「『特集名』特集号」　　解　説

経路積分解析と統計的サンプリングを用いた非線形確率最適制御問題の解法

佐藤 訓志*

1. はじめに

近年の制御要求の多様化・高度化により，動的に変化する未知環境や人間との相互作用などの不確定な変動を考慮したシステムの自律化や最適な動作の実現が求められている．このような制御では，システムの非線形性や確率的不確かさを適切に表現できる非線形確率システムと，それらを考慮して評価関数の期待値を最小化する最適制御入力を与える確率最適制御が重要な役割を果たす．

非線形確率最適制御問題の最適フィードバック制御則の獲得は，一般に求解が困難な非線形２階偏微分方程式である確率ハミルトン・ヤコビ・ベルマン (SHJB) 方程式に帰着されるため [1,2]，線形システムの場合と異なり実用的に十分な解法がなかった．これに対し，確率論に基づく経路積分解析と統計学的サンプリング手法に基づく新しいアプローチである経路積分最適制御法が提案された [3]．この手法は，制御対象と評価関数のクラスに特別な条件を設け，この条件の下で SHJB 方程式にある対数変換を施すことで，SHJB 方程式を線形偏微分方程式に変換する．つぎに，ファインマン-カッツの公式 [2,4] により，この線形偏微分方程式の解の確率表現を得る．さらに，この解の確率表現に対して経路積分解析を行うことで，最適制御入力の計算に必要な解の偏導関数の確率表現も与える．得られた確率表現の計算には，統計力学の分野で提案されている様々なサンプリング手法が利用でき，精度と計算コストのバランスを考慮した数値計算が可能であり，並列化も容易である．

経路積分最適制御法 [3] は，近似や数値微分などを用いることなく SHJB 方程式の解とその偏導関数の確率表現が得られるという理論的な利点と，数値計算と親和性が高いという実装上の利点を持っている．しかしながら，この手法は制御対象と評価関数に特別な条件を必要とし，適用が大きく制限される問題があった．そこで筆者らは，逐次近似手法と線形偏微分方程式の知見に基づく反復型経路積分最適制御法を提案した [5]．この方法は，従来法の特別な条件を必要としないにも関わらず，従来法の利点を継承した反復解法となっている．さらに文献 [6] では，逐次近似における反復規則を拡張することで，無限評価区間や入力飽和をもつ確率最適制御や，確定外乱も同時に扱える確率 H_∞ 制御などにも適用可能であることが示され，より広いクラスの非線形確率制御問題に対する経路積分解析に基づく統一的な解法を与えている．

本稿では，まず非線形確率システムに対する確率最適制御問題の定式化について説明する．つぎに，この問題の解法として経路積分最適制御法 [3] を紹介する．さいごに，文献 [6,5] に基づき筆者らが提案した経路積分最適制御法のいくつかの拡張を紹介する．本稿は確率システムの初学者を対象としたため，厳密性はある程度犠牲にしてできる限り測度論や確率論の話は表に出さないようにした．しかし，より進んだ理解のためにはこれらの知識は必須となる．確率システムに関する基礎事項や詳細は文献 [7–11] などを，一般の確率論・確率微分方程式論については文献 [12,13,4,14] などを参照頂きたい．

2. 有限時間確率最適制御問題と SHJB 方程式

本節では，非線形確率最適制御問題の定式化と付随する SHJB 方程式を説明する．本稿では，次式の伊藤型確率微分方程式で表される連続時間確率システムを考える．

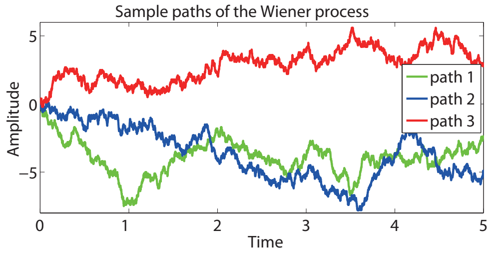
$$\mathrm{d}x = f(x,t)\mathrm{d}t + g(x,t)u\mathrm{d}t + h(x,t)\mathrm{d}w,\ x(0) = x_0 \quad (1)$$

ここで，$x(t)\in\mathbb{R}^n$ はシステムの状態，x_0 は状態の初期値を表し，$u(t)\in\mathbb{R}^m$ は制御入力を表す．関数 $f:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}^n$, $g:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}^{n\times m}$, $h:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}^{n\times r}$ は十分滑らかな関数とし，本節では時間区間 $t\in[0,T]$ 上の確率最適制御問題を考えるため，終端時刻 T までの (1) 式の解の存在と一意性を仮定しておく．そのための十分条件についての詳細は文献 [12,4] などを参照頂きたい．

さて，確率微分方程式 (1) が常微分方程式と本質的に異なる点は，右辺第三項の存在である．ここで $w(t)\in\mathbb{R}^r$ は標準ウィーナ過程またはブラウン運動過程と呼ばれる確率過程である．第１図に標準ウィーナ過

* 大阪大学大学院　工学研究科

Key Words: stochastic optimal control, path integral analysis, stochastic Hamilton-Jacobi-Bellman equation



第 1 図 標準ウィーナ過程の見本過程の例

程の三つの見本過程を示した．ウィーナ過程は実現する見本過程毎に異なる不規則な挙動を示す確率過程であり，システムに存在するノイズ源の役割を持つ．(1) 式中の $h(x,t)$ はこのようなノイズがどのようにシステムに影響を及ぼすかを表す関数であり，本稿ではノイズポートと呼ぶ．このように，確率過程がシステムの動特性に本質的に内在することが，本質的な違いである．もう少し具体的に，標準ウィーナ過程の統計的性質を示す．

- 確率 1 で $w(0)=0$.
- (正規増分性) : 任意の $s,t,\ 0\le s<t$ に対して，$w(t)-w(s)$ は $\mathcal{N}(0,(t-s)I_r)$ に従う．ただし，$\mathcal{N}(m,Q)$ は平均 $m\in\mathbb{R}^r$, 分散共分散行列 $Q\in\mathbb{R}^{r\times r}$ の正規分布を表し，I_r は $r\times r$ 単位行列を表す．
- (独立増分性) : 任意の $s,t,\ 0\le s<t$ に対して，$w(t)-w(s)$ は $w(\tau),\ \tau\in[0,s]$ と独立である．
- 見本路は確率 1 で二乗平均収束の意味で連続である．ただし，確率変数 $X_{(n)}$ が X に二乗平均収束するとは，期待値演算 $E\{\cdot\}$ を用いて $\lim_{n\to\infty}E\{\|X_{(n)}-X\|^2\}=0$ を満たすことである．

本稿では一般性を失うことなく，標準ウィーナ過程を考えるが，任意の分散共分散行列 $Q(x,t)\in\mathbb{R}^{r\times r}$ を扱う際には，$h(x,t)$ を $h(x,t)Q(x,t)^{1/2}$ で置き換えればよい．

それでは，確率システム (1) に対して有限時間非線形確率最適制御問題を定式化していこう．まず，評価関数を定義するために次式の汎関数を定義する．

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

ただし，$\phi:\mathbb{R}^n\to\mathbb{R}$, $V:\mathbb{R}^n\times[0,T]\to\mathbb{R}$ はそれぞれ終端コストと瞬時コストを表す十分滑らかな非負関数であり，$R(x,t)\in\mathbb{R}^{m\times m}$ は入力コストに関する重み行列である．また，表記 $(\cdot)_{t:T}$ は時間区間 $[t,T]$ 上の関数を表す．本稿では，最適制御入力を陽に導出するために，入力コストは $R(x,t)$ に関する入力の二次形式に限定し，$R(x,t)$ は任意の $x\in\mathbb{R}^n$, $t\in[0,T]$ において正定対称行列と仮定する．有限時間非線形確率最適制御問題の目的は，システム (1) に対して評価関数 $\mathit{\Gamma}(x_0,u_{0:T},0)$ を最小にする最適制御入力 $u(t)=u^*(x(t),t)$ を求めることである．

ここで，価値関数を $J(x,t):=\min_{u_{t:T}}\mathit{\Gamma}(x,u_{t:T},t)$ と定義する．価値関数は，与えられた時刻 t における状態 x を初期状態として (1) 式から生成される解過程に関して，達成し得る最小のコストを返す関数である．価値関数の存在は自明ではないが，価値関数 $J(x,t)$ が存在し，$C^{2,1}$ 級関数であると仮定すると，$J(x,t)$ が満たすべき方程式である SHJB 方程式が次式で与えられる [1,2].

$$-\frac{\partial J}{\partial t}=\min_u\left(V+\frac{1}{2}u^{\mathrm{T}}Ru+\frac{\partial J}{\partial x}(f+gu)\right.\\ \left.+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 J}{\partial x^2}hh^{\mathrm{T}}\right\}\right),\quad J(x,T)=\phi(x) \quad (3)$$

評価関数 (2) において，入力コストは正定対称行列に関する入力の二次形式であるため，(3) 式右辺を最小とする入力は一意に決定でき，最適制御入力 u^* は次式となる．

$$u^*(x,t)=-R(x,t)^{-1}g(x,t)^{\mathrm{T}}\frac{\partial J(x,t)}{\partial x}^{\mathrm{T}} \quad (4)$$

(4) 式を (3) 式に代入して入力項を消去することで，SHJB 方程式 (3) は次式のようになる．

$$\frac{\partial J}{\partial t}+V-\frac{1}{2}\frac{\partial J}{\partial x}gR^{-1}g^{\mathrm{T}}\frac{\partial J}{\partial x}^{\mathrm{T}}+\frac{\partial J}{\partial x}f\\ +\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 J}{\partial x^2}hh^{\mathrm{T}}\right\}=0,\quad J(x,T)=\phi(x) \quad (5)$$

結局，本稿で扱う確率最適制御問題は SHJB 方程式 (5) の解 $J(x,t)$ を求める問題に帰着される．$J(x,t)$ が求まれば，最適制御入力が (4) 式で得られる．

SHJB 方程式 (5) と，確定システムにおけるハミルトン・ヤコビ・ベルマン (HJB) 方程式 [15] との違いは，確率解析特有の項である (5) 式左辺第五項の有無であり，HJB 方程式は J に関する非線形 1 階，SHJB 方程式は非線形 2 階の偏微分方程式である．このような項が現れることを直感的に説明する．ウィーナ過程の正規増分性より，ウィーナ過程の微小変化分に関して次式が成り立つ．

$$E\{\mathrm{d}w(t)\}=0,\quad \forall t\ge 0, \quad (6)$$
$$E\{\mathrm{d}w(t)\mathrm{d}w(t)^{\mathrm{T}}\}=\mathrm{d}t\ I_r,\quad \forall t\ge 0 \quad (7)$$

このことから，確率解析の要点として $\mathrm{d}w$ の二乗のオーダーが $\mathrm{d}t$ のオーダーとなるため，テーラー展開において確定システムの解析では考慮する必要がなかった 2 階微分の項の一部を適切に取り扱う必要がある．この要点を念頭におき，動的計画法で現れる次の計算を見てみよう．

$$
\begin{aligned}
&E\{J(x+\mathrm{d}x,t+\mathrm{d}t)-J(x,t)|\ (x,t)\}\\
=&E\left\{\frac{\partial J}{\partial t}\mathrm{d}t+\frac{\partial J}{\partial x}\mathrm{d}x+\frac{1}{2}\mathrm{d}x^{\mathrm{T}}\frac{\partial^2 J}{\partial x^2}\mathrm{d}x\middle|(x,t)\right\}+o(\|\mathrm{d}x\|^2)\\
=&\frac{\partial J}{\partial t}\mathrm{d}t+\frac{\partial J}{\partial x}(f+gu)\mathrm{d}t+\frac{\partial J}{\partial x}hE\{\mathrm{d}w|\ (x,t)\}\\
&+\frac{1}{2}E\left\{\mathrm{d}w^{\mathrm{T}}h^{\mathrm{T}}\frac{\partial^2 J}{\partial x^2}h\mathrm{d}w\middle|(x,t)\right\}+o(|\mathrm{d}t|)
\end{aligned}
\tag{8}
$$

ただし，$o(a)$ は $\lim_{a\to 0}o(a)/a=0$ を意味する．(8) 式右辺第三項は，ウイーナ過程の独立増分性から $\mathrm{d}w(t)$ は $x(t)$ とは独立であり，これと (6) 式から零となる．つぎに，(8) 式右辺第四項は，次式のように計算できる．

$$
\frac{1}{2}\mathrm{tr}\left\{h^{\mathrm{T}}\frac{\partial^2 J}{\partial x^2}hE\left\{\mathrm{d}w\mathrm{d}w^{\mathrm{T}}\middle|(x,t)\right\}\right\}=\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 J}{\partial x^2}hh^{\mathrm{T}}\right\}\mathrm{d}t
$$

直感的にはこの様にして，(5) 式左辺第五項が現れる．

3. 経路積分最適制御法

本節では，経路積分最適制御法 [3] を紹介する．以降ではノイズポート h に関する非退化性を導入するため，文献 [16] の議論に基づき，(1) 式の代わりにノイズにより直接駆動されるサブシステムと，連成により間接駆動されるものとに分けた次式のシステムを考える．

$$
\begin{aligned}
\begin{pmatrix}\mathrm{d}x^u\\ \mathrm{d}x^c\end{pmatrix}&=\begin{pmatrix}f^u(x,t)\\ f^c(x,t)\end{pmatrix}\mathrm{d}t+\begin{pmatrix}0_{(n-n_c)\times m}\\ g^c(x,t)\end{pmatrix}u\mathrm{d}t\\
&\quad+\begin{pmatrix}0_{(n-n_c)\times r}\\ h^c(x,t)\end{pmatrix}\mathrm{d}w,\qquad x(0)=x_0
\end{aligned}
\tag{9}
$$

ただし，$x:=(x^{u\mathrm{T}},x^{c\mathrm{T}})^{\mathrm{T}}\in\mathbb{R}^n$ であり，$x^c\in\mathbb{R}^{n_c}$ はノイズにより直接駆動されるサブシステムの状態，$x^u\in\mathbb{R}^{n-n_c}$ は間接駆動されるサブシステムの状態をそれぞれ表す．$0_{i\times j}$ は $i\times j$ の零行列を表すものとする．ここで，ノイズポート h^c に対して非退化の仮定をおく．

（仮定 1） $h^c(x,t)h^c(x,t)^{\mathrm{T}}$ は，任意の $x\in\mathbb{R}^n$, $t\in[0,T]$ に対して正定行列となる．

SHJB 方程式 (5) に，正定数 λ を任意の設計パラメータとした次の対数変換

$$
J(x,t)=-\lambda\log\psi(x,t) \tag{10}
$$

を施すことで，ψ に関する方程式として次式を得る．

$$
\begin{aligned}
&\frac{\partial\psi}{\partial t}-\frac{V}{\lambda}\psi+\frac{\partial\psi}{\partial x}f+\frac{\lambda}{2\psi}\frac{\partial\psi}{\partial x}gR^{-1}g^{\mathrm{T}}\frac{\partial\psi}{\partial x}^{\mathrm{T}}\\
&-\frac{1}{2\psi}\frac{\partial\psi}{\partial x}hh^{\mathrm{T}}\frac{\partial\psi}{\partial x}^{\mathrm{T}}+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2\psi}{\partial x^2}hh^{\mathrm{T}}\right\}=0\ ,\\
&\psi(x,T)=\exp\left(-\frac{\phi(x)}{\lambda}\right)
\end{aligned}
\tag{11}
$$

ここで，経路積分最適制御法では特別な条件を仮定する．

（仮定 2） 任意の $x\in\mathbb{R}^n$, $t\in[0,T]$ に対して，次式が成立する．

$$
\lambda g^c(x,t)R(x,t)^{-1}g^c(x,t)^{\mathrm{T}}=h^c(x,t)h^c(x,t)^{\mathrm{T}} \tag{12}
$$

【注意 **1**】 ノイズポート h^c を分散共分散行列 Q を陽に表した $h^cQ^{1/2}$ に置き換えると，特別な条件 (12) は $\lambda g^cR^{-1}g^{c\mathrm{T}}=h^cQh^{c\mathrm{T}}$ となる．この条件は，ノイズの共分散が大きい程，最適制御における入力重みを小さくする必要があるという直感的な解釈ができ，この反比例関係の比例定数が λ であることを要請している． □

仮定 2 により，(11) 式の非線形項である左辺第四項と第五項が相殺され，次の線形 2 階偏微分方程式を得る．

$$
\begin{aligned}
&\frac{\partial\psi}{\partial t}-\frac{V}{\lambda}\psi+\frac{\partial\psi}{\partial x}f+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2\psi}{\partial x^2}hh^{\mathrm{T}}\right\}=0,\\
&\psi(x,T)=\exp\left(-\frac{\phi(x)}{\lambda}\right)
\end{aligned}
\tag{13}
$$

(13) 式は線形であるだけでなく，確率論でよく知られるコルモゴロフの後向き方程式と同じ構造を持ち，付録 1. に示すファインマン-カッツの公式 [2,4] から解 ψ の確率表現が次式のように得られる．

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

ここで，$p(\xi_{t:T}|x,t)$ は，$\xi(t)=x$ を初期状態とした時間区間 $[t,T]$ での (1) 式における自由運動，つまり

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

からある解過程 $\xi_{t:T}$ が実現される確率を表し，$E^{p(\xi_{t:T}|x,t)}\{\cdot\}$ は，この実現確率の下での全ての解過程に関する期待値演算を表す．SHJB 方程式の変換で得られた (13) 式の解 $\psi(x,t)$ が，(14) 式より解過程に沿った積分の形で与えられることが経路積分最適制御法の名の由来である．

ここで，解過程の経路コストに関する重み付けをした新しい確率を次式のように定義する．

$$
\begin{aligned}
&q(\xi_{t:T}|x,t):=\\
&\frac{p(\xi_{t:T})}{\psi(x,t)}\exp\left(-\frac{1}{\lambda}\left(\phi(\xi(T))+\int_t^T V(\xi,\tau)\mathrm{d}\tau\right)\right)
\end{aligned}
\tag{16}
$$

(13) 式の解 $\psi(x,t)$ の確率表現 (14) に対して経路積分解析を行うことで，最適制御入力 (4) を計算する際に必要な解の偏導関数，すなわち $\partial\log\psi(x,t)/\partial x^c$ の確率表現も与えられる．これは，偏導関数の計算に解 ψ の数値微を用いる必要がないという大きな利点となる．経路積分解析の実際の計算は文献 [3,5] に委ね，結果のみを示す．

$$\frac{\partial \log\psi(x,t)}{\partial x^c}^{\mathrm{T}}\mathrm{d}t = (h^c h^{c\mathrm{T}})^{-1}h^c(x,t)E^{q(\xi_{t:T}|x,t)}\{\mathrm{d}w(t)\} \tag{17}$$

よって (4), (17) 式と仮定 2 より，経路積分最適制御法において，最適制御入力が次式のように求まる．

$$\begin{aligned} &g^c(x,t)u^*(x,t)\mathrm{d}t \\ &\quad = \lambda g^c(x,t)R(x,t)^{-1}g^c(x,t)^{\mathrm{T}}\frac{\partial \log\psi(x,t)}{\partial x^c}^{\mathrm{T}}\mathrm{d}t \\ &\quad = h^c(x,t)E^{q(\xi_{t:T}|x,t)}\{\mathrm{d}w(t)\} \end{aligned} \tag{18}$$

確率表現における期待値を解析的に得ることは困難であるが，様々なサンプリング手法を利用することで，精度と計算コストのバランスを考慮した数値計算が可能である．ここでは最も基本的なモンテカルロサンプリングによる具体的な計算を示す．$\xi(t)=x$ を満たし，時間区間 $[t,T]$ での自由運動 (15) から生成された解過程 $\xi_{t:T}$ に関する経路コストを $S^k(\xi_{t:T}) := \phi(\xi^k(T)) + \int_t^T V(\xi^k,\tau)\mathrm{d}\tau$ とおき，k を k 本目のサンプル，サンプル数を N とする．このとき，N 本のサンプルから計算される $\psi(x,t)$ と $E^{q(\xi_{t:T}|x,t)}\{\mathrm{d}w(t)\}$ の近似は次式で与えられる．

$$\psi^N(x,t) = \sum_{k=1}^{N}\exp\left(-S^k(\xi_{t:T})/\lambda\right)/N \tag{19}$$

$$E^{q(\xi_{t:T}|x,t)}\{\mathrm{d}w(t)\}^N = \frac{\sum_{k=1}^{N}\mathrm{d}w^k(t)\exp\left(-S^k(\xi_{t:T})/\lambda\right)}{\sum_{k=1}^{N}\exp(-S^k(\xi_{t:T})/\lambda)}$$

サンプリングを利用した数値計算法の特徴として，アルゴリズムは比較的簡単であるが計算量が多くなる．しかし，並列計算が容易であり，マルチコア CPU, GPGPU など計算機の性能向上の恩恵を受けやすい．文献 [3,17] では，ランジュヴァンサンプリング，重点サンプリング，ラプラスサンプリングなどを用いた計算法も述べられており，文献 [18] では経路積分最適制御によるクアッドコプタ 10 機の実時間編隊制御の達成が報告されている．

経路積分最適制御法は，仮定 2 を満たす非線形確率最適制御問題を，制御対象 (9) の自由運動 (15) から生成される解過程のサンプルで解くことができる．特に，最適制御入力が，サンプルに関する特別な重み付き確率 (16) の下でのノイズベクトルの期待値 (18) で与えられるという結果は，従来法にはない知見を与える．また，最適制御入力の実際の計算は，各種のサンプリング法と並列計算により効率化が可能であり，数値計算と親和性も高い．

4.　反復型経路積分最適制御法

3. の経路積分最適制御法は，特別な条件により適用できる制御対象と評価関数のクラスが制限されていた．そこで本節では，筆者らが提案した反復型経路積分確率最適制御法 [5] を紹介する．この方法は，逐次近似手法を新たに導入することで，従来法の利点を継承しながら，特別な条件 (12) や対数変換 (10) を必要としない．反復法となるため計算コストが増えるが，従来法で扱えなかったより広いクラスの確率最適制御問題に適用できる．以降では，まず提案法の概要と反復規則について述べ，提案法の収束に関する結果を示す．反復型経路積分最適制御法における i 回目の反復規則は，解 $J_{(i)}$ に関する次式の線形偏微分方程式のコーシー問題として与えられる．

$$\begin{aligned} &\frac{\partial J_{(i)}}{\partial t} + \frac{\partial J_{(i)}}{\partial x}\hat{f}_{(i)} + \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 J_{(i)}}{\partial x^2}hh^{\mathrm{T}}\right\} + \hat{V}_{(i)} = 0, \\ &J_{(i)}(x,T) = \phi(x) \end{aligned} \tag{20}$$

ここで，$\hat{f}_{(i)} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$, $\hat{V}_{(i)} : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ は，$i-1$ 回目の解 $J_{(i-1)}$ を用いて次式で定義される．

$$\hat{f}_{(i)} = \begin{pmatrix}\hat{f}^u_{(i)} \\ \hat{f}^c_{(i)}\end{pmatrix} := f - gR^{-1}g^{\mathrm{T}}\frac{\partial J_{(i-1)}}{\partial x}^{\mathrm{T}}, \tag{21}$$

$$\hat{V}_{(i)} := V + \frac{1}{2}\frac{\partial J_{(i-1)}}{\partial x}gR^{-1}g^{\mathrm{T}}\frac{\partial J_{(i-1)}}{\partial x}^{\mathrm{T}} \tag{22}$$

本手法では，特別な条件 (12) を用いる代わりに逐次近似手法を導入し，また付録 1. よりファインマン-カッツの公式がコルモゴロフ方程式 (13) よりも広いクラスのコーシー問題に適用できることを利用し，反復規則 (20) を (13) 式とは異なる構造の線形偏微分方程式のコーシー問題として定式化している．そのため，従来法の対数変換 (10) を用いる必要もなくなり，ψ ではなく価値関数 J を直接求める反復解法となっている．このことは，後に 5. で述べる更なる拡張を可能にするという利点ももたらす．ここで，反復規則 (20) の解とその偏導関数に関する以下の結果を示す．

【定理 **1**】（[5]） (9) 式の制御対象と，(2) 式で定義される評価関数 $\Gamma(x_0,u_{0:T},0)$ を考える．仮定 1 が成立し，反復規則 (20) において，$C^{2,1}$ 級関数かつ任意の $x\in\mathbb{R}^n$, $t\in[0,T]$ で $\|\partial J_{(0)}(x,t)/\partial x^c\|^2 < \infty$ を満たす初期関数 $J_{(0)}$ が与えられているものとする．

このとき，任意の $x\in\mathbb{R}^n$, $t\in[0,T]$ における反復規則 (20) の i 回目の解 $J_{(i)}$ の確率表現は次式で与えられる．

$$\begin{aligned} J_{(i)}(x,t) &= E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\left\{\hat{S}_{(i)}(\xi_{t:T})\right\}, \\ \hat{S}_{(i)}(\xi_{t:T}) &:= \phi(\xi(T)) + \int_t^T \hat{V}_{(i)}(\xi(\tau),\tau)\mathrm{d}\tau \end{aligned} \tag{23}$$

ここで，(15) 式の f^u, f^c を (21) 式の $\hat{f}^u_{(i)}, \hat{f}^c_{(i)}$ でそれぞれ置き換えたダイナミクスを i 回目のサンプル生成ダイナミクスと呼ぶ．このとき，(23) 式中の $\hat{p}_{(i)}(\xi_{t:T}|x,t)$ は，初期状態 $\xi(t)=x$ の下で i 回目のサンプル生成ダイナミクスからある解過程 $\xi_{t:T}$ が実現される確率を表す．

さらに，偏導関数 $\partial J_{(i)}/\partial x^c$ の確率表現は，

$$\frac{\partial J_{(i)}(x,t)}{\partial x^c}^{\mathrm{T}}\mathrm{d}t = (h^c h^{c\mathrm{T}})^{-1} h^c E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\{\hat{S}_{(i)}(\xi_{t:T})\mathrm{d}w(t)\} \quad (24)$$

で与えられる．■

定理1の(24)式より，i回目の反復における準最適制御入力は次式で与えられる．

$$u^*_{(i)}(x,t)\mathrm{d}t = -R(x,t)^{-1}g^c(x,t)^{\mathrm{T}}(h^c h^{c\mathrm{T}})^{-1}h^c(x,t) \times E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\{\hat{S}_{(i)}(\xi_{t:T})\mathrm{d}w(t)\} \quad (25)$$

【注意 **2**】 付録1.のファインマン-カッツの公式から，従来法と提案法のコーシー問題(13), (20)式の構造を比較すると，従来法は(A1)式において$\alpha = V/\lambda$, $\beta = 0$に相当し，提案法は$\alpha = 0$, $\beta = \hat{V}_{(i)}$に相当する．このようにコーシー問題の構造が異なるため，(24)式では従来法の結果(17)における重み付け確率(16)が現れないなど，従来法からは得られない結果を定理1は与えている．□

つぎに，反復型経路積分最適制御法における逐次反復の収束に関する結果を示す．まず，SHJB方程式との誤差関数を導入するため，次式の表記を準備する．

$$H(\eta,\zeta) := \frac{\partial \eta}{\partial t} + \frac{\partial \eta}{\partial x}\left(f - gR^{-1}g^{\mathrm{T}}\frac{\partial \zeta}{\partial x}^{\mathrm{T}}\right) + \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 \eta}{\partial x^2}hh^{\mathrm{T}}\right\} + V + \frac{1}{2}\frac{\partial \zeta}{\partial x}gR^{-1}g^{\mathrm{T}}\frac{\partial \zeta}{\partial x}^{\mathrm{T}} \quad (26)$$

(26)式の表記を用いると，反復規則(20)は$H(J_{(i)}, J_{(i-1)}) = 0$, $J_{(i)}(x,T) = \phi(x)$と書け，$H(J^*, J^*) = 0$, $J^*(x,T) = \phi(x)$を満たす反復規則の極限関数J^*が得られれば，J^*はSHJB方程式(5)の解となる．ここで，i回目の反復で得られた解$J_{(i)}$の(x,t)における誤差関数を，

$$d_{(i)}(x,t) := H(J_{(i)}, J_{(i)})(x,t) \quad (27)$$

で定義する．このとき，以下の結果が得られている．

【定理 **2**】（[5]） 定理1の仮定がすべて満たされているものとする．反復規則(20)の下でi回目の反復後，任意の$x \in \mathbb{R}^n$, $t \in [0,T]$において$J_{(i)}(x,t)$と$\partial J_{(i)}(x,t)/\partial x$が利用可能であるとする．このとき，反復規則(20)で得られる解$J_{(i)}(x,t)$は，任意の$x \in \mathbb{R}^n$, $t \in [0,T]$において，$i \to \infty$で各点収束する．さらに，(27)式で定義される誤差関数$d_{(i)}(x,t)$は，任意の$x \in \mathbb{R}^n$, $t \in [0,T]$において，$i \to \infty$で確率1で零に収束する．■

さいごに各反復における具体的な計算式を示す．i回目の反復のサンプル数を$N_{(i)} > 0$，経路コストを$\hat{S}^k_{(i)}(\xi_{t:T}) := \phi(\xi^k(T)) + \int_t^T \hat{V}_{(i)}(\xi^k,\tau)\mathrm{d}\tau$としたときの準最適解$J^{N_{(i)}}_{(i)}$と準最適制御入力$u^{*N_{(i)}}_{(i)}$は次式となる．

$$J^{N_{(i)}}_{(i)}(x,t) = \sum_{k=1}^{N_{(i)}} \hat{S}^k_{(i)}(\xi_{t:T})/N_{(i)}, \quad (28)$$

$$u^{*N_{(i)}}_{(i)}(x,t)\mathrm{d}t = -R(x,t)^{-1}g^c(x,t)^{\mathrm{T}}(h^c h^{c\mathrm{T}})^{-1} \times h^c(x,t)\sum_{k=1}^{N_{(i)}} \hat{S}^k_{(i)}(\xi_{t:T})\mathrm{d}w^k(t)/N_{(i)} \quad (29)$$

定理1の確率表現(23), (25)は，(28), (29)式においてサンプル数$N_{(i)}$を増やすことで，任意の精度で数値計算が可能であるものの，各反復計算で得られる解とその偏導関数には誤差が伴う．本稿では省略したが，文献[5]では定理2の収束性解析において，各反復計算の誤差も陽に考慮した結果が示されている．また，実際の各反復計算において全ての$x \in \mathbb{R}^n$, $t \in [0,T]$における$J_{(i)}(x,t)$と$\partial J_{(i)}(x,t)/\partial x$の値を計算することは不可能であるため，文献[5]ではこのことを考慮して，複数の反復結果を用いた多項式フィッティングや，価値関数の局所二次近似を用いた具体的なアルゴリズムと数値例が示されている．

5. 反復規則の拡張による非線形確率制御の統一的解法

注意2より，ファインマン-カッツの公式は(13), (20)式よりも広いクラスのコーシー問題に適用可能であることから，文献[6]では以下のような定理1の拡張を行った．

【定理 **3**】（[6]） 付録1.のコーシー問題の解の偏導関数$\partial v/\partial x$の確率表現は次式で与えられる．

$$\frac{\partial v(x,t)}{\partial x}^{\mathrm{T}}\mathrm{d}t = (hh^{\mathrm{T}})^{-1}hE^{p(\xi_{t:T}|x,t)}\{S_{FK}(\xi_{t:T})\mathrm{d}w(t)\}$$

■

ファインマン-カッツの公式と定理3を用いて，反復規則を拡張することで，有限時間確率最適制御問題だけでなく，無限評価区間や入力飽和を持つ確率最適制御や，確定外乱も同時に扱える確率H_∞制御などのより広いクラスの非線形確率制御問題に対する経路積分解析に基づく統一的な解法を与えることができる[6]．

ここでは入力飽和付き確率最適制御問題への適用[6]について簡単に紹介する．この結果は，確定システムにおける結果[19]を確率システムへと拡張したものであり，$\mathrm{sat}(\cdot)$を飽和関数とし，評価関数(2)の入力重み行列Rを対角行列に限定すると，次式のSHJB方程式を得る．

$$
\frac{\partial J}{\partial t}+V-\frac{1}{2}\frac{\partial J}{\partial x}gR^{-1}g^{\mathrm{T}}\frac{\partial J}{\partial x}^{\mathrm{T}}+\frac{\partial J}{\partial x}f
+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 J}{\partial x^2}hh^{\mathrm{T}}\right\}+\Psi_{sat}(J)=0,\quad J(x,T)=\phi(x),
$$

$$
\Psi_{sat}(J):=\frac{1}{2}\left(\mathrm{sat}\left(-R^{-1}g^{\mathrm{T}}\frac{\partial J}{\partial x}^{\mathrm{T}}\right)+R^{-1}g^{\mathrm{T}}\frac{\partial J}{\partial x}^{\mathrm{T}}\right)^{\mathrm{T}}
\times R\left(\mathrm{sat}\left(-R^{-1}g^{\mathrm{T}}\frac{\partial J}{\partial x}^{\mathrm{T}}\right)+R^{-1}g^{\mathrm{T}}\frac{\partial J}{\partial x}^{\mathrm{T}}\right) \tag{30}
$$

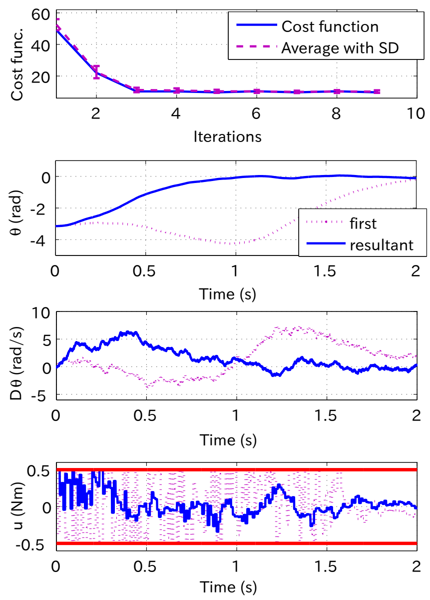
4. の逐次反復において，H, $\hat{V}_{(i)}$, $\hat{S}_{(i)}$, $u^*_{(i)}$ をそれぞれ

$$
\begin{aligned}
&H_{sat}(J_{(i)},J_{(i-1)}):=H(J_{(i)},J_{(i-1)})+\Psi_{sat}(J_{(i-1)}),\\
&\hat{V}_{sat(i)}:=\hat{V}_{(i)}+\Psi_{sat}(J_{(i-1)}),\\
&\hat{S}_{sat(i)}(\xi_{t:T})\ :=\int_t^T\hat{V}_{sat(i)}(\xi,\tau)\mathrm{d}\tau+\phi(\xi(T)),\\
&u^*_{(i)}\mathrm{d}t=\mathrm{sat}\Big(-R^{-1}g^{c\mathrm{T}}(h^c h^{c\mathrm{T}})^{-1}h^c\\
&\qquad\times E^{\hat{p}_{(i)}(\xi_{t:T}|x,t)}\{\hat{S}_{sat(i)}(\xi_{t:T})\mathrm{d}w(t)\}\Big)
\end{aligned}
$$

に置き換えることで，入力飽和付き確率最適制御問題の解法を与えることができる．さいごに，1 リンクマニピュレータに適用した数値例を簡単に紹介する．これは文献 [5] の数値例に新たに入力飽和を考慮したものであり，制御対象の詳細については省略する．鉛直上向きを原点としたリンク角度を θ として状態を $x:=(\theta,\dot{\theta})^{\mathrm{T}}$，関節トルクを制御入力 u とする．初期状態を鉛直下向きの $x_0=(-\pi,0)^{\mathrm{T}}$ として，入力制約 $|u|<0.50$ を満たしながら鉛直上向きに振り上げることを制御目標とする．評価関数は，(2) 式において $T=2.0$s, $\phi=1/2x_T^{\mathrm{T}}\mathrm{diag}\{10,0.50\}x_T$, $V=1/2x^{\mathrm{T}}\mathrm{diag}\{5.0,0.30\}x$, $R=0.80$ とした．初期関数は入力飽和を無視し，原点まわりで線形近似して得られる LQG 問題の価値関数とし，$N_{(i)}=30000, i=1,\dots,9$ とした．第 2 図の 1 番上の図は，9 回の反復結果における評価関数の推移であり，鎖線は 50 回の独立した試行の平均値と標準偏差を表し，実線は 50 回の中のある試行の結果を示している．2 番目から 4 番目の図は，それぞれこの試行における θ, $\dot{\theta}$, u の時間応答であり，点線で 1 回目の反復結果，実線で 9 回目の反復結果を示している．第 2 図より，反復が概ね収束し，得られた応答は入力制約を満たし反復前よりも性能が向上していることが分かる．

6. おわりに

本稿では，非線形確率システムに対する確率最適制御問題について説明し，有用な解法である経路積分最適制御法 [3] を紹介した．つぎに，この方法が持つ特別な条件を排除した反復型経路積分最適制御法 [5] と，より広い確率制御問題に対する解法を与える拡張 [6] について述べた．さいごに，本研究の一部は JSPS 科研費 15K18089 の助成を受けたものであり，ここに謝意を表します．



第 2 図　入力飽和付き確率最適制御のシミュレーション結果

付　　録

付録 1．ファインマン-カッツの公式 [2]

確率微分方程式 (1) と，次式の線形偏微分方程式のコーシー問題を考える．

$$
\frac{\partial v}{\partial t}+\frac{\partial v}{\partial x}f+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 v}{\partial x^2}hh^{\mathrm{T}}\right\}-\alpha(x,t)v+\beta(x,t)=0,
$$
$$
v(x,T)=\phi(x) \tag{A1}
$$

ここで，$\beta:\mathbb{R}^n\times[0,T]$ は連続関数，$\alpha:\mathbb{R}^n\times[0,T]$ は非負連続関数とする．(A1) の解 $v(x,t)$ が存在し $C^{2,1}$ 級関数であるとき, $v(x,t)$ の確率表現は次式で与えられる．

$$
\begin{aligned}
v(x,t)&=E^{p(\xi_{t:T}|x,t)}\{S_{FK}(\xi_{t:T})\},\\
S_{FK}(\xi_{t:T})&:=\int_t^T e^{-\int_t^\tau\alpha(\xi,s)\mathrm{d}s}\beta(\xi,\tau)\mathrm{d}\tau\\
&\quad+e^{-\int_t^T\alpha(\xi,s)\mathrm{d}s}\phi(\xi(T))
\end{aligned}
$$

参 考 文 献

[1] J. Yong and X. Y. Zhou : *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer-Verlag, New York (1999)

[2] H. Pham : *Continuous-time Stochastic Control and Optimization with Financial Applications*, Springer-Verlag (2009)

[3] H. J. Kappen : Path integrals and symmetry breaking for optimal control theory; *J. Statistical Mechanics: Theory and Experiment*, p. P11011 (2005)

[4] B. Øksendal : *Stochastic differential equations, An introduction with applications*, Springer-Verlag, Berlin Heidelberg New York, 5th edition (1998)

[5] S. Satoh, H. J. Kappen and M. Saeki : An iterative method for nonlinear stochastic optimal control based on path integrals; *IEEE Trans. Autom. Contr.*, Vol. 62, No. 1, pp. 262–276 (2017)

[6] S. Satoh and M. Saeki : A unified approach to nonlinear stochastic control based on path integral analysis; *Proc. SICE International Symposium on Control Systems 2015*, pp. (USB) 712-1 (2015)

[7] 砂原 (編)：確率システム理論 I,II, 朝倉書店 (1981)

[8] 大住 (システム制御情報学会 (編))：確率システム入門, システム制御情報ライブラリー 24, 朝倉書店 (2002)

[9] 山本：確率微分方程式の基礎；計測と制御, Vol. 50, No. 11, pp. 937–943 (2011)

[10] 佐藤, 藤本：力学系の性質を利用した非線形確率システムの制御；計測と制御, Vol. 50, No. 11, pp. 981–986 (2011)

[11] 佐藤：非線形確率システムの考え方と応用；第８回横幹連合コンファレンス予稿集, pp. (USB) B–4–1 (2017)

[12] I. Gihman and A. Skorohod : *Stochastic Differential Equations*, Springer-Verlag (1972)

[13] N. Ikeda and S. Watanabe : *Stochastic differential equations and diffusion processes*, North-Holland/Kodansha, Amsterdam/Tokyo (1981)

[14] 津野：ファイナンスの確率積分–伊藤の公式，Girsanov の定理，Black-Scholes の公式–, 共立出版 (2001)

[15] 大塚：非線形最適制御入門, コロナ社 (2011)

[16] 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)

[17] H. J. Kappen : An introduction to stochastic control theory, path integrals and reinforcement learning; *Proc. 9th Granada seminar on computational physics: Cooperative behavior in neural systems*, pp. 149–181 (2007)

[18] V. Gómez, S. Thijssen, A. Symington, S. Hailes and H. J. Kappen : Real-time stochastic optimal control for multi-agent quadrotor swarms; *Proc. 26th Int. Conf. Automated Planning and Scheduling*, pp. 468–476 (2016)

[19] R. Fujimoto and N. Sakamoto : The stable manifold approach for optimal swing up and stabilization of an inverted pendulum with input saturation; *Proc. 18th IFAC World Congress*, pp. 8046–8051 (2011)