sum_{j=1}^{M}\left(\varphi_{t_i}^{EY,(j)} - \tilde{Y}_{t_{i+D}}^{(j)}\right)^2;$

end

- $L_1 += \frac{1}{M}\sum_{j=1}^{M}\left(\varphi_{t_N}^{Y,(j)} - c_t(X_{t_N}^{(j)} - \bar{X}_{t_N})\right)^2;$
- Compute the gradient $\nabla L_1(\Theta^Y, \Theta^Z)$ and $\nabla L_2(\Theta^{EY})$ by backpropagation through time;
- Stop if $L_1(\Theta^Y, \Theta^Z)$ are close to 0, and $L_2(\Theta^{EY})$ converges, return $(\Theta^Y, \Theta^{EY}, \Theta^Z)$;
- Otherwise, update $\Theta_{e+1}^Y, \Theta_{e+1}^Z$ and Θ_{e+1}^{EY} according to SGD algorithm;

end

Again, in the following graphics, we choose $x_0 = 0, c_f = c_t = 1, \sigma = 1, T = 10, \tau = 4, \Delta t = 0.1, M = 4,000$. In approach 3, each of the three LSTM networks approximating $Y_{t_i}, \mathbb{E}[Y_{t_{i+D}}|\mathcal{F}_{t_i}]$ and Z_{t_i} consists of 128 hidden neurons, respectively. In approach 4, the LSTM consists of 128 hidden neurons, and each of the feedforward neural networks has parameters $d_1 = 128, h_1 = 64, h_2 = 128, h_3 = 64, d_2 = 1$. For a specific representative path, the underlying Brownian motion paths are the same for different approaches. **Figure 3** compares one representative optimal trajectory of the state dynamic and

the control via two approaches, and they coincide. **Figure 4** plots the sample average of optimal trajectory of the dynamic and the control, which are trajectories of 0, which is the same as the theoretical mean. Comparing to **Figures 1**, **2**, as well as based on numerous experiments, we find that given a path of Brownian motion, the two algorithms would yield similar optimal trajectory of state dynamic and similar path for the optimal control. From **Figure 6**, the loss L_1 as defined in (3.16) becomes approximately 0.02 in 1,000 epochs for both approach 3 and approach 4. This can also be observed from **Figure 5**, since the red dash line and the blue solid line coincide for both left and right graphs. In addition, from the righthand side of **Figure 6**, we observe the loss L_2 as defined in (3.16) converges to 50 after 400 epochs. This is due to the fact that the conditional expectation can be understood as an orthogonal projection. **Figure 7** plots 64 sample paths of the process $(Z_{t_i})_{t_0 \leq t_i \leq t_N}$, which seems to be a deterministic function since σ is constant in this example. Finally, **Figure 8** shows the convergence of the value function as number of epochs increases. Both algorithms arrive approximately at the same optimal value which is around 6 after 400 epochs. This confirms that the out control problem has a unique solution. In section 5, we show that the MV-FASBDE system is uniquely solvable. It is also observable that the first algorithm converges faster than the second one, since it directly paramerizes the control using one neural network, instead of solving the MV-FABSDE system, which uses three neural networks.

3.3. Numerically Solving the Optimal Control Problem With No Delay

Since the algorithms we proposed embrace the case with no delay, we illustrate the comparison between numerical results and the analytical results. By letting $\tau > T$ we obtain $\alpha_{t-\tau} = 0$ in (3.6), and we aim at solving the following linear-quadratic mean-field control problem by minimizing

$$J(\alpha) = \mathbb{E}\left[\int_0^T \left(\frac{1}{2}\alpha_t^2 + \frac{c_f}{2}(X_t - m_t)^2\right) dt + \frac{c_t}{2}(X_T - m_T)^2\right], \quad (3.17)$$

subject to

$$\begin{aligned} dX_t &= \alpha_t dt + \sigma dW_t, \quad t \in [0, T], \\ X_0 &= x_0. \end{aligned} \quad (3.18)$$

Again, from section 4, we find the optimal control

$$\hat{\alpha}_t = -Y_t,$$

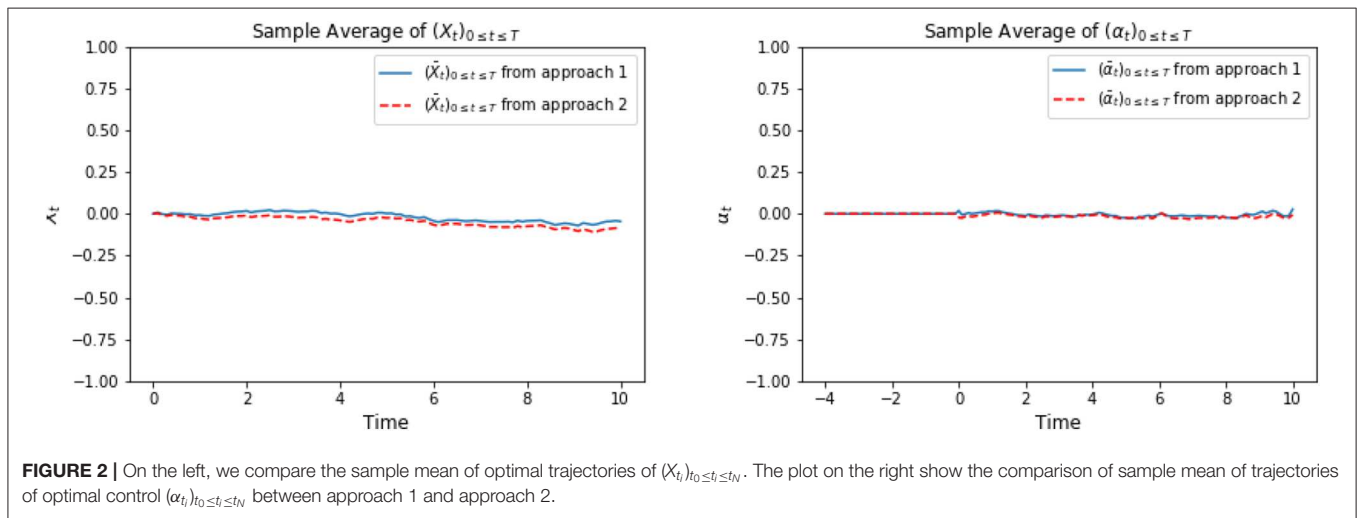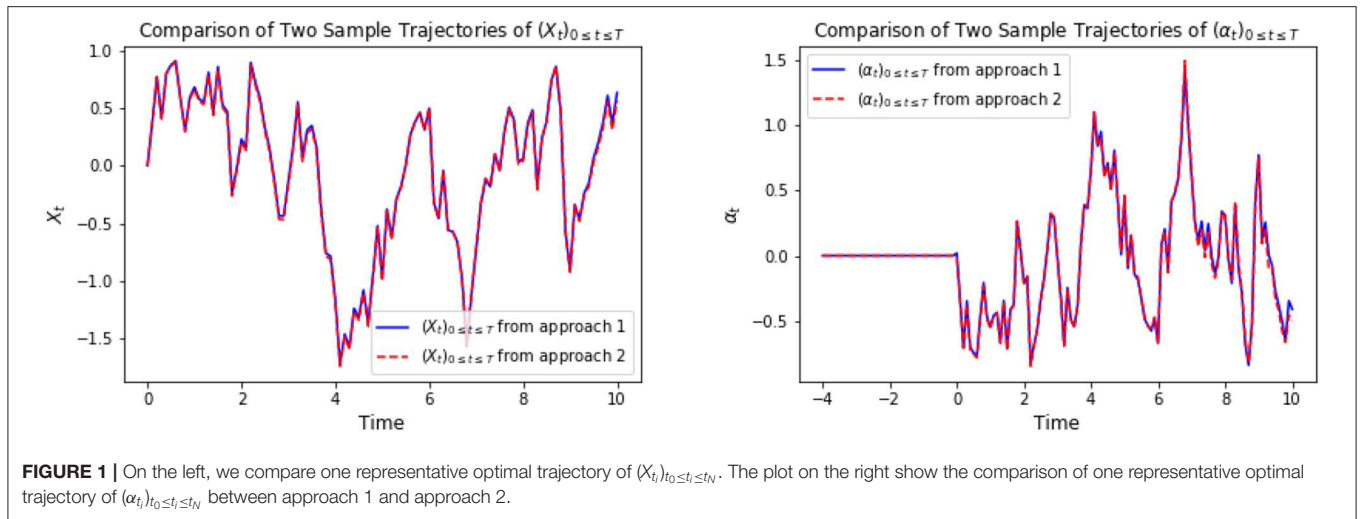where (Y_t, Z_t) is the solution of the following adjoint process,

$$dY_t = -c_f(X_t - m_t)dt + Z_t dW_t. \quad (3.19)$$

Next, we make the ansatz

$$Y_t = \phi_t(X_t - m_t), \quad (3.20)$$

for some deterministic function ϕ_t, satisfying the terminal condition $\phi_T = c_t$. Differentiating the ansatz, the backward equation should satisfy

$$dY_t = (\dot{\phi}_t - \phi_t^2)(X_t - m_t)dt + \phi_t \sigma dW_t, \quad (3.21)$$

FIGURE 1 | On the left, we compare one representative optimal trajectory of $(X_{t_i})_{t_0 \le t_i \le t_N}$. The plot on the right show the comparison of one representative optimal trajectory of $(\alpha_{t_i})_{t_0 \le t_i \le t_N}$ between approach 1 and approach 2.

FIGURE 2 | On the left, we compare the sample mean of optimal trajectories of $(X_{t_i})_{t_0 \le t_i \le t_N}$. The plot on the right show the comparison of sample mean of trajectories of optimal control $(\alpha_{t_i})_{t_0 \le t_i \le t_N}$ between approach 1 and approach 2.

where $\dot{\phi}_t$ denotes the time derivative of ϕ_t. Comparing with (3.19), and identifying the drift and volatility term, ϕ_t must satisfy the scalar Riccati equation,

$$
\begin{cases}
\dot{\phi}_t = \phi_t^2 - c_f, \\
\phi_T = c_t,
\end{cases}
\tag{3.22}
$$

and the process Z_t should satisfy

$$
Z_t = \phi_t \sigma,
\tag{3.23}
$$

which is deterministic. If we choose $x_0 = 0$, $c_f = c_t = 1$, $T = 10$, $\phi_t = 1$ solves the Riccati equation (3.22), so that $Z_t = 1$, $\forall t \in [0, T]$, and from (3.20), the optimal control satisfies

$$
\hat{\alpha}_t = -(X_t - m_t).
\tag{3.24}
$$

Numerically, we apply the two deep learning algorithms proposed in the previous section. The first algorithm

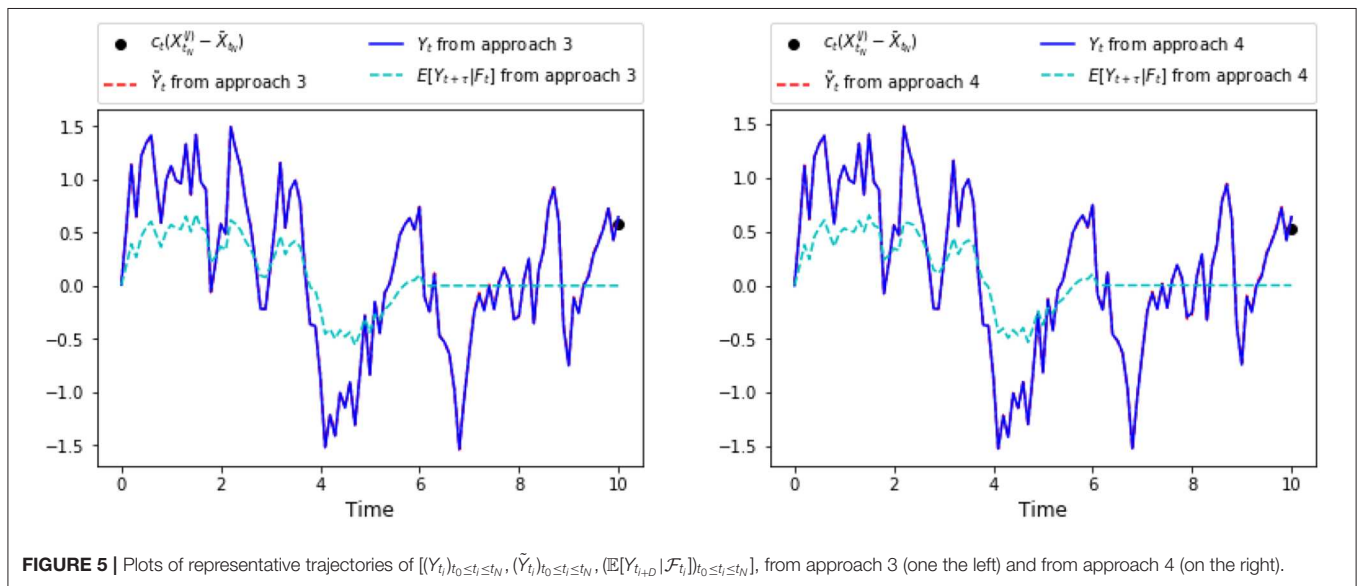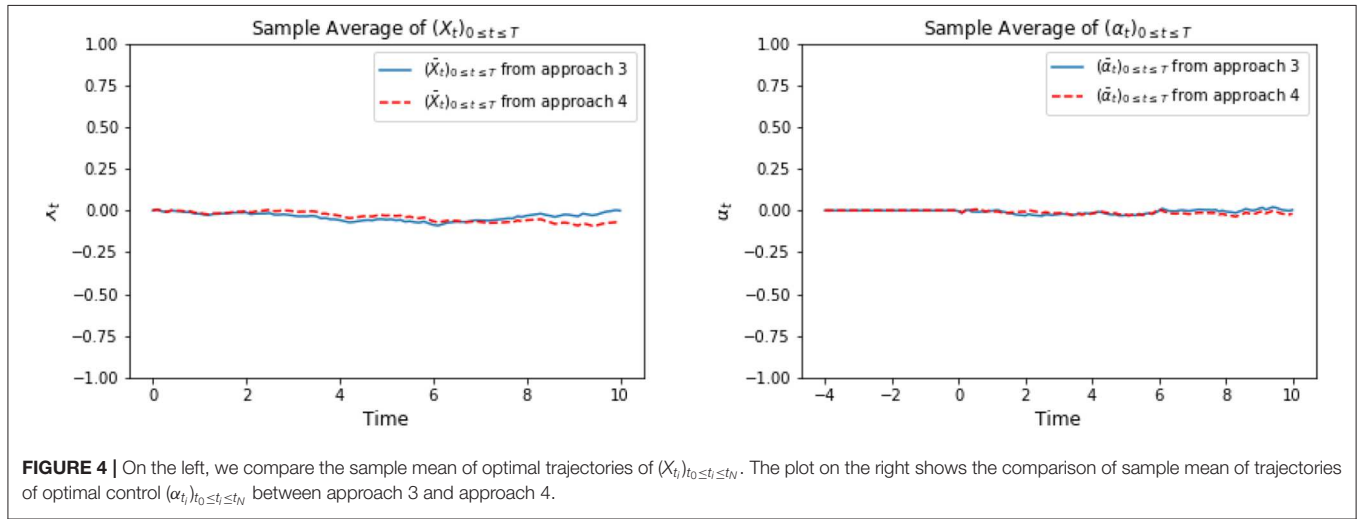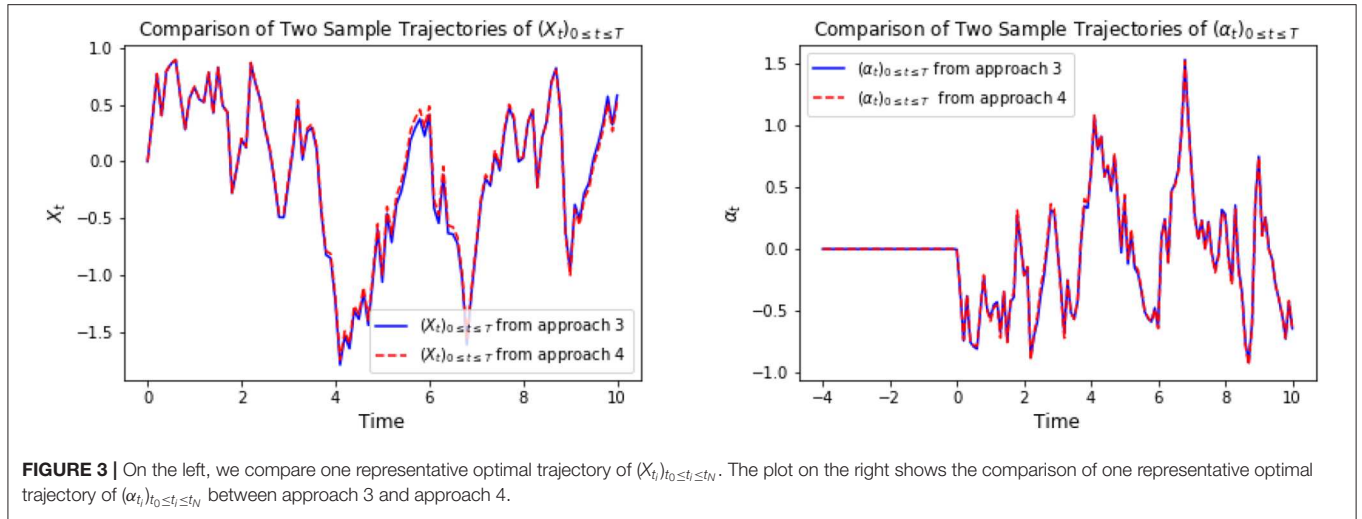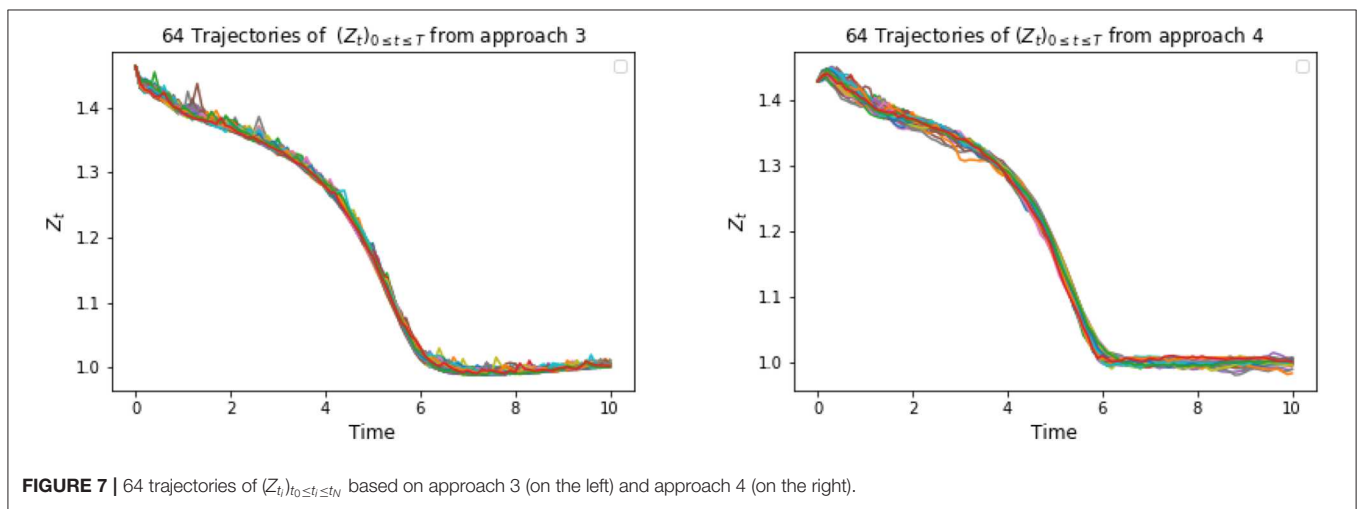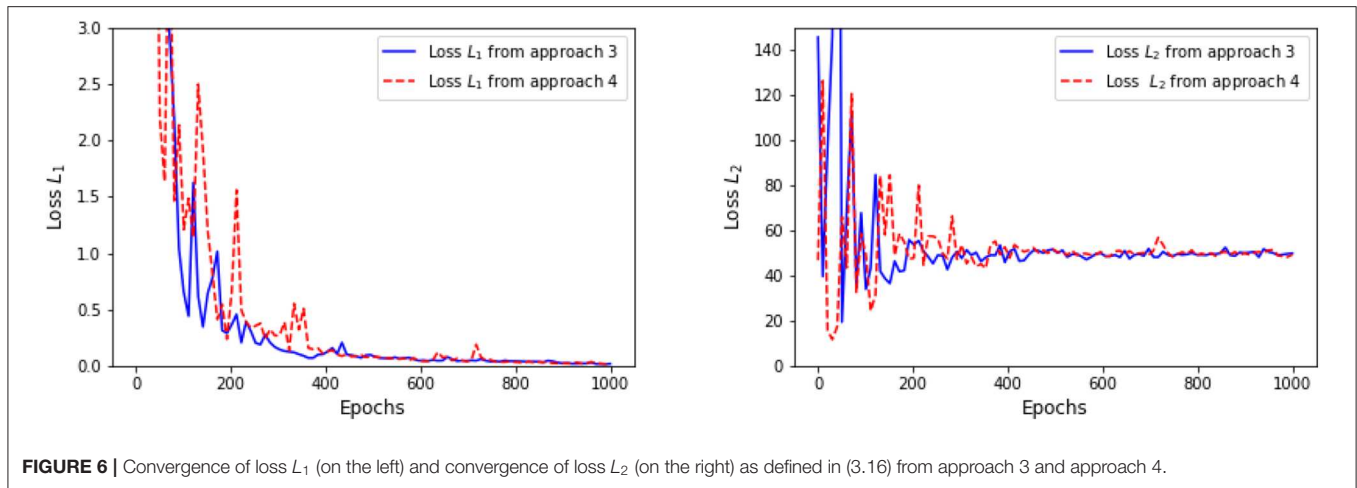directly approximates the control. According to the open loop formulation, we set

$$
(a_{t_i}, c_{t_i}) = \varphi(t_i, (\Delta W_s)_{t_0 \le s \le t_i} | \Phi)
$$
$$
\text{for } \varphi \in \mathbb{LSTM}_{2,h,t} \text{ and}
$$
$$
\Phi = (A_f, A_i, A_o, A_c, U_f, U_i, U_o, U_c, b_f, b_i, b_o, b_c),
$$
$$
\alpha(t_i, (W_s)_{t_0 \le s \le t_i}) \approx \psi(a_{t_i} | \Psi) \text{ for } \psi \in \mathbb{M}_{h,1}^{Id} \text{ and } \Psi = (A, b),
\tag{3.25}
$$

for some $h \in \mathbb{Z}^+$. We remark that the last dense layer is used to match the desired output dimension. The second algorithm numerically solves the forward backward system as in (3.18) and (3.19). From the ansatz (3.20) and the Markovian feature, we approximate $(Y_t, Z_t)_{0 \le t \le T}$ using two feedforward neural networks, i.e.,

$$
Y_{t_i} \approx \psi^1(t_i, X_{t_i} | \Psi^1) \text{ for } \psi^1 \in \mathbb{NN}_{2,1}^l, \ \Psi^1 = (A_0^1, b_0^1, \ldots, A_l^1, b_l^1);
$$
$$
Z_{t_i} \approx \psi^2(t_i, X_{t_i} | \Psi^2) \text{ for } \psi^2 \in \mathbb{NN}_{2,1}^l, \ \Psi^2 = (A_0^2, b_0^2, \ldots, A_l^2, b_l^2).
\tag{3.26}
$$

Figure 9 shows the representative optimal trajectory of $(X_{t_i} - \bar{X}_{t_i})_{t_0 \le t_i \le t_N}$ and $(\alpha_{t_i})_{t_0 \le t_i \le t_N}$ from both algorithms, which are

FIGURE 3 | On the left, we compare one representative optimal trajectory of $(X_{t_i})_{t_0 \le t_i \le t_N}$. The plot on the right shows the comparison of one representative optimal trajectory of $(\alpha_{t_i})_{t_0 \le t_i \le t_N}$ between approach 3 and approach 4.

FIGURE 4 | On the left, we compare the sample mean of optimal trajectories of $(X_{t_i})_{t_0 \le t_i \le t_N}$. The plot on the right shows the comparison of sample mean of trajectories of optimal control $(\alpha_{t_i})_{t_0 \le t_i \le t_N}$ between approach 3 and approach 4.

FIGURE 5 | Plots of representative trajectories of $[(Y_{t_i})_{t_0 \le t_i \le t_N}, (\tilde{Y}_{t_i})_{t_0 \le t_i \le t_N}, (\mathbb{E}[Y_{t_{i+D}} | \mathcal{F}_{t_i}])_{t_0 \le t_i \le t_N}]$, from approach 3 (one the left) and from approach 4 (on the right).

FIGURE 6 | Convergence of loss L_1 (on the left) and convergence of loss L_2 (on the right) as defined in (3.16) from approach 3 and approach 4.

FIGURE 7 | 64 trajectories of $(Z_{t_i})_{t_0 \le t_i \le t_N}$ based on approach 3 (on the left) and approach 4 (on the right).

exactly the same. The symmetry feature can be seen from the computation (3.24). On the left of **Figure 10** confirms the mean of the processes $(X_{t_i} - \bar{X}_{t_i})_{t_0 \le t_i \le t_N}$ and $(\alpha_{t_i})_{t_0 \le t_i \le t_N}$ are 0 from both algorithms. The right picture of **Figure 10** plots the data points of x against α, and we can observe that the optimal control α is linear in x as a result of (3.24), and the slope tends to be -1, since $\phi = 1$ solves the scalar Riccati equation (3.22). Finally, **Figure 11** plots representative optimal trajectories of the solution to the adjoint equations (Y_t, Z_t). On the left, we observe that the adjoint process $(Y_t)_{0 \le t \le T}$ matches the terminal condition, and on the right, $(Z_t)_{0 \le t \le T}$ appears to be a deterministic process of value 1, and this matches the result we compute previously.

4. STOCHASTIC MAXIMUM PRINCIPLE FOR OPTIMALITY

In this section, we derive the adjoint equation associated to our mean field stochastic control problem (2.3) and (2.4). The necessary and sufficient parts of stochastic maximum principle
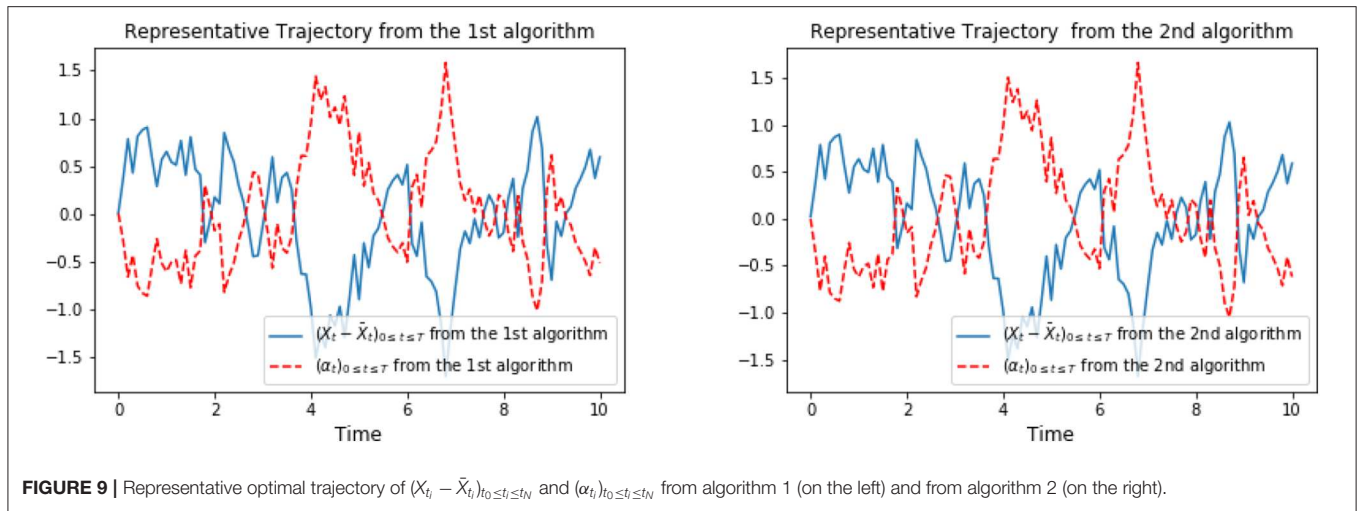

FIGURE 8 | Comparison convergence of objective values as in (3.10) among four approaches.
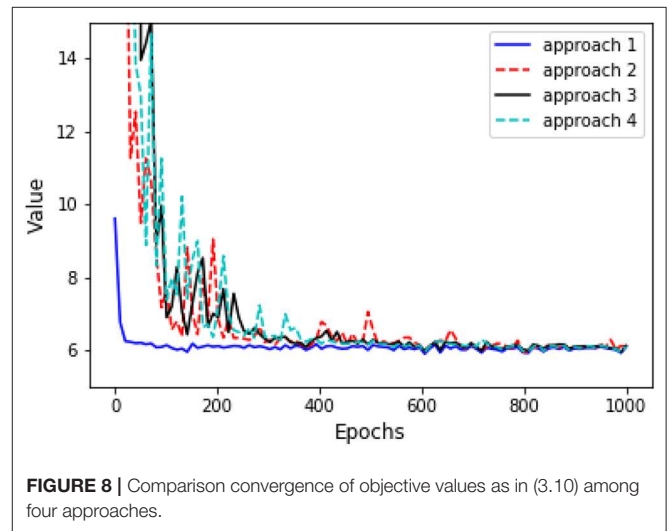
FIGURE 9 | Representative optimal trajectory of $(X_{t_i} - \bar{X}_{t_i})_{t_0 \leq t_i \leq t_N}$ and $(\alpha_{t_i})_{t_0 \leq t_i \leq t_N}$ from algorithm 1 (on the left) and from algorithm 2 (on the right).
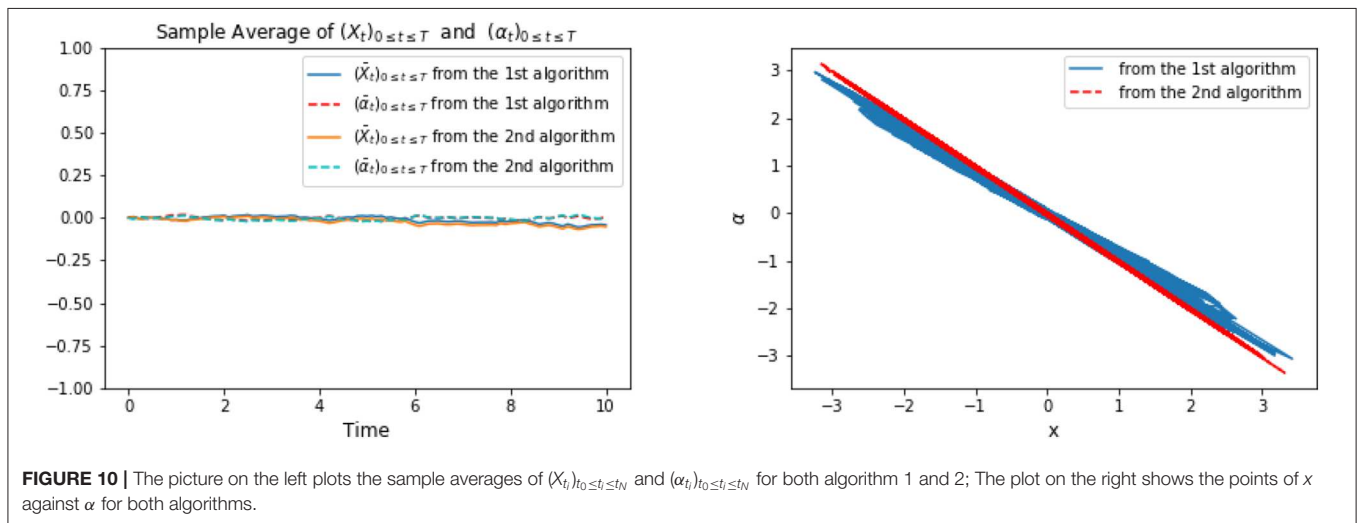

FIGURE 10 | The picture on the left plots the sample averages of $(X_{t_i})_{t_0 \leq t_i \leq t_N}$ and $(\alpha_{t_i})_{t_0 \leq t_i \leq t_N}$ for both algorithm 1 and 2; The plot on the right shows the points of x against α for both algorithms.
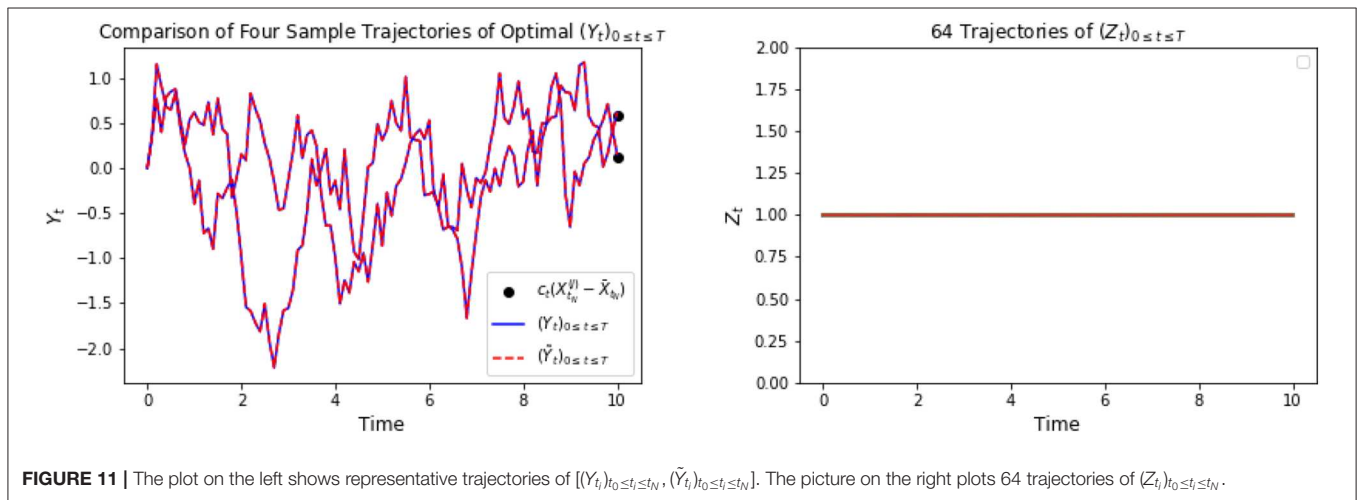

FIGURE 11 | The plot on the left shows representative trajectories of $[(Y_{t_i})_{t_0 \leq t_i \leq t_N}, (\tilde{Y}_{t_i})_{t_0 \leq t_i \leq t_N}]$. The picture on the right plots 64 trajectories of $(Z_{t_i})_{t_0 \leq t_i \leq t_N}$.

have been proved for optimality. We assume

(H4.1) b, σ are differentiable with respect to $(X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau})$; f is differentiable with respect to $(X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha)$; g is differentiable with respect to (X_T, μ_T). Their derivatives are bounded.

In order to simplify our notations, let $\theta_t = (X_t, \mu_t, \alpha_t)$. For $0 < \epsilon < 1$, we denote α^ϵ the admissible control defined by

$$\alpha_t^\epsilon := \alpha_t + \epsilon(\beta_t - \alpha_t) := \alpha_t + \epsilon\Delta\alpha_t,$$

for any $(\alpha)_{0 \le t \le T}$ and $(\beta)_{0 \le t \le T} \in \mathbb{A}$. $X_t^\epsilon := X_t^{\alpha^\epsilon}$ is the corresponding controlled process. We define

$$\nabla X_t := \lim_{\epsilon \to 0} \frac{X_t^\epsilon - X_t^\alpha}{\epsilon}$$

to be the variation process, which should follow the following dynamic for $t \in (0, T]$,

$$
\begin{aligned}
d\nabla X_t = &\Big[\partial_x b(t, \theta_t, \theta_{t-\tau})\nabla X_t + \partial_{x_\tau} b(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} \\
&+ \tilde{\mathbb{E}}[\partial_\mu b(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] + \tilde{\mathbb{E}}[\partial_{\mu_\tau} b(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] \\
&+ \partial_\alpha b(t, \theta_t, \theta_{t-\tau})\Delta\alpha_t + \partial_{\alpha_\tau} b(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dt \\
&+ \Big[\partial_x \sigma(t, \theta_t, \theta_{t-\tau})\nabla X_t + \partial_{x_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} \\
&+ \tilde{\mathbb{E}}[\partial_\mu \sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] + \tilde{\mathbb{E}}[\partial_{\mu_\tau}\sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] \\
&+ \partial_\alpha \sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_t + \partial_{\alpha_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dW_t \qquad (4.1)
\end{aligned}
$$

with initial condition $\nabla X_t = \Delta\alpha_t = 0$, $t \in [-\tau, 0)$. $(\tilde{X}_t, \nabla\tilde{X}_t)$ is a copy of $(X_t, \nabla X_t)$ defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, where we apply differential calculus on functions of measure, see Carmona and Delarue [3] for detail. $\partial_x b, \partial_{x_\tau} b, \partial_\mu b, \partial_{\mu_\tau} b, \partial_\alpha b, \partial_{\alpha_\tau} b$ are derivatives of b with respect to $(X_t, X_{t-\tau}, \mu_t, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau})$, respectively, and we use the same notation for $\partial.\sigma$.

In the meantime, the Gateaux derivative of functional $\alpha \to J(\alpha)$ is given by

$$
\begin{aligned}
&\lim_{\epsilon \to 0} \frac{J(\alpha^\epsilon) - J(\alpha)}{\epsilon} \\
&= \mathbb{E}\Big[\partial_x g(X_T, \mu_T)\nabla X_T + \tilde{\mathbb{E}}[\partial_\mu g(X_T, \mu_T)(\tilde{X}_T)\nabla\tilde{X}_T]\Big] \\
&+ \mathbb{E}\int_0^T \Big[\partial_x f(\theta_t, X_{t-\tau}, \mu_{t-\tau})\nabla X_t + \tilde{\mathbb{E}}[\partial_\mu f(\theta_t, X_{t-\tau}, \mu_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] \\
&+ \partial_{x_\tau} f(\theta_t, X_{t-\tau}, \mu_{t-\tau})\nabla X_{t-\tau} + \tilde{\mathbb{E}}[\partial_{\mu_\tau} f(\theta_t, X_{t-\tau}, \mu_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] \\
&+ \partial_\alpha f(\theta_t, X_{t-\tau}, \mu_{t-\tau})(\Delta\alpha_t)\Big]dt
\end{aligned}
$$
$$(4.2)$$

In order to determine the adjoint backward equation of $(Y_t, Z_t)_{0 \le t \le T}$ associated to (2.3), we assume it is of the following form:

$$
\begin{aligned}
dY_t &= -\varphi_t dt + Z_t dW_t, \quad t \in [0, T], \\
Y_T &= \partial_x g(X_T, \mu_T) + \tilde{\mathbb{E}}[\partial_\mu g(X_T, \mu_T)(\tilde{X}_T)], \qquad (4.3) \\
Y_t &= Z_t = 0, \quad t \in (T, T+\tau]
\end{aligned}
$$

Next, we apply integration by part to ∇X_t and Y_t. It yields

$$
\begin{aligned}
d(\nabla X_t Y_t) = &Y_t\Big[\partial_x b(t, \theta_t, \theta_{t-\tau})\nabla X_t + \partial_{x_\tau} b(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} \\
&+ \tilde{\mathbb{E}}[\partial_\mu b(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] + \tilde{\mathbb{E}}[\partial_{\mu_\tau} b(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] \\
&+ \partial_\alpha b(t, \theta_t, \theta_{t-\tau})\Delta\alpha_t + \partial_{\alpha_\tau} b(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dt \\
&+ Y_t\Big[\partial_x \sigma(t, \theta_t, \theta_{t-\tau}) + \partial_{x_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} \\
&+ \tilde{\mathbb{E}}[\partial_\mu \sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] + \tilde{\mathbb{E}}[\partial_{\mu_\tau}\sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] \\
&+ \partial_\alpha \sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_t + \partial_{\alpha_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dW_t - \varphi_t \nabla X_t dt \\
&+ \nabla X_t Z_t dW_t + Z_t\Big[\partial_x \sigma(t, \theta_t, \theta_{t-\tau})\nabla X_t \\
&+ \partial_{x_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} + \tilde{\mathbb{E}}[\partial_\mu \sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] \\
&+ \tilde{\mathbb{E}}[\partial_{\mu_\tau}\sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] + \partial_\alpha \sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_t \\
&+ \partial_{\alpha_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dt
\end{aligned}
$$

We integrate from 0 to T, and take expectation to get

$$
\begin{aligned}
&\mathbb{E}[\nabla X_T Y_T] \\
&= \mathbb{E}\int_0^T Y_t\Big[\partial_x b(t, \theta_t, \theta_{t-\tau})\nabla X_t + \partial_{x_\tau} b(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} \\
&+ \tilde{\mathbb{E}}[\partial_\mu b(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] + \tilde{\mathbb{E}}[\partial_{\mu_\tau} b(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] \\
&+ \partial_\alpha b(t, \theta_t, \theta_{t-\tau})\Delta\alpha_\tau + \partial_{\alpha_\tau} b(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dt \\
&- \mathbb{E}\int_0^T \varphi_t \nabla X_t dt + \mathbb{E}\int_0^T Z_t\Big[\partial_x \sigma(t, \theta_t, \theta_{t-\tau})\nabla X_t \\
&+ \partial_{x_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\nabla X_{t-\tau} + \tilde{\mathbb{E}}[\partial_\mu \sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_t)\nabla\tilde{X}_t] \\
&+ \tilde{\mathbb{E}}[\partial_{\mu_\tau}\sigma(t, \theta_t, \theta_{t-\tau})(\tilde{X}_{t-\tau})\nabla\tilde{X}_{t-\tau}] + \partial_\alpha \sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_t \\
&+ \partial_{\alpha_\tau}\sigma(t, \theta_t, \theta_{t-\tau})\Delta\alpha_{t-\tau}\Big]dt
\end{aligned}
$$
$$(4.4)$$

Using the fact that $Y_t = Z_t = 0$ for $t \in (T, T+\tau]$, we are able to make a change of time, and by Fubini's theorem, so that (4.4) becomes

$$
\begin{aligned}
\mathbb{E}[\nabla X_T Y_T] = &\mathbb{E}\int_0^T \Big(Y_t\partial_x b(t, \theta_t, \theta_{t-\tau}) + Y_{t+\tau}\partial_{x_\tau} b(t+\tau, \theta_{t+\tau}, \theta_t) \\
&+ \tilde{\mathbb{E}}[\partial_\mu b(t, \tilde{\theta}_t, \tilde{\theta}_{t-\tau})(X_t)] \\
&+ \tilde{\mathbb{E}}[\partial_{\mu_\tau} b(t+\tau, \tilde{\theta}_{t+\tau}, \tilde{\theta}_t)(X_t)]\Big)\nabla X_t dt \\
&+ \mathbb{E}\int_0^T \Big(\partial_\alpha b(t, \theta_t, \theta_{t-\tau}) + \partial_{\alpha_\tau} b(t+\tau, \theta_{t+\tau}, \theta_t)\Big)\Delta\alpha_t dt \\
&- \mathbb{E}\Big[\int_0^T \varphi_t \nabla X_t\Big]dt + \mathbb{E}\int_0^T \Big(Z_t\partial_x \sigma(t, \theta_t, \theta_{t-\tau}) \\
&+ Z_{t+\tau}\partial_{x_\tau}\sigma(t+\tau, \theta_{t+\tau}, \theta_t) + \tilde{\mathbb{E}}[\partial_\mu \sigma(t, \tilde{\theta}_t, \tilde{\theta}_{t-\tau})(X_t)] \\
&+ \tilde{\mathbb{E}}[\partial_{\mu_\tau}\sigma(t+\tau, \tilde{\theta}_{t+\tau}, \tilde{\theta}_t)(X_t)]\Big)\nabla X_t dt \\
&+ \mathbb{E}\int_0^T \Big(\partial_\alpha \sigma(t, \theta_t, \theta_{t-\tau}) + \partial_{\alpha_\tau}\sigma(t+\tau, \theta_{t+\tau}, \theta_t)\Big)\Delta\alpha_t dt
\end{aligned}
$$
$$(4.5)$$

Now we define the Hamiltonian H for $(t, x, \mu, x_\tau, \mu_\tau, y, z, \alpha, \alpha_\tau) \in [0, T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \times A \times A$ as

$$
\begin{aligned}
& H(t, x, \mu, x_\tau, \mu_\tau, \alpha, \alpha_\tau, y, z) \\
& = b(t, x, \mu, x_\tau, \mu_\tau, \alpha, \alpha_\tau) y + \sigma(t, x, \mu, x_\tau, \mu_\tau, \alpha, \alpha_\tau) z \\
& \quad + f(t, x, \mu, x_\tau, \mu_\tau, \alpha) \quad (4.6)
\end{aligned}
$$

Using the terminal condition of Y_T, and plugging (4.5) into (4.2), and setting the integrand containing ∇X_t to zero, we are able to obtain the adjoint equation is of the following form

$$
\begin{aligned}
dY_t = - \Big\{ & \partial_x H(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau}, Y_t, Z_t) \\
& + \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{X}_t, \mu_t, \tilde{X}_{t-\tau}, \mu_{t-\tau}, \tilde{\alpha}_t, \tilde{\alpha}_{t-\tau}, \tilde{Y}_t, \tilde{Z}_t)(X_t)] \\
& + \mathbb{E}[\partial_{x_\tau} H(t + \tau, X_{t+\tau}, \mu_{t+\tau}, X_t, \mu_t, \alpha_{t+\tau}, \alpha_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \\
& + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t + \tau, \tilde{X}_{t+\tau}, \mu_{t+\tau}, \tilde{X}_t, \mu_t, \tilde{\alpha}_t, \tilde{\alpha}_{t-\tau}, \tilde{Y}_{t+\tau}, \tilde{Z}_{t+\tau}) \\
& (X_{t+\tau})]|\mathcal{F}_t] \Big\} dt + Z_t dW_t \\
Y_T = & \partial_x g(X_T, \mu_T) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T, \mu_T)(X_T)].
\end{aligned}
$$

$$(4.7)$$

Theorem 4.1. Let $(\alpha_t)_{0 \leq t \leq T} \in \mathbb{A}$ be optimal, $(X_t)_{0 \leq t \leq T}$ be the associated controlled state, and $(Y_t, Z_t)_{0 \leq t \leq T}$ be the associated adjoint processes defined in (4.7). For any $\beta \in A$, and $t \in [0, T]$,

$$
\begin{aligned}
& \Big(\partial_\alpha H(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau}, Y_t, Z_t) \\
& + \mathbb{E}[\partial_{\alpha_\tau} H(t + \tau, X_{t+\tau}, \mu_{t+\tau}, X_t, \mu_t, \alpha_{t+\tau}, \alpha_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \Big) \\
& (\beta - \alpha_t) \geq 0 \ a.e.
\end{aligned}
$$

$$(4.8)$$

Proof: Given any $(\beta_t)_{0 \leq t \leq T} \in \mathbb{A}$, we perturbate α_t by $\epsilon(\beta_t - \alpha_t)$ and we define $\alpha_t^\epsilon := \alpha_t + \epsilon(\beta_t - \alpha_t)$ for $0 \leq \epsilon \leq 1$. Using the adjoint process (4.7), and apply integration by parts formula to $(\nabla X_t Y_t)$. Then plug the result into (4.2), and the Hamiltonian H is defined in (4.6). Also, since α is optimal, we have

$$
\begin{aligned}
0 \leq & \lim_{\epsilon \to 0} \frac{J(\alpha^\epsilon) - J(\alpha)}{\epsilon} \\
= & \mathbb{E} \int_0^T \Big([\partial_\alpha H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) + \\
& \mathbb{E}[\partial_{\alpha_\tau} H(t + \tau, \theta_{t+\tau}, \theta_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \Big)(\beta_t - \alpha_t) dt \quad (4.9)
\end{aligned}
$$

Now, let $C \in \mathcal{F}_t$ be an arbitrary progressively measurable set, and denote C' the complement of C. We choose β_t to be $\beta_t := \beta \mathbb{1}_C + \alpha_t \mathbb{1}_{C'}$ for any given $\beta \in A$. Then,

$$
\begin{aligned}
\mathbb{E} \int_0^T \Big(& [\partial_\alpha H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) \\
& + \mathbb{E}[\partial_{\alpha_\tau} H(t + \tau, \theta_{t+\tau}, \theta_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \Big)(\beta_t - \alpha_t) \mathbb{1}_C dt \geq 0,
\end{aligned}
$$

$$(4.10)$$

which implies,

$$
\begin{aligned}
& (\partial_\alpha H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) \quad (4.11) \\
& + \mathbb{E}[\partial_{\alpha_\tau} H(t + \tau, \theta_{t+\tau}, \theta_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t])(\beta - \alpha_t) \geq 0. \ a.e.
\end{aligned}
$$

$\square$

Remark 4.2. When we further assume that H is convex in $(\alpha_t, \alpha_{t-\tau})$, then for any $\beta, \beta_\tau \in A$ in Theorem 4.1, we have

$$
\begin{aligned}
& H(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau}, Y_t, Z_t) \\
& \leq H(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \beta, \beta_\tau, Y_t, Z_t), \ a.e.
\end{aligned}
$$

as a direct consequence of (4.8).

Theorem 4.3. Let $(\alpha_t)_{0 \leq t \leq T} \in \mathbb{A}$ be an admissible control. Let $(X_t)_{0 \leq t \leq T}$ be the controlled state, and $(Y_t, Z_t)_{0 \leq t \leq T}$ be the corresponding adjoint processes. We further assume that for each t, given Y_t and Z_t, the function $(x, \mu, x_\tau, \mu_\tau, \alpha, \alpha_\tau) \to H(t, x, \mu, x_\tau, \mu_\tau, \alpha, \alpha_\tau, Y_t, Z_t)$, and the function $(x, \mu) \to g(x, \mu)$ are convex. If

$$
\begin{aligned}
& H(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t, \alpha_{t-\tau}, Y_t, Z_t) \\
& = \inf_{\alpha' \in \mathbb{A}} H(t, X_t, \mu_t, X_{t-\tau}, \mu_{t-\tau}, \alpha_t', \alpha_{t-\tau}', Y_t, Z_t),
\end{aligned}
$$

$$(4.12)$$

for all t, then $(\alpha_t)_{0 \leq t \leq T}$ is an optimal control.

Proof: Let $(\alpha_t')_{0 \leq t \leq T} \in \mathbb{A}$ be a admissible control, and let $(X_t')_{0 \leq t \leq T} = (X_t^{\alpha'})_{0 \leq t \leq T}$ be the corresponding controlled state. From the definition of the objective function as in (2.4), we first use convexity of g, and the terminal condition of the adjoint process Y_t in (4.7), then use the fact that H is convex, and because of (4.12), we have the following

$$
\begin{aligned}
& J(\alpha) - J(\alpha') = \mathbb{E}[g(X_T, \mu_T) - g(X_T', \mu_T')] \\
& + \mathbb{E} \int_0^T [f(t, \theta_t, X_{t-\tau}, \alpha_{t-\tau}) - f(t, \theta_t', X_{t-\tau}', \alpha_{t-\tau}')] dt \\
& \leq \mathbb{E}[\partial_x g(X_T, \mu_T)(X_T - X_T') + \tilde{\mathbb{E}}[\partial_\mu g(X_T, \mu_T)(\tilde{X}_T)(\tilde{X}_T - \tilde{X}_T')]] \\
& + \mathbb{E} \int_0^T [f(t, \theta_t, X_{t-\tau}, \alpha_{t-\tau}) - f(t, \theta_t', X_{t-\tau}', \alpha_{t-\tau}')] dt \\
& = \mathbb{E}[(\partial_x g(X_T, \mu_T) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T, \mu_T)(X_T)])(X_T - X_T')] \\
& + \mathbb{E} \int_0^T [f(t, \theta_t, X_{t-\tau}, \alpha_{t-\tau}) - f(t, \theta_t', X_{t-\tau}', \alpha_{t-\tau}')] dt \\
& = \mathbb{E}[Y_T(X_T - X_T')] + \mathbb{E} \int_0^T [f(t, \theta_t, X_{t-\tau}, \alpha_{t-\tau}) \\
& - f(t, \theta_t', X_{t-\tau}', \alpha_{t-\tau}')] dt = \mathbb{E} \int_0^T \Big[\big(b(t, \theta_t, \theta_{t-\tau}) - b(t, \theta_t', \theta_{t-\tau}') \big) Y_t \\
& + \big(\sigma(t, \theta_t, \theta_{t-\tau}) - \sigma(t, \theta_t', \theta_{t-\tau}') \big) Z_t \Big] dt - \mathbb{E} \int_0^T \Big[\big(\partial_x H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) \\
& + \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{\theta}_t, \tilde{\theta}_{t-\tau}, \tilde{Y}_t, \tilde{Z}_t)(X_t)] \big) (X_t - X_t') \Big] dt \\
& - \mathbb{E} \int_0^T \Big[\big(\mathbb{E}[\partial_{x_\tau} H(t + \tau, \theta_{t+\tau}, \theta_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \\
& + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t + \tau, \tilde{\theta}_{t+\tau}, \tilde{\theta}_t, \tilde{Y}_{t+\tau}, \tilde{Z}_{t+\tau})(X_t)]|\mathcal{F}_t] \big) (X_t - X_t') \Big] dt \\
& + \mathbb{E} \int_0^T \Big[H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) - H(t, \theta_t', \theta_{t-\tau}', Y_t, Z_t) \Big] dt
\end{aligned}
$$

$$+ \ \mathbb{E}\int_0^T \Big[\big(b(t,\theta_t,\theta_{t-\tau}) - b(t,\theta_t,\theta_{t-\tau})\big) Y_t + \big(\sigma(t,\theta_t,\theta_{t-\tau})$$

$$-\sigma(t,\theta_t,\theta_{t-\tau})\big) Z_t \Big]dt \le -\mathbb{E}\int_0^T \Big[\partial_x H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)(X_t - X_t')$$

$$+ \ \tilde{\mathbb{E}}[\partial_\mu H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)(\tilde{X}_t)(\tilde{X}_t - \tilde{X}_t')]dt$$

$$- \ \mathbb{E}\int_0^T \Big[\partial_{x_\tau} H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)(X_{t-\tau} - X_{t-\tau}')$$

$$+ \ \tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)(\tilde{X}_{t-\tau})(\tilde{X}_{t-\tau} - \tilde{X}_{t-\tau}')] \Big]dt$$

$$+ \ \mathbb{E}\int_0^T \Big[H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t) - H(t,\theta_t',\theta_{t-\tau}',Y_t,Z_t) \Big]dt$$

$$\le \ \mathbb{E}\int_0^T \Big[\partial_\alpha H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)(\alpha_t - \alpha_t') + \partial_{\alpha_\tau} H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)$$

$$(\alpha_{t-\tau} - \alpha_{t-\tau}') \Big]dt \le \mathbb{E}\int_0^T \Big(\partial_\alpha H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t)$$

$$+ \ \mathbb{E}[\partial_{\alpha_\tau} H(t+\tau,\theta_{t+\tau},\theta_t,Y_{t+\tau},Z_{t+\tau})|\mathcal{F}_t] \Big)(\alpha_t - \alpha_t')dt \le 0. \quad (4.13)$$

$\square$

5. EXISTENCE AND UNIQUENESS RESULT

Given the necessary and sufficient conditions proven in section 4, we use the optimal control $(\hat{\alpha}_t)_{0 \le t \le T}$ defined by

$$\hat{\alpha}(t,X_t,\mu_t,X_{t-\tau},\mu_{t-\tau},Y_t,Z_t,\mathbb{E}[Y_{t+\tau}|\mathcal{F}_t],\mathbb{E}[Z_{t+\tau}|\mathcal{F}_t])$$
$$= \arg\min_{\alpha \in \mathbb{A}} H(t,X_t,\mu_t,X_{t-\tau},\mu_{t-\tau},\alpha_t,\alpha_{t-\tau},Y_t,Z_t), \quad (5.1)$$

to establish the solvability result of the McKean–Vlasov FABSDE (2.3) and (4.7) for $t \in [0,T]$:

$$dX_t = b(t,X_t,\mu_t,X_{t-\tau},\mu_{t-\tau},\hat{\alpha}_t,\hat{\alpha}_{t-\tau})dt$$
$$+ \ \sigma(t,X_t,\mu_t,X_{t-\tau},\mu_{t-\tau},\hat{\alpha}_t,\hat{\alpha}_{t-\tau})dW_t,$$

$$dY_t = -\Big\{ \partial_x H(t,X_t,\mu_t,X_{t-\tau},\mu_{t-\tau},\hat{\alpha}_t,\hat{\alpha}_{t-\tau},Y_t,Z_t)$$

$$+ \ \tilde{\mathbb{E}}[\partial_\mu H(t,\tilde{X}_t,\mu_t,\tilde{X}_{t-\tau},\mu_{t-\tau},\tilde{\hat{\alpha}}_t,\tilde{\hat{\alpha}}_{t-\tau},\tilde{Y}_t,\tilde{Z}_t)(X_t)]$$

$$+ \ \mathbb{E}[\partial_{x_\tau} H(t+\tau,X_{t+\tau},\mu_{t+\tau},X_t,\mu_t,\hat{\alpha}_{t+\tau},\hat{\alpha}_t,Y_{t+\tau},Z_{t+\tau})|\mathcal{F}_t]$$

$$+ \ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t+\tau,\tilde{X}_{t+\tau},\mu_{t+\tau},\tilde{X}_t,\mu_t,\tilde{\hat{\alpha}}_{t+\tau},\tilde{\hat{\alpha}}_t,\tilde{Y}_{t+\tau},\tilde{Z}_{t+\tau})$$

$$(X_t)]|\mathcal{F}_t]\Big\}dt + Z_t dW_t \quad (5.2)$$

with initial condition $X_0 = x_0$; $X_t = \hat{\alpha}_t = 0$ for $t \in [-\tau,0)$ and terminal condition $Y_T = \partial_x g(X_T,\mu_T) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T,\mu_T)(X_T)]$. In addition to assumption (H 4.1), we further assume

(H5.1) The drift and volatility functions b and σ are linear in $x,\mu,x_\tau,\mu_\tau,\alpha,\alpha_\tau$. For all $(t,x,\mu,x_\tau,\mu_\tau,\alpha,\alpha_\tau) \in [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathcal{P}_2(\mathbb{R}) \times A \times A$, we assume that

$$b(t,x,\mu,x_\tau,\mu_\tau,\alpha,\alpha_\tau) = b_0(t) + b_1(t)x + \bar{b}_1(t)m$$
$$+ b_2(t)x_\tau + \bar{b}_2(t)m_\tau + b_3(t)\alpha + b_4(t)\alpha_\tau,$$

$$\sigma(t,x,\mu,x_\tau,\mu_\tau,\alpha,\alpha_\tau) = \sigma_0(t) + \sigma_1(t)x + \bar{\sigma}_1(t)m$$
$$+ \sigma_2(t)x_\tau + \bar{\sigma}_2(t)m_\tau + \sigma_3(t)\alpha + \sigma_4(t)\alpha_\tau, \quad (5.3)$$

for some measurable deterministic functions $b_0, b_1, \bar{b}_1, b_2, \bar{b}_2, b_3, b_4, \sigma_0, \sigma_1, \bar{\sigma}_1, \sigma_2, \bar{\sigma}_2, \sigma_3, \sigma_4$ with values in $\mathbb{R}$ bounded by R, and we have used the notation $m = \int x d\mu(x)$ and $m_\tau = \int x d\mu_\tau(x)$ for the mean of measures μ and μ_τ, respectively.

(H5.2) The derivatives of f and g with respect to $(x,x_\tau,\mu,\mu_\tau,\alpha)$ and (x,μ) are Lipschitz continuous with Lipschitz constant L.

(H5.3) The function f is strongly L-convex, which means that for any $t \in [0,T]$, any $x,x',x_\tau,x_\tau' \in \mathbb{R}$, any $\alpha,\alpha' \in A$, any $\mu,\mu',\mu_\tau,\mu_\tau' \in \mathcal{P}_2(\mathbb{R})$, any random variables X and X' having μ and μ' as distribution, and any random variables X_τ and X_τ' having μ_τ and μ_τ' as distribution, then

$$f(t,x',\mu',x_\tau',\mu_\tau',\alpha') - f(t,x,\mu,x_\tau,\mu_\tau,\alpha)$$
$$- \ \partial_x f(t,x,\mu,x_\tau,\mu_\tau,\alpha)(x' - x)$$
$$- \ \partial_{x_\tau} f(t,x,\mu,x_\tau,\mu_\tau,\alpha)(x_\tau' - x_\tau)$$
$$- \ \mathbb{E}[\partial_\mu f(t,x,\mu,x_\tau,\mu_\tau,\alpha)(X) \cdot (X' - X)]$$
$$- \ \mathbb{E}[\partial_{\mu_\tau} f(t,x,\mu,x_\tau,\mu_\tau,\alpha)(X_\tau) \cdot (X_\tau' - X_\tau)]$$
$$- \ \partial_\alpha f(t,x,\mu,x_\tau,\mu_\tau,\alpha)(\alpha' - \alpha) \ge \kappa|\alpha' - \alpha|^2. (5.4)$$

The function g is also assumed to be L-convex in (x,μ).

Theorem 5.1. *Under assumptions (H5.1-H5.3), the McKean–Vlasov FABSDE (5.2) is uniquely solvable.*

The proof is based on continuation methods. Let $\lambda \in [0,1]$, consider the following class of McKean–Vlasov FABSDEs, denoted by MV-FABSDE(λ), for $t \in [0,T]$:

$$dX_t = (\lambda b(t,\theta_t,\theta_{t-\tau}) + I_t^b)dt + (\lambda\sigma(t,\theta_t,\theta_{t-\tau}) + I_t^\sigma)dW_t,$$

$$dY_t = -\Big\{\lambda\Big(\partial_x H(t,\theta_t,\theta_{t-\tau},Y_t,Z_t) + \tilde{\mathbb{E}}[\partial_\mu H(t,\tilde{\theta}_t,\tilde{\theta}_{t-\tau},\tilde{Y}_t,\tilde{Z}_t)(X_t)]$$

$$+ \ \mathbb{E}[\partial_{x_\tau} H(t+\tau,\theta_{t+\tau},\theta_t,Y_{t+\tau},Z_{t+\tau})|\mathcal{F}_t]$$

$$+ \ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t+\tau,\tilde{\theta}_{t+\tau},\tilde{\theta}_t,\tilde{Y}_{t+\tau},\tilde{Z}_{t+\tau})(X_t)]|\mathcal{F}_t]\Big)$$

$$+ \ I_t^f\Big\}dt + Z_t dW_t,$$

(5.5)

where we denote $\theta_t = (X_t,\mu_t,\alpha_t)$, with optimality condition

$$\alpha_t = \hat{\alpha}(t,X_t,\mu_t,X_{t-\tau},\mu_{t-\tau},Y_t,Z_t,\mathbb{E}[Y_{t+\tau}|\mathcal{F}_t],$$
$$\mathbb{E}[Z_{t+\tau}|\mathcal{F}_t]), t \in [0,T],$$

and with initial condition $X_0 = x_0$; $X_t = \alpha_t = 0$ for $t \in [-\tau,0)$ and terminal condition

$$Y_T = \lambda\Big\{\partial_x g(X_T,\mu_T) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T,\mu_T)(X_T)]\Big\} + I_T^g,$$

and $Y_t = 0$ for $t \in (T,T+\tau]$, where $(I_t^b,I_t^\sigma,I_t^f)_{0 \le t \le T}$ are some square-integrable progressively measurable processes with values in $\mathbb{R}$, and $I_T^g \in L^2(\Omega,\mathcal{F}_T,\mathbb{P})$ is a square integrable $\mathcal{F}_T$-measurable random variable with value in $\mathbb{R}$.

Observe that when $\lambda = 0$, system (5.5) becomes decoupled standard SDE and BSDE, which has an unique solution. When

setting $\lambda = 1$, $I_t^b = I_t^\sigma = I_t^f = 0$ for $0 \le t \le T$, and $I_T^g = 0$, we are able to recover the system of (5.2).

Lemma 5.2. *Given* $\lambda_0 \in [0,1)$, *for any square-integrable progressively measurable processes* $(I_t^b, I_t^\sigma, I_t^f)_{0 \le t \le T}$, *and* $I_T^g \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, *such that system FABSDE(λ_0) admits a unique solution, then there exists* $\delta_0 \in (0,1)$, *which is independent on* λ_0, *such that the system MV-FABSDE(λ) admits a unique solution for any* $\lambda \in [\lambda_0, \lambda_0 + \delta_0]$.

Proof: Assuming that $(\check{X}, \check{Y}, \check{Z}, \check{\alpha})$ are given as an input, for any $\lambda \in [\lambda_0, \lambda_0 + \delta_0]$, where $\delta_0 > 0$ to be determined, denoting $\delta := \lambda - \lambda_0 \le \delta_0$, we take

$$
\begin{aligned}
I_t^b &\leftarrow \delta[b(t, \check{\theta}_t, \check{\theta}_{t-\tau})] + I_t^b, \\
I_t^\sigma &\leftarrow \delta[\sigma(t, \check{\theta}_t, \check{\theta}_{t-\tau})] + I_t^\sigma, \\
I_t^f &\leftarrow \delta\Big[\partial_x H(t, \check{\theta}_t, \check{\theta}_{t-\tau}, Y_t, Z_t) + \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{\check{\theta}}_t, \tilde{\check{\theta}}_{t-\tau}, \tilde{\check{Y}}_t, \tilde{\check{Z}}_t)(X_t)] \\
&\quad + \mathbb{E}[\partial_{x_\tau} H(t + \tau, \check{\theta}_{t+\tau}, \check{\theta}_t, \check{Y}_{t+\tau}, \check{Z}_{t+\tau})|\mathcal{F}_t] \\
&\quad + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t + \tau, \tilde{\check{\theta}}_{t+\tau}, \tilde{\check{\theta}}_t, \tilde{\check{Y}}_{t+\tau}, \tilde{\check{Z}}_{t+\tau})(\check{X}_t)]|\mathcal{F}_t]\Big] + I_t^f, \\
I_T^g &\leftarrow \delta\Big[\partial_x g(\check{X}_T, \mu_T) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{\check{X}}_T, \mu_T)(\check{X}_T)]\Big] + I_T^g.
\end{aligned}
\tag{5.6}
$$

According to the assumption, let (X, Y, Z) be the solutions of MV-FABSDE(λ_0) corresponding to inputs $(\check{X}, \check{Y}, \check{Z})$, i.e., for $t \in [0, T]$

$$
\begin{aligned}
dX_t &= (\lambda_0 b_t + \delta \check{b}_t + I_t^b)dt + (\lambda_0 \sigma_t + \delta \check{\sigma}_t + I_t^\sigma)dW_t, \\
dY_t &= -\Big\{\lambda_0(\partial_x H_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)] \\
&\quad + \mathbb{E}[\partial_{x_\tau} H_{t+\tau}|\mathcal{F}_t] + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{H}_{t+\tau}(X_t)]|\mathcal{F}_t]) \\
&\quad + \delta(\partial_x \check{H}_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)] + \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau}|\mathcal{F}_t] \\
&\quad + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)]|\mathcal{F}_t]) + I_t^f\Big\}dt + Z_t dW_t,
\end{aligned}
\tag{5.7}
$$

with initial condition, $X_0 = x_0$, $X_s = \alpha_s = 0$ for $s \in [-\tau, 0)$, and terminal condition

$$
Y_T = \lambda_0\Big(\partial_x g_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{g}_T(X_T)]\Big) + \delta\Big(\partial_x \check{g}_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]\Big) + I_T^g,
\tag{5.8}
$$

and $Y_t = Z_t = 0$ for $t \in (T, T+\tau]$, where we have used simplified notations,

$$
\begin{aligned}
b_t &:= b(t, \theta_t, \theta_{t-\tau}); \quad \check{b}_t := b(t, \check{\theta}_t, \check{\theta}_{t-\tau}); \\
\sigma_t &:= \sigma(t, \theta_t, \theta_{t-\tau}); \quad \check{\sigma}_t := \sigma(t, \check{\theta}_t, \check{\theta}_{t-\tau}); \\
\partial_x H_t &:= \partial_x H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t);
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)] &:= \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{\theta}_t, \tilde{\theta}_{t-\tau}, \tilde{Y}_t, \tilde{Z}_t)(X_t)] \\
\mathbb{E}[\partial_{x_\tau} H_{t+\tau}|\mathcal{F}_t] &:= \mathbb{E}[\partial_{x_\tau} H(t + \tau, \theta_{t+\tau}, \theta_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t]; \\
\mathbb{E}[\tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)]|\mathcal{F}_t] &:= \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t + \tau, \tilde{\theta}_{t+\tau}, \tilde{\theta}_t, \tilde{Y}_{t+\tau}, \tilde{Z}_{t+\tau}) \\
&\quad (X_t)]|\mathcal{F}_t]; \\
\partial_x g_T &:= \partial_x g(X_T, \mu_T); \quad \tilde{\mathbb{E}}[\partial_\mu \tilde{g}_T(X_T)]] \\
&:= \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T, \mu_T)(X_T)]
\end{aligned}
$$

similar notation for $\partial_x \check{H}_t$, $\tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)]$, $\mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau}|\mathcal{F}_t]$,

and $\mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)]|\mathcal{F}_t]$.
$\qquad$ (5.9)

We would like to show that the map $\Phi : (\check{X}, \check{Y}, \check{Z}, \check{\alpha}) \to \Phi(\check{X}, \check{Y}, \check{Z}, \check{\alpha}) = (X, Y, Z, \alpha)$ is a contraction. Consider $(\Delta X, \Delta Y, \Delta Z, \Delta \alpha) = (X - X', Y - Y', Z - Z', \alpha - \alpha')$, where $(X', Y', Z', \alpha') = \Phi(\check{X}', \check{Y}', \check{Z}', \check{\alpha}')$. In addition, for the following computation, we have used simplified notation:

$$
\begin{aligned}
\Delta b_t &:= b(t, \theta_t, \theta_{t-\tau}) - b(t, \theta_t', \theta_{t-\tau}'); \\
\Delta \check{b}_t &:= b(t, \check{\theta}_t, \check{\theta}_{t-\tau}) - b(t, \check{\theta}_t', \check{\theta}_{t-\tau}'); \\
\Delta \sigma_t &:= \sigma(t, \theta_t, \theta_{t-\tau}) - \sigma(t, \theta_t', \theta_{t-\tau}'); \\
\Delta \check{\sigma}_t &:= \sigma(t, \check{\theta}_t, \check{\theta}_{t-\tau}) - \sigma(t, \check{\theta}_t', \check{\theta}_{t-\tau}') \\
\partial_x g_T &:= \partial_x g(X_T, \mu_T) - \partial_x g(X_T', \mu_T); \\
\Delta \tilde{\mathbb{E}}[\partial_\mu \tilde{g}_T(X_T)]] &:= \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T, \mu_T)(X_T)] \\
&\quad - \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T', \mu_T)(X_T')] \\
\Delta \partial_x H_t &:= \partial_x H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) \\
&\quad - \partial_x H(t, \theta_t', \theta_{t-\tau}', Y_t, Z_t) \\
\Delta \tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)] &:= \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{\theta}_t, \tilde{\theta}_{t-\tau}, \tilde{Y}_t, \tilde{Z}_t)(X_t)] \\
&\quad - \tilde{\mathbb{E}}[\partial_\mu H(t, \tilde{\theta}_t', \tilde{\theta}_{t-\tau}', \tilde{Y}_t, \tilde{Z}_t)(X_t')] \\
\Delta \mathbb{E}[\partial_{x_\tau} H_{t+\tau}|\mathcal{F}_t] &:= \mathbb{E}[\partial_{x_\tau} H(t + \tau, \theta_{t+\tau}, \theta_t, Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \\
&\quad - \mathbb{E}[\partial_{x_\tau} H(t + \tau, \theta_{t+\tau}', \theta_t', Y_{t+\tau}, Z_{t+\tau})|\mathcal{F}_t] \\
\Delta \mathbb{E}[\tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)]|\mathcal{F}_t] &:= \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t + \tau, \tilde{\theta}_{t+\tau}, \tilde{\theta}_t, \tilde{Y}_{t+\tau}, \\
&\quad \tilde{Z}_{t+\tau})(X_t)]|\mathcal{F}_t] - \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} H(t + \tau, \tilde{\theta}_{t+\tau}', \tilde{\theta}_t', \\
&\quad \tilde{Y}_{t+\tau}, \tilde{Z}_{t+\tau})(X_t')]|\mathcal{F}_t]
\end{aligned}
$$

similar notation for $\Delta \partial_x \check{H}_t$, $\Delta \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)]$, $\Delta \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau}|\mathcal{F}_t]$,

and $\Delta \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)]|\mathcal{F}_t]$.
$\qquad$ (5.10)

Applying integration by parts to $\Delta X_t Y_t$, we have

$$
\begin{aligned}
&d(\Delta X_t Y_t) \\
&= Y_t\Big\{[\lambda_0 \Delta b_t + \delta \Delta \check{b}_t]dt + [\lambda_0 \Delta \sigma_t + \delta \Delta \check{\sigma}_t]dW_t\Big\} \\
&\quad - \Delta X_t\Big\{\lambda_0(\partial_x H_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)] + \mathbb{E}[\partial_{x_\tau} H_{t+\tau}|\mathcal{F}_t] \\
&\quad + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{H}_{t+\tau}(X_t)]|\mathcal{F}_t]) + \delta(\partial_x \check{H}_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)] \\
&\quad + \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau}|\mathcal{F}_t] + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)]|\mathcal{F}_t])\Big\}dt \\
&\quad + \Delta X_t Z_t dW_t + (\lambda_0 \Delta \sigma_t + \delta \Delta \check{\sigma}_t)Z_t dt.
\end{aligned}
\tag{5.11}
$$

After integrating from 0 to T, and taking expectation on both sides, we obtain

$$\mathbb{E}[\Delta X_T Y_T]$$

$$= \lambda_0 \mathbb{E} \int_0^T \Bigg(\Delta b_t Y_t + \Delta \sigma_t Z_t - \Delta X_t (\partial_x H_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)]$$

$$+ \mathbb{E}[\partial_{x_\tau} H_{t+\tau} | \mathcal{F}_t] + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{H}_{t+\tau}(X_t)] | \mathcal{F}_t]) \Bigg) dt$$

$$+ \delta \mathbb{E} \int_0^T \Bigg(\Delta \check{b}_t Y_t + \Delta \check{\sigma} Z_t - \Delta X_t (\partial_x \check{H}_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)]$$

$$+ \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau} | \mathcal{F}_t] + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)] | \mathcal{F}_t]) \Bigg) dt$$

$$\tag{5.12}$$

In the meantime, from the terminal condition of Y_T given in (5.8), and since g is convex, we also have

$$\mathbb{E}[\Delta X_T Y_T]$$

$$= \mathbb{E}\Bigg[\Delta X_T \Big(\lambda_0 (\partial_x g_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{g}_T(X_T)]) + \delta(\partial_x \check{g}_T$$

$$+ \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]) + I_T^g \Big) \Bigg] \geq \lambda_0 \mathbb{E}[g(X_T, \mu_T) - g(X_T' - \mu_T')]$$

$$+ \delta \Delta X_T (\partial_x \check{g}_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]) + \Delta X_T I_T^g$$

$$\tag{5.13}$$

Following the proof of sufficient part of maximum principle and using (5.12), and (5.13), we find

$$\lambda_0 (J(\alpha) - J(\alpha'))$$

$$= \lambda_0 \mathbb{E}[g(X_T, \mu_T) - g(X_T', \mu_T')]$$

$$+ \lambda_0 \mathbb{E} \int_0^T [f(t, \theta_t, X_{t-\tau}, \mu_{t-\tau})$$

$$- f(t, \theta_t', X_{t-\tau}', \mu_{t-\tau}')] dt$$

$$\leq \mathbb{E}[\Delta X_T Y_T] - \delta \Delta X_T (\partial_x \check{g}_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]) - \Delta X_T I_T^g$$

$$+ \lambda_0 \mathbb{E} \int_0^T [f(t, \theta_t, X_{t-\tau}, \mu_{t-\tau}) - f(t, \theta_t', X_{t-\tau}', \mu_{t-\tau}')] dt$$

$$= \lambda_0 \mathbb{E} \int_0^T \Bigg[\Delta b_t Y_t + \Delta \sigma_t Z_t$$

$$- \Delta X_t (\partial_x H_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)] + \mathbb{E}[\partial_{x_\tau} H_{t+\tau} | \mathcal{F}_t]$$

$$+ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{H}_{t+\tau}(X_t)] | \mathcal{F}_t]) \Bigg] dt$$

$$+ \delta \mathbb{E} \int_0^T \Bigg[\Delta \check{b}_t Y_t + \Delta \check{\sigma}_t Z_t - \Delta X_t (\partial_x \check{H}_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)]$$

$$+ \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau} | \mathcal{F}_t] + \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{H}_{t+\tau}(X_t)] | \mathcal{F}_t]) \Bigg] dt$$

$$+ \lambda_0 \mathbb{E} \int_0^T [H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) - H(t, \theta_t', \theta_{t-\tau}', Y_t, Z_t)] dt$$

$$- \lambda_0 \mathbb{E} \int_0^T (\Delta b_t Y_t + \Delta_t \sigma Z_t) dt - \delta \Delta X_T (\partial_x \check{g}_T$$

$$+ \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]) - \Delta X_T I_T^g$$

$$= \lambda_0 \mathbb{E} \int_0^T \Bigg[H(t, \theta_t, \theta_{t-\tau}, Y_t, Z_t) - H(t, \theta_t', \theta_{t-\tau}', Y_t, Z_t)$$

$$- \Delta X_t (\partial_x H_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{H}_t(X_t)] + \mathbb{E}[\partial_{x_\tau} H_{t+\tau} | \mathcal{F}_t]$$

$$+ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{H}_{t+\tau}(X_t)] | \mathcal{F}_t]) \Bigg] dt + \delta \mathbb{E} \int_0^T \Bigg[\Delta \check{b}_t Y_t$$

$$+ \Delta \check{\sigma} Z_t - \Delta X_t (\partial_x \check{H}_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)] + \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau} | \mathcal{F}_t]$$

$$+ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)] | \mathcal{F}_t]) \Bigg] dt$$

$$- \delta \Delta X_T (\partial_x \check{g}_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]) - \Delta X_T I_T^g$$

$$\leq -\mathbb{E} \int_0^T \lambda_0 \kappa |\Delta \alpha_t|^2 dt + \delta \mathbb{E} \int_0^T \Bigg[\Delta \check{b}_t Y_t + \Delta \check{\sigma}_t Z_t$$

$$- \Delta X_t (\partial_x \check{H}_t + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(\check{X}_t)] + \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau} | \mathcal{F}_t]$$

$$+ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(\check{X}_t)] | \mathcal{F}_t]) \Bigg] dt - \delta \Delta X_T (\partial_x \check{g}_T$$

$$+ \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)]) - \Delta X_T I_T^g$$

$$\tag{5.14}$$

Reverse the role of α and α', we also have

$$\lambda_0 (J(\alpha') - J(\alpha))$$

$$\leq -\mathbb{E} \int_0^T \lambda_0 \kappa |\Delta \alpha_t'|^2 dt + \delta \mathbb{E} \int_0^T \Bigg[\Delta \check{b}_t' Y_t' + \Delta \check{\sigma}_t' Z_t'$$

$$- \Delta X_t' (\partial_x \check{H}_t' + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t'(\check{X}_t')] + \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau}' | \mathcal{F}_t]$$

$$+ \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}'(\check{X}_t')] | \mathcal{F}_t]) \Bigg] dt - \delta \Delta X_T' (\partial_x \check{g}_T$$

$$+ \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T')]) - \Delta X_T' I_T^g$$

$$\tag{5.15}$$

Summing (5.14) and (5.15), using the fact that b and σ have the linear form, using change of time and Lipschitz assumption, it yields

$$2 \lambda_0 \kappa \mathbb{E} \int_0^T |\Delta \alpha_t|^2 dt$$

$$\leq \delta \mathbb{E} \int_0^T \Bigg[\Delta \check{b}_t \Delta Y_t + \Delta \check{\sigma} \Delta Z_t - \Delta X_t (\Delta \partial_x \check{H}_t$$

$$+ \Delta \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(X_t)] + \Delta \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau} | \mathcal{F}_t]$$

$$+ \Delta \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(X_t)] | \mathcal{F}_t]) \Bigg] dt + \delta \Delta X_T (\partial_{x'} \check{g}_T'$$

$$- \partial_x \check{g}_T + \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T'(\check{X}_T)] - \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{g}}_T(\check{X}_T)])$$

$$\leq \frac{1}{2} \mathbb{E} \int_0^T \Bigg[\epsilon (|\Delta X_t|^2 + |\Delta Y_t|^2 + |\Delta Z_t|^2)$$

$$+ \frac{1}{\epsilon} \delta^2 \Big(|\Delta \check{b}_t|^2 + |\Delta \check{\sigma}|^2 + |\Delta \partial_x \check{H}_t + \Delta \tilde{\mathbb{E}}[\partial_\mu \tilde{\check{H}}_t(X_t)]$$

$$+ \Delta \mathbb{E}[\partial_{x_\tau} \check{H}_{t+\tau} | \mathcal{F}_t] + \Delta \mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau} \tilde{\check{H}}_{t+\tau}(X_t)] | \mathcal{F}_t]|^2 \Big) \Bigg] dt$$

$$+ \frac{1}{2}\left(\epsilon|\Delta X_T|^2 + \frac{1}{\epsilon}\delta^2\Big|\partial_{x'}\check{g}'_T - \partial_x\check{g}_T + \tilde{\mathbb{E}}[\partial_\mu\check{\tilde{g}}'_T(\check{X}_T)]\right.$$
$$\left. - \tilde{\mathbb{E}}[\partial_\mu\check{\tilde{g}}_T(\check{X}_T)]\Big|^2\right) \leq \frac{1}{2}\epsilon\mathbb{E}\left[\int_0^T \epsilon(|\Delta X_t|^2 + |\Delta Y_t|^2\right.$$
$$+ |\Delta Z_t|^2 + |\Delta\alpha_t|^2)dt + |\Delta X_T|^2\Big]$$
$$+ \frac{1}{2}\delta\frac{C}{\epsilon}\mathbb{E}\left[\int_0^T(|\Delta\check{X}_t|^2 + |\Delta\check{Y}_t|^2 + |\Delta\check{Z}_t|^2\right.$$
$$+ |\Delta\check{\alpha}_t|^2)]dt + |\Delta\check{X}_T|^2\Big], \tag{5.16}$$

Next, we apply Ito's formula to ΔX_t^2,

$$d\Delta X_t^2$$
$$= 2\Delta X_t dX_t + d\langle X, X\rangle_t$$
$$= 2\Delta X_t(\lambda_0\Delta b_t + \delta\Delta\check{b}_t)dt + 2\Delta X_t(\lambda_0\Delta\sigma_t \tag{5.17}$$
$$+ \delta\Delta\check{\sigma}_t)dW_t + \left(\lambda_0\Delta\sigma_t + \delta\Delta\check{\sigma}_t\right)^2 dt$$

Then integrate from 0 to T, and take expectation,

$$\mathbb{E}[|\Delta X_t|^2]$$
$$= 2\lambda_0\mathbb{E}\int_0^t |\Delta X_s\Delta b_s|ds + 2\delta\mathbb{E}\int_0^t |\Delta X_s\Delta\check{b}_s|ds$$
$$+ \mathbb{E}\int_0^t |\lambda_0\Delta\sigma_s + \delta\Delta\check{\sigma}_s|^2 ds$$
$$\leq \lambda_0\mathbb{E}\int_0^t (|\Delta X_s|^2 + |\Delta b_s|^2)ds + \mathbb{E}\int_0^t (|\Delta X_s|^2 + \delta^2|\Delta\check{b}_s|^2)ds$$
$$+ \mathbb{E}\int_0^t (2\lambda_0^2|\Delta\sigma_s|^2 + 2\delta^2|\Delta\check{\sigma}_s|^2)ds$$
$$\leq C\mathbb{E}\int_0^{t+\tau}(|\Delta X_s|^2 + |\Delta\alpha_s|^2)ds + \delta C\mathbb{E}\int_0^{t+\tau}(|\Delta\check{X}_s|^2 + |\Delta\check{\alpha}_s|^2)ds \tag{5.18}$$

From Gronwall's inequality, we can obtain

$$\sup_{0\leq t\leq T}\mathbb{E}[|X_t|^2] \leq C\mathbb{E}\int_0^T |\Delta\alpha_t|^2 dt + \delta C\mathbb{E}\int_0^T(|\Delta\check{X}_t|^2 + |\Delta\check{\alpha}_t|^2)dt \tag{5.19}$$

Similarly, applying Ito's formula to $|\Delta Y_t|^2$, and taking expectation, we have

$$\mathbb{E}\left[|\Delta Y_t|^2 + \int_t^T |\Delta Z_s|^2 ds\right]$$
$$= 2\lambda_0\mathbb{E}\int_t^T \Big|\Delta Y_t\Big(\Delta\partial_x H_t + \Delta\tilde{\mathbb{E}}[\partial_\mu\tilde{H}_t(X_t)]$$
$$+ \Delta\mathbb{E}[\partial_{x_\tau}H_{t+\tau}|\mathcal{F}_t] + \Delta\mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau}\tilde{H}_{t+\tau}(X_t)]|\mathcal{F}_t]\Big)\Big|$$
$$+ 2\delta\mathbb{E}\int_t^T \Big|\Delta Y_t\Big(\Delta\partial_x\check{H}_t + \Delta\tilde{\mathbb{E}}[\partial_\mu\check{\tilde{H}}_t(\check{X}_t)] + \Delta\mathbb{E}[\partial_{x_\tau}\check{H}_{t+\tau}|\mathcal{F}_t]$$
$$+ \Delta\mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau}\check{\tilde{H}}_{t+\tau}(\check{X}_t)]|\mathcal{F}_t]\Big)\Big| + \mathbb{E}|\Delta Y_T|^2$$

$$\leq\mathbb{E}\int_t^T\left(\frac{1}{\epsilon}|\Delta Y_t|^2 + \epsilon\lambda_0^2\Big|\Delta\partial_x H_t + \Delta\tilde{\mathbb{E}}[\partial_\mu\tilde{H}_t(X_t)]\right.$$
$$+ \Delta\mathbb{E}[\partial_{x_\tau}H_{t+\tau}|\mathcal{F}_t] + \Delta\mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau}\tilde{H}_{t+\tau}(X_t)]|\mathcal{F}_t]\Big|^2\right)dt$$
$$+ \mathbb{E}\int_t^T\left(|\Delta Y_t|^2 + \delta^2\Big|\Delta\partial_x\check{H}_t + \Delta\tilde{\mathbb{E}}[\partial_\mu\check{\tilde{H}}_t(\check{X}_t)]\right. \tag{5.20}$$
$$+ \Delta\mathbb{E}[\partial_{x_\tau}\check{H}_{t+\tau}|\mathcal{F}_t] + \Delta\mathbb{E}[\tilde{\mathbb{E}}[\partial_{\mu_\tau}\check{\tilde{H}}_{t+\tau}(\check{X}_t)]|\mathcal{F}_t]\Big|^2\right)dt$$
$$+ \mathbb{E}\Big|\lambda_0\Big(\Delta\partial_x g_T + \Delta\tilde{\mathbb{E}}[\partial_\mu\tilde{g}_T(X_T)]\Big)$$
$$+ \delta\Big(\Delta\partial_x\check{g}_T + \Delta\tilde{\mathbb{E}}[\partial_\mu\check{\tilde{g}}_T(\check{X}_T)]\Big)\Big|^2$$

Choose $\epsilon = 96\max\{R^2, L\}$, and from assumption (H5.1 - H5.2) and Gronwall's inequality, we obtain a bound for $\sup_{0\leq t\leq T}\mathbb{E}|\Delta Y_t|^2$; and then substitute the it back to the same inequality, we are able to obtain the bound for $\int_0^T E|Z_t|^2 dt$. By combining these two bounds, we deduce that

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|Y_t|^2 + \int_0^T |Z_t|^2 dt\right]$$
$$\leq C\mathbb{E}\left(\sup_{0\leq t\leq T}|\Delta X_t|^2 + \int_0^T |\Delta\alpha_t|^2 dt\right)$$
$$+ \delta C\mathbb{E}\left[\sup_{0\leq t\leq T}\left(|\Delta\check{X}_t|^2 + |\Delta\check{Y}_t|^2\right)\right. \tag{5.21}$$
$$+ \int_0^T\left(|\Delta\check{Z}_t|^2 + |\Delta\check{\alpha}_t|^2\right)dt\Big]$$

Finally, combining (5.19) and (5.21), and (5.16), we deduce

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|\Delta X_t|^2 + \sup_{0\leq t\leq T}|\Delta Y_t|^2 + \int_0^T\left(|\Delta Z_t^2| + |\Delta\alpha_t|^2\right)dt\right]$$
$$\leq\delta C\mathbb{E}\left[\sup_{0\leq t\leq T}|\Delta\check{X}_t|^2 + \sup_{0\leq t\leq T}|\Delta\check{Y}_t|^2 + \int_0^T\left(|\Delta\check{Z}_t|^2 + |\Delta\check{\alpha}_t|^2\right)dt\right] \tag{5.22}$$

Let $\delta_0 = \frac{1}{2C}$, it is clear that the mapping Φ is a contraction for all $\delta \in (0, \delta_0)$. It follows that there is a unique fixed point which is the solution of MV-FABSDE(λ) for $\lambda = \lambda_0 + \delta$, $\delta \in (0, \delta_0)$. $\square$

Proof of Theorem 5.1: For $\lambda = 0$, FABSDE(0) has a unique solution. Using Lemma 5.2, there exists a $\delta_0 > 0$ such that FBSDE(δ) has a unique solution for $\delta \in [0, \delta_0]$, assuming $(n-1)\delta_0 < 1 \leq n\delta_0$. Following by a induction argument, we repeat Lemma 5.2 for n times, which gives us the existence of the unique solution of FABSDE(1). $\square$

6. CONCLUSION

Overall, we presented a comprehensive study of a general class of mean-field control problems with delay effect. The state dynamics is described by a McKean–Vlasov stochastic delayed differential equation. We derive the adjoint process

associated to the dynamics, which is in the form of an anticipated backward stochastic differential equation of McKean–Vlasov type. We also prove a version of stochastic maximum principle, which gives necessary and sufficient conditions for the optimal control. Furthermore, we prove the existence and uniqueness of this class of forward anticipated backward stochastic differential equations under suitable conditions.

However, due to the lack of explicit solutions, numerical methods are needed. The non-linear nature of the problem due to the McKean–Vlasov aspect combined with non-Markovianity due to delay rule out classical numerical methods. Our study show that deep learning methods can deal with these obstacles. we proposed two algorithms based on machine learning to numerically solve the control problem. One is to directly approximate the control using a neural network, while the loss is given by the objective function in the control problem. The other algorithm is based on the system of forward and backward stochastic differential equations. We approximate the adjoint processes $(Y, Z.)$ and the conditional expectation of the adjoint process $\mathbb{E}[Y_{.+\tau}|\mathcal{F}_.]$ using neural networks. In this case, there are two loss functions that we need to minimize as shown in (3.16). The first loss is associated with the adjoint process $Y_.$,

and the other one is related to $\mathbb{E}[Y_{.+\tau}|\mathcal{F}_.]$. After minimizing the losses, the optimal control can be readily obtained from the adjoint processes. In addition, **Figure 8** illustrates that the optimal value to the control problem, that is computed from different methods, converges. Moreover, our method also works when the control problem has no delay. As a sanity check, we provide an example of control problem without delay effect, and we show that the solution obtained from the algorithm matches the explicit solution.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

1. Carmona R, Fouque JP, Sun LH. Mean field games and systemic risk. *Commun Math Sci.* (2015) **13**:911–33. doi: 10.4310/CMS.2015.v13.n4.a4
2. Carmona R, Fouque JP, Mousavi SM, Sun LH. Systemic risk and stochastic games with delay. *J Optim Appl.* (2018) **179**:366–99. doi: 10.1007/s10957-018-1267-8
3. Carmona R, Delarue F. *Probabilistic Theory of Mean Field Games with Applications I & II.* Springer International Publishing (2018).
4. Bensoussan A, Frehse J, Yam SCP. *Mean Field Games and Mean Field Type Control Theory.* New York, NY: Springer-Verlag (2013).
5. Peng S, Yang Z. Anticipated backward stochastic differential equations. *Ann Probab.* (2009) **37**:877–902. doi: 10.1214/08-AOP423
6. Zhang J. *Backward Stochastic Differential Equations.* New York, NY: Springer-Verlag (2017).
7. Chen L, Wu Z. Maximum principle for the stochastic optimal control problem with delay and application. *Automatica.* (2010) **46**:1074–80. doi: 10.1016/j.automatica.2010.03.005
8. Peng S, Wu Z. Fully coupled forward-backward stochastic differential equations and applications to optimal control. *SIAM J Control Optim.* (1999) **37**:825–43. doi: 10.1137/S0363012996313549
9. Bensoussan A, Yam SCP, Zhang Z. Well-posedness of mean-field type forward-backward stochastic differential equations. *Stochastic Process Appl.* (2015) **125**:3327–54. doi: 10.1016/j.spa.2015.04.006
10. Ma J, Yong J. *Forward-Backward Stochastic Differential Equations and their Applications.* Berlin; Heidelberg: Springer-Verlag (2007).
11. Ma J, Protter P, Yong J. Solving forward-backward stochastic differential equations explicitly–a four step scheme. *Probab Theory Relat Fields.* (1994) **98**:339–59.
12. Weinan E, Han J, Jentzen A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward

stochastic differential equations. *Commun Math Stat.* (2017) **5**:349–80. doi: 10.1007/s40304-017-0117-6
13. Raissi M. Forward-backward stochastic neural networks: deep learning of high-dimensional partial differential equations. (2018). *arxiv[Preprint].arxiv:1804.07010.*
14. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) **9**:1735–80.
15. Sutton RS, McAllester DA, Singh SP, Mansour Y. Policy gradient methods for reinforcement learning with function approximation. In: *Advances in Neural Information Processing Systems (NIPS).* (1999). p. 1057–63. Available online at: https://papers.nips.cc/paper/1713-policy-gradient-methods-for-reinforcement-learning-with-function-approximation
16. Han J, Weinan E. *Deep Learning Approximation for Stochastic Control Problems, Deep Reinforcement Learning Workshop.* Conference on Neural Information Processing Systems (NIPS) (2016).
17. Carmona R, Delarue F, Lachapelle A. Control of McKean-Vlasov dynamics versus mean field games. *Math Fin Econ.* (2013) **7**:131–66. doi: 10.1007/s11579-012-0089-y
18. Fouque JP, Zhang Z. Mean field game with delay: a toy model. *Risks.* (2018) **6**:1–17. doi: 10.3390/risks6030090