

A Time-Reversal Control Synthesis for Steering the State of Stochastic Systems

Yuhang Mei^{*}, Amirhossein Taghvaei^{*}, Ali Pakniyat[†]

Abstract—This paper presents a novel approach for steering the state of a stochastic control-affine nonlinear system to a desired target within a finite time horizon. Our method leverages the time-reversal of diffusion processes to construct the required feedback control law. Specifically, the control law is the so-called score function associated with the time-reversal of random state trajectories that are initialized at the target state and are simulated backwards in time. A neural network is trained to approximate the score function, enabling applicability to both linear and nonlinear stochastic systems. Numerical experiments demonstrate the effectiveness of the proposed method across several benchmark examples.

I. INTRODUCTION

Steering the state of a stochastic system to a target state or distribution is a fundamental problem in stochastic control and useful in applications such as stochastic thermodynamics [1], [2], [3], machine learning [4], [5], [6], and robotics [7], [8], [9], [10], [11]. One prominent framework for studying this problem is the theory of Schrödinger bridges for diffusion processes [12], [13], which aims to find an optimal control law that drives the system from a given initial distribution to a specified target distribution over a finite horizon. Exact solutions are available for linear stochastic systems with Gaussian initial and target distributions [14], [15], [16], [17], and numerical procedures extend these results to non-Gaussian [18], [19] and nonlinear drift [20] cases. Beyond Schrödinger bridge theory, alternative approaches using the stochastic maximum principle and convex duality have been proposed to derive optimal control policies that steer the state toward desired target points [21], [22] or distributions [23], [24]. However, the reliance on optimality in both these and Schrödinger-based approaches makes controller identification for nonlinear settings computationally expensive, as it requires solving complex partial differential equations.

In contrast, this paper shifts attention away from optimality (and initial-state dependence) toward designing a feedback law that guarantees finite-time attractability of a chosen target. Specifically, we consider a stochastic process X that is governed by a control-affine nonlinear stochastic system (1) and propose a framework for deriving feedback control laws that steer the state X_t toward a target state $X_T = x_f$, at a finite time horizon T , without requiring optimality relative to the initial condition.

Yuhang Mei and Amirhossein Taghvaei are supported by the National Science Foundation (NSF) award EPCN-2318977 and EPCN-2347358.

^{*}Department of Aeronautics & Astronautics, University of Washington, Seattle; yuhangm@uw.edu, amirtag@uw.edu.

[†]Department of Mechanical Engineering, University of Alabama; apakniyat@ua.edu.

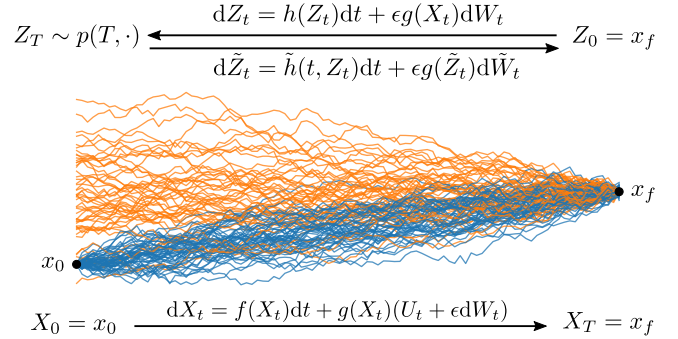


Fig. 1: Illustration of the proposed time-reversal methodology to steer the state X_t from x_0 to x_f . The auxiliary process Z_t is simulated backward in time from x_f . The dynamics of the time-reversal $\tilde{Z}_t := Z_{T-t}$ is forward in time, with an absorption property to x_f . This dynamics is used to design the control law $U_t = k(t, X_t)$. The expressions for the functions h , $\tilde{h}$, and k appear in Section III.

Our approach is inspired by time-reversal theory of diffusions [25], [26], [27], which has recently gained attention in machine learning through diffusion-based generative models for images [28], [29], [30], [31]. In these models, one first simulates a stochastic differential equation (SDE) that gradually adds noise to transform complex data (e.g., images) into a simpler, typically Gaussian distribution. The trajectory of this noising process is then used to learn the so-called score function, which in turn is used to run the SDE in reverse. Sampling from the Gaussian and applying this reverse process (the “denoising” procedure) generates new samples resembling the original data. From a control-theoretic viewpoint, the learned score function serves as a control law that transforms the Gaussian distribution to the target data distribution. A key benefit of this procedure is its computational tractability as it involves solving a regression problem rather than resorting to dynamic programming or the maximum principle, though at the cost of losing optimality.

Inspired by these diffusion generative models, we draw an analogy for our setting by simulating the state of stochastic system backward in time from a given target state x_f (or a small Gaussian distribution around it). Reversing this backward process in the forward-time direction ensures convergence to the target, with the learned score function serving as the feedback control law. To provide an intuitive understanding of our approach, we outline the key steps involved in constructing the proposed feedback control law, accompanied with an illustration in Figure 1.

- 1) **Auxiliary Process Z :** We construct an auxiliary process Z , initialized at the desired terminal state x_f . This process generates a probability density function $p(t, x)$ with a delta distribution at the desired terminal location at its initial time, i.e., $p(0, x) = \delta_{x_f}(x)$ while producing spread-out distributions at other times. This serves as the foundation for defining a time-reversed process.
- 2) **Time-Reversed Auxiliary Process $\tilde{Z}$:** By reversing time for the auxiliary process Z , we obtain a new process $\tilde{Z}_t = Z_{T-t}$ with the probability density function $\tilde{p}(t, x) = p(T-t, x)$ possessing a delta distribution at the desired terminal location (matching the target state) at its terminal time, i.e., $\tilde{p}(T, x) = \delta_{x_f}(x)$. The governing dynamics of $\tilde{Z}$ possess an absorption property for the desired terminal state, i.e., almost all sample paths are attracted to and absorbed by the desired state. This property is critical for ensuring that trajectories converge to the target state.
- 3) **Construction of Feedback Control Law for X :** The feedback law synthesis for X leverages the absorption property of the time-reversed process $\tilde{Z}$, to yield closed loop dynamics that enforce almost sure convergence to the target state within a fixed time horizon. Crucially, if the initial condition x_0 of the original process X lies within the support of the initial distribution of $\tilde{Z}$, then x_0 belongs to the fixed-time stochastic region of attraction of the desired terminal state almost surely.

Relation to diffusion bridges: A diffusion bridge is defined as the law of the diffusion processes X conditioned on the event that $X_0 = x_0$ and $X_T = x_f$. Using Doob's h -transform [32, Sec. 7.5], the conditioned process satisfies the same SDE as the original diffusion but with an added drift that can be interpreted as a control law steering the state to the target. However, this control law is distinct from the one we propose in this paper. In particular, our control law is not generally of gradient form, whereas the Doob h -transform yields the gradient of the log-transition density. Interestingly, the two control laws coincide in the linear Gaussian setting, suggesting a deeper connection between the approaches. From a computational standpoint, there is a recent work [33] that applies time-reversal and score matching to simulate diffusion bridges. However, their proposed method requires applying time reversal and score-function approximation twice [33, Sec. 2.3], with potential error accumulation, while our method requires only a single such operation. A detailed numerical comparison to fully characterize the trade-offs is planned for future investigation.

The paper is organized as follows. Section II presents the problem formulation and the necessary background on the time-reversal theory of diffusions. Section III presents our proposed time-reversal methodology for control synthesis, along with the analysis of the linear Gaussian setting. Section IV presents numerical experiments on several benchmark examples, demonstrating the performance of the proposed method in both linear and nonlinear settings.

A. Notation

The notation $\frac{\partial}{\partial x}$ is used to denote the derivative with respect to the x variable. For example, for a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\frac{\partial f}{\partial x} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes the gradient of f with respect to x . And for a smooth vector-field $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\frac{\partial f}{\partial x} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ denotes the Jacobian. The probability density function (PDF) of a multivariate Gaussian distribution with mean vector m and covariance matrix Σ is denoted by $\mathcal{N}(x; \mu, \Sigma)$. For a positive definite matrix P , the weighted 2-norm of a vector $x \in \mathbb{R}^n$, is denoted by $\|x\|_P$, and defined as $\|x\|_P := \sqrt{x^\top P x}$.

II. PROBLEM FORMULATION AND BACKGROUND

A. Problem setup

Consider a control system governed by a control-affine stochastic differential equation (SDE)

$$dX_t = f(X_t)dt + g(X_t)(U_t dt + \epsilon dW_t), \quad X_0 = x_0, \quad (1)$$

where $X_t \in \mathbb{R}^n$ is the state, $U_t \in \mathbb{R}^m$ is the control input, $W_t \in \mathbb{R}^m$ is the standard Wiener process that denotes the process noise, $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ are drift and diffusion functions, and $\epsilon > 0$ is a positive parameter that denotes the strength of the noise. Let $\mathcal{F}_t = \sigma(W_s; 0 \leq s \leq t)$ be the filtration generated by the Wiener process. The control input U_t is constrained to be adapted to the filtration $\mathcal{F}_t$. We are interested in the problem of designing the control input that steers the state of the system to a given target state.

Problem 1: Given a target state $x_f \in \mathbb{R}^n$ and a fixed terminal time $T > 0$, find a feedback control law $\{U_t = k(t, X_t); t \in [0, T]\}$, for a function $k : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$, such that $X_T = x_f$ almost surely.

Note that the $\mathcal{F}_t$ -adaptability condition is automatically satisfied when U_t is expressed as a function of X_t . We also consider a relaxed version of the problem, where the equality is relaxed to a bound on the average distance from the target. This relaxation allows us to obtain control laws that are not singular as $t \rightarrow T$. See Section III-B.

Problem 2: Given a target state $x_f \in \mathbb{R}^n$, a fixed terminal time $T > 0$, and error tolerance $\delta > 0$, find a feedback control law $\{U_t = k(t, X_t); t \in [0, T]\}$, for a function $k : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$, such that $\mathbb{E}[\|X_T - x_f\|^2] \leq \delta$.

Remark 1 (Modelling assumptions): The control-affine structure of the dynamic model in (1) is critical for the applicability of the proposed method. This model is common in robotics, aerospace, and finance [34], [35], [36]. The time-invariant assumption for the functions f and g is made for ease of presentation and can be relaxed to be time-varying.

Our solution methodology for these problems is based on the time-reversal of diffusion theory, which we review next.

B. Time-Reversal of Diffusions

Consider a diffusion process $Z := \{Z_t \in \mathbb{R}^n; 0 \leq t \leq T\}$ governed by the following SDE

$$dZ_t = h(Z_t)dt + \epsilon g(Z_t)dW_t, \quad Z_0 = z, \quad (2)$$

where $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a drift function, and z is an initial condition, that are later designed as part of our methodology, in Section III-A. We make the following assumption about the functions h and g .

Assumption 1: The functions h and g are smooth and globally Lipschitz.

Assumption 1 implies that SDE (2) admits a unique strong solution [37, Thm. 5.2.1.]. This allows us to define the time-reversal of Z_t according to

$$\tilde{Z} := \{\tilde{Z}_t = Z_{T-t}; 0 \leq t \leq T\}.$$

The time-reversal theory is concerned with obtaining the SDE for the reversed process $\tilde{Z}_t$. In order to do so, we follow the approach presented in [26] which, in addition to Assumption 1, requires a suitable integrability condition on the density of Z_t . In particular, letting $p(t, x)$ denote the probability density function of Z_t , it is required that

$$\int_{t_0}^T \int_{\Omega} |p(t, x)|^2 + \|g(x)^\top \frac{\partial p}{\partial x}(t, x)\|^2 dx dt < \infty, \quad (3)$$

for any open bounded set $\Omega \subset \mathbb{R}^n$ and $t_0 > 0$. This condition is valid under the following assumption about the drift and diffusion functions.

Assumption 2: For all $x \in \mathbb{R}^n$, the subspace generated by the Lie-brackets¹ of the elements of $\{h(x), g_1(x), \dots, g_m(x)\}$, with g_i the i -th column of g , spans the entire space $\mathbb{R}^n$.

Assumption 2 is known as the Hörmander condition that ensures the generator associated with the diffusion process (2) is hypoelliptic, implying that Z_t admits a smooth density $p(t, x)$ for any $t > 0$ [38], [39]. As a result, condition (3) is satisfied. With a smooth density, one can define the so-called score function $s : (0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose i -th component is given by

$$s_i(t, x) := \frac{1}{p(t, x)} \sum_{j=1}^n \frac{\partial}{\partial x_j} (G_{i,j}(x) p(t, x)), \quad (4)$$

where $G_{i,j}(x)$ is the (i, j) -entry of the matrix $G(x) := g(x)g(x)^\top \in \mathbb{R}^{n \times n}$. Then, according to [26, Thm. 2.1], the reversed process $\tilde{Z}_t$ satisfies the SDE

$$d\tilde{Z}_t = -h(\tilde{Z}_t) dt + \epsilon^2 s(T-t, \tilde{Z}_t) dt + \epsilon g(\tilde{Z}_t) d\tilde{W}_t, \quad (5)$$

where $\tilde{W}_t \in \mathbb{R}^m$ is the standard m -dimensional Wiener process. Note that the reversed process satisfies the condition $\tilde{Z}_T = Z_0 = z$. Therefore, the dynamics generated by the score function and the drift term has an absorption property for the state z at terminal time T . This is the basis for our methodology in Section III-A.

Remark 2: The SDE for the time-reversal process may also be obtained using Girsanov theorem, as in [27], but this alternative approach requires the diffusion function to be non-singular, i.e. $G(x) := g(x)g(x)^\top$ be uniformly positive-definite for all $x \in \mathbb{R}^n$ [25], [27]. This assumption is

¹The Lie-bracket of two smooth vector-fields $g_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $g_j : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined as $[g_i, g_j](x) = \frac{\partial g_j}{\partial x}(x)g_i(x) - \frac{\partial g_i}{\partial x}(x)g_j(x)$

restrictive due to the role that the function g plays in the stochastic control problem (1). In particular, the columns of g are the directions that the control input affects the state. Therefore, a positive-definite assumption on G implies full control authority which is restrictive in many control applications. Assumption 2 is a weaker assumption that allows the construction of the time-reversal SDE for a degenerate diffusion. It is also in agreement with the local accessibility condition in geometric control theory for the deterministic version of the system (1), when $\epsilon = 0$ [40], [41].

The formula for the score function (4) depends on the density $p(t, x)$ which is explicitly available only in the special linear Gaussian setting (see Section III-C). In a general nonlinear and non-Gaussian setting, the score function is approximated as the solution to a stochastic optimization problem, as described in the next subsection.

C. Score function approximation

It is numerically useful to note that the score function is the solution to the minimization problem $\min_{\psi} J(\psi)$ where

$$J(\psi) := \mathbb{E} \left[\frac{1}{2} \|\psi(t, Z_t)\|^2 + \sum_{i,j=1}^n G_{i,j}(Z_t) \frac{\partial \psi_i}{\partial x_j}(t, Z_t) \right], \quad (6)$$

and the expectation is both over $t \sim \text{Unif}[0, T]$ and Z_t , solution to (2). This follows by writing the expectation as the integral with respect to the density $p(t, x)$, application of integration by parts on the second term, and expressing the objective function (6) as

$$\mathbb{E} \left[\frac{1}{2} \|\psi(t, Z_t) - s(t, Z_t)\|^2 \right] + (\text{constant}),$$

where the constant is independent of ψ [42, Thm. 1]. The optimization (6) is known as implicit score matching [43]. This optimization procedure is later used in the construction of our numerical algorithm in Section IV-A.

D. Probability transition kernels

This section introduces notations and definitions for probability transition kernels associated with SDEs (2) and (5) that are later used in the proof of Propositions 1 and 2. Let $\kappa_{t,s}(x'|x)$ denote the probability transition kernel from time s to time t that is associated with SDE (2), i.e. the conditional probability density function of $Z_t = x'$ given $Z_s = x$, for any $t \geq s > 0$. Similarly, let $\tilde{\kappa}_{t,s}(x'|x)$ denote the probability transition kernel for SDE (5). Using the probability transition kernel, the joint probability density function of (Z_t, Z_s) satisfies

$$\int_{B_x \times B_{x'}} P_{Z_t, Z_s}(x', x) dx^\top dx' = \int_{B_x \times B_{x'}} \kappa_{t,s}(x'|x) p(s, x) dx^\top dx'$$

for arbitrary Borel sets $B_x, B_{x'} \subset \mathbb{R}^n$, implying

$$P_{Z_t, Z_s}(x', x) = \kappa_{t,s}(x'|x) p(s, x).$$

Similarly, for $(\tilde{Z}_t, \tilde{Z}_s)$ we have

$$P_{\tilde{Z}_t, \tilde{Z}_s}(x', x) = \tilde{\kappa}_{t,s}(x'|x) p(T-s, x),$$

where we used the fact that $\tilde{Z}_s = Z_{T-s}$ with probability density $p(T-s, x)$. The identity $(Z_T, Z_{T-t}) = (\tilde{Z}_0, \tilde{Z}_t)$ implies the equality $P_{Z_T, Z_{T-t}}(x', x) = P_{\tilde{Z}_t, \tilde{Z}_0}(x, x')$, concluding the relationship

$$\kappa_{T, T-t}(x'|x)p(T-t, x) = \tilde{\kappa}_{t, 0}(x|x')p(T, x') \quad (7)$$

between the two kernels κ and $\tilde{\kappa}$.

III. PROPOSED METHODOLOGY

A. Solution to problem 1

We propose to solve Problem 1 using the time-reversal theory that is presented in Section II-B. We start with the process Z_t from (2) and initialize it at $Z_0 = x_f$. With this initialization, the reversed process $\tilde{Z}_t := Z_{T-t}$ satisfies the terminal condition $\tilde{Z}_T = Z_0 = x_f$, and its dynamics has an absorption property for x_f . Therefore, in order to solve Problem 1, we design h so that SDE (5) for $\tilde{Z}_t$ takes a form similar to (1) for X_t . Using the decomposition of the score function as

$$s(t, x) = g(x)k^*(t, x) + \mathbf{g}(x) \quad (8)$$

where the i -th component of $k^* : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are defined as

$$k_i^*(t, x) := \frac{1}{p(t, x)} \sum_{j=1}^n \partial_{x_j} (g_{j,i}(x)p(t, x)), \quad (9)$$

$$\mathbf{g}_i(t, x) := \sum_{j=1}^n \sum_{k=1}^m g_{j,k}(x) \partial_{x_j} g_{i,k}(x), \quad (10)$$

the SDE (5) takes the form

$$\begin{aligned} d\tilde{Z}_t = & (-h(\tilde{Z}_t) + \epsilon^2 \mathbf{g}(\tilde{Z}_t)) dt \\ & + g(\tilde{Z}_t)(\epsilon^2 k^*(T-t, \tilde{Z}_t) dt + \epsilon d\tilde{W}_t). \end{aligned}$$

Designing the function h and control law k according to

$$h(x) := -f(x) + \epsilon^2 \mathbf{g}(x), \quad (11)$$

$$k(t, x) := \epsilon^2 k^*(T-t, x), \quad (12)$$

yields the following expressions for SDEs of Z_t and $\tilde{Z}_t$:

$$dZ_t = (-f(Z_t) + \epsilon^2 \mathbf{g}(Z_t)) dt + \epsilon g(Z_t) dW_t, \quad Z_0 = x_f \quad (13a)$$

$$d\tilde{Z}_t = f(\tilde{Z}_t) dt + g(\tilde{Z}_t)(k(t, \tilde{Z}_t) dt + \epsilon d\tilde{W}_t). \quad (13b)$$

Moreover, using the control law (12) in (1) concludes the following SDE for X_t :

$$dX_t = f(X_t) dt + g(\tilde{X}_t)(k(t, X_t) dt + \epsilon dW_t), \quad X_0 = x_0. \quad (13c)$$

In summary, we have constructed three stochastic processes:

- 1) The process Z_t that solves (13a) from initial condition $Z_0 = x_f$. The density of this process is denoted by $p(t, x)$;
- 2) The process $\tilde{Z}_t := Z_{T-t}$ that solves (13b) and satisfies the condition $\tilde{Z}_T = Z_0 = x_f$. The control law $k(t, x)$ is defined in (12) and (9);

- 3) The process X_t that solves (13c) starting from the initial condition $X_0 = x_0$.

Note that, the SDE for $\tilde{Z}_t$ has the same form as the SDE for X_t and the control law steers the process $\tilde{Z}_t$ to the target state $\tilde{Z}_T = x_f$. However, it remains to be shown that the control law steers X_t to $X_T = x_f$, despite the difference in the initial conditions; $\tilde{Z}_0 = Z_T$ is random with probability density function $p(T, x)$, whereas $X_0 = x_0$ is deterministic. The next result shows that X_t reaches the target state $X_T = x_f$ whenever the initial condition satisfies $p(T, x_0) > 0$.

Proposition 1: Let $p(t, x)$ denote the probability density function of Z_t defined according to (13a) and define the control law $k(t, x)$ according to (12) and (9). If the initial condition $X_0 = x_0$ satisfies $p(T, x_0) > 0$, then, problem 1 is solved with the feedback control law $k(t, x)$.

Proof: Let $\kappa_{t,s}(x'|x)$ and $\tilde{\kappa}_{t,s}(x'|x)$ be the probability transition kernels for SDEs (13a) and (13b), respectively, as defined in Section II-D. The probability transition kernel associated with the SDE (13c) is also $\tilde{\kappa}_{t,s}(x'|x)$ due to the fact that SDEs (13b) and (13c) are identical. Therefore, with the initial condition $X_0 = x_0$, the probability density function of X_t becomes equal to $\tilde{\kappa}_{t,0}(x|x_0)$. The goal is to show that $\tilde{\kappa}_{t,0}(x|x_0)$ approaches the Dirac delta distribution $\delta_{x_f}(x)$ (in the weak sense) as t approaches T . The identity (7) implies

$$\begin{aligned} \tilde{\kappa}_{t,0}(x|x_0) &= \frac{\kappa_{T, T-t}(x_0|x)p(T-t, x)}{p(T, x_0)} \\ &= \frac{\kappa_{T, T-t}(x_0|x)p(T-t, x)}{\kappa_{T,0}(x_0|x_f)} \end{aligned}$$

where we used the assumption that $p(T, x_0) > 0$ and the fact that $p(T, x) = \kappa_{T,0}(x|x_f)$ due the initial condition $Z_0 = x_f$. Taking the limit as $t \rightarrow T$ and using the fact that $p(T-t, x)$ approaches $\delta_{x_f}(x)$ concludes the result. ■

B. Avoiding singularity of the control law

The feedback control law (12) becomes singular in the limit as t approaches the terminal time T because the distribution $p(T-t, x)$ approaches the Dirac delta distribution $\delta_{x_f}(x)$. The singularity is unavoidable when an almost sure constraint $X_T = x_f$ is required. In order to avoid the singularity, we consider Problem 2 where the almost sure equality is relaxed to a bound on the average distance to the target. We propose to solve problem 2 using the time-reversal procedure presented in Section III-A, with the difference that the process Z_0 is now initialized from a Gaussian distribution around x_f , i.e. $\mathcal{N}(\cdot; x_f, \sigma^2)$ with $\sigma > 0$. We prove that, with a small enough σ , the resulting control law is able to solve problem 2 while maintaining non-singularity, in the linear Gaussian setting, with error bounds established in Proposition 2, while deferring the analysis of error bounds for the general nonlinear case to future work.

C. Analysis of the linear Gaussian setting

In this section, we study the proposed time-reversal method in the linear Gaussian setting, where

$$f(x) = Ax, \quad g(x) = B,$$

for matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$. In this case, f is linear and g is constant, implying that both functions are smooth and globally Lipschitz. Therefore, Assumption 1 is satisfied. Furthermore, Assumption 2 will also hold whenever (A, B) is controllable.

Under such setting, the SDEs (13a), (13b), and (13c) take the form

$$dZ_t = -AZ_t dt + \epsilon B dW_t, \quad Z_0 = x_f, \quad (14a)$$

$$d\tilde{Z}_t = A\tilde{Z}_t dt + B(k(t, \tilde{Z}_t) dt + \epsilon d\tilde{W}_t), \quad \tilde{Z}_T = x_f, \quad (14b)$$

$$dX_t = AX_t dt + B(k(t, X_t) dt + \epsilon dW_t), \quad X_0 = x_0. \quad (14c)$$

The probability density of Z_t is Gaussian, for all $t > 0$, because (14a) is linear and the initial state Z_0 is deterministic. In particular, $p(t, x) = \mathcal{N}(x; m_t, \Sigma_t)$ where the mean and covariance are given by

$$m_t = e^{-At} x_f, \\ \Sigma_t = \epsilon^2 \int_0^t e^{-A(t-s)} B B^\top e^{-A^\top(t-s)} ds.$$

Substitution of the Gaussian distribution formula in (9), and the use of (12), yields the following formula for the feedback control law:

$$k(t, x) = -\epsilon^2 B^\top \Sigma_{T-t}^{-1} (x - m_{T-t}), \quad (15)$$

for $t \in [0, T]$. This is similar to the control law that appears in [21, Eq. (11)] and [44], obtained through a stochastic optimal control formulation.

It is worth remarking that in the limit as $t \rightarrow T$, the covariance $\Sigma_{T-t} \rightarrow 0$, thus the control law becomes singular, as described in Section III-B. To resolve the singularity issue, we initialize $Z_0 \sim \mathcal{N}(\cdot; x_f, \sigma^2 I)$. Under this initial condition, the distribution of Z_t remains Gaussian $\mathcal{N}(x; m_t, Q_t)$ with the same mean as before, but with a new covariance that is given by $Q_t = \Sigma_t + \sigma^2 e^{-At} e^{-A^\top t}$. The resulting feedback control law is

$$\hat{k}(t, x) = -\epsilon^2 B^\top Q_{T-t}^{-1} (x - m_{T-t}). \quad (16)$$

The new feedback control law remains nonsingular as $t \rightarrow T$. However, there is no guarantee that it would steer X_t to the target state x_f . The following proposition characterizes the error $\mathbb{E}[\|X_T - x_f\|^2]$ when the control law (16) is used instead of (15).

Proposition 2: Under the feedback control law defined in (16), the expected squared error between the terminal state and the target state is given by

$$\mathbb{E}[\|X_T - x_f\|^2] = \sigma^4 \|e^{TA} x_0 - x_f\|_{M^{-2}}^2 + \sigma^2 (n - \text{Tr}(M^{-1})), \quad (17)$$

where $M := \sigma^2 I + \epsilon^2 \int_0^T e^{As} B B^\top e^{A^\top s} ds$. Moreover, for any $\delta > 0$, there exists a small enough $\sigma > 0$ that solves Problem 2.

Proof: Let κ and $\tilde{\kappa}$ denote the probability transition kernels associated with SDEs (14a) and (14b), respectively, similar to the proof of Proposition 1. In terms of the kernels, the probability distribution of X_T is equal to $\tilde{\kappa}_{T,0}(\cdot|x_0)$, because the SDE (14c) and (14b) have the same form and

Algorithm 1 Time-Reversal Control Synthesis

- 1: **Input:** sample size N , step-size Δt , variance σ , deterministic control input u_t , function class Ψ .
 - 2: $\{Z_0^i\}_{i=1}^N \sim \mathcal{N}(x_f, \sigma^2 I)$
 - 3: **for** $t \in \{\Delta t, 2\Delta t, \dots, T - \Delta t, T\}$ **do**
 - 4: $\{\Delta W_t^i\}_{i=1}^N \sim N(0, \Delta t I_n)$
 - 5: $Z_{t+\Delta t}^i = Z_t^i + (-f(Z_t^i) + \epsilon^2 g(Z_t^i) + g(Z_t^i) u_t) \Delta t + \epsilon g(Z_t^i) \Delta W_t^i$
 - 6: **end for**
 - 7: $k^*(t, \cdot) = \arg \min_{k \in \Psi} \frac{1}{N} \sum_{i=1}^N [\frac{1}{2} \|g(Z_t^i) k(t, Z_t^i) + g(Z_t^i)\|^2 + \sum_{j,l}^n (G_{j,l}(Z_t^i) \partial_{x_l} (g(Z_t^i) k(t, Z_t^i) + g(Z_t^i))_j)]$
 - 8: **Output:** $\{k^*(t, \cdot)\}_{t \in \{0, \Delta t, \dots, T\}}$
-

$X_0 = x_0$. Then, upon the application of the time-reversal relationship (7), and the fact that $p(t, x) = \mathcal{N}(x; m_t, Q_t)$,

$$\begin{aligned} \tilde{\kappa}_{T,0}(x|x_0) &= \frac{\kappa_{T,0}(x_0|x)p(0,x)}{p(T,x_0)} \\ &= \frac{\mathcal{N}(x_0; e^{-AT} x, \Sigma_T) \mathcal{N}(x; x_f, \sigma^2 I)}{\mathcal{N}(x_0; m_T, Q_T)} \\ &= \mathcal{N}(x; \mu, P) \end{aligned}$$

where

$$\begin{aligned} \mu &= x_f + \sigma^2 M^{-1} e^{TA} (x_0 - e^{-TA} x_f), \\ P &= \sigma^2 (I - \sigma^2 M^{-1}). \end{aligned}$$

Now, to compute the error $\mathbb{E}[\|X_T - x_f\|^2]$, we use the fact that $X_T \sim \mathcal{N}(\cdot; \mu, P)$ and, hence:

$$\mathbb{E}[\|X_T - x_f\|^2] = \|\mu - x_f\|^2 + \text{Tr}(P)$$

which yields (17). Moreover, in the limit as $\sigma \rightarrow 0$, M converges to the controllability grammian matrix which is positive definite under the controllability assumption. Therefore, taking the limit of (17) as $\sigma \rightarrow 0$, concludes that the error converges to zero, implying that for any $\delta > 0$, there exists a $\sigma > 0$ such that the error is smaller than δ . ■

Remark 3: The error (17) comprises both a bias term and a variance term. The variance term is independent of x_0 and x_f , and bounded by $\sigma^2 n$. In contrast, the bias term can be significant when $e^{TA} x_0$ and x_f are far apart. This term arises due to the difference between the mean $\mathbb{E}[Z_T] = e^{-TA} x_f$ and x_0 . To address this issue, we allow our proposed methodology the flexibility to incorporate a deterministic control input u_t when Z_t is simulated, and modify the control law to $U_t = k(t, X_t) + \tilde{u}_t$ with $\tilde{u}_t = -u_{T-t}$ ². The addition of the deterministic input decreases the bias error by bringing the mean of Z_T and x_0 closer. Infact, the difference becomes zero in the linear case when (A, B) is controllable. The details of this modification to the algorithm appears in IV-A.

²Our method can also incorporate a feedback control law in addition to the deterministic input. We do not pursue this in the paper, as computing a feedback law is considerably more challenging than finding a deterministic input.

IV. NUMERICAL RESULTS

A. Numerical Algorithm

In this section, we introduce our proposed numerical algorithm which is based on the methodology described in Section III. The algorithm starts with simulating N random realizations $\{Z_t^i\}_{i=1}^N$ of the process (13a), with the flexibility of considering an additional deterministic control input u_t . The simulation is carried out using the Euler-Maruyama discretization method with the step-size Δt . In order to find the control law (12), we use the decomposition of the score function (8) and modify the score function optimization problem (6) according to $\min_k J(gk + g)$. To solve this optimization problem, we parameterize k with a neural network with a 3-block ResNet architecture where each block consists of 2 linear layers of width 32 and an exponential linear unit (ELU)-type activation function. We use ADAM optimizer to find the parameters of the neural network. The batch is generated by uniformly sampling $K_1 = \lfloor \frac{T}{10\Delta t} \rfloor + 1$ time instants $\{t_1, t_2, \dots, t_{K_1}\}$ from $[0, T]$, and $K_2 = 32$ random samples of the N trajectories $\{Z_{t_1}^i, \dots, Z_{t_{K_1}}^i\}_{i=1}^N$. The details of the algorithm appear in Algorithm 1. The code for reproducing the results is available online³.

The deterministic control input u_t is designed to approximately bring Z_T to the vicinity of x_0 . For example, this control may be obtained by the application of trajectory optimization techniques to the deterministic version of the model (13a), when $\epsilon = 0$. In addition to decreasing the bias error, the control input u_t serves as an ‘‘importance sampling’’ mechanism that guides Z_t to be sampled in areas of the domain where the control law is more relevant to the initial condition x_0 , thus increasing sampling efficiency.

B. Two-dimensional Brownian Bridge

We consider a 2-dimensional system governed by the SDE

$$dX_t = U_t dt + \epsilon dW_t, \quad X_0 = x_0, \quad (18)$$

and let $\epsilon = 0.3$, $x_0 = (0, 0)^\top$, $T = 1$, and $x_f = (2, 2)^\top$. This is a linear Gaussian model with $A = 0$ and $B = I$. We employ the deterministic control $u_t = x_0 - x_f$ which brings $Z_0 = x_f$ to $Z_1 = x_0$ in the deterministic setting. In this case, the resulting control law for X_t takes the form

$$U_t = x_f - x_0 - \frac{\epsilon^2(X_t - (1-t)x_0 - tx_f)}{\epsilon^2(1-t) + \sigma^2}. \quad (19)$$

We apply Algorithm 1 with $u_t = x_0 - x_f$, $N = 1000$, $\sigma = 0$, and $\Delta t = 0.004$. The resulting trajectories $\{Z_t^i\}_{i=1}^N$ and $\{X_t^i\}_{i=1}^N$, along with the control inputs $\{U_t^i\}_{i=1}^N$, are shown in Fig. 2. The result demonstrates that the control law obtained from Algorithm 1 successfully steers all trajectories X_t^i from x_0 to desired target x_f .

In order to quantify the performance of the control law, we introduce the following mean-squared-error (MSE) criteria

$$MSE = \frac{1}{N} \sum_{i=1}^N \|X_T^i - x_f\|_2^2. \quad (20)$$

We investigate the influence of the time step Δt and standard deviation σ on the MSE criteria. The result for varying time step-size is presented in Fig 3a, where we fix $\sigma = 0$. We also show the MSE corresponding to implementing the exact form of the control law (19) and the open-loop control $U_t = x_f - x_0$ as baselines. It is observed that the algorithm performs almost as good as the exact solution, and the MSE decreases as $\Delta t \rightarrow 0$. The result for varying σ is presented in Fig 3b with fixed $\Delta t = 0.004$, where for comparison, the case without using the deterministic input is also included. It is observed that including the deterministic input significantly decreases the MSE when $\sigma > 0$, justifying the remark 3. Moreover, σ acts as a regularizer in the optimization procedure and decreases the difference between Algorithm 1 and the exact solution. Although increasing σ increases MSE, it serves to avoid singularity of the control, which is shown by computing the averaged control energy

$$U_{norm} = \frac{1}{N} \sum_{i=1}^N \int_0^T \|U_t^i\|^2 dt \quad (21)$$

as a function of σ , in Fig 3c. The MSE and U_{norm} are both averaged over 5 independent experiments, where the shaded region represents the range from the minimum to the maximum across experiments.

C. Inverted Pendulum

We consider the stochastic pendulum dynamics with

$$f(x) = \begin{bmatrix} x(2) \\ \sin(x(1)) - 0.01x(2) \end{bmatrix}, \quad g(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$\epsilon = 0.3$, $x_0 = [\pi, 0]^\top$, $x_f = [0, 0]^\top$, and $T = 5$. The first component of the state is the angle where the angle equal to 0 denotes the upward position of the pendulum and π denote the downward. The goal is to bring the pendulum from the downward position to the upward position. We apply Algorithm 1 with $N = 1000$, $\sigma = 0$, $u_t = 0$, and $\Delta t = 0.004$. The resulting trajectories and control inputs are shown in Fig. 4, where the successful steering of the pendulum to the upward position is demonstrated.

V. CONCLUSION

This paper introduces a novel approach for steering nonlinear stochastic control-affine systems to a desired target state within a finite time horizon, leveraging time-reversal theory of diffusions. By constructing feedback control laws based on the score function associated with the reversed dynamics, the proposed method ensures finite-time convergence to the target state. Unlike traditional Schrödinger bridge methods or stochastic optimal control formulations, our approach is computationally efficient and applicable to both linear and nonlinear stochastic systems without relying on optimality relative to the initial condition.

An extension of the theory to address practical challenges related to inevitable singularities in the control law near the terminal time is also presented through relaxation of the almost-sure constraint to a distribution constraint and explicit error bounds of this approach are provided for the

³<https://github.com/YuhangMeiUW/P2P>

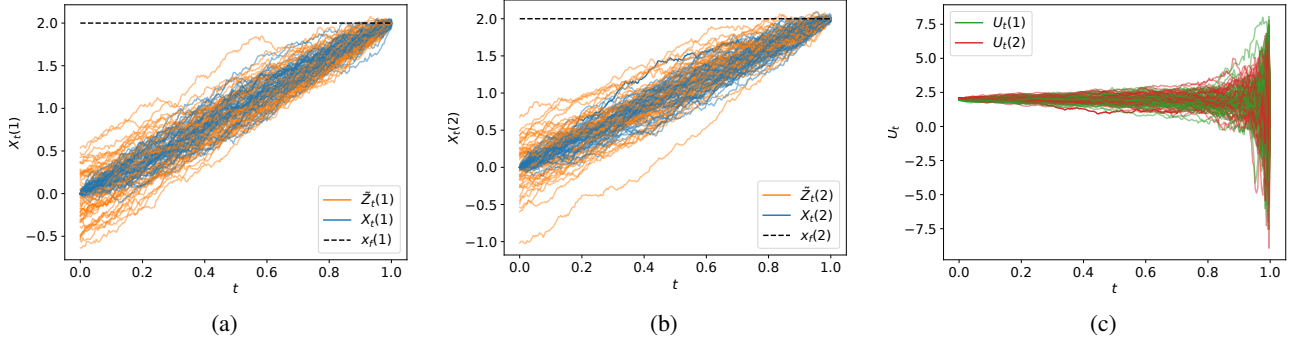


Fig. 2: Numerical result for the application Algorithm 1 to the two-dimensional Brownian bridge example of Section IV-B: (a) First component of X_t and $\tilde{Z}_t$ (b) Second component of X_t and $\tilde{Z}_t$ (c) Control input U_t .

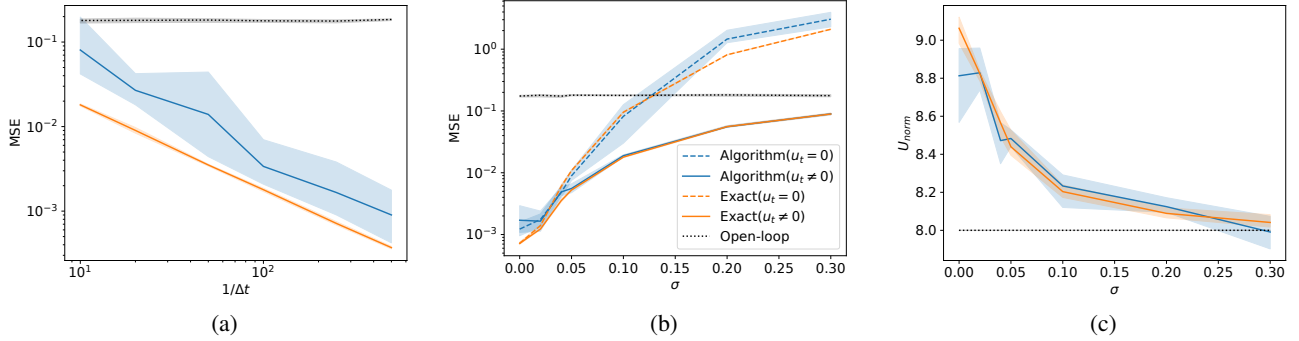


Fig. 3: Numerical error analysis for the application Algorithm 1 to the two-dimensional Brownian bridge example of Section IV-B: (a) Influence of time step-size on MSE (20), with $\sigma = 0$ (b) Influence of σ on MSE, with $\Delta t = 0.004$ (c) Influence of σ on U_{norm} (21), with $\Delta t = 0.004$. The results compare (i) Algorithm 1; (ii) the exact solution (19); and implementing the open loop control $U_t = x_f - x_0$. Panel (b) also includes the case where the deterministic input $u_t = 0$ in both exact solution and Algorithm 1.

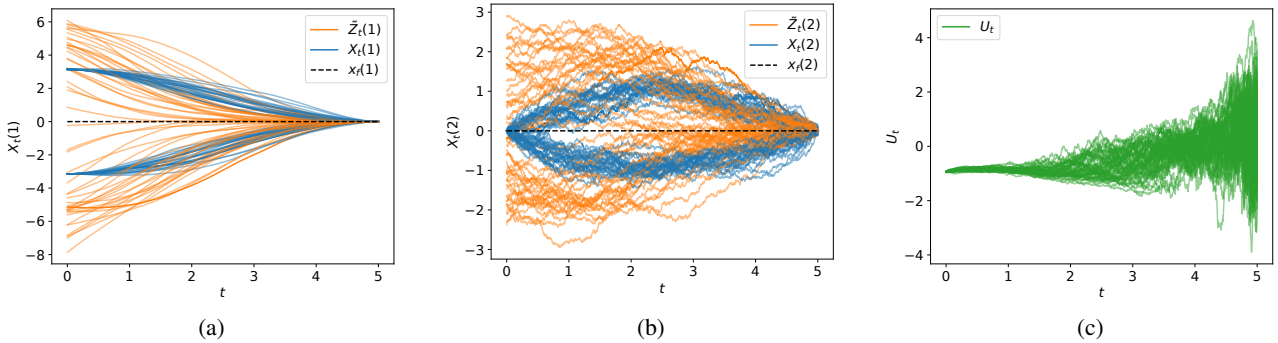


Fig. 4: Numerical result for the application Algorithm 1 to the inverted pendulum example of Section IV-C: (a) First component of X_t and $\tilde{Z}_t$; (b) Second component of X_t and $\tilde{Z}_t$; (c) Control input U_t . The first component of the initial state is equal to $\pm\pi$ and represents the downward position of the pendulum, while the first component of the terminal state is equal to 0, representing the upward position.

linear Gaussian case. Numerical experiments demonstrate the effectiveness of the method across benchmark examples, including a Brownian bridge and inverted pendulum dynamics.

Future research includes the extension of the theoretical results for assigning terminal distribution constraints in the general nonlinear system setting, the improving score function approximation techniques, exploring the optimality properties of the proposed control law through its connection

to Doob's h -transform, and numerical verification of our method on real-world control applications in Aerospace [45], [46].

REFERENCES

- [1] L. Peliti and S. Pigolotti, *Stochastic thermodynamics: an introduction*. Princeton University Press, 2021.
- [2] Y. Chen, T. T. Georgiou, and A. Tannenbaum, "Stochastic control and nonequilibrium thermodynamics: Fundamental limits," *IEEE transactions on automatic control*, vol. 65, no. 7, pp. 2979–2991, 2019.

- [3] R. Fu, A. Taghvaei, Y. Chen, and T. T. Georgiou, “Maximal power output of a stochastic thermodynamic engine,” *Automatica*, vol. 123, p. 109366, 2021.
- [4] T. Chen, J. Gu, L. Dinh, E. A. Theodorou, J. Susskind, and S. Zhai, “Generative modeling with phase stochastic bridges,” *arXiv preprint arXiv:2310.07805*, 2023.
- [5] C. Domingo i Enrich, J. Han, B. Amos, J. Bruna, and R. T. Chen, “Stochastic optimal control matching,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 112 459–112 504, 2024.
- [6] G. Rapakoulias, A. R. Pedram, and P. Tsiotras, “Go with the flow: Fast diffusion for gaussian mixture models,” *arXiv preprint arXiv:2412.09059*, 2024.
- [7] S. K. Shah, C. D. Pahlajani, N. A. Lacock, and H. G. Tanner, “Stochastic receding horizon control for robots with probabilistic state constraints,” in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 2893–2898.
- [8] A. Gorodetsky, S. Karaman, and Y. Marzouk, “High-dimensional stochastic optimal control using continuous tensor decompositions,” *The International Journal of Robotics Research*, vol. 37, no. 2-3, pp. 340–377, 2018.
- [9] E. Theodorou, F. Stulp, J. Buchli, and S. Schaal, “An iterative path integral stochastic optimal control approach for learning robotic tasks,” *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 11 594–11 601, 2011.
- [10] K. Okamoto and P. Tsiotras, “Optimal stochastic vehicle path planning using covariance steering,” *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2276–2281, 2019.
- [11] F. Liu, G. Rapakoulias, and P. Tsiotras, “Optimal covariance steering for discrete-time linear stochastic systems,” *IEEE Transactions on Automatic Control*, 2024.
- [12] A. Wakolbinger *et al.*, “Schrödinger bridges from 1931 to 1991,” in *Proc. of the 4th Latin American Congress in Probability and Mathematical Statistics*, Mexico City, 1990, pp. 61–79.
- [13] M. Pavon and A. Wakolbinger, “On free energy, stochastic control, and Schrödinger processes,” in *Modeling, Estimation and Control of Systems with Uncertainty: Proceedings of a Conference held in Sopron, Hungary, September 1990*. Springer, 1991, pp. 334–348.
- [14] Y. Chen, T. T. Georgiou, and M. Pavon, “Optimal steering of a linear stochastic system to a final probability distribution, part i,” *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1158–1169, 2015.
- [15] —, “Optimal steering of a linear stochastic system to a final probability distribution, part ii,” *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1170–1180, 2016.
- [16] A. M. Teter, W. Wang, and A. Halder, “Schrödinger bridge with quadratic state cost is exactly solvable,” *arXiv preprint arXiv:2406.00503*, 2024.
- [17] A. M. Teter, W. Wang, S. Shivakumar, and A. Halder, “Markov kernels, distances and optimal control: A parable of linear quadratic non-gaussian distribution steering,” *arXiv preprint arXiv:2504.15753*, 2025.
- [18] G.-H. Liu, Y. Lipman, M. Nickel, B. Karrer, E. A. Theodorou, and R. T. Chen, “Generalized Schrödinger bridge matching,” *arXiv preprint arXiv:2310.02233*, 2023.
- [19] Y. Mei, M. Al-Jarrah, A. Taghvaei, and Y. Chen, “Flow matching for stochastic linear control systems,” *arXiv preprint arXiv:2412.00617*, 2024.
- [20] K. F. Caluya and A. Halder, “Wasserstein proximal algorithms for the schrödinger bridge problem: Density control with nonlinear drift,” *IEEE Transactions on Automatic Control*, vol. 67, no. 3, pp. 1163–1178, 2021.
- [21] A. Pakniyat and P. Tsiotras, “Steering the state of linear stochastic systems: a constrained minimum principle formulation,” in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 1300–1305.
- [22] —, “Partially observed steering the state of linear stochastic systems,” in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 3780–3785.
- [23] A. Pakniyat, “A convex duality approach for assigning probability distributions to the state of nonlinear stochastic systems,” *IEEE Control Systems Letters*, vol. 6, pp. 3080–3085, 2022.
- [24] —, “Distributionally constrained convex duality optimal control (DC-CDOC) subject to different forms of constraining the terminal state of nonlinear stochastic systems,” in *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE, 2024, pp. 665–670.
- [25] B. D. Anderson, “Reverse-time diffusion equation models,” *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.
- [26] U. G. Haussmann and E. Pardoux, “Time reversal of diffusions,” *The Annals of Probability*, pp. 1188–1205, 1986.
- [27] P. Cattiaux, G. Conforti, I. Gentil, and C. Léonard, “Time reversal of diffusion processes under a finite entropy condition,” in *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques*, vol. 59, no. 4. Institut Henri Poincaré, 2023, pp. 1844–1881.
- [28] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
- [29] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” *arXiv preprint arXiv:2010.02502*, 2020.
- [30] Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” *Advances in neural information processing systems*, vol. 32, 2019.
- [31] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” in *International Conference on Learning Representations*, 2021.
- [32] S. Särkkä and A. Solin, *Applied stochastic differential equations*. Cambridge University Press, 2019, vol. 10.
- [33] J. Heng, V. De Bortoli, A. Doucet, and J. Thornton, “Simulating diffusion bridges with score matching,” *Biometrika*, p. asaf048, 07 2025. [Online]. Available: <https://doi.org/10.1093/biomet/asaf048>
- [34] H. Zhou, Y. Song, and V. Tzoumas, “Safe non-stochastic control of control-affine systems: An online convex optimization approach,” *IEEE Robotics and Automation Letters*, vol. 8, no. 12, pp. 7873–7880, 2023.
- [35] X. Liu, “Fuel-optimal rocket landing with aerodynamic controls,” *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 1, pp. 65–77, 2019.
- [36] G. C. Calafiore, “Multi-period portfolio optimization with linear control policies,” *Automatica*, vol. 44, no. 10, pp. 2463–2473, 2008.
- [37] B. Øksendal and B. Øksendal, *Stochastic differential equations*. Springer, 2003.
- [38] M. Chaleyat-Maurel and D. Michel, “Hypoellipticity theorems and conditional laws,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 65, no. 4, pp. 573–597, 1984.
- [39] L. Hörmander, “Hypoelliptic second order differential equations,” 1967.
- [40] A. J. Krener, “A generalization of Chow’s theorem and the bang-bang theorem to nonlinear control problems,” *SIAM Journal on Control*, vol. 12, no. 1, pp. 43–52, 1974.
- [41] H. J. Sussmann, “A general theorem on local controllability,” *SIAM Journal on Control and Optimization*, vol. 25, no. 1, pp. 158–194, 1987.
- [42] A. Hyvärinen and P. Dayan, “Estimation of non-normalized statistical models by score matching,” *Journal of Machine Learning Research*, vol. 6, no. 4, 2005.
- [43] W. Tang and H. Zhao, “Score-based diffusion models via stochastic differential equations—a technical tutorial,” *arXiv preprint arXiv:2402.07487*, 2024.
- [44] Y. Chen and T. Georgiou, “Stochastic bridges of linear systems,” *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 526–531, 2015.
- [45] I. Exarchos, E. A. Theodorou, and P. Tsiotras, “Optimal thrust profile for planetary soft landing under stochastic disturbances,” *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 1, pp. 209–216, 2019.
- [46] T. Lew, R. Bonalli, and M. Pavone, “Chance-constrained sequential convex programming for robust trajectory optimization,” in *2020 European Control Conference (ECC)*. IEEE, 2020, pp. 1871–1878.