Chapter 9
Theoretical Guarantees for Satisfaction of Terminal State Constraints for Nonlinear Stochastic Systems

Ali Pakniyat

Abstract In several engineering applications, it is desired to bring a system from an initial configuration to a specific terminal configuration. A motivational example is the vertical landing of reusable rockets which are required to come to full stop at an exact location on the landing platform in an upright configuration with all linear and angular velocities coming to zero. While in a deterministic setting, one can study these problems and provide theoretical guarantees for the satisfaction of the terminal state requirements, e.g., by employing the Pontryagin Minimum Principle (PMP), no such guarantees can be provided for exact satisfaction of terminal state constraints in a stochastic setting and, inevitably, one needs to seek alternative expressions of the desired requirements and establish guarantees for those alternatives. This article presents two novel approaches, each with an alternative expression of the terminal state requirement, and each providing theoretical guarantees for optimality and the satisfaction of the associated terminal state constraints. The first approach is to impose a constraint on the conditional expectations of the terminal state at all future times in which case the associated optimality conditions are expressed in the form of the Terminally Constrained Stochastic Minimum Principle (TC-SMP). The second approach is to impose a terminal state constraint as the matching of the probability distribution of the terminal state with a desired probability distribution in which case the associated optimality conditions are expressed using Hamilton-Jacobi (HJ) type equations. Numerical examples are provided to illustrate the results.

Keywords Hamilton-Jacobi inequalities · Necessary conditions · Nonlinear systems · Optimal control · Probability constraints · Probability distribution control · Steering theory · Stochastic maximum principle · Stochastic processes · Stochastic systems

A. Pakniyat (✉)
Department of Mechanical Engineering, University of Alabama, Tuscaloosa, AL, USA
e-mail: apakniyat@ua.edu

D. Azimov (ed.), *Proceedings of the IUTAM Symposium on Optimal Guidance and Control for Autonomous Systems 2023*, IUTAM Bookseries 40,
https://doi.org/10.1007/978-3-031-39303-7_9

9.1 Introduction

In several missions for aerial and underwater vehicles, the system's state is required to be pinpoint delivered to a desired destination state. A motivational example is the landing of a reusable rocket, e.g., the booster rocket of SpaceX Falcon 9, which is required to come to full stop conditions at an exact location on the landing platform in an upright configuration with all linear and angular velocities coming to zero. In the absence of dynamic uncertainty, one can formulate the problem as a deterministic optimal control problem with a fixed terminal state and invoke powerful theoretical results such as the Pontryagin Minimum Principle (PMP) to obtain the optimal input signal together with the associated optimal trajectory. It is worth remarking that the celebrated Hamilton Jacobi Bellman (HJB) equation is not applicable for this problem since the terminal (boundary) condition for the HJB equation is not well defined (i.e., it turns into a singular function).

In the presence of a stochastic diffusion, these state steering problems are more challenging and have been the subject of a limited number of studies. More precisely, the majority of studies assume linearity of the dynamics and quadratic forms for the cost, so that the associated probabilities take the form of Gaussian distributions. In this case, and in the absence of any additional state constraints, the dynamics of the mean state process and the covariance state process can be shown to be decoupled. Within an infinite time horizon setting, the problem has been formulated as the association of a steady-state distribution with its mean being at the desired terminal location, and a comprehensive study over the assignable covariances for the infinite horizon problem is presented in [16, 21, 52, 53]. For linear stochastic systems over finite time horizons, a similar philosophy is taken in both continuous time and discrete time settings, and the associated distribution assignment methodologies are studied by [1–3, 9–11, 15, 17, 31, 32]. The accommodation of input constraints is considered in [2], and convex relaxations for linear systems subject to chance constraints, which are probabilistic constraints that impose a maximum probability of constraint violation, are studied in [31, 32]. Within the same class of linear quadratic systems, the accommodation of information obtained as the time progresses is presented in the form of a model predictive control (MPC) based approach in [25, 28, 30, 45].

Extensions of the probability distribution assignment to nonlinear systems has been presented for feedback-linearizable systems [6], and implementation through iterative linearization is proposed in [43] and via differential dynamic programming approximations [54]. More recently, new results have emerged via reformulations of the probability assignment problem as nonlinear Schrödinger bridge problem [7, 29]. However, a key limitation of the current nonlinear Schrödinger bridge results is that the Itô differential equation governing the dynamics of the nonlinear stochastic systems must take the special form of a gradient flow. In contrast, for stochastic systems whose dynamics are governed by a general class of nonlinear Itô differential equation, whose costs take general nonlinear forms, and whose desired probability distributions are permitted to take general (not necessarily Gaussian) distributions, the necessary optimality conditions are established in [35] in the form of Hamilton-

Jacobi problems where the optimization objective, in addition to the value function evaluation at the initial conditions, includes an extra term which is the integral of the product of the value function at the terminal time and the desired probability distribution. The methodology used in [35] is based upon the accommodation of terminal distribution constraints on the convex duality method for optimal control problems which was initiated by Vinter and Lewis [50, 51] for deterministic control systems and, later, by Fleming and Vermes for piecewise deterministic [47] and stochastic [14] processes. The fundamental idea of this approach is the introduction of a weak formulation that embeds the original (strong) problem into a convex linear program over the space of Radon measures. Upon establishing the equivalence of the two problems, new necessary and sufficient optimality condition are obtained by invoking the Fenchel-Rockafellar duality theorem. This approach is particularly useful in characterization of optimal policies in certain desirable classes of controls by investigating the extreme points of the set of Hamilton-Jacobi problems (see e.g. [5, 13, 22, 57]). For deterministic control systems, convex duality based numerical algorithms are established in [12, 23, 40] for continuous systems, and in [27, 46, 56] for hybrid systems.

A fundamental limit of methodologies based on the assignment of probability distributions is that the studied probabilities are conditioned on the filtration at the initial time. In contrast, as proposed in [38, 39] the employment of the Stochastic Minimum Principle (SMP) yields a natural accommodation of filtration-adapted updates because the same adaptation requirement must be provided for the adjoint process. In other words, the optimal input expressed in terms of the adjoint process is adapted to the current time filtration, since the solution of the backward stochastic differential equation (BSDE) for the adjoint process must remain adapted to the same forward filtration. This important characteristic provides an opportunity to impose terminal state constraints at all times, as apposed to the current literature where constraints are imposed on probability distributions as viewed at the initial time. In order to solve the associated problem, we invoke the Stochastic Maximum Principle (SMP) presented in [41] and, in particular, the version with terminal state constraints [41, Theorem 5], henceforth called the Terminally Constrained Stochastic Minimum Principle (TC-SMP). While, in general, obtaining numerical solutions to the BSDEs of the adjoint process are computationally expensive, for a class of linear stochastic systems with quadratic costs, we derive analytical solutions to the adjoint equation in terms of the system's state transition matrix, its controllability Gramian and the solution of a differential matrix Riccati equation. Moreover, the accommodation of various information structures on the TC-SMP is studied in [38] and, further, in this article.

The objective of this article is the presentation of a general framework within which theoretical guarantees are presented for the satisfaction of a large class of terminal state constraints for nonlinear stochastic systems, aw well as the illustration of the two key methodologies of the convex duality based HJ inequalities and the TC-SMP with analytic examples. The structure of the article is as follows.

Section 9.2 discusses the characterization of optimal solutions to the deterministic problem of steering the state towards a desired value, and Sect. 9.3 elaborates the discussion to the case with dynamic uncertainties and discusses how one can impose constraint on the state of stochastic systems without violating causality requirements. Section 9.4 presents the problem of constraining the family of conditional expectations of the terminal state and presents the Terminally Constrained Stochastic Minimum Principle (TC-SMP), its associated numerical algorithm and its specialization to linear stochastic systems with quadratic costs. Three examples are presented in this section to further illustrate the results. Section 9.5 presents the problem of constraining the probability distribution of the terminal state and establishes the associated optimality conditions, identifying the corresponding value function as the optimal solution to a family of Hamilton-Jacobi problems where the optimization objective, in addition to the value function evaluation at the initial conditions, includes an extra term which is the integral of the product of the value function at the terminal time and the desired probability distribution. A numerical algorithm based upon this methodology is presented and the results are illustrated via two numerical examples. Concluding remarks are presented in Sect. 9.6.

9.2 Deterministic Case Revisited

Let us first recall the fixed end point optimal control problem for the deterministic case, i.e., in the absence of dynamic uncertainties. For the nonlinear control system with the dynamics

$$\dot{x}_t = f(x_t, u_t, t) \tag{9.1}$$

subject to a given initial state $x_{t_0} = x_0$, and given a fixed finite time horizon $[t_0, t_f]$, we would like to find an input signal input signal $[u] := \{u_s : t_0 \leq s \leq t_f\}$ which brings the state to a desired value $x_{t_f} = x_f \in \mathbb{R}^n$. Furthermore, it is often the case that among all such inputs, one would like to find an optimal input which minimizes the cost

$$J(t_0, x_0, [u]; x_f) = \int_{t_0}^{t_f} \ell(x_s, u_s, s)\mathrm{d}s \tag{9.2}$$

For these problems, one can identify optimal inputs from the necessary optimality conditions of the Minimum Principle [42] which states that for the optimal input process $[u^*] \equiv \{u_s^* : t_0 \leq s \leq t_f\}$ and along the corresponding optimal trajectory $[x^*] = \{x_t^* : t_0 \leq t \leq t_f, \text{ where } x_t^* = x_0 + \int_{t_0}^{t} f(x_s^*, u_s^*, s)\mathrm{d}s\}$ there exist a constant $\gamma \in \{0, 1\}$ and an adjoint process $[\lambda^*] = \{\lambda_s^* : t_0 \leq s \leq t_f\}$ such that the Hamiltonian defined by

$$H(x, u, \lambda, \gamma, t) := \gamma\, \ell(x, u, t) + \lambda^\top f(x, u, t) \tag{9.3}$$

is minimized with respect to u, i.e.,

$$H(x_t^*, u_t^*, \lambda_t^*, \gamma, t) \leq H(x_t^*, u, \lambda_t^*, \gamma, t), \qquad \text{for all } u \in U \subset \mathbb{R}^m \tag{9.4}$$

where

$$\dot{x}_t^* = \left.\frac{\partial H}{\partial \lambda}\right|_{(x_t^*, u_t^*, \lambda_t^*, \gamma, t)} \equiv f(x_t^*, u_t^*, t) \tag{9.5}$$

$$\dot{\lambda}_t^* = \left.\frac{-\partial H}{\partial x}\right|_{(x_t^*, u_t^*, \lambda_t^*, \gamma, t)} \equiv -\gamma \frac{\partial \ell(x_t^*, u_t^*, t)}{\partial x} - \frac{\partial f(x_t^*, u_t^*, t)}{\partial x}^\top \lambda_t^* \tag{9.6}$$

subject to $x_{t_0}^* = x_0$ and $x_{t_f}^* = x_f$. If the problem is normal (see, e.g., [24]), it is possible to satisfy the necessary conditions with the constant γ taken to be 1.

In contrast to the Minimum Principle which is well suited for fixed terminal state problems, Dynamic Programming [4] and the associated Hamilton Jacobi Bellman (HJB) equation (see, e.g., [55]) is not able to handle terminal state constraints. As noted by [48] for terminally constrained problems of the type $x_{t_f} \in S \subset \mathbb{R}^n$, "the value function has at most only a subsidiary role. This is because, unless stringent conditions are imposed on the data, we cannot expect any longer that the value function will be defined on a sufficiently large subset of $\mathbb{R} \times \mathbb{R}^n$, or be sufficiently regular, for it to serve as a Carathéodory function". When the set S becomes a singleton, i.e., $x_{t_f} \in \{x_f\}$, then the HJB equation

$$\frac{\partial V(t, x)}{\partial t} + \inf_{u \in U} \left[\left(\frac{\partial V(t, x)}{\partial x} \right)^\top f(x, u, t) + \ell(x, u, t) \right] = 0 \tag{9.7}$$

becomes subject to the singular terminal condition

$$V(t_f, x) = \begin{cases} 0, & x = x_f, \\ \infty, & x \neq x_f. \end{cases} \tag{9.8}$$

However, by invoking convex duality relation between the space of measures and that of continuous functions, it is possible [48–51] that the value function can be identified as the upper envelope (i.e., supremum) of the smooth subsolutions of the Hamilton-Jacobi inequalities [48]

$$\begin{aligned} V(t_0, x_0) = \sup_{v \in C^1([t_0, t_f] \times \mathbb{R}^n)} \Bigg\{ & v(t_0, x_0) : \\ & \frac{\partial v(t, x)}{\partial t} + \left(\frac{\partial v(t, x)}{\partial x} \right)^\top f(x, u, t) + \ell(x, u, t) \geq 0, \\ & v(t_f, x) \leq 0, \text{ for } x = x_f. \Bigg\} \end{aligned} \tag{9.9}$$

While the above result is promising in the sense that it shows the ability of convex duality method to provide a solution to the deterministic fixed endpoint problem, its special form in its deterministic setting hinders the true nature of this identification and it cannot be directly extended for its stochastic equivalent problem. As a matter of fact, if instead of the above identification, [48] presented their results as

$$V(t_0, x_0) = \sup_{v \in C^1([t_0,t_f]\times\mathbb{R}^n)} \Big\{ v(t_0, x_0) - v(t_f, x_f) : \\ \frac{\partial v(t,x)}{\partial t} + \left(\frac{\partial v(t,x)}{\partial x}\right)^{\top} f(x,u,t) + \ell(x,u,t) \geq 0, \\ v(t_f, x) \leq 0, \text{ for all } x \in \mathbb{R}^n. \Big\} \tag{9.10}$$

then the discovery of its stochastic counterpart would have been made much earlier than [35] and Theorem 3 in this article.

9.3 The Stochastic Versions of the "Fixed" Endpoint Problem

We now formulate the stochastic equivalent of the deterministic fixed endpoint problem.

Let $\big(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t=t_0}^{t_f}, \mathsf{P}\big)$ be a filtered probability space with $\mathscr{F}_t$ being an increasing family of sub σ-algebras of $\mathscr{F}$ such that $\mathscr{F}_{t_0}$ contains all the P-null sets, and $\mathscr{F}_{t_f} = \mathscr{F}$ for a fixed terminal time $t_f < \infty$. Consider a nonlinear stochastic systems governed by the controlled Itô differential equation

$$\mathsf{d}x_s = f(s, x_s, u_s)\,\mathsf{d}s + g(s, x_s)\,\mathsf{d}w_s, \tag{9.11}$$

where, at each $s \in [t_0, t_f]$, the system's state is denoted by $x_s \in \mathbb{R}^n$, its input is denoted by $u_s \in U \subset \mathbb{R}^m$, and the realization of a standard Wiener process is denoted by $w_s \in \mathbb{R}^d$. The input value set U is assumed to be convex and compact and the functions f and g are considered to be Lipschitz continuous functions over, respectively, $[t_0, t_f] \times \mathbb{R}^n \times U$ and $[t_0, t_f] \times \mathbb{R}^n$, which satisfy either boundedness or linear growth conditions.

Let $[u] := \{u_s : t_0 \leq s \leq t_f\}$ denote a nonanticipative, U-valued, input process such that $u_s \in U$ is progressively measurable with respect to $\mathscr{F}_s$ for all $s \in [t_0, t_f]$. We denote by $\mathcal{U}$ the set of all such inputs. We remark that the underlying policy for the determination of the input process $[u]$ can take any form, as long as the policy remains causal, that is, u_s does not depend on future values of the noise or the state. For instance, within a comprehensive closed loop policy, u_s is permitted to depend on s, $[x]_{t_0}^s$, $[w]_{t_0}^s$, $[u]_{t_0}^s$ and expectations of their future values under the $\mathscr{F}_s$ filtration.

However, it can be shown that there is no loss of optimality in feedback structures where u_s depends only on s, x_s and expectations of the cost gradient under the $\mathscr{F}_s$ filtration.

In this paper, we consider only the case with complete and accurate observations of the state. Thus, for time instances t and s such that $t_0 \leq t \leq s \leq t_f$, within the interval $[t_0, t_f]$, and under the filtration $\mathscr{F}_t$, the variable x_s is treated as a *deterministic* variable whenever $s \leq t$, and is treated as a *random* variable whenever $s > t$. However, at all future instances, $s \in (t, t_f]$, the state value x_s remains a random variable under the filtration $\mathscr{F}_t$. We define the notation

$$\mathbb{E}^{[u]}_{\mathscr{F}_t}[x_s] := \mathbb{E}\big[x_s \big| \mathscr{F}_t; [u]_t^{t_f}\big] \equiv \mathbb{E}\big[x_s \big| \mathscr{F}_t; [u]_t^{s}\big], \tag{9.12}$$

for the expected value[1] of x_s at $s \in [t, t_f]$, under the filtration $\mathscr{F}_t$ and given the input process $[u]_t^{t_f}$, where the last equality (conditioning on $[u]_t^s$ instead of $[u]_t^{t_f}$) is a consequence of the causality of the controlled process in (9.11).

The associated optimal control problem corresponds to the minimization of the cost

$$J\big(t, x_t; [u]_t^{t_f}\big) := \mathbb{E}^{[u]}_{\mathscr{F}_t}\left[\int_t^{t_f} \ell(x_s, u_s)\mathsf{d}s + L\big(x_{t_f}\big)\right]. \tag{9.13}$$

subject to appropriate terminal state constraint where, in the above cost, ℓ is a continuous function with polynomial growth.

A naïve approach for the consideration of terminal state constraints is to impose the constraint $x_{t_f} = x_f$ on the stochastic system (9.11). However, such a constraint violates causality as under each filtration $\mathscr{F}_t$ at time $t \in [t_0, t_f]$, future values of the state, i.e., $\{x_s : s \in (t, t_f]\}$ are random variable and, hence, it is not possible (i.e., there does not exists any nonanticipative input process $[u]$ such that the random variable x_{t_f} becomes deterministic so that it then satisfies a constraint such as $x_{t_f} = x_f$.

Another approach for the consideration of terminal state constraints is to impose the constraint $x_{t_f} \overset{a.s.}{=} x_f$ on the stochastic system (9.11). This constraint can be equivalently expressed as $\mathsf{P}(x_{t_f} = x_f) = 1$ or $\mathsf{P}(x_{t_f} \neq x_f) = 0$. While this constraint is both mathematically well-posed and practically desirable, a major challenge is the absence of theoretical guarantees of this time for stochastic processes.

In this article, we presents two novel approaches for the consideration of terminal state constraints on nonlinear stochastic systems, and each providing theoretical guarantees for optimality and the satisfaction of the associated terminal state constraints.

The first method is to impose a terminal state constraint as

$$\mathbb{E}^{[u]}_{\mathscr{F}_t}[x_{t_f}] = x_f, \tag{9.14}$$

[1] Since x_s becomes deterministic for $s = t$, and thus, $\mathbb{E}^{[u]}_{\mathscr{F}_t}[x_s] \equiv \mathbb{E}^{[u]}_{\mathscr{F}_t}[x_t] = x_t$ whenever $s = t$.

at all time instances $t \in [t_0, t_f]$, in order for the state to be steered to a desired value $x_f \in \mathbb{R}^n$

The second method is to impose a terminal state constraint as $x_{t_f}^{[u]} \sim \mathsf{p}_d$, i.e., we require the probability distribution of the terminal state to take the desired form p_d. This, by definition, signifies that for every Borel set $B_x \in \mathbb{R}^n$,

$$\mathsf{P}^{[u]}\big(x_{t_f} \in B_x\big) = \int_{B_x} \mathsf{p}_d(\mathsf{d}x), \tag{9.15}$$

where $\mathsf{P}^{[u]}(\cdot)$ denotes the probability of an event given the input $[u]$.

9.4 Constraining the Family of Conditional Expectations of the Terminal State

As mentioned earlier, in this method we impose a family terminal state constraints in the form of

$$\mathbb{E}^{[u]}_{\mathscr{F}_t}[x_{t_f}] = x_f, \tag{9.16}$$

at all time instances $t \in [t_0, t_f]$, in order for the state to be steered to a desired value $x_f \in \mathbb{R}^n$. Our strategy to solve this problem hinges on the restriction of the class of controllers to those yielding the expected value of the terminal state matching the desired value *under filtrations at all future times*. The novelty of this approach lies within its change of viewpoint, from the conventional conditioning the probability distributions on the information available at the design time, to the less explored, and mathematically more elaborate, approach of conditioning these probabilities on the family of σ-algebras of all possible scenarios for future uncertainties.

In the absence of the constraint (9.16), there are several versions of the Stochastic Maximum Principle (SMP), see, e.g., [55] for historical remarks on the development of the SMP. For the case of problems with terminal constraints, [41] first considered the case of a terminal constraint in the form of the total expectation of a nonlinear function of the state, including the terminal state constraint (9.16) consisting of n individual constraints $\mathbb{E}^{[u]}_{\mathscr{F}_t}\big[x_{t_f}^{(i)}\big] = \mu_f^{(i)}$, $(i = 1, 2, \cdots, n)$. However, the implementation of [41] still uses conditioning under the total expectation, which is equivalent to the filtration $\mathscr{F}_{t_0}$, instead of imposing the expectation on the σ-algebra of all potential realizations of the information at time t which are contained in $\mathscr{F}_t$.

9.4.1 Terminally Constrained Stochastic Minimum Principle (TC–SMP)

Theorem 1 [39] For the system (9.11), the optimal input for the cost (9.13) subject to the constraint (9.16) is determined from

$$u_s^* = \underset{u \in U \subset \mathbb{R}^m}{\operatorname{argmin}} \Big\{ \ell(x_s, u) + \lambda_s^\top f(x_s, u) \Big\}, \tag{9.17}$$

where the adjoint pair (λ_s, Λ_s), $s \in [t, t_f]$ are governed by the backward stochastic differential equation

$$\mathsf{d}\lambda_s = -\left(\frac{\partial f(x_s^*, u_s^*)}{\partial x} \lambda_s + \frac{\partial \ell(x_s^*, u_s^*)}{\partial x} \right) \mathsf{d}s + \Lambda_s \mathsf{d}w_s, \tag{9.18}$$

subject to the terminal condition

$$\lambda_{t_f} = \alpha \frac{\partial L(x_{t_f}^*)}{\partial x} + \beta, \tag{9.19}$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^n$ are constants which are not simultaneously zero. □

It shall be remarked that, as is conventional for backward stochastic processes, the second-order adjoint process Λ is implicitly defined as the (unique) process leading to the statisfaction of the terminal condition in (9.19). However, under sufficient smoothness of the functions, it has been shown (see e.g. [26]) that

$$\Lambda_s = g(x_s)^\top \frac{\partial}{\partial x} \lambda\big(s, x_s\big) \tag{9.20}$$

This is due to the fact that, whenever there exist a twice continuously differentiable function $\tilde{\lambda}$ such that $\lambda_s = \tilde{\lambda}(s, x_s)$, a direct application of Itô's formula yields

$$\begin{aligned} \mathsf{d}\tilde{\lambda}(s, x_s) = & \left(\frac{\partial \tilde{\lambda}(s, x_s)}{\partial s} + \frac{\partial \tilde{\lambda}(s, x_s)}{\partial x}^\top f(x_s, u_s) \right. \\ & \left. + \frac{1}{2} \mathrm{tr}\left(g(x_s)^\top g(x_s) \frac{\partial^2 \tilde{\lambda}(s, x_s)}{\partial x^2} \right) \right) \mathsf{d}s + g(x_s)^\top \frac{\partial \tilde{\lambda}(s, x_s)}{\partial x} \mathsf{d}w_s \\ & \overset{\lambda_s = \tilde{\lambda}(s, x_s)}{\overset{\downarrow}{=}} \mathsf{d}\lambda_s \overset{(9.18)}{=} -\left(\frac{\partial f(x_s, u_s)}{\partial x} \lambda_s + \frac{\partial l(x_s, u_s)}{\partial x} \right) \mathsf{d}s + \Lambda_s \mathsf{d}w_s \end{aligned} \tag{9.21}$$

which requires that

$$
\begin{aligned}
&-\frac{\partial l(x_s,u_s)}{\partial x}-\frac{\partial f(x_s,u_s)}{\partial x}\tilde{\lambda}(s,x_s)=\frac{\partial \tilde{\lambda}(s,x_s)}{\partial s}\\
&+\frac{\partial \tilde{\lambda}(s,x_s)}{\partial x}^{\top} f(x_s,u_s)+\frac{1}{2}\mathrm{tr}\Big(g(x_s)^{\top}g(x_s)\,\frac{\partial^2\tilde{\lambda}(s,x_s)}{\partial x^2}\Big), \qquad (9.22)\\
&\Lambda_s=g(x_s)^{\top}\frac{\partial \tilde{\lambda}(s,x_s)}{\partial x}. \qquad (9.23)
\end{aligned}
$$

It shall be remarked that (9.22) is a partial differential equation (PDE).[2] Thus, it is possible to obtain the solution to the adjoint equation (9.18) from the solution of (9.22). However, it is possible (and in certain applications, it is more numerically efficient) to directly solve the adjoint equation (9.18) while invoking (9.23) to accelerate numerical integrations.

A general solution methodology for the TC-SMP is as follows.

- Step 1: Select $\epsilon > 0$, set the iteration counter to $k = 0$, and initiate the algorithm with an arbitrary control policy π^k, such that $u_s = \pi^k(s, x_s)$. Discretize the time interval $[t_0, t_f]$ into a set of discrete times $\{t_0, t_1, \ldots, t_i, \ldots, t_N = t_f\}$.
- Step 2: For M realizations of the Wiener process w, generate sample paths for the state forward in time using the policy π^k:

$$
x_{t_{i+1}}=x_{t_i}+f\big(x_{t_i},\pi^k(t_i,x_{t_i})\big)\Delta t_i+g\big(x_{t_i}\big)\Delta w_i,\quad x_{t_0}=x_0.
$$

- Step 3: For each sample path, compute the associated adjoint processes backward in time using

$$
\begin{aligned}
\lambda^k_{t_N}&=\alpha\frac{\partial L(x_{t_N})}{\partial x}+\beta,\\
\Lambda^k_{t_N}&=\alpha g\big(x_{t_i}\big)^{\top}\frac{\partial^2 L(x_{t_N})}{\partial x^2}\\
\lambda^k_{t_{i-1}}&=\lambda^k_{t_i}-\mathbb{E}^{\pi^k}_{t_{i-1},x_{t_{i-1}}}\Bigg[\Bigg(\frac{\partial f\big(x_{t_i},\pi^k(t_i,x_{t_i})\big)}{\partial x}\lambda^k_{t_i}\\
&\qquad+\frac{\partial \ell\big(x_{t_i},\pi^k(t_i,x_{t_i})\big)}{\partial x}\Bigg)\Delta t_i-\Lambda^k_{t_i}\Delta w_i\Bigg],\\
\Lambda^k_{t_{i-1}}&\approx\frac{1}{\Delta t_{i-1}}\mathbb{E}^{\pi^k}_{t_{i-1},x_{t_{i-1}}}\big[\Delta w_{i-1}\lambda^k_{t_i}\big],
\end{aligned}
$$

[2] Notice that in the classical (free end-point) LQG case where $f(x,u^*)=Ax+Bu^*=Ax-BR^{-1}B^{\top}\lambda$ and $l(x,u)=\frac{1}{2}x^{\top}Qx+\frac{1}{2}u^{\top}Ru=x^{\top}Qx+\lambda^{\top}BR^{-1}B^{\top}\lambda$, the conjecture $\lambda=\Pi x$ leads to the celebrated Riccati equation. More specifically, $\frac{\partial l}{\partial x}=Qx$, $\frac{\partial f}{\partial x}=A^{\top}$, $\frac{\partial \lambda}{\partial s}=\dot{\Pi}x$, $\frac{\partial \lambda}{\partial x}=\Pi$ and $\frac{\partial^2\lambda}{\partial x^2}=0$. Substitution of these expressions into (9.22) yields

$$
-Qx-A^{\top}\Pi x=\dot{\Pi}x+\Pi^{\top}(Ax-BR^{-1}B^{\top}\Pi x)+0=\big(\dot{\Pi}+\Pi A-\Pi BR^{-1}B^{\top}\Pi\big)x,
$$

which, after cancelling out the x-factor, is the Riccati equation.

- Step 4: Update the policy according to

$$\pi^{k+1}(t_i, x) = \underset{u\in\mathbb{R}^m}{\operatorname{argmin}}\left\{\ell(x, u) + f(x, u)^\top \lambda^k_{t_i}\right\}.$$

- Step 5: If $\left|J\left(\pi^k\right) - J\left(\pi^{k-1}\right)\right| < \epsilon$, stop. Otherwise, increment k by and 1 and go to Step 1.

9.4.2 TC-SMP for Linear Quadratic Problems

In this section, we present the analytical solutions to the the TC-SMP in Theorem 1 for linear stochastic systems with quadratic cost. To this end, let the dynamics (9.11) be of the form

$$\mathsf{d}x_s = \left(A_s x_s + B_s u_s\right)\mathsf{d}s + D_s \mathsf{d}w_s, \tag{9.24}$$

where the time varying parameters in the system dynamics $A \in L^\infty([t_0, t_f]; \mathbb{R}^{n\times n})$, $B \in L^\infty([t_0, t_f]; \mathbb{R}^{n\times m})$, $D \in L^\infty([t_0, t_f]; \mathbb{R}^{n\times k})$, are essentially bounded measurable matrix functions of time.

For simplicity, we assume that the cost (9.13) is a quadratic function of the input and the terminal state, that is,

$$J\left(t, x_t, [u]_t^{t_f}\right) := \frac{1}{2}\mathbb{E}^{[u]}_{\mathscr{F}^t}\left[\int_t^{t_f} u_s^\top R_s u_s \mathsf{d}s + (x_{t_f} - \mu_f)^\top H_f\,(x_{t_f} - \mu_f)\right], \tag{9.25}$$

with $R \in L^\infty([t_0, t_f]; \mathcal{S}^{m\times m})$, $R_s > 0$, for all $s \in [t_0, t_f]$, and $H_f \in \mathcal{S}^{n\times n}$, $H_f \geq 0$, where $\mathcal{S}^{m\times m}$ denotes the space of $m \times m$-dimensional symmetric matrices.

We assume that the system (A_s, B_s) is controllable,[3] and that the system is noise controllable,[4] equivalently, $\mathrm{Im}(D_s) \subset \mathrm{Im}(B_s)$, for all $s \in [t_0, t_f]$, that is,

$$\forall w \in \mathbb{R}^k, \exists u \in \mathbb{R}^m \text{ s.t. } B_s u = D_s w. \tag{9.26}$$

Theorem 2 [38, 39] For the system (9.24) and the cost (9.25) subject to the constraint (9.16), the optimal input is determined by

[3] Hence, the Gramian (9.29) is full rank.

[4] As a requirement for solvability of the Riccati equations (9.30) and (9.31).

$$
\begin{aligned}
u_s^* = -R_s^{-1} B_s^\top \Phi(t_f, s)^\top \big[\mathcal{G}(t, t_f)\big]^{-1} \big(\Phi(t_f, t)x_t - \mu_f\big) \\
- R_s^{-1} B_s^\top \Pi(s; t_f) \Bigg(x_s - \Phi(t; s)x_t \\
+ \mathcal{G}(t, s)\Phi(s; t_f)^\top \big[\mathcal{G}(t, t_f)\big]^{-1} \Big(\Phi(t; t_f)x_t - \mu_f\Big)\Bigg),
\end{aligned}
\tag{9.27}
$$

where $\Phi(t; s) \in \mathbb{R}^{n\times n}$ is the state transition matrix from t to s for the system (9.24), which is the solution of

$$
\dot{\Phi} \equiv \frac{\partial \Phi(t; s)}{\partial s} = A_s \Phi, \qquad \Phi(t; t) = I_{n\times n}, \tag{9.28}
$$

and where

$$
\mathcal{G}(\tau, t) := \int_t^\tau \Phi(s; \tau) B_s R_s^{-1} B_s^\top \Phi(s; \tau)^\top \mathrm{d}s, \tag{9.29}
$$

is the controllability Gramian (see e.g., [8, Theorem 6.1]) over the horizon $[t, \tau] \subset [t_0, t_f]$, and $\Pi(s; t_f)$ is the solution of the following Riccati equation

$$
\dot{\Pi}_s \equiv \frac{\mathrm{d}}{\mathrm{d}s}\Pi(s; t_f) = \Pi_s B_s R_s^{-1} B_s^\top \Pi_s - \Pi_s A_s - A_s^\top \Pi_s, \tag{9.30}
$$

subject to the terminal condition

$$
\Pi(t_f; t_f) = H_f. \tag{9.31}
$$

□

9.4.3 Numerical Illustrations

In order to illustrate the results of Theorem 1 and its specialization to linear quadratic Gaussian problems, Theorem 2, let us consider the following examples.

Example 1 Consider the scalar case of a linear stochastic system with the dynamics

$$
\mathrm{d}x_s := (a x_s - b u_s)\mathrm{d}s + d\,\mathrm{d}w_s, \tag{9.32}
$$

with a, b, d scalar constants, and consider the problem of steering the state to the desired value $\mu_f \in \mathbb{R}$ by enforcing

$$
\mathbb{E}^{[u]}_{\mathscr{F}_t}[x_{t_f}] = \mu_f, \qquad (16)
$$

at all $t \in [t_0, t_f]$, with the cost

$$J\big(t, x_t, [u]_t^{t_f}\big) := \frac{1}{2}\mathbb{E}_{\mathscr{F}_t}^{[u]}\Bigg[\int_t^{t_f} r\, u_s^2 \mathsf{d}s + h(x_{t_f} - \mu_f)^2\Bigg]. \tag{9.33}$$

where $r > 0$ and $h \in \mathbb{R}_{\geq 0} \setminus \{2ar/b^2\}$.[5]

For this problem, we can analytically represent Φ, $\mathcal{G}$ and Π as

$$\Phi(t; s) = e^{a(s-t)}, \tag{9.34}$$

$$\mathcal{G}(t, \tau) = \frac{b^2}{2a} e^{2a\,t_f}\left(e^{-2a\,t} - e^{-2a\,\tau}\right), \tag{9.35}$$

$$\Pi(s; t_f) = \frac{2ar}{b^2\left(1 - \frac{h}{h - \frac{2ar}{b^2}} e^{\frac{b^2}{r}(t_f - s)}\right)}. \tag{9.36}$$

and, therefore, the optimal input (9.27) becomes

$$\begin{aligned} u_s^* = & \frac{-2a}{br\left(e^{2a(t_f-t)} - 1\right)} e^{a(t_f-s)}\left(e^{a(t_f-t)}x_t - \mu_f\right) \\ & - \frac{2a}{b\left(1 - \frac{h}{h - \frac{2ar}{b^2}} e^{\frac{b^2}{r}(t_f - s)}\right)}\Bigg(x_s - e^{a(s-t)}x_t \\ & + e^{a(t_f-s)} \frac{e^{-2at} - e^{-2as}}{e^{-2at} - e^{-2at_f}}\left(e^{a(t_f-t)}x_t - \mu_f\right)\Bigg). \end{aligned} \tag{9.37}$$

Let $a = b = d = r = h = 1$, and the time horizon be $[t_0, t_f] = [0, 2]$. For the steering towards the desired stated $\mu_f = 5$, from the initial condition $x_0 = -3$, the optimal input satisfying (9.16) for all $t \in [0, 2]$, and the associated trajectories for 50 sample paths are illustrated in Fig. 9.1. As can be seen from the figure, all trajectories are almost surely driven to the required terminal state at $t = t_f$. The associated probability distributions for the state and the input processes from the implementation of the TC-SMP on this system are illustrated in Fig. 9.2. It can be observed that the state distribution, starting from a lumped delta distribution at the initial condition, is steered towards a delta distribution at the desired terminal state while in intermediate times the state distribution does not remain lumped. In contrast, the input distribution, starting from a lumped delta distribution, does not remain lumped at any later time, which shows that the satisfaction of the desired terminal state comes at the expense of input uncertainty at later times over a wide range of values.

[5] The exclusion of this value, which occurs only if $a > 0$, is due to the appearance of $h - 2ar/b^2$ as a denominator in (9.37).

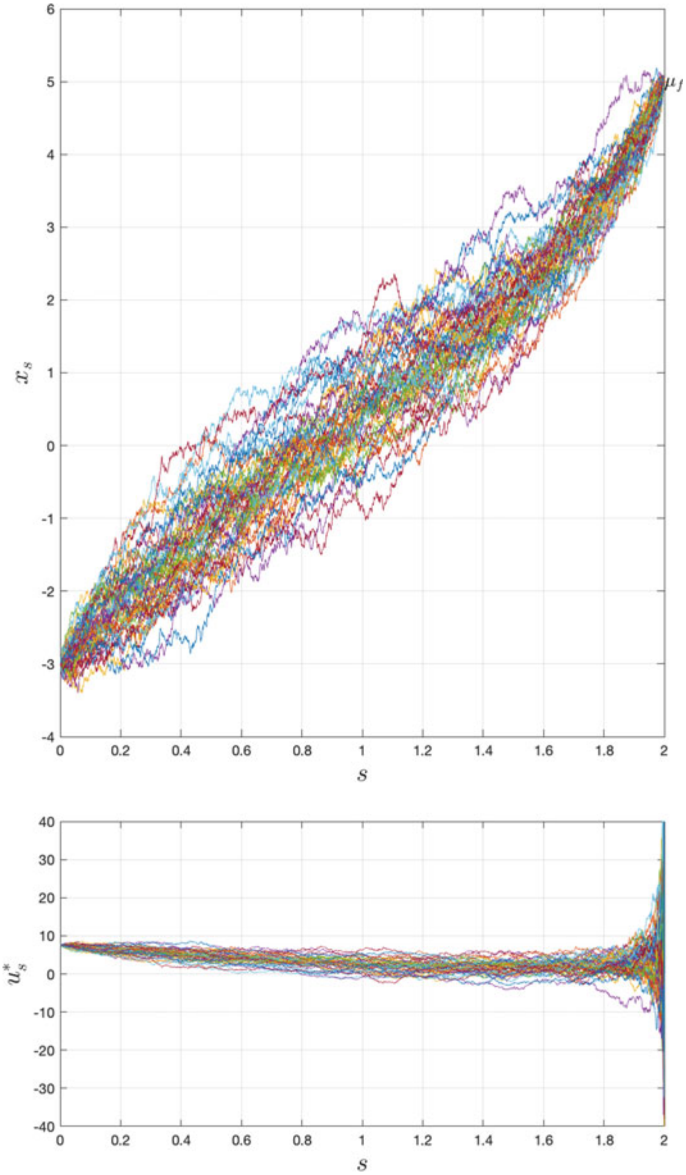


Fig. 9.1 Sample paths associated with the implementation of the terminally constrained stochastic minimum principle (TC-SMP) on the system in Example 1

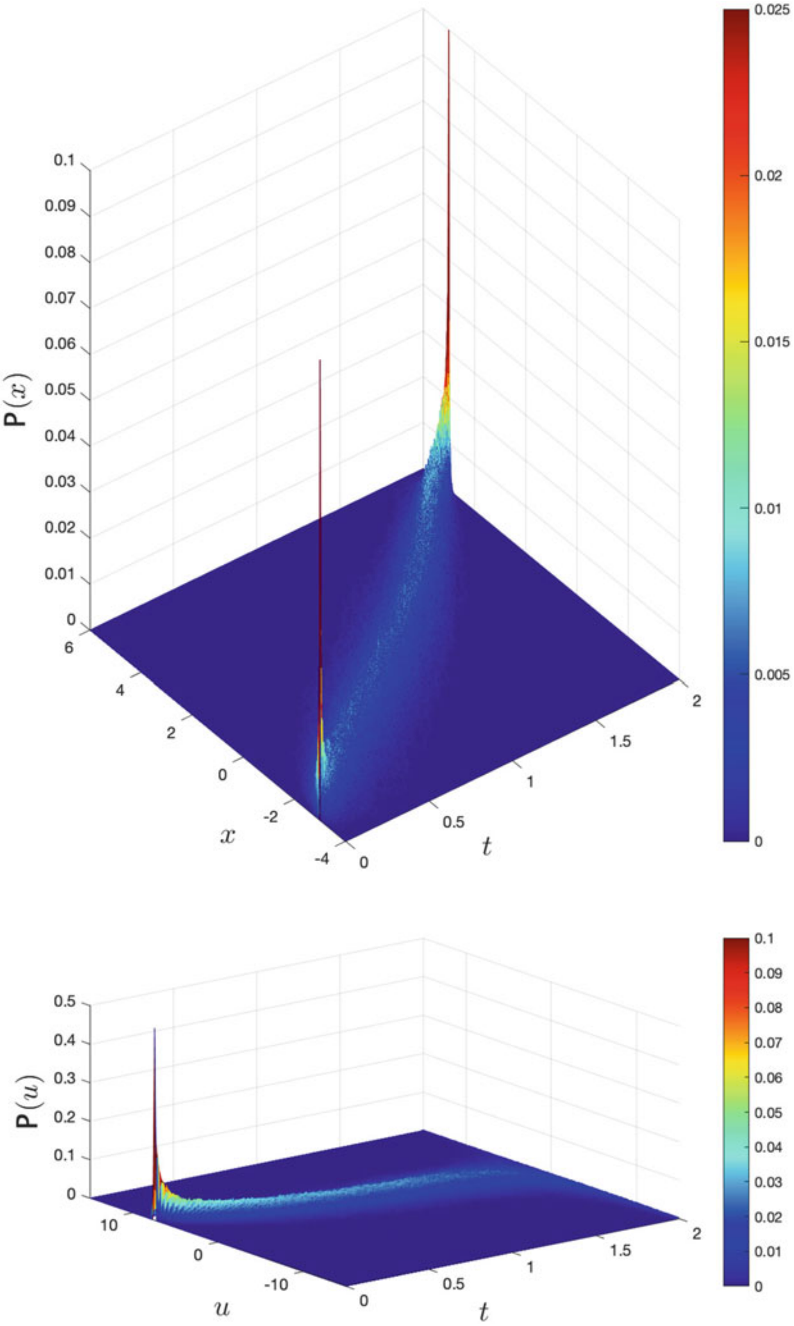


Fig. 9.2 Probability distributions associated with the implementation of the terminally constrained stochastic minimum principle (TC-SMP) on the system in Example 1

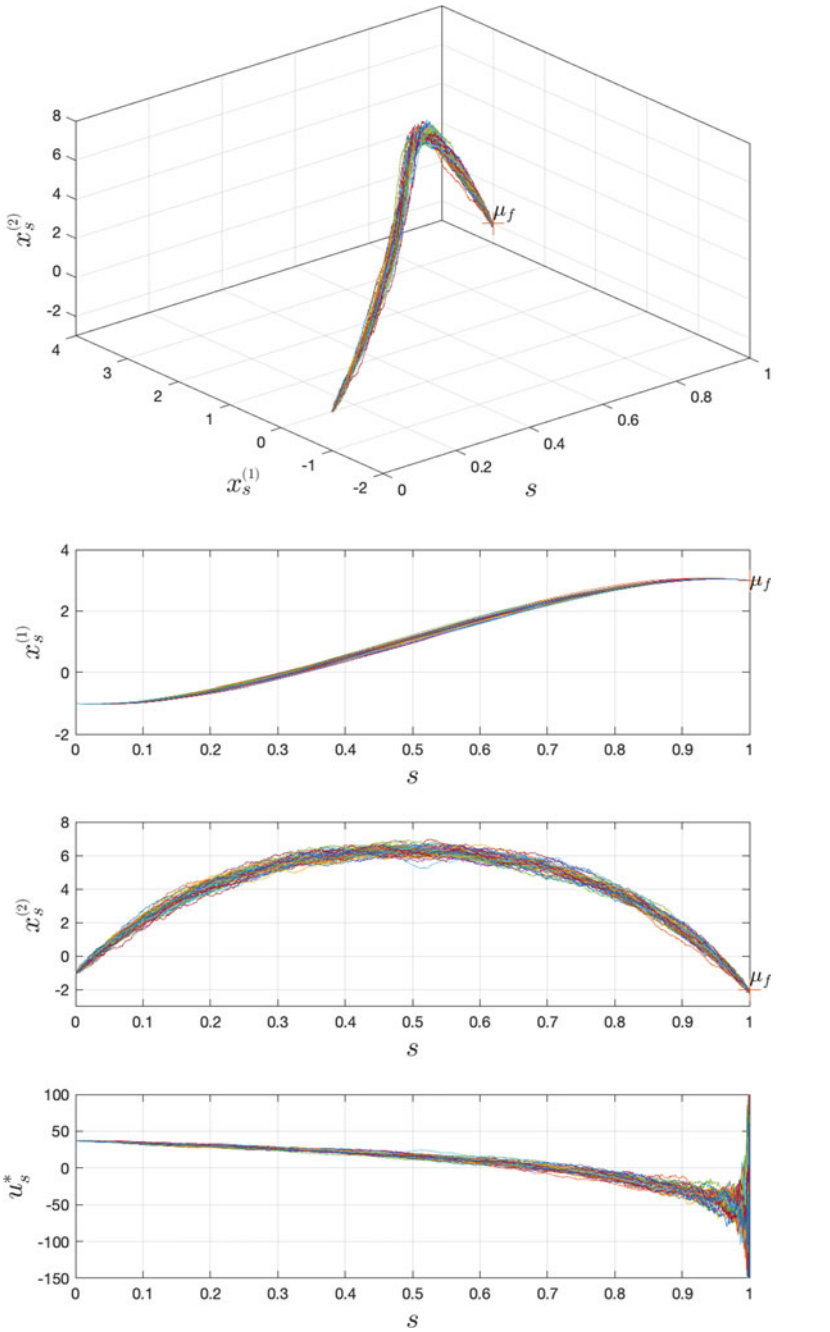


Fig. 9.3 The implementation of the terminally constrained stochastic minimum principle (TC-SMP) on the system in Example 2

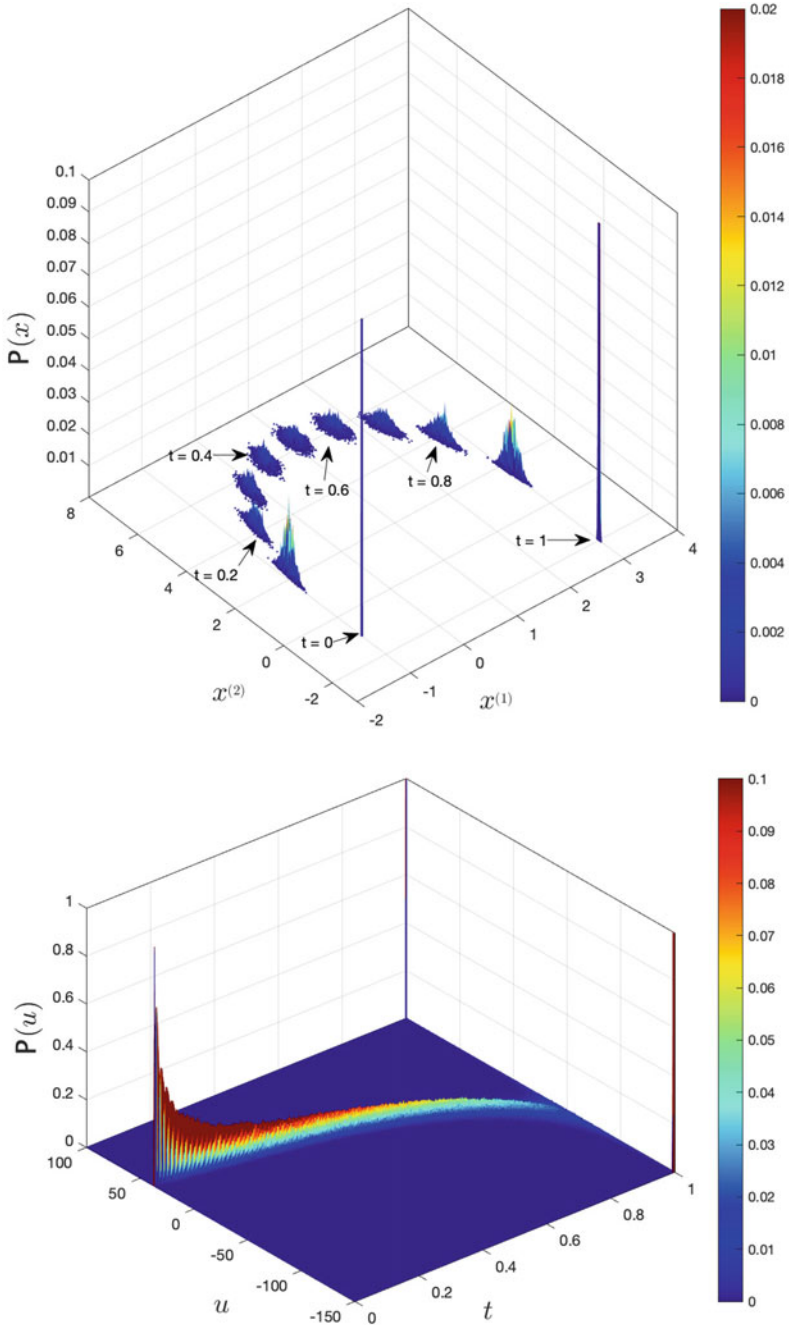


Fig. 9.4 The evolution of probability distributions of the state (top) and the input process (bottom) associated with the implementation of the terminally constrained stochastic minimum principle (TC-SMP) on the system in Example 2

Example 2 Consider the system governed by

$$\mathsf{d}x_s = \left(\begin{bmatrix} 0 & 1 \\ 2 & -3 \end{bmatrix}\begin{bmatrix} x_s^{(1)} \\ x_s^{(2)} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_s\right)\mathsf{d}s + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\mathsf{d}w_s, \tag{9.38}$$

over the time horizon $[t_0, t_f] = [0, 1]$, starting from the initial condition $x_0 = [-1, -1]^\top$, and steered towards the desired terminal state by enforcing

$$\mathbb{E}^{[u]}_{\mathscr{F}_t}[x_1] = \begin{bmatrix} 3 \\ -2 \end{bmatrix}, \tag{9.39}$$

at all $t \in [0, 1]$, and consider the associated optimal control problem with the cost

$$J\big(t, x_t, [u]\big) := \mathbb{E}^{[u]}_{\mathscr{F}_t}\Bigg[\int_t^{t_f} \frac{1}{2}u_s^2 \mathsf{d}s + \frac{1}{2}\left\|x_{t_f} - \mu_f\right\|^2\Bigg]. \tag{9.40}$$

The implementation of the TC-SMP for 50 Sample paths are illustrated in Fig. 9.3 and the evolution of the associated probability distributions for the state process and the input process are displayed in Fig. 9.4. In order to better illustrate the evolution of the probability distribution of the state process, the associated 3 dimensional snapshots of the state distribution at times $t \in \{0, 0.1, \ldots, 0.9, 1\}$ are displayed. As can be observed in these figures, the TC-SMP steers all trajectory realizations to the final desired state. It is worth remarking that the enforcement of the terminal state constraint (9.16) especially when $t \to t_f$, forces the controller to consume as much large values as required, that for x_t realizations away from μ_f, this requires the consumption of large input values. This can also be deduced from expression (9.27) for the optimal input, by noting that $\mathcal{G}(t, t_f) \to 0$ and therefore its inverse becomes as $t \to t_f$. It shall, however, be remarked that the associated singularity is isolated by noting that $t \leq s \leq t_f$

9.5 Constraining the Probability Distribution of the Terminal State

As mentioned earlier, in this method, we impose a terminal state constraint as $x^{[u]}_{t_f} \sim \mathsf{p}_d$, i.e., we require the probability distribution of the terminal state to take the desired form p_d. This, by definition, signifies that for every Borel set $B_x \in \mathbb{R}^n$,

$$\mathsf{P}^{[u]}\big(x_{t_f} \in B_x\big) = \int_{B_x} \mathsf{p}_d(\mathsf{d}x), \tag{9.41}$$

where $\mathsf{P}^{[u]}(\cdot)$ denotes the probability of an event given the input $[u]$. Hence, the value function at the initial time and state is defined as

$$V(t_0, x_0) = \inf_{[u]\in\mathcal{U}} \left\{ \mathbb{E}^{[u]}\left[\int_{t_0}^{t_f} \ell\,(x_s, u_s)\,\mathsf{d}s \right] \text{s.t. } x_T^{[u]} \sim \mathsf{p}_d \right\} \qquad (P)$$

In the absence of the constraint (9.41), one can invoke the convex duality approach of [14] to identify the value function as the upper envelope of the smooth subsolutions of the Hamilton-Jacobi and the associated boundary value inequalities. However, the presence of the constraint requires a more elaborate version of the convex duality approach established in [35] which is presented below.

9.5.1 Convex Duality and the Associated Hamilton–Jacobi (HJ) Inequalities

Theorem 3 [35] For every $x_0 \in \mathbb{R}^n$ and given a desired terminal distribution p_d, the optimal cost (P) is obtained as

$$\begin{aligned} V(t_0, x_0) = \sup_{v\in C^2([t_0,t_f]\times\mathbb{R}^n)} \Bigg\{ & v(t_0, x_0) - \int_{\mathbb{R}^n} v(t_f, x)\mathsf{p}_d(\mathsf{d}x), \\ \text{s.t. } & \frac{\partial v(t,x)}{\partial t} + \left[\frac{\partial v(t,x)}{\partial x}\right]^\top f(t,x,u) \\ & + \frac{1}{2}\text{tr}\Big(g(t,x)^\top g(t,x)\,\frac{\partial^2 v(t,x)}{\partial x^2}\Big) + \ell(t,x,u) \geq 0, \\ & \text{for all } (t,x,u) \in [t_0, t_f] \times \mathbb{R}^n \times U \Bigg\}. \end{aligned} \qquad (9.42)$$

□

Based upon the Theorem 3, a general procedure for the numerical solution can be developed as follows.

- Step 1: Set the iteration counter to $k = 0$, and initiate the algorithm with an arbitrary terminal cost function $L^k(x)$.
- Step 2: Solve the HJB equation[6]

[6] If a classical solution does not exist, one needs to consider an additional supremization over subsolutions of the HJB, i.e., the family of functions (indexed by another iteration j, satisfying the HJ inequalities $\frac{\partial v_j^k(t,x)}{\partial t} + \min_{u\in U}\left\{\left(\frac{\partial v_j^k(t,x)}{\partial x}\right)^\top f(x,u,t) + \ell(x,u,t)\right\} \geq 0$, but subject to the equality conditions $v_j^k(t_f, x) = L^k(x)$ for all j. However, it can be shown that the suprimizing function (over all j) converges to the viscosity solution of the HJB equation (9.43).

$$\frac{\partial v^k(t,x)}{\partial t} + \min_{u\in U}\left\{\left(\frac{\partial v^k(t,x)}{\partial x}\right)^{\top} f(x,u,t) + \ell(x,u,t)\right\} = 0, \quad (9.43)$$

subject to $v^k(t_f, x) = L^k(x)$.

- Step 3: Evaluate $v^k(t_0, x_0) - \int_{\mathbb{R}^n} v^k(t_f, x)\mathsf{p}_d(\mathsf{d}x)$.
- Step 4: Update $L^{k+1}(x)$ using an ascent direction[7] for the cost $v^k(t_0, x_0) - \int_{\mathbb{R}^n} v^k(t_f, x)\mathsf{p}_d(\mathsf{d}x)$.

9.5.2 Numerical Illustration

In order to illustrate the results of Theorem 3 and the difficulties and challenges associated with its algorithmic implementation let us consider the following examples.

Example 3 First, consider the the scalar system

$$\mathsf{d}x_s = (x_s + u_s)\mathsf{d}s + \mathsf{d}w_s, \quad (9.44)$$

with the total cost

$$J(t_0, x_0, [u]) = \mathbb{E}\int_0^T \frac{1}{2}u_s^2 \mathsf{d}s. \quad (9.45)$$

and subject to the constraint $x_{t_f} \sim \mathsf{p}_d$ for the desired probability distribution $\mathsf{p}_d = \mathcal{N}(\mu_d, \sigma_d)$. In order to illustrate (9.42), consider the family of functions $\{v_\gamma\}_{\gamma\in\mathbb{R}_+\times\mathbb{R}}$, where for each $\gamma \equiv (h_\gamma, \mu_\gamma)$, the function

$$v_\gamma(t,x) = \frac{1}{2}\pi(t)x^2 + \beta(t)x + \alpha(t), \quad (9.46)$$

is constructed from the Riccati equations

$$\dot{\pi}(t) = \big(\pi(t)\big)^2 - 2\pi(t), \qquad \pi(T) = h_\gamma, \quad (9.47)$$

$$\dot{\beta}(t) = -\big(1-\pi(t)\big)\beta(t), \qquad \beta(T) = -h_\gamma\mu_\gamma, \quad (9.48)$$

$$\dot{\alpha}(t) = \frac{1}{2}\big(\beta(t)\big)^2 - \frac{1}{2}\pi(t), \qquad \alpha(T) = \frac{1}{2}h_\gamma\mu_\gamma^2. \quad (9.49)$$

Then for every $\gamma \equiv (h_\gamma, \mu_\gamma) \in [0,\infty)\times\mathbb{R}$, the corresponding function $v_\gamma(t,x) \in C^\infty([t_0, t_f]\times\mathbb{R})$ satisfies

[7] Due to the computationally expensive nature of the cost, and the infinite dimensionality of the space of terminal costs, derivative-free methods such as Bayesian optimization shall be used in this general procedure.

$$\frac{\partial v_\gamma(t,x)}{\partial t} + \Big[\frac{\partial v_\gamma(t,x)}{\partial x}\Big](x+u) + \frac{1}{2}\frac{\partial^2 v_\gamma(t,x)}{\partial x^2} + \frac{1}{2}u^2 \geq 0, \tag{9.50}$$

for all $(t,x,u) \in [t_0, t_f] \times \mathbb{R} \times \mathbb{R}$.

For $x_0 = 0$, $T = 2$ and $\mathsf{p}_d = \mathcal{N}(3,1)$, the values of $v_\gamma(0,x_0) - \int_{\mathbb{R}^n} v_\gamma(T,x)$ are displayed over the region $(h_\gamma, \mu_\gamma) \in [0,5] \times [2.5, 3.5]$ in Fig. 9.5. As observed in the simulation, for this family of HJ-subsolutions, the maximum occurs at $(h_{\gamma^*}, \mu_{\gamma^*}) = (0.96, 3)$. As shown in [9], for LQG problems with Gaussian desired distributions, the value function is indeed quadratic and takes the Riccati form, and hence the desired value function V coincides with v_{γ^*} up to the constant $\int_{\mathbb{R}^n} v_{\gamma^*}(T,x)$.

Example 4 Now consider consider the the scalar system

$$\mathsf{d}x_s = (x_s + u_s)\mathsf{d}s + \mathsf{d}w_s, \tag{9.51}$$

with the total cost

$$J(t_0, x_0, [u]) = \mathbb{E}\int_0^T \frac{1}{2}u_s^2 \mathsf{d}s. \tag{9.52}$$

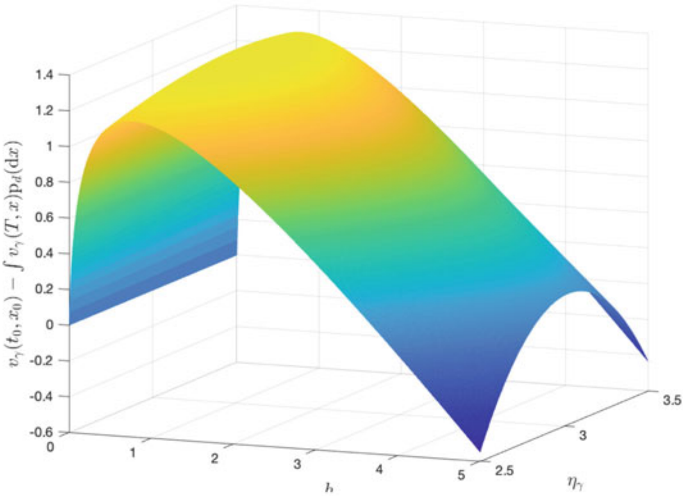
and subject to the constraint $x_{t_f} \sim \mathsf{p}_d$ where the desired probability distribution p_d is not Gaussian, e.g., when it's a mixture of two Gaussian distributions, e.g., $\mathsf{p}_d = \frac{1}{2}\mathcal{N}(\mu_1^d, \sigma_1^d) + \frac{1}{2}\mathcal{N}(\mu_2^d, \sigma_2^d)$. These problems, despite their LQ form of the dynamics and cost, cannot be solved by the conventional covariance control methodologies, e.g., [9]. In contrast, the results of Theorem 1 can be implemented in the following way to identify the value function and the corresponding optimal policy, as explained below.

Consider the family of functions $\{v_\gamma\}$ where for each $\gamma = (\eta_\gamma, \rho_\gamma, h_1^\gamma, \mu_1^\gamma, h_2^\gamma, \mu_2^\gamma)$, the function

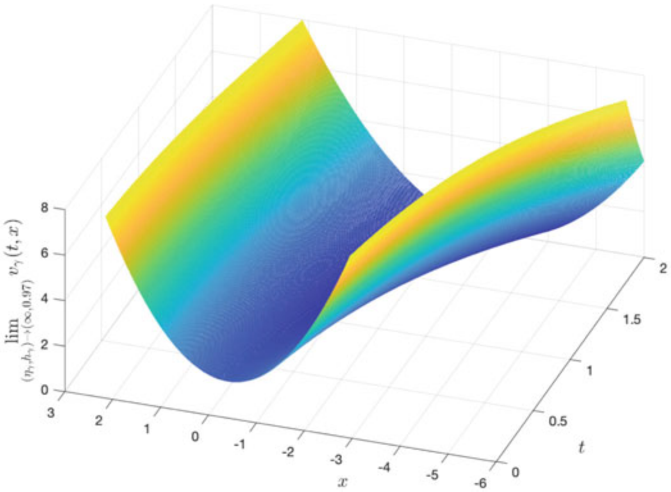
$$v_\gamma(t,x) = \frac{-1}{\eta_\gamma}\ln\Bigg(e^{-\eta_\gamma\rho_\gamma\left(\frac{1}{2}\pi_1(t)x^2+\beta_1(t)x+\alpha_1(t)\right)} + e^{-\eta_\gamma(1-\rho_\gamma)\left(\frac{1}{2}\pi_2(t)x^2+\beta_2(t)x+\alpha_2(t)\right)}\Bigg) \tag{9.53}$$

where $\pi_i, \beta_i, \alpha_i, i = 1,2$, satisfy (9.47)–(9.49) with $(h_\gamma, \mu_\gamma) = (h_i^\gamma, \mu_i^\gamma)$. Then it can be verified (see, e.g., [44]) that (9.53) satisfies the HJ-inequality (9.50) for all $(t,x,u) \in [t_0, t_f] \times \mathbb{R} \times \mathbb{R}$.

In order to restrict the search [as the primary purpose of the example is to illustrate the characterization of the value function by (9.42)] we consider the symmetric case where $\mu_1^d = -\mu_2^d = \mu_d$, $\sigma_1^d = \sigma_2^d$, and $x_0 = (\mu_1^\gamma + \mu_2^\gamma)/2 = 0$, thus $\rho_\gamma = 1/2$, and $\mu_1^\gamma = \mu^d$, and $\mu_2^\gamma = -\mu^d$. In particular, we consider the case with $x_0 = 0$, $T = 2$ and the desired distribution $\mathsf{p}_d = \frac{1}{2}\mathcal{N}(3,1) + \frac{1}{2}\mathcal{N}(-3,1)$, and hence we restrict attention to the sequence of functions parameterized by $\gamma = (\eta_\gamma, \frac{1}{2}, h_\gamma, 3, h_\gamma, -3) \equiv$

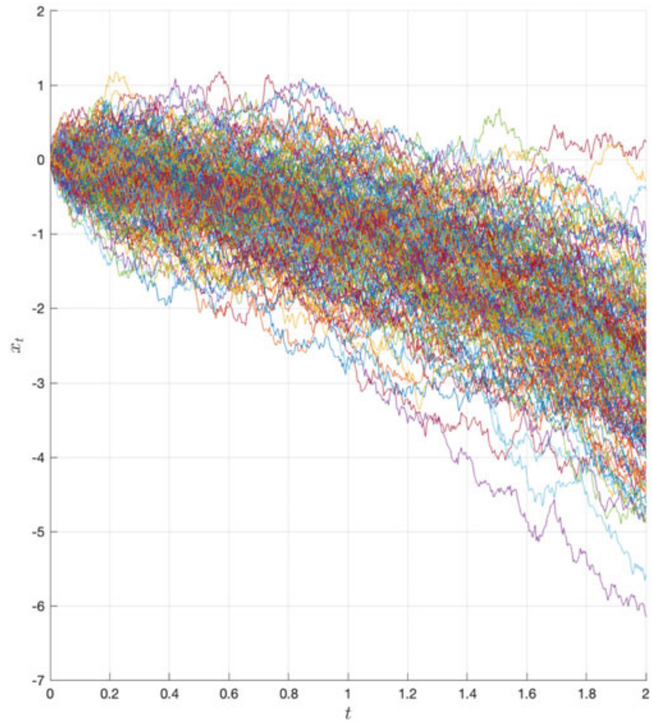


(a)

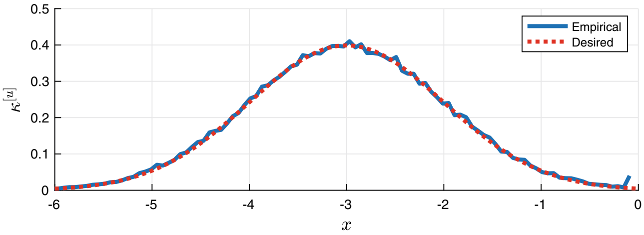


(b)

Fig. 9.5 The identification of the parameters of the value function for $x_0 = 0, T = 2$ and the desired distribution $\mathsf{p}_d = \mathcal{N}(3, 1)$ employing (9.42) and the class of functions $\{v_\gamma\}$ defined by (9.46). **a** Parameterization. **b** Associated limiting function. **c** Trajectories of 200 optimal sample paths. **d** Distribution pf x_T for 100,000 sample paths versus desired distribution



(c)

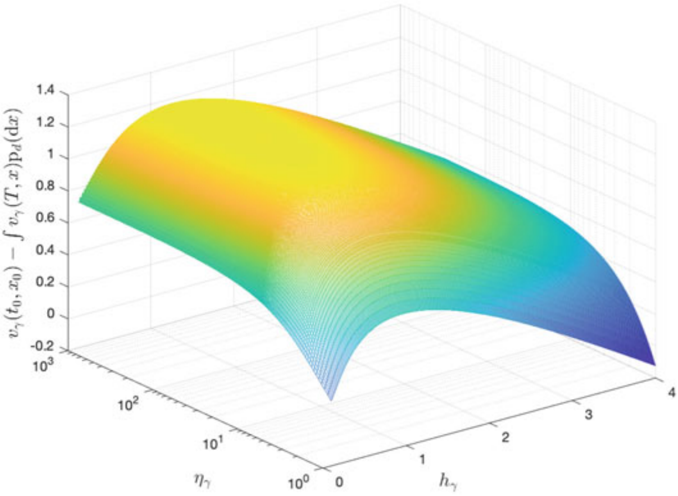


(d)

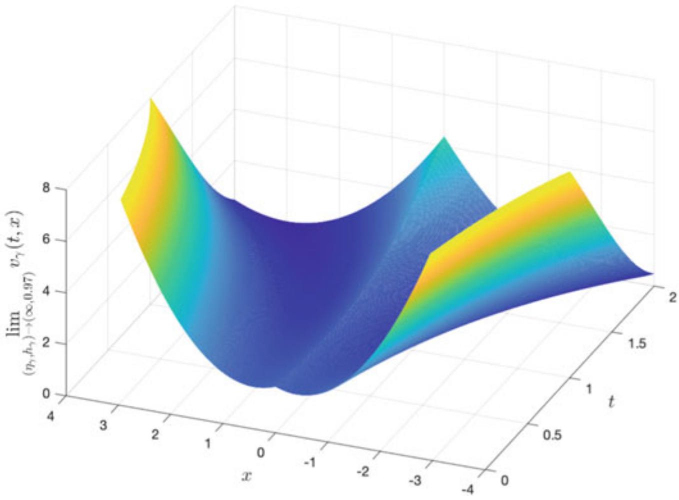
Fig. 9.5 (continued)

(η_γ, h_γ). The corresponding values of $v_\gamma(0, x_0) - \int_{\mathbb{R}^n} v_\gamma(T, x)$ are displayed over the region $(\eta_\gamma, h_\gamma) \in [1, 10^3] \times [0, 4]$ in Fig. 9.6.

As observed in Fig. 9.6a, for the family (9.53) of HJ-subsolutions, the supremum in not attained over the bounded domain and, while $h_{\gamma^*} = 0.97$, the suprimum requires $\eta_\gamma \to \infty$. Indeed, the value function in this case is nonsmooth and is required to be identified from $\lim_{(\eta_\gamma, h_\gamma)\to(\infty, 0.97)} v_\gamma(t, x)$ as displayed in Fig. 9.6b.

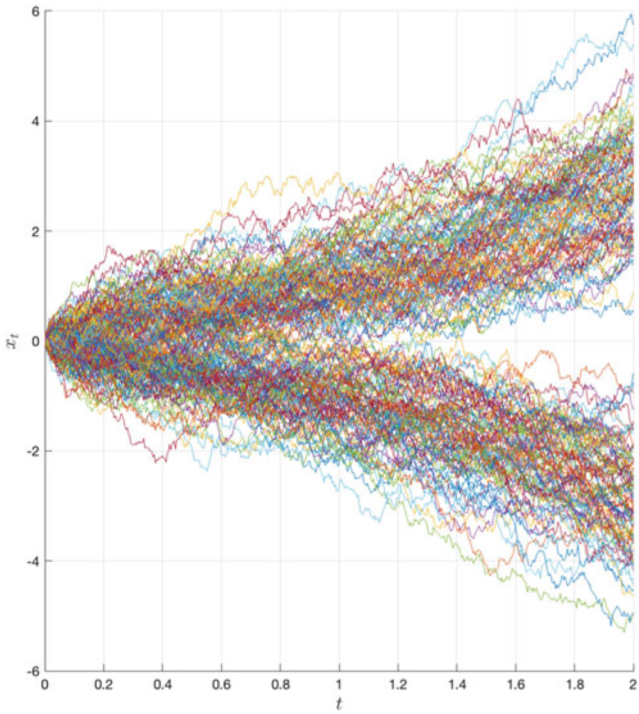


(a)

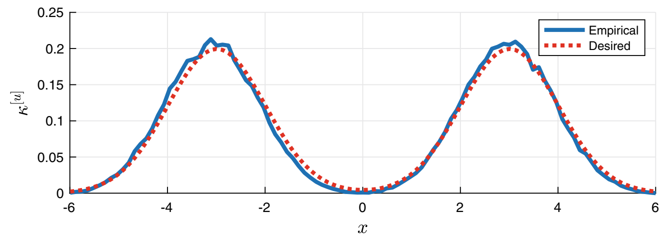


(b)

Fig. 9.6 The identification of parameters (**a**), the associated value function (**b**), sample paths (**c**), and the corresponding distribution (**d**) for $x_0 = 0$, $T = 2$ and $\mathsf{p}_d = \frac{1}{2}\mathcal{N}(3, 1) + \frac{1}{2}\mathcal{N}(-3, 1)$ employing (9.42) and the class of functions $\{v_\gamma\}$ defined by (9.53). **a** Parameterization. **b** Associated limiting function. **c** Trajectories of 200 optimal sample paths. **d** Distribution pf x_T for 100,000 sample paths versus desired distribution



(c)



(d)

Fig. 9.6 (continued)

In order to illustrate that the desired probability distribution is attained, the optimal trajectories of 200 sample paths are displayed in Fig. 9.6c and the empirical distribution of these trajectories obtained from 100,000 sample paths are displayed in Fig. 9.6d.

9.6 Concluding Remarks

This article presents two novel approaches for steering the state of nonlinear stochastic systems towards a desired value. Since equating the random variable of the term state to the desired value violates causality requirements, each of the presented methods provide an alternative causal expression of the terminal state requirement, and for each of these alternatives, theoretical guarantees for optimality and the satisfaction of the associated terminal state constraints are provided. The first approach is to impose a constraint on the first moment (expected value) of of the terminal state and to reimpose this constraint under conditional expectations at all future times. For this case the associated optimality conditions are expressed in the form of the Terminally Constrained Stochastic Minimum Principle (TC-SMP). The second approach is to impose a terminal state constraint as the matching of the probability distribution of the terminal state with a desired probability distribution in which case the associated optimality conditions are expressed using Hamilton-Jacobi (HJ) type equations. While both the TC-SMP and the convex duality based HJ inequalities are power methods in establishing the necessary optimality conditions for their corresponding optimal control problems, it is worth comparing the two.

The TC-SMP performs very successfully in terms of steering of the state towards the desired location an almost all sample paths are being steered to the desired value, however, this is achieved at the expense of increased control effort at the final time. Numerical examples illustrate that the TC-SMP achieves it goal by exploiting the unboundedness of the input value set. In other words, the enforcement of the terminal state constraint (9.16) especially when $t \to t_f$, forces the controller to consume as much large values as required, that for x_t realizations away from μ_f, this requires the consumption of large input values. This can also be deduced from expression (9.27) for the optimal input of the LQG case, by noting that $\mathcal{G}(t, t_f) \to 0$ and therefore its inverse becomes as $t \to t_f$. It shall, however, be remarked that the associated singularity is isolated by noting that $t \leq s \leq t_f$.

The convex duality approach, in contrast, is by definition a methodology for problems subject to terminal constraints of distribution type and, hence, its accuracy in delivering the state to a desired value depends on the expression of the desired probability distribution. Moreover, due to the involvement of the second moment (covariance) and higher order moments of the state distribution, this approach is inevitably time-inconsistent as the desired probability itself is only expressed under the total (and not other conditional) probability measure. The convex duality approach characterizes the value function by a set of Hamilton-Jacobi type conditions. A benefit of this characterization is that it holds true despite potential nonsmoothness of the value function. However, special care must be taken as (i) the sequence of test functions might converge to a value smaller than the optimal cost which suggests that the family of functions does not contain a function characterizing the value function, or (ii) a maximum does not exist but the suprimum (and hence the value function) can be characterized from the limiting behavior of the associated family of functions.

In both of these cases, a notion of suboptimality is required to be developed when the solution is numerically constructed from the associated optimization problem.

Future research directions include the accommodation of chance constraints, which are probabilistic constraints that impose a maximum probability of constraint violation, as a nonlinear systems extension of those established for linear systems in [31, 32], and also the accommodation of hybrid systems features, in particular the presence of controlled and autonomous switchings with exact equality and almost surely equality constraints as switching manifolds as in [33] and nonlinear jump maps as in [34, 36], as well as the development of numerical algorithms for numerical solutions, including the stochastic version of the Hybrid Minimum Principle - Multiple Autonomous Switching Algorithm [37], and Feynman-Kac based algorithms as in [18–20].

Acknowledgements Discussions concerning this work with Panagiotis Tsiotras are gratefully acknowledged.

References

1. Bakolas, E.: Optimal covariance control for discrete-time stochastic linear systems subject to constraints. In: Proceedings of the 55th IEEE Conference on Decision and Control, 1153–1158 (2016)
2. Bakolas, E.: Finite-Horizon covariance control for discrete-time stochastic linear systems subject to input constraints. Automatica **91**, 61–68 (2018)
3. Beghi, A.: On the relative entropy of discrete-time Markov processes with given end-point densities. IEEE Trans. Inf. Theor. **42**(5), 1529–1535 (1996)
4. Bellman, R.: Dynamic programming. Science **153**(3731), 34–37 (1966)
5. Bhatt, A.G., Borkar, V.S.: Occupation measures for controlled Markov processes: characterization and optimality. Ann. Probab. 1531–1562 (1996)
6. Caluya, K.F., Halder, A.: Finite horizon density steering for multi-input state feedback linearizable systems. In: 2020 American Control Conference (ACC), 3577–3582. IEEE (2020)
7. Caluya, K.F., Halder, A.: Wasserstein proximal algorithms for the Schrödinger bridge problem: density control with nonlinear drift. IEEE Trans. Autom. Control **67**(3), 1163–1178 (2021)
8. Chen, C.T.: Linear System Theory and Design. Rinhart and Winston, Orlando (1984)
9. Chen, Y., Georgiou, T.T., Pavon, M.: Optimal steering of a linear stochastic system to a final probability distribution, part I. IEEE Trans. Autom. Control **61**(5), 1158–1169 (2015)
10. Chen, Y., Georgiou, T.T., Pavon, M.: Optimal steering of a linear stochastic system to a final probability distribution, part II. IEEE Trans. Autom. Control **61**(5), 1170–1180 (2015)
11. Chen, Y., Georgiou, T.T., Pavon, M.: Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrodinger bridge. Siam Rev. **63**(2), 249–313 (2021)
12. Chen, Y., Ames, A.D.: Duality between Density Function and Value Function with Applications in Constrained Optimal Control and Markov Decision Process. arXiv:1902.09583 (2019)
13. Cho, M., Stockbridge, R.: Linear programming formulation for optimal stopping problems. SIAM J. Control Optim. **40**(6), 1965–1982 (2002)
14. Fleming, W.H., Vermes, D.: Convex duality approach to the optimal control of diffusions. SIAM J. Control Optim. **27**(5), 1136–1155 (1989)
15. Goldshtein, M., Tsiotras, P.: Finite-Horizon covariance control of linear time-varying systems. In: Proceedings of the 56th IEEE Conference on Decision and Control, 3606–3611 (2017)
16. Grigoriadis, K.M., Skelton, R.E.: Minimum-energy covariance controllers. Automatica **33**(4), 569–578 (1997)

17. Halder, A., Wendel, E.: Finite horizon linear quadratic gaussian density regulator with Wasserstein terminal cost. In: Proceedings of the American Control Conference, 7249–7254 (2016)
18. Hawkins, K.P., Pakniyat, A., Theodorou, E., Tsiotras, P.: Forward-backward rapidly-exploring random trees for stochastic optimal control. In: 2021 60th IEEE Conference on Decision and Control (CDC), 912–917. IEEE (2021a)
19. Hawkins, K.P., Pakniyat, A., Tsiotras, P.:. On the time discretization of the Feynman-Kac forward-backward stochastic differential equations for value function approximation. In: 2021 60th IEEE Conference on Decision and Control (CDC), 892–897. IEEE (2021b)
20. Hawkins, K.P., Pakniyat, A., Tsiotras, P.: Value function estimators for Feynman-Kac forward-backward SDEs in stochastic optimal control. Automatica (to appear in) (2023)
21. Hotz, A., Skelton, R.E.: Covariance control theory. Int. J. Control **46**(1), 13–32 (1987)
22. Kurtz, T., Stockbridge, R.: Existence of Markov controls and characterization of optimal Markov controls. SIAM J. Control Optim. **36**(2), 609–653 (1998)
23. Lasserre, J.B., Henrion, D., Prieur, C., Trélat, E.: Nonlinear optimal control via occupation measures and LMI-relaxations. SIAM J. Control Optim. **47**(4), 1643–1666 (2008)
24. Lewis, A.D.: The maximum principle of Pontryagin in control and in optimal control. In: Handouts for the course taught at the Universitat Politcnica de Catalunya. Department of Mathematics and Statistics Queen's University [online]. Available http://www.mast.queensu.ca/andrew/teaching/courses.shtml (May 5 2006)
25. Liu, C., Gray, A., Lee, C., Hedrick, J.K., Pan, J.: Nonlinear stochastic predictive control with unscented transformation for semi-autonomous vehicles. In: 2014 American Control Conference, 5574–5579. IEEE (2014)
26. Ma, J., Protter, P., Yong, J.: Solving forward-backward stochastic differential equations explicitly-a four step scheme. Probab. Theory Rel. Fields **98**(3), 339–359 (1994)
27. Majumdar, A., Vasudevan, R., Tobenkin, M.M., Tedrake, R.: Convex optimization of nonlinear feedback controllers via occupation measures. Int. J. Robot. Res. **33**(9), 1209–1230 (2014)
28. Mesbah, A., Streif, S., Findeisen, R., Braatz, R.D.: Stochastic nonlinear model predictive control with probabilistic constraints. In: Proceedings of the American Control Conference, 2413–2419 (2014)
29. Nodozi, I., Halder, A.: Schrödinger meets Kuramoto via Feynman-Kac: minimum effort distribution steering for noisy nonuniform Kuramoto oscillators. In: 2022 IEEE 61st Conference on Decision and Control (CDC), 2953–2960. IEEE (2022)
30. Okamoto, K., Tsiotras, P.: Stochastic model predictive control for constrained linear systems using optimal covariance steering (2019a). http://arxiv.org/abs/1905.13296
31. Okamoto, K., Goldshtein, M., Tsiotras, P.: Optimal covariance control for stochastic systems under chance constraints. IEEE Control Syst. Lett. **2**(2), 266–271 (2018)
32. Okamoto, K., Tsiotras, P.: Optimal stochastic vehicle path planning using covariance steering. IEEE Robot. Autom. Lett. **4**(3), 2276–2281 (2019)
33. Pakniyat, A., Caines, P.E.: On the stochastic minimum principle for hybrid systems. In: Proceedings of the 55th IEEE Conference on Decision and Control, Las Vegas, NV, USA, 1139–1144 (2016)
34. Pakniyat, A., Caines, P.E.: Hybrid optimal control of an electric vehicle with a dual-planetary transmission. Nonlin. Anal. Hybrid Syst. **25**, 263–282 (2017)
35. Pakniyat, A.: A convex duality approach for assigning probability distributions to the state of nonlinear stochastic systems. IEEE Control Syst. Lett. **6**, 3080–3085 (2022)
36. Pakniyat, A., Caines, P.E.: On the relation between the minimum principle and dynamic programming for classical and hybrid control systems. IEEE Trans. Autom. Control **62**(9), 4347–4362 (2017). https://doi.org/10.1109/TAC.2017.2667043
37. Pakniyat, A., Caines, P.E.: On the hybrid minimum principle: the Hamiltonian and adjoint boundary conditions. IEEE Trans. Autom. Control **66**(3), 1246–1253 (2020)
38. Pakniyat, A., Tsiotras, P.: Partially observed steering the state of linear stochastic systems. In: Proceedings of the 60th IEEE Conference on Decision and Control (CDC), 3780–3785. https://doi.org/10.1109/CDC45484.2021.9683692

39. Pakniyat, A., Tsiotras, P.: Steering the state of linear stochastic systems: a constrained minimum principle formulation. In: Proceedings of the 2021 IEEE American Control Conference (ACC), 1300–1305 (2021b)
40. Pauwels, E., Henrion, D., Lasserre, J.B.: Linear conic optimization for inverse optimal control. SIAM J. Control Optim. **54**(3), 1798–1825 (2016)
41. Peng, S.: A general stochastic maximum principle for optimal control problems. SIAM J. Control Optim. **28**(4), 966–979 (1990)
42. Pontryagin, L., Boltyanskii, V., Gamkrelidze, R., Mishchenko, E.: The Mathematical Theory of Optimal Processes, vol. 4. Wiley Interscience, New York (1962)
43. Ridderhof, J., Okamoto, K., Tsiotras, P.: Nonlinear uncertainty control with iterative covariance steering. In: 2019 IEEE 58th Conference on Decision and Control (CDC), 3484–3490. IEEE (2019)
44. Salhab, R., Malhamé, R.P., Le Ny, J.: Collective stochastic discrete choice problems: a min-LQG dynamic game formulation. IEEE Trans. Autom. Control **65**(8), 3302–3316 (2019)
45. Sehr, M.A., Bitmead, R.R.: Particle model predictive control: tractable stochastic nonlinear output-feedback MPC. IFAC-PapersOnLine **50**(1), 15361–15366 (2017)
46. Vasudevan, R., Gonzalez, H., Bajcsy, R., Sastry, S.S.: Consistent approximations for the optimal control of constrained switched systems-part 1: a conceptual algorithm. SIAM J. Control Optim. **51**(6), 4463–4483 (2013)
47. Vermes, D.: Optimal control of piecewise deterministic Markov processes. Stochast. Int. J. Probab. Stochast. Process. **14**(3), 165–207 (1985)
48. Vinter, R.B.: Dynamic programming for optimal control problems with terminal constraints. In: Recent Mathematical Methods in Dynamic Programming, pp. 190–202. Springer, Berlin Heidelberg (1985)
49. Vinter, R.: Convex duality and nonlinear optimal control. SIAM J. Control Optim. **31**(2), 518–538 (1993)
50. Vinter, R.B., Lewis, R.M.: A necessary and sufficient condition for optimality of dynamic programming type, making no a priori assumptions on the controls. SIAM J. Control Optim. **16**(4), 571–583 (1978)
51. Vinter, R.B., Lewis, R.M.: The equivalence of strong and weak formulations for certain problems in optimal control. SIAM J. Control Optim. **16**(4), 546–570 (1978)
52. Xu, J.H., Skelton, R.: An improved covariance assignment theory for discrete systems. IEEE Trans. Autom. Control **37**(10), 1588–1591 (1992)
53. Yasuda, K., Skelton, R.E., Grigoriadis, K.M.: Covariance controllers: a new parametrization of the class of all stabilizing controllers. Automatica **29**(3), 785–788 (1993)
54. Yi, Z., Cao, Z., Theodorou, E., Chen, Y.: Nonlinear covariance control via differential dynamic programming. In: 2020 American Control Conference (ACC), 3571–3576. IEEE (2020)
55. Yong, J., Zhou, X.Y.: Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer-Verlag, New York (1999)
56. Zhao, P., Mohan, S., Vasudevan, R.: Optimal control of polynomial hybrid systems via convex relaxations. IEEE Trans. Autom. Control **65**(5), 2062–2077 (2019)
57. Zhu, H.: Convex duality for finite-fuel problems in singular stochastic control. J. Optim. Theory Appl. **75**(1), 155–181 (1992)