Nonlinear Stochastic Receding Horizon Control: Stability, Robustness and Monte Carlo Methods for Control Approximation*

Francesco Bertoli[†]   Adrian N. Bishop[‡]

Abstract

This work considers the stability of nonlinear stochastic receding horizon control when the optimal controller is only computed approximately. A number of general classes of controller approximation error are analysed including deterministic and probabilistic errors and even controller sample and hold errors. In each case, it is shown that the controller approximation errors do not accumulate (even over an infinite time frame) and the process converges exponentially fast to a small neighbourhood of the origin. In addition to this analysis, an approximation method for receding horizon optimal control is proposed based on Monte Carlo simulation. This method is derived via the Feynman-Kac formula which gives a stochastic interpretation for the solution of a Hamilton-Jacobi-Bellman equation associated with the true optimal controller. It is shown, and it is a prime motivation for this study, that this particular controller approximation method practically stabilises the underlying nonlinear process.

1 Introduction

Receding horizon optimal control (RHC) is a strategy for controlling a dynamical system over an (possibly) infinite horizon where the control input at any instant is derived by solving a finite horizon optimal control problem over a fixed length horizon from that instant forwards. An introduction to RHC can be found in [29, 22]. RHC is a natural extension of finite-horizon optimal control and a natural simplification of infinite-horizon optimal control. The term model predictive control is often used interchangeably with RHC.

The contribution of this work is:

1 We study the stability of continuous-time receding horizon control of nonlinear stochastic systems when the optimal controller computation is only approximate. In particular, we consider a number of classes of controller approximation error including deterministic and probabilistic errors. We also consider controller sample and hold errors, that arise due to real-time computing limitations etc.

2a We outline a (Monte Carlo) simulation algorithm for approximating the optimal receding horizon control for nonlinear stochastic continuous-time systems.

2b We connect the controller approximation technique to the stability analysis detailed in this work and show that this Monte Carlo simulation method for controller approximation stabilises the process (in a sense to be made precise). In particular, the approximation errors do not accumulate nor destabilise the system.

The analysis of RHC for nonlinear (deterministic) systems started, largely, with the analysis of Mayne et. al. [28, 31]. Even in this early work, stability was considered for RHC in the presence of a number of controller approximation errors. Broad work on this topic in the nonlinear realm is covered in [33, 7, 13]. Much work in this area has focused on the incorporation of (deterministic) model uncertainty [31, 6, 25, 24] and/or controller and state constraints [31]. This latter focus concerning constraints is largely beyond the scope of this study, although we comment on possible extensions in our concluding remarks.

In the stochastic realm, the foundations of optimal control of nonlinear (continuous-time) systems are studied in, e.g., [10, 20, 43, 21, 39]. Computational methods for general nonlinear stochastic RHC are given in, e.g., [34, 21, 16, 30, 23, 14, 35, 8, 3]. In continuous-time cases, optimal nonlinear RHC has been shown [42] to be stabilising under the assumption that an optimal controller is applied (exactly). In this work, we extend [42] by showing that stability is retained even in the presence of controller approximation errors. We consider a number of general classes of controller approximation error, including deterministic and probabilistic errors and also controller sample and hold errors. To the best of our knowledge, there has been no investigation on the stochastic stability of (nonlinear stochastic) RHC in the presence of controller approximation errors.

In this work we also outline an approximation method for computing the optimal RHC for nonlinear stochastic continuous-time systems. This method is based on Monte Carlo integral approximation and originates in the work of Kappen [15, 16] where such techniques were applied in finite-horizon optimal control for nonlinear stochastic systems. The broad idea is that a solution to the Hamilton-Jacobi-Bellman partial differential equation associated with a typical optimal control problem [10] can be formulated in terms of an expectation over a stochastic trajectory defined by an uncontrolled stochastic differential equation (SDE). Indeed, this relationship between partial differential equations and so-called path-integrals is just a consequence of Feynman-Kac's formula [10]. This expectation (or path-integral) can then be approximated via Monte Carlo simulation, and since this defines the solution to the Hamilton-Jacobi-Bellman equation, it is a short leap from there to the optimal controller (or its approximation). This numerical algorithm for optimal control has received interest in, e.g., [40, 38, 36, 41, 32] where a number of generalisations (and applications) have been investigated. To the best of our knowledge, no investigation of the stochastic stability of this approximation method has been considered.

A contribution of this work is the analysis of this Monte Carlo based controller approximation method, as it applies to RHC of nonlinear stochastic continuous-time systems. In particular, we relate this approximation method to the more general stability analysis provided in this work and we show that this method stabilises the system (in a specific sense to be defined). This stability analysis justifies application of this control algorithm over an extended, possibly infinite, time interval.

The remainder of this work is organised as follows: In Section 2 we outline the basic nonlinear stochastic RHC problem and some related notation. In Section 3 we consider the stability of the nonlinear stochastic RHC regime. In particular, we note the stabilisation properties of the optimal (ideal) controller and we analyze the stability properties of a number of controller approximation methods. In Section 4 we introduce the Monte Carlo based algorithm for controller approximation and we relate this algorithm and its stability properties to the results given in the previous section. In Section 5 we provide some concluding remarks and comment on a number of possible extensions.

2 Nonlinear Stochastic Receding Horizon Optimal Control

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a complete probability space equipped with the natural filtration $(\mathscr{F}_t)_{t \geqslant 0}$ generated by a fixed, standard, Wiener process $W_t(\omega) : [0, \infty) \times \Omega \to \mathbb{R}^d$. We consider a nonlinear controlled process $X_t^{0, x_0, u}(\omega) : [0, \infty) \times \Omega \to \mathbb{R}^n$

$$dX_t^{0, x_0, u} = f(X_t^{0, x_0, u}, u_t)dt + g(X_t^{0, x_0, u})dW_t \tag{1}$$

with $X_0^{0, x_0, u} = x_0 \in \mathbb{R}^n$. We assume $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n \times d}$ to be continuous. The stochastic integrals in this paper are to be read in the Ito sense [1]. Moreover we assume that

$$|f(x, u) - f(y, u)| + |g(x) - g(y)| \leqslant c_1 |x - y|, \quad \forall (x, y, u) \in \mathbb{R}^n \times \mathbb{R}^n \times U$$

$$|f(x, u) - f(x, v)| \leqslant c_1 |u - v|, \quad \forall (x, u, v) \in \mathbb{R}^n \times U \times U$$

for some finite constant $c_1 > 0$. Let $t \geqslant s \geqslant 0$, then the superscripts $X_t^{s, x, u}$ denote that the initial state at $s \geqslant 0$ is x and the control history is $(u_t)_{t \geqslant s}$.

Fix a time interval $[t_0, t_1]$. Then a control $u_t(\omega) : [t_0, t_1] \times \Omega \to U$ is said to be admissible if it is Borel measurable and

$$\mathbb{E}\left[\int_{t_0}^{t_1} |u(X_s^{t_0, x, u})|^q ds \right] < \infty, \quad \forall x \in \mathbb{R}^n, \ q \geqslant 1$$

We denote by $\mathscr{U}_{[t_0, t_1]}$ the class of admissible controls on $[t_0, t_1]$. These conditions are sufficient for the existence of a unique, continuous, (strong) solution to the stochastic process; e.g. see [1, 39].

Here we consider control and stabilisation to (a neighbourhood of) the origin; any other desired set point can be substituted via a simple change of coordinates. To this end, we fix $f(0, u(0)) = 0$, i.e. the origin is an equilibrium point for the nominal deterministic system.

Let $T > 0$ be fixed. We associate with (1) the following receding horizon cost functional

$$w(t, s, x, u) := \mathbb{E}\left[\phi(X_{T+t}^{t+s, x, u}) + \int_{t+s}^{T+t} \ell(X_r^{t+s, x, u}, u_r)dr \right], \quad \text{for } s \in [0, T]$$

where $\ell : \mathbb{R}^n \times \mathbb{R}^m \to [0, \infty)$ and $\phi : \mathbb{R}^n \to [0, \infty)$ are non-negative continuous functions that satisfy

$$c_2 |x|^p \leqslant \phi(x) \leqslant c_3 (1 + |x|^p), \quad \forall x \in \mathbb{R}^n$$

and

$$c_2 (|x|^p + |u|^p) \leqslant \ell(x, u) \leqslant c_3 (1 + |x|^p + |u|^p), \quad \forall (x, u) \in \mathbb{R}^n \times U$$

for some finite (independent) constants $c_2, c_3 > 0$ and $p \geqslant 1$. Further, $\phi(0) = 0$ and $\ell(0, u(0)) = 0$.

We define a value functional as

$$v(t,s,x) := \inf_{u_r \in \mathscr{U}_{[t+s,t+T]}} w(t,s,x,u) = \inf_{u_r \in \mathscr{U}_{[t+s,t+T]}} \mathbb{E}\left[\phi(X_{T+t}^{t+s,x,u}) + \int_{t+s}^{T+t} \ell(X_r^{t+s,x,u}, u_r)dr\right] \tag{2}$$

and denote by $\overline{u}_r(x)$, if it exists, the optimal control, i.e. the admissible control process over the finite horizon $[t, T+t]$ that minimizes (2). In (one-step) RHC it is necessary just to compute $\overline{u}_r(x)$ for $r = t$ at which point the cost functional (and thus the value functional) changes to capture the receding horizon.

The value functional is time-invariant with respect to the first argument, in the sense that

$$\begin{aligned} v(t,s,x) &= \inf_{u_r \in \mathscr{U}_{[t+s,t+T]}} \mathbb{E}\left[\phi(X_{T+t}^{t+s,x,u}) + \int_{t+s}^{T+t} \ell(X_r^{t+s,x,u}, u_r(x))dr\right] \\ &= \inf_{u_r \in \mathscr{U}_{[s,T]}} \mathbb{E}\left[\phi(X_T^{s,x,u}) + \int_s^T \ell(X_r^{s,x,u}, u_r(x))dr\right] = v(0,s,x) \end{aligned}$$

Note that when viewed over $s \in [0, T]$ the value function represents the so-called value-to-go over the fixed finite horizon $[t, T+t]$. Going forward, we often write $v(x)$ in place of $v(t,0,x)$ or $v(s,x)$ in place of $v(t,s,x)$ when dealing with the value-to-go function.

With the modelling hypotheses adopted thus far, we have the following key lemma.

Lemma 1. *There exist a pair of positive constant c_4, c_5, depending only on p, T, c_1, c_2, c_3, such that*

$$c_4|x|^p \leqslant v(x) \leqslant c_5(1+|x|^p), \quad \forall\, x \in \mathbb{R}^n \tag{3}$$

and thus $v(x) \to \infty$ with $|x| \to \infty$.

Proof. The proof of the lemma is given in the appendix. $\qquad\square$

Going forward we write $\partial_x v(x)$ and $\partial_{xx} v(x)$ for the Jacobian vector and Hessian matrix respectively. Furthermore,

$$\partial_s v(s,x) := \frac{\partial v(s,x)}{\partial s} = \frac{\partial}{\partial s} \inf_{u_r \in \mathscr{U}_{[s,T]}} \mathbb{E}\left[\phi(X_{T+t}^{s,x,u}) + \int_s^T \ell(X_r^{s,x,u}, u_r(x))dr\right]$$

where $\partial_s v(s,x)$ is defined on $s \in [0, T]$ for any $t \geqslant 0$. We often write $\partial_s v(x)$ with $s \in [0, T]$ in place of $\partial_s v(s,x)$ for brevity.

Under certain conditions, at any time $s \in [0, T]$ the following Hamilton-Jacobi-Bellman (HJB) equation can be associated with the general value functional

$$-\partial_s v(s,x) = \inf_{u \in \mathscr{U}} \left[\ell(x,u) + f(x,u)^\top \partial_x v(s,x) + \tfrac{1}{2}\text{tr}\left[g(x)g(x)^\top \partial_{xx} v(s,x)\right]\right]$$

with a terminal boundary condition $v(T,x) = \phi(x)$. This association is in the sense that a suitably smooth solution to the HJB equation, if it exists, coincides with the value-to-go [10]. Note that in (one-step) RHC we are only interested in the solution of the HJB equation $v(s,x)$ at time $s = 0$ on the interval $s \in [0, T]$.

We note that assumptions introduced in this work are assumed to hold from the point at which they are introduced throughout the remainder of the work. The modelling hypotheses, e.g. on the functions f, g, ϕ, ℓ, etc. are assumed to hold throughout the remainder until the point they are refined (typically specialised) and from which point the refinement is supposed to hold.

Assumption 1. *We assume that the modelling assumptions outlined to this point are augmented (where and how necessary) to ensure $v(s,x) : [0, T] \times \mathbb{R}^n \to [0, \infty)$ is once continuously differentiable in $s \in [0, T]$ and twice continuously differentiable in x for all $x \in \mathbb{R}^n \setminus \{0\}$ and that $v(s,x)$ is a solution to the corresponding associated HJB equation.*

Sufficient conditions for this assumption to hold (in addition to the modelling hypotheses introduced thus far) are given in, e.g., [10]. Typically, these sufficient conditions take the form of further boundedness or regularity assumptions on the system and cost functions and/or their (partial) derivatives and are not overly restrictive[1]. We could also move from considering classical solutions of the HJB equation to generalised or viscosity solutions [10].

---

[1]It is noteworthy that while a classical solution to the HJB equation arising in deterministic optimal control is not typical, it is well-known [19, 10, 20, 39] that the stochastic optimal control problem is quite generally 'more regular'. Indeed, under an assumption of uniform parabolicity, i.e. uniform positive-definiteness of $g(x)g(x)^\top$, it generally follows that a classical (unique) solution to the HJB equation will exist in the stochastic setting (under mild regularity assumptions on the model/cost); see Chapter IV.4 in [10] or Krylov [20]. Under the modelling hypotheses introduced in this work, uniform parabolicity follows if $g(x)$ is full-rank (there is no real loss of generality here) and if $g(x)$ is bounded below (which is compatible with the hypotheses thus far). Separately, with an added Lipschitz assumption on the cost (compatible with the hypotheses herein), a classical solution to the HJB equation not only exists but is indeed Lipschitz [43, 4]. This Lipschitz setting is commonly assumed when studying the characteristics of stochastic optimal control; e.g. see [43]. Note, we do not generally require (or ask) for this Lipschitz property here. In any case, assumptions of this type concerning the existence of a classical solution are common in the analysis of both deterministic [28, 13] and stochastic optimal control [43].

Given the admissible optimal control process $\overline{u}_s(x)$ over the finite horizon $s \in [0, T]$ then the optimal value function over the finite horizon from any $t \geqslant 0$ to $T + t$ is

$$v(x) = \mathbb{E}\left[\phi(X_{T+t}^{t,x,\overline{u}}) + \int_t^{T+t} \ell(X_s^{t,x,\overline{u}}, \overline{u}_s(x))ds\right]$$

and, given Assumption 1, the value-to-go satisfies the following HJB equation

$$-\partial_s v(s,x) = \ell(x, \overline{u}_s(x)) + f(x, \overline{u}_s(x))^\top \partial_x v(s,x) + \tfrac{1}{2}\text{tr}\left[g(x)g(x)^\top \partial_{xx} v(s,x)\right] \tag{4}$$

on $s \in [0, T]$ with the terminal boundary condition $v(T, x) = \phi(x)$. We will use the following assumption.

Assumption 2. *We assume $\partial_s v(x)|_{s=t} \geqslant 0$ at any time $t \geqslant 0$ where $s \in [t, T + t]$.*

This assumption implies the optimal cost, when viewed at the start of a finite horizon, is increasing with decreasing horizon lengths[2]. One way to interpret this is that if the horizon length is reduced then the control action has less time to stabilize the system and thus the terminal cost is likely to be greater even though the running cost might be reduced.

Finally, we highlight again that in optimal RHC, at any time $t \geqslant 0$, the applied control is just $u_t^*(x) = \overline{u}_s(x)|_{s=0}$ and the remaining, finite horizon, controls $\overline{u}_s(x)$ over $0 < s \leqslant T$ are discarded.

3 Stability of Nonlinear Stochastic Receding Horizon Control and Robustness to Controller Errors

It has been shown in [42] that nonlinear stochastic receding horizon control stabilises the system to the origin if the true optimal control is applied (and under comparable assumptions and hypotheses to those considered here). In this section, we generalise this result to the case in which approximations in computing the optimal control are naturally employed[3].

Going forward we write $\mathscr{B}_\delta := \{x \in \mathbb{R}^n : |x| \leqslant \delta\}$ for the $\delta > 0$ ball around the origin and we use the shorthand $\{v < c\}$ to denote the level set $\{x \in \mathbb{R}^n : v(x) < c\}$ for $c \geqslant 0$. For any $\delta \geqslant 0$ define $m_\delta := \inf\{c \geqslant 0 \mid \mathscr{B}_\delta \subseteq \{v < c\}\}$. Informally stated, we study stability to $\{v < m_\delta\}$ where, from (1), it follows that $\{v < m_\delta\}$ is contained in a ball around the origin with radius tending to 0 as $\delta \to 0$.

3.1 Deterministic Control Errors

We introduce the following control signal $\widehat{u}_t(x) \in \mathcal{U}$ with

$$|\widehat{u}_t(x) - u_t^*(x)| \leqslant \epsilon, \quad \forall(x, t) \in \mathbb{R}^n \times [0, \infty)$$

for some sufficiently small $\epsilon > 0$. We denote by $\widehat{X}_t^{0,x_0,\widehat{u}}$ trajectories of (1) driven by $\widehat{u}_t(x)$ with $\widehat{X}_0 = X_0 = x_0$. If $\widehat{u}_t(x) \to u_t^*(x)$ then $\widehat{X}_t \to X_t^{0,x_0,u^*}$ for all $t \geqslant 0$; i.e. we recover the optimally controlled process in some suitable sense (to be made precise). The goal in this subsection is to prove that if $|\widehat{u}_t(x) - u_t^*(x)|$ is small then $\widehat{X}_t$ behaves similarly to X_t.

We will require the following assumption.

Assumption 3. *There exists a constant $c_6 > 0$ such that $|\partial_x v(x)| \leqslant c_6(1 + |x|^p)$, $\forall x \in \mathbb{R}^n$.*

For generality, we have stated our requirement that $|\partial_x v(x)| \leqslant c_6(1+|x|^p)$ as an assumption. Nevertheless, results of this type, i.e. results concerning estimates/bounds of the derivative of the value function, are well studied[4] and conditions on f, g, ℓ, and ϕ under which this assumption is guaranteed to hold are readily available [10, 20].

Theorem 1. *Suppose Assumptions 1, 2 and 3, and the modelling hypotheses hold. Define $\beta := c_5(1 + \frac{1}{\delta^p})$ and $\lambda := \frac{c_2}{\beta}$. Solutions of the SDE (1) driven by the optimal control u_t^* satisfy:*

- *if $x_0 \in \{v < m_\delta\}$ then, with probability one, X_t^{0,x_0,u^*} will never exit $\{v < m_\delta\}$ and $\mathbb{E}|X_t^{0,x_0,u^*}|^p \leqslant \frac{m_\delta}{c_4}$, $\forall t \geqslant 0$;*

---

[2]This assumption is common and discussed further in [42] and the references therein. It is proven to hold in [26] under quite typical modelling constraints (compatible with the modelling hypotheses presented here).

[3]In [42], stability of the origin (under the exact controller) is studied under the classical requirement [17] that $g(0) = 0$, i.e. that the origin is an 'exact' equilibrium for the diffusion, and thus the noise 'goes to zero' at this point. Here, we consider the practical case in which approximations are made when computing the optimal control, and we study stability to some neighbourhood of the origin. Thus, we also do not require $g(0) = 0$.

[4]For example, it is proven in [10] that $|\partial_x v(x)| \leqslant c_6(1 + |x|^p)$ holds under the given modelling hypotheses adopted in this work (on the cost/dynamics) with essentially the additional assumption $|\partial_x \ell(x, u)| \leqslant c(1 + |x|^p + |u|^p)$, $c > 0$. Thus, we already have the basic conditions on the dynamics/cost to ensure this assumption holds. See Chapter IV.8 in [10] and also Chapter 3 and Chapter 4 in [20]. Even stronger results have been proven [10, 39] implying this assumption holds trivially when one begins imposing Lipschitz conditions on the cost $\ell(x, u)$.

- if $x_0 \notin \{v < m_\delta\}$, it holds

$$\mathbb{E}|X_t^{0,x_0,u^*}|^p \leqslant \frac{1}{c_4}\left(\beta e^{-\lambda t}|x_0|^p + m_\delta\right), \qquad \forall t \geqslant 0,$$

and, with probability one, X_t^{0,x_0,u^*} hits $\{v < m_\delta\}$ in finite time.

These two points imply that almost all solutions to (1) driven by the optimal control u_t^* are exponentially stable to a ball around the origin and almost all trajectories remain within this ball.

Moreover there exists $\overline{\epsilon} > 0$ such that if

$$0 < \epsilon < \overline{\epsilon} \quad and \quad |\widehat{u}_t(x) - u_t^*(x)| \leqslant \epsilon, \quad \forall x \in \mathbb{R}^n$$

then solutions of the SDE (1) driven by the control law $\widehat{u}_t$ satisfy the following:

- if $x_0 \in \{v < m_\delta\}$, then with probability one, $\widehat{X}_t^{0,x_0,\widehat{u}}$ never exits $\{v < m_\delta\}$ and $\mathbb{E}|\widehat{X}_t^{0,x_0,\widehat{u}}|^p \leqslant \frac{m_\delta}{c_4}$, $\forall t$;

- if $x_0 \notin \{v < m_\delta\}$, there exists a constant $\lambda \geqslant \theta_\epsilon > 0$ such that

$$\mathbb{E}|\widehat{X}_t^{0,x_0,\widehat{u}}|^p \leqslant \frac{1}{c_4}\left(\beta e^{-\theta_\epsilon t}|x_0|^p + m_\delta\right), \qquad \forall t \geqslant 0,$$

and the constant θ_ϵ satisfies $\lim_{\epsilon \to 0} \theta_\epsilon = \lambda$. Further, with probability one, $\widehat{X}_t^{0,x_0,\widehat{u}}$ hits $\{v < m_\delta\}$ in finite time.

These two points imply that almost all solutions to (1) driven by the approximate control $\widehat{u}_t$ are exponentially stable to a ball around the origin and almost all trajectories remain within this ball. Such solutions converge exponentially fast under $\widehat{u}_t$ but slower than under u_t^*. In the limit $\epsilon \to 0$ we recover the stability properties of the optimal controller.

Proof. Going forward, we use the shorthand $\widehat{X}_t$ for the process (1) driven by $\widehat{u}_t(\widehat{X}_t)$. We prove only the second half of the theorem concerning the approximate controller $\widehat{u}_t$. Statements on the optimal control follow with $\epsilon = 0$.

Under the hypotheses of the theorem it follows that $c_4|x|^p \leqslant v(x) \leqslant c_5(1 + |x|^p)$ for some $c_2, c_4, c_5 > 0$. Let $\mathscr{L}$ denote the infinitesimal generator [17] of X_t^{0,x_0,u^*}. Then

$$\mathscr{L}v := f(x, u_t^*(x))^\top \partial_x v + \tfrac{1}{2}\mathrm{tr}\left[g(x)g(x)^\top \partial_{xx} v\right]$$

is a function of $x \in \mathbb{R}^n$ derived by applying the infinitesimal generator to $v(x)$. From (4), we have $\mathscr{L}v = -\ell(x, u_t^*(x)) - \partial_s v(s, x)|_{s=t}$ which by the modelling hypotheses and Assumption 2 is strictly negative definite $\mathscr{L}v < -c_2|x|^p$ for all $x \in \mathbb{R}^n \setminus \{0\}$ and $\mathscr{L}v = 0$ at $x = 0$.

Now, Ito's formula yields

$$dv(\widehat{X}_t) = \langle f(\widehat{X}_t, \widehat{u}_t(\widehat{X}_t)), \partial_x v(\widehat{X}_t)\rangle dt + \tfrac{1}{2}\mathrm{tr}[g(\widehat{X}_t)g(\widehat{X}_t)^\top \partial_{xx} v(\widehat{X}_t)]dt + \langle g(\widehat{X}_t)dW_t, \partial_x v(\widehat{X}_t)\rangle$$

Adding and subtracting $\langle f(\widehat{X}_t, u_t^*(\widehat{X}_t)), \partial_x v(\widehat{X}_t)\rangle$, and using the HJB equation (4), we obtain

$$dv(\widehat{X}_t) = \mathscr{L}v(\widehat{X}_t)dt + \langle f(\widehat{X}_t, \widehat{u}_t(\widehat{X}_t)) - f(\widehat{X}_t, u_t^*(\widehat{X}_t)), \partial_x v(\widehat{X}_t)\rangle dt + \langle g(\widehat{X}_t)dW_t, \partial_x v(\widehat{X}_t)\rangle$$

Set

$$\widehat{\mathscr{L}v}(x) := \mathscr{L}v(x) + \langle f(x, \widehat{u}_t(x)) - f(x, u_t^*(x)), \partial_x v(x)\rangle$$

where $\widehat{\mathscr{L}v}(x)$ is the infinitesimal generator of $\widehat{X}_t$ applied to $v(x)$, for any $x \in \mathbb{R}^n$.

Recall that the infinitesimal generator is a purely local construction [17] which allows us to consider $\widehat{\mathscr{L}v}(x)$ and $\mathscr{L}v(x) < -c_2|x|^p$ at the same point in space-time.

By using the Lipschitz condition on $f(x, u)$ and the Cauchy-Schwartz inequality we obtain

$$\begin{aligned}
\widehat{\mathscr{L}v}(x) &\leqslant \mathscr{L}v(x) + c_1|\partial_x v||\widehat{u}_t(x) - u_t^*(x)| \\
&\leqslant -c_2|x|^p + \epsilon c_1|\partial_x v| \\
&\leqslant -c_2|x|^p + \epsilon c_1 c_6(1 + |x|^p) = (-c_2 + \epsilon c_1 c_6)|x|^p + \epsilon c_1 c_6
\end{aligned}$$

We define

$$\alpha_\epsilon := c_2 - (\frac{1}{\delta^p} + 1)\epsilon c_1 c_6 \qquad and \qquad \theta_\epsilon := \frac{\alpha_\epsilon}{\beta}$$

There exists $\overline{\epsilon} > 0$ small enough so $\alpha_\epsilon > 0$, $\forall \epsilon < \overline{\epsilon}$. Moreover we see that $\lim_{\epsilon \to 0} \alpha_\epsilon = c_2 \Rightarrow \lim_{\epsilon \to 0} \theta_\epsilon = \lambda$. Going forward we write $\alpha = \alpha_\epsilon$ for simplicity. It is easy to check that on $\{x \in \mathbb{R}^n : |x| > \delta\}$ we have

$$v(x) \leqslant \beta|x|^p \quad and \quad \widehat{\mathscr{L}v}(x) \leqslant -\alpha|x|^p \tag{5}$$

We define $V(t,x) := v(x)e^{\frac{\alpha t}{\beta}}$. Then, on the set $\{x \in \mathbb{R}^n : |x| > \delta\}$, it follows

$$\widehat{\mathscr{L}V}(t,x) = e^{\frac{\alpha t}{\beta}}\left(\frac{\alpha}{\beta}v(x) + \widehat{\mathscr{L}v}(x)\right) \leqslant 0 \tag{6}$$

where $\widehat{\mathscr{L}V}(t,x)$ is the infinitesimal generator of $\widehat{X}_t$ applied to $V(t,x)$, for any $x \in \mathbb{R}^n$ at $t \geqslant 0$.

Assume now $x_0 \in \{v < m_\delta\}$. Given $t \geqslant 0$, define the stopping times

$$\tau_1 := \inf\{s \geqslant 0 \mid \widehat{X}_s \notin \{v < m_\delta\}\} \wedge t \quad \text{and} \quad \tau_2 := \inf\{s \geqslant \tau_1 \mid \widehat{X}_s \in \{v < m_\delta\}\} \wedge t$$

i.e. τ_1, τ_2 are, respectively, the first exit and re-entry time of the process $\widehat{X}_t$ in $\{v < m_\delta\}$ before t. Going forward we write $V(\widehat{X}_t)$ in place of $V(t, \widehat{X}_t)$ for simplicity. By definition, for any $t_2 \geqslant t_1$, we have

$$V(\widehat{X}_{t_2}) - V(\widehat{X}_{t_1}) = \int_{t_1}^{t_2} \widehat{\mathscr{L}V}(\widehat{X}_s)ds + \int_{t_1}^{t_2} e^{\frac{\alpha}{\beta}s}\partial_x v(\widehat{X}_s)^\top g(\widehat{X}_s)dW_s.$$

We know $\int_{t_1}^{t_2} e^{\frac{\alpha}{\beta}s}\partial_x v(\widehat{X}_s)^\top g(\widehat{X}_s)dW_s$ is a martingale. Then, the optional sampling theorem [9] implies

$$\mathbb{E}\left[V(\widehat{X}_{\tau_2}) - V(\widehat{X}_{\tau_1})\right] = \mathbb{E}\left[\int_{\tau_1}^{\tau_2} \widehat{\mathscr{L}V}(\widehat{X}_s)ds\right] \leqslant 0$$

where the last inequality follows from (6). We also note that, by definition, $V(\widehat{X}_{\tau_2}) \geqslant V(\widehat{X}_{\tau_1})$. Therefore we have $V(\widehat{X}_{\tau_2}) = V(\widehat{X}_{\tau_1})$ almost surely and consequently $\tau_1 = \tau_2$ almost surely. Thus, if the process starts in the set $\{v < m_\delta\}$ it can never exit this set. It follows that

$$\mathbb{E}|\widehat{X}_t|^p \leqslant \frac{1}{c_4}\mathbb{E}\left[v(\widehat{X}_t)\right] \leqslant \frac{m_\delta}{c_4}$$

and proof of the first point is complete.

Now assume $x_0 \notin \{v < m_\delta\}$, fix t and define the following stopping time

$$\tau := \inf\{s \geqslant 0 \mid \widehat{X}_s \in \{v < m_\delta\}\} \wedge t$$

Note that we have already considered the case $t \geqslant \tau \Rightarrow \widehat{X}_t \in \{v < m_\delta\}$. Now write

$$V(\widehat{X}_t) = v(x_0) + V(\widehat{X}_\tau) - v(x_0) + V(\widehat{X}_t) - V(\widehat{X}_\tau)$$

and take the expectation of both sides. Arguing as before $\mathbb{E}[V(\widehat{X}_\tau) - v(x_0)] \leqslant 0$. Moreover

$$\mathbb{E}\left[V(\widehat{X}_t) - V(\widehat{X}_\tau)\right] = \mathbb{E}\left[e^{\frac{\alpha}{\beta}t}v(\widehat{X}_t) - e^{\frac{\alpha}{\beta}\tau}v(\widehat{X}_\tau)\right] \leqslant m_\delta \mathbb{E}\left[e^{\frac{\alpha}{\beta}t} - e^{\frac{\alpha}{\beta}\tau}\right] \leqslant m_\delta e^{\frac{\alpha}{\beta}t}$$

Hence, using the two inequalities just shown, (5) and (3) we have

$$\mathbb{E}|\widehat{X}_t|^p \leqslant \frac{1}{c_4}\mathbb{E}[v(X_t)] = \frac{e^{-\frac{\alpha}{\beta}t}}{c_4}\mathbb{E}\left[V(\widehat{X}_t)\right] \leqslant \frac{e^{-\frac{\alpha}{\beta}t}}{c_4}(v(x_0) + m_\delta e^{\frac{\alpha}{\beta}t}) \leqslant \frac{\beta}{c_4}|x_0|^p e^{-\frac{\alpha}{\beta}t} + \frac{m_\delta}{c_4}$$

and proof of so-called exponential p-stability is complete.

Now, we have already shown $\mathbb{E}\left[V(\widehat{X}_\tau) - V(x_0)\right] = \mathbb{E}\left[\int_0^\tau \mathscr{L}V(\widehat{X}_s)ds\right] \leqslant 0$ which implies $\mathbb{E}\left[v(\widehat{X}_\tau)e^{\gamma\tau}\right] \leqslant v(x_0)$. Further,

$$\mathbb{E}\left[v(\widehat{X}_\tau)e^{\gamma\tau}\right] = \mathbb{E}\left[v(\widehat{X}_\tau)e^{\gamma\tau}\mathbb{1}_{\{\tau \neq t\}} + v(\widehat{X}_\tau)e^{\gamma\tau}\mathbb{1}_{\{\tau = t\}}\right] \geqslant \mathbb{E}\left[v(\widehat{X}_\tau)e^{\gamma\tau}\mathbb{1}_{\{\tau = t\}}\right] > m_\delta e^{\gamma t}\mathbb{P}(\tau = t)$$

and thus $m_\delta e^{\gamma t}\mathbb{P}(\tau = t) < v(x_0)$ for all t. This implies $\mathbb{P}(\tau = t) \to 0$ as $t \to \infty$ which in turn implies that $\mathbb{P}(\tau < \infty) = 1$. From this and inequality (6) it follows that $V(\tau, \widehat{X}_\tau)$ is a positive supermartingale. From Theorem 5.1 in [17] it follows that $V(\tau, \widehat{X}_\tau)$ converges almost surely to a finite limit (dependent on x_0) as $t \to \infty$. Then, from (5) we have

$$|\widehat{X}_\tau|^p \leqslant \frac{(\sup_t V(\tau, \widehat{X}_\tau))}{c_4}e^{-\frac{\alpha}{\beta}\tau}$$

with probability one. Letting $t \to \infty$ proves that almost all solutions converge exponentially fast toward $\{v < m_\delta\}$. Results of this type are known [17], i.e. where p-th moment exponential stability implies almost sure exponential stability. $\qquad\square$

3.2 Probabilistic Control Errors

We now turn our attention to the case where the perturbed controller has a Gaussian distribution,

$$\widehat{u}_t(x) \sim \mathcal{N}(u_t^*(x), \Sigma(x))$$

and the evolution of the nonlinear controlled process $X_t^{0,x_0,u}(\omega) : [0,\infty) \times \Omega \to \mathbb{R}^n$ follows

$$dX_t^{0,x_0,u} = f(X_t^{0,x_0,u})dt + h(X_t^{0,x_0,u})u_t dt + g(X_t^{0,x_0,u})dW_t \tag{7}$$

with the existing modelling hypotheses holding. Here $h : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ is continuous with

$$|f(x) - f(y)| + |g(x) - g(y)| + |h(x) - h(y)| \leqslant c_1|x - y|, \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$$

for some finite constant $c_1 > 0$. We denote by $\widehat{X}_t^{0,x_0,\widehat{u}}$ trajectories of (7) driven by $\widehat{u}_t(x)$ with $\widehat{X}_0 = X_0 = x_0$.

In this subsection we seek a result analogous to Theorem 1 under the proposed probabilistic controller error model. The goal is to show that if $\Sigma(x) \to 0$ for all $x \in \mathbb{R}^n$ then $\widehat{u}_t(x) \to u_t^*(x)$ and $\widehat{X}_t \to X_t^{0,x_0,u^*}$ for all $t \geqslant 0$; i.e. we recover the optimally controlled process in some suitable sense.

As before we need a further assumption on the derivatives of the value function.

Assumption 4. *One of the two following condition holds:*

- *there exists a constant $c_7 > 0$ such that $|\partial_{xx}v(x)| \leqslant c_7(1 + |x|^{p-2})$, $\forall x \in \mathbb{R}^n$;*

- *there exists a constant $c_7 > 0$ such that $|\partial_{xx}v(x)| \leqslant c_7(1 + |x|^p)$, $\forall x \in \mathbb{R}^n$ and $h(x)$ is bounded.*

As with Assumption 3, we have stated our requirement that $|\partial_{xx}v(x)| \leqslant c_7(1 + |x|^p)$ as an assumption (for the sake of generality). Yet similarly again, results of this type, i.e. results concerning estimates/bounds of the second derivative of the value function, are well studied in the literature[5] [10, 20].

The following is the main result of this subsection.

Theorem 2. *Suppose Assumption 1, 2 and 4 and the relevant modelling hypotheses hold. Define $\beta := c_5(1 + \frac{1}{\delta^p})$ and $\lambda := \frac{c_2}{\beta}$. Solutions of (7) driven by the optimal control u_t^* satisfy the relevant convergence results in Theorem 1.*

Moreover, there exists $\overline{\epsilon} > 0$ such that if $0 < \epsilon < \overline{\epsilon}$ and the following holds $\forall x \in \mathbb{R}^n$

- *$\widehat{u}_t(x) \sim \mathcal{N}(u_t^*(x), \Sigma(x))$;*

- *$0 < \Sigma(x) = \Sigma(x)^\top$ and for any norm $|\Sigma(x)| \leqslant \epsilon$*

then solutions of the SDE (7) driven by the approximated controller $\widehat{u}_t$ satisfy the following:

- *if $x_0 \in \{v < m_\delta\}$, then with probability one, $\widehat{X}_t^{0,x_0,\widehat{u}}$ never exits $\{v < m_\delta\}$ and $\mathbb{E}|\widehat{X}_t^{0,x_0,\widehat{u}}|^p \leqslant \frac{m_\delta}{c_4}$, $\forall t$;*

- *if $x_0 \notin \{v < m_\delta\}$, there exists a constant $\lambda \geqslant \theta_\epsilon > 0$ such that*

$$\mathbb{E}|\widehat{X}^{0,x_0,\widehat{u}}(t)|^p \leqslant \frac{1}{c_4}\left(\beta e^{-\theta_\epsilon t}|x_0|^p + m_\delta\right), \qquad \forall t \geqslant 0,$$

and θ_ϵ obeys $\lim_{\epsilon \to 0}\theta_\epsilon = \lambda$ (where λ is the convergence rate of the optimal control; see Theorem 1). Further, with probability one, $\widehat{X}_t^{0,x_0,\widehat{u}}$ hits $\{v < m_\delta\}$ in finite time.

Thus, almost all solutions to (7) driven by the approximate control $\widehat{u}_t$ are exponentially stable to a ball around the origin and almost all trajectories remain within this ball. Such solutions converge exponentially fast under $\widehat{u}_t$ but slower than under u_t^*. As $\epsilon \to 0$ we recover the stability properties of the optimal controller.

Proof. As before, denote by $\widehat{X}_t$ the process (7) driven by the approximated control $\widehat{u}_t(\widehat{X}_t)$. We quickly find

$$d\widehat{X}_t = f(\widehat{X}_t)dt + h(\widehat{X}_t)u_t^* dt + h(\widehat{X}_t)(\widehat{u}_t - u_t^*)dt + g(\widehat{X}_t)dW_t$$

Since $\Sigma(x)$ is (symmetric) positive-definite we have $\Sigma^{1/2}\Sigma^{1/2} = \Sigma(x)$ where $\Sigma^{1/2}$ exists and is unique. Then $\widehat{u}_t(x) \sim \mathcal{N}(u_t^*(x), \Sigma(x))$ implies $(\widehat{u}_t - u_t^*)dt = \Sigma^{1/2}dY_t$ where Y_t is a standard Brownian motion [1]. The two Brownian motions Y_t and W_t are realised on two different spaces: we have already fixed Ω and we denote by Ω' the space associated with the probabilistic controller approximation such that $[W_t^\top, Y_t^\top]^\top$ defines a fixed Brownian motion on $\Omega \times \Omega'$.

---

[5]As with Assumption 3 it is proven in [10] that $|\partial_{xx}v(x)| \leqslant c_7(1 + |x|^p)$ holds under the modelling hypotheses adopted in this work (on the cost/dynamics), with essentially the additional assumption that $|\partial_x \ell(x, u)| \leqslant c(1 + |x|^p + |u|^p)$ and $|\partial_{xx}\ell(x, u)| \leqslant c(1 + |x|^p + |u|^p)$, $c > 0$. See Chapter IV.9 in [10] and also Chapter 3 and Chapter 4 in [20]. Again, stronger results have been proven [10, 39] implying this assumption holds trivially when one imposes Lipschitz conditions on the cost $\ell(x, u)$, which is common in similar analysis [43].

Let $v(x) = \mathbb{E}[\phi(X_T) + \int_0^T \ell(X_s, u_s^*)ds]$ where the process X_t defining $v(x)$ is defined by (7) driven with the optimal control $u_t^*(x)$. We consider

$$\widehat{\mathscr{L}}v = \langle f(x) + h(x)u_t^*, \partial_x v \rangle + \frac{1}{2}\mathrm{tr}[g(x)g(x)^\top \partial_{xx}v] + \frac{1}{2}\mathrm{tr}[\Sigma(x)h(x)h(x)^\top \partial_{xx}v]$$

where $\widehat{\mathscr{L}}v(x)$ is the infinitesimal generator of $\widehat{X}_t$ applied to $v(x)$, for any $x \in \mathbb{R}^n$ at $t \geqslant 0$. Again, $\widehat{\mathscr{L}}v(x)$ should be viewed as a function of $x \in \mathbb{R}^n$.

We know that

$$\langle f(x) + h(x)u_t^*, \partial_x v \rangle + \frac{1}{2}\mathrm{tr}[g(x)g(x)^\top \partial_{xx}v] < -c_2|x|^p$$

from the proof of Theorem 1, i.e. $\mathscr{L}v(x) < -c_2|x|^p$. Owing to Assumption 4 we have, for some positive constant c,

$$\frac{1}{2}\mathrm{tr}[\Sigma(x)h(x)h(x)^\top \partial_{xx}v] \leqslant \epsilon c(|x|^p + 1)$$

and therefore,

$$\widehat{\mathscr{L}}v(x) \leqslant -c_2|x|^p + \epsilon c|x|^p + \epsilon c$$

Define $\alpha_\epsilon := c_2 - (\frac{1}{\delta^p} + 1)\epsilon c$ and the proof now follows exactly that of Theorem 1 and we omit the repetition for brevity. $\qquad \square$

3.3 Mixed Type Errors and Sampled Control

We now state a simple corollary that takes into account a mixed probabilistic and deterministic controller error.

Corollary 1. *Suppose we are working under (7) and Assumptions 1 to 4 and the modelling hypotheses outlined thus far hold. Define $\beta := c_5(1 + \frac{1}{\delta^p})$ and $\lambda := \frac{c_2}{\beta}$. There exist $\overline{\epsilon_1}, \overline{\epsilon_2} > 0$ such that if $0 < \epsilon_1 < \overline{\epsilon_1}$ and $0 < \epsilon_2 < \overline{\epsilon_2}$ and*

- *$(\widehat{u}_t(x) - u_t^*(x)) \sim \mathcal{N}(\mu(x), \Sigma(x))$,*

- *$|\mu(x)| \leqslant \epsilon_1$,*

- *$0 < \Sigma(x) = \Sigma(x)^\top$ and for any norm $|\Sigma(x)| \leqslant \epsilon_2$,*

holds $\forall x \in \mathbb{R}^n$, then solutions of the SDE (7) driven by the approximate control law $\widehat{u}_t$ satisfy:

- *if $x_0 \in \{v < m_\delta\}$, then with probability one, $\widehat{X}_t^{0,x_0,\widehat{u}}$ never exits $\{v < m_\delta\}$ and $\mathbb{E}|\widehat{X}_t^{0,x_0,\widehat{u}}|^p \leqslant \frac{m_\delta}{c_4}$, $\forall t \geqslant 0$;*

- *if $x_0 \notin \{v < m_\delta\}$, put $\epsilon = (\epsilon_1, \epsilon_2)$. There exists a constant $\lambda \geqslant \theta_\epsilon > 0$ such that*

$$\mathbb{E}|\widehat{X}_t^{0,x_0,\widehat{u}}|^p \leqslant \frac{1}{c_4}\left(\beta e^{-\theta_\epsilon t}|x_0|^p + m_\delta\right), \qquad \forall\, t \geqslant 0,$$

  and θ_ϵ obeys $\lim_{\epsilon \to 0}\theta_\epsilon = \lambda$ (where λ is the convergence rate of the optimal control; see Theorem 1). Also, with probability one, $\widehat{X}_t^{0,x_0,\widehat{u}}$ hits $\{v < m_\delta\}$ in finite time, i.e. almost all solutions converge exponentially fast toward $\{v < m_\delta\}$.

Proof. If $\mu(x) \in \mathscr{B}_{\epsilon_1} = \{x \in \mathbb{R}^n : |x| \leqslant \epsilon_1\}$ then $(\widehat{u}_t(x) - u_t^*(x)) \sim \mathcal{N}(\mu(x), \Sigma(x))$ implies $(\widehat{u}_t - u_t^*)dt = \mu(x) + \Sigma(x)^{1/2}dY_t$ where Y_t is a standard Brownian motion. It is then easily seen that the error is split in two parts, one part formed by the added Brownian motion and the other part formed by the deterministic error $\mu(x)$ with $|\mu(x)| \leqslant \epsilon_1$, $\forall x \in \mathbb{R}^n$. The proof of both Theorem 1 and 2 apply readily in this case and we omit the details for brevity. $\qquad \square$

In many practical scenarios it is impossible to compute the optimal control instantaneously and one must instead resort to a sample and hold approach to control whereby the control is computed at discrete-time increments and held constant in between such times. Stability results for such approaches have been considered, e.g., in deterministic settings [28] and stochastic settings [27]. We now provide a related stability result.

Proposition 1. *Consider the more general controlled process (1) and suppose Assumptions 1, 2, and the relevant modelling hypotheses hold. Suppose that under a given control law $u_t(x)$ the solution $X_t^{0,x_0,u}$ to the SDE (1) with initial condition x_0 satisfies*

$$\mathbb{E}|X^{0,x_0,u}(t)|^p \leqslant ce^{-\lambda t}|x_0|^p + m, \qquad \forall\, t \geqslant 0 \tag{8}$$

for some positive constants λ, c and m. Now fix a time step $\Delta > 0$ and let the time interval $t \in [0, \infty)$ be discretised according to $t_0 = 0$, $t_1 = \Delta$, $t_2 = 2\Delta$, ..., $t_k = k\Delta$. Consider the control law defined by $\widehat{u}_t(x_t) = u_{t_k}(x_{t_k})$ for $t \in [t_k, t_{k+1})$, i.e. the control $\widehat{u}_t$ is held constant over small time intervals with a value given by the control u_t at the beginning of each interval. Then there exists a constant step size $\overline{\Delta} > 0$ and constants $M_1, M_2 > 0$ such that

$$\mathbb{E}|\widehat{X}^{0,x_0,\widehat{u}}(t)|^p \leqslant M_1 e^{-\frac{\lambda}{4}t}|x_0|^p + M_2, \qquad \forall\, t \geqslant 0$$

for all $0 < \Delta \leqslant \overline{\Delta}$ and where we denote by $\widehat{X}_t^{0,x_0,\widehat{u}}$ trajectories of (1) driven by $\widehat{u}_t(x)$ with $\widehat{X}_0 = X_0 = x_0$.

Proof. Consider the two stochastic differential equations of the form (1) but driven by the two different controls defined in the statement of the theorem

$$
\begin{aligned}
dX_t^{0,x_0,u} &= f(X_t^{0,x_0,u}, u_t)dt + g(X_t^{0,x_0,u}, u_t)dW_t \\
d\widehat{X}_t^{0,x_0,\widehat{u}} &= f(\widehat{X}_t^{0,x_0,\widehat{u}}, \widehat{u}_t)dt + g(\widehat{X}_t^{0,x_0,\widehat{u}}, \widehat{u}_t)dW_t
\end{aligned}
$$

Then, since both processes share a common initial point, it is straightforward to show that the two Euler-Maruyama time-discretisations of both processes are identical. That is, by induction on $k \in \mathbb{N}$ we have

$$
\begin{aligned}
\widehat{Z}_{t_{k+1}}^{0,x_0,\widehat{u}} &= \widehat{Z}_{t_k}^{0,x_0,\widehat{u}} + f(\widehat{Z}_{t_k}^{0,x_0,\widehat{u}}, \widehat{u}_{t_k})\Delta + g(\widehat{Z}_{t_k}^{0,x_0,\widehat{u}}, \widehat{u}_{t_s})W_\Delta \\
&= Z_{t_k}^{0,x_0,u} + f(Z^{0,x_0,u}, u_{t_k})\Delta + g(Z_{t_k}^{0,x_0,u}, u_{t_k})W_\Delta = Z_{t_{k+1}}^{0,x_0,u}
\end{aligned}
$$

where $W_\Delta \sim \mathcal{N}(0, \Delta \cdot I)$ and $Z_0 = \widehat{Z}_0 = x_0$.

There is a known result [12] which states that p-th moment exponential stability of a stochastic differential equation implies p-th moment exponential stability of its Euler-Maruyama simulation and vice-versa (if the time-step $\Delta > 0$ is sufficiently small). Thus, with minor modifications to the main result of [12] it follows[6] that if (8) holds, then for a sufficiently small step size Δ, the Euler-Maruyama approximation $Z_t^{0,x_0,u}$ of $X_t^{0,x_0,u}$ satisfies

$$
\mathbb{E}|Z_t^{0,x_0,u}|^p \leqslant c|x_0|^p e^{-\frac{1}{2}\lambda t} + M, \qquad \forall\, t \geqslant 0
$$

for some $M > 0$. The same holds for $\widehat{Z}_t^{0,x_0,\widehat{u}}$ as this discrete-time process is identical to $Z_t^{0,x_0,u}$. Again, with slight modification to the results in [12] it follows that if Δ is small enough then

$$
\mathbb{E}|\widehat{X}_t^{0,x_0,\widehat{u}}|^p \leqslant M_1 e^{-\frac{\lambda}{4}t}|x_0|^p + M_2, \qquad \forall\, t \geqslant 0
$$

for some positive constants M_1, M_2. This completes the proof. $\qquad\square$

A straightforward consequence of Proposition 1 is that the convergence results given thus far concerning the various controller approximation errors will continue to hold even if the control is computed only at discrete-time instants and held constant in the interval between such instants (provided that the time elapsed between each updates is small).

The next result brings everything together.

Corollary 2. *Suppose the assumptions of either Theorem 1, Theorem 2 or Corollary 1 hold. Suppose also that $\widehat{u}_t$ is an approximately optimal control law satisfying the requirements of the respective result; e.g. $|\widehat{u}_t - u_t^*| < \epsilon < \overline{\epsilon}$ in Theorem 1 etc. Fix $\delta > 0$, we know that there exists a constant $\theta_\epsilon > 0$, satisfying the statement of the respective result, such that*

$$
\mathbb{E}|\widehat{X}^{0,x_0,\widehat{u}}(t)|^p \leqslant \frac{1}{c_4}\left(\beta e^{-\theta_\epsilon t}|x_0|^p + m_\delta\right), \qquad \forall\, t \geqslant 0
$$

Now suppose that $\widehat{u}_t(x)$ is computed at discrete times t_k with $t_0 = 0$, $t_1 = \Delta$, $t_2 = 2\Delta$, ..., $t_k = k\Delta$ and held constant on the interval $t \in [t_k, t_{k+1})$ as described in Proposition 1. Then there exists a constant step size $\overline{\Delta} > 0$ and constants $M_1, M_2 > 0$ such that

$$
\mathbb{E}|\widehat{X}^{0,x_0,\widehat{u}}(t)|^p \leqslant M_1 \beta e^{-\frac{\theta_\epsilon}{4}t}|x_0|^p + M_2 m_\delta, \qquad \forall\, t \geqslant 0
$$

for all $0 < \Delta \leqslant \overline{\Delta}$.

4 Monte Carlo Methods for Approximately Optimal Stochastic Control

In this section we outline an approximation method to compute the optimal nonlinear stochastic RHC. This method relies on simulating a stochastic process that is related to the original controlled system but that is independent of the control signal. The approximation method outlined in this section was first considered by Kappen [15, 16] for finite-horizon optimal control and then subsequently studied, applied, and generalised in, e.g., [40, 38, 36, 41, 3, 37].

Recall that we are considering the nonlinear controlled process $X_t^{0,x_0,u}(\omega) : [0,\infty) \times \Omega \to \mathbb{R}^n$ defined by

$$
dX_t^{0,x_0,u} = f(X_t^{0,x_0,u})dt + h(X_t^{0,x_0,u})u_t dt + g(X_t^{0,x_0,u})dW_t \tag{9}
$$

---

[6]The result in [12] must be modified since here we consider exponential stability to a ball of the origin (not the origin itself as in [12]). Thus, instead of the strong result of [12], we merely want exponential stability to the ball for a SDE to imply exponential stability to a (possibly different) ball for its Euler-Maruyama simulation (and vice-versa). The fact this relaxation is true follows easily (intuitively) given the strong result in [12]. It is causally unsurprising. Details on the modifications required to relax [12] as stated are available upon request (but needlessly distract the proof otherwise).

with the existing modelling hypotheses holding. Here, $h(x)$ and $g(x)$ (which may be non-square) are assumed (with no real loss of generality) to have full rank. Note, $h(x)$ full rank implies the existence and uniqueness of a left-inverse, i.e. a function $h^{-1}(x) : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ such that $h^{-1}h(x) = \mathrm{I}, \forall x \in \mathbb{R}^n$. Associate with (9) the following receding cost functional

$$w(t, s, x, u) := \mathbb{E}\left[\phi(X_{t+T}^{t+s,x,u}) + \int_{t+s}^{t+T} \tfrac{1}{2} u_r^\top R u_r + \ell(X_r^{t+s,x,u}) dr\right]$$

at any time $t \geqslant 0$ with $s \in [0, T]$ and where the cost on the control input is now quadratic and $R \in \mathbb{R}^{m \times m}$ is a constant positive definite matrix. We define the value-to-go functional as

$$v(t, s, x) := \inf_{u_r \in \mathcal{U}_{[t+s,t+T]}} w(t, s, x, u) = \inf_{u_r \in \mathcal{U}_{[t+s,t+T]}} \mathbb{E}\left[\phi(X_{t+T}^{t+s,x,u}) + \int_{t+s}^{t+T} \tfrac{1}{2} u_r^\top R u_r + \ell(X_r^{t+s,x,u}) dr\right] \tag{10}$$

where $\mathcal{U}_{[t+s,t+T]}$ is the set of admissible controls in the interval $[t + s, t + T]$.

The HJB equation associated with the value functional (10) is

$$-\partial_s v(s, x) = \inf_{u \in \mathcal{U}}\left[\ell(x) + \tfrac{1}{2} u^\top R u + \left[f(x) + h(x)u\right]^\top \partial_x v(s, x) + \tfrac{1}{2}\mathrm{tr}\left[g(x)g(x)^\top \partial_{xx} v(s, x)\right]\right]$$

with a terminal boundary $v(T, x) = \phi(x)$. The optimal control on the interval defined by $s \in [0, T]$ is just $u_{t+s}^*(x) = -R^{-1}h(x)^\top \partial_x v(s, x)$ for all $x \in \mathbb{R}^n$. In (one-step) RHC we are only interested in the solution $v(s, x)$ at $s = 0$. We have

$$u_t^*(x) = -R^{-1}h(x)^\top \partial_x v(x), \quad \forall x \in \mathbb{R}^n$$

Substituting the optimal control back into the HJB equation gives

$$-\partial_s v(s, x) = \ell(x) - \tfrac{1}{2}(\partial_x v(s, x))^\top h(x) R^{-1} h(x)^\top \partial_x v(s, x) + f(x)^\top \partial_x v(s, x) + \tfrac{1}{2}\mathrm{tr}\left[g(x)g(x)^\top \partial_{xx} v(s, x)\right]$$

which is a nonlinear partial differential equation. However, we note the following log-transform of $v(s, x)$

$$\psi(s, x) = \exp\left[\frac{-v(s, x)}{\gamma}\right]$$

for all $x \in \mathbb{R}^n$, $s \in [0, T]$ and for some finite $\gamma > 0$. This transform arises in a number of stochastic control scenarios [10]. We often write $\psi(x)$ in place of $\psi(0, x)$. We note the following required assumption.

Assumption 5. *We assume that there exists $\gamma \in \mathbb{R}$ such that $\gamma h(x) R^{-1} h(x)^\top = g(x)g(x)^\top$.*

This assumption[7] is standard in the path integral formulation of optimal control [16], but it also appears more generally in the stochastic optimal control literature [10]. This assumption allows us [10, 16] to write

$$-\partial_s \psi(s, x) = -\frac{1}{\gamma}\ell(x)\psi(s, x) + f(x)^\top \partial_x \psi(s, x) + \tfrac{1}{2}\mathrm{tr}\left[g(x)g(x)^\top \partial_{xx}\psi(s, x)\right]$$

which is a linear partial differential equation on $[0, T]$ with terminal condition $\psi(T, x) = \exp[-\phi(x)/\gamma]$. It now follows by the Feynman-Kac formula that the solution to the above PDE at $(0, x)$ is given by

$$\psi(x) = \mathbb{E}\left[\exp\left(-\frac{1}{\gamma}\phi(Z_{T+t}^{t,x}) - \frac{1}{\gamma}\int_t^{T+t}\ell(Z_s^{t,x})ds\right)\right]$$

where now $Z_s^{t,x}(\omega) : [t, T + t] \times \Omega \to \mathbb{R}^n$ is a nonlinear (uncontrolled) process satisfying

$$dZ_s^{t,x} = f(Z_s^{t,x})ds + g(Z_s^{t,x})dW_s \tag{11}$$

with initial condition $Z_t^{t,x} = x$. Note that

$$u_t^*(x) = -R^{-1}h(x)^\top \partial_x v(x) = \gamma R^{-1} h(x)^\top \partial_x \log\psi(x)$$

Now, given the solution for $\psi(x)$ derived via the Feynman-Kac formula, it is informally straightforward to devise a Monte Carlo approximation for the control; e.g. one can first simulate sample paths of (11), then form a Monte Carlo approximation of the integral for $\psi(x)$, and approximate the spatial derivative of $\psi(x)$ via differencing. Going forward we explore a more formal Monte Carlo approximation circumventing the need for crude numerical (spatial) differentiation. Firstly, we need the following result.

---

[7]This assumption is satisfied in many applications of stochastic control; e.g. in machine learning and robotics [16, 40, 38, 36, 41, 3]. This assumption requires the dimension of the noise and control to be equal and for the noise and control to act on the same subspace. Then, the cost of control can be related to the noise variance as shown. The interpretation of this relationship is that along directions where the noise variance is small, the control is deemed more expensive while, conversely, in those directions in which the noise has larger variance the control is cheap [16]. Indeed, this may be desirable in practice since it forces control energy to be spent mostly in those directions in which the noise level may be problematic [36].

Proposition 2. *Suppose Assumptions 1 and 5, and the modelling hypotheses hold. Then*

$$u_t^*(x) = -R^{-1}h(x)^\top \partial_x v(x) = \lim_{r \to 0} \frac{1}{r} \frac{\mathbb{E}\left[e^{-\frac{1}{\lambda}\left(\phi(Z_{t+T}^{t,x}) + \int_t^{t+T} \ell(Z_s^{t,x})ds\right)} \int_0^r h^{-1}(Z_s^{t,x})g(Z_s^{t,x})dW_s\right]}{\mathbb{E}\left[e^{-\frac{1}{\lambda}\left(\phi(Z_{t+T}^{t,x}) + \int_t^{T+t} \ell(Z_s^{t,x})ds\right)}\right]} \tag{12}$$

where the expectations are integrals over paths defined by the SDE (11) with initial condition $Z_t^{t,x} = x$.

Proof. This result appears in [37] with $h(x) = g(x)$ and it is straightforward to generalise. $\qquad \square$

The controller form in Proposition 2 (and variations of such) is often referred to as the path integral formulation of optimal control [16]. At this stage, it may appear as though the reformulated optimal controller has been significantly complicated. However, the optimal control as given in Proposition 2 is well suited to Monte Carlo approximation.

The Monte Carlo approach to RHC is given by Algorithm 1. Note also that we consider two time-discretizations defined by $\Delta_1 > 0$ and $\Delta_2 > 0$ respectively. The first, Δ_1, captures the sample and hold application in which the control is computed at discrete-time steps and held constant over those intervals; i.e. we approximate the optimal control $u_t^*(x)$ by $\widehat{u}_t(\widehat{x}_t) = \widehat{u}_{t_k}(\widehat{x}_{t_k})$ over $t \in [t_k, t_{k+1}) = [k\Delta_1, (k+1)\Delta_1)$. We denote by $\widehat{X}_t^{0,x_0,\widehat{u}}$ trajectories of (9) with $\widehat{X}_0 = X_0 = x_0$ driven by $\widehat{u}_t(x_t)$. The second time-discretization, Δ_2, is found solely within Algorithm 1 and defines the time-step employed during the numerical simulation of (11) used to actually compute $\widehat{u}_{t_k}(\widehat{x}_{t_k})$ at each t_k.

Algorithm 1: Optimal Control Approximation via Monte Carlo Simulation

---

Given at time t=0:

1. Model hypotheses: $f(\cdot)$, $g(\cdot)$, $h(\cdot)$, $\phi(\cdot)$, $\ell(\cdot)$, R, T, and γ.
2. Initial starting point: $x_0 \in \mathbb{R}^n$.
3. Discretization of time $t \in [0, \infty)$ via $t_0 = 0$, $t_1 = \Delta_1$, $t_2 = 2\Delta_1$, ..., $t_k = k\Delta_1$.
4. Discretization of the interval $[0, T]$ with step-size Δ_2 such that $T/\Delta_2 = K \in \mathbb{N}$.
5. Parameter approximating the limit $r > 0$ such that $r/\Delta_2 = R \in \mathbb{N}$.

Available at time t_k

1. Current state: $\widehat{x}_{t_k} \in \mathbb{R}^n$.

At time t_k do:

1. Simulate N times the following discrete-time approximation of (11)

$$Z_{t_{s+1}}^{0,\widehat{x}_{t_k}} = Z_{t_s}^{0,\widehat{x}_{t_k}} + f(Z_{t_s}^{0,\widehat{x}_{t_k}})\Delta_2 + g(Z_{t_s}^{0,\widehat{x}_{t_k}})W_{t_s}^{\Delta_2}$$

   over $t_s \in \{0, \Delta_2, \ldots, s\Delta_2, \ldots K\Delta_2\}$ where $W_{t_s}^{\Delta_2} \sim \mathcal{N}(0, \Delta_2 \cdot \mathrm{I})$. Simulation can be parallelised.

2. Let

$$z_{0:K}^{0,\widehat{x}_{t_k}}(i) := \{z_0^{0,\widehat{x}_{t_k}}(i) = \widehat{x}_{t_k}, z_2^{0,\widehat{x}_{t_k}}(i), \ldots, z_K^{0,\widehat{x}_{t_k}}(i)\}$$

   be the ordered set of sample points along the simulated discretised trajectory on the ith simulation run.

3. For each sampled trajectory $i \in \{1, \ldots, N\}$ compute

$$\widehat{W}(i) = \sum_{j=1}^{R} h^{-1}(z_{j-1}^{0,\widehat{x}_{t_k}}(i))g(z_{j-1}^{0,\widehat{x}_{t_k}}(i))\left(w_j^{\Delta_2}(i) - w_{j-1}^{\Delta_2}(i)\right)$$

   where $w_j^{\Delta_2}(i)$ are the sample points of $W_{t_j}^{\Delta_2}$ used previously to generate the trajectory $z_{0:K}^{0,\widehat{x}_{t_k}}(i)$.

4. For each sampled trajectory $i \in \{1, \ldots, N\}$ compute

$$\eta(i) = \phi(z_K^{0,\widehat{x}_{t_k}}(i)) + \sum_{j=0}^{K-1} \ell(z_j^{0,\widehat{x}_{t_k}}(i))\Delta_2$$

5. Compute

$$\widehat{u}_{t_k}(\widehat{x}_{t_k}) = \frac{1}{\sum_{i=1}^N \exp[-\frac{1}{\gamma}\eta(i)]} \sum_{i=1}^N \exp[-\frac{1}{\gamma}\eta(i)] \frac{\widehat{W}(i)}{r}$$

   which gives a (naive) Monte Carlo approximation of the optimal control. Let $\widehat{u}_t(\widehat{x}_t) = \widehat{u}_{t_k}(\widehat{x}_{t_k})$ over $t \in [t_k, t_{k+1}) = [k\Delta_1, (k+1)\Delta_1)$.

This algorithm is easily implementable. The numerical approximation of the stochastic differential equation (11) is known as the Euler-Maruyama method and is the simplest numerical scheme for approximating stochastic differential equations. This numerical approximation may be generalised [18] although care must be taken to

ensure that sufficient gains warrant the sharp increase in complexity that accompanies higher-order numerical approximation schemes.

The error in computing the approximate control signal at the discrete time sites is a mix of the error introduced due to the Monte Carlo sampling (known as the statistical error) and the error introduced due to the approximation of the limit and the time-discretisation (known as the discretisation error); see also [11, 2]. At those specific discretised time sites we note the following result.

Proposition 3. *Suppose Assumptions 1 and 5 and the modelling hypotheses employed to this point hold. Suppose also that the system and value functionals are sufficiently regular. Given $x \in \mathbb{R}^n$, suppose Algorithm 1 is used to compute $\hat{u}_t(x)$. Then there exists a positive constant μ_{Δ_2}, a function $\mu(x)$ satisfying $|\mu(x)| \leqslant \mu_{\Delta_2}$ and a matrix $\Sigma(x)$ such that*

$$\sqrt{N}\left(\hat{u}_t(x) - u_t^*(x) - \mu(x)\right) \to \mathcal{N}(0, \Sigma(x))$$

where convergence is 'in distribution' with the number, N, of Monte Carlo runs; see Algorithm 1. Also, $\lim_{\Delta_2 \to 0} \mu_{\Delta_2} = 0$.

Proof. Let u_t^* denote the optimal control defined by (12). For a fixed $r > 0$ approximate the limit defining

$$u_{t,r}^* = \frac{1}{r} \frac{\mathbb{E}\left[e^{-\frac{1}{\lambda}\left(\phi(Z_{t+}^{t,x}T) + \int_t^{t+T} \ell(Z_\tau^{t,x})d\tau\right)} \int_0^r h^{-1}(Z_\tau^{t,x}) g(Z_\tau^{t,x}) dW_\tau\right]}{\mathbb{E}\left[e^{-\frac{1}{\lambda}\left(\phi(Z_{t+T}^{t,x}) + \int_t^{T+t} \ell(Z_\tau^{t,x})d\tau\right)}\right]}$$

For r small enough we have $|u_t^*(x) - u_{t,r}^*(x)| < \epsilon$ with ϵ to be chosen later. Let $\tilde{u}_t^*$ be the approximation to $u_{t,r}^*$ found purely as a result of the discretized path approximation (associated with step-size Δ_2). Then, given the convergence results for the Euler-Maruyama method [18, 2], it follows that for Δ_2 small enough, there exists a constant c such that $|u_{t,r}^*(x) - \tilde{u}_t^*(x)| \leqslant c\Delta_2$ for all $r > 0$. Using the triangular inequality,

$$|u_t^*(x) - \tilde{u}_t^*(x)| \leqslant c\Delta_2 + \epsilon =: \mu_{\Delta_2} \tag{13}$$

Choosing $\epsilon \simeq \Delta_2$ yields $\lim_{\Delta_2 \to 0} \mu_{\Delta_2} = 0$. We note that $\hat{u}_t(x)$ is a Monte Carlo approximation of $\tilde{u}_t^*(x)$. We denote by $W_{0:K}$ a realised sequence of the discretized Brownian motion associated with the Euler-Maruyama discretization of (11) and by $\tilde{Z}_{0:K}^{0,x}(W_{0:K})$ the discrete path associated with it. Call $\mathbb{P}$ the natural measure on the path space $\{W_{0:K}\}$. Define

$$G(W_{0:K}) := \exp\left[-\frac{1}{\gamma}\left(\phi(\tilde{Z}_K^{0,x}(W_{0:K})) + \sum_{j=1}^{K-1} \ell(\tilde{Z}_j^{0,x}(W_{0:K}))\Delta_2\right)\right]$$

and consider the path measure $\mathbb{Q}$ obtained by the relation

$$d\mathbb{Q} = \frac{G(W_{0:K})}{\mathbb{E}_{\mathbb{P}}[G(W_{0:K})]} d\mathbb{P}$$

Define the function $F(W_{0:K}) = \sum_{j=1}^{R}(W_j - W_{j-1})$, i.e. the sum of the first R Brownian increments. We have that

$$r\hat{u}_t(x) = \mathbb{E}_{\mathbb{Q}}[F(W_{0:K})]$$

When simulating paths in Algorithm 1, we simulate from the measure $d\tilde{\mathbb{Q}} := G(W_{0:K})d\mathbb{P}$ and use self-normalized importance sampling to compute $r\hat{u}_t(x)$. We know [5], that self-normalized importance sampling is asymptotically unbiased and moreover a central limit theorem holds if

$$\int [1 + F^2]\left(\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}\right)^2 d\tilde{\mathbb{Q}} < \infty$$

Here, we have

$$\int [1 + F^2]\left(\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}\right)^2 d\tilde{\mathbb{Q}} = \frac{1}{\mathbb{E}_{\mathbb{P}}[G(W_{0:K})]^2}\left(\int (1 + F^2)d\tilde{\mathbb{Q}}\right)$$

Moreover $\int F^2 d\tilde{\mathbb{Q}} = \mathbb{E}_{\mathbb{P}}[F^2 G] < \infty$ thanks to the fact that G is bounded. Therefore we have

$$r\sqrt{N}\left(\hat{u}_t(x) - \tilde{u}_t^*(x)\right) \to \mathcal{N}(0, \Sigma(x))$$

where $\Sigma(x) = \int \left(\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}\right)^2 [F - \mathbb{Q}(F)]^2 d\tilde{\mathbb{Q}}$. Convergence is in the sense of distribution with N. Now divide by r, add and subtract $u_t^*(x)$, call $\mu(x) = u_t^*(x) - \tilde{u}_t^*(x)$ and use (13) to prove the convergence result. $\qquad\square$

The asymptotic bias in the preceding error result can be reduced by decreasing Δ_2 or via a reduction in the horizon length T. The variance can be reduced by increasing N or through some variation of naive sampling such as improved importance sampling or additionally via particle methods and resampling schemes [5, 16, 32] etc. The role of the parameter r with respect to the variance and the bias in the error approximation can be important; see [2] for a first study of this issue. Note also that

$$\Sigma(x) = \frac{\mathrm{Var}_\mu(F)}{\mathbb{E}_\mathbb{P}[G]}$$

and therefore the variance is intimately connected to $\mathbb{E}_\mathbb{P}[G]$, i.e. the interplay between the cost and dynamics of the uncontrolled SDE. Such performance questions may be explored in future work; see also [37, 2].

Going forward with the analysis we use the following assumption.

Assumption 6. *Suppose that, for N big enough, $(\widehat{u}_t(x) - u_t^*(x)) \sim \mathcal{N}(\mu(x), \frac{1}{N}\Sigma(x))$ and $\Sigma(x)$ is bounded.*

This assumption is just an invocation of the central limit type of result in Proposition 3 (which states that with N increasing, the distribution of the random part of the control approximation can be assumed Gaussian)[8].

We can now state the main stability result of this section.

Theorem 3. *Suppose Assumptions 1, 2, 3, 4, 5, 6 and the modelling hypotheses outlined to this point hold. Define $\beta := c_5(1 + \frac{1}{\delta^p})$ and $\lambda := \frac{c_2}{\beta}$. Given $x \in \mathbb{R}^n$, suppose Algorithm 1 is used to compute $\widehat{u}_t(x)$. With $\Delta_1, \Delta_2 > 0$ small enough and N large enough, there exits $\lambda \geqslant \theta > 0$ and a pair of positive constants $M_1, M_2 > 0$ such that*

$$\mathbb{E}|\widehat{X}^{0,x_0,\widehat{u}}(t)|^p \leqslant M_1 \beta e^{-\frac{\theta}{4}t}|x_0|^p + M_2 m_\delta, \qquad \forall\, t \geqslant 0$$

and $\lim_{\Delta_1, \Delta_2 \to 0, N \to \infty} \theta = \lambda$ (where λ is the rate of convergence of the optimal control; see Theorem 1).

Proof. Since the error $\widehat{u}_{t_k} - u_{t_k}^*$ is of the mixed type, we call on Corollary 1. From Proposition 3 it follows that there exists Δ_2 small enough so that the deterministic part of the controller approximation error is small. Similarly, from Proposition 3 it follows that there exists N large enough so that the variance $\Sigma(x)$ is small. Assumption 6 imposes normality on the error distribution. Corollary 1 applies immediately. Picking Δ_1 small enough to invoke Corollary 2 completes the proof. $\qquad\square$

It is practically feasible that such parameters, N (or Δ_2), could be made large (small) enough to ensure the variance (bias) in the approximation error is viable (and this is seen in practice where Algorithm 1 has been readily applied). However, the strong proof of the preceding theorem relies on the (approximating) assumption that $\widehat{u}_t - u_t^*$ is Gaussian. The error in this assumption, i.e. between the actual distribution and a Gaussian, decreases with increasing N, and can be quantified via Berry-Esseen type bounds [5]. In any case, the preceding theorem is rigorous under this assumption and, in practice, it provides justification for the stabilisation properties (seen in applications) of Algorithm 1 for large enough N. A stronger result with finite N would be difficult to obtain given that stochastic differential equations driven by arbitrary random processes (non-Brownian) are well beyond the scope of this work.

5   Concluding Remarks

In this work we explored the asymptotic stability of nonlinear stochastic RHC when the optimal controller is computed only approximately. We considered a number of general classes of controller approximation error including deterministic and probabilistic errors and even controller sample and hold errors. We also overviewed an approximation method for computing the optimal RHC for nonlinear stochastic continuous-time systems. This method is based on Monte Carlo integration approximation and originates in the work of Kappen [15, 16].

While we study the stability of various RHC approximations, we did not consider any measure of performance. For example, it would be of interest to analyze (path-wise) the running cost error that arises due to the approximation of the optimal controller. Inverse optimality and optimality gaps as studied in [26] would also be of interest here.

The incorporation of state constraints in RHC is common [29]. We note that it may be natural in some cases to incorporate state constraints in the Monte Carlo based approximation algorithm detailed herein. For example, state constraints may be enforced by simply restricting the evolution of the sampled trajectories (e.g. via dictating that certain regions of the state space hold zero probability).

Efficient sampling and Monte Carlo simulation [16, 32] that reduces the variance and thus the error in the Monte Carlo based controller approximation is of interest. Other computational aspects of this approximation are of interest, particularly as they apply to high-dimensional implementation.

---

[8]The point of this assumption is to impose normality on the distribution of the error $\widehat{u}_t - u_t^*$. Regardless of the distribution, it is true that the variance of the error decreases proportionally with increasing N (at the rate $1/N$) and that the bias decreases continuously with Δ_2. In practice, with N large enough, any error in applying this assumption is small and can be quantified via bounds of the Berry-Esseen type [5].

Finally, we mention that extensions which account for partial-information feedback may be important, particularly in the stochastic framework where true state feedback is overly restrictive. In this setting, the coupling of stochastic RHC, and particularly the Monte Carlo approximation algorithm, with sequential Monte Carlo estimation/filtering (e.g. particle filtering [5]) would be a natural topic for further study.

References

[1] L. Arnold. *Stochastic Differential Equations: Theory and Applications.* John Wiley and Sons, New York, 1974.

[2] F. Bertoli and A.N. Bishop. An error analysis in the limit approximation in path integral control. *Technical Note*, 2015.

[3] J. Bierkens and H.J. Kappen. Explicit solution of relative entropy weighted control. *Systems & Control Letters*, 72(10):36–43, October 2014.

[4] R. Buckdahn, J. Huang, and J. Li. Regularity properties for general HJB equations: A backward stochastic differential equation method. *SIAM Journal on Control and Optimization*, 50(3):1466–1501, 2012.

[5] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models.* Springer, New York, N.Y., 2005.

[6] G. De Nicolao, L. Magni, and R. Scattolini. On the robustness of receding-horizon control with terminal constraints. *IEEE Transactions on Automatic Control*, 41(3):451–453, March 1996.

[7] G. De Nicolao, L. Magni, and R. Scattolini. Stabilizing receding-horizon control of nonlinear time-varying systems. *IEEE Transactions on Automatic Control*, 43(7):1030–1036, July 1998.

[8] J.P. De Villiers, S.J. Godsill, and S.S. Singh. Particle predictive control. *Journal of Statistical Planning and Inference*, 141(5):1753–1763, 2011.

[9] J.L. Doob. *Stochastic Processes.* John Wiley & Sons, New York, N.Y., 1953.

[10] W.H. Fleming and H.M. Soner. *Controlled Markov Processes and Viscosity Solutions.* Springer, New York, 2nd edition, 2006.

[11] M.B. Giles. Multilevel Monte Carlo Path Simulation. *Operations Research*, 56(3):607–617, 2008.

[12] D.J. Higham, X. Mao, and A.M. Stuart. Exponential mean-square stability of numerical solutions to stochastic differential equations. *London Mathematical Society: Journal of Computation and Mathematics*, 6:297–313, 2003.

[13] A. Jadbabaie, J. Yu, and J. Hauser. Unconstrained receding-horizon control of nonlinear systems. *IEEE Transactions on Automatic Control*, 46(5):776–783, May 2001.

[14] N. Kantas, J.M. Maciejowski, and A. Lecchini-Visintini. Sequential Monte Carlo for Model Predictive Control. In *Nonlinear Model Predictive Control: Towards New Challenging Applications*, pages 263–273. Springer-Verlag, Berlin, Germany, 2009.

[15] H.J. Kappen. Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, 95(20):200201, November 2005.

[16] H.J. Kappen. Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11011, November 2005.

[17] R. Khasminskii. *Stochastic Stability of Differential Equations.* Springer-Verlag, Berlin, Germany, 2nd edition, 2011.

[18] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, Berlin, Germany, 3rd edition, 1999.

[19] N.V. Krylov. Control of a solution of a stochastic integral equation. *Theory of Probability & Its Applications*, 17(1):114–130, 1972.

[20] N.V. Krylov. *Controlled Diffusion Processes.* Springer, New York, N.Y., 2008.

[21] H. Kushner and P.G. Dupuis. *Numerical methods for stochastic control problems in continuous time.* Springer, New York, N.Y., 2001.

[22] W.H. Kwon and S. Han. *Receding Horizon Control: Model Predictive Control for State Models*. Springer-Verlag, London, 2006.

[23] W. Li and E. Todorov. Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. *International Journal of Control*, 80(9):1439–1453, September 2007.

[24] L. Magni, G. De Nicolao, R. Scattolini, and F. Allgöwer. Robust model predictive control for nonlinear discrete-time systems. *International Journal of Robust and Nonlinear Control*, 13(3-4):229–246, 2003.

[25] L. Magni, H. Nijmeijer, and A.J. Van Der Schaft. A receding–horizon approach to the nonlinear H_∞ control problem. *Automatica*, 37(3):429–435, March 2001.

[26] L. Magni and R. Sepulchre. Stability margins of nonlinear receding-horizon control via inverse optimality. *Systems & Control Letters*, 32(4):241–245, April 1997.

[27] M. Mahmood and P. Mhaskar. Lyapunov-based model predictive control of stochastic nonlinear systems. *Automatica*, 48(9):2271–2276, September 2012.

[28] D.Q. Mayne and H. Michalska. Receding horizon control of nonlinear systems. *IEEE Transactions on Automatic Control*, 35(7):814–824, July 1990.

[29] D.Q. Mayne, J.B. Rawlings, C.V. Rao, and P.O.M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, June 2000.

[30] W.M. McEneaney. *Max-plus methods for nonlinear control and estimation*. Springer, New York, N.Y., 2006.

[31] H. Michalska and D.Q. Mayne. Robust receding horizon control of constrained nonlinear systems. *IEEE Transactions on Automatic Control*, 38(11):1623–1633, November 1993.

[32] M. Morzfeld. Implicit sampling for path integral control. In *Proceedings of the 2014 American Control Conference*, pages 1839–1844, Portland, USA, June 2014.

[33] T. Parisini and R. Zoppoli. A receding-horizon regulator for nonlinear systems and a neural approximation. *Automatica*, 31(10):1443–1451, October 1995.

[34] T. Parisini and R. Zoppoli. Neural approximations for infinite-horizon optimal control of nonlinear stochastic systems. *IEEE Transactions on Neural Networks*, 9(6):1388–1408, June 1998.

[35] D. Stahl and J. Hauth. PF-MPC: Particle filter-model predictive control. *Systems & Control Letters*, 60(8):632–643, August 2011.

[36] E.A. Theodorou, J. Buchli, and S. Schaal. A generalized path integral control approach to reinforcement learning. *The Journal of Machine Learning Research*, 11:3137–3181, November 2010.

[37] S. Thijssen and H. J. Kappen. Path integral control and state-dependent feedback. *Physical Review E*, 91:032104, March 2015.

[38] E. Todorov. Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences*, 106(28):11478–11483, July 2009.

[39] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*. Springer, New York, N.Y., 2012.

[40] B. Van Den Broek, W. Wiegerinck, and H.J. Kappen. Graphical model inference in optimal control of stochastic multi-agent systems. *Journal of Artificial Intelligence Research*, 32:95–122, 2008.

[41] B. Van Den Broek, W. Wiegerinck, and H.J. Kappen. Risk sensitive path integral control. In *Proc. of the 26th Conference on Uncertainty in Artificial Intelligence*, Catalina Island, California, July 2010.

[42] F. Wei and A. Lecchini-Visintini. On the stability of receding horizon control for continuous-time stochastic systems. *Systems & Control Letters*, 63:43–49, January 2014.

[43] J. Yong and X.Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*. Springer, New York, 1999.

Appendix: Proof of Lemma 1

We start with the lower bound. Recall that $v(x) = \mathbb{E}[\phi(X_T) + \int_0^T \ell(X_s, u_s^*)ds]$ where we use the shorthand X_s for the solution of the system (1) driven by the optimal control with initial condition x. We have

$$\mathbb{E}[\phi(X_T)] \geqslant c_2 \mathbb{E}|X_T|^p \geqslant c_2 |\mathbb{E}[X_T]|^p$$

using the modelling hypotheses first and Jensen's inequality second. Moreover, we have

$$
\begin{aligned}
\mathbb{E}\left[\int_0^T \ell(X_s, u_s^*)ds\right] &\geqslant c_2 \mathbb{E}\left[\int_0^T (|X_s|^p + |u_s^*|^p)ds\right] \\
&\geqslant c_2 2^{1-p} \mathbb{E}\left[\int_0^T (|X_s| + |u_s^*|)^p ds\right] \\
&\geqslant \frac{c_2 2^{1-p}}{c_1} \mathbb{E}\left[\int_0^T |f(X_s, u_s^*)|^p ds\right] = \frac{c_2 2^{1-p}}{T^p c_1} \mathbb{E}\left[\int_0^T |Tf(X_s, u_s^*)|^p ds\right]
\end{aligned}
$$

where we have used the modelling hypotheses on $\ell(x, u)$, together with the fact that $|f(x, u)| \leqslant c_1(|x| + |u|)$. Now it follows that

$$
\begin{aligned}
\mathbb{E}\left[\int_0^T \ell(X_s, u_s)ds\right] &\geqslant \frac{c_2 2^{1-p}}{T^{p-1} c_1} \mathbb{E}\left[\left|\int_0^T f(X_x, u_s^*)ds\right|^p\right] \\
&\geqslant \frac{c_2 2^{1-p}}{T^{p-1} c_1} \left|\mathbb{E}\left[\int_0^T f(X_x, u_s^*)ds\right]\right|^p \\
&= \frac{c_2 2^{1-p}}{T^{p-1} c_1} \left|\mathbb{E}[X_T - x]\right|^p = \frac{c_2 2^{1-p}}{T^{p-1} c_1} \left|x - \mathbb{E}[X_T]\right|^p
\end{aligned}
$$

where we used Jensen's inequality twice. Putting together the bounds on $\mathbb{E}[\phi(X_T)]$ and $\mathbb{E}[\int_0^T \ell(X_s, u_s)ds]$ gives

$$
\begin{aligned}
v(x) &\geqslant c_2 |\mathbb{E}[X_T]|^p + \frac{c_2 2^{1-p}}{T^{p-1} c_1} |x - \mathbb{E}[X_T]|^p \\
&\geqslant c\left(|\mathbb{E}[X_T]| + |x - \mathbb{E}[X_T]|\right)^p \\
&\geqslant c\left(|\mathbb{E}[X_T] + x - \mathbb{E}[X_T]|\right)^p = c|x|^p
\end{aligned}
$$

for some constant $c > 0$, where we used the fact the all norms are equivalent on a finite dimensional vector space followed by the triangle inequality. This completes the proof for the lower bound.

We now turn to the upper bound and recall that, given an arbitrary admissible control law, the cost functional $w(t, x, u)$ associated with (1) is

$$w(0, x, u) := \mathbb{E}\left[\phi(X_T) + \int_0^T \ell(X_s, u_s)ds\right]$$

We immediately have

$$v(x) := \inf_{u_s \in \mathcal{U}_{[0,T]}} w(0, x, u) \leqslant w(0, x, 0)$$

where $w(0, x, 0)$ denotes the cost found after applying a constant zero control $u_t \equiv 0$. Going forward, write $X_t := X_t^{0,x,0}$ for the solution of (1) with $X_0 = X_0^{0,x,0} = x$ and a constant zero control $u_t \equiv 0$. Then

$$dX_t = f(X_t, 0)dt + g(X_t, 0)dW_t$$

Note that if $u_t = 0$ is not admissible we may substitute some other (sub-optimal) constant control signal. Then, for all $p \geqslant 1$ with $t \geqslant 0$, its known [39] that given the existence of solutions to (1) it holds that

$$\mathbb{E}|X_t|^p \leqslant c(1 + |x|^p)e^{ct}$$

for some finite $c > 0$. This, together with the assumptions on the cost, gives

$$
\begin{aligned}
w(0, x, 0) &= \mathbb{E}\left[\phi(X_T) + \int_0^T \ell(X_s, 0)ds\right] \\
&\leqslant \mathbb{E}\left[c_3(1 + |X_T|^p)\right] + \mathbb{E}\left[\int_0^T c_3(1 + |X_s|^p + 0)ds\right] \\
&\leqslant c_3\left(1 + T + |x|^p e^{cT} + \int_0^T |x|^p e^{cs}ds\right) \leqslant c_3\left(1 + T + |x|^p e^{cT} + |x|^p \frac{e^{cT} - 1}{c}\right)
\end{aligned}
$$

which, after gathering constants, completes the proof concerning the upper-bound.

Bringing everything together, it follows that there exists a pair of positive constants c_4, c_5, depending only on p, T, c_1, c_2, c_3, such that $c_4 |x|^p \leqslant v(x) \leqslant c_5(1 + |x|^p)$, $\forall x \in \mathbb{R}^n$ and $v(x) \to \infty$ with $|x| \to \infty$. $\qquad \square$