entropy

MDPI

Article

Forward-Backward Sweep Method for the System of HJB-FP Equations in Memory-Limited Partially Observable Stochastic Control

Takehiro Tottori [1,*] and Tetsuya J. Kobayashi [1,2,3,4]

[1] Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8654, Japan
[2] Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
[3] Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8654, Japan
[4] Universal Biology Institute, The University of Tokyo, Tokyo 113-8654, Japan
* Correspondence: takehiro_tottori@sat.t.u-tokyo.ac.jp

Abstract: Memory-limited partially observable stochastic control (ML-POSC) is the stochastic optimal control problem under incomplete information and memory limitation. To obtain the optimal control function of ML-POSC, a system of the forward Fokker–Planck (FP) equation and the backward Hamilton–Jacobi–Bellman (HJB) equation needs to be solved. In this work, we first show that the system of HJB-FP equations can be interpreted via Pontryagin's minimum principle on the probability density function space. Based on this interpretation, we then propose the forward-backward sweep method (FBSM) for ML-POSC. FBSM is one of the most basic algorithms for Pontryagin's minimum principle, which alternately computes the forward FP equation and the backward HJB equation in ML-POSC. Although the convergence of FBSM is generally not guaranteed in deterministic control and mean-field stochastic control, it is guaranteed in ML-POSC because the coupling of the HJB-FP equations is limited to the optimal control function in ML-POSC.

Keywords: decision-making; optimal control; stochastic control; incomplete information; memory limitation; mean-field control

Citation: Tottori, T.; Kobayashi, T.J. Forward-Backward Sweep Method for the System of HJB-FP Equations in Memory-Limited Partially Observable Stochastic Control. *Entropy* **2023**, *25*, 208. https://doi.org/10.3390/e25020208

Academic Editors: Mohammad Reza Rahimi Tabar and Adrian-Mihail Stoica

Received: 5 November 2022
Revised: 9 January 2023
Accepted: 16 January 2023
Published: 21 January 2023

1. Introduction

In many practical applications of the stochastic optimal control theory, several constraints need to be considered. In the cases of small devices [1,2] and biological systems [3–8], for example, incomplete information and memory limitation become predominant because their sensors are extremely noisy and their memory resources are severely limited. To take into account one of these constraints, incomplete information, partially observable stochastic control (POSC) has been extensively studied in the stochastic optimal control theory [9–13]. However, because POSC cannot take into account the other constraint, memory limitation, it is not practical enough for designing memory-limited controllers for small devices and biological systems. To resolve this problem, memory-limited POSC (ML-POSC) has recently been proposed [14]. Because ML-POSC formulates noisy observation and limited memory explicitly, ML-POSC can take into account both incomplete information and memory limitation in the stochastic optimal control problem.

However, ML-POSC cannot be solved in a similar way as completely observable stochastic control (COSC), which is the most basic stochastic optimal control problem [15–18]. In COSC, the optimal control function depends only on the Hamilton–Jacobi–Bellman (HJB) equation, which is a time-backward partial differential equation given a terminal condition (Figure 1a) [15–18]. Therefore, the optimal control function of COSC can be obtained by solving the HJB equation backward in time from the terminal condition, which is called the value iteration method [19–21]. In contrast, the optimal control function of ML-POSC

depends not only on the HJB equation but also on the Fokker–Planck (FP) equation, which is a time-forward partial differential equation given an initial condition (Figure 1b) [14]. Because the HJB equation and the FP equation interact with each other through the optimal control function in ML-POSC, the optimal control function of ML-POSC cannot be obtained by the value iteration method.

To propose an algorithm to solve ML-POSC, we first show that the system of HJB-FP equations can be interpreted via Pontryagin's minimum principle on the probability density function space. Pontryagin's minimum principle is one of the most representative approaches to the deterministic optimal control problem, which converts it into the two-point boundary value problem of the forward state equation and the backward adjoint equation [22–25]. We formally show that the system of HJB-FP equations is an extension of the system of adjoint and state equations from the deterministic optimal control problem to the stochastic optimal control problem.

The system of HJB-FP equations also appears in the mean-field stochastic control (MFSC) [26–28]. Although the relationship between the system of HJB-FP equations and Pontryagin's minimum principle has been briefly mentioned in MFSC [29–31], its details have not yet been investigated. In this work, we investigate it in more detail by deriving the system of HJB-FP equations in a similar way to Pontryagin's minimum principle. We note that our derivations are formal, not analytical, and more mathematically rigorous proofs remain future challenges. However, our results are consistent with many conventional results and also provide a useful perspective in proposing an algorithm.

We then propose the forward-backward sweep method (FBSM) for ML-POSC. FBSM is an algorithm to compute the forward FP equation and the backward HJB equation alternately, which can be interpreted as an extension of the value iteration method. FBSM has been proposed in Pontryagin's minimum principle of the deterministic optimal control problem, which computes the forward state equation and the backward adjoint equation alternately [32–34]. Because FBSM is easy to implement, it has been used in many applications [35,36]. However, the convergence of FBSM is not guaranteed in deterministic control except for special cases [37,38] because the coupling of adjoint and state equations is not limited to the optimal control function (Figure 1c). In contrast, we show that the convergence of FBSM is generally guaranteed in ML-POSC because the coupling of the HJB-FP equations is limited only to the optimal control function (Figure 1b).

FBSM is called the fixed-point iteration method in MFSC [39–42]. Although the fixed-point iteration method is the most basic algorithm to solve MFSC, its convergence is not guaranteed for the same reason as deterministic control (Figure 1d). Therefore, ML-POSC is a special and nice class of optimal control problems where FBSM or the fixed-point iteration method is guaranteed to converge.

This paper is organized as follows: In Section 2, we formulate ML-POSC. In Section 3, we derive the system of HJB-FP equations of ML-POSC from the viewpoint of Pontryagin's minimum principle. In Section 4, we propose FBSM for ML-POSC and prove its convergence. In Section 5, we apply FBSM to the linear-quadratic-Gaussian (LQG) problem. In Section 6, we verify the convergence of FBSM by numerical experiments. In Section 7, we discuss our work. In Appendix A, we briefly review Pontryagin's minimum principle of deterministic control. In Appendix B, we derive the system of HJB-FP equations of MFSC from the viewpoint of Pontryagin's minimum principle. In Appendix C, we show the detailed derivations of our results.

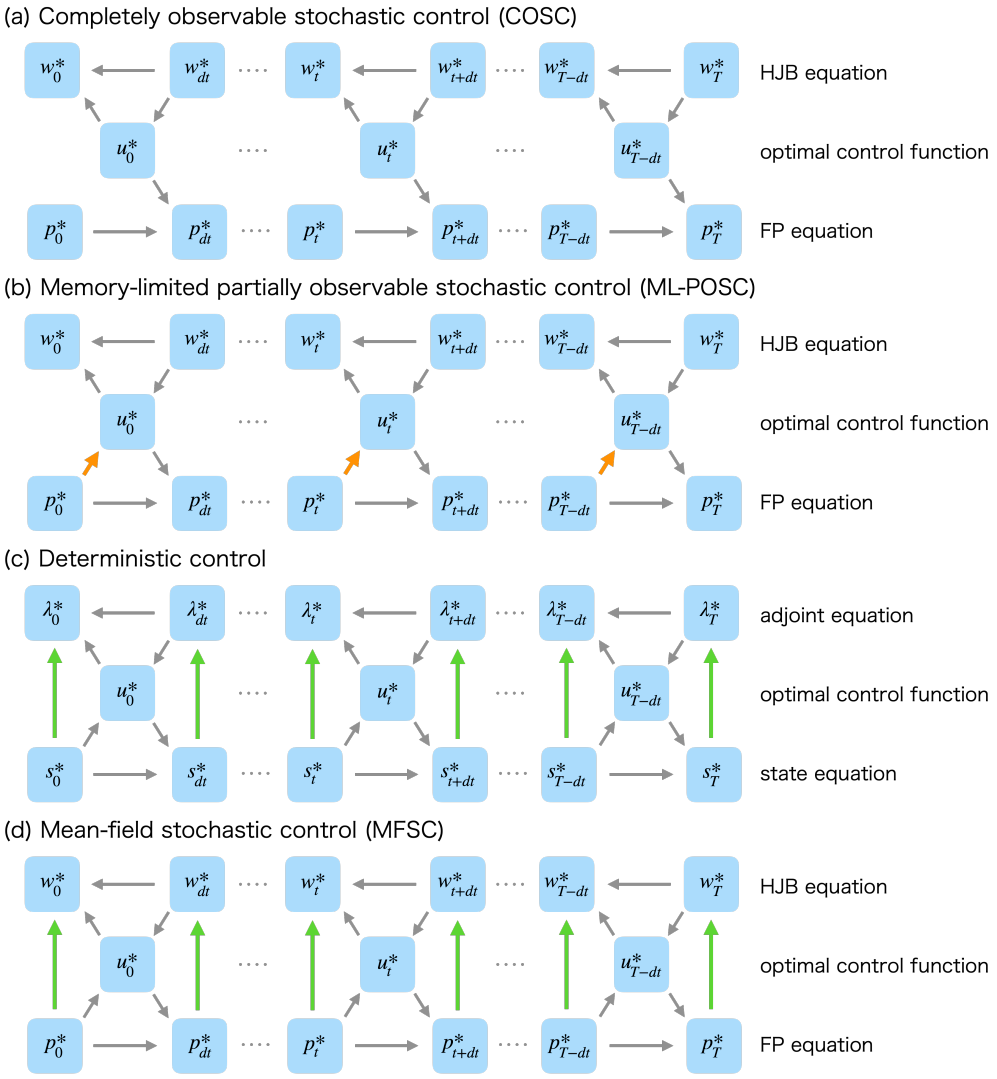

Figure 1. Schematic diagram of the relationship between the backward dynamics, the optimal control function, and the forward dynamics in (**a**) COSC, (**b**) ML-POSC, (**c**) deterministic control, and (**d**) MFSC. w^*, p^*, λ^*, and s^* are the solutions of the HJB equation, the FP equation, the adjoint equation, and the state equation, respectively. u^* is the optimal control function. The arrows indicate the dependence of variables. The variable at the head of an arrow depends on the variable at the tail of the arrow. (**a**) In COSC, because the optimal control function u^* depends only on the HJB equation w^*, it can be obtained by solving the HJB equation w^* backward in time from the terminal condition, which is called the value iteration method. (**b**) In ML-POSC, because the optimal control function u^* depends on the FP equation p^* as well as the HJB equation w^* (orange), it cannot be obtained by the value iteration method. In this paper, we propose FBSM for ML-POSC, which computes the HJB equation w^* and the FP equation p^* alternately. Because the coupling of the HJB equation w^* and the FP equation p^* is limited only to the optimal control function u^*, the convergence of FBSM is guaranteed in ML-POSC. (**c**) In deterministic control, because the coupling of the adjoint equation λ^* and the state equation s^* is not limited to the optimal control function u^* (green), the convergence of FBSM is not guaranteed. (**d**) In MFSC, because the coupling of the HJB equation w^* and the FP equation p^* is not limited to the optimal control function u^* (green), the convergence of FBSM is not guaranteed.

2. Memory-Limited Partially Observable Stochastic Control

In this section, we briefly review the formulation of ML-POSC [14], which is the stochastic optimal control problem under incomplete information and memory limitation.

2.1. Problem Formulation

This subsection outlines the formulation of ML-POSC [14]. The state of the system $x_t \in \mathbb{R}^{d_x}$ at time $t \in [0, T]$ evolves by the following stochastic differential equation (SDE):

$$dx_t = b(t, x_t, u_t)dt + \sigma(t, x_t, u_t)d\omega_t, \tag{1}$$

where x_0 obeys $p_0(x_0)$, $u_t \in \mathbb{R}^{d_u}$ is the control, and $\omega_t \in \mathbb{R}^{d_\omega}$ is the standard Wiener process. In COSC [15–18], because the controller can completely observe the state x_t, it determines the control u_t based on the state x_t as $u_t = u(t, x_t)$. By contrast, in POSC [9–13] and ML-POSC [14], the controller cannot directly observe the state x_t and instead obtains the observation $y_t \in \mathbb{R}^{d_y}$, which evolves by the following SDE:

$$dy_t = h(t, x_t)dt + \gamma(t)d\nu_t, \tag{2}$$

where y_0 obeys $p_0(y_0)$, and $\nu_t \in \mathbb{R}^{d_\nu}$ is the standard Wiener process. In POSC [9–13], because the controller can completely memorize the observation history $y_{0:t} := \{y_\tau | \tau \in [0, t]\}$, it determines the control u_t based on the observation history $y_{0:t}$ as $u_t = u(t, y_{0:t})$. In ML-POSC [14], by contrast, because the controller cannot completely memorize the observation history $y_{0:t}$, it compresses the observation history $y_{0:t}$ into the finite-dimensional memory $z_t \in \mathbb{R}^{d_z}$, which evolves by the following SDE:

$$dz_t = c(t, z_t, v_t)dt + \kappa(t, z_t, v_t)dy_t + \eta(t, z_t, v_t)d\xi_t, \tag{3}$$

where z_0 obeys $p_0(z_0)$, $v_t \in \mathbb{R}^{d_v}$ is the control, and $\xi_t \in \mathbb{R}^{d_\xi}$ is the standard Wiener process. The memory dimension d_z is determined by the available memory size of the controller. In addition, the memory noise ξ_t represents the intrinsic stochasticity of the memory to be used. Therefore, unlike the conventional POSC, ML-POSC can explicitly take into account the memory size and noise of the controller. Furthermore, because the memory dynamics (3) depends on the memory control v_t, it can be optimized through the memory control v_t, which is expected to realize the optimal compression of the observation history $y_{0:t}$ into the limited memory z_t. In ML-POSC [14], the controller determines the state control u_t and the memory control v_t based on the memory z_t as follows:

$$u_t = u(t, z_t), \ v_t = v(t, z_t). \tag{4}$$

The objective function of ML-POSC is given by the following expected cumulative cost function:

$$J[u, v] := \mathbb{E}_{p(x_{0:T}, y_{0:T}, z_{0:T}; u, v)}\left[\int_0^T f(t, x_t, u_t, v_t)dt + g(x_T)\right], \tag{5}$$

where f is the cost function, g is the terminal cost function, $p(x_{0:T}, y_{0:T}, z_{0:T}; u, v)$ is the probability of $x_{0:T}$, $y_{0:T}$, and $z_{0:T}$ given u and v as parameters, and $\mathbb{E}_p[\cdot]$ is the expectation with respect to the probability p. Because the cost function f depends on the memory control v_t, ML-POSC can explicitly take into account the memory control cost, which is also impossible with the conventional POSC.

ML-POSC is the problem of finding the optimal state control function u^* and the optimal memory control function v^* that minimize the expected cumulative cost function $J[u, v]$ as follows:

$$u^*, v^* := \arg\min_{u,v} J[u, v]. \tag{6}$$

ML-POSC first formulates the finite-dimensional and stochastic memory dynamics explicitly, then optimizes the memory control by considering the memory control cost. As a result, unlike the conventional POSC, ML-POSC is a practical framework for memory-limited controllers where the memory size, noise, and cost are imposed and non-negligible.

The previous work [14] has shown the validity and effectiveness of ML-POSC. In the LQG problem of conventional POSC, the observation history $y_{0:T}$ can be compressed into the Kalman filter without a loss of performance [10,18,43]. Because the Kalman filter is finite-dimensional, it can be interpreted as the finite-dimensional memory z_t and discussed in terms of ML-POSC. The previous work [14] has proven that the optimal memory dynamics of ML-POSC become the Kalman filter in this problem, which indicates that ML-POSC is a consistent framework with the conventional POSC. Furthermore, the previous work [14] has demonstrated the effectiveness of ML-POSC in the LQG problem with memory limitation and in the non-LQG problem by numerical experiments.

2.2. Problem Reformulation

Although the formulation of ML-POSC in the previous subsection is intuitive, it is inconvenient for further mathematical investigations. To address this problem, we reformulate ML-POSC in this subsection. The formulation in this subsection is simpler and more general than that in the previous subsection.

First, we define an extended state s_t as follows:

$$s_t := \begin{pmatrix} x_t \\ z_t \end{pmatrix} \in \mathbb{R}^{d_s}, \tag{7}$$

where $d_s = d_x + d_z$. The extended state s_t evolves by the following SDE:

$$ds_t = \tilde{b}(t, s_t, \tilde{u}_t)dt + \tilde{\sigma}(t, s_t, \tilde{u}_t)d\tilde{\omega}_t, \tag{8}$$

where s_0 obeys $p_0(s_0)$, $\tilde{u}_t \in \mathbb{R}^{d_{\tilde{u}}}$ is the control, and $\tilde{\omega}_t \in \mathbb{R}^{d_{\tilde{\omega}}}$ is the standard Wiener process. ML-POSC determines the control $\tilde{u}_t \in \mathbb{R}^{d_{\tilde{u}}}$ based on the memory z_t as follows:

$$\tilde{u}_t = \tilde{u}(t, z_t). \tag{9}$$

The extended state SDE (8) includes the previous SDEs (1)–(3) as a special case because they can be represented as follows:

$$\begin{aligned} ds_t = &\begin{pmatrix} b(t, x_t, u_t) \\ c(t, z_t, v_t) + \kappa(t, z_t, v_t)h(t, x_t) \end{pmatrix} dt \\ &+ \begin{pmatrix} \sigma(t, x_t, u_t) & O & O \\ O & \kappa(t, z_t, v_t)\gamma(t) & \eta(t, z_t, v_t) \end{pmatrix} \begin{pmatrix} d\omega_t \\ dv_t \\ d\xi_t \end{pmatrix}, \end{aligned} \tag{10}$$

where $p_0(s_0) = p_0(x_0)p_0(z_0)$.

The objective function of ML-POSC is given by the following expected cumulative cost function:

$$J[\tilde{u}] := \mathbb{E}_{p(s_{0:T};\tilde{u})}\left[\int_0^T \tilde{f}(t, s_t, \tilde{u}_t)dt + \tilde{g}(s_T)\right]. \tag{11}$$

where $\tilde{f}$ is the cost function and $\tilde{g}$ is the terminal cost function. It is obvious that this objective function (11) is more general than that in the previous subsection (5).

ML-POSC is the problem of finding the optimal control function $\tilde{u}^*$ that minimizes the expected cumulative cost function $J[\tilde{u}]$ as follows:

$$\tilde{u}^* := \arg\min_{\tilde{u}} J[\tilde{u}]. \tag{12}$$

In the following sections, we mainly consider the formulation of this subsection because it is simpler and more general than that in the previous subsection. Moreover, we omit $\tilde{\cdot}$ for simplicity of notation.

3. Pontryagin's Minimum Principle

If the control u_t is determined based on the extended state s_t as $u_t = u(t, s_t)$, ML-POSC is the same problem with COSC of the extended state, and its optimality conditions can be obtained in the conventional way [15–18]. In reality, however, because ML-POSC determines the control u_t based only on the memory z_t as $u_t = u(t, z_t)$, its optimality conditions cannot be obtained in a similar way as COSC. In the previous work [14], the optimality conditions of ML-POSC were obtained by employing a mathematical technique of MFSC [30,31].

In this section, we obtain the optimality conditions of ML-POSC by employing Pontryagin's minimum principle [22–25] on the probability density function space (Figure 2 (bottom right)). The conventional approach in ML-POSC [14] and MFSC [30,31] can be interpreted as a conversion from Bellman's dynamic programming principle (Figure 2 (top right)) to Pontryagin's minimum principle (Figure 2 (bottom right)) on the probability density function space.

In Appendix A, we briefly review Pontryagin's minimum principle in deterministic control (Figure 2 (left)). In this section, we obtain the optimality conditions of ML-POSC in a similar way as Appendix A (Figure 2 (right)). Furthermore, in Appendix B, we obtain the optimality conditions of MFSC in a similar way as Appendix A (Figure 2 (right)). MFSC is more general than ML-POSC except for the partial observability. In particular, the expected Hamiltonian is non-linear with respect to the probability density function in MFSC, while it is linear in ML-POSC.

Although our derivations are formal, not analytical, and more mathematically rigorous proofs remain future challenges, our results are consistent with the conventional results of COSC [15–18], ML-POSC [14], and MFSC [26–28,30,31], and also provide a useful perspective in proposing an algorithm.

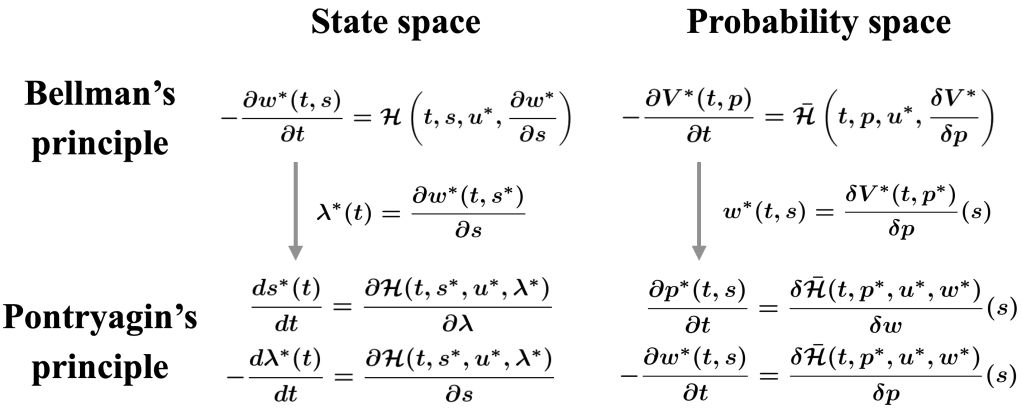

Figure 2. The relationship between Bellman's dynamic programming principle (**top**) and Pontryagin's minimum principle (**bottom**) on the state space (**left**) and on the probability density function space (**right**). The left-hand side corresponds to deterministic control, which is briefly reviewed in Appendix A. The right-hand side corresponds to ML-POSC and MFSC, which are shown in Section 3 and Appendix B, respectively. The conventional approach in ML-POSC [14] and MFSC [30,31] can be interpreted as the conversion from Bellman's dynamic programming principle (**top right**) to Pontryagin's minimum principle (**bottom right**) on the probability density function space.

3.1. Preliminary

In this subsection, we show a useful result in obtaining Pontryagin's minimum principle. Given arbitrary control functions u and u', $J[u] - J[u']$ can be calculated as follows:

$$J[u] - J[u'] = \int_0^T \left(\mathbb{E}_{p(t,s)}\left[\mathcal{H}(t,s,u,w')\right] - \mathbb{E}_{p(t,s)}\left[\mathcal{H}(t,s,u',w')\right] \right) dt, \tag{13}$$

where $\mathcal{H}$ is the Hamiltonian, which is defined as follows:

$$\mathcal{H}(t,s,u,w) := f(t,s,u) + \mathcal{L}_u w(t,s). \tag{14}$$

$\mathcal{L}_u$ is the backward diffusion operator, which is defined as follows:

$$\mathcal{L}_u w(t,s) := \sum_{i=1}^{d_s} b_i(t,s,u)\frac{\partial w(t,s)}{\partial s_i} + \frac{1}{2}\sum_{i,j=1}^{d_s} D_{ij}(t,s,u)\frac{\partial^2 w(t,s)}{\partial s_i \partial s_j}, \tag{15}$$

where $D(t,s,u) := \sigma(t,s,u)\sigma^\top(t,s,u)$. $w'(t,s)$ is the solution of the following Hamilton–Jacobi–Bellman (HJB) equation driven by u':

$$-\frac{\partial w'(t,s)}{\partial t} = \mathcal{H}(t,s,u',w'), \tag{16}$$

where $w'(T,s) = g(s)$. $p(t,s)$ is the solution of the following Fokker–Planck (FP) equation driven by u:

$$\frac{\partial p(t,s)}{\partial t} = \mathcal{L}_u^\dagger p(t,s), \tag{17}$$

where $p(0,s) = p_0(s)$. $\mathcal{L}_u^\dagger$ is the forward diffusion operator, which is defined as follows:

$$\mathcal{L}_u^\dagger p(t,s) := -\sum_{i=1}^{d_s}\frac{\partial(b_i(t,s,u)p(t,s))}{\partial s_i} + \frac{1}{2}\sum_{i,j=1}^{d_s}\frac{\partial^2(D_{ij}(t,s,u)p(t,s))}{\partial s_i \partial s_j}. \tag{18}$$

$\mathcal{L}_u^\dagger$ is the conjugate of $\mathcal{L}_u$ as follows:

$$\int w(t,s)\mathcal{L}_u^\dagger p(t,s)ds = \int p(t,s)\mathcal{L}_u w(t,s)ds. \tag{19}$$

We derive Equation (13) in Appendix C.1.

3.2. Necessary Condition

In this subsection, we show the necessary condition of the optimal control function of ML-POSC. It corresponds to Pontryagin's minimum principle on the probability density function space (Figure 2 (bottom right)). If u^* is the optimal control function of ML-POSC (12), then the following equation is satisfied:

$$u^*(t,z) = \arg\min_u \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u,w^*)],\ a.s.\ {}^\forall t \in [0,T],\ {}^\forall z \in \mathbb{R}^{d_z}, \tag{20}$$

where $w^*(t,s)$ is the solution of the following HJB equation driven by u^*:

$$-\frac{\partial w^*(t,s)}{\partial t} = \mathcal{H}(t,s,u^*,w^*), \tag{21}$$

where $w^*(T,s) = g(s)$. $p_t^*(x|z) := p^*(t,s)/\int p^*(t,s)dx$ is the conditional probability density function of state x given memory z, and $p^*(t,s)$ is the solution of the following FP equation driven by u^*:

$$\frac{\partial p^*(t,s)}{\partial t} = \mathcal{L}_{u^*}^\dagger p^*(t,s), \tag{22}$$

where $p^*(0,s) = p_0(s)$. We derive this result in Appendix C.2.

In deterministic control, Pontryagin's minimum principle can be expressed by the derivatives of the Hamiltonian (Figure 2 (bottom left)). Similarly, the system of HJB-FP Equations (21) and (22) can be expressed by the variations of the expected Hamiltonian

$$\bar{\mathcal{H}}(t,p,u,w) := \mathbb{E}_{p(s)}[\mathcal{H}(t,s,u,w)] \tag{23}$$

as follows:

$$\frac{\partial p^*(t,s)}{\partial t} = \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta w}(s), \tag{24}$$

$$-\frac{\partial w^*(t,s)}{\partial t} = \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s), \tag{25}$$

where $p^*(0,s) = p_0(s)$ and $w^*(T,s) = g(s)$ (Figure 2 (bottom right)). Therefore, the system of HJB-FP equations can be interpreted via Pontryagin's minimum principle on the probability density function space.

3.3. Sufficient Condition

Pontryagin's minimum principle (20) is only a necessary condition and generally not a sufficient condition. Pontryagin's minimum principle (20) becomes a necessary and sufficient condition if the expected Hamiltonian $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u. We obtain this result in Appendix C.3.

3.4. Relationship with Bellman's Dynamic Programming Principle

From Bellman's dynamic programming principle on the probability density function space (Figure 2 (top right)) [14], the optimal control function of ML-POSC is given by the following equation:

$$u^*(t,z,p) = \arg\min_u \mathbb{E}_{p(x|z)}\left[\mathcal{H}\left(t,s,u,\frac{\delta V^*(t,p)}{\delta p}(s)\right)\right], \tag{26}$$

where $V^*(t,p)$ is the value function on the probability density function space, which is the solution of the following Bellman equation:

$$-\frac{\partial V^*(t,p)}{\partial t} = \mathbb{E}_{p(s)}\left[\mathcal{H}\left(t,s,u^*,\frac{\delta V^*(t,p)}{\delta p}(s)\right)\right], \tag{27}$$

where $V^*(T,p) = \mathbb{E}_{p(s)}[g(s)]$. More specifically, the optimal control function of ML-POSC is given by $u^*(t,z) = u^*(t,z,p^*)$, where p^* is the solution of the FP Equation (22).

Because the Bellman Equation (27) is a functional differential equation, it cannot be solved even numerically. To resolve this problem, the previous work [14] converted the Bellman Equation (27) into the HJB Equation (21) by defining

$$w^*(t,s) := \frac{\delta V^*(t,p^*)}{\delta p}(s), \tag{28}$$

where p^* is the solution of FP Equation (22). This approach can be interpreted as the conversion from Bellman's dynamic programming principle (Figure 2 (top right)) to Pontryagin's minimum principle (Figure 2 (bottom right)) on the probability density function space.

3.5. Relationship with Completely Observable Stochastic Control

In the COSC of the extended state, the control u_t is determined based on the extended state s_t as $u_t = u(t,s_t)$. Therefore, in the COSC of the extended state, Pontryagin's minimum principle on the probability density function space is given by the following equation:

$$u^*(t,s) = \arg\min_u \mathcal{H}(t,s,u,w^*),\ a.s.\ {}^\forall t \in [0,T],\ {}^\forall s \in \mathbb{R}^{d_s}, \tag{29}$$

where $w^*(t,s)$ is the solution of the HJB Equation (21). Because this proof is almost identical to that of Section 3.2, it is omitted in this paper.

While the optimal control function of ML-POSC (20) depends on the FP equation and the HJB equation, the optimal control function of COSC (29) depends only on the HJB equation. From this nice property of COSC, Equation (29) is not only a necessary condition but also a sufficient condition without assuming the convexity of the expected Hamiltonian. We derive this result in Appendix C.4.

This result is consistent with the conventional result of COSC [15–18]. Unlike ML-POSC and MFSC, COSC can be solved by Bellman's dynamic programming principle on the state space. In COSC, Pontryagin's minimum principle on the probability density function space is equivalent to Bellman's dynamic programming principle on the state space. Because Bellman's dynamic programming principle on the state space is a necessary and sufficient condition, Pontryagin's minimum principle on the probability density function space may also become a necessary and sufficient condition.

4. Forward-Backward Sweep Method

In this section, we propose FBSM for ML-POSC and then prove its convergence by employing the interpretation of the system of HJB-FP equations by Pontryagin's minimum principle introduced in the previous section.

4.1. Forward-Backward Sweep Method

In this subsection, we propose FBSM for ML-POSC, which is summarized in Algorithm 1. FBSM is an algorithm to compute the forward FP equation and the backward HJB equation alternately. More specifically, in the initial step of FBSM, we initialize the control function $u^0_{0:T-dt}$ and obtain $p^0_{0:T}$ by computing the FP equation forward in time from the initial condition. In the backward step, we obtain $w^1_{0:T}$ by computing the HJB equation backward in time from the terminal condition and simultaneously update the control function from $u^0_{0:T-dt}$ to $u^1_{0:T-dt}$ by minimizing the conditional expected Hamiltonian. In the forward step, we obtain $p^2_{0:T}$ by computing the FP equation forward in time from the initial condition and simultaneously update the control function from $u^1_{0:T-dt}$ to $u^2_{0:T-dt}$ by minimizing the conditional expected Hamiltonian. By iterating the backward and forward steps, the objective function of ML-POSC $J[u^k_{0:T-dt}]$ monotonically decreases and finally converges to the local minimum at which the control function of ML-POSC $u^k_{0:T-dt}$ satisfies Pontryagin's minimum principle.

Pontryagin's minimum principle is only a necessary condition of the optimal control function, not a sufficient condition. Therefore, the control function obtained by FBSM is not necessarily the global optimum except in the case where the expected Hamiltonian is convex. Nevertheless, the control function obtained by FBSM is expected to be superior to most control functions because it is locally optimal.

FBSM has been used in deterministic control [32,34,35,38] and MFSC [39–42]. However, the convergence of FBSM for these problems is not guaranteed because the backward dynamics depend on the forward dynamics even without the optimal control function (Figure 1c,d). In contrast, the convergence of FBSM is guaranteed in ML-POSC because the backward HJB equation does not depend on the forward FP equation without the optimal control function (Figure 1b). More specifically, in FBSM for ML-POSC, the objective function $J[u^k_{0:T-dt}]$ monotonically decreases and finally converges to Pontryagin's minimum principle. In the following subsections, we prove this nice property of FBSM for ML-POSC.

Algorithm 1 Forward-Backward Sweep Method (FBSM)

//— Initial step —//
$k \leftarrow 0$
$p_0^k(s) \leftarrow p_0(s)$
for $t = 0$ to $T - dt$ **do**
  Initialize $u_t^k(z)$
  $p_{t+dt}^k(s) \leftarrow p_t^k(s) + \mathcal{L}_{u_t^k}^\dagger p_t^k(s)dt$
end for
while $J[u_{0:T-dt}^k]$ do not converge **do**
  if k is even **then**
    //— Backward step —//
    $w_T^{k+1}(s) \leftarrow g(s)$
    for $t = T - dt$ to 0 **do**
      $u_t^{k+1}(z) \leftarrow \arg\min_u \mathbb{E}_{p_t^k(x|z)}\left[\mathcal{H}(t, s, u, w_{t+dt}^{k+1})\right]$
      $w_t^{k+1}(s) \leftarrow w_{t+dt}^{k+1}(s) + \mathcal{H}(t, s, u_t^{k+1}, w_{t+dt}^{k+1})dt$
    end for
  else
    //— Forward step —//
    $p_0^{k+1}(s) \leftarrow p_0(s)$
    for $t = 0$ to $T - dt$ **do**
      $u_t^{k+1}(z) \leftarrow \arg\min_u \mathbb{E}_{p_t^{k+1}(x|z)}\left[\mathcal{H}(t, s, u, w_{t+dt}^k)\right]$
      $p_{t+dt}^{k+1}(s) \leftarrow p_t^{k+1}(s) + \mathcal{L}_{u_t^{k+1}}^\dagger p_t^{k+1}(s)dt$
    end for
  end if
  $k \leftarrow k + 1$
end while
return $u_{0:T-dt}^k$

4.2. Preliminary

In this subsection, we show an important result in proving the convergence of FBSM for ML-POSC. We suppose that $u_{0:t-dt,t+dt:T-dt} := \{u_0, ..., u_{t-dt}, u_{t+dt}, ..., u_{T-dt}\}$ is given and only u_t is optimized as follows:

$$u_t^* := \arg\min_{u_t} J[u_{0:T-dt}]. \tag{30}$$

In ML-POSC, u_t^* can be calculated as follows:

$$u_t^*(z) = \arg\min_{u_t} \mathbb{E}_{p_t(x|z)}[\mathcal{H}(t, s, u_t, w_{t+dt})], \ a.s. \ ^\forall z \in \mathbb{R}^{d_z}, \tag{31}$$

where $w_{t+dt}(s)$ is the solution of the following time-discretized HJB equation driven by $u_{t+dt:T-dt}$:

$$w_\tau(s) = w_{\tau+dt}(s) + \mathcal{H}(\tau, s, u_\tau, w_{\tau+dt})dt, \ \tau \in \{t + dt, ..., T - dt\}, \tag{32}$$

where $w_T(s) = g(s)$. $p_t(x|z) := p_t(s)/\int p_t(s)dx$ is the conditional probability density function of state x given memory z, and $p_t(s)$ is the solution of the following time-discretized FP equation driven by $u_{0:t-dt}$:

$$p_{\tau+dt}(s) = p_\tau(s) + \mathcal{L}_{u_\tau}^\dagger p_\tau(s)dt, \ \tau \in \{0, ..., t - dt\}, \tag{33}$$

where $p_0(s)$. Equation (31) is obtained by the similar way to Pontyragin's minimum principle in Appendix C.5 and also by the time discretization method in Appendix C.6.

Importantly, w_{t+dt} does not depend on u_t in ML-POSC (Figure 3a) while λ_{t+dt} and w_{t+dt} depend on u_t in deterministic control (Figure 3b) and MFSC (Figure 3c), respectively. Therefore, u_t^* can be obtained without modifying w_{t+dt} in ML-POSC, which is essentially different from deterministic control and MFSC. From this nice property, the convergence of FBSM is guaranteed in ML-POSC.

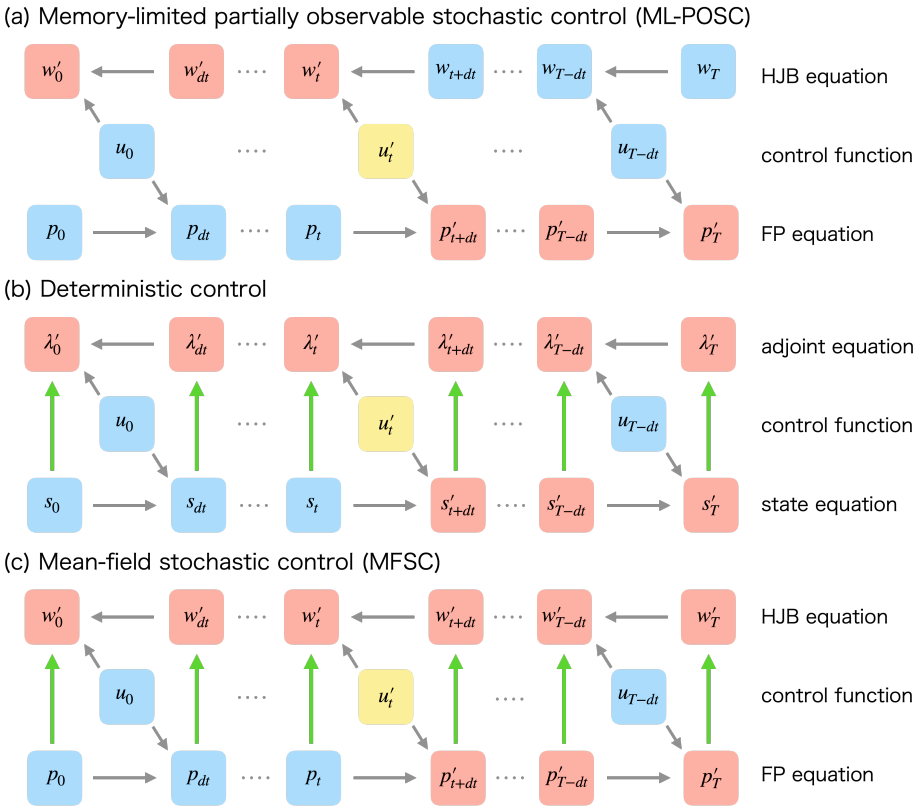

Figure 3. Schematic diagram of the effect of updating the control function to the forward and backward dynamics in (**a**) ML-POSC, (**b**) deterministic control, and (**c**) MFSC. $w_{0:T}$, $p_{0:T}$, $\lambda_{0:T}$, and $s_{0:T}$ are the solutions of the HJB equation, the FP equation, the adjoint equation, and the state equation, respectively. $u_{0:T-dt}$ is a given control function. The arrows indicate the dependence of variables. The variable at the head of an arrow depends on the variable at the tail of the arrow. (**a**) In ML-POSC, while the update from u_t to u'_t (yellow) changes $w_{0:t}$ and $p_{t+dt:T}$ to $w'_{0:t}$ and $p'_{t+dt:T}$, respectively (red), it does not change $p_{0:t}$ and $w_{t+dt:T}$ (blue). From this property, the convergence of FBSM is guaranteed in ML-POSC. (**b**) In deterministic control, the update from u_t to u'_t (yellow) changes $\lambda_{t+dt:T}$ to $\lambda'_{t+dt:T}$ as well (red) because the adjoint equation depends on the state equation (green). Because FBSM does not take into account the change of $\lambda_{t+dt:T}$, the convergence of FBSM is not guaranteed in deterministic control. (**c**) In MFSC, the update from u_t to u'_t (yellow) changes $w_{t+dt:T}$ to $w'_{t+dt:T}$ as well (red) because the HJB equation depends on the FP equation (green). Because FBSM does not take into account the change of $w_{t+dt:T}$, the convergence of FBSM is not guaranteed in MFSC.

4.3. Monotonicity

In FBSM for ML-POSC, the objective function is monotonically non-increasing with respect to the update of the control function at each time step. More specifically,

$$J[u^k_{0:t-dt}, u^{k+1}_{t:T-dt}] \leq J[u^k_{0:t}, u^{k+1}_{t+dt:T-dt}] \tag{34}$$

is satisfied in the backward step, and

$$J[u^{k+1}_{0:t-dt}, u^k_{t:T-dt}] \geq J[u^{k+1}_{0:t}, u^k_{t+dt:T-dt}] \tag{35}$$

is satisfied in the forward step. We prove this result in Appendix C.7. Furthermore, in FBSM for ML-POSC, the objective function is monotonically non-increasing with respect to the update of the control function at each iteration step as follows:

$$J[u^{k+1}_{0:T-dt}] \leq J[u^k_{0:T-dt}]. \tag{36}$$

Equation (36) is obviously satisfied from Equations (34) and (35).

4.4. Convergence to Pontryagin's Minimum Principle

We assume that $J[u_{0:T-dt}]$ has a lower bound. From Equation (36), FBSM for ML-POSC is guaranteed to converge to the local minimum. Furthermore, we assume that if the candidate of u_t^{k+1} includes u_t^k, then set u_t^{k+1} at u_t^k. Under these assumptions, FBSM for ML-POSC converges to Pontryagin's minimum principle (20). More specifically, if $J[u_{0:T-dt}^{k+1}] = J[u_{0:T-dt}^{k}]$ holds, $u_{0:T-dt}^{k+1}$ satisfies Pontryagin's minimum principle (20). We prove this result in Appendix C.8.

Therefore, unlike deterministic control and MFSC, in FBSM for ML-POSC, the objective function $J[u_{0:T-dt}^k]$ monotonically decreases and finally converges to the local minimum at which the control function $u_{0:T-dt}^k$ satisfies Pontryagin's minimum principle (20).

5. Linear-Quadratic-Gaussian Problem

In this section, we apply FBSM to the LQG problem of ML-POSC [14]. In the LQG problem of ML-POSC, the system of HJB-FP equations is reduced from partial differential equations to ordinary differential equations.

5.1. Problem Formulation

In the LQG problem of ML-POSC, the extended state SDE (8) is given as follows [14]:

$$ds_t = (A(t)s_t + B(t)u_t)dt + \sigma(t)d\omega_t, \tag{37}$$

where s_0 obeys the Gaussian distribution $p_0(s_0) := \mathcal{N}(s_0|\mu_0, \Lambda_0)$ where μ_0 is the mean vector and Λ_0 is the precision matrix. The objective function (11) is given as follows:

$$J[u] := \mathbb{E}_{p(s_{0:T};u)}\left[\int_0^T \left(s_t^\top Q(t)s_t + u_t^\top R(t)u_t\right)dt + s_T^\top P s_T\right], \tag{38}$$

where $Q(t) \succeq O$, $R(t) \succ O$, and $P \succeq O$. The LQG problem of ML-POSC is the problem of finding the optimal control function u^* that minimizes the objective function $J[u]$ as follows:

$$u^* := \arg\min_u J[u]. \tag{39}$$

5.2. Pontryagin's Minimum Principle

In the LQG problem of ML-POSC, Pontryagin's minimum principle (20) can be calculated as follows [14]:

$$u^*(t,z) = -R^{-1}B^\top(\Pi K(\Lambda)(s-\mu) + \Psi\mu), \ a.s. \ ^\forall t \in [0,T], \ ^\forall z \in \mathbb{R}^{d_z}, \tag{40}$$

where $K(\Lambda)$ is defined as follows:

$$K(\Lambda) := \begin{pmatrix} O & \Lambda_{xx}^{-1}\Lambda_{xz} \\ O & I \end{pmatrix}, \tag{41}$$

where $\mu(t)$ and $\Lambda(t)$ are the mean vector and the precision matrix of the extended state, respectively, which correspond to the solution of the FP Equation (22). We note that $\mathbb{E}_{p_t(z|x)}[s] = K(\Lambda)(s-\mu) + \mu$ is satisfied. $\mu(t)$ and $\Lambda(t)$ are the solutions of the following ordinary differential equations (ODEs):

$$\dot{\mu} = \left(A - BR^{-1}B^\top\Psi\right)\mu, \tag{42}$$

$$\dot{\Lambda} = -\left(A - BR^{-1}B^\top\Pi K(\Lambda)\right)^\top\Lambda - \Lambda\left(A - BR^{-1}B^\top\Pi K(\Lambda)\right) - \Lambda\sigma\sigma^\top\Lambda, \tag{43}$$

where $\mu(0) = \mu_0$ and $\Lambda(0) = \Lambda_0$. $\Psi(t)$ and $\Pi(t)$ are the control gain matrices of the deterministic and stochastic extended state, respectively, which correspond to the solution of the HJB Equation (21). $\Psi(t)$ and $\Pi(t)$ are the solutions of the following ODEs:

$$-\dot{\Psi} = Q + A^\top \Psi + \Psi A - \Psi B R^{-1} B^\top \Psi, \tag{44}$$

$$-\dot{\Pi} = Q + A^\top \Pi + \Pi A - \Pi B R^{-1} B^\top \Pi + (I - K(\Lambda))^\top \Pi B R^{-1} B^\top \Pi (I - K(\Lambda)), \tag{45}$$

where $\Psi(T) = \Pi(T) = P$. The ODE of Ψ (44) is the Riccati equation [16–18], which also appears in the LQG problem of COSC. In contrast, the ODE of Π (45) is the partially observable Riccati equation [14], which appears only in the LQG problem of ML-POSC. The above result is obtained in [14].

The ODE of Ψ (44) can be solved backward in time from the terminal condition. Using Ψ, the ODE of μ (42) can be solved forward in time from the initial condition. In contrast, the ODEs of Π (45) and Λ (43) cannot be solved in a similar way as the ODEs of Ψ (44) and μ (42) because they interact with each other, which is a similar problem to the system of HJB-FP equations.

5.3. Forward-Backward Sweep Method

In the LQG problem of ML-POSC, FBSM is reduced from Algorithm 1 to Algorithm 2. $\mathcal{F}(\Lambda, \Pi)$ and $\mathcal{G}(\Lambda, \Pi)$ are defined by the right-hand sides of the ODEs of Λ (43) and Π (45), respectively, as follows:

$$\mathcal{F}(\Lambda, \Pi) := -\Big(A - BR^{-1}B^\top \Pi K(\Lambda)\Big)^\top \Lambda - \Lambda\Big(A - BR^{-1}B^\top \Pi K(\Lambda)\Big) - \Lambda\sigma\sigma^\top\Lambda,$$

$$\mathcal{G}(\Lambda, \Pi) := Q + A^\top \Pi + \Pi A - \Pi B R^{-1} B^\top \Pi + (I - K(\Lambda))^\top \Pi B R^{-1} B^\top \Pi (I - K(\Lambda)).$$

This result is obtained in Appendix C.9. Importantly, in the LQG problem of ML-POSC, FBSM computes the ODEs of Λ (43) and Π (45) instead of the FP Equation (22) and the HJB Equation (21).

Algorithm 2 Forward-Backward Sweep Method (FBSM) in the LQG problem

```
//— Initial step —//
k ← 0
Λ_0^k ← Λ_0
for t = 0 to T − dt do
    Initialize Π_{t+dt}^k
    Λ_{t+dt}^k ← Λ_t^k + F(Λ_t^k, Π_{t+dt}^k)dt
end for
while J[u_{0:T−dt}^k] do not converge do
    if k is even then
        //— Backward step —//
        Π_T^{k+1} ← P
        for t = T − dt to 0 do
            Π_t^{k+1} ← Π_{t+dt}^{k+1} + G(Λ_t^k, Π_{t+dt}^{k+1})dt
        end for
    else
        //— Forward step —//
        Λ_0^{k+1} ← Λ_0
        for t = 0 to T − dt do
            Λ_{t+dt}^{k+1} ← Λ_t^{k+1} + F(Λ_t^{k+1}, Π_{t+dt}^k)dt
        end for
    end if
    k ← k + 1
end while
return u_{0:T−dt}^k
```

6. Numerical Experiments

In this section, we verify the convergence of FBSM in ML-POSC by performing numerical experiments on the LQG and non-LQG problems. The setting of the numerical experiments is the same as the previous work [14].

6.1. LQG Problem

In this subsection, we verify the convergence of FBSM for ML-POSC by conducting a numerical experiment on the LQG problem. We consider state $x_t \in \mathbb{R}$, observation $y_t \in \mathbb{R}$, and memory $z_t \in \mathbb{R}$, which evolve by the following SDEs:

$$dx_t = (x_t + u_t)dt + d\omega_t, \tag{46}$$

$$dy_t = x_t dt + d\nu_t, \tag{47}$$

$$dz_t = v_t dt + dy_t, \tag{48}$$

where x_0 and z_0 obey the standard Gaussian distributions, y_0 is an arbitrary real number, $\omega_t \in \mathbb{R}$ and $\nu_t \in \mathbb{R}$ are independent standard Wiener processes, and $u_t = u(t, z_t) \in \mathbb{R}$ and $v_t = v(t, z_t) \in \mathbb{R}$ are the controls. The objective function to be minimized is given as follows:

$$J[u, v] := \mathbb{E}_{p(x_{0:10}, y_{0:10}, z_{0:10}; u, v)} \left[\int_0^{10} \left(x_t^2 + u_t^2 + v_t^2 \right) dt \right]. \tag{49}$$

Therefore, the objective of this problem is to minimize the state variance with small state and memory controls.

This problem corresponds to the LQG problem, which is defined by (37) and (38). By defining $s_t := (x_t, z_t) \in \mathbb{R}^2$, $\tilde{u}_t := (u_t, v_t) \in \mathbb{R}^2$, and $\tilde{\omega}_t := (\omega_t, \nu_t) \in \mathbb{R}^2$, the SDEs (46)–(48) can be rewritten as follows:

$$ds_t = \left(\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} s_t + \tilde{u}_t \right) dt + d\tilde{\omega}_t, \tag{50}$$

which corresponds to (37). Furthermore, the objective function (49) can be rewritten as follows:

$$J[\tilde{u}] := \mathbb{E}_{p(s_{0:10}; \tilde{u})} \left[\int_0^{10} \left(s_t^\top \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} s_t + \tilde{u}_t^\top \tilde{u}_t \right) dt \right], \tag{51}$$

which corresponds to (38).

We apply the FBSM of the LQG problem (Algorithm 2) to this problem. $\Pi^0(t)$ is initialized by $\Pi^0(t) = O$. To solve the ODEs of $\Pi^k(t)$ and $\Lambda^k(t)$, we use the fourth-order Runge–Kutta method. Figure 4 shows the control gain matrix $\Pi^k(t) \in \mathbb{R}^{2\times 2}$ and the precision matrix $\Lambda^k(t) \in \mathbb{R}^{2\times 2}$ obtained by FBSM. The color of each curve represents the iteration k. The darkest curve corresponds to the first iteration $k = 0$, and the brightest curve corresponds to the last iteration $k = 50$. Importantly, $\Pi^k(t)$ and $\Lambda^k(t)$ converge with respect to the iteration k.

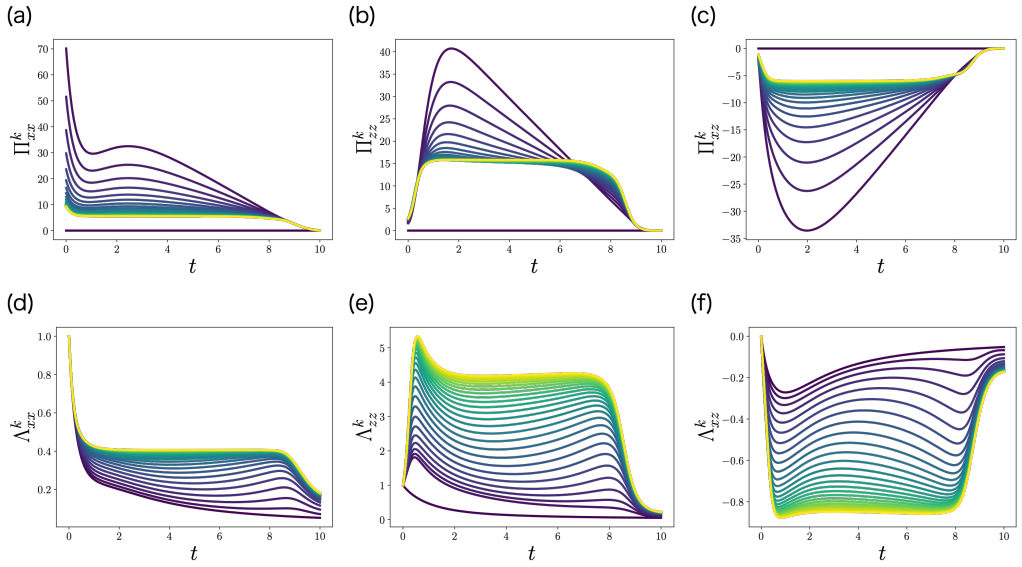

Figure 4. The elements of the control gain matrix $\Pi^k(t) \in \mathbb{R}^{2\times 2}$ (**a**–**c**) and the precision matrix $\Lambda^k(t) \in \mathbb{R}^{2\times 2}$ (**d**–**f**) obtained by FBSM (Algorithm 2) in the numerical experiment of the LQG problem of ML-POSC. Because $\Pi^k_{zx}(t) = \Pi^k_{xz}(t)$ and $\Lambda^k_{zx}(t) = \Lambda^k_{xz}(t)$, $\Pi^k_{zx}(t)$ and $\Lambda^k_{zx}(t)$ are not visualized. The darkest curve corresponds to the first iteration $k = 0$, and the brightest curve corresponds to the last iteration $k = 50$. $\Pi^0(t)$ is initialized by $\Pi^0(t) = O$.

Figure 5a shows the objective function $J[u^k]$ with respect to iteration k. The objective function $J[u^k]$ monotonically decreases with respect to iteration k, which is consistent with Section 4.3. This monotonicity of FBSM is the nice property of ML-POSC that is not guaranteed in deterministic control and MFSC. The objective function $J[u^k]$ finally converges, and u^k satisfies Pontryagin's minimum principle from Section 4.4.

Figure 5b–d compare the performance of the control function u^k at the first iteration $k = 0$ and the last iteration $k = 50$ by performing a stochastic simulation. At the first iteration $k = 0$, the distributions of state and memory are unstable, and the cumulative cost diverges. In contrast, at the last iteration $k = 50$, the distributions of state and memory are stabilized and the cumulative cost is smaller. This result indicates that FBSM improves the performance in ML-POSC.

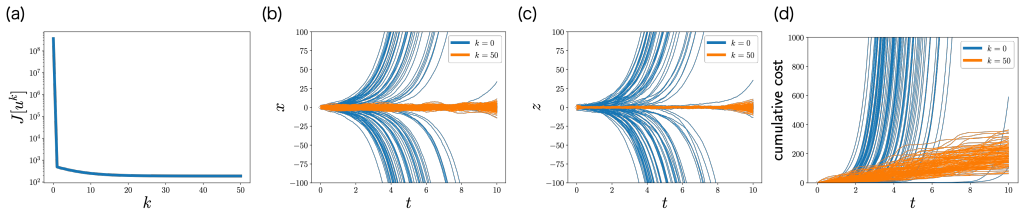

Figure 5. Performance of FBSM in the numerical experiment of the LQG problem of ML-POSC. (**a**) The objective function $J[u^k]$ with respect to the iteration k. (**b**–**d**) Stochastic simulation of state x_t (**b**), memory z_t (**c**), and the cumulative cost (**d**) for 100 samples. The expectation of the cumulative cost at $t = 10$ corresponds to the objective function (49). Blue and orange curves correspond to the first iteration $k = 0$ and the last iteration $k = 50$, respectively.

Although Figure 5b–d look similar to Figure 2d–f in the previous work [14], they are comparing different things. While Figure 5b–d demonstrate the performance improvement by the FBSM iteration, the previous work [14] compares the performance of the partially observable Riccati Equation (45) with that of the conventional Riccati Equation (44).

6.2. Non-LQG Problem

In this subsection, we verify the convergence of FBSM in ML-POSC by conducting a numerical experiment on the non-LQG problem. We consider state $x_t \in \mathbb{R}$, observation $y_t \in \mathbb{R}$, and memory $z_t \in \mathbb{R}$, which evolve by the following SDEs:

$$dx_t = u_t dt + d\omega_t, \tag{52}$$
$$dy_t = x_t dt + d\nu_t, \tag{53}$$
$$dz_t = dy_t, \tag{54}$$

where x_0 and z_0 obey the Gaussian distributions $p_0(x_0) = \mathcal{N}(x_0|0, 0.01)$ and $p_0(z_0) = \mathcal{N}(z_0|0, 0.01)$, respectively. y_0 is an arbitrary real number, $\omega_t \in \mathbb{R}$ and $\nu_t \in \mathbb{R}$ are independent standard Wiener processes, and $u_t = u(t, z_t) \in \mathbb{R}$ is the control. For the sake of simplicity, memory control is not considered. The objective function to be minimized is given as follows:

$$J[u] := \mathbb{E}_{p(x_{0:1}, y_{0:1}, z_{0:1}; u)} \left[\int_0^1 \left(Q(t, x_t) + u_t^2 \right) dt + 10 x_1^2 \right], \tag{55}$$

where

$$Q(t, x) := \begin{cases} 1000 & (0.3 \leq t \leq 0.6, 0.1 \leq |x| \leq 2.0), \\ 0 & (others). \end{cases} \tag{56}$$

The cost function is high in $0.3 \leq t \leq 0.6$ and $0.1 \leq |x| \leq 2.0$, which represents the obstacles. In addition, the terminal cost function is the lowest at $x = 0$, which represents the desirable goal. Therefore, the system should avoid the obstacles and reach the goal with a small control. Because the cost function is non-quadratic, it is a non-LQG problem.

We apply the FBSM (Algorithm 1) to this problem. $u^0(t, z)$ is initialized by $u^0(t, z) = 0$. To solve the HJB equation and the FP equation, we use the finite-difference method. Figure 6 shows $w^k(t, s)$ and $p^k(t, s)$ obtained by FBSM at the first iteration $k = 0$ and at the last iteration $k = 50$. From Appendix C.6, $w^k(t, s)$ is given as follows:

$$w^k(t, s) = \mathbb{E}_{p(s_{t+dt:1} | s_t = s; u^k)} \left[\int_t^1 \left(Q(\tau, x_\tau) + (u_\tau^k)^2 \right) d\tau + 10 x_1^2 \right]. \tag{57}$$

Because $u^0(t, z) = 0$, $w^0(t, s)$ reflects the cost function corresponding to the obstacles and the goal (Figure 6a–e). In contrast, because $u^{50}(t, z) \neq 0$, $w^{50}(t, s)$ becomes more complex (Figure 6f–j). In particular, while $w^0(t, s)$ does not depend on memory z, $w^{50}(t, s)$ depends on memory z, which indicates that the control function $u^{50}(t, z)$ is adjusted by the memory z. We note that $w^0(1, s)$ (Figure 6e) and $w^{50}(1, s)$ (Figure 6j) are the same because they are given by the terminal cost function as $w^0(1, s) = w^{50}(1, s) = 10x^2$. Furthermore, while $p^0(t, s)$ is a unimodal distribution (Figure 6k–o), $p^{50}(t, s)$ is a bimodal distribution (Figure 6p–t), which can avoid the obstacles.

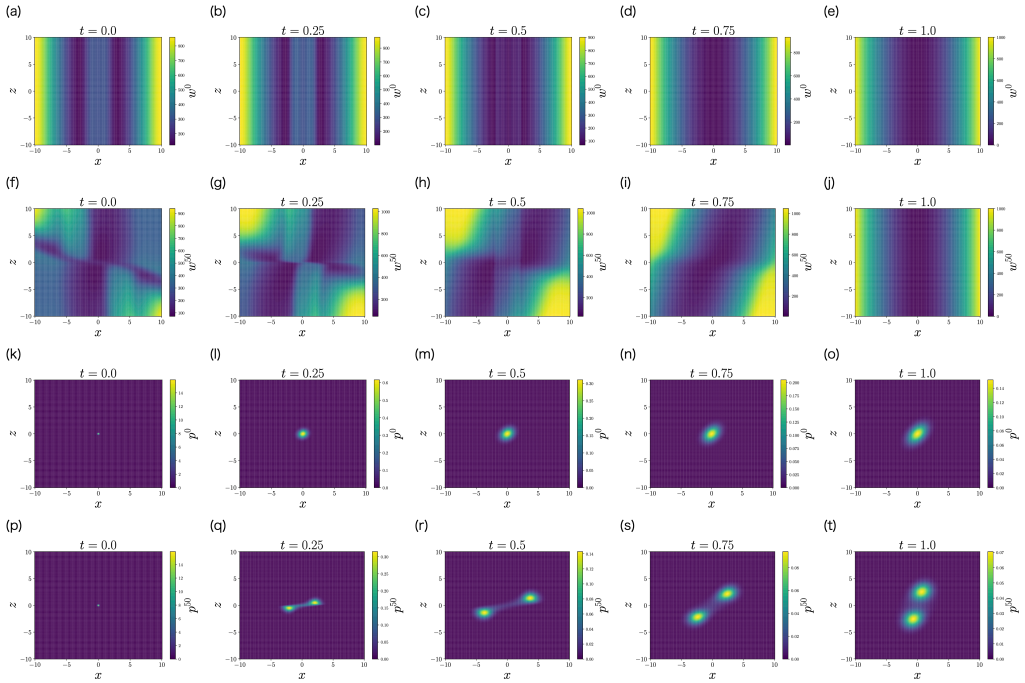

Figure 6. The solutions of the HJB equation $w^k(t,s)$ (**a**–**j**) and the FP equation $p^k(t,s)$ (**k**–**t**) at the first iteration $k = 0$ (**a**–**e**,**k**–**o**) and at the last iteration $k = 50$ (**f**–**j**,**p**–**t**) of FBSM (Algorithm 1) in the numerical experiment of the non-LQG problem of ML-POSC. $u^0(t,z)$ is initialized by $u^0(t,z) = 0$.

Figure 7a shows the objective function $J[u^k]$ with respect to iteration k. The objective function $J[u^k]$ monotonically decreases with respect to iteration k, which is consistent with Section 4.3. This monotonicity of FBSM is the nice property of ML-POSC that is not guaranteed in deterministic control and MFSC. The objective function $J[u^k]$ finally converges, and its u^k satisfies Pontryagin's minimum principle from Section 4.4.

Figure 7b,c compare the performance of the control function u^k at the first iteration $k = 0$ and the last iteration $k = 50$ by conducting the stochastic simulation. At the first iteration $k = 0$, the obstacles cannot be avoided, which results in a higher objective function. In contrast, at the last iteration $k = 50$, the obstacles can be avoided, which results in a lower objective function. This result indicates that FBSM improves the performance in ML-POSC.

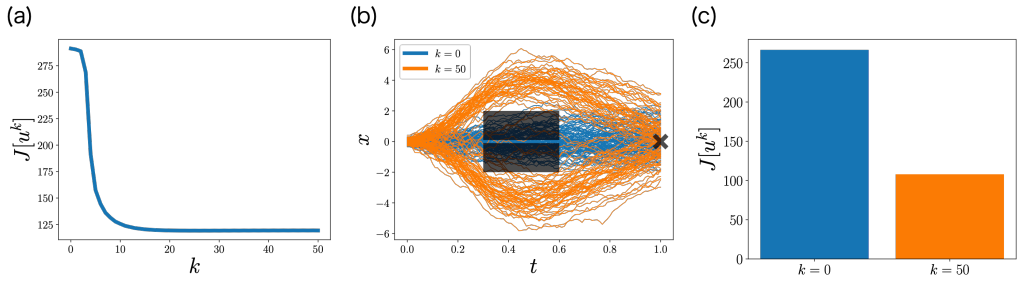

Figure 7. Performance of FBSM in the numerical experiment of the non-LQG problem of ML-POSC. (**a**) The objective function $J[u^k]$ with respect to the iteration k. (**b**) Stochastic simulation of the state x_t for 100 samples. The black rectangles and the cross represent the obstacles and the goal, respectively. Blue and orange curves correspond to the first iteration $k = 0$ and the last iteration $k = 50$, respectively. (**c**) The objective function (55), which is computed from 100 samples.

Although Figure 7b,c look similar to Figure 3a,b in the previous work [14], they are comparing different things. While Figure 7b,c demonstrate the performance improvement by the FBSM iteration, the previous work [14] compares the performance of ML-POSC with the local LQG approximation of the conventional POSC.

7. Discussion

In this work, we first showed that the system of HJB-FP equations corresponds to Pontryagin's minimum principle on the probability density function space. Although the relationship between the system of HJB-FP equations and Pontryagin's minimum principle has been briefly mentioned in MFSC [29–31], its details have not yet been investigated. We addressed this problem by deriving the system of HJB-FP equations in a similar way to Pontryagin's minimum principle. We then proposed FBSM to ML-POSC. Although the convergence of FBSM is generally not guaranteed in deterministic control [32,34,35,38] and MFSC [39–42], we proved the convergence in ML-POSC by noting the fact that the update of the current control function does not affect the future HJB equation in ML-POSC. Therefore, ML-POSC is a special and nice class where FBSM is guaranteed to converge.

Our derivation of Pontryagin's minimum principle on the probability density function space is formal, not analytical. Therefore, more mathematically rigorous proofs should be pursued in future work. Nevertheless, because our results are consistent with the conventional results of COSC [15–18], ML-POSC [14], and MFSC [26–28,30,31], they would be reliable except for special cases. Furthermore, our results provide a unified perspective on FBSM in deterministic control [32,34,35,38] and the fixed-point iteration method in MFSC [39–42], which have been studied independently. It clarifies the different properties of ML-POSC from deterministic control and MFSC, which ensures the convergence of FBSM.

The regularized FBSM has recently been proposed in deterministic control, which is guaranteed to converge even in the general deterministic control [44,45]. Our work gives an intuitive reason why the regularized FBSM is guaranteed to converge. In the regularized FBSM, the Hamiltonian is regularized, which makes the update of the control function smaller. When the regularization is sufficiently strong, the effect of the current control function on the future backward dynamics would be negligible. Therefore, the regularized FBSM of deterministic control would be guaranteed to converge for a similar reason to the FBSM of ML-POSC. However, the convergence of the regularized FBSM is much slower because the stronger regularization makes the update of the control function smaller. The FBSM of ML-POSC does not suffer from such a problem because the future backward dynamics already do not depend on the current control function without regularization.

Our work gives a hint about a modification of the fixed-point iteration method to ensure convergence in MFSC. Although the fixed-point iteration method is the most basic algorithm in MFSC, its convergence is not guaranteed [39–42]. Our work showed that the fixed-point iteration method is equivalent to the FBSM on the probability density function space. Therefore, the idea of regularized FBSM may also be applied to the fixed-point iteration method. More specifically, the fixed-point iteration method may be guaranteed to converge by regularizing the expected Hamiltonian.

In FBSM, we solve the HJB equation and the FP equation using the finite-difference method. However, because the finite-difference method is prone to the curse of dimensionality, it is difficult to solve high-dimensional ML-POSC. To resolve this problem, two directions can be considered. One direction is the policy iteration method [21,46,47]. Although the policy iteration method is almost the same as FBSM, only the update of the control function is different. While FBSM updates the system of HJB-FP equations and the control function simultaneously, the policy iteration method updates them separately. In the policy iteration method, the system of HJB-FP equations becomes linear, which can be solved by the sampling method [48–50]. Because the sampling method is more tractable than the finite-difference method, the policy iteration method may allow high-dimensional ML-POSC to be solved. Furthermore, the policy iteration method has recently been studied in MFSC [51–53]. However, its convergence is not guaranteed except for special cases in MFSC. In a similar way to FBSM, the convergence of the policy iteration method may be guaranteed in ML-POSC.

The other direction is machine learning. Neural network-based algorithms have recently been proposed in MFSC, which can solve high-dimensional problems efficiently [54,55]. By extending these algorithms, high-dimensional ML-POSC may be solved efficiently. Fur-

thermore, unlike MFSC, the coupling of the HJB-FP equations is limited only to the optimal control function in ML-POSC. By exploiting this nice property, more efficient algorithms may be devised for ML-POSC.

Author Contributions: Conceptualization, Formal analysis, Funding acquisition, Writing—original draft, T.T. and T.J.K.; Software, Visualization, T.T. All authors have read and agreed to the published version of the manuscript.

Funding: The first author received a JSPS Research Fellowship (Grant No. 21J20436). This work was supported by JSPS KAKENHI (Grant No. 19H05799) and JST CREST (Grant No. JPMJCR2011).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COSC	Completely Observable Stochastic Control
POSC	Partially Observable Stochastic Control
ML-POSC	Memory-Limited Partially Observable Stochastic Control
MFSC	Mean-Field Stochastic Control
FBSM	Forward-Backward Sweep Method
HJB	Hamilton-Jacobi-Bellman
FP	Fokker-Planck
SDE	Stochastic Differential Equation
ODE	Ordinary Differential Equation
LQG	Linear-Quadratic-Gaussian

Appendix A. Deterministic Control

In this section, we briefly review Pontryagin's minimum principle in deterministic control [22–25].

Appendix A.1. Problem Formulation

In this subsection, we formulate deterministic control [22–25]. The state of the system $s_t \in \mathbb{R}^{d_s}$ at time $t \in [0, T]$ evolves according to the following ordinary differential equation (ODE):

$$\frac{ds_t}{dt} = b(t, s_t, u_t), \tag{A1}$$

where the initial state is s_0, and the control is $u_t = u(t) \in \mathbb{R}^{d_u}$. The objective function is given by the following cumulative cost function:

$$J[u] := \int_0^T f(t, s_t, u_t)dt + g(s_T), \tag{A2}$$

where f is the cost function and g is the terminal cost function. Deterministic control is the problem of finding the optimal control function u^* that minimizes the cumulative cost function $J[u]$ as follows:

$$u^* := \arg\min_u J[u]. \tag{A3}$$

Appendix A.2. Preliminary

In this subsection, we show a useful result in deriving Pontryagin's minimum principle. Given arbitrary control functions u and u', $J[u] - J[u']$ can be calculated as follows [16]:

$$J[u] - J[u'] = \int_0^T \left(\mathcal{H}(t, s_t, u_t, \lambda'_t) - \mathcal{H}(t, s'_t, u'_t, \lambda'_t) - \left(\frac{\partial \mathcal{H}(t, s'_t, u'_t, \lambda'_t)}{\partial s} \right)^\top (s_t - s'_t) \right) dt + g(s_T) - g(s'_T) - \left(\frac{\partial g(s'_T)}{\partial s} \right)^\top (s_T - s'_T), \tag{A4}$$

where $\mathcal{H}$ is the Hamiltonian, which is defined as follows:

$$\mathcal{H}(t, s, u, \lambda) := f(t, s, u) + \lambda^\top b(t, s, u). \tag{A5}$$

λ'_t is the solution of the following adjoint equation driven by u':

$$-\frac{d\lambda'_t}{dt} = \frac{\partial \mathcal{H}(t, s'_t, u'_t, \lambda'_t)}{\partial s}, \tag{A6}$$

where $\lambda'_T = \partial g(s'_T)/\partial s$. s_t and s'_t are the solutions of the state Equation (A1) driven by u and u', respectively.

In the following, we derive Equation (A4). $J[u] - J[u']$ can be calculated as follows:

$$\begin{aligned} J[u] - J[u'] &= \left[\int_0^T f(t, s_t, u_t) dt + g(s_T) \right] - \left[\int_0^T f(t, s'_t, u'_t) dt + g(s'_T) \right] \\ &= \left[\int_0^T \Big(\mathcal{H}(t, s_t, u_t, \lambda'_t) - (\lambda'_t)^\top b(t, s_t, u_t) \Big) dt + g(s_T) \right] \\ &\quad - \left[\int_0^T \Big(\mathcal{H}(t, s'_t, u'_t, \lambda'_t) - (\lambda'_t)^\top b(t, s'_t, u'_t) \Big) dt + g(s'_T) \right] \\ &= \int_0^T \big(\mathcal{H}(t, s_t, u_t, \lambda'_t) - \mathcal{H}(t, s'_t, u'_t, \lambda'_t) \big) dt \\ &\quad - \int_0^T (\lambda'_t)^\top \big(b(t, s_t, u_t) - b(t, s'_t, u'_t) \big) dt + g(s_T) - g(s'_T). \end{aligned} \tag{A7}$$

From the state Equation (A1),

$$\begin{aligned} J[u] - J[u'] &= \int_0^T \big(\mathcal{H}(t, s_t, u_t, \lambda'_t) - \mathcal{H}(t, s'_t, u'_t, \lambda'_t) \big) dt \\ &\quad - \int_0^T (\lambda'_t)^\top \frac{d(s_t - s'_t)}{dt} dt + g(s_T) - g(s'_T). \end{aligned} \tag{A8}$$

From the integration by parts and $s_0 - s'_0 = 0$,

$$\begin{aligned} J[u] - J[u'] &= \int_0^T \big(\mathcal{H}(t, s_t, u_t, \lambda'_t) - \mathcal{H}(t, s'_t, u'_t, \lambda'_t) \big) dt \\ &\quad + \int_0^T \left(\frac{d\lambda'_t}{dt} \right)^\top (s_t - s'_t) dt + g(s_T) - g(s'_T) - (\lambda'_T)^\top (s_T - s'_T). \end{aligned} \tag{A9}$$

From the adjoint Equation (A6), Equation (A4) is obtained.

Appendix A.3. Necessary Condition

In this subsection, we show the necessary condition of the optimal control function of deterministic control. It corresponds to Pontryagin's minimum principle on the state space

(Figure 2 (bottom left)). If u^* is the optimal control function of deterministic control (A3), then the following equation is satisfied [16]:

$$u^*(t) = \arg\min_u \mathcal{H}(t, s_t^*, u, \lambda_t^*), \ ^\forall t \in [0, T], \tag{A10}$$

where λ_t^* is the solution of the following adjoint equation driven by u^*:

$$-\frac{d\lambda_t^*}{dt} = \frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s}, \tag{A11}$$

where $\lambda_T^* = \partial g(s_T^*)/\partial s$. s_t^* is the solution of the following state equation driven by u^*:

$$\frac{ds_t^*}{dt} = \frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial \lambda}, \tag{A12}$$

where $s_0^* = s_0$. Because $\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)/\partial \lambda = b(t, s_t^*, u_t^*)$, Equation (A12) is consistent with Equation (A1).

In the following, we show that Equation (A10) is the necessary condition of the optimal control function of deterministic control. We define the control function:

$$u^\varepsilon(t) := \begin{cases} u^*(t) & t \in [0, T] \backslash E_\varepsilon, \\ u(t) & t \in E_\varepsilon, \end{cases} \tag{A13}$$

where $E_\varepsilon := [t', t' + \varepsilon] \subseteq [0, T]$, and $^\forall u : [0, T] \to \mathbb{R}^{d_u}$. From Equation (A4), $J[u^\varepsilon] - J[u^*]$ can be calculated as follows:

$$\begin{aligned} J[u^\varepsilon] - J[u^*] &= \int_0^T \left(\mathcal{H}(t, s_t^\varepsilon, u_t^\varepsilon, \lambda_t^*) - \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*) - \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s} \right)^\top (s_t^\varepsilon - s_t^*) \right) dt \\ &\quad + g(s_T^\varepsilon) - g(s_T^*) - \left(\frac{\partial g(s_T^*)}{\partial s} \right)^\top (s_T^\varepsilon - s_T^*) \\ &= \int_0^T \left(\mathcal{H}(t, s_t^\varepsilon, u_t^*, \lambda_t^*) - \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*) - \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s} \right)^\top (s_t^\varepsilon - s_t^*) \right) dt \\ &\quad + g(s_T^\varepsilon) - g(s_T^*) - \left(\frac{\partial g(s_T^*)}{\partial s} \right)^\top (s_T^\varepsilon - s_T^*) \\ &\quad + \int_{E_\varepsilon} (\mathcal{H}(t, s_t^\varepsilon, u_t, \lambda_t^*) - \mathcal{H}(t, s_t^\varepsilon, u_t^*, \lambda_t^*)) dt. \end{aligned} \tag{A14}$$

Letting $\varepsilon \to 0$,

$$\begin{aligned} J[u^\varepsilon] - J[u^*] &= \int_0^T \left(\left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s} \right)^\top (s_t^\varepsilon - s_t^*) - \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s} \right)^\top (s_t^\varepsilon - s_t^*) \right) dt \\ &\quad + \left(\frac{\partial g(s_T^*)}{\partial s} \right)^\top (s_T^\varepsilon - s_T^*) - \left(\frac{\partial g(s_T^*)}{\partial s} \right)^\top (s_T^\varepsilon - s_T^*) \\ &\quad + (\mathcal{H}(t', s_{t'}^*, u_{t'}, \lambda_{t'}^*) - \mathcal{H}(t', s_{t'}^*, u_{t'}^*, \lambda_{t'}^*)) dt \\ &= (\mathcal{H}(t', s_{t'}^*, u_{t'}, \lambda_{t'}^*) - \mathcal{H}(t', s_{t'}^*, u_{t'}^*, \lambda_{t'}^*)) dt. \end{aligned} \tag{A15}$$

Because u^* is the optimal control function, the following inequality is satisfied:

$$0 \leq J[u^\varepsilon] - J[u^*] = (\mathcal{H}(t', s_{t'}^*, u_{t'}, \lambda_{t'}^*) - \mathcal{H}(t', s_{t'}^*, u_{t'}^*, \lambda_{t'}^*)) dt. \tag{A16}$$

Therefore, Equation (A10) is the necessary condition of the optimal control function of deterministic control.

Appendix A.4. Sufficient Condition

Pontryagin's minimum principle (A10) is a necessary condition and generally not a sufficient condition. Pontryagin's minimum principle (A10) becomes a necessary and sufficient condition if the Hamiltonian $\mathcal{H}(t, s, u, \lambda)$ is convex with respect to s and u and the terminal cost function $g(s)$ is convex with respect to s.

In the following, we show this result. We define the arbitrary control function ${}^{\forall}u : [0, T] \to \mathbb{R}^{d_u}$. From Equation (A4), $J[u] - J[u^*]$ is given by the following equation:

$$J[u] - J[u^*] = \int_0^T \left(\mathcal{H}(t, s_t, u_t, \lambda_t^*) - \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*) - \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s} \right)^\top (s_t - s_t^*) \right) dt + g(s_T) - g(s_T^*) - \left(\frac{\partial g(s_T^*)}{\partial s} \right)^\top (s_T - s_T^*). \tag{A17}$$

Since $\mathcal{H}(t, s, u, \lambda)$ is convex with respect to s and u and $g(s)$ is convex with respect to s, the following inequalities are satisfied:

$$\mathcal{H}(t, s_t, u_t, \lambda_t^*) \geq \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*) + \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial s} \right)^\top (s_t - s_t^*) + \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial u} \right)^\top (u_t - u_t^*), \tag{A18}$$

$$g(s_T) \geq g(s_T^*) + \left(\frac{\partial g(s_T^*)}{\partial s} \right)^\top (s_T - s_T^*). \tag{A19}$$

Hence, the following inequality is satisfied:

$$J[u] - J[u^*] \geq \int_0^T \left(\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial u} \right)^\top (u_t - u_t^*) dt. \tag{A20}$$

Because u^* satisfies (A10), the following stationary condition is satisfied:

$$\frac{\partial \mathcal{H}(t, s_t^*, u_t^*, \lambda_t^*)}{\partial u} = 0. \tag{A21}$$

Hence, the following inequality is satisfied:

$$J[u] - J[u^*] \geq 0. \tag{A22}$$

Therefore, Equation (A10) is the sufficient condition of the optimal control function of deterministic control if $\mathcal{H}(t, s, u, \lambda)$ is convex with respect to s and u and $g(s)$ is convex with respect to s.

Appendix A.5. Relationship with Bellman's Dynamic Programming Principle

From Bellman's dynamic programming principle on the state space (Figure 2 (top left)) [16], the optimal control function of deterministic control is given by the following equation:

$$u^*(t, s) = \arg\min_u \mathcal{H}\left(t, s, u, \frac{\partial w^*(t, s)}{\partial s} \right), \tag{A23}$$

where $w^*(t, s)$ is the value function on the state space, which is the solution of the following Hamilton-Jacobi-Bellman (HJB) equation:

$$-\frac{\partial w^*(t, s)}{\partial t} = \mathcal{H}\left(t, s, u^*, \frac{\partial w^*(t, s)}{\partial s} \right), \tag{A24}$$

where $w^*(T,s) = g(s)$. More specifically, the optimal control function of deterministic control is given by $u^*(t) = u^*(t, s_t^*)$, where s_t^* is the solution of the state Equation (A12).

The HJB Equation (A24) can be converted into the adjoint Equation (A11) by defining

$$\lambda_t^* := \frac{\partial w^*(t, s_t^*)}{\partial s}, \tag{A25}$$

where s_t^* is the solution of the state Equation (A12). This approach can be interpreted as the conversion from Bellman's dynamic programming principle (Figure 2 (top left)) to Pontryagin's minimum principle (Figure 2 (bottom left)) on the state space.

In the following, we obtain this result. First, we define

$$\Lambda^*(t,s) := \frac{\partial w^*(t,s)}{\partial s}. \tag{A26}$$

By differentiating the HJB Equation (A24) with respect to s, the following equation is obtained:

$$-\frac{\partial \Lambda^*(t,s)}{\partial t} = \frac{\partial \mathcal{H}(t,s,u^*,\Lambda^*)}{\partial s} + \left(\frac{\partial \Lambda^*(t,s)}{\partial s}\right)^\top b(t,s^*,u^*), \tag{A27}$$

where $\Lambda^*(T,s) = \partial g(s)/\partial s$. Then the derivative of $\lambda_t^* = \Lambda^*(t,s_t^*)$ with respect to t can be calculated as follows:

$$\frac{d\lambda_t^*}{dt} = \frac{\partial \Lambda^*(t,s_t^*)}{\partial t} + \left(\frac{\partial \Lambda^*(t,s_t^*)}{\partial s}\right)^\top \frac{ds_t^*}{dt}. \tag{A28}$$

By substituting Equation (A27) into Equation (A28), the following equation is obtained:

$$-\frac{d\lambda_t^*}{dt} = \frac{\partial \mathcal{H}(t,s,u^*,\lambda^*)}{\partial s} - \left(\frac{\partial \Lambda^*(t,s_t^*)}{\partial s}\right)^\top \underbrace{\left(\frac{ds_t^*}{dt} - b(t,s^*,u^*)\right)}_{(*)}. \tag{A29}$$

From the state Equation (A12), $(*) = 0$ is satisfied. Therefore, $\lambda^*(t)$ satisfies the adjoint Equation (A11).

Appendix B. Mean-Field Stochastic Control

In this section, we show that the system of HJB-FP equations in MFSC corresponds to Pontryagin's minimum principle on the probability density function space. Although the relationship between the system of HJB-FP equations and Pontryagin's minimum principle has been mentioned briefly in MFSC [29–31], its details have not yet been investigated. In this section, we address this problem by deriving the system of HJB-FP equations in the similar way as Appendix A. Although our derivations are formal, not analytical, our results are consistent with the conventional results of MFSC [26–28,30,31].

Appendix B.1. Problem Formulation

In this subsection, we formulate MFSC [26–28]. The state of the system $s_t \in \mathbb{R}^{d_s}$ at time $t \in [0,T]$ evolves by the following stochastic differential equation (SDE):

$$ds_t = b(t,s_t,p_t,u_t)dt + \sigma(t,s_t,p_t,u_t)d\omega_t, \tag{A30}$$

where s_0 obeys $p_0(s_0)$, $p_t(s) := p(t,s)$ is the probability density function of the state s, $u_t(s) := u(t,s) \in \mathbb{R}^{d_u}$ is the control, and $\omega_t \in \mathbb{R}^{d_\omega}$ is the standard Wiener process. The objective function is given by the following expected cumulative cost function:

$$J[u] := \mathbb{E}_{p(s_{0:T};u)}\left[\int_0^T f(t,s_t,p_t,u_t)dt + g(s_T,p_T)\right], \tag{A31}$$

where f is the cost function, g is the terminal cost function, $p(s_{0:T}; u)$ is the probability of $s_{0:t} := \{s_\tau | \tau \in [0, t]\}$ given u as a parameter, and $\mathbb{E}_p[\cdot]$ is the expectation with respect to probability p. MFSC is the problem of finding the optimal control function u^* that minimizes the expected cumulative cost function $J[u]$ as follows:

$$u^* := \arg\min_u J[u]. \tag{A32}$$

Appendix B.2. Preliminary

In this subsection, we show a useful result in deriving Pontryagin's minimum principle. Given arbitrary control functions u and u', $J[u] - J[u']$ can be calculated as follows:

$$\begin{aligned} J[u] - J[u'] = \int_0^T \Bigg(& \bar{\mathcal{H}}(t, p, u, w') - \bar{\mathcal{H}}(t, p', u', w') \\ & - \int \frac{\delta \bar{\mathcal{H}}(t, p', u', w')}{\delta p}(s)\big(p(t,s) - p'(t,s)\big) ds \Bigg) dt \\ & + \bar{g}(p) - \bar{g}(p') - \int \frac{\delta \bar{g}(p')}{\delta p}(s)\big(p(T,s) - p'(T,s)\big) ds, \end{aligned} \tag{A33}$$

where $\bar{\mathcal{H}}$ and $\bar{g}$ are the expected Hamiltonian and terminal cost function, respectively, which are defined as follows:

$$\bar{\mathcal{H}}(t, p, u, w) := \mathbb{E}_{p(s)}[\mathcal{H}(t, s, p, u, w)], \tag{A34}$$

$$\bar{g}(p) := \mathbb{E}_{p(s)}[g(s, p)]. \tag{A35}$$

$\mathcal{H}$ is the Hamiltonian, which is defined as follows:

$$\mathcal{H}(t, s, p, u, w) := f(t, s, p, u) + \mathcal{L}_u w(t, s). \tag{A36}$$

$\mathcal{L}_u$ is the backward diffusion operator, which is defined as follows:

$$\mathcal{L}_u w(t, s) := \sum_{i=1}^{d_s} b_i(t, s, p, u) \frac{\partial w(t, s)}{\partial s_i} + \frac{1}{2} \sum_{i,j=1}^{d_s} D_{ij}(t, s, p, u) \frac{\partial^2 w(t, s)}{\partial s_i \partial s_j}, \tag{A37}$$

where $D(t, s, p, u) := \sigma(t, s, p, u)\sigma^\top(t, s, p, u)$. w' is the solution of the following Hamilton-Jacobi-Bellman (HJB) equation driven by u':

$$-\frac{\partial w'(t, s)}{\partial t} = \frac{\delta \bar{\mathcal{H}}(t, p', u', w')}{\delta p}(s), \tag{A38}$$

where $w'(T, s) = (\delta \bar{g}(p')/\delta p)(s)$. p is the solution of the following Fokker-Planck (FP) equation driven by u:

$$\frac{\partial p(t, s)}{\partial t} = \mathcal{L}_u^\dagger p(t, s), \tag{A39}$$

where $p(0, s) = p_0(s)$. p' is the solution of the FP Equation (A39) driven by u'. $\mathcal{L}_u^\dagger$ is the forward diffusion operator, which is defined as follows:

$$\mathcal{L}_u^\dagger p(t, s) := -\sum_{i=1}^{d_s} \frac{\partial (b_i(t, s, p, u) p(t, s))}{\partial s_i} + \frac{1}{2} \sum_{i,j=1}^{d_s} \frac{\partial^2 (D_{ij}(t, s, p, u) p(t, s))}{\partial s_i \partial s_j}. \tag{A40}$$

$\mathcal{L}_u^\dagger$ is the conjugate of $\mathcal{L}_u$ as follows:

$$\int w(t, s) \mathcal{L}_u^\dagger p(t, s) ds = \int p(t, s) \mathcal{L}_u w(t, s) ds. \tag{A41}$$

In the following, we derive Equation (A33). $J[u] - J[u']$ can be calculated as follows:

$$\begin{aligned} J[u] - J[u'] &= \mathbb{E}_{p(s_{0:T})}\left[\int_0^T f(t, s_t, p_t, u_t)dt + g(s_T, p_T)\right] \\ &\quad - \mathbb{E}_{p'(s_{0:T})}\left[\int_0^T f(t, s_t, p'_t, u'_t)dt + g(s_T, p'_T)\right] \\ &= \mathbb{E}_{p(s_{0:T})}\left[\int_0^T (\mathcal{H}(t, s_t, p_t, u_t, w') - \mathcal{L}_{u_t} w'(t, s_t))dt + g(s_T, p_T)\right] \\ &\quad - \mathbb{E}_{p'(s_{0:T})}\left[\int_0^T \Big(\mathcal{H}(t, s_t, p'_t, u'_t, w') - \mathcal{L}_{u'_t} w'(t, s_t)\Big)dt + g(s_T, p'_T)\right] \\ &= \int_0^T (\bar{\mathcal{H}}(t, p, u, w') - \bar{\mathcal{H}}(t, p', u', w'))dt \\ &\quad - \int_0^T \Big(\mathbb{E}_{p(t,s)}\left[\mathcal{L}_u w'(t, s)\right] - \mathbb{E}_{p'(t,s)}\left[\mathcal{L}_{u'} w'(t, s)\right]\Big)dt + \bar{g}(p) - \bar{g}(p'). \end{aligned} \tag{A42}$$

Because $\mathcal{L}_{u_t}$ and $\mathcal{L}_{u'_t}$ are the conjugates of $\mathcal{L}^\dagger_{u_t}$ and $\mathcal{L}^\dagger_{u'_t}$, respectively,

$$\begin{aligned} J[u] - J[u'] &= \int_0^T (\bar{\mathcal{H}}(t, p, u, w') - \bar{\mathcal{H}}(t, p', u', w'))dt \\ &\quad - \int_0^T \int \Big(\mathcal{L}^\dagger_u p(t, s) - \mathcal{L}^\dagger_{u'} p'(t, s)\Big) w'(t, s)dsdt + \bar{g}(p) - \bar{g}(p'). \end{aligned} \tag{A43}$$

From the FP Equation (A39),

$$\begin{aligned} J[u] - J[u'] &= \int_0^T (\bar{\mathcal{H}}(t, p, u, w') - \bar{\mathcal{H}}(t, p', u', w'))dt \\ &\quad - \int_0^T \int \frac{\partial(p(t, s) - p'(t, s))}{\partial t} w'(t, s)dsdt + \bar{g}(p) - \bar{g}(p'). \end{aligned} \tag{A44}$$

From the integration by parts and $p(0, s) - p'(0, s) = p_0(s) - p_0(s) = 0$,

$$\begin{aligned} J[u] - J[u'] &= \int_0^T (\bar{\mathcal{H}}(t, p, u, w') - \bar{\mathcal{H}}(t, p', u', w'))dt \\ &\quad + \int_0^T \int (p(t, s) - p'(t, s)) \frac{\partial w'(t, s)}{\partial t} dsdt \\ &\quad + \bar{g}(p) - \bar{g}(p') - \int (p(T, s) - p'(T, s)) w'(T, s)ds. \end{aligned} \tag{A45}$$

From the HJB Equation (A38), Equation (A33) is obtained.

Appendix B.3. Necessary Condition

In this subsection, we show the necessary condition of the optimal control function of MFSC. It corresponds to Pontryagin's minimum principle on the probability density function space (Figure 2 (bottom right)). If u^* is the optimal control function of MFSC (A32), then the following equation is satisfied:

$$u^*(t, s) = \arg\min_u \mathcal{H}(t, s, p^*, u, w^*), \ a.s. \ ^\forall t \in [0, T], \ ^\forall s \in \mathbb{R}^{d_s}, \tag{A46}$$

where w^* is the solution of the following HJB equation driven by u^*:

$$-\frac{\partial w^*(t, s)}{\partial t} = \frac{\delta \bar{\mathcal{H}}(t, p^*, u^*, w^*)}{\delta p}(s), \tag{A47}$$

where $w^*(T,s) = (\delta \bar{g}(p^*)/\delta p)(s)$. p^* is the solution of the following FP equation driven by u^*:

$$\frac{\partial p^*(t,s)}{\partial t} = \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta w}(s), \tag{A48}$$

where $p^*(0,s) = p_0(s)$.

In the following, we show that Equation (A46) is the necessary condition of the optimal control function of MFSC. We define the control function

$$u^\varepsilon(t,z) := \begin{cases} u^*(t,s) & (t,s) \in ([0,T] \times \mathbb{R}^{d_s}) \backslash (E_{\varepsilon_1} \times F_{\varepsilon_2}), \\ u(t,s) & (t,s) \in E_{\varepsilon_1} \times F_{\varepsilon_2}, \end{cases} \tag{A49}$$

where $E_{\varepsilon_1} := [t', t'+\varepsilon_1] \subseteq [0,T]$, $F_{\varepsilon_2} := [s', s'+\varepsilon_2] \subseteq \mathbb{R}^{d_s}$, and ${}^\forall u : [0,T] \times \mathbb{R}^{d_s} \to \mathbb{R}^{d_u}$. From Equation (A33), $J[u^\varepsilon] - J[u^*]$ can be calculated as follows:

$$\begin{aligned} J[u^\varepsilon] - J[u^*] = & \int_0^T \Bigg(\bar{\mathcal{H}}(t,p^\varepsilon,u^\varepsilon,w^*) - \bar{\mathcal{H}}(t,p^*,u^*,w^*) \\ & - \int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p^\varepsilon(t,s) - p^*(t,s))ds \Bigg) dt \\ & + \bar{g}(p^\varepsilon) - \bar{g}(p^*) - \int \frac{\delta \bar{g}(p^*)}{\delta p}(s)(p^\varepsilon(T,s) - p^*(T,s))ds \\ = & \int_0^T \Bigg(\bar{\mathcal{H}}(t,p^\varepsilon,u^*,w^*) - \bar{\mathcal{H}}(t,p^*,u^*,w^*) \\ & - \int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p^\varepsilon(t,s) - p^*(t,s))ds \Bigg) dt \\ & + \bar{g}(p^\varepsilon) - \bar{g}(p^*) - \int \frac{\delta \bar{g}(p^*)}{\delta p}(s)(p^\varepsilon(T,s) - p^*(T,s))ds \\ & + \int_{E_{\varepsilon_1}} \int_{F_{\varepsilon_2}} (\mathcal{H}(t,s,p^\varepsilon,u,w^*) - \mathcal{H}(t,s,p^\varepsilon,u^*,w^*))p^\varepsilon(t,s)dsdt. \end{aligned} \tag{A50}$$

Letting $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$,

$$\begin{aligned} J[u^\varepsilon] - J[u^*] = & \int_0^T \Bigg(\int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p^\varepsilon(t,s) - p^*(t,s))ds \\ & - \int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p^\varepsilon(t,s) - p^*(t,s))ds \Bigg) dt \\ & + \int \frac{\delta \bar{g}(p^*)}{\delta p}(s)(p^\varepsilon(T,s) - p^*(T,s))ds - \int \frac{\delta \bar{g}(p^*)}{\delta p}(s)(p^\varepsilon(T,s) - p^*(T,s))ds \\ & + (\mathcal{H}(t',s',p^*,u,w^*) - \mathcal{H}(t',s',p^*,u^*,w^*))p^*(t',s')dsdt \\ = & (\mathcal{H}(t',s',p^*,u,w^*) - \mathcal{H}(t',s',p^*,u^*,w^*))p^*(t',s')dsdt. \end{aligned} \tag{A51}$$

Because u^* is the optimal control function, the following inequality is satisfied:

$$0 \leq J[u^\varepsilon] - J[u^*] = (\mathcal{H}(t',s',p^*,u,w^*) - \mathcal{H}(t',s',p^*,u^*,w^*))p^*(t',s')dsdt. \tag{A52}$$

Therefore, Equation (A46) is the necessary condition of the optimal control function of MFSC.

Appendix B.4. Sufficient Condition

Pontryagin's minimum principle (A46) is a necessary condition and generally not a sufficient condition. Pontryagin's minimum principle (A46) becomes a necessary and sufficient condition if the expected Hamiltonian $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u and the expected terminal cost function $\bar{g}(p)$ is convex with respect to p.

In the following, we show this result. We define the arbitrary control function ${}^{\forall}u : [0,T] \times \mathbb{R}^{d_s} \to \mathbb{R}^{d_u}$. From Equation (A33), $J[u] - J[u^*]$ is given by the following equation:

$$\begin{aligned} J[u] - J[u^*] = \int_0^T \Bigg(& \bar{\mathcal{H}}(t,p,u,w^*) - \bar{\mathcal{H}}(t,p^*,u^*,w^*) \\ & - \int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p(t,s) - p^*(t,s))ds \Bigg) dt \\ & + \bar{g}(p) - \bar{g}(p^*) - \int \frac{\delta \bar{g}(p^*)}{\delta p}(s)(p(T,s) - p^*(T,s))ds. \end{aligned} \tag{A53}$$

Because $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u and $\bar{g}(p)$ is convex with respect to p, the following inequalities are satisfied:

$$\begin{aligned} \bar{\mathcal{H}}(t,p,u,w^*) \geq \bar{\mathcal{H}}(t,p^*,u^*,w^*) + \int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p(t,s) - p^*(t,s))ds \\ + \int \left(\frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta u}(s) \right)^\top (u(t,s) - u^*(t,s))ds, \end{aligned} \tag{A54}$$

$$\bar{g}(p) \geq \bar{g}(p^*) + \int \frac{\delta \bar{g}(p^*)}{\delta p}(s)(p(T,s) - p^*(T,s))ds. \tag{A55}$$

Hence, the following inequality is satisfied:

$$J[u] - J[u^*] \geq \int_0^T \mathbb{E}_{p^*(t,s)} \left[\left(\frac{\partial \mathcal{H}(t,s,p^*,u^*,w^*)}{\partial u} \right)^\top (u(t,s) - u^*(t,s)) \right] dt. \tag{A56}$$

Because u^* satisfies Equation (A46), the following stationary condition is satisfied:

$$\frac{\partial \mathcal{H}(t,s,p^*,u^*,w^*)}{\partial u} = 0. \tag{A57}$$

Hence, the following inequality is satisfied:

$$J[u] - J[u^*] \geq 0 \tag{A58}$$

Therefore, Equation (A46) is the sufficient condition of the optimal control function of MFSC if the expected Hamiltonian $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u and the expected terminal cost function $\bar{g}(p)$ is convex with respect to p.

Appendix B.5. Relationship with Bellman's Dynamic Programming Principle

From Bellman's dynamic programming principle on the probability density function space (Figure 2 (top right)) [56–58], the optimal control function of MFSC is given by the following equation:

$$u^*(t,s,p) = \arg\min_u \mathcal{H}\left(t,s,p,u, \frac{\delta V^*(t,p)}{\delta p}(s)\right), \tag{A59}$$

where $V^*(t,p)$ is the value function on the probability density function space, which is the solution of the following Bellman equation:

$$-\frac{\partial V^*(t,p)}{\partial t} = \mathbb{E}_{p(s)}\left[\mathcal{H}\left(t,s,p,u^*, \frac{\delta V^*(t,p)}{\delta p}(s)\right)\right], \tag{A60}$$

where $V^*(T,p) = \mathbb{E}_{p(s)}[g(s)]$. More specifically, the optimal control function of MFSC is given by $u^*(t,s) = u^*(t,s,p^*)$, where p^* is the solution of the FP Equation (A48).

Because the Bellman Equation (A60) is a functional differential equation, it cannot be solved even numerically. To resolve this problem, the previous works [30,31] converted the Bellman Equation (A60) into the HJB Equation (A47) by defining

$$w^*(t,s) := \frac{\delta V^*(t,p^*)}{\delta p}(s), \tag{A61}$$

where p^* is the solution of FP Equation (A48). This approach can be interpreted as the conversion from Bellman's dynamic programming principle (Figure 2 (top right)) to Pontryagin's minimum principle (Figure 2 (bottom right)) on the probability density function space.

Appendix C. Derivation of Main Results

Appendix C.1. Derivation of Result in Section 3.1

In this subsection, we derive Equation (13). $J[u] - J[u']$ can be calculated as follows:

$$\begin{aligned}
J[u] - J[u'] &= \mathbb{E}_{p(s_{0:T})}\left[\int_0^T f(t,s_t,u_t)dt + g(s_T)\right] - \mathbb{E}_{p'(s_{0:T})}\left[\int_0^T f(t,s_t,u'_t)dt + g(s_T)\right] \\
&= \mathbb{E}_{p(s_{0:T})}\left[\int_0^T \left(\mathcal{H}(t,s_t,u_t,w') - \mathcal{L}_{u_t}w'(t,s_t)\right)dt + g(s_T)\right] \\
&\quad - \mathbb{E}_{p'(s_{0:T})}\left[\int_0^T \Big(\mathcal{H}(t,s_t,u'_t,w') - \mathcal{L}_{u'_t}w'(t,s_t)\Big)dt + g(s_T)\right] \\
&= \int_0^T \Big(\mathbb{E}_{p(t,s)}\left[\mathcal{H}(t,s,u,w')\right] - \mathbb{E}_{p'(t,s)}\left[\mathcal{H}(t,s,u',w')\right]\Big)dt \\
&\quad - \int_0^T \Big(\mathbb{E}_{p(t,s)}\left[\mathcal{L}_u w'(t,s)\right] - \mathbb{E}_{p'(t,s)}\left[\mathcal{L}_{u'}w'(t,s)\right]\Big)dt \\
&\quad + \mathbb{E}_{p(T,s)}[g(s)] - \mathbb{E}_{p'(T,s)}[g(s)].
\end{aligned} \tag{A62}$$

Because $\mathcal{L}_{u_t}$ and $\mathcal{L}_{u'_t}$ are the conjugates of $\mathcal{L}^\dagger_{u_t}$ and $\mathcal{L}^\dagger_{u'_t}$, respectively,

$$\begin{aligned}
J[u] - J[u'] &= \int_0^T \Big(\mathbb{E}_{p(t,s)}\left[\mathcal{H}(t,s,u,w')\right] - \mathbb{E}_{p'(t,s)}\left[\mathcal{H}(t,s,u',w')\right]\Big)dt \\
&\quad - \int_0^T \int \Big(\mathcal{L}^\dagger_u p(t,s) - \mathcal{L}^\dagger_{u'}p'(t,s)\Big)w'(t,s)dsdt \\
&\quad + \mathbb{E}_{p(T,s)}[g(s)] - \mathbb{E}_{p'(T,s)}[g(s)].
\end{aligned} \tag{A63}$$

From the FP Equation (17),

$$\begin{aligned}
J[u] - J[u'] &= \int_0^T \Big(\mathbb{E}_{p(t,s)}\left[\mathcal{H}(t,s,u,w')\right] - \mathbb{E}_{p'(t,s)}\left[\mathcal{H}(t,s,u',w')\right]\Big)dt \\
&\quad - \int_0^T \int \frac{\partial(p(t,s) - p'(t,s))}{\partial t}w'(t,s)dsdt \\
&\quad + \mathbb{E}_{p(T,s)}[g(s)] - \mathbb{E}_{p'(T,s)}[g(s)].
\end{aligned} \tag{A64}$$

From the integration by parts and $p(0,s) - p'(0,s) = p_0(s) - p_0(s) = 0$,

$$\begin{aligned}
J[u] - J[u'] &= \int_0^T \Big(\mathbb{E}_{p(t,s)}\left[\mathcal{H}(t,s,u,w')\right] - \mathbb{E}_{p'(t,s)}\left[\mathcal{H}(t,s,u',w')\right]\Big)dt \\
&\quad + \int_0^T \int \big(p(t,s) - p'(t,s)\big)\frac{\partial w'(t,s)}{\partial t}dsdt \\
&\quad + \mathbb{E}_{p(T,s)}[g(s)] - \mathbb{E}_{p'(T,s)}[g(s)] - \int \big(p(T,s) - p'(T,s)\big)w'(T,s)ds.
\end{aligned} \tag{A65}$$

From the HJB Equation (16), Equation (13) is obtained.

Appendix C.2. Derivation of Result in Section 3.2

In this subsection, we show that Equation (20) is the necessary condition of the optimal control function of ML-POSC. It corresponds to Pontryagin's minimum principle on the probability density function space. We define the control function

$$u^{\varepsilon}(t,z) := \begin{cases} u^*(t,z) & (t,z) \in ([0,T] \times \mathbb{R}^{d_z}) \backslash (E_{\varepsilon_1} \times F_{\varepsilon_2}), \\ u(t,z) & (t,z) \in E_{\varepsilon_1} \times F_{\varepsilon_2}, \end{cases} \tag{A66}$$

where $E_{\varepsilon_1} := [t', t' + \varepsilon_1] \subseteq [0,T]$, $F_{\varepsilon_2} := [z', z' + \varepsilon_2] \subseteq \mathbb{R}^{d_z}$, and $^{\forall}u : [0,T] \times \mathbb{R}^{d_z} \to \mathbb{R}^{d_u}$. From Equation (13), $J[u^{\varepsilon}] - J[u^*]$ can be calculated as follows:

$$\begin{aligned} J[u^{\varepsilon}] - J[u^*] &= \int_0^T \Big(\mathbb{E}_{p^{\varepsilon}(t,s)}[\mathcal{H}(t,s,u^{\varepsilon},w^*)] - \mathbb{E}_{p^{\varepsilon}(t,s)}[\mathcal{H}(t,s,u^*,w^*)] \Big) dt \\ &= \int_{E_{\varepsilon_1}} \int_{F_{\varepsilon_2}} \Big(\mathbb{E}_{p_t^{\varepsilon}(x|z)}[\mathcal{H}(t,s,u,w^*)] - \mathbb{E}_{p_t^{\varepsilon}(x|z)}[\mathcal{H}(t,s,u^*,w^*)] \Big) p_t^{\varepsilon}(z) dz dt. \end{aligned}$$

Letting $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$,

$$J[u^{\varepsilon}] - J[u^*] = \Big(\mathbb{E}_{p^*_{t'}(x'|z')}\left[\mathcal{H}(t',s',u,w^*)\right] - \mathbb{E}_{p^*_{t'}(x'|z')}\left[\mathcal{H}(t',s',u^*,w^*)\right] \Big) p^*_{t'}(z') dz dt.$$

Because u^* is the optimal control function, the following inequality is satisfied:

$$0 \leq J[u^{\varepsilon}] - J[u^*] = \Big(\mathbb{E}_{p^*_{t'}(x'|z')}\left[\mathcal{H}(t',s',u,w^*)\right] - \mathbb{E}_{p^*_{t'}(x'|z')}\left[\mathcal{H}(t',s',u^*,w^*)\right] \Big) p^*_{t'}(z') dz dt.$$

Therefore, Equation (20) is the necessary condition of the optimal control function of ML-POSC.

Appendix C.3. Derivation of Result in Section 3.3

In this subsection, we show that Equation (20) is the sufficient condition of the optimal control function of ML-POSC if the expected Hamiltonian $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u. We define the arbitrary control function $^{\forall}u : [0,T] \times \mathbb{R}^{d_z} \to \mathbb{R}^{d_u}$. From Equation (13), $J[u] - J[u^*]$ is given by the following equation:

$$J[u] - J[u^*] = \int_0^T \Big(\mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u,w^*)] - \mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u^*,w^*)] \Big) dt. \tag{A67}$$

Because $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u, the following inequality is satisfied:

$$\begin{aligned} \mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u,w^*)] &= \bar{\mathcal{H}}(t,p,u,w^*) \\ &\geq \bar{\mathcal{H}}(t,p^*,u^*,w^*) + \int \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s)(p(t,s) - p^*(t,s)) ds \\ &\quad + \int \left(\frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta u}(z) \right)^{\top} (u(t,z) - u^*(t,z)) dz. \end{aligned} \tag{A68}$$

Because

$$\begin{aligned} \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta p}(s) &= \frac{\delta}{\delta p} \left(\int p(s) \mathcal{H}(t,s,u^*,w^*) ds \right) \Big|_{p=p^*} \\ &= \mathcal{H}(t,s,u^*,w^*), \end{aligned} \tag{A69}$$

$$\begin{aligned} \frac{\delta \bar{\mathcal{H}}(t,p^*,u^*,w^*)}{\delta u}(z) &= \frac{\delta}{\delta u} \left(\int p_t^*(z) \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u,w^*)] dz \right) \Big|_{u=u^*} \\ &= p_t^*(z) \frac{\partial \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u^*,w^*)]}{\partial u}, \end{aligned} \tag{A70}$$

the above inequality can be calculated as follows:

$$\begin{aligned}\mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u,w^*)] &\geq \int p^*(t,s)\mathcal{H}(t,s,u^*,w^*)ds + \int \mathcal{H}(t,s,u^*,w^*)(p(t,s)-p^*(t,s))ds \\ &\quad + \int p_t^*(z)\left(\frac{\partial \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u^*,w^*)]}{\partial u}\right)^\top (u(t,z)-u^*(t,z))dz \\ &= \mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u^*,w^*)] \\ &\quad + \mathbb{E}_{p_t^*(z)}\left[\left(\frac{\partial \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u^*,w^*)]}{\partial u}\right)^\top (u(t,z)-u^*(t,z))\right]. \end{aligned} \tag{A71}$$

Hence, the following inequality is satisfied:

$$J[u]-J[u^*] \geq \int_0^T \mathbb{E}_{p_t^*(z)}\left[\left(\frac{\partial \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u^*,w^*)]}{\partial u}\right)^\top (u(t,z)-u^*(t,z))\right]dt. \tag{A72}$$

Because u^* satisfies Equation (20), the following stationary condition is satisfied:

$$\frac{\partial \mathbb{E}_{p_t^*(x|z)}[\mathcal{H}(t,s,u^*,w^*)]}{\partial u} = 0. \tag{A73}$$

Hence, the following inequality is satisfied:

$$J[u]-J[u^*] \geq 0 \tag{A74}$$

Therefore, Equation (20) is the sufficient condition of the optimal control function of ML-POSC if $\bar{\mathcal{H}}(t,p,u,w)$ is convex with respect to p and u. .

Appendix C.4. Derivation of Result in Section 3.5

In this subsection, we show that Equation (29) is the sufficient condition of the optimal control function of COSC without assuming the convexity of the expected Hamiltonian. We define the arbitrary control function ${}^\forall u : [0,T]\times\mathbb{R}^{d_s} \to \mathbb{R}^{d_u}$. From Equation (13), $J[u]-J[u^*]$ is given by the following equation:

$$J[u]-J[u^*] = \int_0^T \left(\mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u,w^*)] - \mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u^*,w^*)]\right)dt. \tag{A75}$$

From (29), the following inequality is satisfied:

$$J[u]-J[u^*] \geq \int_0^T \left(\mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u^*,w^*)] - \mathbb{E}_{p(t,s)}[\mathcal{H}(t,s,u^*,w^*)]\right)dt = 0. \tag{A76}$$

Therefore, Equation (29) is the sufficient condition of the optimal control function of COSC.

Appendix C.5. Derivation of Result in Section 4.2 by the Similar Way as Pontyragin's Minimum Principle

In this subsection, we derive Equation (31) from Equation (30) by the similar way as Pontyragin's minimum principle. From Equation (13), the following equality is satisfied:

$$\begin{aligned} &J[u_{0:t-dt}, u_t, u_{t+dt:T-dt}] - J[u_{0:t-dt}, u_t^*, u_{t+dt:T-dt}] \\ &= \left(\mathbb{E}_{p_t(s)}[\mathcal{H}(t,s,u_t,w_{t+dt})] - \mathbb{E}_{p_t(s)}[\mathcal{H}(t,s,u_t^*,w_{t+dt})]\right)dt \\ &= \mathbb{E}_{p_t(z)}\left[\mathbb{E}_{p_t(x|z)}[\mathcal{H}(t,s,u_t,w_{t+dt})] - \mathbb{E}_{p_t(x|z)}[\mathcal{H}(t,s,u_t^*,w_{t+dt})]\right]dt. \end{aligned} \tag{A77}$$

Therefore, Equation (31) is equivalent with Equation (30).

Appendix C.6. Derivation of Result in Section 4.2 by the Time Discretized Method

In this subsection, we derive Equation (31) from Equation (30) by the time discretized method. Equation (30) can be calculated as follows:

$$\begin{aligned}
u_t^* &= \arg\min_{u_t} J[u_{0:T-dt}] \\
&= \arg\min_{u_t} \mathbb{E}_{p(s_{0:T};u_{0:T-dt})}\left[\int_0^T f(\tau, s_\tau, u_\tau)d\tau + g(s_T)\right] \\
&= \arg\min_{u_t} \mathbb{E}_{p(s_{t:T};u_{0:T-dt})}\left[\int_t^T f(\tau, s_\tau, u_\tau)d\tau + g(s_T)\right] \\
&= \arg\min_{u_t} \mathbb{E}_{p(s_{t:T};u_{0:T-dt})}\left[f(t, s_t, u_t)dt + \int_{t+dt}^T f(\tau, s_\tau, u_\tau)d\tau + g(s_T)\right] \\
&= \arg\min_{u_t} \mathbb{E}_{p_t(s_t)}\left[f(t, s_t, u_t)dt + \mathbb{E}_{p(s_{t+dt:T}|s_t;u_{t:T-dt})}\left[\int_{t+dt}^T f(\tau, s_\tau, u_\tau)d\tau + g(s_T)\right]\right] \\
&= \arg\min_{u_t} \mathbb{E}_{p_t(s_t)}\left[f(t, s_t, u_t)dt + \mathbb{E}_{p(s_{t+dt}|s_t;u_t)}[w_{t+dt}(s_{t+dt})]\right],
\end{aligned} \tag{A78}$$

where $p_t(s)$ is the solution of the FP Equation (33) driven by $u_{0:t-dt}$, and $w_{t+dt}(s)$ is defined as follows:

$$w_{t+dt}(s) := \mathbb{E}_{p(s_{t+2dt:T}|s_{t+dt}=s;u_{t+dt:T-dt})}\left[\int_{t+dt}^T f(\tau, s_\tau, u_\tau)d\tau + g(s_T)\right]. \tag{A79}$$

From Ito's lemma,

$$\begin{aligned}
u_t^* &= \arg\min_{u_t} \mathbb{E}_{p_t(s_t)}[f(t, s_t, u_t)dt + w_{t+dt}(s_t) + \mathcal{L}_{u_t} w_{t+dt}(s_t)dt] \\
&= \arg\min_{u_t} \mathbb{E}_{p_t(s_t)}[f(t, s_t, u_t)dt + \mathcal{L}_{u_t} w_{t+dt}(s_t)dt] \\
&= \arg\min_{u_t} \mathbb{E}_{p_t(s)}[\mathcal{H}(t, s, u_t, w_{t+dt})].
\end{aligned} \tag{A80}$$

Because control u_t is a function of memory z in ML-POSC, the minimization by u_t can be exchanged with the expectation by $p_t(z)$ as follows:

$$u_t^*(z) = \arg\min_{u_t} \mathbb{E}_{p_t(x|z)}[\mathcal{H}(t, s, u_t, w_{t+dt})]. \tag{A81}$$

Therefore, Equation (31) is derived from Equation (30). Finally, we prove that $w_t(s)$ is the solution of the HJB Equation (32) driven by $u_{t+dt:T-dt}$. $w_t(s)$ can be calculated as follows:

$$\begin{aligned}
w_t(s) &= \mathbb{E}_{p(s_{t+dt:T}|s_t=s;u_{t:T-dt})}\left[\int_t^T f(\tau, s_\tau, u_\tau)d\tau + g(s_T)\right] \\
&= f(t, s, u_t)dt + \mathbb{E}_{p(s_{t+dt}|s_t=s;u_t)}[w_{t+dt}(s_{t+dt})] \\
&= f(t, s, u_t)dt + w_{t+dt}(s) + \mathcal{L}_{u_t} w_{t+dt}(s)dt \\
&= w_{t+dt}(s) + \mathcal{H}(t, s, u_t, w_{t+dt})dt,
\end{aligned} \tag{A82}$$

where $w_T(s) = g(s)$. Therefore, $w_t(s)$ defined by Equation (A79) is the solution of the HJB Equation (32) driven by $u_{t+dt:T-dt}$.

Appendix C.7. Derivation of Result in Section 4.3

In this subsection, we mainly derive the inequality of the forward step (35). The inequality of the backward step (34) can be derived in a similar way. In the forward step, $u_{0:t-dt}^{k+1}$ and $u_{t+dt:T-dt}^{k}$ are given, and u_t^{k+1} is defined by

$$u_t^{k+1}(z) := \arg\min_{u_t} \mathbb{E}_{p_t^{k+1}(x|z)}\left[\mathcal{H}\left(t, s, u_t, w_{t+dt}^k\right)\right]. \tag{A83}$$

From the equivalence of Equations (30) and (31), the following equation is satisfied:

$$u_t^{k+1} = \arg\min_{u_t} J[u_{0:t-dt}^{k+1}, u_t, u_{t+dt:T-dt}^{k}]. \tag{A84}$$

Therefore, the inequality of the forward step (35) is satisfied.

Appendix C.8. Derivation of Result in Section 4.4

In this subsection, we show that FBSM for ML-POSC converges to Pontryagin's minimum principle (20). More specifically, we prove that if $J[u_{0:T-dt}^{k+1}] = J[u_{0:T-dt}^{k}]$ holds, $u_{0:T-dt}^{k+1}$ satisfies Pontryagin's minimum principle (20). We mainly consider the forward step. We can make a similar discussion in the backward step. If $J[u_{0:T-dt}^{k+1}] = J[u_{0:T-dt}^{k}]$ holds, then $J[u_{0:t}^{k+1}, u_{t+dt:T-dt}^{k}] = J[u_{0:t-dt}^{k+1}, u_{t:T-dt}^{k}]$ holds from Equation (35). Because $J[u_0^{k+1}, u_{dt:T-dt}^{k}] = J[u_{0:T-dt}^{k}]$ holds, $u_0^{k+1} = u_0^k$ holds. Then, because $J[u_0^k, u_{dt}^{k+1}, u_{2dt:T-dt}^{k}] = J[u_{0:T-dt}^{k}]$ holds, $u_{dt}^{k+1} = u_{dt}^k$ holds. Iterating this procedure from $t = 0$ to $t = T - dt$, $u_{0:T-dt}^{k+1} = u_{0:T-dt}^{k}$ holds. Therefore, because the HJB equation and the FP equation depend on the same control function $u_{0:T-dt}^{k+1} = u_{0:T-dt}^{k}$, $u_{0:T-dt}^{k+1}$ satisfies Pontryagin's minimum principle (20).

Appendix C.9. Derivation of Result in Section 5.3

In this subsection, we show that FBSM is reduced from Algorithm 1 to Algorithm 2 in the LQG problem of ML-POSC.

We first consider the initial step. We assume that the control function is initialized by

$$u^0(t,z) = -R^{-1}B^\top\Big(\Pi^0 K(\Lambda^0)(s-\mu) + \Psi\mu\Big), \tag{A85}$$

where Π^0 is arbitrary and Λ^0 is the solution of $\dot{\Lambda}^0 = \mathcal{F}(\Lambda^0, \Pi^0)$ given $\Lambda^0(0) = \Lambda_0$. When the control function is initialized by (A85), the solution of the FP equation is given by the Gaussian distribution $p_t^0(s) := \mathcal{N}(s|\mu, \Lambda^0)$, where μ is the solution of (42) and Λ^0 is the solution of $\dot{\Lambda}^0 = \mathcal{F}(\Lambda^0, \Pi^0)$ given $\Lambda^0(0) = \Lambda_0$.

We then consider the backward step. When the solution of the FP equation is given by the Gaussian distribution $p_t^k(s) := \mathcal{N}(s|\mu, \Lambda^k)$, the solution of the HJB equation is given by the quadratic function $w_t^{k+1}(s) = s^\top \Pi^{k+1} s + (\alpha^{k+1})^\top s + \beta^{k+1}$, where Π^{k+1}, α^{k+1}, and β^{k+1} are the solutions of the following ODEs:

$$-\dot{\Pi}^{k+1} = \mathcal{G}(\Lambda^k, \Pi^{k+1}), \tag{A86}$$

$$\begin{aligned} -\dot{\alpha}^{k+1} &= (A - BR^{-1}B^\top \Pi^{k+1})^\top \alpha^{k+1} \\ &\quad - 2(I - K(\Lambda^k))^\top \Pi^{k+1} BR^{-1}B^\top \Pi^{k+1}(I - K(\Lambda^k))\mu, \end{aligned} \tag{A87}$$

$$\begin{aligned} -\dot{\beta}^{k+1} &= \mathrm{tr}\Big(\Pi^{k+1}\sigma\sigma^\top\Big) - \frac{1}{4}(\alpha^{k+1})^\top BR^{-1}B^\top \alpha^{k+1} \\ &\quad + \mu^\top (I - K(\Lambda^k))^\top \Pi^{k+1} BR^{-1}B^\top \Pi^{k+1}(I - K(\Lambda^k))\mu, \end{aligned} \tag{A88}$$

where $\Pi^{k+1}(T) = P$, $\alpha^{k+1}(T) = 0$, and $\beta^{k+1}(T) = 0$.

We finally consider the forward step. When the solution of the HJB equation is given by the quadratic function $w_t^k(s) = s^\top \Pi^k s + (\alpha^k)^\top s + \beta^k$, the solution of the FP equation is given by the Gaussian distribution $p_t^{k+1}(s) := \mathcal{N}(s|\mu, \Lambda^{k+1})$, where μ is the solution of (42) and Λ^{k+1} is the solution of $\dot{\Lambda}^{k+1} = \mathcal{F}(\Lambda^{k+1}, \Pi^k)$ given $\Lambda^{k+1}(0) = \Lambda_0$. Therefore, FBSM is reduced from Algorithm 1 to Algorithm 2 in the LQG problem of ML-POSC. The details of these calculations are almost the same with [14].

References

1. Fox, R.; Tishby, N. Minimum-information LQG control Part II: Retentive controllers. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 5603–5609. [CrossRef]
2. Fox, R.; Tishby, N. Minimum-information LQG control part I: Memoryless controllers. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 5610–5616. [CrossRef]
3. Li, W.; Todorov, E. An Iterative Optimal Control and Estimation Design for Nonlinear Stochastic System. In Proceedings of the 45th IEEE Conference on Decision and Control, San Diego, CA, USA, 13–15 December 2006; pp. 3242–3247. [CrossRef]
4. Li, W.; Todorov, E. Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. *Int. J. Control.* **2007**, *80*, 1439–1453. [CrossRef]
5. Nakamura, K.; Kobayashi, T.J. Connection between the Bacterial Chemotactic Network and Optimal Filtering. *Phys. Rev. Lett.* **2021**, *126*, 128102. [CrossRef] [PubMed]
6. Nakamura, K.; Kobayashi, T.J. Optimal sensing and control of run-and-tumble chemotaxis. *Phys. Rev. Res.* **2022**, *4*, 013120. [CrossRef]
7. Pezzotta, A.; Adorisio, M.; Celani, A. Chemotaxis emerges as the optimal solution to cooperative search games. *Phys. Rev. E* **2018**, *98*, 042401. [CrossRef]
8. Borra, F.; Cencini, M.; Celani, A. Optimal collision avoidance in swarms of active Brownian particles. *J. Stat. Mech. Theory Exp.* **2021**, *2021*, 083401. [CrossRef]
9. Davis, M.H.A.; Varaiya, P. Dynamic Programming Conditions for Partially Observable Stochastic Systems. *SIAM J. Control.* **1973**, *11*, 226–261. [CrossRef]
10. Bensoussan, A. *Stochastic Control of Partially Observable Systems*; Cambridge University Press: Cambridge, UK, 1992. [CrossRef]
11. Fabbri, G.; Gozzi, F.; Święch, A. Stochastic Optimal Control in Infinite Dimension. In *Probability Theory and Stochastic Modelling*; Springer International Publishing: Cham, Switzerland, 2017; Volume 82. [CrossRef]
12. Wang, G.; Wu, Z.; Xiong, J. *An Introduction to Optimal Control of FBSDE with Incomplete Information*; Springer Briefs in Mathematics; Springer International Publishing: Cham, Switzerland, 2018. [CrossRef]
13. Bensoussan, A.; Yam, S.C.P. Mean field approach to stochastic control with partial information. *ESAIM Control. Optim. Calc. Var.* **2021**, *27*, 89. [CrossRef]
14. Tottori, T.; Kobayashi, T.J. Memory-Limited Partially Observable Stochastic Control and Its Mean-Field Control Approach. *Entropy* **2022**, *24*, 1599. [CrossRef]
15. Kushner, H. Optimal stochastic control. *IRE Trans. Autom. Control.* **1962**, *7*, 120–122. [CrossRef]
16. Yong, J.; Zhou, X.Y. *Stochastic Controls*; Springer: New York, NY, USA, 1999. [CrossRef]
17. Nisio, M. Stochastic Control Theory. In *Probability Theory and Stochastic Modelling*; Springer: Tokyo, Japan, 2015; Volume 72. [CrossRef]
18. Bensoussan, A. Estimation and Control of Dynamical Systems. In *Interdisciplinary Applied Mathematics*; Springer International Publishing: Cham, Switzerland, 2018; Volume 48. [CrossRef]
19. Kushner, H.J.; Dupuis, P.G. *Numerical Methods for Stochastic Control Problems in Continuous Time*; Springer: New York, NY, USA, 1992. [CrossRef]
20. Fleming, W.H.; Soner, H.M. *Controlled Markov Processes and Viscosity Solutions*, 2nd ed.; Number 25 in Applications of Mathematics, Springer: New York, NY, USA, 2006. [CrossRef]
21. Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; Wiley-Interscience: Hoboken, NJ, USA, 2014.
22. Pontryagin, L.S. *Mathematical Theory of Optimal Processes*; CRC Press: Boca Raton, FL, USA, 1987.
23. Vinter, R. *Optimal Control*; Birkhäuser Boston: Boston, MA, USA, 2010. [CrossRef]
24. Lewis, F.L.; Vrabie, D.; Syrmos, V.L. *Optimal Control*; John Wiley & Sons: New York, NY, USA, 2012.
25. Aschepkov, L.T.; Dolgy, D.V.; Kim, T.; Agarwal, R.P. *Optimal Control*; Springer International Publishing: Cham, Switzerland, 2016. [CrossRef]
26. Bensoussan, A.; Frehse, J.; Yam, P. *Mean Field Games and Mean Field Type Control Theory*; Springer Briefs in Mathematics; Springer: New York, NY, USA, 2013. [CrossRef]
27. Carmona, R.; Delarue, F. *Probabilistic Theory of Mean Field Games with Applications I*; Number Volume 83 in Probability Theory and Stochastic Modelling; Springer Nature: Cham, Switzerland, 2018. [CrossRef]
28. Carmona, R.; Delarue, F. *Probabilistic Theory of Mean Field Games with Applications II*; Volume 84, Probability Theory and Stochastic Modelling; Springer International Publishing: Cham, Switzerland, 2018. [CrossRef]
29. Carmona, R.; Delarue, F. The Master Equation for Large Population Equilibriums. In *Stochastic Analysis and Applications 2014*; Crisan, D., Hambly, B., Zariphopoulou, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; Volume 100, pp. 77–128. [CrossRef]
30. Bensoussan, A.; Frehse, J.; Yam, S.C.P. The Master equation in mean field theory. *J. Math. Pures Appl.* **2015**, *103*, 1441–1474. [CrossRef]
31. Bensoussan, A.; Frehse, J.; Yam, S.C.P. On the interpretation of the Master Equation. *Stoch. Process. Their Appl.* **2017**, *127*, 2093–2137. [CrossRef]
32. Krylov, I.; Chernous'ko, F. On a method of successive approximations for the solution of problems of optimal control. *USSR Comput. Math. Math. Phys.* **1963**, *2*, 1371–1382. [CrossRef]

33. Mitter, S.K. Successive approximation methods for the solution of optimal control problems. *Automatica* **1966**, *3*, 135–149. [CrossRef]
34. Chernousko, F.L.; Lyubushin, A.A. Method of successive approximations for solution of optimal control problems. *Optim. Control. Appl. Methods* **1982**, *3*, 101–114. [CrossRef]
35. Lenhart, S.; Workman, J.T. *Optimal Control Applied to Biological Models*; Chapman and Hall/CRC: New York, NY, USA, 2007. [CrossRef]
36. Sharp, J.A.; Burrage, K.; Simpson, M.J. Implementation and acceleration of optimal control for systems biology. *J. R. Soc. Interface* **2021**, *18*, 20210241. [CrossRef]
37. Hackbusch, W. A numerical method for solving parabolic equations with opposite orientations. *Computing* **1978**, *20*, 229–240. [CrossRef]
38. McAsey, M.; Mou, L.; Han, W. Convergence of the forward-backward sweep method in optimal control. *Comput. Optim. Appl.* **2012**, *53*, 207–226. [CrossRef]
39. Carlini, E.; Silva, F.J. Semi-Lagrangian schemes for mean field game models. In Proceedings of the 52nd IEEE Conference on Decision and Control, Firenze, Italy, 10–13 December 2013; pp. 3115–3120. ISSN: 0191-2216. [CrossRef]
40. Carlini, E.; Silva, F.J. A Fully Discrete Semi-Lagrangian Scheme for a First Order Mean Field Game Problem. *SIAM J. Numer. Anal.* **2014**, *52*, 45–67. [CrossRef]
41. Carlini, E.; Silva, F.J. A semi-Lagrangian scheme for a degenerate second order mean field game system. *Discret. Contin. Dyn. Syst.* **2015**, *35*, 4269. [CrossRef]
42. Lauriere, M. Numerical Methods for Mean Field Games and Mean Field Type Control. *arXiv* **2021**, arXiv:2106.06231.
43. Wonham, W.M. On the Separation Theorem of Stochastic Control. *SIAM J. Control.* **1968**, *6*, 312–326. [CrossRef]
44. Li, Q.; Chen, L.; Tai, C.; E, W. Maximum Principle Based Algorithms for Deep Learning. *J. Mach. Learn. Res.* **2018**, *18*, 1–29.
45. Liu, X.; Frank, J. Symplectic Runge–Kutta discretization of a regularized forward–backward sweep iteration for optimal control problems. *J. Comput. Appl. Math.* **2021**, *383*, 113133. [CrossRef]
46. Bellman, R. *Dynamic Programming*; Princeton University Press: Princeton, NJ, USA, 1957.
47. Howard, R.A. *Dynamic Programming and Markov Processes*; John Wiley: Oxford, UK, 1960.
48. Kappen, H.J. Linear Theory for Control of Nonlinear Stochastic Systems. *Phys. Rev. Lett.* **2005**, *95*, 200201. [CrossRef]
49. Kappen, H.J. Path integrals and symmetry breaking for optimal control theory. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P11011. [CrossRef]
50. Satoh, S.; Kappen, H.J.; Saeki, M. An Iterative Method for Nonlinear Stochastic Optimal Control Based on Path Integrals. *IEEE Trans. Autom. Control.* **2017**, *62*, 262–276. [CrossRef]
51. Cacace, S.; Camilli, F.; Goffi, A. A policy iteration method for Mean Field Games. *arXiv* **2021**, arXiv:2007.04818.
52. Laurière, M.; Song, J.; Tang, Q. Policy iteration method for time-dependent Mean Field Games systems with non-separable Hamiltonians. *arXiv* **2021**, arXiv:2110.02552.
53. Camilli, F.; Tang, Q. Rates of convergence for the policy iteration method for Mean Field Games systems. *arXiv* **2022**, arXiv:2108.00755.
54. Ruthotto, L.; Osher, S.J.; Li, W.; Nurbekyan, L.; Fung, S.W. A machine learning framework for solving high-dimensional mean field game and mean field control problems. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9183–9193. [CrossRef]
55. Lin, A.T.; Fung, S.W.; Li, W.; Nurbekyan, L.; Osher, S.J. Alternating the population and control neural networks to solve high-dimensional stochastic mean-field games. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2024713118. [CrossRef]
56. Laurière, M.; Pironneau, O. Dynamic programming for mean-field type control. *C. R. Math.* **2014**, *352*, 707–713. [CrossRef]
57. Laurière, M.; Pironneau, O. Dynamic programming for mean-field type control. *J. Optim. Theory Appl.* **2016**, *169*, 902–924. [CrossRef]
58. Pham, H.; Wei, X. Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM Control. Optim. Calc. Var.* **2018**, *24*, 437–461. [CrossRef]