ESAIM: COCV 32 (2026) 9
https://doi.org/10.1051/cocv/2025091

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

ADVANCED CONTROL STRATEGIES FOR STOCHASTIC SYSTEMS USING PDF OPTIMISATION

Randa Herzallah*

Abstract. This paper presents an innovative probabilistic control framework for continuous-time stochastic systems. Unlike traditional control approaches that optimise deterministic control strategies, our framework directly optimises the probability density function (PDF) of the control signal, allowing for a more adaptable and robust response to stochastic variations. By integrating stochastic differential equations with the Hamilton–Jacobi–Bellman equation and utilising the Fokker–Planck dynamics, our method offers a precise and dynamic approach to managing uncertainty. The framework minimises the Kullback–Leibler divergence to align the system's joint state and control distribution with a desired joint target distribution, ensuring effective control even in unpredictable environments. A novel algorithm iteratively refines the control PDF based on real-time feedback, further enhancing the system's alignment with the target behaviour. The proposed method is demonstrated on an Ornstein–Uhlenbeck process, showcasing its effectiveness in steering the system's state distribution toward desired outcomes and underscoring its broad applicability to stochastic systems.

Mathematics Subject Classification. 93E20, 49L20, 35Q84, 60H10.

Received February 13, 2025. Accepted November 20, 2025.

1. Introduction

Uncertainty modelling is a growing research area with important implications across various domains, including finance, engineering, and environmental sciences. Stochastic processes, often represented by stochastic differential equations (SDEs), are key to understanding and managing systems influenced by randomness and uncertainty. These models describe the unpredictable behaviour caused by noise and offer a detailed framework for both analysis and control.

The endeavour to control stochastic systems began diligently in the 1960s with seminal contributions from Richard Bellman and Rudolf Kalman [1, 2]. Their work established the mathematical foundation for solving linear-quadratic (LQ) problems within deterministic settings. These methods were effective, but they relied on deterministic cost functionals and assumed linear and Gaussian noise. Although these assumptions made the analysis more manageable and tractable, they did not account for the complex behaviours of more intricate or non-linear systems.

Later, the field saw significant advancements through the works of Harold Kushner [3] and William Wonham [4]. They expanded the LQ theory to the stochastic domain and adapted it to accommodate the inherent

---

Keywords and phrases: Stochastic control, probability density function, Kullback–Leibler divergence, Fokker–Planck equation, Hamilton–Jacobi–Bellman equation.

Warwick Mathematics Institute, Warwick University, Coventry CV4 7AL, UK.

* Corresponding author: Randa.Herzallah@warwick.ac.uk

uncertainties of stochastic processes. This advancement was a major step in control theory. It helped move from deterministic models to stochastic frameworks that more accurately reflect real-world complexities. However, despite these advancements, traditional stochastic control approaches still have their limitations. These methods often rely on deterministic cost functions [4–9], which may inadvertently overlook the nuanced dynamics present in real-world scenarios. Although these approaches are robust in certain applications, they might not fully address the complex interactions and variations seen in highly stochastic environments.

As methodologies advanced, dynamic programming methods have become increasingly important. This foundational approach led to the formulation of the Hamilton–Jacobi–Bellman (HJB) equation, which has been instrumental in defining optimal control strategies under uncertainty [10, 11]. However, as systems became more complex, the computational demands of traditional grid-based solvers for the HJB equation posed significant challenges. The Fokker–Planck (FP) control framework, which focuses on the evolution of the system's PDF, offers an alternative approach [12]. It integrates well with the dynamic programming principles found in the HJB equation. Under certain assumptions, the HJB and FP approaches are shown to be equivalent, providing a robust dual strategy for managing the complexities of stochastic systems [13]. The applications of the FP framework in various models, including extensions to mean-field frameworks [14, 15], have demonstrated its versatility and effectiveness in real-world scenarios.

As computational demands continued to challenge traditional methods, approaches such as Monte Carlo schemes [16, 17] and path integral control [18, 19] have become increasingly relevant. Although not entirely new, these methods represent alternative strategies that have received more attention with modern computational advancements. The path integral method which utilises the calculus of variations and probabilistic path integrals, is particularly suited for continuous stochastic processes. These methods, along with Kullback–Leibler (KL) control strategies [20, 21], aim to align the behaviour of actual systems with desired probability distributions. They offer advanced control mechanisms that adapt dynamically to observed state fluctuations. However, both approaches face challenges in continuous-time settings, often requiring discretisation to effectively transition from theoretical models to practical applications [22]. Moreover, a recent review of stochastic linear-quadratic (SLQ) control [8] outlines how modern formulations, including both finite and infinite-horizon problems, often rely on solving stochastic Riccati differential equations (SRDEs) in continuous time. While these SRDEs can be derived analytically in some cases, practical solutions still commonly involve numerical time discretisation, particularly in high-dimensional or uncertain settings.

Although these recent developments in stochastic control methods have expanded the field, a significant gap remains in the continuous-time domain. This gap is particularly due to the lack of fully probabilistic control concepts, which have already been explored in discrete-time settings and have shown promise [23, 24]. While existing methods are effective at addressing specific aspects of stochastic control, they often do not extend seamlessly to continuous-time systems without requiring simplifications that may compromise the integrity of the stochastic modelling.

To address the limitations in existing methods, our work introduces a novel fully probabilistic control framework specifically designed for continuous-time stochastic systems. Unlike traditional approaches that optimise with respect to deterministic control signals, our framework optimises directly with respect to the PDF of the control signal. This represents a fundamental shift from deterministic to probabilistic control, allowing for a more accurate and flexible strategy that better captures the stochastic nature of the system. Additionally, in contrast to KL control and path integral approaches, which typically require discretisation and rely on sampling trajectories or approximating cost-to-go functions, our framework remains continuous in time. It avoids trajectory sampling altogether by formulating the control problem at the level of evolving PDFs, allowing the optimisation to occur directly in distribution space rather than over individual trajectories. Compared to recent advancements in SLQ control, particularly those based on SRDEs, our method departs by modelling the control problem as a joint PDF evolution task. Specifically, we employ the Fokker–Planck equation to characterise the joint dynamics of the system state and control distributions. Rather than relying on a deterministic cost function as in SLQ, we employ the Kullback–Leibler divergence (KLD) as a cost metric to quantify the discrepancy between the evolving joint distribution and a prescribed target distribution. This enables richer representations that account not only for the system's expected behaviour but also for higher-order uncertainty characteristics. By integrating the dynamics of SDEs with the FP equation, our framework offers a more comprehensive approach to modelling and controlling stochastic systems. It contributes to both the theoretical development and practical application of control strategies, providing a method that is better suited to handle the complexities and uncertainties of real-world stochastic environments.

The paper is structured as follows. Section 2 defines the problem and provides the mathematical formulation of probabilistic control for stochastic systems. Section 3 then offers a theoretical framework for deriving optimal control policies for general stochastic systems. The application of Girsanov's Theorem to compute the KLD is introduced in Section 4. Section 5 focuses on the optimal control of linear Gaussian systems, detailing the application of the proposed methods to this specific class of systems. An algorithm for implementing the proposed probabilistic control in Gaussian linear systems is presented in Section 5.1. The practical utility of the framework is demonstrated on an Ornstein–Uhlenbeck (OU) Process in Section 6. Finally concluding remarks are given in Section 7.

2. Problem definition and formulation of probabilistic control for stochastic systems

In the study of complex systems influenced by randomness, SDEs are important for capturing the dynamics that involve both deterministic trends and stochastic variations. These equations are particularly relevant in fields where uncertainty plays a significant role, including financial markets, environmental modelling, and engineering systems. This section introduces a novel framework to the optimal control of such systems. Within this framework the aim is to minimise the deviation of the system's behaviour from a prespecified desired behaviour using the KLD as a cost function.

SDEs integrate random fluctuations directly into system dynamics through noise terms, providing a realistic representation of systems where outcomes are driven by known forces and random environmental inputs. The general form of the SDE studied here is represented as follows:

$$\mathrm{d}x_t = f(x_t, u_t)\,\mathrm{d}t + \sigma(x_t, u_t)\,\mathrm{d}W_t, \tag{2.1}$$

where $x_t \in \mathbb{R}^n$ is the state vector, $u_t \in \mathbb{R}^m$ is the control vector, $f(x_t, u_t)$ is the drift term, $\sigma(x_t, u_t) \in \mathbb{R}^{n\times q}$ is the diffusion term, and $\mathrm{d}W_t$ represents the increment of a q-dimensional standard Brownian motion. The control u_t is applied to influence the state dynamics and is chosen from a set of admissible controls.

The unpredictability introduced by the diffusion term $\sigma(x_t, u_t)$ necessitates a control approach that accounts for the probabilistic nature of state evolution. Such an approach should focus on managing the deviations of the joint PDF of the system dynamics and control input from a desired joint PDF rather than predicting exact outcomes. Therefore, in this study, we will design the control strategy at time t as a randomised control strategy $c(u_t|x_t)$, effectively adapting to the inherent uncertainties.

Assumption 2.1. To ensure the well-posedness of the problem and the existence of solutions, we make the following assumptions:

- The functions $f(x_t, u_t)$ and $\sigma(x_t, u_t)$ are Lipschitz continuous and differentiable with respect to their arguments.
- The control policy $c(u_t \mid x_t)$ is a Markovian probability density supported on an admissible set $U \subseteq \mathbb{R}^m$ (in this paper we take $U = \mathbb{R}^m$ for analytical clarity).

Let $s(x, t)$ denote the probability density of the state x at time t. When the control u is drawn from the Markov policy $c(u_t \mid x_t)$, the density s evolves according to the FP equation:

$$\begin{aligned}\frac{\partial s(x,t)}{\partial t} &= -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left[\left(\int f_i(x_t, u_t)\, c(u_t \mid x_t)\,\mathrm{d}u_t\right) s(x,t)\right] \\ &\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i\,\partial x_j}\left[\left(\int \left(\sigma(x_t,u_t)\sigma(x_t,u_t)^\top\right)_{ij} c(u_t \mid x_t)\,\mathrm{d}u_t\right) s(x,t)\right].\end{aligned} \tag{2.2}$$

This equation describes drift and diffusion of the state when actions are randomised according to $c(u_t \mid x_t)$.

Our contribution is to use the KLD as an objective that measures the discrepancy between the system's actual joint conditional density and a predefined target joint conditional density. By minimising this divergence, we design controls that more accurately steer the system's behaviour toward the desired dynamics. Let the transition probability density from state x_t at time t to state $x_{t+\mathrm{d}t}$ at time $t+\mathrm{d}t$ under control u_t be denoted by $s(x_{t+\mathrm{d}t}|x_t, u_t)$. For a reference system with policy $c_I(u_t \mid x_t)$, let $s_I(x_{t+\mathrm{d}t} \mid x_t, u_t)$ denote the corresponding reference transition density. Then the KLD between the actual and desired joint conditionals at time t is:

$$\widetilde{J}(x_t,t) = \mathcal{D}\left(s(x_{t+\mathrm{d}t}, u_t|x_t) \parallel s_I(x_{t+\mathrm{d}t}, u_t|x_t)\right) = \int\int s(x_{t+\mathrm{d}t}, u_t|x_t) \ln\left(\frac{s(x_{t+\mathrm{d}t}, u_t|x_t)}{s_I(x_{t+\mathrm{d}t}, u_t|x_t)}\right) \mathrm{d}x_{t+\mathrm{d}t}\, \mathrm{d}u_t. \tag{2.3}$$

By applying the chain rule for probability densities, the joint density can be factorised into transition densities of system's state and control policies, yielding:

$$\widetilde{J}(x_t,t) = \mathcal{D}\big(c(\cdot \mid x_t) \,\big\|\, c_I(\cdot \mid x_t)\big) + \mathbb{E}_{u\sim c(\cdot|x_t)}\Big[\mathcal{D}\big(s(\cdot \mid x_t, u) \,\big\|\, s_I(\cdot \mid x_t, u)\big)\Big]. \tag{2.4}$$

For small $\mathrm{d}t$ the divergence between the transition kernels is $O(\mathrm{d}t)$. We therefore define the local KLD rate:

$$\mathcal{D}(x_t,u_t) := \lim_{\mathrm{d}t\to 0} \frac{1}{\mathrm{d}t}\, \mathcal{D}\big(s(\cdot \mid x_t, u_t) \,\big\|\, s_I(\cdot \mid x_t, u_t)\big), \tag{2.5}$$

so that $\mathcal{D}(s(\cdot \mid x_t, u_t) \,\|\, s_I(\cdot \mid x_t, u_t)) = \mathcal{D}(x_t,u_t)\,\mathrm{d}t + o(\mathrm{d}t)$. Taking the incremental cost on $[t, t+\mathrm{d}t]$ to be the sum of a policy term and a transition term with rate $\mathcal{D}(x_t,u_t)$ leads to the running cost rate:

$$J(x_t,t) := \mathcal{D}\big(c(\cdot \mid x_t) \,\big\|\, c_I(\cdot \mid x_t)\big) + \mathbb{E}_{u\sim c(\cdot|x_t)}\big[\mathcal{D}(x_t,u_t)\big], \tag{2.6}$$

The objective is to design a randomised controller $c(u_t|x_t)$ that minimises the expected cumulative running cost. Thus, the value function $V(x_t,t)$ is defined as:

$$V(x_t,t) = \min_{c(u_l|x_l)_{l\geq t}^{T}} \mathbb{E}\left[\int_t^T J(x_l,l)\,\mathrm{d}l \,\middle|\, x_t\right], \tag{2.7}$$

To solve the optimisation problem, we employ the Bellman principle of optimality. This principle states that irrespective of the initial state and decision, the subsequent decisions must constitute an optimal policy with respect to the state resulting from the initial decision. This principle allows us to decompose the value function at an infinitesimal time step $\mathrm{d}t$ as follows:

$$V(x_t,t) = \min_{c(u_l|x_l)_{l=t}^{t+\mathrm{d}t}} \mathbb{E}\left[\int_t^{t+\mathrm{d}t} J(x_l,l)\,\mathrm{d}l + \min_{c(u_l|x_l)_{l=t+\mathrm{d}t}^{T}} \mathbb{E}\left[\int_{t+\mathrm{d}t}^T J(x_l,l)\,\mathrm{d}l \,\middle|\, x_{t+\mathrm{d}t}\right] \,\middle|\, x_t\right]. \tag{2.8}$$

Assuming that $J(x_l,l)$ remains approximately constant over the infinitesimally small interval $\mathrm{d}t$, the first integral simplifies to $J(x_t,t)\,\mathrm{d}t$. This simplification, along with the Bellman principle and the stochasticity of the dynamics, yields the following expectation-based recursion:

$$V(x_t,t) = \min_{c(u_t|x_t)} \left\{J(x_t,t)\,\mathrm{d}t + \mathbb{E}\left[V(x_{t+\mathrm{d}t}, t+\mathrm{d}t) \mid x_t\right]\right\}. \tag{2.9}$$

Since the value function $V(x_t,t)$ depends on both the state x_t and time t, we apply Itô's Lemma to $V(x_t,t)$. For notational brevity, when derivatives appear we suppress the explicit (x_t,t) dependence and write $V_t := \partial_t V(x_t,t)$,

$V_x := \nabla_x V(x_t, t) \in \mathbb{R}^n$, and $V_{xx} := \nabla_x^2 V(x_t, t) \in \mathbb{R}^{n\times n}$. Hence:

$$\mathrm{d}V = V_t\,\mathrm{d}t + (V_x)^\top \mathrm{d}x_t + \frac{1}{2}\,\mathrm{tr}\big[V_{xx}\,\mathrm{d}x_t(\mathrm{d}x_t)^\top\big]\,. \tag{2.10}$$

Substituting $\mathrm{d}x_t = f(x_t, u_t)\,\mathrm{d}t + \sigma(x_t, u_t)\,\mathrm{d}W_t$ gives:

$$\mathrm{d}V = \big(V_t + V_x^\top f(x_t, u_t) + \tfrac{1}{2}\,\mathrm{tr}[\,V_{xx}\,\sigma(x_t, u_t)\sigma(x_t, u_t)^\top]\big)\,\mathrm{d}t + V_x^\top \sigma(x_t, u_t)\,\mathrm{d}W_t. \tag{2.11}$$

Since the control is Markov (so u_l depends only on the current state x_l) and $V \in C^{1,2}$, the process $H_l := \sigma(x_l, u_l)^\top V_x(x_l, l)$ is non-anticipative and square–integrable on any finite horizon (by Asm. 1 together with the regularity of V). Hence the Itô integral $\int_t^{t+\mathrm{d}t} H_l^\top\,\mathrm{d}W_l$ is a square-integrable martingale with zero conditional mean given x_t [25]. Thus, taking conditional expectations in Itô's formula yields:

$$\mathbb{E}[\mathrm{d}V \mid x_t] = \big(V_t + V_x^\top f(x_t, u_t) + \tfrac{1}{2}\,\mathrm{tr}[\,V_{xx}\,\sigma(x_t, u_t)\sigma(x_t, u_t)^\top]\big)\,\mathrm{d}t. \tag{2.12}$$

By substituting the expected change of the value function from equation (2.12), and the immediate cost rate from equation (2.6), into the dynamic programming step given in equation (2.9), and then taking the limit as $dt \to 0$, we derive the HJB equation:

$$\begin{aligned} 0 = V_t + \min_{c(u_t\mid x_t)} \Bigg\{ \int c(u_t \mid x_t)\Big[\ln\frac{c(u_t \mid x_t)}{c_I(u_t \mid x_t)} + \mathcal{D}(x_t, u_t) + V_x^\top f(x_t, u_t) \\ + \tfrac{1}{2}\,\mathrm{tr}\big(V_{xx}\,\sigma(x_t, u_t)\sigma(x_t, u_t)^\top\big)\Big]\,\mathrm{d}u_t\Bigg\}. \end{aligned} \tag{2.13}$$

To simplify this equation further, we introduce the function $\beta(u_t, x_t)$ which aggregates the control policy's effects on both the state probabilities and the dynamics under control:

$$\beta(u_t, x_t) = \mathcal{D}(x_t, u_t) + V_x^\top f(x_t, u_t) + \tfrac{1}{2}\,\mathrm{tr}\big(V_{xx}\,\sigma(x_t, u_t)\sigma(x_t, u_t)^\top\big). \tag{2.14}$$

Using this definition, the HJB equation can be rewritten as:

$$0 = V_t + \min_{c(u_t\mid x_t)} \left\{ \int c(u_t \mid x_t)\left[\ln\left(\frac{c(u_t \mid x_t)}{c_I(u_t \mid x_t)}\right) + \beta(u_t, x_t)\right]\mathrm{d}u_t\right\}. \tag{2.15}$$

This equation implies a further simplified relationship, highlighting the optimal control strategy in terms of the minimum expected cost:

$$-V_t = \min_{c(u_t\mid x_t)} \int c(u_t \mid x_t)\ln\left(\frac{c(u_t \mid x_t)}{c_I(u_t \mid x_t)\exp(-\beta(u_t, x_t))}\right)\mathrm{d}u_t. \tag{2.16}$$

This completes the formulation of the cost function, which will be utilised in the derivation of the optimal randomised controller. The derived HJB equation provides the foundation for determining the randomised control strategies needed to align the system state PDF with a prespecified desired PDF.

3. Optimal probabilistic control strategies

The HJB equation, introduced in the previous section, is essential for determining optimal control laws in continuous-time stochastic systems. This equation correlates the expected total cost with the system's behaviour under the influence of control measures. It emphasises the need for control laws to be adaptable, due to the inherent uncertainties of the environment. The optimal control strategy, detailed in the following theorem, utilises the probabilistic dynamics of the system. This ensures that control actions achieve the desired outcomes and at the same time respond effectively to unexpected changes and uncertainties. As a result, the overall performance of the system is optimised over time.

Theorem 3.1. *The PDF of the optimal control law, $c(u_t|x_t)$ that minimises the cost-to-go function stated in equation (2.16) can be shown to be given by:*

$$c(u_t|x_t) = \frac{c_I(u_t|x_t)\exp[-\beta(u_t, x_t)]}{\int_U c_I(u_t|x_t)\exp[-\beta(u_t, x_t)]\mathrm{d}u_t}, \tag{3.1}$$

where $\beta(u_t, x_t)$ is defined in equation (2.14).

Remark 3.2. If the control input is subject to a hard bound $u_t \in U \subset \mathbb{R}^m$, define the reference density $c_I(u_t \mid x_t)$ so that $c_I(u_t \mid x_t) = 0$ for $u_t \notin U$. Because the optimal density in equation (3.1) shares the support of c_I, it is automatically zero outside U; equivalently, the normalising integral in that equation can be taken over U. All results in Theorem 3.1 remain valid, since the proofs depend only on the support of c_I. For *soft* saturation one may keep $U = \mathbb{R}^m$ and choose c_I with a small covariance so that u_t lies within the desired range with high probability.

Proof. The minimisation in (2.16) reduces to the convex problem:

$$\begin{aligned} \min_{c(\cdot|x_t)} \quad & \int_U c(u_t \mid x_t)\Big[\ln \tfrac{c(u_t|x_t)}{c_I(u_t|x_t)} + \beta(u_t, x_t)\Big]\,\mathrm{d}u_t \\ \text{s.t.} \quad & \int_U c(u_t \mid x_t)\,\mathrm{d}u_t = 1, \quad c(u_t \mid x_t) \geq 0 \;\; \forall u_t \in U, \\ & c(u_t \mid x_t) = 0 \text{ whenever } c_I(u_t \mid x_t) = 0, \end{aligned}$$

i.e., $c(\cdot \mid x_t)$ is supported on the support of $c_I(\cdot \mid x_t)$. Introduce a Lagrange multiplier $\lambda(x_t, t)$ for the normalisation constraint and consider:

$$\mathcal{L}[c, \lambda] = \int_U c(u_t \mid x_t)\Big[\ln \tfrac{c(u_t|x_t)}{c_I(u_t|x_t)} + \beta(u_t, x_t)\Big]\,\mathrm{d}u_t + \lambda(x_t, t)\bigg(\int_U c(u_t \mid x_t)\,\mathrm{d}u_t - 1\bigg).$$

The first variation in the direction δc is:

$$\delta\mathcal{L} = \int_U \delta c(u_t \mid x_t)\,\Big[\ln \tfrac{c(u_t|x_t)}{c_I(u_t|x_t)} + 1 + \beta(u_t, x_t) + \lambda(x_t, t)\Big]\,\mathrm{d}u_t.$$

Optimality requires $\delta\mathcal{L} = 0$ for all δc, hence pointwise on the support of c_I:

$$\ln \tfrac{c(u_t|x_t)}{c_I(u_t|x_t)} + 1 + \beta(u_t, x_t) + \lambda(x_t, t) = 0,$$

which implies:

$$c(u_t \mid x_t) = c_I(u_t \mid x_t)\,\exp\big(-\beta(u_t, x_t) - 1 - \lambda(x_t, t)\big).$$

Define the normalizing constant:

$$Z(x_t, t) := \int_U c_I(u_t \mid x_t)\,\exp\big(-\beta(u_t, x_t)\big)\,\mathrm{d}u_t \quad \text{(assumed finite and nonzero).}$$

Enforcing $\int_U c(\cdot \mid x_t) = 1$ gives $\exp(-1-\lambda(x_t,t)) = Z(x_t,t)^{-1}$, and therefore:

$$c(u_t \mid x_t) = \frac{c_I(u_t \mid x_t)\, \exp\big[-\beta(u_t,x_t)\big]}{\int_U c_I(u_t \mid x_t)\, \exp\big[-\beta(u_t,x_t)\big]\, \mathrm{d}u_t},$$

which is (3.1). Strict convexity of the objective (the KL term is strictly convex in c, the β term is linear) ensures a unique minimiser on the support of c_I (up to c_I-null sets). ☐

The theorem provides a universal solution for controlling continuous stochastic systems. This method works without relying on any specific form of the generative probabilistic model describing the system dynamics. Therefore, it applies to any stochastic system, regardless of whether its dynamics follow the FP equation's assumptions.

4. Kullback–Leibler divergence and Girsanov's theorem

Following the development of the HJB equation for SDEs, this section examines the roles of KLD and Girsanov's Theorem. These tools are important for formulating and evaluating stochastic control objectives under uncertainty.

In our framework, the function $\beta(u_t, x_t)$ measures the impact of the system's dynamics and the difference between the actual and reference behaviours. A key element in this formulation is the KLD which compares the transition probability densities $s(x_{t+\mathrm{d}t}|x_t,u_t)$ and $s_I(x_{t+\mathrm{d}t}|x_t,u_t)$. To evaluate this KLD in the continuous-time context, we employ Girsanov's theorem. The transition probability $s(x_{t+\mathrm{d}t}|x_t,u_t)$ indicates the likelihood of the system transitioning from one state to another as described by equation (2.1), repeated here:

$$\mathrm{d}x_t = f(x_t,u_t)\mathrm{d}t + \sigma(x_t,u_t)\mathrm{d}W_t. \tag{4.1}$$

The objective in our framework is to design a randomised controller $c(u_t|x_t)$ that aligns this transition probability as closely as possible with a specified target transition probability, $s_I(x_{t+\mathrm{d}t}|x_t,u_t)$. The target dynamics under the reference measure are:

$$\mathrm{d}x_t = f_I(x_t,u_t)\mathrm{d}t + \sigma(x_t,u_t)\mathrm{d}\tilde{W}_t, \tag{4.2}$$

where $f_I(x_t,u_t)$ represents the desired control drift.

According to Girsanov's theorem, the Radon–Nikodym derivative of the controlled process measure s with respect to the reference process measure s_I is:

$$\frac{\mathrm{d}s}{\mathrm{d}s_I} = \exp\left(\int_0^t \theta_l^T\, \mathrm{d}W_l + \frac{1}{2}\int_0^t \|\theta_l\|^2\, \mathrm{d}l\right), \tag{4.3}$$

where θ_t relates the drifts of the two processes:

$$\theta_t = \sigma^\dagger(x_t,u_t)\left(f(x_t,u_t) - f_I(x_t,u_t)\right). \tag{4.4}$$

Here, $\sigma^\dagger(x_t,u_t)$ denotes the Moore–Penrose pseudoinverse of $\sigma(x_t,u_t)$. When $\sigma(x_t,u_t)$ has full row rank (so that $\sigma(x_t,u_t)\sigma^T(x_t,u_t)$ is invertible), this pseudoinverse reduces to $\sigma^\dagger(x_t,u_t) = \sigma^T(x_t,u_t)(\sigma(x_t,u_t)\sigma^T(x_t,u_t))^{-1}$ and then $\sigma^{\dagger^\top}\sigma^\dagger = (\sigma\sigma^\top)^{-1}$. If $\sigma(x_t,u_t)$ is not full row rank, we keep the general pseudoinverse notation $\sigma^\dagger(x_t,u_t)$, and all formulas below remain valid. The KLD between the two densities over an infinitesimal interval $\mathrm{d}t$ is then calculated as:

$$\mathcal{D}(s||s_I) = E_s\left[\ln\left(\frac{\mathrm{d}s}{\mathrm{d}s_I}\right)\right]. \tag{4.5}$$

Using Girsanov's theorem and a short-time expansion, this divergence satisfies:

$$\mathcal{D}\big(s\|s_I\big) \;=\; \frac{1}{2}\,\big\|\sigma^{\dagger}(x_t,u_t)\big(f(x_t,u_t)-f_I(x_t,u_t)\big)\big\|^2\,\mathrm{d}t \;+\; o(\mathrm{d}t). \tag{4.6}$$

Equivalently, the associated transition KLD rate is:

$$\lim_{\mathrm{d}t\to 0^+}\frac{1}{\mathrm{d}t}\,\mathcal{D}\big(s\|s_I\big) \;=\; \frac{1}{2}\,\big\|\sigma^{\dagger}(x_t,u_t)\big(f(x_t,u_t)-f_I(x_t,u_t)\big)\big\|^2. \tag{4.7}$$

Taking this rate and substituting into the definition of $\beta(u_t,x_t)$ from equation (2.14), we obtain the following expression:

$$\beta(u_t,x_t)=\frac{1}{2}\,\big\|\sigma^{\dagger}(x_t,u_t)\big(f(x_t,u_t)-f_I(x_t,u_t)\big)\big\|^2+V_x^{\top}f(x_t,u_t)+\frac{1}{2}\,\mathrm{tr}\big[V_{xx}\,\sigma(x_t,u_t)\sigma^{\top}(x_t,u_t)\big]. \tag{4.8}$$

This expression captures, per–unit–time, both the discrepancy between the actual and reference dynamics and the expected evolution of the value function due to the system's dynamics.

5. OPTIMAL CONTROL OF LINEAR GAUSSIAN SYSTEMS

This section applies the theoretical principles discussed previously for controlling stochastic equations of the form given in equation (2.1) to a specific class of stochastic systems characterised by linear Gaussian dynamics. The system's dynamics are determined by a linear drift and a diffusion term that is independent of state and control:

$$\begin{aligned} f(x_t,u_t)&=(\tilde{A}x_t+\tilde{B}u_t),\\ \sigma(x_t,u_t)&=\sigma_t. \end{aligned} \tag{5.1}$$

Given the linear drift and constant diffusion, and since $\sigma(x_t,u_t)=\sigma_t$ is a deterministic and square-integrable function of time, the Itô integral $\int_0^t \sigma_l\,\mathrm{d}W_l$ is a zero-mean Gaussian random variable [25]. Because the drift $f(x_t,u_t)=\tilde{A}x_t+\tilde{B}u_t$ is linear, the solution x_t is an affine transformation of this Gaussian integral, and hence the transition distribution $s(x_{t+\mathrm{d}t}\mid u_t,x_t)$ is multivariate normal:

$$s(x_{t+\mathrm{d}t}|u_t,x_t)\sim\mathcal{N}(\mu_t,\Sigma_t), \tag{5.2}$$

where the mean μ_t and covariance Σ_t evolve according to the differential equations:

$$\begin{aligned} \frac{\mathrm{d}\mu_t}{\mathrm{d}t}&=\tilde{A}\mu_t+\tilde{B}u_t,\\ \frac{\mathrm{d}\Sigma_t}{\mathrm{d}t}&=\tilde{A}\Sigma_t+\Sigma_t\tilde{A}^T+\sigma_t\sigma_t^T. \end{aligned} \tag{5.3}$$

Here μ_t represents the conditional expected value of the state x_t given the control inputs, while Σ_t quantifies the conditional variance of the state around its mean.

For this class of linear SDEs, the target PDF is assumed to follow a normal distribution, given by:

$$s_I(x_{t+\mathrm{d}t}\mid u_t,x_t)\sim\mathcal{N}(\mu_t^I,\Sigma_t), \tag{5.4}$$

where μ_t^I represents the mean of the desired state distribution. This describes the conditional distribution of the desired target state dynamics, which are modelled by the following SDE:

$$\mathrm{d}x_t = (A_I x_t + B_I u_t + x_t^r)\mathrm{d}t + \sigma_t \mathrm{d}\tilde{W}_t. \tag{5.5}$$

This model specification indicates that the target state dynamics, represented by the drift term, are influenced by the control input, u_t, the state, x_t, and an additional term x_t^r. The term x_t^r introduces a factor that can capture desired state behaviours independently of the current state and control input, offering flexibility in defining target trajectories. Consequently, this model represents a generalised framework for the desired behaviour of a stochastic system, accommodating a wide range of possible dynamics and control strategies.

By substituting equations (5.1) and (5.5) into equation (4.4), we reformulate the function $\beta(u_t, x_t)$, as defined in equation (4.8), in the following manner:

$$\begin{aligned}\beta(u_t, x_t) = {} & \frac{1}{2}(\tilde{A}x_t + \tilde{B}u_t - A_I x_t - B_I u_t - x_t^r)^T \sigma_t^{\dagger^T} \sigma_t^{\dagger}(\tilde{A}x_t + \tilde{B}u_t - A_I x_t - B_I u_t - x_t^r) \\ & + V_x^T(\tilde{A}x_t + \tilde{B}u_t) + \frac{1}{2}\mathrm{tr}[V_{xx}\sigma_t\sigma_t^T].\end{aligned} \tag{5.6}$$

This equation can be rewritten by introducing the definitions $A = \tilde{A} - A_I$, and $B = \tilde{B} - B_I$, which group similar terms together. This leads to the following form:

$$\beta(u_t, x_t) = \frac{1}{2}(Ax_t + Bu_t - x_t^r)^T \sigma_t^{\dagger^T} \sigma_t^{\dagger}(Ax_t + Bu_t - x_t^r) + V_x^T(\tilde{A}x_t + \tilde{B}u_t) + \frac{1}{2}\mathrm{tr}[V_{xx}\sigma_t\sigma_t^T], \tag{5.7}$$

Moreover, we define the ideal controller's PDF, $c_I(u_t|x_t)$, which, for the assumed linear and Gaussian context, is taken to be Gaussian:

$$c_I(u_t|x_t) \sim \mathcal{N}(u_t^r, \Gamma), \tag{5.8}$$

where u_t^r and Γ represent the mean and covariance matrix, respectively, of the ideal distribution of the controller.

With these Gaussian models, we can compute the optimal randomised controller from equation (3.1) as established in Theorem 3.1. This is stated in the following theorem.

Theorem 5.1. *The optimal control strategy that minimises the KLD as defined in equation* (2.4) *is:*

$$c(u_t|x_t) \sim \mathcal{N}(\nu_t, \Gamma_t), \tag{5.9}$$

where:

$$\begin{aligned}\nu_t &= -K_t x_t - L_t, \\ K_t &= \Gamma_t\Big(\tilde{B}^T P_t + B^T \sigma_t^{\dagger^T} \sigma_t^{\dagger} A\Big), \\ L_t &= \Gamma_t\Big(-\Gamma^{-1}u_r - B^T \sigma_t^{\dagger^T} \sigma_t^{\dagger} x_t^r + \tilde{B}^T q_t\Big), \\ \Gamma_t &= (\Gamma^{-1} + B^T \sigma_t^{\dagger^T} \sigma_t^{\dagger} B)^{-1},\end{aligned} \tag{5.10}$$

and,

$$\dot{P}_t = -\Big(A^T \sigma_t^{\dagger^T} \sigma_t^{\dagger} A + \tilde{A}^T P_t + P_t \tilde{A} - \Big[P_t \tilde{B} + A^T \sigma_t^{\dagger^T} \sigma_t^{\dagger} B\Big] \Gamma_t \Big[\tilde{B}^T P_t + B^T \sigma_t^{\dagger^T} \sigma_t^{\dagger} A\Big]\Big),$$

$$\begin{aligned}
\dot{q}_t &= -\bigg(- A^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r + \tilde{A}^T q_t - 2\Big[P_t \tilde{B} + A^T {\sigma_t^\dagger}^T \sigma_t^\dagger B\Big]\Gamma_t\Big[- \Gamma^{-1} u_r - B^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r + \tilde{B}^T q_t\Big]\bigg), \\
\dot{r}_t &= -\bigg(- \Big[- u_r^T \Gamma^{-1} - {x_t^r}^T {\sigma_t^\dagger}^T \sigma_t^\dagger B + q_t^T \tilde{B}\Big]\Gamma_t\Big[- \Gamma^{-1} u_r - B^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r + \tilde{B}^T q_t\Big] \\
&\quad + \frac{1}{2}\Big\{u_r^T \Gamma^{-1} u_r + {x_t^r}^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r + tr[P_t \sigma_t \sigma_t^T]\Big\}\bigg). \qquad (5.11)
\end{aligned}$$

Proof. The derivation of the optimal randomised controller begins by evaluating $\beta(u_t, x_t)$, as prescribed in equation (5.7). For this calculation, we start with the ansatz:

$$V(x_t, t) = \frac{1}{2} x_t^T P_t x_t + x_t^T q_t + r_t, \qquad (5.12)$$

so that:

$$\begin{aligned}
V_x &= P_t x_t + q_t, \\
V_{xx} &= P_t. \qquad (5.13)
\end{aligned}$$

Substituting this into equation (5.7) gives:

$$\beta(u_t, x_t) = \frac{1}{2}(A x_t + B u_t - x_t^r)^T {\sigma_t^\dagger}^T \sigma_t^\dagger (A x_t + B u_t - x_t^r) + (x_t^T P_t + q_t^T)(\tilde{A} x_t + \tilde{B} u_t) + \frac{1}{2}\mathrm{tr}[P_t \sigma_t \sigma_t^T]. \qquad (5.14)$$

Simplifying yields:

$$\begin{aligned}
\beta(u_t, x_t) &= \frac{1}{2}\Big[(B u_t - x_t^r)^T {\sigma_t^\dagger}^T \sigma_t^\dagger (B u_t - x_t^r) + 2 x_t^T A^T {\sigma_t^\dagger}^T \sigma_t^\dagger (B u_t - x_t^r) + x_t^T A^T {\sigma_t^\dagger}^T \sigma_t^\dagger A x_t\Big] \\
&\quad + (x_t^T P_t + q_t^T)(\tilde{A} x_t + \tilde{B} u_t) + \frac{1}{2}\mathrm{tr}[P_t \sigma_t \sigma_t^T]. \qquad (5.15)
\end{aligned}$$

The optimal randomised controller can then be computed from equation (3.1). Using equations (5.8) and (5.15), the numerator of equation (3.1), denoted as num, is computed as follows:

$$\begin{aligned}
&\mathrm{num} \propto c_I(u_t|x_t)\exp[-\beta(u_t, x_t)] \\
&= \exp\bigg[- \frac{1}{2}\Big\{(u_t - u_r)^T \Gamma^{-1}(u_t - u_r) + (B u_t - x_t^r)^T {\sigma_t^\dagger}^T \sigma_t^\dagger (B u_t - x_t^r) + 2 x_t^T A^T {\sigma_t^\dagger}^T \sigma_t^\dagger (B u_t - x_t^r) \\
&\quad + x_t^T A^T {\sigma_t^\dagger}^T \sigma_t^\dagger A x_t + 2(x_t^T P_t + q_t^T)(\tilde{A} x_t + \tilde{B} u_t) + \mathrm{tr}[P_t \sigma_t \sigma_t^T]\Big\}\bigg]. \qquad (5.16)
\end{aligned}$$

Expanding and isolating terms involving the control signal u_t from those that do not, the expression simplifies to:

$$\begin{aligned}
&\mathrm{num} \propto \exp\bigg[- \frac{1}{2}\Big\{u_t^T(\Gamma^{-1} + B^T {\sigma_t^\dagger}^T \sigma_t^\dagger B) u_t + u_t^T\Big(- 2\Gamma^{-1} u_r - 2B^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r + 2B^T {\sigma_t^\dagger}^T \sigma_t^\dagger A x_t + 2\tilde{B}^T P_t x_t + 2\tilde{B}^T q_t\Big)\Big\}\bigg] \\
&\times \exp\bigg[- \frac{1}{2}\Big\{u_r^T \Gamma^{-1} u_r + {x_t^r}^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r - 2 x_t^T A^T {\sigma_t^\dagger}^T \sigma_t^\dagger x_t^r + x_t^T A^T {\sigma_t^\dagger}^T \sigma_t^\dagger A x_t + 2 x_t^T P_t \tilde{A} x_t + 2 q_t^T \tilde{A} x_t + \mathrm{tr}[P_t \sigma_t \sigma_t^T]\Big\}\bigg]. \\
&\qquad (5.17)
\end{aligned}$$

Completing the square for the control signal u_t in the first exponential term, we find:

$$\text{num} \propto \exp\left[-\frac{1}{2}\left\{u_r^T\Gamma^{-1}u_r + x_t^{r^T}\sigma_t^{\dagger^T}\sigma_t^{\dagger}x_t^r - 2x_t^TA^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}x_t^r + x_t^TA^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}Ax_t + 2x_t^TP_t\tilde{A}x_t + 2q_t^T\tilde{A}x_t + \mathrm{tr}[P_t\sigma_t\sigma_t^T]\right\}\right]$$
$$\times \exp\left[-\frac{1}{2}\left\{(u_t-\nu_t)^T\Gamma_t(u_t-\nu_t) + Z_t\right\}\right], \tag{5.18}$$

where Γ_t, and ν_t are as defined in equation (5.10) and:

$$Z_t = -\left(-\Gamma^{-1}u_r - B^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}x_t^r + B^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}Ax_t + \tilde{B}^TP_tx_t + \tilde{B}^Tq_t\right)^T\Gamma_t\left(-\Gamma^{-1}u_r - B^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}x_t^r\right.$$
$$\left. + B^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}Ax_t + \tilde{B}^TP_tx_t + \tilde{B}^Tq_t\right). \tag{5.19}$$

The denominator (den) of equation (3.1) can then be computed by integrating the numerator given in (5.18) with respect to u_t. This yields:

$$\text{den} \propto \exp\left[-\frac{1}{2}\left\{u_r^T\Gamma^{-1}u_r + x_t^{r^T}\sigma_t^{\dagger^T}\sigma_t^{\dagger}x_t^r - 2x_t^TA^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}x_t^r + x_t^TA^T\sigma_t^{\dagger^T}\sigma_t^{\dagger}Ax_t + 2x_t^TP_t\tilde{A}x_t + 2q_t^T\tilde{A}x_t\right.\right.$$
$$\left.\left. + \mathrm{tr}[P_t\sigma_t\sigma_t^T]\right\}\right]\exp\left[-\frac{1}{2}Z_t\right]. \tag{5.20}$$

Using the expressions for both num and den as derived in equations (5.18) and (5.20), we apply them to equation (3.1) to compute the conditional PDF of the control signal, u_t. This yields:

$$c(u_t|x_t) \propto \exp\left[-\frac{1}{2}(u_t-\nu_t)^T\Gamma_t(u_t-\nu_t)\right]. \tag{5.21}$$

This completes the proof of the controller's randomised form stated in the theorem.

To verify the Riccati equation, and the linear and constant coefficients of the value function, substitute the optimal control distribution from (5.21) together with the numerator in (5.18) into the HJB equation. Equating coefficients of equal degree then yields the differential system (5.11): the quadratic terms give the Riccati equation, the linear terms determine the linear component of the value function, and the constant terms give the baseline cost in the value function's differential equation. □

Theorem 5.1 is consistent with classical stochastic control. In particular, Riccati relations appear when computing the feedback law, yielding the gain K_t and the shift L_t as in the standard stochastic optimal control setting. The additional first and fourth terms in our Riccati relation arise because we penalise, through the KL divergence, the difference between the joint PDF of state and control induced by the randomised controller and a prescribed reference PDF.

The main difference from conventional approaches is that our method is fully probabilistic. Traditional formulations optimise a deterministic objective and return a deterministic control law. In contrast, we optimise a control law that is a probability distribution. As shown in Theorem 5.1 and equation (5.9), this yields an explicit PDF for the optimal randomised controller, which is well suited to intrinsically stochastic dynamics that require a PDF to describe their time evolution. In this sense, the proposed method generalises and improves upon deterministic control by working directly with probability distributions.

5.1. Algorithm for implementing the proposed probabilistic control in Gaussian linear systems

Algorithm 1 provides a detailed procedure for implementing the proposed probabilistic control framework, specifically designed for Gaussian linear systems.

Algorithm 1 Probabilistic control for gaussian linear systems

```
Input: System matrices $A$, $B$, noise intensity $\sigma_t$, initial state $x_0$, time horizon $T$, time step $dt$, parameters of the target drift function $A_I$, $B_I$, and $x_t^r$, and covariance of ideal controller $\Gamma$.
Output: Optimised distribution of the control signal $c(u_t \mid x_t)$, and evolution of the PDF of the state $x_t$.
Initialise: $P_{t=0} = 0$, $q_{t=0} = 0$, $r_{t=0} = 0$, $x_{t=0} = x_0$.
for each time $t$ from 0 to $T$ in steps of $\mathrm{d}t$ do:
    Compute the target mean $x_t^r$ at time $t$.
    Update the parameters of the value function $P_t$, $q_t$ and $r_t$ using equation (5.11).
    Integrate to update:
        $P_{t+\mathrm{d}t} = P_t + \dot{P}_t \, \mathrm{d}t$,
        $q_{t+\mathrm{d}t} = q_t + \dot{q}_t \, \mathrm{d}t$,
        $r_{t+\mathrm{d}t} = r_t + \dot{r}_t \, \mathrm{d}t$.
    Compute the optimal control parameters, $\Gamma_t$, $K_t$, and $L_t$ using equation (5.10).
    Compute the mean of the optimised random control signal using equation (5.10).
    Use the mean of the optimised random control signal to update the drift of the system and compute the system state value $x_{t+\mathrm{d}t}$.
end for
```

6. SIMULATION STUDY: ORNSTEIN–UHLENBECK PROCESS

In this section we apply the proposed fully probabilistic controller to the OU process and compare its performance with SLQ [8] baseline. This classical stochastic model is used to describe the velocity of a particle influenced by friction and random forces. This example is ideal for demonstrating the effectiveness of the control method due to the inherent stochastic nature of the process.

$$\mathrm{d}x_t = \big(-\gamma x_t + u_t\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

where x_t represents the velocity of the particle, γ is the friction coefficient, σ denotes the intensity of random fluctuations, and u_t is the control input. The process, characterised by the FP equation, is known for its mean-reverting property, resulting in a Gaussian distribution with time-varying mean and variance. The objective of the control framework is to derive a randomised controller $c(u_t \mid x_t)$ to minimise the KLD defined in equation (2.4) hence ensuring that the PDF of the state x_t closely aligns with a desired PDF over time. The desired Gaussian PDF is characterised by a sinusoidal mean function:

$$v(t) = 2\sin\left(\frac{\pi t}{5}\right)$$

and a fixed variance.

In our simulation, the friction coefficient γ was set to 1, reflecting the system's tendency to revert to its mean state. The noise intensity σ was set to 0.2, representing the level of stochastic fluctuations in the system. The covariance of the ideal controller's PDF was chosen to be 0.0001, which affects the responsiveness of the control input. The simulation was conducted over a time horizon T of 5 units with a time step $\mathrm{d}t$ of 0.01 units.

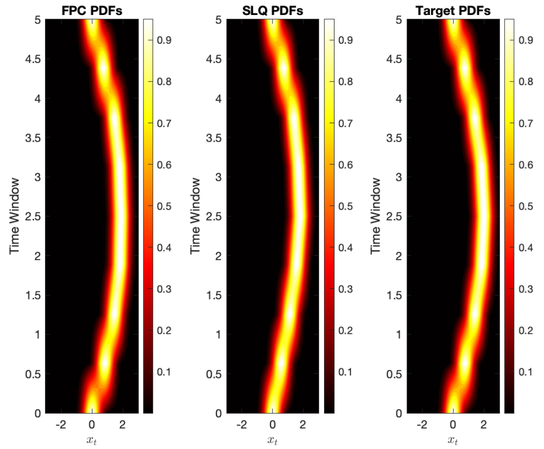

FIGURE 1. Time evolution of state PDFs. Left: proposed FPC; middle: SLQ baseline; right: target distribution. Brighter regions indicate higher probability density.

For the FPC approach, implementation follows the procedure outlined in Algorithm 1. It involved solving the Riccati equation and applying the optimal control law to the system to generate the computed PDFs. For the SLQ control method, we implement a standard quadratic tracking cost $J = \mathbb{E}\int_0^T \Big[q(x_t - v(t))^2 + r\, u_t^2\Big]\, \mathrm{d}t$, with weighting $r = 0.01$ and $q = 1.0$.

Figure 1 compares the PDFs obtained with the fully probabilistic controller (left) and the SLQ baseline (centre) against the desired target distribution (right). As can be seen, both controllers closely follow the target mean trajectory, as evidenced by the alignment of the high-density ridges. However, the fully probabilistic controller exhibits a slightly better match in terms of spread, particularly around the mid-horizon region (*e.g.*, time window $2.5 - 4$), where its contour width appears more consistent with the target distribution. The SLQ controller, by contrast, produces slightly narrower contours in that interval, suggesting a mild underestimation of uncertainty. Overall, while both controllers maintain good tracking performance, the fully probabilistic controller provides a closer match to the target in both mean and variance across the time horizon.

Figure 2 compares the time-slice PDFs produced by the fully probabilistic controller (blue, solid), the SLQ baseline (red, dashed) and the target distribution (black, dotted). Both controllers track the evolving mean of the target distribution well across all time steps. However, a key difference emerges in the variance. In the mid-horizon range (approximately $t = 1$ to $t = 3.5$), the SLQ controller consistently produces slightly narrower distributions compared to the target, indicating a modest underestimation of uncertainty. The fully probabilistic control, by contrast, maintains a spread that more closely matches the target throughout, particularly at time steps around $t \approx 2.5$, $t \approx 3.0$, and $t \approx 3.5$, where its curves nearly overlay the dotted reference. Toward the start and end of the horizon (*e.g.*, $t = 0$ and $t = 5$), all three distributions converge, showing strong alignment. Overall, the FPC demonstrates more accurate tracking of both the mean and variance of the desired distribution, while the SLQ solution remains slightly conservative in its estimation of spread.

To complement the visual evidence, we computed three summary statistics over the full $5-$unit time horizon. First, the mean-squared tracking error $\mathrm{MSE} = \dfrac{1}{T}\displaystyle\int_0^T \Big(\mathbb{E}[x_t] - v(t)\Big)^2\, \mathrm{d}t$ which measures how well the controlled state mean follows the target mean $v(t)$. Second, the time-averaged KLD, $\overline{\mathcal{D}} = \dfrac{1}{T}\displaystyle\int_0^T \mathcal{D}\big[s(x_t)\,\|\,s_I(x_t)\big]\, \mathrm{d}t$ which quantifies the shape mismatch between the controlled and reference PDFs. Finally, the normalised control effort, $E_u = \dfrac{1}{T}\displaystyle\int_0^T u_t^2\, \mathrm{d}t$. As summarised in Table 1, the fully probabilistic controller yields a lower tracking error and

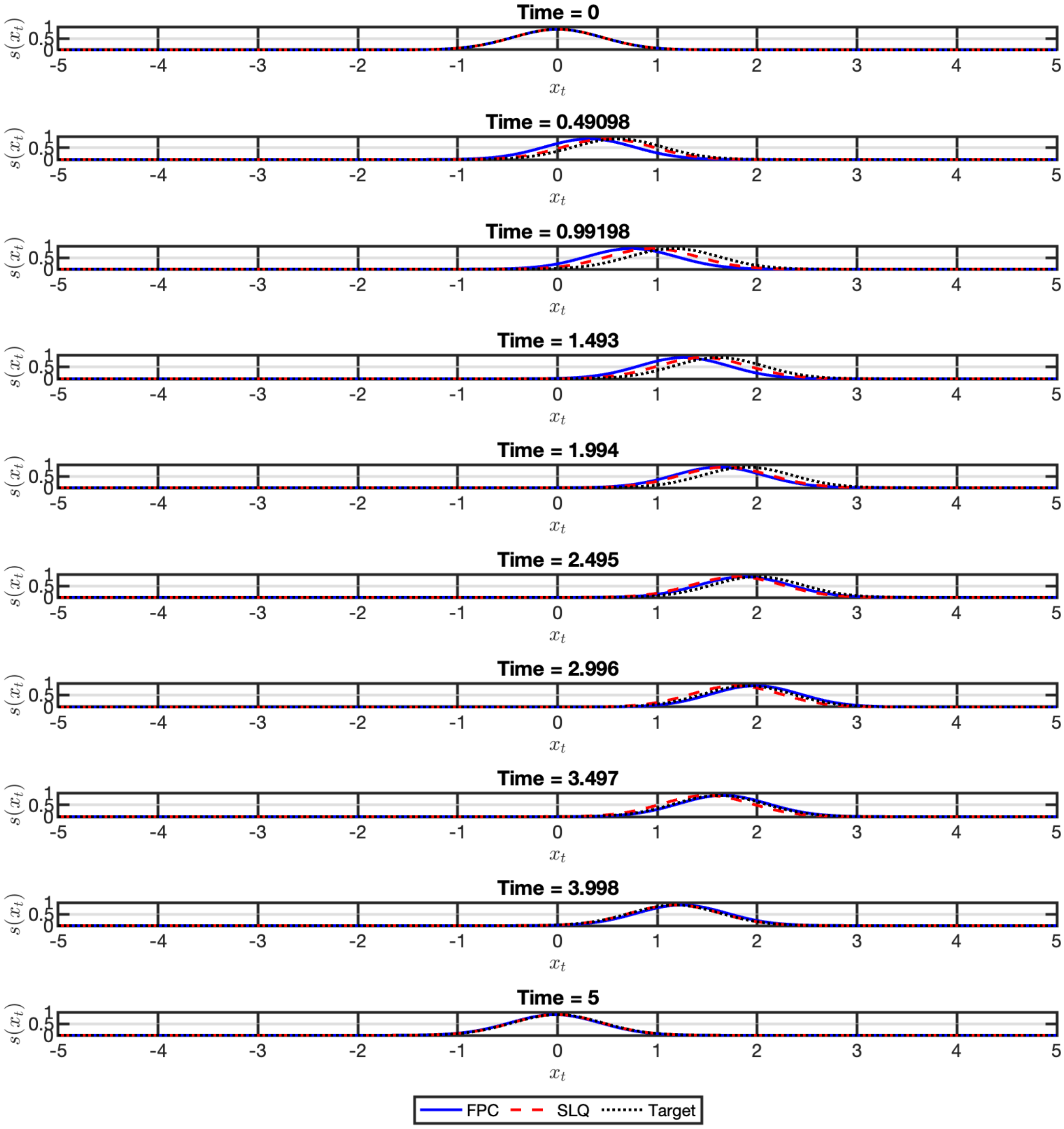

Figure 2. PDF cross-sections at selected time points. Solid blue line: FPC; dashed red line: SLQ; dotted black line: target distribution.

a smaller average KL divergence while requiring less control energy than the SLQ baseline, quantitatively confirming the qualitative trends seen in the figures.

7. Conclusion

This paper presented a novel probabilistic control framework designed specifically for stochastic continuous-time systems. Our approach is based on characterising the systems evolutions and controller using PDFs, and then using the KLD to measure discrepancies between desired closed system and actual system behaviours. The result is an optimised randomised controller that better reflects and considers the inherent uncertainty and stochasticity of real-world control systems.

TABLE 1. Performance metrics over the 5-s horizon.

Controller	MSE	$\overline{\mathcal{D}}$	E_u
FPC	0.0205	0.3436	13.09
SLQ	0.0208	0.3494	16.62

In particular, based on the SDEs that describe the dynamics of stochastic systems, we reformulated the HJB principle in terms of joint PDFs. Our probabilistic formulation yielded a more accurate control strategies that better capture the stochastic nature of the systems. The derived probabilistic strategies as given in Theorem 3.1 are shown to be universal and suitable for any generative probabilistic model describing the system dynamics. It applies to any stochastic system regardless of whether its dynamics follow the FP equation's assumptions.

To demonstrate the theoretical development of our approach, we applied the general solution to linear Gaussian systems. Because these systems are both linear and driven by Gaussian noise, we were able to derive explicit, closed-form expressions for the optimal probabilistic control. Specifically, we have shown that the mean of the optimal controller corresponds to a standard state feedback control law, while the covariance is derived analytically in terms of system parameters such as the noise intensity (σ_t), the control matrix (B), and the target controller distribution (Γ). This demonstration illustrated how the general framework can yield exact solutions when applied to well-structured systems. It also shows how our method extends traditional control strategies by producing a full probability distribution over control actions, rather than a single deterministic path.

The simulation results on the OU process have shown that our probabilistic framework can successfully steer the system's state distribution toward a desired target. Across different time windows, the computed distributions closely match the targets, demonstrating the method's ability to manage uncertainty effectively. We also compared our method to SLQ control method. The comparison results showed that our approach achieves lower mean squared error and better alignment with the target distribution, while maintaining competitive control effort.

To maintain a clear focus on establishing the theoretical groundwork, this paper concentrated on the general formulation (Thm. 3.1) and its analytic specialisation to linear–Gaussian dynamics. For nonlinear or high-dimensional systems, although a general closed form solution of the controller can be obtained as outlined in Theorem 3.1, its solution needs to be computed numerically. These problems are more complex due to the curse of dimensionality and the multiple integrations involved in the computation of optimal control strategies. To stay focused on the core theoretical contributions, we leave these challenges for future research. We plan to investigate low-rank or sparsity-exploiting Riccati solvers, moment-projection and neural-Galerkin schemes, as well as experimental validation and real-time implementations of the fully probabilistic framework.

DATA AVAILABILITY STATEMENT

The data used in this study were generated from numerical simulations using the model and parameters described in the paper. No external datasets were used and no experimental data were collected.

REFERENCES

[1] R. Bellman, I. Glicksberg and O. Gross, Some Aspects of the mathematical Theory of Control Processes. Rand Corporation, Santa Monica, California (1958).

[2] R.E. Kalman, Contributions to the theory of optimal control. *Bol. Soc. Mat. Mex.* **5** (1960) 102–119.

[3] H.J. Kushner, Optimal stochastic control. *IRE Trans. Autom. Control* **7** (1962) 120–122.

[4] W.M. Wonham, On a matrix Riccati equation of stochastic control. *SIAM J. Control* **6** (1968) 681–697.

[5] C.D. Charalambous and R.J. Elliott, Examples of optimal control for nonlinear stochastic control problems with partial information. *Proceedings of the 34th IEEE Conference on Decision and Control (CDC)*, vol. 3. (1995) 2187–2192.

[6] J. Huang and Z. Yu, Solvability of indefinite stochastic Riccati equations and linear quadratic optimal control problems. *Syst. Control Lett.* **68** (2014) 68–75.

[7] J. Sun, X. Li and J. Yong, Open-loop and closed-loop solvability for stochastic linear quadratic optimal control problems. *SIAM J. Control Optim.* **54** (2016) 2274–2308.

[8] J. Sun and J. Yong, Stochastic linear-quadratic optimal control problems – recent developments. *Annu. Rev. Control* **56** (2023) 100899.

[9] H. Zhang and X. Zhang, Stochastic linear quadratic optimal control problems with expectation-type linear equality constraints on the terminal states. *Syst. Control Lett.* **177** (2023) 105551.

[10] J.L. Lions, On the Hamilton–Jacobi–Bellman equations. *Acta Appl. Math.* **1** (1983) 17–41.

[11] R.C. Seydel, Existence and uniqueness of viscosity solutions for QVI associated with impulse control of jump-diffusions. *Stoch. Processes Appl.* **119** (2009) 3719–3748.

[12] M. Annunziato and A. Borzí, Optimal control of probability density functions of stochastic processes. *Math. Model. Anal.* **15** (2010) 393–407.

[13] M. Annunziato and A. Borzi, A Fokker–Planck control framework for stochastic systems. *EMS Surv. Math. Sci.* **5** (2018) 65–98.

[14] A. Bensoussan, J. Frehse and P. Yam, Mean field games and mean field type control theory. Springer Briefs in Mathematics. Springer, New York (2013). MR 3134900

[15] J.M. Lasry and P.L. Lions, Mean field games. *Jpn. J. Math.* **2** (2007) 229–260. MR 2295621

[16] J. Heng, A.N. Bishop, G. Deligiannidis and A. Doucet, *Controlled sequential Monte Carlo. Ann. Statist.* **48** (2020) 2904–2929.

[17] L.R. Ray and R.F. Stengel, A Monte Carlo approach to the analysis of control system robustness. *Automatica* **29** (1993) 229–236.

[18] H.J. Kappen, Path integrals and symmetry breaking for optimal control theory. *J. Statist. Mech. Theory Exp.* **2005** (2005) P11011.

[19] E. Theodorou, F. Stulp, J. Buchli and S. Schaal, An iterative path integral stochastic optimal control approach for learning robotic tasks. *IFAC Proc. Vol.* **44** (2011) 11594–11601. 18th IFAC World Congress.

[20] E.A. Theodorou, K. Dvijotham and E. Todorov, From information theoretic dualities to path integral and Kullback–Leibler control: continuous and discrete time formulations. *Proceedings of the American Control Conference (ACC)* (2013).

[21] E.A. Theodorou and E. Todorov, Relative entropy and free energy dualities: Connections to path integral and KL control. *Proceedings of the 51st IEEE Conference on Decision and Control (CDC)*. Maui, Hawaii, USA (2012) 1466–1473.

[22] K. Itô and K. Kashima, Kullback–Leibler control for discrete-time nonlinear systems on continuous spaces. *SICE J. Control Meas. Syst. Integr.* **15** (2022) 119–129.

[23] R. Herzallah and M. Kárný, Fully probabilistic control design in an adaptive critic framework. *Neural Netw.* **24** (2011) 1128–1135.

[24] M. Kárný, Towards fully probabilistic control design. *Automatica* **32** (1996) 1719–1722.

[25] B. Øksendal, Stochastic Differential Equations: An Introduction with Applications, 6th edn. Springer (2003).