Journal of Machine Learning Research 18 (2017) 1-37 Submitted 11/16; Revised 6/17; Published 10/17

Stability of Controllers for Gaussian Process Dynamics

Julia Vinogradska[1,2] JULIA.VINOGRADSKA@DE.BOSCH.COM
Bastian Bischoff[1] BASTIAN.BISCHOFF@DE.BOSCH.COM
Duy Nguyen-Tuong[1] DUY.NGUYEN-TUONG@DE.BOSCH.COM
Jan Peters[2,3] MAIL@JAN-PETERS.NET

[1] *Corporate Research, Robert Bosch GmbH*
Robert-Bosch-Campus 1
71272 Renningen

[2] *Intelligent Autonomous Systems Lab, Technische Universität Darmstadt*
Hochschulstraße 10
64289 Darmstadt

[3] *Max Planck Institute for Intelligent Systems*
Spemannstraße 10
72076 Tübingen

Editor: George Konidaris

Abstract

Learning control has become an appealing alternative to the derivation of control laws based on classic control theory. However, a major shortcoming of learning control is the lack of performance guarantees which prevents its application in many real-world scenarios. As a step towards widespread deployment of learning control, we provide stability analysis tools for controllers acting on dynamics represented by Gaussian processes (GPs). We consider differentiable Markovian control policies and system dynamics given as (i) the mean of a GP, and (ii) the full GP distribution. For both cases, we analyze finite and infinite time horizons. Furthermore, we study the effect of disturbances on the stability results. Empirical evaluations on simulated benchmark problems support our theoretical results.

Keywords: Stability, Reinforcement Learning, Control, Gaussian Process

1. Introduction

Learning control based on Gaussian process (GP) forward models has become a viable approach in the machine learning and control theory communities. Many successful applications impressively demonstrate the efficiency of this approach (Deisenroth et al., 2015; Pan and Theodorou, 2014; Klenske et al., 2013; Maciejowski and Yang, 2013; Nguyen-Tuong and Peters, 2011; Engel et al., 2006; Kocijan et al., 2004). In contrast to classic control theory

methods, learning control does not presuppose a detailed understanding of the underlying dynamics but tries to infer the required information from data. Thus, relatively little expert knowledge about the system dynamics is required and fewer assumptions, such as a parametric form and parameter estimates, must be made. Employing Gaussian processes as forward models for learning control is particularly appealing as they incorporate uncertainty about the system dynamics. GPs infer a distribution over all plausible models given the observed data instead of one (possibly erroneous) model and, thus, avoid severe modeling errors. Furthermore, obtaining a task solution by a learning process can significantly decrease the involved manual effort.

Unfortunately, performance guarantees rarely exist for arbitrary system dynamics and policies learned from data. An important and well-established type of performance guarantee is *(asymptotic) stability*. A *stability region* in the state space ensures, that all trajectories starting in this region converge to the target. Classic control theory offers a rich variety of stability analysis, e.g., for linear, nonlinear, and stochastic systems (Khalil, 2014; Khasminskii and Milstein, 2011; Skogestad and Postlethwaite, 2005; Kushner, 1967). In a preliminary study (Vinogradska et al., 2016), we introduced a tool to analyze the stability of learned policies for closed-loop control systems with transition dynamics given as a GP. This tool handled two types of dynamics: dynamics given as (i) the mean of a GP, and (ii) the full GP distribution. While the first case results in a deterministic closed-loop system, uncertainty is present in the second case. As propagating distributions through a GP is analytically intractable, we proposed a novel approach for approximate multi-step-ahead predictions based on numerical quadrature. Finally, we were able to analyze asymptotic stability of the deterministic dynamics in case (i), but in case (ii) only finite time horizons were considered. Furthermore, the results were limited to GPs with squared exponential kernel.

In this paper, we substantially extend our previous work (Vinogradska et al., 2016) and move to a solid foundation for a stability analysis tool. For this detailed study on stability of closed-loop control systems, we consider dynamics given as (i) the mean of a GP, and (ii) the full GP distribution. In both cases, we analyze finite as well as infinite time horizons. For dynamics given as the mean of a GP, we obtain a region of starting points in the state space such that the target state is reached up to a given tolerance at the (finite) time horizon. For infinite time horizons, we construct a stability region in the state space. Furthermore, we study the behavior of dynamics given by the mean of a GP when disturbances are present. Here, we derive criteria for the disturbance such that the closed-loop control system remains stable. For full GP dynamics, we propose an algorithm to find a region of starting points in the state space such that the probability of the state to be in the target region at the (finite) time horizon is at least a given minimum probability. As the GP predictive distribution at any query point is Gaussian, the probability to transition to a point arbitrarily far from the current point is greater than zero and, thus, the system will eventually leave any compact set. Hence, for full GP dynamics the notions of deterministic analysis do not apply. Here, we extend the previous results (Vinogradska et al., 2016) substantially and analyze asymptotic behavior of closed-loop control systems with full GP dynamics. Employing methods from Markov chain theory, we show that for many choices of the prior the system converges to a unique, invariant limiting distribution.

This paper lays a foundation for theoretical stability analysis for control of probabilistic models learned from data. The proposed approaches provide stability guarantees for many

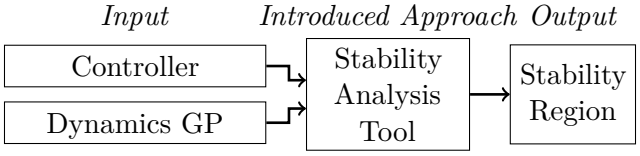


Figure 1: We present a tool to analyze stability for a given controller and Gaussian process (GP) dynamics model. Our tool analytically derives a stability region where convergence to the goal is guaranteed. Thus, we can provide stability guarantees for many of the existing GP based learning control approaches.

existing learning control approaches based on GPs. The paper is organized as follows: first, we briefly review related work and introduce the problem to be addressed. Section 2 provides necessary background for the stability analysis. In Section 3, we analyze closed-loop control systems with dynamics given as the mean of a GP. Subsequently, we study the case when the dynamics are given as the full GP distribution in Section 4. Section 5 provides empirical evaluations on benchmark control tasks. A conclusion is given in Section 6.

1.1 Related Work

To date, only very few special cases of system dynamics and policies learned from data have been analyzed with respect to stability. For example, Perkins and Barto (2003) and Nakanishi et al. (2002) analyze the scenario of an agent switching between given safe controllers. Kim and Ng (2005) monitor stability while learning control for the special case of a linear controller and linear dynamics. Recently, stability properties of autoregressive systems with dynamics given as the mean of a GP were analyzed by Beckers and Hirche (2016), who derive an upper bound on the number of equilibrium points for squared exponential and some special cases of polynomial covariance functions. Furthermore, the authors compute an invariant set and show the boundedness of all system trajectories. However, no control inputs are considered in this work and no asymptotic stability results are given. Furthermore, the full GP distribution is not considered in (Beckers and Hirche, 2016). To the best of our knowledge, stability analysis for closed-loop control with probabilistic models, such as GPs, and arbitrary policies has not been addressed so far.

In classic control theory, the first formal analysis of closed-loop dynamics dates back to the 19th century (Routh, 1877; Hurwitz, 1895; Lyapunov, 1892). Lyapunov's approach allows to analyze stability for nonlinear deterministic systems $\dot{x} = f(x)$. It studies the system behavior around *equilibrium points*, i.e., points, where all derivatives $\dot{x}$ of the state x vanish and the system comes to a hold. An equilibrium point x_e is *stable*, if for every $\varepsilon > 0$ a $\delta > 0$ exists such that $\|x(t) - x_e\| < \epsilon$ for every solution $x(t)$ of the differential equation with $\|x(t_0) - x_e\| < \delta$ and $t \geq t_0$. The equilibrium point is *asymptotically stable*, if it is stable and δ can be chosen such that $\|x(t) - x_e\| \to 0$ as $t \to \infty$, see (Khalil, 2014). One approach to proving stability of x_e is to find a *Lyapunov function*, i.e., a non-negative function, that vanishes only at x_e and decreases along system trajectories. This approach can be applied to nonlinear dynamics. However, finding a Lyapunov function is typically challenging as there exists no general method for this task.

Additionally to the difficulty of finding a suitable Lyapunov function, classical Lyapunov approaches are highly challenging as proving nonnegativity of a nonlinear function is in general intractable. In (Parrilo, 2000), a relaxation of the positive definiteness problem for polynomials was introduced: a polynomial is clearly nonnegative, if it can be written as a sum of squares (SOS) of polynomials. Fortunately, checking whether a polynomial can be written as SOS can be formulated as a semidefinite program (SDP) and can be solved in polynomial time. Thus, for polynomial system dynamics one can search for polynomial Lyapunov functions conveniently by solving the corresponding SDP. This SOS relaxation has been successfully applied to a number of nonlinear stability problems as finding the region of attraction of a system (Chesi, 2004; Majumdar et al., 2014), robust analysis (Topcu et al., 2010b) or controller design (Moore and Tedrake, 2014). While there are some limitations to these methods (Majumdar et al., 2014; Ahmadi and Parrilo, 2011), SOS approaches are of high practical importance. However, SOS methods can only be applied directly to polynomial dynamics. Nonpolynomial dynamics must be recast as polynomial systems by an (automated) procedure (Papachristodoulou and Prajna, 2005). This recasting procedure introduces new variables and constraints for each nonpolynomial term and can, thus, significantly increase the size of the SDP that must be solved. Especially for nonparametric and nonpolynomial models as, e.g., a GP with squared exponential or Matérn kernels, recasting often leads to SDPs that are computationally infeasible (Topcu et al., 2010a).

For consideration of uncertainty in the dynamics, stochastic differential equations (SDEs) have been introduced (Adomian, 1983). SDEs are differential equations where some terms are stochastic processes. In the presence of uncertainty, x_e is *stable in probability*, if for every $\varepsilon > 0$ and $p > 0$, there exists $\delta > 0$ such that $\mathrm{P}\{\|x(t) - x_e\| > \varepsilon\} < p$ for $t > t_0$ and solutions $x(t)$ with $\|x(t_0) - x_e\| < \delta$. Furthermore, x_e is *asymptotically stable in probability*, if it is stable in probability and $\mathrm{P}\{\|x(t) - x_e\| > \varepsilon\} \to 0$ for a suitable δ, see (Khasminskii and Milstein, 2011). Stability follows from the existence of a *supermartingale*, a stochastic Lyapunov function analogue. However, supermartingales exist only if the noise vanishes at the equilibrium point. Relaxing the supermartingale criterion allows for stability statements for a finite time horizon (Steinhardt and Tedrake, 2012; Kushner, 1966). Kushner's (1966) approach requires a supermartingale to be found manually, which is challenging in most cases. Steinhardt and Tedrake (2012) address this difficulty by constructing a supermartingale via SOS programming. This approach is suitable for systems with polynomial dynamics and shares some of the difficulties of SOS approaches as described above.

When controlling a system, uncertainty may arise from modeling inaccuracies, the presence of (external) disturbances and the lack of data, as some system parameters are not known in advance but only during operation. Robustness of nonlinear systems to structured or unstructured uncertainty has been considered in classical control theory, e.g., via the small gain theorem (Zhou and Doyle, 1998). However, these approaches typically lead to an infinite number of nonlinear matrix inequalities and cannot be solved analytically. Another approach is to take uncertainty into account when designing a controller. Thus, a controller must be suitable for a family of systems that is specified, e.g., by bounds for model parameters or for nonparametric uncertainties as bounds for the operator norm of the unknown dynamics. Robust control (Skogestad and Postlethwaite, 2005; Zhou and Doyle, 1998) designs a single controller that is provably stable for all systems in the specified family – often at cost of overall controller performance. Adaptive control (Narendra and Annaswamy, 2012; Tao,

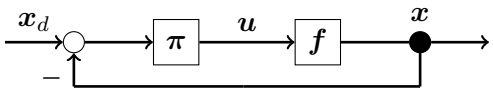


Figure 2: A closed-loop control structure with controller $\boldsymbol{\pi}$ and the system dynamics $\boldsymbol{f}$. We study stability for two types of dynamics: (i) the mean of the GP and (ii) the full GP predictive distribution.

2003) instead adjusts control parameters online in order to achieve prespecified performance, which can be computationally demanding. Both schemes require stability analysis of the dynamics system, e.g., via Lyapunov functions as described above.

1.2 Problem Statement

The goal of this paper is to provide an automated tool to analyze stability of a closed-loop structure when the dynamics model is given as a GP, see Figure 1. As inputs, our tool expects a control policy and a GP dynamics model. It checks the stability of the corresponding closed-loop structure and returns a stability region. Trajectories starting in this region are guaranteed to converge to the target (in probability). Subsequently, we discuss the different components of the tool in more detail.

We consider controllers which depend only on the current state and are differentiable with respect to the state. The dynamics model is given as a GP with stationary covariance function. This covariance function is assumed to have bounded derivatives with respect to all input dimensions and bounded prior mean. We consider a discrete-time system $\boldsymbol{x}^{(t+1)} = \boldsymbol{f}(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)})$ with $\boldsymbol{x}^{(t)} \in \mathbb{R}^D$, $\boldsymbol{u}^{(t)} \in \mathbb{R}^F$, $t = 1, 2, \ldots$ and a controller $\boldsymbol{\pi} \colon \mathbb{R}^D \to \mathbb{R}^F$, whose objective is to move the system to a desired state $\boldsymbol{x}^d$, see Figure 2. In this paper, we study two possible cases for the dynamics $\boldsymbol{f}$: (i) the mean of a GP and (ii) the full GP predictive distribution. Note, that in the second case, distributions have to be propagated through the GP resulting in non-Gaussian state distributions.

2. Preliminaries

In this section, we introduce basic concepts for the proposed stability analysis. First, we briefly review Gaussian process regression, as GPs are employed to describe the considered dynamics. Second, we recap numerical quadrature, as it is the basis for the uncertainty propagation method proposed in Section 4.

2.1 Gaussian Process Regression

Given noisy observations $\mathscr{D} = \{(\boldsymbol{z}^i, y^i = f(\boldsymbol{z}^i) + \varepsilon^i) \mid 1 \leq i \leq N\}$, where $\varepsilon^i \sim \mathcal{N}(0, \sigma_n^2)$, the prior on the values of f is $\mathcal{N}(0, K(Z,Z) + \sigma_n^2 I)$. The covariance matrix $K(Z,Z)$ is defined by the choice of covariance function k as $[K(Z,Z)]_{ij} = k(\boldsymbol{z}^i, \boldsymbol{z}^j)$. One commonly employed covariance function is the squared exponential

$$k(\boldsymbol{z}, \boldsymbol{w}) = \sigma_f^2 \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{w})^\intercal \Lambda^{-1} (\boldsymbol{z} - \boldsymbol{w})\right),$$

with signal variance σ_f^2 and squared lengthscales $\Lambda = \text{diag}(l_1^2, \ldots, l_{D+F}^2)$ for all input dimensions. Given a query point $\boldsymbol{z}_*$, the conditional probability of $f(\boldsymbol{z}_*)$ is

$$f(\boldsymbol{z}_*) \mid \mathscr{D} \sim \mathcal{N}\big(\boldsymbol{k}(\boldsymbol{z}_*, Z)\boldsymbol{\beta},\ k(\boldsymbol{z}_*, \boldsymbol{z}_*) - \boldsymbol{k}(\boldsymbol{z}_*, Z)(K(Z,Z) + \sigma_n^2 I)^{-1}\boldsymbol{k}(Z, \boldsymbol{z}_*)\big) \tag{1}$$

with $\boldsymbol{\beta} = (K(Z,Z) + \sigma_n^2 I)^{-1}\boldsymbol{y}$. The hyperparameters, e.g. $\sigma_n^2, \sigma_f^2, \Lambda$ for the squared exponential kernel, are estimated by maximizing the log marginal likelihood of the data (Rasmussen and Williams, 2005).

In this paper, $\boldsymbol{f}$ models system dynamics. It takes state-action pairs $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{u})^\intercal$ and outputs successor states $\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})$. As these outputs are multivariate, we train conditionally independent GPs for each output dimension. We write $\sigma_{n,m}^2$, $\sigma_{f,m}^2, \Lambda_m$ for the GP hyperparameters in output dimension m and k_m for the corresponding covariance function.

2.2 Numerical Quadrature

Numerical quadrature approximates the value of an integral

$$\int_a^b f(\boldsymbol{x}) d\boldsymbol{x} \approx \sum_{i=1}^{p} w_i f(\boldsymbol{\xi}_i)$$

given a finite number p of function evaluations. A widely used class of quadrature rules are interpolatory quadrature rules, which integrate all polynomials up to a certain degree exactly. In this paper, we employ Gaussian quadrature rules, where the evaluation points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_p$ are chosen to be the roots of certain polynomials from orthogonal polynomial families. They achieve the highest accuracy possible for univariate interpolatory formulæ (Süli and Mayers, 2003). For multivariate integrals, the quadrature problem is significantly harder. While many formulæ for the univariate case can straightforwardly be generalized to multivariate integrals, they often suffer from the curse of dimensionality. However, quadrature methods that scale better and are feasible for up to 20 dimensions have been developed. See (Skrainka and Judd, 2011) for an overview.

3. Stability of GP Mean Dynamics

A wide variety of dynamics can be modeled well by the mean of a GP (Micchelli et al., 2006) due to the universal approximation property of many kernels. Thus, stability of systems with GP mean forward dynamics is an interesting problem. Please note that stability of closed-loop control systems with dynamics given as the mean of a GP is a classical problem from nonlinear control with a particular choice of dynamics function. Classical approaches to solve this problem have been discussed in Section 1.1 and range from direct Lyapunov approaches to the more recent SOS approach to construct polynomial Lyapunov functions. However, for GP mean dynamics, finding a Lyapunov function directly seems even more challenging than for a classical ODE derived from expert knowledge about the physical system. On the other hand, constructing a polynomial Lyapunov function as the solution of an SOS problem is typically computationally infeasible for these dynamics systems, as the nonparametric nature of GP mean functions results in a large number of nonpolynomial terms that significantly increase the complexity of the underlying semidefinite program (see Section 1.1). Thus, we propose an alternative approach to address this problem.

In this section, we provide tools to check the stability of a closed-loop control structure with a given differentiable Markovian control policy π and GP mean dynamics. For this class of forward dynamics, the next state $\boldsymbol{x}^{(t+1)}$ is given as the mean of the GP predictive distribution at the inputs $\boldsymbol{x}^{(t)}, \boldsymbol{u} = \boldsymbol{\pi}(\boldsymbol{x}^{(t)})$. Firstly, we consider infinite time horizons. In this case, we attempt to find a region of starting points in the state space, such that trajectories starting in this region are guaranteed to converge to the desired point $\boldsymbol{x}^d$ as $t \to \infty$. To obtain this result, we analyze the sensitivity of the GP predictive mean to variations in the input space and derive upper bounds for the distance of the predictions at two different points. This analysis can be used to find a region around the target point $\boldsymbol{x}^d$, where the next state is closer to $\boldsymbol{x}^d$ than the point before. It follows straightforwardly, that the full metric ball of maximal radius, which lies completely inside this region, is a stability region. All trajectories that start inside of this metric ball converge to the target point as $t \to \infty$. Furthermore, we will show that in this metric ball, the distance of the current state to the target decreases exponentially. We will exploit this statement to derive some finite time horizon and robustness results. For finite time horizons, we aim to find a region in the state space, such that the state does not deviate more than a given tolerance from the target when the time horizon is reached. The robustness analysis deals with asymptotic stability of the GP mean dynamics when disturbances are present. We derive conditions for these disturbances, such that convergence to the target state $\boldsymbol{x}^d$ is still guaranteed.

In the following, we will briefly review the notion of stability employed for the analysis and derive the main result of this section: an algorithm to find a stability region. After proving correctness of this algorithm, we will derive statements on finite time stability of the GP mean dynamics and on infinite time stability of the disturbed system.

3.1 Stability Notion

If the system dynamics $\boldsymbol{f}$ is given by the mean of a GP, the resulting closed-loop structure is deterministic. To assess the quality of a controller $\boldsymbol{\pi}$, which aims to stabilize the system at the reference point $\boldsymbol{x}^d$, we propose an algorithm to find a stability region of the closed-loop system.

Definition 1 *The reference point $\boldsymbol{x}^d$ is* stable*, if for every $\varepsilon > 0$ there exists $\delta > 0$, such that $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| < \varepsilon$ for $t = 1, 2, \ldots$ and $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| < \delta$. If $\boldsymbol{x}^d$ is stable and there exists a $\delta_0 > 0$ such that $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| \to 0$ for $t \to \infty$, $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| < \delta_0$, $\boldsymbol{x}^d$ is* asymptotically stable*. A subset X^c of the state space is a* stability region*, if $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| \to 0$ as $t \to \infty$ for all $\boldsymbol{x}^{(0)} \in X^c$.*

Please note that $\boldsymbol{\pi}$ is implicit in this definition, as the states $\boldsymbol{x}^{(t)}$ for $t = 1, 2, \ldots$ depend on $\boldsymbol{\pi}$ via $\boldsymbol{x}^{(t+1)} = \boldsymbol{f}(\boldsymbol{x}^{(t)}, \boldsymbol{\pi}(\boldsymbol{x}^{(t)}))$. This definition matches Lyapunov's stability notion (Khalil, 2014). We will present an algorithm that checks asymptotic stability of $\boldsymbol{x}^d$ by constructing a stability region X^c.

When the time horizon considered is finite, we are interested in finding all starting states, such that the distance to the target is at most a given $\varepsilon > 0$.

Definition 2 *A subset X^c of the state space is an* (ε, T)-stability region*, if $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| < \varepsilon$ holds for all $\boldsymbol{x}^0 \in X^c$ and $t \geq T$.*

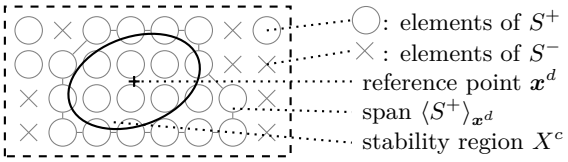


Figure 3: Basic idea of Algorithm 1 to construct a stability region: for a finite set of states S, employ upper bound from Lemma 3 to check whether the successor state is closer to $\boldsymbol{x}^d$, obtaining sets S^+ and S^-. Return metric ball of maximal radius that fits into the span $\langle S^+ \rangle_{\boldsymbol{x}^d}$ as stability region X^c. For $\boldsymbol{x}^{(0)} \in X^c$, the controlled system never leaves X^c and converges to $\boldsymbol{x}^d$ as $t \to \infty$.

In the following, we propose an algorithm to verify asymptotic stability of $\boldsymbol{x}^d$ as in Definition 1 and prove its correctness. Subsequently, we show how to find finite time statements as in Definition 2.

3.2 Algorithm Sketch

In this section, we derive an algorithm, that can find a stability region of a closed-loop control structure with GP mean dynamics. To find a stability region, we analyze how the distance between the current state and the target evolves. The basic idea of the algorithm involves *positively invariant sets*, i.e., sets that once entered, the system will never leave again (Blanchini, 1999). More precisely, we aim to find all $\boldsymbol{x}$, such that

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| < \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| \tag{2}$$

for $\boldsymbol{x}^{(t)} = \boldsymbol{x}$ and a fixed $0 < \gamma < 1$. Assume there is a region containing $\boldsymbol{x}^d$ where Eq. (2) holds. For all states in this region, the distance to the target point decreases in one time step. If it is possible to fit a full metric ball centered at the target point $\boldsymbol{x}^d$ in this region, then all trajectories starting in the ball will never leave it. In addition, the distance of the current state to the reference point $\boldsymbol{x}^d$ will decrease in every step by at least factor γ. Thus, it follows immediately from the existence of such a ball that the target point is asymptotically stable. However, finding a region in closed form where Eq. (2) holds, is usually intractable. Instead, we follow an algorithmic approach to construct a stability region.

The proposed algorithm (Alg. 1) finds a region where Eq. (2) holds if one exists. It employs an upper bound for $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\|$, that depends only on $\boldsymbol{x}^{(t)}$. With this upper bound we can check whether Eq. (2) holds for a (finite) set of points. However, to find a continuous state space region where Eq. (2) holds, this upper bound must allow for generalization from discrete points. More precisely, we need an upper bound of $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\|$ that depends on $\boldsymbol{x}^{(t)}$ smoothly, thus, having bounded gradients, and can be handled conveniently. We employ an upper bound derived from sensitivity analysis of the GP mean to variations in the input space. This bound depends on $\boldsymbol{x}^{(t)}$ in a way that can be exploited to compute a finite set of points S, e.g., a grid that is sufficient to consider as follows: For any grid point, we check if Eq. (2) holds, obtaining the set S^+ with the grid points that fulfill Eq. (2) and the rest in S^-. Let $\langle S^+ \rangle_{\boldsymbol{x}^d}$ be the polygon spanned by the connected component of

Algorithm 1 Stability region X^c for GP mean dynamics

Input: dynamics GP $\boldsymbol{f}$, control policy $\boldsymbol{\pi}$, $\boldsymbol{x}^d$, γ
Output: stability region X^c

```
Construct grid S with Lemma 4, S⁺ ← ∅, X^c ← ∅
for x^(0) ∈ S do
    Compute upper bound C(x^(0), x^d, π) ≥ ‖x^(1) − x^d‖ with Lemma 3
    if C < γ ‖x^(0) − x^d‖ then S⁺ ← S⁺ ∪ {x} fi
od
if ⟨S⁺⟩_{x^d} exists then
    Fit metric ball B(x^d) ⊆ ⟨S⁺⟩_{x^d}
    X^c ← B(x^d) fi
return X^c
```

S^+, which contains $\boldsymbol{x}^d$. We can choose S such that Eq. (2) holds for all points in $\langle S^+ \rangle_{\boldsymbol{x}^d}$, see Figure 3. Algorithm 1 gives an overview, Sec. 3.3 provides technical details and proves correctness of the approach.

3.3 Correctness of the Algorithm

In this section, we elaborate on the computation steps of Algorithm 1 and prove that the returned state space region X^c is a stability region of the closed-loop system in the sense of Definition 1. Fundamental to Algorithm 1 is the upper bound for $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\|$, which can be obtained as follows. We denote $\hat{\boldsymbol{x}} \coloneqq (\boldsymbol{x}, \boldsymbol{\pi}(\boldsymbol{x}))^\intercal$ and recall that $\boldsymbol{f}(\hat{\boldsymbol{x}})$ is the GP predictive mean at $\hat{\boldsymbol{x}}$, see Eq. (1).

Lemma 3 *Let $\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^d \in \mathbb{R}^{D+F}$ and $B \in \mathbb{R}^{D \times D}$ be a positive definite matrix. The distance of GP predictive means $\boldsymbol{f}(\hat{\boldsymbol{x}})$ and $\boldsymbol{f}(\hat{\boldsymbol{x}}^d)$ in the metrics induced by B is bounded by*

$$\|\boldsymbol{f}(\hat{\boldsymbol{x}}) - \boldsymbol{f}(\hat{\boldsymbol{x}}^d)\|_B^2 \leq (\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}^d)^\intercal M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})(\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}^d) =: C(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})$$

with a symmetric matrix M. This matrix can be constructed explicitly and depends on $\boldsymbol{x}, \boldsymbol{x}^d$ and the policy $\boldsymbol{\pi}$.

Proof This statement is obviously true for $\boldsymbol{x} = \boldsymbol{x}^d$, so let $\boldsymbol{x} \neq \boldsymbol{x}^d$. Evaluating

$$\|\boldsymbol{f}(\hat{\boldsymbol{x}}) - \boldsymbol{f}(\hat{\boldsymbol{x}}^d)\|_B^2 = \sum_{m,m'=1}^{D} \sum_{i,k=1}^{N} b_{mm'}(k_m(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^i) - k_m(\hat{\boldsymbol{x}}^d, \hat{\boldsymbol{x}}^i))(k_{m'}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^k) - k_{m'}(\hat{\boldsymbol{x}}^d, \hat{\boldsymbol{x}}^k))\beta_{mi}\beta_{m'k} \quad (3)$$

we realize the need for an upper bound of $k_m(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^i) - k_m(\hat{\boldsymbol{x}}^d, \hat{\boldsymbol{x}}^i)$. Recall that the covariance functions k_m are differentiable with bounded derivatives with respect to $\boldsymbol{x}$ and also, that the control policy π is differentiable with respect to the state $\boldsymbol{x}$. Thus, we can integrate the gradient field of k_m along a curve $\boldsymbol{\tau}$ from $\hat{\boldsymbol{x}}^d$ to $\hat{\boldsymbol{x}}$. As this path integral does not depend on the particular curve $\boldsymbol{\tau}$, we may choose $\boldsymbol{\tau} = \boldsymbol{\tau}^{D+F} \dots \boldsymbol{\tau}^1$ as the curve along the edges of the hypercube defined by $\hat{\boldsymbol{x}}^d$ and $\hat{\boldsymbol{x}}$, i.e., $\boldsymbol{\tau}_p^j(t) = \hat{x}_p$ if $p \leq j - 1$, $\boldsymbol{\tau}_p^j(r) = \hat{x}_p^d + r(\hat{x}_p - \hat{x}_p^d)$ with

$r \in [0;1]$ if $p = j$, and $\boldsymbol{\tau}_p^j(t) = \hat{x}_p^d$ otherwise. This definition yields

$$k_m(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^i) - k_m(\hat{\boldsymbol{x}}^d, \hat{\boldsymbol{x}}^i) = \sum_{j=1}^{D+F} \int_{\hat{\boldsymbol{x}}_j^d}^{\hat{\boldsymbol{x}}_j} \frac{\partial k_m(\boldsymbol{\chi}, \hat{\boldsymbol{x}}^i)}{\partial \chi_j} \Bigg|_{\boldsymbol{\chi}=\boldsymbol{\tau}^j} d\chi_j \tag{4}$$

and we compute the partial derivatives $\partial k_m(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^d)/\partial \hat{\boldsymbol{x}}_j$ in all state-action space dimensions $1 \leq j \leq D + F$.

We rewrite the sum in Eq. (3) by substituting Eq. (4) and the partial derivatives of the covariance functions. To find an upper bound for this sum, we estimate the occurring integrals by the product of integration interval length and an upper or lower mean value according to the sign of the respective summand. The necessary upper and lower mean values can be obtained via Riemannian upper and lower sums or the upper and lower bounds for the partial derivatives of k with respect to $\boldsymbol{x}$. Sorting the summands by products of integration interval lengths $(\hat{x}_j - \hat{x}_j^d)(\hat{x}_p - \hat{x}_j^d)$, this sum can be rewritten as a quadratic form. The entries of M can be chosen to form a symmetric matrix by making $M_{pj} = M_{jp}$ half of the coefficient of $(\hat{x}_j - \hat{x}_j^d)(\hat{x}_p - \hat{x}_p^d)$. ■

For any point in the state action space, we can find an upper bound for the distance of its prediction and the target point. Note that this distance estimated by Lemma 3 depends heavily on the eigenvalues of $M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})$. This fact is exploited to compute a grid S, which constitutes the first step of Algorithm 1. As M is constructed as a symmetric matrix, the eigenvalue problem is well conditioned, i.e., when M is perturbed, the change in the eigenvalues of M is at most as large as the perturbation.

Lemma 4 *Let $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^D$ and $M(\boldsymbol{x}, \boldsymbol{x}^d, \pi)$, B be as defined in Lemma 3. If $\|\boldsymbol{f}(\hat{\boldsymbol{x}}) - \boldsymbol{f}(\hat{\boldsymbol{x}}^d)\|_B < c\|\boldsymbol{x} - \boldsymbol{x}^d\|$ for $c < 1$, there exist Δ_j such that $\|\boldsymbol{f}(\hat{\boldsymbol{z}}) - \boldsymbol{f}(\hat{\boldsymbol{x}}^d)\|_B < c\|\boldsymbol{z} - \boldsymbol{x}^d\|$, for all $\boldsymbol{z}$ with $|\hat{z}_j - \hat{x}_j| < \Delta_j$, $1 \leq j \leq D + F$.*

Proof Solving for the eigenvalues of $M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})$, the set $Q_{\boldsymbol{x}} \coloneqq \{\boldsymbol{v} \in \mathbb{R}^{D+F} \mid (\boldsymbol{v} - \hat{\boldsymbol{x}}^d)^\intercal M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})(\boldsymbol{v} - \hat{\boldsymbol{x}}^d) < a\}$ can be determined. It is symmetric to the axes defined by the eigenvectors of $M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})$ and meets them at $\pm a^{-1}\sqrt{\lambda}$. For $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{\Delta}$ we want to ensure $\boldsymbol{z} \in Q_{\boldsymbol{z}}$, if $\boldsymbol{x} \in Q_{\boldsymbol{x}}$. Thus, we estimate how much the eigenvalues of $M(\boldsymbol{z}, \boldsymbol{x}^d, \boldsymbol{\pi})$ differ from those of $M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})$. As $M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})$ is symmetric, the eigenvalue problem has condition $\kappa(\lambda, M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})) = 1$ for any eigenvalue λ. Thus, $|\partial\lambda/\partial\hat{x}_j| \leq \|\partial M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})/\partial\hat{x}_j\|$. Computing $\|\partial M(\boldsymbol{x}, \boldsymbol{x}^d, \boldsymbol{\pi})/\partial\hat{x}_j\|$ or upper bounds for this expression allows solving for all Δ_j. ■

We are now able to compute a grid S, such that it is sufficient to check Eq. (2) for all grid points to retrieve a (continuous) stability region. While the grid width may become small, there is a lower bound for it. As the GP falls back to the prior far away from training data, the entries of M are bounded for bounded mean priors and, being continuous, Lipschitz. Thus, a lower bound exists.

Theorem 5 *The region X^c returned by Algorithm 1 is a stability region. All trajectories starting in X^c move closer to the desired point $\boldsymbol{x}^d$ in each step. Convergence to $\boldsymbol{x}^d$ is guaranteed for all points in X^c as $t \to \infty$.*

Proof We exploit Lemma 4 to compute a grid S as the first step in Algorithm 1. Employing Lemma 3, the algorithm checks for all points in S whether Eq. (2) holds, obtaining the sets S^+ and S^-, respectively. Lemma 4 ensures that Eq. (2) also holds for all points inside the polygon $\langle S^+ \rangle_{\boldsymbol{x}^d}$. For every point in $\langle S^+ \rangle_{\boldsymbol{x}^d}$, the next state is closer to the target point $\boldsymbol{x}^d$. Fitting a full metric ball $\mathcal{B}(\boldsymbol{x}^d)$ in $\langle S^+ \rangle_{\boldsymbol{x}^d}$, we recover an invariant set X^c. More precisely, the trajectories starting in X^c move closer to $\boldsymbol{x}^d$ with every time step. ■

Theorem 5 provides an asymptotic stability result for closed-loop control systems with dynamics given as the mean of a GP and can, thus, can be applied to a broad class of dynamics systems. However, the procedure is based on a grid in the state space and as a consequence suffers from the curse of dimensionality. This limitation is not surprising, as it is well known that estimating stability regions of high degree is NP-hard (Ahmadi et al., 2013). However, several approaches can help to scale the proposed method to higher dimensions. Instead of computing a global grid width $\boldsymbol{\Delta}$ as in Lemma 4, the proof can straightforwardly be adapted to compute local grid widths. Such local grid widths depend on the magnitude of the derivatives of the GP mean with respect to the inputs and, thus, can be substantially larger than the global one, e.g., when moving away from training data or where the dynamics is very slow. With this technique, the number of grid points can be greatly reduced. Furthermore, as we are constructing an inner approximation to the full stability region, the order of the grid points processed by Algorithm 1 can be modified to avoid redundant computations. When starting at the equilibrium point and expanding the considered grid points spherically according to the chosen distance metrics, computation can be stopped as soon as the first grid point is tested to lie outside the stability region. This first negative test point limits the stability region and all grid points with the same or greater distance to $\boldsymbol{x}^d$ need not be considered anymore.

3.4 Finite Time Horizons and Robustness

In this section, we will employ the proposed Algorithm 1 to derive results for finite time horizons and stability in the presence of disturbances.

3.4.1 Finite Time GP Mean Stability

Above, we proposed an algorithm to check asymptotic stability of the target $\boldsymbol{x}^d$ and to find a stability region, if one exists. In this section, we are interested in finding all starting states which deviate from the target by at most a given tolerance $\varepsilon > 0$ after $T < \infty$ timesteps. State space regions that fulfill this criterion were introduced in Definition 2 as (ε, T)-*stability regions*.

The main result of this section follows from the infinite time horizon results of Sec. 3.3. Recall that Algorithm 1 computes a state space region X^c and $\gamma < 1$, such that

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| < \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\|$$

for all $\boldsymbol{x}^{(t)} \in X^c$. We can exploit this result to find a state space region $X^c_{\varepsilon,T}$, such that $\|\boldsymbol{x}^{(T)} - \boldsymbol{x}^d\| < \varepsilon$ for all $\boldsymbol{x}^{(0)} \in X^c_{\varepsilon,T}$.

Lemma 6 *Let $\varepsilon > 0$, $T < \infty$ and $X^c \subseteq X, \gamma < 1$ such that $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| < \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\|$ for all $t = 1, 2, \ldots$ and $\boldsymbol{x}^{(0)} \in X^c$. Then $\|\boldsymbol{x}^{(T)} - \boldsymbol{x}^d\| < \varepsilon$ for all $\boldsymbol{x}^{(0)} \in X^c$ with $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| < \varepsilon / \gamma^T$.*

Proof Choosing $\boldsymbol{x}^{(0)} \in X^c$ and iteratively employing $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| < \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\|$ we get

$$\|\boldsymbol{x}^{(T)} - \boldsymbol{x}^d\| < \gamma^T \|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| \overset{!}{<} \varepsilon.$$

Solving for ε, we obtain $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| < \varepsilon/\gamma^T$ and, thus, $X^c_{\varepsilon,T} \coloneqq \mathcal{B}_{\varepsilon/\gamma^T}(\boldsymbol{x}^d)$ is an (ε, T)-stability region. ■

Lemma 6 enables finite time horizon statements. For a given tolerance ε we can compute a region $X^c_{\varepsilon,T}$ such that $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| < \varepsilon$ for all $t \geq T$ if $\boldsymbol{x}^{(0)} \in X^c_{\varepsilon,T}$. Such stability statements are relevant for many applications as controllers are often designed for finite system runtimes.

3.4.2 Robustness of GP Mean Dynamics

In this section, we will analyze the behavior of closed-loop control systems with dynamics given as the mean of a GP when disturbances are present. More precisely, we assume that the next state is given as

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{f}\left(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\right) + \boldsymbol{r}^{(t)} \tag{5}$$

with the disturbance $\boldsymbol{r}^{(t)} < \infty$ at time t, that does not depend on the system state. We aim to find conditions for $\boldsymbol{r}^{(t)} < \infty$ such that $\boldsymbol{x}^{(t)}$ still converges to $\boldsymbol{x}^d$ as $t \to \infty$. Let $R^{(t)} \coloneqq \|\boldsymbol{r}^{(t)}\|$ be the magnitude of the disturbance at time t. Based on Equation (2) we derive an upper bound for $R^{(t)}$. Recall that Algorithm 1 returns a stability region X^c, which is a full metric ball and let ρ be the radius of this ball. Convergence to the target state is guaranteed, as long as the system does not leave X^c and $R^{(t)} \to 0$ as $t \to \infty$.

Theorem 7 *Let X^c be a full metric ball of radius ρ that is a stability region of the undisturbed closed-loop control system with GP mean dynamics. For $\boldsymbol{x}^{(0)} \in X^c$, the system in Eq. (5) converges to the target state $\boldsymbol{x}^d$, if*

$$R^{(t)} \leq \rho - \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| \tag{6}$$

for all t and there exists some T_0 such that

$$R^{(t)} < (c - \gamma) \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| \tag{7}$$

with a constant $c < 1$ holds for all $t \geq T_0$.

Proof Firstly, we will show that $\boldsymbol{x}^{(t+1)} \in X^c$, if $\boldsymbol{x}^{(t)} \in X^c$ and Eq. (6) holds. Employing Equation (5), we get

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| = \|\boldsymbol{f}(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}) + \boldsymbol{r}^{(t)} - \boldsymbol{x}^d\| \leq \|\boldsymbol{f}(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}) - \boldsymbol{x}^d\| + R^{(t)} \tag{8}$$

and assuming $\boldsymbol{x}^{(t)} \in X^c$, it follows

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| \leq \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| + R^{(t)}. \tag{9}$$

It follows $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| \leq \rho$, if $R^{(t)} \leq \rho - \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\|$ and, thus, $\boldsymbol{x}^{(t+1)} \in X^c$. By induction, $\boldsymbol{x}^{(t)} \in X^c$ for all $t > 0$, if $\boldsymbol{x}^{(0)} \in X^c$ and Equation (6) holds for all t.

If the disturbance $\boldsymbol{r}^{(t)}$ vanishes after a finite number of time steps, it is sufficient to ensure Equation (6) to show convergence to the target state. After this finite number of

disturbed steps, the system state will still lie in X^c and for all time steps to come, the dynamics will resemble the non-disturbed dynamics. Convergence is then guaranteed by Theorem 5.

Next, we will show that $\boldsymbol{x}^{(t)}$ converges to $\boldsymbol{x}^d$, if the disturbance $\boldsymbol{r}^{(t)}$ decreases sufficiently fast, as stated in Equation (7). For $\boldsymbol{x}^{(t)} \in X^c$, Equation (9) holds. To ensure convergence to $\boldsymbol{x}^d$, we aim to decrease the distance to the target state with every time step. Thus,

$$\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| \leq \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| + R^{(t)} \overset{!}{<} c \, \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| \tag{10}$$

for a constant $c < 1$. Solving for $R^{(t)}$, we obtain Eq. (7). ■

This theorem provides a criterion for the disturbance such that the closed-loop control system will still converge to the target state. However, in practice, Equations (6) and (7) may be difficult to verify, as they involve $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\|$ for all t. We derive a closed-form upper bound for the magnitude $R^{(t)}$ of the disturbance $\boldsymbol{r}^{(t)}$.

Lemma 8 *The Inequality (6) holds for all t, if*

$$R^{(t)} \leq \rho - \left(\gamma^t \|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| + \sum_{k=1}^{t-1} \gamma^k R^{(k)} \right). \tag{11}$$

The inequality (7) holds for $t \geq T_0$, if

$$R^{(t)} \leq (c^{t+1} - \gamma c^t) \|\boldsymbol{x}^{T_0} - \boldsymbol{x}^d\|. \tag{12}$$

Proof Both inequalities can be obtained by iteratively applying Eqs. (6) and (7).

$$\begin{aligned}
\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| &\leq \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| + R^{(t)} \\
&\leq \gamma \left(\gamma \|\boldsymbol{x}^{(t-1)} - \boldsymbol{x}^d\| + R^{(t-1)} \right) + R^{(t)} \\
&\leq \gamma^t \|\boldsymbol{x}^{(0)} - \boldsymbol{x}^d\| + \sum_{k=1}^{t-1} \gamma^k R^{(k)} + R^{(t)} \overset{!}{<} \rho
\end{aligned}$$

and analogously

$$\begin{aligned}
\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^d\| &\leq \gamma \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^d\| + R^{(t)} \\
&< \gamma (c \|\boldsymbol{x}^{(t-1)} - \boldsymbol{x}^d\|) + R^{(t)} \\
&< \gamma c^t \|\boldsymbol{x}^{(T_0)} - \boldsymbol{x}^d\| + R^{(t)} \overset{!}{<} c^{t+1} \|\boldsymbol{x}^{(T_0)} - \boldsymbol{x}^d\|.
\end{aligned}$$

■

Employing Lemma 8, we can analyze stability for disturbed system trajectories, if the magnitude of the disturbance (or an upper bound for the magnitude) is known for every time step.

4. Stochastic Stability of GP Dynamics

In this section, we study closed-loop systems with dynamics given as a full GP distribution. For any query point $\boldsymbol{x}^{(t)}$ a GP predicts the next state $\boldsymbol{x}^{(t+1)}$ to be normally distributed. If, however, $\boldsymbol{x}^{(t)}$ is not a point, but a distribution, the integral

$$p\left(\boldsymbol{x}^{(t+1)}\right) = \int_{\mathbb{R}^D} p\left(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)}\right) p\left(\boldsymbol{x}^{(t)}\right) d\boldsymbol{x}^{(t)} \tag{13}$$

determines the next state distribution. Note that $p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)})$ is Gaussian with respect to $\boldsymbol{x}^{(t+1)}$. The next state distribution $p(\boldsymbol{x}^{(t+1)})$, however, is not Gaussian, even if $p(\boldsymbol{x}^{(t)})$ is. Generally, $p(\boldsymbol{x}^{(t+1)})$ is analytically intractable and only approximations, e.g., via moment matching (Quiñonero-Candela et al., 2003) or linearization (Ko and Fox, 2008), can be computed. These methods suffer from severe inaccuracies in many cases, e.g., they cannot handle distributions with multiple modes. Also, no pointwise approximation error bounds are available. Thus, these methods are unsuitable for stability analysis. In this paper, we propose numerical quadrature (Sec. 4.2.1) to approximate $p(\boldsymbol{x}^{(t+1)})$, instead. This technique yields significantly better results, e.g., it can handle distributions with multiple modes. In addition, error analysis is readily available (Wasowicz, 2006; Masjed-Jamei, 2014) and can be employed to derive stability guarantees for a finite time horizon.

We will also analyze the closed-loop system behaviour for infinite time horizons. The proposed numerical quadrature approximation converges to a stationary limiting distribution. Unfortunately, the error bounds based on quadrature error analysis do not apply when infinitely many timesteps are taken. However, we will show that closed-loop control systems with GP dynamics expose the same infinite time horizon behavior – they converge to a stationary distribution.

In the following, we will discuss *finite time stochastic stability*, introduce an algorithm to find a stability region based on numerical quadrature, and prove its correctness. Subsequently, we will elaborate on the construction of good quadrature rules for multi-step-ahead predictions. Finally, we will analyze the system behavior for infinite time horizons.

4.1 Stability Notion

Consider a deterministic system, that is locally, but not globally, asymptotically stable. Adding noise to the system may render the target point unstable. Especially when noise is unbounded (e.g., Gaussian), all trajectories will eventually leave any ball around the target point with probability one. For this reason, in the study of SDEs, other stability notions than Lyapunov's, are common. In particular, bounding the probability to drift away from the target over a finite time T is desirable, as in the following definition (Kushner, 1966).

Definition 9 *Let Q_1, Q_2 be subsets of the state space X, with Q_2 open and $Q_1 \subset Q_2 \subset X$. The system is* finite time stable *with respect to $Q_1, Q_2, 1-\lambda, T$, if $\boldsymbol{x}^{(0)} \in Q_1$ implies $P\{\boldsymbol{x}^{(t)} \in Q_2\} \geq 1-\lambda$ for all $t \leq T$.*

However, we are interested in finding a set Q_s of initial conditions, such that the goal Q is reached within time T with a desired probability (cf. Steinhardt and Tedrake 2012).

Definition 10 *The set Q_s is a stability region with respect to the target region Q, time horizon T and success probability $1-\lambda$, if $P\{\boldsymbol{x}^{(T)} \in Q\} \geq 1-\lambda$ holds for all $\boldsymbol{x}^{(0)} \in Q_s$ with $\lambda > 0$ and the target region $Q \subset X$.*

Algorithm 2 Stability region for GP dynamics

Input: dynamics GP $\boldsymbol{f}$, control policy $\boldsymbol{\pi}$, time horizon T, target region Q, approximation error tolerance e_{tol}, desired success probability $1-\lambda$

Output: stability region X^c

Construct grid S with Lemma 12
$S^+ \leftarrow \emptyset$
Compute quadrature $\boldsymbol{w}, \boldsymbol{\xi}$ based on Eq. (18) and Lemma 11, such that $\|\epsilon_T\|_{\mathcal{C}(X)} < e_{\text{tol}}$
$\boldsymbol{\phi} \leftarrow (\boldsymbol{f}(\boldsymbol{\xi}_1, \boldsymbol{\pi}(\boldsymbol{\xi}_1)), \ldots, \boldsymbol{f}(\boldsymbol{\xi}_N, \boldsymbol{\pi}(\boldsymbol{\xi}_N)))^\intercal$
$\boldsymbol{m} \leftarrow (\int_Q \phi_1(\boldsymbol{x})d\boldsymbol{x}, \ldots, \int_Q \phi_N(\boldsymbol{x})d\boldsymbol{x})^\intercal$
for $\boldsymbol{x} \in S$ **do**
  $p(\boldsymbol{x}^{(1)}) \leftarrow \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\pi}(\boldsymbol{x}))$
  $\boldsymbol{x}^{(T)} \leftarrow \boldsymbol{\alpha}^{(T)}\boldsymbol{\phi}$
  $P_{min}\{\boldsymbol{x}^{(T)} \in Q\} \leftarrow \boldsymbol{\alpha}^{(T)}\boldsymbol{m} - \text{vol}(Q)\|\epsilon_T\|_{\mathcal{C}(X)}$
  if $P_{min}\{\boldsymbol{x}^{(T)} \in Q\} > 1-\lambda$ **then** $S^+ \leftarrow S^+ \sqcup \{\boldsymbol{x}\}$ **fi**
od
return $X^c \leftarrow \langle S^+ \rangle$

This definition focuses on reaching a target Q after time T, whereas Definition 9 bounds the exit probability from a region Q_2 within time $0 \leq t \leq T$. The methods proposed in this paper can be employed to analyze stability in the sense of both definitions. In the following, we will present an algorithm to find a stability region according to Definition 10.

4.2 Algorithm Sketch

We will now discuss how to find a stability region as in Definition 10 for a closed-loop system with GP dynamics. To analyze system behavior, the capability to compute next state distributions is crucial. As discussed previously, Eq. (13) is analytically intractable and approximation methods must be employed. We propose numerical quadrature to approximately propagate distributions. We will show that numerical quadrature approximates state distributions as Gaussian mixture models. Fortunately, computation of multi-step-ahead predictions becomes convenient and fast even for long time horizons. However, relying on approximations of the state distribution is not sufficient for stability guarantees. The error introduced by approximation and error propagation must be bounded to recover reliable statements on the probability for $\boldsymbol{x}^{(T)}$ to be in a set Q. This error bound $\epsilon_t(\boldsymbol{x})$ at time t can be obtained following one of the available quadrature error analyses, e.g., by Masjed-Jamei (2014).

With numerical quadrature and quadrature error analysis, we can compute a lower bound for the *success probability*, i.e., the probability for $\boldsymbol{x}^{(T)}$ to be in the target set Q. A priori, this enables checking success probabilities for a finite set of points. Fortunately, as in the case of GP mean dynamics, the statement can be generalized to continuous regions. Our algorithm constructs a grid in the state space and computes success probabilities for all grid points. We prove that a stability region can be inferred from these grid point results.

In the following, we elaborate on numerical quadrature as approximate inference method and, studying quadrature error propagation, prove the correctness of the proposed approach. The presented tool is outlined in Algorithm 2.

4.2.1 Numerical Quadrature Uncertainty Propagation

In most applications, especially when the states have physical interpretations, the state space is bounded. For this reason, we assume $X = [a_1, b_1] \times \cdots \times [a_D, b_D]$ and solve

$$F[p(\boldsymbol{x}^{(t)})] \coloneqq \int_X p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)}) p(\boldsymbol{x}^{(t)}) d\boldsymbol{x}^{(t)}. \tag{14}$$

We propose numerical quadrature to approximate this integral. We choose a composed Gaussian product quadrature rule, which will be detailed in Section 4.4. For the rest of this section, it is sufficient to note that our quadrature rule provides a set of evaluation points $\mathbb{X}$ and positive weights w_n for all nodes $\boldsymbol{\xi}^n \in \mathbb{X}$. Integral (14) is then approximated by

$$F[p(\boldsymbol{x}^{(t)})] \approx \sum_{\boldsymbol{\xi}^n \in \mathbb{X}} w_n p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)} = \boldsymbol{\xi}^n) p(\boldsymbol{x}^{(t)} = \boldsymbol{\xi}^n), \tag{15}$$

resulting in a weighted sum of Gaussian distributions. The approximate state distribution at time $t+1$ can be given by

$$p(\boldsymbol{x}^{(t+1)}) \approx \boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(t+1)} \tag{16}$$

with $\alpha_n^{(t+1)} \coloneqq w_n p(\boldsymbol{x}^{(t)} = \boldsymbol{\xi}^n)$, $\phi_n(\boldsymbol{x}) \coloneqq p(\boldsymbol{x}^{(t+1)} = \boldsymbol{x} \mid \boldsymbol{x}^{(t)} = \boldsymbol{\xi}^n)$. Note that the Gaussian basis functions $\phi_n(\boldsymbol{x})$ do not change over time, so the state distribution at time t is represented by the weight vector $\boldsymbol{\alpha}^{(t)}$. To propagate any distribution multiple steps through the GP, the basis functions ϕ_n must be calculated only once and the task reduces to sequential updates of the weight vector $\boldsymbol{\alpha}$. As $p(\boldsymbol{x}^{(t)}) \approx \boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(t)}$, the weight vector $\boldsymbol{\alpha}^{(t+1)}$ is given by

$$\boldsymbol{\alpha}^{(t+1)} = \operatorname{diag}(\boldsymbol{w}) \Phi \boldsymbol{\alpha}^{(t)} = (\operatorname{diag}(\boldsymbol{w}) \Phi)^t \boldsymbol{\alpha}^{(1)} \tag{17}$$

with the matrix Φ, $\Phi_{ij} = \phi_j(\boldsymbol{\xi}^i)$ with $1 \leq i, j \leq n$, which contains the basis function values at all grid points. In practice, it is helpful to normalize the matrix $\operatorname{diag}(\boldsymbol{w})\Phi$ such that each column sums to 1. In this case, unit vectors will be mapped to unit vectors. This ensures that our approximate state distribution is in fact a probability density, i.e., integrates to 1.

Our algorithm aims to find a stability region Q_s, i.e., a region where the success probability is at least $1 - \lambda$ for a given time horizon T, target region Q, and $\lambda > 0$. Computing $\boldsymbol{m} = \int_Q \boldsymbol{\phi}(\boldsymbol{x}) d\boldsymbol{x}$, the probability for $\boldsymbol{x}^{(T)}$ to be in Q is approximately $\boldsymbol{m}^\intercal \boldsymbol{\alpha}^{(T)}$.

4.3 Correctness of the Algorithm

We will now show that the region returned by Algorithm 2 is a stability region as in Definition 10. In Section 4.2.1, we proposed numerical quadrature to approximate GP predictions when the input is a distribution. However, to obtain stability guarantees it is not sufficient to consider approximate solutions. For a lower bound on the success probability, additional knowledge about the approximation error and how it is propagated through the dynamics GP is essential.

When approximate inference steps are cascaded for multi-step-ahead predictions, errors are introduced and propagated through the dynamics. Typically, error propagation fortifies initial errors with every time step. The difference of approximation and true distribution after a finite time is bounded. Fortunately, numerical quadrature allows to control this error bound by refining the used quadrature rule.

The derivation of an upper bound for the approximation error after a finite number of time steps relies heavily on quadrature error analysis, i.e., upper bounds for the errors that are made in one time step when numerical quadrature is employed. Several error analyses for numerical quadrature are available (Masjed-Jamei, 2014; Davis et al., 2014; Wasowicz, 2006; Evans, 1994) with the classic analyses relying on higher order derivatives of the function to be integrated. We will employ the error analysis by Masjed-Jamei (2014) that requires first order derivatives only. More precisely (Masjed-Jamei 2014, Theorem 2.1), for a differentiable function f with $\nu_1(x) \leq f'(x) \leq \nu_2(x)$ with continuous functions ν_1, ν_2 the quadrature error is upper and lower bounded

$$m \leq \sum_{i=1}^{N} w_i f(\xi_i) - \int_a^b f(x)dx \leq M \tag{18}$$

with the bounds

$$m = \sum_{i=0}^{N} \int_{\xi_i}^{\xi_{i+1}} \left(\nu_1(t) \frac{v_i(t) + |v_i(t)|}{2} + \nu_2(t) \frac{v_i(t) - |v_i(t)|}{2} \right) dt$$

$$M = \sum_{i=0}^{N} \int_{\xi_i}^{\xi_{i+1}} \left(\nu_1(t) \frac{v_i(t) - |v_i(t)|}{2} + \nu_2(t) \frac{v_i(t) + |v_i(t)|}{2} \right) dt$$

for $\xi_0 = a$, $\xi_{N+1} = b$ and $v_i = t - \left(a + \sum_{j=1}^{i} w_i\right)$, $i = 0, \ldots, n$. In the following, we employ this result to derive an error bound for the pointwise approximation error of the proposed multi-step-ahead predictions based on numerical quadrature.

Lemma 11 *Let $\boldsymbol{x}^{(0)} \in X$ and ϵ_T be the pointwise approximation error $\epsilon_T(\boldsymbol{x}) = p(\boldsymbol{x}^{(T)} = \boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x})^\intercal \boldsymbol{\alpha}^{(T)}$ at time T. There exists an upper bound for $\|\epsilon_T\|_{\mathcal{C}(X)}$ that can be controlled by the choice of quadrature rule.*

Proof The statement obviously holds for $T = 0$ and for $T = 1$, as $p(\boldsymbol{x}^{(1)})$ is Gaussian and computed exactly. If $T \geq 2$, an error is introduced as $p(\boldsymbol{x}^{(t)})$ cannot be computed analytically for $t \geq 2$. Employing numerical quadrature to compute $p(\boldsymbol{x}^{(2)})$, we obtain $p(\boldsymbol{x}^{(2)}) = \boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(2)} + \epsilon_2$ with our approximation $\boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(2)}$ and the unknown initial quadrature error $\epsilon_2(\boldsymbol{x})$. However, the error analysis by Masjed-Jamei (2014, Theorem 2.1) gives us a bound for $\|\epsilon_2\|_{\mathcal{C}(X)} := \max_{\boldsymbol{x} \in X} |\epsilon_2(\boldsymbol{x})|$. When propagating further, an unknown, bounded error term and an approximation term must be handled. For the approximation term, we get $F[\boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(t)}] = \boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(t+1)} + \boldsymbol{\epsilon}^\intercal \boldsymbol{\alpha}^{(t)}$ with quadrature error bounds $\boldsymbol{\epsilon} = (\epsilon_i(\boldsymbol{x}))_i$ for the integrals (14), when setting $p(\boldsymbol{x}^{(t)}) = \phi_i$. The error term $\epsilon_t(\boldsymbol{x})$ is propagated to $F[\epsilon_t]$ with $\max_{\boldsymbol{x} \in X} |F[\epsilon_t]| \leq C \|\epsilon_t\|_{\mathcal{C}(X)}$ and $C = \max_{\boldsymbol{x} \in X} \int p(\boldsymbol{x}^{(t+1)} \,|\, \boldsymbol{x}^{(t)}) d\boldsymbol{x}^{(t)}$. Finally, the error at time T is bounded by $\|\epsilon_T\|_{\mathcal{C}(X)} \leq \|\boldsymbol{\epsilon}\|^\intercal \boldsymbol{\alpha}^{(T-1)} + \sum_{t=2}^{T-2} C^t \|\boldsymbol{\epsilon}\|^\intercal \boldsymbol{\alpha}^{(t)} + C^{t-1} \|\epsilon_2\|$, where $\boldsymbol{\epsilon}$ and ϵ_2 can

be made arbitrarily small by refining the quadrature rule. ■

Lemma 11 theoretically enables stability guarantees for finite time horizons T. However, as with any approximation of the distribution (14), this error bound grows exponentially with T, while quadrature error decreases polynomially with function evaluations. This drawback limits real-word application of Lemma 11 to small T, but we found that typically, approximation behaves far better than the worst-case bound.

For any starting point, we can now compute the approximate state distribution at time T and a lower bound on the success probability, i.e., the probability for $\boldsymbol{x}^{(T)}$ to be in the target region Q. It remains to show that it is sufficient to compute the success probability for a discrete set of starting states and generalize to the underlying continuous state space region.

Lemma 12 *Let $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^D$ be starting states and $T \in \mathbb{N}$. If $P\{\boldsymbol{x}^{(T)} \in Q\} \geq 1 - \lambda$, there exist Δ_j such that $P\{\boldsymbol{z}^{(T)} \in Q\} \geq 1 - \lambda$ for all $\boldsymbol{z}$ with $|z_j - x_j| < \Delta_j$, $1 \leq j \leq D$.*

Proof Firstly, we note that the absolute error bound does not depend on the starting state. The state distribution $p(\boldsymbol{x}^{(1)})$ is the GP prediction at the query point $\boldsymbol{x}^{(0)}$. Thus, for $T = 1$ the claim follows from Lemma 4. When $t \geq 2$, only an approximation of the state distribution is known. The mixture model weights at time $t = 2$ depend only on the values of $p(\boldsymbol{x}^{(1)})$ at the set of evaluation points $\mathbb{X}$. Thus, the lemma follows for $T = 2$. If $t > 2$, the difference of the approximate distributions for the starting points $\boldsymbol{x}$ and $\boldsymbol{z}$ is

$$\boldsymbol{\phi}^\intercal(\boldsymbol{\alpha}_{\boldsymbol{x}}^{(t)} - \boldsymbol{\alpha}_{\boldsymbol{z}}^{(t)}) = \boldsymbol{\phi}^\intercal(\operatorname{diag}(\boldsymbol{w})\Phi)^{t-2}(\boldsymbol{\alpha}_{\boldsymbol{x}}^{(2)} - \boldsymbol{\alpha}_{\boldsymbol{z}}^{(2)}).$$

Thus, the difference in approximate probability mass is linear in $(\boldsymbol{\alpha}_{\boldsymbol{x}}^{(2)} - \boldsymbol{\alpha}_{\boldsymbol{z}}^{(2)})$. This fact concludes the proof. ■

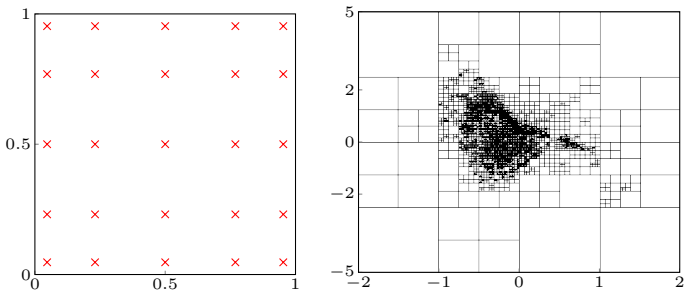
Finally, we are able to prove the main result of this section, the correctness of Algorithm 2.

Theorem 13 *The region X^c returned by Algorithm 2 is a stability region in the sense of Definition 10. For all starting points $\boldsymbol{x}^{(0)} \in X^c$, the probability mass in the target region Q at time T is at least $1 - \lambda$.*

Proof For the first step in Algorithm 2 we exploit Lemma 12 to construct a grid S. For S, stability follows for the region around the grid points S^+ with success probability greater than $1-\lambda$. Secondly, employing Lemma 11 we determine a quadrature rule that ensures the pointwise error at time T to be at most e_{tol}. Finally, we compute approximate success probabilities for all grid points, as in Sec. 4.2.1. Subtracting the maximum error mass $e_{\text{tol}} \operatorname{vol}(Q)$ in target region Q, we obtain all grid points S^+, which have a success probability of at least $1-\lambda$. ■

4.4 Construction of Numerical Quadratures for Uncertainty Propagation

In the sections above, we introduced numerical quadrature to approximate GP predictions at uncertain inputs and showed that this approximate inference method can be used to derive finite time horizon stability guarantees. The presented theoretical results are valid for any choice of quadrature rule that allows for quadrature error analysis. However, in practice, the



(a) Product Quadrature (b) Mountain Car

Figure 4: Construction of suitable quadrature rules. Plot (a) shows the nodes of a Gaussian product quadrature rule on the unit square. Here, the quadrature rule for the unit square was constructed as an outer product of the Gaussian quadrature rule with 5 nodes. The state space partition for the mountain car system obtained with Algorithm 3 is shown in (b). For each rectangle in this partition, a Gaussian product quadrature such as (a) is employed.

choice of a suitable quadrature rule is crucial to the applicability of the proposed algorithm. Thus, in the following, we will elaborate on how to find "good" quadrature rules for a given GP dynamics and control policy.

Gaussian product quadrature extends univariate Gaussian quadrature to a multivariate rule using a product grid of evaluation points. This construction has some desirable properties, such as positive quadrature weights and readily available quadrature nodes. However, being an outer product of one-dimensional rules, it suffers from the curse of dimensionality. In this paper, we apply numerical quadrature to integrals as in Equation (14). Typically, the system trajectories are not uniformly spread over the state space. Instead, they are concentrated in a significantly smaller region. We exploit this observation to improve the efficiency of product grid quadrature and cope with the curse of dimensionality. For this purpose, we partition the state space $X = X_1 \sqcup \cdots \sqcup X_L$ and apply a Gaussian product quadrature to each obtained subregion X_l. The next state distribution, cf. Equation (14), can be written as

$$F[p(\boldsymbol{x}^{(t)})] \coloneqq \int_X p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)}) p(\boldsymbol{x}^{(t)}) d\boldsymbol{x}^{(t)} = \sum_{l=1}^{L} \int_{X_l} p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)}) p(\boldsymbol{x}^{(t)}) d\boldsymbol{x}^{(t)}$$

and, applying a numerical quadrature rule with nodes $\mathbb{X}_l$ for each integral, we get

$$F[p(\boldsymbol{x}^{(t)})] \approx \sum_{l=1}^{L} \sum_{\boldsymbol{\xi}^n \in \mathbb{X}_l} w_{nl} p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)} = \boldsymbol{\xi}^n) p(\boldsymbol{x}^{(t)} = \boldsymbol{\xi}^n).$$

Setting $\mathbb{X} = \mathbb{X}_1 \sqcup \cdots \sqcup \mathbb{X}_L$, we recover Equation (15). Thus, composed quadrature rules can be handled just as a single Gaussian product rule. We exploit this to construct quadrature rules which are efficient for the computation of next state distributions for our particular GP

Algorithm 3 Construction of composed quadrature rules

Input: dynamics GP $\boldsymbol{f}\colon \left(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\right) \mapsto \boldsymbol{x}^{(t+1)}$, control policy $\boldsymbol{\pi}\colon \boldsymbol{x} \mapsto \boldsymbol{\pi}_{\boldsymbol{\theta}}(\boldsymbol{x})$ with parameters $\boldsymbol{\theta}$, state space X, maximum partition size $L_{\max}$
Output: composed quadrature rule with nodes $\mathbb{X}$ and weight vector $\boldsymbol{w}$

```
Initialize partition X = X_1 ⊔ ··· ⊔ X_s and quadrature 𝕏 = 𝕏_1 ⊔ ··· ⊔ 𝕏_s with s < L_max
while s < L_max do
    Sample starting state x^(0), policy parameters θ*
    Compute mean rollout τ_0 = x^(0), ..., τ_H with dynamics GP f and policy π_θ*
   for l = 1, ..., s do
      if vol(X_l) / (|𝕏_l| min_{τ_i ∈ X_l} Var(τ_i)) < 1 then subdivide X_l, add nodes to 𝕏 fi
   od
od
return quadrature nodes 𝕏 and weights w
```

dynamics model and controller. In other words, we aim to find a partition $X = X_1 \sqcup \cdots \sqcup X_L$ of the state space, such that integral (14) is approximated well by the resulting quadrature rule. For this purpose, we maintain a partition of the state space, sample mean trajectories $\boldsymbol{\tau} = \boldsymbol{\tau}_0, \ldots, \boldsymbol{\tau}_H$ and subdivide regions X_l that were visited and fulfill a certain criterion. More precisely, we subdivide X_l if X_l does not contain enough quadrature nodes to integrate predictive distributions from the dynamics GP well. To estimate whether X_l should be divided, we introduce

$$\rho_l(\boldsymbol{\tau}) \coloneqq \frac{\operatorname{vol}(X_l)}{|\mathbb{X}_l| \min_{\boldsymbol{\tau}_i \in X_l} \operatorname{Var}(\boldsymbol{\tau}_i)}, \tag{19}$$

where $\operatorname{Var}(\boldsymbol{\tau}_i)$ denotes the variance of the GP predictive distribution at the point $\boldsymbol{\tau}_i$, and subdivide X_l if $\rho_l(\boldsymbol{\tau})$ is greater than 1. This criterion relates the higher order derivatives of GP predictions which fall inside X_l with the quadrature node density in X_l. Constructing a composed quadrature rule with this approach will concentrate most quadrature nodes in state space regions that are visited frequently when following system trajectories. Algorithm 3 summarizes our approach and Figure 4 illustrates the constructed quadrature rules.

4.5 GP Dynamics with Infinite Time Horizons

In Section 4.2, we proposed numerical quadrature to approximate next state distributions for closed-loop control systems, where the dynamics is given as a full Gaussian process. For this approximate inference, pointwise error bounds are available. Based on this error analysis, we were able to derive stability guarantees for finite time horizons. In this section, we analyze the case when the time horizon is infinite. Unfortunately, error bounds as in Section 4.3 cannot be given in this case. As we will see, the numerical quadrature approximation converges to a stationary limit distribution, i.e., this distribution does not depend on the starting state, as t goes to infinity. As the quadrature error cannot be bounded in this case, there is no guarantee that the system dynamics will converge to the same distribution. However, we will show that closed-loop control systems with GP dynamics also converge to a stationary distribution as $t \to \infty$ for many choices of the prior. We will first analyze

the behavior of the proposed numerical quadrature approximation and subsequently study asymptotic behavior of closed-loop control systems with GP dynamics.

4.5.1 Asymptotic Behavior of NQ Approximated GP Dynamics

In this section we study the properties of numerical quadrature as approximate inference when the time horizon is infinite. Recall that with numerical quadrature the state distribution is approximated by

$$p(\boldsymbol{x}^{(t)}) \approx \boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(t)} \tag{16 revisited}$$

with Gaussian basis functions $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)^\intercal$, that do not change over time, and the weight vector $\boldsymbol{\alpha}^{(t)}$ at time t. As stated in Equation (17), the weight vector $\boldsymbol{\alpha}^{(t)}$ is updated by

$$\boldsymbol{\alpha}^{(t+1)} = \operatorname{diag}(\boldsymbol{w})\Phi\boldsymbol{\alpha}^{(t)} \tag{17 revisited}$$

with the positive matrix $\operatorname{diag}(\boldsymbol{w})\Phi$ as in Section 4.2.1. More precisely, $\operatorname{diag}(\boldsymbol{w})\Phi$ is a left stochastic matrix. Thus, any column vector $\boldsymbol{\beta}$ with norm $\|\boldsymbol{\beta}\|_1 = 1$ is mapped to a unit vector, i.e. $\|\operatorname{diag}(\boldsymbol{w})\Phi\boldsymbol{\beta}\|_1 = 1$. In this case, Equation (17) represents a power iteration which leads to the following theorem.

Theorem 14 *Let $(\mathbb{X}, \boldsymbol{w})$ be a quadrature rule with positive weights $\boldsymbol{w}$ on the state space X and $\boldsymbol{f}\colon (\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}) \mapsto \boldsymbol{x}^{(t+1)}$ be a dynamics GP. The approximate state distribution $p(\boldsymbol{x}^{(t)}) \approx \boldsymbol{\phi}^\intercal \boldsymbol{\alpha}^{(t)}$ obtained by cascading approximate one step ahead predictions with the quadrature $(\mathbb{X}, \boldsymbol{w})$ converges to a unique stationary distribution, independent of the starting distribution, as $t \to \infty$.*

Proof Consider Equation (17). The matrix $\operatorname{diag}(\boldsymbol{w})\Phi$ has norm one, as each of its columns sums to one. Thus, as the starting weights $\boldsymbol{\alpha}^{(0)}$ form a stochastic vector, Equation (17) takes the form of a power iteration, see (Golub and Van Loan, 2013, Section 7.3). By the Perron-Frobenius theorem the largest magnitude eigenvalue λ_1 of $\operatorname{diag}(\boldsymbol{w})\Phi$ is positive and simple. More precisely, the eigenvalue of $\operatorname{diag}(\boldsymbol{w})\Phi$ with the highest absolute value is $\lambda_1 = 1$. The corresponding dominant eigenvector $\boldsymbol{v}_1$ is positive. No other eigenvectors of $\operatorname{diag}(\boldsymbol{w})\Phi$ are non-negative. Thus, the power iteration in Equation (17) converges to the dominant eigenvector, if $\boldsymbol{\alpha}^{(0)}$ has a nonzero projection on the eigenspace of $\lambda_1 = 1$. As $\boldsymbol{\alpha}^{(0)}$ is a non-negative unit vector and $\boldsymbol{v}_1$ is positive, the projection to the dominant eigenspace is ${\boldsymbol{v}_1}^\intercal \boldsymbol{\alpha}^{(0)} > 0$. Thus, the numerical quadrature approximation of $p(\boldsymbol{x}^{(t)})$ converges to $\boldsymbol{\phi}^\intercal \boldsymbol{v}_1$. ■

This theorem shows that the numerical quadrature approximation to the system state converges to a stationary distribution. The limiting distribution corresponds to the dominant eigenvector of the iteration matrix $\operatorname{diag}(\boldsymbol{w})\Phi$ and does not depend on the starting state. However, no bound for the approximation error can be obtained for the limit $t \to \infty$. Thus, in contrast to finite time horizon analysis, no statement about the true system state follows from this result. In the following, we will study asymptotic behavior of closed-loop control systems with GP dynamics to see whether the (exact) system state converges to a limiting distribution as well.

4.5.2 Asymptotic Behavior of GP Dynamics

We have seen that employing the numerical quadrature approximation for next state distributions, closed-loop control systems with GP dynamics converge to a limiting distribution.

In this section, we will show that these closed-loop control systems expose the same behavior even if no approximations are employed. More precisely, a closed-loop control system with dynamics given as a GP will finally converge to a limiting distribution that is independent of the starting state for many common choices of the prior. To obtain this result, we apply Markov Chain theory as in (Meyn and Tweedie, 2009).

Consider a Markov chain $\mathbf{\Sigma} = \{\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \ldots\}$ with the random variables $\mathbf{\Sigma}_t$ taking values in $\mathbb{R}^D$ and a *Markov kernel* $P\colon \mathbb{R}^D \times \mathcal{B}(\mathbb{R}^D) \to \mathbb{R}$, i.e., $P(\cdot, A)$ is a non-negative measurable function on $\mathbb{R}^D$, $P(\boldsymbol{x}, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R}^D)$ and

$$\mathbb{P}_\mu[\mathbf{\Sigma}_0 \in A_0, \cdots, \mathbf{\Sigma}_n \in A_n] = \int_{\boldsymbol{y}_0 \in A_0} \cdots \int_{\boldsymbol{y}_{n-1} \in A_{n-1}} \mu(d\boldsymbol{y}_0) P(\boldsymbol{y}_0, d\boldsymbol{y}_1) \cdots P(\boldsymbol{y}_{n-1}, A_n) \tag{20}$$

for any initial distribution μ and $n \in \mathbb{N}$. We will employ some definitions and results from (Meyn and Tweedie, 2009) to prove our claim.

Definition 15 *The following definitions will help describing Markov chains and the sets they (re-)enter.*

1. *For $A \in \mathcal{B}(\mathbb{R}^D)$ and $\Sigma_0 \in A$, the* first return time *τ_A is given as*

$$\tau_A \coloneqq \min\{n \geq 1 \colon \mathbf{\Sigma}_n \in A\}.$$

   We also define the return time probability *$L(\boldsymbol{x}, A) \coloneqq \mathbb{P}_{\boldsymbol{x}}(\tau_A < \infty)$.*

2. *We call $\mathbf{\Sigma}$ φ-*irreducible*, if there exists a measure φ on $\mathcal{B}(\mathbb{R}^D)$, such that $\varphi(A) > 0$ implies $L(\boldsymbol{x}, A) > 0$ for all $\boldsymbol{x} \in X$ and $A \in \mathcal{B}(\mathbb{R}^D)$. ψ is a* maximal irreducibility measure*, if $\mathbf{\Sigma}$ is ψ-irreducible and $\varphi \prec \psi$, i.e., $\psi(A) = 0 \Rightarrow \varphi(A) = 0$, for all irreducibility measures φ of $\mathbf{\Sigma}$.*

3. *We say that the set $C \in \mathcal{B}(\mathbb{R}^D)$ is* small*, if there exists a probability measure ν, such that $P^n(\boldsymbol{x}, A) \geq \delta\nu(A)$ with some integer n, $\delta > 0$ for all $\boldsymbol{x} \in C$ and all $A \in \mathcal{B}(\mathbb{R}^D)$.*

For finite or countable state spaces, a Markov kernel defines the probability to move from one state to another. The state space partitions into *communication classes*, such that all states in one class can be reached with positive probability from any other state in the same class. If there is only one communication class, the Markov chain is called irreducible – its analysis cannot be further reduced to the consideration of multiple chains with smaller state spaces. The communication structure has an immense effect on the long term behavior of a Markov chain. However, when considering general state spaces (e.g., continuous as in our case), the probability to reach one single state will be zero. The concept of communication is thus generalized to Borel sets via irreducibility measures. Every irreducible Markov chain has maximum irreducibility measures, see Theorem 16. While these are not unique, all maximum irreducibility measures of a Markov chain are equivalent, i.e., they have the same null sets. Small sets play an important role in the analysis of Markov chains on general state spaces, as they behave analogously to single states in discrete state spaces. Sets which consist of one single point, are small. However, it can be shown that for irreducible Markov chains, any set that will be entered with positive probability contains a non-trivial small set (Meyn and Tweedie, 2009, Theorem 5.2.2).

Theorem 16 *Let $\mathbf{\Sigma}$ be φ-irreducible. Then there exists a maximal irreducibility measure ψ, which is equivalent to*

$$\psi'(A) \coloneqq \int_X \varphi(d\boldsymbol{y}) K_{a\frac{1}{2}}(\boldsymbol{y}, A)$$

with the transition kernel $K_{a\frac{1}{2}}(\boldsymbol{y}, A) = \sum_{n=1}^{\infty} P^n(\boldsymbol{x}, A) 2^{-(n+1)}$.

Proof The proof is given in (Meyn and Tweedie, 2009, Proposition 4.2.2). ■

We aim to analyze the asymptotic behavior of Markov chains that are defined by our closed-loop control structure with GP dynamics. One type of asymptotic behavior that is well-known from Markov chains on discrete state spaces is periodicity. The notion of periodicity can be extended to Markov chains on general state spaces as follows.

Definition 17 *Let $\mathbf{\Sigma}$ be irreducible with maximal irreducibility measure ψ. The* period *of $\mathbf{\Sigma}$ is the largest integer $d > 0$, such that there exist non-empty, pairwise disjoint sets $X_1, \cdots, X_d \subseteq \mathbb{R}^D$ with*

$$P(\boldsymbol{x}, X_{i+1}) = 1 \text{ for all } \boldsymbol{x} \in X_i, 1 \leq i \leq d-1 \text{ and } P(\boldsymbol{x}, X_1) = 1 \text{ for } \boldsymbol{x} \in X_d$$

and

$$\psi\left(\bigcup_{i=1}^{d} X_i\right) = 1.$$

If $d = 1$ we call $\mathbf{\Sigma}$ aperiodic*, if $d > 1$ the Markov chain $\mathbf{\Sigma}$ is* periodic*.*

Next, we apply the definitions above to establish some properties of closed-loop control systems with GP dynamics. Fortunately, the topology of $\mathbb{R}^D$ and the Gaussian state predictions obtained from a GP largely facilitate the analysis.

Lemma 18 *The Markov chain $\mathbf{\Sigma}$ on $\mathbb{R}^D$ defined by a GP forward dynamics is irreducible with respect to the Lebesgue measure λ and aperiodic. Also, all compact sets are small.*

Proof The transition probabilities are Gaussian, i.e.,

$$P(\boldsymbol{x}, \cdot) = \mathcal{N}(\boldsymbol{m}(\boldsymbol{x}), S(\boldsymbol{x}))$$

and the Markov kernel has a Gaussian density $q\colon \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$. Thus, $q(\boldsymbol{x}, \boldsymbol{y}) > 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^D$ and for any $A \subseteq \mathbb{R}^D$ with $\lambda(A) > 0$ it holds

$$P(\boldsymbol{x}, A) = \int_A q(\boldsymbol{x}, \boldsymbol{y}) \lambda(d\boldsymbol{y}) > 0,$$

so $\mathbf{\Sigma}$ is irreducible with respect to the Lebesgue measure λ.

Assume $\mathbf{\Sigma}$ has period $d > 1$, i.e., $P(\boldsymbol{x}, X_i) = 1$ for $\boldsymbol{x} \in X_{i-1}$, then $1 = P(\boldsymbol{x}, X_i) = \int_{X_i} q(\boldsymbol{x}, \boldsymbol{y}) \lambda(d\boldsymbol{y})$ and, as $q > 0$, it follows that the complement $\overline{X}_i$ of X_i is a λ-null set, i.e., $\lambda(\overline{X}_i) = 0$. However, as X_i, X_j are disjoint for $j \neq i$ it holds $X_j \subseteq \overline{X}_i$. It follows $\lambda(X_j) = 0$ for $j \neq i$. Analogously as for X_i, $\lambda(\overline{X}_j) = 0$. As λ is a non-trivial measure, this is a contradiction. Thus, $d = 1$ and $\mathbf{\Sigma}$ is aperiodic.

Finally, as P has a continuous, positive density, there is a lower bound $\delta > 0$ for this density on any compact set. Thus, all compact sets are small. ■

Above, we showed that the Markov chains considered are aperiodic. The Markov chain can still re-enter sets infinitely often with positive probability. A stronger form of this property is Harris recurrence. As we will see later, Harris recurrence along with some other properties implies the existence and uniqueness of an invariant state distribution.

Definition 19 *Let $A \in \mathcal{B}(\mathbb{R}^D)$ and consider the event*

$$\{\boldsymbol{\Sigma} \in A \text{ infinitely often}\} \coloneqq \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{\boldsymbol{\Sigma}_k \in A\}.$$

The set A is called Harris recurrent*, if*

$$Q(\boldsymbol{x}, A) \coloneqq \mathbb{P}_{\boldsymbol{x}}\{\boldsymbol{\Sigma} \in A \text{ infinitely often}\} = 1.$$

The Markov chain $\boldsymbol{\Sigma}$ is Harris recurrent, if it is ψ-irreducible and $A \in \mathcal{B}(X)$ is Harris recurrent if $\psi(A) > 0$.

We say that the measure μ is an invariant measure *for the Markov kernel P, if*

$$\mu(A) = \int_X \mu(d\boldsymbol{x}) P(\boldsymbol{x}, A),$$

or shortly $\mu P = \mu$.

The Markov chain $\boldsymbol{\Sigma}$ is positive Harris recurrent*, if it is Harris recurrent and admits an invariant probability measure.*

Finally, we can state some conditions for the existence and uniqueness of an invariant state distribution as well as convergence of the system state to this distribution (in total variation).

Theorem 20 *If the chain $\boldsymbol{\Sigma}$ is φ-irreducible, aperiodic and positive Harris recurrent, then it admits a unique invariant probability measure μ^* and for all probability measures ν it holds*

$$\|\nu P^n - \mu^*\|_{TV} \coloneqq \sup_{A \in \mathcal{B}(\mathbb{R}^D)} |\mu^*(A) - \nu(A)| \to 0$$

as $n \to \infty$.

Proof The proof can be found in (Meyn and Tweedie, 2009, Theorem 13.0.1). ■

To employ this theorem to our case of closed-loop control systems with GP dynamics, it remains to establish Harris recurrence. However, it can be challenging to verify the definition directly. Fortunately, drift criteria can be employed. In the following, we will write $Pf \coloneqq \int P(\boldsymbol{x}, d\boldsymbol{y}) f(\boldsymbol{y})$ for any function $f \colon \mathbb{R}^D \to \mathbb{R}$ for notational convenience.

Theorem 21 *Let P be a φ-irreducible and aperiodic Markov kernel. If there exists a function $V \colon \mathbb{R}^D \to [0, \infty)$ with*

$$PV \leq V + c\mathbb{1}_A \tag{21}$$

for a small set A, a constant $c < \infty$, and the sets $\{\boldsymbol{x} \mid V(\boldsymbol{x}) \leq r\}$ are small for all $r \in \mathbb{R}$, then P is Harris recurrent. P is positive Harris recurrent, if Eq. (21) can be modified to

$$PV \leq V - \varepsilon + c\mathbb{1}_A \tag{22}$$

for some $\varepsilon > 0$.

Proof The proof can be found in (Meyn and Tweedie, 2009, Theorems 9.1.8, 11.0.1). ■

Such a function V is called *drift function*. It is the stochastic analogue to a Lyapunov function. Inequality (21) states that whenever the system state is outside of the small set A, V is expected to decrease in the following step. Thus, the system tends towards the minimum of V, which is a recurrent behavior. The speed of convergence can be incorporated in a drift criterion as well.

Definition 22 *The Markov Kernel P has* geometric drift *towards a set A, if there exists a function $V\colon \mathbb{R}^D \to [1,\infty]$, finite for some $\boldsymbol{x}$, and $\beta < 1, c < \infty$, such that*

$$PV \leq \beta V + c\mathbb{1}_A. \tag{23}$$

We say that a φ-irreducible Markov Kernel P with invariant probability measure μ^* is V-geometrically ergodic *on a set A for a function $V\colon \mathbb{R}^D \to [1,\infty]$, if*

$$\sup_{|f|\leq V} \left| \int f d(P^n(\boldsymbol{x},\cdot) - \mu^*) \right| \leq cV(\boldsymbol{x})\rho^n \tag{24}$$

for all $\boldsymbol{x} \in A$, a constant $c < \infty$ and $\rho < 1$ and $V(\boldsymbol{x}) < \infty$ for all $\boldsymbol{x} \in A$.

We will employ geometric drift functions to prove the convergence of Markov chains generated by closed-loop control systems with GP forward dynamics at an exponential convergence speed, irrespective of the starting state. Studying the drift function $V(\boldsymbol{x}) = 1 + \|\boldsymbol{x}\|^2$, the following criterion for geometric ergodicity can be derived.

Theorem 23 *A Markov kernel P with Gaussian transition probabilites has V-geometric drift towards a compact set if and only if*

$$\limsup_{\|\boldsymbol{x}\|\to\infty} \frac{\|\boldsymbol{m}(\boldsymbol{x})\|^2 + \operatorname{tr}(S(\boldsymbol{x}))}{\|\boldsymbol{x}\|^2} < 1 \tag{25}$$

with the mean map $\boldsymbol{m}(\boldsymbol{x})$ and variance map $S(\boldsymbol{x})$. It follows that P has a unique invariant probability measure μ^* and is V-geometrically ergodic.

Proof The proof is given in (Hansen, 2003, Lemma 3.2, Corollary 3.3). ■

Finally, we can prove the main result of this section, the convergence of many closed-loop control systems with GP dynamics towards a unique, invariant limiting distribution that is independent of the starting state.

Theorem 24 *Any discrete-time closed-loop control system with state space $\mathbb{R}^D$, where the forward dynamics is given as GP with zero-mean prior and stationary covariance function, is geometrically ergodic, i.e., the system has a unique invariant probability measure and the Markov chains generated by the system converge to its stationary distribution with exponential speed, see Equation (24).*

Proof We have to verify Eq. (25). As the GP predictions fall back to the zero mean prior far away from the training data, $\|\boldsymbol{m}(\boldsymbol{x})\|^2 + \operatorname{tr}(S(\boldsymbol{x}))$ converges to $\sum_{k=1}^{D} \sigma_{k,f}^2$ as $\|\boldsymbol{x}\| \to \infty$. Thus,

$$\limsup_{\|\boldsymbol{x}\|\to\infty} \frac{\|\boldsymbol{m}(\boldsymbol{x})\|^2 + \operatorname{tr}(S(\boldsymbol{x}))}{\|\boldsymbol{x}\|^2} = 0 < 1,$$

which concludes the proof. ■

This theorem gives us a strong result about the asymptotic behavior of closed-loop control systems with dynamics given as a GP with stationary covariance function and zero mean prior. This result applies to many learning control approaches in the literature.

5. Empirical Evaluation

In this section, we evaluate the previously obtained theoretical results on two benchmark tasks: mountain car and inverted pendulum. Moreover, the performance of the proposed uncertainty propagation is compared to the state-of-the-art moment matching approach and Monte Carlo sampling. We begin with a brief introduction of the test-beds.

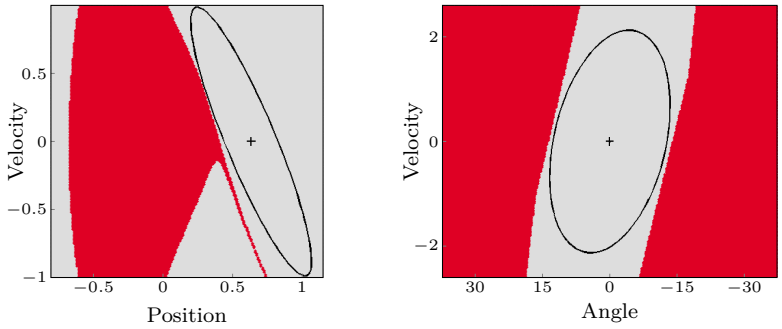
Mountain Car. A car with limited engine power has to reach a desired point in the mountainscape (Sutton and Barto, 1998). The state space has two dimensions: position and velocity of the car. We analyze stability of a PD-controller $\pi((x, \dot{x})^\intercal) = K_p x + K_d \dot{x}$. The gains are chosen as $K_p = 25$ and $K_d = 1$ and the control signal is limited to $u_{\max} = 4$. The GP dynamics model was trained on 250 data points from trajectories with random starting points and control gains.

Inverted Pendulum. In the inverted pendulum task, the goal is to bring the pendulum to an upright position with limited torque (see Doya, 2000) and balance it there. The system state has two dimensions: pendulum angle and velocity. We evaluate stability of a PD-controller with $K_p = 6$, $K_d = 3$ and control limit $u_{\max} = 1.2$. The dynamics GP was trained on 200 points from rollouts with random starting points and control gains.

Cart-Pole. In the cart-pole domain (Deisenroth et al., 2015), a cart with an attached free-swinging pendulum is running on a track of limited length. The goal is to swing the pendulum up and balance it, with the cart in the middle of the track. The state space has four dimensions: position of the cart x, velocity of the cart $\dot{x}$, angle of the pendulum ϑ and the angular velocity $\dot{\vartheta}$. A horizontal force with $u_{\max} = 10$ can be applied to the cart. To demonstrate the proposed approach we analyze stability of the cart-pole system for a PD-controller with randomly chosen gains. The dynamics GP was trained on 250 points from rollouts with sampled starting state and control gains.

5.1 Stability of GP Predictive Mean Dynamics

To evaluate the presented theory on stability of the closed-loop control system with GP mean dynamics, a stability region is determined as described in Section 3. We compare this region to the true stability region, empirically obtained as follows. A grid on the state space is defined and rollouts from every grid point are computed. After a sufficiently long time (1000s), we check whether the state has converged to the target point. Thus, we empirically determine a region of starting points, where the system converges to the desired state. Figure 5 shows the obtained regions for the mountain car and pendulum systems. In both cases the theoretically obtained stability region, which is marked by an ellipsoid, is a subset of the empirically determined region. This effect is due to our analysis yielding a full metric ball centered around the target point, although the full stability region is not necessarily convex. Also, trajectories which first move away from the target point, but finally



(a) Mountain Car (b) Inverted Pendulum

Figure 5: Stability regions for GP mean dynamics on the two benchmark tasks. This figure shows empirically obtained stability regions in gray, points that did not converge in red. The ellipsoid indicates the stability region returned by Algorithm 1.

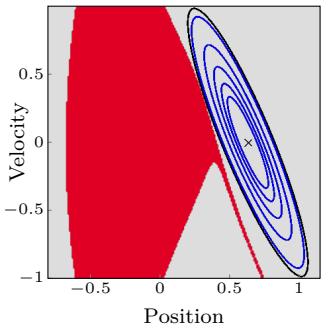


Figure 6: Finite time stability of GP mean dynamics. The black ellipsoid marks the (infinite time) stability region obtained with Algorithm 1. The blue ellipsoids mark the (ε, T)-stability regions with $\varepsilon = 0.05$ and $T = 10, 20, 30, 40, 50$ timesteps (inner to outer ellipsoid). For any starting point inside of one blue ellipsoid, the distance to the target is less than ε after T timesteps.

converge to it, are not considered in the presented theory. Instead, all trajectories starting in the theoretically obtained stability region move towards the target point contractively.

Next, we compute finite time (ε, T)-stability regions as in Lemma 6. Figure 6 shows the obtained regions for $\varepsilon = 0.05$ and different choices of T. By construction, these are full metric balls around the reference point and lie completely inside of the stability region obtained with Algorithm 1.

Finally, we consider robustness of the GP mean dynamics. We introduce disturbances and apply the convergence criteria from Theorem 7 and Lemma 8. We sample the starting time and duration as well as the magnitude of the disturbance. The direction of the disturbance is then sampled independently for every time step where a disturbance is present. The results are shown in Figure 7. Note that robustness results such as Theorem 7 and Lemma 8 must

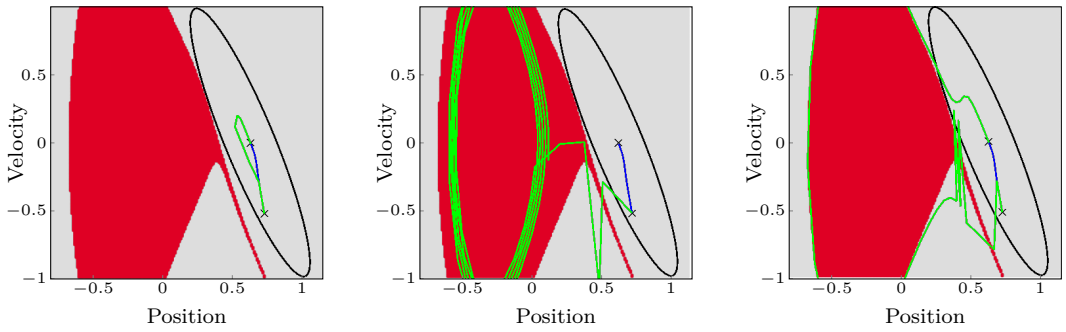
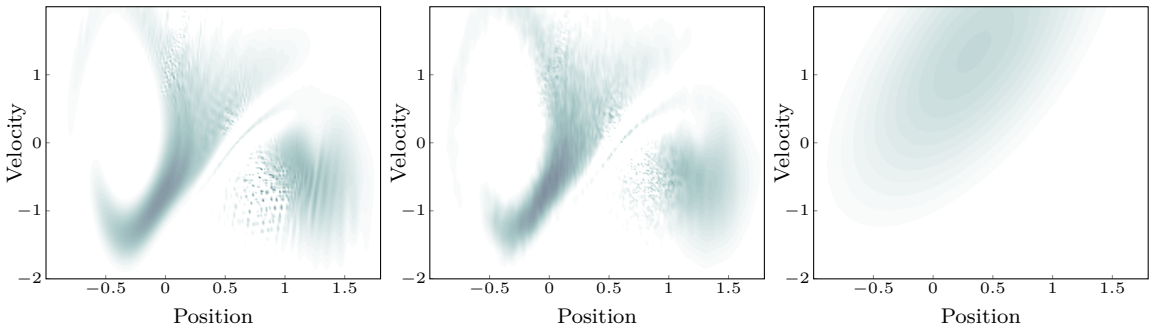


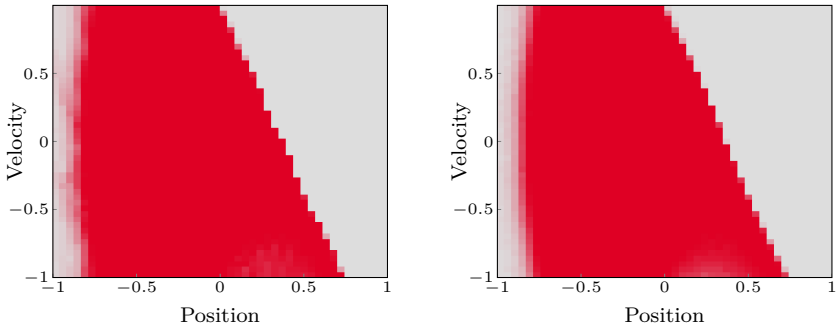
Figure 7: Robustness of GP mean dynamics. Disturbances were sampled for a finite time period. The undisturbed trajectory is colored blue, the distrubed trajectories are colored green. Our robustness criterion from Theorem 7 and Lemma 8 guarantees the convergence of the disturbed trajectory in the left picture. For both other pictures, our criterion cannot guarantee convergence of the disturbed trajectory to the target. As such criteria must be based on worst case estimates of the disturbance, some disturbed trajectories still converge to the target (as in the right picture).



(a) Our Numerical Quadrature (b) Monte Carlo Sampling (c) Moment Matching

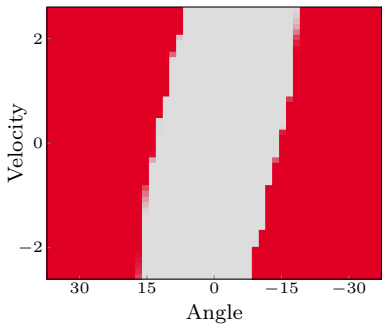
Figure 8: This figure illustrates multi-step-ahead prediction when the input is a distribution. Starting with a normally distributed state centered around the inflection point of the right slope in the mountain car domain, we compute a rollout incorporating uncertainty. The plots show the state distribution after T=30 time steps obtained with our numerical quadrature approach (a), moment matching (c) and the reference Monte Carlo sampling result (b).

consider the worst case scenario. Thus, there are trajectories where the robustness criteria are not able to guarantee convergence, although they converge to the target state.

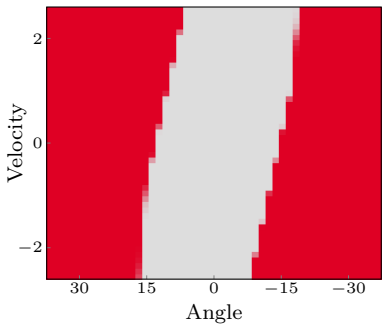


(a) Mountain Car, Our NQ

(b) Mountain Car, MC Sampling



(c) Pendulum, Our NQ



(d) Pendulum, MC Sampling

Figure 9: Success probabilities for dynamics given as GP full distribution and a time horizon of 100s. The color scale encodes the success probability from zero (red) to one (gray). The mountain car ((a) and (b)) and inverted pendulum ((c) and (d)) results were obtained with the proposed numerical quadrature (NQ) and Monte Carlo (MC).

5.2 Numerical Quadrature Uncertainty Propagation

The key to stability analysis for closed-loop systems with dynamics given as full GP distribution is the prediction at uncertain inputs. We compare the performance of the presented approximate inference to the state-of-the-art moment matching (MM) method and Monte Carlo (MC). Consider the following scenario: in the mountain car domain, we position the car on the right slope with some positive velocity. Furthermore, we introduce small Gaussian uncertainty about the starting state. We employ a constant control signal, that is too small to bring the car up directly. We compute rollouts, propagating state distributions through the GP with (i) the presented numerical quadrature (NQ), (ii) MM as in (Deisenroth, 2010), and (iii) MC. The resulting distributions for a time horizon of 3s are shown in Figure 8. The MM approximation differs significantly from MC and NQ results, concentrating most of the probability mass, where the MC approximation has very low probability density. The NQ result closely matches the distribution obtained with MC.

5.3 Stability of Gaussian Process Dynamics

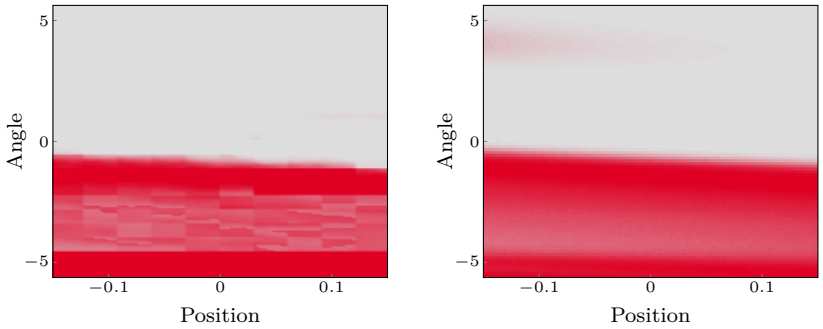
Employing numerical quadrature, we determine success probabilities for the two test-beds and a time horizon of 100s. To obtain the stability region for a particular choice of λ, all pixels whose color corresponds to a value higher then or equal to λ must be selected. As Figure 9 shows, the obtained stability regions match the empirical MC results. The error bound from Lemma 11 can demand for extremely fine quadrature rules to obtain a stability region estimate that matches the true stability region closely. Please note that quadrature rules with less nodes also allow for a guaranteed stability region, however this region can be significantly smaller than the true stability region. For long time horizons, the requirements to obtain a very accurate inner approximation of the stability region can become computationally infeasible. However, we found that the real-world results are substantially better than this worst-case bound. We also experienced computation time ($\approx$ 120s) for NQ to be a fraction of the time required for long time MC predictions. Of course, this will not hold for systems with many state dimensions and our particular setup of product quadrature rules, as these rules suffer from the curse of dimensionality. However, there are various approaches to overcome this drawback of NQ (Heiss and Winschel, 2008; Novak and Ritter, 1996; Xiao and Gimbutas, 2010; Ryu and Boyd, 2015) and our analysis holds for arbitrary quadrature rules.

To show how the proposed approach can be scaled to higher dimensions, we compute success probabilities for the four-dimensional cart-pole system. We employ Algorithm 3 to construct a suitable quadrature rule. However, the quadrature rule CN:3-1 (Stroud, 1971; code from Burkardt, 2014) is applied to each subregion instead of the Gaussian product rule as for the previous examples. This quadrature rule has 8 nodes as opposed to 16 nodes for the Gaussian product rule of the same exactness. Please note also that all necessary computations for the proposed approach can be executed in parallel. Thus, we conduct these computations on a GPU, which leads to a significant speedup and overall computation time comparable to the 2D examples ($\approx$ 140s). The result is shown in Figure 10. As for the other benchmark problems, the obtained stability region closely matches the true stability region of the system.

To evaluate the results on convergence of GP dynamics to a stationary distribution, we considered the dynamics shown in the left plot of Figure 11. The GP mean dynamics has two attractors and their regions of attraction partition the interval $[-1; 1]$. However, when considering the uncertainty, the system has an invariant probability measure (shown in Figure 11 on the right) to which it converges irrespective of the starting point.

6. Conclusion

Gaussian Processes provide a flexible, non-parametric, Bayesian approach to modeling unknown system dynamics from observations. Thus, GP dynamics are a viable approach for learning model-based control. There has been remarkable research in this field that advanced these methods to become an appealing alternative to classic control. For widespread real-world application of learning control based on GP forward dynamics models, performance guarantees are crucial, especially in safety-critical applications. However, stability analysis of closed-loop control with GP forward dynamics learned from data has barely been analyzed so far. In this paper, we laid a foundation for stability analysis of closed-loop control with



(a) Cart-Pole, Our NQ (b) Cart-Pole, MC Sampling

Figure 10: Success probabilities for the cart-pole system. The dynamics are given as GP full distribution and the time horizon is 100s. The color scale encodes the success probability from zero (red) to one (gray). The plots show the results for $\dot{x} = 0, \dot{\theta} = 0$ obtained with numerical quadrature (a) and Monte Carlo sampling (b).

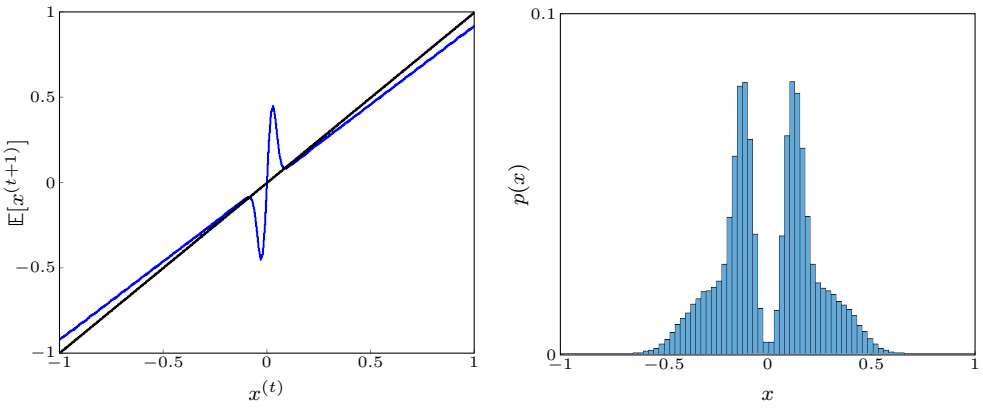
GP dynamics. Subsequently, we conclude the paper with a short summary of the main contributions and a brief outlook on possible future work in this direction.

6.1 Summary of Contributions

In this paper, we analyzed stability of closed-loop control systems with Gaussian process forward model dynamics. We considered two possible types of system dynamics: (i) the mean and (ii) the full GP predictive distribution.

In the first case, we studied asymptotic stability as well as finite time horizons and robustness to disturbances. We presented an algorithm to construct a region in the state space such that trajectories starting inside this region are guaranteed to converge to the target point and stay there for all times. For finite time horizons, we showed how to find a state space region such that the target state will be reached at time horizon up to a certain tolerance. Studying robustness, we derived a criterion for disturbances such that the system remains asymptotically stable. The theoretical results have been evaluated on benchmark problems and compared to empirically obtained results.

In the second case, we introduced a novel approach based on numerical quadrature to approximately propagate uncertainties through a GP. In contrast to other state-of-the-art methods, our approach can model complex distributions with multiple modes. Evaluation results closely match the true distribution approximated by extensive sampling. We used the introduced approximate inference method to derive finite-time stability guarantees based on quadrature error analysis. Empirical Monte Carlo results confirm our theoretical results on the two benchmark problems. Furthermore, we considered the system behavior when the time horizon is infinite. We showed that, applying numerical quadrature to propagate distributions through the GP, the system state converges to a limiting distribution that does not depend on the starting state. Motivated by this result, we studied the system behavior



(a) Forward Dynamics Function

(b) Invariant Limiting Distribution

Figure 11: Asymptotic behavior of closed-loop systems with GP dynamics. Plot (a) shows the mean dynamics function with the mean $m(x)$ of the GP prediction at the query point x on the y axis. When considering the mean dynamics, there are two fixpoints $x_1 = -0.1$ and $x_2 = 0.1$. For any starting point $-1 < x < 0$, the mean trajectory converges to x_1 and analogously for all $0 < x < 1$ the trajectories converge to x_2. Now we consider the full GP dynamics and take the uncertainty into account. Plot (b) shows a histogram of the probability density after 10000 time steps when starting at $x = -0.1$, obtained with Monte Carlo. This result matches the histogram obtained when starting at $x = 0.1$ or at any other point. Thus, the system converges to a unique, invariant limiting distribution.

for infinite time horizons without applying any approximate inference steps. We succeeded to show that closed-loop control systems with dynamics given as full GP distribution converge to a unique and invariant limiting distribution that does not depend on the starting state for many choices of the covariance function. Overall, the proposed methods provide stability guarantees for many existing learning control approaches based on GPs.

6.2 Discussion and Next Steps

While the proposed methods apply to many interesting examples of learning control in the literature, the analysis could be extended in several directions. Firstly, for closed-loop control with GP dynamics, we considered GPs with stationary covariance functions and zero mean prior. To include other choices of the prior, e.g., nonstationary covariance functions and unbounded mean priors, novel criteria to ensure Harris recurrence must be found. Another interesting question is how the stability of GP dynamics is affected by disturbances in the input space. While some results could straightforwardly be followed for dynamics given as the mean of a GP, the task seems a lot more challenging when uncertainty is included. Many of the subtle, intricate arguments applied to show convergence of the system to a limiting distribution must be carefully reconsidered when disturbances are present.

Secondly, the obtained results are stability results for closed-loop control systems with dynamics given as a GP. The GP training data have a huge impact on the quality of the GP dynamics model when compared to the physical system the data was collected from. In this direction, criteria to evaluate the quality of a learned model would further support the application of learning control for real-world problems.

Another interesting research direction is encouraged by the presented result on asymptotic stability of closed-loop control systems with GP dynamics. This paper proves the existence of a limiting distribution the system will converge to. This distribution is unique and invariant. However, little more is known about the limiting distribution. Further research in this direction could help for a better understanding of the behavior of learned dynamics and also be directly employed in learning control – possibly by optimizing the limiting distribution itself instead of local approximations as, e.g., system trajectories.

References

G. Adomian. *Stochastic systems*. Mathematics in Science and Engineering. Elsevier Science, 1983.

A. A. Ahmadi and P. A. Parrilo. Converse results on existence of sum of squares lyapunov functions. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 6516–6521, Dec 2011.

A. A. Ahmadi, A. Majumdar, and R. Tedrake. Complexity of ten decision problems in continuous time dynamical systems. In *2013 American Control Conference*, pages 6376–6381, 2013.

T. Beckers and S. Hirche. Stability of gaussian process state space models. In *Proceedings of the European Control Conference (ECC)*, 2016.

F. Blanchini. Set invariance in control. *Automatica*, 35(11):1747 – 1767, 1999.

J. Burkardt. Stroud – numerical integration in m dimensions. `https://people.sc.fsu.edu/~jburkardt/m_src/stroud/stroud.html`, 2014.

G. Chesi. Estimating the domain of attraction for uncertain polynomial systems. *Automatica*, 40(11):1981–1986, 2004.

P.J. Davis, P. Rabinowitz, and W. Rheinbolt. *Methods of Numerical Integration*. Computer Science and Applied Mathematics. Elsevier Science, 2014.

M.P. Deisenroth. *Efficient Reinforcement Learning Using Gaussian Processes*. Karlsruhe series on intelligent sensor actuator systems. KIT Scientific Publ., 2010.

M.P. Deisenroth, D. Fox, and C.E. Rasmussen. Gaussian processes for data-efficient learning in robotics and control. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(2):408–423, 2015.

K. Doya. Reinforcement learning in continuous time and space. *Neural Computation*, 12: 219–245, 2000.

Y. Engel, P. Szabo, and D. Volkinshtein. Learning to control an octopus arm with gaussian process temporal difference methods. In Y. Weiss, B. Schölkopf, and J.C. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 347–354. MIT Press, 2006.

G.A. Evans. The estimation of errors in numerical quadrature. *International Journal of Mathematical Education in Science and Technology*, 25(5):727–744, 1994.

G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.

N.R. Hansen. Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli*, 9(4):725–743, 08 2003.

F. Heiss and V. Winschel. Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1):62 – 80, 2008.

A. Hurwitz. Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt. *Mathematische Annalen*, 46(2):273–284, 1895.

H.K. Khalil. *Nonlinear control*. Prentice Hall, 2014.

R. Khasminskii and G.N. Milstein. *Stochastic Stability of Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2011.

H.J. Kim and A.Y. Ng. Stable adaptive control with online learning. In L.K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 977–984. MIT Press, 2005.

E.D. Klenske, M.N. Zeilinger, B. Schölkopf, and P. Hennig. Nonparametric dynamics estimation for time periodic systems. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pages 486–493. IEEE, 2013.

J. Ko and D. Fox. GP-BayesFilters: Bayesian filtering using gaussian process prediction and observation models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008.*, 2008.

J. Kocijan, R. Murray-Smith, C.E. Rasmussen, and A. Girard. Gaussian process model based predictive control. In *American Control Conference, 2004. Proceedings of the 2004*, volume 3, pages 2214–2219. IEEE, 2004.

H.J. Kushner. Finite time stochastic stability and the analysis of tracking systems. *Automatic Control, IEEE Transactions on*, 11(2):219–227, 1966.

H.J. Kushner. *Stochastic Stability and Control*. Mathematics in science and engineering. Academic Press, 1967.

A.M. Lyapunov. *General Problem of the Stability Of Motion*. Doctoral dissertation, Univesity of Kharkov, 1892. Englisch Translation by A.T. Fuller, Taylor & Francis, London 1992.

J.M. Maciejowski and X. Yang. Fault tolerant control using gaussian processes and model predictive control. In *Control and Fault-Tolerant Systems (SysTol), 2013 Conference on*, pages 1–12. IEEE, 2013.

A. Majumdar, A. A. Ahmadi, and R. Tedrake. Control and verification of high-dimensional systems with dsos and sdsos programming. In *53rd IEEE Conference on Decision and Control*, pages 394–401, 2014.

M. Masjed-Jamei. New error bounds for gauss-legendre quadrature rules. *Filomat*, 28(6): 1281–1293, 2014.

S. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

Charles A. Micchelli, Yuesheng Xu, and Haizhang Zhang. Universal kernels. *Journal of Machine Learning Research*, 7:2651–2667, December 2006.

J. Moore and R. Tedrake. Adaptive control design for underactuated systems using sums-of-squares optimization. In *2014 American Control Conference*, pages 721–728, June 2014.

J. Nakanishi, J.A. Farrell, and S. Schaal. A locally weighted learning composite adaptive controller with structure adaptation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002.*, pages 882–889 vol.1, 2002.

K.S. Narendra and A.M. Annaswamy. *Stable Adaptive Systems*. Dover Books on Electrical Engineering. Dover Publications, 2012.

D. Nguyen-Tuong and J. Peters. Model learning in robotics: a survey. *Cognitive Processing*, (4), 2011.

E. Novak and K. Ritter. High dimensional integration of smooth functions over cubes. *Numerische Mathematik*, 75(1):79–97, 1996.

Y. Pan and E. Theodorou. Probabilistic differential dynamic programming. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1907–1915. Curran Associates, Inc., 2014.

A. Papachristodoulou and S. Prajna. *Analysis of Non-polynomial Systems Using the Sum of Squares Decomposition*, pages 23–43. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

P.A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, 2000.

T.J. Perkins and A.G. Barto. Lyapunov design for safe reinforcement learning. *Journal of Machine Learning Research*, 3:803–832, March 2003.

J. Quiñonero-Candela, A. Girard, J. Larsen, and C.E. Rasmussen. Propagation of uncertainty in bayesian kernel models - application to multiple-step ahead forecasting. In *International Conference on Acoustics, Speech and Signal Processing*, pages 701–704, vol. 2, 2003.

C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

E.J. Routh. *A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion*. Macmillan and Company, 1877.

E.K. Ryu and S.P. Boyd. Extensions of gauss quadrature via linear programming. *Foundations of Computational Mathematics*, 15(4):953–971, 2015.

S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control: Analysis and Design*. John Wiley & Sons, 2005.

B.S. Skrainka and K.L. Judd. High performance quadrature rules: How numerical integration affects a popular model of product differentiation. *Available at SSRN 1870703*, 2011.

J. Steinhardt and R. Tedrake. Finite-time regional verification of stochastic nonlinear systems. In H.F. Durrant-Whyte, N. Roy, and P. Abbeel, editors, *Robotics: Science and Systems VII*, pages 321–328. MIT Press, 2012.

A.H. Stroud. *Approximate calculation of multiple integrals*. Prentice-Hall series in automatic computation. Prentice-Hall, 1971.

E. Süli and D.F. Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, 2003.

R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

G. Tao. *Adaptive Control Design and Analysis (Adaptive and Learning Systems for Signal Processing, Communications and Control Series)*. John Wiley & Sons, Inc., New York, NY, USA, 2003.

U. Topcu, A. Packard, P. Seiler, and G. Balas. Help on sos [ask the experts]. *IEEE Control Systems*, 30(4):18–23, 2010a.

U. Topcu, A. K. Packard, P. Seiler, and G. J. Balas. Robust region-of-attraction estimation. *IEEE Transactions on Automatic Control*, 55(1):137–142, 2010b.

J. Vinogradska, B. Bischoff, D. Nguyen-Tuong, A. Romer, H. Schmidt, and J. Peters. Stability of controllers for gaussian process forward models. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 545–554, 2016.

S. Wasowicz. On error bounds for gauss-legendre and lobatto quadrature rules. *Journal of Inequalities in Pure & Applied Mathematics*, 7(3):Paper No. 84, 7 p., 2006.

H. Xiao and Z. Gimbutas. A numerical algorithm for the construction of efficient quadrature rules in two and higher dimensions. *Computers & Mathematics with Applications*, 59(2): 663 – 676, 2010.

K. Zhou and J.C. Doyle. *Essentials of Robust Control*. Prentice Hall Modular Series for Eng. Prentice Hall, 1998.