An iterative method for nonlinear stochastic optimal control based on path integrals

Satoshi Satoh, *Member, IEEE* , Hilbert J. Kappen *Senior Member, IEEE* , and Masami Saeki, *Member, IEEE*

***Abstract*—This paper proposes a new iterative solution method for nonlinear stochastic optimal control problems based on path integral analysis. First, we provide an iteration law for solving a stochastic Hamilton-Jacobi-Bellman (SHJB) equation associated to this problem, which is a nonlinear partial differential equation (PDE) of second order. Each iteration procedure of the proposed method is represented by a Cauchy problem for a linear parabolic PDE, and its explicit solution is given by the Feynman-Kac formula. Second, we derive a suboptimal feedback controller at each iteration by using the path integral analysis. Third, the convergence property of the proposed method is investigated. Here, some conditions are provided so that the sequence of solutions for the proposed iteration converges, and the SHJB equation is satisfied. Finally, numerical simulations demonstrate the effectiveness of the proposed method.**

***Index Terms*—Stochastic optimal control, Stochastic systems, Nonlinear control, Path integral.**

I. Introduction

Optimal control theory has been widely developed, since it brought significant achievements in aerospace engineering in 1950's. Nowadays, it plays fundamental roles not only in Engineering but also in many fields. Particularly, stochastic optimal control dealing with uncertainties is utilized for motion planning of robots interacting with dynamic environments, vibration suppression in engineering structures, reaction control in chemical plants, option pricing and portfolio allocation in economics, decision making in biophysics, and so on. The purpose of stochastic optimal control is to control the plant system with uncertainties so that the expectation of some performance index, called cost (reward) function is minimized (maximized).

When a linear stochastic system with additive white Gaussian noise and a quadratic cost function, i.e., the so-called LQG problem is considered, it is well known that the optimal feedback controller can be obtained by solving the Riccati equation [1], [2], [3], and this problem is currently easily solved. While the LQG theory has brought many practical benefits, nonlinear optimal control theory has not yet led to sufficient practical use. According to Bellman's principle of optimality, the optimal feedback controller for a nonlinear stochastic optimal control problem is given by solving a nonlinear partial differential equation (PDE) of second order, called the stochastic Hamilton-Jacobi-Bellman (SHJB) equation [4], [5], [6]. [1] Due to the existence of nonlinearity and second order partial derivatives, the SHJB equation is quite difficult to be solved. Although some useful solution methods for the deterministic Hamilton-Jacobi-Bellman (HJB) equation, which is a first order nonlinear PDE, have been recently proposed, they are not directly applied to the SHJB equation. For example, although the literature [8] reduces solving the HJB equation to finding a submanifold in the cotangent bundle associated with the target HJB equation, the theory of jet bundles is necessary for the SHJB equation. Besides, iterative method with successive approximation is an effective approach to solve nonlinear PDEs. The literature [9] uses a stochastic extension of the approach for the HJB equation in [10], where the SHJB equation is reduced to a sequence of linear PDEs called the generalized HJB. The generalized HJB is approximately solved by using the Galerkin method. The authors in [9] have proposed a meshless finite element method, and it enables efficient numerical computation. However, this method requires an assumption that for any initial condition, some PDE corresponding to the SHJB equation converges to a stationary solution. In addition, although the authors claim that their iteration algorithm converges, there is no convergence analysis. In the literature [11], convergence analysis of successive approximation for the SHJB equation is provided. It gives a condition for a solution to each iteration so that the successive approximation converges to the true solution to the SHJB equation. However, this paper does not provide any concrete algorithm to solve each iteration. Moreover, although each iteration is typically calculated approximately, e.g., with finite basis functions or finite samples, the approximation error of each solution is not considered in [11]. In another approach, an iterative local optimization method around a nominal trajectory based on quadratic approximation of the value function has been proposed in [12]. This method is a stochastic extension of the differential dynamic programming

S. Satoh is with the Division of Mechanical Systems and Applied Mechanics, Faculty of Engineering, Hiroshima University, 1-4-1, Kagamiyama, Higashi-Hiroshima 739-8527, Japan e-mail: s.satoh@ieee.org (see http://home.hiroshima-u.ac.jp/satoh/index.html).

H. J. Kappen is with the Department of Biophysics, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands e-mail: b.kappen@science.ru.nl

M. Saeki is with the Division of Mechanical Systems and Applied Mechanics, Faculty of Engineering, Hiroshima University e-mail: saeki@hiroshima-u.ac.jp

Manuscript received XXX; revised XXX.

[1] It is also well known in deterministic optimal control that Pontryagin's minimum principle (PMP) yields the optimal feedforward controller by solving a two-point boundary value problem. There are also some results in the stochastic case, e.g., [7], [4], [6]. However, since we consider the optimal feedback controller, this paper does not deal with the PMP framework.

approach [13]. When the nominal trajectory is close to the optimal one, this method is practically useful. However, since it is only valid around the optimal trajectory (note that it is generally unknown in advance), it cannot obtain the solution to the SHJB equation for all states and times. Moreover, although the convergence analysis for the deterministic differential dynamic programming has been investigated, e.g, in [14], there is no rigorous convergence analysis in [12] for the stochastic case. In this way, there are still many challenges to satisfactorily solve the SHJB equation, and it has been a serious obstacle to practical success of nonlinear stochastic optimal control.

We aim to solve this problem by taking a different approach from the conventional methods. The proposed method is motivated by path integral optimal control method proposed by the one of the authors in [15], [16]. The key points of this method are summarized as follows. First, this method imposes a particular assumption on the plant system and a cost function to be minimized. Second, an exponential transformation is applied to the SHJB equation. Under the assumption, nonlinear terms in the transformed PDE successfully cancel out, and thus a linear PDE is newly obtained. Since, moreover, the resultant linear PDE has the same structure as the Kolmogorov backward equation, the explicit solution is given by the Feynman-Kac formula [17], [6] as the expectation of a functional along a sample path. By applying the path integral analysis to this solution, the authors in [15], [16] reveal an interesting insight that the optimal controller coincides with the expectation of the noise under some weighted probability of the path cost. This result leads to a practical benefit that the optimal controller is effectively calculated by using several sampling methods in statistical physics. As the issue of the method, however, the particular assumption necessary for the conventional path integral methods [15], [16], [18] restricts the applicable class of the plant systems and cost functions.

The main contribution of this paper is to propose a new iterative solution method applicable to a wider class of nonlinear stochastic optimal control problems without the particular assumption. The proposed method is named iterative stochastic optimal control based on path integrals, shortly, ISOC-PI. First, we provide an iteration law for ISOC-PI, which solves the SHJB equation. Each iteration procedure of ISOC-PI is represented by a Cauchy problem for a linear parabolic PDE. Although this PDE is not of the Kolmogorov backward equation, the explicit solution is also given by the Feynman-Kac formula [6]. Second, we derive a suboptimal feedback controller at each iteration by using the path integral analysis in [15], [16]. We show that the resultant controller forms the expectation of the noise under some probability of the path cost. Note that it is not a trivial result, since the solution has a different path integral representation from the conventional methods. The outline of each iteration is that we generate sample paths and calculate corresponding path costs in parallel, and then calculate a weighted average of the noise with the path cost among those sample paths, which gives a suboptimal feedback controller. Third, the convergence property of ISOC-PI is also investigated. We provide convergence conditions such that the sequence of solutions for the proposed iteration converges, and the SHJB equation is satisfied. In this analysis, the influence of error caused by approximately solving each iteration is adequately considered. Consequently, ISOC-PI can iteratively solve the SHJB equation, and can provide the optimal feedback controller for a nonlinear stochastic optimal control problem. ISOC-PI has the following advantages compared to the conventional path integral methods [15], [16], [18], and other existing results, e.g., [9], [11], [12]:

- ISOC-PI does not require the particular assumption necessary for the conventional path integral methods. Thus, it is applicable to a wider class of nonlinear stochastic optimal control problems;
- ISOC-PI provides not only the solution to the SHJB equation, i.e., the value function, but also the corresponding optimal feedback controller by using the path integral analysis. The numerical differentiation of the value function, which is undesirable particularly in dealing with stochastic systems, is not required;
- the convergence property of ISOC-PI is rigorously analyzed; and
- ISOC-PI is easy to be numerically implemented, and is well adapted to parallel computing.

II. Preliminaries

A. Nonlinear stochastic optimal control

We consider a stochastic optimal control problem on the time interval $t \in [0, T]$ with any constant $T > 0$. Consider a class of nonlinear stochastic systems described by the following Itô stochastic differential equation:

$$\mathrm{d}x = f(x,t)\,\mathrm{d}t + g(x,t)u\,\mathrm{d}t + h(x,t)\,\mathrm{d}w, \quad x(0) = x_0, \tag{1}$$

where the state is denoted by $x(t) \in \mathbb{R}^n$, and the control input is denoted by $u(t) \in U \subset \mathbb{R}^m$, where U is specified as a compact set of admissible controls. $w(t) \in \mathbb{R}^r$ denotes a Wiener process defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, such that

$$E^{\mathcal{P}}\{\mathrm{d}w(t)\,\mathrm{d}w(t)^\top\} = Q\,\mathrm{d}t\ , \quad \forall t \in [0, T]. \tag{2}$$

Here, $E^{\mathcal{P}}\{\cdot\}$ denotes the expectation with respect to the probability measure $\mathcal{P}$, and the matrix $Q \in \mathbb{R}^{r\times r}$ represents the covariance matrix. $\mathcal{F}$ is a sigma algebra of the observable random events and a filtration $\{\mathcal{F}_t\}$ represents an increasing family of σ-algebras with $\mathcal{F}_t \subset \mathcal{F},\ \forall t \in [0, T]$. We suppose that $\{\mathcal{F}_t\}$ is right-continuous and complete.

In this paper, the control input $u(\cdot)$ is supposed to be a progressively measurable, and square-integrable process such that

$$E^{\mathcal{P}}\left\{\int_0^T u(t)^\top u(t)\,\mathrm{d}t\right\} < \infty$$

holds. The functions $f : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$, $g : \mathbb{R}^n \times [0, T] \to \mathbb{R}^{n\times m}$ and $h : \mathbb{R}^n \times [0, T] \to \mathbb{R}^{n\times r}$ are supposed to be sufficiently differentiable. This paper assumes the existence and uniqueness of a strong solution to the system (1) on $[0, T]$. One sufficient condition for this assumption is that f, g and h satisfy the local Lipschitz condition and linear growth condition [19]. Besides, the literature [20] provides the pair

of the local Lipschitz condition and monotone condition as a weaker sufficient condition.

We define the infinitesimal operator [21], [17], which will be utilized for the convergence analysis of the proposed method in Section III-B. For the system (1) with a control input u and a class $C^{2,1}$ function $F(x,t)$, the infinitesimal operator $\mathcal{L}_u(F)$ is defined as

$$\mathcal{L}_u(F) := \frac{\partial F}{\partial t} + \frac{\partial F}{\partial x}(f+gu) + \frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 F}{\partial x^2} hQh^\top\right\}. \quad (3)$$

In addition, $\mathcal{L}_u(F)(x,t)$ denotes the value of the right hand side of Eq. (3) at (x,t). In this paper, for a vector a, a_j denotes its j th element, and for a matrix A, $[A]_{j,k}$, $[A]_{j,:}$ and $[A]_{:,k}$ denote its (j,k) th element, j th row and k th column, respectively. We define $\partial F(x,t)/\partial x$ as an n-dimensional row vector such that its i th column is given by $\partial F(x,t)/\partial x_i$. We also define $\partial^2 F(x,t)/\partial x^2$ as an $n \times n$ matrix such that its (i,j) th element is given by $\partial^2 F(x,t)/\partial x_j \partial x_i$. Then, Itô's formula [22] gives the expectation of the time variation of the function $F(x,t)$ along a sample path $x(t)$ as

$$E^{\mathcal{P}}\{F(x(t),t)\} = F(x_0,0) + E^{\mathcal{P}}\left\{\int_0^t \mathcal{L}_u(F)(x(s),s)\,\mathrm{d}s\right\}. \quad (4)$$

Next, consider the following functional to define the cost function to be minimized:

$$\Gamma(x_t, u_{t:T}, t) = E^{\mathcal{P}}\left\{\phi(x(T)) + \int_t^T V(x,\tau) + \frac{1}{2}u(\tau)^\top R(x,\tau)u(\tau)\,\mathrm{d}\tau \;\middle|\; x(t)=x_t\right\}, \quad (5)$$

where sufficiently differentiable non-negative functions ϕ and V denote the terminal cost and the instantaneous cost, respectively. V is also supposed to be integrable. A symmetric positive definite matrix $R(x,t)$ represents the weight with respect to the control cost. The notation $u_{t:T}$ denotes an input signal on the time interval $[t,T]$. The objective of the paper is to find the optimal feedback controller $u = u^*(x,t)$ for the system (1) such that the cost function $\Gamma(x_0, u_{0:T}, 0)$ is minimized. Here, the value function is defined as

$$J(x,t) := \min_{u_{t:T}} \Gamma(x, u_{t:T}, t).$$

According to Bellman's principle of optimality, the stochastic Hamilton-Jacobi-Bellman (SHJB) equation [4], [5], [6], which $J(x,t)$ should satisfy, is given by

$$-\frac{\partial J}{\partial t} = \min_u \left(V + \frac{1}{2}u^\top R u + \frac{\partial J}{\partial x}(f+gu) + \frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J}{\partial x^2} hQh^\top\right\}\right), \quad J(x,T) = \phi(x). \quad (6)$$

Suppose that there exists a differentiable solution to the SHJB equation (6). Then, the optimal feedback controller $u^*(x,t)$ is given as the one that minimizes the right hand side of Eq. (6). Since the cost function is quadratic with respect to the control input, and $R(x,t)$ is positive definite, the optimal control is uniquely determined. By taking the gradient with respect to u of the inside of the parenthesis on the right hand side, and setting it to zero, the optimal controller is given by

$$u^*(x,t) = -R(x,t)^{-1} g(x,t)^\top \frac{\partial J(x,t)}{\partial x}^\top. \quad (7)$$

By substituting Eq. (7) into Eq. (6), the SHJB equation is reduced to the following partial differential equation (PDE):

$$\frac{\partial J}{\partial t} + V - \frac{1}{2}\frac{\partial J}{\partial x} g R^{-1} g^\top \frac{\partial J}{\partial x}^\top + \frac{\partial J}{\partial x} f + \frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J}{\partial x^2} hQh^\top\right\} = 0, \quad J(x,T) = \phi(x). \quad (8)$$

Consequently, the optimal control problem is reduced to solving the SHJB equation (8). Once the solution J is obtained, the optimal feedback controller is given by Eq. (7). However, it is generally quite difficult to solve the SHJB equation (8), because it is a second order nonlinear PDE with respect to J. One of the authors has proposed a new stochastic optimal control method based on statistical physics approach, called path integral optimal control, in [15]. This method provides an efficient solution for a class of nonlinear stochastic optimal control problems, and is referred to in subsection II-B.

B. Path integral optimal control

In what follows, we partition the system (1) into the subsystem which is directly driven by the noise, and the other one as:

$$\begin{pmatrix}\mathrm{d}x^u \\ \mathrm{d}x^c\end{pmatrix} = \begin{pmatrix}f^u(x,t) \\ f^c(x,t)\end{pmatrix}\mathrm{d}t + \begin{pmatrix}0_{(n-n_c)\times m} \\ g^c(x,t)\end{pmatrix} u\,\mathrm{d}t + \begin{pmatrix}0_{(n-n_c)\times r} \\ h^c(x,t)\end{pmatrix}\mathrm{d}w, \quad (9)$$

where $x := (x^{u\top}, x^{c\top})^\top \in \mathbb{R}^n$ and $x^c \in \mathbb{R}^{n_c}$ denotes the noise-driven state and $x^u \in \mathbb{R}^{n-n_c}$ denotes the other one, respectively, and $0_{i\times j}$ represents the $i \times j$ zero matrix. We suppose that the following assumption holds.

Assumption 1: For all $x \in \mathbb{R}^n$ and $t \in [0,T]$, the matrix $h^c(x,t)Qh^c(x,t)^\top$ is positive definite.

In addition, the conventional path integral optimal control [15], [16], [18] require the following particular assumption.

Assumption 2: For all $x \in \mathbb{R}^n$ and $t \in [0,T]$, there exists a positive constant $\lambda > 0$ such that the following equation holds:

$$\lambda g^c(x,t)R^{-1}(x,t)g^c(x,t)^\top = h^c(x,t)Qh^c(x,t)^\top \quad (10)$$

with the weighting matrix R in Eq. (5), and the covariance matrix Q in Eq. (2).

Remark 1: A possible interpretation of the condition (10) is that the larger the noise effect becomes, the smaller the control cost has to be. In addition, the condition also requires that the control and noise must act in the same dimensions.

By applying the following transformation

$$J(x,t) = -\lambda \log \psi(x,t) \quad (11)$$

to the SHJB equation (8), we have

$$\frac{\partial\psi}{\partial t}-\frac{V}{\lambda}\psi+\frac{\partial\psi}{\partial x}f+\frac{\lambda}{2\psi}\frac{\partial\psi}{\partial x}gR^{-1}g^\top\frac{\partial\psi}{\partial x}^\top-\frac{\lambda}{2\psi}\frac{\partial\psi}{\partial x}h\frac{Q}{\lambda}h^\top\frac{\partial\psi}{\partial x}^\top + \frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2\psi}{\partial x^2}hQh^\top\right\}=0\ ,\quad \psi(x,T)=\exp\left(-\frac{\phi(x)}{\lambda}\right). \tag{12}$$

Under Assumption 2, the nonlinear terms, i.e., the fourth and fifth terms on the left hand side of Eq. (12) cancel out. It leads to the second order linear PDE as

$$\frac{\partial\psi}{\partial t}-\frac{V}{\lambda}\psi+\frac{\partial\psi}{\partial x}f+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2\psi}{\partial x^2}hQh^\top\right\}=0,\qquad \psi(x,T)=\exp\left(-\frac{\phi(x)}{\lambda}\right). \tag{13}$$

Since the resultant linear PDE (13) has the same structure as the Kolmogorov backward equation, the path integral optimal control provides the explicit solution to the PDE (13) based on the Feynman-Kac formula ([6], Theorem 1.3.17), see also [17], [18], as

$$\psi(x,t)=E^{p(\xi_{t:T}|x,t)}\left\{\exp\left(-\frac{1}{\lambda}\left(\phi(\xi(T))+\int_t^T V(\xi,\tau)\,\mathrm{d}\tau\right)\right)\right\}, \tag{14}$$

where $p(\xi_{t:T}|x,t)$ represents the probability that a sample path $\xi_{t:T}$ is realized under the uncontrolled dynamics of the system (9) with $\xi(t)=x$ on $[t,T]$, that is,

$$\begin{pmatrix}\mathrm{d}\xi^u\\ \mathrm{d}\xi^c\end{pmatrix}=\begin{pmatrix}f^u(\xi,t)\\ f^c(\xi,t)\end{pmatrix}\mathrm{d}t+\begin{pmatrix}0_{(n-n_c)\times r}\\ h^c(\xi,t)\end{pmatrix}\mathrm{d}w\ ,\quad \xi(t)=x. \tag{15}$$

$E^{p(\xi_{t:T}|x,t)}\{\cdot\}$ represents the expectation with respect to the probability $p(\xi_{t:T}|x,t)$.

The probability $p(\xi_{t:T}|x,t)$ can be represented by the state transition probability as

$$p(\xi_{t:T}|x,t)=\lim_{\mathrm{d}t\to0}\prod_{s=t}^{T-\mathrm{d}t}p(\xi(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s),\quad \xi(t)=x. \tag{16}$$

Since the system is partitioned into the directly noise-driven subsystem and the other one, according to the same argument as in [18], the state transition probability can also be separated as

$$\begin{aligned}&p(\xi(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\\ &=p(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\times p(\xi^u(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\\ &\propto p(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s),\end{aligned} \tag{17}$$

where the last relation comes from the fact that $\xi^u(s+\mathrm{d}t)$ is deterministically realized under the subsystem $\mathrm{d}\xi^u = f^u(\xi,t)\,\mathrm{d}t$ with $\xi(s)$, and thus $p(\xi^u(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)$ is the Dirac delta function. Eqs. (15), (16) and (17) yield

$$\begin{aligned}&p(\xi_{t:T}|x,t)\propto p(\xi^c_{t:T}|x,t)\\ &=\lim_{\mathrm{d}t\to0}\prod_{s=t}^{T-\mathrm{d}t}p(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\ ,\quad \xi(t)=x,\end{aligned} \tag{18}$$

$$p(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)=\left(\frac{1}{2\pi}\right)^{\frac{n_c}{2}}\frac{1}{\sqrt{\det(\Xi^c(\xi(s),s))\,\mathrm{d}t}} \tag{19}$$

$$\times\exp\left(-\frac{1}{2\,\mathrm{d}t}(\xi^c(s+\mathrm{d}t)-\xi^c(s)-f^c(\xi,s)\,\mathrm{d}t)^\top\Xi^{c-1}\times(\xi^c(s+\mathrm{d}t)-\xi^c(s)-f^c(\xi,s)\,\mathrm{d}t)\right),$$

where

$$\Xi^c(\xi,t):=h^c(\xi,t)Qh^c(\xi,t)^\top. \tag{20}$$

Assumption 2 implies that $\Xi^c=\lambda g^cR^{-1}g^{c\top}$ holds.

From Eq. (14), we define a new weighted probability $q(\xi_{t:T}|x,t)$ with respect to the path cost on $[t,T]$ as

$$\begin{aligned}&q(\xi_{t:T}|x,t)\\ &:=\frac{p(\xi_{t:T}|x,t)}{\psi(x,t)}\exp\left(-\frac{1}{\lambda}\left(\phi(\xi(T))+\int_t^T V(\xi,\tau)\,\mathrm{d}\tau\right)\right).\end{aligned} \tag{21}$$

Then, although the details of calculations are omitted, one of the main contribution of the path integral optimal control method is to derive the fact by path integral analysis with Eqs. (14), (18), (19) and (21) that

$$\frac{\partial\log\psi(x,t)}{\partial x^c}^\top\mathrm{d}t=\Xi^c(x,t)^{-1}h^c(x,t)E^{q(\xi_{t:T}|x,t)}\{\mathrm{d}w(t)\}. \tag{22}$$

Consequently, the optimal feedback controller is obtained by Eqs. (7), (11) and (22) and Assumption 2 as

$$\begin{aligned}&g^c(x,t)u^*(x,t)\,\mathrm{d}t\\ &=\lambda g^c(x,t)R(x,t)^{-1}g^c(x,t)^\top\frac{\partial\log\psi(x,t)}{\partial x^c}^\top\mathrm{d}t\\ &=h^c(x,t)E^{q(\xi_{t:T}|x,t)}\{\mathrm{d}w(t)\}.\end{aligned} \tag{23}$$

In the literature [15], [16], some efficient computation methods for the optimal controller (23) are proposed based on Monte Carlo sampling, Langevin sampling, Importance sampling, Laplace approximation, and so on.

III. Main results

In this section, we propose a new iterative solution method for nonlinear stochastic optimal control problems based on the conventional path integral methods. As the motivation of the proposed method, the conventional path integral methods [15], [16], [18] have the issue that the particular assumption, namely Assumption 2, restricts the applicable class of the plant systems and cost functions. By equipping the successive approximation scheme to solve nonlinear PDEs, the proposed method does not require Assumption 2, nor the exponential transformation (11). The proposed iteration law enables us to solve the SHJB equation (8) iteratively by using a stochastic

representation of the solution to a Cauchy problem for a linear parabolic PDE. Moreover, by applying the path integral analysis to the resultant solution, the corresponding optimal feedback controller is explicitly derived without the numerical differentiation. We also investigate a convergence property of the proposed iteration procedure.

A. Iterative stochastic optimal control based on path integrals (ISOC-PI)

In what follows, we consider a wider class of stochastic optimal control problems with the cost function $\Gamma(x_0, u_{0:T}, 0)$ in (5) and the plant system (9) without Assumption 2. Now, we describe the iteration law for the proposed method named iterative stochastic optimal control based on path integrals, shortly, ISOC-PI. Suppose that a class $C^{2,1}$ function $J_{(0)}$ is appropriately given as an initial function, and that at the i th iteration, the functions $J_{(i-1)}$ and $\partial J_{(i-1)}/\partial x^c$ are already obtained. Then, the i th iteration procedure of ISOC-PI is given by solving the following Cauchy problem for a linear parabolic PDE with respect to the i th solution $J_{(i)}$:

$$\begin{aligned}&\frac{\partial J_{(i)}}{\partial t} + \frac{\partial J_{(i)}}{\partial x}\hat{f}_{(i)} + \frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J_{(i)}}{\partial x^2} hQh^\top\right\} + \hat{V}_{(i)} = 0,\\ &J_{(i)}(x,T) = \phi(x),\end{aligned} \tag{24}$$

where the functions $\hat{f}_{(i)} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $\hat{V}_{(i)} : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ are defined as

$$\begin{aligned}\hat{f}_{(i)} &:= f - gR^{-1}g^\top \frac{\partial J_{(i-1)}}{\partial x}^\top \\ &= \begin{pmatrix} f^u \\ f^c - g^c R^{-1} g^{c\top} \frac{\partial J_{(i-1)}}{\partial x^c}^\top \end{pmatrix} =: \begin{pmatrix} \hat{f}^u_{(i)} \\ \hat{f}^c_{(i)} \end{pmatrix},\end{aligned} \tag{25}$$

$$\begin{aligned}\hat{V}_{(i)} &:= V + \frac{1}{2}\frac{\partial J_{(i-1)}}{\partial x} g R^{-1} g^\top \frac{\partial J_{(i-1)}}{\partial x}^\top \\ &= V + \frac{1}{2}\frac{\partial J_{(i-1)}}{\partial x^c} g^c R^{-1} g^{c\top} \frac{\partial J_{(i-1)}}{\partial x^c}^\top.\end{aligned} \tag{26}$$

We will show some features of the iteration law of ISOC-PI (24). First, each iteration is not of a Kolmogorov backward equation as Eq. (13), but is a Cauchy problem that the Feynman-Kac formula [6] can also be applied to. Therefore, we can obtain a stochastic representation of its solution $J_{(i)}$. Second, the partial derivative $\partial J_{(i)}/\partial x^c$, which is necessary for the next iteration and also for deriving a suboptimal controller corresponding to $J_{(i)}$, can be obtained from the path integral analysis. The next theorem proves the claim:

Theorem 1: Consider the iteration of ISOC-PI in Eq. (24) with the plant system (9) and the cost function $\Gamma(x_0, u_{0:T}, 0)$ in (5). Suppose that the conditions for the existence and uniqueness of the solutions mentioned in Subsection II-A, and Assumption 1 hold. Moreover, an initial function $J_{(0)}$ is given such that it is a class $C^{2,1}$ and $\|\partial J_{(0)}(x,t)/\partial x^c\|^2 < \infty$ for all $x \in \mathbb{R}^n$ and $t \in [0,T]$.

Then, the solution to the i th iteration of ISOC-PI in Eq. (24) is given by

$$J_{(i)}(x,t) = E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\left\{\hat{S}_{(i)}(\xi_{t:T})\right\}, \tag{27}$$

$$\hat{S}_{(i)}(\xi_{t:T}) := \int_t^T \hat{V}_{(i)}(\xi(\tau),\tau)\,\mathrm{d}\tau + \phi(\xi(T)). \tag{28}$$

Here, $\hat{p}_{(i)}(\xi_{t:T}|x,t)$ represents the probability that a sample path $\xi_{t:T}$ is realized under the i th sample generating dynamics with $\xi(t) = x$ on $[t,T]$, which is given as the uncontrolled dynamics (15) by substituting f with $\hat{f}$ in Eq. (25), that is

$$\begin{pmatrix}\mathrm{d}\xi^u \\ \mathrm{d}\xi^c\end{pmatrix} = \begin{pmatrix}\hat{f}^u_{(i)}(\xi,t) \\ \hat{f}^c_{(i)}(\xi,t)\end{pmatrix}\mathrm{d}t + \begin{pmatrix}0_{(n-n_c)\times r} \\ h^c(\xi,t)\end{pmatrix}\mathrm{d}w\ ,\quad \xi(t) = x. \tag{29}$$

$\hat{S}_{(i)}(\xi_{t:T})$ in Eq. (28) represents the weighted total cost associated with the sample path $\xi_{t:T}$, where the instantaneous cost is $\hat{V}_{(i)}(x,t)$ in Eq. (26) instead of $V(x,t)$ in Eq. (5).

Moreover, the partial derivative of $J_{(i)}$ with respect to x^c is given by

$$\begin{aligned}&\frac{\partial J_{(i)}(x,t)}{\partial x^c}^\top \mathrm{d}t \\ &= \Xi^c(x,t)^{-1} h^c(x,t) E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\{\hat{S}_{(i)}(\xi_{t:T})\,\mathrm{d}w(t)\},\end{aligned} \tag{30}$$

where $\Xi^c(x,t)$ is defined in Eq. (20). Consequently, the suboptimal controller $u^*_{(i)}(x,t)$ corresponding to the i th solution $J_{(i)}(x,t)$ is given by

$$\begin{aligned}u^*_{(i)}(x,t)\,\mathrm{d}t = &-R(x,t)^{-1} g^c(x,t)^\top \Xi^c(x,t)^{-1} h^c(x,t) \\ &\times E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\{\hat{S}_{(i)}(\xi_{t:T})\,\mathrm{d}w(t)\}.\end{aligned} \tag{31}$$

Proof: See Appendix A. ∎

The resultant suboptimal controller in Eq. (31) forms the expectation of the noise vector weighted by the total cost. Note that there is no change of measure as is the case in the conventional methods in Eqs. (21) and (23). Also, note that Theorem 1 provides not only the i th solution $J_{(i)}(x,t)$, but also its partial derivative $\partial J_{(i)}(x,t)/\partial x^c$, and it can avoid the numerical differentiation of $J_{(i)}$ in calculating the suboptimal controller $u^*_{(i)}(x,t)$. Since the numerical differentiation is undesirable particularly in dealing with stochastic systems, this is a big advantage of ISOC-PI.

B. Convergence analysis of the ISOC-PI

This subsection investigates a convergence property of the ISOC-PI. It will be proved partly based on the argument in [11]. However, the differences between our analysis and the literature are as follows. First, the literature does not provide the concrete procedure to solve each iteration. Second, although each iteration is typically calculated approximately, e.g., with finite basis functions or finite samples, the literature does not take the approximation error of each iteration into account in the argument.

For notational convenience, we define the following notation:

$$\begin{aligned} H(\eta,\zeta) &:= \frac{\partial \eta}{\partial t}+\frac{\partial \eta}{\partial x}\left(f-gR^{-1}g^\top\frac{\partial \zeta}{\partial x}^\top\right) \\ &\quad+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2\eta}{\partial x^2}hQh^\top\right\}+V+\frac{1}{2}\frac{\partial\zeta}{\partial x}gR^{-1}g^\top\frac{\partial\zeta}{\partial x}^\top \\ &= \frac{\partial \eta}{\partial t}+\frac{\partial \eta}{\partial x^u}f^u+\frac{\partial \eta}{\partial x^c}\left(f^c-g^cR^{-1}g^{c\top}\frac{\partial \zeta}{\partial x^c}^\top\right) \\ &\quad+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2\eta}{\partial x^{c2}}h^cQh^{c\top}\right\}+V+\frac{1}{2}\frac{\partial\zeta}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial\zeta}{\partial x^c}^\top. \end{aligned} \tag{32}$$

In addition, $H(\eta,\zeta)(x,t)$ denotes the value of the right hand side of Eq. (32) at (x,t). Then, by using the notation (32), the iteration law for ISOC-PI (24) is rewritten as

$$\begin{aligned} &\text{The iteration law (24)} \\ &\iff \quad H(J_{(i)},J_{(i-1)})=0\ , \quad J_{(i)}(x,T)=\phi(x), \end{aligned} \tag{33}$$

Then, in order to explicitly consider the calculation error in each iteration possibly caused by approximations, we define the following error functions:

$$\epsilon_{(i)}(x,t) := H(J_{(i)},J_{(i-1)})(x,t) \tag{34}$$

$$d_{(i)}(x,t) := H(J_{(i)},J_{(i)})(x,t). \tag{35}$$

Here, $\epsilon_{(i)}$ represents the error of $J_{(i)}$ caused by the calculation of the i th iteration, while $d_{(i)}$ represents the difference between $J_{(i)}$ and the true solution to the SHJB equation. From the definition, $d_{(i)}\equiv 0$ implies that $J_{(i)}$ satisfies the SHJB equation. We also equip the notation of the L_1 norm on $\Omega\times[t,T]$ for an integrable stochastic process defined on $[t,T]$ as $\|\cdot\|_{L_1(\Omega\times[t,T])} := E^{\mathcal{P}}\{\int_t^T|\cdot|\,\mathrm{d}\tau\}$.

The following theorem shows that the solution to the original SHJB equation (8) can be iteratively obtained by the proposed ISOC-PI method.

Theorem 2: Consider the iteration of ISOC-PI in Eq. (24) with the plant system (9) and the cost function $\Gamma(x_0,u_{0:T},0)$ in (5). Suppose that all assumptions in Theorem 1 hold. Also, suppose that after any i th iteration, $J_{(i)}(x,t)$ and $\partial J_{(i)}(x,t)/\partial x^c$ are available at any $x\in\mathbb{R}^n$ and $t\in[0,T]$, and that at any i th iteration, the L_1 norm on $\Omega\times[t,T]$ of the error function $\epsilon_{(i)}$ defined in Eq. (34), that is, $\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}$ can be arbitrary small at any $t\in[0,T]$.

Then, for each $x\in\mathbb{R}^n$ and $t\in[0,T]$, if $\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}\to 0$ as $i\to\infty$, $J_{(i)}(x,t)$ converges pointwise, and the error $d_{(i)}(x,t)$ defined in Eq. (35) converges pointwise to zero almost surely. The SHJB equation (8) is satisfied almost everywhere.

Proof: Fix an arbitrary integer i. For notational simplicity, we define $\Delta J_{(i+1)} := J_{(i+1)}-J_{(i)}$. First, let us investigate the relation between the errors $d_{(i+1)}$ and $d_{(i)}$ as follows:

$$\begin{aligned} d_{(i+1)} &= H(J_{(i+1)},J_{(i+1)}) \\ &= \frac{\partial J_{(i+1)}}{\partial t}+\frac{\partial J_{(i+1)}}{\partial x}f-\frac{1}{2}\frac{\partial J_{(i+1)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i+1)}}{\partial x^c}^\top \\ &\quad+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J_{(i+1)}}{\partial x^{c2}}h^cQh^{c\top}\right\}+V \\ &= \frac{\partial J_{(i)}}{\partial t}+\frac{\partial(\Delta J_{(i+1)})}{\partial t}+\frac{\partial J_{(i)}}{\partial x}f+\frac{\partial(\Delta J_{(i+1)})}{\partial x}f-\frac{1}{2}\frac{\partial J_{(i)}}{\partial x^c}g^c \\ &\quad\times R^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top-\frac{1}{2}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}^\top \\ &\quad-\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J_{(i)}}{\partial x^{c2}}h^cQh^{c\top}\right\} \\ &\quad+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2(\Delta J_{(i+1)})}{\partial x^{c2}}h^cQh^{c\top}\right\}+V \\ &= d_{(i)}-\frac{1}{2}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}^\top \\ &\quad+\mathcal{L}_{u^*_{(i)}}(\Delta J_{(i+1)}), \end{aligned} \tag{36}$$

where in the last equality, $d_{(i)}=H(J_{(i)},J_{(i)})$ is used, and the infinitesimal operator $\mathcal{L}_{u^*_{(i)}}(\cdot)$ is defined in Eq. (3).

Second, from the definition of $\epsilon_{(i)}$ in Eq. (34), we have

$$\begin{aligned} &\epsilon_{(i+1)} = H(J_{(i+1)},J_{(i)}) = \frac{\partial J_{(i+1)}}{\partial t}+\frac{\partial J_{(i+1)}}{\partial x}f \\ &-\frac{\partial J_{(i+1)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J_{(i+1)}}{\partial x^{c2}}h^cQh^{c\top}\right\} \\ &+\frac{1}{2}\frac{\partial J_{(i)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top+V \\ &\iff \frac{\partial J_{(i+1)}}{\partial t}+\frac{\partial J_{(i+1)}}{\partial x}f+\frac{1}{2}\operatorname{tr}\left\{\frac{\partial^2 J_{(i+1)}}{\partial x^{c2}}h^cQh^{c\top}\right\}+V= \\ &\epsilon_{(i+1)}+\frac{\partial J_{(i+1)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top-\frac{1}{2}\frac{\partial J_{(i)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top. \end{aligned} \tag{37}$$

By substituting Eq. (37) into the first equality in Eq. (36), we have the relation between the errors $d_{(i+1)}$ and $\epsilon_{(i+1)}$ as

$$\begin{aligned} d_{(i+1)} &= \epsilon_{(i+1)}-\frac{1}{2}\frac{\partial J_{(i+1)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i+1)}}{\partial x^c}^\top \\ &\quad+\frac{\partial J_{(i+1)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top-\frac{1}{2}\frac{\partial J_{(i)}}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial J_{(i)}}{\partial x^c}^\top \\ &= \epsilon_{(i+1)}-\frac{1}{2}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}^\top. \end{aligned} \tag{38}$$

It follows from Eq. (36) that

$$\begin{aligned} \mathcal{L}_{u^*_{(i)}}(\Delta J_{(i+1)}) &= d_{(i+1)}-d_{(i)} \\ &\quad+\frac{1}{2}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial(\Delta J_{(i+1)})}{\partial x^c}^\top. \end{aligned} \tag{39}$$

By substituting the representation (38) twice for $i+1$ and for i into the first and second terms in Eq. (39), we have

$$\begin{aligned} &\mathcal{L}_{u^*_{(i)}}(\Delta J_{(i+1)}) \\ &=\frac{1}{2}\frac{\partial(\Delta J_{(i)})}{\partial x^c}g^cR^{-1}g^{c\top}\frac{\partial(\Delta J_{(i)})}{\partial x^c}^\top+\epsilon_{(i+1)}-\epsilon_{(i)}. \end{aligned} \tag{40}$$

Since the solution $J_{(i)}$ of the iteration law (24) is derived by the Feynman-Kac formula (see, Eq. (27)), $J_{(i)}$ with any integer i satisfies the boundary condition of the PDE (8), that is, $J_{(i)}(x,T)=\phi(x),\ \forall i$ holds. It implies that

$$\Delta J_{(i+1)}(x,T)=\phi(x)-\phi(x)=0\ \ ,\quad \forall x\in\mathbb{R}^n. \tag{41}$$

For each $x\in\mathbb{R}^n$ and $t\in[0,T]$, it follows from Eq. (4) that

$$\begin{aligned}&E^{\mathcal{P}}\{\Delta J_{(i+1)}(\xi(T),T)|\xi(t)=x\}=\Delta J_{(i+1)}(x,t)\\&+E^{\mathcal{P}}\left\{\int_0^T\mathcal{L}_{u^*_{(i)}}(\Delta J_{(i+1)})(\xi(\tau),\tau)\,\mathrm{d}\tau\middle|\xi(t)=x\right\}.\end{aligned} \tag{42}$$

By substituting Eqs. (40) and (41) into Eq. (42), we have

$$\begin{aligned}&\Delta J_{(i+1)}(x,t)\\&=-E^{\mathcal{P}}\left\{\int_t^T\mathcal{L}_{u^*_{(i)}}(\Delta(J_{(i+1)})(\xi(\tau),\tau)\,\mathrm{d}\tau\middle|\xi(t)=x\right\}\\&=-E^{\mathcal{P}}\left\{\int_t^T\frac{1}{2}\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}g^c(\xi(\tau),\tau)R^{-1}(\xi(\tau),\tau)\right.\\&\quad\times g^c(\xi(\tau),\tau)^\top\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}^\top+\epsilon_{(i+1)}(\xi(\tau),\tau)\\&\quad\left.-\epsilon_{(i)}(\xi(\tau),\tau)\,\mathrm{d}\tau\middle|\xi(t)=x\right\}\\&\le-\frac{1}{2}E^{\mathcal{P}}\left\{\int_t^T\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}g^c(\xi(\tau),\tau)R^{-1}(\xi(\tau),\tau)\right.\\&\quad\times g^c(\xi(\tau),\tau)^\top\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}^\top\,\mathrm{d}\tau\middle|\xi(t)=x\right\}\\&\quad+\|\epsilon_{(i+1)}\|_{L_1(\Omega\times[t,T])}+\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}.\end{aligned} \tag{43}$$

Now, suppose that the first expectation in the last inequality (43) is not zero for all i. Then, the positive definiteness of R implies that the integrand is non-negative, and thus the first term in the inequality (43) is negative for all i. Hence, there exists $\bar{\epsilon}_{(i)}(x,t)>0$ for each i such that if $\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}\le\bar{\epsilon}_{(i)}(x,t),\ \forall i$ are satisfied, the following condition holds:

$$\Delta J_{(i)}(x,t)<0,\ \forall i. \tag{44}$$

Note that the upper bounds $\bar{\epsilon}_{(i)}$'s might respectively depend on (x,t). Since V is non-negative, and the second term in Eq. (26) is so due to the positive definiteness of R, $\hat{V}_{(i)}$ is a non-negative function for each i. From this fact and the non-negativity of ϕ, Eqs. (27) and (28) yield

$$J_{(i)}(x,t)\ge 0\ ,\quad \forall x\in\mathbb{R}^n,\ t\in[0,T],\ i. \tag{45}$$

From Eqs. (44) and (45), $\{J_{(i)}(x,t)\}_{i=1}^{\infty}$ is a non-negative and monotonically decreasing sequence. Therefore, for each $x\in\mathbb{R}^n$ and $t\in[0,T]$, $J_{(i)}(x,t)$ converges pointwise if $\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}\to 0$ as $i\to\infty$.

Moreover, under this condition, from Eq. (43) we have

$$\begin{aligned}&\lim_{i\to\infty}E^{\mathcal{P}}\left\{\int_t^T\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}g^c(\xi(\tau),\tau)R^{-1}(\xi(\tau),\tau)\right.\\&\left.\times g^c(\xi(\tau),\tau)^\top\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}^\top\,\mathrm{d}\tau\middle|\xi(t)=x\right\}=0.\end{aligned} \tag{46}$$

Equation (46) implies

$$\begin{aligned}&\lim_{i\to\infty}\frac{\partial(\Delta J_{(i)}(x,t))}{\partial x^c}g^c(x,t)R^{-1}(x,t)g^c(x,t)^\top\\&\quad\times\frac{\partial(\Delta J_{(i)}(x,t))}{\partial x^c}^\top=0\end{aligned} \tag{47}$$

holds at (x,t) almost surely. From Eqs. (38) and (47), we can conclude that for each $x\in\mathbb{R}^n$ and $t\in[0,T]$, $d_{(i)}(x,t)$ converges pointwise to zero almost surely if $\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}\to 0$ as $i\to\infty$. Then, from the definition of the error function $d_{(i)}$ in Eq. (35), the SHJB equation (8) is satisfied almost everywhere.

On the contrary, if there exists some integer $i>1$ such that the first expectation in Eq. (43) becomes zero, that is,

$$\begin{aligned}&E^{\mathcal{P}}\left\{\int_t^T\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}g^c(\xi(\tau),\tau)R^{-1}(\xi(\tau),\tau)\right.\\&\left.\times g^c(\xi(\tau),\tau)^\top\frac{\partial(\Delta J_{(i)}(\xi(\tau),\tau))}{\partial x^c}^\top\,\mathrm{d}\tau\middle|\xi(t)=x\right\}=0,\quad\exists i.\end{aligned} \tag{48}$$

Then, from Eq. (43), we have

$$\lim_{\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])},\ \|\epsilon_{(i+1)}\|_{L_1(\Omega\times[t,T])}\to 0}\Delta J_{(i+1)}(x,t)=0. \tag{49}$$

Besides, Eq. (48) implies

$$\frac{\partial(\Delta J_{(i)}(x,t))}{\partial x^c}g^c(x,t)R^{-1}(x,t)g^c(x,t)^\top\frac{\partial(\Delta J_{(i)}(x,t))}{\partial x^c}^\top=0 \tag{50}$$

holds at (x,t) almost surely. From Eqs. (38) and (50), we can conclude that for each $x\in\mathbb{R}^n$ and $t\in[0,T]$, $d_{(i)}(x,t)$ converges pointwise to zero almost surely as $\|\epsilon_{(i)}\|_{L_1(\Omega\times[t,T])}\to 0$. Then, the definition of the error function $d_{(i)}$ in Eq. (35), the SHJB equation (8) is satisfied almost everywhere. This fact with Eq. (49) implies that $J_{(i)}(x,t)$ converges pointwise to the solution to the SHJB equation (8) almost surely. ■

IV. Discussions on computation

There are points to be noted on the actual computation of the proposed ISOC-PI method. First, the solution in Eq. (27) and controller in Eq. (31) of ISOC-PI are respectively given as expectations in Theorem 1. In Subsection IV-A, we provide a computation method for them based on a statistical sampling.

Second, each iteration of ISOC-PI generates one suboptimal path. Therefore, after the i th iteration, we do not know every value of $\partial J_{(i)}(x,t)/\partial x$ on $\forall x\in\mathbb{R}^n,\ t\in[0,T]$, but only know $\partial J_{(i)}(x,t)/\partial x\mid_{x=\bar{\xi}_{(i)}(t)}$ along the i th suboptimal path $\bar{\xi}_{(i)}$. However, the next iteration may require $\partial J_{(i)}(x,t)/\partial x$ at a different state, for calculations of $\hat{f}_{(i+1)}(x,t)$ in Eq. (25) and $\hat{V}_{(i)}(x,t)$ in Eq. (26). Regarding this, in Subsections IV-B and IV-C, we provide two approximation methods to cope with this computational issue.

Finally, we provide a concrete algorithm for computation of ISOC-PI in Subsection IV-D.

A. Computation method for the solution and controller based on Monte Carlo sampling

As the conventional path integral methods [15], [16], several statistical sampling methods such as Monte Carlo sampling, Langevin sampling, Importance sampling, Laplace approximation, and so on can be used for efficient computation in ISOC-PI. Here, we provide a basic computation method of the solution in Eq. (27) and controller in Eq. (31) by using Monte Carlo sampling with the sample number $N_{(i)} > 0$ as

$$\hat{S}^k_{(i)}(x,t) := \int_t^T \hat{V}_{(i)}(\xi^k,\tau)\,\mathrm{d}\tau + \phi(\xi^k(T)) \tag{51}$$

$$J^{N_{(i)}}_{(i)}(x,t) = \frac{\sum_{k=1}^{N_{(i)}} \hat{S}^k_{(i)}(x,t)}{N_{(i)}} \tag{52}$$

$$\frac{\partial J^{N_{(i)}}_{(i)}(x,t)}{\partial x^c}\,\mathrm{d}t = \Xi^c(x,t)^{-1}h^c(x,t)\frac{\sum_{k=1}^{N_{(i)}} \hat{S}^k_{(i)}(x,t)\mathrm{d}w^k(t)}{N_{(i)}}, \tag{53}$$

$$u^{*N_{(i)}}_{(i)}(x,t)\,\mathrm{d}t = -R(x,t)^{-1}g^c(x,t)^\top \frac{\partial J^{N_{(i)}}_{(i)}(x,t)}{\partial x^c}\,\mathrm{d}t, \tag{54}$$

where ξ^k represents the k th sample path of the total $N_{(i)}$ samples under the i th sample generating dynamics (29). According to the property of Monte Carlo sampling, both approximations Eqs. (52) and (54) respectively converge to $J_{(i)}(x,t)$ in Eq. (27) and $u^*_{(i)}(x,t)$ in Eq. (31) almost surely as the total sampling number $N_{(i)} \to \infty$.

B. Approximate method based on curve fitting

Here, we consider an approximate method of the ISOC-PI based on curve fitting. Let $L_{(i)}$ denote the total number of suboptimal paths at the i th iteration. In this method, we generate $L_{(i)}$ suboptimal paths by repeating the i th iteration of the ISOC-PI method $L_{(i)}$ times. Let $\bar{\xi}^l_{(i)}$ be the l th path of the $L_{(i)}$ suboptimal paths. Then, we have $L_{(i)}$ time series data

$$\frac{\partial J_{(i)}(x,t)}{\partial x_j}\Big|_{x=\bar{\xi}^l_{(i)}(t)}, \quad t\in[0,T], \quad (l=1,\dots,L_{(i)}).$$

for each j, $(j=1,\dots,n)$. From those data, we construct an approximation of $\partial J_{(i)}(x,t)/\partial x_j$, which will be used for the next iteration, by a polynomial curve fitting. By setting the highest orders O_x and O_t of x and t, a polynomial curve fitting method gives polynomial coefficients $K^{i,j}_{o_{x_1}\dots o_{x_n}o_t}$, and we obtain the following polynomial approximation:

$$\frac{\partial J_{(i)}(x,t)}{\partial x_j} \approx \sum_{\substack{o_{x_1}+\dots+o_{x_n}\le O_x \\ o_t \le O_t}} K^{i,j}_{o_{x_1}\dots o_{x_n}o_t}\, x_1^{o_{x_1}}\dots x_n^{o_{x_n}} t^{o_t}. \tag{55}$$

C. Approximate method based on local quadratic approximation of the value function

Next, we consider an approximate method of the ISOC-PI based on local quadratic approximation of the value function. A local quadratic approximation technique of the value function is proposed in [12]. The literature [23] uses our previous version of the ISOC-PI, which is reported in [24], and it provides an approximate method based on [12] for our method [24]. However, the approximate method in [23] is insufficient in that it does not consider the noise effect at all. This subsection is motivated by the two literature [12], [23], and we provide some modification of them for an approximate method of the ISOC-PI. The resultant recurrence formulae are slightly different from those in [12]. Our formulation is written in a continuous-time manner instead of a discrete-time manner. Besides, the authors in [12] consider the deviations around a deterministic nominal trajectory, while we consider the deviations around a stochastic sample path. This causes different treatments of some noise terms and the terminal condition in deriving the recurrence formulae.

Here, suppose that an i th suboptimal path on $t\in[0,T]$ has been already obtained. In this subsection, we consider an arbitrary iteration i, and drop the subscript (i) for notational simplicity. Let $\bar{u}$ and $\bar{\xi}$ denote a pair of sequences of the optimal control input and optimal path on $t\in[0,T]$. The objective here is that for given (x,t), we approximately calculate the partial derivative $\partial J(x,t)/\partial x$ by using the data $\bar{u}$ and $\bar{\xi}$. The local quadratic approximation describes $J(x,t)$ as $J(x,t) \approx J(\bar{\xi}(t),t) + \delta J(\delta x,t;\bar{\xi},\bar{u})$ with $\delta x := x - \bar{\xi}(t)$, where $\delta J(\delta x,t;\bar{\xi},\bar{u})$ is represented as $\delta J(\delta x,t;\bar{\xi},\bar{u}) = 1/2\delta x^\top \mathbf{S}(t;\bar{\xi},\bar{u})\delta x + \mathbf{s}(t;\bar{\xi},\bar{u})^\top\delta x + s(t;\bar{\xi},\bar{u})$ with some $\mathbf{S}(t;\bar{\xi},\bar{u})\in\mathbb{R}^{n\times n}$, $\mathbf{s}(t;\bar{\xi},\bar{u})\in\mathbb{R}^n$ and $s(t;\bar{\xi},\bar{u})\in\mathbb{R}$. Based on the arguments in [12], we can derive recurrence formulae of $\mathbf{S}(t;\bar{\xi},\bar{u}), \mathbf{s}(t;\bar{\xi},\bar{u})$ and $s(t;\bar{\xi},\bar{u})$. For notational simplicity, we define the followings:

$$\begin{aligned}
&\mathbf{v}(x,u,t) := V(x,t) + \frac{1}{2}u^\top R(x,t)u\\
&\mathbf{F}(x,u,t) := f(x,t) + g(x,t)u\\
&\mathbf{G}(t;\bar{\xi},\bar{u}) := \frac{\partial^2\mathbf{v}(\bar{\xi}(t),\bar{u}(t),t)}{\partial u\partial x}\,\mathrm{d}t + \left(\frac{\partial\mathbf{F}(\bar{\xi}(t),\bar{u}(t),t)}{\partial u}\,\mathrm{d}t\right)^\top\\
&\quad\times\mathbf{S}(t;\bar{\xi},\bar{u})\left(\frac{\partial\mathbf{F}(\bar{\xi}(t),\bar{u}(t),t)}{\partial x}\,\mathrm{d}t + I_n\right)\\
&\mathbf{H}(t;\bar{\xi},\bar{u}) := \frac{\partial^2\mathbf{v}(\bar{\xi}(t),\bar{u}(t),t)}{\partial^2 u}\,\mathrm{d}t + \left(\frac{\partial\mathbf{F}(\bar{\xi}(t),\bar{u}(t),t)}{\partial u}\,\mathrm{d}t\right)^\top\\
&\quad\times\mathbf{S}(t;\bar{\xi},\bar{u})\left(\frac{\partial\mathbf{F}(\bar{\xi}(t),\bar{u}(t),t)}{\partial u}\,\mathrm{d}t\right)\\
&\mathbf{g}(t;\bar{\xi},\bar{u}) := \frac{\partial\mathbf{v}(\bar{\xi}(t),\bar{u}(t),t)}{\partial u}^\top\,\mathrm{d}t + \left(\frac{\partial\mathbf{F}(\bar{\xi}(t),\bar{u}(t),t)}{\partial u}\,\mathrm{d}t\right)^\top\\
&\quad\times\mathbf{s}(t;\bar{\xi},\bar{u}),
\end{aligned}$$

where we denote $\bar{u}(\bar{\xi}(t),t)$ as just $\bar{u}(t)$. Then, the recurrence

formulae are given by

$$
\begin{aligned}
-d\mathbf{S} &= \frac{\partial^2 \mathbf{v}}{\partial x^2}\,\mathrm{d}t + \left(\frac{\partial \mathbf{F}}{\partial x}\,\mathrm{d}t\right)^\top \mathbf{S}\left(\frac{\partial \mathbf{F}}{\partial x}\,\mathrm{d}t\right) + \left(\frac{\partial \mathbf{F}}{\partial x}\,\mathrm{d}t\right)^\top \mathbf{S} \\
&\quad + \mathbf{S}\left(\frac{\partial \mathbf{F}}{\partial x}\,\mathrm{d}t\right) - \mathbf{G}^\top \mathbf{H}^{-1}\mathbf{G} + \sum_{j,k=1}^{r} \frac{\partial [h]_{:,j}}{\partial x}^\top \mathbf{S} \frac{\partial [h]_{:,k}}{\partial x}[Q]_{k,j}\,\mathrm{d}t, \\
\mathbf{S}(T;\bar{\xi},\bar{u}) &= \frac{\partial^2 \phi(\bar{\xi}(T))}{\partial x^2} \qquad (56) \\
-d\mathbf{s} &= \frac{\partial \mathbf{v}}{\partial x}^\top \mathrm{d}t + \left(\frac{\partial \mathbf{F}}{\partial x}\,\mathrm{d}t\right)^\top \mathbf{s} - \mathbf{G}^\top \mathbf{H}^{-1}\mathbf{g}, \\
\mathbf{s}(T;\bar{\xi},\bar{u}) &= \frac{\partial \phi(\bar{\xi}(T))}{\partial x}^\top \qquad (57) \\
-ds &= -\frac{1}{2}\mathbf{g}^\top \mathbf{H}^{-1}\mathbf{g}\ , \qquad s(T;\bar{\xi},\bar{u}) = 0. \qquad (58)
\end{aligned}
$$

Here, for a matrix A, $[A]_{j,k}$ denotes the (j,k) th element of A and $[A]_{j,:}$ and $[A]_{:,k}$ denote the j th row and k th column, respectively.

After the i th iteration of ISOC-PI, for a resultant pair of sequences $\bar{\xi}_{(i)}$ and $u^*_{(i)}$ on $t \in [0,T]$, we can calculate the corresponding $\mathbf{S}(t;\bar{\xi}_{(i)},u^*_{(i)}), \mathbf{s}(t;\bar{\xi}_{(i)},u^*_{(i)})$ and $s(t;\bar{\xi}_{(i)},u^*_{(i)})$ from Eqs. (56), (57) and (58). Then, for a given (x,t), $\partial J_{(i)}(x,t)/\partial x$ is approximately calculated as

$$
\begin{aligned}
\frac{\partial J_{(i)}(x,t)}{\partial x}^\top &\approx \frac{\partial J_{(i)}(\bar{\xi}_{(i)}(t),t)}{\partial x}^\top \\
&\quad + \mathbf{S}(t;\bar{\xi}_{(i)},u^*_{(i)})(x - \bar{\xi}_{(i)}(t)) + \mathbf{s}(t;\bar{\xi}_{(i)},u^*_{(i)}). \qquad (59)
\end{aligned}
$$

D. Algorithm of ISOC-PI with Monte Carlo sampling

We provide a concrete algorithm named ISOC-PI with Monte Carlo. Here, consider the plant system (9) and the cost function $\Gamma(x_0,u_{0:T},0)$ in (5). Suppose that the conditions for the existence and uniqueness of the solutions mentioned in Subsection II-A, and Assumption 1 hold.

[Algorithm 1:] (ISOC-PI with Monte Carlo)

Given: A positive integer N_I as the the number of iterations of ISOC-PI. A sufficiently short sampling period $dt > 0$ for the plant dynamics (1). The terminal time $T := N_T\,\mathrm{d}t$ with some positive integer N_T. An initial function $J_{(0)}$ such that it is a class $C^{2,1}$ and $\|\partial J_{(0)}(x,\tau)/\partial x^c\|^2 < \infty$, $\forall x \in \mathbb{R}^n, \tau \in \{0,\mathrm{d}t,\ldots,T\}$.

If the curve fitting in Subsection IV-B is used, positive integers O_x and O_t as the highest orders of x and t, and a positive integer $L_{(i)}$ as the total number of suboptimal paths.

Obtain: Optimal path $\bar{\xi}$, solution $J(\bar{\xi}(\tau),\tau)$, its partial derivative $\partial J(\bar{\xi}(\tau),\tau)/\partial x^c$ and controller $u^*(\bar{\xi}(\tau),\tau)$

for $i = 1$ to N_I **do**

Choose the total number of samples $N_{(i)} > 0$, and set $\bar{\xi}_{(i)}(0) = x_0$.

Execute the following procedure, and obtain an i th suboptimal path $\bar{\xi}_{(i)}$, solution $J^{N_{(i)}}_{(i)}$, its partial derivative $\partial J^{N_{(i)}}_{(i)}/\partial x^c$ and controller $u^{*N_{(i)}}_{(i)}$ of the i th iteration of ISOC-PI (24):

for $j = 1$ to N_T **do**

for $k = 1$ to $N_{(i)}$ **do**

Generate a sample $\xi^k(\tau), \tau = j\,\mathrm{d}t,\ldots,N_T\,\mathrm{d}t$ under the i th sample generating dynamics (29) with $\xi^k(j\,\mathrm{d}t) = \bar{\xi}_{(i)}(j\,\mathrm{d}t)$.

if $i > 1$ and the curve fitting in Subsection IV-B is used **then**

Use Eq. (55) to calculate $\partial J_{(i-1)}(\xi^k(\tau),\tau)/\partial x^c$.

else if $i > 1$ and the local quadratic approximation in Subsection IV-C is used **then**

Use Eq. (59) to calculate $\partial J_{(i-1)}(\xi^k(\tau),\tau)/\partial x^c$.

end if

end for

Calculate $\hat{S}^k_{(i)}(\bar{\xi}_{(i)}(j\,\mathrm{d}t),j\,\mathrm{d}t)$, $J^{N_{(i)}}_{(i)}(\bar{\xi}_{(i)}(j\,\mathrm{d}t),j\,\mathrm{d}t)$, $\partial J^{N_{(i)}}_{(i)}(\bar{\xi}_{(i)}(j\,\mathrm{d}t),j\,\mathrm{d}t)/\partial x^c$ and $u^{*N_{(i)}}_{(i)}(\bar{\xi}_{(i)}(j\,\mathrm{d}t),j\,\mathrm{d}t)$ by using Eqs. (51), (52), (53) and (54), respectively.

Observe $\bar{\xi}_{(i)}((j+1)\,\mathrm{d}t)$ from the plant dynamics (1) with $\bar{\xi}_{(i)}(j\,\mathrm{d}t)$ and $u = u^{*N_{(i)}}_{(i)}(\bar{\xi}_{(i)}(j\,\mathrm{d}t),j\,\mathrm{d}t)$.

$\bar{\xi}_{(i)}(j\,\mathrm{d}t) \Leftarrow \bar{\xi}_{(i)}((j+1)\,\mathrm{d}t)$

end for

if the curve fitting in Subsection IV-B is used **then**

Repeat the above procedure $L_{(i)}$ times, and calculate the polynomial coefficients $K^{i,:}_{o_{x_1}\ldots o_{x_n}o_t}$.

end if

end for

return $\bar{\xi} \Leftarrow \bar{\xi}_{(N_I)}$, $J \Leftarrow J^{N_{(N_I)}}_{(N_I)}$, $\partial J/\partial x^c \Leftarrow \partial J^{N_{(N_I)}}_{(N_I)}/\partial x^c$ and $u^* \Leftarrow u^{*N_{(N_I)}}_{(N_I)}$

V. Numerical examples

This section exhibits applications of ISOC-PI. First, in Subsection V-A, we start with an illustrative toy problem. Here, we consider a one-dimensional stochastic bilinear system in terms of the state and the noise. Second, in Subsection V-B, we consider a one-link robot manipulator in the presence of noise. All simulations in this section have been executed by using the Euler-Maruyama method with the time step of 1×10^{-3} s. The control is implemented via a sample and zero-order hold with the sampling interval of 1×10^{-2} s.

A. Illustrative toy problem of a one-dimensional stochastic bilinear system

Since it is generally difficult to analytically calculate the value function and optimal controller of a stochastic optimal control problem, we start with an illustrative toy problem. Let us consider the following one-dimensional stochastic bilinear system in terms of the state and the noise:

$$\mathrm{d}x = gu\,\mathrm{d}t + hx\,\mathrm{d}w,$$

where $x,u,w \in \mathbb{R}$, and g and h are constants. The cost function to be minimized is defined as

$$E^{\mathcal{P}}\left\{\frac{\Lambda_e x(T)^2}{2} + \int_0^T \frac{Ru(\tau)^2}{2}\,\mathrm{d}\tau \;\middle|\; x(0) = x_0\right\},$$

where Λ_e and R are positive constants, respectively. Note that the conventional path integral method mentioned in Subsection II-B is not applicable to this problem, since Assumption 2 does not hold.

The corresponding SHJB equation (8) is obtained as

$$\frac{\partial J}{\partial t}-\frac{1}{2}\frac{g^2}{R}\left(\frac{\partial J}{\partial x}\right)^2+\frac{1}{2}\frac{\partial^2 J}{\partial x^2}h^2x^2Q=0, \quad J(x,T)=\frac{\Lambda_e x^2}{2}, \tag{60}$$

where $Q > 0$ denotes the covariance in Eq. (2). We can analytically solve the SHJB equation (60), and the solution $J(x,t)$ is given by

$$J(x,t)= \begin{cases} \frac{1}{2}\frac{\Lambda_e Rh^2Q}{\Lambda_e g^2-e^{h^2Q(t-T)}(\Lambda_e g^2-Rh^2Q)}x^2, & \text{if } Rh^2Q\le\Lambda_e g^2 \\ \frac{1}{2}\frac{\Lambda_e Rh^2Q}{\Lambda_e g^2+e^{h^2Q(t-T)}(Rh^2Q-\Lambda_e g^2)}x^2, & \text{Otherwise.} \end{cases} \tag{61}$$

Thus, the optimal controller is followed from Eqs. (7), (61) as

$$u^*(x,t)= \begin{cases} -\frac{\Lambda_e gh^2Q}{\Lambda_e g^2-e^{h^2Q(t-T)}(\Lambda_e g^2-Rh^2Q)}x, & \text{if } Rh^2Q\le\Lambda_e g^2 \\ -\frac{\Lambda_e gh^2Q}{\Lambda_e g^2+e^{h^2Q(t-T)}(Rh^2Q-\Lambda_e g^2)}x, & \text{Otherwise.} \end{cases} \tag{62}$$

We compare the sample paths with controllers given by the proposed ISOC-PI with Algorithm 1 with the curve fitting, and with the optimal controller (62), respectively. The concrete parameters used in the simulation are $g = 1$, $h = 1$, $Q = 1$, $\Lambda_e = 1$ and $R = 0.2$. Those parameters imply that $Rh^2Q \le \Lambda_2 g^2$ holds in Eqs. (61) and (62). We execute 5 iterations of the proposed method ($N_I = 5$), and this set of iterations is referred to as 1 optimization trial. We repeat 10 optimization trials. The total sample number at each iteration is $N_{(i)} = 5000$ $(i = 1,\ldots,5)$. The terminal time is $T = 0.5$ s, the initial condition is $x(0) = 2$, and the initial function is chosen to be $J_{(0)}(x,t) \equiv 0$. Figure 1 shows the first iteration results in one of the 10 optimization trials, and the optimal ones under the same noise sequence, respectively. The top figures exhibit the time responses of the resultant path under the first suboptimal controller, and the optimal path under the optimal controller (62), respectively. The time sequences of the first suboptimal input and the optimal one are shown in the middle figures, respectively. The bottom figures exhibit the time sequences of the value function at the first iteration $J_{(1)}(x,t)$ and the true value function $J(x,t)$ in Eq. (61), respectively. Figure 2 shows those results in the fifth iteration in the same optimization trial as well as Fig. 1. Figure 3 displays changes in the value function at the initial condition $J_{(i)}(0,x_0)$ versus iterations $i = 1,\ldots,5$ in the solid line with circle. It also shows the average of each $J_{(i)}(0,x_0)$ of 10 samples with 95% CI (confidence interval) in the dotted line with circle. Besides, the solid line with diamond exhibits the optimal value of the true value function $J(0,x_0)$ in Eq. (61). Finally, in Fig. 4, the left figure exhibits the surface of $\partial J(x,t)/\partial x$ on $-5 \le x \le 5$ and $0 \le t \le T$. The right figure shows an approximated surface generated by a polynomial curve fitting of the first order of x and the third order of t to 10 time sequences of $\partial J_{(5)}(x,t)/\partial x$ along the sample paths at the fifth iteration of 10 optimization trials, which means $O_x = 1, O_t = 3$ and $L_{(i)} = 10(i = 1,\ldots,5)$. From

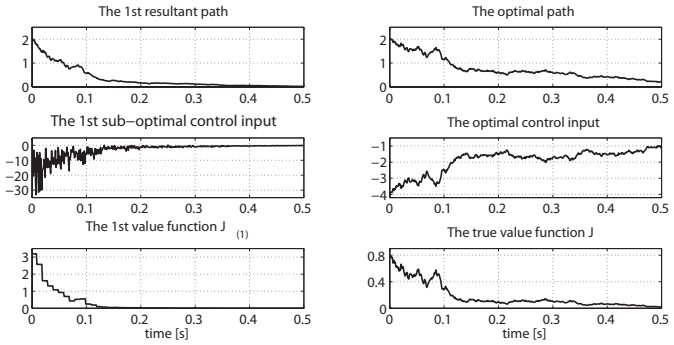

Fig. 1. The first iteration results (left), and the optimal ones under the same noise sequence (right)

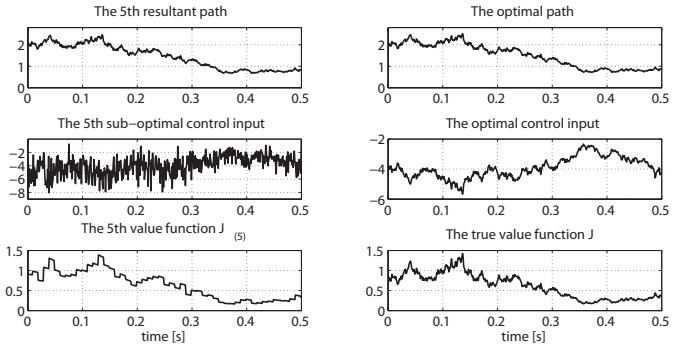

Fig. 2. The fifth iteration results (left), and the optimal ones under the same noise sequence (right)

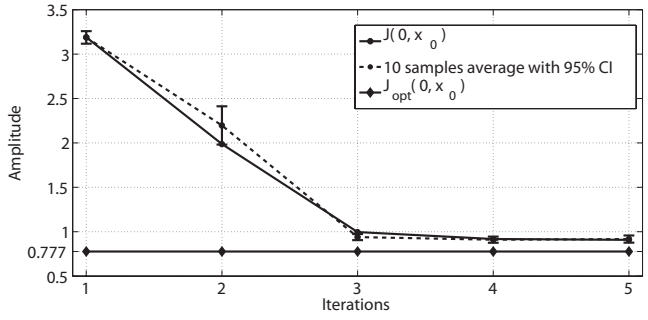

Fig. 3. The changes in $J_{(i)}(0,x_0)$ versus iterations $i = 1,\ldots,5$ (solid line with circle), the average of each $J_{(i)}(0,x_0)$ of 10 samples with 95% CI (dotted line with circle), and the optimal value of $J(0,x_0)$ (solid line with diamond)

Figs. 1 and 2, the controller performance is improved as the iteration proceeds. At the fifth iteration, the time history of the resultant path exhibits good consistency with the optimal path, and the time history of the value function is also close to that of the true value function. Although the high frequency components between the resultant controller and the optimal one are different, their low frequency components behave similarly. Figure 3 shows that the value function monotonically decreases and converges along the iteration. It is the reason

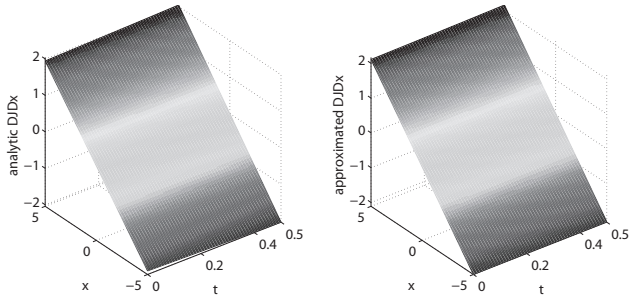

Fig. 4. The surface of $\partial J(x,t)/\partial x$ (left), and that generated by a polynomial curve fitting of $\partial J_{(5)}(x,t)/\partial x$ with $O_x = 1, O_t = 3$ and $L_{(5)} = 10$ (right).

TABLE I
PHYSICAL PARAMETERS

m	Mass of the link	[kg]
l	Length of the link	[m]
l_c	Length to the center of gravity	[m]
I	Inertia of the link	[kgm^2]
g	Gravity acceleration	[m/s^2]

why the error remains compared to the true value function that we use finite sample numbers in optimization trials in this simulation. Figure 4 exhibits that the resultant surface represents the true $\partial J(x,t)/\partial x$ well. From Fig. 4, it is fairly expected that we can generate an approximated surface of $\partial J(x,t)/\partial x$ and therefore the optimal feedback controller. Once such a surface is generated, we can immediately obtain the optimal control input at any x and t as mentioned in Subsection IV-B.

B. Swing-up control of a one-link robot manipulator

Let us consider a one-link robot manipulator moving on a vertical plane depicted in Fig. 5. As in the figure, the joint angle of the link is denoted by θ, and the control torque is denoted by u, respectively. The physical parameters of this apparatus are summarized in Table I.

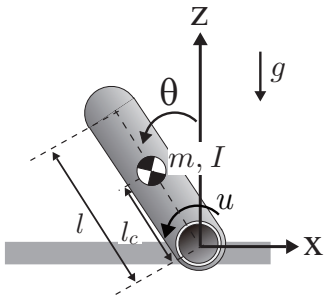

Fig. 5. One-link robot manipulator

As the control objective, we consider the swing-up control of the manipulator. We aim to obtain the optimal controller such that the manipulator swings up from the pendant position to the upright position, and remains at this position even in the presence of noise.

The dynamics of this apparatus with the state $x := (\theta, \dot{\theta})^\top$ is described by a system of the form (9) as

$$\mathrm{d}x = \begin{pmatrix} x_2 \\ \frac{ml_c g \sin x_1}{ml_c^2+I} \end{pmatrix} \mathrm{d}t + \begin{pmatrix} 0 \\ \frac{1}{ml_c^2+I} \end{pmatrix} u\,\mathrm{d}t + \begin{pmatrix} 0_{1\times r} \\ h^c(x,t) \end{pmatrix} \mathrm{d}w. \tag{63}$$

In this simulation, we choose the noise port $h^c(x,t) \in \mathbb{R}^{1\times r}$ as

$$h^c(x,t) = \frac{1}{ml_c^2 + I} \begin{pmatrix} k_{h1} & k_{h2}x_2 \end{pmatrix} \tag{64}$$

with $r = 2$ and some constants k_{h1} and k_{h2}. Equation (64) describes persistent system noise and uncertainty in viscous friction. Regarding to the aforementioned control objective, the cost function to be minimized is defined as

$$E^{\mathcal{P}}\left\{ \frac{1}{2}x(T)^\top \Lambda_e x(T) + \int_0^T \frac{1}{2}x(\tau)^\top \Lambda_x x(\tau) + \frac{1}{2}Ru(\tau)^2\,\mathrm{d}\tau \;\middle|\; x(0) = x_0 \right\}, \tag{65}$$

where symmetric positive definite matrices Λ_e and Λ_x, and a positive constant R denote weighting matrices and constant with respect to the terminal, instantaneous and input costs, respectively. As in the case of the previous example, the conventional path integral method is not applicable to this problem.

Although the corresponding SHJB equation is obtained from Eq. (8), it cannot be analytically solved unlike the previous example. In this example, we generate an optimal controller by using the ISOC-PI with Algorithm 1 with the local quadratic approximation, and we investigate behaviors of the resultant optimal paths. We also design the LQG controller for the linearized system of Eq. (63) around the origin, and compare the performance of ISOC-PI and that of LQG. The concrete parameters used in the simulation are $m = 1$ kg, $l = 1, l_c = 0.5$ m and $I = 8.3 \times 10^{-2}$ kgm^2. We choose the parameters of the noise port (64) as $k_{h1} = 1$ and $k_{h2} = -0.1$. The covariance matrix of the noise is set to $Q = I_2$. The design parameters are chosen as $\Lambda_x = \mathrm{diag}\{1, 0.1\}$, $\Lambda_e = \mathrm{diag}\{4, 4\}$ and $R = 2$, respectively. First, we solve the LQG problem for the linearized system and the same cost function (65), and obtain a value function J_{LQG} and a corresponding LQG controller u_{LQG}. Then, we execute 5 iterations of the proposed method, and this set of iterations is referred to as 1 optimization trial. We repeat 10 optimization trials. The total sample number at each iteration is $N_{(i)} = 8000$ $(i = 1, \ldots, 5)$. The terminal time is $T = 1.5$ s, the initial condition is $x(0) = (-\pi, 0)^\top$, which represents the pendant position, and the initial function is chosen to be $J_{(0)}(x,t) = J_{LQG}(x,t)$.

The simulation results are shown in Figs. 6 to 8. Figure 6 displays changes in the value function at the initial condition $J_{(i)}(0, x_0)$ versus iterations $i = 1, \ldots, 5$ in the solid line with circle. It also shows the average of each $J_{(i)}(0, x_0)$ of 10 samples with 95% CI in the dotted line with circle. From Fig. 6, since the value function monotonically decreases until the fourth iteration and then slightly increases at the fifth iteration because of using finite sample numbers, we adopt the fourth

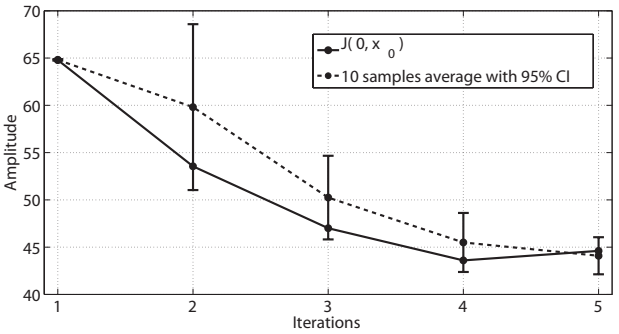

Fig. 6. The changes in $J_{(i)}(0, x_0)$ versus iterations $i = 1, \ldots, 5$ (solid line with circle) and the average of each $J_{(i)}(0, x_0)$ of 10 samples with 95% CI (dotted line with circle)

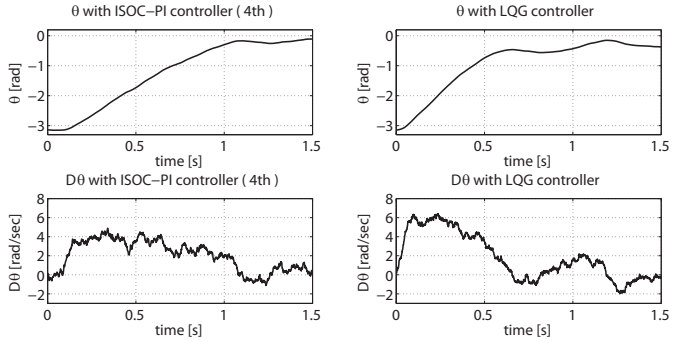

Fig. 7. The fourth iteration results of θ and $\dot{\theta}$ (left), and those with the LQG controller under the same noise sequence (right)

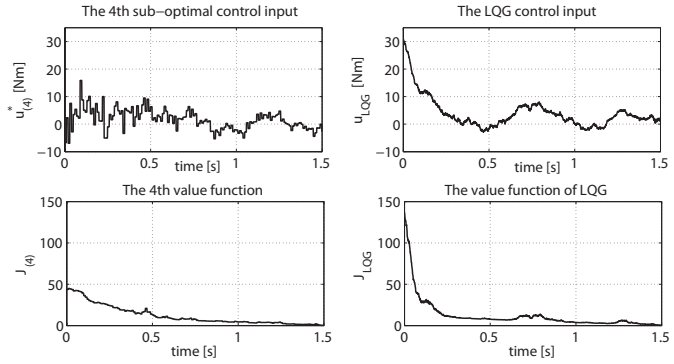

Fig. 8. The fourth iteration results of $u^*(x,t)$ and $J(x,t)$ (left), and those with the LQG controller under the same noise sequence (right)

controller $u^*_{(4)}$ as the resultant controller. We confirm that the resultant controllers from the 10 optimization trials achieve a similar performance as is implied from the average result in Fig. 6. Figures 7 and 8 show the fourth iteration results in one of the 10 optimization trials and the results with the LQG controller. The top figures in Fig. 7 exhibit the time response of θ under the resultant controller, and that under the LQG controller, respectively. The time sequences of $\dot{\theta}$ and that with the LQG controller are respectively shown in the bottom figures. The top figures in Fig. 8 exhibit the time sequences of the resultant controller and that of the LQG controller, respectively. The bottom figures show the time sequences of the resultant value function, i.e., $J_{(4)}(x,t)$, and the value function obtained from the LQG, $J_{LQG}(x,t)$, respectively. Figures 7 and 8 show that the resultant controller obtained by ISOC-PI has a better performance with less control input than the LQG controller.

VI. Further progresses

Although this paper only considers a nonlinear stochastic optimal control with a fixed time interval, recently we have had further progresses of this framework. Some of them has been already reported in conference proceedings [25], [26].

First, we have succeeded to deal with other types of optimal control problems: infinite time interval optimal control with average cost and with discounted cost, and first exit time optimal control, where the terminal time is described by a Markov random time. Second, we have strictly taken input saturations with prescribed saturation functions into account in nonlinear stochastic optimal control. Here, we have newly derived an SHJB equation considering input saturations based on the argument in [27], and provided an iterative solution method for the resultant equation. Third, we have also provided a solution method for a nonlinear stochastic H_∞ control problem [28], [29] based on our framework. Thus, the proposed iterative solution method for SHJB equation based on path integral analysis has a potential for solving various types of nonlinear stochastic control problems. The results provided in this paper are fundamental to those extensions.

Besides, one of the authors has demonstrated a real time performance for high dimensional systems in [30], where a real time application of the path integral control to a 20 dimensional control task with 10 quadrotors using about 10.000 samples per control step. The report [30] shows that this method is well adapted to parallel computing and the computation can be easily accelerated using GPUs.

VII. Conclusion

This paper has proposed a new iterative solution method for nonlinear stochastic optimal control problems based on path integral analysis. First, we have provided an iteration law for solving a corresponding SHJB equation to a stochastic optimal control problem. Each iteration procedure of the proposed method is represented by a Cauchy problem for a linear parabolic PDE, and its explicit solution is given by the Feynman-Kac formula. Second, we have derived a suboptimal feedback controller at each iteration by using the path integral analysis. The resultant controller forms the expectation of the noise under some probability of the path cost, and is effectively calculated by using several sampling methods. Third, the convergence property of the proposed method has been investigated, and convergence conditions have been clarified. Under these conditions, it is shown that the sequence of solutions for the proposed iteration converges, and the SHJB equation is satisfied. Since this result is a qualitative convergence property, investigation of quantitative properties such as the estimates of necessary samples and steps is an important future work. Finally, numerical simulations demonstrate the effectiveness of the proposed method.

APPENDIX A
PROOF OF THEOREM 1

We prove the first assertion of the theorem. Here, fix an arbitrary integer i. Since the PDE in the i th iteration of ISOC-PI (24) relates to the infinitesimal operator $\mathcal{L}_0$ in Eq. (3) associated with the i th sample generating dynamics (29), the stochastic representation of its solution $J_{(i)}(x,t)$ is given by Eqs. (27) and (28) from the Feynman-Kac formula ([6], Theorem 1.3.17).

Then, we prove the other assertion of the theorem. We derive the suboptimal controller $u^*_{(i)}(x,t)$ by using the path integral analysis. Consider the probability $\hat{p}_{(i)}(\xi_{t:T}|x,t)$, and then it can be represented by the state transition probability as

$$\hat{p}_{(i)}(\xi_{t:T}|x,t)=\lim_{\mathrm{d}t\to 0}\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s),\ \ \xi(t)=x.$$

Since the dynamics (29) is partitioned into the directly noise-driven subsystem and the other one, according to the same argument as in Eq. (17), the state transition probability can also be separated, and therefore there exists a proportionality constant $k_{p(i)}$ satisfying

$$\hat{p}_{(i)}(\xi_{t:T}|x,t)=k_{p(i)}\hat{p}_{(i)}(\xi^c_{t:T}|x,t). \tag{66}$$

Then, from Eqs. (27) and (66), the solution to the i th iteration of ISOC-PI in Eq. (24) is rewritten as

$$J_{(i)}(x,t)=k_{p(i)}\lim_{\mathrm{d}t\to 0}\sum_{\xi_{t+\mathrm{d}t:T}}\left[\left(\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\right)\left(\sum_{s=t}^{T-\mathrm{d}t}\hat{V}_{(i)}(\xi(s),s)\,\mathrm{d}t+\phi(\xi(T))\right)\right]. \tag{67}$$

By using Eqs. (66) and (67), we can calculate $\partial J_{(i)}(x,t)/\partial x^c$ as

$$\begin{aligned}
\frac{\partial J_{(i)}(x,t)}{\partial x^c}&=k_{p(i)}\lim_{\mathrm{d}t\to 0}\sum_{\xi_{t+\mathrm{d}t:T}}\left[\frac{\partial}{\partial x^c}\left(\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\right)\left(\sum_{s=t}^{T-\mathrm{d}t}\hat{V}_{(i)}(\xi(s),s)\,\mathrm{d}t+\phi(\xi(T))\right)+\left(\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\right)\frac{\partial}{\partial x^c}\left(\sum_{s=t}^{T-\mathrm{d}t}\hat{V}_{(i)}(\xi(s),s)\,\mathrm{d}t\right)\right]\\
&=k_{p(i)}\lim_{\mathrm{d}t\to 0}\sum_{\xi_{t+\mathrm{d}t:T}}\left[\frac{\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)}{\hat{p}_{(i)}(\xi^c(t+\mathrm{d}t),t+\mathrm{d}t|x,t)}\right.\\
&\times\left(\sum_{s=t}^{T-\mathrm{d}t}\hat{V}_{(i)}\,\mathrm{d}t+\phi(\xi(T))\right)\frac{\partial\hat{p}_{(i)}(\xi^c(t+\mathrm{d}t),t+\mathrm{d}t|x,t)}{\partial x^c}\\
&\left.+\left(\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\right)\frac{\partial\hat{V}_{(i)}(x,t)\,\mathrm{d}t}{\partial x^c}\right]\\
&=\lim_{\mathrm{d}t\to 0}\sum_{\xi_{t+\mathrm{d}t:T}}\left[k_{p(i)}\prod_{s=t}^{T-\mathrm{d}t}\hat{p}_{(i)}(\xi^c(s+\mathrm{d}t),s+\mathrm{d}t|\xi(s),s)\right.\\
&\times\left\{\left(\sum_{s=t}^{T-\mathrm{d}t}\hat{V}_{(i)}\,\mathrm{d}t+\phi(\xi(T))\right)\frac{\partial\log\hat{p}_{(i)}(\xi^c(t+\mathrm{d}t),t+\mathrm{d}t|x,t)}{\partial x^c}\right.\\
&\left.\left.+\frac{\partial\hat{V}_{(i)}(x,t)\,\mathrm{d}t}{\partial x^c}\right\}\right]\\
&=\lim_{\mathrm{d}t\to 0}E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\left\{\left(\sum_{s=t}^{T-\mathrm{d}t}\hat{V}_{(i)}\,\mathrm{d}t+\phi(\xi(T))\right)\right.\\
&\left.\times\frac{\partial\log\hat{p}_{(i)}(\xi^c(t+\mathrm{d}t),t+\mathrm{d}t|x,t)}{\partial x^c}+\frac{\partial\hat{V}_{(i)}(x,t)\,\mathrm{d}t}{\partial x^c}\right\}.
\end{aligned}\tag{68}$$

We continue to calculate the argument of the expectation on the right hand side of the last equality in Eq. (68). Eqs. (19) and (29) yield

$$\begin{aligned}
\frac{\partial\log\hat{p}_{(i)}(\xi^c(t+\mathrm{d}t),t+\mathrm{d}t|x,t)}{\partial x^c}&=-\frac{1}{2\det\Xi^c}\frac{\partial\det\Xi^c}{\partial x^c}\\
&+\frac{1}{\mathrm{d}t}\left(\xi^c(t+\mathrm{d}t)-x^c-\hat{f}^c_{(i)}\,\mathrm{d}t\right)^\top\Xi^{c-1}\left(I_{n_c}+\frac{\partial\hat{f}^c_{(i)}}{\partial x^c}\,\mathrm{d}t\right)\\
&-\frac{1}{2\,\mathrm{d}t}\left(\xi^c(t+\mathrm{d}t)-x^c-\hat{f}^c_{(i)}\mathrm{d}t\right)^\top\\
&\times\left.\frac{\partial\Xi^c(y,t)^{-1}(\xi^c(t+\mathrm{d}t)-x^c-\hat{f}^c_{(i)}\mathrm{d}t)}{\partial y^c}\right|_{y=x}\\
&=-\frac{1}{2}\left(\mathrm{tr}\left\{\Xi^{c-1}\frac{\partial\Xi^c}{\partial x^c_1}\right\},\dots,\mathrm{tr}\left\{\Xi^{c-1}\frac{\partial\Xi^c}{\partial x^c_{n_c}}\right\}\right)\\
&+\frac{1}{\mathrm{d}t}\left(\xi^c(t+\mathrm{d}t)-x^c-\hat{f}^c_{(i)}\,\mathrm{d}t\right)^\top\Xi^{c-1}\left(I_{n_c}+\frac{\partial\hat{f}^c_{(i)}}{\partial x^c}\,\mathrm{d}t\right)\\
&-\frac{1}{2\,\mathrm{d}t}\left(\xi^c(t+\mathrm{d}t)-x^c-\hat{f}^c_{(i)}\mathrm{d}t\right)^\top\\
&\times\left.\frac{\partial\Xi^c(y,t)^{-1}(\xi^c(t+\mathrm{d}t)-x^c-\hat{f}^c_{(i)}\mathrm{d}t)}{\partial y^c}\right|_{y=x},
\end{aligned}\tag{69}$$

where I_j represents $j\times j$ identity matrix. Here, the last equality comes from the fact that for any nonsingular matrix $A(\alpha)$ with

a scalar parameter α, the following equality holds:

$$\frac{\mathrm{d}\left(\det A(\alpha)\right)}{\mathrm{d}\alpha} = \det A(\alpha)\,\mathrm{tr}\left\{A(\alpha)^{-1}\frac{\mathrm{d}A(\alpha)}{\mathrm{d}\alpha}\right\}.$$

From Eqs. (29), (68) and (69), we have

$$\begin{aligned}\frac{\partial J_{(i)}(x,t)}{\partial x^c}\,\mathrm{d}t &= E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\Bigg\{\left(\int_t^T \hat{V}_{(i)}\,\mathrm{d}\tau + \phi(\xi(T))\right)\\ &\times\left(-\frac{1}{2}\left(\mathrm{tr}\left\{\Xi^{c-1}\frac{\partial \Xi^c}{\partial x_1^c}\right\},\ldots,\mathrm{tr}\left\{\Xi^{c-1}\frac{\partial \Xi^c}{\partial x_{n_c}^c}\right\}\right)\mathrm{d}t+\right.\\ &\left.\mathrm{d}w(t)^\top h^{c\top}\Xi^{c-1}-\frac{1}{2}\,\mathrm{d}w(t)^\top h^{c\top}\frac{\partial \Xi^c(y,t)^{-1}h^c\,\mathrm{d}w(t)}{\partial y^c}\bigg|_{y=x}\right)\Bigg\}\\ &+ o(\mathrm{d}t),\end{aligned} \tag{70}$$

where $o(\cdot)$ denotes the landau notation.

Now, we consider the j th element of the last term in the expectation in Eq. (70). We have

$$\begin{aligned}&E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\Bigg\{-\frac{1}{2}\Bigg[\mathrm{d}w(t)^\top h^{c\top}\\ &\qquad\times\frac{\partial \Xi^c(y,t)^{-1}h^c(x,t)\,\mathrm{d}w(t)}{\partial y^c}\bigg|_{y=x}\Bigg]_j\Bigg\}\\ &=\frac{1}{2}E^{\mathcal{P}}\left\{\mathrm{d}w(t)^\top h^{c\top}\Xi^{c-1}\frac{\partial \Xi^c(x,t)}{\partial x_j^c}\Xi^{c-1}h^c\,\mathrm{d}w(t)\middle| x,t\right\}\\ &=\frac{1}{2}\mathrm{tr}\left\{h^{c\top}\Xi^{c-1}\frac{\partial \Xi^c(x,t)}{\partial x_j^c}\Xi^{c-1}h^c E^{\mathcal{P}}\left\{\mathrm{d}w(t)\,\mathrm{d}w(t)^\top\middle| x,t\right\}\right\}\\ &=\frac{1}{2}\,\mathrm{tr}\left\{\Xi^{c-1}\frac{\partial \Xi^c(x,t)}{\partial x_j^c}\Xi^{c-1}h^c Q h^{c\top}\,\mathrm{d}t\right\}\\ &=\frac{1}{2}\,\mathrm{tr}\left\{\Xi^{c-1}\frac{\partial \Xi^c(x,t)}{\partial x_j^c}\right\}\mathrm{d}t,\end{aligned} \tag{71}$$

where the first equality comes from the relation $\mathrm{d}A(\alpha)^{-1}/\,\mathrm{d}\alpha = -A(\alpha)^{-1}(\mathrm{d}A(\alpha)/\,\mathrm{d}\alpha)A(\alpha)^{-1}$, and the third equality comes from Eq. (2) with the fact that $\mathrm{d}w(t)$ is independent of (x,t). Since Eq. (71) implies that the first term and the last one in the last equality in Eq. (69) cancel out, Eq. (70) is reduced to

$$\begin{aligned}&\frac{\partial J_{(i)}(x,t)}{\partial x^c}^\top\,\mathrm{d}t = \Xi^c(x,t)^{-1}h^c(x,t)\\ &\times E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\left\{\left(\int_t^T \hat{V}_{(i)}(\xi,\tau)\,\mathrm{d}\tau+\phi(\xi(T))\right)\mathrm{d}w(t)\right\},\end{aligned}$$

which implies Eq. (30). Therefore, the i th suboptimal controller is obtained as

$$\begin{aligned}u_{(i)}^*(x,t)\,\mathrm{d}t &= -R(x,t)^{-1}g^c(x,t)^\top\Xi^c(x,t)^{-1}h^c(x,t)\\ &\times E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\left\{\left(\int_t^T \hat{V}_{(i)}(\xi,\tau)\,\mathrm{d}\tau+\phi(\xi(T))\right)\mathrm{d}w(t)\right\}.\end{aligned} \tag{72}$$

Since Eq. (72) coincides with Eq. (31), the other assertion of the theorem is proved.

ACKNOWLEDGMENT

This work was supported by JSPS Grant-in-Aid for Young Scientists (B) (No. 24760338). One of the author would like to thank Dr. N. Sakamoto and Dr. Y. Umemura for their helpful advice on numerical computation. One of the author would like to thank Dr. S. Thijssen for useful discussion.

REFERENCES

[1] H. J. Kushner, "Optimal stochastic control," *IRE Trans. Autom. Contr.*, vol. 7, pp. 120–122, 1962.
[2] M. Aoki, *Optimization of Stochastic Systems*. New York: Academic Press, 1967.
[3] W. M. Wonham, "On the separation theorem of stochastic control," *SIAM J. Contr.*, vol. 6, no. 2, pp. 312–326, 1968.
[4] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*. New York: Springer-Verlag, 1999.
[5] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed. New York: Springer-Verlag, 2006.
[6] H. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*. Springer-Verlag, 2009.
[7] N. E. Karoui and L. Mazliak, Eds., *Backward Stochastic Differential Equations*. Longman, 1997.
[8] N. Sakamoto and A. J. van der Schaft, "Analytical approximation methods for the stabilizing solution of the Hamilton-Jacobi equation," *IEEE Trans. Autom. Contr.*, vol. 53, no. 10, pp. 2335–2350, 2008.
[9] M. Kumar, S. Chakravorty, and J. L. Junkins, "Computational nonlinear stochastic control based on the Fokker-Planck-Kolmogorov equation," in *Proc. AIAA Guidance, Navigation and Control Conf.*, 2008, pp. AIAA–2008–6477.
[10] R. Beard, G. Saridis, and J. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.
[11] F.-Y. Wang and G. N. Saridis, "On successive approximation of optimal control of stochastic dynamical systems," in *Modeling Uncertainty*, M. Dror, P. Lécuyer, and F. Szidarovszky, Eds. Springer-Verlag, 2005, vol. 46, pp. 333–358.
[12] W. Li and E. Todorov, "Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system," *Int. J. Control*, vol. 80, no. 9, pp. 1439–1453, 2007.
[13] D. Jacobson and D. Mayne, *Differential Dynamic Programming*. New York: American Elsevier Publ. Co., Inc., 1970.
[14] L. Z. Liao and C. A. Shoemaker, "Convergence in unconstrained discrete-time differential dynamic programming," *IEEE Trans. Autom. Contr.*, vol. 36, no. 6, pp. 692–706, 1991.
[15] H. J. Kappen, "Path integrals and symmetry breaking for optimal control theory," *J. Statistical Mechanics: Theory and Experiment*, p. P11011, 2005.
[16] ——, "An introduction to stochastic control theory, path integrals and reinforcement learning," in *Proc. 9th Granada seminar on computational physics: Cooperative behavior in neural systems*, 2007, pp. 149–181.
[17] B. Øksendal, *Stochastic differential equations, An introduction with applications*, 5th ed. Berlin Heidelberg New York: Springer-Verlag, 1998.
[18] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *J. Machine Learning Research*, vol. 11, pp. 3153–3197, 2010.
[19] I. Gihman and A. Skorohod, *Stochastic Differential Equations*. Springer-Verlag, 1972.
[20] X. Mao and L. Szpruch, "Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients," *J. Computational and Applied Mathematics*, vol. 238, pp. 14–28, 2013.
[21] H. J. Kushner, *Stochastic Stability and Control*. Academic Press, 1967.
[22] K. Itô, "On a formula concerning stochastic differentials," *Nagoya Math. J.*, vol. 3, pp. 55–65, 1951.
[23] K. Yasunami, T. Matsubara, and K. Sugimoto, "An approximate method of iterative path integral stochastic optimal control based on local linear quadratic approximation," in *Proc. SICE 13th Annual Conference on Control Systems.*, 2013, pp. 7B3-4, (in Japanese).
[24] S. Satoh, "Iterative path integral method for nonlinear stochastic optimal control," in *Workshop on statistical physics of inference and control theory*, Granade, Spain, 2012, Video Lecture : http://videolectures.net/cyberstat2012_satoh_optimal_control/.

[25] S. Satoh and M. Saeki, "A solution method for stochastic optimal control with markov times based on path integrals," in *Proc. the 58th Annual Conf. of the Institute of Systems, Control and Information Engineers*, 2014, pp. (CD–ROM) 332–6, (in Japanese).
[26] ——, "A unified approach to nonlinear stochastic control based on path integral analysis," in *Proc. SICE International Symposium on Control Systems 2015*, 2015, pp. (USB) 712-1.
[27] R. Fujimoto and N. Sakamoto, "The stable manifold approach for optimal swing up and stabilization of an inverted pendulum with input saturation," in *Proc. 18th IFAC World Congress*, 2011, pp. 8046–8051.
[28] D. Hinrichsen and A. J. Pritchard, "Stochastic H^∞," *SIAM J. Control Optim.*, vol. 36, no. 5, pp. 1504–1538, 1998.
[29] W. Zhang, B. S. Chen, H. Tang, L. Sheng, and M. Gao, "Some remarks on general nonlinear stochastic H_∞ control with state, control, and disturbance-dependent noise," *IEEE Trans. Autom. Contr.*, vol. 59, no. 1, pp. 237–242, 2014.
[30] V. Gómez, S. Thijssen, A. Symington, S. Hailes, and H. J. Kappen, "Real-time stochastic optimal control for multi-agent quadrotor swarms," 2015, arXiv:1502.04548.

PLACE PHOTO HERE

Satoshi Satoh received the B.Sc., M.Sc. and Ph.D. degrees in engineering from Nagoya University, Japan, in 2005, 2007 and 2010, respectively. He is currently an assistant professor of Faculty of Engineering, Hiroshima University, Japan. He held a visiting research position at Radboud University Nijmegen, The Netherlands in 2011. He received the 10th Asian Control Conference The Best Paper Award in 2015, the SICE Young Author's Award in 2010, and the IEEE Robotics and Automation Society Japan Chapter Young Award in 2008. His research interests include nonlinear control theory and its application to stochastic systems and robotics.

PLACE PHOTO HERE

Hilbert J. Kappen received the Ph.D. degree in particle physics in 1987 from Rockefeller University, New York, NY, USA. From 1987 until 1989, he was a Scientist at the Philips Research Laboratories, Eindhoven, The Netherlands. Presently, he is Professor of physics at SNN University of Nijmegen, The Netherlands, conducting research in machine learning and computational neuroscience. He is author of over 150 publications.

PLACE PHOTO HERE

Masami Saeki received his B.S., M.S., and Ph.D. degrees from Kyoto University, Japan, in 1976, 1978, and 1982, respectively. Since 1992, he has been a Professor at the Faculty of Engineering, Hiroshima University. His research interests include robust control design and its application.